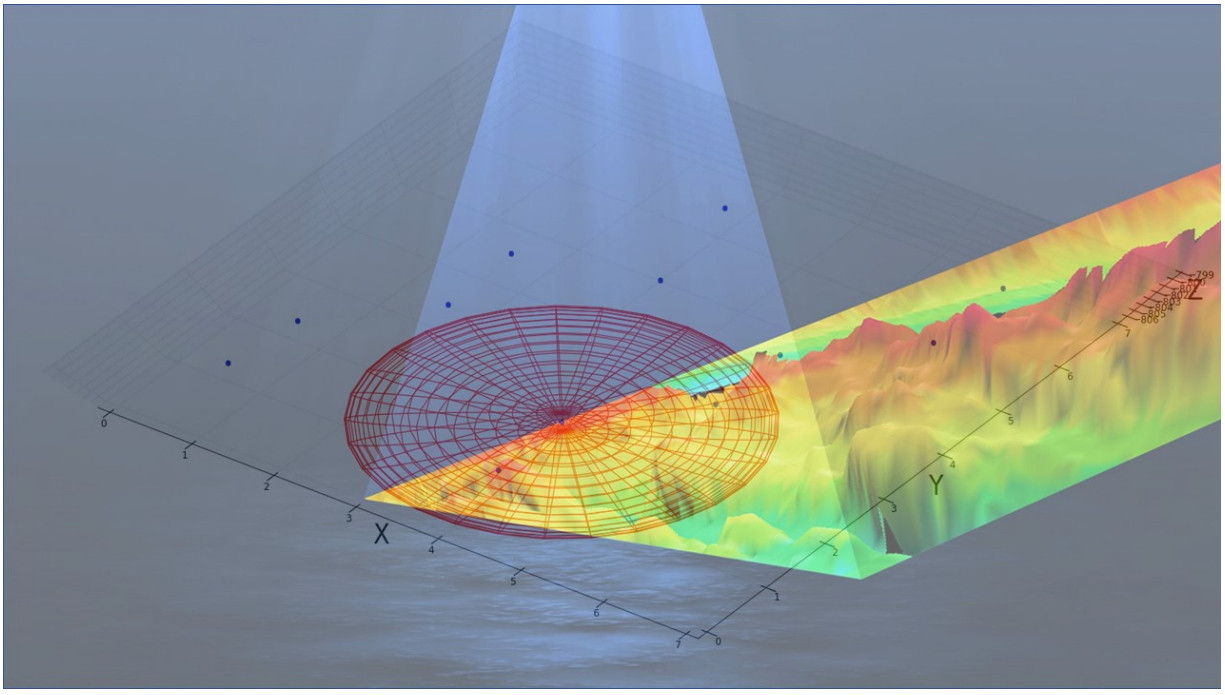
Analyse af metoder til efterbehandling af multibeam data



ANE AAGAARD RASMUSSEN
OPMÅLING OG KORTLÆGNING
4. SEMESTER
KANDIDAT I LANDINSPEKTØRVIDENSKAB
SKOLEN FOR ARKITEKTUR, DESIGN OG PLANLÆGNING
AALBORG UNIVERSITET
8. JUNI 2018

AALBORG UNIVERSITET
STUDENTERRAPPORT

Skolen for Arkitektur,
Design og Planlægning
Niels Jernes Vej 10-12
9220 Aalborg Øst
www.sadp.aau.dk

Titel:

Analyse af metoder til efterbehandling af multibeam data

Projekt:

Speciale

Projekt periode:

1. februar 2018 - 8. juni 2018

Grupemedlemmer:

Ane Aagaard Rasmussen

Vejleder:

Karsten Jensen

Sideantal: 101

Bilagsantal: 2

Synopsis:

I dette projekt er det undersøgt, hvordan multibeam data bearbejdes automatisk. Det er endvidere undersøgt, om der er én metode til behandlingen af data, der er bedre end andre. For at besvare dette er det undersøgt, hvilke softwarepakker EIVA, QPS og CARIS kan tilbyde på nuværende tidspunkt, og det er herefter undersøgt, hvilke algoritmer og metoder der ligger til baggrund for de forskellige software. Her har de interessante algoritmer og metoder været CUBE, S-CAN og EC-3D. For at give en forståelse for de tre metoder er de gennemgået teoretisk med en række illustrationer, som skal medvirke til en bedre forståelse.
Efter den teoretiske del foretages der en analyse af de tre metoder. Her anvendes der to forskellige datasæt til at teste metoderne - et fra Danmark og et fra Grønland. Det vælges at anvende netop disse datasæt, fordi der er forskel i havbundens udformning, og fordi det ønskes at undersøge, om en metode er bedre at anvende på én type data frem for en anden.
Konklusionen er, at der ikke er en metode, der alene kan håndtere udelukkende rå data. CUBE ses dog som den bedste løsning, idet den er hurtig, objektiv og håndterer data pænt. Der vil dog formentlig kræves en grov rensning af data inden anvendelse af CUBE. S-CAN og EC-3D klarer også rensningen af data nogenlunde, men en kombination af flere metoder vil nok ses som den bedste løsning, selvom dette ikke er testet i projektet.

Rapportens indhold er frit tilgængeligt, men offentliggørelse (med kildeangivelse) må kun ske efter aftale med forfatterne.

Abstract

This Master's thesis is investigation different processing tools to process multibeam data. The essence is that the method or algorithm should be automatic, so the manual interaction with the cleaning process is limited.
The problem statement of the project is:
How multibeam data can be processed automatically and which algorithm and methods is the best to use for different kinds of data?
To answer the question the thesis will first investigate some different products from different manufactures and figure out which processing tolls there are on the marked. Here it will be highlighted in some tables which software there are, and afterwards the thesis will investigate some specific hydrographic processing software from EIVA, QPS and CARIS. Following is has to be investigated which algorithm and methods that are behind the software. This is needed to be known, to figure out how it can be concluded which algorithm or method is better for one type of data than another.
After knowing the different methods and algorithms some test has been performed, on different kind of datasets. Two datasets have been used, one from Denmark and one from Greenland. All of the different methods and algorithms has been tested on both dataset and evaluated based standard deviations and subjectivity. It has been concluded that the CUBE algorithm is the best way to process both types of data, but that it could be a good idea to clean the data rough before using the algorithm. Further some of the methods in EC-3D could be a good way to process the data, but where the methods should be used on top of each other.

Forord

Dette speciale er udarbejdet af Ane Rasmussen på Landinspektør uddannelsen. Specialet er resultat af 10. semester på uddannelsen og dermed afslutningen af denne. Specialet er skrevet i kandidat specialiseringen Opmåling og Kortlægning. Projektperioden har varet fra d. 1 februar til aflevering d. 8. juni 2018. Projektet forsvares mundtligt efter aflevering.
I forbindelse med specialet har der været eksterne samarbejder med henholdsvis EIVA A/S og Geodatastyrelsen. En speciel tak skal rettes til Ole Kristensen og Ole Lund Frederiksen fra EIVA som har bistået specialet ved at besvare en lang række spørgsmål og har haft en god tålmodighed. Der skal også rettes en speciel tak til Belen Jimenez fra Geodatastyrelsen til forståelse af data samt mange gode samtaler om forskellige typer af software.
Der skal selvfølgelig også rettes en særlig stor tak til vejleder ved universitetet Karsten Jensen for at hjælpe igennem projektet gennem god vejledning.

Læsevejledning

Til læseren anbefales det, at læse projektet i kronologisk rækkefølge, da projektet er opbygget efter denne struktur. Hele projektet er angivet med fortløbende numre. Det betyder, at kapitler starter fra 1 og fremefter, afsnit er noteret som kapitelnummer og derefter et fortløbende afsnitsnummer, eksempelvis 3.2 hvis det er andet afsnit i kapitel 3. Det samme gælder for figurer og tabeller, som noteres seperat. Det betyder at der både kan være et afsnit 3.1, en figur 3.1 og en tabel 3.1.
Det antages at læseren har en forståelse for opmåling, dog ikke nødvendigvis hydrografisk opmåling. Basale notationer og definitioner vil blive anvendt i projektet.
Kildehenvisninger sker efter Harvard metoden. Dette betyder, at der gennem projektet refereres til kilder som [Efternavn, Årstal][evt. sidetal]. Dette refererer til en litteraturliste bagerst i projektet, hvor yderligere detaljer om kilden kan findes.
Der er vedlagt et elektronisk bilag til projektet indeholdende alle figurer i projektet. Det kan være svært at se disse på papir, og derfor er de vedlagt elektronisk. De er navngivet på baggrund af deres figurnummer i projektet.

Ane Kirstine Aagaard Rasmussen

Forsidebilledet er fra https://i.ytimg.com/vi/_ ww2PflbWD8/maxresdefault.jpg

Indhold

1 Introduktion **1**

2 Problemformulering **3**

3 Metode **5**

4 Databearbejdningsprogrammer **9**
4.1 EIVA software 12
4.2 QPS software 13
4.3 CARIS software 15

5 Algoritmer og metoder **17**
5.1 CUBE 19
5.2 S-CAN 27
5.3 EC-3D 31

6 Analyse af algoritmer **45**
6.1 CUBE anvendt på dansk og grønlandsk data 47
6.2 S-CAN anvendt på dansk og grønlandsk data 57
6.2.1 S-CAN component 57
6.2.2 S-CAN score 64
6.3 EC-3D anvendt på dansk og grønlandsk data 69
6.3.1 Nearest Neighbour 70
6.3.2 Neighbour In Sphere 73
6.3.3 Z Summation In Circle 79
6.3.4 Distance To Average Surface 83
6.3.5 K Nearest Neighbour 87
6.3.6 K Nearest Neighbour, Z 90

7 Konklusion **95**

8 Diskussion **101**

Litteratur **103**

A Digitalt billag med figurer **105**

B Illustration af CUBE overflader - Teoretisk **107**

1 | Introduktion

Der samles meget data ind. Rigtig meget data. Der er sket meget i den teknologiske udvikling, som medvirker til, at der samles store mængder af data ind. Det er både hurtigere og lettere at samle data ind nu end for 50 år siden. Det kan være, at din telefon hele tiden har slået placering til, og at den derfor altid indsamler din position og lagrer den information. Det kan være, at du besøger en hjemmeside, hvorefter du får reklamer for denne hjemmeside i den følgende tid. Det kan være at du "liker"noget, og noget lignende dukker op i den følgende tid. Der lagres meget data om vores almindelig gøren og færden, som kan anvendes til mange ting. Der samles dog ikke kun data ind om os, men vi kan også være med til at informere og give data videre. Det kan være, at du informerer kommunen om en løs flise gennem en app, hvor appen lagrer informationen om den løse flise samt din position. Det medvirker til, at vi kan hjælpe hinanden gennem teknologien og vores data.
Når al denne data indsamles, skal den lagres og håndteres, og det er vigtigt, at de data der bliver indsamlet også bliver håndteret, sådan at informationen bliver brugt korrekt. Problemet er dog, at mange datasæt efterhånden er begyndt at blive store. Rigtig store. Når en bygning eksempelvis laserscannes dannes der flere millioner punkter, ligesom når Danmark bliver laserscannet fra fly, eller når der indsamles data om havbunden fra skib. Ved at indsamle data om havbunden fra skib, anvendes der multibeam, men tidligere var det mere almindeligt at anvende single beam og for endnu længere tid siden et lod. Ved singlebeam anvendes et ekkolod, hvor ekkoloddet er placeret under båden, og der bliver målt én dybde under det ekkolod, hvorimod et multibeam ekkolod kan måle omkring 500 punkter på tværs af sejlretningen afhængig af valg af multibeam. Dette giver altså 500 gange så store datasæt, og derfor kræver det noget mere at behandle disse data.
Når der arbejdes med multibeam data, er der tale om datasæt på op til flere millioner punkter, og det er derfor meget tidskrævende, at skulle behandle alle disse punkter. Det er ikke altid, at behandlingsprocesserne følger ligeså hurtigt med den teknologiske udvikling som instrumenter til indsamling af data. De store mængder af data kan være uhåndgribelige og ikke være til at finde hoved og hale i. Derfor er det vigtigt, at der er en form for struktur i behandlingsprocessen, sådan at data behandles korrekt. Den nuværende bearbejdningsprocess er meget tidskrævende, og derfor handler det om at optimere de enkelte processer så meget som muligt. Når der indsamles data med multibeam, skal data renses, hvilket kan være en omstændelig process og fungerer som en flaskehals. Her skal al støj fjernes, og når der er flere millioner punkter, giver det mange punkter at undersøge. Det kan derfor være interessant at undersøge hvordan støj og fejlbehæftede punkter kan fjernes fra datasættet med så lidt inddragelse fra brugeren som mulig, da det er meget tidskrævende at skulle undersøge hver enkelt punkt.

2 | Problemformulering

Der findes mange forskellige måder, at håndtere store mængder af data på. Der er mange forskellige typer af software til at behandle data, hvilket selvfølgelig også er tilfældet inden for hydrografisk software. De mange typer af software kan gøre det svært at finde rundt i, hvilke software der vil være ideel at anvende til at behandle en bestemt type af data med. For indeholder de forskellige typer af software de samme metoder til at behandle data med? Og er nogle bedre end andre?
Indsamles der hovedsageligt data i et område der er overvejende flad og ikke så stor variation i havbunden, kan det være, at en behandlings metode er bedre at anvende frem for en anden. Det samme kan være tilfældet, hvis der hovedsageligt skal indsamles og behandles data i meget kuperede områder. Det kan også være, at formålet med data er at identificere objekter på havbunden som eksempelvis rør, eller at formålet er, at finde vandrette områder på havbunden til placering af eksempelvis vindmøller.
Der er mange muligheder, men i dette projekt vil det kun undersøges hvordan data renses og gøres klar til anvendelse, eksempelvis produktion af søkort. Det leder frem til følgende problemformulering.

Hvordan efterbehandles multibeam data automatisk, og hvilke algoritmer og metoder ses som mest hensigtsmæssig at anvende til forskellige typer af data?

- **Teoretisk:**
  - **Hvilke behandlingsprogrammer findes der?**
  - **Hvilke algoritmer anvendes?**
  - **Hvordan virker algoritmerne?**
- **Analytisk:**
  - **Hvordan reagerer de forskellige algoritmer på forskellige typer af input?**
  - **Ses en algoritme som mere anvendeligt på en type af data frem for et andet?**

Projektet skal altså forsøge at undersøge, hvilke typer af software der findes, hvordan de er opbygget med de bagvedliggende metoder og derefter anvende metoderne til at vurdere om en metode vil være mere hensigtsmæssig at anvende til en bestemt type data frem for en anden. Det næste kapitel vil beskrive hvordan ovenstående problemformulering forsøges besvaret.

3 | Metode

Dette metode kapitel omhandler de metoder, der anvendes til at besvare problemformuleringen. Afsnittet er med til at redegøre for de tanker, der ligger til grund for projektet og hver enkelt delspørgsmål til problemformuleringen. Derudover har kapitlet til formål at illustrere den overordnede struktur.

Som beskrevet i indledningen samles der mere og mere data ind, fordi teknologien bliver bedre og bedre. Det er blevet lettere at samle store mængder data ind på kort tid end for nogle år tilbage, men der opstår også en række problemer med at håndtere og behandle de store mængder data. Det er derfor relevant at undersøge, hvordan de store mængder data kan behandles.
Alt data er forskelligt og behandling af data skal passe til den bestemte type af data. Det giver ikke mening at behandle data, hvis ikke kvaliteten af data kan forbedres, og derfor er det relevant at undersøge, hvordan data bedst og måske også hurtigst behandles, sådan at behandlingen ikke bliver en flaskehals i processen til at præsentere data i eksempelvis et søkort.

Projektstruktur

Det projektet skal forsøge at besvare er problemformuleringen, som lyder:

Hvordan efterbehandles multibeam data automatisk og hvilke algoritmer og metoder ses som mest hensigtsmæssig at anvende til forskellige typer af data?

Som en del af problemformuleringen er der opstillet en række delspørgsmål, og det følgende vil beskrive, hvordan hvert enkelt delspørgsmål forsøges besvaret.

1. Hvilke behandlingsprogrammer findes der?
Til at undersøge hvilke behandlingsprogrammer der findes, anvendes der søgninger på internettet. Søgninger på internettet giver mange resultater og kan medvirke til at der konstrueres en tabel over mulige programmer. Her er www.geo-matching.com en vigtig kilde, da det er en hjemmeside, der sammenligner både hardware og software fra mange forskellige producenter indenfor blandt andet opmåling og positionering. Inde på den hjemmeside findes en række software pakker til behandling af hydrografisk data, men andre steder undersøges dog også. Udover www.geo-matching.com vil der også anvendes tidligere viden omkring hydrografiske behandlingsprogrammer. Efter at det er undersøgt hvilke software der er på markedet, vil der udvælges nogle få, idet alle software pakker ikke kan behandles i dette projekt. Udvælgelsen af denne software vil blive baseret på umiddelbare indtryk af software samt tidligere viden og erfaring indenfor området.

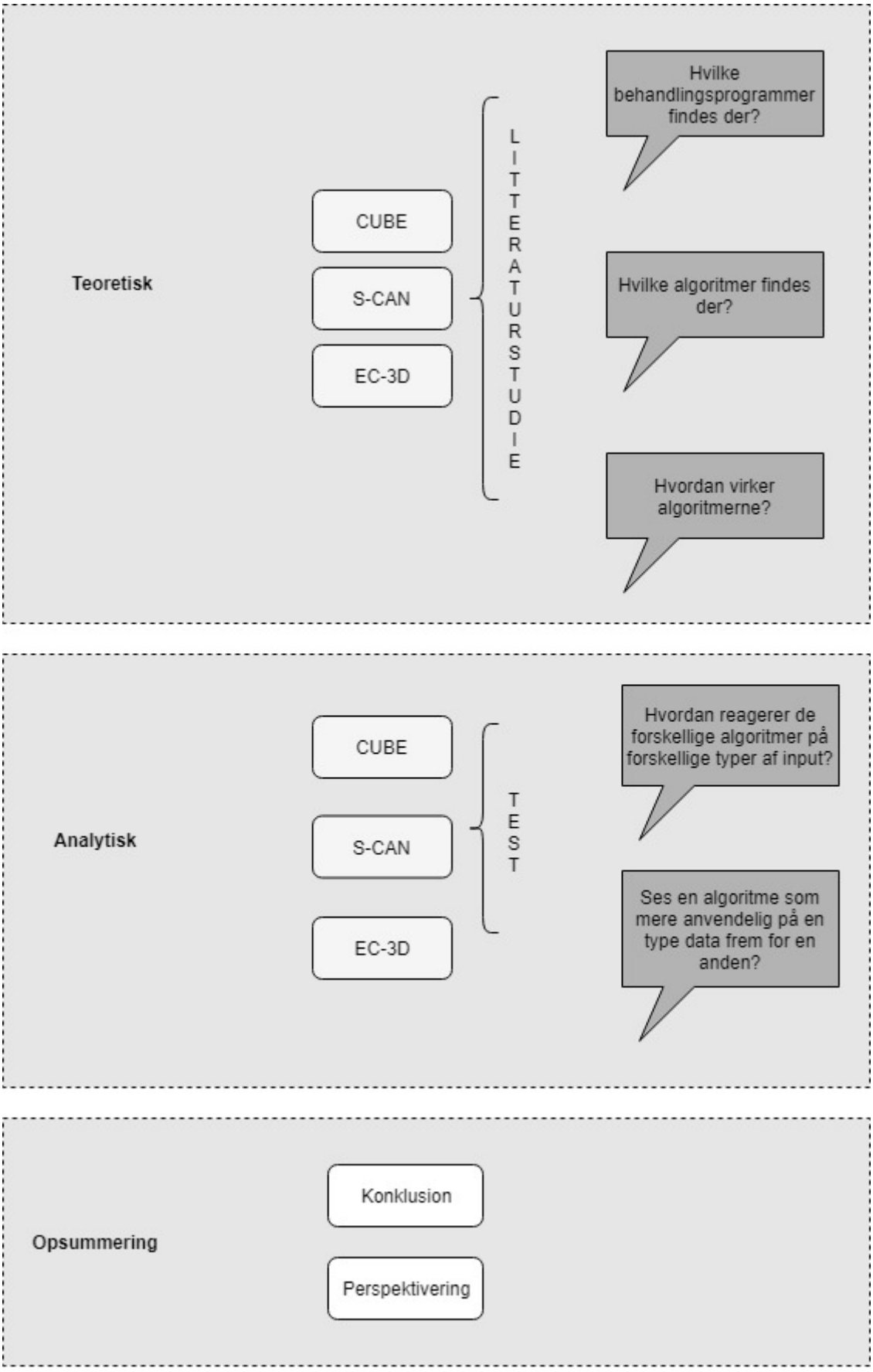
2. Hvilke algoritmer anvendes?
For at finde ud af, hvilke algoritmer der anvendes vil de udvalgte softwarepakkers manualer og software producents hjemmesider anvendes. Manualer og hjemmesider skal undersøges grundigt for at finde specifikt ud af hvilke metoder der anvendes, da det ikke er sikkert at der kan tilgås adgang til softwaren.

3. Hvordan virker algoritmerne?
Til at finde ud af hvordan de forskellige algoritmer fungerer, undersøges der igen litteratur. Der vil blive søgt på flere artikler fra udviklerne af algoritmerne, som skal medvirke til en forståelse for algoritmerne. Der findes litteratur der beskriver algoritmerne i forskellige niveauer. Nogle artikler vil være komplekse og nogle vil være lidt mere forsimplede, hvor det herefter handler om at kombinere det mere forsimplede og det komplekse, sådan at læseren kan forstå hvordan algoritmerne virker. Til at underbygge hvordan algoritmerne virker, vil der blive illustreret eksempler, som skal hjælpe læseren til en bedre forståelse for algoritmen.

4. Hvordan reagerer de forskellige algoritmer på forskellige typer af input?
Der kan anvendes mange tilgange til at undersøge, om de forskellige algoritmer og metoder reagerer forskelligt på forskellige typer af input. Der kan i projektet ikke undersøges alle typer af data, og det vælges at foretage test på et datasæt fra Danmark og et fra Grønland. Baggrunden for dette valg er, at havbundene i de to lande er forskellige, da havbunden i Danmark er overvejende horisontal, flad og uden meget variation, hvor den grønlandske havbund derimod er mere kuperet og kan variere meget inden for kortere afstande. Det vil blive undersøgt, hvordan algoritmerne og metoderne reagerer ved at anvende dem på begge datasæt. Herefter vil de behandlede data blive subjektiv vurderet og sammenlignet med de rå data.

5. Ses en algoritme som mere anvendeligt på en type af data frem for et andet?
Vurderingen om en algoritme ses som mere anvendelig på en bestemt type data vil udelukkende ske på baggrund af resultaterne fra de test, der er fortaget, for at se hvordan algoritmerne reagerer. Her skal konklusionerne fra de forskellige test medvirke til, at der kan vurderes, om der findes en algoritme eller metode, der er mere anvendelig end en anden. Det er ikke sikkert, at den samme algoritme eller metode er hensigtsmæssig at anvende til begge typer af data, men det kan være en mulighed. Der vil ikke blive foretaget yderligere test på de enkelte datasæt end dem, der er med til at besvare delspørgsmål nummer 4.

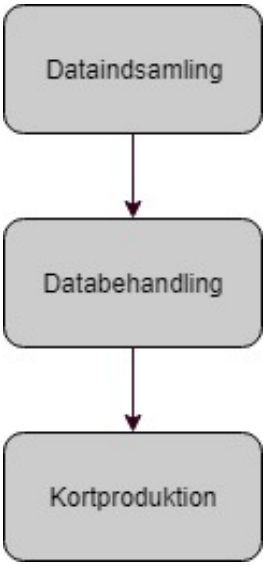
På Figur 3.1 ses et overblik over strukturen for for projektet. Heraf ses det, at de første tre delspørgsmål er baseret på litteraturstudie og er den teoretiske del af projektet. Derefter foretages den mere analytiske del af projektet, som er at teste de forskellige algoritmer og metoder på de forskellige typer af data. Til sidst opsummeres projektet i en konklusion og der foretages en perspektivering.



Figur 3.1. Struktur over projektet

4 | Databearbejdningsprogrammer

Der findes mange forskellige metoder til at behandle multibeam data, og der findes mange virksomheder, som arbejder med at udvikle software til behandling af disse data. De virksomheder, der udvikler software til behandling af multibeam data arbejder som oftest også med at udvikle software til dataindsamling og den efterfølgende kortproduktion. Det er dog ikke alle virksomheder, der har en komplet løsning indenfor software for alle tre trin i processen illustreret på Figur 4.1.



Figur 4.1. Processen til søopmåling og -kortproduktion

Figur 4.1 beskriver processen for at kunne lave et søkort. Først indsamles data hvilket blandt andet kan være fra skib, og ønsker læseren at læse mere om hvordan denne dataindsamling kan finde sted, henvises der til [Rasmussen, 2018]. Når data er indsamlet skal det behandles, og der er mange forskellige behandlingsmetoder, hvilket netop er det, som dette projekt omhandler. Her skal alle fejlbehæftede punkter gerne findes for efterfølgende at slettes fra punktskyen, dog uden at der slettes valid data. Når data er renset, er det klar til at kunne påbegynde kortproduktionen. Denne kortproduktion vil ligesom dataindsamlingen heller ikke blive beskrevet i dette projekt.

Der er virksomheder, som kun beskæftiger sig med at udvikle software til bestemte dele af denne process, men de fleste virksomheder har som oftest software til flere dele af processen, idet der er en naturlig sammenhæng. I Tabel 4.1 ses en række producenter samt deres produkter.

Producent	Software
Chesapeake Technology, Inc.	SonarWiz Post-Processing: Bathymetry SonarWiz Post-Processing: Sub-bottom SonarWiz 7 Office Post-Processing Suite SonarWiz 7 Single Beam Acquisition and Processing
Triton Imaging Inc	Triton Perspective-MBE Triton Perspective-SSS
Teledyne CARIS	CARIS HIPS and SIPS CARIS Onboard
BeamworX	AutoClean AutoPatch
EIVA	EIVA NaviSuite Kuda – Post-Processing
HYPACK	HYPACK Acoustic HYPACK MAX HYPACK Office HYPACK SUB-BOTTOM HYSWEEP HYSWEEP OFFICE
Stema Systems	Silas Processing
Eye4Software B.V.	Hydromagic
QPS	QPS Qimera
Teledyne Marine	Teledyne PDS

Tabel 4.1. Forskellig efterbehandlingssoftware, [https://geo-matching.com/en/category/hydrographic-processing-software]

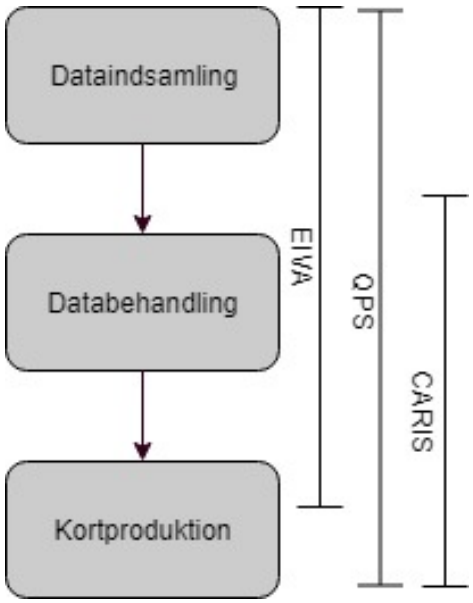
Informationerne i tabellen er fundet gennem hjemmesiden www.geo-matching.com, som er en hjemmeside til at finde og sammenligne over 2000 produkter fra 500 forhandlere inden for opmåling og positionering. Sammenligning omfatter både hardware og software fra forskellige virksomheder, som arbejder med geografiske data. Hjemmesiden sammenligner langt mere end blot hydrografiske efterbehandlingsprogrammer, men for at finde ude af mere om den, henvises der til hjemmesiden, www.geo-matching.com. Til at konstruere Tabel 4.1 er det undersøgt, hvilke software der var indenfor kategorien "Hydrographic Processing Software".
Udover det software, der er opstillet i Tabel 4.1, findes der dog andre producenter, der ikke er i søgekartoteket ved www.geo-matching.com. Blandt andet vides det, at der findes Kongsberg, som også har udarbejdet en liste over hydrografisk databehandlings software. Størstedelen af det software der er på denne liste, findes dog også i Tabel 4.1, men lidt software er ikke gengangere og de er listet op i Tabel 4.2.

Producent	Software
Kongsberg	SIS MDM
Geocap	Seafloor

Tabel 4.2. Yderligere efterbehandlingsprogrammer, [Kongsberg, Ukendt]

Det kan altså ske, at der findes yderligere software udover det nævnte i Tabel 4.1 og Tabel 4.2, men tabellerne dækker mange af de producenter og efterbehandlingsprogrammer, der findes. Det er umuligt i dette speciale at undersøge og beskrive al software, hvilket betyder, at der skal udvælges

noget software til nærmere undersøgelser. Softwaren udvælges ud fra et tidligere kendskab til producenter. EIVA udvælges, idet det er en dansk virksomhed, der er godt med i forhold til udviklingen og at der allerede er et kendskab til denne producent. CARIS udvælges også, idet det vides at deres kortproduktions software er anvendt flere steder og det er derfor interessant at undersøge deres software til databearbejdning. Til sidst udvælges QPS idet der også er et begrænset kendskab til denne producent og det vides, at de tilbyder hele pakken af software til at indsamle data, til behandling af data og til kortproduktionen.
På Figur 4.2, ses det, hvilke dele af processen, som de tre udvalgte virksomheder producerer software indenfor.



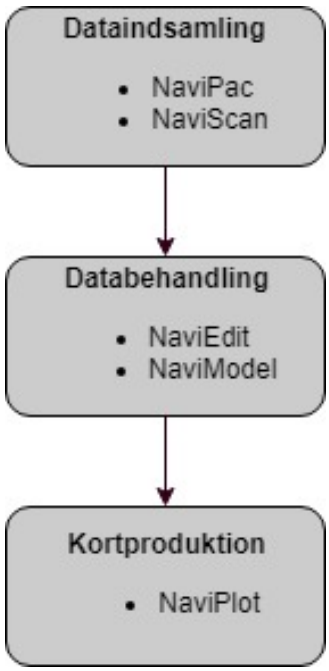
Figur 4.2. Hvordan de forskellige virksomheder dækker software

Der er altså udvalgt tre producenter og deres software, hvilket vil blive undersøgt i de følgende afsnit. Det er vigtigt at understrege at producenterne er valgt udelukkende på et eksisterende kendskab til deres eksistens. Der er altså ikke opstillet krav om hvilke kvalifikationer, som de skal opfylde, og er ikke valgt på baggrund af indholdet og deres anvendte algoritmer og metoder.
I det følgende vil softwarepakkerne fra de tre producenter blive præsenteret.

4.1 EIVA software

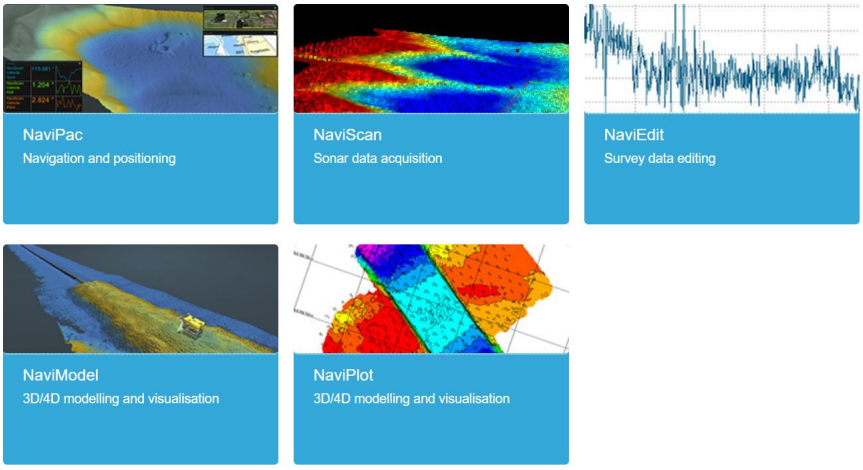
EIVA er en dansk virksomhed som udvikler software til indsamling og behandling af data indsamlet til søs. EIVA opererer inden for mange forskellige områder som f.eks. olie, gas, vindmøller, boreplatforme, vandmiljø og opmåling af havbunden.

I dette projekt er det interessante område opmåling og kortlægning af havbunden. Her har EIVA en række programmer til at indsamle, behandle og til dels at kortlægge data. Den måde EIVA har opbygget deres programmer på, ses på Figur 4.3.



Figur 4.3. EIVA software

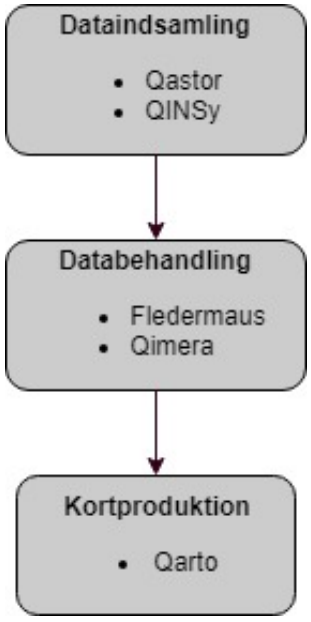
Her ses det, at **NaviPac** og **NaviScan** anvendes til dataindsamlingen.
NaviEdit anvendes til at foretage editeringer i data inden en punktsky eksporteres. Alle editeringer sker på sensor niveau, hvilket betyder, at der korrigeres for hver enkelt sensor, eksempelvis GPS højde, inden at alle sensorerne integreres til samlet output. Denne editering i indsamlede data påvirker så den punktsky som dannes og eksporteres for derefter at åbne i NaviModel.
NaviModel er det software, der anvendes til at efterbehandle eller modellere data, og ses derfor som det mest relevante for dette projekt. EIVA tilbyder et stykke software som giver flere forskellige muligheder for valg af metode, som kan anvendes på baggrund af hvad brugeren synes er mest hensigtsmæssigt.
Til sidst har EIVA også kortprogrammet **NaviPlot**, som kan anvendes til at konstruere simple kort. Programmet kræver i modsætning til de øvrige ikke en licens, og er til konstruktion af simple kort. EIVA tilbyder altså programmer til hele processen med at samle data, behandle dem og præsentere dem på et kort. De interessante software fra EIVA er i dette projekt NaviModel, der spiller en stor rolle for behandlingen af den indsamlede data.



Figur 4.4. EIVA software, [https://www.eiva.com/products/eiva-software/product-guide]

4.2 QPS software

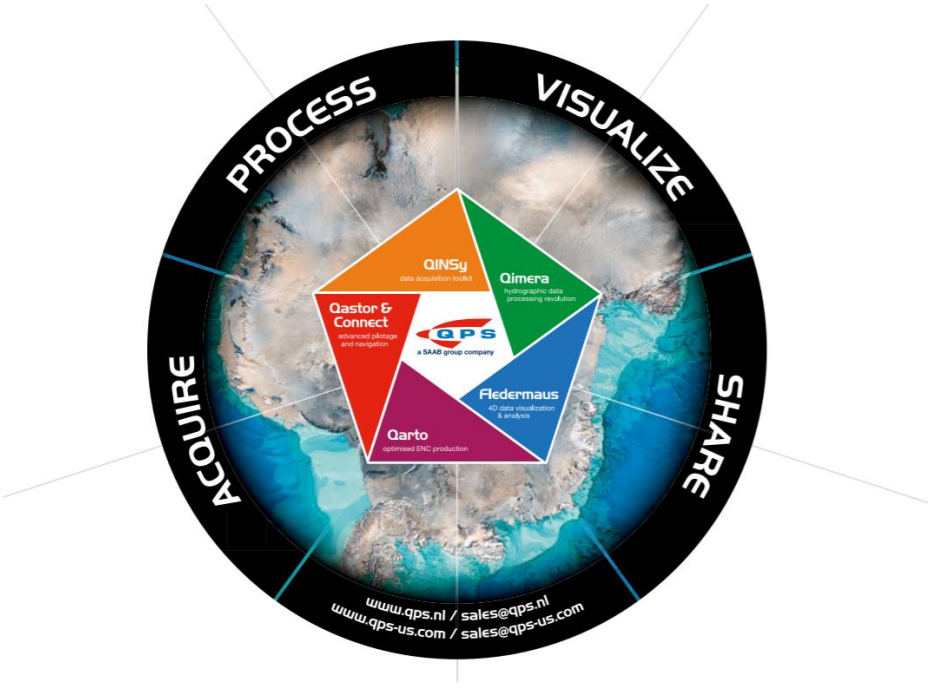
QPS er en hollandsk virksomhed, som udvikler software vedrørende dataindsamling til søs og behandling af denne data. QPS står for Quality Positioning Services og er en software virksomhed, som blev grundlagt i 1986. QPS har hovedkvarter i Holland, men har også kontorer i USA, Canada og England. Det følgende afsnit baseres på [QPS, 2017].
QPS's fokus er på at integrere flere opmålingssensorer samt at udvikle software til opmåling, kortlægning af havbunden og produktion af elektroniske søkort(ENC - electronic navigation charts).
Ligesom EIVA har QPS også en software pakke tilpasset til de forskellige processer. Et overblik over dette, kan ses på Figur 4.5.



Figur 4.5. QPS software

Qastor er et Electronic Chart System, som muliggør navigation, lodsning og præcis dokning. Qa-

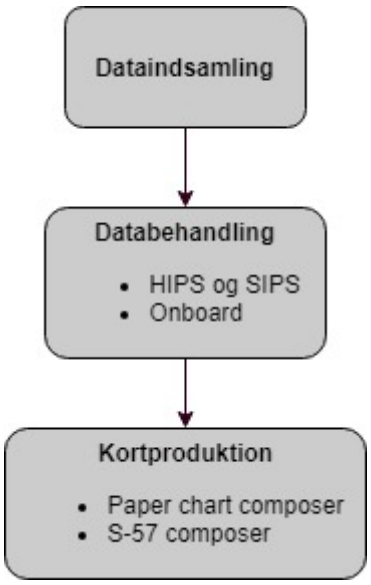
stor anvendes af navigatøren til at navigere og blandt andet til at se retning og hastighed. Foruden dette visualiseres også søkort, hvilket blandt andet viser dybdekurver og dybder.
QINSy er et navigations- og positioneringssoftware, der anvendes ombord på eksempelvis borerigge, konstruktions fartøjer og opmålingsskibe. Derudover er QINSy også i stand til at efterbehandle data, hvor der blandt andet også kan tilkobles flere sensorer, som eksempelvis ekstra multibeam.
Efterbehandlingen af data er en vigtig del af programmet, og udviklerne har lagt meget vægt på, at det skal være tidsbesparende og mindske sandsynligheden for at skulle ud og genopmåle. Derfor kan programmet tilbyde funktioner, der kvalitetstjekker realtime, men også hvor offset i sensorer, lydhastigheds-refraktion, korrektion for tidevand og DTM punkter afledes. Når alt dette er afledt og beregnet, fås en DTM, over det område som er blevet opmålt. På den måde kan det løbende vurderes om de indsamlede data umiddelbart har en passende kvalitet. Softwaren har derudover en række andre funktioner, som er meget anvendelige i databehandlingen. Eksempelvis kan softwaren beregne usikkerheder vertikalt og horisontalt (Total horizontal/vertical uncertainty(THU/TVU)).
Fledermaus er software, der er udviklet til at visualisere i 4D. Programmet anvendes til at analysere og visualisere, for derefter at kunne dele data. Derudover tilbyder programmet effektivitet i søkortlægning, geologisk fortolkning og vurdering af habitater. Det kan til dels ses som hydrografernes GIS program, og Fledermaus tilbyder faktisk direkte forbindelse til et ESRI "workspace"i ArcGIS for maritime.
Qimera er en applikation som anvendes til at behandle det indsamlede data med.Det kan ses som et stykke software, der er udarbejdet efter de vigtigste elementer fra QINSy og Fledermaus. Qimera anvendes derfor både i efterbehandling af data og derefter til visualisering og analysering. Da Qimera er et produkt, som er udarbejdet på baggrund af både QINSy og Fledermaus, ses dette som værende interessant at undersøge for dette projekt.
Qarto anvendes til at producere elektroniske søkort. Det ses derfor, at QPS tilbyder en pakkeløsning, hvor hele processen til at lave søkort er indeholdt.



Figur 4.6. QPS software, [QPS, 2017]

4.3 CARIS software

Teledyne CARIS, for det meste bare omtalt CARIS, har i 35 år udviklet software designet som GIS til det maritime område, og har derfor stor erfaring inden for området. Det er udviklet i samarbejde med hydrografer og universiteter, sådan at værktøjerne i programmet tilbyder det, som brugerne efterspørger.
CARIS har umiddelbart ikke software til dataindsamlingen, men omfatter software til både behandling af data og til kortproduktion. Et oversigt over hvordan en del af CARIS's software ser ud ses på Figur 4.7. [CARIS, 2018a]

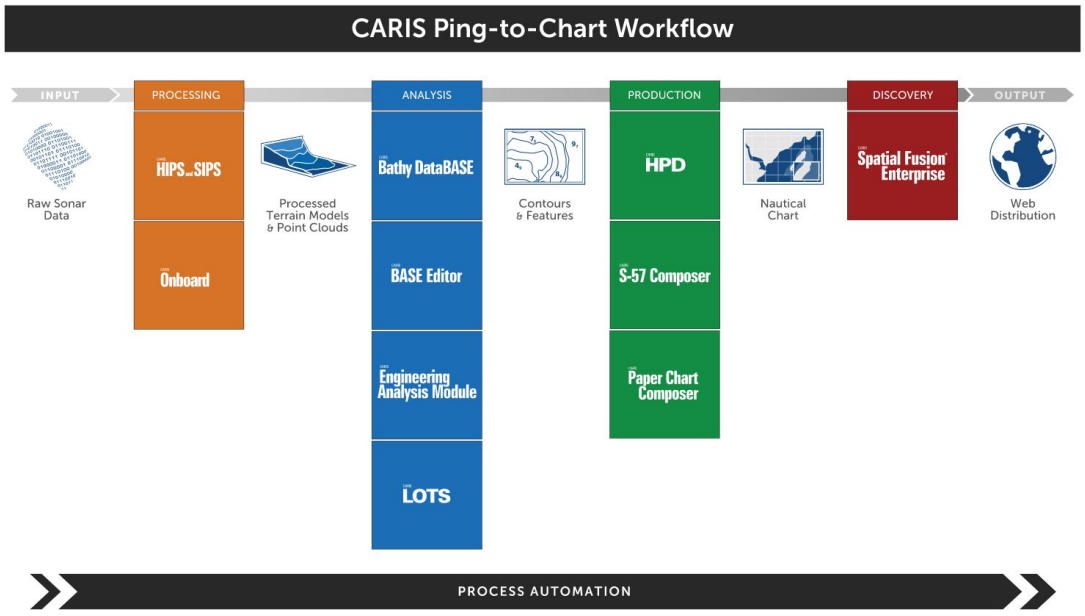


Figur 4.7. CARIS software

Til efterbehandling af data er der to stykker software, HIPS og SIPS som det ene og CARIS Onboard som det andet.
CARIS Onboard er et forholdsvis nyt produkt fra CARIS, og er en real-time databehandlingspakke, som er udviklet med henblik på at behandle data indsamlet fra undervandsfartøjer eller ubemandede fartøjer.[CARIS, 2018a]
HIPS og SIPS er et system til at behandle hydrografisk data, som tilbyder den seneste teknologi, og kan anvendes på store mængder af data og sikrer, ifølge producenten selv, tilfredsstillende resultater. Udover at HIPS og SIPS kan anvendes på store mængder af data kan programmet også anvendes til at sammenkoble data i forskellige opløsninger(forskellig grid størrelse). Derudover kan det også generere en række forskellige output såsom en elevations model, geologiske mosaikker og kontur linjer. Med hurtig generering af ovenstående kan en plot composer anvendes ombord på båden til hurtigt at konstruere kort, såfremt det ønskes. [CARIS, Ukendt]
De to programmer **Paper Chart Composer** og **S-57 Composer** anvendes til kortproduktion. Paper Chart Composer anvendes til produktion af analoge papir søkort, mens S-57 composer anvendes til at producere elektroniske søkort.[CARIS, 2018b]

CARIS har altså ikke en komplet løsning til hele processen for at konstruere søkort, men har nogle mere avancerede kortproduktionsprogrammer end eksempelvis EIVA.
CARIS arbejder med begrebet Ping-to Chart, hvilket beskriver processen fra en opmåling til et færdigt kort. Alle deres programmer kan anvendes alene, eller i samarbejde med andre

virksomheders programmer, men CARIS ser det bedst at anvende alle deres programmer samlet for at få mest ud af data. En figur over Ping-to-Chart ses på Figur 4.8.[CARIS, Ukendt]



Figur 4.8. CARIS Ping-to-Chart, [http://www.caris.com/ping-to-chart/]

For at læse mere om andre programmer i forbindelse med Ping-to-Chart henvises der til CARIS's egen hjemmeside.
Der findes altså en række interessante software systemer, der kan anvendes til at efterbehandle data indsamlet med multibeam. Udover ovenstående, findes der yderligere en række software, hvilke dog ikke er beskrevet eller vil blive beskrevet i dette projekt.
De programmer, det kunne være interessant at undersøge yderligere er:

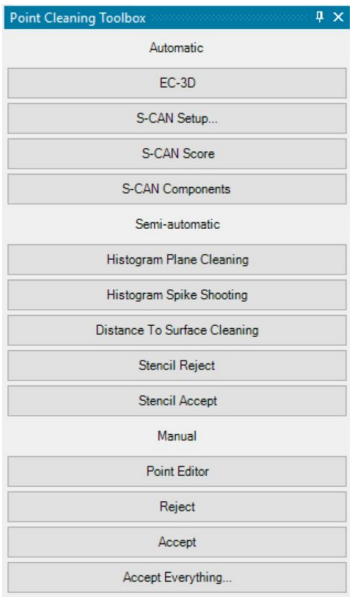
- NaviModel
- Qimera
- HIPS og SIPS

Det er interessant at undersøge, hvilke algoritmer og metoder, der ligger til grund for softwaren, for efterfølgende at kunne undersøge de forskellige efterbehandlingsmetoder. Det næste kapitel vil derfor undersøge, hvilke algoritmer og metoder, der anvendes i softwaren samt en beskrivelse af disse.

5 | Algoritmer og metoder

Efter at have undersøgt nogle af de mange softwarepakker, der findes til at behandle multibeam data med, er det interessant at undersøge, hvilke algoritmer og metoder der ligger bag softwaren. Dette kapitel har derfor til formål, først at undersøge og derefter beskrive hvilke algoritmer der anvendes. Kapitlet er derfor meget teoretisk.
Kapitlet er opdelt på den måde, at det først undersøges hvilke algoritmer, der anvendes for hver enkelt software som så opsummeres det i en tabel. Derefter beskrives algoritmerne og metoderne mere detaljeret.

EIVA's **NaviModel** giver brugeren en række valgmuligheder, hvad angår efterbehandling af data. Når data importeres i NaviModel genereres der en terræn model på baggrund af de rå data. Ud fra denne DTM kan der foretages en række efterbehandlinger, og algoritmen CUBE kan vælges at anvende til at generere en anden terræn model. Derudover findes der i EIVA's software yderligere metoder til at behandle data med, hvilket er vist på Figur 5.1.



Figur 5.1. Oversigt over mulige metoder til efterbehandling af data

Disse databearbejdningsmetoder kan ud fra Figur 5.1 ses at være både automatiske, semi-automatiske og manuelle. Den manuelle valgmulighed giver brugeren mulighed for at editere i enkelte punkter.
EIVA har altså flere metoder til at behandle data med, og det er interessant at undersøge, hvordan disse metoder virker. De metoder fra EIVA der vil blive undersøgt fremadrettet er altså *CUBE*,

*S-CAN*og *EC-3D*, da det er disse metoder, der er automatiske og derfor interessante for dette projekt.

Det hollandsk udviklede **Qimera** er QPS's pendant til EIVA's NaviModel. Programmet er en sammenkobling af grundelementerne fra to andre QPS programmer og har flere værktøjer til automatisk at rense data. Disse er *Spline filter* og *CUBE* filter. [QPS, 2019] Der vil dog kun undersøges CUBE af disse to fremadrettet i dette projekt, idet der ikke har været mulighed for senere at teste Spline filtret.

Caris HIPS og SIPS er også et behandlingsprogram ligesom Qimera og NaviModel, og anvender også algoritmer til behandling af data. HIPS og SIPS tilbydes i forskellige pakker, hvor enten kun HIPS, SIPS eller begge dele indgår. På nedenstående figur ses det, hvilke funktioner der er i forskellige software pakker.

	SIPS Essential	HIPS Essential	HIPS & SIPS Essential	HIPS Professional	HIPS & SIPS Professional	
Single beam bathymetry		⬣	⬣	⬣	⬣	Sensor data
Multibeam bathymetry		⬣	⬣	⬣	⬣	
Side Scan imagery	⬣		⬣		⬣	
Multibeam imagery	⬣		⬣		⬣	
Water Column Imagery (WCI)/Multiple Detections				⬣	⬣	
Lidar				⬣	⬣	
Multi-user processing				⬣	⬣	
Calibration		⬣	⬣	⬣	⬣	
Post-processed auxilliary data				⬣	⬣	
Total Propagated Uncertainty (TPU) and CUBE		⬣	⬣	⬣	⬣	
IHO QC tools				⬣	⬣	
Variable Resolution (VR) Surfaces				⬣	⬣	
Contouring and soundings		⬣	⬣	⬣	⬣	
Additional GIS tools				⬣	⬣	Features

Figur 5.2. Oversigt over softwarens egenskaber [http://www.caris.com/essential/]

Af ovenstående figur, ses det, at CUBE anvendes i HIPS og derved også i HIPS og SIPS, hvilket er det relevante til dette projekt.

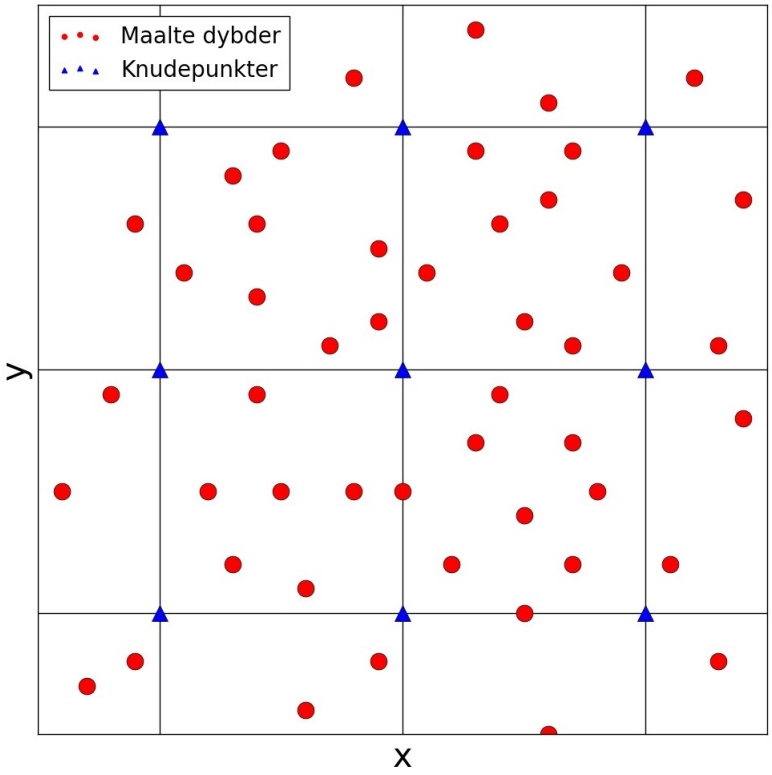
De samlede algoritmer, der anvendes er følgende:

CUBE: Combined Uncertainty and Bathymetric Estimator
S-CAN: SCALGO Combinatorial Anti Noise
EC-3D: EIVA Cleaning 3D

Nu er det interessant at undersøge hvordan disse algoritmer virker og er opbygget, hvilket de følgende afsnit vil beskrive.

5.1 CUBE

CUBE, **C**ombined **U**ncertainty and **B**athymetric **E**stimator, er en DTM generator, der estimerer dybder samt usikkerheder for en række knudepunkter i et batymetrisk grid, se eksempel på Figur 5.3. Algoritmen er udviklet at Brian R. Calder fra University of New Hampshire og er kompleks.

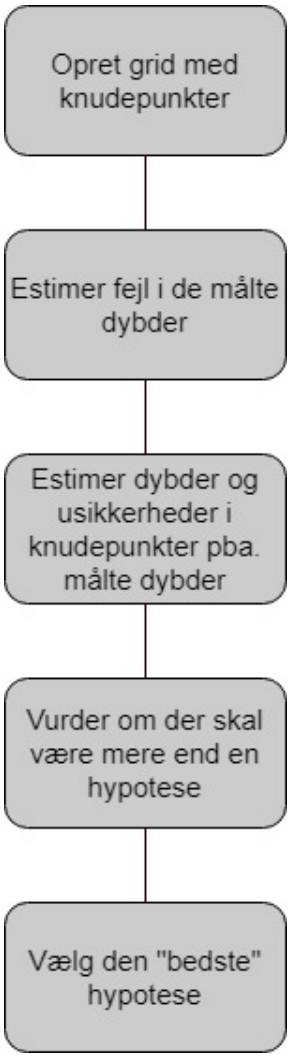


Figur 5.3. Illustration i 2D af knudepunkter og målte dybder. Punkterne er fiktive og tilfældig placeret.

Formålet med CUBE, er at bestemme den sande dybde samt dennes usikkerhed til en række knudepunkter. Ved at forsøge at besvare spørgsmålet om hvad den sande dybde er, adskiller CUBE sig fra en række andre algoritmer som også anvendes til at bearbejde multibeam data. Andre algoritmer forsøger generelt at besvare spørgsmålet om, hvor god den enkelte måling er, hvilket VISE er et eksempel på. Her vurderes det på baggrund af et punkts nabolag hvorvidt punktet er en outlier eller ej, men det vurderes ikke om hele nabolaget samt punktet er placeret i en forkert dybde. Det CUBE gør, er at forøge at besvare spørgsmålet om hvor god dybden er i et givent punkt, samt at estimere en usikkerhed for denne dybde, og dermed besvare spørgsmålet hvor godt dybden i punktet kendes.
Når der er estimeret dybder og usikkerheder i alle knudepunkter dannes en overflade og herfra vurderes det så, på baggrund af hydrografisk viden hvorvidt denne overflade er passende til at repræsentere havbunden eller ej. Det et ikke altid let at vurdere, om denne model af havbunden er repræsentativ for området, og derfor kan mere data anvendes til at opdatere estimaterne i hver enkelt knudepunkt. På den måde danner mere end én opmålt dybde grundlag for estimatet af dybden. Ved at anvende mere end én opmålt dybde til at estimere knudepunktet skabes formentlig også en mindre usikkerhed for knudepunktets dybde.

Når CUBE anvendes skal der tages højde for en række faktorer, der skal inddrages i algoritmen. Det er eksempelvis nødvendigt at kende de fejl, der er i målingerne foretaget med multibeam, for at kunne estimere usikkerheden for en estimeret dybde. Hver gang en dybde inddrages til at estimere dybden i et knudepunkt, bidrager målingens usikkerhed nemlig til den estimerede usikkerhed for knudepunktets dybde. Det skal også vurderes, hvordan en opmålt dybde kan anvendes til at estimere den "sande" dybde i knudepunktet. Derudover skal der opstilles en grænse, for hvornår et punkt har en statistisk variation fra de resterende data, frem for at være en reel outlier. Disse faktorer spiller ind i hvordan CUBE algoritmen virker og beskrivelse af dette vil ske i det følgende. Algoritmen er meget kompleks og derfor er alt matematikken udeladt for blot at beskrive algoritmen. Matematikken er tung at læse og forstå, og såfremt det ønskes at læses, henvises der til [Calder and Mayer, 2003][Appendix A, s. 16]. I det følgende vil der først gives et kort overblik over algoritmen, hvorefter det bliver skrevet mere detaljeret.

Til CUBE anvendes der rå data som input, og de målte dybder anvendes til at opdatere estimaterne i knudepunkterne. Det er dog ikke sikkert, at alle de målte dybder ikke er fejlbehæftede, og derfor indsættes der en form for kontrol i modellen, der skal undersøge hvorvidt de nyindkommende dybder er i overensstemmelse med det nuværende estimat for dybden i knudepunktet. Såfremt den nye information omkring dybden er i overensstemmelse med det eksisterende estimat implementeres dybden i modellen. Er dybden ikke i overensstemmelse med estimatet i knudepunktet betegnes det som inkonsistent. Når data er inkonsistent betyder det, at dybden er beliggende over en specificeret grænse, der er afhængig af fejl i den enkelte dybde måling. Såfremt data er inkonsistent skal det ikke implementeres, idet det kan være tegn på at der er fejl i målte dybder. Der oprettes i så fald en ny model til at repræsentere de(n) inkonsistente måling(er) og derved oprettes der en hypotese om flere forskellige modeller i det givne knudepunkt. Ved at foretage denne kontrol, giver det mulighed for, at der kan repræsenteres flere potentielle estimater(hypoteser) i et knudepunkt. Det skal så efterfølgende vurderes, hvilken hypotese der er mest sandsynlig i hvert knudepunkt. Den mest sandsynlige hypotese kan vælges på forskellige baggrunde, hvilket eksempelvis kan være at vælge den hypotese hvor der er flest målte dybder, der stemmer overens med den estimerede dybde eller at vælge den hypotese, der har kortest afstand til nabodybderne. Når der er valgt den mest sandsynlig hypotese i alle knudepunkterne er der således konstrueret en mængde af punktestimater og usikkerheder for opmålingsområdet, der teoretisk repræsenterer et statistisk estimat for en dybde i et bestemt lokation. Efterfølgende genereres en DTM model over området på baggrund af estimaterne. En figur over processen for CUBE kan ses på Figur 5.4. Det følgende vil beskrive her af de forskellige skridt i flowchartet.



Figur 5.4. Process i CUBE

Opret grid med knudepunkter

Først oprettet et grid med knudepunkterne. Gridstørrelsen afhænger af hvor stor opmålingsområdet er, men også hvilken opløsning der ønskes at arbejde med. Algoritmen stiller ingen krav til placering eller afstand mellem knudepunkterne, så længe de er placeret tæt nok på hinanden til at give et repræsentativ billede af havbunden. Hvis der eksempelvis måles med 500 beams på tværs af sejlretningen og en frekvens på 50Hz vil der skulle vælges en passende gridstørrelse. Størrelsen på griddet vil så afhænge af hvor dybt der er, idet dette påvirker afstanden mellem de opmålte dybder. Er der ikke så dybt, vil der være en større punkttæthed frem for hvis der er dybt.
Afstanden mellem punkterne på havbunden afhænger også af den beamformning, der laves. Her kan det vælges, at måle med en fast vinkel eller med en fast afstand mellem punkterne på havbunden. Måles der med en fast vinkel vil afstanden mellem punktene på havbunden øges, des længere væk fra nadirpunktet der kommes. Tabel 5.1 beskriver henholdsvis den maksimale og den minimale afstand mellem to punkter(d_{MAX} og d_{MIN} samt afstanden, hvis der anvendes den beamformning hvor afstanden mellem punkterne på havbunden er lige stor(d). Det skal bemærkes at tabellen er beregnet ud fra normaltilfældet og med en åbningsvinkel på 65° hvilket er vinklen fra lodlinjen og på tværs af sejlretningen, og giver derfor en samlet opmålingsvinkel på 130°.

Samtidig regnes der med 500 beams.

Dybde	$\mathbf{d_{MIN}, m}$	$\mathbf{d_{MAX}, m}$	$\mathbf{d, m}$
25 m	0,1	0,6	0,2
100 m	0,5	2,5	0,9
200 m	1,0	5,0	1,7
400 m	2,1	10,1	3,4
700 m	3,7	17,6	6,0

Tabel 5.1. Den maksimale, den minimale og den gennemsnitlige afstand mellem punkterne på havbunden

Dette gælder på tværs af sejlretningen. Langs sejlretningen vil tætheden af punkter afhænge af den hastighed, der sejles med samt hvilken frekvens, der måles med. Sejles der med en hastighed på 8 knob svarer det til ca. $4,12$ m/s. Måles der med en frekvens på 50 Hz betyder at der måles punkter med en afstand på $0,1$ m.

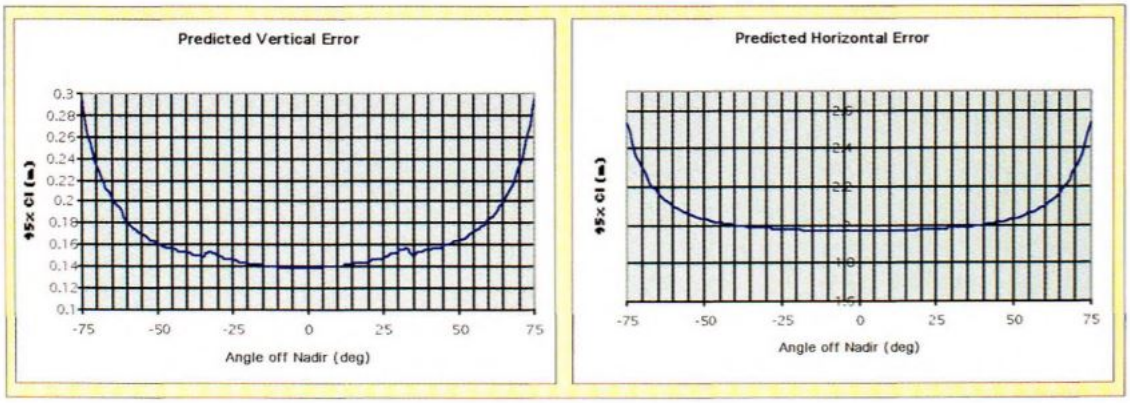
Eksempler fra [Calder and Mayer, 2003] angiver eksempelvis data med en afstand mellem knudepunkterne op henholdsvis 2 m og 4 m på lavt vand. Dette kan ses som hensigtsmæssigt, idet punkterne her ligger tættere på hinanden, men vil ikke ses som hensigtsmæssigt at anvende på dybt vand, idet afstanden mellem punkterne er større. Der er derfor ikke sikkert, at knudepunkter kan blive fyldt med nok information fra deres nabolag til at kunne estimere en "korrekt"dybde.

Knudepunkternes lokation er fikset og defineret i et projekteret koordinatsystem og grundet dette, kan der kun betragtes én korrekt dybde i et knudepunkt, da der til et givent punkt aldrig vil kunne forefindes to dybder. Da der kun kan forefindes én dybde i et knudepunkt, betyder det, at der kun skal estimeres én ukendt konstant, hvilket gør estimeringen let, frem for at skulle estimere flere værdier i knudepunktet.
Fordi knudepunktets lokation er absolut kendt, skal der ikke tages hensyn til den horisontale usikkerhed, da denne er ikke eksisterende eller i hvert fald lille. Det betyder, at det kun er den vertikale usikkerhed, der skal tages højde for knudepunkterne.
Det antages, at alle målte dybder er korrekte og fejlfi, og ved denne antagelse kan det fremadrettet være ligegyldigt i hvilken rækkefølge dybderne anvendes til at estimere dybder og usikkerheder i knudepunkterne. Når rækkefølgen er ligegyldig, kan data behandles løbende under dataindsamlingen eller det kan behandles på en gang når alt data er indsamlet.
Til at estimere den "sande" dybde i et knudepunkt tilknyttes en række målte dybder i umiddelbar nærhed til knudepunktet. Estimatet for knudepunktets dybde samt usikkerhed bestemmes ved anvendelse af et Kalman filter, der er en rekursiv algoritme, som estimerer på baggrund af en tidsserie.

Estimer fejl i de målte dybder

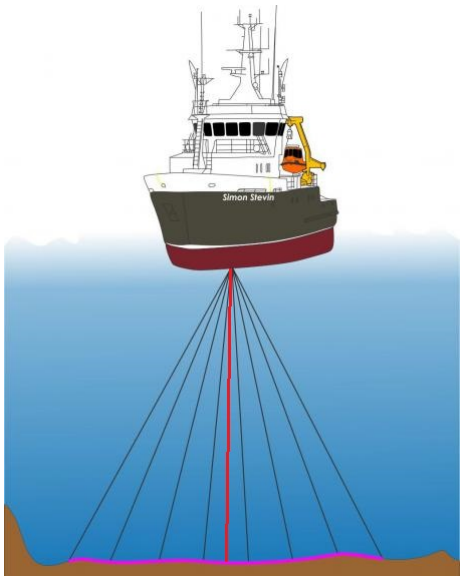
For at kunne anvende de målte dybder til at estimere dybden i knudepunktet, er det nødvendigt at have et estimat for usikkerheden i de målte dybder. Dette sker gennem fejlforplantning, hvor der på baggrund af fejl i multibeam ekkoloddet og hjælpesensorer såsom GPS og IMU estimeres en fejl forbundet til hver målt dybde. Til at anvende fejlforplantning og derved bestemme et estimat for fejlen hørende til hver målt dybde er der mange faktorer, der spiller ind såsom det lokale koordinatsystem på fartøjet, der beskriver hvor de forskellige sensorer er placeret i forhold til

hinanden. På Figur 5.5 ses et typisk output fra fejlmodellen. Hvordan dette output ser ud, afhænger dog af det udstyr, der måles med, og derfor ses figuren blot som et eksempel.



Figur 5.5. Typisk fejlmodel for et multibeam system i lavt vand. Dybden i dette eksempel er ca. 25 m. Modellerne viser usikkerheden i det horisontale og det vertikale plan for punkter i en vinkel fra nadir punktet, [Calder, 2003]

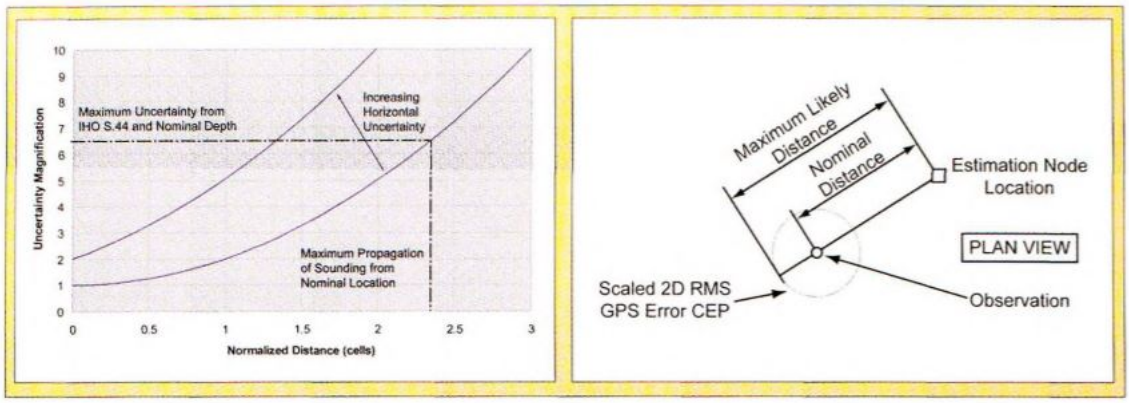
Fejlmodellen præsenteret på Figur 5.5 beskriver, hvordan fejlene ses som være mere indflydelsesrige, des længere væk der kommes fra nadir punktet. Nadirpunktet er det punkt, der er lodret fra ekkolodet og er illustreret på Figur 5.6.



Figur 5.6. Illustration af nadirpunktet, som er præsenteret med rød. Nadir punktet er det punkt på havbunden, som lodlinjen fra ekkoloddet rammer, [http://www.vliz.be/en/multibeam-en](manipuleret)

Fejlmodellen på Figur 5.5 er fundamental for hele CUBE, idet det er den model, der medvirker til at vurdere, hvornår et punkt anses som værende fejlbehæftet. Modellen er derfor vigtig, men det skal også tages in mente, at modellen kun er repræsentativ indenfor en afstand af den målte dybde. Fejlmodellen på figuren beskriver både den horisontale fejl og den vertikale fejl, men da knudepunkterne er fikset og kun har en vertikal fejl skal fejlen for de målte dybder formidles på en sådan måde at den kan sammenholdes med knudepunktets usikkerhed. Dette sker, ved at

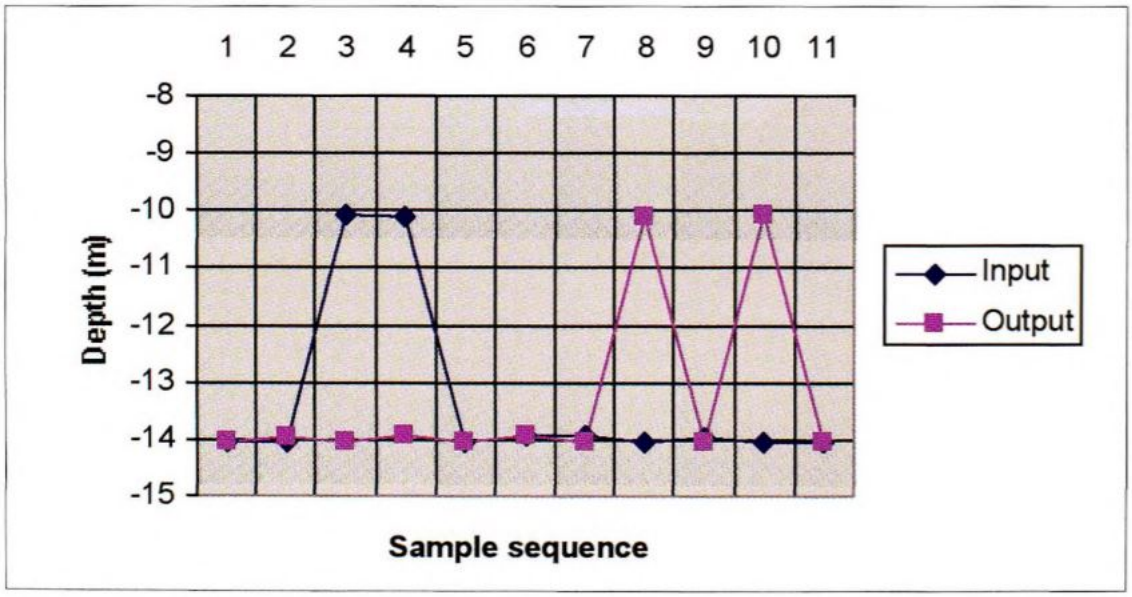
kombinere den horisontale og den vertikale usikkerhed til én samlet usikkerhed for den målte dybde, og de to usikkerheder kan nu sammenlignes.
I algoritmen tages der også hensyn til hvor langt væk en målt dybde er i forhold til knudepunktet. Derfor må usikkerheden for en målt dybde øges som en funktion af afstanden til knudepunktet. Dette gøres ved at skalere den vertikale usikkerhed for den målte dybde med en faktor, der øges kvadratisk med afstanden. Det vil sige at dybden vægtes lavere og lavere des længere væk fra knudepunktet den målte dybde er placeret. Dette giver mening, idet sandsynligheden for en afvigelse i havbunden er større, des længere væk fra knudepunktet der er målt. Hvad angår den horisontale usikkerhed tages denne i betragtning, ved at antage at den målte dybde må være i en maksimal afstand fra knudepunktet. Denne afstand baseres på en fast brøkdel af den horisontale usikkerhed hørende til den målte dybde, som tillægges den reelle afstand. Hvis målingen har en stor usikkerhed vil det medvirke til, at skaleringsfaktor til at beregne den maksimale afstand øges, hvilket gør, at der rapporteres en større usikkerhed til knudepunktet, da punktet kommer længere væk.
Ved at håndtere usikkerhederne på denne måde, gives målte dybder, der ligger længere væk en højere usikkerhed og dermed lavere vægt, og den horisontale usikkerhed bidrager med en større afstand hvilket yderligere medvirker til, at usikkerheden øges hurtigt. Des større usikkerhed des lavere vægt.
På Figur 5.7 ses det, hvordan udbredelse af fejl forholder sig.



Figur 5.7. Udbredelse af usikkerheder

Estimer dybder og usikkerheder i knudepunkter på baggrund af målte dybder

Til at estimere dybder og usikkerheder i knudepunkterne på baggrund af de målte dybder anvendes et Kalman filter. Kalman filtret er en rekursiv algoritme, hvilket betyder, at det estimerer på baggrund af tidligere estimater, men anvender også nye observationer til at beregne estimater. Såfremt nye observationer er i uoverensstemmelse med estimatet i knudepunktet, vil det påvirke det nye estimat, men dette tager Kalman filtret til dels højde for gennem en vægt.
Når de målte dybder er blevet tilknyttet til et knudepunkt anvendes der først et andet filter, til at bytte om på rækkefølgen af målingerne. Det betyder, at uregelmæssige dybder kan forsinkes i processen, hvilket medvirker til at hele CUBE algoritmen bliver mere robust, da algoritmen på den måde til at begynde med beskyttes mod uregelmæssigheder. Anvendelsen af filtret ses på Figur 5.8, og her ses det, hvordan rækkefølgen af målingerne er byttet om, sådan at det uregelmæssige inputmålinger(3 og 4) er forsinket, så de først implementeres i Kalman filtret som senere observationer(8 og 10).

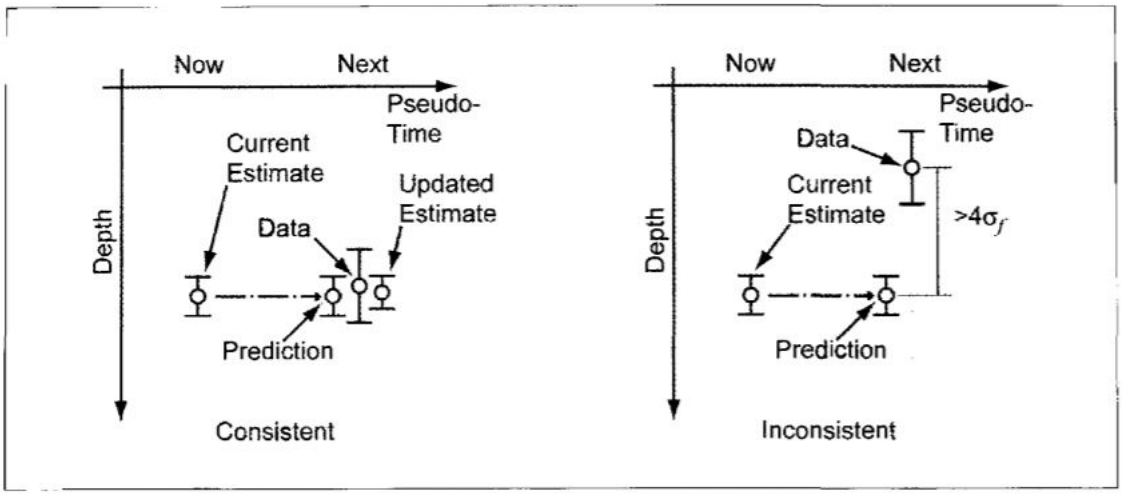


Figur 5.8. Ombytning rækkefølge på målte dybder,[Calder, 2003]

Efter at input data er sorteret, kan Kalman filtret anvendes. I Kalman filtret anvendes der en vægt kaldet Kalman Gain, som angiver, hvordan en observation skal vægtes i forhold til et nuværende estimat. Det vil sige, at når en dybde inddrages i Kalman filtret vægtes denne måling på baggrund af usikkerheden knyttet til målingen i forhold til usikkerheden for det nuværende estimat. Har den målte dybde en noget højere usikkerhed tilknyttet til sig end usikkerheden for det nuværende estimat vil den målte dybde ikke tillæges så meget vægt og omvendt. Det er her filtret tager højde for punkter, der er placeret længere væk, ved at de har fået tillagt en større vægt tidligere og derfor bliver tillagt mindre betydning i anvendelsen af Kalman filtret.
Når der er foretaget et nyt estimat for dybden i knudepunktet kan de rå målinger reelt set kasseres, men CUBE algoritmen er implementeret sådan, at målingerne dog beholdes i en database til eventuel yderligere analyse.

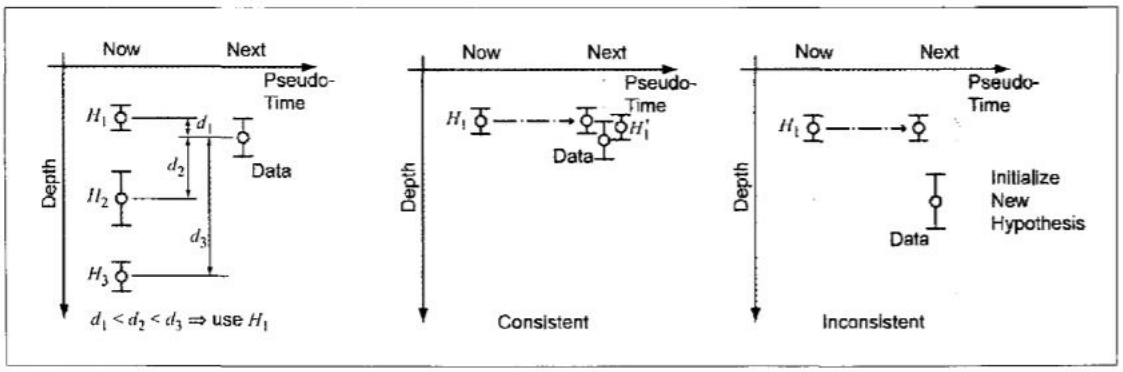
Vurder om der skal være mere end en hypotese

CUBE er opbygget sådan, at den forventer én konstant dybde i et knudepunkt, hvilket dog ikke overholdet i praksis, da der eksempelvis observeres outliers, som kan indikere, at der er flere dybder i en given lokalitet.For at håndtere outliers arbejder CUBE med hypoteser, hvor nulhypotesen er, at der er konsistens i data, hvilket vil sige, at en målt dybde er i overensstemmelse med det nuværende estimat for dybden i knudepunktet. Med overensstemmelse menes der, at den målte dybde ligger indenfor en specificeret grænse bestemt af den målte dybdens forudsagte fejl. Såfremt der ikke er konsistens i data kan det konkluderes, at der er tilstrækkelig indikation for at en dybde er inkonsistent. Dette er illustreret på Figur 5.9.



Figur 5.9. Model over hvornår måling og estimat enten er i overensstemmelse med hinanden eller ej

Når det er konstateret, at der er tale om inkonsistens, er der tale om at der ikke er overensstemmelse mellem dybderne og det er derfor ikke repræsentativt at inddrage dem i den modellen for havbunden. Derfor antages det, at der er tale om et andet potentielt estimat for dybden i knudepunktet, og der oprettes en hypotese om knudepunktets sande dybde. Løbende kan der konstateres endnu flere potentielle estimater for dybderne i knudepunkterne, og derved kan der skabes flere hypoteser for et knudepunkt. Alle disse hypoteser behandles på samme måde, og der skal ikke træffes en afgørelse om, hvilken der er den "bedste" hypotese som indeholder den sande dybde før til sidst. Der kan dog opstå nogle teoretiske udfordringer når det skal vælges, hvilken hypotese indkommende data skal knyttes til. Dette løser CUBE ved vurdere om målingen i en statistisk signifikant grad er tæt på et af estimaterne, og er den det, vil den indgår i beregningen for næste estimat. Stemmer målingen ikke i en statistisk signifikant grad med nogle af de aktuelle estimater oprettes der så en ny hypotese, se Figur 5.10.



Figur 5.10. Valg af hypotese

På Figur 5.10(midt) ses det, hvordan der ikke oprettes en ny hypotese såfremt data er konsistent, hvorimod der initialiseres en ny hypotese, når data er inkonsistent Figur 5.10(højre).

Vælg den "bedste" hypotese

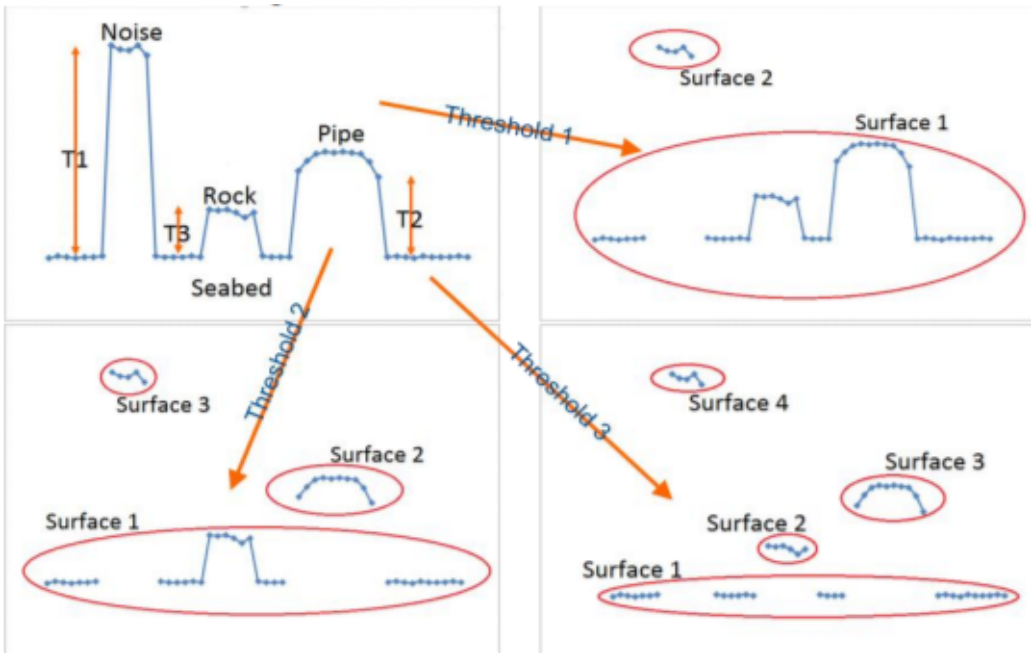
Når der konstateret inkonsistens i data, dannes der flere hypoteser, hvilket kan medvirke til en robusthed i algoritmen, men det giver også en række udfordringer, da det skal vurderes efterfølgende, hvilken hypotese der er den bedste og indeholder den sande dybe. Der er flere måder at vælge den "bedste" hypotese på.
Den *første* metode til at vælge en hypotese på er at vælge den hypotese, der har integreret flest data punkter. Dette er med til at understøtte dybden, men metoden kan dog ses som ikke anvendelig i de tilfælde, hvor der er et betydeligt støjindhold i data. Dette kan eksempelvis være hvis der er en større mængde af data samlet på et sted, der ikke er havbunden. Eksempelvis hvis en hval har svømmet under ekkoloddet, mens der er foretaget målinger.
En *anden* metode er, at tage en mængde af nabo knudepunkter, der kun har én hypotese tilknyttet til sig. Disse knudepunkter må derved anses som havende en sand dybde tilknyttet, og kan derfor ses som vejledende til at finde den mest sandsynlige dybde i knudepunktet med flere hypoteser tilknyttet. I knudepunktet vælges så den hypotese, hvor dybden er tættest på de vejledende knudepunkters dybde.
En *tredje* metode er at konstruere kontekst ved at anvende en anden overflade. Dette kan potentielt være en overflade med en lavere opløsning fra en tidligere opmåling eller det kan være en overflade dannet af rå data fra samme opmåling som nuværende data behandles fra. Overfladen skal kun ses som vejledende, og det er ikke en nødvendighed, at overfladen er hydrografisk korrekt. Hvis bare overfladen er på nogenlunde den samme lokation, kan den være CUBE behjælpelig med at finde ud af, hvilken hypotese der skal anvendes, og dermed hvilken dybde der skal betragtes som den sande.
Med henblik på dybden, er CUBE i stand til at opstille en række informationer om dybder, usikkerheder for dybdeestimaterne, antal hypoteser i hvert knudepunkt samt hvor sikker algoritmen er på den valgte hypotese. Hver af disse informationer kan præsenteres som et kort, hvorfra det kan indikeres for brugeren om der er områder, som kræver brugerinddragelse. Eksempler på overflader kan ses i Bilag B.
CUBE er altså en algoritme, der danner en overflade på baggrund af en række estimerede dybder samt tilhørende usikkerheder. Algoritmen er automatisk og brugerinddragelsen er lav, hvilket gør den ideel til at behandle store mængder af data.

5.2 S-CAN

S-CAN er et plug-in udviklet til EIVA's NaviModel i tæt samarbejde med MADALGO(Center for Massive Data Algorithms) fra Århus universitet, og har til formål at frasortere støj fra data indsamlet med multibeam. S-CAN står for SCALGO Combinatorial Anti Noise. S-CAN er på nuværende tidspunkt ved at blive erstattet af EC-3D, men vil dog stadig være tilgængelig og anvendelig for brugeren, dog uden at blive udviklet eller vedligeholdt yderligere fremover. Afsnittet baseres på [Dall, 2011].
Ideen bag S-CAN er at beregne en såkaldt "noise score" for hvert enkelt datapunkt i datasættet, sådan at brugeren let og interaktivt kan rense hele eller dele af datasættet ved at fjerne punkter, som har en høj "noise score". Scoreværdien for hver enkelt måling indekseres i et indledende trin. Det efterfølgende trin med at udvælge områder med forskellige grænseværdier er udviklet for at effektivisere processen, og her er brugerinddragelse nødvendig. S-CAN består af to værktøjer, som begge forsøger at filtrere støj væk fra punktskyen - S-CAN Component og S-CAN Score.

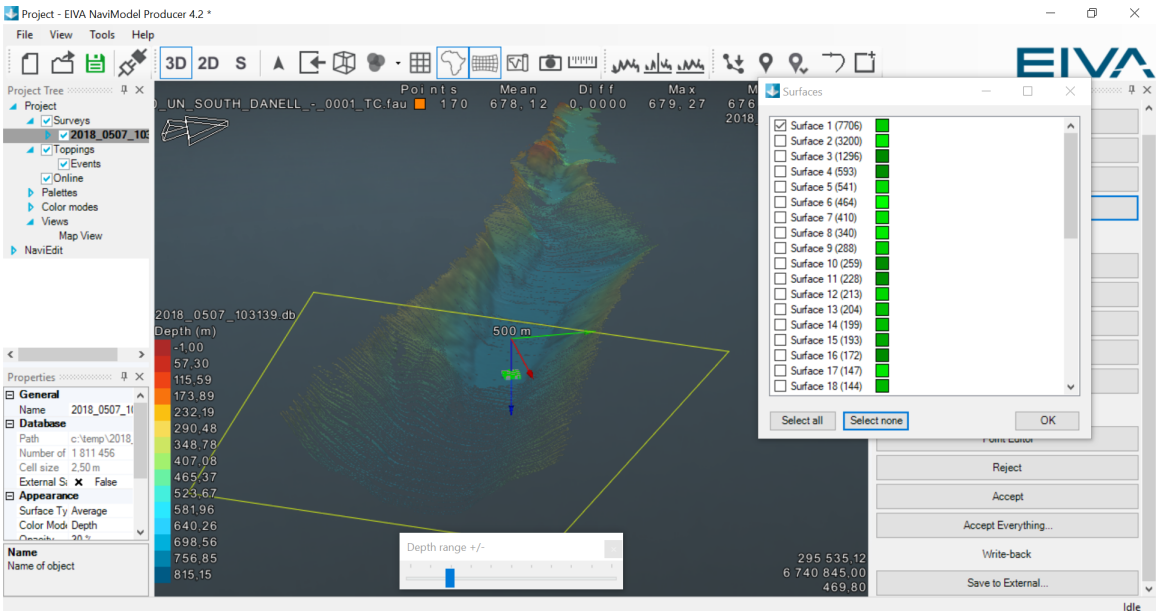
S-CAN Component

S-CAN component inddeler inputobservationerne i flere delmængder af observationer, der opfylder et krav om en maksimal tålelig grænse (grænseværdien) mellem nabopunkter. Disse mængder af observationer betegnes i det følgende som overflader. Ved at danne disse overflader med en valgt grænseværdi, vil minimumshøjden mellem hver overflader, og derved naboerne, være større end grænseværdien. Såfremt en høj grænseværdi vælges, vil der dannes få overflader, og omvendt vil der dannes flere overflader, hvis en lavere grænseværdi vælges. Dette er illustreret på Figur 5.11.



Figur 5.11. Princippet i S-CAN Component, [Dall, 2011]

I eksemplet på Figur 5.11 ses det, at når en stor grænseværdi vælges, T1, inddeles der i to overflader. Vælges der derimod en lavere grænseværdi, ses det, at der dannes flere overflader.
Ved at sætte grænseværdien til T1, ses det, at støjen er beliggende i sin egen overflade, og nemt kan sorteres fra. Sættes grænseværdien lavere til T2, ses det, at røret får sin egen overflade og hvis grænseværdien sættes til T3, dannes der fire overflader, hvor stenen også får sin egen overflade.

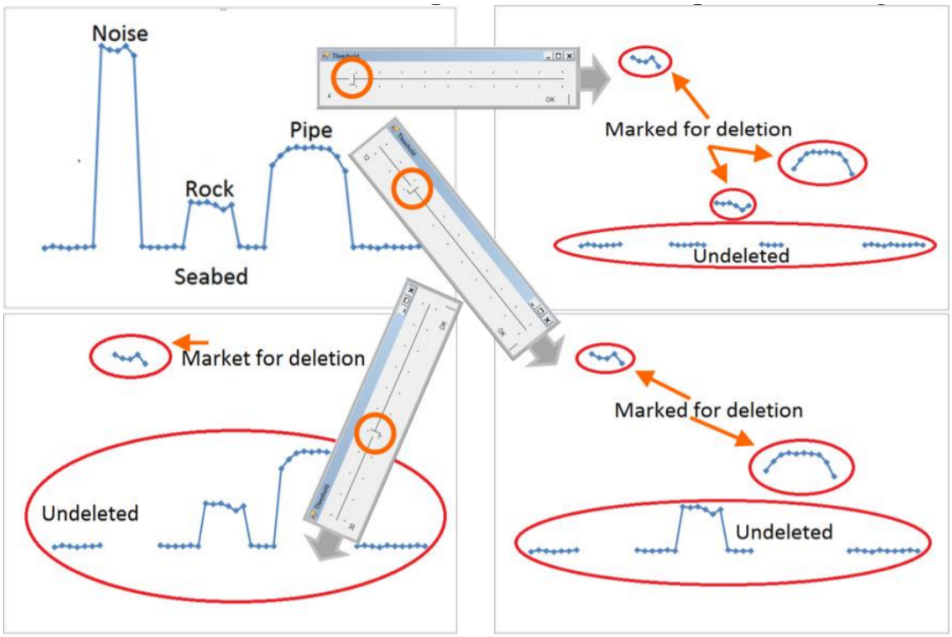


Figur 5.12. S-CAN Component i NaviModel

Når overfladerne er dannet på baggrund af en bruger valgt grænseværdi, opstilles en liste over de forskellige overflader, se Figur 5.12. Denne liste beskriver blandt andet, hvor mange punkter der er i hver overflade. Ved at opstille listen kan de forskellige overflader slås til og fra, hvorefter der igen er brug for at brugeren inddragelse. Brugeren vælger således hvilke overflader, der skal bevares, og hvilke der skal noteres som støj og slettes. Såfremt den valgte grænseværdi ikke passer til at efterbehandle det pågældende data gentages processen med en anden grænseværdi.

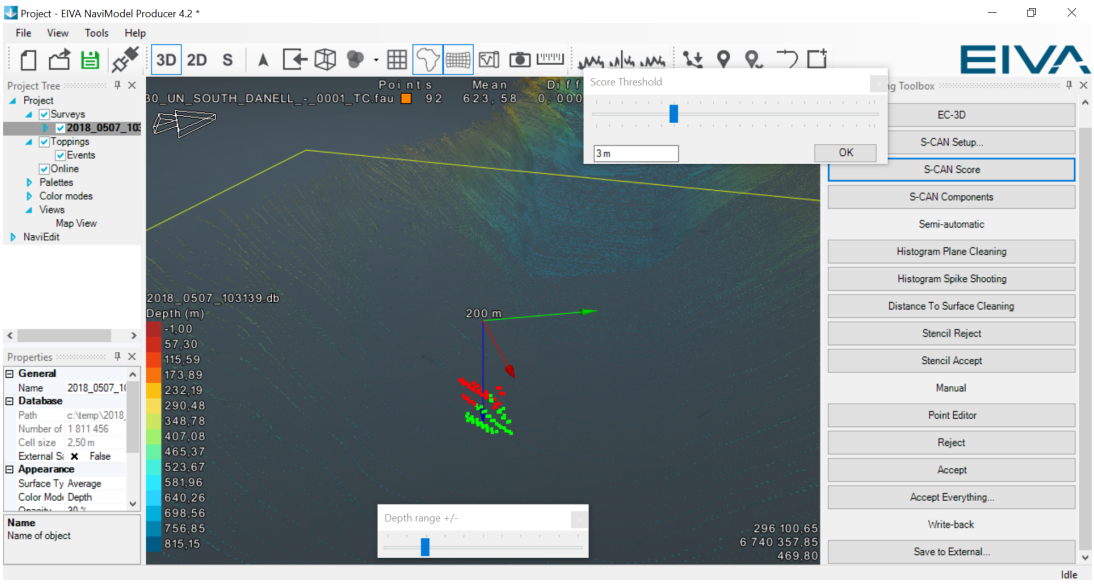
S-CAN Score

Hvis dette værktøj sammenlignes med S-CAN component, kan det bedst beskrives som et værktøj, der beregner for alle grænseværdier til at starte med. På den måde, skal brugeren ikke selv angive en ønsket grænseværdi, men blot justere på en bar for at finde en passende grænseværdi, se Figur 5.13.



Figur 5.13. Princippet i S-CAN Score, [Dall, 2011]

Det at der kan justeres på baren for at undersøge grænseværdier, se Figur 5.14, gør, at brugeren kan vurdere den passende grænseværdi for et givent område, men måske også erfare, at én grænseværdi måske ikke passer til hele området. Derfor kan det ske, at der skal foretages rensning af data, med forskellige grænseværdier, hvilket så efterfølgende kan gøres med S-CAN component.



Figur 5.14. S-CAN Score i NaviModel

Score varianten af S-CAN er generelt hurtigere end component varianten. Score varianten ses, som værende bedst i situationer, hvor der kun er en overflade, hvorimod component varianten ses som værende mest hensigtsmæssig at anvende, når der er større variationer i havbunden. Oftest ses en kombination af de to varianter som den bedste løsning, hvor score varianten først andevendes,

da den er mest effektiv og derefter anvendes component varianten til de mere komplekse områder. Score varianten kan dog svare til component varianten, såfremt score varianten anvendes på udvalgte områder. Dette kan være hurtigere, idet score varianten giver brugeren mulighed for at finde en passende grænseværdi.

S-CAN består altså at to stykker værktøj, som begge hjælper til at rense data. Det grundlæggende bag de to typer værktøj er det samme, ved at de beregne en mængde overflader baseret på en grænseværdi. Forskellen er kun at der i component varianten vælges en grænseværdi af brugeren selv, mens der ved score varianten justeres på grænseværdien gennem en bar. Det kan både være en fordel og en ulempe at anvende component varianten. Denne variant giver brugeren mulighed for at tilvælge overflader og på den måde finsortere lidt mere i data, men jo flere overflader der dannes, des mere tid skal der bruges fra brugeren, og det er netop tiden og brugerinddragelsen, der gerne skulle nedbringes ved at automatisere databehandlingsprocessen.

5.3 EC-3D

Ligesom S-CAN er EC-3D et værktøj i EIVA's NaviModel. EC-3D står for EIVA Cleaning 3D og er et værktøj, der indeholder forskellige rensningsmetoder til at efterbehandle indsamlet data for havbunden og laser data indsamlet over vand i 3D. EC-3D erstatter S-CAN i den nyeste version af NaviModel. S-CAN er stadig tilgængelig i den nyeste version af NaviModel, men værktøjet opdateres og vedligeholdes ikke længere. Afsnittet baseres på [EIVA, 2017] og [EIVA, 2018].
Den grundlæggende ide med EC-3D er, at det for hvert punkt i punktskyen, vurderes hvorvidt punktet er støj eller valid data. Til at bestemme dette findes der flere forskellige metoder, som er:

- Nearest Neighbour Cleaning
- Neighbour In Sphere Cleaning
- Z Summation In Circle Cleaning
- Distance To Average Surface Cleaning
- K Nearest Neigbour Cleaning

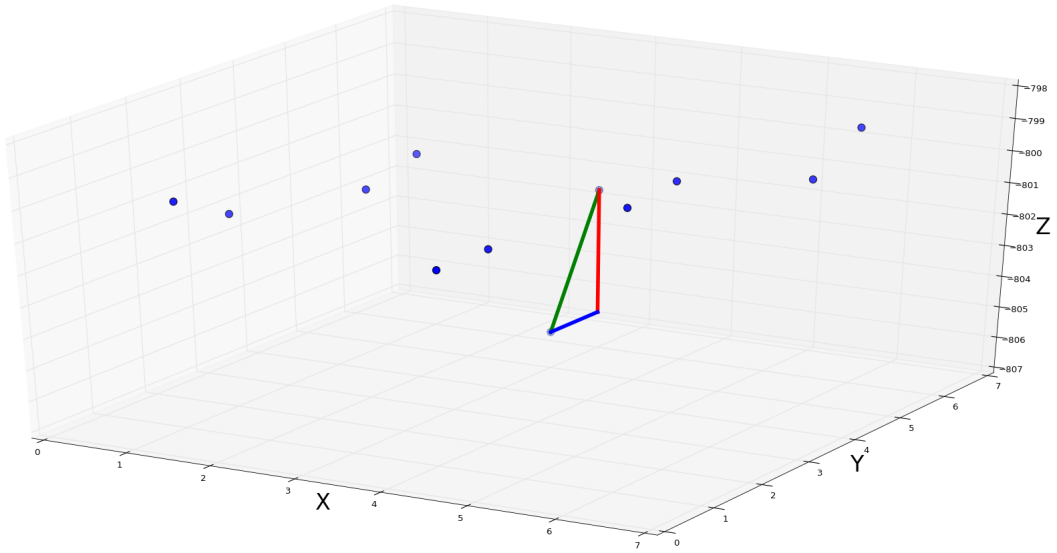
De forskellige metoder er anvendelige i forskellige situationer, og nogle metoder passer måske bedre til én situation end en anden. Alle metoderne kan dog anvendes i sammenhæng med de andre metoder, og på den måde kan en kombination af metoderne anvendes for muligvis at opnå et bedre resultat. For at få en fornemmelse for hvornår og hvordan de forskellige metoder kan anvendes, er det nødvendigt at undersøge, hvordan de virker. De følgende afsnit er baseret på manualen for NaviModel 4.2, [EIVA, 2018].
Inden de næste afsnit læses, er der vigtigt at understrege for læseren, at koordinater i illustrationerne er fiktive, og kun lavet for at give læseren en ide om, hvordan metoderne fungerer. Derudover anvendes begrebet nabo, hvilket kan defineres på forskellige måder. EIVA definerer en nabo til et punkt P, som det punkt K, der er tættest på P. Dette betyder at afstanden til punkt K fra punkt P er den mindste afstand. Hvis der er 5 naboer, er det så de fem nærmeste punkter til punkt P, der er naboer. Med nærmeste menes der minimumsafstanden i 3D. I gennemgangen af de forskellige metoder tilknyttes der eksempler, og læseren skal gøres opmærksom på, at det er det samme datasæt der er anvendt i hvert eksempel. Datasættet er fra Grønland og der skal yderligere gøres opmærksom på at havbunden er stærkt kuperet. Inden for 30 m kan dybden variere med 50 m.

Nearest Neighbour Cleaning

Ideen med metoden er, at der beregnes en afstand til den nærmeste nabo, og herfra vælger brugeren en øvre og en nedre grænseværdi for hvor stor denne afstand mellem punkt og naboen må være. Afstanden er i 3D, og beregnes som:

$$d = \sqrt{(x_{p_i} - x_{p_n})^2 + (y_{p_i} - y_{p_n})^2 + (z_{p_i} - z_{p_n})^2}$$

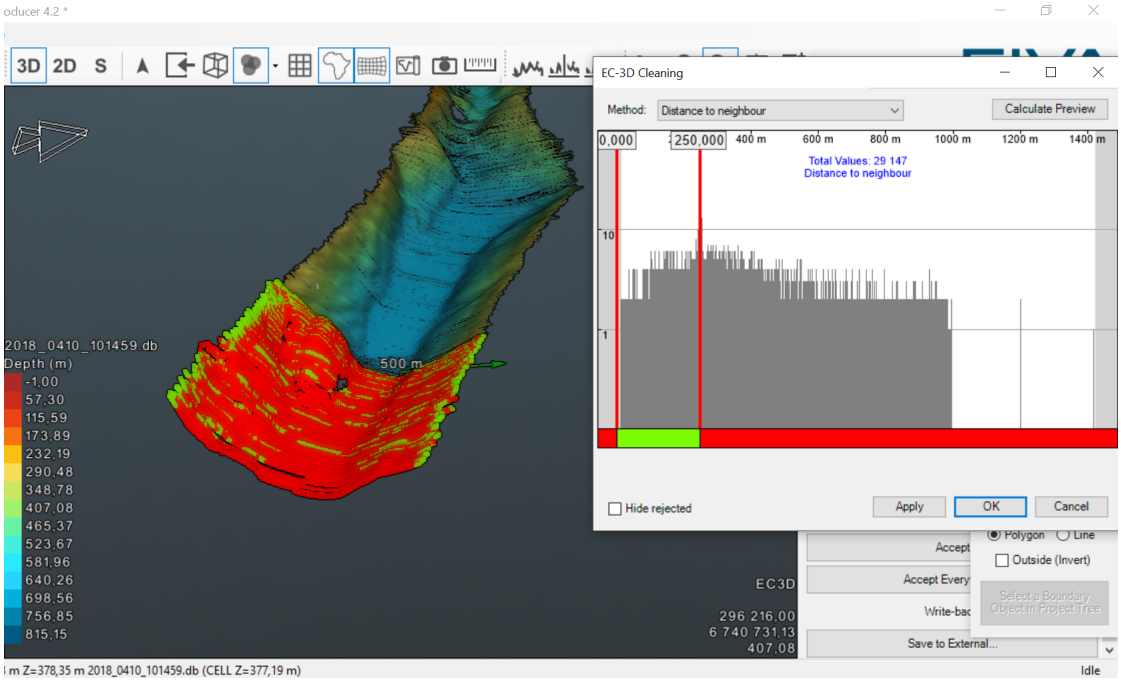
Hvor p_i er et vilkårligt punkt og p_n er punkt p_i's nærmeste nabo. På Figur 5.15 illustreres afstanden mellem to punkter i 1D(Rød), 2D(Blå) og 3D(Grøn). I denne metode er der tale om afstanden mellem punkter i 3D, hvorfor det her er afstanden markeret med grøn på Figur 5.15, som anvendes.



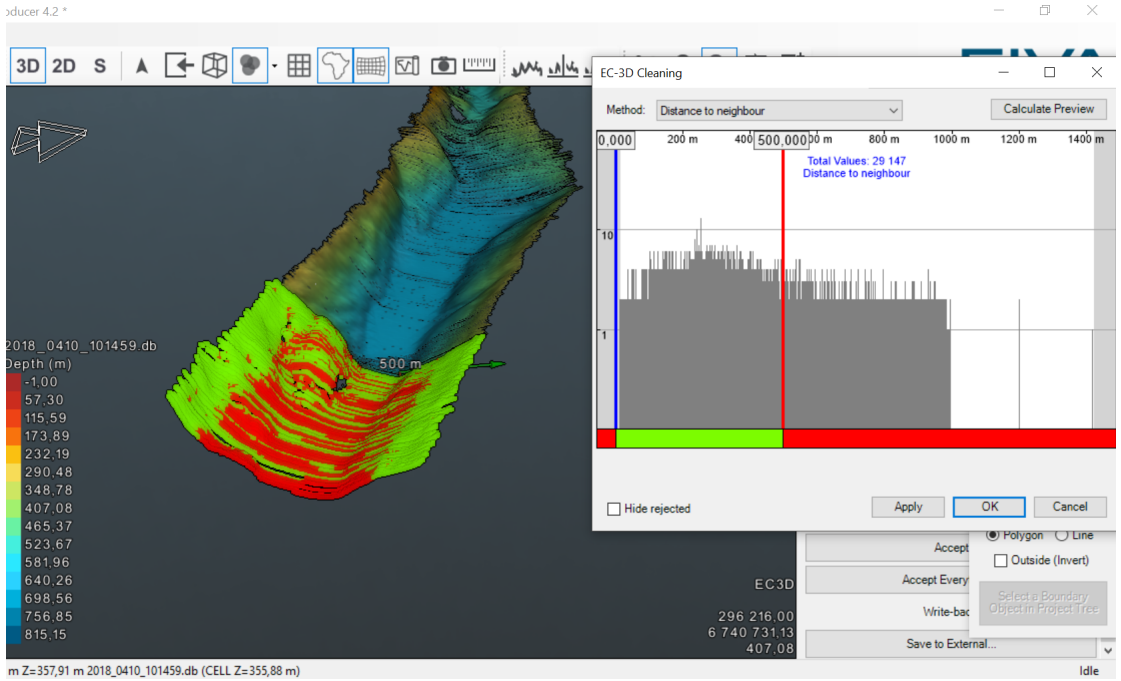
Figur 5.15. Illustration af afstande i 1D(rød), 2D(blå) og 3D(grøn)

På Figur 5.16 og Figur 5.17 angiver de grønne punkter de målinger, der er indenfor de to grænseværdier, mens de røde punkter angiver de målinger, der ikke er inden for de angivne grænseværdier. Det ses af de to figurer, Figur 5.16 og Figur 5.17, at holdes den nedre grænse på 0 og øges den øvre grænse, så vil antallet af grønne punkter også øges, og datasættet vil blive præget af grønne punkter frem for røde punkter. Yderligere ses det også, at NaviModel genererer et histogram, som viser, hvor mange punkter der er indenfor de givne afstande. På baggrund af dette histogram kan den øvre og den nedre grænseværdi ændres, sådan at det passer bedst muligt til det anvendte datasæt.

Metoden kan anvendes til at finde punkter, der overstiger en bestemt afstand til dets næreste nabo. Metoden kan derfor ses mest anvendelig til at finde målinger, der ligger alene i datamængden.



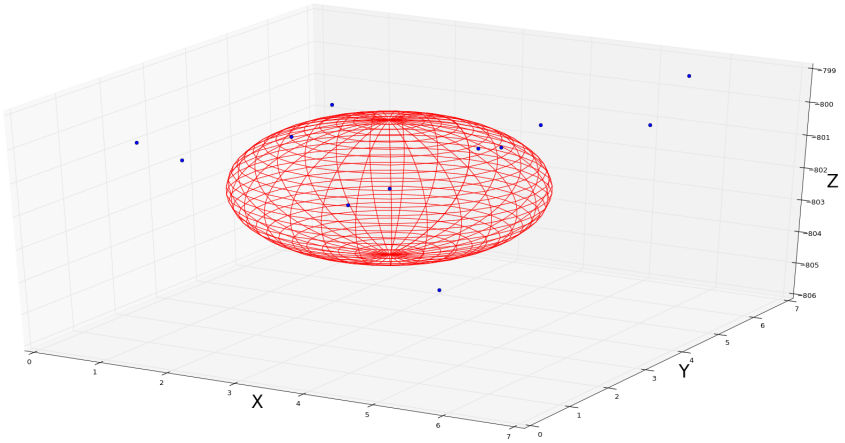
Figur 5.16. Nearest Neighbour Cleaning, hvor $T_{min} = 0m$ og $T_{max} = 250m$



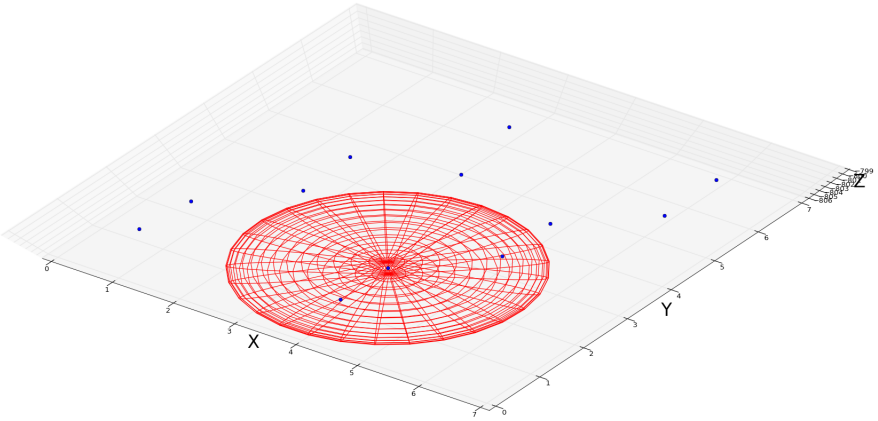
Figur 5.17. Nearest Neighbour Cleaning, hvor $T_{min} = 0m$ og $T_{max} = 500m$

Neighbour In Sphere Cleaning

Denne metode tæller antallet af naboer indenfor en sfære fra et punkt. For at metoden virker, skal den radius, der vælges til sfæren, være stor nok til at indeholde den nærmeste nabo. Yderligere skal der til metoden angives to grænseværdier, hvilke beskriver den øvre og den nedre grænse for antallet af tilladte naboer. Det er mest normalt kun at angive og justere den nedre grænse, da der normalvis stilles et minimumskrav til antal af naboer, og ved ikke at angive den maksimale grænse, sættes denne automatisk det det størst antal naboer i sfæren.
På Figur 5.18 og Figur 5.19, er metoden illustreret.



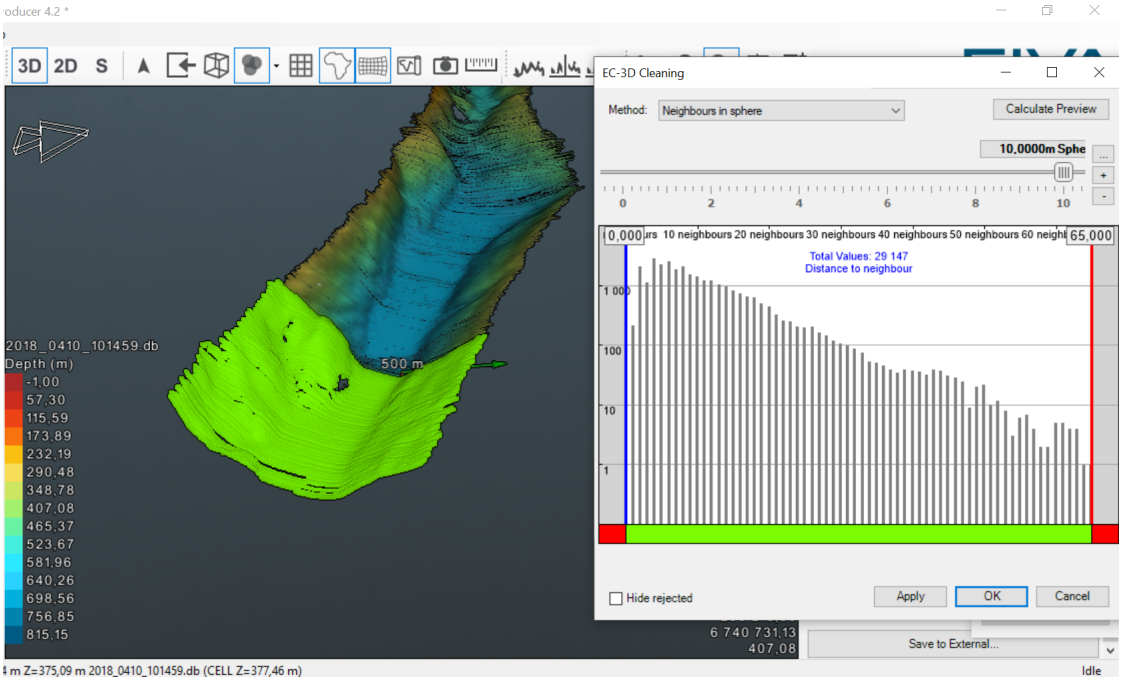
Figur 5.18. Illustration af Neighbour In Sphere Cleaning



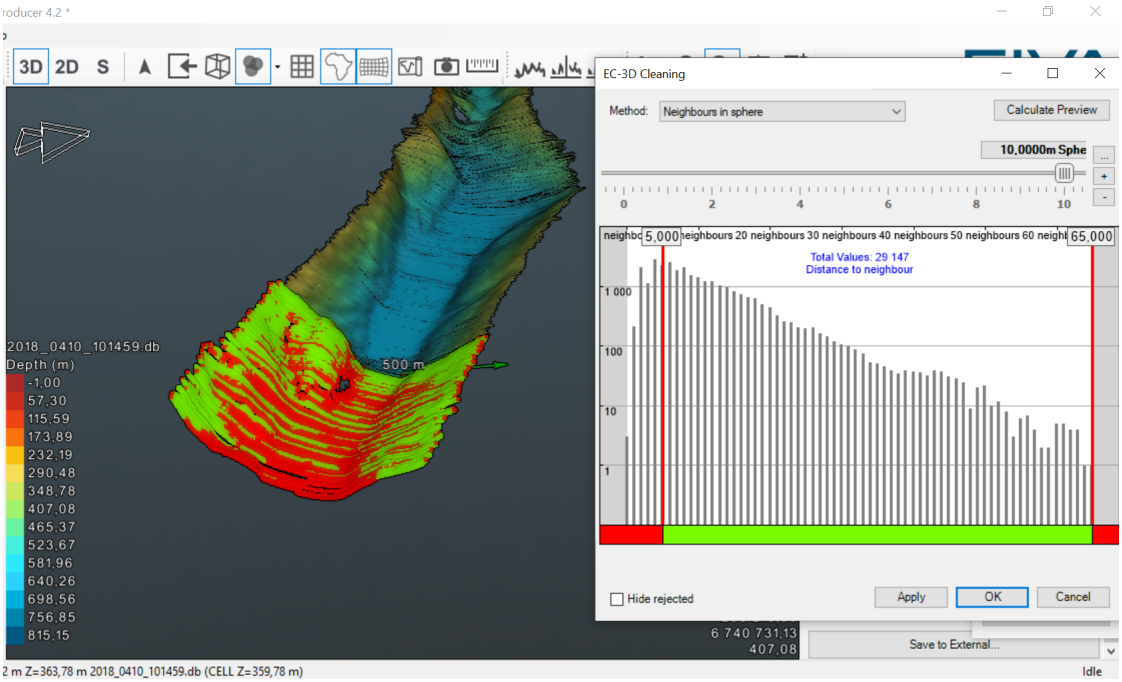
Figur 5.19. Illustration af Neighbour In Sphere Cleaning

På Figur 5.18 og Figur 5.19 ses det, at vælges der, til dette eksempel, en sfære med en radius på 2 fra det undersøgte punkt, og at der er to punkter indenfor sfæren. Såfremt de to punkter er naboer til punktet i centrum af sfæren, er antallet af naboer for dette punkt således 2.
Metoden ses anvendelig til at finde punker, som ikke har så mange naboer, eller punkter som har mange naboer.

I eksemplet på Figur 5.20 og Figur 5.21 ses det, at vælges der en radius på 10 m, reduceres antallet af grønne punkter når minimums antallet af naboer sættes op. På Figur 5.20 er den nedre grænse for antal af naboer sat til at være 0, hvorved alle punkter er med, da den øvre grænse tilpasses datasættet, da der ikke er sat en øvre grænseværdi, og det maksimale antal af naboer derfor automatisk vælges. Sættes den nedre grænse til 5 naboer, ses det, hvordan en stor del at målingerne bliver markeret røde, og derfor ikke er indenfor grænseværdierne, se Figur 5.21.



Figur 5.20. Neighbourg In Sphere Cleaning, $r = 10m$, $T_{min} = 0$ og $T_{max} = 65$



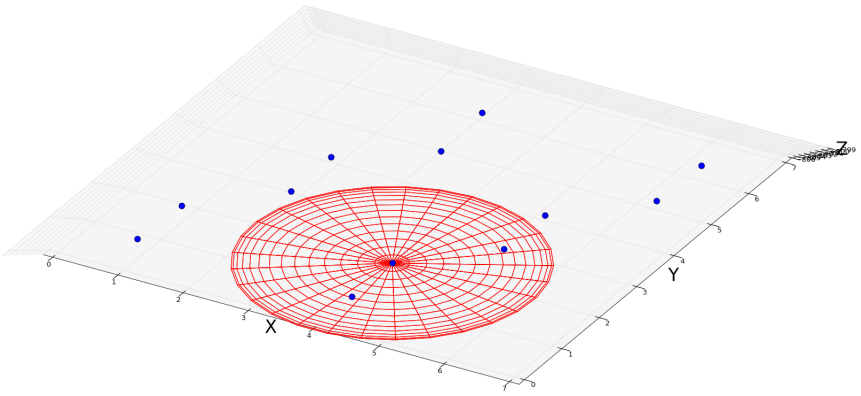
Figur 5.21. Neighbourg In Sphere Cleaning, $r = 10m$, $T_{min} = 5$ og $T_{max} = 65$

Z Summation In Circle Cleaning

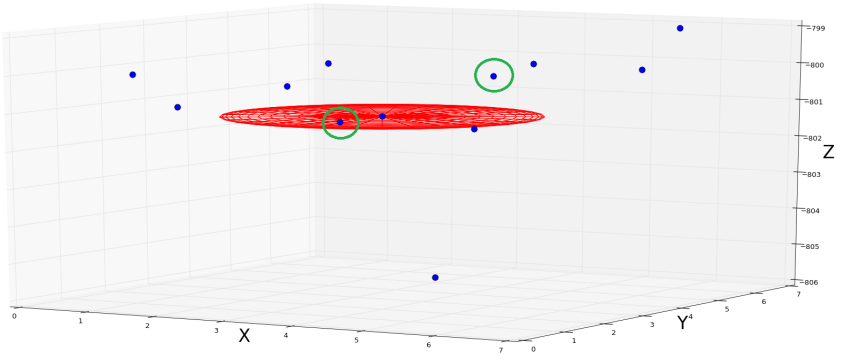
I denne metode vælger brugeren en radius til en cirkel i det horisontale plan, som er stor nok til at indeholde den nærmeste nabo, men heller ikke indeholder for mange punkter. Derefter beregnes z-differencen til alle naboer indenfor denne cirkel. Z-differencen er dybdeforskellen og er altså den afstand, der er mellem to dybder på z-aksen. Udover en radius til cirklen, kan to grænseværdier angives af brugeren i et histogram. Den ene grænse angiver den mindste summerede værdi og den anden grænse angiver den højeste summerede værdi. Jo større værdi, des større summerede afstand mellem dybderne.
På Figur 5.22 og Figur 5.23 er metoden illustreret. På Figur 5.22 ses punkterne og den dannede cirkel med en radius på 2, i en tilnærmelsesvis horisontal plan. Dette medvirker til at illustrere hvilke punkter, der er indenfor den angivne radius. På Figur 5.23 ses punkterne mere i det vertikale plan, hvor de forskellige dybder kan ses, frem for positionerne. De punkter, der er markeret med en grøn cirkel, er de punkter, der er beliggende indenfor cirklen i det horisontale plan. Nu beregner metoden så den summerede forskel i dybderne ud fra:

$$\Delta z_{sum} = \sum_{i=1}^{n} (z_k - z_i)$$

Hvor z_k angiver dybden til det punkt, hvorom cirklen er konstrueret og z_i angiver dybden for de n naboer, der er beliggende inden for den angivne radius.

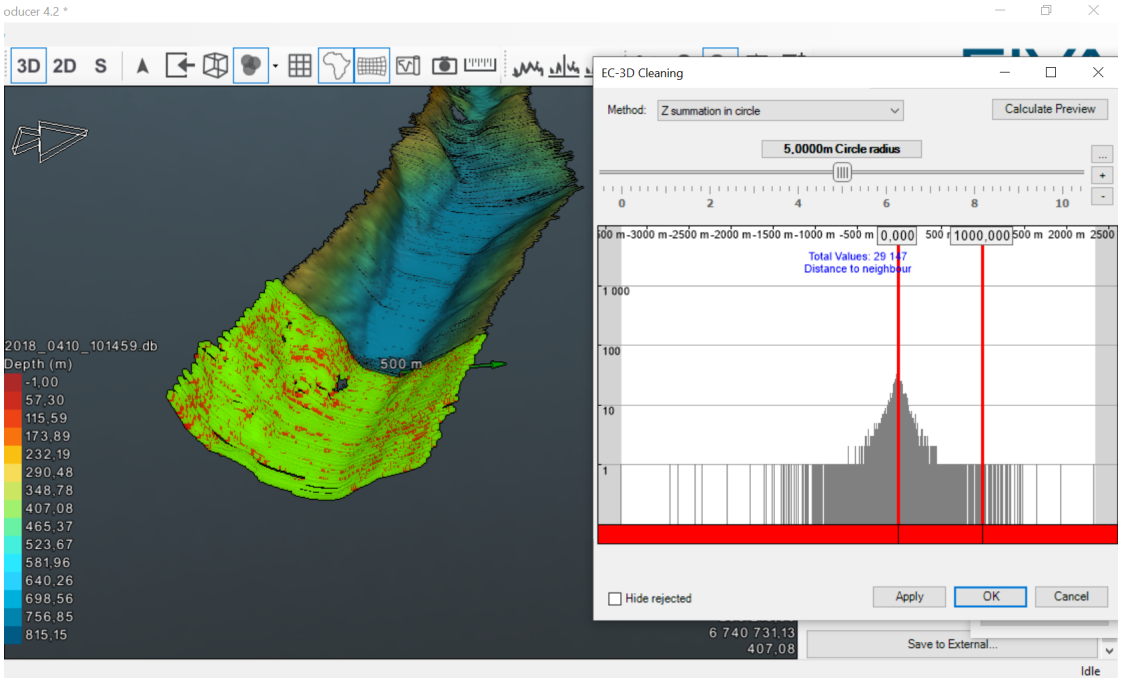


Figur 5.22. Illustration af Z Summation In Circle Cleaning

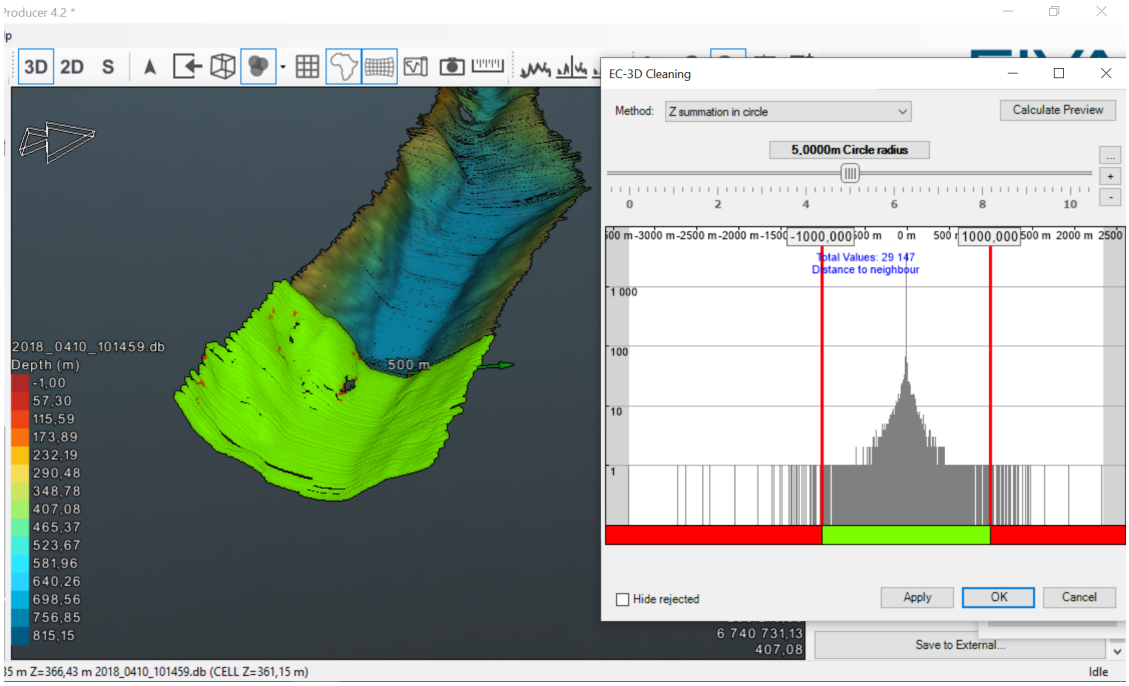


Figur 5.23. Illustration af Z Summation In Circle Cleaning

På Figur 5.24 er der fastholdt en radius på 5 m, en nedre grænse på 0 m og en øvre grænse ændres på 1000 m. Af histogrammet ses det, at der er en centrering omkring 0m. Det mest optimale er ikke at sætte den ene grænseværdi til 0 m, da grænseværdierne angiver afstande, hvilke kan være både positive og negative. Derfor vil det mest optimale være at angive en grænseværdi under 0 m samt en over 0 m. Ved at sætte en nedre grænseværdi til -1000 m og en øvre grænseværdi til 1000 m vil datasættet se ud som på Figur 5.25.



Figur 5.24. Z summation in circle, $r = 5m$, $T_{min} = 0$ og $T_{max} = 1000$

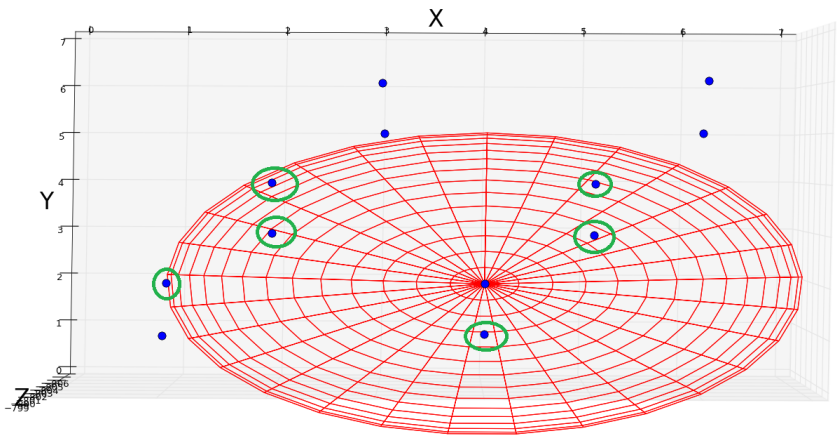


Figur 5.25. Z summation in circle, $r = 5m$, $T_{min} = -1000$ og $T_{max} = 1000$

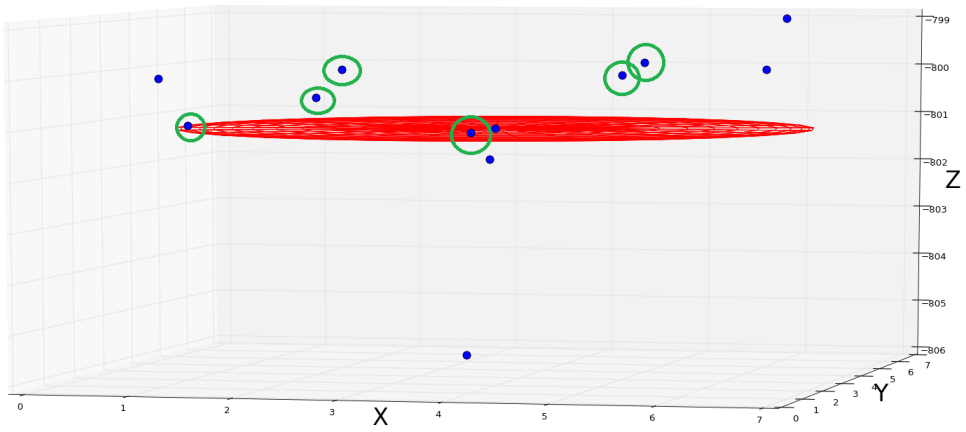
Distance To Average Surface Cleaning

Denne metode beregner en dybdeforskel fra et punkt til en gennemsnitlig dybde, hvor den gennemsnitlige dybde er beregnet ud fra dybder i en specificeret radius fra et punkt. Her angiver brugeren to grænseværdier, som i dette tilfælde angiver den længste og den korteste afstand fra den gennemsnitlige dybde.

For at finde den gennemsnitlige dybde, skal det først undersøges, hvilke punkter der er beliggende inden for en given radius i det horisontale plan. Dette sker ved at der konstrueres en cirkel i den horisontale plan, med en brugerdefineret radius, hvilket er illustreret på Figur 5.26.



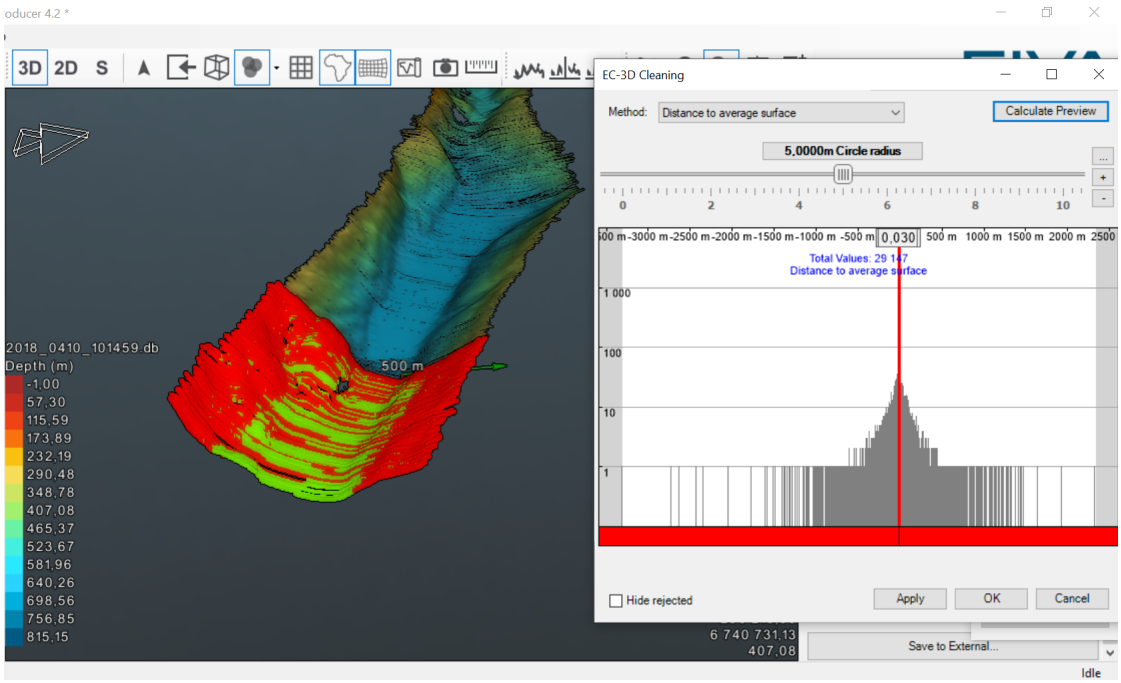
Figur 5.26. Illustration af Distance To Average Surface Cleaning



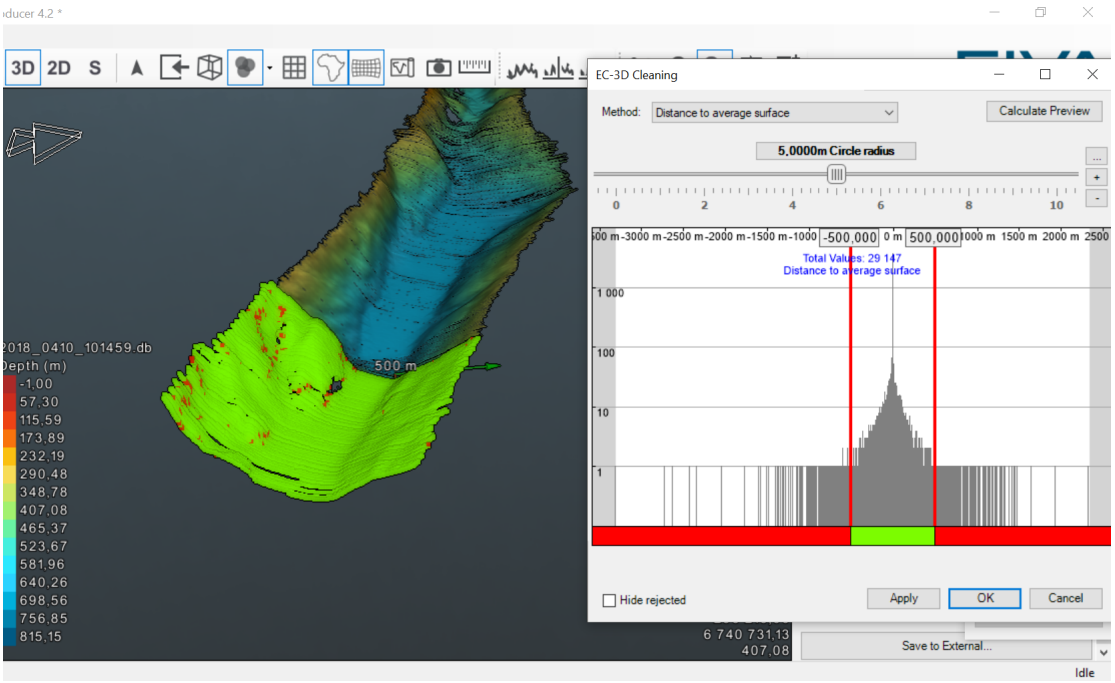
Figur 5.27. Illustration af Distance To Average Surface Cleaning

På Figur 5.26 og Figur 5.27 er der markeret en række punkter med en grøn ring, og det er disse punkter, der er beliggende indenfor en radius på 3. På baggrund af disse punkters dybder, beregnes en gennemsnitsdybde. Herfra går metoden så ud på at beregne dybdeforskellen fra punktet, hvorom cirklen er genereret, til denne gennemsnitsdybde. Metoden kan være en god at anvende til at finde dele af havbunden, der tilnærmelsesvis er vandret, da metoden netop fokuserer på forskel til en gennemsnitsdybde. Når grænseværdierne er lave, vil punkter tæt på denne gennemsnitsdybde blive markeret som grøn.

På Figur 5.28 og Figur 5.29 ses det, hvordan grænseværdierne ændrer antallet af grønne punkter. Det ses, ligesom tidligere når der beregnes afstande, at der vælges både en positiv og en negativ værdi som grænseværdi. Når grænseværdierne er tæt på hinanden, ses det, at den vandrette havbund anses som værende valid, hvilket kan være fordi, at havbunden her er vandret, og derved at de fleste punkter vil ligge tæt på en gennemsnits dybde. Når der kommer mere variation i havbunden, er det sværere at generere en gennemsnitlig vandret overflade, der passer til punkterne med varierende dybde.



Figur 5.28. Distance To Average Surface Cleaning, $r = 5m$, $T_{min} = -0.03m$ og $T_{max} = 0.03m$



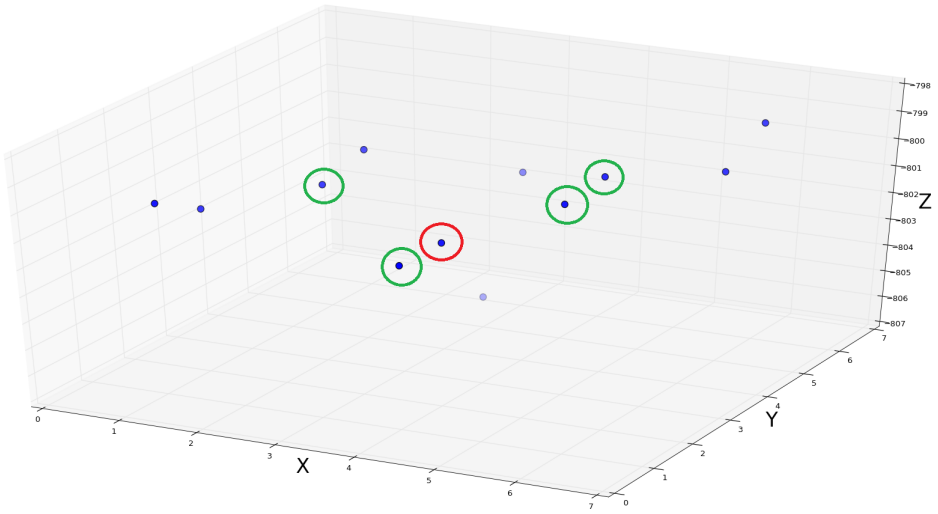
Figur 5.29. Distance To Average Surface Cleaning, $r = 5m$, $T_{min} = -500m$ og $T_{max} = 500m$

K Nearest Neigbour Cleaning

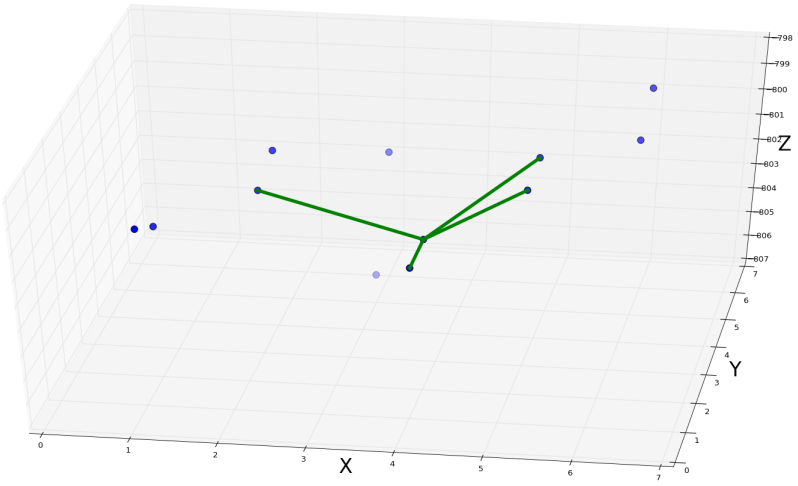
Til denne metode samles de k nærmeste naboer for hvert punkt, og den gennemsnitlige afstand til de k punkter beregnes ud fra:

$$T = \frac{\sum_{i=1}^{k} d_i}{k}$$

Hvor d_i er afstanden mellem et punkt og dets nabo. Dette summeres så op for alle k naboer, hvorefter der divideres med antallet af naboer(k) for at få den gennemsnitlige afstand. Den øvre og den nedre grænseværdi for den gennemsnitlige afstand kan angives af brugeren, og normalt er det den maksimale afstand, der ændres på. På Figur 5.30 ses en række punkter. Her indikerer det punkt markeret med rød cirkel, det punkt der skal undersøges, hvor de grønne punkter angiver de 4 nærmeste naboer til punktet.



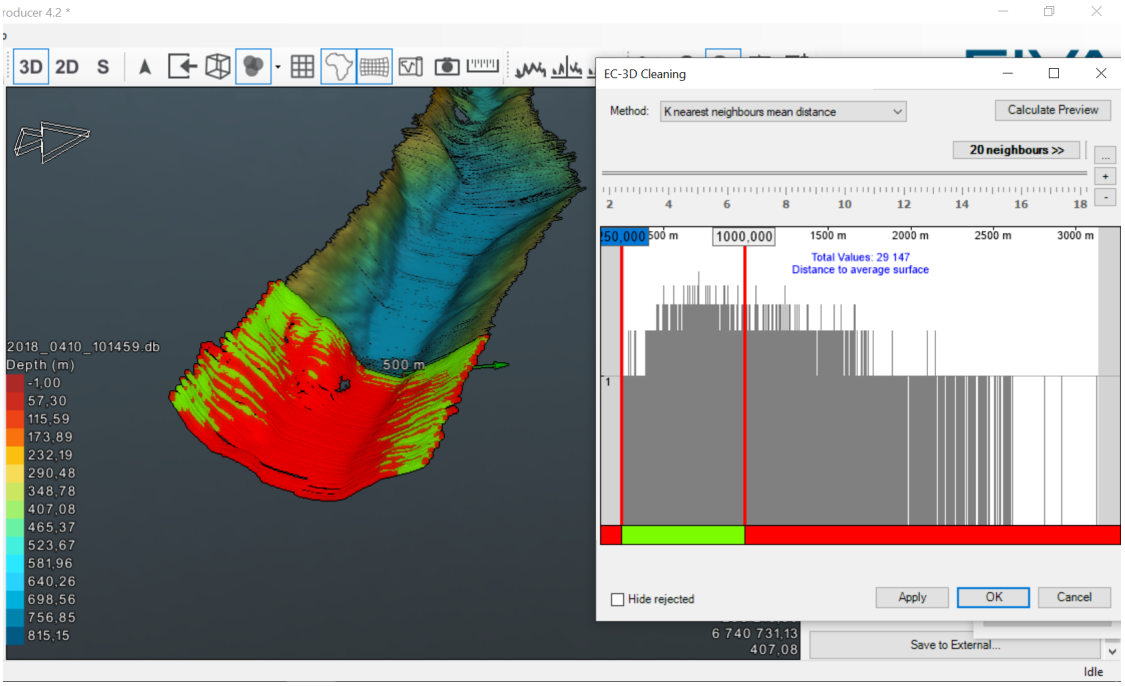
Figur 5.30. Illustration af K Nearest Neigbour Cleaning



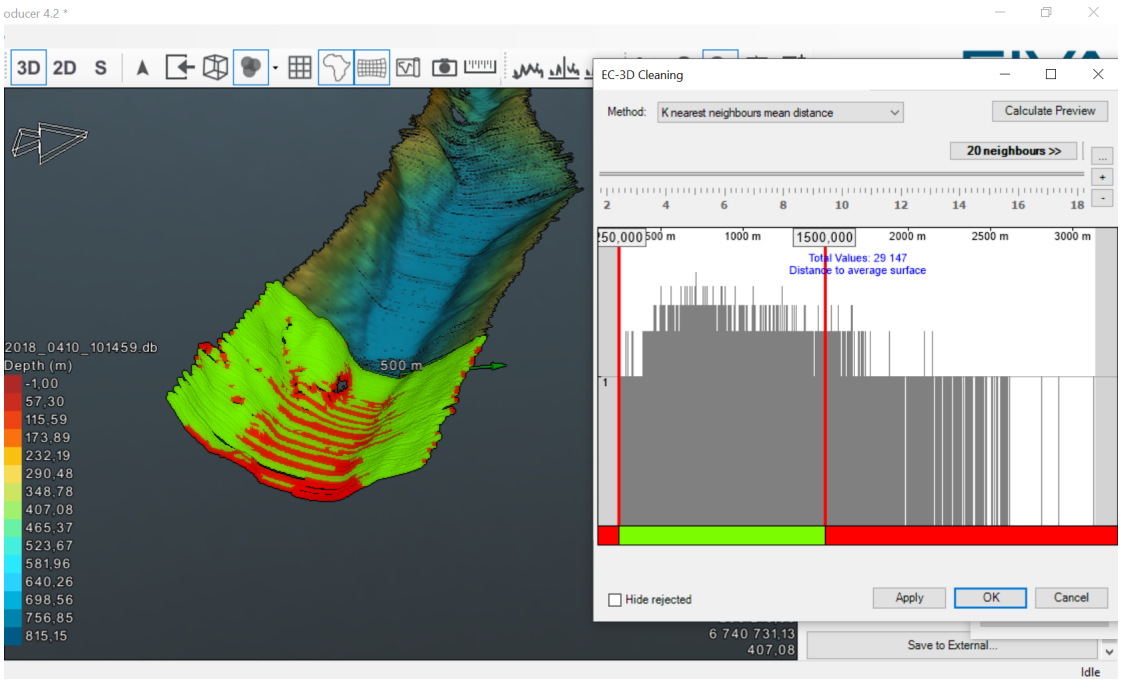
Figur 5.31. Illustration af K Nearest Neigbour Cleaning

På Figur 5.31 er afstandene til de 4(k) punkter illustreret, og metoden anvender disse afstande til at finde en gennemsnitlig afstand til de 4(k) naboer.

I eksemplet på Figur 5.32 og Figur 5.33 er der valgt et antal naboer på 20, og det ses tydeligt på figurerne, hvordan antallet at grønne punkter ændres, i takt med at den øvre grænseværdi ændres. I dette tilfælde er grænseværdien ændret fra 1000 m til 1500 m, og der ses en klar forskel på de to figurer. Derudover ses det, at det i dette tilfælde er den flade bund samt kanten af opmålingen, der stadig er markeret rød ved at øge den øvre grænseværdi. Dette kan være grundet at punkttætheden ikke er så stor her, og at der derfor ikke er så mange naboer tæt på punktet, hvilket vil kunne påvirke den gennemsnitlige afstand.



Figur 5.32. K Nearest Neigbour Cleaning, $K = 20$,$T_{min} = 250m$ og $T_{max} = 1000m$



Figur 5.33. K Nearest Neigbour Cleaning, $K = 20$,$T_{min} = 250m$ og $T_{max} = 1500m$

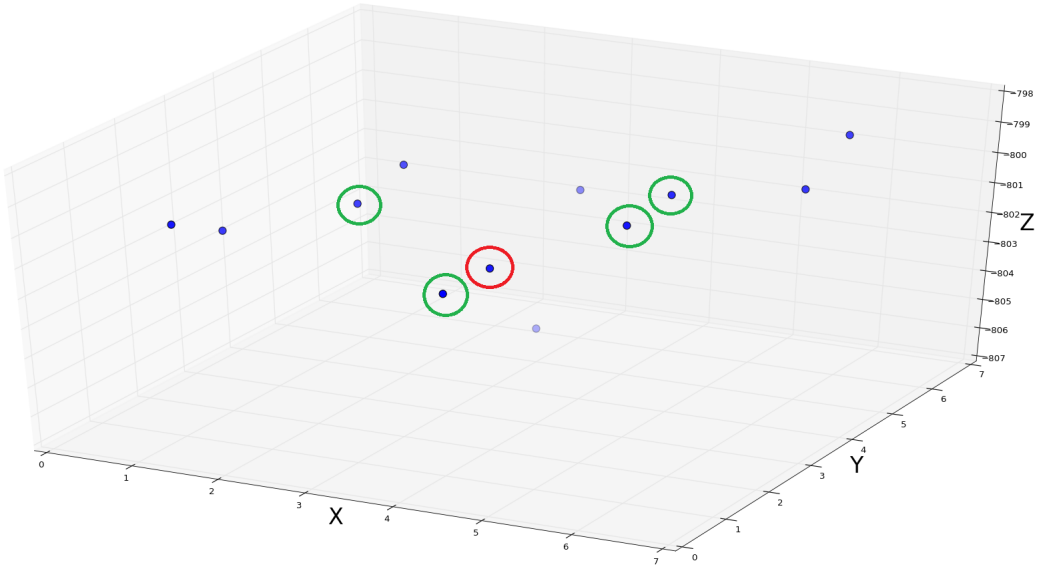
K Nearest Neigbour Cleaning, Z

Denne metode er nogenlunde den samme, som den lige beskrevet. Forskellen mellem de to varianter er, at denne udgave anvender afstanden i z-aksen, altså dybden, mens den beskrevet overfor anvender afstanden i 3D. Det vil sige, at den gennemsnitlige afstand i z-retningen beregnes som

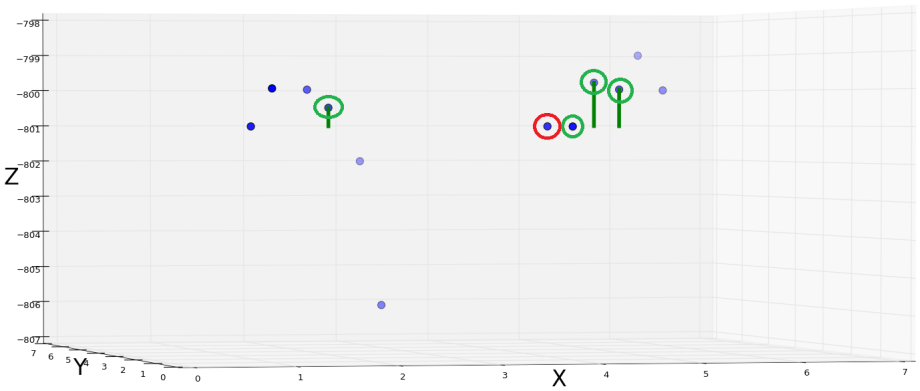
$$T = \frac{\sum_{i=1}^{k} \Delta z_i}{k}$$

hvor Δz_i angiver dybdeforskellen mellem et punkt og en nabo.

Anvendes det samme eksempel, som lige beskrevet i *K Nearest Neigbour Cleaning* er punkerne som på Figur 5.34. I stedet for at beregne alle afstandene i 3D og beregne den gennemsnitlige afstand, beregnes afstanden nu i 1D, altså i dybden.



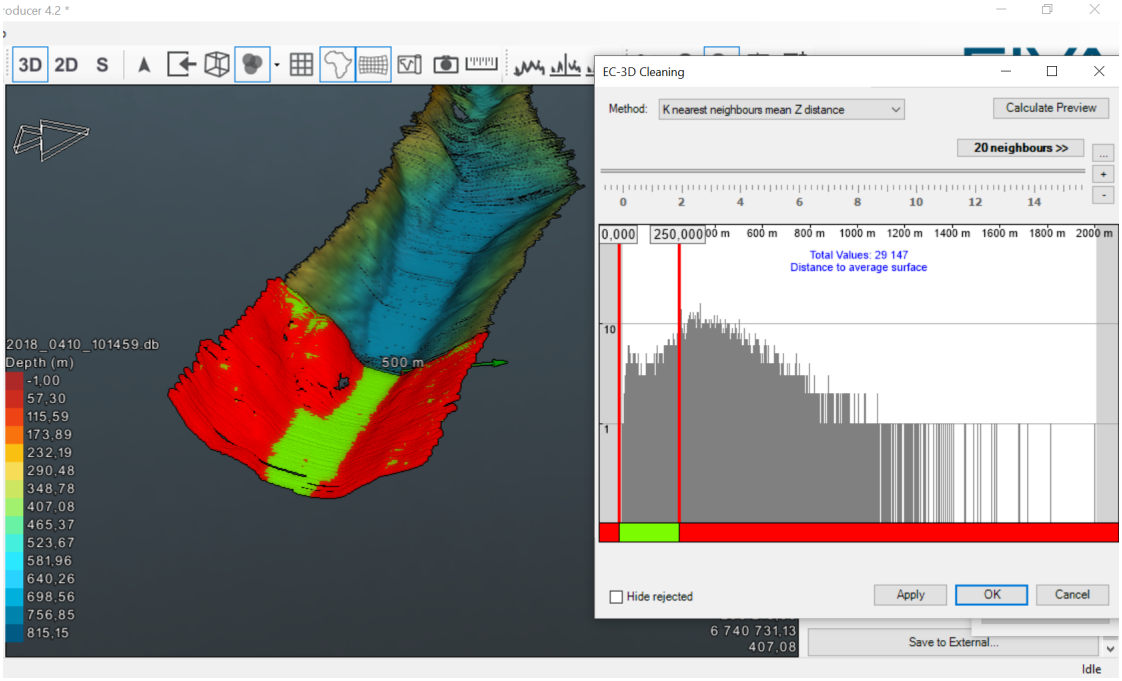
Figur 5.34. Illustration af K Nearest Neigbour Cleaning



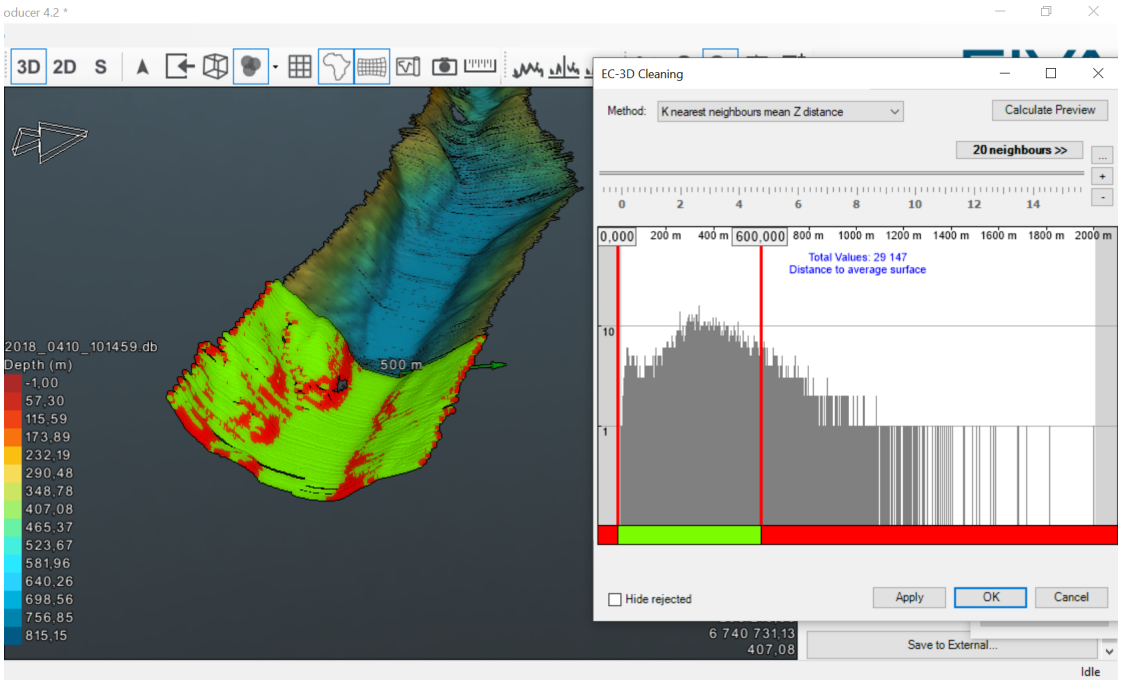
Figur 5.35. Illustration af K Nearest Neigbour Cleaning, Z

Af Figur 5.35 ses variationen i dybden mellem punktet markeret med rød og dets fire naboer. Det ses, at et af punkterne er beliggende i samme dybde, og der er derfor ingen dybdeforskel illustreret.

Anvendes metoden ser det ud som på Figur 5.36 og Figur 5.37. Det ses, at de mere vandrette dele af havbunden er markeret med grøn på Figur 5.36, hvilket giver mening, da variationen i z ikke er så stor her, sammenlignet med det resterende af det undersøgte område. Her er variationen i dybden større, men ændres den øvre grænse fra 250 m til 600 m, ses det, at flere punkter bliver markeret med grøn. Metoden kan derfor med fordel anvendes på jævne og vandrette dele af havbunden.



Figur 5.36. K Nearest Neigbour Cleaning, Z, $K = 20$,$T_{min} = 0m$ og $T_{max} = 250m$



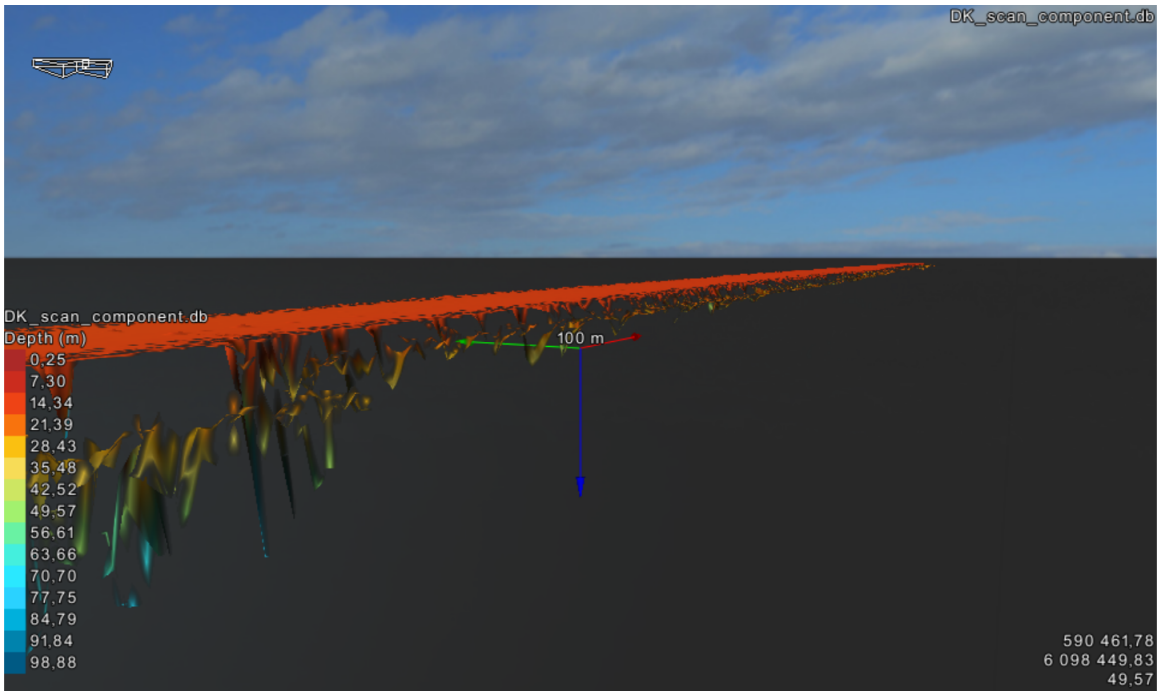
Figur 5.37. K Nearest Neigbour Cleaning, $K = 20$,$T_{min} = 0m$ og $T_{max} = 600m$

EC-3D er altså et værktøj, som indeholder en række forskellige metoder til at rense data. Metoderne kan anvendes på forskellige typer af data, hvor nogle metoder måske passer bedre til én type af data end en anden. Derudover kan metoderne anvendes i forbindelse med hinanden og en kombination af de forskellige metoder, kan muligvis medvirke til et mere fyldestgørende resultat end ved kun at anvende én af metoderne.

6 | Analyse af algoritmer

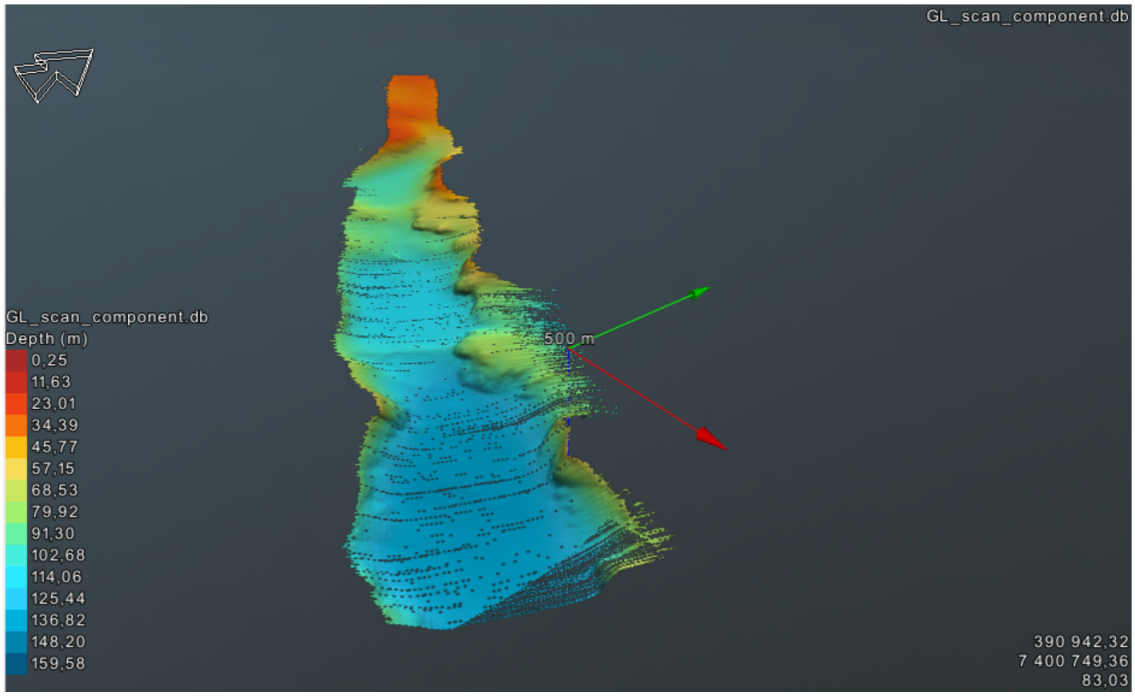
Dette kapitel omhandler forskellige test af de forskellige algoritmer og metoder omtalt i projektet. Ideen er, at anvende de forskellige algoritmer på både danske og grønlandske data, for at kunne vurdere hvordan algoritmerne behandler data afhængig af, hvilken type data der gives som input. Det kan eksempelvis ske, at en algoritme er bedre at anvende til én type data end en anden. Til at teste de forskellige metoder til at behandle data, anvendes EIVA's NaviModel 4.2. Dette software anvendes, idet det indeholder alle de metoder til efterbehandling, der er beskrevet i projektet.
Det er valgt, at anvende et dansk og et grønlandsk datasæt for at have to datasæt, der er i kontrast til hinanden. Den danske havbund er overvejende flad, og den grønlandske havbund er noget mere kuperet og varierer mere i dybden. I det grønlandske datasæt kan dybden variere 50 m indenfor en horisontal afstand på 30 m. Alle test vil i det følgende blive udført på rå data.
Kapitlet vil først anvende CUBE på begge typer data med forskellige parametre til anvendelse af algoritmen. Her vil der eksempelvis justeres på antallet af målinger, der som minimum skal knyttes til et knudepunkt. Til at vurdere CUBE anvendes en række kort som algoritmen producerer, som eksempelvis et der viser antal af hypoteser, et der viser styrke af hypoteser samt et der viser varianser.
Herefter testes det, hvordan både S-CAN component og S-CAN score behandler de rå data, og til sidst vil EC-3D blive anvendt på de rå data. Det vil sige, at alle de beskrevne metoder til at bearbejde data vil blive testet.

De data der anvendes til de forskellige test, er det samme både for det danske og det grønlandske. Det vil sige, at alle test udføres på det samme danske datasæt hele vejen igennem, og tilsvarende med det grønlandske. Dette skal medvirke til, at læseren får et kendskab til de forskellige datasæt, men også for bedre at kunne sammenligne testene på tværs.
De danske data der anvendes er opmålt af Forsvaret d. 19. april 2017 lige syd for Fyn. Den målte linje er omkring $5,5$ km lang, 100 m bred har en dybde på omkring 15 m. På Figur 6.1 ses et udsnit af en DTM model for det danske data. Her ses det blandt andet, hvor flad den danske havbund er, men der ses også nogle variationer i kanten af opmålingen, hvilket formentlig er en række fejlbehæftede målinger.



Figur 6.1. Screenshot af de danske data

Det grønlandske datasæt er opmålt af Forsvaret d. 14. juni 2017. Opmålingen er foretaget i det vestlige Grønland syd for Holsteinsborg/Sisimiut og nord for Itivdleq. Det opmålte område er omkring 3 km lang, varierer i bredde fra 200 m til 450 m, og har også varierende dybder fra 26 m til omkring 140 m. På Figur 6.2 ses det, at havbunden i Grønland varierer noget mere, end den gør i Danmark, men dog at målingen her, ikke umiddelbart er lige så påvirket af fejl som den danske opmåling, hvilket kan skyldes anvendelse af andet udstyr og fartøj.



Figur 6.2. Screenshot af de grønlandske data

Begge de anvendte datasæt er præsenteres som en DTM over området med en cellestørrelse på 2 m, og modellen er dannet ud fra rå data. For at kunne vurdere kvaliteten af data løbende i kapitlet undersøges det, hvilke standard afvigelser der er for data, og det er i NaviModel muligt at se, hvor mange celler, der har en standard afvigelse, der ligger under bestemte grænser. Standard afvigelserne bliver altså beregnet for hver DTM celle og ikke på hele datasættet. Det følgende er output fra NaviModel og er for både det danske og det grønlandske data.

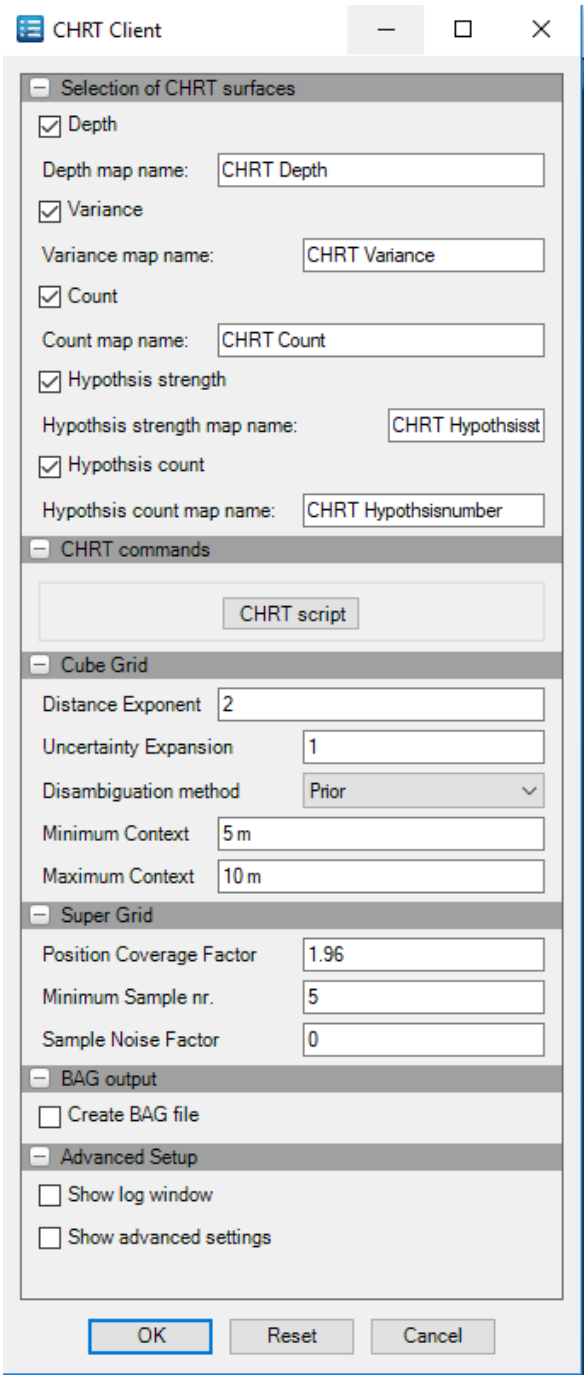
Standard afvigelse	Danmark	Grønland
<0,10 m	91,08%	75,08%
<0,20 m	94,93%	84,40%
<0,30 m	95,87%	90,01%
<0,40 m	96,26%	93,18%
<0,50 m	96,48%	95,28%
<0,60 m	96,62%	96,64%
<0,70 m	96,69%	97,60%
<0,80 m	96,82%	98,17%
<0,90 m	97,08%	98,54%
<1,00 m	97,33%	98,80%

Tabel 6.1. Procentdel af cellerne i de rå datasæt, der er under en given standard afvigelse for henholdsvis Danmark og Grønland

Af Tabel 6.1 ses det, at der ved de grønlandske data er en større procentdel af cellerne, der er har en standard afvigelse på mindre 1 m, men at der for det danske er en langt større andel af cellerne, der har en standard afvigelse på under $0,05$ m.Dette kan skyldes, at der i Grønland er større variationer i havbunden, hvilket kan medvirke til, at standard afgivelserne øges for hver celle. I dette tilfælde er cellestørrelsen 2 m, og er dybden varierende indenfor denne celle, vil det påvirke standard afvigelsen. Outputtet vil afhænge af cellestørrelsen, så et andet resultat vil opnås for samme datasæt hvis en anden cellestørrelse vælges.

6.1 CUBE anvendt på dansk og grønlandsk data

Dette afsnit vil omhandle CUBE, hvor algoritmen vil blive testet på både det grønlandske og det danske datasæt. Der vil i afsnittet kun blive anvendt CUBE på rå data, da det ønskes at se, hvordan algoritmen håndterer data med støj. Testen vil blive vurderet ved en subjektiv vurdering af de output, der kommer fra algoritmen. DTM cellestørrelsen vil variere ved test af de to datasæt. Når CUBE anvendes kan det vælges hvilke kort, der ønskes dannet, samt hvor stor en mængde målinger, der som minimum skal anvendes til at estimere dybden og usikkerheden i knudepunkterne. Parametrene, som kan indstilles til CUBE, ses på Figur 6.3.



Figur 6.3. Parametre, som kan indstilles til anvendelse ad CUBE

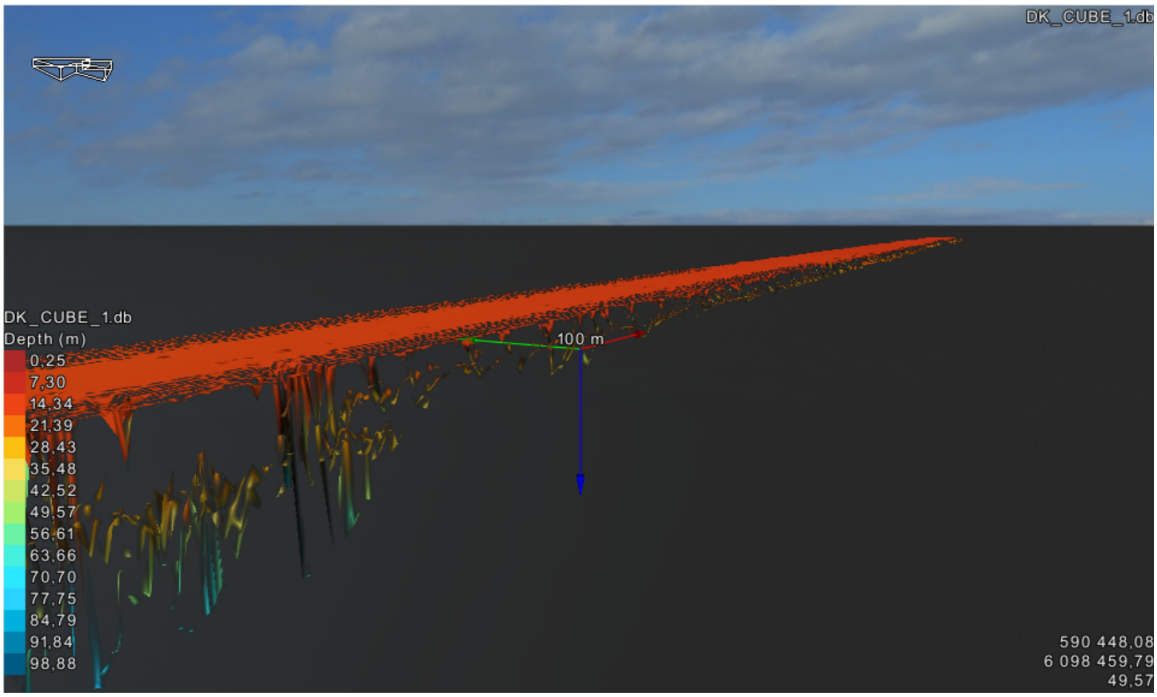
Her ses det, at der er seks overskrifter, hvor der under hver overskrift kan justeres på forskellige parametre. **Selection of CHRT surfaces** giver mulighed for at vælge hvilke kort eller lag, der skal dannes på baggrund af CUBE og giver mulighed for at navngive dem.

- *Depth* er en overflademodel, der viser dybderne.
- *Variance* er et lag, der viser varianserne pr cm^2.
- *Count* er et lag, der viser antallet af punkter pr m^2.
- *Hypothsis strength* er et lag, der viser styrken af den valgte hypotese.
- *Hypothsis count* er et lag, der viser antallet af hypoteser i det konstruerede grid.

De to sidste lag er med til at give en indikation for, hvornår der skal ske en brugerinddragelse fra en hydrograf. Antallet af hypoteser kan medvirke til at skabe tvivl om dybden i knudepunktet, og det kan ske, at CUBE vælger en hypotese, som virker mest sandsynlig, men som i virkeligheden er forkert. Dette kan eksempelvis være, hvis der er et vrag beliggende på havbunden, som har en mast, der stikker op. I så fald kan det ske, at hypotesen indeholdene mastetoppen vil blive kategoriseret som svag af CUBE og en anden stærkere hypotese vil blive valgt. Derfor kan begge lag være med til at finde de områder, hvor der er behov for at en hydrograf undersøger nærmere.
CHRT commands giver mulighed for at se og ændre på forskellige kommandoer i algoritmen.
Cube Grid og Super Grid giver mulighed for at indstille forskellige parametre for algoritmen. Den værdi, der indstilles på er minimum sample nr. som angiver det mindste antal målinger, der skal tilknyttes et knudepunkt.
BAG output giver mulighed for at danne en BAG fil.
Advanced Setup giver mulighed for at vise et log vindue og at vise avancerede indstillinger.

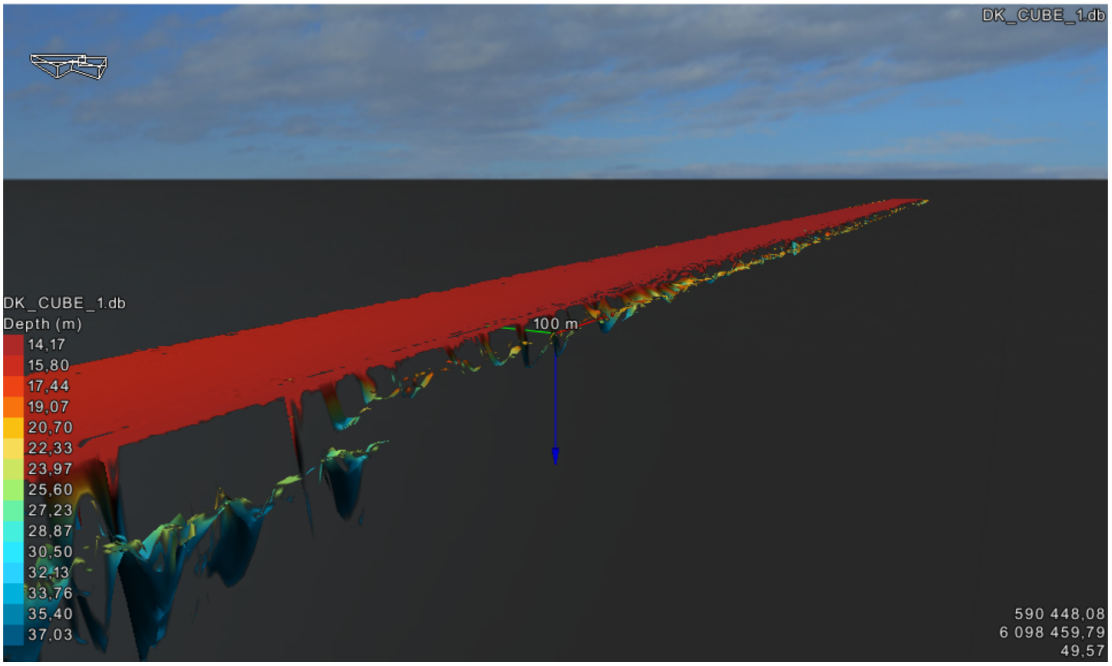
Danmark

Til at teste CUBE på det danske datasæt anvendes der auto detect til cellestørrelse. Her vurderer NaviModel på baggrund af punkttætheden en passende cellestørrelse, som i dette tilfælde er dette $1,37$ m, og giver en DTM som er illustreret på Figur 6.4.



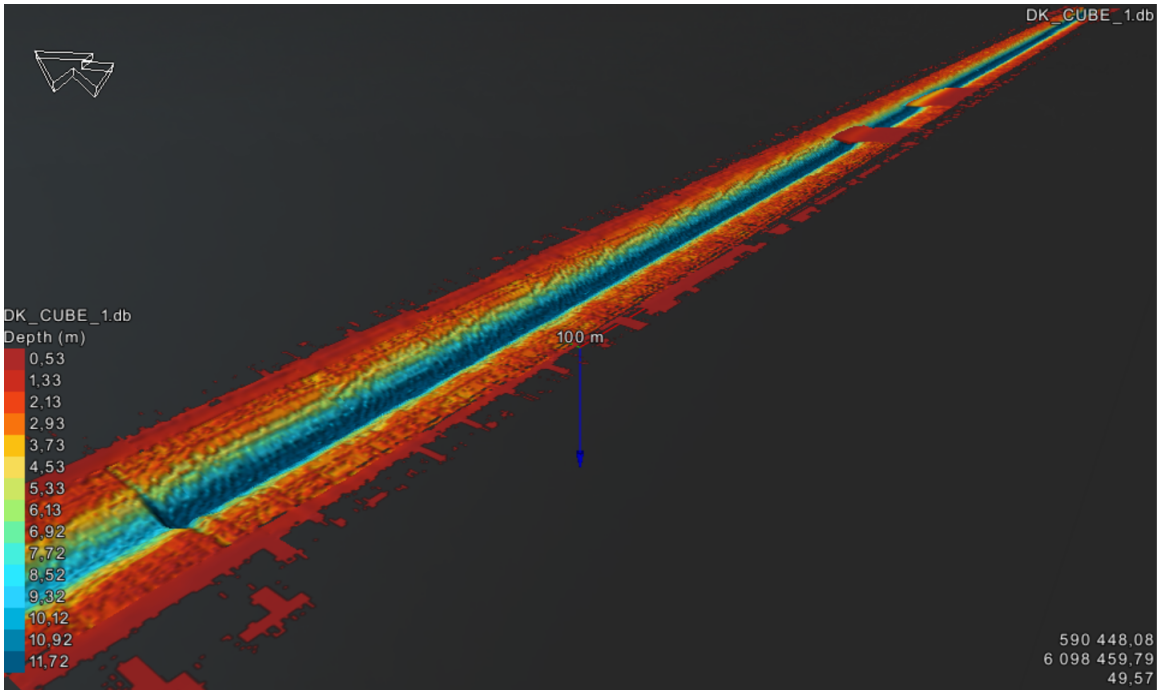
Figur 6.4. Det danske datasæt før anvendelse af CUBE

Til at anvende CUBE kan det bestemmes, hvilke kort der skal dannes, og hvor mange målinger der som minimum skal knyttes til et knudepunkt. I dette tilfælde vælges det, at anvende et minimums antal målinger på 10, samt at danne alle de mulige kort. På Figur 6.5 ses et udsnit af den DTM, der viser dybderne efter at CUBE er anvendt. Området er det samme som på Figur 6.4.



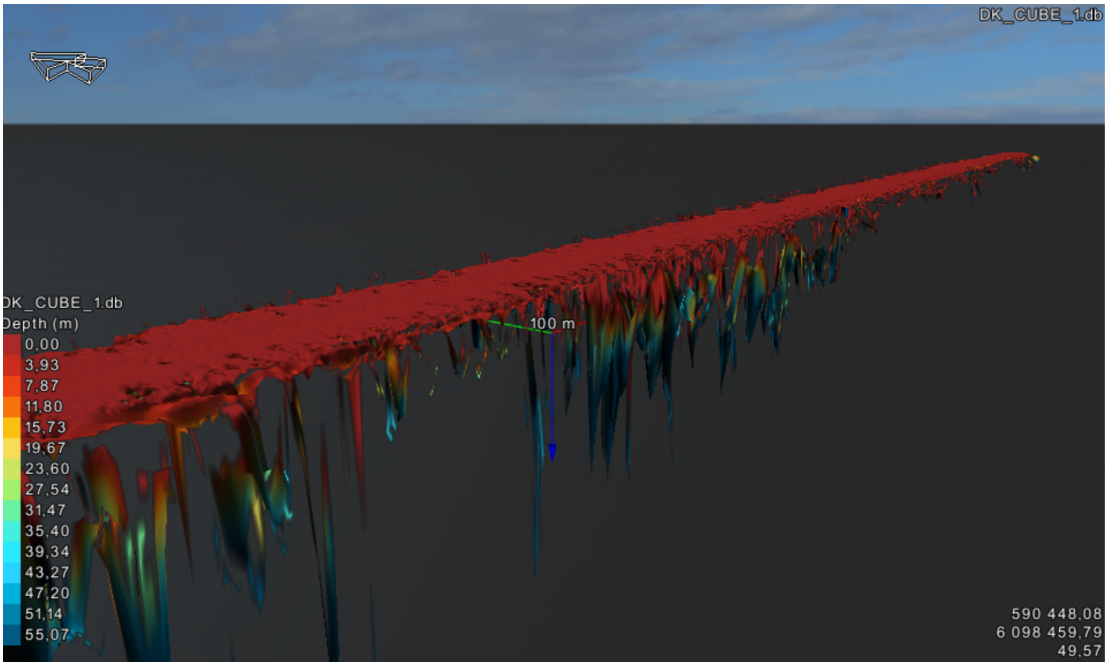
Figur 6.5. Det danske datasæt efter anvendelse af CUBE

Ved at sammenligne Figur 6.4 og Figur 6.5 ses det, at der er sket forbedringer med datasættet. CUBE estimerer dybder i knudepunkterne, men hvis knudepunkterne kun får tilknyttet information fra outliers, som alle er beliggende nogenlunde samme sted, kan den ikke skelne mellem, hvad der er "korrekt" data, og hvad der er fejlbehæftet data.
En model af målinger pr m^2 ses på Figur 6.6. Her indikerer de kolde farver en stor værdi, hvilket altså betyder at der er mange punkter. Signaturforklaringen på figuren står til at angive Depth, men angiver antallet af punkter pr m^2.



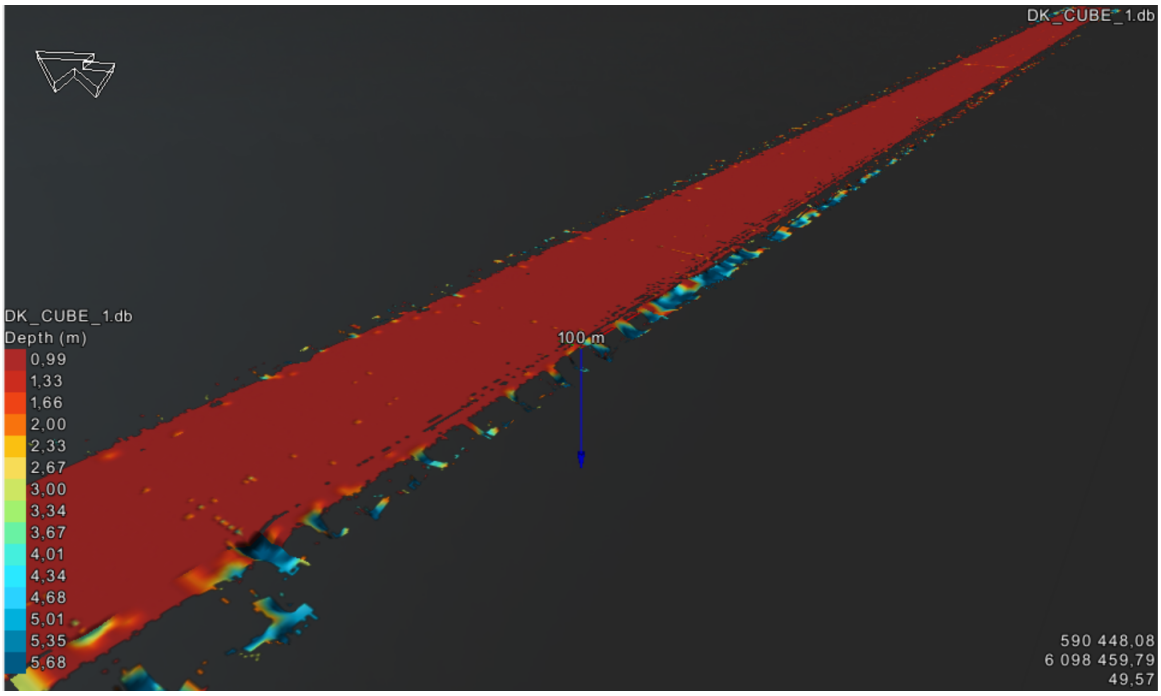
Figur 6.6. Illustration af antal målinger i knudepunkterne

Der dannes en række andre kort eller lag af CUBE, og undersøges disse, ses det blandt andet, at variansen som forventet et størst i kanten af opmålingen, se Figur 6.7. Af signaturen på figuren ses det, at der står Depth, men signaturen angiver varians.



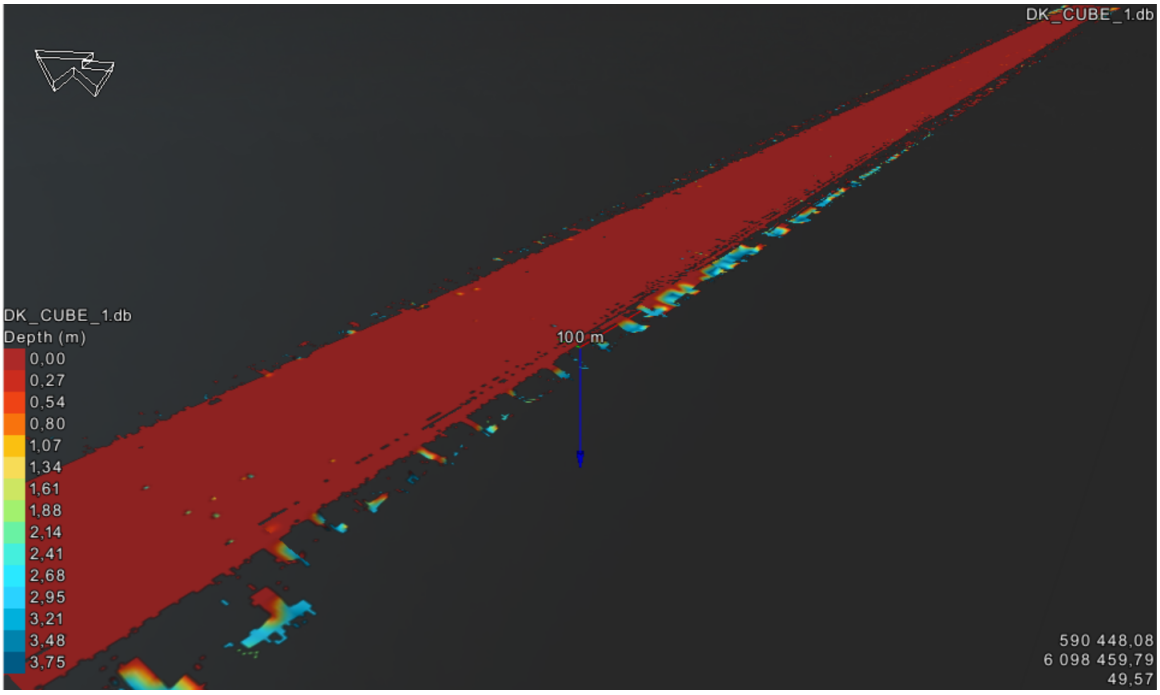
Figur 6.7. Varianser i det danske datasæt efter anvendelse af CUBE

Da CUBE er en DTM generator, kan der ikke sammenlignes standard afvigelser for datasættet før og efter, idet standard afvigelsen bliver beregnet ud fra punkterne, der genererer modellen. I og med at CUBE kun genererer en DTM slettes der ingen punkter, og der opnås dermed ikke mindre standard afvigelser i de enkelte celler, men den forbliver den samme.
Undersøges CUBE i forhold til hypoteser, kan det ses af Figur 6.8, at det er i kanten af den målte linje, hvor der er flere hypoteser i knudepunkterne. Des mere blå og kold i farven området er markeret des flere hypoteser er der. Der er også flere steder på linjen, hvor der er flere hypoteser, men det kan ses af det endelige resultat på Figur 6.5, at der er kommet er fornuftigt resultat ud af det. Signaturen på figuren står til at angive Depth, men der er tale om antal hypoteser.



Figur 6.8. Illustration af antal hypoteser i knudepunkterne

Undersøges det så, hvilken styrke der er i hypoteserne, ses det af Figur 6.9, at overfladen er overvejende rød, hvilet betyder en stærk styrke, og at der ikke kræves yderligere i det meste af området. Der er dog nogle områder i kanten, der er mere kolde i farverne, hvilket indikerer en svagere styrke af hypotesen, og dette giver mening i forhold til, at det er her, der er mange outliers og færre data. Signaturen på figuren står til at angive Depth, men angiver styrken af den valgte hypotese.

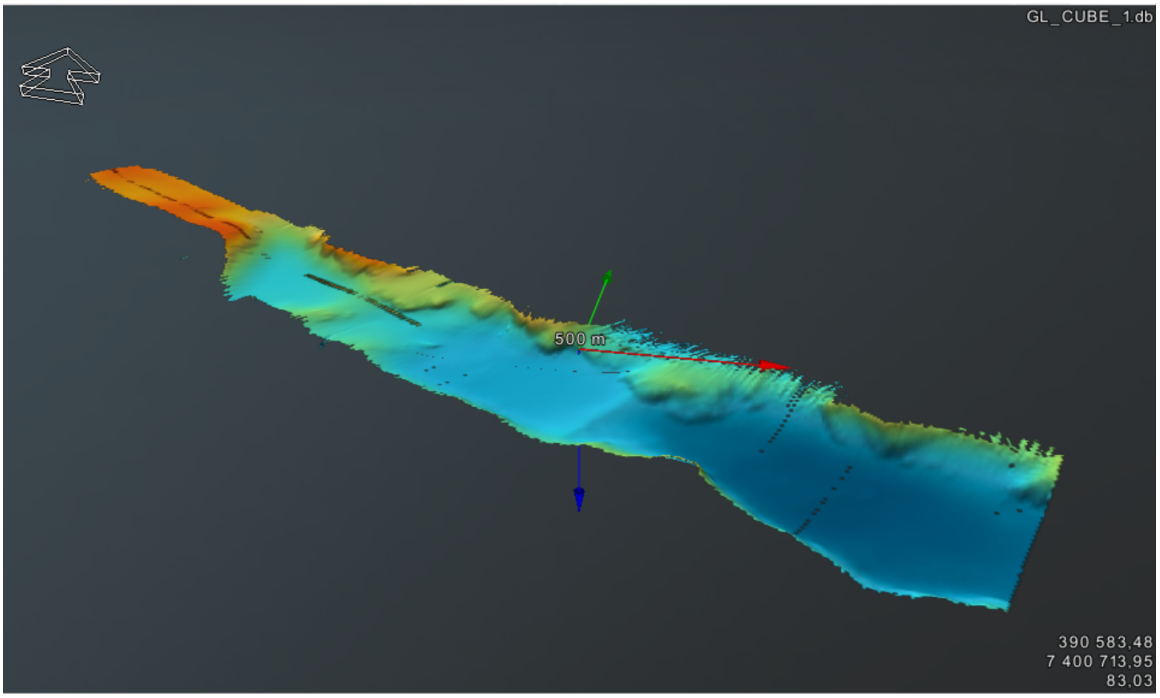


Figur 6.9. Illustration af styrken for hypoteser i knudepunkterne

CUBE er en hurtig måde til at behandle data på og virker godt, men algoritmen har svært ved at håndtere større samlinger af outliers. Dette er også beskrevet i diverse litteratur omkring CUBE, men i litteraturen var det ikke beskrevet hvor meget støj, der skulle til for at påvirke algoritmen. Det kunne derfor være interessant at undersøge, hvordan algoritmen arbejder, når der fjernes de værste outliers. Dette bliver dog ikke gjort i dette projekt.

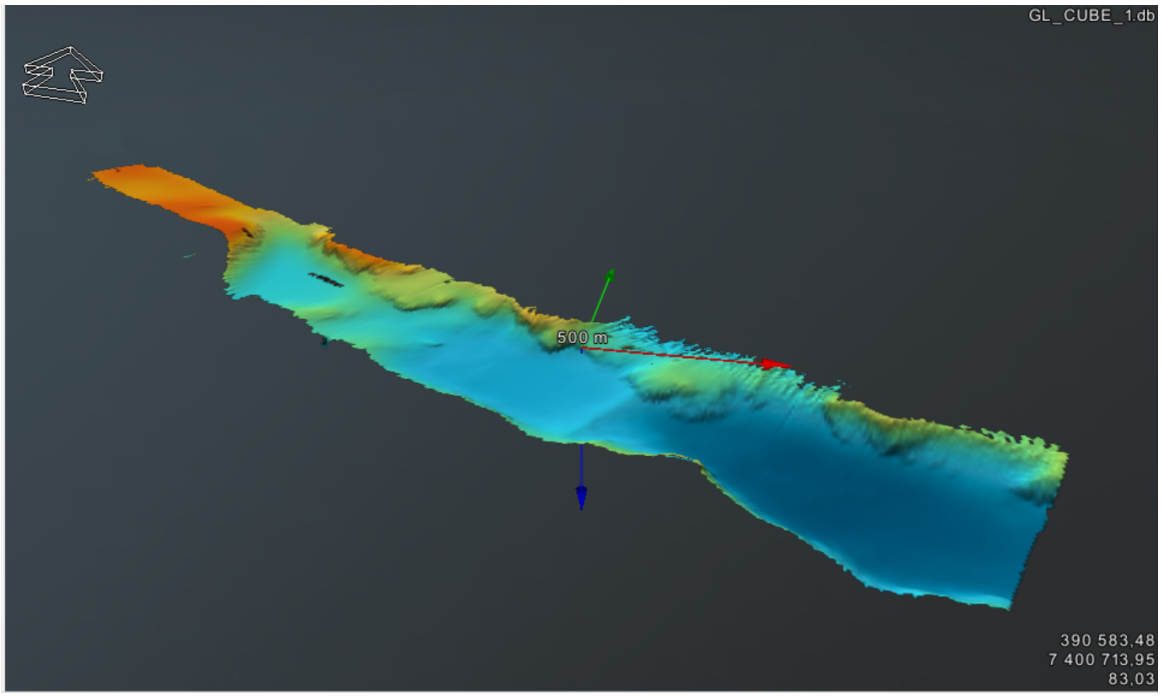
Grønland

Til at teste det grønlandske datasæt anvendes også auto detect til at finde en passende cellestørrelse til modellen, hvilken i NaviModel bestemmes til at være $5,98$ m, og modellen for de rå data ses på Figur 6.10.



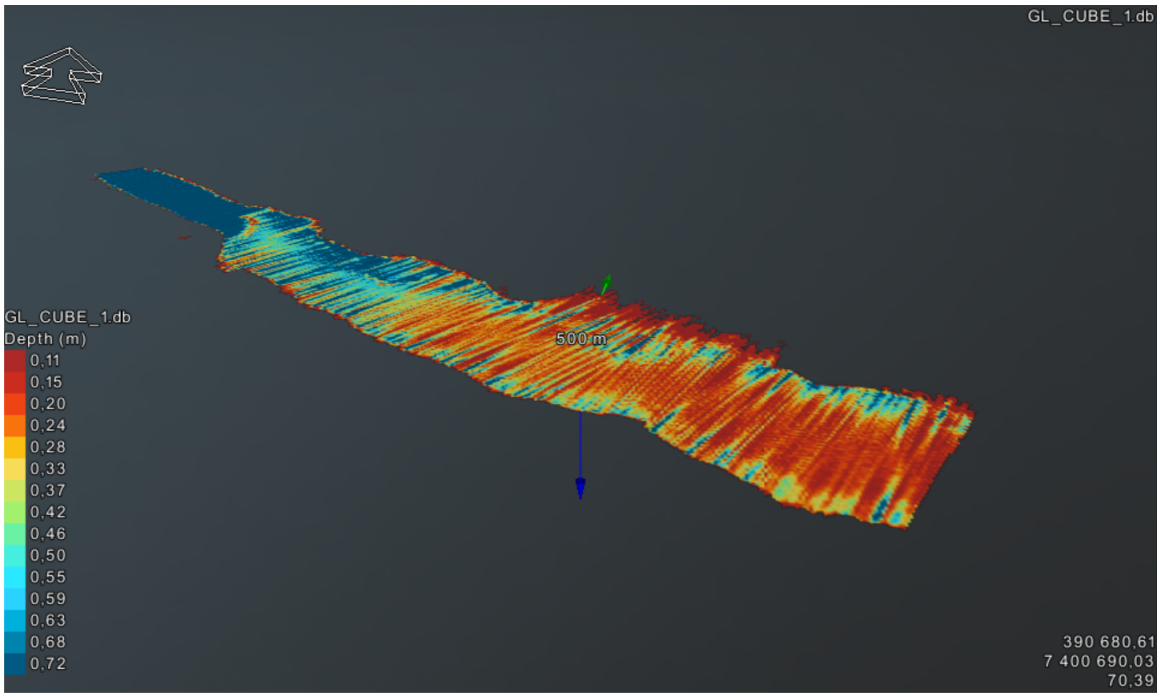
Figur 6.10. Det grønlandske datasæt før anvendelse af CUBE

Minimumskravet til antal af målinger, der skal knyttes til knudepunkterne sættes til 10. På baggrund af disse parametre generer CUBE den model, som er illustreret på Figur 6.11.



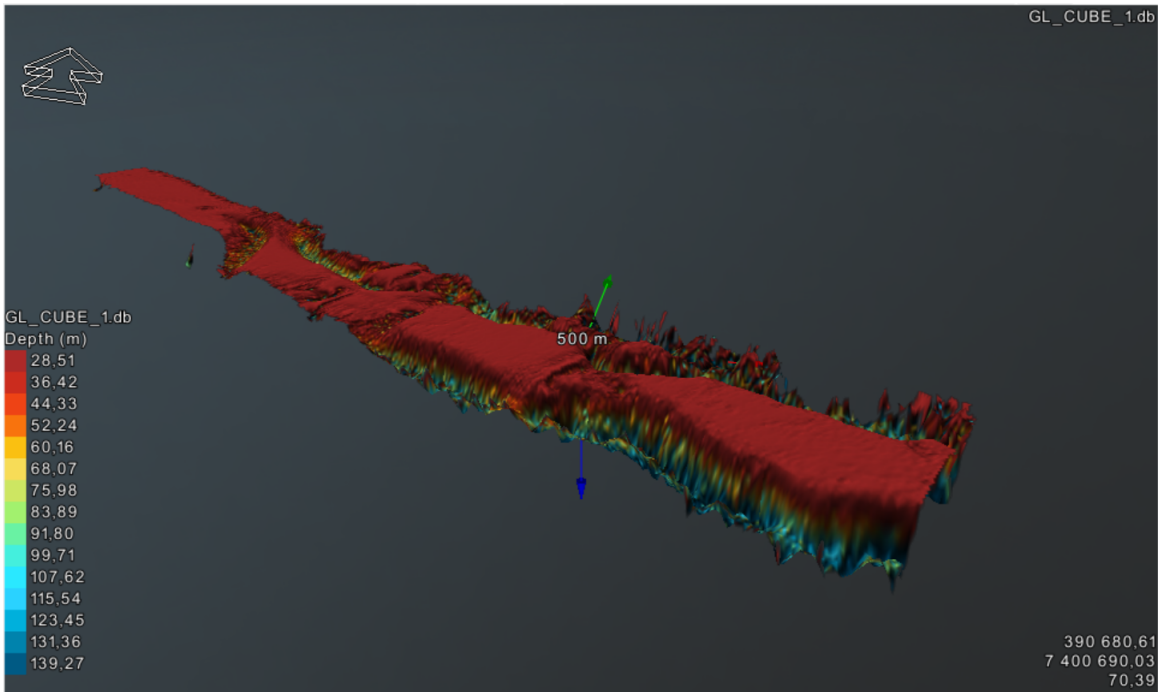
Figur 6.11. Det grønlandske datasæt efter anvendelse af CUBE

Det kan ses af Figur 6.11, at der er sket lidt forbedringer i forhold til de variationer, der skabes af nogle outliers, men det ses også, at det ikke er alle outliers, der håndteres lige vel, idet der stadig er få tilbage efter anvendelse af CUBE. Der er dog sket forbedringer i data.
Antallet af målinger ses på Figur 6.12, og her ses det, at punkttætheden af målinger er større på lavere vand. I signaturen står det angivet, at der indekseres efter dybden, men det er antal punkter.



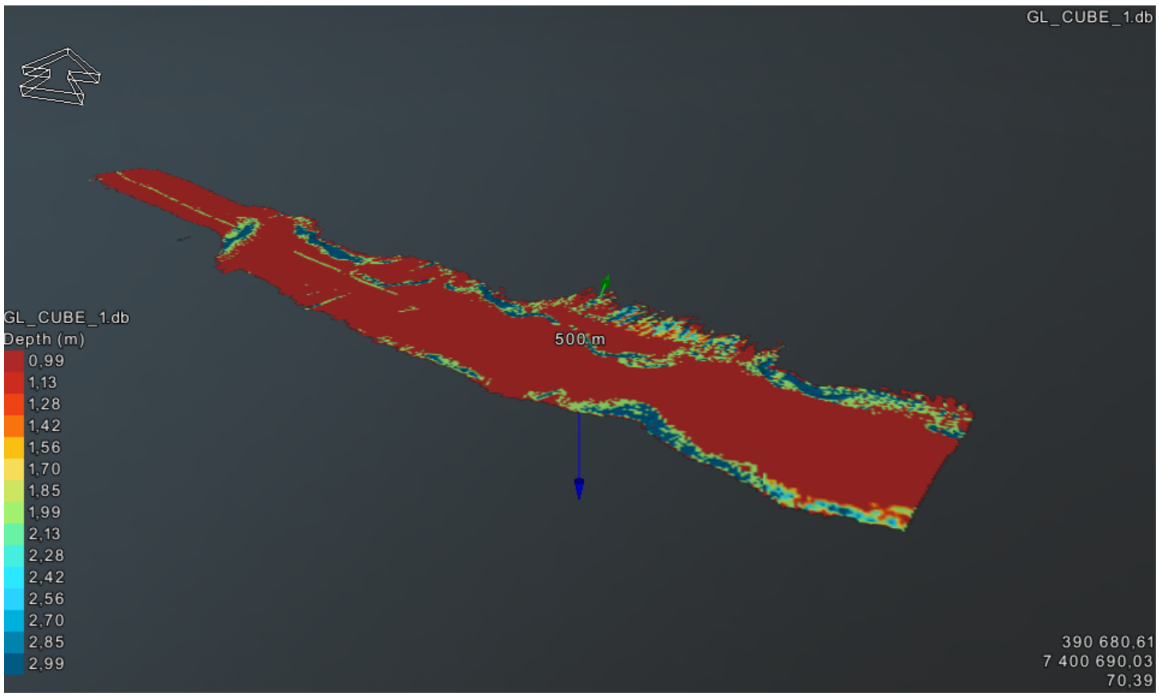
Figur 6.12. Illustration af antal målinger i knudepunkterne

På Figur 6.13 ses varianserne, og det ses som forventet, at varianserne er størst i de områder, hvor havbunden varierer i dybde. Dette giver mening, idet at variansen bliver beregnet pr m^2 og at der er forskelligt hvordan hver m^2 er varierende i dybde. Nogle steder er der store variationer i dybden, men der andre næsten igen er. Signaturen står til at angive Depth, men angiver variansen.



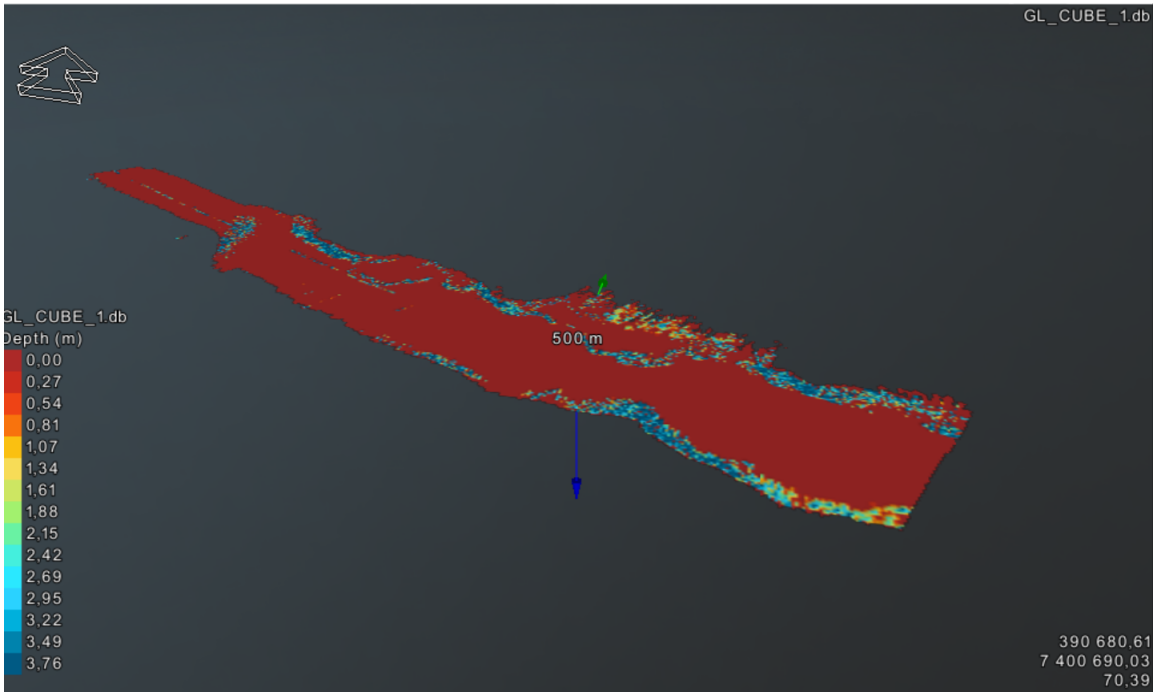
Figur 6.13. Varianser i det grønlandske datasæt efter anvendelse af CUBE

Ved at undersøge antallet af hypoteser, ses det også, at CUBE opretter flere hypoteser til de områder, der varierer i dybden, og at det er de samme områder, som har en høj varians. Det giver mening, at der tildeles flere hypoteser til disse områder, da knudepunkterne kan få tildelt målinger i en varierende dybde, som derfor kan påvirke estimaterne for knudepunkterne. Signaturen står til at angive Depth, men angiver antal hypoteser.



Figur 6.14. Illustration af antal hypoteser i knudepunkterne

Undersøges styrken af hypoteserne, ses det af Figur 6.15 at de samme områder som har mange hypoteser og høj varians, også har en svag styrke af hypoteserne. Sammenholdes dette så med resultatet Figur 6.11 ser det dog ud til, at der er valgt et rimeligt estimat for knudepunktets dybde, selvom at styrken for hypotesen er svagere end for så mange andre hypoteser. Signaturen står til at angive Depth, men angiver styrken af den valgte hypotese.



Figur 6.15. Illustration af styrken for hypoteser i knudepunkterne

Igen er metoden hurtig at anvende, og genererer en fornuftig DTM, men algoritmen har igen svært ved at håndtere outliers der ligger samlet.

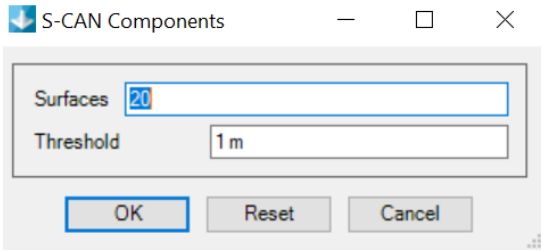
CUBE er en hurtig metode til at generere en terræn model for havbunden, men der skal formenligt en grov rensning til for at kunne generere en god model. Den tid, det tager at anvende CUBE afhænger selvfølgelig af hvilken computer og hvilket datasæt der anvendes. Til anvendelse af CUBE på disse datasæt er der ca anvendt 30-60 sekunder på at behandle data. CUBE gør et stort arbejde for brugeren, men der kræves stadig lidt brugerinteraktion til at behandle større samlinger af outliers. Disse kan hurtigt fjernes inden anvendelse af CUBE, selv manuelt. Derfor ses CUBE som klart anvendelig til at rense multibeam data.

6.2 S-CAN anvendt på dansk og grønlandsk data

I denne test ønskes det at teste, hvordan S-CAN behandler de to datasæt. Testen vil blive udført ved at anvende både S-CAN score og S-CAN component på begge datasæt og sammenligne for hver enkelt metode. De to datasæt vil begge blive anvendt med en cellestørrelse på 2 m. Testen vil blive vurderet på baggrund af en oversigt over standard afvigelser i cellerne samt en subjektiv vurdering af, hvilke målinger metoden vælger henholdsvis at beholde og at slette.

6.2.1 S-CAN component

Til at teste de to datasæt med S-CAN component gives de samme input til værktøjet. Det vil sige, at der vil blive dannet lige mange overflader og den samme grænseværdi vil blive angivet. Vælges det at sætte antal af overflader til 20 og en grænseværdi på 1 m, vil det sige, at når der er mere end 1 m vertikalt mellem målinger, vil der blive oprettet en overflade. I ovenstående tilfælde, vil der så blive dannet 20 overflader, og hvis der forefindes nogle målinger, der ikke passer ind i nogle af de overflader, vil de blive lagt i den kategori, der hedder "Remaning", altså resterende.



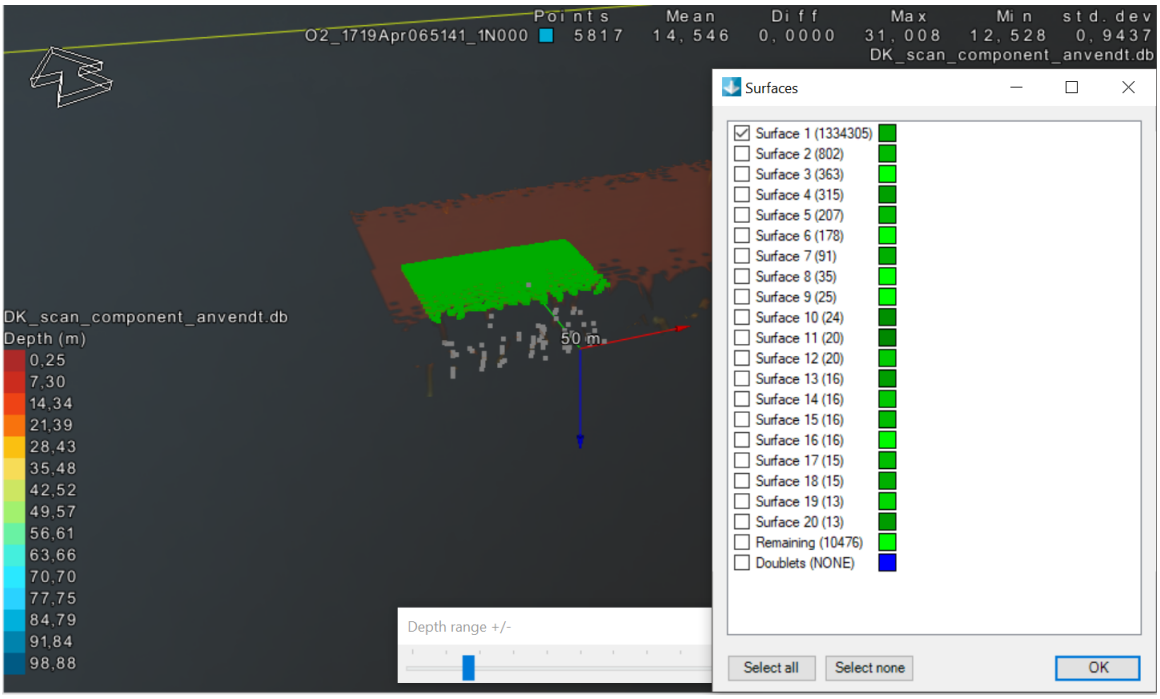
Figur 6.16. Værdier til S-CAN component

Det kan både være en fordel og en ulempe at have mange overflader. Fordelen ved mange overflader er, at såfremt der er dannet få overflader, og der er mange målinger, kan det ske, at der vil blive dannet en overflade med mange resterende målinger i. Disse målinger burde måske ikke alle burde være i den samme overflade, som derved kan indeholde målinger der ønskes slettet, men også nogle målinger der skulle bevares. For at løse dette problem betyder det, at metoden enten igen skal benyttes på hele opmålingsområdet, eller på et mindre område. Havde der været dannet flere overflader, var det ikke sikkert at denne problemstilling var opstået, da målingerne ville være placeret i flere overflader.
En ulempe ved mange overflader, er at det kan være tidskrævende at skulle slå overfladerne til og

fra for at undersøge, om de skal bevares eller ej. Dette er selvfølgelig brugerens sikkerhed for at ingen forkert data bliver slettet.

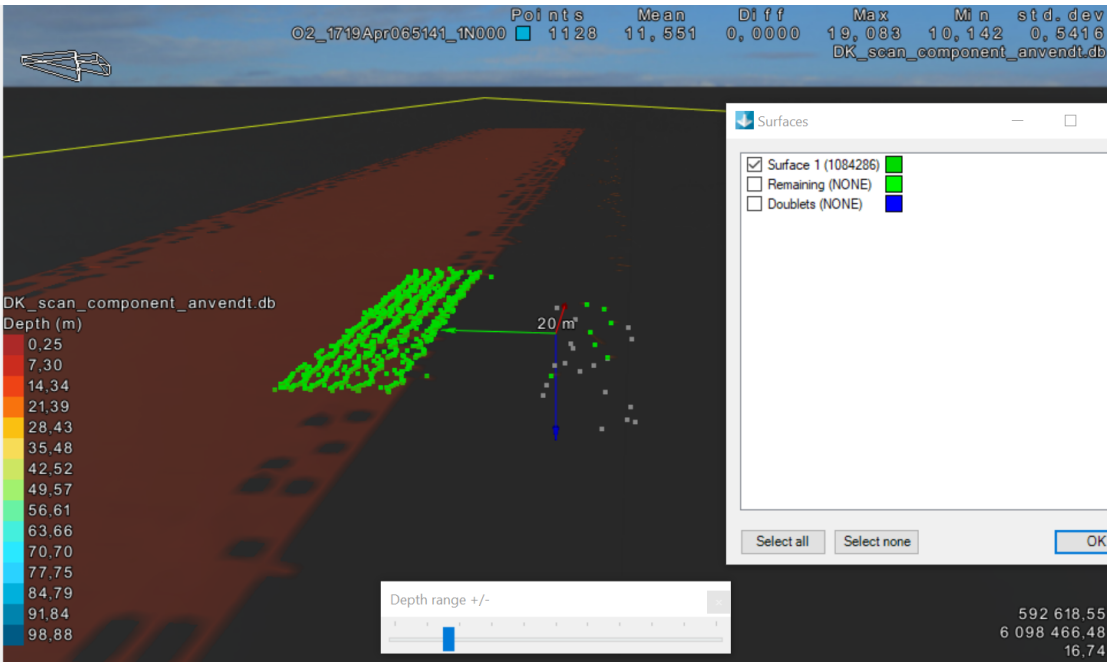
Danmark

Først anvendes S-CAN component på det danske data. Det vælges at sætte antallet af overflader til 20 og grænseværdien til 1 m. Når disse værdier er indtastet startes metoden, og outputtet som ses på Figur 6.17 fremkommer.



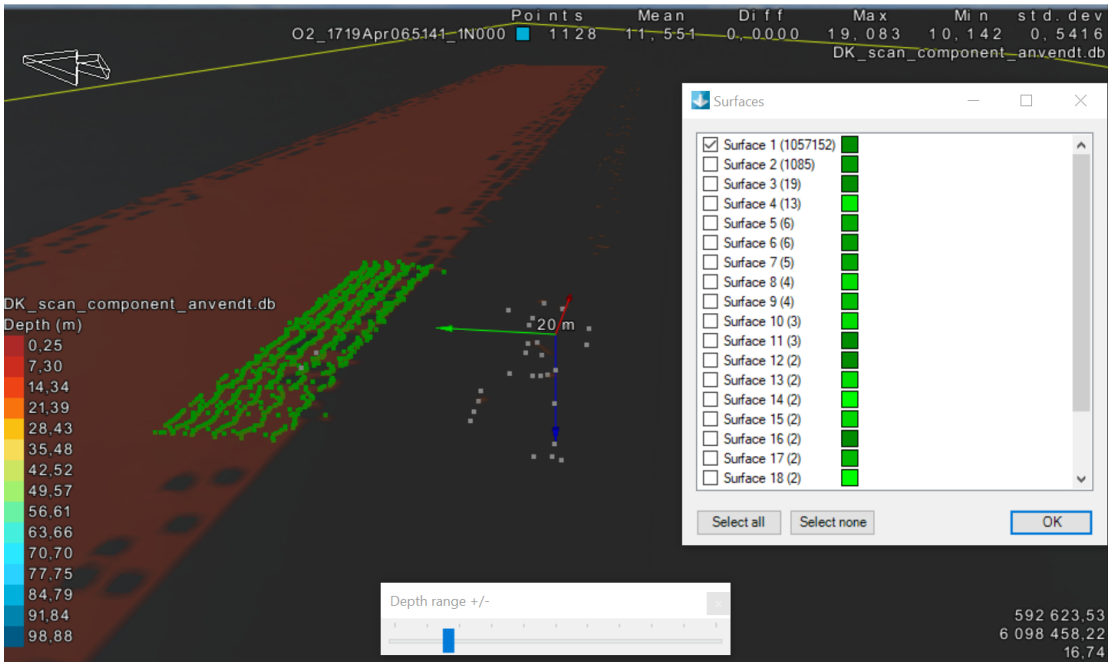
Figur 6.17. Oversigt over de konstruerede overflader, samt 3D model, der illustrer de accepterede overfader

På Figur 6.17 illustrerer de grønne punkter, målingerne i de overflader det er valgt at acceptere og de grå punkter, viser de målinger, der ikke er accepteret og er beliggende i andre overflader. I dette tilfælde er der kun valgt én overflade, som indeholder de fleste målinger fra opmålingen. Det er nu muligt at tilvælge andre overflader, såfremt det af brugeren vurderes at de indeholder målinger, der skal besvares.
Brugeren anvender nu markøren til at bevæge sig rundt i data for at vurdere om der er flere overflader, der skal accepteres. Dette kan brugeren vurdere idet at området med grønne og grå punkter følger med rundt, og derfra kan tilføje eller fravælge overfladerne i forskellige geografiske områder. I dette tilfælde vælges det kun at beholde én overflade, hvilken er overflade 1. Det kan lade sig gøre kun at vælge én overflade, idet den danske havbund er overvejende flad og at de fleste målinger derfor vil være beliggende inden for den samme overflade.
Efter at data er blevet behandlet, ses det, at der stadig er nogle målinger, der muligvis kan være påvirket af fejl, se Figur 6.18.



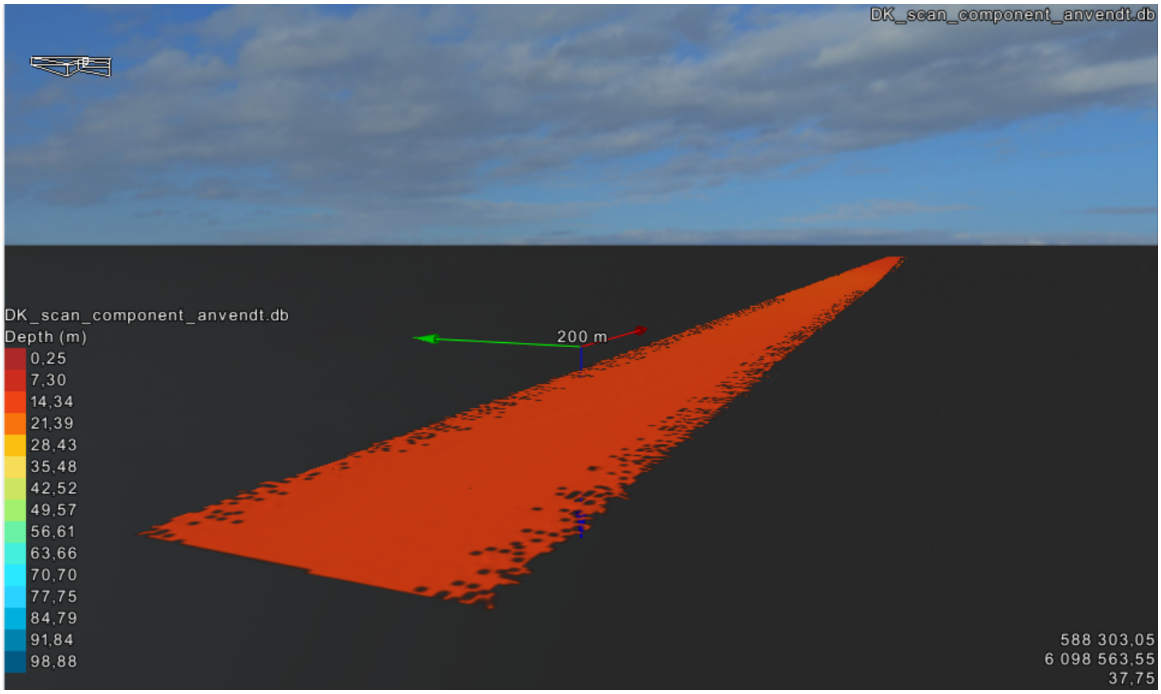
Figur 6.18. Det danske datasæt efter at S-CAN component er anvendt med en grænseværdi på $1,0$m

Til dette kan S-CAN component anvendes igen, men hvor grænseværdien nedsættes og kun anvendes i et specificeret område. Anvendes metoden på hele datasættet, kan det ske, at der slettes målinger, der ikke er fejlbehæftede og derfor ikke skal slettes. Det tilfælde, der er repræsenteret på Figur 6.18 er i kanten af den opmålte linje og er gældende i over halvdelen af datasættet. Derfor anvendes metoden igen på denne del af datasættet, men hvor grænseværdien nedsættes til $0,8$ m. At sætte grænseværdien til $0,8$ m er kun et gæt, og det vides derfor ikke hvordan det påvirker datasættet.



Figur 6.19. Det danske datasæt efter at S-CAN component er anvendt med en grænseværdi på $0,8$m

På Figur 6.19 ses data efter at der er anvendt en grænseværdi på $0,8$m. Det ses, at der er en forbedring i frasorteringen af fejlbehæftede målinger. Det vurderes herefter, at data er blevet renset tilfredsstillende, og der foretages ikke yderligere på datasættet. På Figur 6.20 ses det danske data renset med anvendelse af S-CAN component.



Figur 6.20. Det danske datasæt efter at S-CAN component er anvendt

Undersøges det hvordan standard afvigelserne har ændret sig efter rensning af data, ses det generelt, at procentdelen af celler der er mindre end en given standard afvigelse er blevet større.

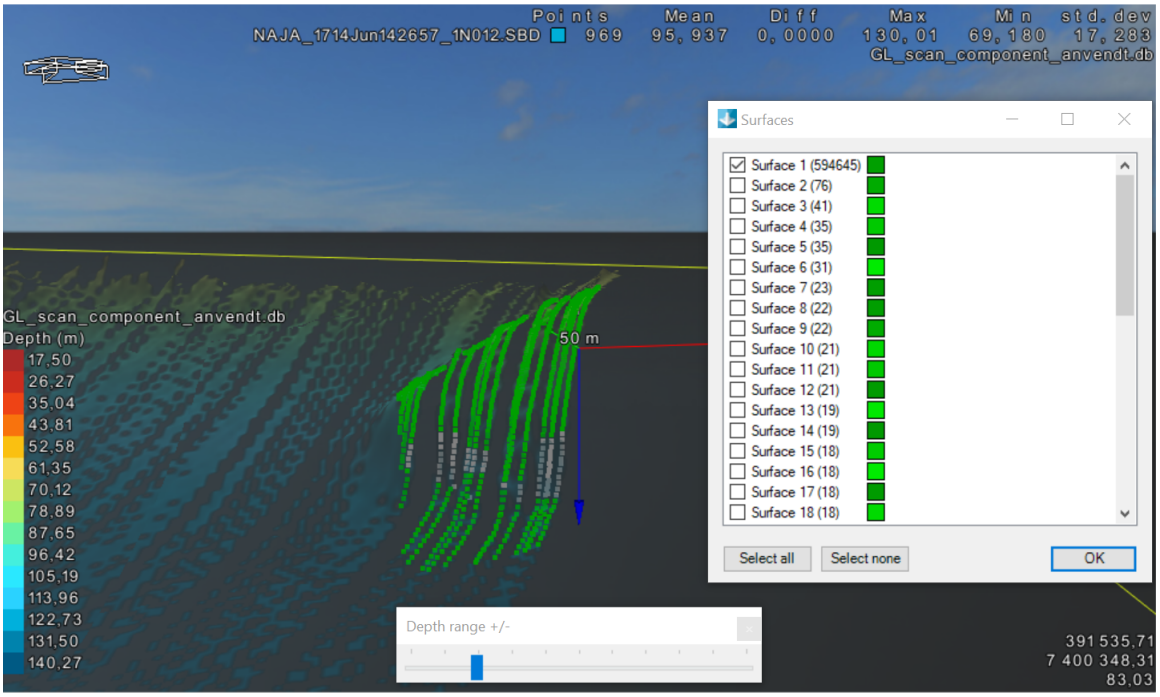
Standard afvigelse	**Danmark, før**	**Danmark, efter**
<0,10 m	91,08%	94,07%
<0,20 m	94,93%	98,22%
<0,30 m	95,87%	99,24%
<0,40 m	96,26%	99,65%
<0,50 m	96,48%	99,85%
<0,60 m	96,62%	99,96%
<0,70 m	96,69%	99,99%
<0,80 m	96,82%	100,00%
<0,90 m	97,08%	100,00%
<1,00 m	97,33%	100,00%

Tabel 6.2. Procentdel af cellerne, der er under en given standard afvigelse for henholdsvis Danmark efter anvendelse af S-CAN component

I det danske datasæt kan dette ses som en forbedring af data, idet havbunden er overvejende flad. Det skal dog nævnes, at såfremt der er noget beliggende på havbunden eller at havbunden varierer vil det ikke nødvendigvis kunne konkluderes, at kvaliteten af data er blevet forbedret ved større forandringer i usikkerhederne.

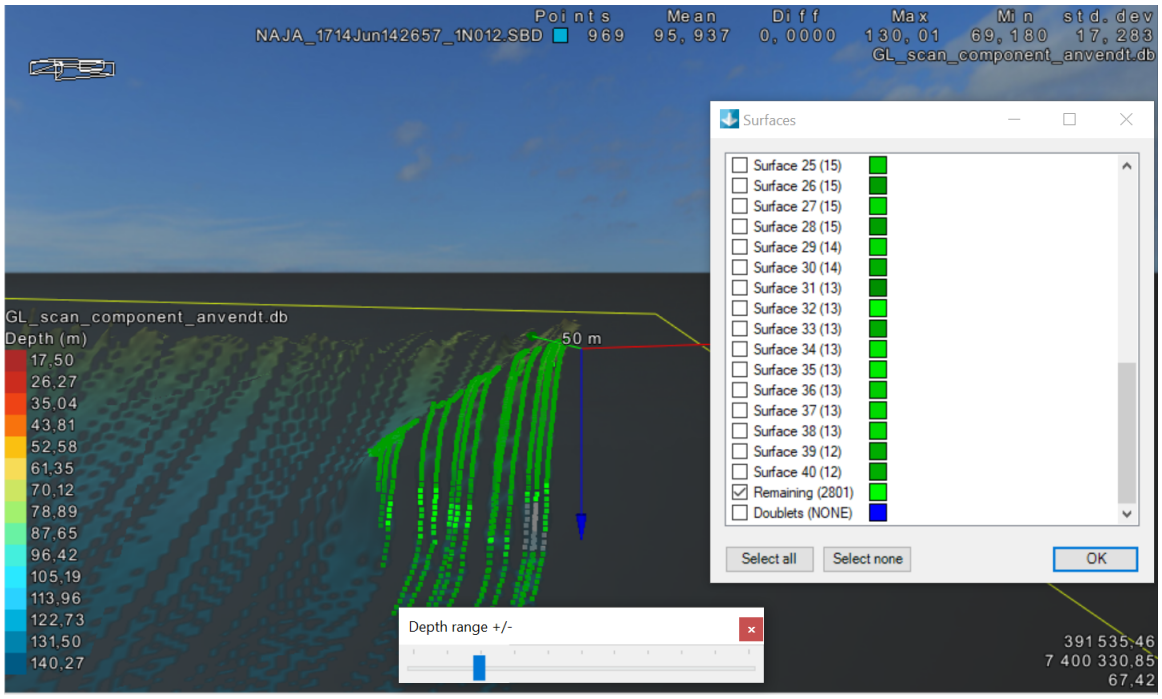
Grønland

Nu anvendes S-CAN component på det grønlandske data. Her vælges det, at sætte antallet af overflader til 40, idet der i Grønland er en mere varieret havbund, og det forventes derfor, at der er behov for flere overflader. Grænseværdien sættes ligesom ved behandling af det danske datasæt til 1 m. Når værdierne er indtastet startes metoden og outputtet ses på Figur 6.21.



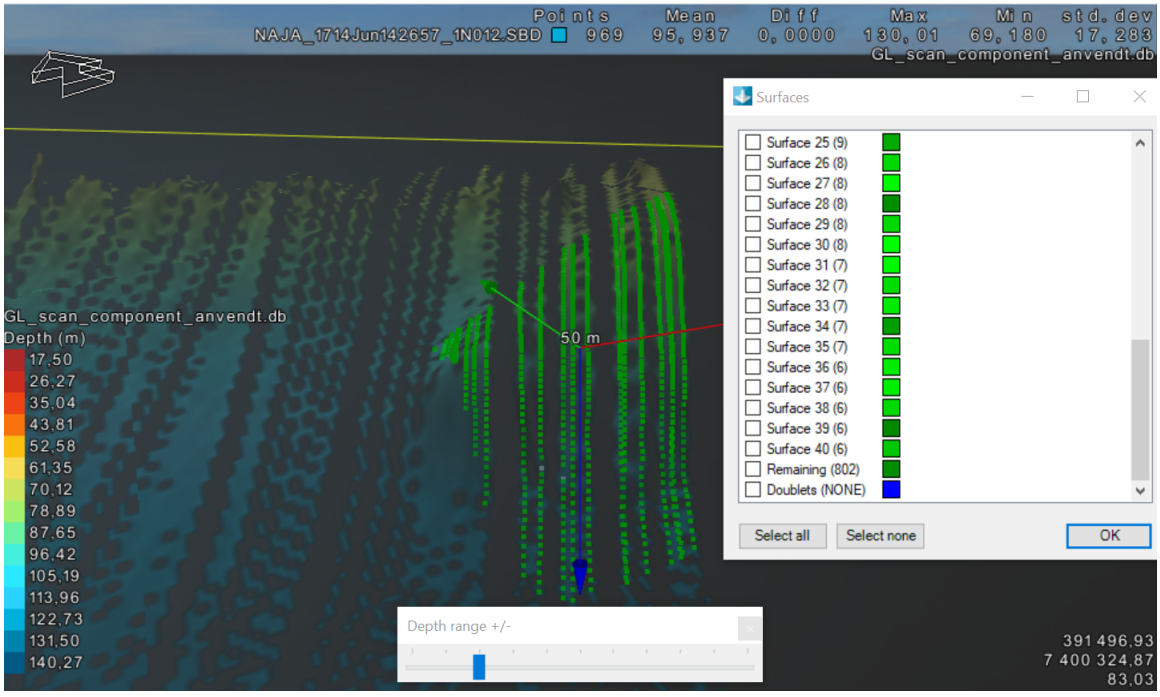
Figur 6.21. Oversigt over de konstruerede overflader, samt 3D model, der illustrer de accepterede overfader

Af Figur 6.21 ses det, at der ved behandling af det grønlandske data opstår en situation, som ikke opstod i forbindelse med behandling af de danske data. Det ses, at en mængde af målinger, der ikke umiddelbart er fejlbehæftede er markeret til sletning. Derfor skal de(nne) overflade(r) aktiveres, sådan at målingerne ikke slettes.
Mange af målingerne er placeret i den overflade der betegnes som "Remaining", se Figur 6.22. Derfor skal der enten ændres antal af overflader, sådan at der kan accepteret målinger der ikke er fejlbehæftede eller også skal grænseværdien ændres.



Figur 6.22. Oversigt over de konstruerede overflader, samt 3D model, der illustrer de accepterede overfader

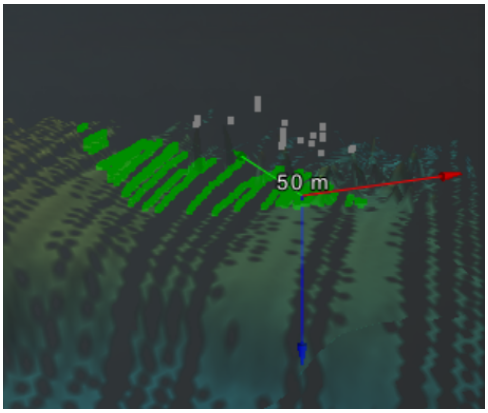
Det vælges, at ændre grænseværdien, idet der er stor variation i havbundes dybder, og det formodes, at det er begrænset hvor mange fejlbehæftede målinger, der vil blive medtaget. Det vælges stadig at beholde antallet af overflader. Efter at inputværdier er givet, ses det, at der er sket en forbedring.



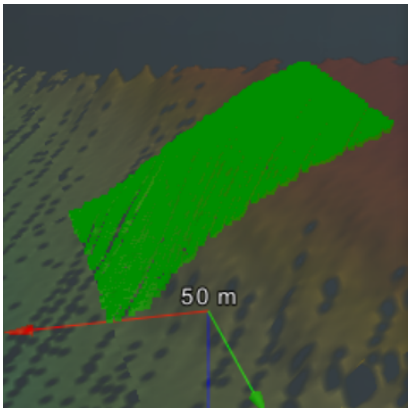
Figur 6.23. Det grønlandske datasæt efter at S-CAN component er anvendt med en grænseværdi på 2m

Af Figur 6.23 ses det, at der stadig er nogle få målinger, der stadig ikke er markeret i den primære

overflade, og de er placeret i overfladen "Remaining". Undersøges det resterende datasæt ses det yderligere, at metoden fint frasorterer fejlbehæftede målinger, og beholder de korrekte målinger under jævne stigninger i havbunden.



Figur 6.24. Frasortering af støj



Figur 6.25. Jævn variation af havbunden

Problemet med S-CAN component opstår, idet der er en stærk kupering i havbunden, som eksemplet på Figur 6.23 indikerer. Hvis antallet af overflader øges til 80 og grænseværdien stadig er 2 m er der stadig målinger fra eksemplet på Figur 6.23 der er i overfladen "Remaining", men antallet i overfladen mindskes til 611 frem for 2801 i tilfældet med 40 overflader og en grænseværdi på 1 m. Det vælges så at acceptere overflade 1 og "Remaining", for derefter at anvende metoden igen, sådan at overfladen "Remaining"deles op i flere overflader. Ved at skabe så stort et antal overflader som 80 skabes ikke et overblik over data. Der accepteres data hvor metoden er anvendt med 80 overflader og en grænseværdi på 2 m.
Selv efter anvendelse af metoden igen er det svært at frasortere støj uden at slette korrekte målinger. Metoden ses derfor ikke hensigtsmæssig at anvende på kuperede områder.

I og for sig er data ikke renset færdigt, men det undersøges alligevel, hvordan standard afvigelserne har ændret sig fra det oprindelige datasæt til efter den første anvendelse af metoden.

Standard afvigelse	Grønland, før	Grønland, efter
<0,10 m	75,08%	75,13%
<0,20 m	84,40%	84,54%
<0,30 m	90,01%	90,06%
<0,40 m	93,18%	93,23%
<0,50 m	95,28%	95,33%
<0,60 m	96,64%	96,69%
<0,70 m	97,60%	97,65%
<0,80 m	98,17%	98,23%
<0,90 m	98,54%	98,60%
<1,00 m	98,80%	98,86%

Tabel 6.3. Procentdel af cellerne, der er under en given standard afvigelse for Grønland efter anvendelse af S-CAN component

Det ses, at der ikke er sket den helt store påvirkning på standard afvigelserne, hvilket dog heller ikke var forventet. Standard afvigelserne i de enkelte celler vil være større i de grønlandske data end i de danske, men det grønlandske datasæt var heller ikke ligeså fejlbehæftet som det danske.

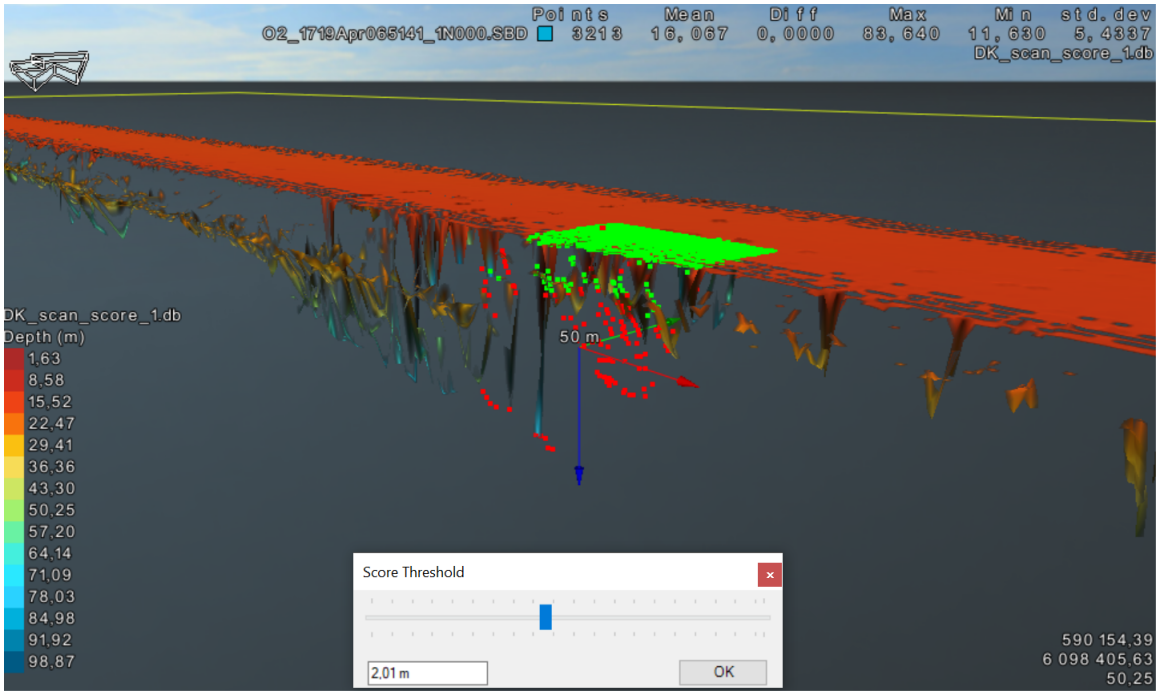
Det er klart, at standard afvigelsen bliver mindre ved at der fjernes outliers, da de påvirker standard afvigelsen til det højere, men påvirkningen afhænger dog også af hvor mange målinger der er i cellen i forhold til antallet af outliers. Når der i det grønlandske datasæt ikke fjernes så mange outliers, kan det heller ikke forventes, at der sker de store ændringer i standard afvigelserne.

6.2.2 S-CAN score

Når S-CAN score anvendes dannes der ligesom i S-CAN component en række overflader, og forskellen er blot, at til S-CAN score gives der ingen inputværdier, og grænseværdien justeres på en bar. Da S-CAN component og score ligger meget op af hinanden, må det forventes, at der til dels ses det samme resultat som i ovenstående, men score udgaven giver mulighed for at tilpasse grænseværdien mere til specifikke områder. Til testen hentes henholdsvis danske og grønlandske data ind og grænseværdien justeres, sådan at data renses bedst muligt ved anvendelse af metoden.

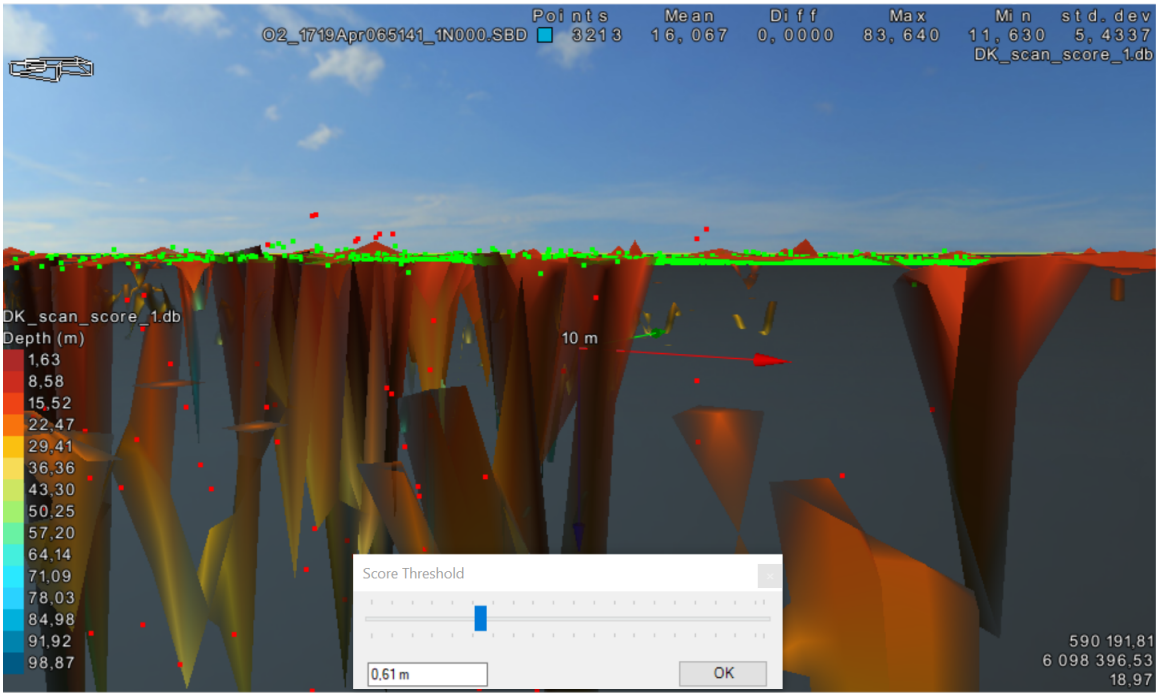
Danmark

Igen testes det danske datasæt først. Her vælges det, at behandle hele datasættet med den samme grænseværdi, ligesom det var tilfældet ved component varianten. Sættes grænseværdien til 1 m vil der opnås samme resultat som efter første anvendelse af S-CAN component, idet der her blev valgt kun at acceptere en overflade.



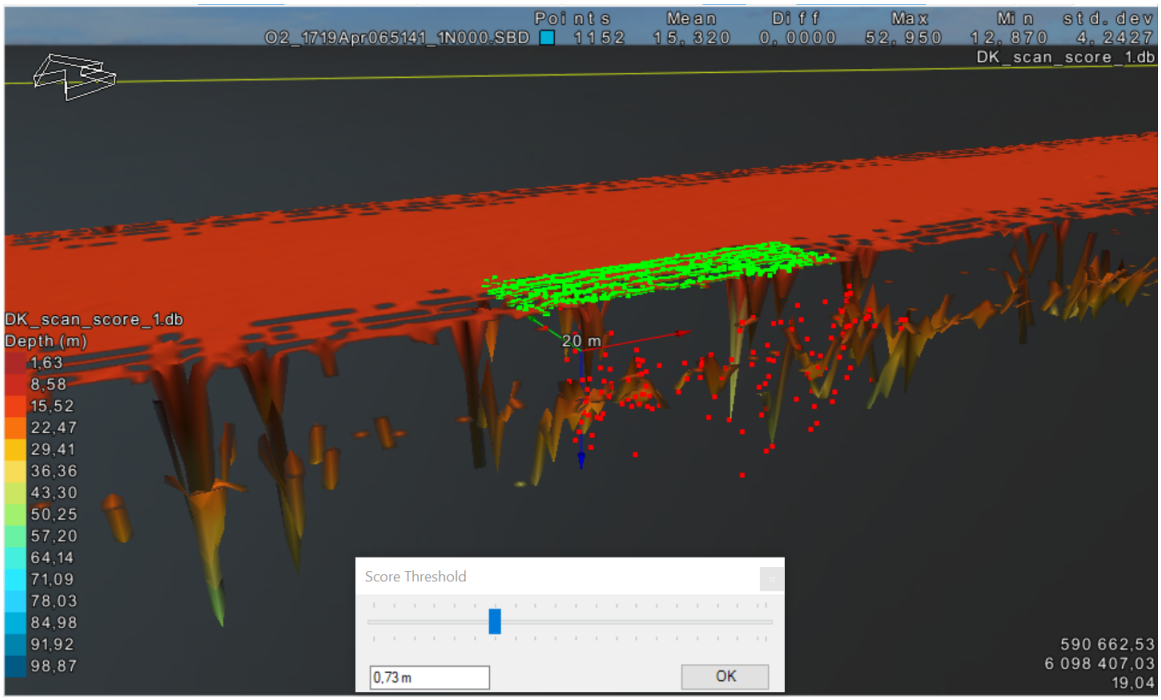
Figur 6.26. 3D model af data med illustration af accepterede og ikke-accepterede dybder

Af Figur 6.26 ses det, at der er en bar til at justere grænseværdien på. Som default er værdien til at starte med sat til $2,01$ m, hvilket ikke ses som passende baseret på de røde og de grønne markeringer. De røde indikerer som tidligere nævnt de målinger, der ikke accepteres, hvor de grønne viser de accepterede målinger. Det ses altså, at grænseværdien skal nedsættes for at være passende.



Figur 6.27. 3D model af data med illustration af accepterede og ikke-accepterede dybder

Nedsættes grænseværdien til $0,61$ m kan det ses, at der er flere målinger, der ikke accepteres. Dette er nødvendigvis ikke godt, idet der kan være faktiske variationer i havbunden, som på den måde vil blive slettet. Derfor skal grænseværdien sættes op, for at disse variationer bliver bevaret.



Figur 6.28. 3D model af data med illustration af accepterede og ikke-accepterede dybder

Det vurderes, at $0,73$ m er en passende grænseværdi for hele datasættet og data renses på denne baggrund. Herefter undersøges det, hvordan standard afgivelserne ser ud. Ligesom ved component

udgaven, ses det, at procentdelen af cellerne der er mindre end en given standard afvigelse, igen er større.

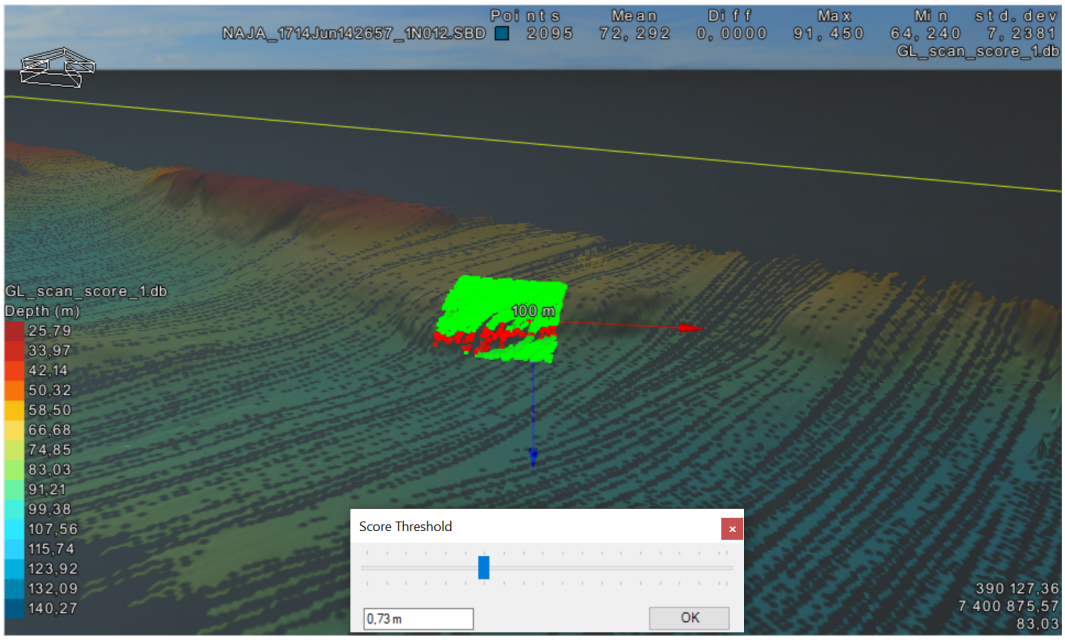
Standard afvigelse	Danmark, før	Danmark, efter
<0,10 m	91,08%	94,15%
<0,20 m	94,93%	98,32%
<0,30 m	95,87%	99,33%
<0,40 m	96,26%	99,72%
<0,50 m	96,48%	99,89%
<0,60 m	96,62%	99,98%
<0,70 m	96,69%	99,99%
<0,80 m	96,82%	100,00%
<0,90 m	97,08%	100,00%
<1,00 m	97,33%	100,00%

Tabel 6.4. Procentdel af cellerne, der er under en given standard afvigelse for Danmark efter anvendelse af S-CAN score

Igen skyldes de lave standard afvigelser, at havbunden er overvejende flad. Det ses yderligere, hvis der sammenlignes med Tabel 6.2, at der opnås nogenlunde det samme resultat efter anvendelse af S-CAN component og score. Dette skyldes den flade havbund. Da der er anvendt nogenlunde samme grænseværdi vil resultaterne også blive nogenlunde det samme. Det skal dog nævnes, at S-CAN score er hurtigere at anvende end S-CAN component, idet det hurtigere giver brugeren en fornemmelse af hvad grænseværdien skal være.

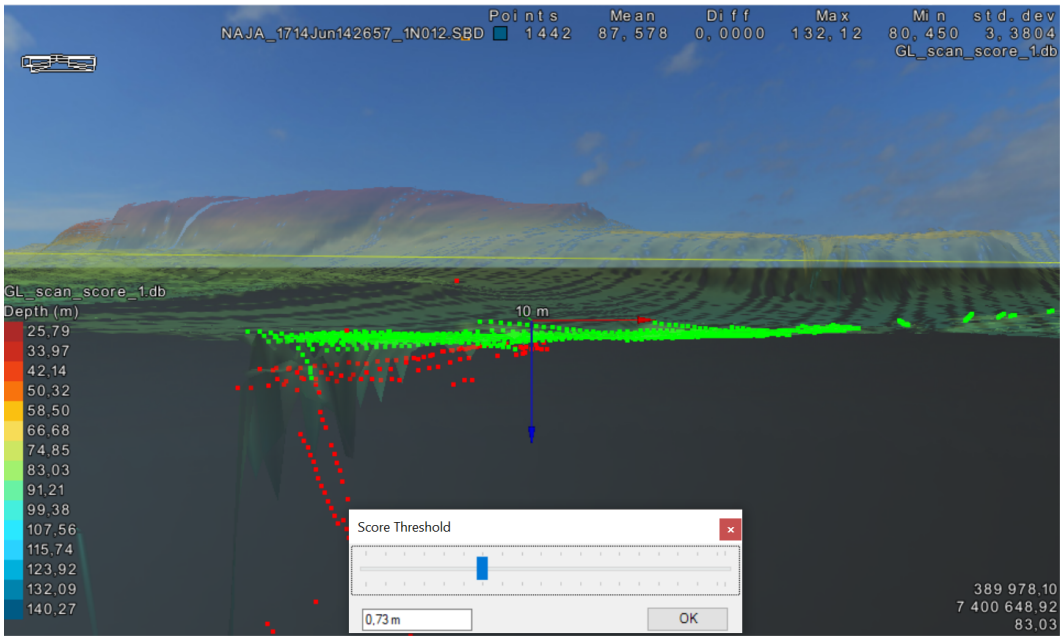
Grønland

Nu udføres samme test på det grønlandske data som på det danske. Ved anvendelse af metoden anvendes en default grænseværdi på $0,73$ m, hvilket på Figur 6.29 ses ikke at være hensigtsmæssigt i det viste område.



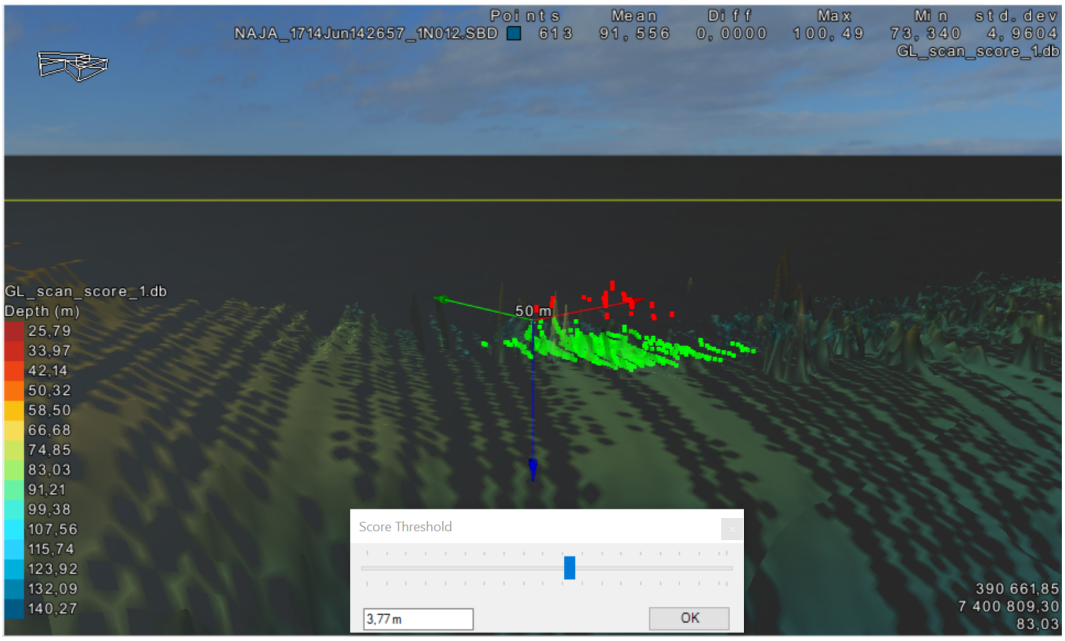
Figur 6.29. 3D model af data med illustration af accepterede og ikke-accepterede dybder

Selvom grænseværdien anses som værende for lille i dette område, kan der være områder hvor den er passende. Et eksempel på dette kan ses på Figur 6.29, hvor det ses, at ved at sætte grænseværdien op, vil flere fejlbehæftede målinger muligvis accepteres.



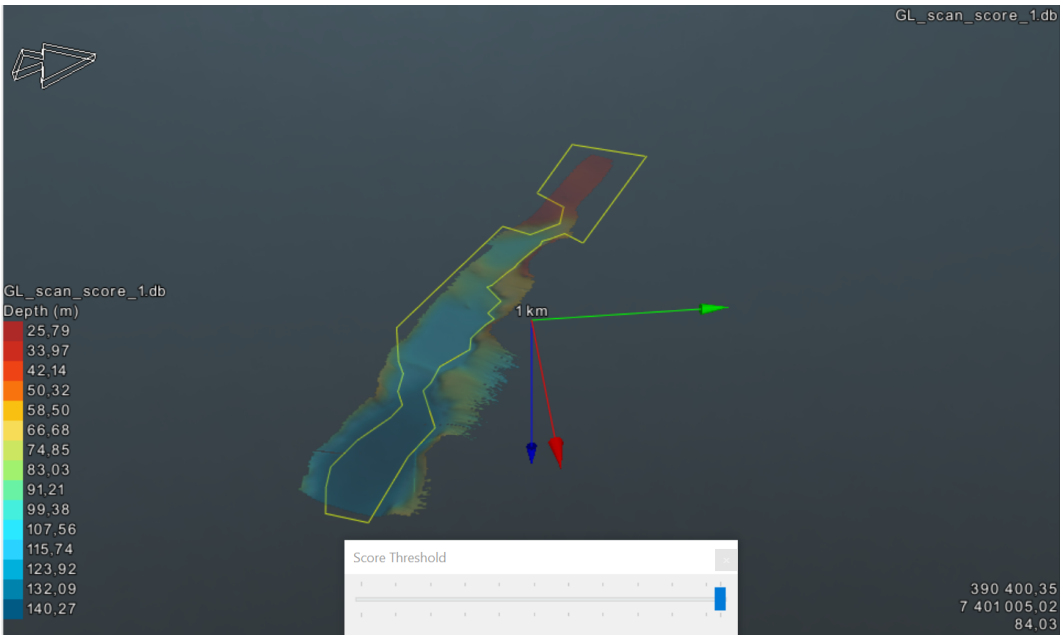
Figur 6.30. 3D model af data med illustration af accepterede og ikke-accepterede dybder

Derfor vælges det, først at grov sortere i data, sådan at de værste outliers fjernes. Dette gøres ved at tilpasse grænseværdien således, at de større variationer i havbunden accepteres. Den grænseværdi der ses som passende i dette tilfælde er på $3,77$ m. På den måde er de værste outliers frasorteret fra data, se Figur 6.31, men det vides, at der stadig er en række fejlbehæftede målinger tilbage i data.



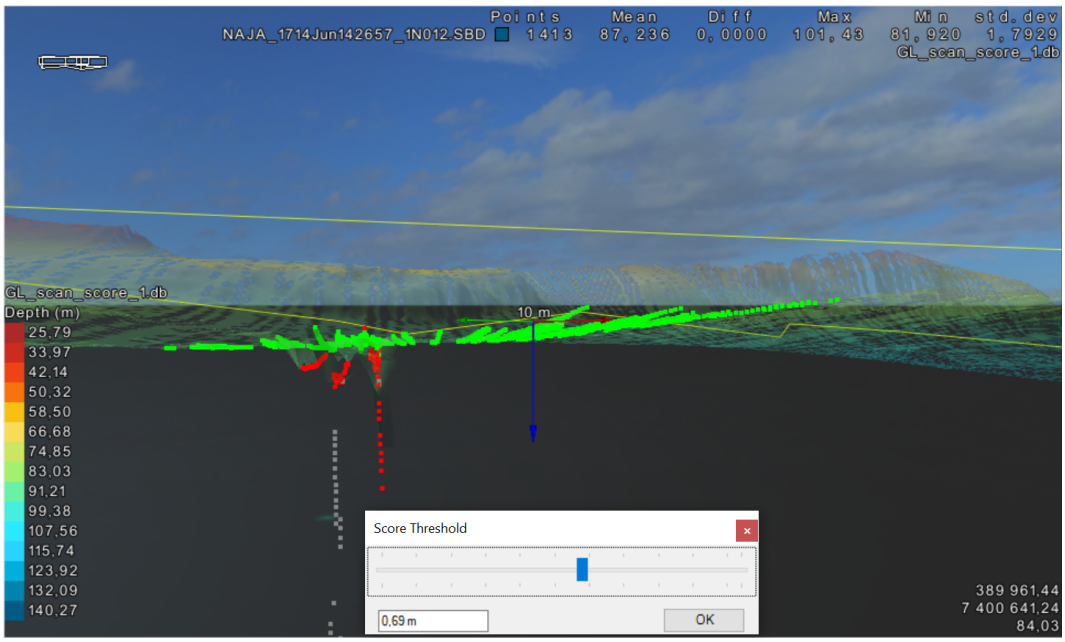
Figur 6.31. Illustration af grov sortering i data

For at fjeren de resterende fejlbehæftede målinger anvendes metoden endnu engang. Denne gang anvendes den dog ikke på hele datasættet, men kun på de områder, hvor der ikke er store variationer i havbunden.



Figur 6.32. Det oprettede polygon, hvor der indenfor polygonen foretages rensning med S-CAN score

Når S-CAN score anvendes oprettes der så en polygon, som ikke medtager de kuperede områder, sådan at en rensning af det resterende kan foretages. Polygonen kan ses på Figur 6.32. Polygonen er konstrueret meget grov, men fordi der er foretaget en grov rensning af data, ses det ikke som nødvendigt at gøre den mere detaljeret. Nu anvendes S-CAN score på dette område for at finsortere i data.



Figur 6.33. 3D model af data med illustration af accepterede og ikke-accepterede dybder

Det vurderes, at en passende grænseværdi er $0,69$ m. På den måde frasorteres noget støj, se eksempel på Figur 6.33, som er beliggende i de mere jævne områder af havbunden.

I og for sig er data ikke renset færdigt, da der stadig kan være nogle outliers beliggende i mere kuperede områder. Undersøges det hvordan standard afvigelserne har ændret sig fra det oprindelige datasæt til efter den første anvendelse af metoden.

Standard afvigelse	Grønland, før	Grønland, efter
<0,10 m	75,08%	75,18%
<0,20 m	84,40%	84,62%
<0,30 m	90,01%	90,15%
<0,40 m	93,18%	93,32%
<0,50 m	95,28%	95,43%
<0,60 m	96,64%	96,78%
<0,70 m	97,60%	97,74%
<0,80 m	98,17%	98,32%
<0,90 m	98,54%	98,68%
<1,00 m	98,80%	98,94%

Tabel 6.5. Procentdel af cellerne, der er under en given standard afvigelse for Grønland efter anvendelse af S-CAN score

Konklusionen er igen den samme som ved anvendelse af S-CAN component. Der er ikke sket så stor effekt i standard afvigelserne, men dette var også forventet på baggrund af en varierende havbund.
Ligesom S-CAN component ses score varianten mere hensigtsmæssig at anvende på de danske data, da den har svært ved at håndtere variationer i havbunden. S-CAN er altså generelt mere anvendelig til flade havbunde eller hvis havbunden varierer jævnt, og ikke har pludselige ændringer inden for et kort horisontalt område som det grønlandske data besidders.

6.3 EC-3D anvendt på dansk og grønlandsk data

I dette afsnit ønskes det at teste, hvordan de to datasæt behandles med EC-3D. Her vil der blive undersøgt de seks metoder til at behandle data, hvilke er følgende:

1. Nearest Neighbour Cleaning
2. Neighbour In Sphere Cleaning
3. Z Summation In Circle Cleaning
4. Distance To Average Surface Cleaning
5. K Nearest Neigbour Cleaning
6. K Nearest Neigbour Cleaning, Z

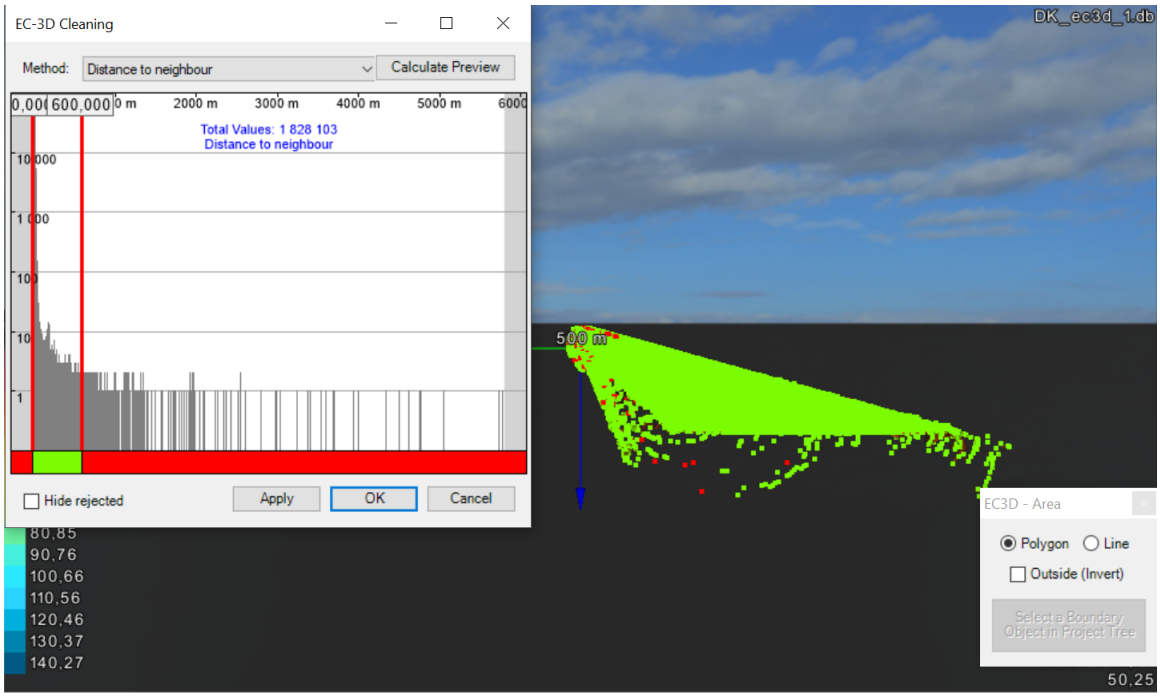
Testen af metoderne vil ske individuelt, og der vil ikke blive foretaget en kombination af de forskellige metoder. Til anvendelse benyttes henholdsvis det danske og det grønlandske datasæt som DTM med en cellestørrelse på 2 m. Testen vurderes både subjektivt og på baggrund af standard afvigelser ligesom ved S-CAN.

6.3.1 Nearest Neighbour

For at anvende metoden Nearest Neighbour skal der angives to grænseværdier som input. Der skal angives en minimums og en maksimal grænseværdi, som skal danne baggrund for hvilke punkter, der skal accepteres, og hvilke der ikke skal accepteres. Disse grænseværdier kan dog ændres løbende i et histogram, hvor det samtidig er illustreret i en 3D model, hvilke målinger der accepteres, og hvilke der ikke accepteres.

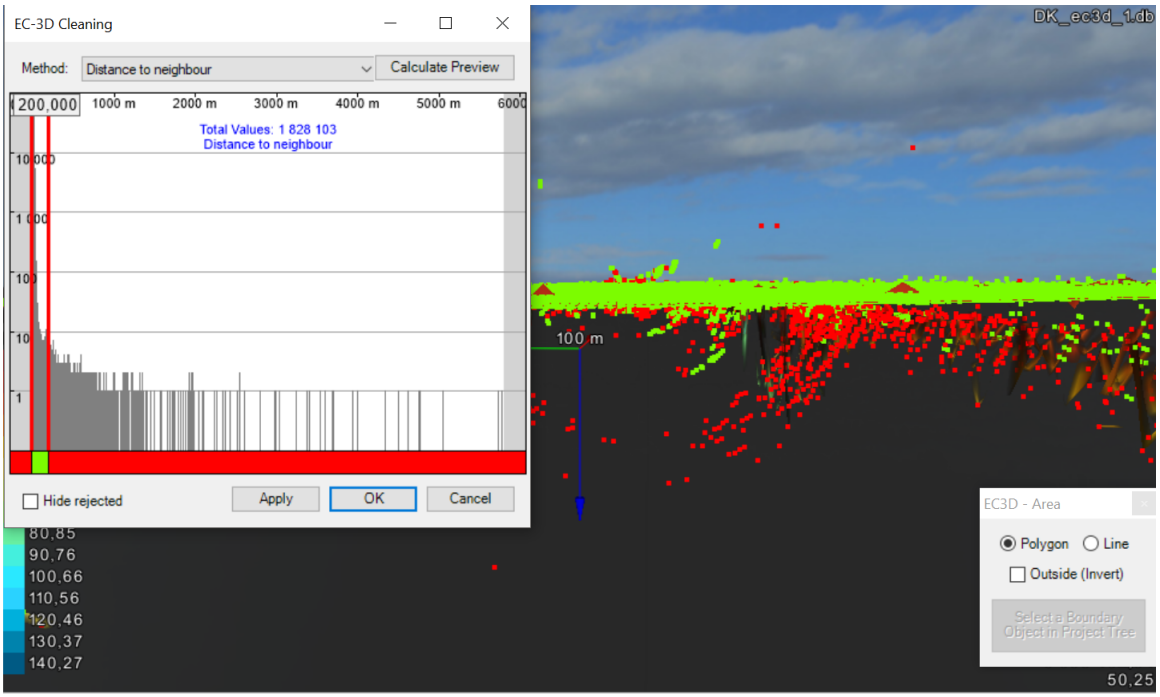
Danmark

Til at teste metoden på det danske datasæt sættes inputværdierne blot som et gæt, og grænseværdierne sættes derfor til henholdsvis 0 m og 600 m.



Figur 6.34. Nearest Neighbour metoden hvor $T_{min} = 0$m og $T_{max} = 600$m

Af Figur 6.34 ses det at disse grænseværdier ikke er passende til at behandle det danske datasæt. Derfor må grænseværdierne ændres, sådan at de er mere passende, for at kunne vurdere, om metoden er hensigtsmæssig at anvende på datasættet.



Figur 6.35. Nearest Neighbour metoden hvor $T_{min} = 0$m og $T_{max} = 200$m

Ved at nedsætte den øvre grænseværdi fra 600 m til 200 m ses det af Figur 6.35, at der sker en forbedring, men at metoden har svært ved at håndtere outliers, der ligger samlet. Det ses af Figur 6.35 at hvis outliers ligger lidt spredt, håndterer metoden det fint, og de markeres røde, men såfremt der ligger to outliers ved siden af hinanden, er de markeret med grøn. Dette skyldes, at metoden beregner en afstand til den nærmeste nabo, hvilken selvfølgelig vil være kort mellem to outliers, der er tæt på hinanden. Derfor går afstanden ikke over grænsen, og målingerne ses derfor som valide. Derfor vil det heller ikke hjælpe at anvende metoden endnu engang, da de igen vil blive markeret grønne. At anvende metoden endnu engang, kan dog muligvis medvirke til, at andre outliers vil blive fjernet, men hele datasættet kan ikke renses færdigt. Derfor vælges det, ikke at forsætte med at anvende metoden. Dette vil selvfølgelig også have en indvirkning på standard afvigelserne.

Standard afvigelse	Danmark, før	Danmark, efter
<0,10 m	91,08%	92,16%
<0,20 m	94,93%	96,19%
<0,30 m	95,87%	97,17%
<0,40 m	96,26%	97,58%
<0,50 m	96,48%	97,80%
<0,60 m	96,62%	97,95%
<0,70 m	96,69%	98,02%
<0,80 m	96,82%	98,18%
<0,90 m	97,08%	98,54%
<1,00 m	97,33%	98,86%

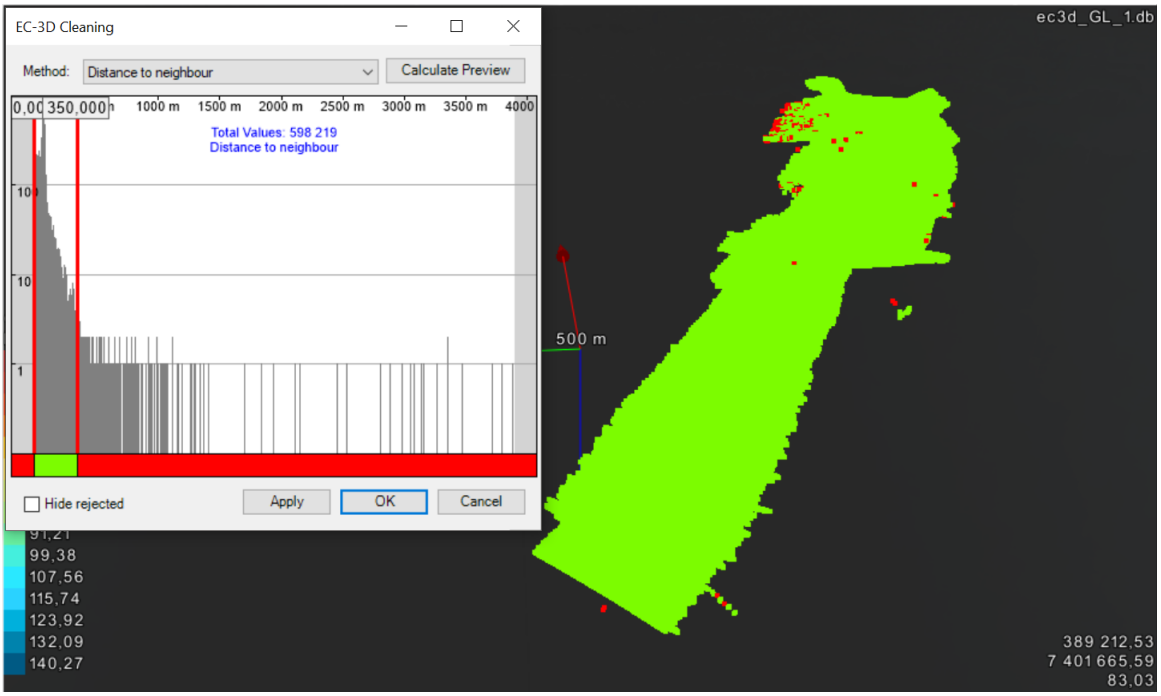
Tabel 6.6. Procentdel af cellerne, der er under en given standard afvigelse for henholdsvis Danmark efter anvendelse af Nearest Neighbour metoden

Som det ses i Tabel 6.6 er der som forventet ikke sket de helt store påvirkninger i standard

afvigelserne. Dette skyldes som skrevet at der er flere outliers, der ikke bliver slettet, da de har en anden outlier som nabo. Derfor ses metoden ikke som anvendelig til at behandle det danske datasæt.

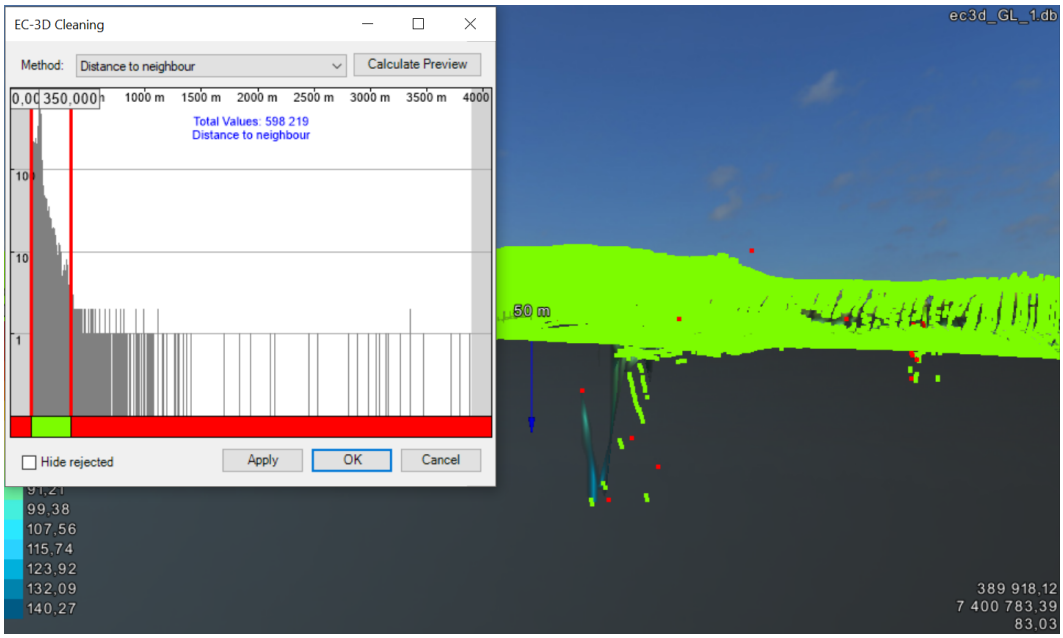
Grønland

Til det grønlandske data anvendes grænseværdierne på henholdsvis 0 m og 350.



Figur 6.36. Nearest Neighbour metoden hvor $T_{min} = 0$m og $T_{max} = 350$m

Af Figur 6.36 ses det at data overvejende er grønt og altså accepteret. Dette skyldes at der ikke er så mange outliers i data som i det danske, og der altså ikke er så mange målinger der skal frasorteres i det grønlandske datasæt. Generelt frasorterer metoden enkelte outliers fint, men ligesom situationen i det danske datasæt har metoden svært ved at frasortere outliers, der er beliggende ved siden af en anden outlier. Dette kan eksempelvis ses på Figur 6.37.



Figur 6.37. Nearest Neighbour metoden hvor $T_{min} = 0$m og $T_{max} = 350$m

Undersøges standard afvigelserne kan det formodes at der er sket en lille forbedring i disse, men det forventes dog at den er begrænset.

Standard afvigelse	**Grønland, før**	**Grønland, efter**
<0,10 m	75,08%	75,08%
<0,20 m	84,40%	84,49%
<0,30 m	90,01%	90,02%
<0,40 m	93,18%	93,20%
<0,50 m	95,28%	95,31%
<0,60 m	96,64%	96,67%
<0,70 m	97,60%	97,63%
<0,80 m	98,17%	98,21%
<0,90 m	98,54%	98,58%
<1,00 m	98,80%	98,85%

Tabel 6.7. Procentdel af cellerne, der er under en given standard afvigelse for Grønland efter anvendelse af Nearest Neighbour metoden

Som formodet er der ikke sket de store ændringer i forhold til standard afvigelser hvilket også var forventet. Metoden er mest hensigtsmæssig at anvende til at finde enkelte outliers, der ikke er beliggende ved siden af andre. Dette forklarer også hvorfor metoden ikke ses som anvendelig til at behandle hele datasæt hverken for Danmark eller Grønland.

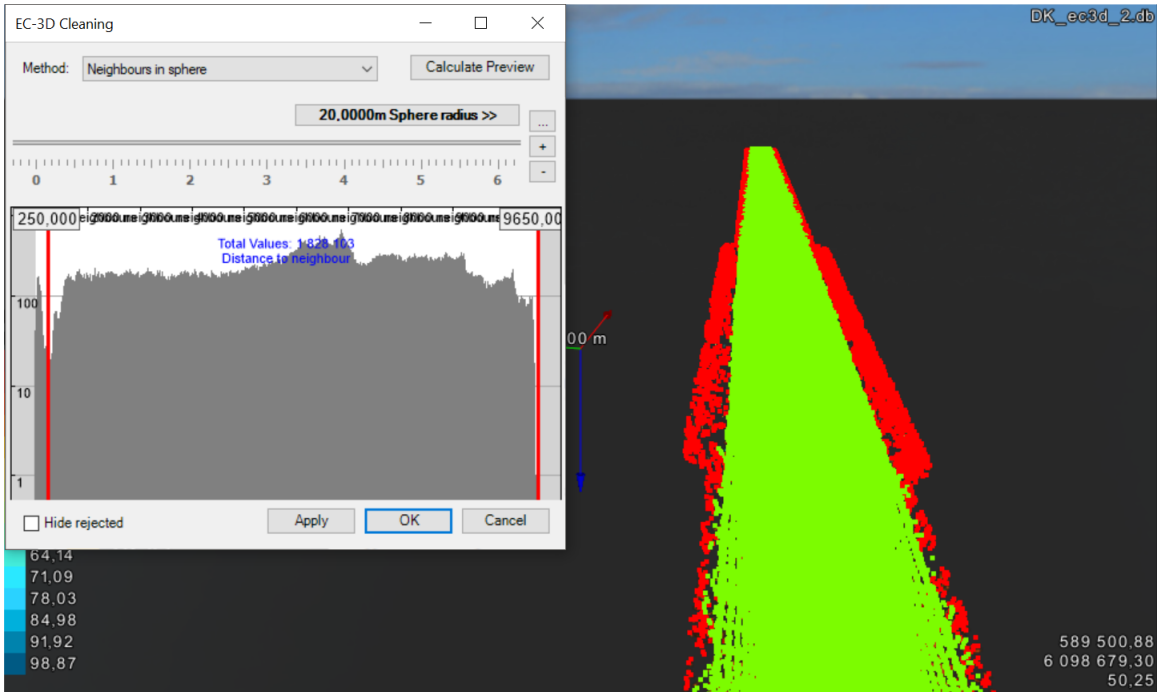
6.3.2 Neighbour In Sphere

Metoden Neighbour I Spere kræver en række input for at kunne anvendes. Først skal der anvendes en radius for den valgte sfære og derefter skal der vælges en minimums og en maksimal grænseværdi, der ligesom ved Nearest Neighbour skal danne baggrund for accepterede og ikke

accepterede målinger. Grænseværdierne kan ændres i et histogram, men hvis sfærens radius skal ændres, skal der beregnes på ny.

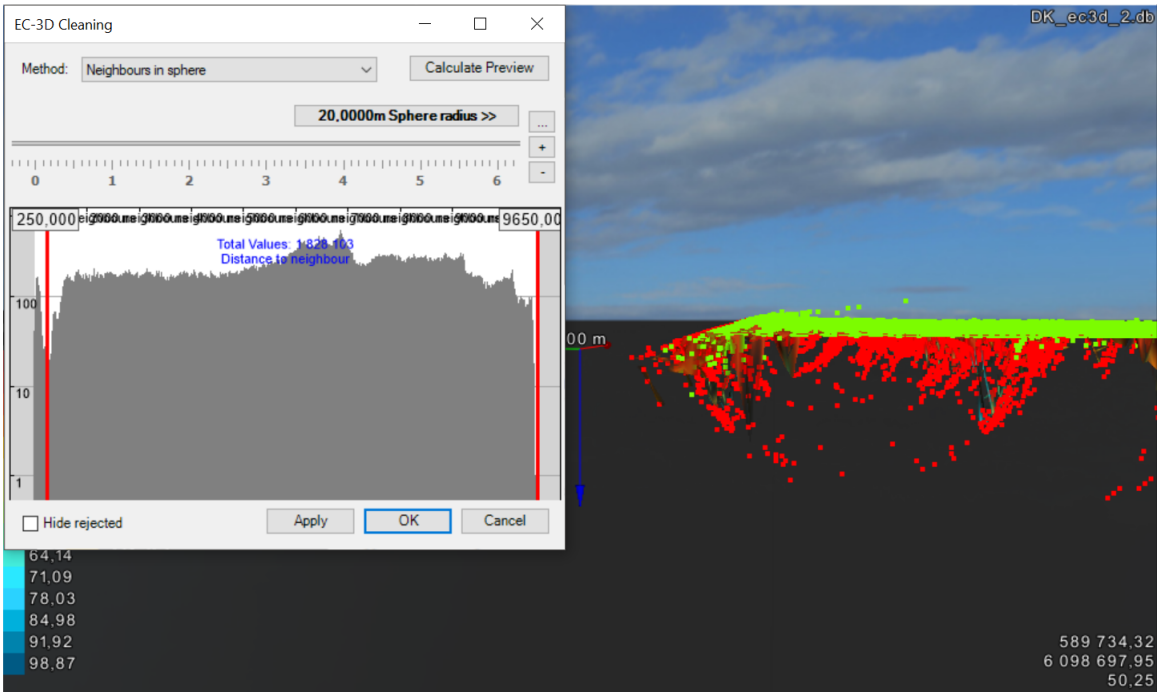
Danmark

For at anvende metoden indtastes startværdier til denne. Det vælges, at anvende en radius på 20 m, samt en nedre grænse på 250. Den øvre grænse sættes til det maksimale antal naboer inden for sfæren.



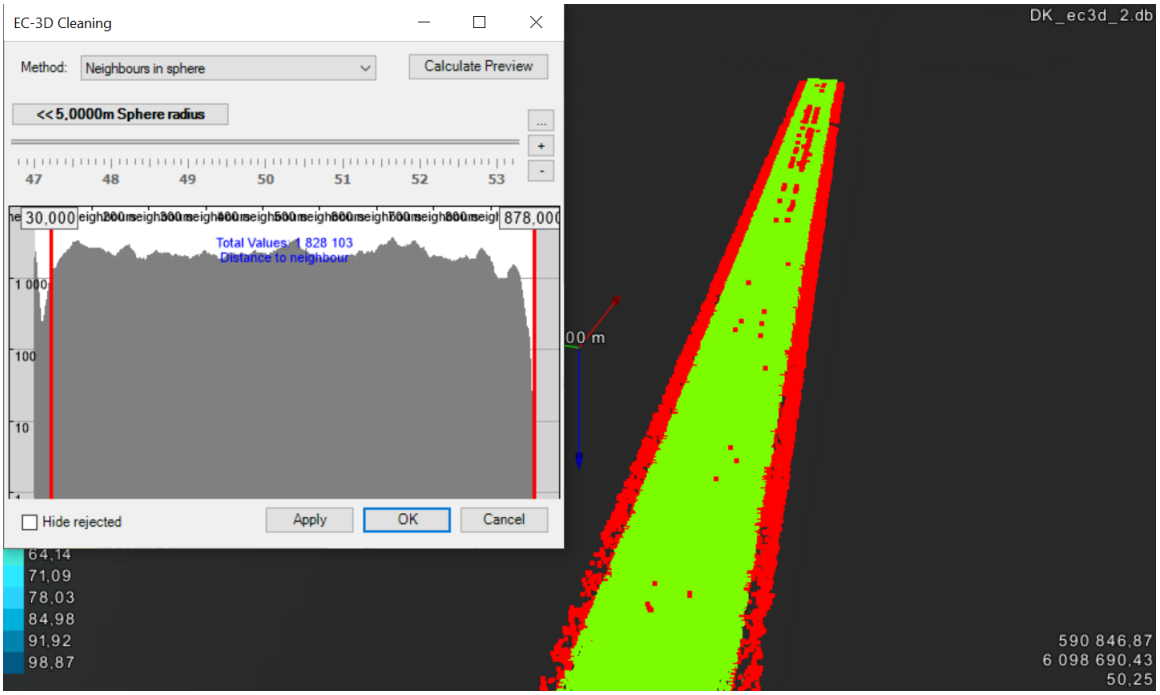
Figur 6.38. Neighbour In Sphere metoden hvor $r = 20$ m, $T_{min} = 250$ naboer og $T_{max} = 9650$ naboer

Det ses af Figur 6.38 at metoden til dels fungerer til dette datasæt. Det støj der er i kanten af opmålingsområdet bliver frasorteret, og tilsvarende med de målinger der er beliggende under den målte havbund. Dette er dog ikke tilfældet med de målinger, der beliggende over den målte havbund, hvilket ses på Figur 6.39.



Figur 6.39. Neighbour In Sphere metoden hvor $r = 20$ m, $T_{min} = 250$ naboer og $T_{max} = 9650$ naboer

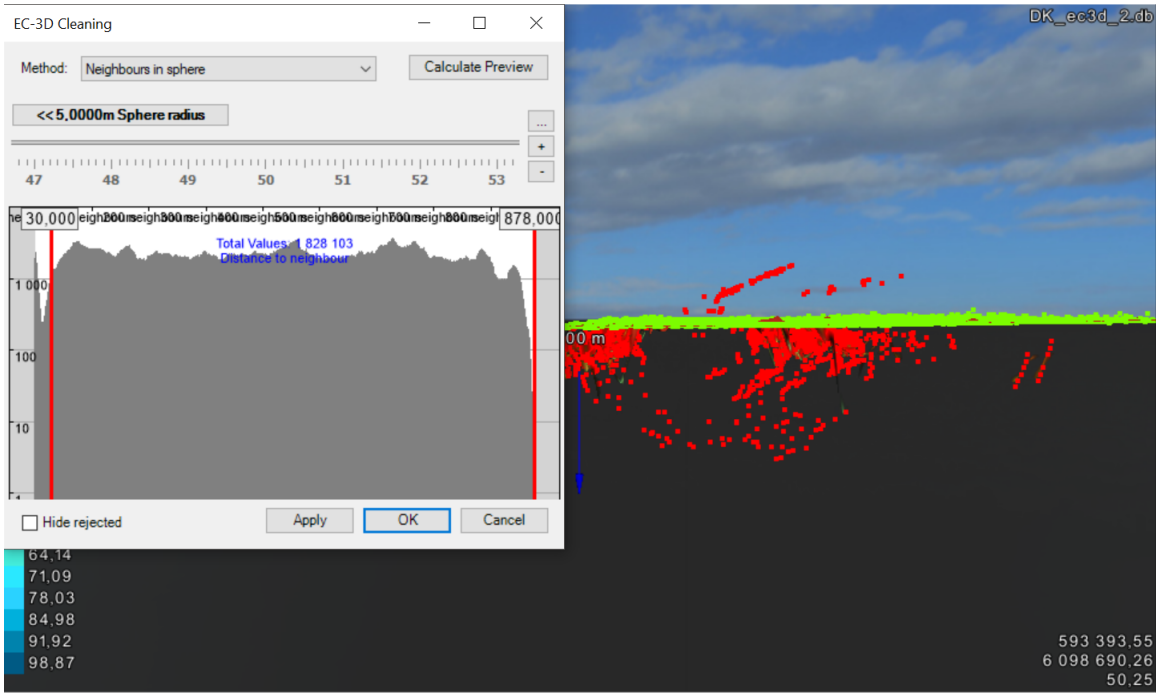
Det vurderes, at metoden ikke renser data tilfredsstillende. Det prøves derfor at ændre radius i sfæren fra 20 m til 5 m. Radiussen nedsættes, idet der forhåbentlig kan medvirke til bedre rensning, da der skabes et mindre område som der renses for, og det på den måde repræsenterer de enkelte områder bedre.



Figur 6.40. Neighbour In Sphere metoden hvor $r = 5$ m, $T_{min} = 30$ naboer og $T_{max} = 878$ naboer

På Figur 6.40 ses det at metoden igen pænt frasorterer de outliers der er i kanten af

opmålingsområdet, men det ses også at der er nogle målinger markeret med rød i midten. Kigges der på Figur 6.41, ses det, at de røde punkter netop er en del af de outliers, der er beliggende over havbunden. Derfor er metoden mere anvendelig til at rense data når radiussen for sfæren nedsættes.



Figur 6.41. Neighbour In Sphere metoden hvor $r = 5$ m, $T_{min} = 30$ naboer og $T_{max} = 878$ naboer

Denne metode er altså mere anvendelig til denne type af data end Nearest Neighbour. Undersøges standard afvigelserne ses resultatet i Tabel 6.8.

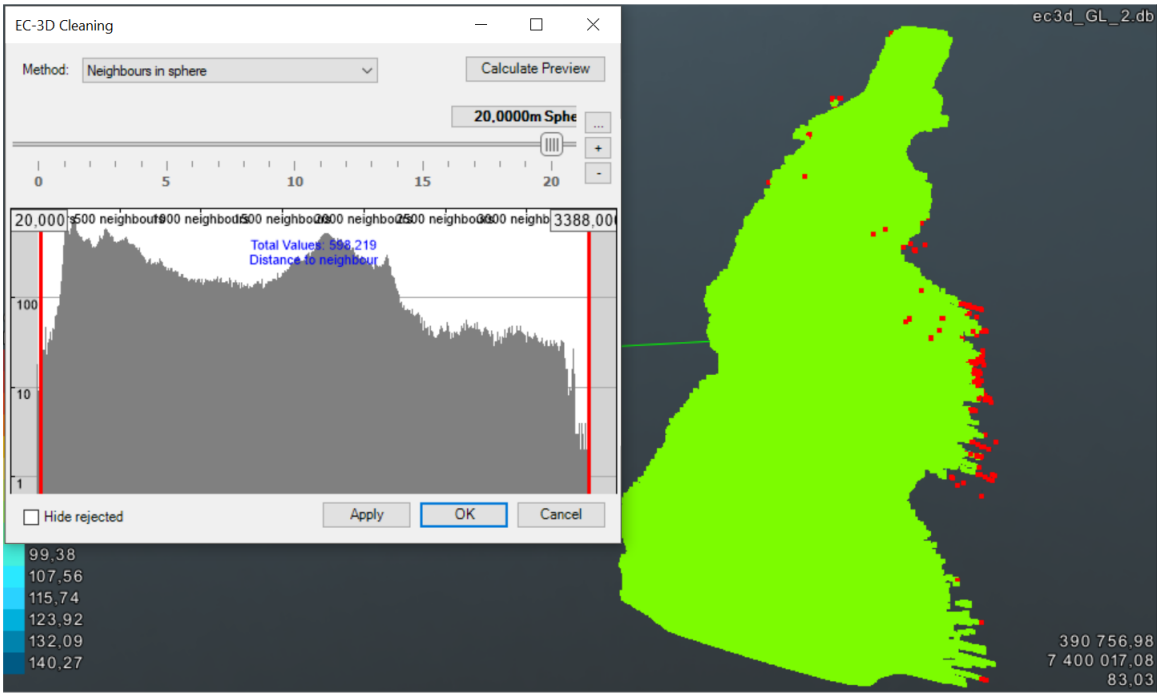
Standard afvigelse	**Danmark, før**	**Danmark, efter**
<0,10 m	91,08%	94,31%
<0,20 m	94,93%	98,13%
<0,30 m	95,87%	99,08%
<0,40 m	96,26%	99,51%
<0,50 m	96,48%	99,74%
<0,60 m	96,62%	99,88%
<0,70 m	96,69%	99,92%
<0,80 m	96,82%	99,94%
<0,90 m	97,08%	99,95%
<1,00 m	97,33%	99,96%

Tabel 6.8. Procentdel af cellerne, der er under en given standard afvigelse for henholdsvis Danmark efter anvendelse af Neighbour In Sphere metoden

Af Tabel 6.8 kan de ses, at standard afvigelserne for cellerne er blevet mindre, og også mindre end ved anvendelse af Nearest Neighbour metoden, hvilket er i overensstemmelse med det visuelle. Metoden ses altså som anvendelig til denne type data.

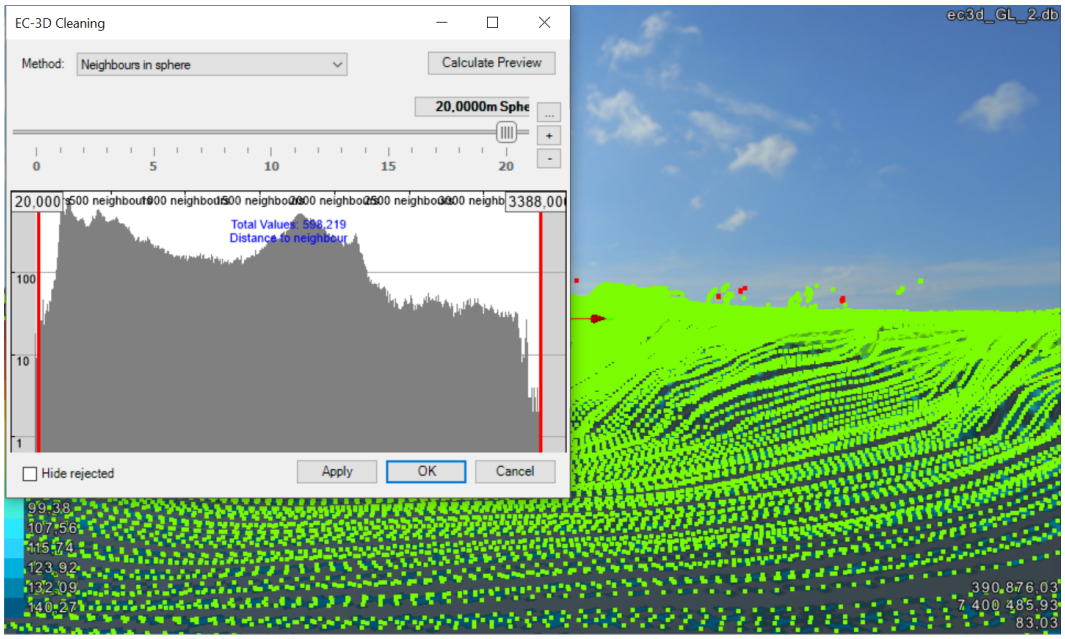
Grønland

Til metoden anvendes der en radius på 20 m ligesom ved det danske. Den øvre grænse sættes igen til at være det maksimale antal af naboer inden for sfæren og den nedre grænse sættes til 20. Sættes den nedre grænse højere, vil korrekt data begynde ikke at blive accepteret.



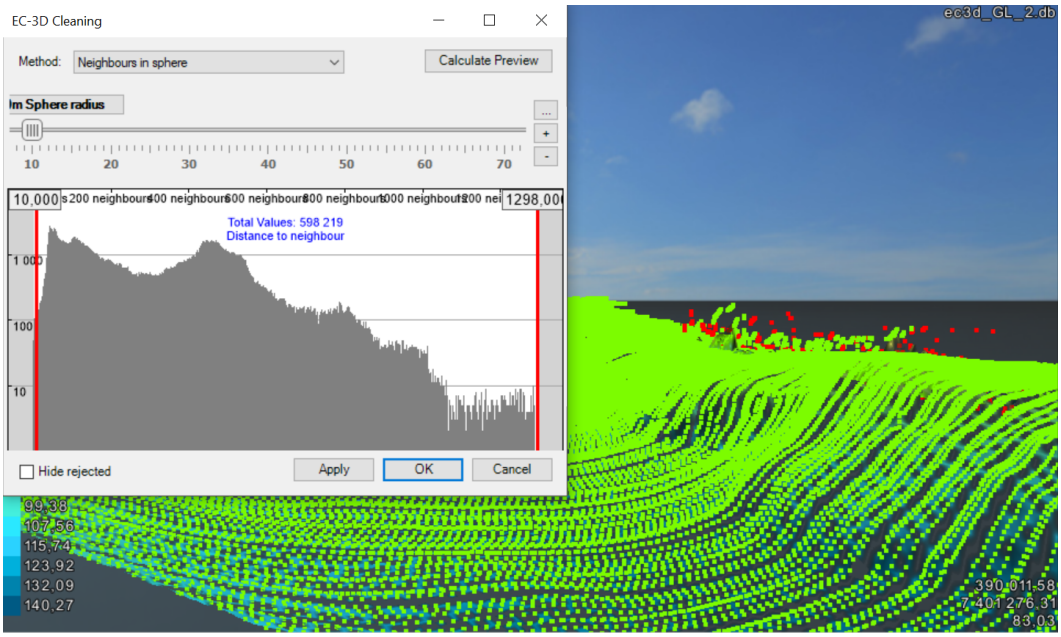
Figur 6.42. Neighbour In Sphere metoden hvor $r = 20$ m, $T_{min} = 20$ naboer og $T_{max} = 3388$ naboer

Af Figur 6.42 ses det, at data igen umiddelbart er meget grøn. Der er nogle enkelte røde punkter, hvilke alle er formentlig fejlbehæftede målinger som skal slettes. Der er dog stadig en række fejlbehæftede målinger, som der ikke er blevet medtaget i målinger til sletning, se eksempel på Figur 6.43.



Figur 6.43. Neighbour In Sphere metoden hvor $r = 20$ m, $T_{min} = 20$ naboer og $T_{max} = 3388$ naboer

Det prøves derfor at ændre radius, og denne ændres til 10 m med en nedre grænse på 10 naboer. Det giver resultatet på Figur 6.44.



Figur 6.44. Neighbour In Sphere metoden hvor $r = 10$ m, $T_{min} = 10$ naboer og $T_{max} = 1298$ naboer

Det hjælper lidt at nedsætte radius, men nedsættes radius for meget vil flere "korrekte" målinger blive slettet, og derfor sættes den ikke længere ned end 10 m. Fra 3D modellen virker det til, at data er blevet renset lidt bedre ved at nedsætte radius, men det kan ikke vurderes, hvorvidt data er renset tilfredsstillende. Dette vil formentlig også blive reflekteret i standard afvigelserne.

Standard afvigelse	Grønland, før	Grønland, efter
<0,10 m	75,08%	75,09%
<0,20 m	84,40%	84,52%
<0,30 m	90,01%	90,04%
<0,40 m	93,18%	93,22%
<0,50 m	95,28%	95,32%
<0,60 m	96,64%	96,68%
<0,70 m	97,60%	97,65%
<0,80 m	98,17%	98,22%
<0,90 m	98,54%	98,59%
<1,00 m	98,80%	98,85%

Tabel 6.9. Procentdel af cellerne, der er under en given standard afvigelse for Grønland efter anvendelse af Neighbour In Sphere metoden

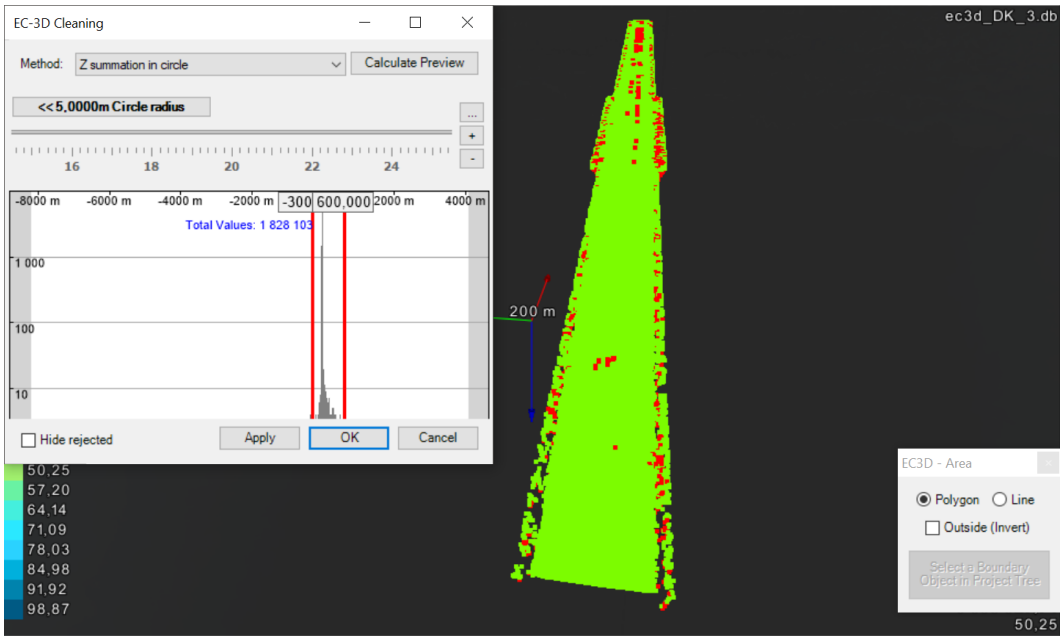
Denne metode ses heller ikke som hensigtsmæssig at anvende alene til at rense grønlandsk data.

6.3.3 Z Summation In Circle

Som input til metoden Z Summation In Circle kræves der en radius til den cirkel som skal dannes, samt to grænseværdier ligesom i de andre metoder. Ligesom ved Neighbour in Sphere kan grænseværdierne løbende ændres, men vil radiussen på cirklen ændres, skal der beregnes på ny.

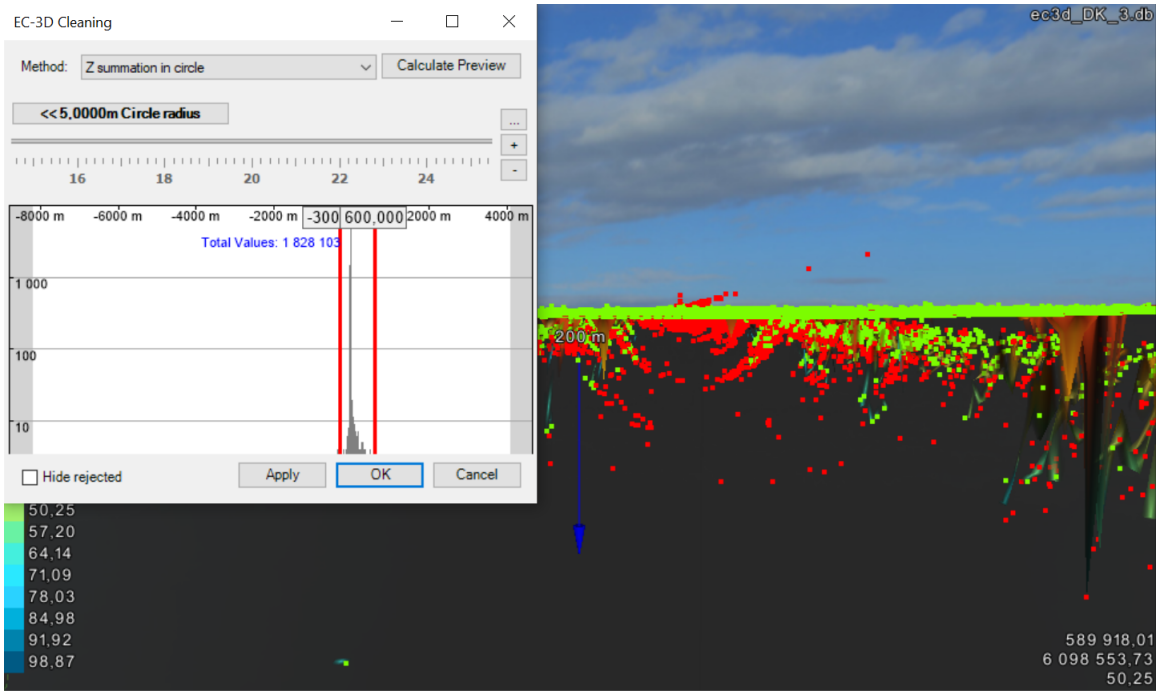
Danmark

Til det danske datasæt sættes radius på cirklen til 5 m til at starte med. Herefter vælges grænseværdierne henholdsvis -300 m og 600 m.



Figur 6.45. Z Summation In Circle metoden hvor $r = 20$ m, $T_{min} = -300$ m og $T_{max} = 600$ m

Af Figur 6.45 ses det, at metoden umiddelbart ikke ser særlig hensigtsmæssig ud til dette datasæt. Ses datasættet fra siden af, ser det ud som på Figur 6.46.



Figur 6.46. Z Summation In Circle metoden hvor $r = 20$ m, $T_{min} = -300$ m og $T_{max} = 600$ m

Af Figur 6.46 ses det, at metoden beholder mange målinger i kanten af opmålingsområdet som ikke skal beholdes. Dette kan skyldes, at de er beliggende i nogenlunde samme dybde som havbunden, og derfor når der summeres indenfor cirklen ikke påvirker det samlede resultat så meget. Metoden ses ikke som anvendelig for behandling af denne type data. Det er yderligere forsøgt at ændre på radius, men konklusionen for datasættet forbliver den samme.
Hvad angår standard afvigelserne, så forventes det, at de bliver en smule lavere, men ikke i et acceptabelt niveau baseret på 3D plottet.

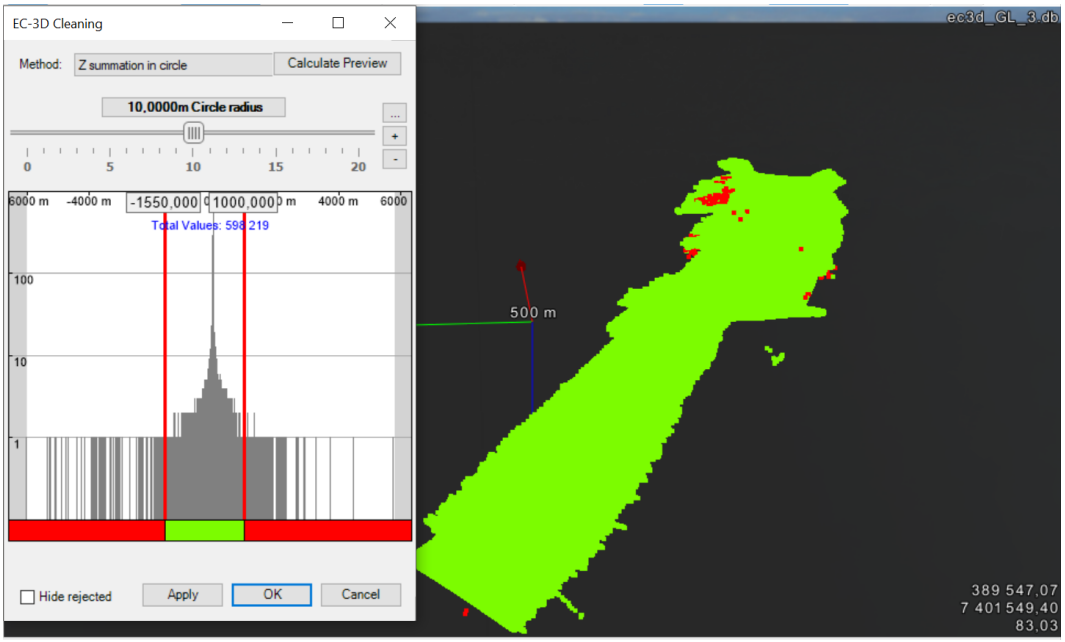
Standard afvigelse	Danmark, før	Danmark, efter
<0,10 m	91,08%	92,55%
<0,20 m	94,93%	96,49%
<0,30 m	95,87%	97,45%
<0,40 m	96,26%	97,86%
<0,50 m	96,48%	98,08%
<0,60 m	96,62%	98,22%
<0,70 m	96,69%	98,28%
<0,80 m	96,82%	98,40%
<0,90 m	97,08%	98,68%
<1,00 m	97,33%	98,93%

Tabel 6.10. Procentdel af cellerne, der er under en given standard afvigelse for henholdsvis Danmark efter anvendelse af Z Sumamtion In Circle metoden

Af Tabel 6.10 ses et resultat som var forventet på baggrund af 3D modellen, og metoden ses ikke som anvendelig til et datasæt med en flad havbund med meget støj.

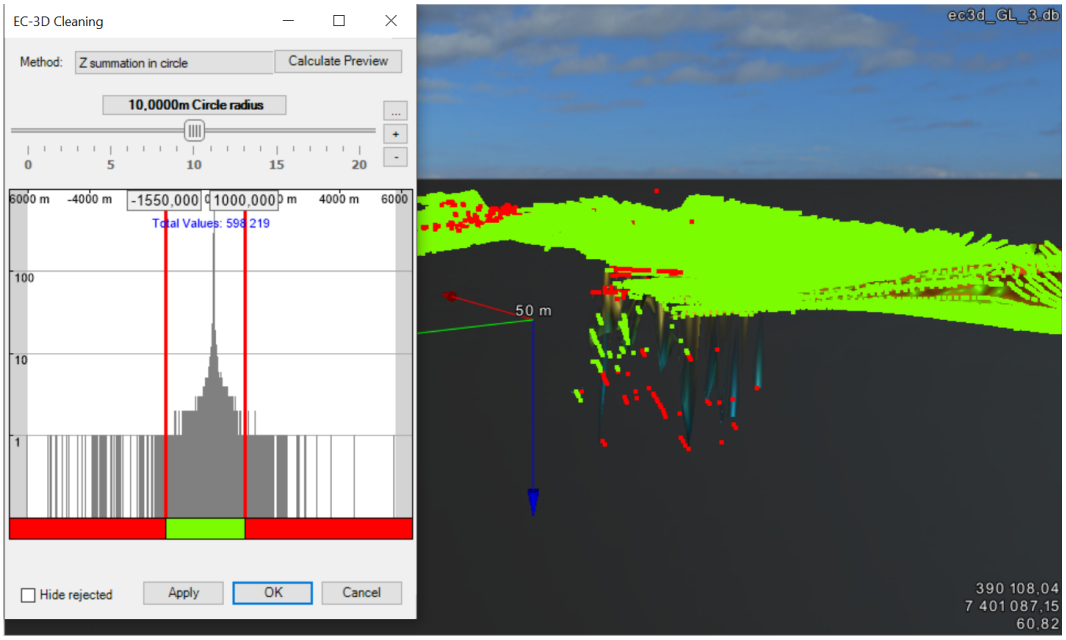
Grønland

Til anvendelse af metoden på det grønlandske datasæt sættes der en raduis på cirklen til 10 m. Grænseværdierne sættes til henholdsvis -1550 m og 1000 m.



Figur 6.47. Z Summation In Circle metoden hvor $r = 10$ m, $T_{min} = -1550$ m og $T_{max} = 1000$ m

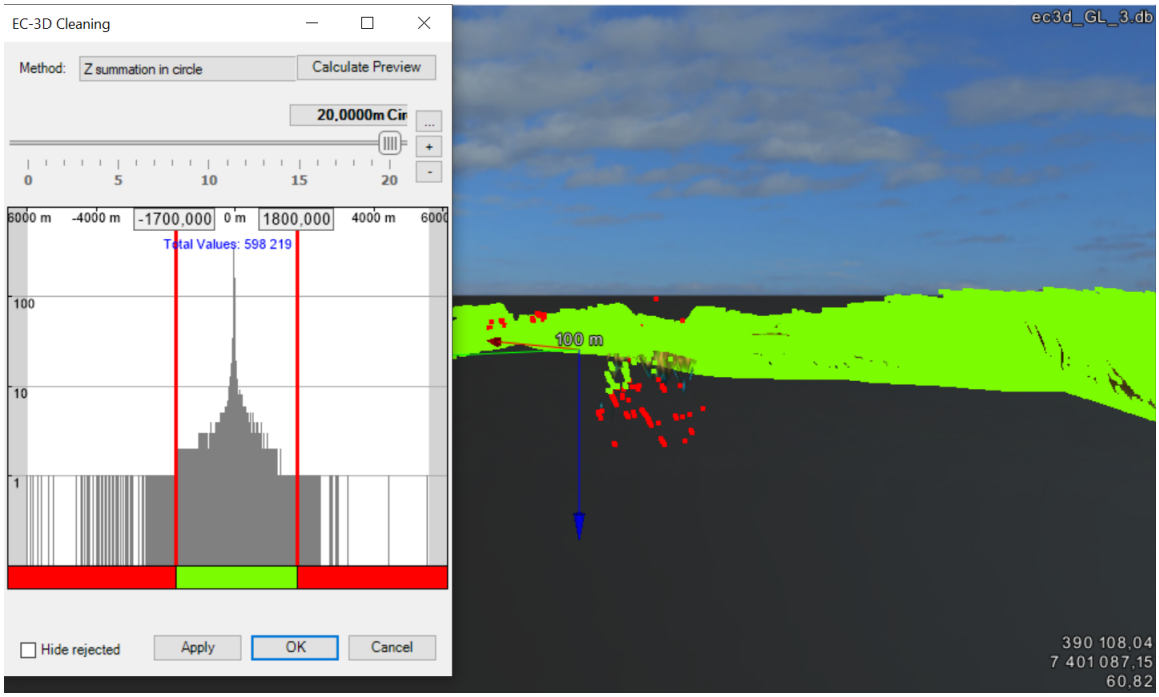
Igen er data overvejende grønt, men der er nogle røde markeringer områder både over og under det der formodes at være den korrekte havbund. På Figur 6.49 ses det, at der er en række outliers, der ikke er markeret røde og derved accepteres selvom de ikke skal.



Figur 6.48. Z Summation In Circle metoden hvor $r = 10$ m, $T_{min} = -1550$ m og $T_{max} = 1000$ m

På baggrund af at flere outliers markeres til at blive beholdt prøves det at justere på nogle

parametre, og cirklen radius sættes til 20 m og grænseværdierne sættes til -1700 m og 1800 m. Igen vurderes dette ikke tilstrækkeligt baseret på eksempelvis tilfældet illustreret på Figur 6.49.



Figur 6.49. Z Summation In Circle metoden hvor $r = 20$ m, $T_{min} = -1700$ m og $T_{max} = 1800$ m

Såfremt alle outliers skal markeres røde vil en række "korrekte" målinger blive markeret rød. Dette skyldes at metoden summerer alle afstande indenfor en given radius, og når havbunden varierer, vil den summerede værdi også blive høj, og målingerne ses derfor ikke som valide.
Igen formodes det ikke at der sker store ændringer i forbindelse med standard afvigelserne.

Standard afvigelse	**Grønland, før**	**Grønland, efter**
<0,10 m	75,08%	75,10%
<0,20 m	84,40%	84,51%
<0,30 m	90,01%	90,03%
<0,40 m	93,18%	93,20%
<0,50 m	95,28%	95,30%
<0,60 m	96,64%	96,67%
<0,70 m	97,60%	97,63%
<0,80 m	98,17%	98,20%
<0,90 m	98,54%	98,57%
<1,00 m	98,80%	98,83%

Tabel 6.11. Procentdel af cellerne, der er under en given standard afvigelse for Grønland efter anvendelse af Z Summation In Circle metoden

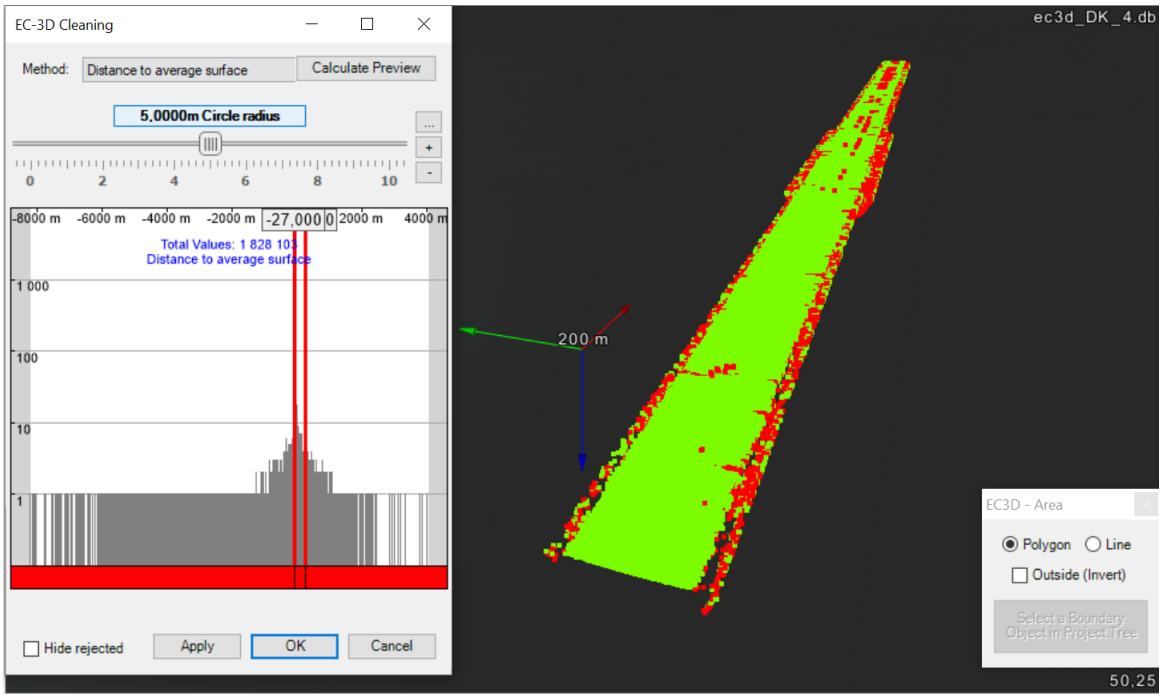
I Tabel 6.11 ses det som forventet at der ikke er sket store ændringer. Der forventes heller ikke at alle procenter vil være over 99, idet standard afvigelserne for det grønlandske data er noget højere end for de danske baseret på den varierende havbund.

6.3.4 Distance To Average Surface

Også til Distance To Average Surface skal der anvendes en radius og to grænseværdier som input. Radius vil så angive indenfor hvor stort et område, der skal anvendes målinger til at danne en overflade, og grænseværdierne angiver hvor meget målinger må afvige fra denne overflade.

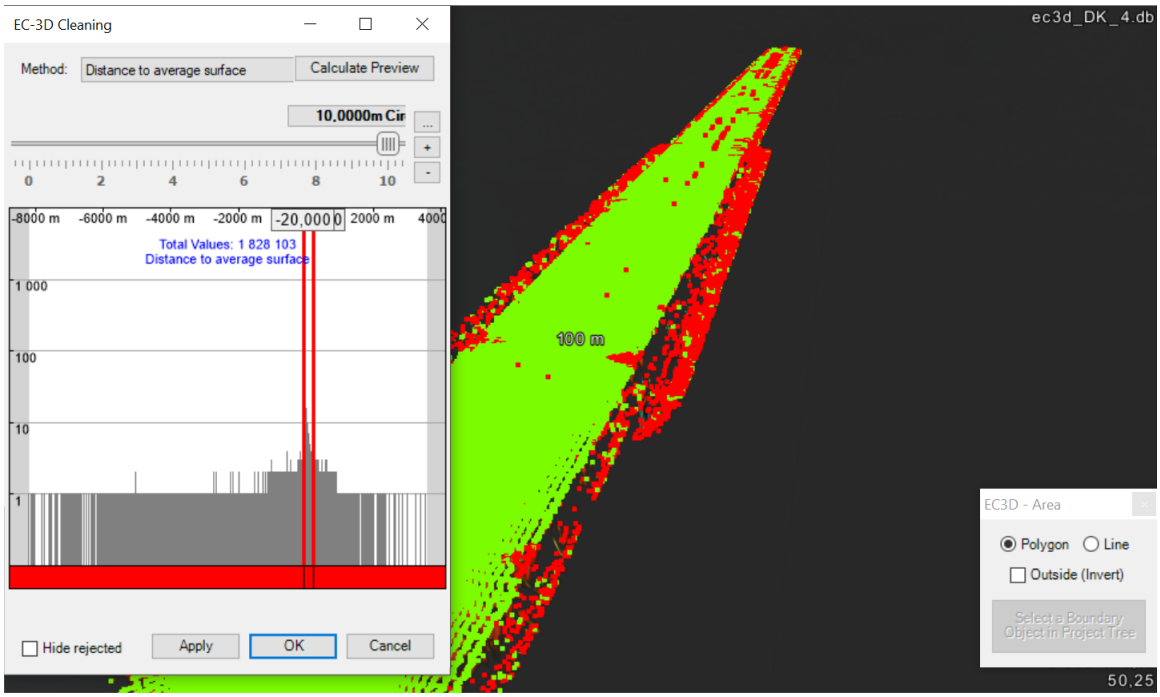
Danmark

Til det danske datasæt anvendes der igen en radius på 5 m. De to grænseværdier sættes efter at der er kommet en illustration af data og sættes til henholdsvis -27 m og 270 m, da det vurderes at dette er de bedst passende til datasættet. Af Figur 6.50 ses det, hvordan metoden ikke umiddelbart ses som anvendelig til denne type data.



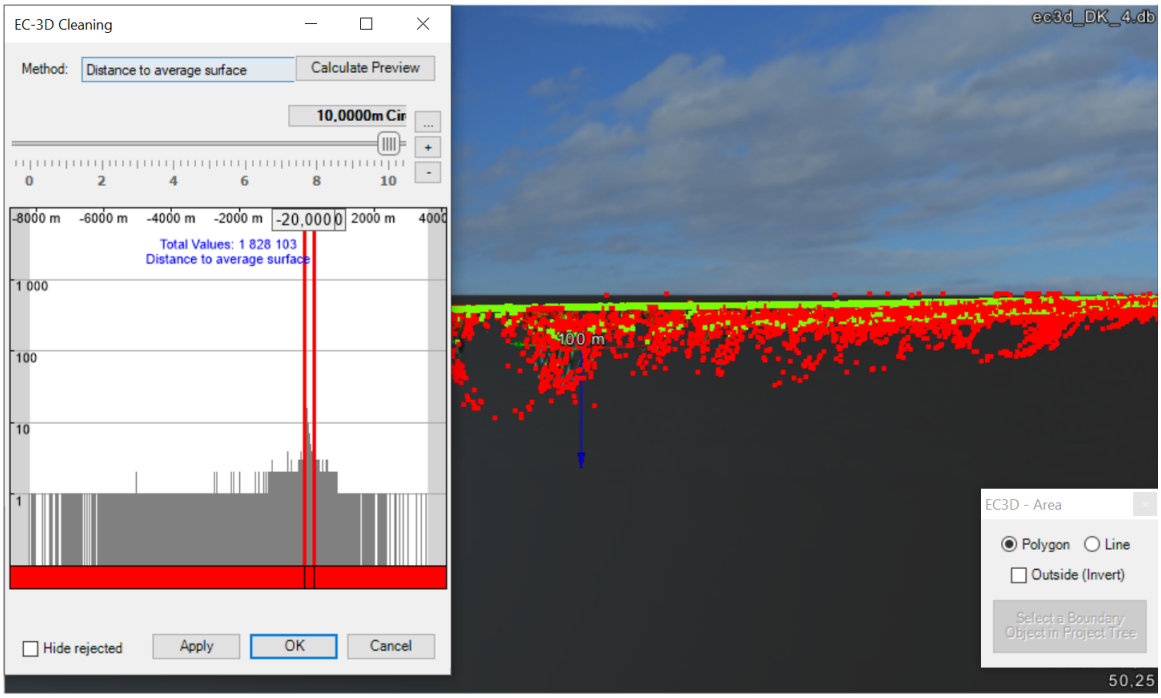
Figur 6.50. Disance To Average Surface hvor $r = 5$ m, $T_{min} = -27$ m og $T_{max} = 270$ m

Flere outliers er markeret til at blive beholdt og korrekt data er så småt begyndt at blive markeret til at skulle slettes. Det kan ske, at dette vil ændre sig ved at ændre radius, og derfor prøves det, at ændre radius til 10m.



Figur 6.51. Disance To Average Surface hvor $r = 10$ m, $T_{min} = -20$ m og $T_{max} = 230$ m

Af Figur 6.51 ses det hvordan der er sket en lille bedring i data, men at der stadig er nogle outliers der er markeret til at skulle beholdes. Roteres datasættet en smule ses outputtet på Figur 6.52, hvor det ses, at der dannes en form for en linje med de outliers, der skal beholdes. Det tyder altså på, at denne række af målinger er i samme niveau som havbunden, og derfor ikke varierer så meget i dybden i forhold til den "korrekte" havbund.



Figur 6.52. Disance To Average Surface hvor $r = 10$ m, $T_{min} = -20$ m og $T_{max} = 230$ m

Metoden ses i sig selv ikke som anvendelig for at behandle den type data, som det danske datasæt er, men kunne være anvendelig i kombination med en anden metode.

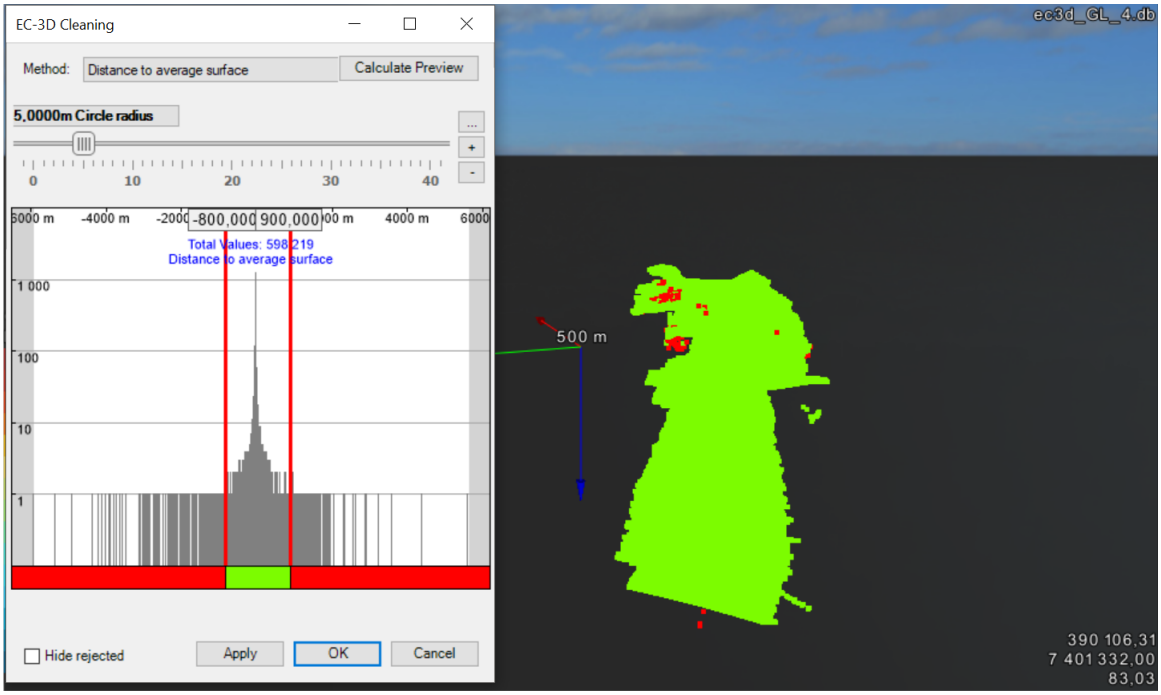
Standard afvigelse	Danmark, før	Danmark, efter
<0,10 m	91,08%	94,50%
<0,20 m	94,93%	98,25%
<0,30 m	95,87%	99,15%
<0,40 m	96,26%	99,52%
<0,50 m	96,48%	99,69%
<0,60 m	96,62%	99,78%
<0,70 m	96,69%	99,81%
<0,80 m	96,82%	99,84%
<0,90 m	97,08%	99,91%
<1,00 m	97,33%	99,97%

Tabel 6.12. Procentdel af cellerne, der er under en given standard afvigelse for henholdsvis Danmark efter anvendelse af Distance To Average Surface metoden

Undersøges standard afvigelserne ses det i Tabel 6.12, at det egentlig ser fint ud, men at der selvfølgelig er plads til forbedringer. Metoden i sig selv renser altså umiddelbart ikke danske data optimalt, men kunne være god i kombination med en anden metode.

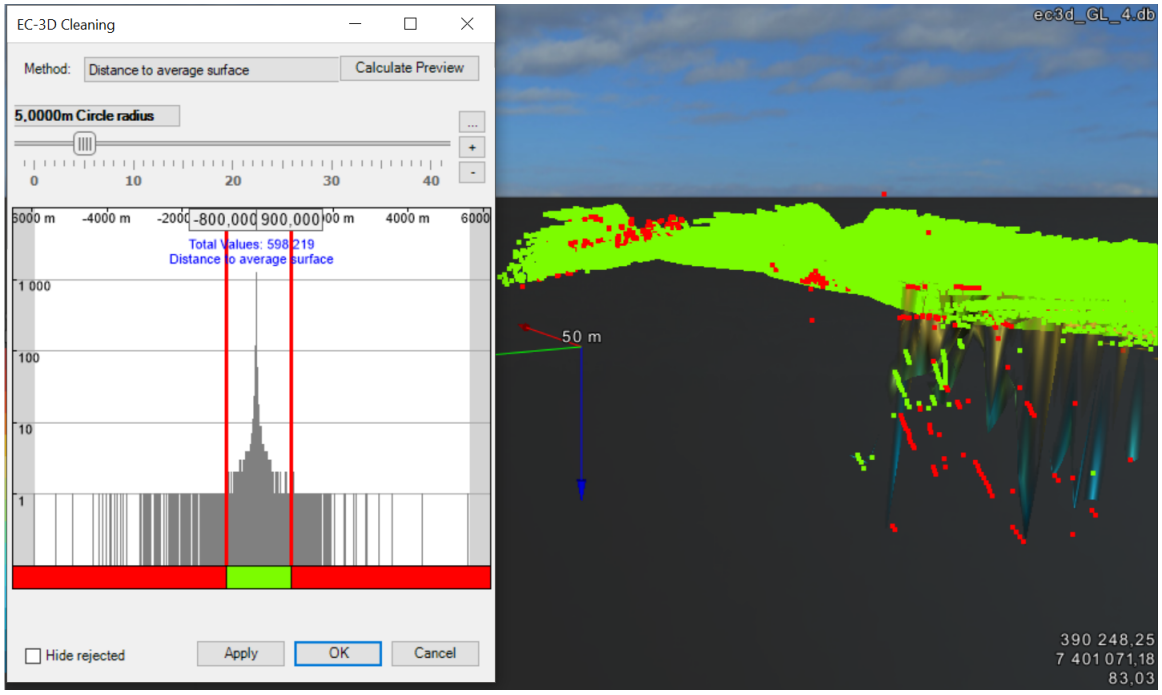
Grønland

Til denne metode anvendes der en radius på 5 m. Grunde til dette er, at der på den måde vil blive dannet flere mindre overflader, som hver især kan være mere repræsentativ for de enkelte områder. Grænseværdierne sættes til henholdsvis -800 m og 900 m.



Figur 6.53. Disance To Average Surface hvor $r = 5$ m, $T_{min} = -800$ m og $T_{max} = 900$ m

Igen er datasættet meget grønt og undersøges det nærmere, ses det, at mange outliers faktisk er markeret med rød, se eksempel på Figur 6.54. Der er dog stadig nogle, der ikke er markeret med rød, men umiddelbart virker det til at metoden har medtaget en del outliers.



Figur 6.54. Disance To Average Surface hvor $r = 5$ m, $T_{min} = -800$ m og $T_{max} = 900$ m

Metoden anvendes med disse parametre, da det vurderes, at det er det bedste resultat med denne metode. Det giver et overblik over standard afvigelserne som vist i Tabel 6.13.

Standard afvigelse	**Grønland, før**	**Grønland, efter**
<0,10 m	75,08%	75,13%
<0,20 m	84,40%	84,55%
<0,30 m	90,01%	90,07%
<0,40 m	93,18%	93,24%
<0,50 m	95,28%	95,35%
<0,60 m	96,64%	96,71%
<0,70 m	97,60%	97,67%
<0,80 m	98,17%	98,25%
<0,90 m	98,54%	98,62%
<1,00 m	98,80%	98,89%

Tabel 6.13. Procentdel af cellerne, der er under en given standard afvigelse for Grønland efter anvendelse af Distance To Average Surface metoden

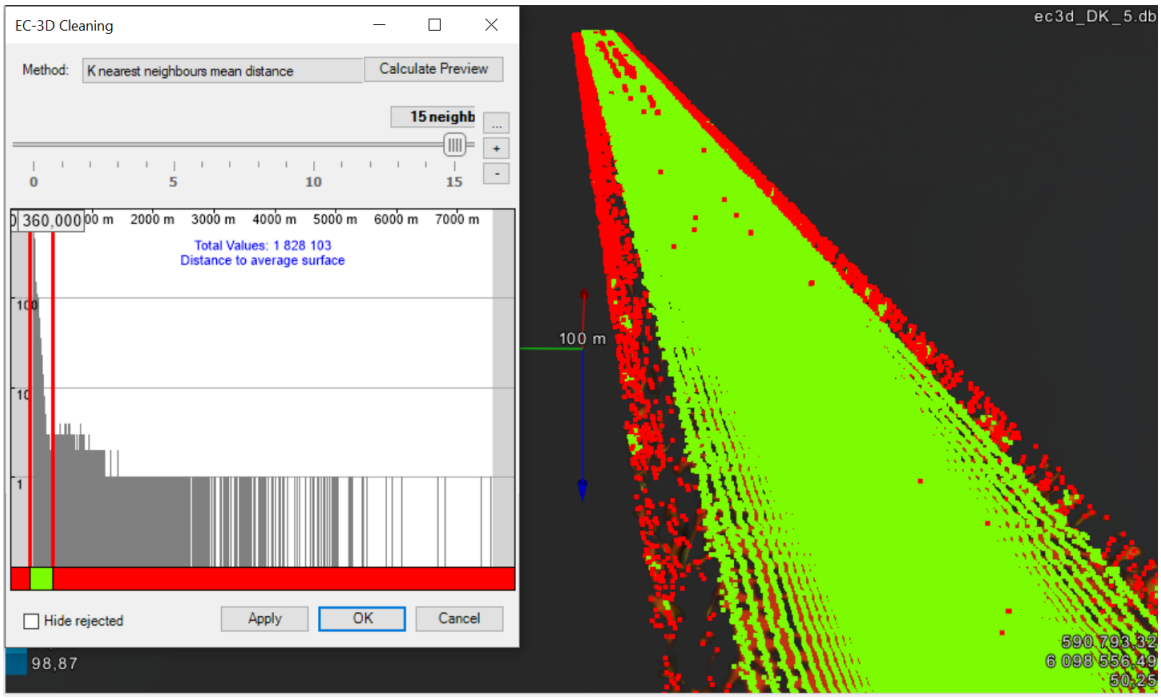
Baseret på standard afvigelser ligner det ikke, at der er sket så mange ændringer, men i det grafiske kan der ses en forskel. Metoden ses derfor til dels som anvendelig i det grønlandske datasæt, men vil ligesom så mange af de andre metoder, nok være bedst at anvende sammen med en anden metode.

6.3.5 K Nearest Neighbour

Til denne metode skal der indtastes input værdier, der angiver hvor mange naboer, der skal medtages i beregningerne samt grænseværdierne for hvor stor en afvigelse der må være fra den gennemsnitlige afstand.

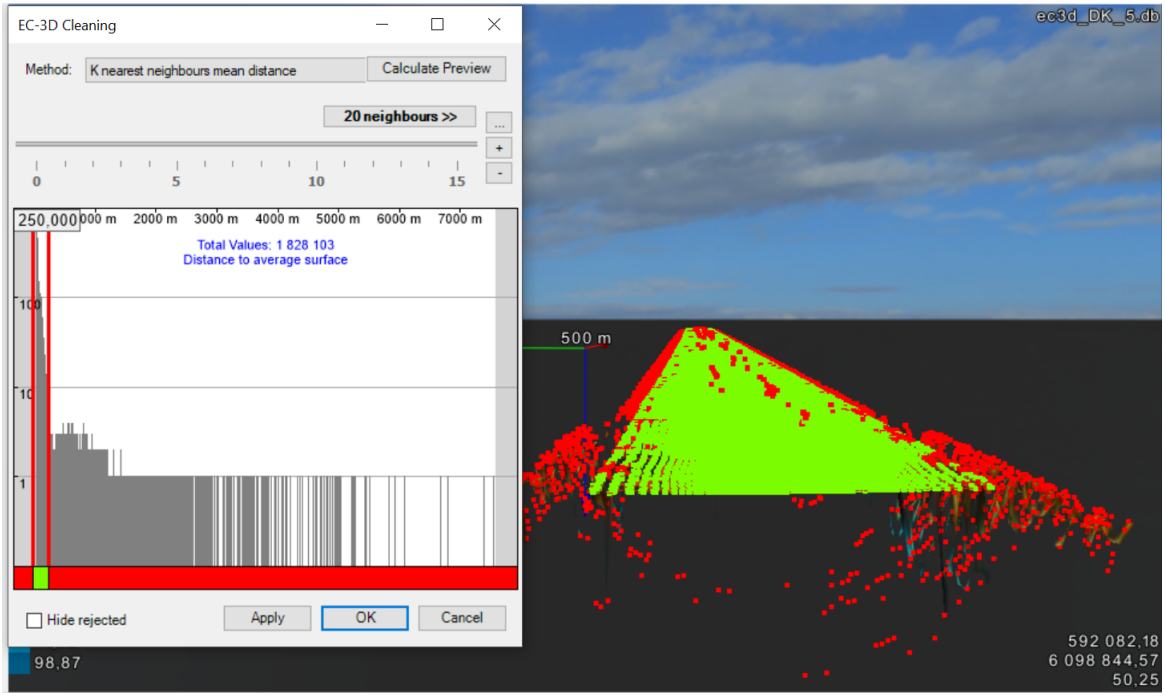
Danmark

Til at påbegynde denne metode anvendes der som input værdier et antal af naboer på 15. Den øvre og den nedre grænse angives ikke, men justeres i histogrammet så snart der er lavet et preview for data, hvor data illustreres. Det giver outputtet på Figur 6.55.



Figur 6.55. K Nearest Neighbour hvor $k = 15$, $T_{min} = 0$ m og $T_{max} = 360$ m

Data ser fine ud, men det ses, at der nogle grønne punkter ude i siden og det vides, at disse er outliers. Derfor vælges det at undersøge, hvad der sker hvis antallet af naboer øges. På den måde burde det være muligt at kunne frasortere outliers i kanten af opmålingsområdet. Grunden til dette er, at målinger i kanten ikke har så mange naboer tæt på sig, som en måling i midten af den målte linje. På den måde vil nogle af de naboer en måling i kanten får tildelt være beliggende et stykke væk fra målingen, og afstanden vil derfor øges. Det vælges, at øge antallet af naboer til 20.



Figur 6.56. K Nearest Neighbour hvor $k = 20$, $T_{min} = 0$ m og $T_{max} = 250$ m

Efter at antallet af naboer er blevet øget ses det af Figur 6.56 at data er blevet renset. Faktisk ser det ikke umiddelbart ud til, at der er nogle outliers der ikke er taget med. Dette skulle gerne reflekteres i standard afvigelserne.

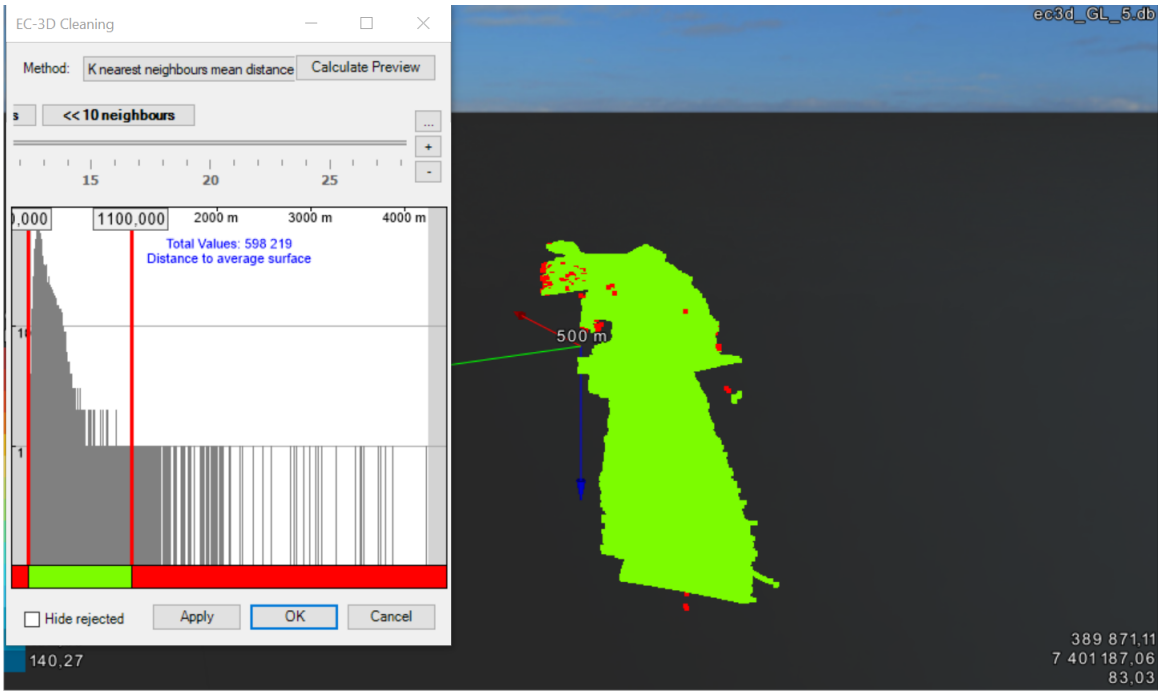
Standard afvigelse	Danmark, før	Danmark, efter
<0,10 m	91,08%	94,13%
<0,20 m	94,93%	98,18%
<0,30 m	95,87%	99,13%
<0,40 m	96,26%	99,56%
<0,50 m	96,48%	99,79%
<0,60 m	96,62%	99,93%
<0,70 m	96,69%	99,96%
<0,80 m	96,82%	99,98%
<0,90 m	97,08%	99,98%
<1,00 m	97,33%	99,98%

Tabel 6.14. Procentdel af cellerne, der er under en given standard afvigelse for henholdsvis Danmark efter anvendelse af K Nearest Neighbour metoden

Af Tabel 6.14 ses det også at standard afvigelserne er fine og at metoden er anvendelig til den danske data. Det kan dog ske, at der kan foretages forbedringer ved kombination med en anden metode.

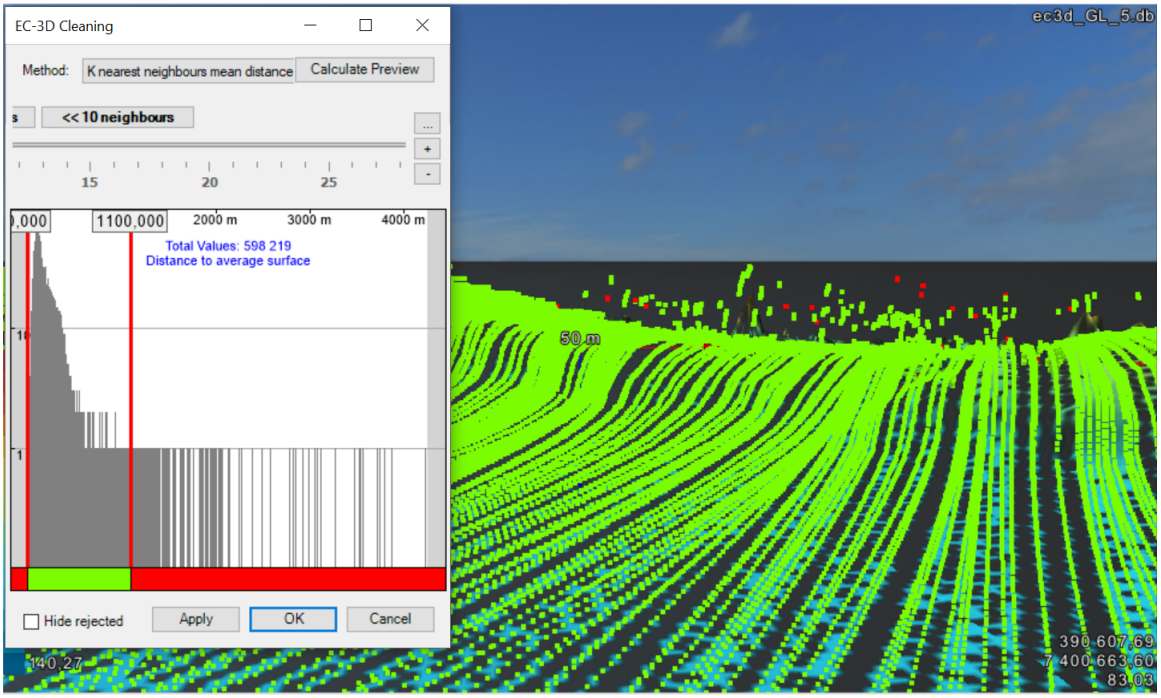
Grønland

Til denne test anvendes der et antal af naboer på 15 og grænseværdierne sættes til 0 m som den nedre og 1100 m som den øvre.



Figur 6.57. K Nearest Neighbour hvor $k = 10$, $T_{min} = 0$ m og $T_{max} = 1100$ m

Efter anvendelse af metoden ses det, at store dele af datasættet er overvejende grønt og accepteret. Justeres der mere på den øvre grænseværdi, vil "korrekt" data blive markeret rød, hvilket ikke er intentionen ved at rense data. Disse grænseværdier passer altså med at ingen "korrekt" data slettes, men sørger også for, at nogle fejlbehæftede slettes, se Figur 6.58.



Figur 6.58. K Nearest Neighbour hvor $k = 10$, $T_{min} = 0$ m og $T_{max} = 1100$ m

Det er forsøgt at ændre antallet af naboer, men konklusionen bliver den samme. Mange outliers

bliver ikke markeret til sletning, uden at "korrekt" data også slettes.

Standard afvigelse	Grønland, før	Grønland, efter
<0,10 m	75,08%	75,09%
<0,20 m	84,40%	84,50%
<0,30 m	90,01%	90,02%
<0,40 m	93,18%	93,20%
<0,50 m	95,28%	95,30%
<0,60 m	96,64%	96,66%
<0,70 m	97,60%	97,62%
<0,80 m	98,17%	98,19%
<0,90 m	98,54%	98,56%
<1,00 m	98,80%	98,82%

Tabel 6.15. Procentdel af cellerne, der er under en given standard afvigelse for Grønland efter anvendelse af K Nearest Neighbour metoden

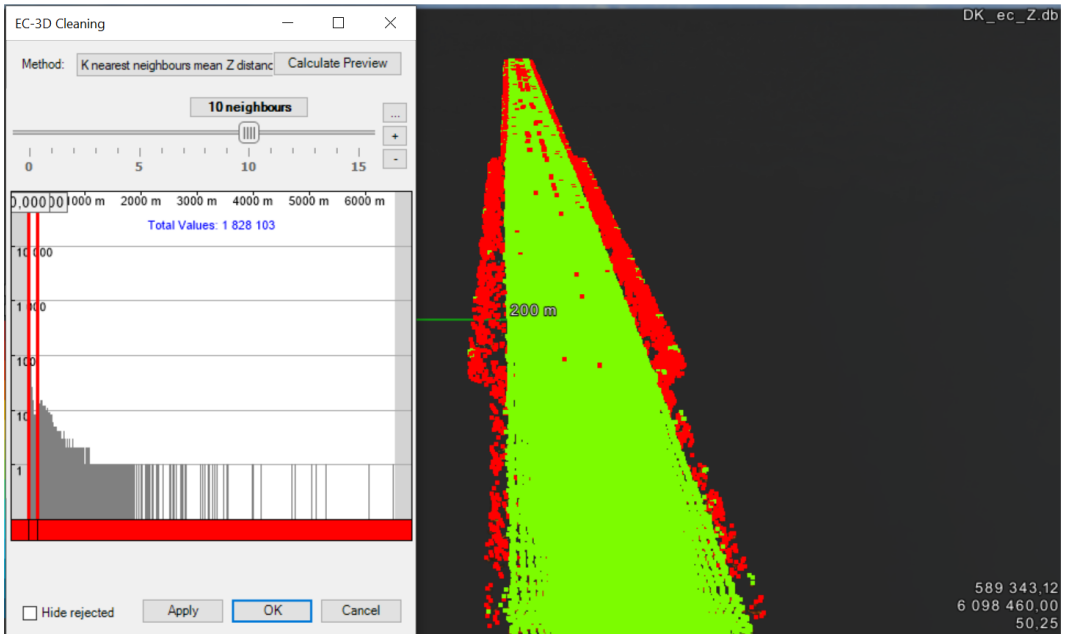
Tabellen, der viser procentdelen af celler, der er under en given standard afvigelse, har igen ikke ændret sig så meget, hvilket heller ikke var forventeligt.

6.3.6 K Nearest Neighbour, Z

Denne metode skal have de samme input værdier som K Nearest Neighbour, da metoderne næsten er identiske. I stedet for at anvende en 3D afstand, anvendes dog blot dybdeforskellen.

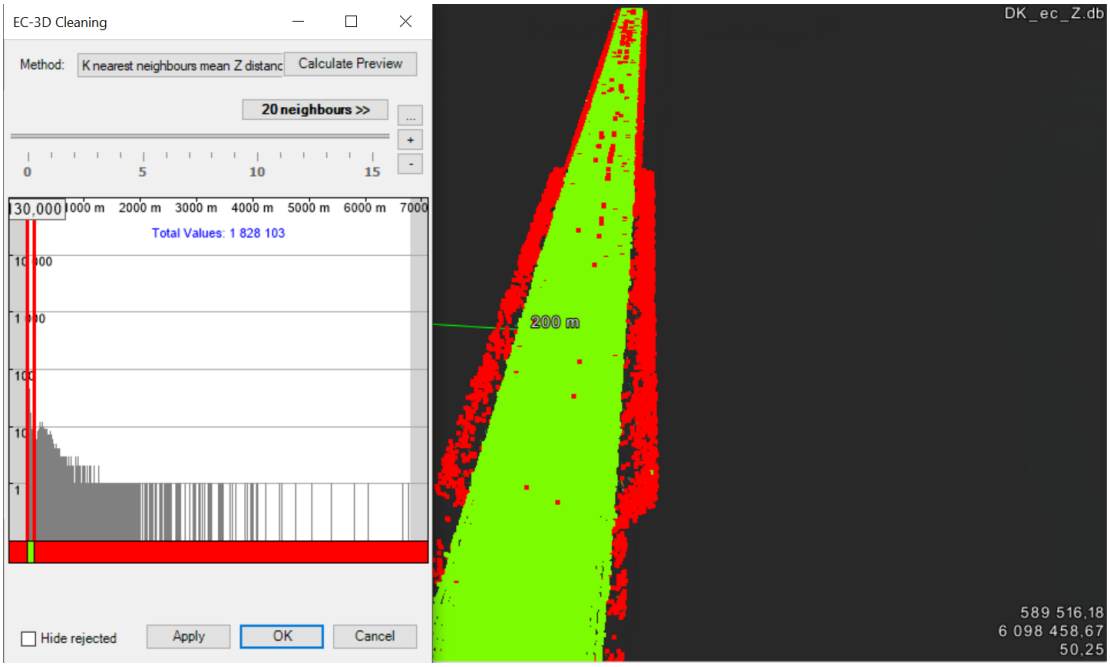
Danmark

Først testes det danske datasæt, og de input værdier, der anvendes er en antal af naboer på 15, samt en nedre grænse på 0 m og en øvre grænse på 150 m. Det giver outputtet på Figur 6.59.



Figur 6.59. K Nearest Neighbour, Z hvor $k = 10$, $T_{min} = 0$ m og $T_{max} = 150$ m

Det ses at data umiddelbart renses fint, men at der stadig er nogle outliers i kanten af opmålingsområder, der er markeret grønne. Dette skyldes formenligt at de er beliggende i samme dybde som havbunden, hvilket hovedparten af alle punkterne er.



Figur 6.60. K Nearest Neighbour, Z hvor $k = 20$, $T_{min} = 0$ m og $T_{max} = 250$ m

Ændres antallet af naboer til 20 og den øvre grænseværdi til 130 fås outputtet på Figur 6.60, hvilket ser noget bedre ud. Her er der ikke umiddelbart markeret nogle valide punkter til sletning, og det ligner at alle outliers er markeret til sletning. Standard afvigelserne undersøges også.

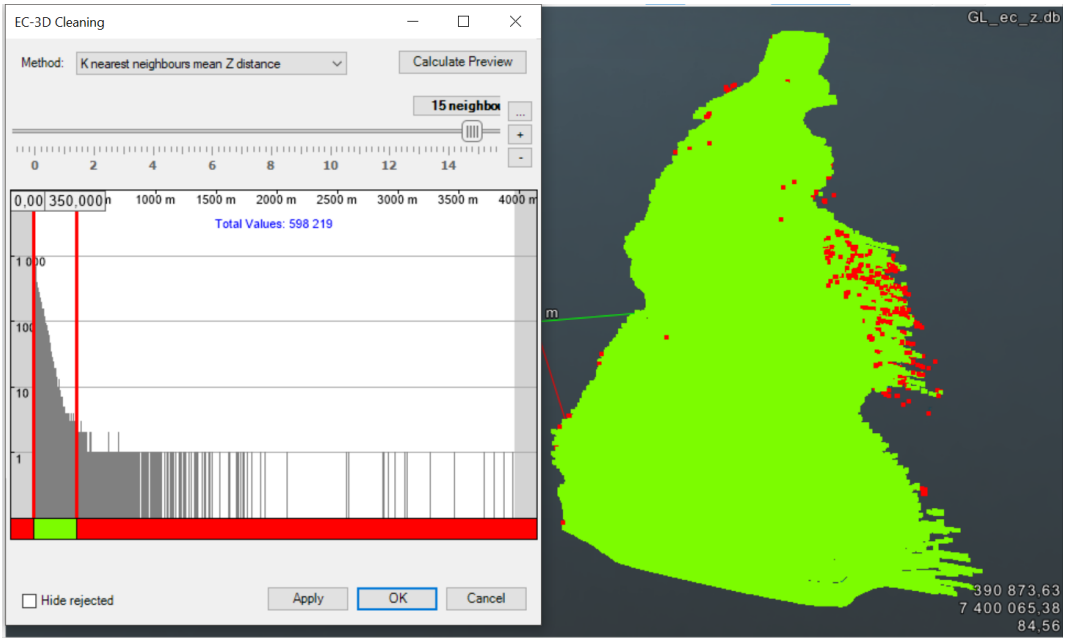
Standard afvigelse	**Danmark, før**	**Danmark, efter**
<0,10 m	91,08%	93,97%
<0,20 m	94,93%	98,10%
<0,30 m	95,87%	99,12%
<0,40 m	96,26%	99,55%
<0,50 m	96,48%	99,77%
<0,60 m	96,62%	99,89%
<0,70 m	96,69%	99,91%
<0,80 m	96,82%	99,92%
<0,90 m	97,08%	99,92%
<1,00 m	97,33%	99,92%

Tabel 6.16. Procentdel af cellerne, der er under en given standard afvigelse for henholdsvis Danmark efter anvendelse af K Nearest Neighbour, Z metoden

Standard afvigelserne ser fornuftige ud, og metoden ses som anvendelig til at behandle det danske datasæt.

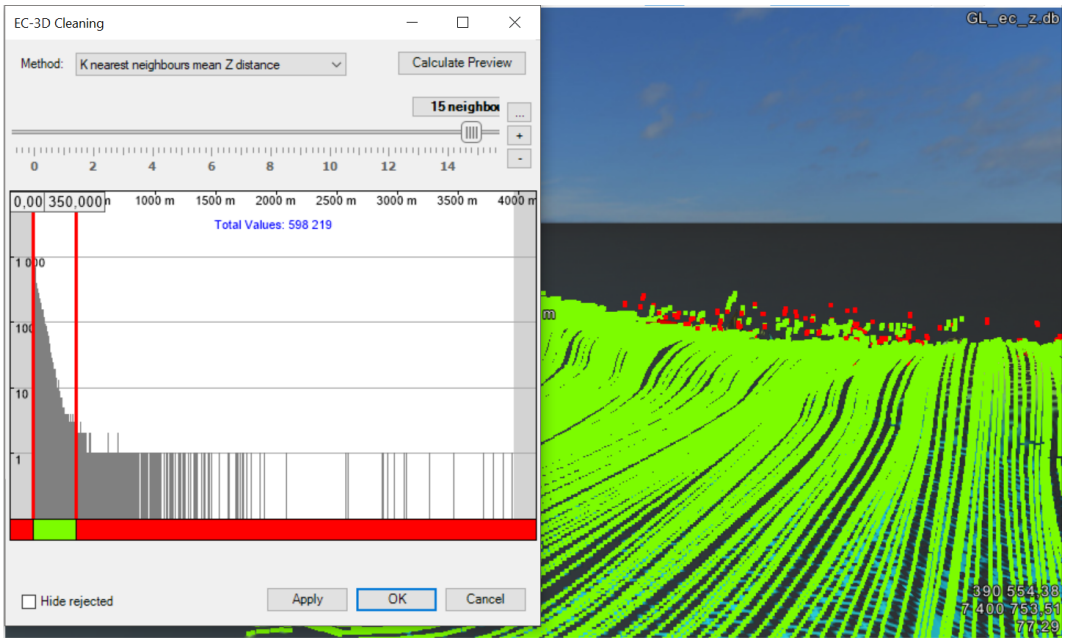
Grønland

Til at teste det grønlandske data anvendes der 15 naboer og den nedre grænseværdi sættes til 0 m og 350 m som den øvre. Outputtet kan ses på Figur 6.61.



Figur 6.61. K Nearest Neighbour hvor $k = 15$, $T_{min} = 0$ m og $T_{max} = 350$ m

Her ses det, at data overvejende er grøn. Nedsættes den øvre grænse vil nogle af de kuperede områder blive markeret røde. Dette er fordi der her er store variationer i dybden, og det er nok derfor at metoden ikke kan håndtere dem. Det betyder, at der er en række outliers der bliver markret grønne, og derfor accepteres, se Figur 6.62.



Figur 6.62. K Nearest Neighbour hvor $k = 10$, $T_{min} = 0$ m og $T_{max} = 1100$ m

Det ser altså ikke ud som om at metoden er anvendelig til denne type data, men det kan være, at et bedre resultat vil kunne opnås ved at kombinere metoden med en anden. Standard afvigelserne ses i Tabel 6.17.

Standard afvigelse	**Grønland, før**	**Grønland, efter**
<0,10 m	75,08%	75,11%
<0,20 m	84,40%	84,54%
<0,30 m	90,01%	90,06%
<0,40 m	93,18%	93,24%
<0,50 m	95,28%	95,34%
<0,60 m	96,64%	96,70%
<0,70 m	97,60%	97,66%
<0,80 m	98,17%	98,24%
<0,90 m	98,54%	98,61%
<1,00 m	98,80%	98,87%

Tabel 6.17. Procentdel af cellerne, der er under en given standard afvigelse for Grønland efter anvendelse af K Nearest Neighbour, Z metoden

Det ses at standard afvigelserne ikke er forbedret meget, hvilket dog heller ikke var forventeligt.

Det er ved anvendelse af de forskellige metoder i EC-3D erfaret, at der ikke er en metode, der ses som anvendelig til en bestemt type data. Meget tyder på at metoderne er bedst ved at anvende en kombination af forskellige metoder afhængig af hvilket datasæt der arbejdes med. Nogle af metoderne virker nogenlunde, men der er generelt, at ingen af dem kan stå alene. I hvert fald ikke i de testede datasæt. Til at anvende metoderne går der noget tid med at finde passende input værdier. Dette kan dog skyldes uerfaren anvendelse med programmet og data, men både S-CAN og EC-3D kan godt tage et par minutter at finjustere sådan at et ønsket resultat opnås.

7 | Konklusion

Formålet med dette projekt har været at besvare problemstillingen som lyder

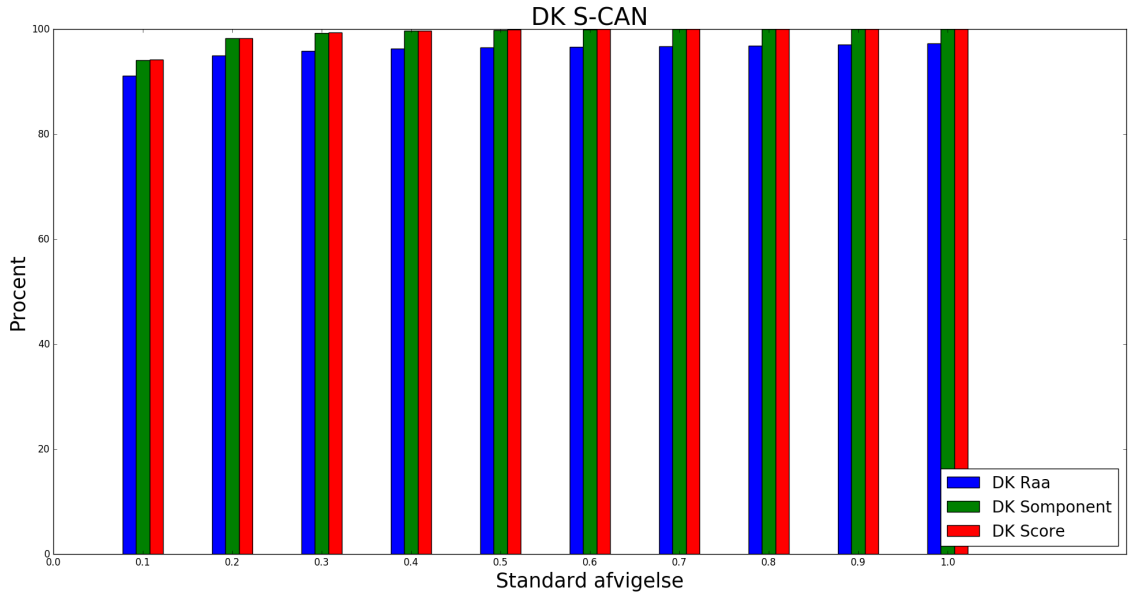
Hvordan efterbehandles multibeam data automatisk, og hvilke algoritmer og metoder ses som mest hensigtsmæssige at anvende til forskellige typer af data

For at besvare problemstillingen er det først litterært undersøgt, hvilke muligheder der er indenfor behandling af multibeam data. Her er det undersøgt, hvilke programmer der findes på markedet til at behandle multibeam data. Efter at denne undersøgelse er fortaget, er der udvalgt tre producenter til yderligere undersøgelser. Dette har medvirket, til at EIVA, QPS og CARIS er undersøgt nærmere for at finde ud af, hvilke metoder og algoritmer der ligger som baggrund for software, og der er erfaret, at der anvendes CUBE, S-CAN og EC-3D.
CUBE er en algoritme, der generer en DTM over området på baggrund af en række estimater. Her opdeles hele opmålingsområdet i et grid, hvor der er en række knudepunkter. Det er i disse knudepunkter, der på baggrund af målte dybder og deres usikkerheder estimeres én dybde og én usikkerhed. Algoritmen går altså ind og estimerer den "sande" dybde til en given lokation.
S-CAN er udviklet af EIVA i samarbejde med SCALGO og kommer i to varianter. De to varianter minder meget om hinanden, og de danner en række overflader på baggrund af en grænseværdi. Her fastsættes en grænseværdi, hvor punkter indenfor denne værdi samles i en overflade. Punkter, der overstiger grænseværdien lægges i en anden overflade, og på den måde skal det vurderes, hvilke overflader der skal beholdes.
EC-3D er udviklet af EIVA og består af seks forskellige metoder til databehandling. Værktøjet består af de seks metoder, som hver især behandler data på hver sin måde. Her behandles data eksempelvis efter afstand til nærmeste nabo, antal naboer inden for en radius, summerede dybder inden for en radius eller afstand til en gennemsnitsdybde.
Efter at de forskellige metoder og algoritmer til at behandle data er beskrevet, fortages der en række tests. Her anvendes et dansk datasæt og en grønlandsk datasæt til at teste de forskellige metoder. Alle algoritmer og metoder testes på begge datasæt, hvilket vil sige, at hver algoritme eller metode bliver testet to gange. En gang på det danske datasæt og en gang på det grønlandske. Det følgende vil opsummere resultaterne for hver enkelt datasæt.
Ved anvendelse af CUBE blev det konkluderet, at algoritmen er hurtig og effektiv til at behandle data. Dog kan der ske forbedringer, såfremt at data grovsorteres inden at CUBE anvendes. Litteraturen til CUBE omtaler også denne ulempe ved algoritmen, hvor den har svært ved at håndtere outliers, der ligger samlet i et område. Når outliers ligger på den måde, kan algoritmen ikke vurdere, hvad der er valid data, og hvad der er fejlbehæftet data.

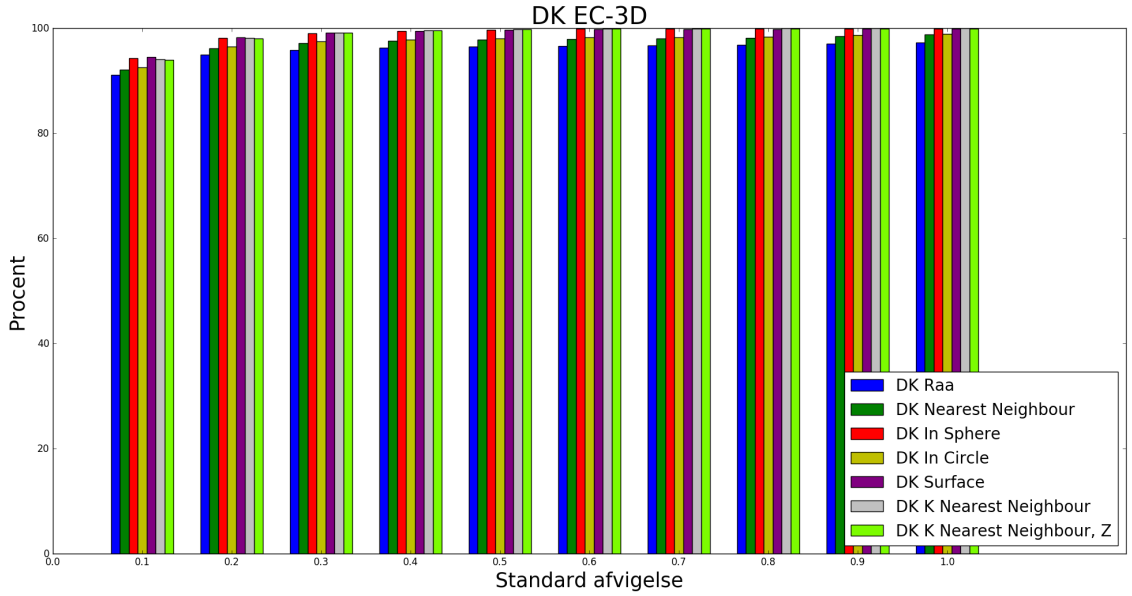
Danmark

Det danske datasæt er blevet testet med de forskellige algoritmer og metoder. I diagrammerne på Figur 7.1 og Figur 7.2, ses det hvordan andelen af celler under en given standard afvigelse har ændret sig efter anvendelse af de forskellige metoder. Af Figur 7.1 ses resultaterne for standard afvigelserne ved at anvende S-CAN. Det ses, at der sker en forbedring i standard afvigelserne ved at anvende både S-CAN component og S-CAN score. Overordnet ses det, at der ikke er den store forskel i om data behandles med den ene eller anden metode indenfor S-CAN, men der er dog mindre variationer i andelen af celler, der er under en given sandard afvigelse. Det kan dog konkluderes, at begge metoder kan ses som værende anvendelige til behandle data med.
Ved test af EC-3D ses det af Figur 7.2 at andelen af celler under en given standard afvigelse forbedres ved at anvende metoderne. Der ses yderligere nogle variationer i resultaterne fra de forskellige tests.
Ved anvendelse af *Nearest Neighbour* bliver standard afvigelserne bedre, men der er stadig en række outliers der ikke slettes, hvilke er dem, der har en anden outlier som nabo.
Metoden *Neighbour In Sphere* giver et godt resultat og er er anvendelig for at rense data af samme type som det danske datastæt. Her er procentdelen lidt større end ved Nearest Neighbour.
Da *Z Summation In Circle* blev anvendt, blev det vurderet, at metoden ikke blev set som passende, idet mange outliers blev medtaget. Hvis alle outliersne skulle medtages ville valid data også blive markeret til sletning. Det kan ses af diagrammet, at procentdelen af celler under en given standard afvigelse generelt er lidt lavere ved anvendelse af denne metode frem for en anden. Dog med forbedringer i forhold til de rå data.
Ved anvendelse af *Didstance To Average Surface* ses det, at standard afvigelserne igen øges, og at metoden renser data nogenlunde. Der skabes en linje af outliers i kanten af opmålingen, der er beliggende i samme dybde som havbunden. Metoden ses derfor ikke som anvendelig til at bruge alene for datasættet.
K Nearest Neighbour er en god metode at anvende til denne type af data, men det kan ske, at rensningen af data kunne blive endnu bedre ved at kombinere med en anden metode. Af diagrammet Figur 7.2 ses det også hvordan at den grå bar nærmer sig 100%. Det samme er gældende for den lysegrønne, hvilken præsenterer *K Nearest Neighbour, Z.* Denne metode ses til dels også som anvendelig til at behandle det danske data.

Der er altså flere metoder, som til dels renser data fint, men det må også erkendes, at der næsten for hver metode er lidt outliers tilbage. Derfor vil det generelt være en god ide at foretage en kombination af forskellige metoder. Generelt vurderes det, at det danske datasæt bedst ville behandles ved at grovsortere i data, eksempelvis med *K Nearest Neighbour, (Z)* eller S-CAN for derefter at anvende CUBE på datasættet.



Figur 7.1. Opsummering af standard afvigelses fordelingen for det danske data



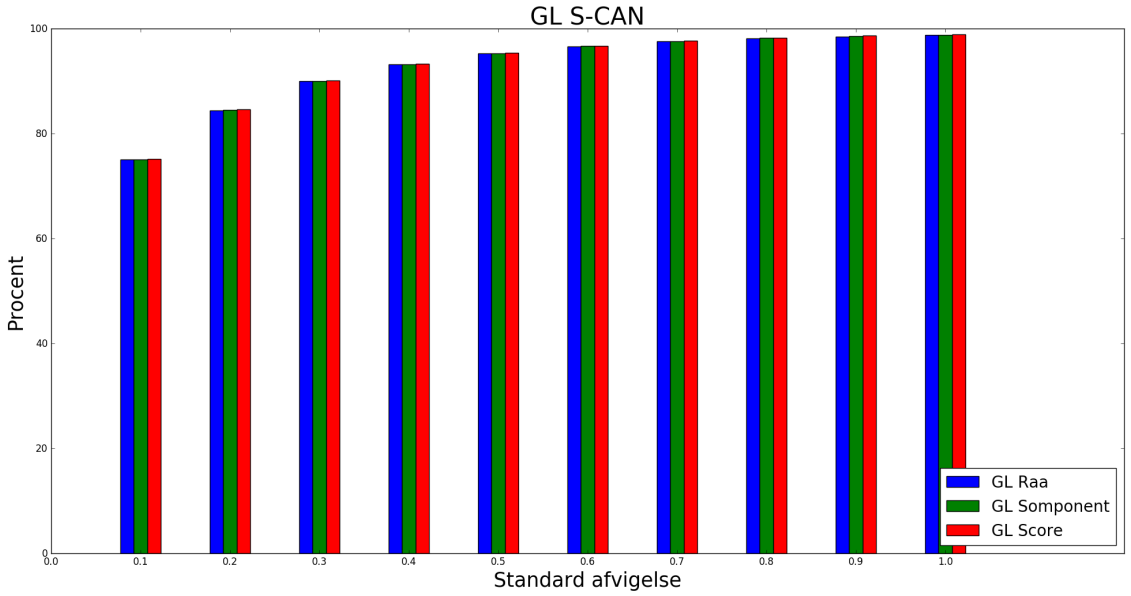
Figur 7.2. Opsummering af standard afvigelses fordelingen for det danske data

Grønland

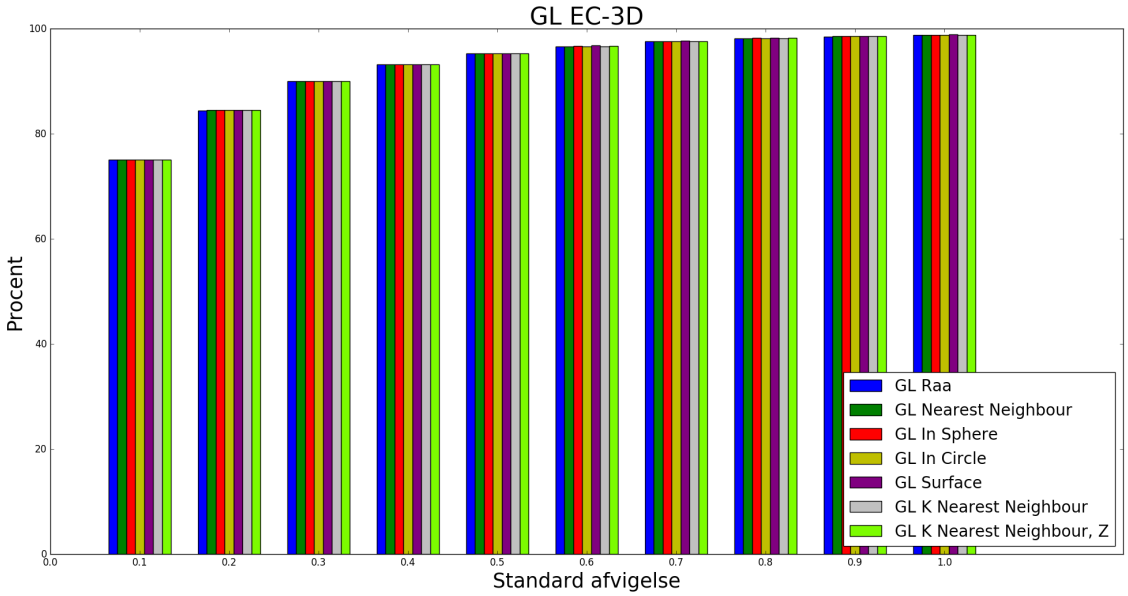
På samme måde som for det danske datasæt er der også opstillet diagrammer for det grønlandske, for at kunne sammenholde resultater fra de forskellige test. Diagrammerne ses på Figur 7.3 ogFigur 7.4.
Ved anvendelsen af S-CAN, se Figur 7.3, ses det, at der egentlig ikke er så meget forskel på anvendelsen af henholdsvis S-CAN component og S-CAN score. I de test der blev fortaget, blev data altså renset nogenlunde ens, men begge metoder krævede en del inddragelse af brugeren, samt at metoderne blev anvendt mere end en gang. Dette skyldes nogle af de stærkt kuperede områder, som metoderne har svært ved at håndtere, da havbunden nærmest bevæger sig lodret. Det kunne dog konkluderes, at de værste outliers kunne fjernes fra de mere flade områder af havbunden, men at det var svært at fjerne outliers i de kuperede områder.
I diagrammet på Figur 7.4 ses andelen af celler under en given standard afvigelse, hvor det ses, at de er ret ens. Der kan dog anes nogle mindre variationer.
Ved anvendelsen af *Nearest Neighbour* blev data ikke renset fyldestgørende. En større mængde outiers blev ikke fjernet, idet de er beliggende samme sted, og derfor har hinanden som nabo. Derfor blev metoden ikke vurderet som passende til at rense de grønlandske data.
Neighbour In Sphere har samme problem som Nearest Neighbour. En større samling af outliers bliver ikke markeret til sletning, uden at valid data også vil blive slettet. Derfor ses denne metode heller ikke som anvendelig.
Anvendelsen af *Z Summation In Circle* blev vurderet til ikke at kunne anvendes alene, da det samme problem som ved de to andre metoder opstår. En mulighed kunne være at anvende flere metoder i forbindelse med hinanden.
Distance To Average Surface ses til dels som anvendelig for det grønlandske data. Igen medtages der en del outliers, hvilket igen medvirker til at data ikke renses helt færdigt.
Anvendes *K Nearest Neighbour* er der mange outliers, der ikke markeres røde til sletning. Dette er formenligt fordi, at flere outliers har hinanden som naboer.
Til sidst anvendes *K Nearest Neighbour, Z* som vurderes til at være en af de bedste metoder til at behandle det grønlandske data. Dog behandles data ikke færdig, da outlier problemet også er at finde her.

Alle de anvendte metoder får problemer med at behandle alle outliers, uden at det påvirker valid data. Det kan ske, at en kombination af de forskellige metoder vil være bedst at anvende for at kunne behandle data færdigt og korrekt. Derudover kan en mulig løsning være, at grovsortere i data og derefter anvende CUBE, ligesom det blev forslået med det danske datasæt.

Konklusionen for projektet er derfor, at hvis data grovsorteres for de værste outliers og derefter behandles med CUBE, vil det formenlig være den mest hensigtsmæssige metode at anvende på begge typer af data. Ved anvendelse af EC-3D ses det, at der kan være forskel på, om der behandles data, der er overvejende flad i havbunden eller om havbunden er kuperet. Dette er dog kun ved behandling med én af metoderne. Det er ikke undersøgt, hvordan resultaterne vil forholde sig, hvis en kombination af metoderne blev anvendt. Ved anvendelse af S-CAN blev der heller ikke vurderet nogen betydningsfuld forskel ved anvendelsen af det danske og det grønlandske datasæt. CUBE ses blandt andet som den bedste løsning efter en grovsortering fordi den er hurtigere at anvende end de andre metoder, men især fordi den ikke vurderes subjektivt. CUBE giver det samme resultat hver gang for et givent datasæt, og er altså mere objektiv.



Figur 7.3. Opsummering af standard afvigelses fordelingen for det grønlandske data



Figur 7.4. Opsummering af standard afvigelses fordelingen for det grønlandske data

8 | Diskussion

I projektet er det undersøgt hvordan en begrænset mængde af metoder virker og de er anvendt til forskellige tests. Denne diskussion vil forsøge at diskutere andre mulige metoder til at besvare problemformuleringen på, men også reflektere over konklusionen i projektet.

Problemformuleringen er besvaret på den måde, at der er udvalgt en tre producenter, som laver behandlingsprogrammer til multibeam data. Ud fra disse tre, er det så undersøgt, hvilke algoritmer og metoder der anvendes, men der kan være mange flere metoder. I projektet er der opstillet en tabel, hvori der er præsenteret 12 producenter og kun tre er udvalgt til videre undersøgelse. De tre udvalgte producenter deler en række algoritmer og det er muligt, at de resterende også anvender disse algoritmer. Modsat er det også muligt, at de ikke gør, og at der findes andre og måske bedre algoritmer og metoder til at behandle data i andet software. Derfor kan besvarelsen af projektet, ikke ses som ensbetydende med, at CUBE er den bedste metode at anvende. Det er muligt at dette er tilfældet, men for at besvare problemformuleringen helt konkret, burde andre producenters software også have været undersøgt. Dette kunne dog af tidsmæssige årsager ikke lade sig gøre for projektets tidsramme.

I konklusionen af projektet blev det vurderet, at en grovsortering af data samt anvendelse af CUBE blev set som den bedste løsning. Her skal der tages højde for, at ingen af metoderne er forsøgt anvendt i kombination med hinanden. Alle algoritmer og metoder er anvendt separat, og det vides derfor ikke, om en kombination af flere metoder fra EC-3D kunne medvirke til et bedre resultat end ved anvendelse af den forslåede løsning. Det kunne derfor være interessant at teste en kombination af de forskellige metoder på et eller flere datasæt. På den måde vil det muligvis kunne vurderes, hvorvidt at data kunne renses fyldestgørende. En anden ide kunne være, at teste CUBE med anvendelse af S-CAN eller EC-3D. Dette kunne være interessant fordi konklusionen faktisk kun bygger på overvejelser og at der faktisk ikke ligger en test til grund for den.

Yderligere kunne det være interessant at undersøge den tid, der bliver anvendt på behandling af data. Den automatiske process skal medvirke til at data renses optimalt og at rensning af data ikke bliver en flaskehals i hele processen. Derfor kunne det være interessant at undersøge, hvor meget tid der bliver anvendt for at kunne gøre det op med et resultat. Det kan være, at et datasæt kan renses 95% indenfor 2 min, men at det vil tage 10 min at rense datasættet 99%, og er det så værd at bruge 8 minutter på de sidste procent?

Litteratur

Calder, **2003**. Brian Calder. *Automatic Statistical Processing of Multibeam Echosounder Data*. https://scholars.unh.edu/cgi/viewcontent.cgi?article=1980&context=ccom, 2003.

Calder ogMayer, **2003**. Brian Calder og Larry Mayer. *Automatic Processing of High-Rate, High-Density Multibeam Echosounder Data*. https://scholars.unh.edu/cgi/viewcontent.cgi?referer=https://www.google.dk/&httpsredir=1&article=1109&context=ccom, 2003.

CARIS, **Ukendt**. CARIS. *CARIS HIPS and SIPS*. http://www.caris.com/downloads/brochures/hipssips-en.pdf, Ukendt.

CARIS, **2018a**. CARIS. *About Teledyne CARIS*. http://www.caris.com/company.cfm, 2018.

CARIS, **2018b**. CARIS. *Software Products*. http://www.caris.com/products/, 2018.

Dall, **2011**. Lars Dall. *Efficient Sonar Data Cleaning*. http://download.eiva.dk/online-training/NaviModel3%20manuals/Efficient%20Sonar%20Data%20Cleaning.pdf, 2011.

EIVA, **2018**. EIVA. *Manual - NaviModel 4.2*, 2018.

EIVA, **2017**. EIVA. *New cleaning option excels in not just 2D but also 3D sonar and laser data*. https://www.eiva.com/about/the-eiva-log/show/new-cleaning-option-excels-in-not-just-2d-but-also-3d-sonar-and-laser-dat Action=1&PID=7483, 2017.

Kongsberg, **Ukendt**. Kongsberg. *Hydrographic data processing software*. https://www.km.kongsberg.com/ks/web/nokbg0240.nsf/AllWeb/C8069F640B8F569FC12574E100328AEE?OpenDocument, Ukendt.

QPS, **2017**. QPS. *QPS - acquire, process, visualize, share*. http://www.qps.nl/download/attachments/14778751/brochure+QPS+ENG+v5+web.pdf?version=1&modificationDate=1492200528283, 2017.

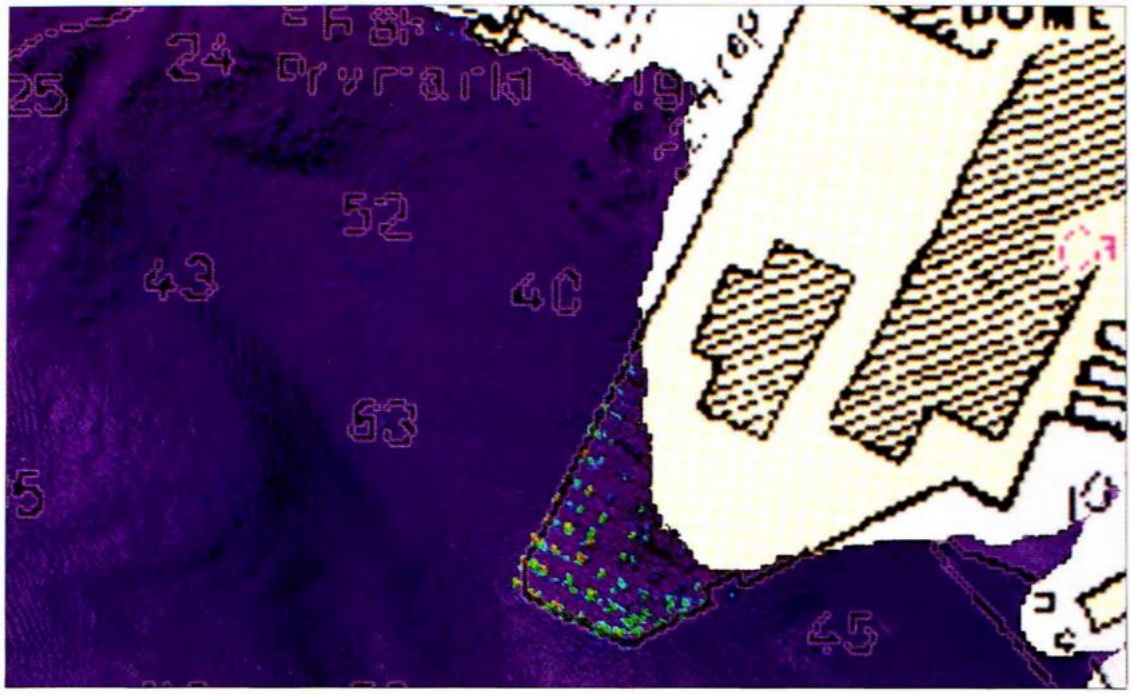
QPS, **2019**. QPS. *Data cleaning in Qimera*. http://www.qps.nl/display/qimera/data_cleaning, 2019.

Rasmussen, **2018**. Ane Aagaard Rasmussen. *Detektering af spikes i multibeam data*, Aalborg Universitet, 2018.

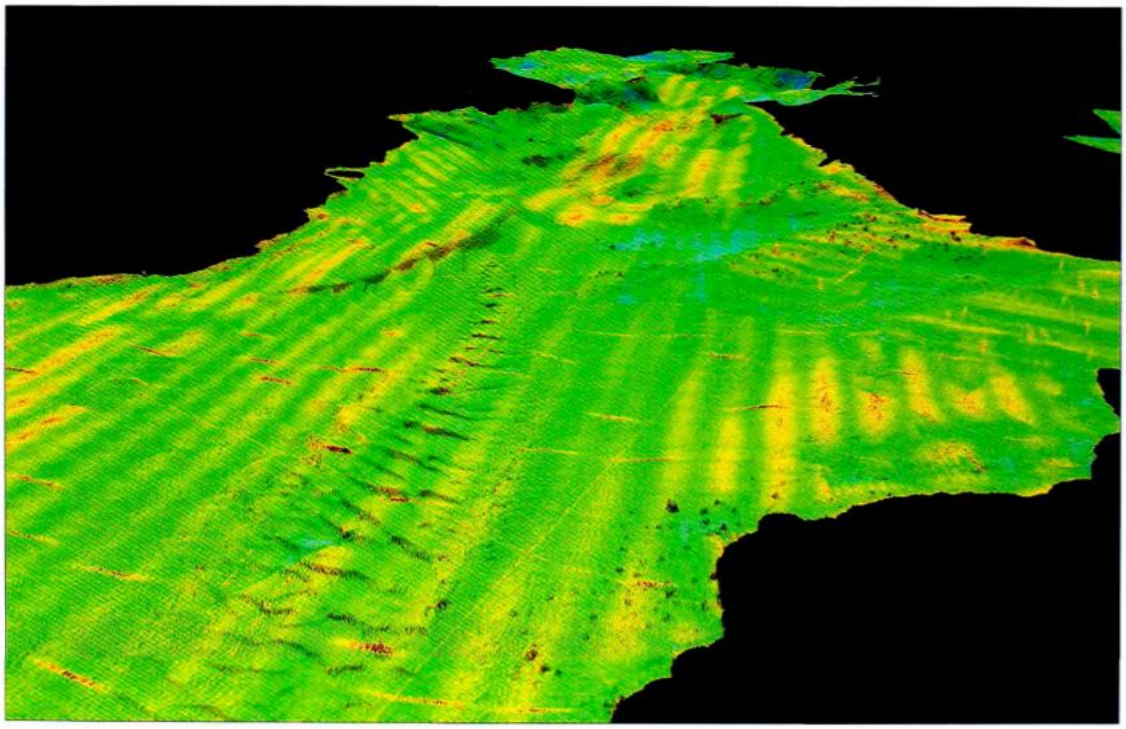
A | Digitalt billag med figurer

Dette bilag er et digitalt bilag, der er vedlagt projektet. Bilaget er en mappe kaldet figurer, og indeholder de figurer, der er anvendt i projektet. Inddelingen er på den måde, at der er en overordnet mappe kaldet figurer, som så har undermapper inddelt efter kapitelnumre. I disse undermapper er alle figurer for hvert kapitel, og figurerne er navngivet med deres nummer i projektet.

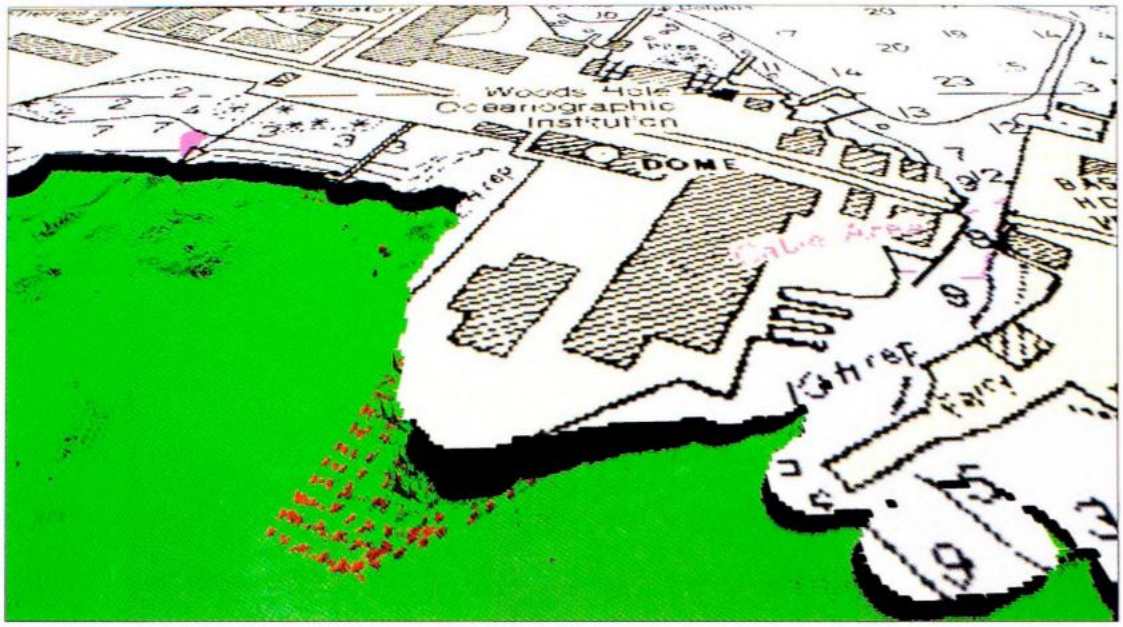
B Illustration af CUBE overflader - Teoretisk



Figur B.1. Antal af hypoteser i hvert knudepunkt. Varme farver indikerer at der er flere hypoteser i knudepunktet



Figur B.2. Overflade farvet efter usikkerhed i knudepunkter I ET 95% konfidens interval. Varme farver indikerer højere usikkerheder



Figur B.3. Styrke af hypotese. Grøn indikerer stor bekræftelse af den valgte hypotese og rød indikerer lav bekræftelse og indikerer at der er andre plausible løsninger

List of Corrections