

Bilaga 3: Metoder för modellering av effekter av fysisk påverkan på makrovegetation och epibentisk fauna

Bilaga till Nyström Sandman m. fl. 2024. Mänsklig påverkan och effekter på bentisk miljö. Metoder för bedömning av havsbottnens integritet i svenska hav. Naturvårdsverket.

Innehållsförteckning

1	Metodik.....	3
1.1	Inferens via statistik och maskininlärning.....	3
1.2	Mål och metod vid kausal inferens och dos-respons-analys	4
1.3	Krav på observationsdata	5
2	Metodbeskrivning.....	7
2.1	Val av parameter att göra effektstudier på	7
2.2	Val av prediktorvariabler och databehandling.....	7
2.3	Justeringar vid olika modellkörningar/analyser	11
2.4	Dataegenskaper och regressionsfamilj	11
2.5	Statistiska metoder	12
2.6	Maskininlärning	13
2.6.1	Metoder	13
2.6.2	Träning och modellernas betydelse.....	13
2.7	Analys och visualisering.....	14
2.8	Utvärdering av metoderna	20
3	Metod för en kombinerad påverkansmodell	22
A.C1	Ingående komponenter	22
A.C2	Metod för bottenträning	22
	Preparera indata.....	22
	Beräkna trälningsintensitet, för SAR och ev. subSAR om data finns.....	23
	Behandla resultaten	23
	Metod för sammanslagning.....	23
4	Referenser	24

1 Metodik

Den övergripande problematiken är denna: Eftersom vi bara har observationer, inga laborativa resultat, att tillgå, och dessa observationer inte är slumpmässigt utlagda i rummet och miljön, kan man inte skatta effekten av påverkan med naiv statistik. Dels beror effekterna till stor del på komplicerade och okända samband med miljön, dels reflekterar observationerna inte fördelningen av olika miljöer, och observationerna är inte heller jämnt fördelade eller slumpvis stratifierade mellan olika grad av fysisk påverkan i olika typmiljöer. Vi behöver metoder som dels låter oss ta hänsyn till olika effekter i olika miljöer, dels för att kunna vikta observationerna så att man tar större hänsyn till sällsynta kombinationer av miljö och påverkan.

Att etablera samband mellan miljö, påverkan och effekt kräver antingen ganska avancerade statistiska metoder eller maskininläring. I pilotstudierna används olika metoder från båda av dessa ramverk.

1.1 Inferens via statistik och maskininläring

Fokus för de statistiska metoderna är *inferens*, dvs. i detta fall förklaring av hur fysisk störning påverkar ett utfall, i aktuella fallet tillståndet hos vegetation och fauna. Tonvikten är resultat på populationsnivå men även stratifierat/nedbrutet på olika miljöegenskaper och/eller olika doser i en dos-responsmodell.

Maskininläring har istället fokus på *prediktion*, dvs. förmågan att hitta mönster och förutse tillstånd. Via modellering och prediktion ligger tonvikten för effekter på individnivå. En väl och med noggrann styrning genomförd maskininläring bär även på stor generell förklaringspotential genom att prediktionen identifierar "rätt" associationer (kausala och inte bara korrelativa) vilket innebär att maskininläringen därmed närmar sig den statistiska inferensen resultatmässigt. Förenklat uttryckt kan man säga att statistisk inferens förenklar samband och erhåller robusta medelvärden, medan maskininläring snarare utarbetar komplicerade samband som beskriver eller förutsäger ett utfall med stor precision men som behöver kontrolleras noga om man avser att förklara faktiska samband.

Eftersom maskininläring normalt letar efter rent korrelativa samband har man traditionellt istället använt statistiska metoder inom kausal inferens. Statistiska metoder har dock en begränsning i att de oftast är additiva, dvs. adderar effekter från olika miljövariabler genom parametrering, dvs matematiska samband. Maskininläring kan i kontrast till detta dels identifiera interaktionseffekter av godtycklig dimensionalitet, dels etablera icke-linjära/stokastiska effekter eftersom träningen inom ML följer data, inte en parametrerad kurva. Inom ML är den största faran överinläring eller *overfitting*, då modellen följer observerade data allt för precist. Effekter skall ju trots allt begripas på populationsnivå, med eller utan stratifiering. För att undvika detta införs å ena sidan regularisering, vilket gör modellen mindre exakt, å andra sidan sparas oftast en datamängd som används för validering av modellens generalitet. Man får då ett mått på hur modellen kan förutsäga observationer som inte ingår i modellen.

Beroende på de både ramverkens styrkor har i fallstudierna i detta projekt därför maskininläring använts sida vid sida med statistisk inferens för att identifiera gemensamma trender. Genom en sådan harmonisering kan man tillmäta resultaten större relevans eftersom de jobbar "från två håll", dels genom robust (men trubbig) statistik, dels genom explorativa associationer med förutsägelsekraft inom ML.

Målet för de båda fallstudierna är att etablera "responskurvor" som visar hur en indikator svarar på ett ökat fysiskt påverkanstryck. Förhoppningen har varit att de statistiska resultaten skall visa samma bild som en aggregering (stratifierad medelvärdesbildning) från maskininläringen.

Eftersom den övergripande frågeställning i Carambha rör kumulativa effekter är det den totala "kartan" med påverkanstryck som används (kallad "påverkanszoner" i Törnqvist et al. 2020).

Både inom de statistiska metoderna och inom momenten med maskininlärning inkluderas ett antal miljövariabler så som substrat, temperatur och vågexponering i inferensen och prediktionen. Inom de statistiska metoderna "kontrolleras för" dessa miljövariabler, inom maskininlärning vägs de på liknande sätt in som bidragande till predikerad förekomst/tillstånd. En viktig skillnad är att man inom statistisk inferens endast vill väga in parametrar som korrelerar både med orsak och verkan, *treatment* och *effect*. På så vis reduceras bias genom att miljövariablerna tillåts förklara andelar av utfallet. Inom maskininlärning adderas snarare så många miljövariabler som möjligt och maskinen låter miljövariabler med hög korrelation med utfallet förklara detta opportunistiskt.

Slutligen är en viktig skillnad mellan statistisk inferens och maskininlärning att eftersom man inom maskininlärning får resultat på observationsnivå och då kan studera effekter på enskilda observationer eller via godtycklig stratifiering/aggregering. Detta gör det möjligt att exempelvis plotta kartor över modellerade effekter. Genom möjligheten att prediktera individuella effekter kan man även med maskininlärning skapa prediktioner över effekter på ej observerade lokaler, eller via scenarier estimerar effekter om miljövariabler eller kausala faktorer (t.ex. fysisk påverkan) ändras.

För en introduktion till kausal inferens, se exempelvis (Austin, 2011; Austin & Stuart, 2015; Hainmueller, 2012; Robins et al., 2000; Thoemmes & Ong, 2016).

1.2 Mål och metod vid kausal inferens och dos-respons-analys

När man söker skatta kausala effekter genom hur stort bidrag en variabel ger för ett visst utfall, t.ex. en medicins effekt för en population, görs sådana analyser genom att randomisera fördelningen av experimentsubjekt och kontrollsubjekt så att man får två (eller flera) neutrala mängder att jämföra medelvärde för. Detta gäller för experimentella studier inom vilka man direkt kan undersöka utfall av olika experiment jämfört med kontrollgruppen.

I innevarande studie måste vi förlita oss på observationsdata som från början är skevt fördelade. Dels finns många observationer i vissa landskapsavsnitt och färre i andra avsnitt, dels finns det betydande korrelation eller beroenden mellan exponeringen (fysisk påverkan), miljövariabler (t.ex. vågexponering) och utfallet (abundans, täckningsgrad, makrofytindex). Exempelvis bygger man bryggor i skyddade vikar där arter som trivs i skyddade miljöer påträffas, och substrat och andra miljövariabler beror förstås av denna skyddade miljö. Och vissa påverkanstryck som exempelvis muddring är överrepresenterade just i grunda skyddade vikar. Detta ger upphov till flera utmaningar. Först och främst måste man kontrollera för miljövariablerna om man försöker jämföra "påverkade" och "opåverkade" (kontrollgruppen) områden. Sedan måste man också vara beredd på att det är svårt att lösa ut miljövariablernas inbördes bidrag eftersom de är starkt korrelerade (multicollinear). Djup, temperatur, istäcke och vågverkan på botten är exempel på fyra starkt korrelerade variabler.

För att få bukt med snedfördelningen av observationer (över- respektive underrepresentation) viktas eller sovras man vanligtvis observationerna enligt förekomst (*propensity score matching/subsetting*) (Hirano & Imbens, 2004). En vanlig metod är att göra en regression mellan utfall och kontrollerande variabler så att "sällsynta kombinationer" viktas upp och vanliga viktas ned och man får en vikt för varje observation som bygger på sannolik förekomst (*inverse probability weights, IPW*). Det kan ses som en tredje problematik att veta vilka variabler man skall använda för denna "artificiella stratifiering".

Vi har dessutom en fjärde problematik. Vi har nämligen ingen tydligt "påverkad" grupp eller en tydlig "kontrollgrupp", den fysiska påverkan är inte binär och det är därmed inte möjligt att göra en binär

analys av "average treatment effects", ATE på populationsnivå (PATE), även om vi utför balansering via IPW eller andra metoder.

Vi behöver därför metoder för att skatta en kontinuerlig dos-responsfunktion. Vissa sådana metoder bygger på metoder från binära undersökningar och skapar "syntetiska subgrupper" baserat på miljövariablerna och det man får ut av dessa metoder är exempelvis en "average partial effect" APE, förenklat uttryckt en medelderivata (exempelvis "-x % per y") framräknad inom de många matchningarna subgrupperna emellan. Denna kan sedan multipliceras med exponeringen för att få ut en generaliserad aggregerad effekt. En nackdel är då förstås att resultatet blir linjärt och det är inte säkert att dos-responsfunktionen är linjär i verkligheten.

En annan metod är att etablera en kontinuerlig dos-responsfunktion som, givet varje intervall av exponering, undersöker hur utfallet svarar på denna exponering givet observationernas viktning (IPW). I detta sammanhang talar man även om GPS, *generalized propensity scores*, som vikt för observationerna och via GPS etableras i metoderna nedan en s.k. *conditional distribution function* givet olika nivåer på exponering (Wu et al., 2024). Viktningen sker då alltså inte bara mot miljövariablerna utan även mot intensiteten i exponeringen så att svaret blir "rättvist", dvs. oberoende av miljövariabler och exponeringsfrekvens. När proverna viktats på detta sätt kan man teckna en kontinuerlig dos-responsfunktion som visar, "allt annat lika", hur responsen blir för olika exponeringsintervall.

1.3 Krav på observationsdata

För att kausal inferens och modellering av dos-respons skall vara rättvisande finns dock ett antal viktiga antaganden som behöver vara uppfyllda vad gäller observationsdata. Om villkoren inte uppfylls riskerar man att introducera bias, dvs i praktiken mer eller mindre generell över- eller underskattning som beror av en eller fler faktorer och/eller miljöparametrar. I innevarande fallstudie är följande antaganden viktiga att adressera (VanderWeele & Hernán, 2013):

1. *Stable Unit Treatment Value Assumption* (SUTVA). Materialet skall var jämförbart, homogent och rättvist dokumenterat på en gemensam skala. Vidare förutsätter SUTVA att observationer inte påverkar varandra. I miljön bryts detta antagande till någon del då det finns rumslig autokorrelation via dispersion; t.ex. fröspridning och larvspridning. Men bedömningen är att på aktuell rumslig skala bör det inte skapa så stora bias i analysen i synnerhet om det finns tillräckligt många observationer.
2. *Ignorability assumption*. Detta antagande går ut på att vi kontrollerar för relevanta miljövariabler, dvs. det finns inga okända bakomliggande faktorer som samtidigt påverkar exponering och utfall och kan introducera systematiska bias. Det är okänt om det finns några okända variabler som man egentligen skulle behöva ta ställning till.
3. *Positivity assumption*. Detta antagande menar att för att ett vägt resultat skall vara meningsfullt skall det finnas en positiv chans att alla typer av observationer utsätts för exponeringen.

För punkt 1, SUTVA-kriteriet ovan, innebär det att vi måste ha ett material som håller jämn kvalitet. Vi har valt två dataset som insamlats av samma personer med samma metodik inom loppet av några år, i två angränsande län. Dock finns det en problematik i att arterna ingår i rumsliga samband med exempelvis fröspridning och ekologiska kaskadeffekter. Observationerna är alltså aldrig oberoende. I större studier kan möjligtvis denna typ av problematik adresseras med att det är alldeles för omfattande för denna studie.

För punkt 2 innebär det att vi måste samla in miljödata (modellprediktorer) som omfattar de huvudsakliga miljöparametrarna som ligger till grund för förekomst och rikedom av olika arter (salinitet, substrat, temperatur osv). Till grund för denna process togs det fram en s.k. DAG (*Directed Acyclic Graph*) (Digitale et al., 2022) från vilken man kan härleda de miljövariabler som man behöver justera för. Se vidare nedan.

För punkt 3 uppstår ett problem genom att fysisk störning är mycket skevt fördelat i rummet och störningsfaktorerna korrelerar starkt med vissa arter (man muddrar exempelvis i grunda vikar där grunda vikars arter finns). Denna punkt löses dels statistiskt som berörts ovan, dels genom *propensity scoring* (Ebrahim Valojerdi & Janani, 2018), uppviktning av sällsynta kombinationer eller sovring av frekventa, dels genom justering för de olika miljötyperna, i synnerhet viktigt vid maskininlärningsmetoderna. För att uppfylla detta villkor görs i modellkörningarna nedan även en överlappsanalys (*common support*) så att endast observationer vars miljökombinationer förekommer i alla olika exponerade grupper tas med. Men i synnerhet i fallstudien gällande djup epibentisk fauna som utsätts för trålning ser vi att det är stora problem med representativitet då trålning bara förekommer i vissa miljöer, inte andra.

Dessa punkter gör dessutom att det ställs krav på stora mängder observationsdata; givet miljöns komplexitet och observationernas karaktär av "expertbedömning" (okulär granskning) krävs det mycket data för att dels få många observationer i olika exponeringsintervall (inom *common support*), med tillräckligt många observationer så att stratifierade medelvärden blir meningsfulla och för att inte missar i dokumentationen (brus i modellen snarare än bias) skall inverka negativt.

Sammantaget innebär dessa osäkerheter att vi dels behöver vara noga och kritiska när vi bedömer resultaten, men även att vi behöver genomföra analyserna i andra områden med andra data för att identifiera eventuella bias och få ner osäkerheten. Och att vi bör prova olika metoder och kontrollera för olika variabler och studera utfallen.

2 Metodbeskrivning

2.1 Val av parameter att göra effektstudier på

Frågan är förstas vilka parametrar man skall följa effekterna på. Den övergripande tematiken är effekter på makrovegetation och effekter på epibentisk fauna. Utgångspunkten är förstas tillgängliga observationsdata.

För makrovegetation fanns i observationsdata information om:

- Täckningsgrad för en stor mängd enskilda arter (se appendix A)
- Mängden död vegetation (täckningsgrad)

Dessa parametrar testades med avseende på hur de svarar på modellen över fysisk påverkan.

Dessutom räknades index över följande parametrar ut, baserat på observationsdata:

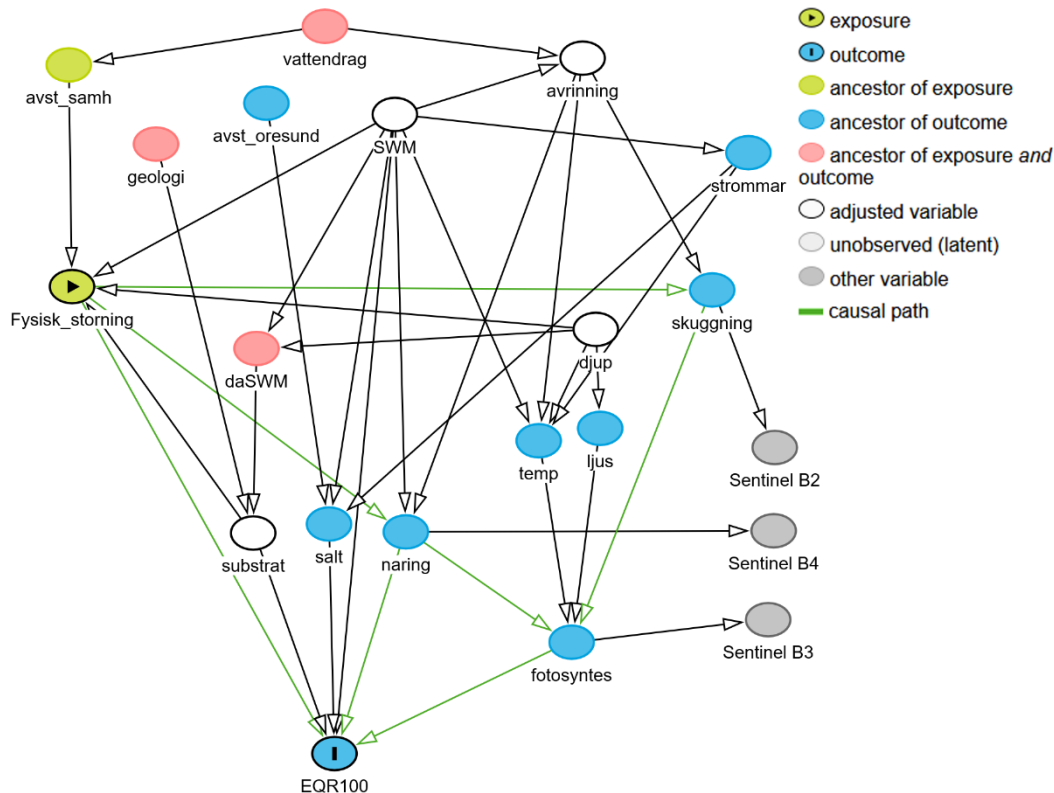
- Total täckningsgrad
- Artrikedom (antal arter)
- Biodiversitet (Shannons och Simpsons index)
- Makrofytindex (MI_a , MI_b , EQR_1) (Hansen, 2012)

Även dessa parametrar testades med avseende på hur de svarar på modellen av fysisk påverkan. Att testa makrofytindex kan ses som ett komplement till arbetet med att använda makrofytindex för att undersöka störning på grunda mjukbottnar (Hansen & Snickars, 2014); när tidigare arbeten fokuserar på att härleda störning från artsammansättning testas här istället om artsammansättningen svarar statistiskt på fysisk störning. Om svaret blir ja stärker detta modellen över makrofytindex och vi får en metod med vilken vi på sikt kanske kan skatta påverkan på en landskapsnivå.

För epibentisk fauna låg fokus på två arter som tidigare undersökts med avseende på bottentrålning och som förekommer i tillräckligt stor omfattning ($n=656$ respektive $n=775$) i tillgängliga observationsdata; piprensarna *Pennatula phosphorea* of *Virgularia mirabilis*. Abundansen för dessa (individer per 25 m²) finns noterad i observationsdata.

2.2 Val av prediktorvariabler och databehandling

Initialt gjordes ett hundratal modellkörningar med olika kombinationer av prediktorvariabler och modelleringsmetod. Normalt inom maskininlärning är att låta metoden välja ut de mest förklarande variablerna, dvs. de som ger en modell med minst fel (exempelvis låg RMSE eller hög R^2 för regression eller hög *balanced accuracy*, TSS (*true skill statistics*) eller Kappa för sannolikhetsmodeller). Det visade sig snart att även om modellerna hade ganska låga nivåer av fel hade de helt olika responskurvor, och de statistiska analyserna var oerhört instabila och koefficienterna varierade avsevärt och bytte ofta t.om. tecken, dvs. gick från negativt till positivt bidrag och tvärt om!

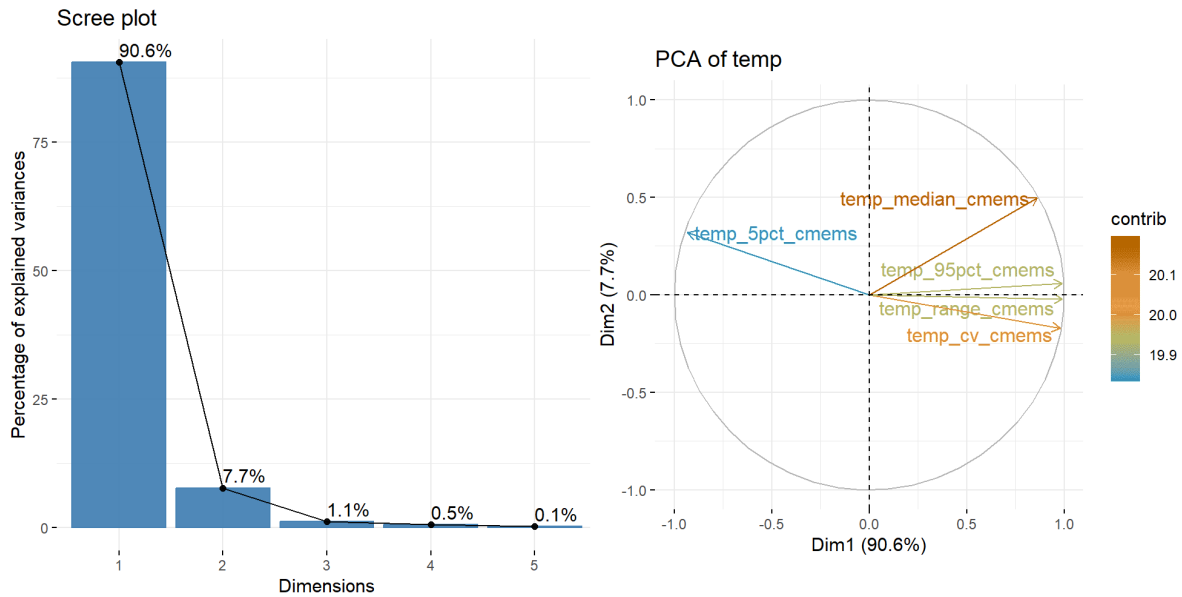


Figur 1. DAG över valet av miljövariabler att justera för, gällande makrovegetation. I förslaget ovan är det de justerade; SWM (vågexponering). Avrinning (avstånd från vattendrag), djup och substrat som behövs. Resten kan användas för att med maskininlärning bättre prediktera utfallet men responskurvorna och den statistiska analysen blir då behäftad med bias.

För den mest komplicerade situationen, makrovegetation, gjordes därför ett försök med att skapa några DAG:s (*Directed Acyclic Graphs*), för att skilja ut variabler att justera för. Via sådana grafer, som är till för kausal inferens med statistiska metoder, går det att identifiera de variabler man behöver ta hänsyn till då de direkt eller indirekt ingår i orsakssamband mellan den kausala faktorn (här: fysisk påverkan) och utfallet (här exempelvis täckningsgrad av vegetation eller abundans av sjöpennor) (Textor et al., 2016). Det man söker är antingen *confounders*, *colliders*, *mediators* eller *backdoor paths* (Pearl & Mackenzie, 2018). Justerar man för fel variabler kan man introducera bias genom att effekten felaktigt minskas, ökas och/eller blir fel i olika intervall av justerade variabler.

Det visade sig att merparten av de variabler som initialt tagits med i modellen inte behövdes eller t.o.m. skapade bias och oförutsägbart beteende. Justeras endast för de få utvalda variablerna minskar modellens precisa förutsägelsekraft (prediktion) men stratifierat på dessa variabler blir istället utfallet justerat för bias och därmed mest korrekt.

Fortfarande var de statistiska modellerna lite oberäknliga och visade sinsemellan stora skillnader. Det förutsätts även att variabler i statistiska modeller är korrelerade. Genom en korrelationsanalys visade det sig att många variabler var starkt korrelerade och därför genomfördes en principalkomponentanalys, PCA, för att minska den flerdimensionella korrelationen, *multicollinearity* (James et al., 2013).



Figur 2. Exempel på PCA. Fem variabler reducerade till två med 98,3 % av variansen kvar, och problem med korrelationer reducerade.

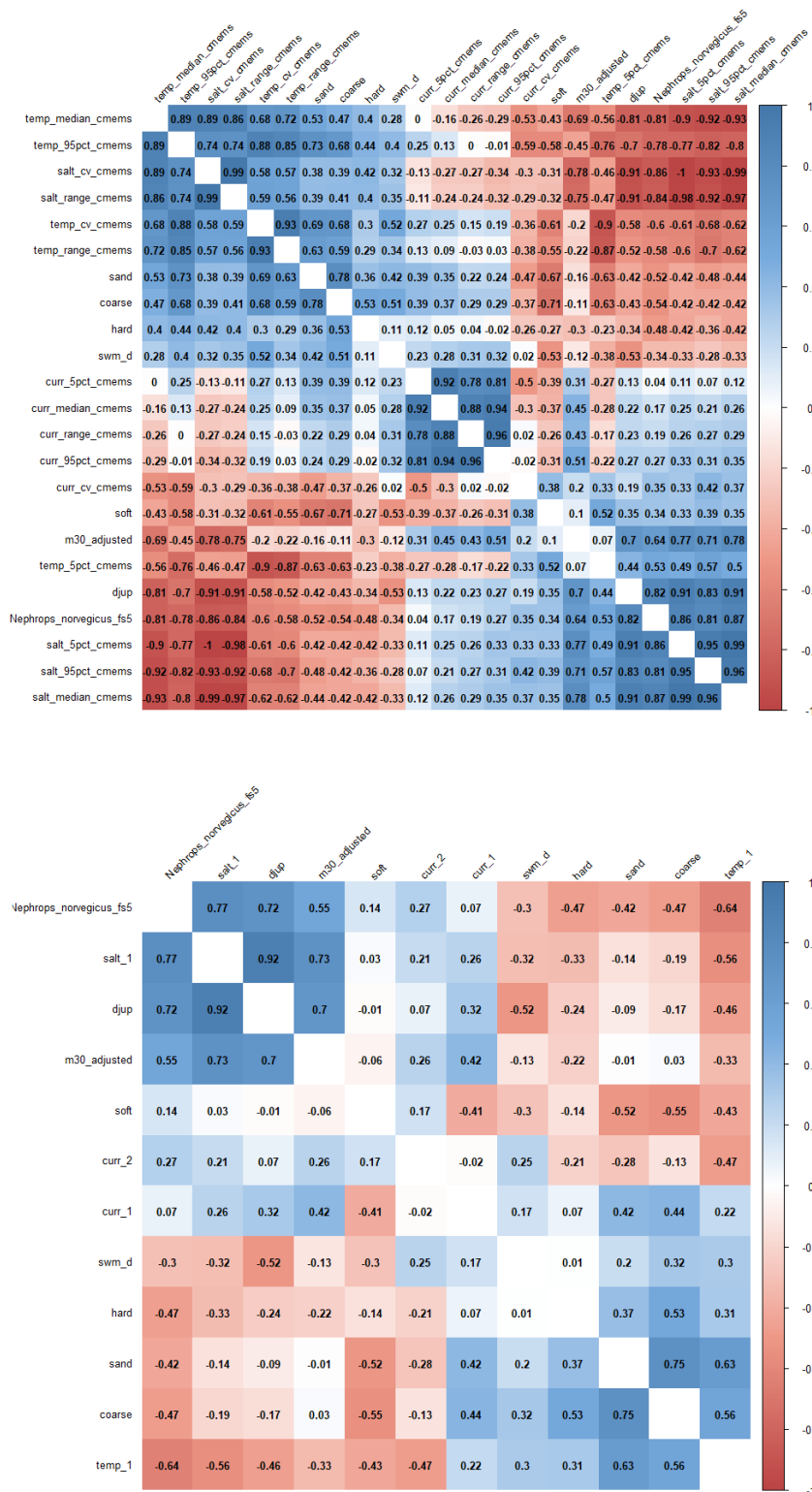
I processen valdes de kombinationer av variabler som sammantaget minst förklarade 80% av variansen i kombinationen. Detta betyder att en viss förklaringsgrad går förlorad.

Genom PCA reduceras antalet variabler och en nackdel med detta är att axlarna i de nya kombinerade variablerna inte längre är tolkningsbara för det mänskliga ögat. Responskurvor och förklaringsgrad för dessa syntesvariabler blir därmed inte möjliga att tolka på något meningsfullt sätt. Denna tolkning har ju dock inte varit fokus i dessa pilotstudier, utan istället tolkningen av effekter från fysisk störning.

Korrelationen mellan miljövariablerna (för exemplet epibentisk fauna i Kattegatt) innan respektive efter PCA-reducering framgår av figuren nedan. Fortfarande kvarstår en hel del korrelation men då variablerna är viktiga för att förklara miljön bör de inte reduceras mer.

En anledning varför så många miljövariabler som möjligt behålls i fallet som rör epibentisk fauna är att den kausala modellen måste kunna kompensera för relationen mellan förekomst av fysisk påverkan (i detta fall trålning) och förekomst av studerad parameter (här sjöpennor). Förvisso ingår en modell av havskraftor i analysen (mer om detta senare under avsnittet om epibentisk fauna) men de makroekologiska sambanden mellan trålning, miljö och sjöpennor är så pass oklar att det är bättre att ha kvar de miljövariabler som vi, i brist på en bra DAG, tror kan påverka dessa relationer. En DAG över relationen mellan trålning och sjöpennor är alltså svår att göra eftersom man trålar av olika anledningar och beroende på vilka arter som man trålar efter bör man justera för olika miljövariabler.

Correlation (spearman) between training data covariates

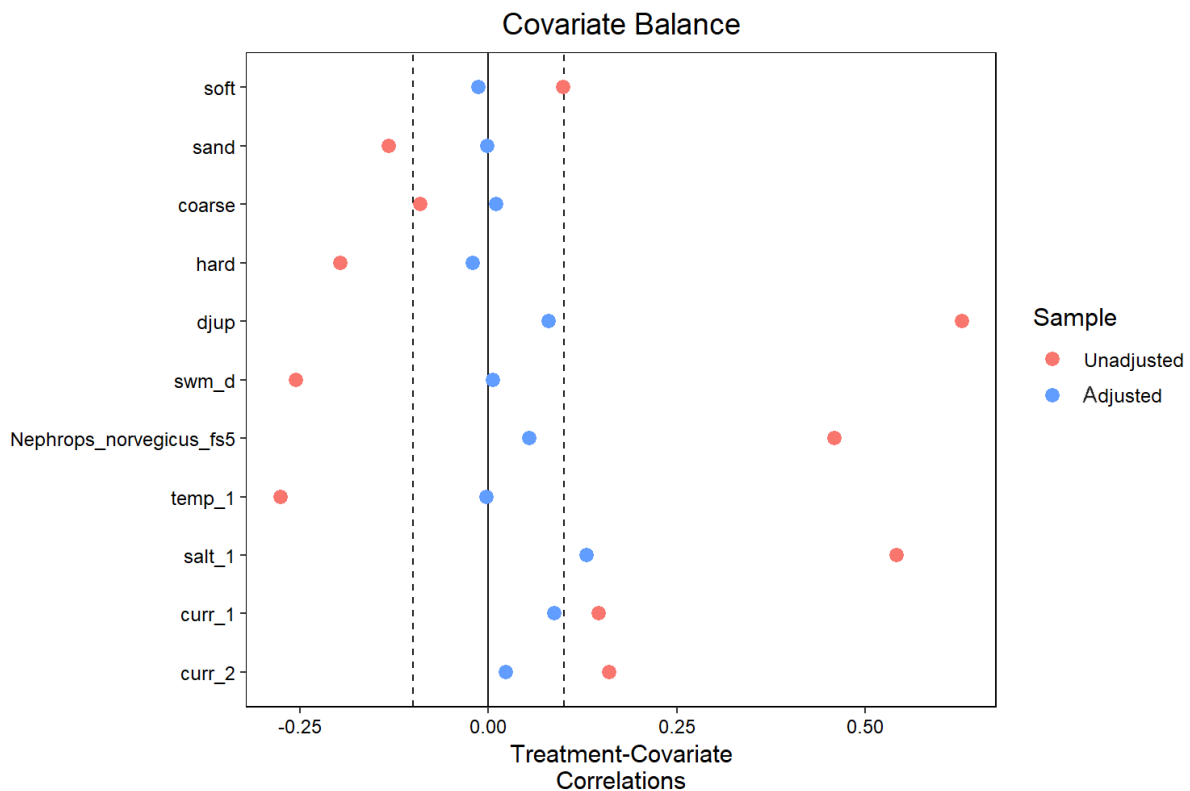


Figur 3. Korrelation (Spearman) mellan miljövariablerna före och efter PCA-analys. Analysen gäller sjöpennor/djup epibentisk fauna.

2.3 Justeringar vid olika modellkörningar/analyser

Beroende på vilka modeller och metoder som användes gjordes följande justeringar:

- **Normalisering.** För att erhålla effekter (APE/AME) på en jämförbar skala normaliserades alla variabler till mellan 0–1.
- **Common support.** Vid dos-responsmodellering (metoderna BART och GAM) togs den delmängd av observationerna bort som låg utanför den gemensamma miljövariabelytan (*covariate space*) för olika doser av påverkan.
- **Covariate balancing.** Alla statistiska metoder har någon form av balansering av observationerna. Ovan har exempelvis IPW, *inverse propensity weights*, beskrivits. Men det har även testats andra metoder så som *energy balancing*, *nonparametric covariate balancing propensity score weighting* m.fl. Det är inte enkelt avgjort vilka metoder som fungerar bäst för olika dataset men genom att visualisera/kvantifiera balansen i tabell och diagram kan man få ett grepp om vilka metoder som ger en mer rättvis balans, så att träningsdata och, med avseende på både miljövariabler och effekt, är jämnt fördelade för olika exponeringsintervall av i detta fallet fysisk påverkan.

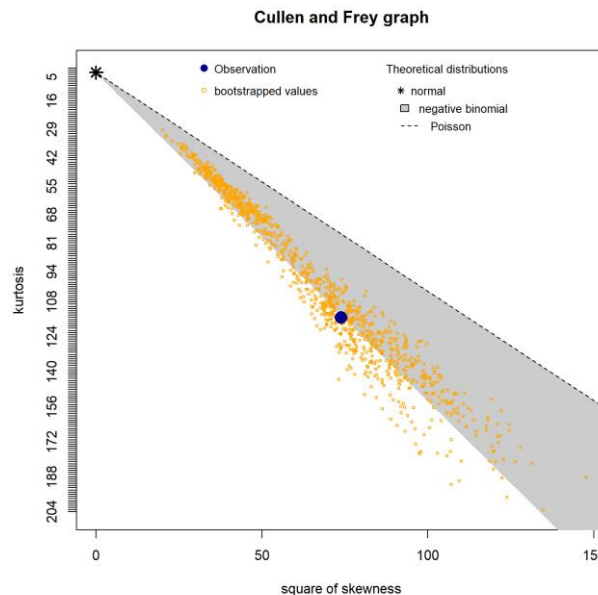


Figur 4. Exempel på balansering av miljövariabler. PCA-reducerade prediktorer för djup epibentisk fauna. Målet är för de justerade vikterna är att få balansen så nära noll som möjligt, helst inom zonen $\pm 0,1$. I exemplet ovan misslyckades detta för salt_1.

2.4 Dataegenskaper och regressionsfamilj

Både observationerna av sjöpennor och täckningsgrad för enskilda arter är skevt fördelade, har hög kurtosis och skewness. Det är därför svårt att göra en bra modell över dessa data när det är ett fåtal outliers med höga värden och en stor mängd nollvärden (zero inflation). Som modellfamilj (regression) respektive model objective kan man välja negbin eller tweedie för att till någon del minska problemen. Modellerna har dock ändå svårt att prediktera alla nollor och alla outliers med höga värden. Modellerna blir därmed något "trubbigare" än verklighetens observationer. Av denna

anledning – systematiska felprediktioner – är det mycket viktigt att kontrafaktiska effektstudier genomförs genom att ställa kontrafaktiska modellresultat mot faktiska modellresultat, inte mot faktiska observationer.



Figur 5. Både data över makrofytter (på artnivå) och över täckningsgraden för sjöpennor (med exemplet *Pennatula phosphorea* i figuren) följer närmast en negativ binomial fördelning. Av denna anledning valdes "negbin" som link-funktion för antalet sjöpennor och "tweedie" som funktion för kontinuerlig täckningsgrad av vegetation. (Delignette-Muller & Dutang, 2015)

2.5 Statistiska metoder

Följande enklare metoder används för att uppskatta kontinuerliga effekter av fysisk påverkan i de två fallstudierna:

1. Propensity-score-viktad GLM (Generalized Linear Models) med vikter hämtade genom IPW (Lumley, 2004). Multipel linjär regression som ger medeleffekt, *average marginal/partical effect*, APE/AME (Använda paket: WeightIt, svyglm/svydeign)
2. Propensity-score-viktad GLM med vikter hämtade genom modell baserad på boosted regression trees (Zhu et al., 2015). Multipel linjär regression som ger APE/AME (Använda paket: TwangContinuous/gbm, WeightIt, svyglm/svydeign)
3. *Bayesian Causal Effect Estimator* (Talbot & Beaudoin, 2022). Denna metod bygger på att genom MCMC-simulering inom *bayesian regression trees*, BART, räkna ut AME/APE som i detta fall är en full sannolikhetsdistribution, inte ett diskret medelvärde och konfidensintervall. (Använda paket: BCEE)
4. *CausalForest*: Metod baserad på Random Forest genom gruppering av observationer och jämförelse av utfall i de olika grupperna (Athey et al., 2019).

Dessa metoder ger alltså ett medelvärde på derivatan för ändring-per-enhet, t.ex. förlust av abundans per grad av påverkanstryck. En stor nackdel är att responsen inte är linjär och i synnerhet om modellen är dåligt parametrerad eller det finns andra bias och brister i exempelvis sampling som viktningen/balanseringen inte lyckats kompensera kan en felaktig respons i delar av intervallet ta ut effekter under andra delar av intervallet. Därmed blir det tydligt att det behövs metoder för att modellera och granska responsen under hela intervallet av påverkan.

Av dessa anledningar implementerades följande mer komplicerade statistiska metoder för att skapa kontinuerliga dos-responskurvor:

1. Kubisk (spline) multippel linjär regression med interaktionseffekter som ger respons och slope (AME/APE) per exponeringsintervall efter att ha viktat observationerna genom energy balance (paket: WeightIt, lm, margineffects, cobalt).
2. BART (Bayesian Additive Regression Trees) genom Markov Chain Monte Carlo (MCMC), med IPW-viktning och common-support (Galagate, 2016). Denna metod ger medelrespons per exponeringsintervall, plus prediktionsintervall. (Använda paket: causaldrf)
3. GAM, Generalized Additive Models (Wood, 2017), viktad genom IPW. Denna metod ger medelrespons per exponeringsintervall. (Använda paket: causaldrf)
4. GPCERF (*Gaussian Processes for Estimating Causal Exposure Response Curves*) (Ren et al., 2023). Modellering av gaussisk respons byggd på BART och MCMC. Denna metod ger medelrespons per exponeringsintervall, plus prediktionsintervall i form av probabilitetsdistributionen från BART. (Använda paket: GPCERF)

Statistikrapporter (loggfiler) från alla körningar, inklusive visualisering, finns i appendix D.

Målet både för metoder som ger medeleffekt och de som ger en kontinuerlig effekt är att skapa "dubbelt robusta" (*doubly robust*) resultat. Detta innebär alltså att det i praktiken skapas två regressioner per modell, en för relation mellan exponering och miljö, samt en mellan miljö och utfall. Genom denna procedur reduceras bias och problemet med misspecificering av modellen och bias i sampling minskar.

2.6 Maskininlärning

2.6.1 Metoder

För detta moment har två olika metoder för maskininlärning, ML, använts. Båda bygger på "ensembles of weak models" som i dessa fall innebär att många beslutsträd (hundratals) identifieras empiriskt utifrån träningsdata och adderas till en medelvärdesbildad modell över regressionsutfallet. (Använt paket: xgboost). Metoderna är:

- XGBOOST – eXtreme Gradient BOOSTing (Chen & Guestrin, 2016). Metoden identifierar genom s.k. *stochastic gradient descent* olika intervall i träningsdata som kan associeras med respons och genom upprepade beslutsträd söks den kombination som minimerar modellens fel.
- Random Forest (Breiman, 2001). Till skillnad från XGBOOST slumpar metoden ut ett stort antal brytpunkter i träningsdata (miljövariablerna) och beräknar effekten av om ett beslutsträd använder sig av dessa brytpunkter för att forma en responskurva. Dessa adderas på liknande sätt som inom XGBOOST till en ensemble, eller medelvärdesmodell.

2.6.2 Träning och modellernas betydelse

För att modellerna inte skall ge överprediktion (via *overfitting*) finns flera hyperparametrar som anger hur metoderna väljer ut delar av materialet att träna på, och vad som krävs för ett beslut (regularisering). Dessutom utförs en flerfaldig intern korsvalidering, inom vilken metoden körs flera gånger (i detta fallet 5) på en delmängd av materialet (i detta fall 4/5 av allt material), varpå modellen utvärderas på den resterande delen (i detta fall resterande 1/5 observationer).

För att slutligen testa modellens allmängiltighet, i detta fallet hur väl del motsvarar den totala mängden observationer, utförs sedan en validering mot en undanhållen delmängd av

observationsdata som ej används för träning. Denna delmängd är ofta mellan 20 och 40% av observationerna.

Modellmetoderna körs över ett antal hyperparameterinställningar och för varje kombination av parametrar alltså 5 ggr via korsvalidering, totalt ger det upp emot ett hundratal olika modeller, varvid den som har bäst allmängiltighet (minst fel vid validering mot undanhållna data) behålls.

En viktig aspekt är att man inom ML sällan eller aldrig använder sig av viktning av punkterna visavi deras representativitet i materialet (*propensity scores*, IPW o dyl.). Detta beror bl.a. på att maskininlärningens beslutsregler inte är uppbyggda på ett sådant sätt att en linjär viktning av punkterna får en linjärt utjämnande effekt, som sker inom statistisk kausal inferens.

Eftersom modellerna tar fram för maximal förklaringsgrad visar de alltså effekter på de samplade punkterna och inte för en tänkt population eller en "rent allmän" effekt. En styrka istället är att man kan bryta ner en modell i stratifieringar för olika miljötyper och då istället mer i detalj kan identifiera effekter exempelvis i olika miljöavsnitt.

Eftersom ML för att nå maximal förklaringsgrad söker opportunistiskt (*greedy methods*) använder de sig mest av de med responsen mest associerade variablerna, ofta i komplicerad interaktion. Associationen har heller ingen form utan identifieras och skulpteras av metoden. Resultaten kan alltså bilda responskurvor av mycket komplicerad karaktär även när det inte föreligger teknisk överinlärning (*overfitting*). I synnerhet är detta ett problem om observationsdata ligger skevt fördelade i miljön och då miljövariabler är korrelerade då kan metoden välja fel associationer. Genom att granska responskurvorna kan man ganska lätt se om metoden har "gissat fel" i sin associativa inferens.

Inom ML är det därför mycket viktigt att välja ut vilka miljöparametrar (prediktorer) som används och det finns även några speciella direktiv som begränsar mängden fel. I innevarande projekt har s.k. *monotonic constraints* tillämpats, med vilka man kan ange att exempelvis identifierade associationer mellan ökad bottenstrålning och ökad förekomst av känsliga arter är direkt fel. Modellmetoden väljer då bort dessa association och letar efter alternativa förklaringar. Det finns även andra metoder för att styra vilken interaktion som tillåts, exempelvis om temperatur och salinitet skall kunna interagera och på detta sätt påverka en art, medan temperatur och bottenstrat inte skall det. Att begränsa sådana samband är dock svårt och skulle kräva separata forskningsprojekt så dessa har inte tagits med.

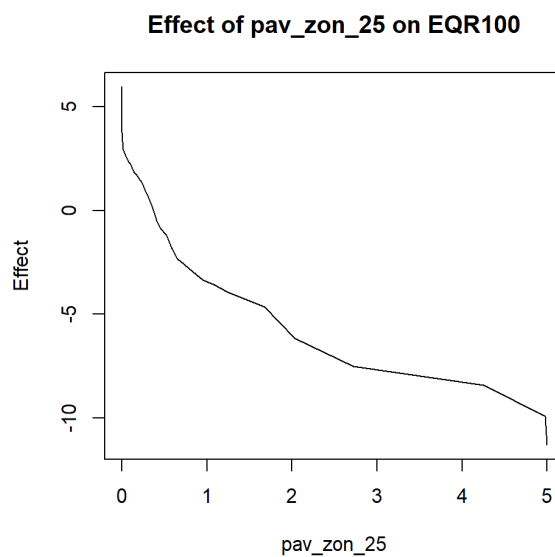
Vid granskning av responskurvor framgår inte bara om modellen verkar ha etablerat en felaktig kurva för den sökta modellerade parametern givet "dosen" i modellen (här: fysisk störning) utan man granskar lämpligen modellens resultat via samtliga responskurvor. Detta är förstås viktigt när modellen inte bara skall prediktera ett utfall, utan dessutom behöver etablera ett kvantitativt samband mellan miljövariabler och utfall.

2.7 Analys och visualisering

Den tekniska utvärderingen av en maskininlärningsmetod beskriver bara hur väl den antingen kan beskriva utfall enligt träningsdata (validering med träningsdata) eller mer allmänt, för data som ej ingår i träningen (validering med valideringsdata). Utvärderingen avslöjar dock inte om modellen etablerat felaktiga samband mellan miljö och utfall. Utvärderingen säger heller inget om andra dolda fel, exempelvis hur väl den förklarar ej modellerade värden (intrapolering, extrapolering, "unknown unknowns") och inte heller om effekten från vissa intervall i miljövariabler har fångats in bättre eller sämre. Sämre resultat kan exempelvis nås om det finns fåtal träningspunkter som ger upphov till "brus" ("known unknowns") i vissa intervall då modellen kan "offra" korrektheten i dessa till förmån för andra intervall som totalt sett minskar felet i modellen.

Sådana kvaliteter kan istället avslöjas mer informellt genom okulär granskning av modellens resultat. I fallstudierna används följande grafiska metoder:

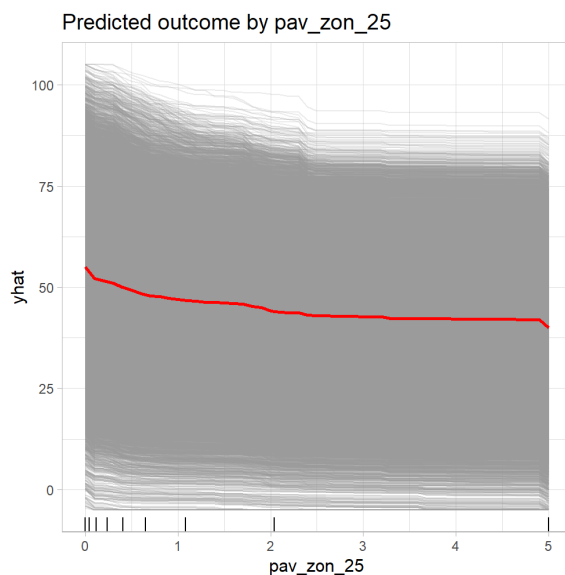
ALE, *Accumulated Local Effects* (Apley & Zhu, 2020), är en grafisk motsvarighet till statistikens Average Marginal Effects fast nedbrutet på olika intervall av påverkanszoner. ALE visar alltså *skillnad* i utfall för modellerad parameter hos observationsdata enligt hur parametern varierar med en annan (kausal) variabel för respektive parameterintervall. Effekten som syns i grafen beräknas för de observationsdata som förekommer i intervallet, i en analysruta (s.k. *moving window*) och återspeglar alltså modellerad faktisk skillnad om den kausala faktorn ändras. Ofta centreras y-axeln vid medianvärdet av den kausala faktorn så y-axelns absoluta värde har ingen direkt betydelse.



Figur 6. Exempel på ALE. I grafen sjunker EQR100 för observerade data med drygt 15 i intervallet 0 till 5 för den kausala faktorn (*pav_zon_25*).

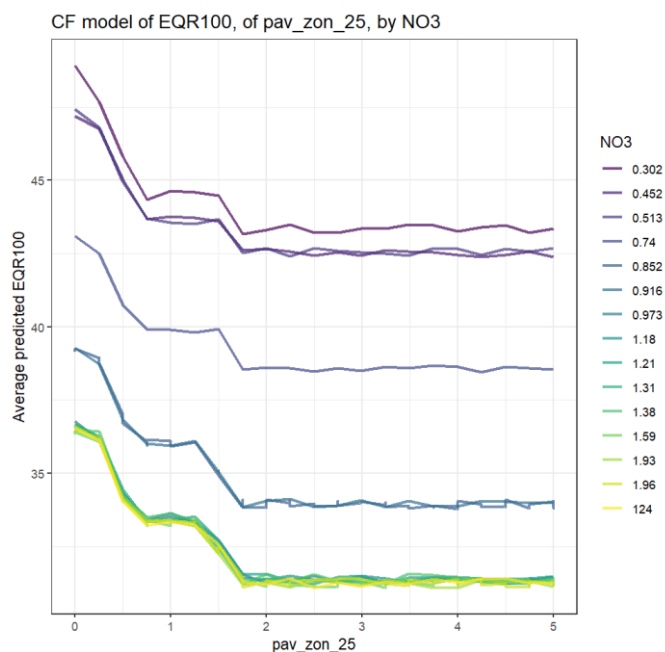
PDP, *Partial Dependency Plot* (Greenwell et al., 2018), visar absolut medelpredikerat utfall, "yhat", (rött i bilden nedan) med *Individual Conditional Expectiles*-kurvor, ICE (grå i bilden). De senare visar hur prediktionen varierar om man kontrafaktiskt ändrar (simulerar) kausal faktor för varje observation och gör om prediktionen. Y-axeln visar alltså absolut medelvärdet för den modellerade effekten på observationsdata och utfallet för varje enskild observation om påverkan ändras.

PDP kan alltså ses som en graf över kontrafaktisk simulering och därur beräknat absolut medelvärde, medan ALE istället visar lokala relativa trender från de faktiska observationernas värden, utan simulering mer än lokalt, i sökfönstret (som typiskt utgör någon procent av intervallet).



Figur 7. Exempel på PDP. I grafen sjunker medelvärdet av det modellerade värdet (yhat, rött i grafen) för observerade data från ca 55 till ca 40 i intervallet 0 till 5 för den kausala faktorn (pav_zon_25), medan kontrafaktiska simulerade resultat för varje enskild observation i intervallet på pav_zon_25 syns som tunna grå linjer. I grafen drygt 12 000 linjer.

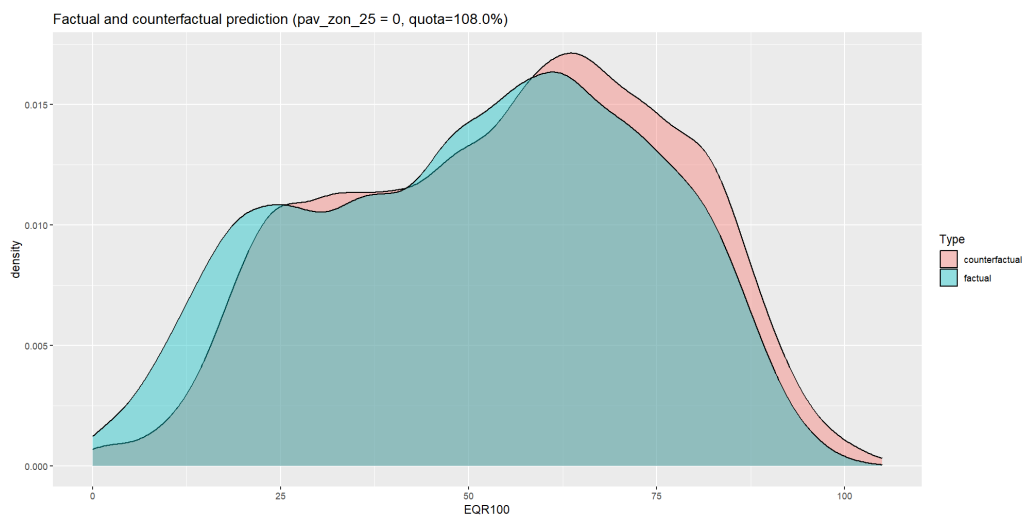
Kontrafaktisk prediktion är en teknik med vilken man precis som vid PDP simulerar olika utfall givet modellen och träningsdata och där man varierar en eller fler miljövariabler. Resultaten går sedan att illustrera grafiskt eller bryta ner i olika grupper så att modellens föreslagna samband kan granskas och modellerade effekter studeras.



Figur 8. Exempel på responskurvor från kontrafaktisk prediktion, här hur EQR100 varierar med halten NH4 enligt olika grad av fysisk störning (pav_zon_25). Notera hur påverkanszonen och NH4 samvarierar genom att sänka EQR100 när båda variablerna ökar i värde; högt värde på NO3 och högt värde på pav_zon_25 ger lägst EQR100 enligt modellen.

Effektberäkning enligt modellen görs även genom att summera skillnad i predikterade resultat från modellen för observationsdata med dels faktiska värden på kausal faktor, dels denna faktor satt till något kontrafaktiskt värde. Exempelvis kan man med denna metod jämföra dagens situation med en

situation om man sätter fysisk påverkan till noll och ber modellen prediktera utfallet med denna kontrafaktiska påverkan.



Figur 9. Kontrafaktisk effektberäkning på observationsdata. Grafen visar hur EQR100 skulle ändras, enligt modellen, om man tog bort alla fysiska påverkanstryck, dvs. reducerade påverkanszonerna till 0. Enligt aktuell modell skulle total EQR100 för observationspunkterna öka med 8 %. I graferna bildar dagens förekomst (factual) den turkosa ytan och den kontrafaktiska (noll påverkan) bildar den laxrosa ytan

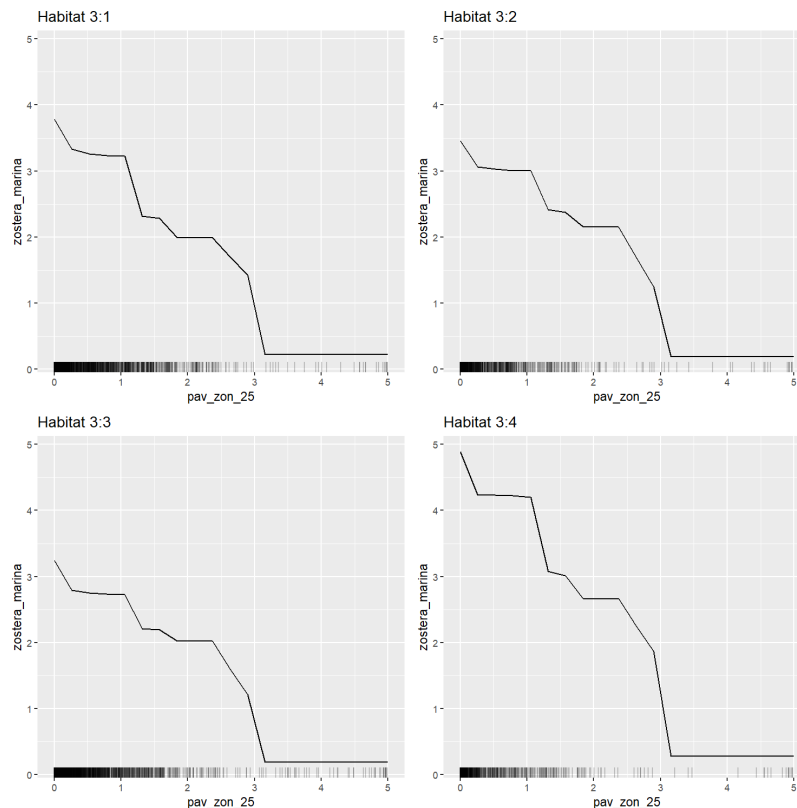
Stratifierad respons innebär att skillnaden mellan faktiska och kontrafaktiska utfall givet modellen bryts ner i olika strata. Genom detta kan man exempelvis undersöka hur modellen menar att en art påverkas negativt givet olika typmiljöer, djupintervall, vågexponering osv.

Inom fallstudie 1 kommer fyra habitat att användas som grund för stratifiering. Dessa består av områden grundare än 3 meter och som är indelade i olika klasser enligt ytnära vågexponering (Isaeus, 2004). Se tabell nedan

Tabell 1. Indelning av abiotiska habitat inom fallstudie 1 för stratifiering av effekter på olika miljötyper.

Vågexponering	Habitat
Ultra till extremt skyddat (<5 000 m ² s ⁻¹)	3:1
Mycket skyddat (5 000–10 000 m ² s ⁻¹)	3:2
Skyddat (10 000–100 000 m ² s ⁻¹)	3:3
Exponerat (>100 000 m ² s ⁻¹)	3:4

För exempel på hur en stratifierad respons kan se ut, nedbrutet på dessa fyra habitat, se Figur 10 nedan.



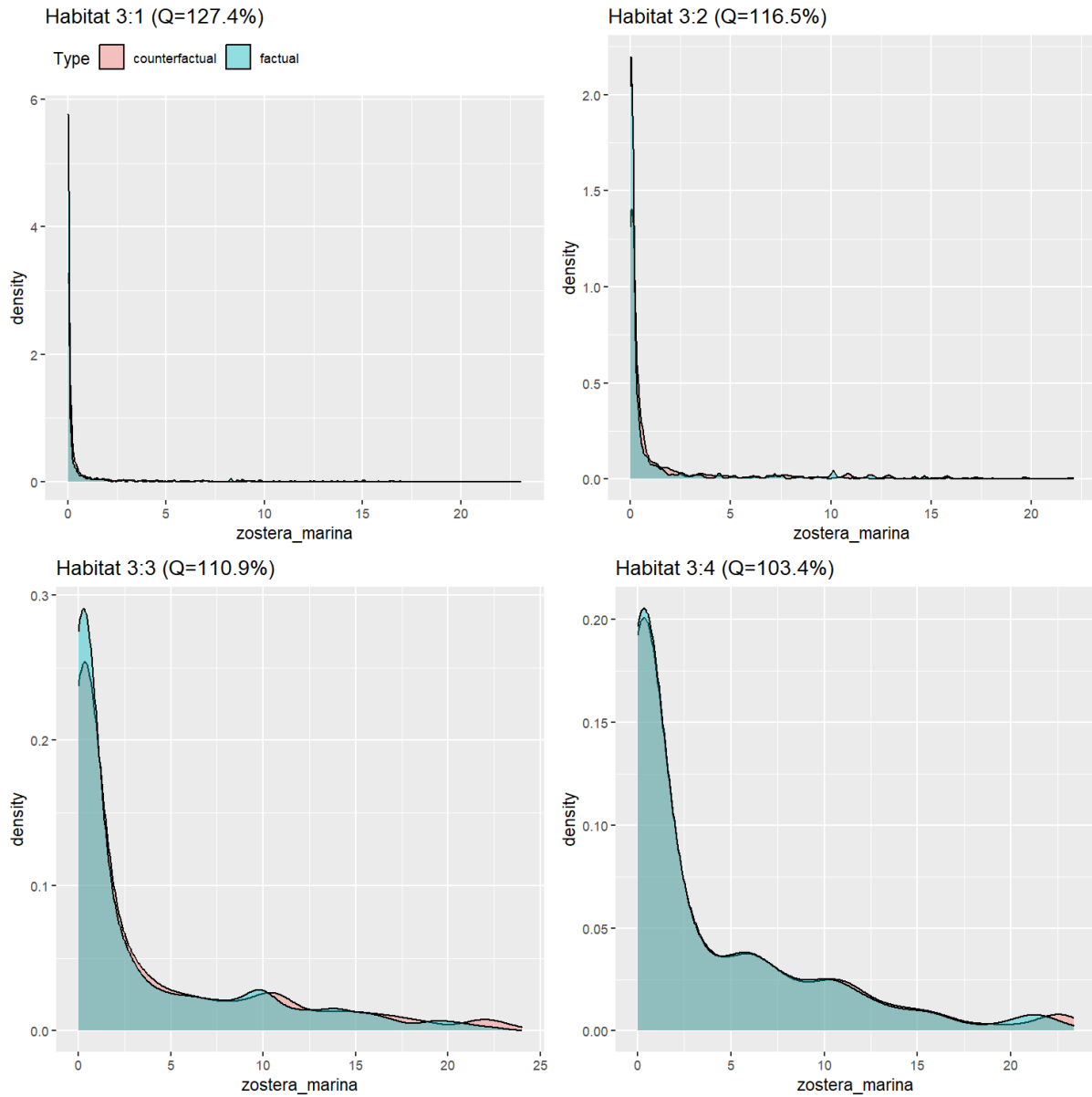
Figur 10. Exempel på stratifierad respons. I grafen visas medelvärdet av täckningsgrad av ålgräs ($n=10298$) i Östergötland och Sörmland, nedbrutet på fyra miljötyper; habitat 3:x anger djup 0-3 meter där x anger vågexponeringsklass enligt Eunis, 1 = mycket skyddat, 4 = exponerat. Grafen visar att *Zostera marina* föredrar exponerad kust och avtar med påverkanszon (pav_zon_25) men också att det saknas observationsdata i exponerade områden med hög påverkan (svarta vertikala streck=perentiler observationer).

Stratifierad effekt innebär en analys och kvantifiering av skillnader mellan faktiska och kontrafaktiska värden på den kausala variabeln, som i exemplen ovan, där effekterna summerats liksom i Figur 9 ovan, fast nedbrutet på någon miljötyp, i exemplet nedan i enlighet med Figur 10 ovan.

Nedan presenteras ett exempel på hur effekter kan delas upp för olika miljötyper. I exemplet framgår hur modellerad täckningsgrad av ålgräs, *Zostera marina*, varierar med och utan fysisk störning. Modellen härleder alltså sitt resultat från dagens situation med fysisk störning och slutleder sedan hur situationen skulle se ut om den fysiska störningen reducerades till noll, genom att "lära sig" vilken täckningsgrad lokaler utan fysisk störning har, givet olika miljöegenskaper.

Sådana effekter kan sedan förstås summeras/aggregeras administrativt eller rumsligt så att det går att bilda sig en uppfattning om inte bara hur stor effekten från påverkan är, utan även var den är stor och hur stor den är sammantaget över en administrativ enhet, habitat, region osv.

Carambha Bilaga 3



Figur 11. Exempel på stratifierad effektanalys, enligt vilken *Zostera marina* skulle kunna öka med mellan 3,4 och 27,4% om fysisk påverkan helt togs bort från Östergötland och Sörmland, givet modellens identifierade effekter. I graferna bildar dagens förekomst (factual) den turkosa ytan och den kontrafaktiska (noll påverkan) bildar den laxrosa ytan.

2.8 Utvärdering av metoderna

Metoderna ger ibland ganska olika resultat. En stor del av detta dilemma kommer sig av att observationsdata och även miljövariablerna egentligen inte är helt lämpade för uppgiften, varför brist i data trigger brister i de olika metoderna och olika typ av bias inträder.

Men en annan orsak till de skiftande resultaten är även att alla, men i synnerhet de mer avancerade metoderna, kan parametreras på olika sätt, vilket för de enklare metoderna oftast resulterar i mindre skillnader i resultat men för de mer avancerade metoderna kan resultaten skifta radikalt beroende på hur metoderna styrs. Trots att exempelvis en statistisk utvärdering kan ge liknande värden.

En kort utvärdering av metoderna följer nedan.

WeightIt/svyglm och **twangContinuous**: Robusta metoder, enkla att förstå. Ger oftast liknande svar. Det går att experimentera med hur PS-viktningen görs men det ger oftast mindre skillnader i resultaten. De ger dock bara medelrespons, inte varierande respons över ett intervall. Slutsats: **Rekommenderas.**

Bayesian Causal Exposure Model: Teoretiskt sett borde denna vara en bayesisk variant av ovanstående, och dess dubbla robusta karaktär redovisas analytiskt i processen. Metoden ger mycket stabila resultat som är starkt korrelerade med ovanstående och har några fler parametrar som går att styra, varför man kan få experimentera en del. Även denna metod ger bara medelrespons. Slutsats: **Rekommenderas.**

Causal Forest: Mycket avancerad metod som går att styra på många olika sätt. Minsta lilla parameterändring kan ge radikalt olika svar. Metoden ger dock möjlighet att både prediktera individuella utfall och ta fram olika typ av medelrespons, men det är mycket svårt att tygla metoder. Slutsats: **Måste utredas vidare.**

Continuous spline: Den variant som användes här baseras på att låta alla variabler blint ingå i interaktion med varandra. Detta ger upphov till stor risk för opålitlighet i modellen, även vid låg multicollinearitet. Den är även väldigt känslig för övrig parametrering. Om denna metod skall användas bör man bättre undersöka hur man väljer ut variabler som skall interagera. Eftersom det finns mer avancerade metoder (se nedan) som ger mer rättvisa resultat bör nog snarare dessa användas. Slutsats: **Avvakta.**

Conditional dose-response function, GPCERF: Metoden är inte bara tekniskt sett robust (doubly robust) utan då metodiken bygger på att foga samman ett antal oberoende delmodeller över exponeringsgradienten ger den resultat som är ärliga, då de istället för att släta över brusiga samband med en fin kurva istället snarare framhäver modellbrus. Det går dessutom att enkelt välja olika intervall för dessa delmodeller, eller brytpunkter i kurvan, varför man experimentellt kan väga detaljupplösning med slagiga kurvor eller en mer generaliserad modell, hela tiden med full kontroll över inter bara medelresponsen i intervallen utan även det sannolika utfallet inom en satt konfidensnivå. Eftersom konfidensintervallet visas i grafen kan man enkelt skatta om det, "de hackiga kurvorna till trots", kan rymmas en sann responskurva inom detta hackiga intervall. Den grafiska metoden och de ärliga kurvorna gör att man även kan utvärdera underliggande data genom att identifiera de intervall inom vilka kurvorna beter sig underligt. En nackdel är att modellen kan ta

ganska lång tid att köra, vid stora datamängder och hyperparameteroptimering rör det sig om flera timmar. Slutsats: **Rekommenderas.**

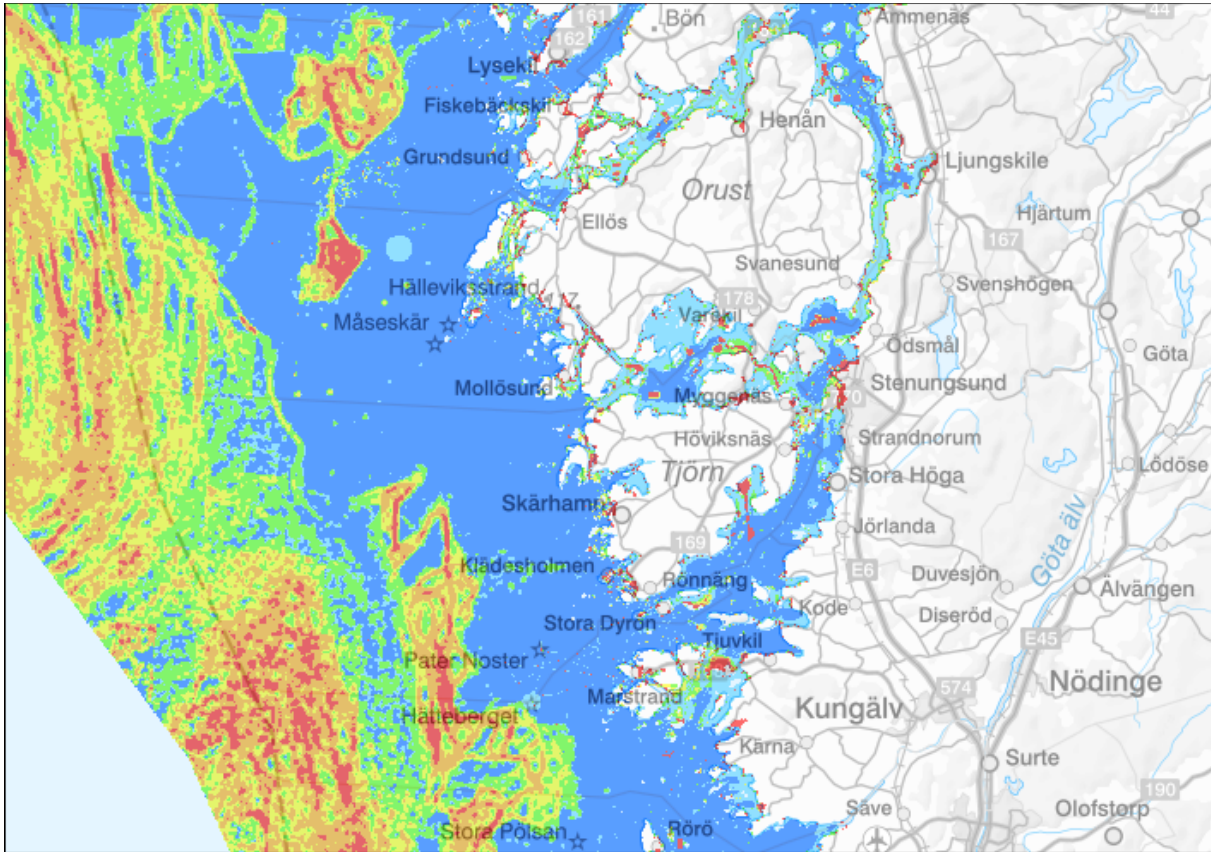
Continuous dose-response model, BART. En metod som liknar den ovanstående på flera sett. Den har dock inte samma avancerade metoder för att hantera bias och ger inte lika pålitliga resultat som GPCERF. I valet mellan metoderna är det bättre att använda GPCERF. Slutsats: **Avvakta.**

Continuous dose-response model, GAM. Jämfört med de två metoderna ovan ger denna variant mycket skiftande och osäkra resultat och är inte heller ärlig eftersom den formar mjuka splines när data egentligen förslås mycket mer komplicerade förhållanden. Slutsats: **Avvakta.**

3 Metod för en kombinerad påverkansmodell

I denna del beskrivs hur en modell för bottenstrålning kan tas fram med hjälp av data från Helcom och Ospar, samt hur den kan kombineras med underlag från den nationella miljöövervakningen, fysisk påverkan (Törnqvist m.fl., 2020) så att ett enhetligt skikt med påverkanstryck erhålls.

Denna heltäckande modell kan sedan användas tillsammans med sensitivitet hos inventerade eller modellerade arter och habitat för att skatta biologisk och fysisk förlust enligt ovan.



Figur 12. Exempel på kombinerad påverkansmodell, med bottenstrålning och fysisk påverkan i grundområden, området utanför Orust.

A.C1 Ingående komponenter

- Grunda havsområden; påverkanskartläggning från MÖ (Törnqvist m.fl., 2020).
- Djupa havsområden; påverkanskartläggning från HMD (se A.C2 nedan).

A.C2 Metod för bottenstrålning

Preparera indata

1. Indata Helcom + ICES
2. Projicera, klipp och rastrea. Använd Sweref99TM, 10 meter upplösning. Vattenmask från nationella marktäckedata, NMD ("Nationella Marktäckedata (NMD)", u.å.).
3. Räkna ut medelstrålning per år, ta ut år i rapportintervallrt, håll isär SAR och subSAR om det finns data för detta.
4. Följ Helcoms försiktighetsprincip. Låt det värsta året ange den trälade årsmängden.
5. Slå ihop dataset till ett heltäckande.

Beräkna trålningsintensitet, för SAR och ev. subSAR om data finns.

1. Indata VMS-pings; välj ut rätt år om det går, om metadata finns
2. Ta bort punkter från trålfria områden (brus/felsignaler)
3. Rastrera i 10 m upplösning
4. Beräkna point density av trålnings-pings, 250 m sökradie
5. Summera punktdensitet inom varje Ices-ruta till rutan, rastrera denna
6. Dividera varje pixels punktdensitet med summan av punktdensitet i varje ruta. Vi får då varje pixels del av hela rutans SAR
7. Multiplicera
8. Multiplicera varje Ices-rutas SAR med punktdensitet inom varje Ices-ruta.
9. Mutiplicera föregående resultat med divisionen från punkt 6. Resultatet blir ett raster där varje pixel har sin andel av Ices-rutans SAR, och där den totala trålningen i rutan är det SAR som tidigare rapporterats.
10. Eftersom punkterna inte perfekt visar trålade lokaler kommer det att bli en del artefakter, i synnerhet om det är få punkter i ett område. Sätt därför ett takvärde för maximal SAR genom exempelvis raster-kalkylatorn, förslagsvis $\max(\text{SAR}) = 20$.

Behandla resultaten

1. Klassindela trålningsintensiteten enligt sensitivitetens sex intervaller. Resultatet blir då ett eller två (om subSAR finns) dataskikt med sex klasser, värde 0 för ingen eller nästan ingen trålning, t.o.m. 5 för mycket intensiv trålning.

Metod för sammanslagning

1. Så ihop trålningsmodellen med de grunda havsområdenas påverkansmodell genom addition.
2. Sätt takvärde för 5, dvs. om summan av de båda modellerna är större än 5 så är max 5.
3. **Gör en medelvärdesanalys, sökradien 125 meter som "mjukar" upp resultatet.**
4. Skapa ett skikt med permanent förlust; objekt från tidigare kartläggningar av permanent karaktär (muddringar, byggnationer osv.) samt motsvarande objekt från aktuell kartläggningen av fysisk störning. Rastrera dessa objekt och tilldela värdet 6.
5. Lägg på dessa rasterade permanenta till resultatet via en mosaik-funktion, eller raster-kalkylatorn. Klass 6 är då att betrakta som "direkt under tryck", dvs. bortgrävt, utfyllt, överbyggt osv.

Resultatet utgör ett enhetligt skikt med påverkan i grunda och djupa områden med påverkanszon 0-5 samt nya och gamla permanenta objekt med värde 6 som anger direkt under tryck, direkt förlust.

4 Referenser

Apley, D. W., & Zhu, J. (2020). Visualizing the Effects of Predictor Variables in Black Box Supervised Learning Models. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 82(4), 1059–1086. <https://doi.org/10.1111/rssb.12377>

Athey, S., Tibshirani, J., & Wager, S. (2019). Generalized random forests. *The Annals of Statistics*, 47(2), 1148–1178. <https://doi.org/10.1214/18-AOS1709>

Austin, P. C. (2011). An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies. *Multivariate Behavioral Research*, 46(3), 399–424. <https://doi.org/10.1080/00273171.2011.568786>

Austin, P. C., & Stuart, E. A. (2015). Moving towards best practice when using inverse probability of treatment weighting (IPTW) using the propensity score to estimate causal treatment effects in observational studies. *Statistics in Medicine*, 34(28), 3661–3679. <https://doi.org/10.1002/sim.6607>

Breiman, L. (2001). Random Forests. *Machine Learning*, 45(1), 5–32. <https://doi.org/10.1023/A:1010933404324>

Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794. <https://doi.org/10.1145/2939672.2939785>

Delignette-Muller, M. L., & Dutang, C. (2015). fitdistrplus: An R Package for Fitting Distributions. *Journal of Statistical Software*, 64, 1–34. <https://doi.org/10.18637/jss.v064.i04>

Digitale, J. C., Martin, J. N., & Glymour, M. M. (2022). Tutorial on Directed Acyclic Graphs. *Journal of Clinical Epidemiology*, 142, 264–267. <https://doi.org/10.1016/j.jclinepi.2021.08.001>

Ebrahim Valojerdi, A., & Janani, L. (2018). A brief guide to propensity score analysis. *Medical Journal of the Islamic Republic of Iran*, 32, 122. <https://doi.org/10.14196/mjiri.32.122>

Galagate, D. (2016). CAUSAL INFERENCE WITH A CONTINUOUS TREATMENT AND OUTCOME: ALTERNATIVE ESTIMATORS FOR PARAMETRIC DOSE-RESPONSE FUNCTIONS WITH APPLICATIONS. <http://hdl.handle.net/1903/18170>

Greenwell, B. M., Boehmke, B. C., & McCarthy, A. J. (2018). A Simple and Effective Model-Based Variable Importance Measure (arXiv:1805.04755). *arXiv*. <https://doi.org/10.48550/arXiv.1805.04755>

Hainmueller, J. (2012). Entropy Balancing for Causal Effects: A Multivariate Reweighting Method to Produce Balanced Samples in Observational Studies. *Political Analysis*, 20(1), 25–46. <https://doi.org/10.1093/pan/mpr025>

Hansen, J. P. (2012). Benthic vegetation in shallow inlets of the Baltic Sea: Analysis of human influences and proposal of a method for assessment of ecological status. Department of Botany, Stockholm University. <https://urn.kb.se/resolve?urn=urn:nbn:se:su:diva-126535>

Hansen, J. P., & Snickars, M. (2014). Applying macrophyte community indicators to assess anthropogenic pressures on shallow soft bottoms. *Hydrobiologia*, 738(1), 171–189. <https://urn.kb.se/resolve?urn=urn:nbn:se:su:diva-107411>

- Hirano, K., & Imbens, G. W. (2004). The Propensity Score with Continuous Treatments. In *Applied Bayesian Modeling and Causal Inference from Incomplete-Data Perspectives* (pp. 73–84). John Wiley & Sons, Ltd. <https://doi.org/10.1002/0470090456.ch7>
- Isaeus, M. (2004). Factors structuring Fucus communities at open and complex coastlines in the Baltic Sea. <https://urn.kb.se/resolve?urn=urn:nbn:se:su:diva-89>
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An Introduction to Statistical Learning: With Applications in R*.
- Lumley, T. (2004). Analysis of Complex Survey Samples. *Journal of Statistical Software*, 9, 1–19. <https://doi.org/10.18637/jss.v009.i08>
- Pearl, J., & Mackenzie, D. (2018). *The Book of Why: The New Science of Cause and Effect*. Penguin UK.
- Ren, B., Wu, X., Braun, D., Pillai, N., & Dominici, F. (2023). A Bayesian Gaussian Process for Estimating a Causal Exposure Response Curve in Environmental Epidemiology (arXiv:2105.03454). arXiv. <https://doi.org/10.48550/arXiv.2105.03454>
- Robins, J. M., Hernán, M. Á., & Brumback, B. (2000). Marginal Structural Models and Causal Inference in Epidemiology. *Epidemiology*, 11(5), 550. https://journals.lww.com/epidem/fulltext/2000/09000/marginal_structural_models_and_causal_inference_in.11.aspx
- Talbot, D., & Beaudoin, C. (2022). A generalized double robust Bayesian model averaging approach to causal effect estimation with application to the study of osteoporotic fractures. *Journal of Causal Inference*, 10(1), 335–371. <https://doi.org/10.1515/jci-2021-0023>
- Textor, J., van der Zander, B., Gilthorpe, M. S., Liškiewicz, M., & Ellison, G. T. (2016). Robust causal inference using directed acyclic graphs: The R package ‘dagitty.’ *International Journal of Epidemiology*, 45(6), 1887–1894. <https://doi.org/10.1093/ije/dyw341>
- Thoemmes, F., & Ong, A. D. (2016). A Primer on Inverse Probability of Treatment Weighting and Marginal Structural Models. *Emerging Adulthood*, 4(1), 40–59. <https://doi.org/10.1177/2167696815621645>
- VanderWeele, T. J., & Hernán, M. A. (2013). Causal Inference Under Multiple Versions of Treatment. *Journal of Causal Inference*, 1(1), 1–20. <https://doi.org/10.1515/jci-2012-0002>
- Wood, S. N. (2017). *Generalized Additive Models: An Introduction with R, Second Edition* (2nd ed.). Chapman and Hall/CRC. <https://doi.org/10.1201/9781315370279>
- Wu, X., Mealli, F., Kioumourtoglou, M.-A., Dominici, F., & Braun, D. (2024). Matching on Generalized Propensity Scores with Continuous Exposures. *Journal of the American Statistical Association*, 119(545), 757–772. <https://doi.org/10.1080/01621459.2022.2144737>
- Zhu, Y., Coffman, D. L., & Ghosh, D. (2015). A Boosting Algorithm for Estimating Generalized Propensity Scores with Continuous Treatments. *Journal of Causal Inference*, 3(1), 25–40. <https://doi.org/10.1515/jci-2014-0022>