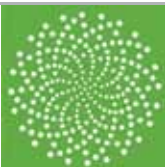
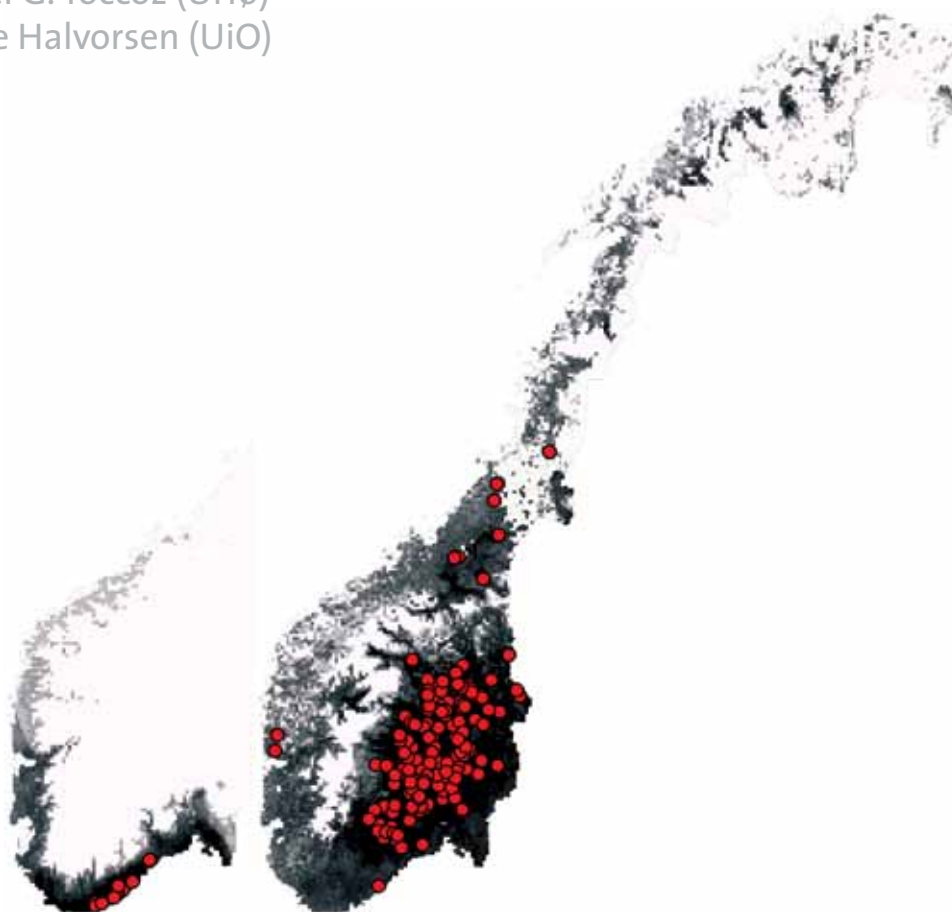


Prediksjonsmodeller som verktøy for kartlegging, overvåking og forvaltning av biologisk mangfold – anvendelse, utviklingspotensial og utfordringer

Naturhistorisk museum Rapport | 1 | ISSN 1891-8050 | ISBN nr. 978-82-7970-010-4



Jogeir N. Stokland (UiO)
Vegar Bakkestuen (UiO)
Trine Bekkby (NIVA)
Eli Rinde (NIVA)
Olav Skarpaas (NINA)
Anne Sverdrup Thygeson (NINA)
Nigel G. Yoccoz (UiTø)
Rune Halvorsen (UiO)



Prediksjonsmodeller som verktøy for kartlegging, overvåking og forvaltning av biologisk mangfold – anvendelse, utviklingspotensial og utfordringer

Jogeir N. Stokland (UiO)
Vegar Bakkestuen (UiO)
Trine Bekkby (NIVA)
Eli Rinde (NIVA)
Olav Skarpaas (NINA)
Anne Sverdrup Thygeson (NINA)
Nigel G. Yoccoz (UiTø)
Rune Halvorsen (UiO)

Denne Rapportsserien utgis av

Naturhistorisk museum
Postboks 1172 Blindern
0318 Oslo
www.nhm.uio.no

Redaksjon:

Jogeir N. Stokland (UiO)
Vegar Bakkestuen (UiO)
Trine Bekkby (NIVA)
Eli Rinde (NIVA)
Olav Skarpaas (NINA)
Anne Sverdrup Thygeson (NINA)
Nigel G. Yoccoz (UiTø)
Rune Halvorsen (UiO)

Teknisk editering av denne Rapporten:

Jan Wesenberg

Publiseringsform:

Elektronisk (pdf)

Sitering:

Stokland, J.N., Bakkestuen, V., Bekkby, T., Rinde, E., Skarpaas, O., Thygeson, A.S., Yoccoz, N.G., Halvorsen, R. 2008. Prediksjonsmodeller som verktøy for kartlegging, overvåking og forvaltning av biologisk mangfold – anvendelse, utviklingspotensial og utfordringer. Naturhistorisk museum (Oslo) Rapport 1: 1-72.

ISSN

Innhold

6	Sammendrag
8	Forord
9	1. Innledning
9	1.1 Bakgrunn
9	1.2 Hva er prediksjonsmodellering?
10	1.3 Kort historikk
11	1.4 Formål med rapporten
12	2. Analysemetoder - litteraturgjennomgang
12	2.1 Innledende oversikt
14	2.2 Problemstillinger som er belyst
18	2.3 Kvantitative metoder for modellering
18	2.3.1 Regelbaserte metoder
18	2.3.2 Numeriske metoder
21	2.3.3 Modellbaserte statistiske metoder
24	2.3.4 Oppsummering
24	2.4 Modellseleksjon og prediksjonsevne
27	2.5 Prediksjonskart og feltvalidering
31	3. Metodiske utfordringer
31	3.1 Fundamental og realisert nisje
32	3.2 Romlig skala
35	3.3 Manglende (relevante) miljødata
35	3.4 Feilkilder i datagrunnlaget
36	3.5 Statistisk over- og undertilpasning
37	3.6 Hvor gode prediksjoner kan vi forvente?
38	4. Stedfestet artsinformasjon
38	4.1 Informasjonskilder
38	4.2 Artsgrupper
38	4.3 Dataegenskaper
41	5. Digitale kart med miljøinformasjon
41	5.1 Primære miljøtemakart
44	5.2 Modellerte (avledete) temakart
50	6. Eksempler fra norsk prediksjonsmodellering
56	7. Praktisk bruk i Norge
56	7.1 Kartlegging av rødlistede arter
56	7.2 Kartlegging av naturtyper
57	7.3 Biomangfoldkart og naturindekser
58	7.4 Praktisk forvaltning
59	7.5 Vurdere klimaeffekter
59	7.6 Overvåking
62	8. Utfordringer og utviklingspotensial
62	8.1 Relevante miljødata
64	8.2 Kvalitetssikrede artsdata
64	8.3 Analysekapasitet og feltvalidering
66	Referanser

Sammendrag

Prediksjonsmodellering er et relativt nytt fagfelt der man modellerer arters (og naturtypers) forekomst som funksjon av ulike miljøfaktorer og kan produsere kart over deres forventede geografiske fordeling. Denne rapporten beskriver fagfeltet som en overordnet metodisk tilnærming og utdyper spesifikke metoder som anvendes. Prediksjonsmodeller har stor nytte for bl.a. kartlegging av rødlistede arter, naturtypekartlegging og arealforvaltning (f. eks. kystsoneplanlegging, reservatplanlegging, og restaurering). I rapporten omtales en rekke anvendelser av prediksjonsmodellering, samt hva slags datamateriale man trenger. Anvendelsene som omtales er i hovedsak prediksjon av arter, siden dette arbeidet har kommet lengst. Mange av de samme problemstillingene gjelder også for prediksjon av habitater og naturtyper. Rapporten omtaler også prediksjonsmodelleringens potensielle nytteverdi i Norge og utfordringer (flaskehalser) som begrenser mulighetene for å ta metodikken i bruk i større omfang.

Fire hovedpilarer

Prediksjonsmodellering bygger på fire hovedpilarer: 1) stedfestede artsfunn, 2) digital kartinformasjon for ulike miljøfaktorer, 3) modelleringsmetoder, og 4) feltvalidering av modellene. I korte trekk går prediksjonsmodellering ut på å finne sammenhenger mellom kjente artsforekomster og variabler som karakteriserer miljøforholdene på de kjente lokalitetene, for så å bruke denne kunnskapen til å beregne sannsynligheten for forekomst i områder som ikke er kartlagt. Resultatet av analysearbeidet er en nisjemodell for arten, uttrykt som en matematisk formel (eller algoritme for ikke-parametriske modeller). Vanligvis angir modellen sannsynligheten for at arten skal opptre, gitt ulike kombinasjoner av verdier for miljøfaktorene. Resultatet presenteres ofte som et prediksjonskart, dvs. en geografisk presentasjon av nisjemodellen. Påliteligheten til prediksjonskartet er direkte avhengig av kvaliteten på informasjonen om artsfunnene, kvaliteten på miljødataene som benyttes i analysene og modelleringsmetoden, men også av hvor representative de kjente observasjonene av arten er for artens totale forekomst. Feltkontroller (feltvalidering) er essensielt for å finne ut hvor pålitelige prediksjonene er. Godt planlagt feltvalidering tjener tre spesifikke formål: a) å etterprøve påliteligheten til prediksjonsmodellen

– under hvilke forhold den gir gode og mindre gode prediksjoner, b) å skaffe nye data som gir grunnlag for å lage forbedrede versjoner av modellen, og c) å oppdage nye forekomster av arten.

Anvendelser

Når man har gode artsdata og kart med relevante miljødata viser det seg at økologisk prediksjonsmodellering har en stor evne til å peke ut potensielle steder for ukjente forekomster, der arten ved senere undersøkelser viser seg å forekomme. Et eksempel kan illustrere dette. I Sveits modellerte man forekomsten av den rødlistede fjellplanten *Eryngium alpinum* med høy romlig oppløsning (30 m kartoppløsning) og oppsøkte i etterkant potensielle lokaliteter der arten ikke var kjent fra før. Under feltkontrollen fant man arten på 7 helt nye lokaliteter, noe som var ganske oppsiktsvekkende da arten er lett å oppdage. Denne studien viser også at utvikling av prediksjonskart er et svært kostnadseffektivt verktøy for å kartlegging av utbredelsen til arter, ved å fokusere feltinnsatsen på en optimal måte. Forøvrig gjengir rapporten flere eksempler på anvendelse av prediksjonsmodellering – både fra utlandet og i Norge.

I kapittel 2 omtaler vi en rekke felter hvor prediksjonsmodellering er anvendt, herunder karakterisering av arters nisje og utbredelse

- identifisering av ukjente lokaliteter med rødlistede arter
- basis for forvaltningsplaner: reservatplanlegging, restaurering og re-introduksjon
- evaluere klimaeffekter på arters utbredelse
- predikere ekspansjon av introduserte arter
- modellere biomangfoldindekser og indikatorer
- teste biogeografiske, økologiske og evolusjonære hypoteser

Metodikk og utfordringer

En rekke metoder benyttes til prediksjonsmodellering, og i rapporten beskrives de viktigste (kapittel 2.3). Det finnes ingen 'generelt beste metode', men metodene utfyller hverandre. For eksempel er noen sterkest på store datasett mens andre er bedre til å utnytte

datasett med få observasjoner. Noen metoder trenger både tilstede- og fraværdata, mens andre klarer å utnytte data med kun tilstedeobservasjoner. Ulike metoder gir i hovedtrekk like prediksjoner, men de kan avvike i viktige nyanser. Det er derfor en god strategi å benytte flere metoder på de samme dataene for å fange opp variasjoner i prediksjonene.

I et eget delkapittel peker vi også på noen grunnleggende utfordringer og potensielle begrensninger i metodene. Disse omfatter forhold mellom arters potensielle og realiserte nisje, effekter av romlig skala, feilkilder i datagrunnlaget samt statistisk over- og undertilpasning.

Norsk kompetanse

Flere norske fagmiljøer har initiert arbeid med prediksjonsmodellering, herunder NIVA, NINA, Universitetet i Tromsø og Universitetet i Oslo (Naturhistorisk museum). NIVA er den institusjonen som har kommet lengst i slikt arbeid og har utviklet flere modeller for spesifikke livsmiljøer og arter i marint miljø (tareskog, ålegrassamfunn, m.v.). Også Universitetet i Oslo har initiert slikt arbeid ved å analysere samlingsdata for sopp (tilgjengeliggjort gjennom Global Biodiversity Information Facility – GBIF), samt ved å modellere treslag og vedlevende arter. Ved NINA er det initiert arbeid på prediksjonsmodellering av spesifikke miljøer for rødlistede arter (hule eiker) og ved Universitetet i Tromsø/NINA for alpin flora og fauna. Et eget kapittel (kapittel 6) omtaler prediksjonsmodelleringsarbeid i Norge.

Nytteverdi

Norge har et meget godt utgangspunkt for prediksjonsmodellering: Stor naturvariasjon gjør at arter forekommer i distinkte miljøer; det finnes mye stedfestet artsinformasjon, mye kartfestet miljøinformasjon og mye relevant fagkompetanse. I kapittel 7 omtaler rapporten en del potensielle anvendelser i Norge, herunder:

- kartlegging av rødlistede arter
- kartlegging av naturtyper
- grunnlag for arealforvaltning (reservatplanlegging, restaurering)
- utvikling av kart over biodiversitetsindikatorer, naturindekser
- overvåking av biologisk mangfold

Utfordringer for økt anvendelse

Selv om vi har et godt utgangspunkt for prediksjonsmodellering i Norge og prediksjonsmodellering har et stort nyttepotensial, er det likevel enkelte utfordringer som må løses for at modellering skal kunne tas i bruk i større skala. I et avsluttende kapittel omtaler vi slike utfordringer nærmere. De viktigste av disse er: 1) tilgang til digitale kart over viktige miljøfaktorer med tilstrekkelig romlig oppløsning (bl.a. bedre terrengmodeller og avledninger fra disse, heldekkende og nyanterte kart for havstrømmer, mer nyanterte klimadata, biologisk informative berggrunnskart, arealdekkende informasjon fra bl.a. satellittbilder), 2) tilgang på presist stedfestede artsfunn og artsfunn basert på en god innsamlingsprosedyre (design), 3) behov for mer ressurser og større kapasitet til å utføre prediksjonsmodellering og feltvalidering.

Forord

Denne rapporten er utarbeidet på oppdrag fra det Interdepartementale utvalget for Nasjonalt program for kartlegging og overvåking av biologisk mangfold, delprogram "Kartlegging og overvåking av rødlistede arter". Arbeidet med rapporten er også delvis finansiert av UiO, UiTø, NIVA og NINA. Formålet med rapporten er å gi en kunnskapsoversikt over temaet prediksjonsmodellering basert på en litteraturgjennomgang, samt å vise hvordan metodikken kan brukes i det videre arbeidet med kartlegging og overvåking av biologisk mangfold, herunder rødlistede arter.

Oppdraget ble gitt til Universitetet i Oslo, Naturhistorisk Museum med en anmodning om at også andre fagmiljøer ble trukket inn i arbeidet. På denne bakgrunnen ble fagpersoner med erfaring fra prediksjonsmodellering ved NINA, NIVA og Universitetet i Tromsø invitert til å være med i utarbeidelse av rapporten. Alle fagmiljøene takket ja til å være med på arbeidet, noe som har styrket rapportens faglig bredde.

Under utarbeidelse av rapporten har vi hatt to målgrupper i tankene: a) fagbiologer med interesse for og kunnskap om prediksjonsmodellering, og b) personer i offentlig forvaltning som arbeider med utfordringer i forhold til bevaring av biologisk mangfold. Dette valget av målgrupper betyr at vi skriver for et skolert publikum. Vi har derfor ikke popularisert stoffet nevneverdig. For å finne en balanse har vi de fleste steder ikke gått dypt inn i temaene vi omtaler, men fokusert på hovedprinsipper og gitt referanser til sentrale publikasjoner. De som er interessert i en mer omfattende behandling av ulike temaer henvises til primærkildene. Ettersom rapportens tema er ganske metodisk orientert har vi skrevet et fyldig sammendrag i en mer popularisert form som ikke forutsetter kjennskap til temaet. Videre har vi forsøkt å illustrere rapporten rikelig slik at innholdet skal bli lettere tilgjengelig.

1. Innledning

1.1 Bakgrunn

FN's konvensjon om biologisk mangfold "Rio-konvensjonen" fra 1992 forplikter alle land til å kjenne til og ivareta sitt biologiske mangfold. I Norge ble dette fulgt opp i St. meld. 58 (1996-97) 'Miljøvernpolitikk for en bærekraftig utvikling', der alle landets kommuner ble pålagt å gjennomføre en kartlegging og verdiklassifisering av det biologiske mangfoldet innen kommunens områder. Arbeidet med biologisk mangfold ble videre fulgt opp med forarbeider til et overvåkingsprogram for biologisk mangfold i Norge (DN-rapport 1998-1) som resulterte i St. meld. 42 (2000-2001) om kartlegging og overvåking av biologisk mangfold. I etterkant av denne ble det etablert et nasjonalt program for kartlegging og overvåking av biologisk mangfold, i første fase for perioden 2003-2006 som er videreført for perioden 2007-2010. Dette programmet er delt opp i flere programområder, og i to av dem har prediksjonsmodellering kommet opp som en relevant metodisk tilnærming.

I *marine miljøer* har prediksjonsmodellering vist seg som et svært nyttig verktøy for utvikling av metoder til å kartlegge utbredelse og lokal forekomst av arter og naturtyper. Ved oppstart av nasjonalt program var det stor mangel på kunnskap om viktige marine naturtypers forekomst og utbredelse, og utvikling av habitatmodeller (modeller for naturtypers forekomst) har fått en sentral rolle i den marine delen til kartleggingsprogrammet. Prediksjonsmodellene og oversikt over tidligere registreringer av de ulike marine naturtypene er benyttet som grunnlag for å anslå kartleggingsstatus for de ulike naturtypene og for å beregne kostnader for å fullføre kartleggingen av nasjonalt og regionalt viktige forekomster. Modellene er brukt for planlegging av feltdesign og innsamling av data for kartleggingsarbeidet i de enkelte kommunene, som videre blir brukt til å videreutvikle og forbedre modellene og for å avgrense de lokale forekomstene.

Også for temaet *truede (rødlistede) arter* innenfor nasjonalt program er det foreslått at prediksjonsmodellering kan være et viktig supplement for å øke kunnskapsgrunnlaget. Det er gjentatte ganger påpekt at truede arter fanges opp på en mangelfull måte i pågående kartleggings- og overvåkingsprogrammer og at undersøkelse av alle enkeltarter er uoverkommelig ressurskrevende (Anonym 2002, Riksrevisjonen 2006). Dette henger sammen med at mange av artene har små eller klumpvis fordelte populasjoner

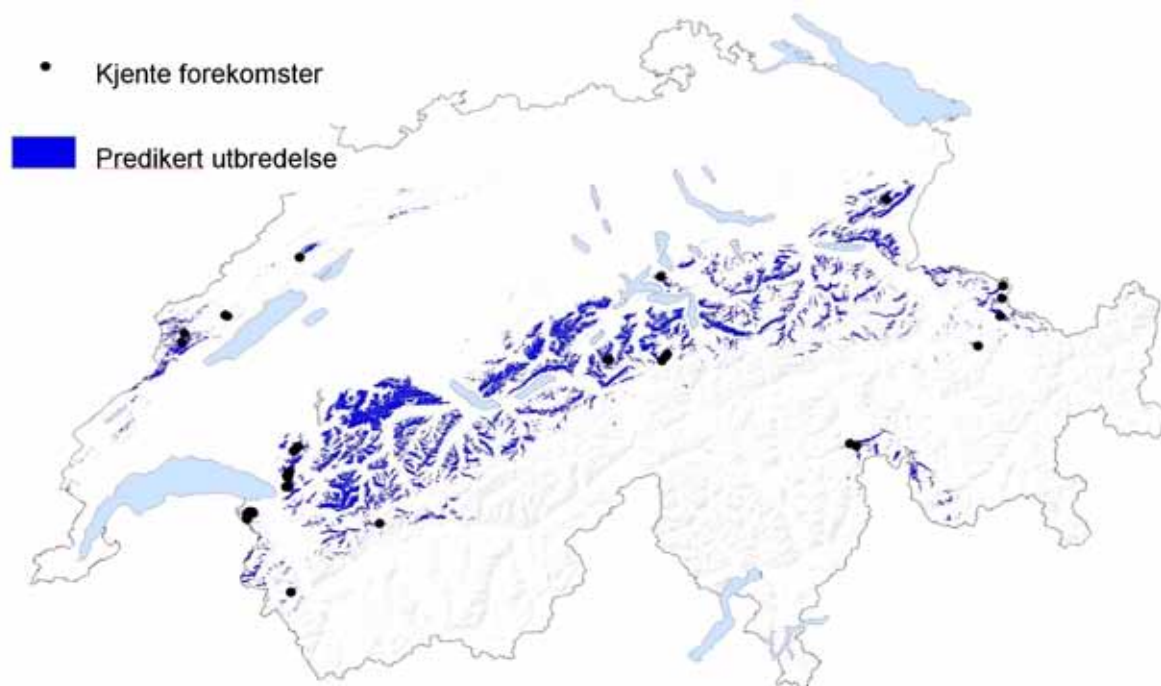
som i liten grad fanges opp med generelle, eksten-sive (arealrepresentative) overvåkingsstrategier. Det er derfor behov for metoder som gjør det mulig å predikere slike arters forekomst, ikke bare for en og en art ad gangen, men for grupper av arter (f.eks. som er knyttet til samme livsmiljøer). Prediksjonsmodellering kan spille en nøkkelrolle i oppbygging av generell kunnskap om truede arter. Ved å benytte prediksjonsmodeller for truede arter tilknyttet spesielle miljøer vil man kunne øke treffsikkerheten ved søk etter nye forekomster og effektivisere kartleggingen samtidig som man får kunnskap som kan generaliseres til også å gjelde andre arter.

1.2 Hva er prediksjonsmodellering?

Kort fortalt innebærer prediksjonsmodellering å bruke kjente forekomster av en art eller naturtype til å utvikle en modell for sammenhenger mellom miljøfaktorer og sannsynligheten for å finne arten eller naturtypen. En rekke kvantitative metoder kan benyttes for å lage slike modeller (se 2.3). Prediksjonsmodellen kan i sin tur brukes til å lage et prediksjonskart for total utbredelse innen et studieområde (figur 1). Man benytter da prediksjonsmodellen sammen med digitale kart over variasjonen i ulike miljøfaktorer innen området for å lage utbredelseskartene på en objektiv måte. Tidligere har man framstilt tilsvarende kart subjektivt ved å skravere områder mellom kjente forekomster under antagelsen om at arten forekommer også der.

Til forskjell fra den subjektive tilnærmingen er prediksjonsmodellene klare og utvetydige, de utnytter informasjon om en rekke miljøfaktorer på en gang, og forekomsten framstilles som en gradert sannsynlighet. Prediksjonskartene kan lages med svært forskjellig detaljeringsgrad; fra global/kontinental skala ned til lokal skala innen en liten kommune. En viktig fordel med prediksjonsmodellene er at resultatene systematisk kan etterprøves.

I tillegg til selve prediksjonskartet viser de fleste metodene også den relative viktigheten av ulike miljøfaktorer og hvilke spesifikke verdier av miljøfaktorene som begünstiger forekomst av arten eller naturtypen. Dette er nødvendig kunnskap for å kunne si noe om hvordan endringer i miljøforholdene vil kunne påvirke arter og naturtypers utbredelse.



Figur 1. Kjente forekomster av fjellplanten *Eryngium alpinum* i Sveits (svarte prikker) og første versjons predikert utbredelse (mørkeblå farge) basert på artens forekomst i forhold til en rekke miljøfaktorer (kilde: Guisan et al. 2006). Dette studiet er omtalt nærmere i kap. 2.2 hvor vi omtaler forbedring av denne prediksjonsmodellen og tilhørende prediksjonskart.

1.3 Kort historikk

Prediksjonsmodellering som kvantitativ metode springer ut fra pionerstudier utført av Mike Austin m.fl. i Australia på 1980-tallet (Austin et al. 1984, Austin et al. 1983, se Guisan and Thuiller 2005 for en oversikt). De studerte blant annet forskjellige plantarters fordeling i forhold til høyde over havet, og fant at arter var avgrenset til ulike høydeintervaller. Disse sammenhengene ble så oversatt til "regler" for arters klimatiske toleransegrenser som i neste tur ble anvendt til å framstille kart over potensiell utbredelse som samsvarte med artenes klimatoleranse. Dette er kjernen i klima-konvolutt metoden (implementert i metoden BIOCLIM) som fremdeles benyttes til en viss grad. Austin var imidlertid fullt klar over at også andre miljøfaktorer var avgjørende for planters geografiske forekomst. Derfor ble ytterligere faktorer brakt inn som grunnlag for å modellere arters utbredelse (Austin and Meyers 1996).

Utover på 1990-tallet fikk prediksjonsmodellering økt vind i seilene som følge av tekniske utviklinger innen geografiske informasjonssystemer (GIS) og tilgjengelighet av digitale kart over forskjellige miljøfaktors romlige fordeling. Parallelt med dette foregikk det også omfattende arbeid med å digitalisere naturhistoriske samlingsdata samt å stedfeste artsforekomstene med geografiske koordinater. Videre hadde mange biologer blitt fortrolig med statistiske analyser av tilstedeværelse-fravær data ("presence-absence" data) ved hjelp av logistisk regresjon innen rammen av stadig bedre statistikkpakker. Resultater fra disse trendene ble knyttet sammen på slutten av 1990-tallet. Siden har en rekke fagpersoner utnyttet den slagkraftige kombinasjonen av stedfestede artsfunn, digitale kart implementert i GIS og statistiske analyser for å predikere arters utbredelse.

Det siste tiåret er prediksjonsmodellering tatt i bruk i forhold til en rekke problemstillinger. Disse spenner fra økologisk, biogeografisk og evolusjo-

nær grunnforskning, via bevaringsbiologi, til direkte anvendelse innen arealforvaltning, naturvern og konsekvensutredninger. En viktig grunn til populariteten innen anvendt biologi er at metodene er tids- og pengebesparende sammenlignet med tradisjonelle innsamlingsmetoder, og de gjør det mulig å dekke store områder der data mangler. Metoden har således vist seg å være av stor nytte for forvaltningen (for anvendelse i Norge, se f. eks. Rinde et al. 2006). Årsaken til metodikkens popularitet innen grunnforskning er at prediktive modeller gir nye muligheter for å analysere flere faktorer i sammenheng og at man lett kan etterprøve resultatene. Dette gir en bedre forståelse av faktorer og prosesser som strukturerer utbredelsen og sammensetningen av arter.

Etter sammenkoblingen av stedfestet artsinformasjon, GIS/miljødata og modellering på slutten av 1990-tallet har det nærmest skjedd en "eksplosjon" i antall analysemetoder som benyttes for å predikere arters forekomst. Dette har de siste årene resultert i flere komparative studier som evaluerer de ulike metodenes prediksjonsevne (f. eks. Elith et al. 2006, Guisan et al. 2007b). Vi har dermed en ganske god forståelse av styrker og svakheter ved alternative metoder, og hele fagfeltet framstår allerede som en ganske moden disiplin. Likevel er det aspekter ved metodene vi trenger å forstå bedre, slik som deres evne til å predikere forekomster på ulike skalaer, vektleggingen av å predikere feil tilstedeværelse og feil fravær i ulike sammenhenger, hvordan man skal forbedre det empiriske modelleringsgrunnlaget, osv. Dette er temaer som er gjenstand for aktiv forskning og som ganske sikkert vil føre til at man bedre vil forstå hvilke metoder som er egent til ulike problemstillinger og ulike typer av data.

1.4 Formål med rapporten

Hovedformålet med denne rapporten er å gi en helhetlig beskrivelse av fagfeltet prediksjonsmodellering. Som nevnt i forordet er fagbiologer med interesse for fagfeltet en sentral målgruppe for rapporten. Vi har derfor valgt å gå ganske grundig inn i metodiske aspekter og å inkludere referanser til viktig litteratur. Vi har avgrenset rapporten til det karakteristiske fokuset for prediksjonsmodellering – nemlig univariat respons knyttet til ulike aspekter av biologisk mangfold. Den responsvariabelen vi sterkest vil fokusere på er forekomst av enkeltarter, men vi vil også omtale modellering av naturtyper og artsrikdom. Rapporten omfatter *ikke* multivariat responsmodellering slik som betinget ordinasjon (*constrained ordination*) og tilsvarende metoder som analyserer forekomst av flere arter og som kan benyttes til å identifisere og predikere utbredelsen til ulike typer samfunnsstrukturer. Vi vil bare så vidt omtale modellering som operativt trinn i primærkartlegging (for eksempel tolking av satellittbilder) og metoder for avledning av nye miljødatakart (hvor en rekke tekniske GIS-operasjoner benyttes). Dette er andre typer modellering som i denne sammenheng blir støttedisipliner.

Det er også et formål med rapporten å vise nyttepotensialet til prediksjonsmodellering. Vi har derfor gitt en del eksempler på hvordan disse metodene er anvendt rundt om i verden. Videre har vi pekt på en rekke områder hvor metodene kan komme til nytte i Norge. Avslutningsvis i rapporten har vi omtalt noen spesifikke utfordringer i forhold til å anvende prediksjonsmodellering i større omfang i Norge.

2. Analysemetoder – litteraturgjennomgang

2.1 Innledende oversikt

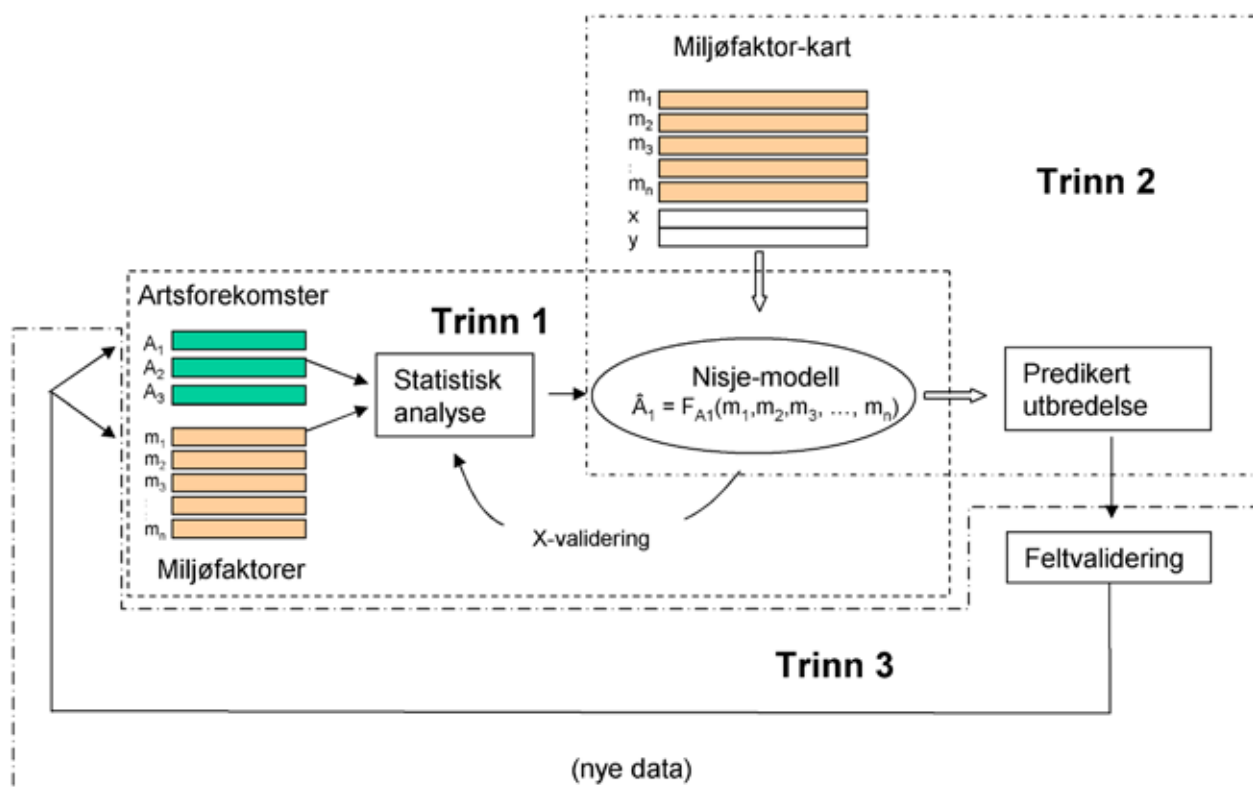
Prediksjonsmodellering består av flere trinn der ulike elementer inngår (figur 2). Når alle trinn og elementer er planlagt og utført som en helhet, har man mulighet til å utvikle gode prediksjonsmodeller. En slik helhetlig tilnærming til innsamling av data og utvikling av prediksjonsmodeller utgjør til sammen en slagkraftig strategi.

I det etterfølgende beskriver vi prediksjonsmodellering for artsforekomster, og da betegnes gjerne prediksjonsmodellen for en *nisjemodell*. I praksis kan nøyaktig samme metodikk benyttes for å predikere forekomster av naturtyper. Da bruker man stedfestede forekomster av naturtyper isteden for artsforekomster og lager en prediksjonsmodell for naturtyper. I dette tilfellet kaller man ikke resultatet for en nisjemodell ettersom begrepet nisje kun brukes for arter.

Trinn 1, nisjemodell

I det første trinnet gjør man en analyse (statistisk, regelbasert, numerisk, se 2.3) av datasett som består av a) artsforekomster (A_i) på et utvalg lokaliteter og b) registrerte miljøfaktorer (m_i) på de samme lokalitetene. Resultatet av analysen er en nisjemodell som uttrykker sannsynligheten for at en art A_i forekommer i forhold til ulike verdikombinasjoner av miljøfaktorene m_i . Prediksjonsevnen til denne nisjemodellen blir ofte evaluert ved kryssvalidering. Dette innebærer at datasettet blir delt i en hoveddel som benyttes til å lage modellen og en valideringsdel som benyttes for å måle hvor god modellen er til å predikere forekomsten av arten på nye lokaliteter (dvs. lokalitetene i valideringsdelen), se 2.4.2.

I prinsippet behøver ikke artsforekomstene og de registrerte miljøfaktorene å være stedfestet med koordinater, så lenge man vet at de stammer fra



Figur 2. De ulike elementene som inngår i en helhetlig strategi for prediksjonsmodellering. De tre stiplede feltene markerer de tre trinnene i prediksjonsmodellering (se tekst for nærmere beskrivelse).

det samme utvalget av lokaliteter. I praksis henter man imidlertid informasjon om miljøfaktorene fra de samme digitale kartene som benyttes for å lage prediksjonskart i trinn 2.

Trinn 2, prediksjonskart

I dette trinnet benytter man digitale kart for studieområdet som viser hvordan miljøfaktorene som ble benyttet i nisjemodellen i trinn 1 varierer innenfor det området man skal lage prediksjonskart for. Det er viktig at hver miljøfaktor er målt, beregnet eller modellert med samme romlige oppløsning (samme pixel-størrelse) og kvalitet som ble benyttet for å lage nisjemodellen. Hvis det er forskjeller mellom miljødata anvendt i trinn 1 og trinn 2, vil sannsynlighetene i prediksjonskartet bli annerledes enn det modellen tilsier.

Ved hjelp av nisjemodellen beregnes sannsynligheten for forekomst av arten for hver verdikombinasjon av miljøfaktorene som er representert i kartene. Resultatet er et heldekkende prediksjonskart som viser sannsynlighet for forekomst (evt. forventet kvantitet) av arten eller naturtypen i hver pixel.

Trinn 3, feltvalidering

Prediksjonskartet produsert i trinn 2 er kun en geografisk visualisering av nisjemodellen fra trinn 1, og har samme grad av kvalitet og usikkerhet som denne. Selv om modellen viser god prediksjonsevne i kryssvalideringen omtalt under trinn 1, er ikke dette en garanti for at modellen viser like god prediksjonsevne i en uavhengig test. Dette kan skyldes systematiske utvalgsskjevheter (mangel på design) i det primære datasettet, unøyaktig stedfesting, feilidentifiseringer, osv. For å få et inntrykk av modellens evne til å predikere artens forekomst (og kartets pålitelighet) er det viktig å gjøre en uavhengig feltvalidering. Man oppsøker da et utvalg av lokaliteter hvor kartet (dvs. modellen) predikerer artens tilstedeværelse og fravær med ulike grader av sannsynlighet, og bruker innsamlede data til å kvantifisere prediksjonsevnen.

Feltvalidering bør ideelt sett designes slik at den tester modellen der den predikerer lav, middels og høy sannsynlighet for tilstedeværelse av arten. Modellen bør også testes ut under et vidt spekter av miljøforhold for å se om modellen predikerer bedre under noen miljøforhold enn andre. Videre bør modellen testes ut i flere geografiske områder, og aller helst i et

helt annet område enn der testdataene stammer fra. Felles for all feltvalidering er at teststedene som oppsøkes bør trekkes ut som et randomisert, stratifisert eller systematisk utvalg, og ikke blir steder som oppsøkes på grunn av enkel tilgjengelighet (for eksempel kort avstand fra vei).

De nye observasjonene som opparbeides under feltvalideringen (både tilstedeværelse og fraværobbservasjoner) kan med fordel legges til datasettet som benyttes for å lage prediksjonsmodellen. Dermed vil man systematisk forbedre modellen ved å lage nye modell-generasjoner på et stadig bedre datagrunnlag (Guisan et al. 2006).

Formål med modellering

Kvantitative modeller kan konstrueres på mange ulike måter, avhengig av formålet (Levins 1966). Økologisk modellering har ofte hatt som formål å *forklare* sammenhenger og øke forståelsen av et system. I denne tradisjonen inngår alt fra teoretiske (deduktive) modeller, f.eks. Levins metapopulasjonsmodell (Levins 1969), som slutter fra mekanismer til mønster, til statistiske (induktive) modeller som tar utgangspunkt i et mønster og gjør slutninger om årsakssammenhenger og effekter. Med parameterverdier estimert fra data, kan slike teoretiske og statistiske modeller være gode beskrivelser ('redescription'; Taylor 2000) av det systemet datasettet representerer. Her er imidlertid formålet å *predikere*, dvs. forutsi observasjoner som ikke inngår i datagrunnlaget for modellen. For å utvikle gode prediksjonsmodeller er det essensielt å teste prediksjoner med uavhengige data, altså validere modellen.

Formålene å forklare og å predikere er til dels sammenfallende. Vi kan likevel ikke slutte at hvis vi kan predikere, så har vi forstått årsakssammenhengene. Det finnes modeller som kan gi gode prediksjoner, uten at vi har full forståelse av årsakssammenhengene. For eksempel er høyde over havet ikke noe økologisk variabel i seg selv – det er nok heller jordtemperatur og fuktighet osv som er relevante (se Körner 2007). Videre er temperatur fra klimadata en god prediktor (forklaringsvariabel) for utbredelsen av mange arter, uten at vi alltid vet nøyaktig hvordan temperaturen i detalj påvirker artene (overlevelse, reproduksjon, osv.). De mest robuste prediksjonene vil vi få fra modeller som baserer seg på en god forståelse av systemet. Har vi forstått de nødvendige og tilstrekkelige sammenhengene (og er i stand til å måle dem)

kan vi også predikere nye observasjoner med stor sikkerhet. Modellen er en 'generative representation' av systemet (Taylor 2000).

Som en slags sammenfatning kan man si at ved *forklarende* modellering er man mest opptatt av selve prediksjonsmodellen (dvs. resultatet av trinn 1 over) – hvilke viktige prediktorvariabler den inneholder og hvordan de influerer på responsen, biologisk tolkning av disse sammenhengene, og datatilpasning/forklart variasjon ('model fit'). Ved *prediktiv* modellering er man særlig opptatt av modellens prediksjonsevne (for eksempel AUC-verdien i kryssvalideringen i trinn 1), hvordan prediksjonskartet ser ut (dvs. resultatet av trinn 2), samt å kunne forbedre modellens prediksjonsevne med nye data (dvs. resultatet av trinn 3) og justerte parameterverdier i nye analyser.

2.2 Problemstillinger som er belyst

Prediksjonsmodellering springer ut fra enkelte pionerstudier på 1980-tallet og tidlig på 1990-tallet. Det er imidlertid de siste 10 årene fagfeltet har fått omfattende anvendelse. Til tross for en relativt kort historikk, har slik modellering blitt tatt i bruk innen en rekke områder og vi ser i dag et sterkt økende antall publikasjoner innen feltet. Vi skal her løfte fram en del sentrale anvendelsesområder og illustrere disse med et utvalg publiserte arbeider.

Karakterisering av arters nisje og utbredelse

Et sentralt fokus for prediksjonsmodellering har vært å karakterisere arters økologiske nisje og geografiske utbredelse. De fleste modellene har vært utviklet for vanlige plante- og dyrearter hvor det er lett å skaffe til veie relevant og tilstrekkelig informasjon. Videre fordrer de fleste metodene ganske store datamengder, og det er knapt meningsfullt å publisere modeller basert på et utilstrekkelig datamateriale.

I et forholdsvis tidlig arbeid utviklet Austin & Meyers (1996) nisjemodeller for ulike arter av eukalyptus, bl.a. med implikasjoner for skogforvaltning. For øvrig er det laget modeller for ulike treslag, slik som et studium av 30 arter av trær i Sveits (Guisan et al. 2007b) og flere løvtreslag i Utah, USA (Zimmermann et al. 2007). Et annet godt eksempel på nisjemodellering er utført av Edwards jr. et al. (2006) som modellerte fire epifyttiske makrolav i Pacific northwest,

USA og benyttet uavhengige observasjonsdata for å evaluere disse modellene. Her hjemme er det utviklet nisjemodeller for blant annet ålegras i marine miljøer (Bekkby et al. *Submitted*) og svalbardryper (Pedersen et al. 2007).

Identifisering av nye forekomster av sjeldne og rødlistede arter

En rekke studier har anvendt prediksjonsmodellering for å finne nye forekomster av sjeldne og truede arter. Elith & Burgman (2002) representerer et tidlig studium der de vurderte prediksjonsevnen for sjeldne karplanter i sentrale høyland i Australia. Et typisk eksempel er et studium av en truet plante i India der man benyttet nylig innsamlede data til å modellere artens økologiske krav og identifiserte flere nye områder hvor arten skulle forventes å opptre (Irfan-Ullah et al. 2007). En noe mer sofistikert tilnærming ble benyttet av Edwards jr. et al. (2006) som modellerte fire lavarter med spesifikk økologi for å finne sjeldnere arter med enda mer spesifikke økologiske krav.

Et case fra Sveits

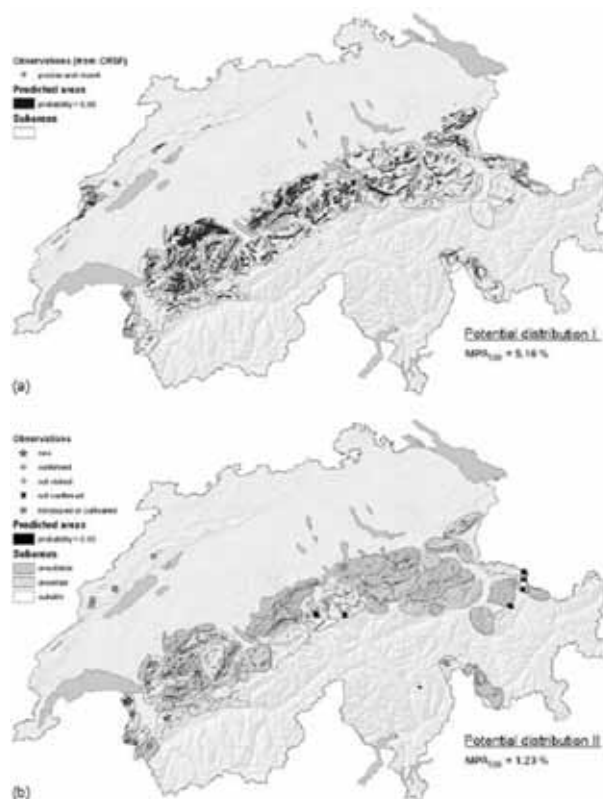
Meningen med denne delen av rapporten er å gi eksempler på anvendelse av prediksjonsmodeller i ulike sammenhenger. For ett studium (Guisan et al. 2006) vil vi dog gjøre et unntak og referere noen sentrale funn mer inngående. Arbeidet omfatter modellering og feltsøk etter den sjeldne og rødlistede karplanten alpestikle *Eryngium alpinum* (en slektning av vår hjemlige strandtorn) i alpine habitater i Sveits. Arten er stor og svært lett å oppdage når man er på stedet hvor den vokser.

I utgangspunktet sammenstilte man alle kjente forekomster med en stedfestingspresisjon som var 25 m eller bedre. Videre genererte man fraværsobservasjoner ved å bruke funnsteder for andre sjeldne plantearter med samme stedfestingspresisjon, der fokusarten ikke var påvist (følgende resonnement ble benyttet: hadde arten vært til stede ville den blitt registrert av observatøren som registrerte en annen sjelden art fordi vedkommende da har kompetanse til å identifisere uvanlige plantearter). Dette ga til sammen et datasett med 92 tilstede-observasjoner og 2380 fravær-observasjoner. Videre la man ned et betydelig arbeid i å etablere en rekke kart over prediktor-variabler med en romlig oppløsning på 25 m. Som

prediktorer benyttet man følgende faktorer: sesongmessig fordeling av nedbør, solinnstråling, antall nedbørsdager i vekstsesongen, temperatursum for vekstsesongen, antall frostdøgn, topografisk posisjon (topp-områder, lier, dalbunner, osv.), stedets vannbalanse og terrengets helningsvinkel. Basert på disse datasettene laget man en prediksjonsmodell vha. GAM og oppnådde en meget høy prediksjonsevne (AUC-verdi lik 0,96, se 2.4 for omtale av AUC).

Nisjemodellen ble overført til et prediksjonskart som identifiserte en rekke delområder (figur 3a) der en stor andel av arealet var klassifisert som velegnet for arten. Det er dette kartet som er anvendt som illustrasjon i figur 1. Basert på prediksjonskartet designet man et inventeringsprogram hvor man innen disse delområdene spesifiserte tilfeldig valgte ruter slik at man dekket omtrent like mange velegnede som ugunstige pixler for arten. Nye forekomster ble stedfestet med GPS, og med regelmessig avstand (minst 500 m) registrerte man punkter uten observasjoner av arten. Et resultat av inventeringsprogrammet var at man fant 7 helt nye lokaliteter, noe som var ganske oppsiktsvekkende da arten er lett oppdagbar. Det nye datasettet ble inkludert i modellen og en fikk en signifikant forbedret prediksjonsmodell (nå med AUC-verdi lik 0,98). I den oppdaterte modellen ble en stor andel av det velegnede arealet fra første modell klassifisert som uegnet for arten (figur 3b).

Vel så interessant, metodisk sett, var en simulering av spillet mellom opprinnelig modell, feltsøk, en oppdatert modellversjon, nytt feltsøk, ... osv. Her fokuserte man på hvor stor feltinnsatsen burde være mellom hver modellversjon for å oppnå en forbedring av modellen. Simuleringen viste at bruken av prediksjonskart effektiviserte søket etter nye forekomster slik at man sparte opp til 70 % av tiden i felt sammenlignet med tilfeldig søk etter arten. Et annet resultat av simuleringen var at den beste strategien var å ha relativt få feltsøk etter arten mellom hver modellversjon og hyppige modelloppdateringer. En slik strategi gjorde at man trengte 6 modellgenerasjoner og feltsesonger for å doble antall nyregistreringer, mens man med mer tilfeldige søk etter arten ville trenge 20 feltsesonger for å doble antall nyregistreringer. Årsaken er at man raskt blir mer treffsikker i å finne arten ved å bygge på en prediksjonsmodell som oppdateres hyppig. Basert på dette studiet anbefalte forfatterne en generell prosedyre som er illustrert i figur 4.



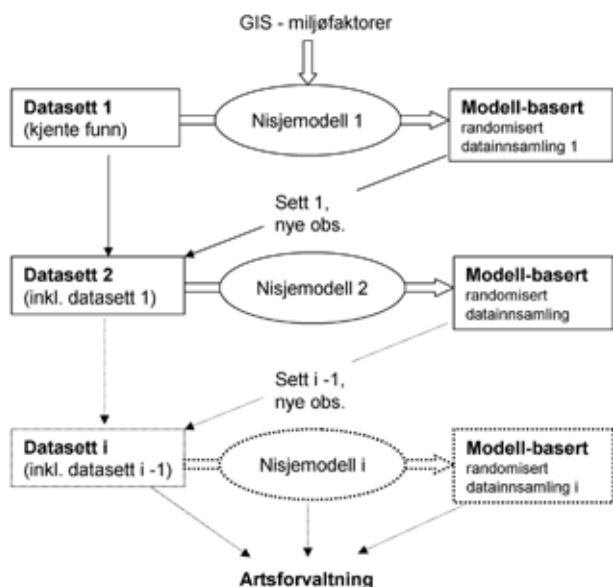
Figur 3. Predikert utbredelse av *Eryngium alpinum* i Sveits. A) Initiell predikert utbredelse – samme kart som i figur 1, og avgrensning av delområder som ble undersøkt i felt. B) Oppdatert predikert utbredelse etter feltvalidering, de skraverte delområdene ble funnet å være uegnet for arten (Kilde: Guisan et al. 2006).

Grunnlag for restaurering og re-introduksjon

En interessant anvendelse av prediksjonsmodellering er å finne egnede lokaliteter for å reintrodusere arter der de har forsvunnet. Et eksempel på slik anvendelse er studiet til Pearce & Lindenmayer (1998) hvor de analyserte hvorvidt aktuelle områder var bioklimatisk egnet for å re-introdusere en fugleart i Australia.

Grunnlag for reservatplanlegging

Prediksjonsmodeller kan brukes som grunnlag for reservatplanlegging. Ferrier (2002) viste at en kombinasjon av fjernmåling og prediksjonsmodellering kunne være en effektiv måte å lokalisere verneom-



Figur 4. Synergieffekt mellom oppdaterte datasett for artsforekomster, prediktive nisjemodeller og modellbaserte randomiserte datainnsamlinger i nye områder. Funnene fra de nye datainnsamlingene legges til tidligere data og gir grunnlag for en oppdatert versjon av den prediktive nisjemodellen. Etter flere slike oppdateringer sitter man med funndata og en nisjemodell som gir så god beskrivelse av artens utbredelse og økologi at den samlede kunnskapen kan benyttes som grunnlag for konkret artsforvaltning. Modifisert etter Guisan et al. (2006).

råder i regioner med sparsomt datagrunnlag for biologisk mangfold. Han understreket imidlertid at påliteligheten til en slik anvendelse bør testes ut i områder med bedre datagrunnlag. En vesentlig mer utviklet reservatplanlegging pågår i Colombia i Syd-Amerika. Her har man laget prediksjonsmodeller for 200 endemiske og truede fuglearter, og fått tydelig fram hvordan diversiteten til denne gruppen varierer innen landet (figur 5). Disse analysene vil brukes til å revidere landets reservatplaner i framtiden.

Araújo *et al.* (2004) viser hvordan prediksjonsmodeller kan benyttes til å vurdere effekten av klimaendringer på forekomst av arter innenfor etablerte eller planlagte reservater. De modellerte utbredelse av 1200 plantearter i Europeiske reservater 50 år fram i tid under to klimascenarier. De fant at 6-11 % av artene kunne gå tapt i de analyserte reservatene.

Et lignende resultat fant Hannah *et al.* (2007) som analyserte arters endrede utbredelsesmønstre under moderate klimaendringer og effekter av permanente reservatgrenser. De konkluderte at det er billigere og mer effektivt å utvide reservater relativt raskt sammenlignet med å utsette slike tiltak.

Grunnlag for studier av klimaeffekter på arters utbredelse

I de senere årene er prediksjonsmodeller i tiltagende omfang brukt til å evaluere potensielle effekter av klimaendringer. Det viser seg nesten uten unntak at ulike klimafaktorer (særlig ulike temperaturvariabler) er viktige prediktorer for arters utbredelse. Dermed kan man modellere artenes forekomst i forhold til dagens klima. Men istedenfor å bruke kart over klimavariasjon i nåtid, så benytter man framtidige klimaprognoser og framstiller prediksjonskart i forhold til disse. Et sentralt studium av denne typen er utført av Thuiller *et al.* (2005) som analyserte 1350 europeiske plantearters utbredelse i forhold til syv ulike klimascenarier. De fant at fjellplanter var særskilt sårbare i forhold til klimaendringer og at ulike deler av Europa hadde forskjellig sårbarhet med hensyn til artstap. For Norges del var den boreale sonen lite sårbar, mens de alpine områdene var mer sårbare, særlig i Nord-Norge. Under omtalen av reservatplanlegging (se over) viste vi også et par eksempler på slik bruk av prediksjonsmodeller. Analyser som fokuserer på klimaendringers effekt på alpine arter viser gjerne at disse får mindre utbredelse fordi de "skyves oppover" mot fjelltopper. Slike utfall bør imidlertid ikke være de eneste. I en modellering av kortnebbgås på Svalbard fant Jensen *et al.* (2008) at denne arten forventes å få utvidet sitt hekkeområde fordi arten velger områder med egnede fødesøk-arealer der maitemperaturen er over en viss grense. Omfanget av slike arealer vil øke på Svalbard i lys av klimaprognosene som ble benyttet.

Tilnærmingen som er beskrevet over innebærer at man predikerer arters forekomst i fremtiden. En lignende tilnærming er å predikere bakover i tid, såkalt "hindcasting". Her benytter man rekonstruert klima i fortiden som grunnlag for å predikere arters tidligere utbredelse. En slik metode ble benyttet av Bonaccorso *et al.* (2006) som lagde nisjemodeller for fuglearter i Amazonas og projiserte deres utbredelse tilbake til tiden for siste isbre-maksimum. Dette ble gjort for å vurdere om istidsrefugier kunne forklare endemismesentra for arter i dag.

Spredning av introduserte arter

Prediksjonsmodellering er ofte brukt til å vurdere introduserte arters spreningspotensial. I korte trekk modellerer man artens nisje ut fra deres kjente utbredelse der de naturlig hører hjemme. Deretter lager man et prediksjonskart over potensiell utbredelse i et annet område der arten er introdusert eller kan komme til å bli introdusert. Slike studier er gjort for en rekke arter og organismegrupper, for eksempel planter i Nord-Amerika (Peterson et al. 2003), en amerikansk froskeart i ulike deler av verden (Ficetola et al. 2007) og for plantearter i Syd-Afrika og andre deler av verden med lignende klima (Richardson and Thuiller 2007). Her er det et stort potensial for prediksjonsmodeller i forhold til å vurdere risikoen for spredning av arter som kan føre til både økologisk og økonomisk skade. Problemstillingen er særlig relevant for marine miljøer der arter i stor grad spres via ballastvann og kan utgjøre en stor trussel for oppdrett og fiskeri.

Identifisering av regionale biodiversitetsmønstre

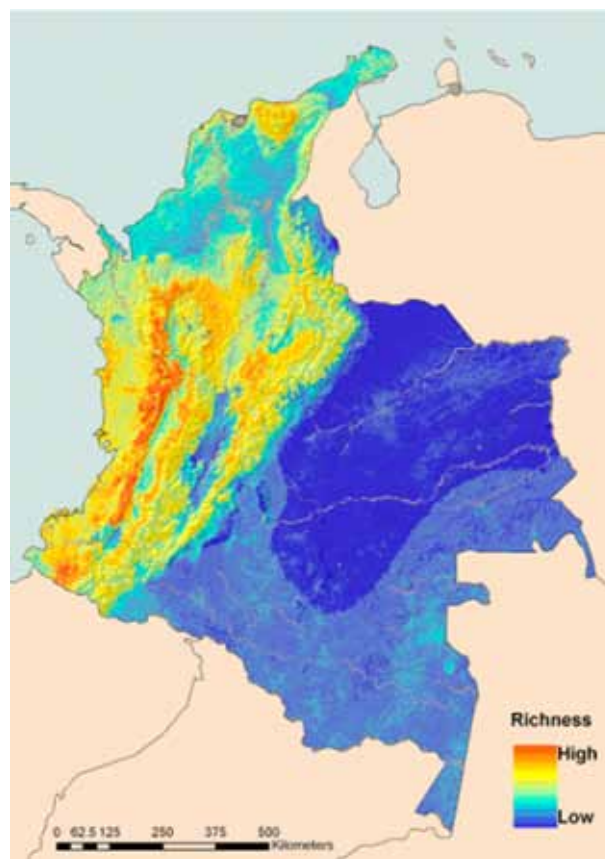
Prediksjonsmodeller er også anvendt til å modellere biodiversitetsmønstre (biodiversitetsindekser og -indikatorer) innen ulike regioner. Et eksempel på dette er forekomst av endemiske og sjeldne fuglearter som vi presenterte over (se figur 5). Andre eksempler er studier av alpin flora i Sveits (Guisan and Theurillat 2000), tempererte skogsamfunn i Sveits (Maggini et al. 2006) og floraen i Sør-Afrika (Thuiller et al. 2006).

Testing av biogeografiske, økologiske og evolusjonære hypoteser

Til slutt vil vi nevne at prediksjonsmodellering også kan anvendes til å teste grunnleggende biologiske hypoteser. Et eksempel på dette er et studium av Martínez-Meyer et al. (2004) som modellerte utbredelse av pattedyr i Nord-Amerika tilbake til siste istid for å teste hypoteser om nisjestabilitet.

2.3 Kvantitative metoder for modellering

Mange kvantitative metoder som egner seg for prediksjonsmodellering har blitt utviklet de siste årene, inkludert metoder for å lage kart (en type romlig pre-



Figur 5. Biodiversitetskart basert på maksimum entropi modellert utbredelse av 200 arter av endemiske og truede fugler i Colombia, Syd-Amerika. Foreløpige resultater fra pågående arbeid av Jorge Velasquez-Tibata, Catherine Graham, Paul Salaman og BIOMAP team. Kilde: http://www.esi-topics.com/nhp/2007/may-07-Elith_Graham.html

diksjon) på grunnlag av prediksjonsmodellene. Det er ikke noe enkel måte å klassifisere disse metodene på, og det er ikke hensikten her å beskrive alle (se Hastie et al. 2001 for en grundig oversikt over de statistiske aspektene). Vi har derfor valgt å fokusere på noen anerkjente metoder for modellering innen tre hovedkategorier: regelbaserte, numeriske og modellbaserte statistiske metoder. Vi går fra metoder som er basert kun på ekspertkunnskap om effekter av prediktorvariabler (regelbaserte metoder) til statistiske metoder som fokuserer kun på prediksjonsmål, og omtaler til slutt statistiske metoder som er basert på en ekspli-

sitt modell om effekten av prediktorvariabler. Et hovedprinsipp bak statistiske metoder er at man prøver å finne et kompromiss mellom en komplisert modell som beskriver et datasett godt men som oftest gir upresise prediksjoner (fordi modellen fokuserer sterkere på å beskrive det spesifikke datasettet enn på å finne enkle sammenhenger med generell gyldighet), og en enkel modell som gir høyere presisjon i prediksjonene men som ofte har forventningsskjevhet i prediksjoner og parametere. Kompliserte modeller er også vanskeligere å tolke/forstå, men naturen er selvsagt ikke nødvendigvis enkel. Hva som er best er ofte avhengig av hva man ser etter. Man kan bruke et kart som sammenligning: for mange detaljer kan hindre at man ser hovedtrekkene, men detaljerte kart kan være nyttige når man er interessert i å se på de fine terrengdetaljene.

2.3.1 Regelbaserte metoder

Regelbaserte metoder (konvoluttmetoder) er en klasse av metoder der man setter opp regler, basert på økologiske tilstander, for hvor arter skal forventes å opptre. For hver økologisk faktor definerer man grenseverdier for en gitt art, og modellen identifiserer de stedene som finnes innenfor det økologiske rommet man på denne måten antar at arten utnytter. Grenseverdier kan være basert på feltobservasjoner langs gradienter i fysiske miljøfaktorer, basert på dokumenterte sammenhenger fra andre steder eller basert på lokal ekspertkunnskap. Kartet for predikert utbredelse er dermed ikke basert på statistisk analyse av sammenhengen mellom forekomst og forklaringsvariabler.

BIOCLIM (Busby 1991) er en slik regelbasert metode som særlig vektlegger effekten av klimafaktorer på arters utbredelse. Denne metoden anvendes fremdeles, for eksempel til å definere utbredelse av arter i lys av framtidige klimaprognoser. Denne metoden er implementert i den gratis programvaren Diva-GIS, se <http://www.diva-gis.org/>.

Utbredelsen til marine arter og naturtyper er i stor grad styrt av geofysiske faktorer. Den enkleste metoden for romlig kartlegging er å finne grenseverdier for arters og naturtypers utbredelse langs de ulike fysiske gradientene, for så å modellere denne utbredelsen romlig i GIS. Grenseverdier er basert på regler (feltobservasjoner eller teoretisk kunnskap) eller ekspertkunnskap. Resultatet blir et romlig binært

(tilstede/ikke tilstede) kart som gir en oversikt over potensielle lokaliteter for de ulike artene og naturtypene. Denne metoden er brukt i arbeidet med kartlegging av marine naturtyper både i Norge (Bekkby et al. 2002, Bekkby and Isæus *Accepted*, Rinde et al. 2006) og Sverige (Bekkby and Rosenberg 2006). Metodikken har gitt et svært nyttig utgangspunkt for planleggingen av kartleggingen av marine naturtyper, der en i utgangspunktet hadde liten kunnskap om utbredelsen til de fleste av naturtypene og i liten grad hadde tilgang til stedfestet informasjon om tilstedeværelse eller fravær av naturtypene. Den modellerte potensielle utbredelsen til utvalgte naturtyper som tareskog, ålegrasenger og bløtbunnsområder er benyttet som grunnlag for å prioritere områder som skal kartlegges for å få oversikt over nasjonalt og regionalt viktige forekomster av ulike naturtyper.

2.3.2 Numeriske metoder

Tremodeller: regresjonstrær og klassifikasjonstrær

Tremodeller representerer en klasse av modeller der en responsvariabel predikeres på grunnlag av verdiene til et sett av prediktorvariabler som kan være en blanding av kontinuerlige og kategoriske variabler. Framgangsmåten er at man iterativt splitter datasettet i to undergrupper (et dikotomt tre) basert på verdiintervaller eller spesifikke kategorier av prediktorene. Dersom responsvariabelen er en kontinuerlig variabel (for eksempel antall individer) får vi et *regresjonstre* som predikerer forventet verdi til responsvariabelen. Dersom responsvariabelen er kategorisk eller binær får vi et *klassifikasjonstre* som predikerer klassetilhørighet for responsvariabelen. Resultatet fra en tremodell vil typisk se ut som i figur 6.

Hovedprinsippet for tremodeller er at metoden evaluerer hele settet av prediktorvariabler og finner den som forklarer størst andel av variasjonen ("devians") i responsvariabelen. Derne velges en splitt-verdi (eller en gruppering av prediktorklassene) slik at de to undergruppene minimerer variasjonen av kvantitative responser eller best klassifiserer kategoriske responser. Denne prosessen gjentas for nye undergrupper inntil modellens forklaringssevne ikke lenger øker eller at antall gjenværende observasjoner blir lavere enn en forhåndsdefinert grenseverdi (for eksempel 5 observasjoner). Det siste gjøres for å unngå at man lager et overtilpasset tre med en observasjon i hver terminal node (grenspiss av treet).

Et sentralt tema i forbindelse med tremodeller er å finne et optimalt antall noder (splittpunkter). Dette gjøres vanligvis ved kryssvalidering der man deler datasettet i 5–10 deler og holder en del ute mens modellen bygges på grunnlag av de andre observasjonene. Prediksjonsevnen til ulike størrelser av treet testes ved å beregne feilklassifiseringsraten for de utelatte datapunktene. I denne prosessen vil trærnes feilklassifisering minke inntil et visst antall splitt-punkter, og deretter øke som en effekt av overtilpasning. Til slutt framstiller man det beste treet grafisk og viser prediksjonsevnen ved hjelp av evalueringsdataene.

Klassifikasjonstrær og regresjonstrær er enkle å ha med å gjøre fordi de ikke stiller noen krav til statistisk fordeling verken for responsvariabel eller prediktorvariabler. Videre er resultatet av slike modeller intuitivt lett å forstå og modellene kan enkelt framstille komplekse interaksjoner av flere miljøfaktorer.

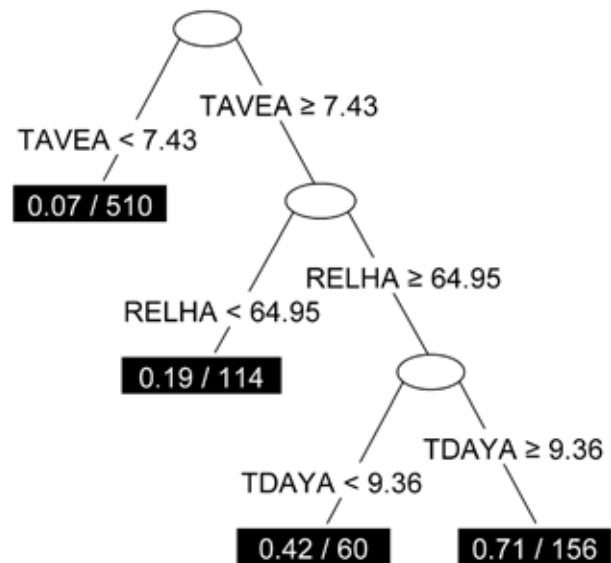
Boosted regression trees (BRT)

Ved å kombinere flere klassifikasjonstrær kan man forbedre prediksjonsevnen signifikant i forhold til bruk av et enkelt klassifikasjonstre. En etablert metode for å utvikle og velge slike kombinasjoner av klassifikasjonstrær er såkalt boosting. Forenklet kan man si at metoden består i å gjentagsvis å beregne et klassifikasjonstre, re-vekta observasjonene i datasettet slik at økt vekt tillegges de observasjonene som er feil klassifisert, beregne et nytt klassifikasjonstre på grunnlag av de på re-vektede dataene osv.. Sluttresultatet er en lineær kombinasjon av klassifikasjonstrærne fra hvert trinn i prosessen.

Denne metoden har to distinkte fortrinn framfor bruken av et enkelt tre: a) feilklassiferingen av dataene går signifikant ned, og b) metoden synes å være relativt immun mot statistisk overtilpasning. Ulempen er at metoden er svært regneintensiv og krever vesentlig større kapasitet enn det normale PC-er har.

Metoden har vært lovpriset som den beste tilgjengelige klassifikasjonsmetoden i det såkalte "maskinlæringsmiljøet" men framsto lenge som en "black box" metode. Et gjennombrudd, fra et statistisk ståsted, kom ved (Friedman et al. 2000) som viste at den rådende boosting-metoden essensielt var en optimeringsmetode for å finne en klassifikator som minimerte en bestemt eksponensiell feilklassifikasjonsfunksjon ("loss function"). Dermed knyttet de en metode som til da hadde framstått som relativt

Lobaria pulmonaria



Figur 6. Trebasert prediksjonsmodell for tilstedeværelse av lungenever (*Lobaria pulmonaria*) i Pacific Northwest, USA. Hver node (todeling) angir verdier for en prediktor i de to sidegrenene. Prediktorene i dette eksempelet er: TAVEA = temperatur (årgjennomsnitt), RELHA = % relativ fuktighet (årgjennomsnitt), TDAYA = dagtid-temperatur (årgjennomsnitt). Verdiene i de terminale nodene (de svarte boksene) angir sannsynlighet for forekomst og antall pixler som klassifiseres til denne kombinasjonen av miljøfaktorer (Kilde: Edwards jr et al. 2005).

obskur til likelihood-estimering innen standard statistikk (Ridgeway 2000).

Litteratur og programvare

Tremodeller er beskrevet i en rekke bøker (for eksempel Crawley 2002, for eksempel Hastie et al. 2001). Videre finnes det svært gode beskrivelser av tremodeller anvendt i økologiske studier, for eksempel De'ath & Fabricius (2000) – marine miljøer, Vayssières et al. (2000) – terrestre planter, og Edwards jr. et al. (2005, 2006) – lav.

Det finnes en rekke varianter av tremodeller i statistikk-pakken R i bibliotekene randomForest, RPART og TREE. Metoder for "boosted regression trees" er å finne i R-biblioteket GBM. CART er en annen statistikk-pakke som er spesialutviklet for tre-modellering.

Maskinlæringsmetoder

"Machine learning", eller "maskinlæring" på norsk, er en metode som klassifiseres under fagfeltet kunstig intelligens ("artificial intelligence" – AI). Begrepet maskinlæring brukes om teknikker som gjør maskinen (dvs programvaren) i stand til å "lære", altså på egen hånd å kunne forbedre sin evne til problemløsning. Hovedarbeidet ligger i å utvikle algoritmer som kan testes og byttes ut med andre i leting etter den mest optimale løsningen. Dette fordrer da at programvaren er i stand til å evaluere resultatene (handlingene) og at den kan belønne gode handlinger og straffe uheldige handlinger. De viktigste anvendelsene av maskinlæring er innen beslutningsstøttesystemer, der man skal lage kortfattede oversikter fra store mengder data (også kjent som "data mining", på norsk "datagraving"), stemmegjenkjenning, automatisk føring av kjøretøy, samt andre områder der anledningen til forprogrammering av maskinen er begrenset.

I prediktiv modellering er det "data mining"-metodene som brukes. De mest kjente maskinlæringsmetodene for prediktiv modellering er DesktopGarp og Maxent. De to metodene er vel tilpasset for bruk i prediksjon basert på forekomstdata fra herbarier og museumssamlinger.

DesktopGarp

Denne metoden setter opp et sett av logiske regler mellom miljøfaktorene og hvordan disse reglene svarer med forekomst av arten. Regelsettet utsettes for tilfeldige endringer og utvikler seg ved hjelp av en genetisk algoritme som velger ut nye regelsett som bedre predikerer artsforekomster. Artsfunnene splittes opp i ett treningssett og to valideringssett. Det ene valideringssettet brukes internt for å la regelsettet utvikle seg (fungerer som utvalgs-kriterium i den genetiske prosessen). Det andre valideringssettet benyttes for kryssvalidering og til å evaluere prediksjonsevnen til resultatet.

I praksis kjøres Garp-analyser flere ganger, og hver gang blir resultatkartet forskjellig fra forrige kjøring

(på grunn av de tilfeldige endringene av regelsettene). Basert på kryssvalideringene kan man f.eks. velge ut de 10 % beste prediksjonskartene og benytte disse som grunnlag for å presentere det endelige prediksjonskartet.

Denne metoden bruker bare tilstede observasjoner som grunnlagsdata, sammen med kart over miljøfaktorer (variasjoner i klima, terreng, geologi osv.) i GIS-format – fortrinnsvis fra standard ESRI produkter. Garp er en typisk "black box" metode der man får et prediksjonskart (egentlig flere alternative kart) som resultat. Resultatkartene forteller bare hvor det er en høy sannsynlighet for å finne arten, men skiller ikke mellom ulike grader av middels-lav sannsynlighet for forekomst. Metoden forteller heller ikke hvilke prediktorer som er viktigst for å forklare artens forekomst eller den relative betydningen av disse.

Software: <http://nhm.ku.edu/desktopgarp/>

Maxent

Maxent er en såkalt "maksimum entropi"-metode og er godt dokumentert i Phillips et al. (2006). Hovedideen i metoden er å estimere en ukjent sannsynlighetsfordeling (dvs artens forekomst) i forhold til et sett av restriksjoner. Restriksjonene som settes i Maxent er at forventet verdi (det sanne gjennomsnittet) i forhold til hver miljøfaktor skal være den samme som det observerte gjennomsnittet i det empiriske observasjonsmaterialet. For øvrig finner metoden en sannsynlighetsfordeling som er maksimalt spredd i forhold til de observerte verdiene (har maksimal entropi) og således ligger nær opp til uniform fordeling. Maskinlæringsaspektet ligger i at metoden prøver ut forskjellige interaksjoner mellom miljøfaktorene. Det er vist matematisk at det finnes en entydig løsning på problemet formulert over, og at den sannsynlighetsfunksjonen man søker å beregne er identisk med en Gibbs-fordeling som maksimerer likelihood (dvs sannsynligheten) for de observasjonene man har i det empiriske materialet.

Metodikken i Maxent har en rekke fordeler. Den er matematisk presis og transparent, og er gjenstand for aktiv videreutvikling innen fagfeltene maskinlæring og statistikk. Av mer praktisk betydning er at metoden bare trenger tilstedeobservasjoner, den produserer et entydig prediksjonskart (uttrykt som kvantitative sannsynligheter), den rangerer den relative betydningen av hver miljøfaktor ved å kvantifisere deres relative innflytelse på prediksjonsevnen

(AUC-verdien), samt viser hvordan spesifikke verdier av miljøfaktorene influerer på sannsynlighetsfordelingen. Sammenlignet med andre metoder kommer Maxent svært godt ut mht prediksjonsevne (dvs evnen til å gi en riktig prediksjon av tilstedeværelse og fravær av arten). Siden Maxent finner en sannsynlighetsfordeling som er nært tilpasset dataene, er det en fare for at metoden gjør en overtilpasning til de spesifikke observasjonene i datasettet. Dette kan motvirkes ved såkalt regularisering der man i stedet for å si at de sanne gjennomsnittene skal være lik de empiriske gjennomsnittene for prediktorene, så skal forskjellen mellom disse størrelsene være mindre enn en viss verdi. Dette gjør at ubetydelige miljøfaktorer blir neglisjert og faren for overtilpasning reduseres betydelig.

Software: Maxent <http://www.cs.princeton.edu/~schapiro/maxent/>

2.3.3 Modellbaserte statistiske metoder

Generaliserte lineære modeller (GLM)

GLM ble først beskrevet tidlig på 1970-tallet som en type modellering som generaliserte mange anerkjente modelleringsmetoder: lineære modeller (for eksempel lineær regresjon eller ANOVA), log-lineære modeller og logistisk regresjon (se McCullagh and Nelder 1989). Disse metodene hadde vært i bruk i økologi før 1970-tallet, men GLM-konseptet og programvaren (GLIM) fikk en rask utvikling etter 1975. I dag er GLM hyppig brukt som en standard metode i prediksjonsmodellering fordi den benytter typiske forekomst data (tilstedeværelse/fravær eller telldata) og fordi modelleringsresultatet forholdsvis enkelt kan presenteres som et prediksjonskart ved hjelp av GIS. Hovedmotivasjonen for å utvikle metoden var fleksibilitet med hensyn til at a) forutsetningen om normalfordelt frekvensfordeling med konstant variasjon kan erstatte med andre fordelinger (for eksempel binomial for tilstedeværelse/fravær data og Poisson for telldata, som har en annen sammenheng mellom gjennomsnitt og variasjon), og b) prediktorvariablene kan ha en ikke-lineær sammenheng med forventet verdi av responsvariabelen.

GLM er beskrevet i en rekke standardbøker om statistisk analyse: McCullagh & Nelder (1989) er fortsatt en grundig (men relativt teoretisk) referanse, Venables & Ripley (2002) og Faraway (2006) viser mange eksempler med bruk av statistikkpakken R.

Hovedprinsippet i GLM er at den statistiske modellen er reformulert slik at man kan skille mellom fordelingen til responsvariabelen (stokastisk del av modellen) og sammenhengen mellom forventet verdi og responsvariabelen (systematisk del av modellen). Den systematiske delen er:

$$G(E[Y]) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p$$

G er en såkalt linkfunksjon (som må være monoton, slik at det til hver unik verdi av $E[Y]$ svarer en unik verdi av $G(E[Y])$). Det er flere viktige grunner til å bruke GLM. Særlig viktig er det at transformasjonen av responsvariabelen ved hjelp av linkfunksjonen kan gjøre det meningsfullt å modellere responsen som en lineær funksjon av prediktorene. Det er tilfellet når forventet verdi av responsvariabelen er begrenset til verdiene 0 og 1 eller til tall mellom 0 og 1 (sannsynligheter). Typiske data av denne typen er tilstede-fravær data hvor man modellerer en sannsynlighet for at arter er til stede. For slike data brukes *logit* link: $\text{logit}(p) = \log(p/(1-p))$. Uansett verdi for den logit-transformerte responsvariabelen vil ikke predikert p ligge utenfor intervallet $[0,1]$. For telldata (f.eks. artsantall), som bare kan anta heltallige og positive verdier bruker man oftest *log* (logaritme) link funksjonen. Ofte er residualvariansen i modellen (variasjon som er ikke forklart av modellen) større enn forventet ved en binomial fordeling (tilstede/fravær data) eller Poisson fordeling (telldata). Dette fenomenet, som kalles overdispersjon (det motsatte, dvs underspersjon, er uvanlig), kan ha to hovedgrunner: positiv romlig autokorrelasjon og/eller at viktige variable ikke er med i modellen. For telldata kan det også forekomme når variasjonen i responsvariabelen øker mer med økende middelvei enn det Poisson fordelingen forutsetter (at variansen øker proporsjonalt med middelvei).

Etter at den modellen som best forklarer sammenhengen mellom den transformerte verdien av Y og prediktorene (se 2.4. om modellseleksjon) er funnet, kan man tilbaketransformere predikert respons og lage kart over sannsynlighet for forekomst av Y .

Dataegenskaper: GLM anvendt på tilstede-fravær observasjoner forutsetter at datagrunnlaget også inneholder fraværsobservasjoner. Det beste er selvsagt å ha ekte fraværsobservasjoner (påvist fravær), men det er vanlig å generere pseudo-fraværsobservasjoner som en løsning. Pseudo-fraværsobservasjoner er gjerne tilfeldig valgte punkter innen det området man har data fra. Disse tilfeldig valgte punktene re-

presenterer dermed fravær av observasjoner og man antar at dette også representerer fravær av arten. I virkeligheten kan slike pseudo-fraværsobservasjoner sammenfalle med reelle forekomster.

Programvareplattform: GLM er en standard funksjon ("workhorse") i R (<http://www.r-project.org/>).

Generaliserte additive modeller (GAM)

Link funksjoner og polynomer i en GLM kan i praksis brukes til å tilpasse en ikke-lineær sammenheng (for eksempel kan et annenordens polynom i en logistisk regresjon beskrive fordelingen av en art med et optimum langs en økologisk gradient), men de mangler fleksibilitet. Som en videre utvikling av GLM ble GAM (Generalised Additive Models) etablert som statistisk metode rundt 1990 (Hastie and Tibshirani 1990). Denne metoden ble tatt aktivt i bruk innen økologisk forskning like etterpå (Yee and Mitchell 1991). I dag er metoden ofte brukt i prediksjonsmodellering, gjerne parallelt med GLM fordi den generaliserer GLM.

Den grundigste beskrivelse i dag er gitt av Wood (2006). Hovedprinsippet i GAM er at man ikke setter opp en spesifikk funksjonssammenheng mellom responsen og prediktorene og estimerer koeffisientene til hver prediktor. Isteden estimeres funksjonssammenhenger direkte fra data gjennom lokal kurvetilpasning til hver prediktor. Det vil si at modellstrukturen ser slik ut:

$$G(E[Y]) = f_1(X_1) + f_2(X_2) + \dots + f_p(X_p).$$

I analysen lister man opp de aktuelle prediktorene og hva slags type kurvetilpasning (f_i) en ønsker, og metoden finner en spesifikk kurvetilpasning mellom responsen og hver av prediktorene. En fare ved denne metoden er at man overtilpasser modellen til de spesifikke dataene. Dette kan reguleres ved å nedjustere hvor tett dataene skal tilpasses slik at man reduserer faren for å tilpasse støy i dataene.

Når man har funnet den modellen som best forklarer sammenhengen mellom Y og kurvetilpasningene til prediktorene har man en oversiktlig situasjon. Responsen er summen av de transformerte prediktor verdier (det er faktisk mulig nå å ha en funksjon av flere variable samtidig, men det er datakrevende; Wood 2006). Det kreves et ekstra trinn for å oversette denne modellen til et prediksjonskart basert på kart over miljøfaktors verdier. Dette løses ved hjelp av applikasjonen GRASP (se 2.5.1).

Programvareplattform: mgcv er biblioteket i R som tilsvarer Wood (Wood 2006).

Partial least squares regression (PLS)

I økologiske studier måles gjerne mange miljøvariabler som potensielle prediktorer, av og til flere variabler enn observasjoner. I tillegg er miljøvariablene ofte korrelerte. Dette gjør at man ikke kan tilpasse vanlige regresjonsmodeller (det er ingen frihetsgrader igjen hvis man har flere variabler enn observasjoner). Dette er en vanlig situasjon i økonometri og kjemometri (hvor man ofte har målt konsentrasjon av veldig mange komponenter) hvor metoden PLS ble utviklet (Wold et al. 1984, se Wold et al. 2001 for en grundig beskrivelse).

PLS ("Partial Least Squares regression" eller "Projection on Latent Structures") er en regresjonsteknikk som reduserer problemene med mange korrelerte prediktorvariabler. Hovedideen med metoden er, som i multivariat analyse, å finne hovedstrukturen i det mangedimensjonale rommet som spennes ut av de originale prediktorvariablene ved å finne et underrom av ortogonale hovedakser som samtidig gir en robust prediksjon av responsvariabelen. Disse prediktoraksene omtales som "latente variabler" (svarer til ordinasjonsakser i ordinasjonsanalyse). Ortogonale prediktorer er per definisjon uavhengige av hverandre. Til forskjell fra modeller med multivariate responser (f.eks. PCA) er det i PLS bare en responsvariabel, og aksene finnes på en slik måte at de forklarer mest mulig av variasjonen i responsen.

PLS kan utføres i R. Pakka 'pls' inneholder basisfunksjoner for modelltilpasning og validering ('pls', 'mvr'), diagnoseplot ('scoreplot', 'loadingplot', 'correlationplot'), mm. Regresjonsmodellene er lineære og uten link-funksjoner (tilsvarende 'lm'), så eventuelle ikke-lineariteter og ikke-kontinuerlige responser (f.eks. tilstede-fravær) må håndteres med transformasjoner. Modellseleksjon er foreløpig ikke godt tilrettelagt i R, og dette gjør at modelleringsarbeidet fort blir tidkrevende. Det finnes imidlertid kommersielt tilgjengelige programmer med bedre muligheter for rask, interaktiv modellseleksjon (f.eks. Unscrambler).

Bayesianske metoder

Bayesianske metoder har de siste årene blitt sterkt videreutviklet, hovedsakelig fordi det har blitt mulig å bruke intensive numeriske prosedyrer ("MCMC" =

Markov Chain Monte Carlo) for å estimere fordelinger av parametrene (istedenfor analytiske formler som er ofte vanskelig å finne). Forskjellen mellom Bayesianiske metoder og de mer "vanlige" statistiske metodene (GLM med "maximum likelihood" estimering = MLE) er ikke selve modellen (f. eks. binomial fordeling og logit link mellom respons og prediktor variable) men at man antar at parametrene i modellen har en "prior" fordeling, og ikke at parametrene har en gitt men ukjent verdi. Bayesianiske metoder er basert på følgende sammenheng mellom prior fordeling $p(\theta)$ og posterior fordeling $p(\theta | \text{data})$ av parametrene (det såkalte Bayes teorem):

$$p(\theta | \text{data}) = p(\theta) p(\text{data} | \theta) / p(\text{data}),$$

hvor $p(\text{data} | \theta)$ er den sannsynlighetsfunksjon som er brukt i for eksempel GLM eller i AIC (se 2.5).

Oftest kan man bruke en såkalt uinformativ prior for θ , for eksempel en fordeling som er mer eller mindre uniform over et stort intervall. Med slike priorer er svaret for enkle modeller (dvs posterior fordeling av parametrene) i praksis det samme som for de mer vanlige metoder. En fordel med Bayesianiske metoder som er relevant her er at det er lettere å tilpasse modeller med forskjellige nivåer av variasjon eller modeller med direkte og indirekte effekter (se Myrsterud et al. 2008 for et eksempel) og å inkludere målingsfeil (se Clark and Gelfand 2006). Nye eksempler inkluderer modellering av romlig autokorrelasjon (Gelfand et al. 2006).

Software: Bayesianiske modeller kan estimeres med 1) egen gratis pakke WinBUGS som har en interface i R (R2winbugs), 2) "arm" R-bibliotek som inkluderer bl.a. Bayesianiske GLM (se Gelman et al 2007) og som er spesielt tilpasset til modeller med flere nivåer, og 3) R-bibliotek BRugs som er basert på open source OpenBUGS og som har mye av de samme mulighetene som WinBUGS.

Modellere romlig autokorrelasjon

De fleste statistiske metoder antar at observasjonene har en tilfeldig variasjon ("feil") som er uavhengige av hverandre. Det betyr for eksempel at i en logistisk regresjonsmodell, skal sannsynligheten for å observere en art i et observasjonspunkt være uavhengig av om arten er tilstede i nabo-observasjonspunktet (sannsynlighetene kan være forskjellige dersom miljøvariablene har forskjellig verdi, men her ser vi på det

som kan observeres gitt disse miljøvariabelverdiene). Antakelsen om uavhengighet er oftest urealistisk for økologiske data (Økland 2007). Dette skyldes for eksempel at spredning kan resultere i bestander som er spredt over relativt stort areal (stort i forhold til observasjonshetenes størrelse). Det samme problemet oppstår ved modellering av dyrearters habitatbruk (fordi dyr beveger seg). Slike prosesser som mobilitet resulterer i en positiv romlig autokorrelasjon, det vil si at naboobservasjoner tenderer til å ha mer like verdier for responsvariabelen enn forventet (dersom alle observasjonene hadde vært uavhengige av hverandre). Det er viktig å merke seg at man kan observere en positiv autokorrelasjon i responsvariabelen også om man har tatt hensyn til prediktorvariabler som er romlig korrelert (Hobert et al. 1997). Det er vanskelig å skille mellom ukjente miljøeffekter og prosesser som er resultatet av biologiske prosesser som for eksempel spredning. En positiv romlig autokorrelasjon som ikke er tatt hensyn til resulterer i såkalt Type I-feil ved statistisk analyse, det vil si at nullhypoteser forkastes oftere enn datamaterialet egentlig gir grunnlag for, konfidensintervaller som er for små eller modeller som er unødvendig kompliserte (Dormann et al. 2007).

Software: bibliotek lme/lme4, GeoR i R.

2.3.4 Oppsummering

Som det framgår av dette kapittelet finnes det et bredt spekter av modelleringsmetoder. I tabell 1 har vi oppsummert noen egenskaper ved de ulike metodene. De aller siste årene har det også kommet flere studier som sammenligner prediksjonsevnen til ulike metoder (Elith et al. 2006, Guisan et al. 2007b) på identiske datasett. Elith et al. (2006) er den klart mest grundige i så henseende.

2.4 Modellseleksjon og prediksjonsevne

Valg av modell(er) er avhengig av to kriterier: 1) at modellen passer i forhold til dataene (dvs at antagelsene som definerer modellen er konsistente med dataene), og 2) at den modellen eller de modellene man bruker til prediksjon er bedre enn andre modeller som passer dataene. Det første kriteriet kan vurderes på grunnlag av en rekke metoder som går under fellesbetegnelsen

diagnostikk. Diagnostikk kan brukes til å vurdere uavhengighet, om sammenhenger er lineære og fordelinger (spesielt varians-gjennomsnitt sammenheng). Mange bøker beskriver disse metodene (Faraway 2006, for eksempel Venables and Ripley 2002), og de er implementerte i R.

2.4.1 Modellseleksjon

Gitt at vi har en rekke modeller som passer dataene, kan man gå videre med modellseleksjon. Dette er en prosedyre som anvendes systematisk i dataanalysen når man skal finne fram til den beste prediksjonsmodellen basert på en gitt metode. Det er nå blitt klart at bruk av statistiske tester, hvor man sammenligner en enkel modell (nullhypotese) med en mer komplisert modell (alternativ hypotese), er lite effektive og i mange tilfeller ikke vil gi de beste prediksjonsmodellene. En bruker nå såkalte statistiske kriterier, som man estimerer for hver modell, som et relativt mål på hvor god modellen er. Et viktig prinsipp er "the principle of parsimony", dvs. balansen mellom god tilpasning av modellen til dataene og antall forklaringsvariabler i modellen (en modell med flere variable kommer alltid til å tilpasse dataene bedre, men estimatene og derfor prediksjonene kommer fort til å være upresise om det er mange variabler). Den beste modellen er altså den som får mest støtte av dataene men som samtidig bruker få prediktorvariabler.

AIC (Akaike's Information Criterion) er nok det vanligst brukte kriteriet for modellseleksjon. Dette baserer seg på "Maximum Likelihood Estimering"

og Kullback-Leibler avstand mellom to fordelinger (se Burnham and Anderson 2002 for en historisk oversikt). Akaike (1973) fant sammenhengen mellom denne avstanden og Fisher's maksimerte "log-likelihood"-funksjon (De Leeuw 1992) og etablerte AIC som et modellseleksjonskriterium (se Burnham and Anderson 2002). AIC og andre lignende kriterier brukes til å velge mellom modeller og gir en indikasjon på hvor gode modellene er i forhold til hverandre.

2.4.2 Prediksjonsevne

Kriterier for modellseleksjon benyttes gjerne for å sammenligne alternative modeller utviklet på samme metodeplattform. Ofte er vi imidlertid interessert i å sammenligne modeller på tvers av metodeplattformer, og da benytter man gjerne kriterier som måler prediksjonsevne. Den vanligste og som ofte er brukt for lineære modeller er R^2 . En definisjon av R^2 er korrelasjon(predikerte verdier, observasjoner)², som viser klart hvorfor den kan brukes som mål på prediksjonsevne. De kan faktisk brukes også for andre modeller (glm, gam), se Zheng & Agresti (2000). Det er to forhold en må være oppmerksomme på ved bruk av R^2 som kriterium for prediksjonsevne: 1) at verdien av R^2 man får fra utvalget overestimerer verdien man får fra et nytt utvalg (dvs. det er et skjevt estimat), og det er best å bruke estimerer som adjusted- R^2 (se Mittlböck and Heinzl 2001 for adjusted R^2 and glm) eller å bruke kryssvalidering, og 2) lave R^2 verdier betyr ikke nødvendigvis at modellen er dårlig dersom det er tilstede-fravær forekomster man predikerer. Dette er

Tabell 1. Oppsummering av noen egenskaper ved de ulike metodene

Metode	Type responsdata	Evaluerer relativ betydning av hver prediktorvariabel	Prediksjonsevne (Elith et al. 2006)
BIOCLIM	tilstede	nei	moderat
Regresjonstrær	tilstede-fravær	nei	ikke vurdert
Boosted Regression Trees	tilstede-fravær	nei	svært god
DesktopGarp	tilstede	nei	god
Maxent	tilstede	ja	svært god
GLM	tilstede-fravær	ja	god
GAM	tilstede-fravær	ja	god
PLS	tilstede-fravær	ikke direkte	ikke vurdert

Tekstboks: Kriterier for modellseleksjon

Definisjon av AIC er:

$$AIC = -2 \log(\text{likelihood}) + 2 n_p,$$

hvor likelihood er maksimert verdi for sannsynlighetsfunksjonen (sannsynligheten for å generere dataene, gitt at modellen er riktig), og n_p er antall parametere i modellen (for eksempel 2 for en enkel regresjonsmodell: $E[y] = \beta_0 + \beta_1 X$; evt. 3 om residualvariansen også er inkludert som parameter). I en glm er $-2 \log(\text{likelihood})$ kalt deviance. En versjon for små datasett (AICc) har blitt utviklet (Burnham and Anderson 2004, Hurvich and Tsai 1989). Den beste modellen er den med laveste verdi for AIC, og ved sammenlikning av modeller er det kun forskjellen i AIC som er viktig (dvs de absolutte AIC-verdiene betyr ingenting alene, kun i forhold til hverandre).

Det er to viktige poenger som man bør huske: 1) det finnes flere kriterier enn bare AIC, og 2) det er ofte en god idé å bruke flere modeller enn kun den modellen som har høyest verdi for AIC. AIC har en rekke sammenhenger med andre kriterier som var i bruk før for lineære modeller, slik som adjusted R^2 og Mallows C_p (Mallows 1973). De senere årene ble det utviklet andre kriterier som er basert på det samme prinsippet som AIC, men som varierer bl.a. i hvordan kompleksitet av modellen er tatt hensyn til. BIC (Bayesian Information Criterion) er for eksempel:

$$BIC = -2 \log(\text{likelihood}) + \log(n) n_p,$$

hvor n er utvalgstørrelsen. BIC gir enklere modeller enn AIC (siden $\log(n)$ oftest er større enn 2). Link & Barker (2006) har nylig diskutert forskjellen mellom AIC og BIC, og dette er fortsatt et aktivt forskningsfelt i statistikk (van der Linde 2005).

Bruk av flere modeller for å få mer robuste prediksjoner er blitt vanlig i klima- og værmeldingsammenheng (Gneiting and Raftery 2005, Räisanen and Ruokolainen 2006), og er foreslått også for prediksjonsmodellering (Araújo and New 2007). Prinsippet er at man regner ut gjennomsnittet av prediksjoner fra flere modeller, men at prediksjonene vektlegges mht hvor god modellen er. AIC og de deriverte AIC vektorer kan brukes (Burnham and Anderson 2002). I klimaforskning, hvor man bruker GCM (Global Circulation Model), er det mye diskusjon om hvordan man kan måle kvalitet av en modell (se Räisanen and Ruokolainen 2006). Siden forskjellige kriterier gir varierende svar, har man oftest valgt å gi de samme vektene til alle modeller.

fordi slike observasjonsdata kun kan få to verdier (0 eller 1) mens modellen gir predikerte verdier mellom 0 og 1 (Cox and Wermuth 1992).

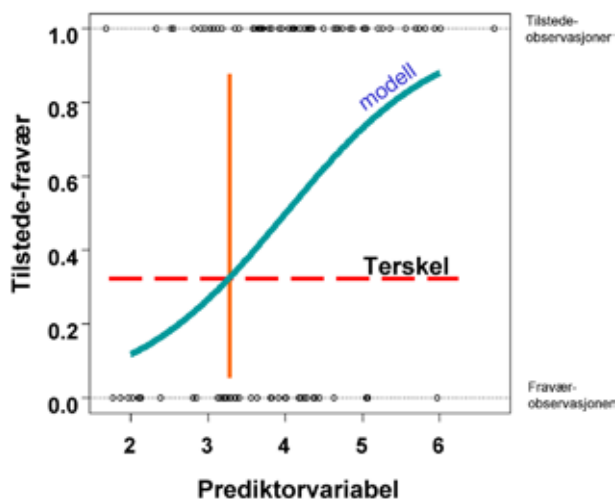
Siden R^2 kan være vanskelig å tolke med 0/1 data, er det utviklet flere kriterier for å sammenligne prediksjonsevne (se Jolliffe and Stephenson 2004). Flere er basert på gruppering av prediksjonene i to klasser (tilstede og fravær) og sammenlikning av disse prediksjonene med de faktiske observasjonene (se figur 7).

De fleste metoder predikerer artenes forekomst som en gradert (kontinuerlig) sannsynlighet, og ikke som enten tilstede eller fravær. De predikerte sannsynlighetene kan klassifiseres i to grupper (tilstede og fravær) ved hjelp av en terskel. Man klassifiserer da predikerte sannsynligheter som er høyere enn terskelen som tilstede, og lavere sannsynligheter til fravær.

Deretter kan man kombinere antall observasjoner som faller i de fire mulige kategoriene (Tabell 2).

De fire tallene a til d kan brukes i en rekke indekser, som for eksempel: sensitivitet = a/P , spesifisitet = d/Q og odds-ratio = ad/bc . Sensitivitet er evnen til å predikere riktig tilstedeværelse, dvs. i hvilken grad modellen predikerer tilstedeværelse når dette er riktig. Spesifisitet er evnen til å predikere riktig fravær, dvs. i hvilken grad modellen unnlater å predikere falske tilstedeværelser. Kappa koeffisient = $[(a + d) - (((a + c)(a + b) + (b + d)(c + d))/N)] / [N - (((a + c)(a + b) + (b + d)(c + d))/N)]$ er en indeks som er mye brukt til å sammenligne modellens prediksjonsevne (se f.eks. Engler et al. 2004, Fielding and Bell 1997).

Verdien til de nevnte indeksene vil variere som en funksjon av det nivået terskelen settes til. En høy



Figur 7. En modell (f.eks. logistisk regresjon) gir prediksjoner som sannsynligheten for at arten er til stedet som en funksjon av en prediktorvariabel.

terskelverdi vil gi få falske predikerte tilstedeværelser men mange falske predikerte fravær, mens det mostatte skjer med en lav terskelverdi.

Ved å variere terskelen fra 0 til 1 kan man se hvordan de to typer feilklassifisering varierer. Man kvantifiserer da hvor stor andel av observasjoner hvor arten er fraværende som er feil predikert (modellen predikerer at arten er til stedet men det er usant; dette er lik 1- spesifisitet), og hvor stor andel av tilstede-observasjonene som er riktig predikert (dvs modellens sensitivitet) (figur 8).

Arealet under kurven (AUC = "area under curve") i figur 8 utgjøres av observasjonene som er riktig predikert. AUC-verdien beregnes som en andel av hele arealet av kvadratet (som representerer alle observasjoner) i figur 8. En god modell er en modell som minimerer begge feilkilder, dvs. har høy sensitivitet og høy spesifisitet (eller lave verdier for 1-spesifisitet). Da vil kurven skyves opp mot øvre venstre hjørne og AUC-

verdien vil nærme seg 1.0. Den stiplede linjen langs diagonalen viser en tilfeldig klassifisering av tilstedeværelse og fravær. En slik modell får en AUC-verdi lik 0,5. AUC-verdien har etter hvert blitt et vanlig kriterium for å karakterisere modellens prediksjonsevne.

Til slutt bør det nevnes at finnes det en rekke numeriske teknikker for å estimere forventningsskjevhet og presisjon i prediksjonsevner. Som skrevet om R^2 er verdien man får ved å bruke utvalget "for gode", dvs. en tror at modellen er bedre enn den faktisk er. For enkelte modeller er det mulig å korrigere for dette (f.eks. ved å bruke adjusted R^2 i stedet for R^2), men dette antar at modellen er riktig. En annen måte som nesten er blitt en standard metode er bruk av kryssvalidering. Kryssvalidering ble først utført på en enkel måte: man fjerner en observasjon fra datasettet, estimere parametrene med de andre observasjonene og predikerer observasjonen som ble utelatt (såkalt "jackknife" eller "leave-one-out" kryssvalidering). Denne metoden har ofte vist seg å være upresis (Hastie et al. 2001), og mange foretrekker nå å dele datasettet i flere grupper (k, oftest mellom 5 og 10), bruker k-1 grupper for å estimere parametrene og ser på prediksjonsevnen i forhold til den siste gruppen (se Boyce et al. 2002). Man kan bruke alle måter å dele opp datasettet i k grupper, eller i hvert fall et tilfeldig utvalg av disse. Dette viser seg å gi gode estimater av prediksjonsevnen. I hvert fall dersom man begrenser prediksjonen til en begrenset region (se Randin et al. 2006 for et eksempel på dårlige prediksjoner mellom regioner).

2.5 Prediksjonskart og feltvalidering

Et hovedresultat fra prediksjonsmodellering er kart over predikert utbredelse for arten som modelleres. Hvordan dette kartet skal forstås er et tema i seg selv (Araújo and Guisan 2006, Pulliam 2000). Her skal vi betrakte kartet som en geografisk visualisering av prediksjonsmodellen og forstå dette som predikert

Tabell 2. Klassifisering av prediksjoner og observasjoner for tilstede-fravær data.

	Predikert tilstede	Predikert fravær	
Observert tilstede	a	b	$P=a+b$
Observert fravær	c	d	$Q=c+d$
	$R=a+c$	$S=b+d$	$N=a+b+c+d$

utbredelse av artens habitat, dvs. hvor vi skal forvente å finne arten. Vi beskriver noen prinsipielle og praktiske aspekter ved utvikling og bruk av prediksjonskart. Særlig vil vi understreke viktigheten av å utføre feltkontroll for å etterprøve og forbedre kartenes pålitelighet.

2.5.1 Framstilling av prediksjonskart

Innledningsvis i dette kapittelet ble framstilling av prediksjonskart omtalt som et eget trinn (trinn 2) i en overordnet strategi for prediksjonsmodellering. Dette trinnet består av å beregne sannsynligheten for artens eller naturtypens forekomst som en funksjon (representert ved prediksjonsmodellen) av miljøfaktorer (prediktorvariabler) representert i digitale kart for modellområdet. Dette resulterer i et heldekkende kart som viser sannsynligheten for forekomst av arten (eller naturtypen) innenfor modellområdet.

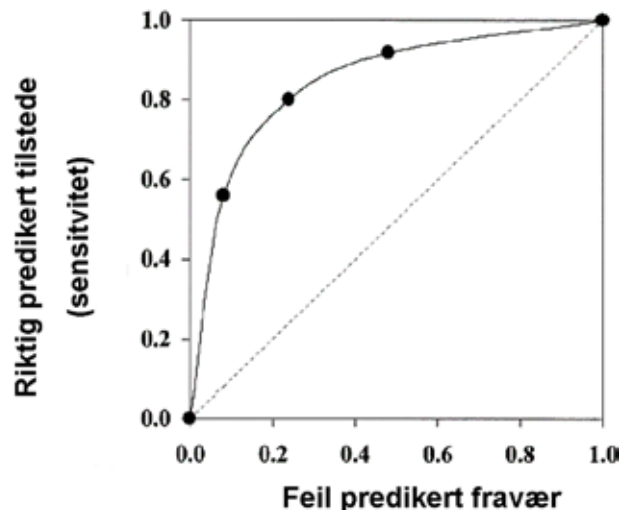
Prinsipielle aspekter

Det er viktig å understreke at påliteligheten til prediksjonskartet (dvs. hvor godt de faktisk stemmer overens med artens reelle forekomst) kan variere fra lav til høy. Denne kvaliteten avhenger av minst fire ulike forhold: a) hvilken modelleringsmetode som er anvendt, b) hvorvidt viktige miljøfaktorer (for arten) inngår som forklaringsvariabler i modellen, c) hvorvidt det finnes et tilstrekkelig antall observasjoner av arten med tilstrekkelig nøyaktig stedfesting, d) hvorvidt observasjonene av arten (både forekomst og fravær) er innhentet over hele spekteret av miljøforhold og kombinasjoner av miljøforhold som arten kan leve under. Nesten bestandig vil en første versjon av modeller for predikert utbredelse være mangelfull pga. disse forholdene. Det er derfor viktig å etterprøve kartenes pålitelighet ved feltvalidering (se 2.5.2).

I figur 9 har vi gjengitt et eksempel som viser prediksjonskart produsert ved bruk av ulike modelleringsmetoder basert på det samme datamaterialet. Effekten av alternative metoder og de øvrige faktorene på resultatet i prediksjonskartet er fremdeles lite evaluert. Dette skyldes dels at større oppmerksomhet har vært rettet mot modellenes prediksjonsevne kvantifisert vha. ROC-analyse og AUC-verdier.

Praktiske aspekter

Det er særlig tre praktiske forhold som spiller inn for



Figur 8. ROC ("Receiver operating characteristic") kurve som viser hvordan sensitivitet og spesifisitet varierer som en funksjon av terskelen. AUC-verdien tilsvarer arealet under kurven og integrerer dermed de to egenskapene over alle terskelverdier.

å kunne lage gode prediksjonskart: a) forekomst av et tilstrekkelig antall artsobservasjoner med tilfredsstillende stedfestingspresisjon, b) forekomst av digitale kart med relevant miljøinformasjon og c) hvor enkelt det er å lage kart basert på de ulike modelleringsmetodene.

I Norge er vi i en gunstig situasjon med hensyn til tilgang på digitaliserte, stedfestede artsfunn. Ved utgangen av 2007 var i alt 2,5 millioner artsfunn gjort tilgjengelige gjennom GBIF-Norge, og av disse er mer enn 95 % stedfestet med koordinater. Det ligger imidlertid en praktisk begrensning i at mange stedfestinger er forholdsvis upresise. De dekker heller ikke på langt nær alle artsgrupper. I kapittel 4 omtaler vi status for digitalisering av norske artsfunn.

Vi er også i en forholdsvis gunstig situasjon i Norge med hensyn til tilgang på digitale kart med miljøinformasjon. Imidlertid er ikke alltid den tilgjengelige miljøinformasjonen like relevant. Her kommer man i et dilemma, fordi man ønsker å ha gode prediktorvariabler knyttet til artsfunnene når man modellerer. Det er da mulig å inkludere disse variablene i de statistiske analysene for etablering av prediksjonsmodellen. Men for å lage kart over predikert utbredelse er man avhengig å ha heldekkende

kart for alle prediktorvariabler som inngår i modellen (med samme romlige oppløsning). Dette betyr at potensielle prediktorvariabler i modellene i virkeligheten begrenser seg til variabler det finnes tilgjengelige digitale kart for. For svært mange miljøfaktorer som er viktige for arters forekomst og utbredelse, finnes ikke heldekkende digitale kart. Dette er i praksis en av de viktigste begrensene faktorene for hvor gode prediksjonskartene kan bli både i Norge og i andre land. Det er imidlertid mulig å komme et stykke lengre ved å utvikle avledet, relevant informasjon fra eksisterende digitale kart. I kapittel 5 omtaler vi status mht. tilgjengelige kart og relevante avledede kart for prediktorvariabler. I kapittel 8 påpeker vi de viktigste utfordringene for å lage slike avledede temakart for bedre å kunne utnytte potensialet til prediksjonsmodellering.

En tredje praktisk side ved produksjon av prediksjonskart er at prediksjonsmodellen må transformeres til prediksjonskart når alle data og nødvendige digitale kart er på plass. De ulike modelleringsmetodene gir forskjellige muligheter for dette. En ytterlighet er Maxent (se 2.3.3) som benytter digitale kart for prediktorene og en fil med koordinater for tilstedeværelse av arten som input, og som produserer både analyseresultater (relativ betydning av ulike miljøfaktorer, AUC-verdier, osv.) og predikert utbredelse som direkte resultat. Noe mer komplisert er GLM der man ofte må dele opp kartet i mindre deler for å beregne sannsynlighet for tilstedeværelse (på grunn av innebygde begrensninger i antall pixler/observasjoner som R håndterer). Enda mer komplisert er metoder som regresjonstrær. Her må man konvertere kartfilene til nye formater og gjøre separate beregninger av sannsynlighet for forekomst utenfor selve modellen.

For GAM (og GLM) er det etablert en egen modul i programpakken S-Plus og R, kalt GRASP, som kan benyttes både til å etablere nisjemodellen og til å lage prediksjonskart. I disse programpakken kan prediksjonskartene kun lages for relativt små områder (få pixler). Men GRASP har utviklet en løsning for prediksjon over store områder med mange pixler ved å produserer LUT ("look-up-tables") filer. Disse kan benyttes i ArcView (via en GRASP "extension") til å lage prediksjonskart over store arealer (Lehmann et al. 2004, Lehmann et al. 2003) såfremt man har GIS-kart for prediktorene som er inkludert i den statistiske modellen. Metoden er brukt f. eks. ved modellering av marine sedimenter og arter, f. eks. for å skille hardbunn fra bløtbunn, og ved modellering av ålegras og

tareskog (Bekkby et al. *Submitted*, Norderhaug et al. 2007, Soldal et al. 2008).

2.5.2 Feltvalidering

Etter at man har laget et kart over predikert utbredelse er det viktig å etterprøve kartets prediksjonsevne ved å oppsøke utvalgte lokaliteter der arten ikke er kjent fra før. I en slik feltvalidering er det viktig å ha et klart formål med feltvalideringen.

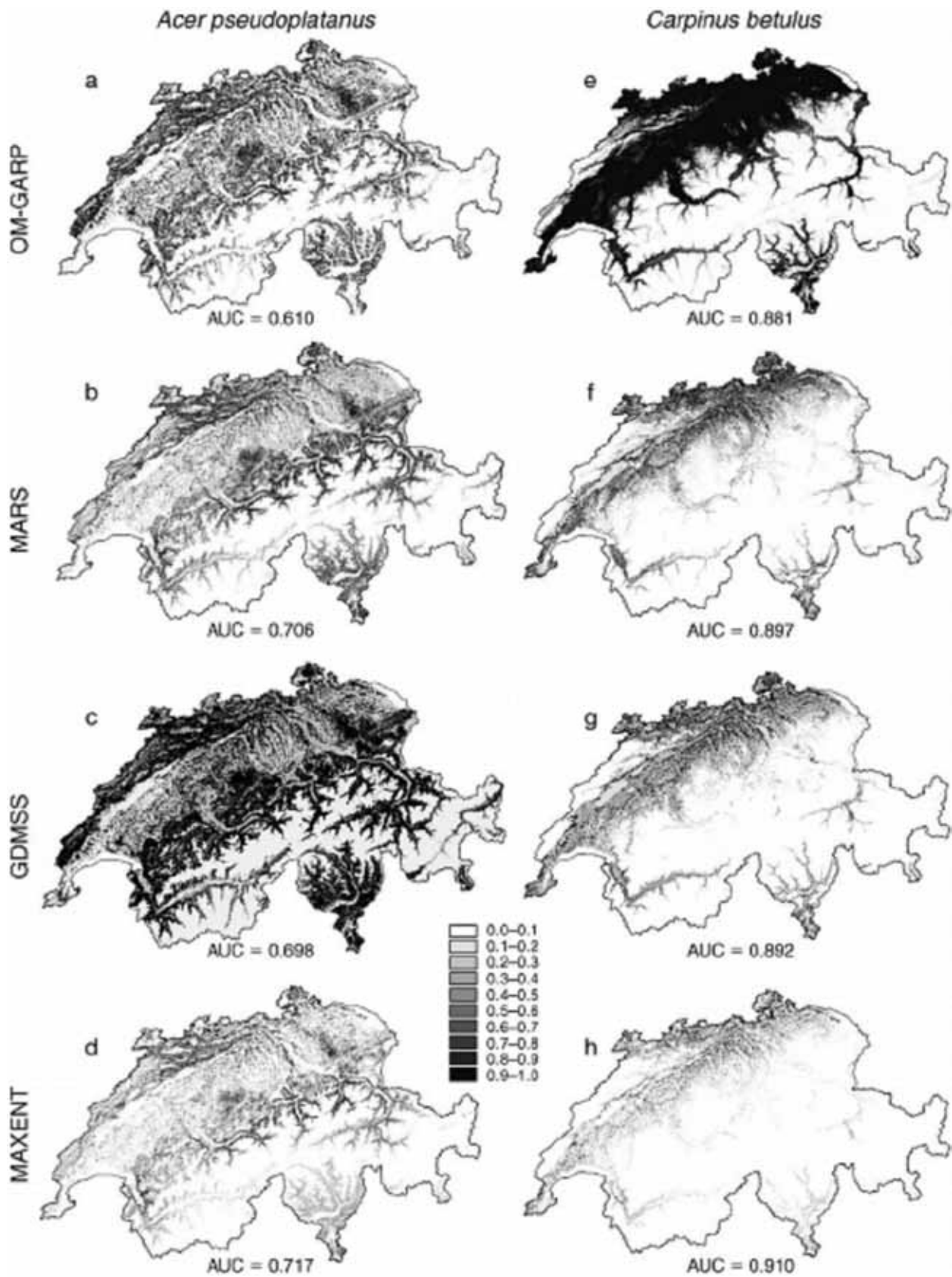
Enkel feltvalidering

Et opplagt formål er å sjekke om modellen faktisk er i stand til å predikere forekomster på steder som ikke er undersøkt. Dette gjøres ganske enkelt ved å oppsøke ikke-registrerte lokaliteter der kartet (dvs. modellen) predikerer høy sannsynlighet for forekomst. Flere arbeider har vist at slik modellstyrt søk etter nye forekomster er vesentlig mer effektive enn tilfeldige søk eller søk basert på personlig erfaring (Guisan et al. 2006, Thompson 2004).

Metodisk feltvalidering

Et mer ambisiøst formål er å utføre feltvalideringen på en slik måte at de nye dataene kan brukes som input til en ny versjon av prediksjonsmodellen. For dette formålet er det viktig å anlegge en metodisk tilnærming der man systematisk søker etter arten på steder der modellen predikerer ulike sannsynligheter for forekomst. En god framgangsmåte kan være å tilstrebe en tredelt stratifisering slik at man oppsøker steder med lav sannsynlighet for forekomst (for eksempel 5-30 %), middels (30-60 %), høy (60-90 %), og eventuelt høyere dersom modellen skulle predikere dette.

Videre er det essensielt at feltvalideringen utføres ved et randomisert eller systematisk valg av lokaliteter (i form av punkter eller søketraseer/ transektlinjer) og ikke benytter kriterier som tilgjengelighet, avstand fra hjemsted, osv. for valg steder med ulik grad av sannsynlighet for forekomst. For feltvalidering av marine bunnlevende arter har det vist seg å være nyttig å benytte nedsenkbart undervannskamera for registrering av arter i tilfeldig valgte punkter eller i punkter med regelmessig avstand langs tilfeldig valgte transektlinjer. Det er viktig å gjøre registreringer av ekte fravær av arten med geografisk stedfesting av observasjonene langs transektlinjene.



Figur 9. Predikert utbredelse for to treslag (lønn *Acer pseudoplatanus* og agnbøk *Carpinus betulus*) i Sveits. Kartene er basert på ulike modelleringsmetoder anvendt på identiske data. Mørkere farge indikerer større sannsynlighet for forekomst. Legg merke til hvor forskjellig metodene predikerer artenes forekomst (kilde: Guisan et al. 2007b).

3. Metodiske utfordringer

Med det store antallet modelleringsmetoder som finnes kan man kanskje tro at prediksjonsmodellering representerer en allmektig verktøykasse som altoverveiende gir pålitelige prediksjoner. Det er langt fra tilfellet. Det er snarere slik at økologisk prediksjonsmodellering befinner seg i en utviklingsfase og at mange forhold kan føre til usikre og endog misvisende prediksjoner. I dette kapitlet skal vi drøfte en del forhold som bidrar til dette.

3.1 Fundamental og realisert nisje

I prinsippet antar vi ved modellering at en art er til stede, med ulik grad av sannsynlighet, der kombinasjonen av miljøfaktorer er mer eller mindre gunstig og fraværende ellers. Den rendyrkede utgaven av denne tankegangen representeres av Grinnell's nisjebegrep (se Pulliam 2000) som sier at arten alltid er til stede under gunstige kombinasjoner av miljøfaktorer og fraværende ellers (se fig 10.A). Med andre ord, artens sannsynlige forekomst er en funksjon av ulike miljøfaktorer spesifikk tilstander, og disse kombinasjonene av miljøfaktorer er i prinsippet et uttrykk for artens fundamentale nisje.

Arter opptrer imidlertid ikke som perfekt forutsigbare responser, og det er flere grunner til dette.

En grunn kan være at arten under visse miljøforhold blir helt eller delvis utkonkurrert av en annen art som er bedre i stand til å utnytte denne kombinasjonen av miljøfaktorer (figur 10.B). Det vi observerer er artens realiserende nisje med en konkurrent til stede. Et velkjent eksempel på dette fra Norge er furuas opptreden på Østlandet hvor vi finner den som dominerende art på tørre og halvtørre partier (åskammer, soleksponerte øvre lisider) og på svært våte, næringsfattige områder (furumyrskog). På Vestlandet opptrer arten også dominerende på steder med fuktigere skogbunn, en voksestedstype som domineres av gran på Østlandet. Bakgrunnen for disse forskjellene er at grana er en sterkere konkurrent enn furu på frisk skogsmark og i skyggefulle terrengposisjoner. Grana har ikke innvandret til Vestlandet og dermed ser vi hvordan furuas fundamentale nisje også omfatter frisk skogsmark.

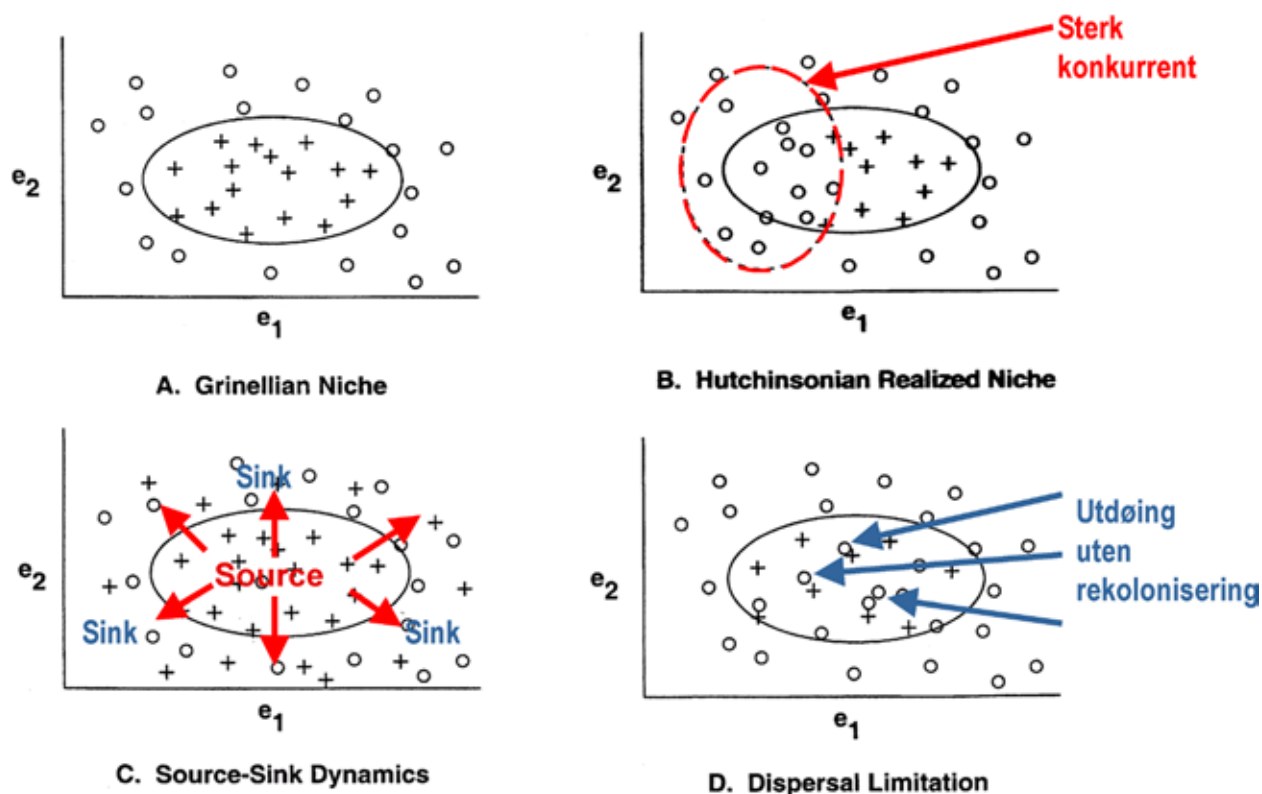
En annen grunn til at en art kan være fraværende fra et område der miljøkravene synes å være oppfylte kan være at arten ikke har evnet/rukket å etablere seg i dette området (som grana på Vestlandet). Videre kan

det skyldes mer lokal metapopulasjonsdynamikk, der arten har dødd ut fra en lokalitet og senere ikke evnet å kolonisere tilbake på grunn av for lang avstand fra eksisterende delpopulasjoner (fig. 10.D). Dette fenomenet er velkjent for en rekke insektsarter. Også ålegras i marine miljøer synes å vise slike tendenser.

Også det motsatte kan skje. En art kan forekomme på steder hvor kombinasjonen av miljøbetingelser ikke skulle tilsi at vi forventer å finne arten. Dette kan forekomme når vi har sterke lokale populasjoner ("source" (eller kilde) populasjoner) som produserer et overskudd av individer som sprer seg til habitater eller lokaliteter hvor arten ikke ville kunne eksistere som en selvreproduserende populasjon over lengre tid (fig. 10.C). En slik situasjon er påvist for kjøttmeis *Parus major* i Finland. Denne arten trenger å produsere et visst antall utfløyne unger per hekkende par for å balansere voksendødelighet. I frodig løvskog produseres langt flere utfløyne unger per par enn dette balansenivået, men i barskog produseres et underskudd av unger. Det er gjort en beregning som viser at en vesentlig del av alle kjøttmeiser i Finland hekker i slike sub-optimale "sink" habitater. Dette viser at habitat generalister ofte kan forekomme i sub-optimale miljøer som ikke alene har tilstrekkelig kvalitet til å opprettholde artens populasjonsnivå.

Alle disse forholdene som forårsaker uoverensstemmelse mellom artens fundamentale nisje og faktiske forekomst under ulike miljøfaktorer kan potensielt skape store problemer ved modellering av artens respons i forhold til miljøfaktorer. Det er faktiske tilstede og fravær-observasjoner som benyttes ved modellering, og det vil derfor være en fare for at man lager modeller som er overtilpasset disse spesifikke observasjonene. Aktuelle løsninger på dette problemet er a) at man ikke nødvendigvis skal etterstrebe maksimal modelltilpasning til data, b) at man aktivt tester modellen med helt uavhengige data som er innhentet fra et annet geografisk område (cfr. Araújo and Guisan 2006).

I lys av problemstillingene omtalt over kan man saktens spørre om det er meningsfullt å skille mellom fundamental og realisert nisje. Alle karakteriseringer av nisjer vil være ufullstendige beskrivelser av de abiotiske og biotiske faktorer som oppfyller en arts økologiske krav. I en bok om nisjeteori gikk Chase & Leibold (2003) bort fra dette skillet og foreslo en oppdatert definisjon av nisjebegrepet: de økologiske forhold som tilfredsstiller en arts minimumskrav slik at reproduksjonsraten for en delpopulasjon er lik eller



Figur 10. Fire prinsipielle sammenhenger mellom en arts nisje og faktiske forekomst i forhold til to miljøgradienter (e_1 og e_2). I hvert diagram markerer den liggende ellipsen artens fundamentale nisje (dvs. der miljøbetingelsene tilsier at arten kan eksistere som en selvreproduserende populasjon), + markerer tilstedeværelse av arten på en lokalitet, og o markerer fravær av arten på en lokalitet. I følge Grinnell's nisjebegrep (A) vil arten finnes på alle steder hvor miljøforholdene er gunstige, og ikke ellers. Hutchinson's realiserste nisje (B) postulerer at arten vil være fraværende fra den (de) delen(e) av nisjerommet som benyttes av en dominerende konkurrent. I situasjoner med source-sink dynamikk (C), vil arten kunne opptre utenfor den fundamentale nisjen ved at nye individer koloniserer fra nærliggende lokaliteter/habitater hvor reproduksjonsraten er høyere enn mortalitetsraten. Både spredningsbarrierer og metapopulasjonsdynamikk (D) kan føre til at arten er fraværende på lokaliteter hvor miljøforholdene er gunstige. Modifisert etter Pulliam (2000).

større enn dødsraten (s. 19). Araujo & Guisan (2006) framhevet på sin side at det er viktig å skille mellom nisjemodeller og romlig eksplisitte modeller av forekomst. Nisjemodeller som baserer seg på kvantifiserte miljøfaktorer gir bare en prediksjon om artens *potensielle habitat*. Når nisjemodeller og romlige eksplisitte kart over miljøfaktorenes variasjon kombineres får man prediksjonskart som viser den *potensielle geografiske utbredelsen* til artene.

3.2 Romlig skala

Begrepet romlig skala står sentralt i nisjemodelleingsarbeidet, fordi responsvariablene (f.eks. forekomsten av en gitt art) framviser variasjon på flere romlige skalaer, og fordi variasjon på disse ulike skalaene har ulike økologiske årsaker. Når vi ønsker å *predikere* variasjon i forekomsten av en art (responsvariabel i nisjemodellen), vil derfor mengden av variasjon i

responsvariabelen, hvilke prediktorvariabler som er relevante og hvor stor andel av totalvariasjonen som lar seg predikere, bestemmes av hvilken romlig skala vi velger å betrakte artsforekomsten på.

Begrepet romlig skala omfatter flere aspekter (Dungan et al. 2002). Det er derfor nødvendig å presisere hva vi mener med romlig skala. I prediksjonsmodellering er det særlig to aspekter som er viktige (Dungan et al. 2002, Wiens 1989):

- Romlig oppløsning (*kornstørrelse, grain*). Den minste enheten vi betrakter et naturfenomen på; f.eks. rutestørrelsen (pixelstørrelsen) vi bruker for å registrere forekomsten av en art.
- Romlig utstrekning (*extent*): Utstrekningen av det området vi har plassert observasjonsheter innenfor når vi betrakter et naturfenomen; f.eks. undersøkelsesområdet.

Skala-nivåer

Den romlige oppløsningen har stor betydning for hvilke miljøfaktorer man kan registrere på en meningsfull måte (figur 11). Den romlige utstrekningen har på den annen side stor betydning for hvor store økologiske forskjeller (økologisk gradientlengde) man fanger opp.

På en nasjonal-regional skala (med en romlig oppløsning på 1-5 km og utstrekning på mange hundre km) er det klimafaktorer som er de viktigste prediktorene. Klimafaktorene bestemmer artenes overordnede utbredelse. Når man øker den romlige oppløsningen krysser man en svært viktig skala-grensesone i området 250–750 m der det finner sted et skifte fra regionale (klimatiske) faktorer til lokale (topografiske og jordbunn/substrattypen) faktorer som viktigste mønsterbestemmende faktorer. Dette er diskutert for moderne kulturlandskap, men også generelt av Økland et al. (2006). Når man analyserer arters forekomst på landskapskala (med en romlig oppløsning på 10-100 m og utstrekning på 10-50 km) kommer faktorer som solinnstråling, løsmassefordeling og markfuktighet inn som sterke prediktorer. Disse er avgjørende for lokale forekomster av arter og naturtyper. En kompleksgradient som normalt oppfattes som lokal, men som kan framvise mønstre på skalaer som overskrider dette grensesone-området er mineralnæringsgradienten, på grunn av den store variasjonen i skala for berggrunn- og løsmassevariasjon.

I marine miljøer er GIS-modellerte terrengeskaper sentralt for å utvikle gode prediksjonsmodeller (Bekkby et al. 2005). Substrattypen er en sentral

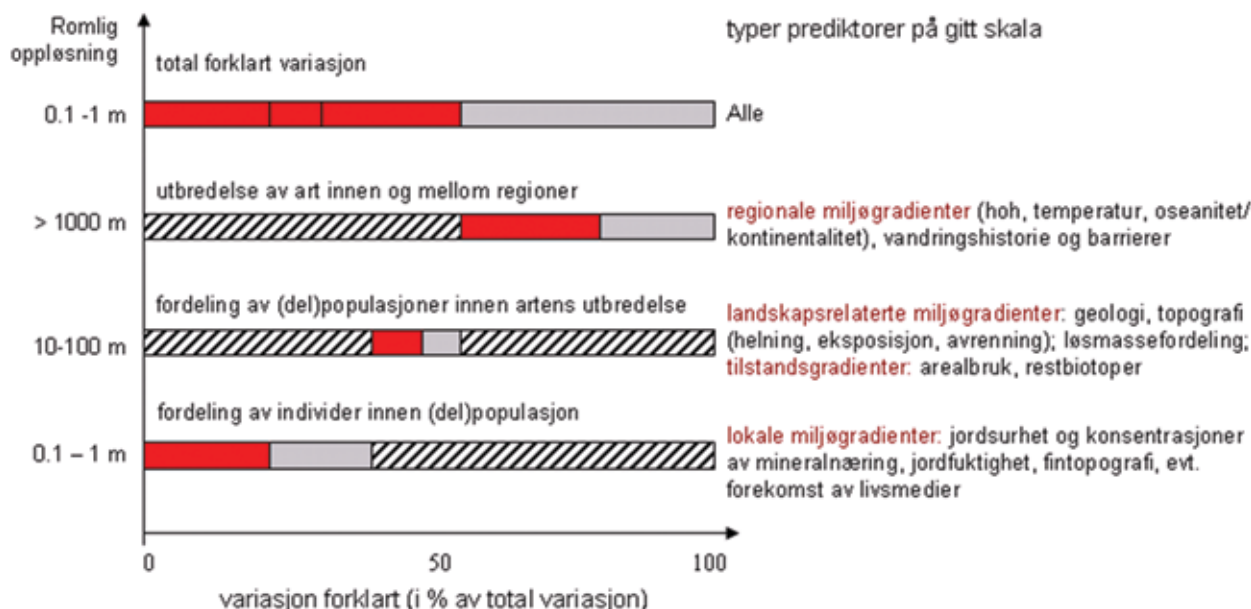
miljøparameter for marine arter og naturtyper, og en oppløsning på 25 m har vist seg å være egnet for å kunne utnytte terrengvariasjonen til å si noe om substrattypen. Denne muligheten blir kraftig redusert ved en grovere oppløsning (Bekkby et al. 2005).

Ved prediksjonsmodellering vil den romlige skalaen bestemme hvor mye variasjon det er i responsvariabelen og hvor mye av denne som kan forklares av et gitt sett prediktorvariabler. Derfor er både valg av romlig skala (oppløsning og utstrekning) og prediktorvariabler helt avgjørende for forklaringssevnen til en nisjemodell. Et viktig spørsmål er om det finnes generelle mønstre i arters forekomst (og fravær) på ulike skalaer og hvilke typer prediktorvariabler som forklarer mønstrene på disse skalaene.

Nisjemodelleringens formål er å predikere så stor andel av tilstedeværelse og fravær (eventuelt kvantitativ forekomst) i responsvariabelen som mulig, på det romlige skalanivået vi er interessert i, eller generelt ved at vi adresserer flere ulike romlige skalanivåer i samme analyse. Figur 11 illustrerer hvordan andelen riktig predikert forekomst for en lite bevegelig organisme som f.eks. en karplanteart vil kunne tilskrives prediktorer på ulike skalanivåer. Det er særlig to skalanivåer hvor det ofte er mulig å forklare mye variasjon: a) i området for variasjon langs lokale miljøgradienter (med registreringsenheter, pixelstørrelse) i området 0,1–1 m², og b) i området for regionale eller klimatiske miljøgradienter (registreringsenheter 1 km² og større). I mellomområdet vil lokal topografi, geologi (både berggrunns- og kvartærgeologi) og i noen grad lokale og regionale miljøgradienter kunne være kilder til variasjon. Studier av artssammensetningen i disse to 'hovedskalanivåene' hører inn under parallelldisiplinene samfunnsøkologi og biogeografi, som egentlig bare er to aspekter av samme disiplin men som adresserer ulike romlige skalaer. For andre organismegrupper kan miljøfaktorer på det mellomliggende landskapsnivået forklare stor variasjon i artenes forekomst.

Skala-mismatch

Ved prediksjonsmodellering kan man lett komme i en situasjon der artsobservasjoner foreligger på en skala, mens miljødata foreligger på en annen skala. En slik "mismatch" kan gå begge veier. Hvis man har artsobservasjoner med en høy romlig oppløsning (for eksempel med stedfestingspresisjon på 3-5 m vha. GPS), er det lite tilfredsstillende å kun benytte predik-



Figur 11. Variasjon i en responsvariabel (f.eks. forekomsten av en gitt karplanteart) på tre romlige skalaer. Romlig skala omfatter faktorer som spenner fra individstørrelse (10 cm) til artens utbredelsesområde (100 mil). Variasjon mellom observasjonenheter innen undersøkelsesområde på et gitt skalanivå er fordelt på forklart variasjon (rødt) og uforklart variasjon (grått), mens skravur angir variasjon på andre skalaer. Se teksten for ytterligere forklaring.

torvariabler med en romlig oppløsning på 1 km. Den kortsiktige "løsningen" i denne situasjonen er å slå sammen alle observasjoner innen hver km² rute for å få skalasamsvar mellom respons og prediktorer. Den motsatte situasjonen oppstår når man har prediktor-data med høy romlig oppløsning (for eksempel kart med 25 m oppløsning), men de aller fleste artsobservasjonene med en stedfestingspresisjon på 1 km². I denne situasjonen må man helt eller delvis gi avkall på å bruke prediktorer med høy romlig oppløsning (evt. slå sammen pixler og lage avledete indekser på det skalanivået der artsobservasjonene foreligger).

En vesentlig bedre, men langt mer arbeids- og kostnadskrevende måte å komme rundt de omtalte problemene er følgende. For upresise artsobservasjoner kan man ofte forbedre stedfestingspresisjonen til artsobservasjonene ved å kontakte observatøren (gitt at vedkommende lever) og få en mer nøyaktig angivelse av funnstedet. Enda bedre er det selvsagt å oppsøke lokaliteten på nytt og få en presis stedfesting. For miljøfaktorer med en grov oppløsning kan man i flere tilfeller gjøre en nedskalering ved å model-

lere miljøfaktorenes systematiske variasjon i forhold til annen kartinformasjon med finere oppløsning. Det er for eksempel utviklet standardmetoder for nedskalering av temperatur fra grov (1-5 km pixler) til fin skala (for eksempel 10-30 m pixler) ved hjelp av en terrengmodell. Tilsvarende remodellering av miljøfaktorer er under utvikling for å få marine miljødata med relevant romlig oppløsning i forhold til vind-, bølge- og havstrømeksponering.

3.3 Manglende (relevante) miljødata

Ofte mangler man relevante miljødata for prediksjonsmodellering av arters utbredelse og lokale forekomst. Dette resulterer i at en stor andel av variasjonen i artens forekomst ikke forklares av den beste modellen man klarer å lage (residualdevians er stor i forhold til total devians). En vanlig situasjon er at man ikke kjenner artens økologi tilstrekkelig, og dermed ikke vet hvilke prediktorvariabler man skal benytte for

å modellere artens forekomst. Det kan finnes viktige miljøfaktorer som er med på å bestemme artens forekomst i naturen, som vi ikke vet om og som vi derfor ikke har inkludert i prediksjonsmodellen.

Et annet forhold er at man kjenner artens økologi, men har likevel ikke informasjon om de essensielle miljøfaktorer fra de stedene hvor arten forekommer (og ikke forekommer) fordi det er svært tidkrevende eller kostbart å registrere disse. Det er imidlertid vel verdt å investere innsats i å opparbeide gode miljødata i datasettet som benyttes til modellutvikling. Nyten kommer i form av vesentlig bedre modeller når essensielle miljøfaktorer er blant prediktorvariablene.

Selv om man har utviklet en god modell basert på relevante miljødata for å predikere en arts forekomst, kommer det en ny begrensning når man skal anvende modellen til å predikere arens forekomst utenom de steder som inngår i det empiriske datasettet. De relevante miljøfaktorene er ganske enkelt ikke kartlagt over større områder, eller de er kartlagt med en annen romlig oppløsning enn den som er relevant for arten. Heldekkende kartlegging av miljøfaktorer over større områder (fra en kommune til hele Norge) blir raskt så kostbart at det ikke finnes betalingsevne (eller -vilje) til å gjennomføre kartleggingen. Man kan da ta et langt skritt tilbake og helt utelate flere av de viktige prediktorvariablene som gjorde modellen god og kun benytte prediktorvariabler som kan hentes ut fra heldekkende kart. Alternativt kan man undersøke om det er mulig å avlede informasjon om viktige miljøfaktorer fra eksisterende digitale kart. Dette er gjerne den nest beste veien å gå og involverer alt fra enkle avledninger til mer komplekse koblinger av informasjon fra ulike kart og satellittbilder/flyfotos. Konkrete avledete kart for miljøfaktorer er nærmere omtalt i kapittel 5.2.

3.4 Feilkilder i datagrunnlaget

3.4.1 Stedfestingspresisjon

Artsforekomster er registrert i Norge i løpet av mer enn hundre år. Gjennom denne tiden har metodene for stedfesting endret seg vesentlig. Før detaljerte kart fantes, ble artsfunn gjerne knyttet til stedsnavn (fylke, kommune, kjente steder) og den faktiske posisjonen i forhold til disse referansene kunne være meget unøyaktig sett med dagens ønske om stedfes-

tingspresisjon med meters nøyaktighet. Ikke desto mindre kan eldre funn være stedfestet med høy nøyaktighet gjennom detaljerte feltnotater. For eksempel kan gamle marine data være mer presise enn senere kartreferanser gjennom kobling til gjenkjennelig referansepunkter som kompassretning og avstand.

Utover på 1960-tallet ble kartblad i M711-serien med UTM-referanser produsert. Dette ga mulighet for stedfesting med 1 km og 100 m nøyaktighet. I denne sammenheng er det viktig å være klar over at det finnes to versjoner av UTM-koordinatsystemet: ED50 og WGS84. Koordinatene i disse versjonene er forskjøvet 100-200 m i både øst-vest og nord-syd retning. Selv om WGS84 kartene (blått rutenett) ble etablert i 1984 er fremdeles ED50-kart (svart rutenett) i bruk. De fleste er fullt klar over betydningen av å angi riktig UTM-datum, men det er nok mange stedfestinger med 100 m presisjon som i realiteten er forskjøvet og dermed har lavere stedfestingspresisjon enn det som er angitt. En annen feilkilde med lignende effekt er at observatører plasserer et artsfunn i naboruten til der funnet egentlig er gjort. Dette er nok en ganske hyppig feil.

I løpet av de siste 5 årene har stadig flere feltbiologer tatt i bruk GPS-posisjonering av artsfunn. Dette gir normalt en presisjon på 1-5 m og høy kvalitet på stedfestingene.

Ved prediksjonsmodellering er det viktig å sortere artsobservasjonene i forhold til hvilken stedfestingspresisjon de har for å unngå problemer knyttet til skala-mismatch (jfr. 3.2). Det viktigste å være klar over i denne sammenhengen er at artsobservasjonene ikke har en stedfestingsnøyaktighet som er vesentlig dårligere enn den romlige oppløsningen til de mest detaljerte kartene med prediktorverdier.

I Norge har vi allerede en stor mengde med digitaliserte, stedfestede artsobservasjoner som er lett tilgjengelig via GBIF, jfr. kap. 4. For omtrent 90 % av disse funnene er stedfestingsnøyaktigheten angitt. Det er imidlertid viktig å være klar over at stedfestingsnøyaktighet kan være noe lavere enn den klassen som er angitt. Det gjelder særlig UTM-baserte koordinater med 100 m presisjon som forklart over.

3.4.2 Skjevheter i artsobservasjoner

En annen feilkilde som forekommer systematisk i forhold til det aller meste av stedfestede artsfunn er at det ikke ligger noen helhetlig innsamlingsprosedyre

(design) til grunn for observasjonene. Artsfunnene er rett og slett gjort der personer med artskunnskap har ferdes. Det er således velkjent at det er en overhyppighet av artsfunn nær de store universitetsbyene, i kort avstand fra vei og nær bosted/hytte til eksperter. Videre er det en tendens til at arter overdokumenteres nær sin yttergrense og tilsvarende underdokumenteres der de forekommer hyppig.

Den samlede effekten er at artsfunn ikke fordeler seg representativt i forhold til artens reelle forekomst, og dette kan slå ut som feil i prediksjonsmodeller (Edwards jr et al. 2006). Som nevnt under 2.5 er det viktig å utføre metodiske feltvalideringer, blant annet for å avsløre slike feilkilder.

3.5 Statistisk over- og undertilpasning

Alle metodene omtalt i kapittel 2 kan benytte et stort antall prediktorer. Videre benytter flere av metodene (GAM, tremodeller, GARP, Maxent) matematiske eller regelbaserte tilpasninger som gjør det mulig å få til en svært tett kobling mellom respons og prediktorer i observasjonsmaterialet. Innen statistikk kalles dette overtilpasning ("overfitting") og representerer en falsk sammenheng og ikke reell kunnskap.

Overtilpasning

Flere av metodene omtalt i 2.3 har innebygde mekanismer for å redusere graden av overtilpasning. I GAM kan man regulere dette ved å justere hvor tett kurvetilpasningen tilpasser datapunktene. I tremodeller reduseres overtilpasning ved å utføre beskjæring ("pruning") av noder som i realiteten bidrar lite til modellens prediksjonsevne. I Maxent reduseres graden av overtilpasning ved å justere verdien til reguleringsparameteren(e). Det er viktig å bygge opp solid erfaring med de ulike prediksjonsmetodene for å unngå/redusere omfanget av overtilpassede modeller.

Et viktig grep for å unngå statistisk overtilpasning er å utføre kryssvalidering (se kapittel 2), som nå ofte omtales som et obligatorisk trinn ved prediksjonsmodellering. Dersom modellen i stor grad er beheftet med overtilpasning til observasjonsmaterialet (trenings datasettet), vil dette normalt vises ved lave AUC-verdier. Dette er bakgrunnen for at det etter hvert er blitt et krav å gjøre kryssvalidering og presentere AUC-verdier eller annen tilsvarende informasjon

for å få publisert prediksjonsmodelleringsarbeider i vitenskapelige tidsskrifter.

Selv om man utfører kryssvalidering på riktig måte, er man likevel ikke garantert å avsløre misvisende modeller. Den typiske måten å kryssvalidere modeller på er å dele det empiriske datasettet i flere deler (for eksempel 5) og suksessivt utelate en del for validering mens man lager modellen basert på hoveddelen av materialet. Dersom hele materialet er beheftet med systematiske samplingsfeil (se 3.4.2), vil de samme feilkildene være representert i så vel trenings datasettet som i valideringsdatasettet, og kryssvalideringen vil ikke avsløre dette. Dette er bakgrunnen for at vi i denne rapporten har framhevet betydningen av å utføre en uavhengig feltvalidering der helt nye data samles inn med en metodisk sunn innsamlingsprosedyre for å få virkelig gode valideringsdata, se 2.5.2. I prinsippet burde de nye valideringsdataene samles inn fra et annet geografisk område enn der trenings datasettet stammer fra (Araújo and Guisan 2006, Randin et al. 2006). Dersom det forekommer sterke korrelasjoner mellom ulike miljøfaktorer i området, vil nye data repetere disse sammenhengene og modellen blir ikke tilstrekkelig utfordret til å avdekke viktige årsakssammenhenger.

Undertilpasning

De fleste som arbeider med modellering er klar over faren for å lage en modell som er overtilpasset med for mange prediktorer og som gir en falsk høy forklaringsstyrke i forhold til det konkrete observasjonsmaterialet. Man bør imidlertid ikke bare fokusere på faren for overtilpasning. Undertilpasning, dvs å bruke en for enkel modell, kan være minst like alvorlig og gi store systematiske feil. Problemet er spesielt alvorlig med små datasett hvor presisjon er relativt dårlig og man bruker en enkel modell. En slik modell vil gi dårlige prediksjoner siden flere variable enn dem som er i modellen kan spille en viktig rolle. Undertilpasning kan også skje om man har et utvalg hvor prediktorvariablene viser lite variasjon (for eksempel om alle observasjoner kommer fra samme høyde).

En typisk situasjon ved modellering er at man har en stor andel variasjon i responsvariabelen som ikke er forklart av modellen. Dette kan skyldes flere forhold, men en vanlig situasjon er at en eller flere miljøfaktorer som virker inn på artens forekomst ikke inngår som forklaringsvariabler i modellen. Dette er grunnen til at vi har lagt så stor vekt på å framheve

betydningen av å lage kart over miljøfaktorer som artene responderer i forhold til (kap. 5.2).

Ekstrapolering

Prediksjonene fra GAM kan brukes til å lage romlige prediksjoner. Noe som kan være problematisk er at man lager prediksjoner utenfor de verditilstandene av prediktorvariablene som ble brukt til å lage modellen. Prediksjonene vil i disse områdene være ekstrapolasjoner, er dermed ikke til å stole på.

Det samme er tilfelle for Maxent. Her viser prediksjonskartet hvilke deler av predikert utbredelse som er knyttet til prediktorverdier som ligger utenfor de verdiintervallene som inngikk i modelleringsgrunnlaget. Disse delene av predikert utbredelse får en egen skravor ("clamped").

3.6 Hvor gode prediksjoner kan vi forvente?

I kapittel 2.4.2 redegjorde vi for AUC-verdien som et mål på modellens prediksjonsevne. En modell med AUC-verdi = 1.0 predikerer perfekt både tilstedeværelse og fravær, mens en modell med AUC-verdi lik 0.5 er likeverdig med tilfeldig gjetting av tilstedeværelse og fravær. Som en grov gruppering av AUC-verdier kan man benytte følgende gradering (Swets 1988):

- 0.5 – 0.6 knapt noen prediksjonsevne
- 0.6 – 0.7 dårlig
- 0.7 – 0.8 middels
- 0.8 – 0.9 god
- 0.9 – 1.0 svært god

I kapittel 2.2 gjenga vi et eksempel (*Eryngium alpinum* fra Sveits) der man hadde laget en prediksjonsmodell med en AUC-verdi på 0,98. Dette representerer en eksepsjonelt god prediksjonsevne som viser hva som er mulig når datagrunnlaget er godt og forholdene for øvrig ligger til rette. Mer typisk er det at man klarer å lage middels til gode prediksjonsmodeller, dvs. AUC-verdier mellom 0,7 og 0,9 (Elith et al. 2006, se også tabell 4).

Det er flere forhold som virker inn på modellens prediksjonsevne. I kapittel 2.3.4 viste vi at *modelleringsmetoden* spiller inn. Videre syns økt *romlig oppløsning* (mer presis stedfesting) av observasjonsdata og prediktorvariabler å øke prediksjonsevnen (Guisan et al. 2007a, Thuiller et al. 2003). Også arters egenskaper spiller inn på hvor presist deres forekomst kan predikeres. Modeller for arter med vid geografisk utbredelse og bred habitatamplitude innen modelleringsområdet får gjerne dårlig til middels prediksjonsevne (AUC-verdier i intervallet 0,6 – 0,8). Modeller for arter med snever utbredelse og/eller smal habitatamplitude (dvs. økologiske spesialister) får gjerne god til svært god prediksjonsevne (AUC-verdier > 0,8) (Guisan et al. 2007b, se også tabell 4).

Til slutt må betydningen av gode data understrekes. Hvis man har presist stedfestede artsfunn og reelle fraværsobservasjoner sammen med en rekke relevante prediktorvariabler (miljødata) med høy romlig oppløsning, så har man et godt utgangspunkt for å lage gode prediksjonsmodeller. Ytterligere kvalitet oppnår man ved i tillegg å utføre en metodisk feltvalidering (se 2.5.2) og ved å opparbeide nye tilstede- og fraværsobservasjoner under ulike økologiske forhold. Da er det realistisk å få til modeller med svært god prediksjonsevne, spesielt for arter som er økologiske spesialister.

4. Stedfestet artsinformasjon

Stedfestede artsfunn er en av hovedpilarene i prediksjonsmodellering. Norge har kommet langt med å digitalisere og gjøre tilgjengelig informasjon om stedfestede artsfunn. I dette kapitlet skal vi kort redegjøre for status på dette området.

4.1 Informasjonskilder

Databaser med stedfestede artsfunn foreligger ved en rekke institusjoner og organisasjoner/ foreninger. All denne informasjonen er langt på vei samordnet og gjort tilgjengelig gjennom to dataportaler: den norske GBIF-noden ved NHM, UiO (<http://www.gbif.no/>) og Artsdatabankens artskart (<http://artskart.artsdatabanken.no/>).

Det er GBIF-systemet som står for selve samordningen av data. Fra GBIF-portalen kan man laste ned artsdata i stort omfang og få detaljert informasjon om artsidentitet, funndato, lokalitetsnavn, geografiske koordinater (med usikkerhetsangivelse), primærdatakilde, med mer. Webkart for artsdata er en svært nyttig innsynsløsning som gjør at man kan se alle observasjoner i forhold til ulike bakgrunnskart, men man kan ikke laste ned data herfra. Kjernen i GBIF-samordningen er et internasjonalt standardformat ("Darwin Core 2") for å dele artsinformasjon i distribuerte datasystemer. Siden opprettelsen av den norske GBIF i 2005 har det pågått en storstilt innsats for å ekstrahere digitaliserte artsdata til dette standardformatet. I dag er praktisk talt alle norske faginstitusjoner knyttet opp mot GBIF og gjør sine data tilgjengelig via dette systemet (se figur 12). I tillegg til data fra norske institusjoner, gir GBIF systemet også tilgang til funn fra Norge som stammer fra informasjonskilder i utlandet.

Hovedtanken bak GBIF er å fungere som en samordningsmekanisme som gir gratis tilgang til artsfunn fra en felles dataportal. De individuelle databasene forblir eiendommen til de respektive vertsinstitusjonene og det er disse som har ansvar for å oppdatere databasene.

4.2 Artsgrupper

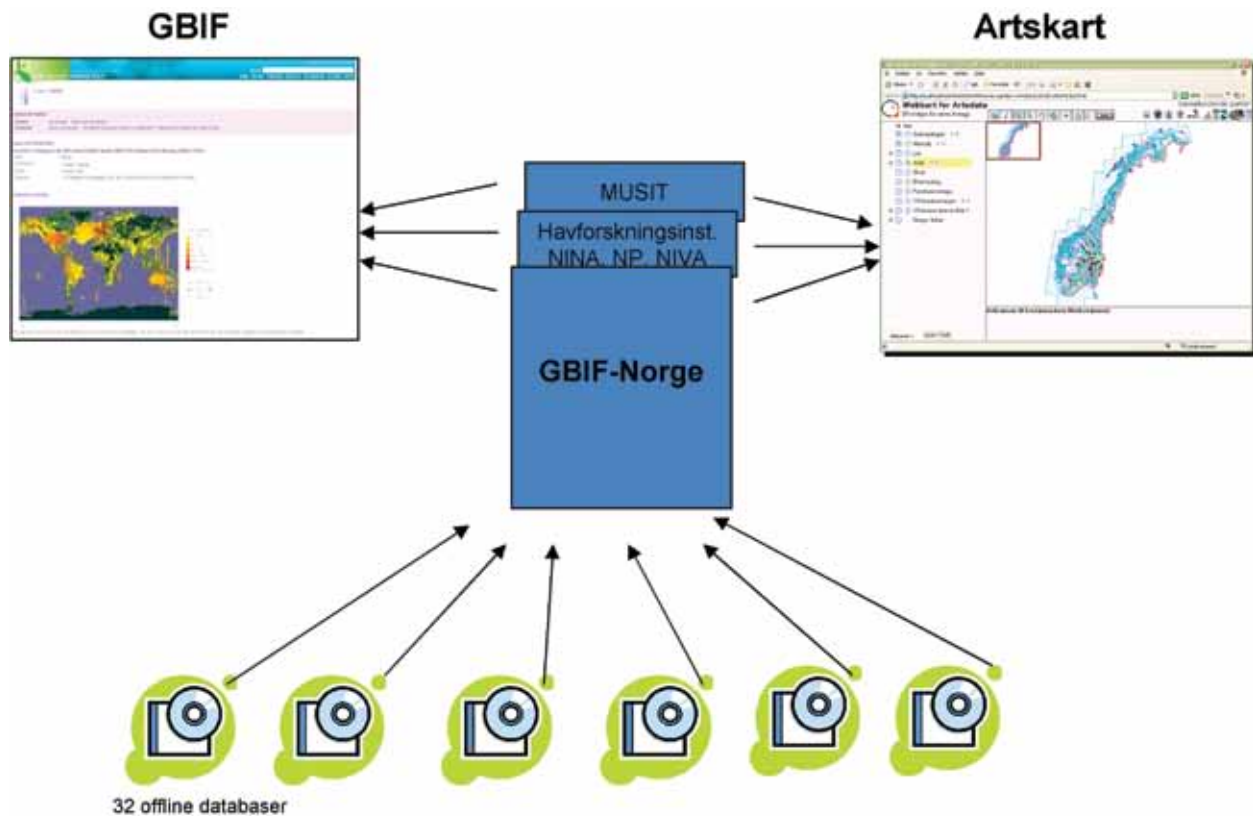
Ved utgangen av november 2007 var 2.4 millioner artsfunn tilgjengelige gjennom GBIF Norge. I tabell

3 ser vi at disse funnene altoverveiende representerer arter med botanisk forankring (særlig karplanter, moser, sopp og lav). Men også sommerfugler er rikelig representert. I alt finnes det om lag 12 mill. artsfunn ved de norske naturhistoriske samlingene, hvorav 3.3 mill. er botaniske, 5.3 mill. er zoologiske og 3.5 mill. er feltobservasjoner (flest botaniske). Dette betyr at vi i framtiden vil få et økende tilfang av data som kan brukes til å utvikle prediksjonsmodeller. I tillegg kommer selvsagt hele tilveksten av nye funn.

4.3 Dataegenskaper

Datakvalitet er av stor betydning for muligheter og begrensninger i modelleringsarbeidet. En viktig dataegenskap er stedfestingspresisjon. Stedfesting vha. GPS gir stor mulighet for framtidig datainnsamling med høy kvalitet på stedfestingen. En stedfestingspresisjon på 1000 m, som er vanlig for eldre kartangivelser av artsfunn, er tilstrekkelig for å modellere arters totale utbredelse i Norge. Slik modellering er allerede initiert ved Naturhistorisk museum, Univ. i Oslo (se figur 16). Med en stedfestingspresisjon på 10 m, som er realistisk med GPS posisjonering, kan man også lage prediksjonsmodeller for arters potensielle forekomst på enkeltlokaliteter.

Innsamlingsprosedyren som ligger til grunn for innsamling av data er en annen dataegenskap som er viktig når man skal bruke artsdata for prediksjonsmodellering. Det er typisk for data fra naturhistoriske samlinger (som representerer hovedinnholdet i GBIF-systemet) at de ikke er basert på en metodisk innsamlingsprosedyre (design), for eksempel en randomisert eller stratifisert protokoll. Artsfunnene stammer ganske enkelt fra steder hvor artsinteresserte personer har vært og hvor de synes det er verdt å rapportere artene fra. Det er velkjent at dette medfører en skjev rapportering i forhold til artenes reelle utbredelse og forekomst. For eksempel er områder omkring universitetsbyer, bosteder og feriesteder for artsinteresserte personer, samt områder nær vei overrepresentert med hensyn til artsobservasjoner. Dette kan lett innebære en feilkilde i forhold til prediksjonsmodellering. Heldigvis er flere av metodene forholdsvis robuste mot slike feilkilder, men dersom forekomster under visse økologiske betingelser ikke blir dokumentert, for eksempel fordi man gjenoppsøker arter på lignende steder som der de er funnet, vil ikke prediksjonsmo-



Figur 12. Dataflyt mellom primærdatabaser og de to dataportalene GBIF og Artskart. Informasjonen fra ulike primærdatabaser konverteres til "Darwin core" som er et felles dataformat, og gjøres tilgjengelig for allmennheten gjennom de to dataportalene. Den norske GBIF-noden samordner en rekke individuelle databaser, mens enkelte store institusjoner deler data direkte via de to portalene.

dellene være i stand til å predikere slike forekomster heller.

Et annet forhold som representerer en slags feilkilde er at naturhistoriske samlingsdata ikke omfatter informasjon om *arters fravær*. Dette er for så vidt svært naturlig ettersom samlingenes funksjon er å

dokumentere arters tilstedeværelse. Men siden naturhistoriske samlinger også arkiverer feltobservasjoner uten belegg, vil det være verdifullt om slike observasjonsarkiver kan utvides til også å omfatte fraværsobservasjoner.

Tabell 3. Antall stedfestede artsfunn tilgjengelig gjennom GBIF Norge ved utgangen av november 2007. Database betegner primærdatakilden ,

Database	artsgruppe	antall funn	stedfestingspresisjon					ukjent
			10 m	100 m	1000 m	10000 m	> 10000 m	
BioFokus_vascular	karplanter	192	73.4	20.8	0.5	0.5	4.6	0
Inventory_vascular	karplanter	904	49.1	10.5	0.6	0.2	39.4	0
KMN_vascular	karplanter	58 711	6.2	42.8	49.5	0.5	0.8	0
KMN_vxl	karplanter	142 691	0.3	10.9	74.1	14.1	0.3	0
O_vascular	karplanter	445 570	1.6	13.8	50.5	14.8	19.1	0
O_vxl	karplanter	1 084 910	0	0.6	61	32.9	5.3	0
TRH_vascular	karplanter	11 004	0.9	7.2	62.8	16	12.8	0
BG_lichens	lav	45 492	2.8	14.8	36.2	17	28.3	0.5
BioFokus_lichens	lav	3 981	21.3	75.4	1.3	0.4	1.4	0
Inventory_lichens	lav	1 376	20	32.1	36.4	10.6	0.6	0
NFRI_lichens	lav	210	0	11.4	41.9	0.9	45.7	0
NLDS	lav	51 065	2.9	12.6	33.3	15.8	33.1	2.1
NLH_lichens	lav	1 391	0	93.7	5.1	0	1	0
NXL	lav	75 495	1.2	20.2	72.1	6.3	0	0
O_lichens	lav	82 808	5	25.2	44.7	23.6	0	1.3
TM_Arthropoda	leddyr	17 477	0	0.9	19.2	56.6	23.1	0
BG_bryophytes	moser	11 660	0	19	5.1	2.8	72.8	0
BioFokus_bryophytes	moser	508	14.7	81.2	0.9	1.3	1.5	0
Inventory_bryophytes	moser	1	0	0	0	0	100	0
O_bryophytes	moser	16 773	0.7	15.9	7.7	4.4	71	0
LEPARB_lepidoptera	sommerfugler	72 813	1.6	67.3	20.1	0.6	10.2	0
O_lepidoptera	sommerfugler	58 233	3.2	52.9	29	1.6	13	0
VM_Lepidoptera	sommerfugler	310	0	0	0	0	0	100
BG_fungi	sopp	12 980	0	3.4	6.1	0.9	89.3	0
BioFokus_fungi	sopp	6 103	39.2	57.9	1.2	0.2	1.3	0
Inventory_fungi	sopp	39	64.1	7.6	0	0	28.2	0
NFRI_fungi	sopp	640	29.8	12.3	1.7	3.1	50	2.9
NSDS	sopp	1 026	0.2	5.3	67	24.6	2.6	0
NXS	sopp	40 262	3.3	20.3	67.3	7.8	1.1	0
O_fungi	sopp	126 815	3.9	19.1	24.6	11.6	40.5	0
O_Fungi_Notes	sopp	27 585	1.4	31.9	47.8	15.7	3	0
VM_Odonata	øyenstikkere	712	6.6	75.4	14.1	1.8	1.9	0
SUM		2 399 737	1.4	12.3	52.7	21.7	11.7	0.1

5. Digitale kart med miljøinformasjon

Digitale kart med miljøinformasjon er den andre hovedpilaren i prediksjonsmodellering. Det er artenes modellerte respons i forhold til miljøfaktorene, slik de er representert i kartene, som gjør det mulig å predikere hvor vi kan forvente å finne artene med ulik grad av sannsynlighet. Vi har mye relevant miljøinformasjon i form av digitale karter i Norge. Disse kartene er imidlertid utviklet for andre bruksformål enn prediksjonsmodellering. Det er karakteristisk at nesten samtlige slike temakart må gjøres om til rasterkart, transformeres, eventuelt også kombineres med andre kart, for å avlede nye kart over miljøfaktorer som er mer relevante til å predikere arters forekomst.

På denne bakgrunn har vi valgt å dele dette kapitlet i to. I kapittel 5.1 presenterer vi kartinformasjon slik de foreligger fra informasjonskildene. I kapittel 5.2 presenterer vi ulike avledninger som er laget, er under utvikling, eller som kan lages for å danne et bedre grunnlag til å predikere arters forekomst.

Alle kartene som omtales i dette kapitlet er tilgjengelige gjennom det nasjonale kartsamarbeidet "Norge Digitalt" (primærkart) eller gjennom de institusjonene som produserer kartavledningene. Per dato er forvaltning og distribusjon av rasterkart under utvikling innen Norge Digitalt, se for øvrig kapittel 8.1.

5.1 Primære miljøtemakart

Klima-, vær-, snø- og vanndata

Meteorologisk Institutt (MI) har i samarbeid med Norges Vassdrags- og Energidirektorat (NVE) utviklet en karttjeneste der en rekke data om snø, vær, vann og klima kan inspiseres (se <http://www.senorge.no/>). Tematisk omfatter karttjenesten følgende data med en tidsoppløsning på døgn, måned og år: vær (døgnetbør, døgntemperatur), snø (bl.a. snødybde), vann (regn og snøsmelting). Tidsmessig dekkes perioden fra 1960 fram til i dag.

Videre er det utviklet en klimaprognoser som viser bl.a. årsmiddeltemperatur, årsmiddelbørsum, fordampning, snødybde for perioden 2071-2100, samt endring fra 1961-1990 normalen til 2071-2100 prognosen.

Disse dataene dekker fastlands-Norge med en pixel-oppløsning på 1 km i UTM-sone 33.

Alle disse dataene kan bestilles for nedlastning for hele eller deler av Norge.

Terrengmodeller og dybdemodeller

En terrengmodell er et kart som viser hvordan høyden/dybden varierer innen et geografisk område. Den tradisjonelle måten å vise terreng- og dybdemodeller er ved koter (høydekurver) på kartet. For å få en heldekkende framstilling av terrengvariasjonen (det er jo ikke angitt høyder/dybder mellom kotene) transformeres et vektorisert kotekart til et rasterkart ved å interpolere høyden mellom kotene. Man kan da bestemme hvilken romlig oppløsning (pixel-størrelse) man ønsker i rasterkartet. I marint miljø har bruk av multistråleekkolodd blitt svært vanlig. Dette gir en enorm oppløsning på registrerte dybde data, og gir mulighet for stor oppløsning på rasterkartet.

Høyden over havet kan i seg selv være en prediktorvariabel, spesielt lokalt innen et område. Men verdier for middeltemperatur, temperatursummer, minimums- og maksimumstemperaturer er mer direkte funksjonelle prediktorer (se klimadata over), som vanligvis er bedre prediktorer enn høyden som sådan. Årsaken til det er samvirkning mellom høyde og andre faktorer som gjør det biologiske utslaget. For eksempel avtar temperaturen generelt med økende høyde over havet, men samme høydenivå vil ikke svare til samme temperatur på ulike steder – blant annet fordi temperaturen avtar mot nord. En terrengmodell kan imidlertid danne grunnlag for beregning av en rekke, biologisk meget signifikante variabler. En transformasjon av terrengmodell er solinnstråling som viser hvilke deler av landskapet som er sterkt solekspontert og hvilke som er skyggefulle. En annen viktig transformasjon er topografisk våthetsindeks ("topographic wetness index") som viser i hvor stor grad vann renner av eller akkumuleres i terrenget. Disse transformasjonene omtales nærmere i kapittel 5.2.

For Norge er det laget en totaldekkende høydemodell med 25 m pixelstørrelse som er beregnet fra N50 (M711-kartserien) med 20 m høydekoter. Det er mulig å lage mer presise høydemodeller fra N5 (økonomisk kartverk) med 5 m høydekoter, eller enda mer presise modeller basert på flybilder der høyden er målt med bedre enn 1 m nøyaktighet. For praktisk prediksjonsmodellering er den eksisterende 25 m-modellen tilstrekkelig i mange tilfeller, men tilgang til mer presise høydemodeller vil være ønskelig for nøyaktig modellering innen geografisk snevre områder.

For sjø finnes det dybdemodeller som angir avstand fra vannoverflaten til bunnen. Det finnes ingen dybdemodeller for ferskvann i Norge. Derimot

finnes det digitale terrengmodeller for Norges kyst og marine områder med 25 m romlig oppløsning (se Rinde et al. 2006). De nye 25 m dybdemodellene ble etablert ved interpolering av dybdedata (dvs. primærdata fra Sjøkartverket, der alle dybdepunkter og -koter ble benyttet) og høydedata fra Felles Kartdata Base (FKB) gjennom Norge digitalt (blant høydedata ble punkter og koter opptil 40 m høyde, samt kystkonturen, benyttet). I Barentshavregionen ble også Olex-data inkludert blant rådata for interpoleringen. Olex dybdedata samles inn systematisk og kontinuerlig av fiskebåter med Olex-utstyr, og ble gjort tilgjengelig gjennom Olex AS med 25 m oppløsning. Den valgte oppløsningen på 25 m er ikke reell for områder dypere enn 50 m pga. datamangel. For noen områder er det etablert dybdemodeller med finere oppløsning. Bl.a. er det for Skagerrak etablert et dybdedatasett med 12,5 m oppløsning, og for deler av Møre-kysten foreligger det datasett med 10 m oppløsning.

Basert på de ulike kildene og prosesseringen til dybdemodellene vil kvaliteten på de tilgjengelige dybdedataene variere mellom ulike områder. Det nye 25 m dybdedatasettet fra Sjøkartverket for Nord-Norge nord for 67. breddegrad vil ha høyest kvalitet siden disse er basert på nye målinger ved bruk av multi-stråledata. De etablerte nye 25 m dybdemodellene vil være av best kvalitet i kystnære områder der det både er tilgang til primærdata fra Sjøkartverket og høydedata på land fra FKB.

Digitalt markslagskart (DMK)

Dette er et høyoppløselig kart (skala 1:5000) som viser arealdekket på fastlandet opp til tregrensen. I den standardiserte terminologien for digitale kart betegnes kartserien AR5 (Bjørkelo 2006). De viktigste arealklassene som kan benyttes for prediksjonsmodellering er en rekke skogkategorier (bartredominert, løvtredominert, blandingsskog; fastmarksskog eller forsumpet skog; og fire produktivitetsskategorier – impediment, lav, middels, høy bonitet), tresatt myr, myr, åkermark, beitemark, "annen jorddekt fastmark" (som kan være alt fra tørrenger/strandenger til veikanter og gårdstun), fjell i dagen, bebygd areal, samferdsel (dvs. vei, jernbane, parkeringsplass, osv.) og vann.

Minste kartfigurstørrelse i DMK er 0,5–5,0 daa, avhengig av arealklasse, hvilket tilsier at et rasterkart med 25 m pixler normalt gir homogene arealer eller en kombinasjon av to arealklasser der én vanligvis dominerer (med mindre markslagsgrensen går gjennom

midten av ruten). DMK foreligger nå som digitaliserte rasterkart for hele Norge. Primærkartet eies og forvaltes av Norsk institutt for skog og landskap.

Noen av arealklassene i DMK er direkte anvendbare som prediktorer i prediksjonsmodellering, mens andre arealklasser må bearbeides videre for å bli biologisk relevante (for eksempel vil "annen jorddekt fastmark" som ligger mindre enn 10 meter fra kystlinjen sannsynligvis være havstrandeng; høyproduktiv, løvdominert skog som ligger inntil elver vil sannsynligvis være gråor-skog, osv.). Det pågår for tiden et utviklingsarbeid ved UiO/NHM for å lage prediksjonskart for treslag basert bl.a. på DMK.

Berggrunn

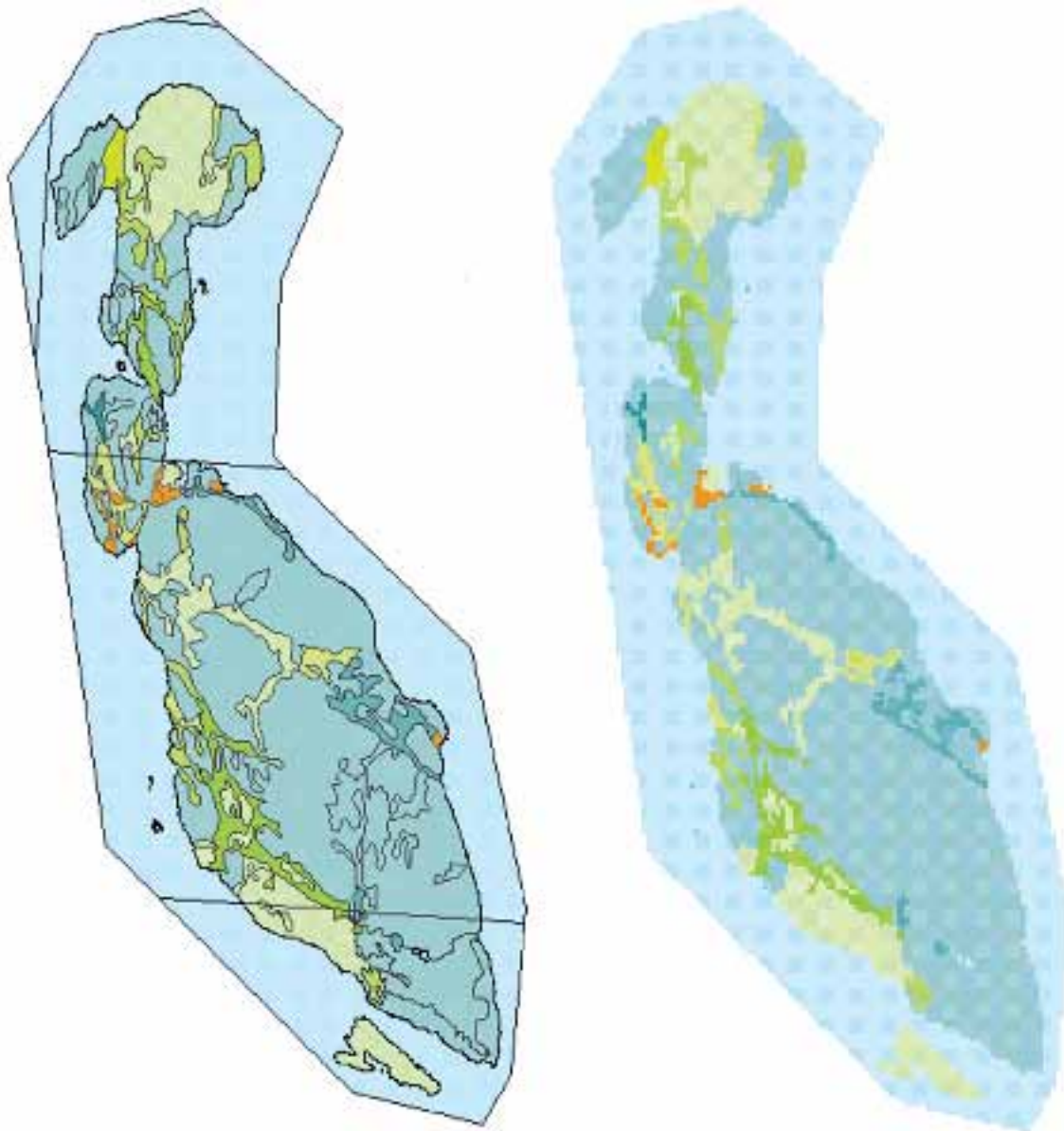
Berggrunnskart viser utbredelsen av de forskjellige bergarter på jordens overflate. De viser også strukturer i bergartene og lokalisering av kjente forekomster av malm, industrimineraler og bygningsstein. Kartleggingen bygger på feltobservasjoner, analyser av innsamlede prøver og tolkning av geofysiske data. Regionalt landsdekkende kart er tilgjengelig i målestokk 1:250 000 og kan lastes ned/bestilles på nettsidene til Norges geologiske undersøkelse (www.ngu.no).

Løsmasser/kvartærgeologi

Digitale kart som viser oversikt over løsmasser i Norge er basert på kvartærgeologiske kart i beste tilgjengelige målestokker (1:20 000, 1:50 000, 1:250 000) i tillegg til et kart i 1:1 million, som foreligger analogt i flere målestokker. Kartene er konvertert til digital form ved hjelp av skanning og vektorisering. I første omgang er kun flater og formelementer (linjer) etablert som digitale datasett. Senere vil også nedlastbar digital punktinformasjon bli tilgjengelig på nettsidene til Norges geologiske undersøkelse (www.ngu.no). Geologidatabasen inneholder opplysninger om jordartstype (løsmassetype), og med utgangspunkt i jordartenes egenskaper er det utviklet avledede temaer. Dette kan være infiltrasjonsegenskaper og grunnvannspotensial. Forholdene ligger også til rette for senere å utvikle andre temaer, f.eks. grunnvarme, avfallsdeponering, grunnforhold og telefare.

Marine geologiske kart

Norges geologiske undersøkelser har kartlagt bunnen med tanke på substrat flere steder langs norskekys-



Figur 13. DMK i vektorisert utgave (til venstre) og som rasterkart (25 m oppløsning) til høyre.

ten. Disse data gir informasjon om kornstørrelse og sedimentklasser. Også andre institutter, bl.a. NIVA, utfører også registreringer av kornstørrelse og penetrasjonsdyp tilknyttet overvåkingsprogrammer og en rekke andre prosjekter i kystsonen. Kornstørrelse og penetrasjonsdyp er gode indikatorer på substratklasse, og gir nødvendig informasjon i forhold til planlegging av feltinnsamlinger for vurdering av økologisk

status.. Slike kart er ikke heldekkende, men har spredd dekning over hele kysten.

Havstrømmer

For hele Norskekysten foreligger det (hos Meteorologisk institutt) informasjon om strøm på et grovt nivå (4 km romlig oppløsning). For noen utvalgte områder

langs norskekysten foreligger det informasjon om strøm på mer detaljert nivå, dvs. 500 m og 100 m oppløsning. Disse er laget enten som enlagsmodeller eller modeller med flere dybdelag, estimerte vha ROMS (Shchepetkin and McWilliams 2005) og spesialdesignet for bruk i marine habitatmodeller. I tillegg er det utviklet detaljerte, og avanserte strømmodeller for enkelte fjorder og mindre områder langs norskekysten, både av Havforskningsinstituttet, NIVA og andre institutter. Det er ikke etablert heldekkende modeller for norskekysten med en tilstrekkelig detaljert romlig oppløsning for habitatmodellering.

Veibase

Vbase er en landsdekkende digital veidatabase som forvaltes av Statens Kartverk. Basen inneholder geometrisk senterlinje for samtlige kjørbare offentlige veier, private veier og skogsbilveier i Norge.

Vbase har en nøyaktighet på +/- 2 meter (noen områder har +/- 5 meter). Ajourholdet skjer gjennom plantegninger/ferdigmålingsdata, fotogrammetrisk konstruksjon eller ved hjelp av GPS montert i bil. Oppdateringen foregår i en årlig syklus. For større veianlegg er Vbase oppdatert når veianlegget åpnes.

Landets kommuner, Statens vegvesens fylkesvise vegkontor, Kartverkets fylkeskartkontor, fylkesmannsembetet ved Landbruksavdelingen, Landbruksdepartementet, Vegdirektoratet og Kartverket sentralt samarbeider om ajourholdet av Vbase.

Moens vegetasjonsregioner/seksjoner

Vegetasjonsatlasen (Moen 1998) er en omfattende ekspertklassifisering av Norges vegetasjon i målestokk 1:1 000 000. Landet er delt inn i vegetasjonssoner (åtte soner fra nemoral til høyalpin) som reflekterer en nord-sør og høydegradient, og seks seksjoner (fra sterkt oseanisk til svakt kontinentalt) som representerer en gradient fra vest til øst, fra et mer humid/oseanisk til mer kontinentalt klima. Kartene er digitalisert og tilgjengelig i ESRI shape format.

Andre temakart

Det finnes ytterligere temakart som kan være relevante som grunnlag for å predikere forekomst av arter. Eksempler på dette er vegetasjonskart, naturtypekart, og kart over verdifulle skogsområder (MiS-kart). Vi har ikke omtalt disse nærmere fordi de sannsynligvis

har begrenset verdi. De viktigste grunnene til dette er: a) at de ikke er fullstendig arealdekkende, b) de er norske spesialkonstruksjoner som ikke kan brukes som prediktorgrunnlag utenfor landets grenser, og c) de har delvis status av å være parallelle til forekomstdata med tilsvarende enheter som vi er interessert i å predikere. Dvs. de hører mer hjemme på responsiden enn på prediktorsiden.

Det er imidlertid ingenting i veien for at den samme miljøfaktoren kan være respons i en analysesituasjon og prediktor i en annen. Hvis vi får et heldekkende kart som viser forekomst av ekstremrikmyr i (deler av) landet, så vil selvsagt dette være en meget verdifull prediktor for rikmyrsarter i (denne delen av) landet.

5.2 Modellerte (avledete) temakart

Som nevnt innledningsvis i dette kapittelet er det nødvendig å avlede nye temakart fra de primære temakartene for å få gode kart over relevante prediktorvariabler. Bakgrunnen for dette er at mange arter reagerer på svært spesifikke miljøfaktorer. Eksempler på dette kan være marine alger som trenger en viss grad av bølgeeksponering for å være konkurransedyktige, bløtbunnsarter som er avhengig av svake havstrømmer for å få tilstrekkelig finkornet sediment, insekter som trenger bakketemperatur godt over 20 grader større deler av sommeren for å fullføre sin livssyklus, lav eller mose-arter som ikke tåler luftfuktighet under et visst nivå flere dager i strekk, karplanter som er avhengig av mer eller mindre kontinuerlig markfuktighet, vedlevende sopp som trenger en viss hyppighet av dødt trevirke i visse dimensjoner og nedbrytningsstadier. Ofte er det slik at vi bare delvis eller endog dårlig forstår nøyaktig hvilke miljøfaktorer som virker inn på artenes konkrete forekomster. Ikke desto mindre vet vi for mange arter omtrent hva slags miljøfaktorer de reagerer på og vi kan ofte lage avledete kart med miljødata som korrelerer med disse faktorene. I det etterfølgende skal vi framheve en del slike kartavledninger.

Klimavariabler

Når det gjelder forhold som temperatur, nedbør, snødekke, osv så er vi knapt interessert i å bruke spesifikke verdier for konkrete datoer (værdato). Derimot

er det svært aktuelt å lage karter som viser en rekke klimaaspekter ut over dem som er beregnet fra MI/NVE. De viktigste av disse kan avledes fra følgende døgnerverdier: daglig minimumstemperatur, daglig maksimumstemperatur, døgnnedbør og snødybde. Den sykliske variasjonen av disse verdiene gjennom året kan brukes til å beregne avledete verdier som:

- årets minimumstemperatur og varighet av frostperioder
- vekstsesonglengde (gitt alternative basistemperaturer), varmesum, maksimumstemperatur
- fordamping (evapotranspirasjon), gjennomsnittlig *dagtemperatur* (ikke bare døgntemperatur)
- nedbørsum for ulike deler av året, antall regndager, nedbørsoverskudd (nedbør minus fordamping)
- luftfuktighet (mettet og omgivende damptrykk)
- snødekkevarighet

Formodentlig er det laget noen slike avledete karter i Norge siden dataene foreligger på en form som gjør dette forholdsvis lett. Det er imidlertid behov for å gjøre en systematisk gjennomgang av hva slags klimakarter som er relevante for prediksjonsmodellering, og hvilke slike karter som er laget allerede, respektivt bør utvikles. Det er også behov for å få informasjon (metadata) som viser nøyaktig hvordan slike karter er beregnet (interpolasjonsrutiner, hvorvidt temperaturer er korrigert for høyde over havet, osv.).

Hydrologisk avdeling i NVE har utarbeidet avrenningskart (rasterkart med 1 f. 1 km oppløsning) som viser årlige middelverdier for avrenning for hele Norge i perioden 1961-1990. Kartet viser hvor stor del av nedbøren som renner ut i vassdragene per kilometerrute. Resten antas å fordampe fra landoverflaten (evapotranspirasjon).

PCA Norge

Det er gjort en analyse av samtlige tilgjengelige regionale GIS data i Norge (54 variabler) som kunne tenkes å ligge til grunn for den regionale variasjonen i vegetasjonssammensetning (biogeoklimatisk variasjon). Tanken bak denne studien (Bakkestuen et al. accepted (conditionally)) var å se på den interne

samvariasjonen mellom variablene, hvilke regionale kompleksgradienter de danner og om gradienter ekstrahert med numeriske metoder sammenfaller med gradientene som ligger til grunn for soneinndelingen i vegetasjonsatlasen (Moen 1998). Analysene er gjort på uttrekk av data for ruter med størrelse 10 f. 10 km, 5 f. 5 km og 1 f. 1 km. Alle variablene ble transformert til null statistisk skjevhet (zero skewness; Økland et al. 2001) og standardisert slik at hver variabel skulle veie likt i analysen. Variabelmatrisen ble analysert med PCA. Resultatene var konsistente over de ulike skalaene (fra 1 f. 1 til 10 f. 10 km) og ga fire tolkbare PCA-akser. Den første akse var direkte sammenlignbar med Moens seksjoner og den andre med Moens soner. En slik objektiv parametrisering av vegetasjonssoner og -seksjoner i form av trinnløse modeller, som jo en PCA akse er, vil gi viktige bidrag som variabler til prediksjonsmodellering.

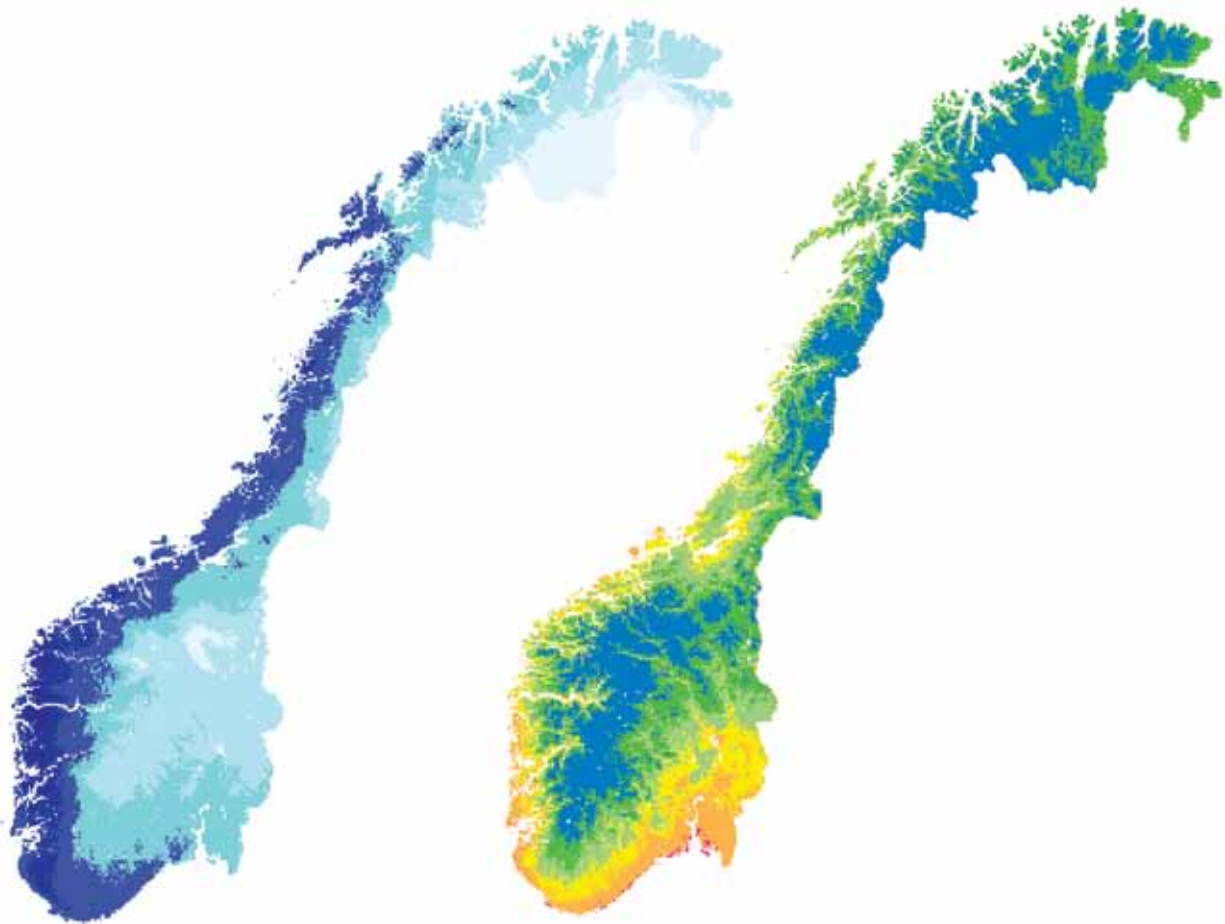
Soleksponering

Terrengmodeller kan, som nevnt over, transformeres til solinnstrålingskart. Det finnes standard prosedyrer for dette hvor man f.eks. kan angi hvilken del av året man ønsker å beregne solinnstråling for. Disse prosedyrene har som standard at man angir et rasterkart (terrengmodellen) som inndata, eventuelt sammen med tidsintervall (fra dato til dato) og tidsintervall i løpet av dagen. Deretter beregnes et nytt rasterkart som angir samlet solinnstråling for dette tidsrommet. De mest avanserte prosedyrene leser breddegraden direkte ut fra inndatafilen (dersom denne er georeferert), mens man i andre prosedyrer må angi midlere breddegrad for kartområdet.

Resultatet av en slik transformering er illustrert i figuren under. Her ser man tydelig hvilke deler av terrenget som er sterkt soleksponert, hvilke som er skygefulle og alle mellomformer. Ettersom topografien resulterer i et så utpreget relieff på den sydlige delen blir kontrastene store her, mens forskjellene er mindre på den nordlige delen. I flatt terreng er forskjellene svært små (og helt fraværende over vannflater).

Topografisk våthetsindeks

Topografien styrer i stor grad variasjoner i hydrologiske forhold som avrenning, jordfuktighet og grunnvannstrømninger. Hydrologisk avdeling i NVE har utarbeidet avrenningskart som viser årlige middelverdier for avrenning for hele Norge i perioden 1961



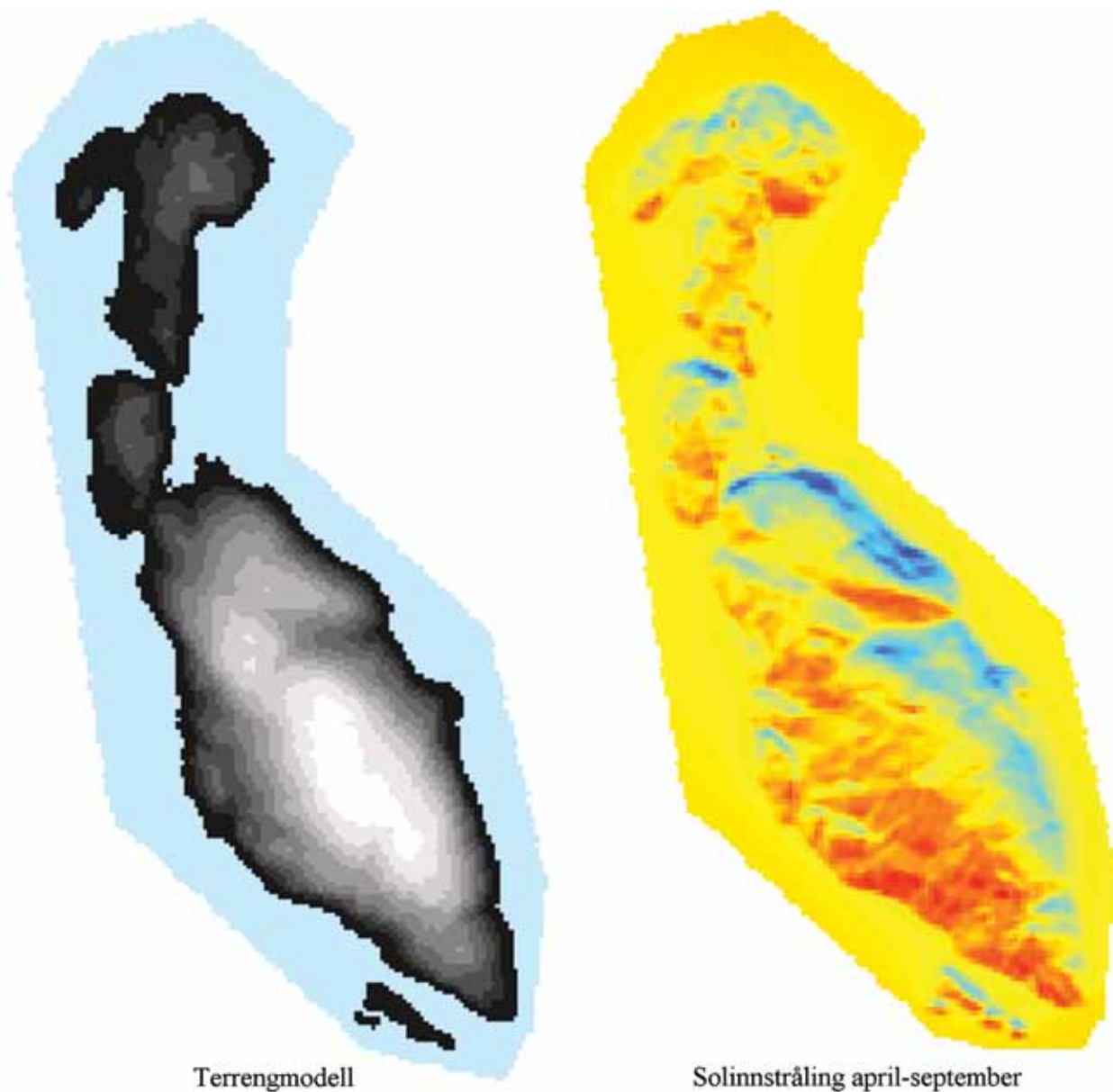
Figur 14. Norgeskart som viser fordelingen av arealer klassifisert i forhold til PCA-akse 1 (til venstre) og PCA-akse 2 (til høyre). Denne klassifiseringen er gjort for å sammenligne PCA analysen med Moens seksjoner og soner.

til 1990. Kartet viser hvor stor del av nedbøren som renner ut i vassdragene i en gitt kilometer rute over hele landet. Resten er forutsatt å fordampe fra land-overflaten (evapotranspirasjon). En annen indeks er topografisk våthetsindeks ("topographic wetness index", TWI) som ble utviklet av Beven og Kirkby (1979). TWI er definert som $\ln(a/\tan\beta)$ hvor a er arealet av nedbørsfeltet som drenerer inn til et definert punkt og β er den lokale helningen i punktet. TWI indeksen har blitt brukt til å studere romlige skalaeffekter av hydrologiske prosesser, identifisere hydrologiske avrenningsmønstre for geokjemisk modellering og for å karakterisere biologiske prosesser som for eksempel

primær produksjon, vegetasjonsgradienter (mønstre) og potensiell skogproduksjon (Sørensen et al. 2005).

Topografisk posisjon

Dette er en indeks som viser hva slags topografisk posisjon de ulike posisjonene i terrenget (pixlene i kartet) har. Hovedprinsippet bak denne indeksen er å beregne i hvor stor grad nabopixlene ligger høyere eller lavere i terrenget. Dermed vil indeksen få fram topper, lisider og forsenkninger. Denne indeksen angir i hvilken grad arealene er eksponert eller beskyttet i forhold til vind, snøavblåsing (vinterstid) og uttørking.



Figur 15. Terrengmodell (25 m oppløsning) for Håøya ved Drøbak i Indre Oslofjord og til høyre avledet solinnstråling for perioden april til september. De røde partiene er sterkest soleksponerte, mens de blå partiene er skyggefulle.

De digitale terrengmodellene kan også benyttes til å avlede andre terrengegenskaper som for eks. skråning og terrengvariasjon. Både topografisk posisjon, skråning og terrengvariasjon gir relevant informasjon i forhold til substrattypen for marine habitater (Bekkby et al. 2002).

Berggrunn

Det eksisterende berggrunnskartet for Norge må tilrettelegges for å kunne anvendes effektivt til prediksjonsmodellering. For det første må det lages rasterkart med en egnet oppløsning. Formodentlig

er 1 f. 1 km mest relevant, men det kan i tillegg være aktuelt med en finere romlig oppløsning for visse deler av landet der mer detaljerte berggrunnskart finnes slik som for Oslofeltet. Videre er det nødvendig å oversette de ulike bergartene til spesifikke kategorier for mineralnæringsstatus, eventuelt også forvittringshastighet. Da vil vi kunne få et godt grunnlag for å modellere forekomst av ulike berggrunsspesialister ("kalkarter", etc.).

Marine miljøfaktorer

Flere studier har dokumentert at dyp (som indikator for lys), terrengstrukturer (f. eks. skråning), fysiske faktorer (f. eks. bølgeeksponering) og andre miljøfaktorer som bunnsubstrat er viktig for utbredelse og dekningsgrad av marine habitater og arter. Når det gjelder algesamfunn spesielt, så er dyp og bølgeeksponering viktig for både utbredelse, dekningsgrad og størrelsen på økosystemene (Bekkby et al. 2002, Bruntse et al. 1999, Crothers 1976, Isæus 2004, Norton et al. 1977, Sjøtun and Fredriksen 1995, Sjøtun et al. 1998, Subrahmanyam 1959), samt for strukturen og arts mangfoldet til assosierte arter (Christie et al. 2007, Christie et al. 2003, Hacker and Steneck 1990, Norderhaug 2004, Norton et al. 1977, Schultze et al. 1990). Strømforhold har også stor innflytelse på utbredelsen til marine arter og naturtyper. I tillegg til å styre utbredelsen av arter og naturtyper påvirker også bølgeeksponering og strøm morfologien til alger i forhold til å minske drag-kreftene og for å maksimere opptaket av lys og næring.

Bølgeeksponering

Som en del av det nasjonale programmet for kartlegging av marint biologisk mangfold, har bølgeeksponering blitt modellert med 25 m oppløsning for hele Norges kyst. Tilsvarende datasett finnes for Møre-kysten med 10 m oppløsning. Beregningene av bølgeeksponeringsverdi er utviklet og beskrevet av (Isæus 2004). Metoden er en "fetchmodell" (maksimal bølgehøyde er avhengig av distansen som vinden har blåst over vannflaten uten å være hindret av land, *fetch*). *Fetch*-verdiene i modellen er justert for å simulere refraksjons- og diffraksjonseffekter rundt holmer og skjær. De beregnede fetch-verdiene for hver av 16 himmelretninger multipliseres med middel vindstyrke til den respektive himmelretningen. Bølgeeksponeringsverdien til hver enkelt rute beregnes deretter

som middelverdien til de 16 himmelretningene. Modellen har blitt revidert til også å inkludere reduksjon av bølgeeksponering med økende dybde (Bekkby et al. submitted).

Marin lysmodellering

Lys er en faktor som også er svært viktig for marine arter og naturtyper. Lyset svekkes nedover i dypet og påvirkes sterkt av partikkelmengden i vannmassene. Lysforholdene på bunnen påvirkes dessuten av plasseringen i forhold til optimal lysinnstråling (relatert til skråning og himmelretning). NIVA har utviklet indekser for dette (Trine Bekkby, upubliserte data) som kan modelleres i ArcView ved bruk av digitale dybdekart og informasjon om siktedyp. Metoden produserer to typer indekser; 1) prosent av overflatelyst som når bunnen, og 2) indeks for bunnens posisjon i forhold til optimal lysinnstrålingsvinkel.

Havbunnssubstrat

Muligheten for å modellere substrattypen basert på dybde data og avledede terrengegenskaper (topografisk posisjon, skråning osv) med ulik oppløsning og kvalitet har blitt testet (Bekkby et al. 2005). Det ble utviklet regelbaserte modeller for å skille løsmasser og hardbunn (fjell) i et utvalgt studieområde ved Møre-kysten. Metoden har blitt videreutviklet til å lage prediksjonskart for sannsynligheten av hardbunn og løsmasser, og ulike sedimentklasser i (upubliserte data, Trine Bekkby, NIVA) for deler av kysten.

Limniske miljøfaktorer

I limniske miljøer er det særlig to miljøfaktorer som er bestemmende for forekomst av arter: vannets bevegelseshastighet (stillestående eller hurtigrennende) og vannkjemi. I tillegg kommer faktorer som sonering i ulike dyp og bunntype (som indirekte bestemmes mye av vannets bevegelseshastighet).

I rennende vann er en rekke arter knyttet til forskjellige strømhastigheter og substrattyper. Dette er vesentlig basert på empirisk kunnskap (dvs. hvor de finnes), og i liten grad på fysiologiske eller eksperimentelle data. Vannets bevegelseshastighet i vassdragene er nesten utelukkende bestemt av topografien og kan således beregnes fra terrengmodellen som vannet strømmer over. I N50 skilles det mellom rennende og stillestående vann. Ved NVE har man laget

et vektorbasert elvenettverk (ELVIS) som følger små-bekker, senterlinje i to-streks elveløp (dvs. vann som flatedata) og innsjøer, se <http://atlas.nve.no>. Dette elvenettverket er videre knyttet sammen med en terrengmodell for å beregne potensialet for småkraftverk. I denne analysen var det fokus på vannføring og fallhøyder, og bevegelseshastighet som sådan ble ikke beregnet. Det må formodentlig gjøres ekstra analyser for å beregne bevegelseshastigheten til vannet langs ulike segmenter av elvenettverket som funksjon av en terrengmodell.

Vannkjemien er i stor grad bestemt av berggrunn og jordvannskjemi i nedbørsfeltet som drenerer vann til ulike deler av vassdraget. En helt spesiell vannkjemi finner vi i kalksjøer som dannes i tjern og innsjøer med avrenning fra kalkholdige bergarter og jordsmonn. Sannsynlighet for forekomst av kransalger kan trolig i stor grad baseres på geologiske kart gjennom sitt krav til kalsium. En del snegler og muslinger har også krav til kalsium.

Fysiske inngrep

For en rekke arter representerer fysiske inngrep og medfølgende tilstandsendringer viktige miljøfaktorer som virker sterkt inn på artens forekomst. Et eksempel er hogst av eldre skog som medfører endrede lys- og fuktighetsforhold, samt sterkt redusert tilgang av død ved siden et velutviklet tresjikt elimineres for mange år framover. Arter kan respondere med økt el-

ler redusert sannsynlighet for forekomst som en følge av slike inngrep avhengig av hva slags skogtilstand de er tilpasset å utnytte.

Fysiske inngrep er ressurskrevende å kartlegge direkte, men det er i prinsippet mulig å gjøre dette basert på satellittbilder. Et annet, mer indirekte, mål på fysiske inngrep er avstand til vei. I Norge er det definert en egen arealklassifisering, "inngrepsfrie naturområder" (INON), som reflekterer avstand til sterk menneskelig påvirkning, se <http://www.dirnat.no/inon>.

Inngrepsfrie naturområder er definert til å være alle områder som ligger mer enn en kilometer i luftlinje fra tyngre tekniske inngrep. Tyngre tekniske inngrep er her definert som veier, jernbaner, kraftlinjer og vannkraftanlegg.

I tillegg til avstanden 1 km benytter INON-konseptet også avstandene 3 og 5 km for å definere soner som i økende grad har preg av villmark.

Det er verdt å merke seg at INON-konseptet kanskje er mest egnet til å inndele naturen sett fra et menneskelig rekreasjonssynspunkt (Skjeggedal et al. 2005). Sett fra et modelleringssynspunkt, kan det være bedre å splitte de ulike typene av tekniske inngrep, da de på forskjellig måte impliserer fysiske inngrep. Avstand til vei er trolig den faktoren som er viktigst, da vei i praksis er eneste transportåre for anleggsmaskiner, osv. Videre kan det være fornuftig å beregne avstander direkte (eventuelt å benytte en finere oppdeling i avstandsklasser enn 1, 3 og 5 km).

6. Eksempler fra norsk prediksjonsmodellering

Prediksjonsmodellering er en ung fagdisiplin internasjonalt, og i Norge er det kun noen få fagmiljøer som har begynt å anvende disse metodene. I dette kapitlet omtaler vi et utvalg av studier for å vise hvilke problemstillinger og arter/ naturtyper metodene er tatt i bruk for.

6.1 Biogeografisk prediksjon av sopp

Herbariedata har hittil vært en undervurdert datakilde for økologisk og biogeografisk forskning, særlig for storsoppene siden disse tilbringer mye av livssyklusen under jordoverflata, og er således ikke lett registrerbare. I en studie har Wollan et al. (in prep) brukt herbariedata for ni selekterte sopparter og korrelert utbredelsesmønstrene mot regionalt tilgjengelig miljøvariable. Artsforekomster ble registrert i 5 x 5 km ruter over hele Norge samtidig som alle rutene også fikk tilegnet verdier fra 75 ulike miljøvariable (prediktorer). Datasettet ble analysert med GLM (logistisk regresjon) og prediktorvariablene kunne forklare mellom 27.9 – 59.8 % av variasjonen i utbredelsesmønstrene. Variabler relatert til temperatur og innstråling ble oftest inkorporert i modellene. Signifikante variable i GLM modellen ble brukt til å lage en prediksjonsmodell for artens totale utbredelse i Norge ved hjelp av Maxent. Prediksjonskartene (figur 16) viser sannsynlige utbredelsesområder. Ulogiske prediksjoner og lav prediksjonsverdi kunne i de fleste tilfellene forklares med utbredelsestype (spredd versus begrenset, kyst versus innland) og/eller økologisk preferanse (for eksempel uegnet substrattype). Konklusjonen er at herbariedata og prediksjonsmodellering er en viktig kilde til informasjon for å forstå arters økologi og utbredelse.

6.2 Naturtyper i marine miljøer

I og med at utbredelsen av arter i sjø i stor grad styres av dyp og eksponering, har man for marine kystområder utviklet en regelbasert modell for prediksjon av EUNIS-klasser basert på disse prediktorene. Gjennom EUs vannrammedirektiv (www.vanndirektivet.no) og det hierarkiske klassifiseringssystemet EUNIS (<http://eunis.eea.eu.int>) utvikles det prinsipper for inndeling av marine områder slik at man lettere kan sammenligne habitater og økologisk status på tvers

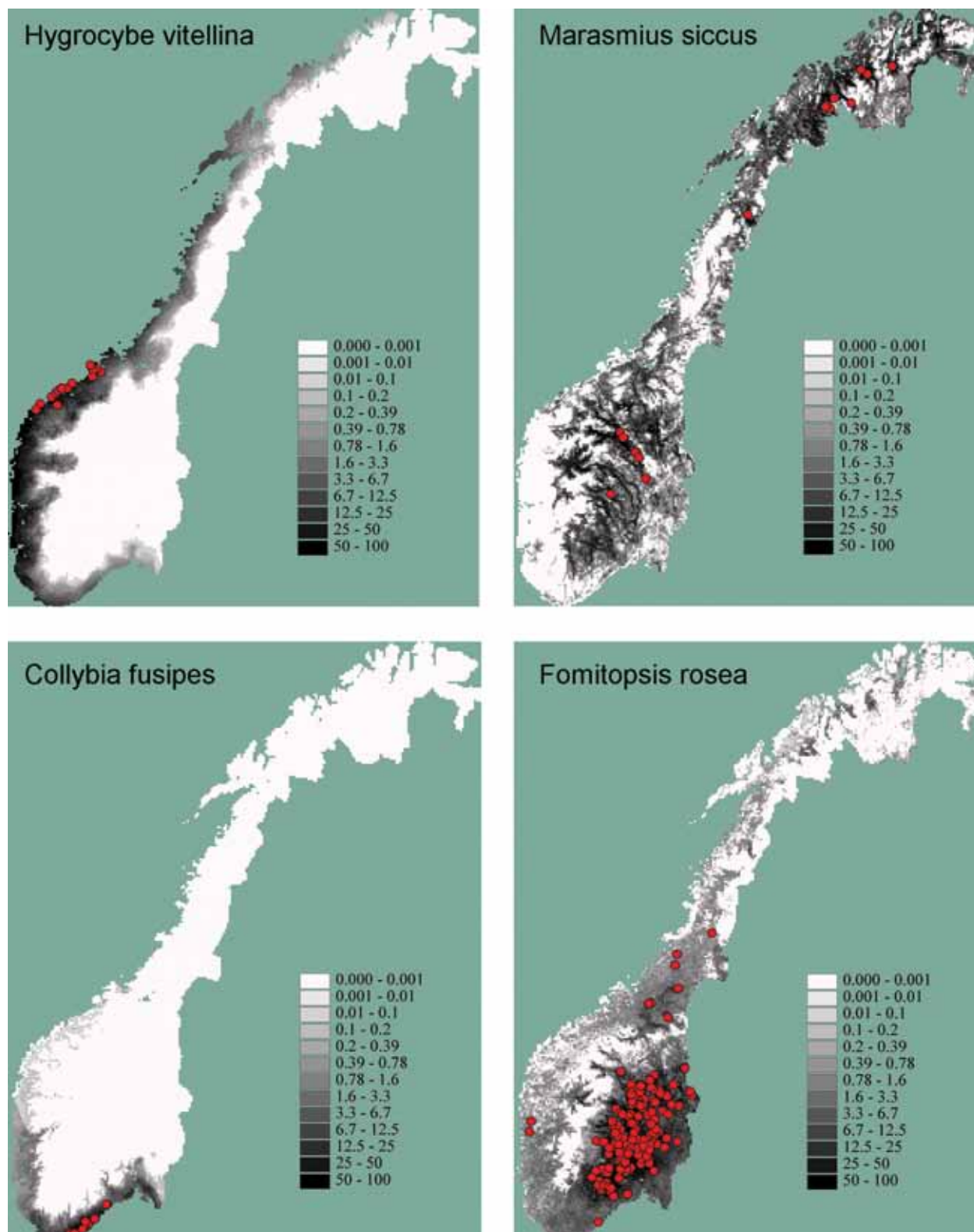
av landegrenser. Stedfesting av EUNIS-klasser på de øverste nivåene (klassifisert etter dybde og eksponeringsgrad) har blitt brukt til å vise geofysiske gradienter og habitatvariabilitet i marint miljø. De predikerte EUNIS-klassekartene har bl.a. dannet grunnlag for å velge ut stasjonsnett for etablering og overvåking av økologisk status mht Vannrammedirektivet.

Kartfestet informasjon om substrattype mangler i stor grad for marine kystområder. For å bøte på denne mangelen er det utviklet prediksjonsmodeller for fordelingen av hardbunn og bløtbunn, samt for sannsynligheten for å finne bløtt sediment egnet for prøvetaking for å kartlegge økologiske status. Som en del av det nasjonale kartleggingsprogrammet har bløtbunnsområder i strandsonen blitt modellert vha regelbasert modellering for deler av norskekysten basert på prediktorene dyp og skråning (Rinde et al. 2006).

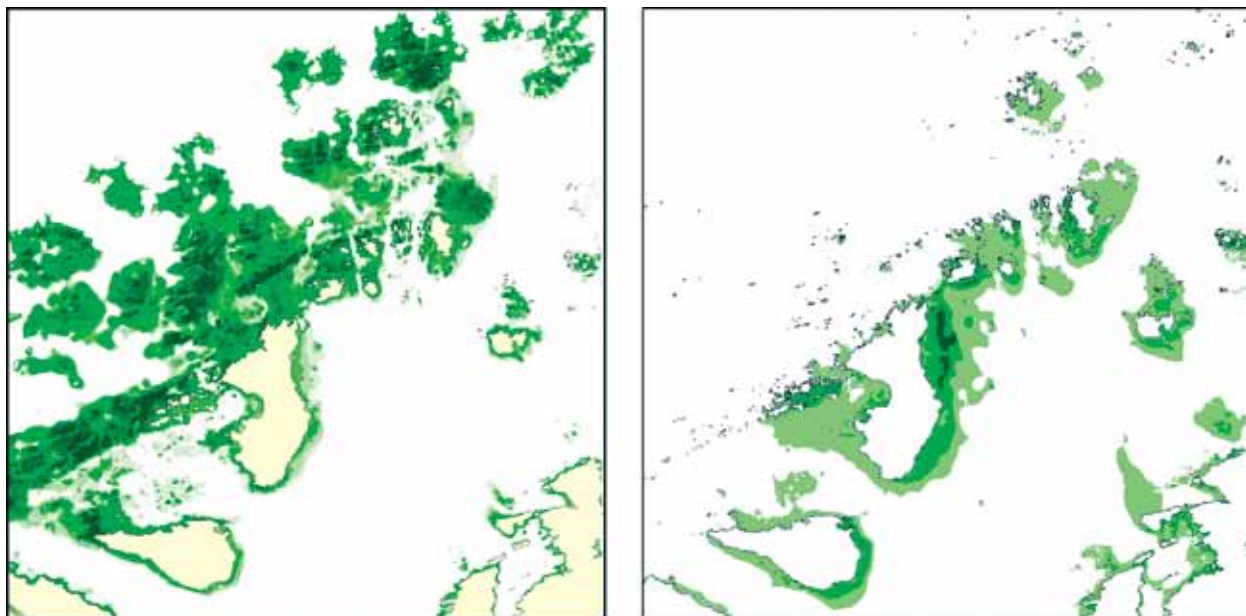
Store og grunne bukter og vikene er et Natura 2000-habitat, dvs et habitat som har en ekstra beskyttelse under EU's Habitatdirektiv. Selv om Norge ikke er forpliktet til EU's Habitatdirektiv, er det ønskelig å følge en klassifisering som er sammenlignbar med det som gjøres i våre naboland. Utbredelsen av naturtypen har derfor blitt modellert vha regelbasert modellering og prediktorene dyp, bølgeeksponering, avstand til land, og størrelse (Bekkby and Isæus *Accepted*).

Prediksjonsmodellering har blitt brukt på et utvalg av arter i marint miljø. Som en del av det nasjonale kartleggingsprogrammet har ålegras og stortareskog blitt modellert vha regelbasert metode for henholdsvis Skagerrak og hele norskekysten vha prediktorene dyp, skråning og bølgeeksponering (Rinde et al. 2006). For deler av kysten har metodene blitt videreutviklet til romlige sannsynlighetsmodeller (Bekkby et al. *Submitted*) basert på analyser av innsamlede feltregistre-ringer. Ulike biodiversitetsindekser har blitt utviklet for bløtbunn. NIVA jobber med å teste om utbredelsesmønstret av disse indeksene kan fanges opp av geofysiske faktorer som dyp, terrengforhold, strøm og bølgeeksponering (upublisert data, pers. med. Trine bekkby).

Prediksjonsmodellering i marint miljø gjøres i hovedsak av forskere ved NIVA, i samarbeid med NINA, UiO, DNMI og AquaBiota Water Research i Sverige. I Norge består kjernen av forskere av i underkant av 10 personer. Arbeidet framover vil bl.a. bestå i å se på mulige klimaeffekter på kråkebollenes nedbeiting av tareskogen ved Nordlandskysten (gjennom prosjekter finansiert av NIVA, Direktoratet for Naturforvaltning



Figur 16. Påviste forekomster (røde symboler) og predikert utbredelse (gradert skala i grått; hvitt er meget usannsynlige forekomster, de mørkeste angir 50-100 % sannsynlighet) for 4 sopparter basert på en romlig oppløsning på 5 km x 5 km. Tre av disse artene (H. vittolina, M. siccus, F. rosea) er rødlistet.



Figur 17. Prediksjonsmodeller for sannsynlig utbredelsen av stortare, *Laminaria hyperborea* (til venstre) og ålegras, *Zostera marina* (til høyre). Jo mørkere grønn, jo større er sannsynligheten for forekomst. Modellene er utviklet for området rundt Sandøy kommune, Møre og Romsdal, med en romlig oppløsning på 10 m. Kilde: Bekkby, upublisert materiale (stortare) og Bekkby, Rinde et al. Subm. (ålegras).

og Norges forskningsråd) og på biologisk mangfold i taeskoen ved ulike miljøforhold (finansiert av Norges forskningsråd fra 2008).

bli brukt i 2008 for å samle data om ca. 30 økologisk viktige arter.

6.3 Ecochange og utbredelse av alpine plantearter

ECOCHANGE (<http://www.ecochange-project.eu/>) er et stort EU-prosjekt som har som hovedmål å gi bedre prediksjoner om endringer i vegetasjon innen de neste 100 år. Prosjektet begynte i 2007 og avsluttes i 2011. Universitetet i Tromsø har medansvar for to relevante problemstillinger: 1) å sammenligne nisjemodeller for plantearter som finnes både i Alpene og i Nord-Norge, slik at man kan vurdere hvor forskjellig nisjen kan være mellom regioner, og 2) å estimere usikkerheten i prediksjonene, spesielt mht. komponenter av prognosemodeller (dvs. klimaprognooser/scenarier, nisjemodeller, osv). Prosjektet kommer til å bruke artslistene ved Museet i Tromsø (i samarbeid med T. Alm) til å utvikle en stratifisert sampling design som vil

6.4 Trær og vedlevende arter

Vedlevende arter (ca. 6000 arter i Norge, spesielt sopp og insekter) utgjør en stor andel av det biologiske mangfoldet i skog. Videre utgjør denne funksjonelle gruppen litt mer enn 50 % av alle rødlistede arter i skog i Sverige (Dahlberg and Stokland 2004) og det er rimelig å anta at andelen er omtrent den samme i Norge. De aller fleste vedlevende artene er økologiske spesialister som bl.a. er spesialisert i forhold til å utnytte død ved fra forskjellige treslag.

På denne bakgrunn ble det i 2007 initiert et 3-årig prosjekt ved Naturhistorisk Museum, Universitetet i Oslo for utvikling av prediksjonsmodeller for forskjellige treslag og et utvalg vedlevende arter på ulike romlige skalaer (romlig oppløsning fra 25 m til 1 km; romlig utstrekning fra 10-100 km² i utvalgte testområder til hele Norge). Tilstede-fravær data fra Landskogs-

takseringen (ca. 11 000 sampelpunkter) benyttes til å modellere treslagenes forekomst. Videre benyttes et randomisert stratifisert datasett med høy stedfestingspresisjon og tilstede-fravær artsdata (160 lokaliteter), og artsdata fra GBIF/Norge (kun tilstede data) til å modellere artsforekomster. Prosjektet legger ned en betydelig innsats i å etablere flere heldekkende rasterkart for prediktorvariabler med detaljert romlig oppløsning (25 m pixelstørrelse) avledet fra DMK og terrengmodell.

Prosjektet vil benytte et utvalg av modelleringsmetoder som har vist seg å ha god til svært god prediksjonsevne (GLM, GAM, Maxent, Regresjons trær). De innledende analysene viser at modellene for treslag har god til svært god prediksjonsevne (AUC-verdier fra 0,76 til 0,97 for GLM-modellene og 0,60 til 0,95 for Maxent-modellene, tabell 4) basert på økologiske data fra Landskogstakseringen. En nærmere inspeksjon av modellene viser at både klimarelaterte prediktorer med grov (1 km) oppløsning og lokale ter-

Tabell 4. Ulike miljøfaktors viktighet i prediksjonsmodeller for ulike treslag basert på data fra Landsskogstakseringen. Kolonnene representerer prediktorvariabler med 1000 m romlig oppløsning (PCA-verdier og klimadata) og 25 m romlig oppløsning (bonitet og solinnstråling i forhold til terrengposisjon). Stjerner indikerer p-verdier for de ulike prediktorene i GLM-modellene (* for $p < 0,05$; ** for $p < 0,01$; *** for $p < 0,001$) og grønnfarge indikerer prediktorenes relative bidrag til AUC-verdien i Maxent-modellene (jo sterkere grønnfarge, dess viktigere i modellen). AUC-verdier angir de enkelte modellenes prediksjonsstyrke. Rogn, selje og hassel var ikke analysert med Maxent da rapporten ble skrevet.

	1000 m							25 m		AUC	
	PCA1	PCA2	Hoh	Nedbør Juni	Nedbør Januar	Temp. Juni	Temp. mini- mum	Boni- tet	Sol- innstr.	GLM	MaxEnt
Furu	***	*	***	*	**	**	***	***	**	0,77	0,66
Gran	***	***	***	***	***	***	**	***	***	0,80	0,60
Osp	**	*	***	***		*	***		***	0,80	0,77
Rogn	***	***	**	**			*	***	***	0,77	
Selje		***	***	*	*		**	***	***	0,76	
Gråor	***	*	***	*	**	***	*	***	*	0,85	0,80
Hegg	***	***	***				**	***		0,88	0,86
Svartor			***							0,86	0,87
Lønn	***	**	**					***	***	0,89	0,94
Hassel	***	***	***	***	*			***	***	0,89	
Eik	***	**	***	**	**			**	***	0,97	0,95
Ask		***				***	***	***	*	0,93	0,93
Lind	**	*	***	***			*		**	0,90	0,92

reng- og jordsmonnsrelaterte prediktorer med fin (ca. 25 m) oppløsning er viktige faktorer som forklarer treslagenes forekomst og fravær. Det er spesielt lovende at lokale faktorer som solinnstråling og produksjonsevne (bonitet) er viktige faktorer ettersom disse er representert i kartene med prediktordata omtalt over. I løpet av vinteren 2007/2008 vil prediksjonsmodeller bli laget for Østlandet, og da vil det bli mulig å lage finskala (ned til 25 m oppløsning) prediksjonskart for forekomst av ulike treslag og assosierte vedlevende arter.

Det er planlagt å gjøre feltvalidering i utvalgte testområder på Østlandet (2008 og 2009), eventuelt også i andre deler av landet (2009). I tillegg til å modellere de fleste treslagene vil prosjektet utvikle prediksjonsmodeller for et utvalg av vedlevende arter knyttet til furu, gran, eik og osp, eventuelt også for arter knyttet til andre treslag.

6.5 Billefauna i hule eiker

For kartlegging, overvåking og forvaltning av det biologiske mangfoldet er det nyttig å kjenne til hvilke områder som huser mange sjeldne arter (hotspots). En stor andel av artene på den norske Rødlista er insekter (45 %), og mange av disse er billearter knyttet til eik (Kålås et al. 2006, Sverdrup-Thygeson et al. 2007). Det betyr at den boreo-nemorale sonen i Sørøst-Norge er en viktig "hotspot" region. For alle praktiske formål (kartlegging, overvåking, forvaltning) er imidlertid en slik avgrensning for grov. Vi trenger å vite hvilke habitater og lokaliteter innenfor regionen som er viktigst. I denne studien brukes statistiske modeller for å predikere forekomst av antall eiketilknyttede rødlistearter på grunnlag av miljøvariabler. Modellene er utviklet med data fra 5 hule eiketrær på hver av 11 lokaliteter kartlagt med finansiering hovedsaklig fra Nasjonalt program for kartlegging og overvåking av biologisk mangfold (Sverdrup-Thygeson et al. 2007, Ødegaard et al. 2006).

Ved hvert tre ble en vindusfelle plassert foran hulromsåpningen og i en av greinene. En rekke miljøvariabler knyttet til enkelttrær (diameter, hulhetsstadium, vitalitet, mm.) og lokaliteter (skog/kulturlandskap, gjengroing, forekomst av eik, lauvtrær, bartrær, mm.) ble registrert. Fordi antallet potensielle prediktorer er stort, og mange av prediktorene er korrelerte, ble det brukt en regresjonsmetode som tar hensyn til

dette: PLS (Partial Least Squares regression – se 2.3). Responsvariabelen er antall av rødlistede biller som er kjent for å ha en økologisk tilknytning til eik.

Et viktig resultat er at skog og kulturlandskap krever forskjellige modeller. Med andre ord er det forskjellige faktorer som styrer forekomsten av rødlistede arter i eik i de to landskapstypene. Skogmodellene er gjennomgående mer komplekse og har dårligere prediksjonsevne enn kulturlandskapsmodellene. Et annet resultat er at trevariabler (f.eks. hulhet og omkrets) er viktige prediktorer for antall rødlistede arter i hulromsfellene, men mindre viktige for greinfellene. Greinfellene er vanskeligere å predikere, trolig fordi de fanger et bredere spekter av arter fra trærnes omgivelser. For prediksjon av antall rødlistede arter i greinfeller, samt antall rødlistede arter totalt på tre- og lokalitetsnivå, ser geografisk posisjon (nord og/eller vest) ut til å ha en viss prediksjonsverdi. God geografisk prediksjon vil imidlertid kreve stedfestet informasjon om landskap (skog, kulturlandskap), forekomst av eik og andre treslag, samt egenskaper ved eiketrær. Noe av informasjonen som trengs er per dags dato ikke tilgjengelig, men vil kunne fremskaffes ved kartlegging og/eller modellering (se kapittel 5). Dette vil fokuseres videre i studiet.

6.6 Karplantemangfold i kulturlandskap

I denne rapporten er nisjemodellering avgrenset til å omfatte modellering av enkeltarters forekomster og mengde, samt artsrikdom på samfunnsnivå. Med artsrikdom menes antall arter registrert innenfor et areal av en gitt størrelse.

Prosjektet 3Q-km (NFR-prosjekt 2001–04) var et samarbeidsprosjekt mellom Naturhistorisk museum (NHM), UiO og Skog og Landskap, som tok utgangspunkt i overvåkingsprogrammet 3Q (Tilstandsrapportering og resultatkontroll i jordbrukets kulturlandskap). I 2001 og 2002 ble 16 overvåkingsflater á 1 km² kartlagt med hensyn til landskapselementer ("flekke" eller "patcher") som ble klassifisert til 18 typer. I hvert landskapselement ble alle karplantearter og en rekke forklaringsvariabler registrert (se Bratli et al. 2006 for mer detaljert informasjon). Til sammen ble 2201 landskapselementer avgrenset innen de 16 overvåkingsflatene. Økland et al. (2006) modellerte artsrikdom på

fem romlige skalanivåer i kulturlandskapet ved hjelp av variasjonskomponentanalyse.

Variasjonen i karplanteartsrikdom i observasjons-enhetene viste at svært mye av variasjonen var forklart av de ca. 30 forklaringsvariablene. På landskaps-element-nivå (2201 flekker) ble 75 % av variasjonen forklart. Flekkgeometrien, og delvis også økologiske faktorer, var de viktigste prediktorene. Også ved sammenslåing av landskapselementer i grupper kunne en høy andel av artsrikdomsvariasjonen predikeres (81–87 %). Den høyere forklaringsprosenten skyldes først og fremst det lave antallet observasjoner i forhold til antall forklaringsvariabler. Med 1 km² flater som observasjonsenhet var det bare flekkgeometrivariabelen areal (andel av flaten som var dekket av kulturlandskapsnaturtyper; dvs. ikke-kulturlandsskapselementer som bebyggelse osv. var fjernet) som forklarte en signifikant andel av variasjonen. Relasjonen mellom artsrikdom og areal var negativ, dvs. omvendt av hva

vi hadde forventet. Vi konkluderte med at denne relasjonen må ha vært et resultat av datasæregenheter, slik at det egentlig ikke fantes noen tolkbar sammenheng mellom artsrikdom og prediktorvariablene på overvåkingsflatenivå.

Undersøkelsen konkluderte dermed med at det er mulig å predikere artsrikdom på flekk-skala i kulturlandskapet ut fra flekkstrukturegenskaper som areal og form og andre egenskaper som type og miljøforhold. Dette stemmer overens med skalaen der romlige basisøkokliner er antatt å være viktige for artenes fordeling (Fig. 11). En grundig analyse av betydningen av areal og form for artsrikdom i moderne kulturlandskap basert på det samme materialet er utført av Heegaard *et al.* (2007). Mangelen på forklarbarhet på overvåkingsflatenivå stemmer også overens med Fig. 11, ettersom alle 16 flater ligger i boreonemoral eller sørboreal sone på Sørøstlandet. Dette tilsier at den regionale variasjonen mellom dem er liten.

7. Praktisk bruk i Norge

I kapittel 2.2 har vi vist eksempler på anvendelse av prediksjonsmodellering i andre deler av verden og i kapittel 6 finner vi eksempler på prediksjonsmodellering fra Norge. I dette kapitlet vil vi adressere noen konkrete utfordringer innen norsk forvaltning og rødlistearbeid, og redegjøre for hvordan prediksjonsmodellering kan bidra på en verdifull måte i dette arbeidet.

Vi vil gjenta fra kapittel 2.5 at påliteligheten til prediksjonsmodeller og prediksjonskart er nært knyttet til kvaliteten på det empiriske datagrunnlaget. Derfor er det viktig å huske at alle prediksjonsmodeller og avledete kart bør underkastes minst én metodisk feltvalidering og oppdatering før de benyttes til praktisk anvendelse.

7.1 Kartlegging av rødlistede arter

Artsutbredelse

Prediksjonsmodellering er velegnet til å lage prediksjonskart med relativt grov (1–10 km) romlig oppløsning, for hele Norge. For dette formålet har en stor andel av de digitaliserte artsfunnene tilstrekkelig god stedfestingspresisjon (1 km eller bedre). Videre vil ulike klimavariabler og de avledete PCA-scorene (se kap. 5.2) være velegnet som prediktorer. For å få til relativt pålitelige prediksjonskart, bør antall stedfestede forekomster være minst 20–30. Det er *mulig* å lage utbredelseskart basert på færre observasjoner, men slike kart bør kun brukes til å styre supplerende kartlegging. Eksempler på noen slike utbredelseskart er vist i kapittel 6.1, hvorav 3 er for rødlistede arter.

Slike utbredelseskarter kan brukes til å beregne potensielt utbredelsesareal for rødlistede arter – et kriterium som benyttes til å vurdere rødlistestatus. For velkjente og godt kartlagte arter (for eksempel fugler) kan det være bedre å bruke kartleggingsdata direkte, men for svært mange andre og mindre kjente arter kan slike prediksjonskart være svært nyttige.

Lokale artsforekomster

Prediksjonsmodellering kan også brukes til å lage prediksjonskart med fin (25–100 m) romlig oppløsning for hele eller deler av artens predikerte utbredelse (romlig utstrekning). For dette formålet er det nødvendig å ha kart med miljødata som har en romlig oppløsning

på 25–100 m. Videre er man avhengig av artsfunn med en stedfestingspresisjon på ca. 10 m.

Slike detaljerte prediksjonskart kan brukes til å målrette og effektivisere kartlegging av rødlistede arter, jfr. Eksemplet fra Sveits som ble utdypet i kapittel 2.2.

En annen anvendelse av finskala prediksjonskart er å anslå det faktiske arealet som ulike rødlistede arter potensielt kan forekomme på (med ulike grader av sannsynlighet). Dermed kan man få et inntrykk av i hvilken grad forskjellige arter er underdokumentert i forhold til sin reelle forekomst. Prediksjonsmodellering synes å være den eneste kjente metoden for å estimere reell forekomst/utbredelse. Det må imidlertid understrekes at slike beregninger kun bør gjøres på grunnlag av prediksjonsmodeller som har vært testet i felt.

7.2 Kartlegging av naturtyper

Ny norsk naturtypeinndeling (NNN)

Norge har en lang tradisjon for å beskrive og kartlegge vegetasjon, basert på ulike kartleggingsnøkler (f. eks. Fremstad 1997, Rekdal and Larsson 2005). På slutten av 1990-tallet startet en storstilt naturtypekartlegging i kommunal regi, med formål å kartfeste forekomsten av et utvalg naturtyper som ble ansett som viktige for det biologiske mangfoldet (Anonym 2001, Anonym 2007). For perioden 2007–10 foregår fullføring av den kommunale kartleggingen i regi av Nasjonalt program for kartlegging og overvåking av biologisk mangfold.

Til tross for en omfattende kartleggingsinnsats i Norge gjennom mange år, finnes ingen rimelig fullstendig nasjonal oversikt (arealstatistikk, kartdatabase) over naturtypers forekomst. Dette skyldes først og fremst at en liten andel av landet faktisk er kartlagt. Litt over 10 % av Norges landareal er vegetasjonskartlagt (Rekdal og Larsson 2005), og i den kommunale kartleggingen av verdifulle naturtyper er det typisk at kartlagte forekomster utgjør mellom 2 og 5 % av kommunens areal. Videre har ikke de ulike kartleggingsnøklerne blitt forsøkt oversatt til et felles naturtypesystem; en forutsetning for at resultater fra kartlegging med ulike nøkler skal kunne presenteres sammen. Oversikt over forekomst og utbredelse av marine naturtyper er i vesentlig grad dårligere stilt

enn for landområdene. Planlegging og utvikling av metoder for en systematisk kartlegging av de antatt viktig-

ste marine naturtypene startet først opp i 2001.

I 2006 initierte Artsdatabanken arbeid med en ny standard naturtypeinndeling for områder under norsk suverenitet, som har foreløpig tittel Ny norsk naturtypeinndeling (NNN). NNN-systemet skal etter planen foreligge i versjon 1.0 ved utgangen av 2008. Systemet vil inneholde en fullstendig arealdekkende inndeling basert på kriterier relatert til fem organisasjonsnivåer for biologisk mangfold (naturtype-nivåer): livsmedium (substrat; organismers umiddelbare levested), økosystem (svarer til habitat i EUNIS systemet), landskapsdel, landskap og region. Et av hovedformålene med NNN er å lage en fullstendig arealdekkende naturtypeinndeling som kan tjene som fellesnøkkel for kartfestet arealinformasjon på land, i ferskvann og i marint miljø. Når en slik fellesnøkkel foreligger, vil det være mulig å samle digital kartfestet arealinformasjon i en og samme database; et første skritt mot en nasjonal oversikt over naturtypers forekomst og utbredelse.

NNN forventes også å bedre kunne målrette innsatsen for kartlegging av naturtyper i Norge, herunder naturtyper der rødlistede arter forekommer hyppig. Det er to trinn i dette arbeidet: 1) identifisering av naturtyper med stor konsentrasjon av rødlistede arter (bedre kobling av forekomster av rødlistede arter mot naturtyper vil være en del av arbeidet med Rødlistearter fram til 2010), og 2) spissing av naturtypekartleggingen mot naturtyper med høy forekomst av rødlistede arter. Prediksjonsmodellering kan spille en viktig rolle i trinn 2. I kapittel 2.1 beskrev vi hvordan arters utbredelse og lokale forekomster kan predikeres basert på kjente forekomster og heldekkende kart over ulike miljøfaktorer. På nøyaktig samme måte kan man bruke kjente forekomster av naturtyper og bruke de samme kartene over miljøfaktorer for å predikere naturtypers totale utbredelse. Akkurat som for arter vil det være nødvendig å foreta feltvalideringer for å påvise ukjente forekomster og for å forbedre prediksjonsmodellene. Denne framgangsmåten burde i løpet av 3-4 modellversjoner (jfr. figur 4) kunne gi pålitelige nasjonale arealestimater for de modellerte naturtypene og tilhørende prediksjonskart som viser hvor en stor andel av de potensielle forekomstene er.

Marine EUNIS-habitatklasser

I det nasjonale programmet er det i tillegg til modeller for spesifikke naturtyper som tareskog og ålegras, også utviklet GIS-modeller for å identifisere over-

ordnete habitatklasser i henhold til det europeiske habitatklassifiseringssystemet EUNIS. Identifisering av utbredelsen til disse habitatklassene er gjort for å lette planleggingen av kartlegging av ålegras også i de nordlige økoregionene der innflytelsen av tidevannsstrømmer og mangelen på denne type data, har gjort det umulig å utvikle gode ålegrasmodeller. I tillegg er disse habitatklasse-kartene utviklet og benyttet som et verktøy for å velge ut stasjoner og områder som referansestasjoner for å klassifisere og overvåke økologisk status tilknyttet arbeidet med Vannrammedirektivet.

7.3 Biomangfoldkart og naturindekser

Miljøverndepartementet og Direktoratet for naturforvaltning har bedt de biologiske fagmiljøene om å utvikle en naturindeks for Norge. Hensikten med indeksen er å beskrive utviklingen i norsk natur som helhet, dvs et slags "bruttonasjonalprodukt" for natur. Det er også ønskelig at indeksen sier noe om utviklingen for ulike naturelementer og ulike geografiske områder. En viktig bakgrunn for indeksen er behovet for å måle utviklingen av biologisk mangfold, med tanke på det politiske målet om å stanse tapet av biologisk mangfold innen 2010. Rammeverket for indeksen er ennå ikke ferdigstilt (S. Nybø et al., under utarbeiding), men kommer etter all sannsynlighet til å basere seg på arealer og tilstand av naturtyper, hvor variabler som relaterer seg til arter (artsrikdom og abundans) blir en av flere måter å måle tilstand på. Det ideelle grunnlaget for å beregne indeksen vil være landsrepresentative overvåkingsdata for utvalgte hovednaturtyper, og et så stort som mulig utvalg av arter innen et bredt utvalg av naturtyper.

Det er håp om at den norske overvåkingen av biologisk mangfold vil forbedres på sikt, men per i dag er den på ingen måte helhetlig og derfor heller ikke tilfredsstillende som grunnlag for en naturindeks. I fravær av gode overvåkingsdata kan prediksjonsmodellering spille en viktig rolle. Det gjelder både prediksjon av geografisk utbredelse og mengde av arter, men også prediksjoner av artsrikdom (for eksempel "hotspots"). For at slike modellprediksjoner skal kunne brukes som "data" i en naturindeks, må det brukes godt dokumenterte og etterprøvbare metoder som lar seg oppdatere med nye grunnlagsdata, og som gjør det mulig å estimere usikkerheten ved prediksjonene.

De fleste av metodene som er drøftet i denne rapporten fyller disse kriteriene.

7.4 Praktisk forvaltning

Resultater fra prediksjonsmodellering kan også anvendes i lokal arealplanlegging og forvaltning. I kapittel 2.2 omtaler vi noen slike anvendelser i forhold til konkrete forvaltningsplaner fra utlandet. Her skal vi kort peke på noen potensielt viktige anvendelser i Norge.

Prediksjonsmodeller er allerede benyttet i flere studier for å vurdere hvorvidt verneområder fanger opp artsmangfoldet på en god måte i dag og i forhold til framtidige klimascenarier, se 2.2. Spesielt vurderinger av framtidig bevaringspotensial har vist at det kan være billigere og mer effektivt å justere reservatgrenser relativt raskt sammenlignet med å utsette slike tiltak.

En annen anvendelse er å benytte predikert forekomst av sårbare arter eller naturtyper som grunnlag for å kreve konsekvensutredninger eller stoppe tiltak. To eksempler fra andre land kan illustrere slik bruk. I Sør-Afrika benyttes predikert forekomst av nesehorn og andre truede pattedyr som grunnlag for å velge mellom (eventuelt skrinlegge) alternative tiltak som veitraseer og turistanlegg. I Mexico har prediksjonkart for naturlig forekommende bomullsarter blitt brukt av myndighetene til å stoppe etablering av bomullsplantasjer som benytter foredlet bomull, innen kjerneområder for naturlig bomull med høy naturlig genetisk variasjon. Tilsvarende kan man i Norge lage prediksjonkart for spesielle arter og naturtyper som viser hvor man bør være spesielt aktsom og foreta nærmere kartlegging dersom tiltak planlegges som kan ha uønsket negativ effekt.

En annen type tiltak er restaurering av naturmiljøer og habitater for arter, herunder utvikling av spesifikke forvaltningsplaner. Dette er særlig aktuelt for områder hvor tidligere arealbruk har hatt andre prioriteringer enn å bevare biologisk mangfold. Det kan ofte være vanskelig å vurdere hvilke konkrete områder som er best egnet med hensyn til restaurering og tiltak. Her kan prediksjonsmodeller basert på artsforekomster i mindre påvirkede områder benyttes til å vurdere kostnadseffektiviteten av ulike alternativer.

Re-introduksjon av utdødde eller sårbare arter er lite brukt som forvaltningstiltak i Norge. Et unntak er

imidlertid gjeninnføringen av moskusfe på Dovrefjell, en art som er kjent fra Norge i siste mellomistid. Videre er edelkreps re-introdusert etter krepsepest i Østfold og ørret i en del sør-norske vann etter fiskedød pga. sur nedbør. Avlsprogrammer for hubro, fjellrev og spesifikke stammer av laks er eksempler på aktiv styrking av sårbare arter gjennom utsetting av individer. Man kan også tenke seg slike tiltak for invertebrater, planter, sopp og lav der man kjenner til tidligere forekomster, men hvor lokalitetene er helt ødelagt som habitat eller der stedfetingen er for upresis til å vite nøyaktig hvor det er aktuelt å sette ut individer. I slike tilfeller kan prediksjonsmodellering hjelpe til med å lokalisere de stedene hvor potensialet for vellykket gjeninnføring er høyest.

For marine områder er det stor mangel på kunnskap om utbredelsen til viktige arter og naturtyper. Kunnskap om utbredelsen til enkeltarter har i stor grad vært knyttet til forskning og overvåking av de kommersielt viktige fiskeartene. For å kunne ta hensyn til viktige naturverdier ved for eksempel utbygging og næringsvirksomhet i kystsonen er det nødvendig å få kartlagt og kartfestet informasjon om hvor de viktige naturverdiene forekommer. Her har prediksjonsmodellering vist seg å være svært nyttig både som verktøy for å planlegge kartlegging av marine arter og naturtyper, men også for å identifisere egnede områder for overvåkingsformål. Tilsvarende som nevnt over for land, vil prediksjonsmodellering kunne bli benyttet til å planlegge og å vurdere utforming og størrelse av verneområder i forhold til framtidige klimascenarier og for vurdering av behovet for konsekvensutredninger i områder med sårbare arter eller naturtyper. Andre relevante problemstillinger er restaurering av for eksempel nedbeitede tareskogsområder og prediksjon av spredning av introduserte arter.

7.5 Vurdere klimaeffekter

Forskning omkring klimaendringer har vært en hovedsatsing i Norge i en årrekke. Denne forskningen har blant annet resultert i flere klimaprognoser for norske landområder mht. framtidige endringer i temperatur, vind, nedbør og snøforhold (jfr. 5.1, <http://www.senorge.no/>). Derimot er kunnskapen om hvilke effekter dette kan få for det biologiske mangfoldet lite utforsket her i landet. Det er gjort en studie av

effekter på europeiske planters utbredelse i forhold til ulike klimascenarier som også omfatter Norge, se 2.2, "Evaluere klimaeffekter på arters utbredelse" (Thuiller et al. 2005). Denne studien benyttet nettopp prediksjonsmodellering som hovedmetode. Det vil være naturlig å følge opp effektforskningen med de mer nyanserte klimaprognosene som nå er utviklet for Norge.

Temaer som opplagt bør studeres nærmere er potensielle endringer som følge av klimaendringer i den alpine floraen (jfr. EcoChange-prosjektet omtalt i 6.3), i skogens utbredelse og sammensetning – både endringer i tregrensen og utbredelse av edelløvskog, og tilsvarende viktige endringer i marine miljøer (jfr. 6.2 som omtaler initiering av slik forskning), osv. I alle disse sammenhengene vil prediksjonsmodellering stå sentralt fordi metodene kan anvende kart for framtidige klimaprognoser som prediksjonsgrunnlag isteden for dagens klimakart. Dermed får man et kvalifisert grunnlag for å si noe om potensielt omfang og størrelsesorden på de endringer man kan forvente.

7.6 Overvåking

Mandatet for den foreliggende utredningen er å vurdere prediksjonsmodelleringens rolle i forhold til et helhetlig program for kartlegging og overvåking av biologisk mangfold i Norge. I andre deler av rapporten har vi vist at slik modellering kan gi viktig supplerende kunnskap om status og utviklingsprognoser for en rekke arter. Her skal vi omtale noen potensielle anvendelser av prediksjonsmodellering med direkte relevans for overvåking.

Internasjonalt er det en økende tendens til å dele overvåking inn i to typer: *surveillance* og *monitoring* (Nichols and Williams 2006, Yoccoz et al. 2001). Den første av disse typene (*surveillance*) omfatter bredt anlagt overvåking som sikter mot å følge utviklingen i et forholdsvis vidt spekter av arter og naturtyper. Den andre typen (*monitoring*) omfatter mer fokusert overvåking som er knyttet opp mot et spesifikt problemområde eller en spesifikk forvaltningsutfordring.

7.6.1 Bredt anlagt overvåking

Norsk forvaltning opererer med tre hovedstrategier og sju delstrategier for kartlegging og overvåking (K&O), jf. Nasjonal plan for kartlegging og overvåking

av biologisk mangfold, se DN (1998), Fremstad og Kålås (2001) og (Økland manuscript):

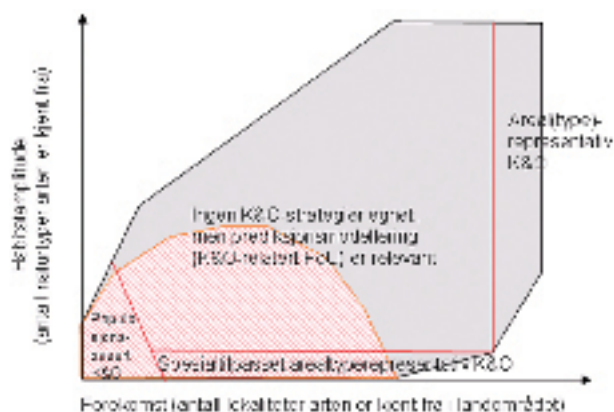
- *ekstensiv* K&O: arealstatistikk for et større landområde
 - *arealdekkende* K&O: arealstatistikk basert på data for hele det aktuelle landområdet
 - *arealrepresentativ* K&O: arealstatistikk basert på et utvalg observasjoner som er representativt for hele det aktuelle landområdet
 - *arealtyperepresentativ* K&O: arealstatistikk basert på et utvalg observasjoner som er representativt for spesifiserte naturtyper innenfor det aktuelle landområdet
- *intensiv* K&O: registreringer av arter og naturtyper i utvalgte observasjonsområder som er typiske for eller inneholder gitte naturtyper
 - *områdebaset* K&O: registreringer med observasjonsområder som observasjonsenhet
 - *arealfigurbaset* K&O: registreringer med arealfigurer (av gitte naturtyper) innen observasjonsområder som observasjonsenhet
 - *fastruteovervåking*: registreringer i fastmerkete observasjonsenheter ('ruter') i gitte naturtyper innen referanseområder
- *spesialobjekt*-K&O
 - *populasjonsbasert* K&O: oppdatert populasjonsstatus for arter med høy gjenoppdagbarhet basert på forekomstinformasjon fra gitte arter på kjente lokaliteter

I denne sammenhengen benyttes gjerne følgende begrepsapparat for kartlegging og overvåking.

Kartlegging defineres som registrering av indikatorvariabler med standardiserte og veldokumenterte metoder, det vil si registreringer utført på en slik måte at helt tilsvarende registreringer kan utføres av andre observatører på et seinere tidspunkt. Gjentatt kartlegging resulterer i en **tidsserie** for indikatorvariabelen.

Overvåking er systematisk gjentatt kartlegging. All overvåking starter med førstegangs kartlegging. Prediksjonsmodellering kan bidra på flere måter sammen med en slik overvåkingsstrategi og således bli et viktig redskap for indirekte informasjon om rødlistearters populasjons- og statusendring (se Fig. 18).

I forhold til ekstensiv, arealtyperepresentativ overvåking kan man lage prediksjonsmodeller for naturtyper som er økologisk distinkte og har lite totalareal (og derfor krever en spesialtilpasset utvalgsplan). Dette kan for eksempel være naturtyper/naturforhold



Figur 18. Sammenheng mellom artsegenskaper (arter med persistente populasjoner) og ulike kartlegging og overvåking (K&O)-strategier sin relevans for vurdering av populasjonsutvikling og utdøingsrisiko. Det grå feltet viser realistiske kombinasjoner av habitatamplitude og forekomst, det skraverte feltet viser kombinasjoner av habitatamplitude og forekomst som er typiske for rødlistede arter. Figuren viser at prediksjonsmodellering vil være eneste realistiske kilde til populasjonsinformasjon for mange arter.

som bekkekløfter, kalkområder, edelløvskogområder og boreale regnskogsområder.

I forhold til intensiv overvåking kan man lage prediksjonsmodeller for utvalgte naturtyper for å lokalisere ytterligere kandidatområder som kan inngå i utvalgsgrunnlaget. Dette er særlig aktuelt for naturtyper som er betinget av spesielle økologiske forhold. Derneest kan man velge ut aktuelle områder for artsovervåking (utvalgte metapopulasjoner). Prediksjonsmodeller er allerede anvendt på en slik måte ved å benytte prediksjonskart for å velge ut marine stasjoner og områder som referansestasjoner for å klassifisere og overvåke økologisk status tilknyttet arbeidet med Vannrammedirektivet (se 7.2).

I forhold til spesialobjekt-overvåking (typisk av arter som både har snever habitatamplitude og få kjente forekomster) kan prediksjonsmodellering bidra til å finne nye potensielle lokaliteter. Videre kan pre-

diksjonskartene benyttes til å anslå totalomfanget av potensielle lokaliteter for slike arter.

7.6.2 Fokusert overvåking

Med fokusert overvåking menes overvåking som er nært koblet til konkret forvaltning eller forskning, f. eks. med bevaringsbiologisk relevans. På denne måten har overvåkingen som funksjon å estimere systemtilstander som er premissgivende for beslutninger i den videre forvaltningen, såkalt "adaptive management" (Nichols and Williams 2006). Alternativt kan overvåkingen ha som formål å gi data til å etterprøve hypoteser om årsakssammenhenger eller utvikling i forhold til en framtidig prognose, for eksempel klimaprognoser. Både for praktisk forvaltning og forskningsbaserte hypoteser er det essensielt å besvare følgende spørsmål ved planlegging av overvåkingsaktiviteten (Yoccoz et al. 2001): **hvorfor** (hensikten med overvåkingen), **hva** (hvilke spesifikke faktorer/tilstandsvariabler skal overvåkes) og **hvordan** (skal overvåkingen designes for å være effektiv i forhold til hensikten).

Overvåking av skogens sunnhetstilstand i Norge, som har pågått siden tidlig på 1980-tallet, er et eksempel på en slik fokusert overvåking. Denne overvåkingen var basert på en velbegrunnet hypotese om at skogen (både trær og bakkevegetasjon) kunne få redusert vitalitet eller endog dø ("skogdød") som følge av langtransportert luftforurensning fra de Britiske øyer og Sentral-Europa. På dette grunnlaget etablerte man en fokusert overvåking for å kvantifisere og følge utviklingen i ulike deler av landet. Dette overvåkingsprogrammet har dokumentert flere effekter, men det ligger utenfor rammen av denne rapporten å gå inn på disse.

Når det gjelder prediksjonsmodelleringens rolle i forhold til slik målrettet overvåking, kan i hvert fall én praktisk anvendelse nevnes. Vi har flere steder i denne rapporten vist at prediksjonsmodeller er brukt til å vise potensielle effekter av framtidige klimaprognoser på arters utbredelse (se 2.2, 6.4, 7.5). Det er svært nærliggende å anvende slike prediksjonsmodeller som kvalifiserte hypoteser om hva som kan skje, og å etablere et overvåkingsprogram i forhold til å teste disse prediksjonene.

8. Utfordringer og utviklingspotensial

Prediksjonsmodellering kan utvilsomt bli et nyttig redskap for forskning og utredning innenfor en rekke områder, slik vi har beskrevet i kapittel 7. Det er imidlertid noen viktige utfordringer som må løses for at prediksjonsmodellering kan bli gjennomførbart i det omfanget behovet tilsier. I dette kapittelet skal vi ta for oss de viktigste utfordringene knyttet til et ønske om økt innsats innen prediksjonsmodellering i Norge. Vi har særlig valgt å framheve utfordringer som kan løses med relativt beskjedne økonomiske midler (i størrelsesordener opp til noen få millioner kroner), men vi omtaler også viktige utfordringer som er mer ressurskrevende å eliminere.

8.1 Relevante miljødata

En viktig forutsetning for at potensialet i prediksjonsmodellering skal kunne utnyttes i fullt monn, er tilgang på gode miljødata for relevante prediktorer. Med gode miljødata menes digitale kart som a) kvantifiserer/klassifiserer miljøfaktorer som er viktige og av direkte betydning for mange arter og naturtyper, b) har en romlig oppløsning (pixelstørrelse) som samsvarer med den skala der arter (individer eller delpopulasjoner) og naturtyper responderer, og c) har en stor geografisk dekning (ideelt hele Norge). I det etterfølgende vil vi løfte fram en del miljøfaktorer der bedre kartinformasjon vil være vesentlig for framtidig prediksjonsmodellering.

Klimadata

Generelt sett har vi gode klimadata for Norge takket være heldekkende klimakart produsert ved DNMI. Det er imidlertid behov for å lage kart for ytterligere klimavariabler med en høy tidsoppløsning (dag til dag variasjon). Variabler som er viktige for organismenes forekomst og utbredelse er gjennomsnittlig *dagtemperatur* (ikke bare døgntemperatur), luftfuktighet (*saturated and ambient vapour pressure*), evapotranspirasjon, nedbøroverskudd, nedbørfrekvens, varmesum (GDD; *Growing Degree Days*), vekstsesonglengde etc.

Disse variablene kan forholdsvis enkelt avledes fra tre grunnleggende klimavariabler der det allerede finnes gode måldata: daglig minimumstemperatur, maksimumstemperatur og nedbør. I tillegg er informasjon om snødekkevarighet viktig. Dagens snødekkedata har en romlig oppløsning på 1 km, og disse må

tilpasses en finere oppløsning basert på avledning (topografisk posisjon) fra terrengmodell. Slike nye kart for klimadata gir meget relevant informasjon og kan produseres til en relativt lav kostnad.

Klimadata er ikke like lett tilgjengelig for sjø, selv om grove overflatemodeller kan lastes ned fra satellitter. Klimadata for sjø er ønskelig for å studere effekten av endrede klimaforhold på marint miljø.

Terrengmodeller med avledninger

Det foreligger en terrengmodell for hele Norge med 25 m oppløsning avledet fra N50-kartserien med 20 m ekvidistanse. Det er behov for å lage tre spesifikke avledninger fra denne modellen: a) solinnstråling (se fig. 15 i kap. 5.2) og hvordan denne varierer gjennom året, b) estimerer for topografisk betinget jordfuktighet (topografisk våthetsindeks), og c) topografisk posisjon. Utviklingsarbeid med dette som formål er initiert i et samarbeid mellom UiO og Skog og landskap for deler av landet, men burde fullføres for hele landet.

Videre er det behov for å lage en ny, mer detaljert terrengmodell basert på N5 (ØK) med 5 m ekvidistanse. Datagrunnlaget foreligger som et vektorisert kotekart, men det må lages en terrengmodell på dette grunnlaget.

Dybdemodeller med avledninger

Også i marint miljø er det behov for en bedre terrengmodell. I dag finnes det en dybdemodell med 25 m oppløsning. Denne modellen er av variabel kvalitet, og det er behov for en enhetlig og nøytral (dvs ikke korrigert mht navigasjon) dybdemodell for hele landet med 25 m oppløsning. For å etablere en landsdekkende 25 m dybdemodell basert på de best tilgjengelige dybde-data, er det nødvendig med en avklaring med Forsvaret og Sjøkartverket, siden det per i dag foreligger en sikkerhetsklausul for tilgang til dybde-data med bedre romlig oppløsning enn 50 m.

Det er også ønskelig å etablere rasterkart for marine solinnstrålingsindekser (jf kap 5.2) for hele norskekysten basert på de tilgjengelige dybdemodelene. Dermed kan man anvende den etablerte prediksjonsmodellen for å identifisere store grunne bukter og vikar (kap 6.2) for hele landet.

Havstrømmer

Heldekkende GIS-modeller av strømforhold i marine

miljøer er essensielt fordi dette er en viktig miljøfaktor for fordeling av ulike bunntyper. Det er vist at modellert bølgeeksponering (langs kysten) har stor evne til å predikere utbredelsen av marine arter og habitater. Strøm er et annet uttrykk for vannets bevegelsesenergi (og også relevant for frie vannmasser), og forårsaker tilsvarende som bølger en kombinert forstyrrelse- og vannutskiftingseffekt. Grad av vannutskifting påvirker både tilgang og opptak av næringssalter for alger og tilgang til fødeorganismer for planktonspisende arter. Strømdata finnes kun for enkelte områder, og en innsats for å få slike data for hele Norge er nødvendig.

Substratkart for sjø

Bunnssubstrat har stor betydning for utbredelsen av arter. Artssammensetningen varierer mye mellom hardbunn og bløtbunn, samt mellom bløtbunn med ulik type kornstørrelse. Denne faktoren vet vi lite om, det foreligger substratkart (stort sett samlet inn av Norges Geologiske Undersøkelse, NGU) foreløpig kun på enkelte steder i norsk kystområder. Slike data kan skaffes via multistråleekkolodd som brukes ved kartlegging (sjømåling) av dybde og substrattypen med høy oppløsning. Slik kartlegging er kostbar, og foreløpig kun planlagt for større områder i nordnorske havområder innen programmet MAREANO. Tilsvarende kartlegging bør etter hvert utføres alle norske kystområder. Kunnskap om dybde og substrattypen er en forutsetning for kunne vite hvilke arter og naturtyper en har langs norskekysten.

Berggrunn og løsavsetninger

Det foreligger et vektorisert berggrunnskart for hele Norge i målestokk 1:250 000 som bør gjøres om til rasterkart til en oppløsning på 100 m pixel-størrelse. Videre er det behov for å 'oversette' de ulike berggrunnstypene til kategorier som er biologisk meningsfulle i form av "forvittringsrater" og mer spesifikt, rater for avgivelse av mineralnæringsstoffer (bl.a. kalsium) til jordsmonnet. For enkelte områder (for eksempel Oslofeltet) foreligger mer detaljerte berggrunnskart. Omgjøring til rasterkart av disse med en høyere oppløsning, f.eks. 25 m, ville være gunstig.

Det foreligger også kart over marin grense for hele Norge. Også dette kartet er det behov for å gjøre om til raster for eksempel med en romlig oppløsning på 25 m.

Arealdekke og kart avledet fra fjernmålingsdata

Vinteren 2007/2008 blir Digitalt markslagskart (DMK) rastret til 25 m romlig oppløsning for hele landet i et samarbeid mellom UiO og Skog & landskap. Dette kartgrunnlaget dekker imidlertid bare arealer opp til tregrensen. Det er behov for tilsvarende arealdekkekart ovenfor tregrensen. Videre er DMK noen steder foreldet, da kartet kan representere arealtilstander opptil 40 år tilbake i tid. Ytterligere et aspekt er at arealdekke tilstanden påvirkes av arealbruk, for eksempel skogens aldersfordeling som følge av hogst (se også neste punkt).

Fjernmålingsdata (satellittbilder og flybilder) er en informasjonskilde som kan utnyttes for å lage bedre kart over arealdekke (inkludert naturtypeforekomster), herunder arealdekke over tregrensen. Vi har et nasjonalt arkiv for satellittbilder og det er utviklet en del rutiner for effektiv (herunder automatisert) tolkning av disse. Det er behov for et systematisk forsknings- og evalueringsarbeid der ulike kart avledet fra satellitt- og andre fjernmåleinformasjonskilder testes for evne til å identifisere ulike naturtyper. Ny norsk naturtypeinndeling vil kunne tjene som en standard for dette evalueringsarbeidet. Gode, relevante og validerte arealdekkekart basert på fjernmåldata bør ha en romlig oppløsning på ca. 25 m og en tidsoppløsning som viser endring for eksempel med 10 års mellomrom. Produksjon av slike kart kan bygge på eksisterende rutiner for visse arealdekke kategorier, men det vil kreve til dels betydelig utviklingsarbeid for andre kategorier.

Fysiske inngrep som avstand til vei

Nye fysiske inngrep blir gjort hele tiden mens spor etter mange gamle inngrep viskes ut gjennom jengroing/suksesjon. Det er knapt mulig å kartfeste alle disse endringene. Derimot er mange fysiske inngrep korrelert med avstand til vei. Det finnes et godt oppdatert kart over det norske veinettet, inkludert skogsbilveier. Dette er imidlertid et vektorisert kart som bør gjøres om til et rasterkart som viser avstand til vei.

Kart for avstand til vei bør være et rasterkart med høy romlig oppløsning (25 m) som viser avstand fra pixel-sentrum til nærmeste vei (evt. ganske nyanserte avstandsintervaller, for eksempel 50 m). Et slikt kart bør oppdateres med 10 års mellomrom.

Kartarkiv

I dag finnes kartinformasjon for viktige prediktorer spredt hos mange dataprodusenter og brukere, og det finnes ingen samlet oversikt over avledete kartdata (som forvaltes av den som har utført eller finansiert avledningen). Det er åpenbart at det vil være store samordningsgevinster ved et felles nasjonalt bibliotek for digital avledet kartinformasjon, f.eks. forvaltet innenfor rammen av Norge Digitalt samarbeidet. Som redskap for prediksjonsmodellering vil en GIS-basert database for geografisk basisinformasjon, GIB, med gode verktøy for uttak av data, være uvurderlig. I et slikt system bør det også inngå kartfestete integrerte miljøindekser. Et eksempel på en slik indeks er modellen for trinnløs plassering av alle punkter på det norske fastlandet langs de regionale gradientene 'sone' og 'seksjon' (jfr. Moen 1998) ved ordinasjonsanalyse av et stort sett av geografiske, klimatiske, geologiske og andre grunnlagsdata (Bakkestuen et al. accepted (conditionally)) med *principal component analysis* (PCA).

8.2 Kvalitetssikrede artsdata

I Norge har vi godt med stedfestede artsfunn av tilstrekkelig høy kvalitet til å utvikle verdifulle prediksjonsmodeller og -kart (jmfør kap. 4). Like fullt vil man raskt oppdage at gode artsdata vil kunne mangle for å utføre prediksjonsmodellering, i hvert fall dersom formålet er å lage kart som skal brukes til konkret arealforvaltning med god romlig oppløsning. Spesifikke behov som vil melde seg omfatter: gjenbesøk på kjente lokaliteter for å få bedre stedfesting ved bruk av GPS (samt aktuell populasjonsstatus); dobbeltsjekke gyldighet av avvikende observasjoner (riktig artsbestemmelse, riktig stedfesting, nøyaktig stedfesting); etablere flere observasjoner av sjeldne arter. Det er utvilsomt et sterkt behov for økt aktivitet når det gjelder kartlegging av truede og rødlistede arter i Norge. Slikt arbeid bør samordnes med prediksjonsmodellering.

Av hensyn til prediksjonsmodellering men også annen bruk av artsinformasjon, bør det prioriteres å arbeide med kvalitetssikring av artsinformasjon (bestemmelseskontroll, digitalisering, stedfesting, feltoppdatering og generell korrekturkontroll av opplysninger). I denne forbindelse er det viktig å utvikle

en tradisjon for nøyaktig stedfesting ved hjelp av GPS blant leverandører av stedfestet artsinformasjon.

8.3 Analysekapasitet og feltvalidering

Ved utgangen av 2007 er det i hvert fall fire fagmiljøer som aktivt arbeider med prediksjonsmodellering basert på metodene som er beskrevet i denne rapporten: NIVA, NINA (både i Oslo og Tromsø), Univ. i Tromsø, samt Univ. i Oslo, Naturhistorisk museum. Det må imidlertid understrekes at det er et lite antall forskere som har tatt disse metodene i bruk og få av disse har prediksjonsmodellering som sin hovedbeskjeftigelse.

Kompetanse- og kapasitetsutvikling

For å videreutvikle kompetansen innen prediksjonsmodellering av arter og naturtyper er det viktig at det utvikles et samlet og sterkt fagmiljø i Norge. At denne rapporten er utarbeidet av flere institusjoner i fellesskap illustrerer at ingen av de nevnte institusjonene har kapasitet (eller ambisjoner) om å dekke feltet alene. Det vil være hensiktsmessig om institusjonenes kompetanse og kapasitet utvikles på en koordinert måte. Dette er bakgrunnen for at det i etterkant av et seminar arrangert av GBIF-Norge ved Naturhistorisk museum, Universitetet i Oslo, høsten 2007 ble etablert et nasjonalt forum for prediksjonsmodellering.

En videre utvikling av prediksjonsmodellering slik at det kan bli et viktig redskap for forskning og forvaltning vil kreve ressurser. Omfanget av naturmangfoldforskning i Norge i dag gir knapt rom for omprioritering innen eksisterende ressursrammer. En satsing vil derfor kreve ekstra ressurser for tilrettelegging av data (miljøvariabelkart), opprettelse og drift av et felles kartarkiv (GIB; se kapittel 8.1), stillinger øremerket analysekompetanse, analysekapasitet og utdanning (universitetsundervisning, stipendiater), samt generelle driftsmidler (f.eks. til å arrangere workshops/kurs, feltarbeid, se under, osv). Det burde være grunnlag for at en del av denne virksomheten kan oppdragsfinansieres.

Feltvalidering

Det er påpekt flere steder i denne rapporten at utvikling av gode prediksjonsmodeller forutsetter feltvalidering. Feltarbeid i forbindelse med prediksjonsmo-

dellering vil imidlertid ha en videre funksjon enn bare å være et ledd i modelleringsarbeidet. Slike feltsøk representerer også en målrettet innsats for å oppdage nye artsforekomster, med vesentlig større sannsynlighet for funn enn tradisjonell leting (se kap. 2.2).

Referanser

- Akaike, H. 1973. Information theory as an extension of the maximum likelihood principle. Pages 267-281 in B. N. Petrov and F. Csaki, eds. *Second International Symposium on Information Theory*. Akademiai Kiado, Budapest.
- Anonym. 2001. Kartlegging av marint biologisk mangfold. *DN-Håndbok* 19: 1-87.
- . 2002. Nasjonalt program for kartlegging og overvåking av biologisk mangfold. Arbeidsgruppe 1. Truete arter/naturtyper, fremmede arter. Direktoratet for naturforvaltning.
- . 2007. Kartlegging av Naturtyper. Verdisetting av biologisk mangfold. *DN-Håndbok* 13: 1-000.
- Araújo, M. B., M. Cabeza, W. Thuiller, L. Hannah, and P. H. Williams. 2004. Would climate change drive species out of reserves? An assessment of existing reserve-selection methods. *Global Change Biology* 10: 1618-1626.
- Araújo, M. B., and A. Guisan. 2006. Five (or so) challenges for species distribution modelling. *Journal of Biogeography* 33: 1677-1688.
- Araújo, M. B., and M. New. 2007. Ensemble forecasting of species distributions. *Trends in Ecology & Evolution* 22: 42-47.
- Austin, M. P., R. B. Cunningham, and P. M. Fleming. 1984. New approaches to direct gradient analysis using environmental scalars and statistical curve-fitting procedures. *Vegetatio* 55: 11-27.
- Austin, M. P., R. B. Cunningham, and R. B. Good. 1983. Altitudinal distribution in relation to other environmental factors of several Eucalypt species in southern New South Wales. *Australian Journal of Ecology* 8: 169-180.
- Austin, M. P., and J. A. Meyers. 1996. Current approaches to modelling the environmental niche of eucalypts: Implications for management of forest biodiversity. *Forest Ecology and Management* 85: 95-106.
- Bakkestuen, V., L. Erikstad, and R. H. Økland. accepted (conditionally). Step-less models for regional biogeoclimatic variation in Norway. *Journal of Biogeography*.
- Bekkby, T., L. Erikstad, V. Bakkestuen, and A. Bjørge. 2002. A landscape ecological approach to coastal zone applications. *Sarsia* 87: 396-408.
- Bekkby, T., L. Erikstad, O. Christensen, and O. Longva. 2005. Effekten av skala og kriterier for inndeling i marine substrattyper. *Vann* 1: 35-43.
- Bekkby, T., and M. Isæus. Accepted. Mapping large shallow inlets and bays – modelling a Natura 2000 habitat with digital terrain and wave exposure models. *ICES Journal of Marine Science*.
- Bekkby, T., E. Rinde, L. Erikstad, V. Bakkestuen, O. Longva, O. Christensen, M. Isæus, and P. E. Isachsen. Submitted. Spatial probability modelling of eelgrass *Zostera marina* L. distribution on the West coast of Norway. *ICES Journal of Marine Science*.
- Bekkby, T., and R. Rosenberg. 2006. Marine habitats utbredelse – terrengmodellering i Gullmarsfjorden. Pages 33 pp. Länsstyrelsen i Västra Götaland län, Vattenvårdsenheten.
- Beven, K. J., and M. J. Kirkby. 1979. A physically based, variable contributing area model of basin hydrology. *Hydrol. Sci. Bull.* 24: 43-69.
- Bjørkelo, K. 2006. AR5 klassifikasjonssystem, foreløpig utgave. Pages 27 pp. Norsk institutt for skog og landskap.
- Bonaccorso, E., I. Koch, and A. T. Peterson. 2006. Pleistocene fragmentation of Amazon species' ranges. *Diversity and distributions* 12: 157-164.
- Boyce, M. S., P. R. Vernier, S. E. Nielsen, and F. K. A. Schmiegelow. 2002. Evaluating resource selection functions. *Ecological Modelling* 157: 281-300.
- Bratli, H., T. Økland, R. H. Økland, W. E. Dramstad, R. Elven, G. Engan, W. Fjellstad, E. Heegaard, O. Pederesen, and H. Solstad. 2006. Patterns of variation in vascular plant species richness and composition in SE Norwegian agricultural landscapes. *Agric. Ecosyst. Environm.* 114: 270-286.
- Bruntse, G., T. E. Lein, R. Nielsen, and K. Gunnarsson. 1999. Response to wave exposure by littoral species in the Faroe Islands. *Frødskeparrit* 47: 181-198.
- Burnham, K. P., and D. R. Anderson. 2002. *Model selection and multimodel inference. A practical information-theoretic approach*. Springer-Verlag, New York.
- . 2004. Multimodel inference - understanding AIC and BIC in model selection. *Sociological Methods & Research* 33: 261-304.
- Busby, J. R. 1991. BIOCLIM - a bioclimatic analysis and prediction system. Pages pp. 64-68 in C. R. Margules and M. P. Austin, eds. *Nature Conservation: Cost Effective Biological Surveys and data Analysis*. CSIRO, Melbourne.
- Chase, J. M., and M. A. Leibold. 2003. *Ecological niches - linking classical and contemporary approaches*. The University of Chicago Press, Chicago.

- Christie, H., N. M. Jørgensen, and K. M. Norderhaug. 2007. Bushy or smooth, high or low; importance of habitat architecture and vertical level for distribution of fauna on kelp. *Journal of Sea Research* In press.
- Christie, H., N. M. Jørgensen, K. M. Norderhaug, and E. Waage-Nielsen. 2003. Species distribution and habitat exploitation of fauna associated to kelp (*Laminaria hyperborea*) at the Norwegian coast. *Journal of the Marine Biological Association of the United Kingdom* 83: 687-699.
- Clark, J. S., and A. E. Gelfand, eds. 2006. *Hierarchical modelling for the environmental sciences*. Oxford University Press, Oxford.
- Cox, D. R., and N. Wermuth. 1992. A comment on the coefficient of determination for binary responses. *The American Statistician* 46: 1-4.
- Crawley, M. J. 2002. *Statistical Computing*. Wiley.
- Crothers, J. H. 1976. On the distribution of some common animals and plants along the rocky shores of West Somerset. *Field Stud.* 4: 469-489.
- Dahlberg, A., and J. N. Stokland. 2004. Vedlevande arter krav på substrat – en sammanställning och analys av 3600 arter. Pages 75 pp. Skogsstyrelsen, rapport 2004: 7.
- De Leeuw, J. 1992. Introduction to Akaike (1973) information theory and an extension of the maximum likelihood principle. Pages 599-609 in S. Kotz and N. L. Johnson, eds. *Breakthroughs in statistics*. Springer-Verlag, London.
- De'ath, G., and K. E. Fabricius. 2000. Classification and regression trees: a powerful, yet simple technique for ecological data analysis. *Ecology* 81: 3178-3192.
- DN. 1998. Plan for overvåking av biologisk mangfold. Pages DN-rapport 1998-1 Direktoratet for Naturforvaltning, Trondheim.
- Dormann, C. F., J. M. McPherson, M. B. Araujo, R. Bivand, J. Bolliger, G. Carl, R. G. Davies, A. Hirzel, W. Jetz, W. D. Kissling, I. Kuhn, R. Ohlemuller, P. R. Peres-Neto, B. Reineking, B. Schroder, F. M. Schurr, and R. Wilson. 2007. Methods to account for spatial autocorrelation in the analysis of species distributional data: a review. *Ecography* 30: 609-628.
- Dungan, J. L., J. N. Perry, M. R. T. Dale, P. Legendre, S. Citron-Pousty, M.-J. Fortin, A. Jakomulska, M. Miriti, and M. S. Rosenberg. 2002. A balanced view of scale in spatial statistical analysis. *Ecography* 25: 626-640.
- Edwards jr, T. C., D. R. Cutler, N. E. Zimmermann, L. Geiser, and J. Alegria. 2005. Model-based stratifications for enhancing the detection of rare ecological events. *Ecology* 86: 1081-1090.
- Edwards jr, T. C., D. R. Cutler, N. E. Zimmermann, L. Geiser, and G. G. Moisen. 2006. Effects of sample survey design on the accuracy of classification tree models in species distribution models. *Ecological Modelling* 199: 132-141.
- Elith, J., and M. A. Burgman. 2002. Predictions and their validation: rare plants in the central highlands, Victoria, Australia. Pages 303-313 in J. M. Scott, P. J. Heglund, J. B. Haufler, M. Morrison, M. G. Raphael, W. B. Wall, and E. Samson, eds. *Predicting species occurrences: issues of accuracy and scale*. Island Press, Covelo, California.
- Elith, J., C. G. Graham, R. P. Anderson, M. Dudík, S. Ferrier, A. Guisan, R. J. Hijmans, F. Huettmann, J. R. Leathwick, A. Lehmann, J. Li, L. Lohmann, B. Loiselle, G. Manion, C. Moritz, M. Nakamura, Y. Nakazawa, J. M. Overton, A. T. Peterson, S. J. Phillips, K. Richardson, R. Scachetti-Pereira, R. E. Schapire, J. Soberon, S. Williams, M. S. Wisz, and N. E. Zimmermann. 2006. Novel methods improve prediction of species' distributions from occurrence data. *Ecography* 29: 129-151.
- Engler, R., A. Guisan, and I. REchsteiner. 2004. An improved approach for predicting the distribution of rare and endangered species from occurrence and pseudo-absence data. *Journal of Applied Ecology* 41: 263-274.
- Faraway, J. J. 2006. *Extending the linear model with R: Generalized linear, mixed effects and nonparametric regression models*. Chapman and Hall/CRC, London.
- Ferrier, S. 2002. Mapping Spatial Pattern in Biodiversity for Regional Conservation Planning: Where to from Here? *Sysyematic Biology* 2: 331-363.
- Ficetola, G. F., W. Thuiller, and C. Miaud. 2007. Prediction and validation of the potential global distribution of a problematic alien invasive species — the American bullfrog. *Diversity and Distributions* 13: 476-485.
- Fielding, A. H., and J. F. Bell. 1997. A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environmental Conservation* 24: 38-49.
- Framstad, E., and J. A. Kålås. 2001. TOV 2000. Nytt program for overvåking av terrestrisk biologisk mangfold - videreutvikling av dagens naturover-

- våking (TOV). NINA Oppdragsmelding 702. 49 s.
- Fremstad, E. 1997. Vegetasjonstyper i Norge. *Norsk Inst. Naturforsk. Temahefte* 12: 1-279.
- Friedman, J., T. Hastie, and R. Tibshirani. 2000. Additive logistic regression: a statistical view of boosting. *Annals of Statistics* 28: 337-374.
- Gelfand, A. E., A. Latimer, S. Wu, and J. A. Silander. 2006. Building statistical models to analyze species distributions. Pages 77-97 in J. S. Clark and A. E. Gelfand, eds. *Hierarchical modelling for the environmental sciences*. Oxford University Press, Oxford.
- Gneiting, T., and A. E. Raftery. 2005. Atmospheric science - Weather forecasting with ensemble methods. *Science* 310: 248-249.
- Guisan, A., O. Broennimann, R. Engler, M. Vust, N. G. Yoccoz, A. Lehmann, and N. E. Zimmermann. 2006. Using niche-based models to improve the sampling of rare species. *Conservation Biology* 20: 501-511.
- Guisan, A., C. H. Graham, J. Elith, F. Huettman, and N. S. D. M. Group. 2007a. Sensitivity of species distribution models to change in grain size. *Diversity and distributions* 13: 332-340.
- Guisan, A., and J.-T. Theurillat. 2000. Assessing alpine plant vulnerability to climate change: a modeling perspective. *Integrated Assessment* 1: 307-320.
- Guisan, A., and W. Thuiller. 2005. Predicting species distribution: Offering more than simple habitat models. *Ecological Letters* 8: 993-1009.
- Guisan, A., N. E. Zimmermann, J. Elith, C. H. Graham, S. J. Phillips, and A. T. Peterson. 2007b. What matters for predicting the occurrences of trees: techniques, data, or species' characteristics? *Ecological Monographs* 77: 615-630.
- Hacker, S. D., and R. S. Steneck. 1990. Habitat architecture and the abundance and body-size-dependent habitat selection of a phytal amphipod. *Ecology* 71: 2269-2285.
- Hannah, L., G. Midgley, S. Andelman, M. B. Araújo, G. Hughes, E. Martinez-Meyer, R. Pearson, and P. Williams. 2007. Protected area needs in a changing climate. *Frontiers in Ecology and the Environment* 5: 131-138.
- Hastie, T., and R. Tibshirani. 1990. *Generalized additive models*. Chapman and Hall, London.
- Hastie, T., R. Tibshirani, and J. Friedman. 2001. *The elements of statistical learning. Data mining, inference and prediction*. Springer-Verlag, New-York.
- Heegaard, E., R. H. Økland, H. Bratli, W. E. Dramstad, G. Engan, O. Pedersen, and H. Solstad. 2007. Regularity of species richness - relationships to patch size and shape. *Ecography* 30: 589-597.
- Hobert, J. P., N. S. Altman, and C. L. Schofield. 1997. Analyses of fish species richness with spatial covariate. *Journal of the American Statistical Association* 92: 846-854.
- Hurvich, C. M., and C.-L. Tsai. 1989. Regression and time series model selection in small samples. *Biometrika* 76: 297-307.
- Irfan-Ullah, M., G. Amarnath, M. S. R. Murthy, and A. T. Peterson. 2007. Mapping the geographic distribution of *Aglaia bourdillonii* Gamble (Meliaceae), an endemic and threatened plant, using ecological niche modeling. *Biodiversity and Conservation* 16: 1917-1925.
- Isæus, M. 2004. Factors structuring *Fucus* communities at open and complex coastlines in the Baltic Sea. *Department of Botany*, . Stockholm University, Sweden.
- Jensen, R. A., J. Madsen, M. O'Connell, M. S. Wisz, H. Tømmervik, and F. Mehlum. 2008. Prediction of the distribution of Arctic-nesting pink-footed geese under a warmer climate scenario. *Global Change Biology* 14: 1-10.
- Jolliffe, I. T., and D. B. Stephenson. 2004. *Forecast verification. A practitioner's guide in atmospheric science*. John Wiley & Sons, New York.
- Körner, C. 2007. The use of 'altitude' in ecological research. *Trends in Ecology and Evolution* 21: 569-574.
- Kålås, J. A., Å. Viken, and T. Bakken. 2006. Norsk Rødliste 2006 - 2006 Norwegian Red List. Artsdatabanken, Trondheim, Norway.
- Lehmann, A., J. R. Leathwick, and J. M. Overton. 2004. GRASP v.3.0 User's Manual. Swiss Centre for Faunal Cartography, Switzerland.
- Lehmann, A., J. M. Overton, and J. R. Leathwick. 2003. GRASP: generalized regression analysis and spatial prediction. *Ecological Modelling* 160: 165-183.
- Levins, R. 1966. The strategy of model building in population biology. *American Scientist* 54: 421-431.
- . 1969. Some demographic and genetic consequences of environmental heterogeneity for biological control. *Bulletin of the American Entomological Society* 15: 237-240.
- Link, W. A., and R. J. Barker. 2006. Model weights and the foundations of multimodel inference. *Ecology* 87: 2626-2635.
- Maggini, R., A. Lehmann, N. E. Zimmermann, and A. Guisan. 2006. Improving generalized regression analysis for the spatial prediction of forest com-

- munities. *Journal of Biogeography* 33: 1729–1749.
- Mallows, C. L. 1973. Some comments on Cp. *Technometrics* 15: 661–675.
- Martínez-Meyer, E., A. T. Peterson, and W. W. Hargrove. 2004. Ecological niches as stable distributional constraints on mammal species, with implications for Pleistocene extinctions and climate change projections for biodiversity. *Global Ecology and Biogeography* 13: 305–314.
- McCullagh, P., and J. A. Nelder. 1989. *Generalized linear models*. Chapman and Hall, London.
- Mittlböck, M., and H. Heinzl. 2001. A note on R-2 measures for Poisson and logistic regression models when both models are applicable. *Journal of Clinical Epidemiology* 54: 99–103.
- Moen, A. 1998. *Nasjonalatlas for Norge: Vegetasjon*. Statens Kartverk, Hønefoss.
- Mysterud, A., N. G. Yoccoz, R. Langvatn, N. Pettorelli, and N. C. Stenseth. 2008. Hierarchical path analysis of deer responses to direct and indirect effects of climate in northern forest. *Philosophical Transactions of the Royal Society B-Biological Sciences* in press.
- Nichols, J. D., and B. K. Williams. 2006. Monitoring for conservation. *Trends in Ecology & Evolution* 21: 668–673.
- Norderhaug, K. M. 2004. Use of red algae as hosts by kelp-associated amphipods. *Marine Biology* 225–230: 144.
- Norderhaug, K. M., M. Isæus, T. Bekkby, F. Moy, and A. Pedersen. 2007. Spatial predictions of *Laminaria hyperborea* at the Norwegian Skagerrak coast. Pages 28 p.
- Norton, T. A., K. Hiscock, and J. A. Kitching. 1977. The ecology of Lough Ine XX. The *Laminaria* forest at Carrigathorna. *Journal of Ecology* 65: 919–941.
- Pearce, J., and D. Lindenmayer. 1998. Bioclimatic analysis to enhance reintroduction biology of the endangered Helmeted Honeyeater (*Lichenostomus melanops cassidix*) in southeastern Australia. *Restoration Ecology* 6: 238–243.
- Pedersen, A. Ø., J. U. Jepsen, N. G. Yoccoz, and E. Fuglei. 2007. Ecological correlates of the distribution of territorial Svalbard rock ptarmigan (*Lagopus muta hyperborea*). *Canadian Journal of Zoology* 85: 122–132.
- Peterson, A. T., M. Papes, and D. A. Kluza. 2003. Predicting the potential invasive distributions of four alien plant species in North America. *Weed Science* 51: 863–868.
- Phillips, S. J., R. P. Anderson, and R. E. Schapire. 2006. Maximum entropy modeling of species geographic distributions. *Ecological Modelling* 190: 231–259.
- Pulliam, H. R. 2000. On the relationship between niche and distribution. *Ecology Letters* 3: 349–361.
- Randin, C. F., T. Dirnbock, S. Dullinger, N. E. Zimmermann, M. Zappa, and A. Guisan. 2006. Are niche-based species distribution models transferable in space? *Journal of Biogeography* 33: 1689–1703.
- Rekdal, Y., and J. Y. Larsson. 2005. Veiledning i vegetasjonskartlegging M 1:20 000 - 50 000. *Norsk Inst. Jord- Skogkartlegging Rapp. 2005*: 1–108.
- Richardson, D. M., and W. Thuiller. 2007. Home away from home — objective mapping of high-risk source areas for plant introductions. *Diversity and Distributions* 13: 299–312.
- Ridgeway, G. 2000. Discussion of “Additive Logistic Regression: a Statistical View of Boosting,” by J. Friedman, T. Hastie, and R. Tibshirani. *Annals of Statistics* 28: 393–400.
- Riksrevisjonen. 2006. Riksrevisjonens undersøkelse av myndighetenes arbeid med kartlegging og overvåking av biologisk mangfold og forvaltning av verneområder. Pages 104 s. Riksrevisjonen.
- Rinde, E., B. Rygg, T. Bekkby, M. Isæus, L. Erikstad, S.-E. Sloreid, and O. Longva. 2006. Dokumentasjon av modellerte marine naturtyper i DN's Naturbase. Førstegenerasjonsmodeller til kommunenes startpakker for kartlegging av marine naturtyper. Pages 32 pp.
- Räisänen, J., and L. Ruokolainen. 2006. Probabilistic forecasts of near-term climate change based on a resampling ensemble technique. *Tellus Series a-Dynamic Meteorology and Oceanography* 58: 461–472.
- Schultze, K., K. Janke, A. Krüß, and W. Weidemann. 1990. The macrofauna and macroflora associated with *Laminaria digitata* and *L. hyperborea* at the island of Helgoland (German Bight, North Sea). *Helgoländer Meeresuntersuchungen* 44: 39–51.
- Shchepetkin, A. F., and J. C. McWilliams. 2005. Regional Ocean Model System: a split-explicit ocean model with a free surface and topography-following vertical coordinate. *Ocean Modelling* 9: 347–404.
- Sjötun, K., and S. Fredriksen. 1995. Growth allocation in *Laminaria hyperborea* (Laminariales, Phaeophyceae) in relation to age and wave exposure. *Marine Ecology Progress Series* 126: 213–222.
- Sjötun, K., S. Fredriksen, and J. Rueness. 1998. Effect of canopy biomass and wave exposure on growth in *Laminaria hyperborea* (Laminariaceae: Phaeophyceae).

- ta). *European Journal of Phycology* 33: 337-343.
- Skjeggedal, T., T. Arnesen, J. Kveli, P. G. Thingstad, G. Wollan, and J. Aastre. 2005. Inngrepsfrienaturområder som verktøy for arealforvaltning. Pages 283 pp.
- Soldal, E., T. Bekkby, E. Rinde, V. Bakkestuen, L. Erikstad, and M. Isæus. 2008. Predictive probability modelling of marine habitats – a case study from the West coast of Norway. *Proceedings from the International Symposium on Integrated Coastal Zone Management, Arendal, Norway, 11-14 June 2007*. Blackwell Sciences.
- Subrahmanyam, R. 1959. Ecological studies on the Fucales. 1. *Pelvetia canaliculata* Dcne. et Thur. Edited for publication from part of the Thesis accepted for Ph. D. . University of Liverpool, UK.
- Sverdrup-Thygeson, A., H. H. Blom, T. E. Brandrud, H. Bratli, O. Skarpaas, and F. Ødegaard. 2007. Kartlegging og overvåking av rødlistearter. Delprosjekt II: Arealer for Rødlistearter - Kartlegging og Overvåking (AR-KO). Faglig framdriftsrapport for 2006. *NINA Rapport 238*. NINA, Trondheim.
- Swets, K. A. 1988. Measuring the accuracy of diagnostic systems. *Science* 240: 1285-1293.
- Sørensen, R., U. Zinko, and J. Seibert. 2005. On the calculation of the topographic wetness index: evaluation of different methods based on field observations. *Hydrol. Earth Sys. Sci. Discuss.* 2: 1807-1834.
- Taylor, P. 2000. Socio-ecological webs and sites of sociality: Levins' strategy of model building revisited. *Biology and Philosophy* 15: 197-210.
- Thompson, W. L. 2004. *Sampling rare or elusive species: concepts, designs and techniques for estimating population parameters*. Island Press Washington, DC.
- Thuiller, W., M. B. Araujo, and S. Lavorel. 2003. Generalized models vs. classification tree analysis: predicting spatial distributions of plant species at different scales. *Journal of Vegetation Science* 14: 669-680.
- Thuiller, W., S. Lavorel, M. B. Araujo, M. T. Sykes, and I. C. Prentice. 2005. Climate change threats to plant diversity in Europe. *Proceedings of the National Academy of Sciences of the United States of America* 102: 8245-8250.
- Thuiller, W., G. F. Midgley, M. Rougeti, and R. M. Cowling. 2006. Predicting patterns of plant species richness in megadiverse South Africa. *Ecography* 29: 733-744.
- van der Linde, A. 2005. DIC in variable selection. *Statistica Neerlandica* 59: 45-56.
- Vayssières, M. P., R. E. Plant, and B. H. Allen-Diaz. 2000. Classification trees: An alternative non-parametric approach for predicting species diatribution. *Journal of Vegetation Science* 11: 679-694.
- Venables, W. N., and B. D. Ripley. 2002. *Modern applied statistics with S*. Springer-Verlag, Berlin.
- Wiens, J. 1989. Spatial scaling in ecology. *Functional Ecology* 3: 385-397.
- Wold, S., A. Ruhe, H. Wold, and W. J. Dunn. 1984. The collinearity problem in linear regression: the Partial Least-Squares (PLS) approach to generalized inverses. *Siam Journal on Scientific and Statistical Computing* 5: 735-743.
- Wold, S., M. Sjostrom, and L. Eriksson. 2001. PLS-regression: a basic tool of chemometrics. *Chemometrics and Intelligent Laboratory Systems* 58: 109-130.
- Wood, S. N. 2006. *Generalized additive models: an introduction with R*. Taylor & Francis, CRC Press, London.
- Yee, T. W., and N. D. Mitchell. 1991. Generalized additive models in plant ecology. *Journal of Vegetation Science* 2: 587-602.
- Yoccoz, N. G., J. D. Nichols, and T. Boulinier. 2001. Monitoring of biological diversity in space and time. *Trends in Ecology and Evolution* 16: 446-453.
- Zheng, B. Y., and A. Agresti. 2000. Summarizing the predictive power of a generalized linear model. *Statistics in Medicine* 19: 1771-1781.
- Zimmermann, N. E., T. C. Edwards jr, G. G. Moisen, T. S. Frescino, and J. A. Blackard. 2007. Remote sensing-based predictors improve distribution models of rare, early successional and broadleaf tree species in Utah. *Journal of Applied Ecology* 44: 1057-1067.
- Ødegaard, F., H. H. Blom, T. E. Brandrud, J. B. Jordal, J. E. Nilsen, J. Stokland, A. Sverdrup-Thygeson, and P. A. Aarrestad. 2006. Kartlegging og overvåking av rødlistearter. Delprosjekt II: Kartlegging og overvåking av prioriterte lokaliteter for rødlistearter. *NINA Rapport 174*. NINA, Trondheim.
- Økland, R. H. 2007. Wise use of statistical tools in ecological field studies. *Folia geobot.* 42: 123-140.
- . manuscript. Faglig grunnlag for en helhetlig nasjonal plan for overvåking av naturmangfold og rollen til Ny norsk naturtypeinndeling (NNN).
- Økland, R. H., H. Bratli, W. E. Dramstad, A. Edvardsen, G. Engan, W. Fjellstad, E. Heegaard, O. Pedersen, and H. Solstad. 2006. Scale-dependent importance of environment, land use and landscape structure for species richness and composition of SE Norwegian

modern agricultural landscapes. *Landscape Ecology* 21: 969-987.

Økland, R. H., T. Økland, and K. Rydgren. 2001. Vegetation-environment relationships of south boreal spruce swamp forests in Østmarka nature reserve, SE Norway. *Sommerfeltia* 29: 1-190.



Naturhistorisk museum (NHM) er landets største i sitt slag med over 65 prosent av landets samlinger av biologisk og geologisk mangfold. Samlingene er hovedkilden til kunnskap om naturens endringer i tid og rom, og utgjør en omfattende ikke-human genbank.

NHM produserer kunnskap for allmennhet, forvaltning og forskere. Museets virksomhet er basert på samling, dokumentasjon, forskning og formidling, i tillegg til undervisning. Botanisk hage er en viktig del av museet og har holdt til på Tøyen siden 1814. NHM driver forskning og oppdragsvirksomhet innen biologi og geologi, blant annet knyttet til artsmangfold og miljøovervåkning.