

## Dataflyt og modellering av indikatorer for naturindeksen

Olav Skarpaas, Joachim P. Töpper, Erlend B. Nilsen og Jens Åström



## **NINAs publikasjoner**

### **NINA Rapport**

Dette er NINAs ordinære rapportering til oppdragsgiver etter gjennomført forsknings-, overvåkings- eller utredningsarbeid. I tillegg vil serien favne mye av instituttets øvrige rapportering, for eksempel fra seminarer og konferanser, resultater av eget forsknings- og utredningsarbeid og litteraturstudier. NINA Rapport kan også utgis på annet språk når det er hensiktsmessig..

### **NINA Temahefte**

Som navnet angir behandler temaheftene spesielle emner. Heftene utarbeides etter behov og serien favner svært vidt; fra systematiske bestemmelsesnøkler til informasjon om viktige problemstillinger i samfunnet. NINA Temahefte gis vanligvis en populærvitenskapelig form med mer vekt på illustrasjoner enn NINA Rapport.

### **NINA Fakta**

Faktaarkene har som mål å gjøre NINAs forskningsresultater raskt og enkelt tilgjengelig for et større publikum. Faktaarkene gir en kort framstilling av noen av våre viktigste forskningstema.

### **Annen publisering**

I tillegg til rapporteringen i NINAs egne serier publiserer instituttets ansatte en stor del av sine vitenskapelige resultater i internasjonale journaler, populærfaglige bøker og tidsskrifter.

# Dataflyt og modellering av indikatorer for naturindeksen

Olav Skarpaas  
Joachim P. Töpper  
Erlend B. Nilsen  
Jens Åström

Skarpaas, O., Töpper, J.P., Nilsen, E.B., Åström, J. 2018. Dataflyt og modellering av indikatorer for naturindeksen. NINA Rapport 1560. Norsk institutt for naturforskning.

Oslo, oktober 2018

ISSN: 1504-3312

ISBN: 978-82-426-3299-9

RETTIGHETSHAVER

© Norsk institutt for naturforskning

Publikasjonen kan siteres fritt med kildeangivelse

TILGJENGELIGHET

Åpen

PUBLISERINGSTYPE

Digitalt dokument (pdf)

KVALITETSSIKRET AV

Erik Framstad

ANSVARLIG SIGNATUR

Forskningssjef Kristin Thorsrud Teien (sign.)

OPPDRAUGSGIVER(E)/BIDRAGSYTER(E)

Miljødirektoratet

OPPDRAUGSGIVERS REFERANSE

M-1003|2018

KONTAKTPERSON(ER) HOS OPPDRAGSGIVER/BIDRAGSYTER

Else Løbersli

FORSIDEBILDE

Issoleie *Ranunculus glacialis*. © Olav Skarpaas

NØKKEWORD

Norge, Sogn og fjordane fylke, Aurland kommune

Fjellfiol, *Viola biflora*

Issoleie, *Ranunculus glacialis*

Naturindeks

Modellering

KEY WORDS

Norway, Sogn og fjordane county, Aurland municipality

Yellow arctic violet, *Viola biflora*

Glacier buttercup, *Ranunculus glacialis*

Nature index

Modelling

#### KONTAKTOPPLYSNINGER

##### **NINA hovedkontor**

Postboks 5685 Torgarden  
7485 Trondheim  
Tlf: 73 80 14 00

##### **NINA Oslo**

Gaustadalléen 21  
0349 Oslo  
Tlf: 73 80 14 00

##### **NINA Tromsø**

Postboks 6606 Langnes  
9296 Tromsø  
Tlf: 77 75 04 00

##### **NINA Lillehammer**

Vormstuguvegen 40  
2624 Lillehammer  
Tlf: 73 80 14 00

##### **NINA Bergen**

Thormøhlensgate 55  
5006 Bergen  
Tlf: 73 80 14 00

[www.nina.no](http://www.nina.no)



## Sammendrag

Skarpaas, O., Töpper, J.P., Nilsen, E.B., Åström, J. 2018. Dataflyt og modellering av indikatorer for naturindeksen. NINA Rapport 1560. Norsk institutt for naturforskning.

Som en del av arbeidet mot en kostnadseffektiv, robust og høyoppløselig naturindeks, er det ønskelig å se nærmere på muligheter for å kombinere data fra ulike kilder ved modellering, og legge til rette for forbedret romlig oppløsning, geografisk dekning og automatisering av naturindeksberegninger. I denne rapporten forsøker vi å ta noen skritt i retning av å svare på disse utfordringene gjennom konseptuelle rammeverk for integrert hierarkisk modellering og dataflytstruktur, illustrert med eksempler.

Vi omtaler kort noen viktige datakilder og deres egenskaper (datastruktur, format) samt muligheter for eksport/import, med fokus på datakilder som er relevante for modelleringsarbeid i dette og beslektede prosjekter (planter, fugl og miljøvariabler). På sikt vil en skriptbasert dataflyt legge til rette for automatisert oppdatering av modeller og indikatorverdier i naturindeksen når datasettene oppdateres. Vi har imidlertid ikke gått langt i retning av standardisering av dataflyt i dette prosjektet, fordi det har blitt klart i løpet av arbeidet at dataflyt til naturindeksen bør være en del av en mer langsiktig og bredere tverrinstitusjonell prosess.

For å illustrere modellering som kan gi heldekkende indikatorverdier, bruker vi fjellfiol *Viola biflora* (tilsvarende øvelser for fugl presenteres i en egen rapport). Fjellfiol er ikke en indikatorart for naturindeksen, men fungerer som et demonstrasjonseksempel på hvordan klimadata, populasjonsdata og forekomstdata kan integreres for å predikere forekomst og populasjonsstørrelse over tid, og gi estimater av indikatorverdi og tilhørende usikkerhet med høy oppløsning.

Vi beskriver den nyeste utviklingen av rutiner for import av modellerte indikatorverdier til naturindeksbasen med R-pakken 'Nlcalc'. I tillegg til det tradisjonelle formatet på indikatorverdier (middelverdi og kvartiler), vil nå også parameteriserte lognormalfordelinger, parameteriserte poissonfordelinger, diskrete fordelinger, og empiriske fordelinger aksepteres som indikatorverdier. Dette øker mulighetene for å bruke resultater fra flere ulike typer modeller, inkludert integrerte hierarkiske modeller, direkte som input til naturindeksen.

Avslutningsvis diskuterer og oppsummerer vi våre resultater og erfaringer i forhold til muligheter for økt datatilgang og automatisert oppdatering av naturindeksen. Vi konkluderer med to anbefalinger: For det første, at rammeverket for dataflyt til naturindeksen bør tas videre i et bredt tverrinstitusjonelt samarbeid mellom sentrale dataaktører og sees i sammenheng med beslektede prosesser og systemer i forvaltningen. For det andre, at modelleringsarbeidet i dette og tidligere prosjekter, som gir grunnlag for å øke oppløsning og geografisk dekning av naturindeksindikatorer, samt å øke antall indikatorer, legges til grunn for en revurdering av naturindeksens indikatorsett for planter fram mot neste hovedoppdatering.

Olav Skarpaas, Joachim P. Töpper, Erlend B. Nilsen, Jens Åström  
NINA, Gaustadalléen 21, 0349 Oslo  
[olav.skarpaas@nina.no](mailto:olav.skarpaas@nina.no)

## Abstract

Skarpaas, O., Töpper, J.P., Nilsen, E.B., Åström, J. 2018. Data flow and modelling of indicators for the nature index. NINA Report 1560. Norwegian Institute for Nature Research.

As part of the process towards a cost-efficient, robust and high-resolution nature index, there is a need to investigate options for utilizing and combining data from different sources by modelling, and to improve spatial resolution and geographical coverage and facilitate automation of nature index calculations. In this report, we make a few steps towards these challenges through a conceptual framework for integrated hierarchical modelling and data flow, illustrated with examples.

We briefly describe some important data sources and their properties (structure, format), as well as possibilities for export/import, with focus on data of relevance for modelling in this and related projects (plants, birds and environmental variables). In due course, a script-based data flow will facilitate automated updating of models and indicators in the nature index when data sources are updated. However, we did not go far in standardizing data flow in this project, because it has become clear during the project that dataflow into the nature index should be part of a long-term broader process involving several institutions.

To illustrate modelling approaches that may give full-cover indicator values, we use the arctic yellow violet (*Viola biflora*) as an example (modelling of birds is presented in a separate report). The arctic yellow violet is not an indicator species in the nature index, but serves as an example to demonstrate how climate data, population data, and occurrence data can be integrated to predict population development over time, and give estimates of indicator values and uncertainties with high geographical resolution.

We describe the latest development of routines for import of modelled indicator values to the nature index database with the R package 'Nlcalc'. In addition to the traditional data format (means and quartiles), parameterized lognormal distributions, Poisson distributions, discrete distributions and empirical distributions can be accepted as indicator values. This increases the possibilities for using results from several kinds of models, including integrated hierarchical models, as direct input to the nature index.

In the final chapter, we discuss and summarize our results and experiences with respect to possibilities for increased data access and automated updating of the nature index. We conclude with two recommendations. Firstly, we recommend that the framework for data flow should be further developed in a broader collaboration between key data managers and be coordinated with related processes and management systems. Secondly, modelling work in this and other projects, which may increase resolution and geographical coverage of nature index indicators, and increase the number of indicators, should be used as a basis for a revision of the plant indicator set towards the next main update of the nature index.

Olav Skarpaas, Joachim P. Töpper, Erlend B. Nilsen, Jens Åström  
NINA, Gaustadalléen 21, 0349 Oslo  
[olav.skarpaas@nina.no](mailto:olav.skarpaas@nina.no)

# Innhold

<b>Sammendrag .....</b>	<b>3</b>
<b>Abstract .....</b>	<b>4</b>
<b>Innhold .....</b>	<b>5</b>
<b>Forord .....</b>	<b>6</b>
<b>1 Innledning.....</b>	<b>7</b>
1.1 Modellering og integrering av datakilder .....	7
1.2 Dataflytstruktur for automatisert oppdatering av naturindeksen .....	9
<b>2 Datakilder og metoder for innhenting av data .....</b>	<b>10</b>
2.1 Datakilder.....	10
2.2 Metoder og skript for innlesing av data .....	11
<b>3 Modellering av indikatorverdier .....</b>	<b>13</b>
3.1 Demografiske data.....	13
3.2 Populasjonsmodellering .....	14
3.3 Framskriving av populasjonsstørrelse.....	15
<b>4 Import av modellerte indikatorverdier til NI-basen .....</b>	<b>18</b>
<b>5 Diskusjon.....</b>	<b>20</b>
5.1 Automatisert dataflyt.....	20
5.2 Integrert modellering for bedre geografisk oppløsning og dekning.....	21
5.3 Oppsummering og anbefalinger .....	21
<b>6 Referanser .....</b>	<b>23</b>

## Forord

Denne rapporten er utarbeidet av NINA med støtte fra Miljødirektoratet, som svar på behovet for å videreutvikle og styrke naturindeksens datagrunnlag (jfr. Miljødirektoratets utlysning, ref 2016/4526). Vi takker for støtten fra Miljødirektoratet, både økonomisk og praktisk.

Prosjektet har tatt for seg dataflytstruktur og modellering av utvalgte planter og fugl. I denne rapporten omtales modelleringsarbeid for planter. Resultater for fugl vil bli omtalt i en egen rapport.

12. oktober 2018

Olav Skarpaas  
Prosjektleder



# 1 Innledning

Som en del av arbeidet mot en kostnadseffektiv, robust og høyoppløselig naturindeks, er det ønskelig å se nærmere på muligheter for å kombinere data fra ulike kilder ved modellering og legge til rette for forbedret romlig oppløsning, geografisk dekning og automatisering av naturindeksberegninger.

I naturindeksen brukes i dag flere store og viktige datakilder, som GBIF/Artskart og NINAs terrestriske naturovervåking (TOV). Disse datakildene er imidlertid ikke godt utnyttet. I tillegg finnes flere andre datakilder som ikke utnyttes fordi strukturene for å omsette alle tilgjengelige data til naturindeksverdier ikke er på plass. Det gjelder strukturer for dataflyt, men ikke minst statistisk holdbare modeller for geografisk heldekkende estimering med tilstrekkelig romlig oppløsning.

I denne rapporten tar vi noen skritt i retning av å svare på noen av disse utfordringene. Målsettingen er todelt. For det første vil vi vurdere hvordan ulike datakilder kan kombineres for å få fram indikatorverdier med bedre romlig dekning og oppløsning. For det andre vil vi drøfte hvordan dataflyt og estimeringsprosedyrer kan tilrettelegges for økt datatilgang og på sikt bidra til automatiserte beregninger.

## 1.1 Modellering og integrering av datakilder

Integrering av data fra ulike kilder krever modellverktøy utover de standardmodeller som er lett tilgjengelig i eksisterende programmer. Et lovende modellrammeverk er såkalte integrerte hierarkiske modeller, som kombinerer forskjellige data for å beskrive ulike deler av en prosess av interesse (her: endringer i indikatorernes tilstandsverdi), og samtidig tar høyde for at slike data er samlet med ulike metoder og dermed har ulike statistiske egenskaper og feilkilder (**Figur 1**). Slike modeller gir store muligheter for bedre presisjon i estimatene, bedre prosessforståelse, økt geografisk dekning og økt romlig oppløsning. Alt dette vil kunne styrke datagrunnlaget for naturindeksen. Vi ser også på programmering av slike modeller, med tilhørende dataflyt, som et svært viktig skritt mot en framtidig automatisering av beregningene i naturindeksen.

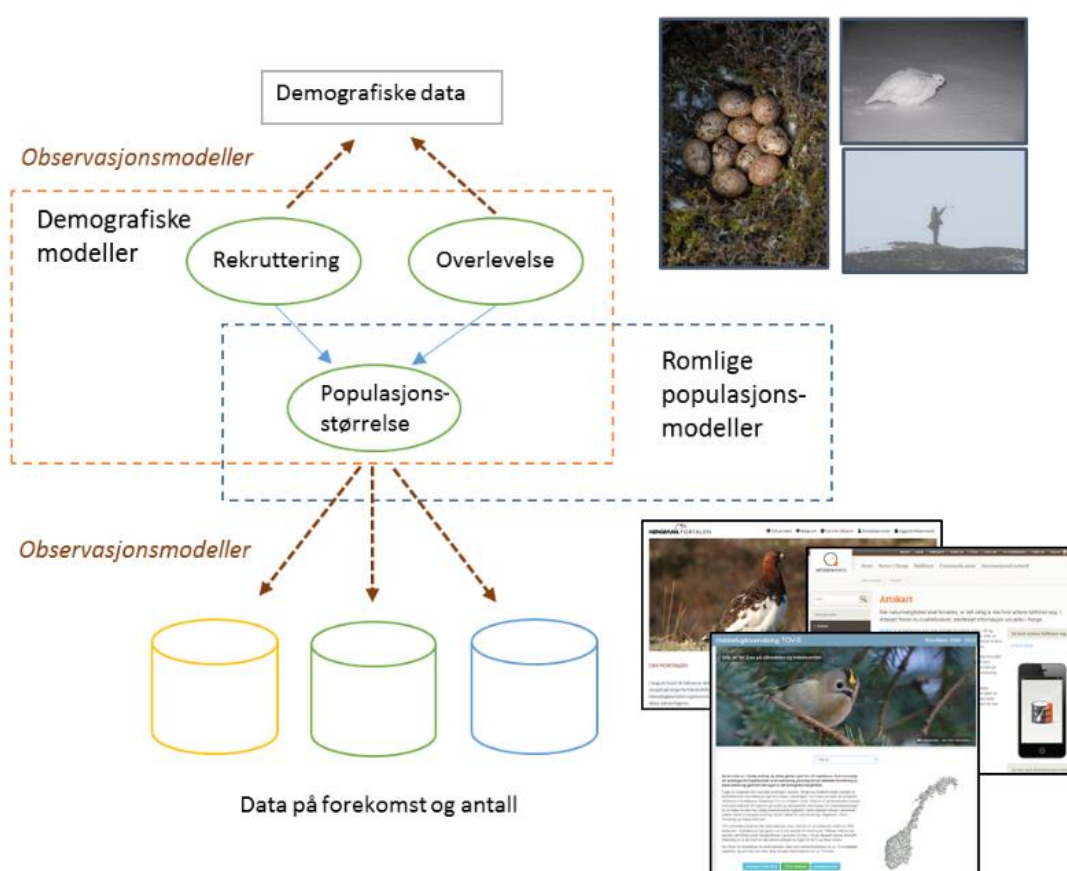
I dette prosjektet har vi jobbet med videreutvikling av integrerte modeller (**Figur 1**) og dataflytstruktur for å kunne levere data til naturindeksen med størst mulig geografisk dekning og høyest mulig oppløsning for planter og fugl.

Innenfor et hierarkisk modellrammeverk er det flere muligheter for modellering og integrering av data for planter. For karplanteindikatorerne i NI 2015 ble det brukt en hierarkisk modell med to nivåer. Det øverste nivået er en observasjonsmodell som beskriver observasjonsvariasjonen som en funksjon av innsamlingsinnsats og underliggende populasjonsstørrelse. Det underliggende nivået er prosessmodellen, som beskriver populasjonsdynamikk i 100x100 km koblete celler over hele Norge. Denne modellen gir geografisk heldekkende informasjon for alle karplanteindikatorerne over tid, men kan forbedres på flere punkter. For det første kan den relativt enkle prosessmodellen videreutvikles slik at den bedre reflekterer plantenes populasjonsdynamikk og spredningsevne. For det andre kan data fra overvåking (f.eks. TOV) og vegetasjons- og populasjonsstudier (diverse datakilder) utnyttes til å estimere parametere i populasjonsmodellene. For det tredje er det ønskelig å effektivisere estimeringsrutiner slik at beregningstiden går ned. Dette vil gjøre det mulig å kjøre modeller med romlig oppløsning under kommunenivå.

For fugl er betydelige mengder relevante data samlet inn i regi av overvåkningsprogrammet for hekkefugl (TOV-E). Som for planter vil vi imidlertid for enkelte fuglearter også kunne utnytte data fra andre kilder, inkludert de regulære TOV-områdene samt takseringsdata for lirype or-

ganisert via Hønsefuglportalen (<http://honsefugl.nina.no/>). For fugl er det langsiktige målet hierarkiske modeller, med en observasjonsmodell som eksplisitt modellerer variasjonen i observasjonssannsynlighet, og en prosessmodell som beskriver den underliggende populasjonsprosessen (**Figur 1**). Den underliggende geografiske oppløsningen vil gjøres så fin som mulig, men vil i praksis trolig variere noe mellom arter avhengig av omfang på tilgjengelig data. På lik linje med planteeksempelet baserer modellene seg på et grid-system med koblede celler. Arbeidet utføres i samarbeid med NINA sin strategiske instituttsatsing (SIS) INTEGRATE (<https://www.nina.no/V%C3%A5re-fagomr%C3%A5der/Strategiske-institutsatsinger/SIS-2016-2019/Integrate>), hvor vi arbeider med tilsvarende modeller, men for andre formål. Med denne koblingen vil man i NI-sammenheng kunne utnytte pågående modellutviklingsarbeid i NINA og dermed få en merverdi ved at NI får tilgang til fugleindikatorer med den høyeste geografiske oppløsningen som samtidig respekterer forventninger om presisjon og representativitet.

I denne rapporten omtales modelleringsarbeid for planter (kap 3). Resultater for fugl vil bli omtalt i en egen rapport (Nilsen m fl., under utarbeiding).



**Figur 1.** Eksempel på integrert hierarkisk modell for fugl, inkludert alle datakilder. Modellen består av prosessmodeller (stiplede bokser) som beskriver demografiske prosesser (rekruttering, overlevelse, mm) og romlige populasjons-prosesser (spredning mm.), og observasjonsmodeller (stiplede linjer) som beskriver hvordan ulike datainnsamlingsmetoder resulterer i henholdsvis demografiske data og data på forekomst og antall individer. Ved hjelp av disse modellene kan man slutte fra observerte data av ulike slag til mer presise og geografisk dekkende estimater av den underliggende prosessen (i NI-sammenheng: indikatorverdier) fordi man støtter seg på observasjoner fra flere ulike kilder og «filtrerer bort» observasjonsfeil med observasjonsmodellen.

## 1.2 Dataflytstruktur for automatisert oppdatering av naturindeksen

I første fase av prosjektet fokuserte vi på sammenstilling og strukturering av data, og programmering av funksjoner for dataflyt fra rådatakilder (inndata) via modellverktøy for integrerte modeller (i R) til NI-basen (utdata) (**Figur 2**).

Vi har satt opp et konseptuelt rammeverk for en automatisert dataflyt- og modelleringsprosess, som fylles med rutiner i form av skript og funksjoner i R (R Development Core Team 2016). Dette er samkjørt med NINAs utviklingsarbeid med naturindeksen under rammeavtalen med Miljødirektoratet (f.eks. Pedersen & Nybø 2015, Pedersen et al. under utarbeiding), men går betydelig videre: Under rammeavtalen utvikles R-pakken for beregning av NI ('Nlcalc', se kap 4), med standardiserte rutiner for dataimport på bestemte formater. I dette prosjektet har vi satt opp rutiner som både høster inndata fra diverse datakilder, formatterer for modellering (se kap 2), og leverer utdata på egnede formater til indikatorverdier i NI-basen og beregning av NI med 'Nlcalc'. På sikt vil en slik scriptbasert dataflyt legge til rette for automatisert oppdatering av modeller og indikatorverdier i naturindeksen når datasettene oppdateres.



**Figur 2.** Prinsippskisse for dataflytstrukturen. Integrerte hierarkiske modeller vil fungere som et viktig bindeledd i en automatisert prosess for å nyttiggjøre heterogene datakilder i naturindeksen, og vil gi geografisk dekkende informasjon med forbedret romlig oppløsning.

## 2 Datakilder og metoder for innhenting av data

Dette kapitlet presenterer noen viktige datakilder og deres egenskaper i form av datastruktur og muligheter for eksport/import, men gir ikke en uttømmende oversikt over alle mulige datakilder til naturindeksindikatorer. Vi fokuserer på datakilder som er relevante for modelleringsarbeid i dette og beslektede prosjekter (f.eks. NINAs strategiske instituttsatsing INTEGRATE). Datatypeene vi jobber med er imidlertid nokså generiske, og vi legger til rette for å kunne håndtere andre typer data ved å gjøre funksjonene for datahåndtering så modulære og fleksible som mulig. Det betyr at vi ikke tester alle tenkelige dataformater og porteringsløsninger, da det har blitt klart gjennom prosessen med prosjektet at det er ønskelig med en bredere nasjonal tilnærming til dataflyt (se kap 5).

### 2.1 Datakilder

For planter finnes det i hovedsak tre typer data som modellene kan støtte seg på: forekomster av arter/individer, data om populasjonsstørrelse/dekning og data om individer og demografisk struktur.

Data om forekomster av arter finnes i stort monn og er åpent tilgjengelig gjennom Artskart (<https://artskart.artsdatabanken.no/>) og GBIF (<https://www.gbif.org/>). Disse dataene har stor geografisk dekning over lang tid (fra før 1800), men angir som regel bare forekomst (ikke populasjonsstørrelse) og varierer i geografisk presisjon, særlig bakover i tid. I tillegg er det som regel ikke informasjon tilgjengelig som sier noe om studiedesign, hvor man lette etter arten uten å finne den, og andre faktorer som kan være viktige når man skal tolke dataene.

Data om populasjonsstørrelser er gjerne mer presist angitt geografisk, men er ofte bare engangsregistreringer (av og til gjentatt etter lang tid) eller korte tidsserier. Slike data er foreløpig mindre åpent tilgjengelig, men finnes for eksempel i NINAs databaser fra terrestrisk naturovervåking (TOV). Der foreligger mengdeangivelser for planter som frekvens- og dekningsdata (art-rute-matriser).

Demografiske rådata finnes også primært i enkelte forskeres og institusjoners databaser, og kan være vanskelig tilgjengelig i praksis, selv om et økende antall datasett og demografiske modeller publiseres åpent på nettet (f.eks. COMPADRE- og COMADRE-databasene med matrisemodeller for henholdsvis planter og dyr, <https://www.compadre-db.org>). Som et eksempel har vi tatt et datasett vi har tilgjengelig fra SeedClim-prosjektet (<https://www.uib.no/en/rg/EECRG/55395/seedclim>). Dette er en aksesdatabase som er tilgjengelig via databasespørringer i R, med individregistreringer for fire arter (fjellfiol *Viola biflora*, myrfiol *Viola palustris*, fjellveronika *Veronica alpina* og legeveronika *Veronica officinalis*). Ingen av disse artene er indikatorer i naturindeksen per i dag, men kan likevel tjene godt som eksempler. Vi har valgt å fokusere på en art, fjellfiol (*Viola biflora*), siden prinsippene for modellering og anvendelse av data vil være de samme for alle artene.

En rekke relevante miljødata er tilgjengelig nasjonalt, f.eks. gjennom Norge digitalt, og etter hvert i økologisk grunnkart, som nå er under etablering. Ved NINA er data om arealdekke, klima, geologi, etc. samlet fra flere kilder og tilrettelagt på et felles format, slik at det er enkelt å bruke og kombinere ulike data. I dette prosjektet vil vi derfor ikke bruke mye tid på miljødata, men viser nedenfor hvordan ulike observasjonsdata for arter kan kombineres med tilgjengelige miljødata, og klargjøres for modellering.

## 2.2 Metoder og skript for innlesing av data

En mye brukt tilnærming til dataimport i R er via standardfunksjonene for import av tekstfiler (funksjoner som 'read.table' o.l.), regneark (pakken 'readxl') eller klipp og lim av data ('clipboard'). Dette innebærer ofte manuelle grep i forbindelse med nedlasting av data, formattering til txt, csv eller liknende, og/eller selve importen til R. For vårt formål kan slik innlesing fungere dersom nedlastings- og bearbeidingsprosessen kan automatiseres. Det finnes også flere muligheter for import av data til R som er aktuelle for datatypene skissert ovenfor.

I R finnes en rekke pakker med rutiner for import av data på vektor- og rasterformat, for eksempel 'sp', 'rgrass7', og 'raster'. Funksjonen 'getData' i pakken 'raster' gir enkel tilgang til globale klimadatasett ('WorldClim'), med et delsett tilpasset biologiske anvendelser ('BioClim'; **Figur 3**).

Forekomstdata for arter er tilgjengelige gjennom nettsidene til Artskart og GBIF (**Figur 3**). Artsdatabanken tilgjengeliggjør i tillegg vektorfiler (shape) med rødliste- og fremmedartslistedata i en zip-fil på nettadressen: <http://kart.artsdatabanken.no/WMS/kartdata/artskart/artskart.zip>.

Import av data fra databaser (med data av ulike slag, f.eks. populasjonsstørrelser eller demografiske data) kan gjøres med databasespørringer. En enkel funksjon for datainnlasting fra en Access database kan se slik ut:

```
getAccessData <- function (db = "dataPath/myData.mdb")
{
  con <- odbcConnectAccess2007(db)
  dbQ <- "SELECT sites.siteID, ..."
  data <- sqlQuery(con,dbQ)
  close(con)
}
```

Denne funksjonen fungerer for en forhåndskjent struktur, som kan brukes til å formulere spørringen mot databasen (tekststrengen dbQ). Funksjonen kan gjøres mer generisk ved å plassere databasespørringen (linjene med dbQ og sqlQuery i funksjonen ovenfor) i egne scriptfiler, som kan tilpasses hver enkelt database. Dette kan også gjøres på ulike måter, for eksempel ved at spørrings-scriptene lagres lokalt:

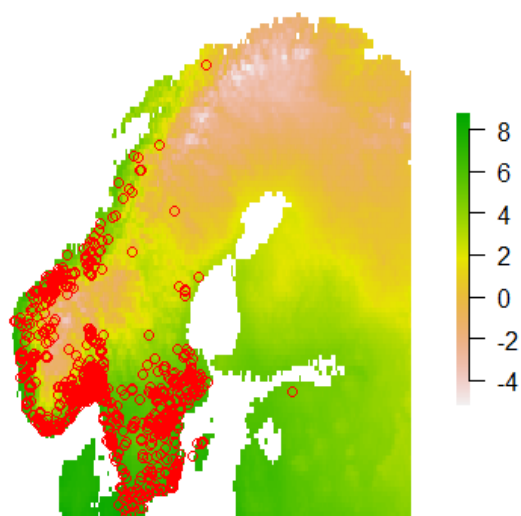
```
getAccessData <- function (db = "dataPath/myData.mdb", dbQueryFile = "loadData.r")
{
  con <- odbcConnectAccess2007(db)
  source(dbQueryFile)
  close(con)
}
```

Alternativt kan spørrings-filene lagres i en R-pakke:

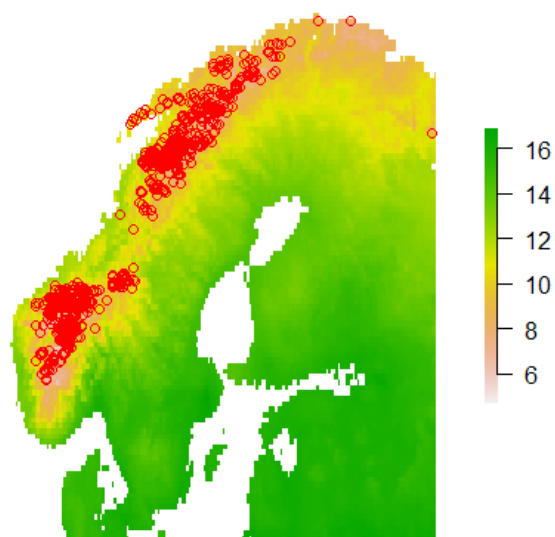
```
getAccessData <- function(db = "dataPath/myData.mdb",dbQueryFile = "loadData.r")
{
  con <- odbcConnectAccess2007(db)
  source(file.path(system.file(package = "MyPkg"), "fileFolder", dbQueryFile))
  close(con)
}
```

Det siste alternativet er mer fleksibelt dersom dataflytstrukturen skal flyttes mellom ulike plattformer og maskiner, men krever kontinuerlig vedlikehold og oppdatering av R-pakken, dersom nye datakilder skal tas i bruk.

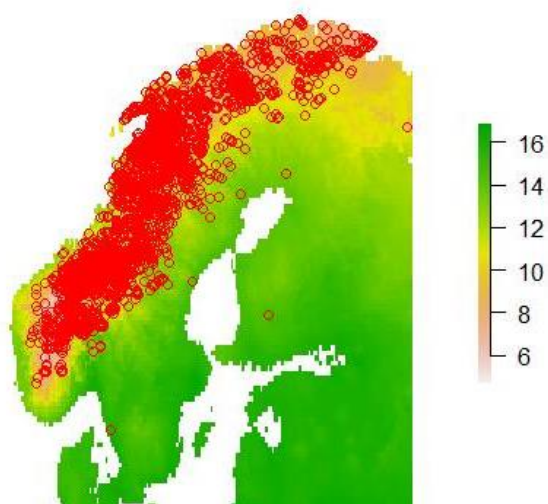
### Alm og årsmiddeltemperatur



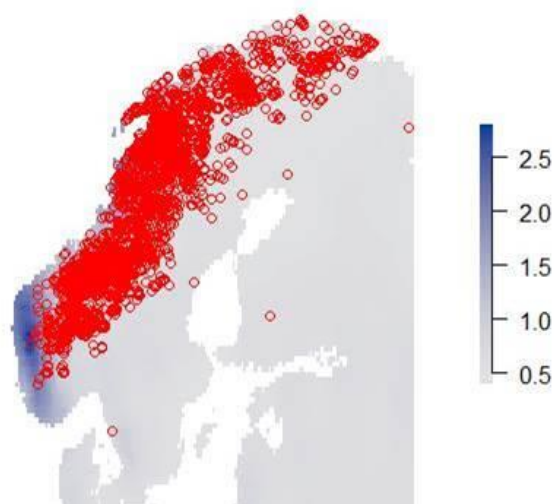
### Issoleie og sommertemperatur



### Fjellfiol og sommertemperatur



### Fjellfiol og årsnedbør



**Figur 3.** Eksempler på observasjonsdata (røde punkter) lastet ned fra GBIF med funksjonen `occ_search` i pakken `rgbif`, plottet på klimakart fra WorldClim, lastet ned med funksjonen `getData` i pakken `raster`. Sommertemperatur i °C og årsnedbør i meter.



### 3 Modellering av indikatorverdier

I dette kapitlet presenterer vi et eksempel på modellering som kan gi heldekkende indikatorverdier. For naturindeksen er modellering både i rom og tid relevant. Geografisk modellering kan gi mer heldekkende og høyoppløselige data for et gitt tidspunkt, men siden observasjoner ikke finnes for alle tidspunkter, vil ekstrapolering i tid også være nyttig, og noen ganger nødvendig (bl.a. for en del karplanteindikatorer hvor referansetilstanden er satt til året 1900). Her har vi derfor valgt å demonstrere metoden ved å modellere populasjonsutvikling over tid med høy romlig oppløsning.

Vi bruker arten fjellfiol (*Viola biflora*) som modellart. Fjellfiol er ikke en indikatorart for naturindeksen, men fungerer som et demonstrasjonseksempel på hvordan klimadata, populasjonsdata og forekomstdata kan integreres for å predikere forekomst og populasjonsstørrelse. Populasjonsdataene er fra fire lokaliteter mellom Hemsedalsfjellet og Vikafjellet (se SeedClim-prosjektet: <http://www.uib.no/en/rg/EECRG/55395/seedclim>) og dekker en nedbørsgradient fra 500 til 2000 mm/år. Dette gjør det mulig å modellere demografiske prosesser som en funksjon av nedbørsmengden og dermed projisere populasjonsvekstrater i rom og tid med bakgrunn i klimadata og -fremskrivninger. Vi har holdt oss til et avgrenset geografisk område (innenfor en kommune) for å spare regnetid, men metoden kan i prinsippet også brukes for andre romlige skalaer (utbredelsen til fjellfiol har også tidligere blitt modellert over et større område med utgangspunkt i metoder basert på forekomst-data; Meineri et al. 2012).

I dette kapitlet beskriver vi metoder og resultater fra modelleringsøvelsen, og i kap. 5.2 diskuterer og oppsummerer vi relevansen for naturindeksen.

#### 3.1 Demografiske data

Populasjonsdataene for fjellfiol ble samlet inn over en periode på åtte år fra 2009 til 2016 (dvs. syv overganger mellom år) på fire lokaliteter, tre alpine og én subalpin. De alpine lokalitetene har en sommergjennomsnittstemperatur på ca. 6°C og årsnedbør på henholdsvis 596, 1321 og 1925 mm, og den subalpine lokaliteten har en sommergjennomsnittstemperatur på 8,7°C og en årsnedbør på 1848 mm. I alle de fire lokalitetene ble det lagt ut tre analyseruter på 25 x 25 cm på tilfeldig vis, men med den ene betingelsen at de måtte inneholde fjellfiol. Hvert år i begynnelsen av august ble det notert skjebne (overjordisk tilstede eller ikke tilstede) samt vegetative og reproduktive mål for alle fjellfiolindivider i analyserutene.

De vegetative målene var antall blader og bredden av det største bladet, som vi brukte for å beregne individets biomasse (heretter 'størrelse') med bakgrunn i en biomasseformel fra Meineri et al. (2014).

De reproduktive målene var om en plante blomstret eller ikke i det respektive året og hvor mange blomster/kapsler den hadde. Antall frø per blomst/kapsel, spiringsfrekvensen, frøplante-etableringsraten og frøbankoverlevelse ble estimert i separate eksperimenter (se Töpper, 2015). Nyoppdagede individer ble inndelt i frøplanter og klonalt avkom etter størrelse og tilstedeværelse av frøblader. Siden klonal vekst skjer underjordisk var det ikke uten videre mulig å fastslå hvilke planter som var foreldrene til klonene. Derfor tildelte vi klonene det nærmeste etablerte individet (ikke frøplanter) som forelder. For alle overlevende planter fra året før ble det dermed notert om de hadde laget kloner eller ikke og hvor mange kloner det eventuelt var.

Fjellfiol er en art med forlenget hvile (vegetativ dormans), der individer kan overleve underjordisk uten å skyte overjordisk plantemateriale i opp til i hvert fall tre år (Evju et al. 2010). Modellering av dormansen er dermed ikke mulig for de første og siste tre overgangene i dette datasettet, bare for den midterste. Derfor valgte vi å behandle dormante individer som 'døde' og 'våkne' individer som nye kloner. Dette gjorde det mulig å jobbe med alle syv overgangene (se Olsen et al. 2016).

## 3.2 Populasjonsmodellering

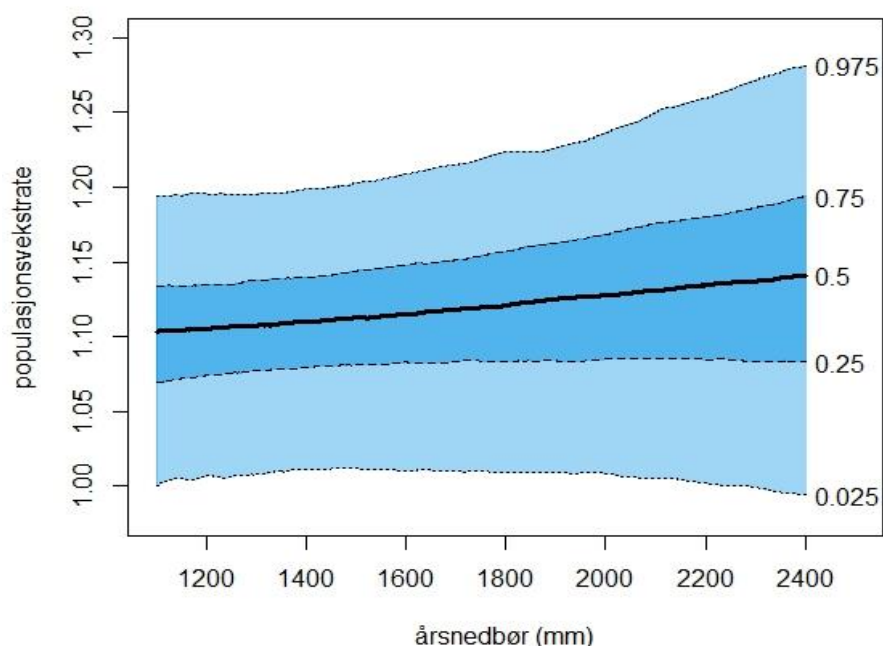
Vi brukte 'integral projection models' (IPM) som verktøy for populasjonsmodelleringen. Der blir populasjonsprosesser som overlevelse, vekst, klonal reproduksjon og seksuell reproduksjon først relatert til et mål for individenes utviklingsstadium (i vårt tilfelle er det individstørrelse) gjennom regresjon. Deretter blir regresjonene brukt til å bygge populasjonsmatriser som kan analyseres videre for blant annet vitale rater (overlevelse, vekst, reproduksjon) og populasjonsvekstrater, populasjonsvekstens følsomhet for underliggende vitale rater og miljøvariabler, med mer (Easterling et al. 2000, Merow et al. 2014, Caswell 2001).

Vi brukte 'generalized linear mixed models' (GLMM; Bates et al. 2015) i R (R Core Team 2016) for regresjonsanalysene, med størrelse og årsnedbør (normalverdi 1961-90) som fixed-effects forklaringsvariabel og lokalitet og plot som random effects. Modellene for vekst (altså størrelse i det påfølgende året) og klonstørrelse ble parametrisert med en normalfordeling, modellene for om individene var seksuelt/klonalt reproduktive eller ikke, ble parametrisert med en binomisk fordeling (og logit-link), og modellene for antall blomster/kapsler og antall kloner ble parametrisert med en poissonfordeling (og log-link). I binomiske modeller og poissonmodeller sjekket vi for 'overdispersion' og modellerte den der det var nødvendig med en 'random effect' på observasjonsnivået.

Regresjonene, samt konstanter for antall frø per kapsel, spiringsrate, frøplantetablering, frøbankoverlevelse og frøplantestørrelse, ble kombinert til populasjonsmatriser for forskjellige nedbørsscenarier ved bruk av R-pakken 'IPMpack' (Metcalf et al. 2013). De respektive populasjonsvekstratene ble estimert via matrisenes dominante egenverdier (Caswell 2001). For matrisebyggingen brukte vi en matrisedimensjon på 101 x 101, der den ytterste raden og kolonnen representerer frøbanken og de resterende 100 x 100 cellene den kontinuerlige delen av IPMen, og størrelsesgrenseverdier som var litt lavere og høyere enn henholdsvis min- og maksverdiene til individstørrelsene i datasettet. Problemer med 'individual eviction' (Williams et al. 2012) ble håndtert gjennom korreksjoner i matrisebyggingen (Metcalf et al. 2013).

Usikkerhet for populasjonsmatrisene og dermed populasjonsvekstratene ble beregnet ved å resample individene med tilbakelegging 1000 ganger (ikke-parametrisk bootstrapping). Dermed fikk vi 1000 verdier for hver regresjonskoeffisient som igjen gjør det mulig å lage 1000 varianter av populasjonsmatrisen og følgelig 1000 estimater av populasjonsvekstraten. Egenskaper ved fordelingen av disse estimatene (for eksempel kvartiler) brukes til å angi usikkerhet.

De underliggende regresjonene i populasjonsmodellen viste i utgangspunktet bare svake og statistisk ikke signifikante sammenhenger mellom de ulike populasjonsprosessene og årsnedbør. Sydd sammen til en populasjonsmodell har årsnedbør likevel en svak økende effekt på populasjonvekstraten (**Figur 4**),



**Figur 4.** Modellert sammenheng mellom årsnedbør og populasjonsvekstrate i fjellfiol. Figuren viser mediantrenden (0,5 kvantilen) samt 0,025, 0,25, 0,75 og 0,975 kvantilene av 1000 bootstrap-gjentak.

### 3.3 Framskriving av populasjonsstørrelse

Vi brukte BioClim-årsnedbørsdata fra WorldClim-databasen for å framskrive fjellfiol-populasjonsstørrelser i 1 x 1 km celler i Aurland kommune frem i tid basert på anslaget for årsnedbør i 2050. Ifølge artskart er fjellfiol godt utbredt i sub-alpine og alpine strøk i Aurland kommune og for illustrasjonseksempelet her antar vi derfor at den finnes i alle celler som oppfyller de klimatiske kravene som arten har (sommertemperatur 4,5 - 8,5°C, årsnedbør er høy nok i hele kommunen).

Vi laget lineære interpolasjoner av årsnedbør og sommertemperatur fra normalperioden 1961-90 (representert gjennom året 1975) til 2050-projeksjonen fra WorldClim for å få en årlig nedbørsverdi og sommertemperaturverdi for hver grid-celle i kommunen. Med utgangspunkt i en startpopulasjon (se nedenfor) kunne vi så beregne populasjonsstørrelsen i hver celle for hvert år  $t+1$  ved å multiplisere populasjonsstørrelse i år  $t$  med populasjonsvekstraten tilhørende nedbørsverdien i år  $t+1$ .

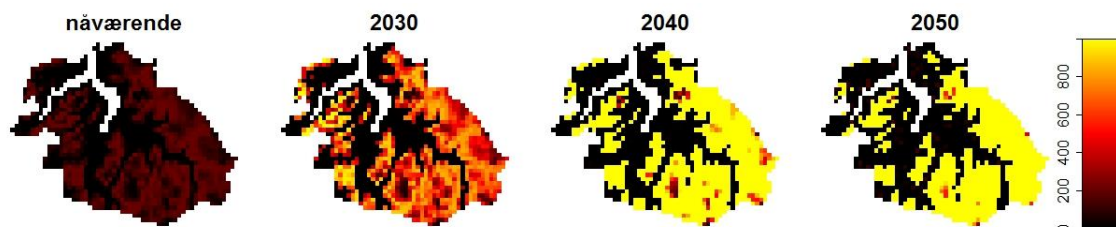
I tillegg korrigerer vi for samtidig økende temperatur i samme tidsforløpet siden oppvarmingen etter hvert vil gjøre det mulig for arten fjellfiol å kolonisere nye, tidligere for kalde, høyalpine områder mens den trolig vil forsvinne i noen områder ved dagens varmere utbredelsesgrense på grunn av økende konkurranse (Olsen et al. 2016). Startpopulasjonsstørrelsene baserer seg på antall individer per rute i populasjonsdatasettet, men ble justert etter cellenes sommertemperatur med bakgrunn i dagens klimatiske fordeling av arten (**Figur 3, Tabell 1**) vurdert på grunnlag av forekomstobservasjoner og -modeller (Artskart; Meineri et al. 2012). Med utgangspunkt i generell økologisk teori om artsresponser og artens forekomstrespons langs temperaturgradienten (Meineri et al. 2012), er det rimelig å anta at tetthetene vil ha et optimum nær midten av utbredelsen langs temperaturgradienten, og avta mot kantene (mer nøyaktige startpopulasjoner vil ofte kunne beregnes fra andre datasett - se diskusjon i kap. 5).

**Tabell 1.** Sommertemperaturavhengige startpopulasjoner av fjellfiol for populasjonsprosjisering i Aurland kommune. De angitte verdiene for N er gjennomsnittsverdier for temperaturspennet de gjelder for. Startpopulasjonene baseres på en lineær interpolasjon av disse gjennomsnittsverdiene.

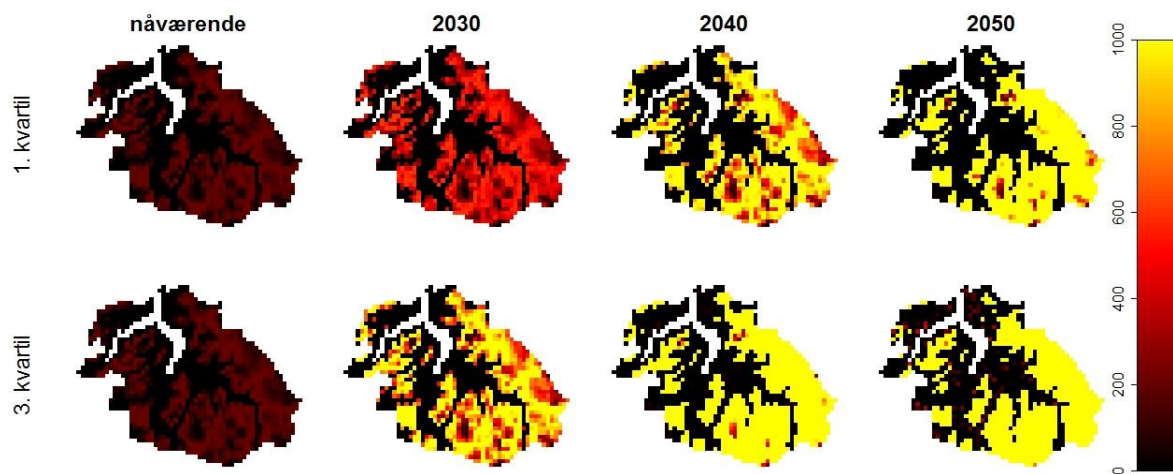
temperatur (°C)	< 4,5	4,5 - 5	5 - 5,5	5,5 - 6,5	6,5 - 7,5	7,5 - 8,5	> 8,5
N (# / m <sup>2</sup> )	0	50	100	200	100	50	0

Heldekkende kart på populasjonsstørrelse for fjellfiol i Aurland kommune ble laget for første, andre og tredje kvartil fra bootstrap-fordelingen av estimatene for dagens utbredelse, 2030, 2040 og 2050. Andre kvartil (medianen) samsvarer indikatorverdien i naturindeksen (vanligvis rapporteres gjennomsnitt, men her er de to nokså like), og første og tredje kvartil representerer nedre og øvre verdi i usikkerhetsintervallet som hittil har vært standard rapportering av usikkerhet for indikatorverdier til naturindeksen.

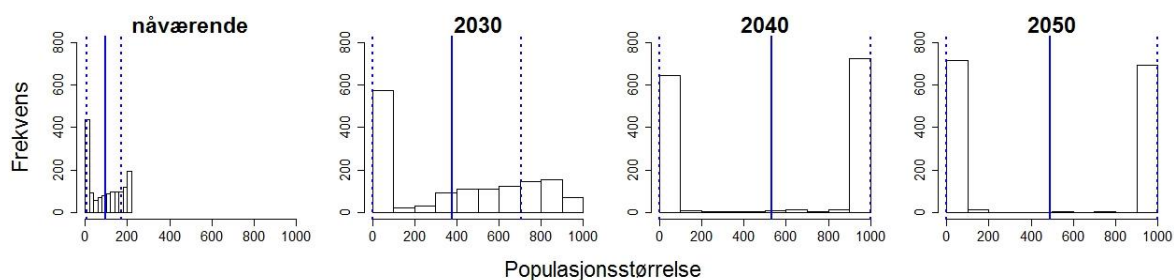
Selv om årsnedbør bare har en svak økende effekt på populasjonvekstraten (**Figur 4**) og nedbørsøkningen fram mot 2050 ikke er voldsom, ser vi store endringer i framskrivingene av populasjonstettheter (**Figur 5**). Dette er fordi populasjonsprosessen er eksponensiell. Variasjonen i framskrivingene innad i kommunen øker også med tiden (**Figur 6**), med divergerende fordelinger (**Figur 7**). Ved slutten av framskrivingen (2050) hadde enkeltceller enten totalt fravær eller forekomst i høy tetthet.



**Figur 5.** Framskrivning av populasjonsstørrelse (individer / m<sup>2</sup>) for fjellfiol i Aurland kommune frem til 2050 basert på 0,5 kvantilen (medianen) for nedbørsavhengig populasjonsvekstrate i **Figur 4**.



**Figur 6.** Framskrivning av populasjonsstørrelse (individer /  $m^2$ ) for fjellfiol i Aurland kommune frem til 2050 basert på 1. og 3. kvartil for nedbørsavhengig populasjonsvekstrate (0,25 og 0,75 kvantilene i **Figur 4**).



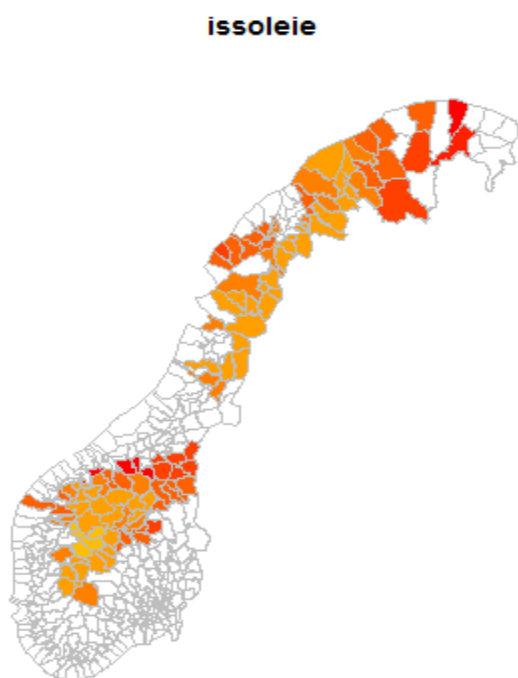
**Figur 7.** Fordeling av populasjonstettheter (individer /  $m^2$ ) i 1 x 1 km celler i Aurland kommune, i nåtid og ved tre framtidige tidspunkter (tilsvarende **Figur 5** og **6**).

## 4 Import av modellerte indikatorverdier til NI-basen

Beregningene av naturindeksen tar referanseverdier, indikatorverdier og usikkerheter som input-data (**Figur 2**), og gir indeksverdier som output (Pedersen & Nybø 2015). Selv om indeksen kan beregnes for et vilkårlig område og en vilkårlig skala, baseres beregningene alltid på en gitt minste romlige skala, en såkalt basisenhet. Input-data for en indikator kan imidlertid angis for større geografiske enheter enn basisenheten, for eksempel når man har modellert populasjoner eller fra en annen kilde har bestandsestimat for hele fylker eller andre forvaltningsenheter. For hver enkelt indikator må en derfor, før eller når data leses til databasen, spesifisere avgrensingen av «indikatorområdene» som det gis input-data for. Prinsippet med basisenheter innebærer imidlertid at hvert indikatorområde må bestå av en eller flere basisenheter. For Norge er basisenheten i skrivende stund kommuner. I de tilfeller input-data forekommer på en skala som ikke uten videre tilfredsstiller begrensingene for hvordan indikatorområder avgrenses, for eksempel fra modellprediksjoner i rasterformat (**Figur 5 og 6**), må disse re-skaleres f.eks. til basisenhet-skalaen (kommuner), før de kan importeres i naturindeksdatabasen.

Her kan modelleringen av karplanter for naturindeksen 2015 tjene som eksempel. Planteindikatorerne ble modellert i et grid med 100x100km celler over hele Norge. Disse resultatene på rasterformat ble så brukt til å beregne indikatorverdier for kommuner, som vist for issoleie i **Figur 8**. Rutinene som ble brukt for disse beregningene, kan også brukes for data med annen oppløsning, slik som fjellfioleksemplet (**Figur 5 og 6**).

Import av indikatorverdier til naturindeksbasen har tidligere blitt tillatt i kun ett format; et punkt-estimat for indikatorverdien, og en nedre og en øvre kvartil for usikkerheten i dette estimatet. Disse verdiene har deretter blitt tolket via en automatisk rutine i form av en sannsynlighetsfordeling, som har blitt brukt internt i beregningen av indeksen. Ved beregningen av Naturindeks 2020 vil imidlertid metodikken kunne akseptere indikatorverdier også i flere formater. Dette gir større fleksibilitet til å bruke modellerte indikatorverdier, og til å imøtekomme andre tilfeller der fordelingen til usikkerheten allerede er kjent.



**Figur 8.** Modellerte heldekkende data for issoleie til naturindeksen 2015 (lysere gulfarger angir høyere verdier). Modellen er kjørt med 100x100km oppløsning. Resultatene er beregnet for hver kommune som snittet av celler som kommunen overlapper med, veid med arealet av overlapp.



I tillegg til det tradisjonelle formatet på indikatorverdier, vil i skrivende stund også parameteriserte lognormalfordelinger, parameteriserte poissonfordelinger, diskrete fordelinger, og empiriske fordelinger aksepteres som indikatorverdier. Skalaen på indikatorverdiene er valgfri, så lenge de overensstemmer med referanseverdiene. Man kan altså angi indikatorverdier som for eksempel tall mellom null og en, eller som antall dyr i en populasjon. Der det angis egendefinerte fordelinger for indikatorverdiene vil disse brukes i beregningen av Naturindeks ved å trekke tilfeldige verdier direkte fra fordelingene. Dette i kontrast til indikatorverdier angitt som punktestimat og nedre og øvre kvartil, som først må gjennomgå et såkalt "elisiteringstrinn", der de tolkes om til en parametrisk fordeling.

Ved modellering av indikatorverdien ved bruk av vanlige, statistiske verktøy vil man ofte ønske å kunne angi indikatorverdien som en lognormalfordeling. I en lineær regresjon, der indikatorverdien  $I$  for basisenhet  $i$  modelleres som  $\log(I_i) \sim \mathcal{N}(\mu_i, \sigma^2)$  og  $\mu_i = \alpha + \beta_i * X_i$ . I dette tilfellet kan man benytte  $\mu_i$  og  $\sigma^2$  direkte som input-verdier til naturindeksdatabasen. Poissonfordelinger håndteres på tilsvarende måte, men her angis bare forventningsverdiene  $\lambda_i$ .

I andre tilfeller kan man ønske å angi en diskret fordeling, slik som når indikatorene kan ta et begrenset antall verdier med gitte sannsynligheter. Man kan for eksempel angi at det er 30% sannsynlighet for at indikatorverdien i en basisenhet er 0,2, og at det er 70% sannsynlighet for at verdien er 0,4. Slike tilfeller kan for eksempel opptre ved beregning av økologisk samfunnsindeks (Åström mfl. 2014). En slik diskret fordeling kan da gis direkte til naturindeksdatabasen.

Andre eksempler på empiriske fordelinger er simulerte, modellbaserte fordelinger som bootstrap-fordelingene i fellfioleksemplet (**Figur 7**) og posteriori-verdier fra en Bayesiansk modell. Selv om slike fordelinger iblant kan tilpasses med en parametrisert fordeling (for eksempel  $\log(I_i) \sim \mathcal{N}(\mu_i, \sigma^2)$ ), fungerer dette mindre bra for skjeve eller mer kompliserte fordelinger. Da kan man isteden gi hele, eller en representativ del av bootstrap- eller posteriori-verdiene som input til naturindeksdatabasen.

NINA har utviklet et API og en R-pakke `Nlcalc` med funksjoner for import og eksport av indikator- og indeks-verdier til og fra naturindeksdatabasen. Pakken er fritt tilgjengelig via Github (<https://github.com/NINAnor/Nlcalc>).

En typisk arbeidsflyt for import av verdier ved hjelp av denne pakken vil være:

1. Etablere en kobling til databasen, gjennom å bruke samme brukernavn og passord som brukes ellers for eksperter på innsynsløsningen til Naturindeks (<http://naturindeks.nina.no/>)
2. Liste indikatorene eksperter har tilgang til å redigere
3. Last ned nåværende verdier for indikatorene
4. Angi nye verdier for et gitt antall indikatorer, indikatorområder, og tidspunkter
5. Importere de oppdaterte verdiene til databasen.

Arbeidsflyten trenger ikke noen interaktivitet fra brukeren, og trinnene 1 til 5 kan dermed automatiseres helt i et skript. En gitt bruker kan kun se og oppdatere indikatorverdier som er knyttet til eksperter, men funksjonaliteten er ikke knyttet til noe spesifikt sted, men er tilgjengelig fra Internett. For videre informasjon av bruk av pakken `Nlcalc` henvises til dokumentasjonen i pakken.

## 5 Diskusjon

### 5.1 Automatisert dataflyt

På sikt er det ønskelig at naturindeksen skal kunne oppdateres ved automatisk nedlasting av data og modellering av indikatorverdier når nye data foreligger. Det er altså ønskelig med en mest mulig automatisert prosess, helst helt uten manuelle mellomledd. Det tar trolig en stund før dette er realisert fullt ut, men det kan likevel være nyttig å ha som visjon. Her vil vi diskutere noen erfaringer fra prosjektet vårt i forhold til denne visjonen.

Det første poenget vi vil ta for oss, er graden av standardisering for effektiv dataflyt. Automatisering krever standardisering, som igjen krever at man gjør noen valg i forhold til datahåndtering som kan legge begrensninger på analyse. Hvordan avveie standardisering mot fleksibilitet?

Det er neppe mulig å sette opp en dataflytstruktur som kan tjene alle formål. Her er formålet å øke datatilgangen til naturindeksen ved å legge til rette for integrert modellering med utgangspunkt i ulike, men bestemte former for biologiske data og miljødata. Det gir mulighet for standardisering, som letter automatiseringen av naturindeksen (og dermed reduserer kostnader og avlaster mange eksperter). Samtidig vil det antakelig være ønskelig for mange bidragsytere til naturindeksen å ta analysene videre i ulike former for forskning, og inkludere nye typer data og/eller modeller som vi ikke berører her. Det er det selvsagt ingenting i veien for, og på sikt kan slikt arbeid gi viktige bidrag til naturindeksen og nye standarder for datahåndtering. Men for naturindeksen er det viktig å etablere en felles plattform med dagens standarder for å kunne bringe mest mulig data sammen på en effektiv måte. Inntil videre er vi imidlertid usikre på i hvilken grad det hierarkiske modelleringsrammeverket vil la seg realisere for ulike typer indikatorer, selv om deler av det vil være relevant for mange. I modelleringen av planter har vi benyttet forekomstdata for å sette rammene for en mer detaljert populasjonsmodell basert på demografiske data. Vi har foreløpig ikke lyktes med å implementere dette i en Bayesiansk modell, men håper og tror at dette vil være mulig med ytterligere forskning på feltet. Fjellfiol og flere andre arter vil være gjenstand for slike studier de nærmeste årene. Sammen med pågående arbeid på fugl (Nilsen et al. under utarbeiding), og fleksibiliteten som bygges opp i Nlcalc (kap 4), håper vi at dette vil gi grunnlag for en enda mer standardisert dataflyt for modellering.

Dette bringer oss til et annet moment i forhold til effektiv dataflyt, nemlig organisering på tvers av institusjoner. I dette prosjektet har vi primært jobbet innenfor NINA, som har ansvaret for den faglige utviklingen av naturindeksen og betydelig relevant modelleringsaktivitet, men vi har også hatt diskusjoner med andre nasjonale miljøer, bl.a. på fagsamlinger for naturindeksen. I løpet av denne prosessen har det blitt klart at dataflytstrukturen for naturindeksen ikke bør utvikles i isolasjon fra annen nasjonal infrastruktur. Data vi trenger til modelleringen forvaltes av flere ulike miljøer og gjøres tilgjengelig gjennom store og tunge aktører som Artsdatabanken, GBIF og Norge digitalt. Disse jobber med dataflyt for mange formål. Naturindeksen bør koble seg opp mot dette arbeidet, og samkjøre med det som nå foregår i relaterte prosesser som utviklingen av rammeverket for økologisk tilstand (Nybø & Evju 2017) og økologisk grunnkart (<http://www.miljodirektoratet.no/no/Tema/Arter-og-naturtyper/Okologisk-grunnkart>). Selv om dette vil kreve mer tid og organisering, vil det på lengre sikt være flere fordeler med at vi etablerer nasjonale systemer og unngår å lage lokale strukturer. For det første, vil gjennomarbeidede nasjonale systemer antakelig være mer robuste enn lokale strukturer, som trolig oftere vil måtte revideres for å tilpasses nye rammebetingelser (f.eks. problemer med ulike dataformater og flytting mellom plattformer, etter som nye behov dukker opp). For det andre reduserer vi sjansene for unødvendig parallelt arbeid med tilrettelegging av data i flere institusjoner. For det tredje vil nasjonale systemer ha større muligheter for å legge til rette for portering av kode og redusere behovet for å flytte på datasett. Dette er stadig mer aktuelt ettersom datamengdene øker. Alt dette taler for at utfordringene med dataflyt til naturindeksen bør løses i tverrinstitusjonelle fora, i samsvar med nasjonal strategi for åpne forskningsdata (Anonym 2017).

## 5.2 Integrert modellering for bedre geografisk oppløsning og dekning

Modelleringsøvelsene med fjellfiol har gitt viktige erfaringer med tanke på modellering av naturindeksindikatorer mer generelt. Fjellfioleksempelen viser at detaljerte data og modeller kan gi høyoppløselige kart med potensielle indikatorverdier for naturindeksen, med tilhørende estimater av usikkerhet. Prisen er en betydelig innsats i modellering. Tidsbruken ved høyoppløselig modellering, særlig med modeller som inkluderer detaljerte populasjonsprosesser, kan være nokså omfattende. Ren regnetid for datamaskinen (en standard laptop) for fjellfiolanalysene i forhold til nedbørsframskrivingen i Aurland frem til 2050 var ca. 2 døgn med 1000 bootstrapgjentak. 10 000 gjentak hadde vært mer robust, men ville ha kostet nesten tre uker regnetid. Selv om slike demografiske modeller er komplekse og krever mye data og regnekraft, viser også andre studier at de gir mer detaljerte og pålitelige prediksjoner (Merow et al. 2011, Pagel & Schurr 2012, Schurr et al. 2012).

En fordel med prosessorienterte og hierarkiske modeller er at de synliggjør og eksplisitt behandler ulike lag av variasjon og usikkerhet. I fjellfioleksempelen er usikkerhet i de demografiske modellene estimert gjennom bootstrappingen, og usikkerhet for gjennomsnittlig indeks per kommune beregnet på grunnlag av romlig varierende populasjonsstørrelse i kommunen. Det er også usikkerhet i klimamålinger og -framskrivinger. Det er ikke håndtert her, men det er mulig å håndtere det på tilsvarende måte som demografisk usikkerhet. Den sammensatte usikkerheten som rapporteres til naturindeksen kan dermed dekomponeres. Det er en fordel, både for tolkningen av indikatorverdier og deres variasjon, men også for å kunne prioritere kunnskapsinnhenting der usikkerheten er størst.

Modellen fra fjellfioleksempelen kan opplagt forbedres på flere punkter. For det første er det klart at en eksponensiell populasjonsvekst ikke er realistisk over lengre tid. Det gir urealistiske lokale populasjonsstørrelser, og er en del av forklaringen på de divergerende tetthetsfordelingene (**Figur 7**). Tetthetsreguleringer er det derfor viktig å inkludere i en neste versjon av modellen, enten i modellstrukturen eller i framskrivingen. Maksimalpopulasjonen kan variere etter lokale forhold, og dette kan trolig modelleres. For det andre er startstørrelse for populasjonsframskriving ikke helt enkelt å fastslå presist. I analyserutene der det finnes data for arten, er populasjonsstørrelsene antakelig større enn i vilkårlige ruter. Dette kan selvsagt undersøkes i felt, enten ved nye feltstudier eller sammenligning med analyseruter fra andre studier, men for naturindeksen er det trolig mindre viktig, så lenge indikatorverdier skaleres med referanseverdier som måles/modelleres på samme skala. Et tredje poeng vi vil trekke fram er at gradientstudier ikke nødvendigvis alltid gir det beste grunnlaget for å spå om fremtiden. I fjellfioleksempelen har nedbørsgradienten svake og ikke-signifikante effekter på populasjonsprosessene (som allikevel gir utslag for populasjonsstørrelse), men samtidig vet vi fra eksperimentell klimamanipulasjon at økt nedbør fører til en sterk positiv effekt på fjellfiol (Tøpper 2015). Grunnen til at gradienten ikke gjenspeiler dette er nok at arten finnes lokalt i nisjer der miljøvariablene på mikroskala er gode. Dette er et relevant poeng for naturindeksen, fordi plantedata typisk ikke er eksperimentelle, men kun observasjoner som relateres til lokalt målte variabler, dvs. miljøgradienter.

Demografiske data og modeller er foreløpig ikke lett tilgjengelig for mange plantearter i Norge. Det finnes ikke noen formell, komplett oversikt, men vi kjenner til tidsserier, modeller og noen eksperimenter hos ulike institusjoner og enkeltforskere. De fleste av artene det gjelder er ikke indikatorer i naturindeksen per i dag, men en del av dem kan trolig fylle kriteriene (Nybø et al. 2015, Pedersen et al. 2018).

## 5.3 Oppsummering og anbefalinger

Vi har presentert et utvalg importrutiner til R (kap 2) med ulike fordeler og ulemper, men gitt at det vil være behov for ulike tilnærminger en stund framover, samt videre utvikling av tverrinstitusjonelt nasjonalt samarbeid om dataflyt, anbefaler vi å ikke «låse» rammeverket for dataflyt og import til naturindeksen. Arbeidet som er lagt ned i Nlcalc for å øke fleksibiliteten i dataimport (kap 4), er viktig for å øke datatilfanget, men full utnyttelse av data forutsetter også videre modellutvikling. Standardisering av dataflyt for automatisering av naturindeksen bør modnes

over lengre tid, i et bredt tverrinstitusjonelt samarbeid mellom sentrale dataaktører (f.eks. Artsdatabanken, GBIF og Norge digitalt), og må sees i sammenheng med beslektede prosesser og systemer i forvaltningen (f.eks. rammeverket for vurdering av økologisk tilstand og økologisk grunnkart).

Modelleringsøvelsene bak denne rapporten (kap 3), samt modelleringsarbeidet som ble lagt ned for karplanter i 2015, gir grunnlag for automatisering av modellering basert på populasjonsdata og forekomstdata, og dermed mulighet for å øke oppløsning og geografisk dekning, samt å øke antall indikatorer. Vi anbefaler derfor å vurdere indikatorsettet for planter på nytt fram mot neste hovedoppdatering av naturindeksen.

## 6 Referanser

Anonym 2017. Nasjonal strategi for tilgjengeliggjøring og deling av forskningsdata. Kunnskapsdepartementet. <https://www.regjeringen.no/no/dokumenter/nasjonal-strategi-for-tilgjengeliggjoring-og-deling-av-forskningsdata/id2582412/>.

Bates, D., Maechler, M., Bolker, B., & Walker, S. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1-48. doi:10.18637/jss.v067.i01.

Caswell, H. 2001. *Matrix population models: construction, analysis, and interpretation*. (2nd ed.). Sunderland, Massachusetts: Sinauer Associates.

Easterling, M. R., Ellner, S. P., & Dixon, P. M. 2000. Size-specific sensitivity: Applying a new structured population model. *Ecology*, 81(3), 694-708.

Evju, M., Halvorsen, R., Rydgren, K., Austrheim, G., & Mysterud, A. 2010. Interactions between local climate and grazing determine the population dynamics of the small herb *Viola biflora*. *Oecologia*, 163(4), 921-933. doi:10.1007/s00442-010-1637-x.

Meineri, E., Skarpaas, O., Vandvik, V. (2012) Modeling alpine plant distributions at the landscape scale: Do biotic interactions matter? *Ecological Modelling*, 231, 1-10.

Meineri, E., Skarpaas, O., Spindelböck, J., Bargmann, T., & Vandvik, V. 2014. Direct and size-dependent effects of climate on flowering performance in alpine and lowland herbaceous species. *Journal of Vegetation Science*, 25(1), 275-286. doi:10.1111/jvs.12062

Merow, C., Dahlgren, J. P., Metcalf, C. J. E., Childs, D. Z., Evans, M. E. K., Jongejans, E., . . . McMahon, S. M. 2014. Advancing population ecology with integral projection models: a practical guide. *Methods in Ecology and Evolution*, 5(2), 99-110.

Merow, C., LaFleur, N., Silander, J. A., Wilson, A. M., & Rubega, M. 2011. Developing Dynamic Mechanistic Species Distribution Models: Predicting Bird-Mediated Spread of Invasive Plants across Northeastern North America. *American Naturalist*, 178(1), 30-43. doi:10.1086/660295.

Metcalf, C. J. E., McMahon, S. M., Salguero-Gomez, R., & Jongejans, E. 2013. IPMpack: an R package for integral projection models. *Methods in Ecology and Evolution*, 4(2), 195-200. doi:10.1111/2041-210x.12001.

Nybø, S. & Evju, M. (red.) 2017. Fagsystem for vurdering av god økologisk tilstand. Forslag fra et ekspertråd. Ekspertrådet for økologisk tilstand, 247 s. <https://www.regjeringen.no/no/dokument/rapportar-og-planar/id438817/>

Nybø, S., Pedersen, B., Skarpaas, O., Aslaksen, I., Bjerke, J.W., Certain, G., Edvardsen, H., Framstad, E., Garnåsjordet, P.A., Granhus, A., Gundersen, H., Henriksen, S., Hovstad, K.A., Jelmert, A., McBride, M.M., Norderhaug, A., Ottersen, G., Oug, E., Pedersen, H.C., Schartau, A.K., Stora-unet, K.O. & van der Meeren, G.I. 2015. Økologisk rammeverk. I Pedersen, B. & Nybø, S. (red.). *Naturindeks for Norge 2015. Økologisk rammeverk, beregningsmetoder, data-lagring og nettbasert formidling*. NINA Rapport 1130. Norsk institutt for naturforskning. s. 11 – 20.

Olsen, S. L., Töpper, J. P., Skarpaas, O., Vandvik, V., & Klanderud, K. 2016. From facilitation to competition: temperature-driven shift in dominant plant interactions affects population dynamics in semi-natural grasslands. *Global Change Biology*, 22, 1915-1926. doi:10.1111/gcb.13241.

Pagel, J., & Schurr, F. M. 2012. Forecasting species ranges by statistical estimation of ecological niches and spatial population dynamics. *Global Ecology and Biogeography*, 21(2), 293-304. doi:10.1111/j.1466-8238.2011.00663.x.

Pedersen, B., Bjerke, J.W., Pedersen, H.C., Brandrud, T.E., Gjershaug, J.O., Hanssen, O., Lyngstad, A. & Øien, D.-I. 2018. Naturindeks for Norge – fjell og våtmark. Evaluering av eksisterende indikatorsett, dets datagrunnlag og behovet for ytterligere tilfang av datakilder. NINA Rapport 1462. Norsk institutt for naturforskning.

Pedersen, B., & Nybø, S. (red) 2015. Naturindeks for Norge 2015. Økologisk rammeverk, beregningsmetoder, datalagring og nettbasert formidling.

R Development Core Team. 2016. R: a language and environment for statistical computing. Vienna, Austria. Available from <http://cran.r-project.org>: R Foundation for Statistical Computing.

Schurr, F. M., Pagel, J., Cabral, J. S., Groeneveld, J., Bykova, O., O'Hara, R. B., . . . Zimmermann, N. E. 2012. How to understand species' niches and range dynamics: a demographic research agenda for biogeography. *Journal of Biogeography*, 39(12), 2146-2162. doi:10.1111/j.1365-2699.2012.02737.x

Töpper, J. P. 2015. Predicting and assessing climate-change impacts on the population dynamics of alpine and lowland forbs (PhD thesis), Universitetet i Bergen.

Williams, J.L., Miller, T.E.X. & Ellner, S.P. 2012 Avoiding unintentional eviction from integral projection models. *Ecology*, 93, 2008–2014.

Åström, S.C.H., Åström, J., Bøhn, K., Gjershaug, J.O., Staverløkk, A., & Ødegaard, F. 2014. Dagsommerfugler og humler som tilstandsindikatorer i Naturindeks for Norge. Oppsummering av aktiviteten i 2014. NINA Rapport 1098. Norsk institutt for naturforskning.





*Norsk institutt for naturforskning, NINA, er en uavhengig stiftelse som forsker på natur og samspillet natur–samfunn.*

*NINA ble etablert i 1988. Hovedkontoret er i Trondheim, med avdelingskontorer i Tromsø, Lillehammer, Bergen og Oslo. I tillegg driver NINA Sæterfjellet avlsstasjon for fjellrev på Oppdal, og forskningsstasjonen for vill laksefisk på lms i Rogaland.*

*NINAs virksomhet omfatter både forskning og utredning, miljøovervåking, rådgivning og evaluering. NINA har stor bredde i kompetanse og erfaring med både naturvitere og samfunnsvitere i staben. Vi har kunnskap om artene, naturtypene, samfunnets bruk av naturen og sammenhenger med de store drivkreftene i naturen.*

ISSN:1504-3312  
ISBN: 978-82-426-3299-9

## Norsk institutt for naturforskning

NINA Hovedkontor

Postadresse: Postboks 5685 Torgarden, 7485 Trondheim

Besøks-/leveringsadresse: Høgskoleringen 9, 7034 Trondheim

Telefon: 73 80 14 00, Telefaks: 73 80 14 01

E-post: [firmapost@nina.no](mailto:firmapost@nina.no)

Organisasjonsnummer 9500 37 687

<http://www.nina.no>



Samarbeid og kunnskap for framtidens miljøløsninger