

Mulig anvendelse af citizen science data til at bestemme ændringer i sandsynlighed for forekomst

Notat fra DCE - Nationalt Center for Miljø og Energi

Dato: 17. december 2015

Christian Damgaard & Morten Strandberg

Institut for Bioscience

Antal sider: 9

Faglig kommentering:
Rasmus Ejrnæs
Kvalitetssikring, centret:
Susanne Boutrup



AARHUS
UNIVERSITET

DCE - NATIONALT CENTER FOR MILJØ OG ENERGI

Tel.: +45 8715 0000
E-mail: dce@au.dk
<http://dce.au.dk>

Indhold

Formål	3
Baggrund	3
Metoder	4
Resultater	7
Diskussion	9
Referencer	9

Formål

At udvikle generiske statistiske metoder, der muliggør en sammenstilling af citizen science data af biodiversitet med overvågningsdatasæt af biodiversitet (fx NOVANA), hvor tilfældig stikprøvetagning er blevet anvendt. Potentielt vil en sådan sammenstilling gøre det muligt at kombinere informationen fra de forskellige datasæt og dermed opnå en højere statistisk styrke i test af fx mulig tilbagegang af en given art.

Baggrund

Antallet af biodiversitetsregistreringer indsamlet af borgere vha. software-applikationer på smartphones er hurtigt stigende. I Danmark, er der fx indsamlet mere end 1.000.000 observationer fra mere end 20.000 brugere med software-applikationen "Naturbasen", som er administreret af Danmarks Fugle og Natur (www.fugleognatur.dk).

Citizen science data er allerede i dag vigtigt input til vores viden om, hvor forskellige arter findes, hvilket eksempelvis bruges i myndighedernes rumlige prioritering af naturindsatser under landdistriktsprogrammet (Brunbjerg et al. 2016). Men da prøvetagningen ikke er tilfældig (Dickinson et al. 2010), er det ikke umiddelbart muligt at foretage en statistisk analyse af fx mulig tilbagegang af en given art.

Der er udviklet Bayesianske metoder med henblik på at sammenstille data fra en ikke-tilfældig prøvetagning med data fra en tilfældig prøvetagning (Elliot 2009). Hvis sådanne metoder bliver videreudviklet, vil de kunne anvendes til at sammenstille citizen science data og museumsoplysninger med overvågningsdatasæt af biodiversitet og dermed teste fx mulig tilbagegang i udbredelsen af en given art.

De udviklede metoder vil være generelt anvendelige for alle typer af biodiversitetsdata, men i dette projekt er det valgt at undersøge udviklingen i udvalgte relativt sjældne arter som typisk findes på tørre heder (EU-habitat kode: 4030). Disse relativt sjældne arter er overvåget i NOVANA (Nielsen et al. 2012), men er typisk ikke hyppige nok til at man med en rimelig statistisk styrke kan afgøre, om deres udbredelse henholdsvis stiger eller falder. Dette er i modsætning til mere hyppige arter, fx rynket rose, hvor man udelukkende ved hjælp af NOVANA data kan påvise signifikante ændringer i udbredelsen (Damgaard et al. 2011).

Et af de statistiske problemer med citizen science data er, at prøvetagningsindsatsen generelt er ukendt og varierer fra år til år, og man kan derfor ikke umiddelbart anvende antallet af observationer til at vurdere om der har været ændringer i udbredelsen over tid.

For at korrigere for prøvetagningsindsatsen vil vi i dette studie afprøve en metode, hvor man sammenholder observationerne af de relativt sjældne arter med observationer af hedelyng, som man med sikkerhed ved findes på de tørre heder.

Metoder

Udvalgte arter

De udvalgte relativt sjældne arter, som typisk findes på tørre heder er: guldblomme, djævelsbid, cypres-ulvefod, og lav skorsonér. Af disse er cypres-ulvefod på rødlisten som "moderat truet" art (endangered, EN). Arterne er udvalgt, fordi de bl.a. findes på heder, og her indikerer gunstig bevaringstilstand, samtidig er de så sjældne at de kun i ringe grad registreres i NOVANA overvågningen. Arterne er sandsynligvis truet af eutrofiering og muligvis også forsuring, hvilket kan føre til at de udkonkurreres på deres voksesteder. De er derfor afhængige af pleje af deres voksesteder, og samtidig kan ensartet intensiv pleje sandsynligvis også være en trussel.

Hedelyng - *Calluna vulgaris*
Tør hede.

Tør hede er kendetegnet ved at findes på sandet næringsfattig jordbund.

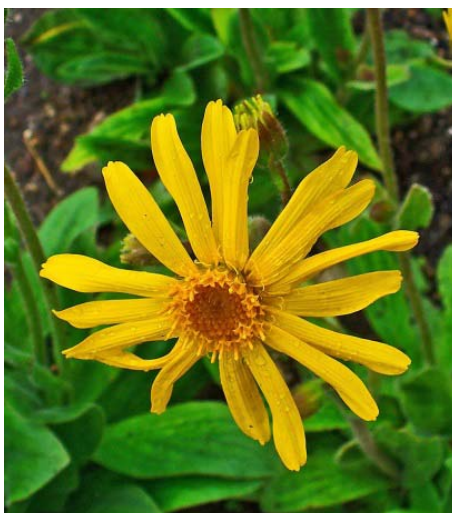
Foto: Williw Wikimedia commons.



Guldblomme – *Arnica montana*.

Vokser på heder, sure overdrev, enge og i egekrat.

Foto: H. Zell Wikimedia commons.



Djævelsbid – *Succisa pratensis*.

Vokser på moser, enge, overdrev og heder.

Foto: Hajotthu Wikimedia commons.



Cypres-ulvefod - *Diphasiastrum tristachyum*.

Vokser på lyncheder.

Foto: Bernd Haynold Wikimedia commons.



Lav skorsonér – *Scorzonera humilis*.

Vokser på sure overdrev og heder.

Foto: Algirdas_ Wikimedia commons.



Citizen science data

Observationsdata for de udvalgte arter samt hedelyng blev rekvireret fra Danmarks Fugle og Natur. De relevante datatyper er i denne sammenhæng (Dato, UTMYlokalitet, UTMXlokalitet, UTMYPrecision, UTMXPrecision, Antal), og af disse blev kun (Dato, UTMYlokalitet, UTMXlokalitet), som angiver observationsdagen samt lokalitetens UTM koordinat anvendt i den videre analyse. Det præcise UTM koordinat for fundstedet var ikke rapporteret i et tilstrækkeligt omfang til at det kunne bruges i denne analyse, og den kvalitati-

ve antalsangivelse vurderes til at være for upræcis til at være direkte anvendelig i en statistisk analyse.

For hver lokalitet siden 2007, hvor hedelyng var observeret, blev det opgjort, hvilke af de udvalgte sjældne arter som også blev observeret. Det resulterende datasæt antages at være binomialt fordelt og sandsynligheden for, at en af de sjældne arter blev observeret, varierer lineært med tiden,

$$y_{i,t} = \text{Bin}(n_{i,t}, \text{logit}(p_0 + p_t t)) \quad (1),$$

hvor $y_{i,t}$ er antallet af lokaliteter i år t hvor den sjældne art er observeret og $n_{i,t}$ er antallet af lokaliteter i år t hvor hedelyng er observeret.

Parametrene i model (1) blev estimeret ved hjælp af en numerisk Bayesiansk MCMC (Markov-chain Monte Carlo) metode og den marginale posterior fordeling af parameteren, som måler den gennemsnitlige ændring per år, p_t , blev summeret ved 2,5%, 50%, og 97,5% percentilerne, samt fordelings middelværdi og standardafvigelse.

NOVANA data

Ligeledes blev ændringen i sandsynligheden for at observere de samme udvalgte arter siden 2004 i NOVANA overvågningsprogrammet (Nielsen et al. 2012) modelleret i en tilsvarende model, hvor man dog tager hensyn til den hierarkiske dataindsamling (Damgaard et al. 2011). I dette tilfælde er de binomiale data dog ikke korrigeret i forhold til hedelyng, men derimod andelen af positive observationer af en af de udvalgte arter i tilfældigt udlagte og regelmæssigt genbesøgte (med GPS-usikkerhed) plots med en radius på fem meter. Disse plots er placeret indenfor en række stationer og kun plots, som blev klassificeret som naturtype 4030 og besøgt tre eller flere gange, blev anvendt i analysen. I alt blev der anvendt 3738 plot. Også i dette tilfælde blev den marginale posterior fordeling af parameteren, som måler den gennemsnitlige ændring per år, p_t , summeret ved 2,5%, 50%, og 97,5% percentilerne, samt fordelings middelværdi og standardafvigelse.

Betydningen af forskellige typer af indsamlingsadfærd

En usikkerhed ved anvendelsen af citizen science data fremkommer som følge af indsamlerens adfærd ved indsamling af data. Som allerede redegjort for er der ikke tale om en videnskabelig indsamlingsstrategi baseret på tilfældig udvælgelse af plots. Indsamlingen har tværtimod en overvægt på indsamling af sjældne arter, hvorved citizen science programmer ideelt set supplerer og komplementerer det videnskabelige indsamlingsprogram. Der er dog mulige fælder i anvendelsen af citizen science data, som skyldes de personlige mål, som deltagerne har. Blandt disse typer af personlige mål er:

1. Mange registreringer
2. Mange arter
3. Mange sjældne arter.

Alle tre typer medfører en overvægt af besøg på artsrige lokaliteter, og dermed også en tendens til, at borgerne lærer af hinanden og derfor besøger de samme lokaliteter, da dette opfylder det personlige mål for alle tre typer. Type 2 og type 3 vil i forfølgelsen af personlige mål have en tendens til en faldende aktivitet over tid, da personlige mål ikke tilfredsstilles af at registrere den samme art flere gange. For sjældne arter vil deltagere med personlige mål af både type 2 og 3

ofte indrapportere, når arter bliver fundet på nye lokaliteter. Endelig vil arter, der er vanskelige at bestemme, ofte være underrepræsenteret.

Til sammen betyder disse typer af indsamlingsadfærd, at der ved databehandlingen kan fremkomme mønstre, som mere reflekterer indsamlingsadfærd end tendenser i arternes forekomst. For eksempel vil type 2 og type 3 medføre, at rapporteringen af en given art fra en lokalitet vil aftage over tid. Dette kan umiddelbart ligne tilbagegang, hvis der ikke i databehandlingen tages højde for lokalitet, og selv hvis der tages højde for lokalitet, er det ikke muligt at skelne mellem tilbagegang og manglende rapportering.

Citizen science programmer kan forsøge at modvirke disse tendenser til kun at forfølge personlige mål ved at indføre incitamenter til at indrapportere flere arter, flere lokaliteter og fra steder med få indrapporteringer. Atlasundersøgelsen "Svampeatlas" (www.svampeatlas.dk), som forløb fra 2009 - 2013 forsøgte sig med en række af sådanne incitamenter, fx kunne man få udbredelseskort fra alle arter og kort over hvor mange arter, der var fundet i 10 *10 km², ligesom der var konkurrencer blandt deltagerne om det højeste antal registreringer, det højeste antal arter og det højeste antal rødlistede arter.

Resultater

Som forventet var hedelyng den hyppigst observerede af de udvalgte arter i Danmarks Fugle og Natur med mere end cirka dobbelt så mange observationer som for djævelsbid, den næst hyppigste af de udvalgte arter (Tabel 1). Der var i alt 1476 observationer af hedelyng, men en del af disse var redundante observationer fra den samme lokalitet og år og i alt 911 observationer (lokalitet * år) blev anvendt i analysen (Tabel 2).

Tabel 1. Antal observationer af udvalgte arter om sommeren (maj til september) i "Fugle og natur" fra 2007 til 2015, samt hvor mange af disse, som blev anvendt i analyserne (ikke redundante).

Art	Observationer	Anvendt i analyse
Hedelyng	1476	911
guldblomme	320	45
djævelsbid	763	116
cypres-ulvefod	39	0
lav skorsonér	327	44

Tabel 2. Antal lokaliteter med hedelyng i "Fugle og natur" fra 2007 til 2015 om sommeren (maj til september) i de forskellige år.

År	Antal lokaliteter med hedelyng
2007	57
2008	49
2009	99
2010	56
2011	70
2012	90
2013	98
2014	262
2015	130

Når ændringen i forekomst over tid sammenlignes med de to forskellige metoder med hver deres datasæt (Tabel 3), får man kvalitativt det samme billede for de tre hyppigste arter ved de to metoder. Ingen af de tre arter (guldblomme, djævelsbid og lav skorsonér) viser en signifikant ændring i forekomst over tid (Tabel 3).

Cypres-ulvefod blev ikke fundet i NOVANA overvågningens 3738 femmeter cirkler på tør hede, men blev observeret 39 gange i Danmarks Fugle og Natur (Tabel 1). Ligeledes var ingen af de 39 observationer af cypres-ulvefod fra lokaliteter, hvor der også var observeret hedelyng. Dette kan umiddelbart virke overraskende, men skyldes med stor sandsynlighed, at indsamlerne har større fokus på at indrapportere sjældne arter end almindelige arter. På Fugle og Naturs hjemmeside kan man se, at 5 ud af 12 af de billeder, der er taget af cypres-ulvefod, også har hedelyng med på billedet, på trods af at der er tale om nærbilleder fokuseret på ulvefoden.

Estimaterne af den årlige ændring i forekomst viser tilsyneladende en signifikant tilbagegang af cypres-ulvefod (Tabel 3; de lave estimerede værdier af den årlige ændring skal sammenholdes med den anvendte logit-transformation på den lave andel), men dette skyldes ikke en reel nedgang af forekomstsandsynligheden, men derimod en reducere af usikkerheden i løbet af årene hvor indrapporteringsindsatsen øgedes (Tabel 2).

Tabel 3. Percentilerne af den marginale posterior fordeling af parameteren, som måler den gennemsnitlige ændring per år, p_t , for citizen science data og NOVANA data.

Art	Citizen science data			NOVANA data		
	2,5%	50%	97,5%	2,5%	50%	97,5%
guldblomme	-0,0989	0,0085	0,1273	-0,3352	-0,0553	0,0803
djævelsbid	-0,0743	-0,0066	0,0574	-0,6571	-0,2407	0,0146
cypres-ulvefod	-49,3	-27,9	-13,6	-	-	-
lav skorsonér	-0,0424	0,0611	0,1826	-0,0259	0,0606	0,1429

For de tre hyppigste arter (guldblomme, djævelsbid, og lav skorsonér) er det principielt muligt at kombinere informationen fra de to metoder, for eksempel ved deres posterior fordelings middelværdi og standardafvigelse (Tabel 4) samt en vægtningsfaktor, men da det ikke er den samme type forekomst som måles, henholdsvis på en lokalitet og i en fem meter cirkel, er der valgt ikke at foretage en sådan kombination.

Tabel 4. Stikprøvestørrelse, samt middelværdi og standardafvigelse af den marginale posterior fordeling af parameteren, som måler den gennemsnitlige ændring per år, p_t , for citizen science data og NOVANA data.

Art	Citizen science data			NOVANA data		
	n	μ	σ	n	μ	σ
guldblomme	911	-0,0099	0,0580	3738	0,0125	0,0467
djævelsbid	911	-0,0071	0,0338	3738	-0,1085	0,0716
cypres-ulvefod	911	-30,57	12,75	3738	-	-
lav skorsonér	911	0,0639	0,0571	3738	0,0601	0,0429

Diskussion

For de tre hyppigste af de udvalgte arter støttede de to metoder hinanden i deres respektive konklusion og man kan derfor konkludere, at anvendelsen af citizen science data gav en øget kvalitativ sikkerhed af den konklusion, som man drog af NOVANA analysen. Dog er det vigtigt at understrege, at det ikke er den samme rumlige skala som undersøges; ved analyse af NOVANA data bestemmes ændringen i sandsynligheden for at arten er i en cirkel med radius fem meter, og ved analyse af citizen science data bestemmes ændringen i sandsynligheden for at arten er på en lokalitet af et ukendt areal.

Den relativt sjældne rødlistede cypres-ulvefod blev ikke fundet i NOVANA overvågningen eller i de lokaliteter, hvor der også var observeret hedelyng. Det er bemærkelsesværdigt, at ingen af de 39 observationer af cypres-ulvefod stammede fra lokaliteter, hvor der også var observeret hedelyng. Hvis der er en tendens til at sjældne planter kun observeres isoleret og ikke som en del af et plantesamfund, vil metoden ikke kunne anvendes på sjældne planter. Måske kunne app'en udvikles således, at ved indtastning af sjældne arter, fx rødlistede arter, opfordredes man til også at indtaste de mest almindelige arter på lokaliteten. Alternativt kunne man overveje andre metoder til at korrigere for variabel indsamlingsindsats.

Det er muligt at få yderligere information af eventuel tilbagegang i forekomstsandsynlighed ved at inddrage andre datatyper, fx ældre floralister, data fra det kommende Atlas Flora Danica, museumsfund, samt opfølgende feltundersøgelser.

Den anvendte metode kan supplere overvågningsdata på andre habitattyper, hvor der optræder karakteristiske arter, og dermed være en nem og billig måde at opnå yderligere information af om fx forekomsten af rødlistede plantearter går tilbage.

Referencer

Brunbjerg, A. K., Bladt, J., Brink, M., Fredshavn, J., Mikkelsen, P., Moeslund, J. E., Nygaard, B., Skov, F. & Ejrnæs, R. (2016): Development and implementation of a high nature value (HNV) farming indicator for Denmark. - *Ecological Indicators* 61, Part 2:274-281.

Damgaard, C., Nygaard, B., Ejrnæs, R. & Kollmann, J. (2011): State-space modeling indicates rapid invasion of an alien shrub in coastal dunes. - *Journal of Coastal Research* 27:595-599.

Dickinson, J. L., Zuckerberg, B. & Bonter, D. N. (2010): Citizen science as an ecological research tool: challenges and benefits. - *Annual Review of Ecology, Evolution, and Systematics* 41:149-172.

Elliot, M. R. (2009): Combining data from probability and non-probability samples using pseudo-weights. - *Survey Practice*.

Nielsen, K. E., Bak, J. L., Bruus Pedersen, M., Damgaard, C., Ejrnæs, R., Fredshavn, J. R., Nygaard, B., Skov, F., Strandberg, B. & Strandberg, M. (2012): NATURDATA.DK - Danish monitoring program of vegetation and chemical plant and soil data from non-forested terrestrial habitat types. - *Biodiversity & Ecology* 4:375.