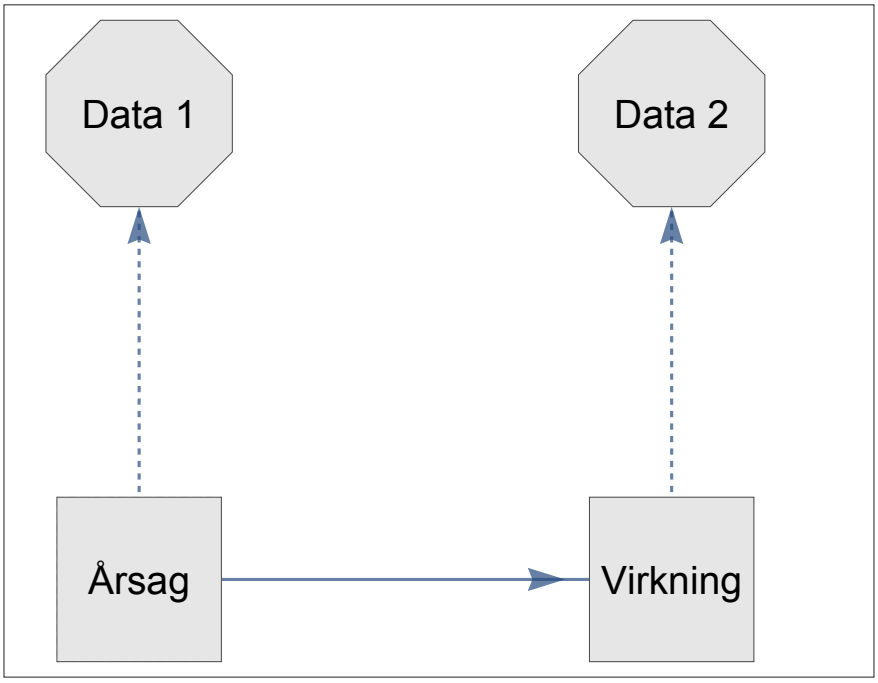
Miljømodellering og usikkerheder

Målte variable i mange anvendte miljømodeler har ofte en betydelig variation på grund af måle- og stikprøvefejl. Sådanne fejl kan resultere i upræcise prognoser, og dermed betyde at den rådgivning som gives med baggrund i modellerne kan være forkert.

Christian Damgaard

Forudsigelser baseret på økologi - og miljømodeller bruges ofte til rådgivning af det administrative og politiske system og spiller således en vigtig rolle i politiske beslutninger vedrørende miljøspørgsmål. I bestræbelsen på at udvikle realistiske modeller og lave så præcise forudsigelser som muligt er mange miljømodeller blevet relativt komplekse med mange inputvariabler og parametre. Dette fokus på realisme har i mange tilfælde ført til en nedprioritering af den statistiske modellering af de data, der er nødvendige for tilpasning (fitning) af modellerne til miljødata, og ofte er modellerne fittet med de observerede middelværdier af inputvariablerne. Dette er et problem, da de målte variable i mange anvendte miljømodeler har en betydelig variation på grund af måle- og stikprøvefejl, hvilket kan give betydelige bias (skævheder) af de estimerede parametre og forudsigelsesintervaller. Det er muligt at tage hensyn til effekten af måle- og stikprøvefejl ved hjælp af hierarkiske modelleringsmetoder (Fig. 1). Imidlertid er disse hierarkiske modelleringsmetoder kun sjældent anvendt i komplicerede miljømodeller.



Figur 1: Princippet i en hierarkisk model. A priori viden bruges til at formulere forskellige fænomener (årsag og virkning), som man vil undersøge, og arbejdshypoteser vedrørende årsag-virkningssammenhængen mellem fænomenerne formuleres i procesligningen (fuld pil). Fænomenerne knyttes til observerede data ved hjælp af måleligninger hvor måle- og stikprøveusikkerheden for de forskellige fænomener modellers (stiplede pile). Her vises et simpelt eksempel med kun to fænomener, men adskillige fænomener og processer kan på lignende måde knyttes sammen i et årsags-virkningsnetværk. Den samlede model fittes fx ved hjælp af Bayesiansk statistik, hvor hver enkelt pil i figuren betragtes som en betinget sandsynlighed . Selve fitningen forgår ved numeriske metoder (Damgaard 2020).

En vigtig antagelse i standard statistisk modellering er at de forklarende variabler er målt uden usikkerhed. Men antag at der er en måle- og stikprøvefejl i stikprøve i, z_i, således at den observerede forklarende variabel er, $x_{obs,i} = x_i + z_i$, hvor x_i er den sande værdi. Derudover antager vi at måle- og stikprøvefejlen ikke har nogen bias og er normalfordelt, $z_i \sim N(0,\tau)$, hvor τ er standardafvigelsen at målefejlen. Resultatet af denne enkle målefejlsmodel er velkendt (fx Carroll et al. 2006), men det kan være illustrativt at tage udgangspunkt i et konkret eksempel, og tusind datasæt ($x_{obs,i}$, y_i), blev derfor simuleret i en simpel lineær model, $y_i = x_i + e_i$, hvor residual variationen (den variation som ikke er forklaret af modellen) antages at være normal fordelt $e_i \sim N(0,\sigma)$ med σ som residual standardafvigelse (Damgaard 2020). Som forventet afhænger den estimerede hældning og skæring med y-aksen tydeligt af størrelsen af målefejlen, hvilket igen betyder at målefejl kan lede til forkerte modelforudsigelser (Fig. 2, Tabel 1). Hvis målefejlen var dobbelt så stor som residualfejlen, var det gennemsnitlige forudsigelsesinterval 7,17 - 8,96, og enkeltværdiforudsigelsesintervallet var 4,40 - 11,74, hvilket skal sammenlignes med 10 som er den forventede værdi (Tabel 1).

Til sammenligning blev datasæt, som blev genereret under den samme målefejlmodel som ovenfor, fittet ved hjælp af hierarkiske modeller (Damgaard 2020). Konklusionen af disse simuleringerne er at det er vigtigt at reducere måle- og stikprøvefejl blandt de forklarende variable ved gentagen prøvetagning, samt at det er muligt at tage hensyn til effekten af målefejl ved at modellere denne i hierarkiske modeller (Figur 1).

For at tilstræbe præcision og realisme er

mange økologi - og miljømodeller komplekse med mange inputvariabler og parametre. Det er dog vigtigt at denne modelkompleksitet ikke står i vejen for en grundig statistisk modellering af data. Som vist ovenfor kan dette føre til forkerte forudsigelser. Heldigvis er der på det seneste udviklet en række softwareløsninger så det er blevet nemmere at anvende hierarkiske modeller til at fitte selv komplicerede modeller.

Litteratur

Carroll, R. J., D. Ruppert, L. A. Stefanski, and C. Crainiceanu. 2006. Measurement Error in Nonlinear Models: A Modern Perspective. 2 edition. CRC.

Damgaard, C. 2020. Measurement uncertainty in ecological and environmental models. Trends in Ecology & Evolution. (https://doi.org/10.1016/j.tree.2020.07.003)

Christian Damgaard, Bioscience, Aarhus Universitet, Vejlsøvej 25, 8600 Silkeborg, Denmark. cfd@bios.au.dk

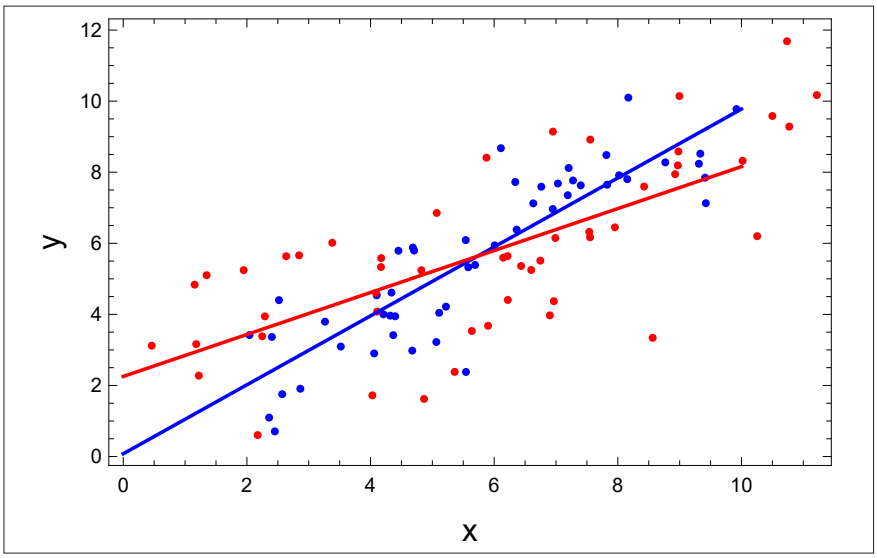


Fig. 2. To realisationer af datamodellen ved forskellige niveauer af måle- og stikprøvefejl, blå: τ=0, rød: τ =2. Linjerne er de fittede linjer. Prøvestørrelse = 50, Residual standardafvigelse (σ) = 1.

Tabel 1. Gennemsnitlige parameterestimater og forudsigelsesintervaller fra 1000 realiseringer af datamodelen ved forskellige niveauer af måle- og samplingsfejl (τ). Prøvestørrelse = 50, residual standardafvigelse σ = 1. De nedre og øvre 95% forudsigelsesintervaller, for både gennemsnittet og enkeltværdier blev foretaget for χ =10 med en forventet værdi på 10.

τ	Skæring	Hældning	Varians	Nedre genm.	Øvre genm.	Nedre enkeltv.	Øvre enkeltv.
0	-0.016±0.013	1.0031±0.0021	0.998±0.006	9.386±0.010	10.644±0.010	7.920±0.012	12.110±0.012
1	0.966±0.016	0.8315±0.0025	1.823±0.012	8.489±0.011	10.072±0.013	6.466±0.015	12.095±0.015
2	2.635±0.014	0.5433±0.0022	3.157±0.020	7.171±0.011	8.964±0.013	4.400±0.017	11.735±0.015
3	3.789±0.013	0.3410±0.0021	4.094±0.022	6.311±0.011	8.086±0.012	3.048±0.018	11.348±0.013

Om diverse konferencer

Flere af årets konferencer er udsat til 2021 på grund af COVID-19.