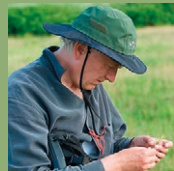


Udvekslingen af CO₂ og metan mellem vandet og luften måles med små sensorer i et flydekammer på et vandløb. Foto: Kenneth Thorø Martinsen.

Forfatterne



Kaj Sand-Jensen er professor ved Ferskvandsbiologisk Sektion, Biologisk Institut, Københavns Universitet. Han forsker i planters biodiversitet og økofysiologi og systemøkologi i søer og vandløb. ksandjensen@bio.ku.dk



Kenneth Thorø Martinsen er postdoc ved Ferskvandsbiologisk Sektion, Biologisk Institut, Københavns Universitet. Han forsker i fisks biodiversitet og kulstof- og iltomsætning i søer og vandløb og opskalering af CO₂- og CH₄-udledning for større regioner. kenneth.martinsen@bio.ku.dk



Theis Kragh er lektor ved Biologisk Institut, Syddansk Universitet. Han forsker i kulstof- og fosforomsætning i søer samt fiskeøkologi. theis.kragh@biology.sdu

Forfatterne støttes lige nu af Danmarks Frie Forskningsfond i studier af udledning og omsætning af drivhusgasser fra søer, vandløb og vådområder.

INTELLIGENTE COMPUTERSYSTEMER I ØKOLOGIEN

Kunstig intelligens finder hastigt stigende anvendelse i videnskaben. Vi har benyttet maskinlæring til at forudsige vandkvaliteten i Danmarks mere end 180.000 små og store søer samt CO₂-indholdet i Skandinaviens vandløb.

Overalt i fysik, kemi og biologi, hvor der optræder komplekse relationer og data, er intelligente computerprogrammer nødvendige. Mennesker kan nemlig ikke overskue mylderet og opdage de mønstre, som computeren registrerer.

I astrofysik må millioner fotos fra teleskoper af stjerner i fjerne galakser tygges igennem af avancerede computere for at se tilbage til universets tidlige historie. I kemien bygger computerne modeller af enzymernes 3-D struktur for at forstå deres evne til at katalysere stoffers reaktion med hinanden. I biologien benyttes DNA-profiler fra nuclevende og uddøde arter til at udlede slægtskab og udviklingslinjer, hvilket har

ført til gennemgribende ændringer i opfattelsen af evolutionen.

Blandt planterne har man for eksempel udledt, at den store familie af kurvblomster med 25-35 tusinde nuclevende arter oprindeligt opstod i Sydamerika ved slutningen af Kridttiden cirka 85 millioner år før nu. Siden flyttede centret for artsdannelse midt i Tertiærtiden, cirka 45 millioner år før nu, til lysåbne, halvtørre egne i Sydafrika, hvor 90% af de nuværende udviklingslinjer opstod. Senere udviklede kurvblomsterne sig yderligere i Asien, hvorfra de spredte sig til Europa.

I modsætning til den indsigt, man har opnået i evolutionen som i eksemplet ovenfor, har man ikke

flyttet sig tilsvarende i forhold til at beskrive og forstå de fleksible relationer i økosystemernes fødenet. Opstilling af økologiske modeller for den samlede fotosyntese og respiration i søer, kilder og koralrev har ellers en glørværdig baggrund tilbage fra amerikanerne Howard Odum and Eugene Odum i 1950'erne.

Stofstrømmene gennem økosystemernes komplicerede fødenet kan måles ved tilsætning af uorganiske eller organiske stoffer mærket med stabile isotoper af kulstof og kvælstof, så deres veje gennem organismer i fødenettene følges over tid. Enkelte sådanne eksperimenter findes, men er udført over kort tid i plaatsække i søer og kystvande. Gennembruddene

Avanceret billedgenkendelse i hverdag og videnskab

Intelligente computersystemer har gjort sit indtog i vores dagligdag i en grad, så vi hurtigt vænner os til nye vilde anvendelser uden at tænke nærmere over det. En af de anvendelser, de fleste har stiftet bekendtskab med, er billedgenkendelse. For eksempel tilbyder mange smartphones ansigtsgenkendelse af ejermænd, så selv det svage lys fra skærmen i buldermørke automatisk låser telefonen op. Samtidigt træder ejeren ud af skyggen og optræder på fotos på nettet fra fortiden. Fællesbilleder fra en afslutning med sangforeningen Morgenrøden, en skoleafslutning og en glemt vild aften med klikken for 20 år siden.

Indenfor biologien er genkendelse af svampe- og plantearter på billeder eller fuglearter ud fra lyd populære anvendelser af avanceret computerteknologi. Den professionelle biologs eksklusive kompetence til artsgenkendelse bliver efterhånden tilgængelig for alle, mens de vanskelige og sjældne tilfælde dog stadig må overlades til ekspertens vurdering og tjek af artens specifikke DNA i store databaser. Vi har også selv benyttet billedgenkendelse i økologisk forskning, hvor 160 fangede og mærkede gedder



Øverst: Billederne viser en gedde fanget i november 2017 og genfanget i maj 2019. Nederst: Til sammenligning er vist et andet individ med tydeligt anderledes mønster.

i Mjels Sø i Sønderjylland fik en selvstændig eksistens i en samlet estimeret bestand i søen på 560 individer. Fotogenkendelse bygger på individernes unikke mønster af bestående lyse striber og pletter på

kroppen. Geddefiskere kan derved bidrage til afdækning af individernes adfærd og udvikling, når de fotograferer, vejer og genudsætter fiskene og sender informationen til hovedcomputeren.

kommer måske i fremtiden ved at anvende teknikken i små søer eller velafgrænsede naturlige økosystemer på land. Men det forudsætter, at der etableres den nødvendige økonomi og ekspertise til store eksperimenter og håndtering af mange data, der dækker fra sæsoner til år.

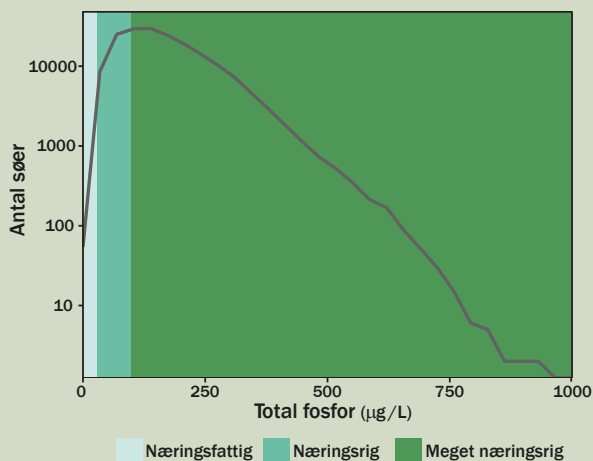
Åbne datakilder og maskinlæring i systemøkologi

Vi har i vores forskning benyttet mængden af tilgængelig information om søerne og deres omgivelser for at kunne forudsige vandkvaliteten i Danmarks mere end 180.000 kortlagte søer. De omfatter hele spændet fra utallige små søer på få kvadratmeter til få store søer på flere kvadratkilometer. Vi har afgræn-

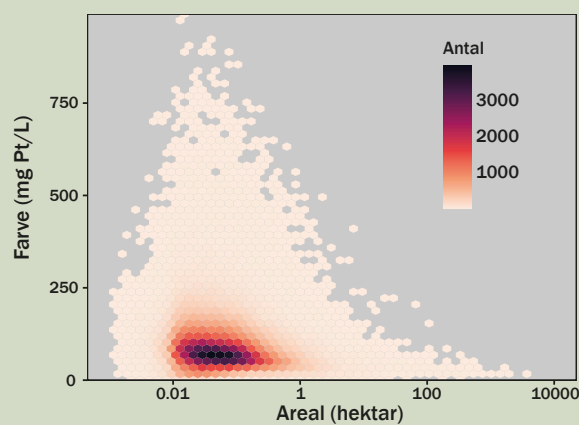
set søernes oplande og benyttet åbne datakilder til at karakterisere klima, jordtype, arealanvendelse og højdekurver i både oplandet og zoner af 50-250 meters bredde, som følger søernes kystlinje. Endvidere indgår søernes areal og form, og fra højdekurver kan vi beskrive terrænets stejlhed, krumning og uregelmæssighed.

For omkring 1000 søer eksisterer gode data over de seneste 20 år til at beregne den gennemsnitlige vandkvalitet på baggrund af søvandets alkalinitet, fosfor, kvælstof, planktonalger, humusstoffer samt pH og sigtddybe (vandets klarhed). Fra pH, vandtemperatur og alkalinitet kan vi beregne vandets CO₂-indhold.

Sådanne sammenstillinger af data fra forskellige kilder resulterer hurtigt i komplekse datasæt. Med mere end 50 forklarende variable for hver af de 8 vandkvalitetsparametre er det nemt at overse vigtige relationer på grund af ikke-lineære sammenhænge og interaktioner mellem forklarende variable. Derfor har vi benyttet 9 forskellige maskinlæringsmodeller, der automatisk kan tilnærme de komplekse sammenhænge mellem søernes vandkvalitet og de forklarende variable. Modellerne varierer i deres fleksibilitet til at beskrive disse sammenhænge. Målet er at træne modeller, der kan generalisere og ramme plet på nye data. Systemerne søger at finde frem til rette svar, mens de lærer af fejtagelser og succes på et



Indhold af total-fosfor for Danmarks mere end 180.000 søer forudsagt med en "Random Forest" model. Kun en lille andel (0,4 %) af søerne kan betegnes som næringsfattige (0-30 µg P/L), mens 24,6 % og 75 % kan betegnes som henholdsvis næringsrige (30-100 µg P/L) og meget næringsrige (>100 µg P/L).



Relationen mellem vandets farve (y-akse) og søareal (x-akse). Farve er et udtryk for vandets brunfarvning med humusstof. Især talrige små søer, særligt skovsøer, kan være meget gul- og brunfarvede, mens det sjældent gør sig gældende for de få store søer.

trænings sæt med 80 % af de 1000 søer; de resterende 20 % anvendes derefter til at vurdere modellernes evne til forudsigelse.

Den bedst fungerende model

"Random Forest" hedder den model, som fungerede bedst i de fleste af vores tilfælde. Den består af mange modeller, såkaldte regressions-træer. Enkeltvis er disse træer relativt dårlige modeller, men når de samles – i en skov – så kan de danne grundlag for stærke modeller. Ved at træne mange regressions-træer på en tilfældig andel af observationerne og udvalg af de forklarende variable, beskriver de enkelte træer forskellige mønstre i data og bliver derved bedre samlet. Tilsvarende, hvis enkeltpersoner skal gætte antallet af bolde i et glas, rammer de sjældent rigtigt, mens gennemsnittet af alle gæt ofte er markant bedre. Det samme gør sig gældende for sådanne "ensemble"-modeller.

For søerne kan de trænede modeller forklare mellem 28 og 60 % af variationen i de forskellige vandkvalitetsforhold alene ud fra oplandets karakter og arealudnyttelse samt søernes areal og form. Modellen forbedres og forklarer 45-78 % af variationen i de enkelte vandkvalitetsparametre ved at træne nye modeller med forudsigelserne af de øvrige vandkvalitetsparametre.

Hvad viser resultaterne? Som ventet steg eutrofieringen – udtrykt ved indholdet af fosfor, kvælstof og planktonalger – med øget andel af landbrug i oplandet. Søerne i de kalkfattige sandområder i Vestjylland havde som ventet lavere alkalinitet og pH end søerne i Østjylland og på Øerne.

Det var imidlertid bemærkelsesværdigt og helt nyt, at de nærmeste omgivers stejlehed, krumning og uregelmæssighed omkring søen markant påvirkede søers indhold af næringsstoffer og alger samt vandets klarhed. Forklaringen er sandsynligvis, at terrænets stejlehed øger tilførslen til søen af næringsrige jordpartikler ved vind- og vanderosion og tillige giver en hurtigere og mere direkte tilførsel af jordvand til søen, så færre næringsstoffer fjernes ved processer i jorden under vandets passage.

Endvidere viste modellerne, at små søer påvirkes mest af tilførsel af humusstof fra omgivelserne, idet randkontakten mellem land og vand er stor, og fortyndingen af stofferne ude i søens lille vandvolumen er begrænset. Så små søer er gennemgående mere brunfarvede af humusstoffer, og de er rigere på CO₂ fra bakteriers nedbrydning af humusstofferne sammenlignet med store søer.

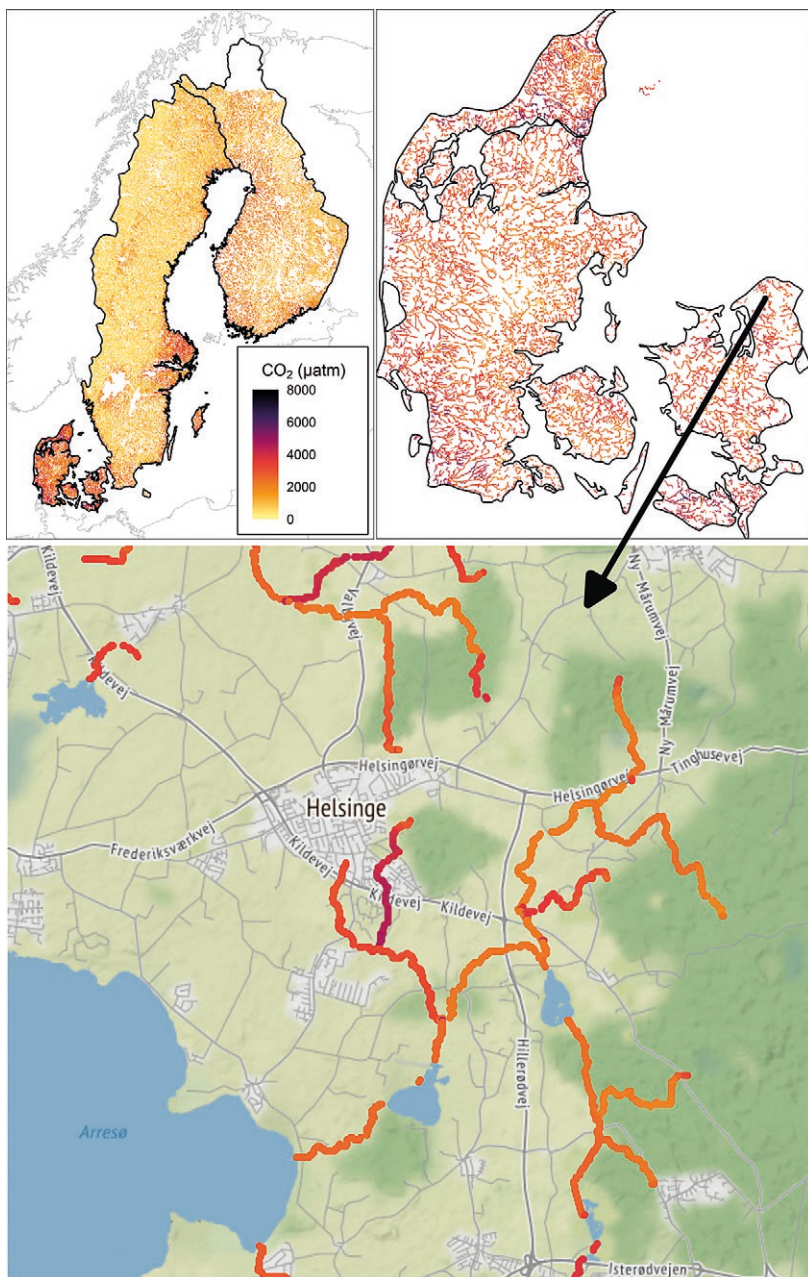
Forudsigelser af vandkvalitet

Et afgørende udbytte af arbejdet er naturligvis, at vi kan benytte modellerne til at forudsige vandkvaliteten i alle de mange danske søer, der ikke findes målinger fra, men hvis beliggenhed, omgivers karakter, areal og form, vi kender fra tilgængelige arealdata. Vi kan derfor forudsige det ukendte uden at have direkte målinger på vandet.

Med flere fremtidige data kan vi sandsynligvis forbedre modellerne yderligere; især ved at inddrage meget små søer, som er underrepræsenteret i materialet. Vi er særligt interesserede i at benytte forudsigelserne af CO₂-indholdet til at beregne afgangningen fra søerne til luften. Det kræver kendskab til den fysiske udvekslingshastighed for CO₂ mellem vandet og luften, der styres af overfladevandets turbulens. Beregning af afgangningen kan direkte tjekkes med de flydekamre med indbyggede automatiske CO₂-sensorer, som vi har udviklet. For modelforudsigelser bør gå hånd i hånd med direkte målinger.

CO₂ i Skandinaviens vandløb

Vi har benyttet samme type frit tilgængelige data om oplandets jordbund, geologi, arealudnyttelse og terrænforhold samt vandløbsnetværkets forløb til at forudsige CO₂-indholdet i vandløbene i Danmark, Sverige og Finland. Her



CO₂-indholdet i vandløb i Danmark, Sverige og Finland (øverst, venstre), alene for Danmark (øverst, højre) og i området omkring Helsingør i Nordsjælland. Frit tilgængelige data og maskinlæringsmodeller giver mulighed for gode forudsigelser af vandkemi i vandløb på både nationalt og lokalt niveau. Danske vandløb har højere CO₂-indhold end svenske og finske. Øvre vandløb er mere CO₂-rige end nedre vandløb i de fleste vandløb, her illustreret omkring Helsingør.

gav Random Forest-modellen, trænet og testet på 2298 vandløbslokaliteter, også de bedste forudsigelser. Den kunne forklare 66 % af variationen i vandløbenes CO₂ i Skandinavien. Efterfølgende blev modellen anvendt på i alt 268.000 km vandløb med bredder over cirka 2,5 m. Fra CO₂-indholdet og vandløbets fysiske forhold beregnede vi også CO₂-afgasning fra vandløbet til atmosfæren.

Vandløbene modtager vand fra de omgivende jorde, og vandets CO₂-indhold varierede som ventet fra lokalitet til lokalitet. Vandløbene var gennemgående overmættet med CO₂ i forhold til ligevægt med atmosfæren. Som gennemsnit var vandet overmættet 7 gange i Danmark, hvor temperaturerne er højere, jordlagene tykkere og mere næringsrige og bakterierne frigivelse af CO₂ ved nedbrydning af det organiske stof derfor er stor. Vandløbene er i

gennemsnit overmættet 3 gange i Finland og 2 gange i Sverige; i Sverige indgår vandløb, som udspringer i fjeldene, hvor jordlagene er tynde og bakterierne nedbrydning lav. CO₂-afgasningen til luften var i gennemsnit 3 gange større fra danske vandløb sammenlignet med finske og svenske; den fysiske udvekslingskonstant var nemlig lavere i de danske vandløb med langsom strøm sammenlignet med de finske og især de svenske med stort fald og hurtig strøm fra højderne.

Vi har nu de første tal for vandløbenes CO₂-afgasning for alle tre lande. Hvad mangler vi så for at anvende tallene i landenes samlede CO₂-regnskab? Dels skal direkte målinger af afgasningen verificere beregningerne, hvilket kan ske med allerede tilgængelige data og benyttelse af de omtalte flydekamre. Derudover er der usikkerhed om det samlede vandløbsareal, blandt andet på grund af sæsonvariation. Store vandløb kan ses fra satellitter, mens data mangler for små vandløb. Her kræves manuelle målinger i terrænet eller overflyvning med droner for at levere de nødvendige data.

Hvad fremtiden bringer?

Forskellige intelligente computerværktøjer vil i stigende grad blive bragt i spil for at opskalere data for udledning og omsætning af drivhusgasser i atmosfæren til nationalt og globalt niveau. Meteorologer og atmosfærefysikere anvender allerede sådanne avancerede modeller i stort omfang. Vi ser også åbenlyse muligheder i fremtidige studier af omsætningsprocesser blandt planter og mikroorganismer i jorden, som er tilsvarende komplekse ude i den dynamiske natur, men som i forskningen i dag især måles under velregulerede, ikke-dynamiske forhold i laboratoriet. Her mener vi, at der kan være et stort potentiale for nye indsigter ved at flytte forskningen ud af laboratoriet og bringe nye målemetoder i spil under relevante naturlige forhold, hvorefter de komplekse data kan analyseres med intelligente computerværktøjer i lighed med danske søers vandkvalitet. ■

Supplerende læsning
Martinsen, K. T. & Sand-Jensen, K. (2022). Predicting water quality from geospatial lake, catchment, and buffer zone characteristics in temperate lowland lakes. *Science of the Total Environment*, 851, 158090.

Martinsen, K. T., Kragh, T., & Sand-Jensen, K. (2020). Carbon dioxide partial pressure and emission throughout the Scandinavian stream network. *Global Biogeochemical Cycles*, 34(12), e2020GB006703.

Kristensen, E., Sand-Jensen, K., Martinsen, K. T., Madsen-Østerby, M., & Kragh, T. (2020). *Fingerprinting pike: The use of image recognition to identify individual pikes*. *Fisheries Research*, 229, 105622.

Jensen, M. B., Bahnson, C. H., Nasrollahi, K. & Moeslund, T. B. (2018). *Deep learning – et gennembrud inden for kunstig intelligens*. *Aktuel Naturvidenskab*, 2, 8-13.

Sand-Jensen, K., Riis, T., Kjær, J. E. & Martinsen, K. T. (2022). *High and variable carbon dioxide concentrations and emissions in lowland streams*. *Journal of Geophysical Research: Biogeosciences* (submitted).

Sand-Jensen, K. & Schou, J. C. (2021). *Kurveblomstfamilien – vi er mest til åbne vidder*. *Urt*, 1, 9-12.