At få computeren til at hjælpe os

Henrik Kragh Sørensen   Mikkel Willum Johansen

20. februar 2023

Henrik Kragh Sørensen og Mikkel Willum Johansen (apr. 2022). „Invitation til de datalogiske fags videnskabsteori“. Lærebog til brug for undervisning ved Institut for Naturfagenes Didaktik, Københavns Universitet. Under udarbejdelse. Kapitel 0, version ().

Indhold

1 Matematiske modeller og modellering

Vi er omgivet af og benytter os konstant af forskellige former for modeller for at navigere i verden. Men for at forstå, hvad modeller er, og hvad de kan, er det nødvendigt at kigge lidt mere filosofisk på selve modelbegrebet og på den slags viden, som modeller muliggør.

En model er generelt en form for repræsentation af et genstandsfelt (domæne, det modellen er en model *af*), hvor visse aspekter er udvalgt frem for andre til repræsentationen. Tænk på et kort over metroen i København: Sådan et metrokort er lavet for at det er let at orientere sig mellem stationerne og se, hvor man kan skifte mellem linje M1 og M3. Men metrokortet er fx ikke målfast, så der kan sagtens være længere at gå mellem Lufthavnen og Kastrup end mellem Rådhuspladsen og Hovedbanegården, selvom begge par er nabostationer på kortet. Altså ser vi, at metrokortet er fokuseret på et formål og derfor har prioriteret visse informationer frem for andre. Og de to forhold er generelt gældende for modeller.

Modeller er kendetegnet ved deres *epistemiske funktion*. Derved adskiller de sig fra andre repræsentationer: En modeljernbane er ikke en model — med mindre vi kan lære noget ud fra den. Solsystemet er ikke en model af atomkernen — men kan tjene som en effektiv analogi. Men en skaleret arkitekt-model af et højhus kan godt være en model, hvis man fx kan undersøge lysindfald ved hjælp af den — og den er samtidig en visualisering, som man kan danne æstetiske indtryk fra.

I videnskaben taler man om forskellige modeller: Mus kan være modeller i kræftforskning, fx hvis musenes organer på vigtige punkter ligner menneskers. Eller rotter kan være modeller for menneskelig adfærd, hvis man kan tillægge rotterne egenskaber som stedsans. Når vi taler om modeller i videnskaben handler det altså om, at modellen *på en eller anden måde* repræsenterer det, som modellen skal være en model af.

En anden måde at forsøge at indskærpe, hvad en (videnskabelig) model er, er ved at forsøge at dele det komplekse begreb op i underbegreber eller eksempler. Vi har allerede omtalt eksempler på skalerede, fysiske modeller og på analogier, der ikke har model-karakter.

I det følgende skal vi behandle flere typer modeller: matematiske modeller, beregningsmodeller og modeller til simulering. Nogle af dem er ret simple, men har alligevel interessante erkendelsesmæssige udfordringer. Andre er meget komplekse, og det rejser blot endnu flere spørgsmål at stille — og forsøge at besvare.

Overordnet kan man inddele modeller i forskellige typer alt efter, hvordan modellens forbindelse til dens domæne er: Vi taler således om 1. *ikoniske modeller*, hvis modellen *ligner* dens domæne, men fx er nedskaleret i størrelse, 2. *analogi-modeller*, hvis model og domæne ikke ligner hinanden, men alligevel deler en form for metaforisk analogi, og 3. abstrakte eller *matematiske modeller*, hvis modellen er en abstrakt og formel repræsentation af sit domæne. Ikoniske modeller kan være enormt brugbare i konstruktioner, hvor man fx kan bygge en model af et hus for at kunne visualisere lysforhold, inden man faktisk begynder at lægge mursten oven på hinanden. Og modellen virker, fordi der er *lighed* mellem modellens dele og domænets dele: modellen *står i stedet for* sit domæne i alle relevante sammenhænge. I forlængelse af denne definition kan man også betragte *virtuelle* modeller (VR) som ikoniske modeller, selvom der ikke foreligger nogen materiel model men derimod en computer-medieret simulation af at bevæge sig rundt i modellen.

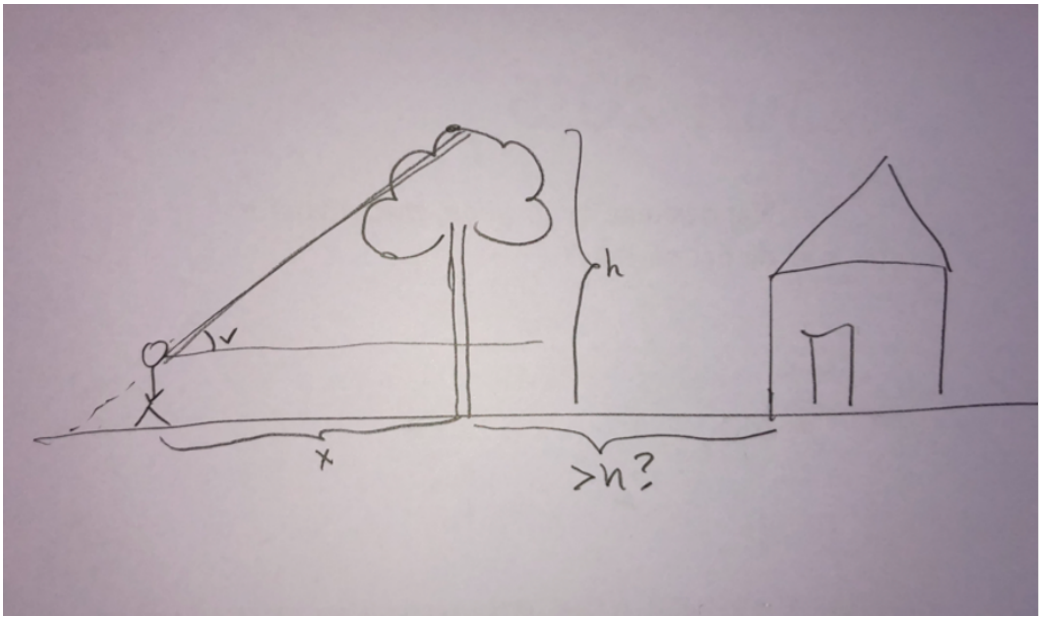
Anderledes ser det ud for analogi-modeller, som fx Niels Bohrs (1885–1962) atommodel, som er baseret på en analogi med vores solsystem: atomkernen er i midten, ligesom solen, og elektronerne kredser i baner om kernen, ligesom planeterne gør om solen. Og analogien rækker videre, således at der ud af modellen næsten følger, at der må være en kraft, der holder atomet sammen, ligesom tyngdekraften holder solsystemet sammen. Denne slags modeller kan være heuristiske 'pumper', der kan give ideer, men det er sværere — uden yderligere antagelser — at argumentere for, at de altid virker og fx kan bruges som årsagsforklaringer.

Endelig er der de matematiske modeller, som vi skal beskæftige os mest med her. De matematiske modeller repræsenterer et 'virkeligt' fænomen (i den bredeste betydning) ved en formel beskrivelse, og logiske relationer i det formelle system skal så svare til faktiske relationer i virkeligheden, hvis modellen ellers er konstrueret ordentligt. På den måde kan man bruge matematiske modeller både til forklaringer og til at kontrollere og forudsige udsnit af virkeligheden.

2 Et legetøjseksempel

Lad os betragte et simpelt eksempel. Lad os forestille os, at jeg skal fælde et træ i min indkørsel, og det naturligt nok er vigtigt for mig, at træet ikke vælter ind i mit hus. Vi identificerer hurtigt, at dette problem kan analyseres ved hjælp af en matematisk model. Og som noget af det første foretager vi os nogle abstraktioner: Vi ser bort fra træets grene, vi er ligeglade med, at huset er et træhus, og garagen på den anden side af huset betyder ikke noget for vores model. Nogle af de ting, vi abstraherer væk, kunne godt være relevante for vores problem, men vi vurderer alligevel, at det er nemmest at se bort fra dem — i hvert fald lige nu. Dermed er vi i stand til at opstille en foreløbig skitse af vores model: træet er skitseret, mit betragtningspunkt og husets placering er angivet, men der er endnu ikke sat matematiske begreber og størrelser på modellen. Men hvis vi betegner afstanden mellem mig og træet med x, højden af træet med h og sigtevinklen til træets top med v, så kan vi udtrykke en relation mellem størrelserne som fx $h = x\tan(v)$, og det gør det muligt for mig at beregne h ud fra x og v, som jeg kan måle (se figur 1).

Henrik Kragh Sørensen og Mikkel Willum Johansen (apr. 2022). „Invitation til de datalogiske fags videnskabsteori". Lærebog til brug for undervisning ved Institut for Naturfagenes Didaktik, Københavns Universitet. Under udarbejdelse. Kapitel 0, version ().



Figur 1: En skitse af den matematiske model, som skal hjælpe mig til at fælde mit træ.

Dermed kunne vi se ud til at være færdige — jeg kan beregne træets højde og sikre mig, at det ikke rammer huset, når jeg fælder det.

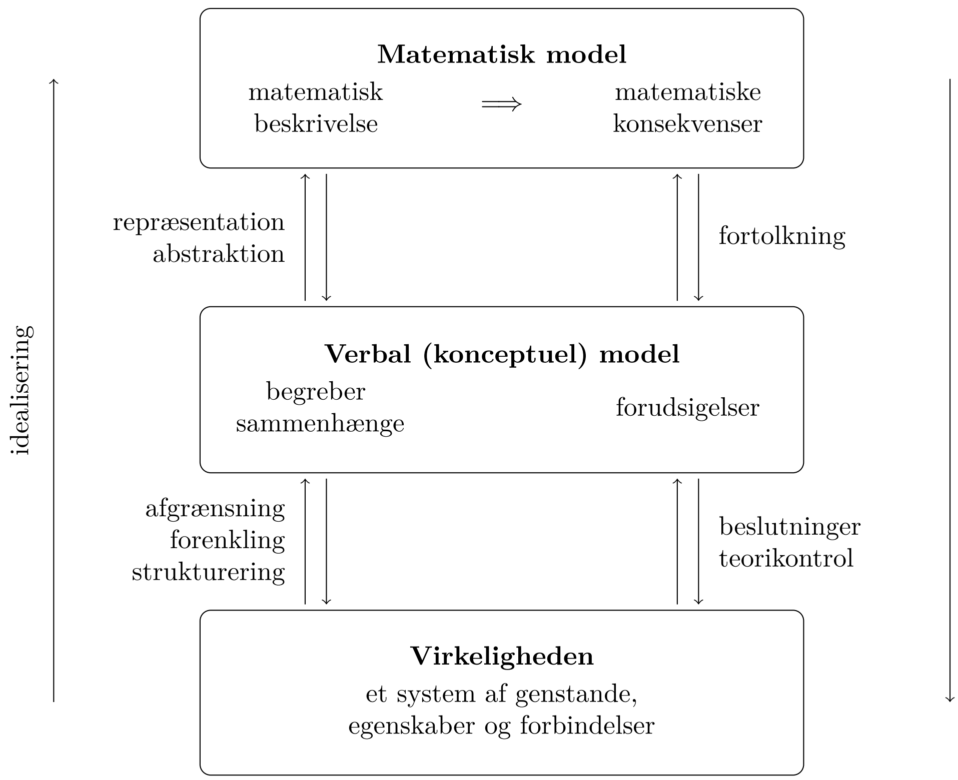
Men det kan være, at det kommer meget tæt på, og jeg bliver bekymret. En grund til bekymring kunne være, at jeg ikke er sikker på, hvor præcise mine målinger er. Så jeg bliver nødt til at sige noget om usikkerheden på min beregning af h. Det kan jeg fx gøre ved at se, hvad der sker, hvis jeg lægger lidt til eller trækker lidt fra x. På den måde kan vi bruge den matematiske model til at regne forskellige scenarier igennem (se nedenfor), hvilket kan være meget effektivt som en slags eksperimenter i situationer, hvor vi ikke bare kan foretage et fysisk eksperiment.

Men der er også nogle idealiseringer, som vi har foretaget i selve modellen, som kan have endog stor indflydelse på modellens forudsigelser. I vores konkrete skematiske form er der taget hensyn til, at jeg er to meter høj, hvorfor vinklen v faktisk skulle måles et andet sted. Og vi har antaget, at træet faktisk vokser vinkelret op, ligesom vi slet ikke har taget hensyn til, at jordens overflade jo er krum. Dette er idealiseringer fordi vi faktisk godt ved, at de er kontrafaktiske antagelser — men de har hjulpet os til at bygge modellen. Hvis det nu viser sig, at modellen er utilstrækkelig af en eller anden grund, så må vi jo genbesøge vores idealiseringer og se, om vi kan forbedre dem.

Selvom dette er et legetøjseksempel, viser det nogle af de principielle udfordringer, som er nødvendige at reflektere over for at lave en faglig-kritisk vurdering af matematiske modeller: Hvilke antagelser er bygget ind i modellen? Hvilke faktiske forhold er der set bort fra? Hvilke matematiske og evt. statistiske begreber og beregninger er involveret? Hvad betyder det for modellens usikkerheder? Hvordan fortolkes og formidles modellens resultater?

3 Modelleringsprocessen

Man kan illustrere de væsentligste dele af den matematiske modelleringsproces som i figur 2, hvor et udsnit af virkeligheden afgrænses til en verbal model, der *matematiseres* til en matematisk beskrivelse. Denne venstre side af figuren handler generelt om *idealisering* ved afgrænsning og abstraktion. Den matematiske beskrivelse vil typisk have form af et system af ligninger eller differentialligninger, og man ønsker at kunne udlede matematiske konsekvenser af beskrivelsen, helst i form af en *eksakt* løsning. Disse konsekvenser fortolkes så igen begrebsligt, og fører til forudsigelser, der kan lede til beslutninger i den virkelige verden. Denne proces i højre side af figuren handler især om at gøre den matematiske indsigt *anvendelig*.



Figur 2: Matematisk modellering er en iterativ og dialektisk proces (Johansen og Sørensen, 2014, s. 181).

Det, vi netop har skitseret, vil være et enkelt gennemløb af processen i figur 2, men det er nogle af de vigtigste indsigter i matematiske modeller inden for de sidste årtier, at denne proces er *dialektisk* og *iterativ*. At den er dialektisk betyder, at hvert skridt i modellen er en afvejning af muligheder og behov: Der er ikke noget vundet ved at udlede en matematisk beskrivelse, som man ikke har redskaberne til at undersøge — så er det bedre at forenkle den begrebslige model yderligere ved at se bort fra nogle faktorer, selvom man ved, at de nok øver indflydelse på den modellerede opførsel. Men der kan også være mange andre hensyn end denne slags *pragmatik* på spil: Hvis man fx er ude efter at bruge modellen til at forstå et fænomen, vil man ofte vægte simplicitet frem for stor forudsigelsespræcision, og den slags værdier (cf. THOMAS KUHNS (1922–1996) paradigmer) spiller også centralt ind i modelleringsprocessen. At modelleringsprocssen er iterativ skyldes, at selv efter nok så mange gennemløb af processen, er den opnåede model kun ’foreløbig’, men forhåbentlig anvendelig viden. Hvis modellens (fortsatte) brug forudsætter større præcision eller en mere nuanceret inddragelse af faktorer, må man udvide eller forfine modellen, og dette vil føre til gentagne gennemløb af processen i figur 2.

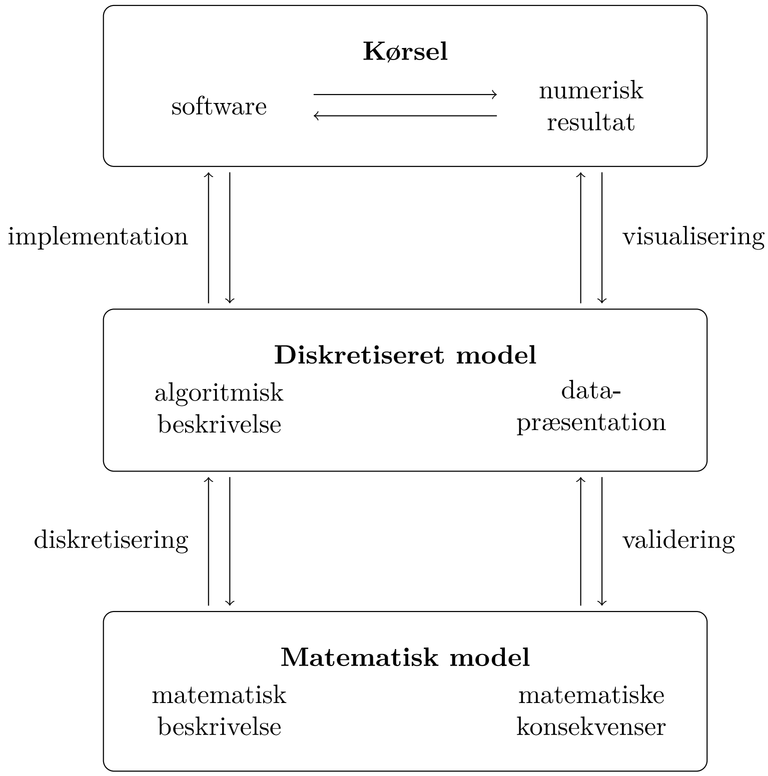
4 Modeller med stor kompleksitet

I nogle af de områder, hvor man særligt bruger modeller, og som samtidig rejser vigtige filosofiske spørgsmål, modellerer man meget komplekse systemer. Et særligt interessant problem opstår, når den komplekse model opstår som en *aggregering* af mindre størrelser: enten aktører eller andre modeller. Sådan nogle komplekse modeller opstår fx i økonomi eller klimamodellering.

I økonomiske modeller er der en stående diskussion omkring relationen mellem de enkelte agenter (fx borgere), som foretager handlinger på *mikroniveau*, og de effekter, som kan aflæses på det overordnede niveau (fx BNP), som kan ses på *makroniveau*. I denne slags modeller består aggregeringen altså i at samle de mange individuelle aktørers handlinger, som er styret af antagelser (og dermed modeller) på mikroniveau, til nogle af de oplysninger på makroniveau, som økonomen er interesseret i.

I klimamodellering kan man sjældent holde den enkelte model op imod en objektiv virkelighed for validering, og i stedet bruger man nogle gange *andre modeller* til at bedømme og korrigere sin model. Hvis man således har en samling af modeller, måske udviklet i forskellige laboratorier og helst ud fra forskellige grundantagelser, så kan man håbe på, at en *aggregeret* model kan være bedre end nogen af de enkelte modeller, dels fordi den måske tager højde for flere faktorer i alt, og dels fordi den typisk vil være mindre sensitiv, hvis den fx er dannet som et gennemsnit af andre modeller. Men hvordan man foretager denne aggregering, og om alle de indgående modeller vægtes ligeligt, er beslutninger med store konsekvenser, som man kan have svært ved at forsvare ud fra generelle principper.

For sådanne komplekse modeller vil man ofte benytte computer-simulationer, hvor man så kan regne på forskellige scenarier for at følge konsekvenserne. På den vis kan man i nogle henseender se computersimulationer som en afart af klassiske eksperimenter, og nogle gange som tankeeksperimenter. Hvis vi opfatter komplekse modeller som alternativer til traditionelle eksperimenter, så udmærker simulationerne og især scenarietænkningen sig jo ved, at man kan foretage 'eksperimenter' *in silico*, som man ikke ville kunne foretage i virkeligheden, enten fordi tidsperspektivet er for langt (fx klimamodeller) eller fordi eksperimentet er dyrt, etisk uforsvarligt eller farligt (fx simulationer af nukleare processer).



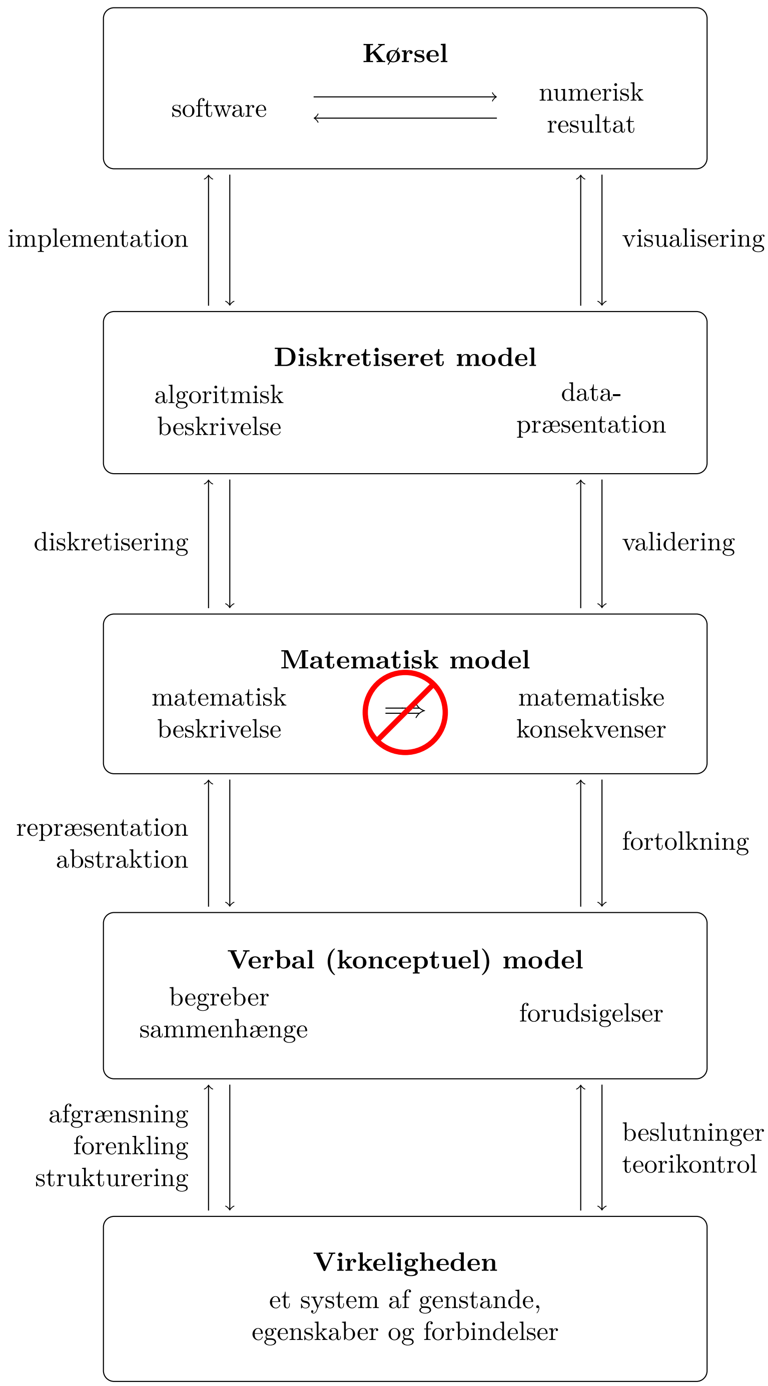
Figur 3: Udvidelse af skemaet om matematisk modellering for at tage hensyn til beregningsmodeller. Denne del af skemaet kan udgøre udforskning af forskellige *scenarier* ved at variere nogle af de indgående parametre, som ellers er konstante i hver gennemregning af modellen.

Ud fra disse repræsentationer af resultaterne er det ofte muligt at foretage en foreløbig og delvis *validering* af beregningsmodellen ved at sammenholde data med betingelser, man har kunnet udlede af den matematiske model. Det kan være helt simple observationer som fortegn etc., men det kan være nok så relevant til at fange *numeriske artefakter*, dvs. opførsel, som ikke hidrører fra modellens teoretiske antagelser og matematiske formulering, men er opstået under implementationen og kørslen af beregningsdelen. Endelig kan resultaterne så lede til forudsigelser og beslutninger omkring det udsnit af virkeligheden, som modellen er bygget til at handle om.

Ligesom det er tilfældet for matematiske modeller mere generelt, er denne modelleringsproces tænkt som både iterativ og dialektisk. Den er først og fremmest *dialektisk* i den forstand, at valg og beslutninger påvirker modelleringsprocessen både 'fremad-' og 'bagudrettet' (alle pile i det samlede skema i figur 4 peger i begge retninger): Kendskab til hvilke diskretiseringer, der kan lede til effektive implementationer, har oftest en indflydelse på, hvilke former for afgrænsning og abstraktion, der giver mening i de forudgående led. Og omvendt vil valget af implementation såsom sprog og skalering af arkitektur selvfølgelig afhænge af egenskaber ved den algoritme, som man ønsker at implementere. Og modellen skal også forstås som *iterativ* i den forstand, at modelleringsprocessen gennemgås flere gange i takt med at resultaterne peger på muligheder for forfinelser og udvidelser. Dette kan være drevet både af utilstrækkeligheder i den tidligere model og af nye muligheder i form af større beregningskraft, flere data, eller nye faktorer, som skal integreres.

Henrik Kragh Sørensen og Mikkel Willum Johansen (apr. 2022). „Invitation til de datalogiske fags videnskabsteori". Lærebog til brug for undervisning ved Institut for Naturfagenes Didaktik, Københavns Universitet. Under udarbejdelse. Kapitel 0, version ().

Men der er også en yderligere form for gentagelse, som gør beregningsmodeller meget brugbare, og som typisk ikke består ved almindelige matematiske modeller. Man kan nemlig variere nogle af de modellerede parametre og køre beregningen igen. Derved kan man få et indblik i disse parametres indvirken på resultaterne, og dette bruges til såkaldte *scenarier*. Dette er en anden måde at bruge beregningsmodeller på, som man kan sammenligne med en slags *udforskende eksperimenteren* (eng: *exploratory experimentation*, se Steinle, 2016), hvorved beregningen bliver mere til et eksperiment end en test af en hypotese.



Figur 4: Samlet skematisk oversigt over processerne i beregningsmodellering.

Modellering af smittespredning: en simpel beregningsmodel

Når man skal modellere sygdomsspredning, er der rigtigt mange faktorer, man kunne forestille sig have indflydelse: sygdommens alvorlighed og spredningsform, befolkningens sammensætning og spredning, tiltag for at begrænse spredningen etc. Smittespredning er derfor et eksempel på et komplekst problem, som man ikke kan gøre sig forhåbninger om at finde en eksakt matematisk beskrivelse af. I stedet benytter man sig af beregningsmodeller, som kan udvikles og tilpasses de specifikke omstændigheder ved en given sygdom.

I kernen af rigtig mange smittemodeller ligger en forholdsvis simpel afgrænsing af problemet, idet man som udgangspunkt antager, at befolkningen er stabil inden for den horisont, hvor modellen skal virke: Der sker ikke ændringer i befolkningens størrelse eller sammensætning, og man antager fx at befolkningen er konstant og homogen. Yderligere antagelser, som man typisk gør sig i denne slags første modeller på befolkningsniveau er, at befolkningen med hensyn til denne sygdom kan inddeles i tre disjunkte kategorier: S som er modtagelige (eng: *susceptible*) for sygdommen, I som er inficerede med sygdommen, og R som enten er døde eller er blevet immune over for sygdommen, og derfor er fjernet fra den relevante population (eng: *removed*). Hvert individ kan altså følge en progression $S \to I \to R$, og modellen kaldes derfor *SIR-modellen.* Dermed er vi på vej til at indskrænke et udsnit af virkeligheden (smittespredning af en given sygdom) til en *verbal model* (befolkningen er konstant og homogen, og sygdommen spredes på en bestemt måde).

For at kunne opstille en matematisk model er vi nødt til at gøre nogle yderligere antagelser og indføre noget notation. For det første kan vi omsætte vores antagelse om konstant befolkning til $N = S(t)+I(t)+R(t)$, hvor t angiver vores tidsparameter. Og hvis vi antager, 1. en modtagelig har risikoen $\frac{I(t)}{N}$ for at møde en smittet og en konstant risiko, β, ved hvert sådan møde for selv at blive smittet, og 2. at smittede dør eller opnår immunitet med en fast sandsynlighed γ, så kan vi begynde at formulere vores matematiske model i termer af ændringshastighederne i de tre kategorier:

$$\frac{dS}{dt} = -\frac{\beta}{N} \cdot I(t) \cdot S(t),$$
$$\frac{dR}{dt} = \gamma \cdot I(t),$$

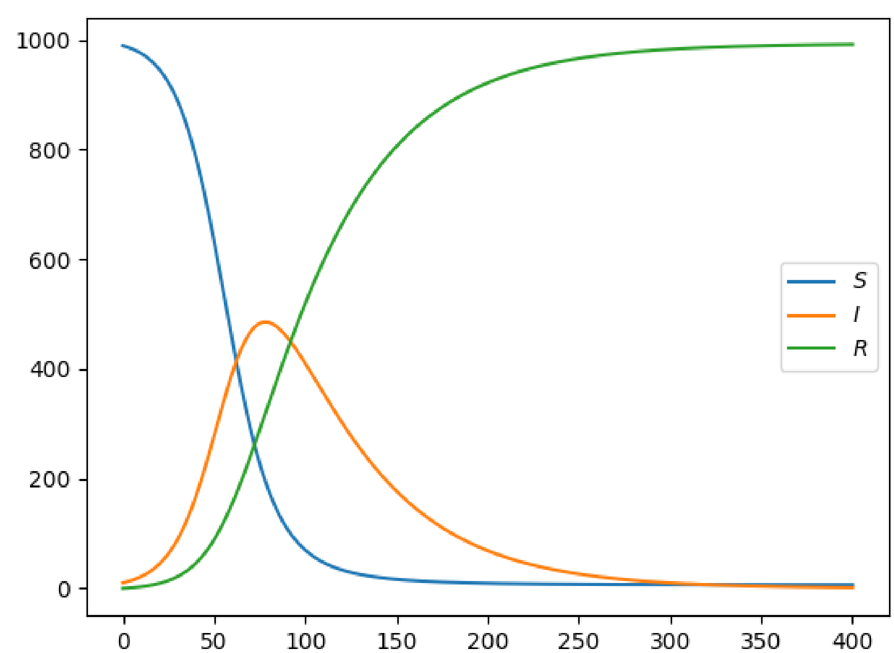
som betegner tilføjede og fjernede smittede, hvorfor

$$\frac{dI}{dt} = \frac{\beta}{N} \cdot I(t) \cdot S(t) - \gamma \cdot I(t).$$

Derved har omsat vores begrebslige model til en *matematisk beskrivelse* i form af tre koblede differentialligninger. Man bemærker her, at der også er en hel del faktorer, som nok konkret er vigtige, som vi ikke har medtaget i vores verbale model; fx problemet omkring latenstid, idet vi antager, at alle inficerede selv smitter videre med konstant sandsynlighed fra første øjeblik.

Nu kunne man tænke sig, at dette sæt koblede differentialligninger tillod en eksakt løsning — men det er generelt ikke tilfældet, og fordi modellen er så forsimplet har man lyst til i næste iteration at tilføje komplesitet, fx ved at gøre parametrene β, γ variable. Derfor tyr man i stedet til beregningsmodellering. Men før man kan oversætte systemet af differentialligninger til noget, en computer kan regne effektivt på, skal den matematiske formulering diskretiseres. Afhængig af problemfeltet findes der forskellige etablerede metoder til at diskretisere, men i det konkrete tilfælde vil det sikkert være så simpelt som at betragte tiden som en diskret følge med konstante spring. Så i stedet for at betragte t som en kontinuert variabel og fx $\frac{dS}{dt}$ som instantan hastighed, betragter vi i stedet tiden som en diskret størrelse og undersøger udviklingen $\Delta S(t) = S(t+\Delta t) - S(t)$, hvor Δt er en fast (og 'lille') konstant. Så er $\frac{\Delta S}{\Delta t}$ en diskret approximation til $\frac{dS}{dt}$, og den matematiske model kan laves om til en *algoritmisk beskrivelse*, hvor vi lader t gennemløbe værdierne $t_0, t_0+\Delta t, t_0+2\Delta t, \ldots, t_1$ og hele tiden opdaterer, fx $S(t+\Delta t) = S(t) + \Delta t \cdot \Delta S(t)$.

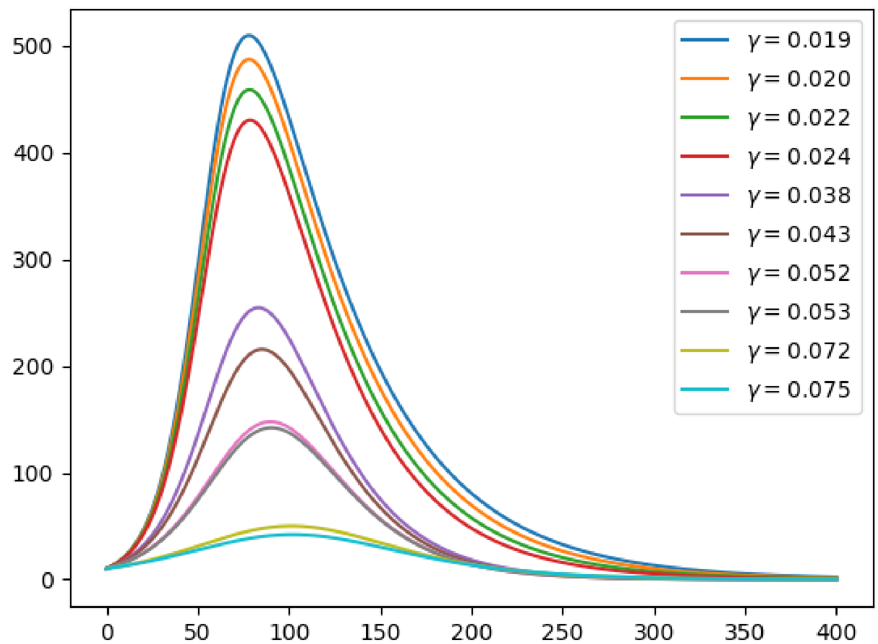
Denne algoritme skal nu implementeres i et programmeringssprog med henblik på afvikling på en given arkitektur, dvs. som et stykke *software*. Og som omtalt i kapitel **??** giver det anledning til en række overvejelser, for både sprog og arkitektur kan være mere eller mindre egnede til den konkrete opgave. Til denne slags simuleringsopgaver vil det generelt være en god ide at benytte fx Python eller Fortran, som er udviklet specifikt til videnskabelige beregninger, og som har masser af biblioteker, der understøtter det. Men hvis man fx valgte at implementere det i Excel eller C eller et andet sprog med en fast, endelig opløsning i repræsentationen af reelle tal, kan man risikere at visse numeriske beregninger går galt på grund af afrundingsfejl. En implementation i Python kan ses i appendix **??**, og når den bliver afviklet på en understøttet arkitektur, giver den anledning til en række talværdier for udviklingen af S, I, R over tid. Dette er det *numeriske resultat* af kørslen. Meget ofte vil det være meget mere ønskeligt at bruge computeren til også at fremstille andre *repræsentationer af resultaterne*, fx gode visuelle repræsentationer af forløbet over tid (en sådan graf er gengivet i figur 5).



Figur 5: Grafisk repræsentation af forløbet af S, I, R for givne parametre.

Modellens resultater kan så holdes sammen med viden om den matematiske model. For eksempel udviser kurverne i figur 5 det forventede overordnede forløb, så vi har ikke umiddelbart grund til at betvivle den numeriske beregning. Og selvfølgelig skal modellens resultater også bringes i spil til at lave de forudsigelser eller beslutninger, som var formålet med at bygge den i første omgang.

I figur 6 er gengivet, hvordan antallet af smittede I afhænger af parametren γ for et antal forskellige (her normal-fordelte) værdier. Dermed kan man fx begynde at estimere denne parameter ud fra fx kendte data eller ud fra kapacitetsbetragtninger, hvilket måske ikke kan gøres eksplicit. Hvis fx kapaciteten af en eller anden grund er på 200 inficerede, kan man ud af scenarierne aflæse, at man skal stræbe efter en værdi $\gamma \geq 0.043$.



Figur 6: Grafisk repræsentation af forløbet af I under forskellige scenarier.

Litteratur

Johansen, Mikkel Willum og Henrik Kragh Sørensen (2014). *Invitation til matematikkens videnskabsteori*. København: Forlaget Samfundslitteratur.

Steinle, Friedrich (2016). *Exploratory Experiments. Amp re, Faraday, and the Origins of Electrodynamics*. Overs. af Alex Levine. Pittsburgh: University of Pittsburgh Press.