**MODELLEREN EN SIMULEREN HOOFDSTUK 4: Modelevaluatie en -validatie**

1. Inleiding

1.1 Inleiding

* Hoe zeker zijn we over de juistheid vh model?
* Wordt het systeem voldoende nauwkeurig beschreven?
* Mogen de veronderstellingen en vereenvoudigingen (eigen aan het model) gemaakt worden?

1.2 Modelevaluatie en modelvalidatie

* Modelevaluatie
  + = het nagaan of het model een goede weergave is van de experimentele data
  + Hiervoor het ‘Goodness-of-fit’ criterium gebruiken
    - = beschrijft hoe goed het model overeenkomt met de experimentele observaties
  + Bij een ‘goed/robuust’ model: (criteria)
    - Kunnen de parameters geschat worden met lage onzekerheid
    - Kan met het geparametriseerd model de experimentele data voldoende nauwkeurig worden bekomen
    - En dit zonder daarbij ruis te modelleren (overfitting vermijden!)
    - ⬄ underfitting: als model de experimentele data onvoldoende nauwkeurig volgt maw indien onvoldoende parameters of foute modelstructuur
    - ⬄ overfitting/ overparametrisatie = als model na parameterschatting de ruis mee modelleert (kan door vb teveel parameters in model te brengen)
* Modelvalidatie
  + = het nagaan of het model nieuwe waarnemingen met grote waarschijnlijkheid kan beschrijven
    - Ook al is er een goede overeenkomst tssn model & exp data waarmee model geparametriseerd is => wil niet zeggen dat model het proces goed beschrijft
    - Want beschrijft het ook nieuwe data met even grote nauwkeurigheid?
  + Probleem treedt op bij: complexe modellen of bij gebruiken van polynomiale modellen die een fysische basis ontbreken
  + Onderzoeken of het (geparametriseerd)model gebruikt kan worden in een gewijzigde situatie, op een ander tijdstip, gedurende een langer tijdsinterval
* Als na modelevaluatie of -validatie geen overeenkomst met realiteit
  + => nieuw of verbeterd model kiezen
* Conclusie: eerst modelevaluatie & daarna modelvalidatie

2. Modelevaluatie

* Voorgaand H: voorbeelden om de goodness-of fit te beoordelen gezien
  + Tijdens schatten van parameters gebruikten we de goodness-of-fit criterium om de voortgang vd parameterschatting door een minimalisatiealgoritme te evalueren
    - Vaak wordt de SSE geminimaliseerd
    - => maw een lage waarde van SSE is een kwantitatieve bepaling vd goodness-of-fit
  + Echter nog andere methoden om modellen te evalueren
* Evaluatiemethoden
  + 1) grafische methodes: helpen bij de visuele interpretatie vd fitting/ overeenkomst
    - Voordelen: de volledige dataset kan tegelijk bekeken worden
      * Het is een eenvoudige weergave over brede range vh verband tssn model & data
  + 2) kwantitatieve of numerieke methodes: gebaseerd op statistische evaluatie & leveren een meer objectieve evaluatie vh model
    - Nadeel: meer beperkt, want gefocust op bep. aspect vd data
      * Vaak alle info in 1 numerieke waarde persen

2.1 Grafische methodes

* 1) Visuele interpretatie van modeloutput
  + => Simuleren met het te evalueren model & nadien visueel vglen vd modeloutput met de experimentele data
  + Vb: dit doet men voor én na parameterschatting (fig ppt 197)
    - Gevolg: visuele interpretatie geeft het effect van het schatten vd parameters op de juistheid vh model weer
    - Figuur
      * Blauw = experimentele data, stippellijn = simulatie vh model voor parameterschatting; rode lijn= simulatie na parameterschatting
      * Voor parameterschatten: klokvormige output vh model is zichtbaar, maar overeenkomst is nog niet voldoende
      * Na parameterschatten (minimale doelfunctie bekomen): goede overeenkomst tssn exp data & model = visueel waarneembaar
  + Vb: bij vglen van eenvoudig vs complexer model (minder of meer parameters)
    - Gevolg: visuele interpretatie geeft idee of het nuttig is om een complexer model te gebruiken (met vb meer parameters) of niet
  + OPM:
    - Bij visuele interpretatie w een groot aantal punten in het interval vd simulatie geplot => niet enkel de modelwaarden op de tijdstippen vd meting
    - De meetwaarden = punten; model = volle of stippellijn
    - Visuele interpretatie is onmogelijk voor zeer compexe modellen & lange en meervoudige datasets die tegelijk moeten bestudeerd w
* Scatter plots
  + => vergelijken modelvoorspellingen en experimentele data na eliminatie vd tijdscomponent
  + => bij goed model liggen datapunten op de bissectrice
  + Figuur ppt p198:
    - 2 modellen worden geëvalueerd voor het voorspellen vd hoeveelheid biomassa in bioreactor op einde van proces
    - Waarden model uitzetten tov waarden vd waarnemingen
    - X-as: gemeten biomassaconcentraties y-as modelwaarden
    - Links: modeloutput wijkt op systematische manier af vd exp waarden
      * Bij lagere conc biomassa => gesimuleerde conc lager dan de gemeten
      * Bij hogere conc biomassa => omgekeerd
      * = trend in de afwijking / systematiek in de afwijking
      * => zal 1 of ander fenomeen/parameter niet in rekening zijn gebracht
    - Rechts: goed model, datapunten liggen op bissectrice
* Plotten van residuals
  + Residual/ afwijking voor 1 bepaald datapunt
    - = de waargenomen meetwaarde, minus de modelwaarde/ de gefitte waarde
    - = de responswaarde (gemeten) minus de voorspelde waarde (model)
    - = de verticale afstand van meetpunt tot model (de rechte lijn) curve
    - Worden weergegeven tov de modeloutput
    - Als model correct is: residuals = de meetfouten vh experiment
  + Figuur ppt p199: modellen die biomassaconcentratie voorspellen
    - X-as: gemeten biomassaconcentratie; y-as = waarden vd residuals
    - Fig 1: ergens fenomeen niet in rekening gebracht => systematisch patroon
    - Fig 2: de afwijking is gewoon ruis op de data = goed => random verdeeld
  + Het plotten vd residuals geeft beeld vd manier waarop deze afwijkingen van model tov metingen verspreid liggen in functie vd onafhankelijke variabele vb tijd
    - Als residuals random verdeeld zijn => voldoet het model
    - Als residuals systematisch patroon vertonen => voldoet model niet
      * Dan zoeken naar een complexer model, meer parameters in rekening brengen
    - Vb: ppt p201: random verdeeld, heteroscedastische (of niet homogene verdeelde) en niet lineaire spreiding vd residuals (voldoet ook niet)

2.2. Kwantitatieve/numerieke methodes

* Absolute methodes
  + = niet-relatieve methodes
  + => baseren op het verschil tssn de geobserveerde & gemeten waarde
  + Typische kwantitatieve methoden zijn gebaseerd op de SSE, MSE, SAE en MAE
    - n= het aantal (onafhankelijke) observaties
* Genormaliseerde methodes
  + = relatieve methodes
  + => de modelevaluatie w genormaliseerd (of relatief) ten opzichte v/e standaard waarde voorgesteld
  + Vb: de Nash-Sutcliffe coefficient voor modelefficientie E
    - n= aantal (onafhankelijke) observaties
    - ygem, exp = het gem. vd geobserveerde waarden
    - = 1- verhouding vd 2 afwijkingen
    - = samengesteld uit de afwijking vd simulatie tov de metingen, gedeeld door de afwijking van de metingen tov het gemiddelde
    - = dezelfde uitdrukking als de determinatiecoeff R2 voor statistische modellen
    - Bij regressiemodellen: E=R2 (met waarden tussen 0 en 1)
    - Bij andere regressiemodellen: E varieert tssn [-infty,1]
      * Zie ppt

Afbeelding met schermafbeelding

Automatisch gegenereerde beschrijving

3. Modelvalidatie

* Model kan gevalideerd worden om na te gaan of het gebruikt kan worden op andere metingen van hetzelfde systeem, of het model kan gevalideerd w om na te gaan of het ook in andere omstandigheden kan gebruikt worden

3.1 Crossvalidatie

* Crossvalidatie
  + = testen v/e model met onafhankelijke data van die waarmee het geëvalueerd werd
  + => Model testen op **hetzelfde systeem**, mbv **verschillende metingen**
  + Nodige voorwaarde: onafhankelijke dataset voorhanden ( is niet steeds het geval)
    - 1) dataset voor de training: parameterschatting & modelevaluatie
    - 2) andere dataset (vb week later) voor validatie
  + Indien er geen onafhankelijke datasets voorhanden zijn => grote dataset opsplitsen in 2 datasets
    - 1) 1 dataset om modelparameters te schatten
    - 2) 1 validatieset om het bekomen geparametriseerd model te valideren
    - Vaak w de dataset in 3 delen opgesplitst
      * 1) data voor training (parameterschatting & modelevaluatie)
      * 2) data voor validatie
      * 3) data voor de verdere evaluatie
  + Welke crossevaluatie gebruikt w is afhankelijk vd hoeveelheid data die beschikbaar zijn & hoe moeilijk het is om de validatietests te runnen
* Extrapolatie
  + Men vertrekt v/e model dat bekomen werd door parameterschatting, waarbij data gebruikt werd v/e bepaald systeem onder bepaalde omstandigheden
  + Men kijkt na of dit model ook gebruikt kan worden voor hetzelfde systeem, onder andere omstandigheden (vb verder in toekomst)
  + => Model testen op **hetzelfde systeem** onder **versch omstandigheden**
  + vb: versch groeimodellen die de groei vd wereldbevolking in toekomst voorspellen
    - => model gebaseerd op zwarte stippellijn
    - => dan nagaan hoe ver model de groei goed volgt
      * => leiden tot versch extrapolaties
* Verschillende systemen
  + Men gaat na of een bep. modelstructuur kan gebruikt w om andere systemen te modelleren
  + => Model testen op **ander systeem** onder **dezelfde of versch omstandigheden**
  + Vb: stel een groeimodel gebruikt voor de groei van een bep micro-organisme te beschrijven in een bioreactor
    - Eens groeimodel voor een micor-org is ontwikkeld => nagaan of model ook kan toegepast w om groei van andere micro org te beschrijven
    - Vb: een groeimodel dat gebruikt werd om de groei van E coli te beschrijven, kan toegepast w om de groei van Lactobacilus casei op een ander substraat te beschrijven
  + Bij gebruik van dezelfde modelstructuur op ander systeem
    - => Moet het model vaak opnieuw geparametriseerd worden
    - vb: elk micro-organisme heeft zijn eigen specifieke groeisnelheid => deze parameter moet opnieuw bepaald w voor elk groeimodel
  + Onderscheid maken tussen
    - Enerzijds model & anderzijds modelstructuur
      * Model: groeimodel van E coli
      * Modelstructuur: het type groeimodellen gebruikt om groei van micro-org te beschrijven
    - Enerzijds model & anderzijds geparametriseerd model
      * Model: het type groeimodellen gebruikt om groei van micro-org te beschrijven
      * Geparametriseerd model: model geparametriseerd voor bep micro-org