



Strål
säkerhets
myndigheten

Swedish Radiation Safety Authority

Författare: Peter Hessling

Forskning

2017:23

Kalibrering för bestämning av
optimal beräkningsmodell

SSM perspektiv

Bakgrund

SSM bedömer i tillsynen avancerade datorberäkningar relaterade till drift av kärnkraftverk, utförda av extern part. Inom kärnkraft kan experimentella undersökningar vara olämpliga. Beräkningar blir därför ofta det enda alternativet att undersöka möjliga konsekvenser av planerade förändringar. Beräkningsmodeller innehåller emellertid normalt ett flertal mer eller mindre grova approximationer och antaganden, vilka orsakar beräkningsfel.

Möjliga beräkningsfel kan tas i beaktande med generella antaganden om beräkningsmodellens osäkerheter. Metoder för osäkerhetskvantifiering överför dessa osäkerheter till uppskattade osäkerheter i de beräknade resultaten. För mätningar kallas motsvarande mätosäkerhet, omfattande mätinstrumentfel men även experimentella begränsningar. Överensstämmelse betyder att osäkerhetsintervallen för beräkningen och mätresultatet överlappar.

En svaghet med osäkerhetskvantifiering är att den baseras på antaganden om modellens parametrar. Dessa kan dock bestämmas indirekt genom att modellen och dess osäkerhet justeras för bästa överensstämmelse med mätresultat, för ett kalibreringsexperiment. Den erhållna optimerade beräkningsmodellen antas sedan ge bästa möjliga förutsägelser.

Projektet föreslår två nya metoder för modellkalibrering. Båda bygger på iterativ kalibrering med deterministiska ensembler. Medan metod 1 baseras på etablerad s.k. maximum likelihood-teknik föreslår metod 2 ett helt nytt koncept som också baseras på en annan tolkning av rapporterad osäkerhet. Metod 1 kräver fullständig information för kalibreringsdata, medan metod 2 använder tillgänglig information. Tveksamma antaganden som krävs för metod 1 kan ofta helt undvikas med metod 2. Modellosäkerheten blir betydligt högre för metod 2 än för metod 1, som i regel blir alldeles för låg.

De två kalibreringsmetoderna utvecklades för ett antal enkla testmodeller. Kalibreringsmetoderna testades sedan för en realistisk CFD- (computational fluid dynamics) modell, för att bestämma den strömningshastighet som tidigare definierats av ett jämförelse- eller s.k. benchmarkprojekt föreslaget av OECD. Kalibreringsdata framtoogs för detta s.k. GEMIX (GEneric MIXing) experiment vid Paul Scherrer Institute i Schweiz.

Syfte

Tidigare har SSM använt deterministisk sampling för osäkerhetskvantifiering. SSMs övergripande mål är att få ett exempel på hur deterministisk sampling kan användas i en metod för modellkalibrering. Syftet är också att konkret påvisa nyttan med kalibrering för en given beräkningsmodell och ett specifikt kalibreringsexperiment, från ett större OECD jämförelseprojekt med SSMs deltagande

Resultat

Den principiella skillnaden mellan metoderna är om modellen antas vara sann (metod 1) eller om kalibreringsdata får avgöra i vilken utsträckning den är korrekt (metod 2). Modellens osäkerhet beskrivs följaktligen på olika sätt, som "osäkerhet för bästa skattning" (metod 1) eller "bästa representation av osäkerhet för

kalibreringsdata”. Kalibreringen av CFD-modellen gav marginella förbättringar av det förväntade resultatet. Däremot korrigerades osäkerheten desto mer. Modellen klarade validering efter kalibrering med metod 2, men blev underkänd med metod 1. Optimeringen av beräkningsmodellen begränsades alltså huvudsakligen till dess osäkerhet.

Behov av ytterligare forskning

Medan metod 1 baserad på traditionell kalibreringsmetodik inte förändrats under projektet, så har metod 2 genomgått ett flertal större justeringar. Dessa har framförallt adresserat konvergens vid iteration. Eftersom metod 2 ger överlägset störst chanser för en given modell att bli godkänd, kan vidare studier och utveckling av metod 2, som ger optimal representation genom iterativ s.k. stratifiering och annealing, anses ha stort alternativt mycket stort värde. Konsekvent användning av kalibrering med metod 2 kan utveckla framtida beräkningsmetodik mot mindre grad av spekulation i förutsättningar. Därmed ge högre trovärdighet för de beräkningsresultat som SSM ibland använder vid beslutsfattande. Beräkningens förväntade kvalitet blir med metod 2 direkt avspeglad i dess beräknade osäkerhet i kontrast mot metod 1 var osäkerhet relaterar till skattningen av beräkningsmodellen och inte kvaliteten på dess förutsägelser.

Projekt information

Kontaktperson SSM: Peter Hedberg

Referens: SSM 2016-2851



Strål
säkerhets
myndigheten

Swedish Radiation Safety Authority

Författare: Peter Hessling
Kapernicus AB, Hällingsjö

2017:23

Kalibrering för bestämning av optimal beräkningsmodell

Datum: Augusti 2017

Rapportnummer: 2017:23 ISSN: 2000-0456

Tillgänglig på www.stralsakerhetsmyndigheten.se

Denna rapport har tagits fram på uppdrag av Strålsäkerhetsmyndigheten, SSM. De slutsatser och synpunkter som presenteras i rapporten är författarens/författarnas och överensstämmer inte nödvändigtvis med SSM:s.

Sammanfattning

En beräkningsmodell utgörs huvudsakligen av två komponenter, ekvationer samt värden på parametrar och förutsättningar. Kalibrering syftar till att justera, eller optimera dessa värden så att modellresultat och mätresultat för ett specifikt kalibreringsexperiment överensstämmer så bra som möjligt. Dessutom bör en osäkerhet för anpassningen bestämmas. Avsikten med kalibrering är att kunna ersätta experiment med systematiskt kontrollerade och optimerade beräkningar, vilket på många sätt kan vara fördelaktigt. Det gäller i synnerhet inom kärnkraft där t.ex. kritiska driftsituationer inte bör undersökas experimentellt eller tester kan vara mycket kostsamma.

Inom projektet har två helt olika kalibreringsmetoder utvecklats. Metod 1 baseras på traditionell maximum likelihood *skattning*. Metod 2 *representerar* kalibreringsdata med beräkningsmodellen. Brist på information ökar modellens osäkerhet (metod 2), eller kräver gissning av det som saknas (metod 1). Ett antal enkla testmodeller har använts för att studera metodernas egenskaper. Båda metoder har visat sig fungera tillfredsställande utifrån deras inneboende begränsningar.

Kalibreringsmetoderna har även tillämpats för att bestämma vätskeströmningshastighet med en realistisk CFD- (computational fluid dynamics) modell. Frågeställningen härrör från ett OECD jämförelseprojekt, med kalibreringsdata framtagna vid Paul Scherrer Institute i Schweiz med ett GEMIX (GENeric MIXing) experiment. Nyttan med kalibrering befanns emellertid vara begränsad i detta fall, eftersom det systematiska modellfelet dominerade. Kalibreringen justerade framförallt osäkerheten för metod 2 så att modellen klarade en validering. Däremot resulterade metod 1 i underkänd modell.

En helt ny kalibreringsmetodik, liknande föreslagen metod 2, kan således krävas för att kalibrering ska kunna bli ett skarpt instrument för att optimera avancerade beräkningsmodeller. Den överdrivet självsäkra beräkningsmodell som den mer traditionella kalibreringsmetoden 1 ger, illustrerar med all tydlighet den gamla sanningen:

Att vara säker är inte alls detsamma som att ha rätt.

Innehållsförteckning

Syfte.....	4
Förutsättningar.....	4
Bakgrund.....	4
Problemdefinition.....	7
Översikt.....	8
Linjäritet.....	10
Linjär regression.....	10
Enkel osäkerhetskvantifiering.....	11
Modellkalibrering.....	12
Identifierbarhet.....	14
Konsistens.....	14
GEMIX kalibrering.....	15
Anpassning och konvertering.....	17
Startensemble.....	17
Resultat.....	18
Traditionell mot ny modellkalibreringsmetodik.....	18
Konsistens.....	20
Kalibrering av deterministisk ensembler.....	22
Fixpunktsiteration — successiv förbättring.....	22
Metod 1: Maximum likelihood skattning.....	23
Metod 2: Stratifiering och annealing.....	24
Grundidé.....	24
Komplikationer.....	25
Ensemblekollaps.....	27
Residual.....	28
Mätfel.....	29
Osäker modell.....	30
Stratifiering.....	31
Annealing.....	31
Utvärdering.....	32

Konsistens	32
Osäkerhetsbidrag	32
Täckningsintervall	33
Sampelstorlek kalibreringsdata	33
Testmodeller	33
Affin modell	33
Chirp modell	38
GEMIX	42
Kalibrering 0 – Start	42
Kalibrering 1	47
Kalibrering 2 – Slutgiltig	52
Parameterensemble (kalibrerad)	56
Parameterstatistik (kalibrerad)	57
Korrelation parametrar (kalibrerade)	58
Validering	59
Giltighet	59
Utvärdering	59
Slutsats	62
Referenser	63

Syfte

Avsikten med aktuellt forskningsuppdrag är att sprida etablerade och utveckla nya metoder för trovärdig kalibrering av beräkningsmodeller. I synnerhet komplexa och stora modeller. Förhoppningen är att optimera modeller med kalibrering för att ge bättre förutsägelser. Metoder för bästa skattning av modellparametrar ska undersökas. Framförallt är deras osäkerheter och beroenden av intresse.

Projektet syftar också specifikt på kalibrering av en vald vätskedynamisk CFD- (computational fluid dynamics) s.k. 'k-epsilon' modell, som använts för att studera ett mixningsproblem definierat i ett jämförelseprojekt ('benchmark') föreslaget av OECD [1].

Förutsättningar

Resurs för beräkning av CFD-modell med givna parametrar tillhandahålls av Strålsäkerhetsmyndigheten. Kalibreringsdata från motsvarande fysiska (GEMIX) experiment att kalibrera beräkningsmodellen mot finns tillgängliga inom OECD-projektet [1], liksom valideringsdata för att kontrollera den kalibrerade modellen.

Bakgrund

Aktuellt kalibreringsprojekt är en efterföljare till det föregående SSM-projektet "Modellosäkerhet i beräkningar", utfört 2014-2015 [2]. Det adresserade osäkerhetskvantifiering med givna modeller och osäkerheter, med användning av Wilks metod [3] och olika varianter av föreslagen deterministisk sampling [4]. Deterministisk sampling utökas med detta projekt till att även omfatta omvänd, eller *invers osäkerhetskvantifiering* [5], dvs bestämning av modeller och osäkerheter utifrån kalibreringsdata från kontrollerade experiment. En skillnad mellan projekten är att i detta projekt adresseras en bestämd beräkningsmodell för vätskeströmning, med givna kalibreringsdata från GEMIX (GEneral MIX) experiment utförda vid Paul Scherrer Institutet i Schweiz.

Medan den naturliga måttstocken i föregående projekt var slumpad (eng. 'random') sampling, är konventionell maximum likelihood skattning med linjäriserade modeller [6][7] utgångspunkten för detta uppdrag. Problemen med samplingsvarians [8] som är centrala för Wilks metod [3] och andra slumpade samplingstekniker [9] kommer ej att behandlas här eftersom det inte finns något motsvarande för deterministiska ensembler. En komplex aspekt är tolkningen av modellosäkerhet och vad som påverkar dess storlek.

Generellt erhålls mycket små osäkerheter med konventionell skattningsmetodik, som dessutom kräver försumbara systematiska fel i förhållande till mätosäkerheten för kalibreringsdata — något som ofta *inte* är uppfyllt i praktiken. Denna fråga är därför allt annat än filosofisk, eftersom ett för litet osäkerhetsintervall ger en alltför 'självssäker' modell som löper mycket hög risk att förkastas, dvs ogiltigförklaras vid validering. Tolkningen blir därför bokstavligen en existensiell fråga för modellen. Hur en trovärdig

och rimlig men inte överdrivet hög modellosäkerhet kan identifieras från osäkra kalibreringsdata är därför den viktigaste frågan att besvara i detta arbete. Det är förmodligen inte hur bra den bästa skattningen är som är huvudfrågan. Bästa skattning har nämligen ingen annan tolkning än att den minimerar det förväntade modellfelet utifrån en given vald norm, mer eller mindre oberoende av identifieringsmetod.

Ifall algoritmeffektiviteten var avgörande för direkt osäkerhetskvantifiering i föregående projekt, så gäller det i ännu högre utsträckning för modellkalibrering (denna studie). Faktum är att för de flesta etablerade metoder så går det inte ens att genomföra en modellkalibrering överhuvudtaget, utan omfattande och drastiska förenklingar av modeller och förutsättningar. Den etablerade metodiken baserad på statistisk skattning må vara begriplig och naturlig på många sätt men lider av en stor nackdel som sällan eller aldrig diskuteras: Den kräver fullständig information, något vi i princip *aldrig* har. Det blir därför mer än en uppmuntran till att fylla alla kunskapsluckor med mer eller mindre tvivelaktiga antaganden, det blir ett *krav*. Antaganden kan tyckas enkla att bortförklara med att det handlar om vår *upplevda* kunskap (eng. state-of-belief) enligt Bayesianisk formalism [10], vilket gör att vi måste *tillåtas* att mer eller fritt beskriva den. Hur vi rättfärdigar antaganden spelar dock ingen som helst roll om vi faktiskt har helt fel. Vår upplevelse kan t.ex. vara att vår planet är helt plan. Det orimliga i denna 'upplevelse' sänker ovillkorligen vår förmåga att förstå omvärlden, precis som en skev uppfattning om kalibreringsdata reducerar den kalibrerade modellens förmåga att förutsäga resultat för andra situationer än den undersökta (vid kalibreringen).

Istället för att kräva mer kunskap så bör vi snarast respektera vår okunskap och paketera den som tvetydighet. Det ger oss mindre information och därmed mer osäkerhet i vår inferens. Tvetydighet (eng. ambiguity) [11] kommer alltså även i detta projekt att få en central betydelse. I detta fall är dess ursprung att vår information om kalibreringsdata är ofullständig eller osäker. Exempelvis vet vi oftast ej hur mätfel *sam*-varierar statistiskt. Oavsett vad vi föreställer oss eller tror, så resulterar det i en faktisk osäkerhet på allt som senare förutsägs med den kalibrerade modellen. Det är vårt aktiva val om vi (som vanligt) väljer att ignorera den med specifika antaganden. Gissar vi information är risken uppenbar att vi helt *slumpmässigt* reducerar modellosäkerheten. Modellen blir därmed alltför självsäker (eng. overconfident). Alltför hög självsäkerhet brukar i regel leda till katastrof, så även i detta fall. Denna osäkerhet för beräknad osäkerhet benämns här 'tvetydighet'. Den specialstuderas i detta projekt utifrån de tämligen lösa antaganden som normalt görs, med syfte att se hur det påverkar modellosäkerheten och därmed storleken på osäkerheten för framtida förutsägelser. Tvetydigheten kan i princip respekteras genom att helt undvika tveksamma antaganden och istället acceptera tillgänglig informations ofullständighet. Ingen(!) etablerad kalibreringsmetod tillåter emellertid det. Av denna orsak påbörjades arbetet med en ny metod som i förslaget för detta projekt benämndes 'Regulariserad stratifiering', med nummer 2. Vad namnet bör vara för att ge en bra beskrivning är ännu ej bestämt. Därför benämns metoden olika i delrapporteringen, dock alltid med nummer 2.

Bayesiansk metodik [10] kan tyckas hantera tvetydighet, men gör faktiskt precis det omvända. Genom att kombinera förkunskap med härledd kunskap så *ökas* informationsmängden så att osäkerheten minskar än mer. Det anses acceptabelt att hävda en viss förkunskap så länge den beskriver vår uppfattning på ett till synes 'rimligt' sätt.

Har den inte ett trovärdigt ursprung, så blir modellosäkerheten ovillkorligen mer oriktig, än om den endast baserats på härledd kunskap. Om kravet på fullständig information generellt är problematiskt så blir det därför än värre för Bayesiansk metodik — såväl förkunskap som härledd kunskap måste vara fullständig. Med andra ord så går Bayes generalisering i omvänd riktning mot syftet i detta projekt att respektera tvetydigheter. Därför utesluts Bayes metodik ur diskussionen.

Principen för ‘maximum entropy’ [12][13] adresserar däremot kunskapsbrist på ett adekvat sätt. Kort kan denna princip sägas maximera den kvarvarande osäkerheten (entropin) när allt vi vet tagits i beaktande. Den förlitar sig på ett kvantitativt mått av informationsmängd som kallas entropi (eng. entropy), föreslaget av Claude Shannon i den mycket kända artikeln “Theory of Communication” [14]. Ju mindre informationsinnehåll, desto högre entropi. Metodens motiv är precis detsamma som för antaganden — att fylla på information när statistiska metoder kräver det. Den avgörande skillnaden är att göra antaganden på ‘känsla och erfarenhet’, eller att göra det på det allra mest försiktiga sättet genom att maximera entropin — så att alltför självsäkra slutsatser i största möjliga mån kan undvikas. Grundsynen att maximera entropin när information saknas harmonierar därför väl med ambitionen i detta projekt.

För modellkalibrering krävs emellertid sällan en syntes av osäkerhet liknande principen för maximal entropi, av ett mycket enkelt skäl. Även med en hög grad av ofullständighet, kan kalibreringsdata nämligen vara tillräckliga för att *entydigt* bestämma en osäker beräkningsmodell. Det som fyller i informationsluckorna, traditionellt med antaganden eller maximering av entropin, är den begränsning i frihetsgrader som modellekvationerna ger för det överbestämda kalibreringsproblemet. Ett överslag illustrerar principen. Antag att kalibreringsdata finns för 1000 punkter. Modellen måste förutsäga kalibreringsdata i precis samma punkter, men med nödvändighet(!) mycket starka beroenden eftersom modellen kanske bara har 10 fria parametrar. Medelvärden och kovarianser för dessa motsvarar inte mer än $10 + 10 \cdot 11/2 = 65$ frihetsgrader att passa in mot kalibreringsdata som har $1000 + 1000 \cdot 1001/2 = 500500$ motsvarande frihetsgrader. Det är kvoten mellan dessa som avgör hur säkra vi blir på vår kalibrerade osäkra modell, i detta fall $500500/65 \approx 7700$. Det finns alltså inte mindre än runt 7700 tvång på varje ‘frihet’ modellen har. I detta fall har vi alltså full information om kalibreringsdata.

I praktiken saknas information om kalibreringsdata. Den antagna hypotesen att vår modellstruktur är korrekt innebär att ett fel, dvs realiserad osäkerhet, i en punkt *måste* ha en statistiskt bestämd relation till fel i närliggande punkter. Att då påstå att felet, dvs osäkerheten i olika punkter är helt oberoende ger modellen mycket stora ‘anpassningsproblem’ och därmed väldigt liten osäkerhet. Vore det sant, skulle användandet av modeller vara ett fantastiskt sätt att reducera osäkerheten för en observation. Modellen skulle fungera som ett synnerligen effektivt brusreducerande lågpasfilter! Det stora problemet är dock att vi inte någonsin kan bevisa att en modellstruktur är absolut sann. Endast att förutsägelser som är statistiskt inkonsistenta med observationer förkastar hela modellen. Därför kan vi inte förlita oss på modellstrukturens korrekthet.

Okorrelerad osäkerhet för mätapparatur är också ett oriktigt men dessvärre helt dominerande påstående. Det är just bestämningen av korrelerade mätfel som är målet för t.ex. all dynamisk kalibrering. Att matcha modellstrukturens egenskaper med påstått okorrelerade kalibreringsdata ger därför på många sätt ett godtyckligt resultat som vanligtvis kraftigt underskattar modellens osäkerhet. Det enda vi vet är att vi inte vet! För exemplet ovan, reducerar denna insikt frihetsgraderna för kalibreringsdata till $1000 + 1000 \approx 2000$, eftersom att endast varianser men inte kovarianser kan anses kända. Detta ger ändå en redundans på $2000/65 \approx 30$ tvång för varje frihet, vilket betyder en marginal på 30 i entydighet och runt $\sqrt{7700/30} = 16$ ggr större samplingsvarians. Det ökar osäkerheten på ett sätt som motsvarar vårt faktiska underlag.

Mätosäkerheten är en osäkerhet för varje observation. Ingen modell kan reducera denna på annat sätt än att just modellen själv bidrar med *ytterligare* kunskap. Det är förkunskap på motsvarande sätt som Bayesiansk skattning tar in statistisk förkunskap. Modellstrukturen bidrar med *deterministisk förkunskap* genom att lägga till *korrelationer*, oftast som en effekt av modellens differential-/differensekvationers beteende. I princip reduceras osäkerheten på motsvarande sätt som för 'Bayes', med ett implicit logiskt 'och'. För att inkludera modellens förkunskap som en osäkerhetsreducerande komponent vid kalibrering så kan vi låta kalibreringsdata anta alla möjliga korrelationer (tvetydighet) och sedan välja det som modellen 'föredrar'. Dvs det den har lättast att anpassa sig till och som därför ger högsta modellosäkerhet. Det är nämligen precis dessa korrelationer som modellen skulle bidra med som förkunskap om den själv fick bestämma. I teorin ger det också en tänkbar metod för att justera kalibreringsdatas korrelation. Det är ekvivalent med att ta bort alla villkor på korrelation/kovarians, vilket är precis vad som görs i metod 2.

Problemdefinition

Modellkalibrering omfattar med nödvändighet många komplexa men generella begrepp som kan behöva förtydligas, eftersom deras betydelser varierar i litteraturen.

Med *sampel* avses här samtliga sampelvärden, inte varje enskild uppsättning modellparametrar. *Ensemble* betecknar en liknande uppsättning sampelvärden men utan att beskriva ursprunget. Beteckningen sampel sammanfaller här med vad som normalt avses inom matematisk statistik. Ensemble är en mer generell benämning som är mer lämpad för deterministisk sampling. Ensemblen kan vidare specificeras med vad den representerar eller refererar till. T.ex. *parameter-ensemble* för modellens parametrar, *modell-ensemble* för modellens resultat, eller *kalibrerings-ensemble* för att representera flera uppsättningar kompletta kalibreringsdata. Ensemblens olika sampelvärden kan också benämnas ensemblens *medlemmar* för att undvika association av vektorer och liknande strukturer till 'värden', vilket inte är naturligt.

En *representation* är en speciellt vald beskrivning av information men behöver inte motsvara typiska eller ens sannolika utfall. En representation ska ses som en abstrakt beskrivning, eller kompakt 'bokföring' av vad vi tror oss veta. Vad den beskriver, hur den

kan användas och vilka krav vi ställer beror på situation. Ett vanligt exempel på representation är Fourier-koefficienterna för en tidsberoende signal. Dessa beskriver signalens information eller innehåll men betyder inte att de harmoniska signaler med oändlig varighet koefficienterna refererar till är troligt förekommande, eller ens kan observeras överhuvudtaget. Vilket de faktiskt heller aldrig kan, eftersom varaktigheten är oändlig.

Variation av en variabel beskrivs i enklaste form av dess *varians*. *Sam-variation* av två variabler beskrivs i enklaste form av *ko-varians*. Kovarians i den generella mening som inkluderar både varians och kovarians ger den minsta möjliga *konsistenta* beskrivning av hur flera variabler samtidigt varierar. *Beroenden* mellan slumpvariabler syftar normalt på deras relativa *statistiska* samvariation. *Korrelation* syftar på väntevärdet av *två* olika variablers *relativa* variationer, alltså lägsta ordnings beskrivning av beroende. Om en variabel ökar, ökar då också oftast den andra variabeln (positiv korrelation) eller minskar den (negativ korrelation)? Korrelation är enklast tänkbara beskrivning av beroenden mellan två variabler.

Översikt

Med ett statistiskt osäkerhetsperspektiv på en komplex beräkning i många observationspunkter blir det tämligen många aspekter att samtidigt hålla i minnet. Det gäller i synnerhet vid kalibrering och speciellt när olika kalibreringsmetoder jämförs, som i detta projekt. En översikt där vanliga definitioner och etablerade metoder introduceras ges därför nedan i Fig. 1, för att placera de studerade metoderna i sitt sammanhang.

Benämningen *enkel osäkerhetskvantifiering* används här för att betona att den ej föregåtts av någon modellkalibrering, utan endast nyttjar förkunskap och annan information, samt antaganden.

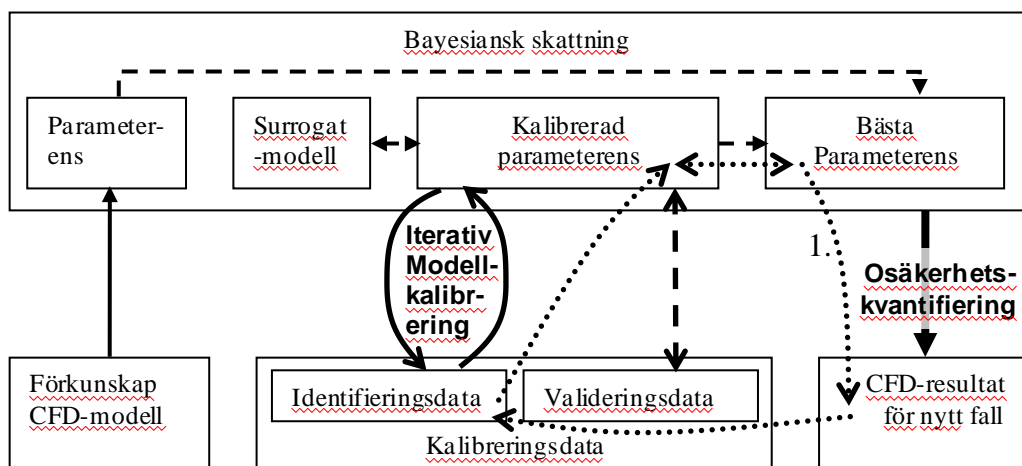


Fig. 1. Översikt. Iterativ modellkalibrering, [enkel] osäkerhetskvantifiering, konsistens av representation (1-2-3-4), samt Bayesiansk skattning [10] (exkluderad i projektet).

Båda metoder 1 och 2 för modellkalibrering i detta projekt är *iterativa*, dvs förbättrar modellen successivt genom upprepad anpassning. Iterationen är en specifik variant av vad som ofta benämns ‘fixpunktsiteration’ i matematik och ‘självkonsistent beräkning’ i fysik, vars konvergens garanteras i Banachs fixpunktssats [15]. Iterationen får dock inte förväxlas med Bayesian skattning [10] som lägger samman statistisk förkunskap och härledd kunskap i *ett* steg, se Fig. 1. I båda fallen finns en ingående *parameterensemble*, ‘nollensemblen’ $\Sigma_{[0]}$.

För Bayesian skattning representerar nollensemblen oberoende information som *allra sist* läggs samman med den härledda ensemblen $\Sigma_{[n]}$, för att erhålla bästa ensemble Σ^* . Den iterativa modellkalibreringen däremot, använder endast $\Sigma_{[0]}$ som en *första* gissning. Den avgör hur många iterationer som krävs för konvergens, dvs för vilket minsta värde på n vi erhåller $\Sigma_{[n+1]} \approx \Sigma_{[n]}$. Det som uppdateras vid iteration k är vår kunskap om modellens beteende via *modellensemblen* $H(r, \Sigma_{[k]})$, i *observationspunkterna* r , där $r = (x, y)$ för GEMIX experimentet [1]. Från denna kunskap beräknas, eller *identifieras* sedan den nya parameterensemblen, $H(r, \Sigma_{[k]}) \Rightarrow \Sigma_{[k+1]}$. Begreppet *kalibrering* inkluderar två delar, experimentet som genererar kalibreringsdata (*karaktärisering*), samt bestämning av modell från dessa (*identifiering*). Modellensemblen samplar beräkningsmodellen $H(r, \theta)$, som beskriver datormodellens ekvationer beräknade för *fältet* r och parametrar θ , på två helt olika sätt: Dels direkt i *observationsrummet* där modellen beräknas (r), dels statistiskt i *parameterrummet* (θ) för alla osäkra parametrar. Att särskilja de två helt olika samplingarna är centralt för att förstå vad modellkalibreringen gör, dvs justerar samplingen i parameterrummet för bästa sampling av resultat i observationsrummet.

Samplingspunkterna i observationsrummet är på förhand given och beror på vilken upplösning som eftersträvas. Ju fler punkter, desto mer redundans och säkrare identifiering. En vanlig tumregel är att antal relevanta justerbara ‘friheter’, eller *frihetsgrader* för kalibreringsdata ska vara minst 10 ggr fler än motsvarande frihetsgrader för modellen. Detta indikerar hur mycket kalibreringsdata som *minst* krävs. En försvårande praktisk aspekt för att kunna illustrera kalibreringsdata är att fältet r oftast omfattar mer än dimension, för GEMIX är det t.ex. två — koordinater längs (x) och tvärs (y) kanalen i blandningszonen. Med flera testfall (Nxxx) blir det i detta fall egentligen tre dimensioner. Även resultatet kan ha flera dimensioner. Skulle kalibreringen inkludera både koncentration och strömningshastighet så får modellens indata tre och utdata två dimensioner, dvs $H(r \in \mathcal{R}^3, \theta) \in \mathcal{R}^2$. Det är förstås relevant för CFD-beräkningen men dimensionaliteten är helt irrelevant för kalibreringen som sådan. Av detta skäl är det vid kalibrering lämpligt att skanna både indata och utdata seriellt till en enda lång sammansatt vektor $H(r \in \mathcal{R}^{N \times 1}, \theta) \in \mathcal{R}^{M \times 1}$, på identiskt lika sätt för kalibreringsdata och modellresultat. Det gör det mindre begripligt fysikaliskt, men förenklar analys och presentation eftersom enkla två-dimensionella grafer kan användas och tensorer undvikas. Statistisk sampling över osäkra parametrar θ ger sedan parameterensembler och

modellensembler på universal matematisk form $H(r \in \mathfrak{R}^{N \times 1}, \Sigma \in \mathfrak{R}^{n \times m}) \in \mathfrak{R}^{M \times m}$, för N observationspunkter, n parametrar och m statistiska sampelevärden.

Linjäritet

Linjäritet har en synnerligen central betydelse för de flesta matematiska operationer, så även osäkerhetskvantifiering. Det är dock lätt att blanda samman vad som avses eftersom eventuellt linjärt beroende kan referera till i varje fall tre helt olika klasser av data:

- Modellens *fältberoende*, dvs variation mellan olika observationspunkter, är nästan aldrig linjärt. I regel avgörs det av modellens differential/differensekvationer, samt variation i rand- och begynnelsevillkor.
- *Linjäritet-i-respons* avser modellens beroende på amplituden för insignal, rand- eller initialvillkor, dvs det som driver modellens resultat.
- *Linjäritet-i-parametrar* är det som är intressant för osäkerhetskvantifiering. Det handlar egentligen om linjärt *beroende*, eller *affinitet* dvs för (a, b) ,
 $f(a, b, x, y) = a \cdot g(x, y) + b \cdot h(x, y)$, snarare än linjäritet som definieras med *additivitet* och *homogenitet*, $f(a \cdot x + b \cdot y) = a \cdot f(x) + b \cdot f(y)$.

För modeller som är linjära-i-parametrar gäller således ($\delta\theta \equiv \theta - \langle \theta \rangle$),

$$f(\theta_1 + \delta\theta_1, \theta_2 + \delta\theta_2, \dots) = \delta\theta_1 \cdot \frac{\partial f}{\partial \theta_1} + \delta\theta_2 \cdot \frac{\partial f}{\partial \theta_2} + \dots \quad (1)$$

Det förenklar enkel osäkerhetskvantifiering dramatiskt och ligger till grund för en *känslighetsanalys*, där känsligheten normalt relaterar till konstant derivata $\partial f / \partial \theta_k$ och därför är begränsad till affina modeller.

Linjär regression

Affina modeller förenklar även kalibreringen, ibland kallad invers osäkerhetskvantifiering. Eftersom alla parametrar inte kan bestämmas från endast ett modellresultat i en punkt, kan inte modellens invers bestämmas direkt. Däremot kan ett antal ($m \geq n$) uppsättningar av värden på modellens (\tilde{H}) basfunktioner för de (n) osäkra parametrarna θ sammanställas i en *regressionsmatrix* R . För bästa modellapproximation \tilde{H} av kalibreringsdata H_C driven av insignal, rand- och begynnelsevillkor d , som är linjära-i-parametrar A , för observationer i punkter r , gäller att,

$$H_C(d, r, \Psi) \approx \tilde{H} \equiv A(d, r)R(\Psi), R = \begin{pmatrix} \Psi_1^{(1)} & \Psi_1^{(2)} & \dots & \Psi_1^{(m)} \\ \Psi_2^{(1)} & \Psi_2^{(2)} & \dots & \\ \vdots & & & \\ \Psi_n^{(1)} & & & \Psi_n^{(m)} \end{pmatrix}, A(d, r)^T = \begin{pmatrix} A_1(d, r) \\ A_2(d, r) \\ \vdots \\ A_n(d, r) \end{pmatrix}. \quad (2)$$

Notera att linjäriteten syftar på att $\tilde{H} \propto A$, men ställer *inga* speciella krav på $A(d, r)$. Specifikt krävs varken linjärt beroende på fältet r eller linjär respons på drivande storheter d så som inflödes hastighet i GEMIX experimentet, förutsatt att eventuella olinjäriteter kan uttryckas med $A(d, r)$. Testmodellen $H(x, \theta) = \sin(\theta_1 + \theta_2 x^3)$ kan t.ex. inte skrivas på denna form, men däremot $H(x, \theta) = \theta_1 x^2 + \theta_2 \sin(x^3)$. Benämningen linjär regression är därför lite förrädisk och betydligt mer generell av vad det kan verka. Linjär regression är metoden att finna bästa approximation genom att minimera [vanligen] summan av de kvadratiske felen i alla punkter r . Eftersom det är en mycket vanlig operation inom statistik och dataanalys [16], har de flesta matematiska programvaror färdiga funktioner för detta.

Det finns varianter av enkel linjär regression som t.ex. viktad anpassning. Ibland kan även inverteringen vålla numeriska problem om samplingspunkterna är för få eller inte är [tillräckligt] linjärt oberoende. Resultatet av att minimera det kvadratiske felet är att ekvation 2 ovan multipliceras med transponatet för att erhålla ett kvadratisk ekvationssystem $H_C R^T \approx A R R^T$, de s.k. *normalekvationerna* och studera egenskaperna/egenvärdena för den kvadratiske matrisen $R R^T$. Det ger lösningen

$$A_{\text{OPT}} = H_C R^T (R R^T)^{-1}. \quad (3)$$

Linjär regression används här för ett flertal ändamål direkt relaterade till modellkalibrering:

1. Bestämning av affin *surrogat*-modell.
(H_C ersätts av full beräkningsmodell H ovan)
2. Bästa anpassning av medelvärden och [ko-]varians, av beräkningsmodell till kalibreringsdata.
3. Stratifiering (beskriven nedan), bestämning av optimal viktning av sampel.
4. Bestämning av invers till icke-kvadratiske matriser, definierad av linjära ekvationssystem.

Surrogatmodeller (1) används ofta vid omfattande slumpad sampling eller andra ineffektiva metoder, för att ge rimliga beräkningstider. Viktad anpassning av en affin modell till kalibreringsdata (2) är huvudmomentet i den traditionella metoden 1 baserad på maximum likelihood skattning. Stratifiering (3) beskriver första steget i metod 2, där statistiken för kalibreringsdata överförs på den samplade beräkningsmodellen, genom att justera sampelviktningen men bevara samplingspunkterna, dvs ensemblen Σ . Inverser till icke-kvadratiske matriser (4) krävs för att undersöka *representationsgraden*, se avsnitt 'Konsistens'.

Enkel osäkerhetskvantifiering

Vid *osäkerhetskvantifiering* överförs statistisk information om modellens parametrar, rand- och begynnelsevillkor och andra förutsättningar till modellens resultat, vanligtvis ett tal, vektor eller fält. Den kanske vanligaste metoden, som också används för att bestämma mätosäkerhet är att linjärisera beräkningsmodellen i dess osäkra parametrar så att den blir

affin. Det motsvarar en s.k. känslighetsanalys. Används linjär regression kan en bättre approximation erhållas genom att anpassa över ett mer relevant omfång av de osäkra parametrarna. Variansen för modellresultaten ges sedan av,

$$\text{var}(H(r)) = \sum_{i,j=1}^n \frac{\partial H}{\partial \theta_i} \frac{\partial H}{\partial \theta_j} \langle \delta \theta_i \delta \theta_j \rangle = \sum_{i,j=1}^n K_i K_j \text{cov}(\theta_i, \theta_j). \quad (4)$$

Notera att antalet bidrag $n(n-1)/2$ från kovarianserna ($i \neq j$) är många fler än antalet n från varianserna ($i = j$). Det medför att beroenden nästan alltid ger stora effekter på resulterande varians.

Av alla kända metoder att kvantifiera osäkerhet är troligtvis slumpad sampling enklast att både förstå och utföra. Slumpad sampling är egentligen en direkt *simulering* av hur vi oftast föreställer oss osäkerhet för en modell: Enligt 'frekventist'-tolkningen [17] föreställer vi oss en antal oberoende upprepade beräkningar. Beroende på slump råkar de ha lite olika värden på modellparametrar och andra förutsättningar men beskriver för övrigt identiskt lika situationer. Det ger ett *statistiskt sampel* av beräkningen sedd som ett slumpobjekt, på samma sätt som en statistiker tar ett sampel från en *population* av t.ex. mänskliga egenskaper. Med population menas 'samtliga möjliga', medan ett slumpat sampel är ett 'troligt begränsat urval'.

Ett sampel av beräkningen betyder här ett sampel (θ) för varje observationspunkt (r). Vid den statistiska analysen av samplet föreställer man sig normalt endast en observationspunkt åt gången. Det är generellt inte helt korrekt eftersom modellen länkar samman resultat i olika punkter, vilket orsakar starka beroenden. Tror vi på modellens riktighet (något annat vore underligt eftersom vi faktiskt avser att använda den) ska resultatet i alla observationspunkter behandlas som en helhet. Det är synnerligen viktigt för modellkalibrering. För enkel osäkerhetskvantifiering syftar statistiska beroenden på hur olika parametrars fel statistiskt förväntas samverka [över alla tänkbara felkonstellationer].

Modellkalibrering

Modellkalibrering gör precis det omvända mot enkel osäkerhetskvantifiering. Statistisk information om resultatet för ett kalibreringsexperiment, s.k. kalibreringsdata överförs till en modell med tillhörande osäkerhet. Därför kallas den ofta för omvänd invers osäkerhetskvantifiering. Kalibrering har den uppenbara fördelen gentemot en enkel osäkerhetskvantifiering att informationen är *observerbar*. Statistiskt förväntade värden kan därför direkt skattas med statistiska metoder.

Traditionell kalibrering görs ofta genom att göra en skattning av bästa modell och dess osäkerhet med s.k. maximum likelihood-metodik [6][7]. Den börjar med att studera statistiken, eller multivariata sannolikhetstätheten $f(\Delta)$ för populationen av modellens H avvikelser mot kalibreringsdata H_C , för alla punkter r givet att alla parametrar θ är kända,

$$\Delta(r) = H(r, \theta) - H_C(r). \quad (5)$$

Sedan ställs en hypotes om att modellen är perfekt, vilket gör att Δ (här antagen kolumnvektor) endast består av mätfel. Normalt antas de vara slumpmässiga och ofta normalfördelade med kovarians $\text{cov}(H_C)$,

$$f_\Delta(\Delta) \propto \exp\left[-\frac{1}{2} \Delta^T \text{cov}^{-1}(H_C) \Delta\right]. \quad (6)$$

Problemet är nu att modellens parametrar är okända. Täthetsfunktionen $f(\Delta)$ är därför egentligen ingen täthetsfunktion alls, eftersom den beror på θ . För att tydliggöra detta kallar man den istället för 'likelihood-funktion' och ändrar beteckning, $f(\Delta) \rightarrow l(\theta, \Delta)$. Om parametrarnas rätta värden $\bar{\theta}$ vore kända skulle likelihood-funktionen övergå i sannolikhetstäthet, dvs $l(\bar{\theta}, \Delta) \rightarrow f(\Delta)$. Med ändliga kalibreringsdata är det p.g.a. samplingsvariens teoretiskt omöjligt att bestämma $\bar{\theta}$. Istället får vi nöja oss med en bästa skattning θ^* , dvs $l(\theta^*, \Delta) \approx f(\Delta)$. Alla andra parametervärden skulle lägga till en residual, dvs systematiskt modellfel och därmed öka Δ , dvs minska $l(\theta, \Delta)$. För att bestämma θ^* blir det därför naturligt att maximera $l(\theta, \Delta)$. Därav namnet 'maximum-likelihood'. Ofta är det numeriskt fördelaktigt att maximera logaritmen av likelihood-funktionen, $L(\theta, \Delta) = \log[l(\theta, \Delta)]$. Det blir en enkel kvadratisk form med normalfördelningsantagande, som i ekvation 6.

När väl θ^* har blivit funnen genom maximering, så omtolkas likelihood-funktionen $L(\delta\theta \equiv \theta - \theta^*, \text{cov}(\theta^*))$ till att bero på θ^* och $\text{cov}(\theta^*)$. De är bestämda av medelvärdet $\langle H_C \rangle$ och kovariansen $\text{cov}(H_C)$ av kalibreringsresultatet H_C . Med andra ord byts variabler från observation till modell. En sådan 'vändning' är egentligen inte tillåten. Det 'logiska glappet' kan emellertid förstås genom Bayesiansk skattning. Vad vi önskar är att bestämma modellens täthetsfunktion givet mätdata H_C , dvs den villkorliga (eng. *conditional*) sannolikhetstätheten $f_\theta(\theta|H_C)$. Ett betydligt enklare problem är det omvända, dvs bestämma mätningens täthetsfunktion givet rätt modell, $f_C(H_C|\theta)$. Då kan vi använda hypotesen i ekv. 6 om modellfelet direkt. Maximum likelihood metoden antar sedan att dessa är lika, $f_\theta(\theta|H_C) = f_C(H_C|\theta)$, vilket inte alls behöver vara sant. Bayes visade istället att,

$$f_\theta(\theta|H_C) = f_C(H_C|\theta) \cdot f_\theta(\theta) / f_C(H_C), \quad (7)$$

där $f_C(H_C|\theta)$ numera benämns just *likelihood-funktion* och beskriver kunskap härledd från ett kalibreringsexperiment, medan *förkunskapen* (eng. *prior knowledge*) $f_\theta(\theta)$ är känd *före*, eller mer korrekt *oberoende av* kalibreringen. Endast om $f_\theta(\theta)$ är likformig (konstant) och *omsluter hela* $f_C(H_C|\theta)$ reduceras detta till antagandet för maximum likelihood metoden, $f_\theta(\theta|H_C) = f_C(H_C|\theta)$.

Maximum likelihood metoden blir tillräckligt enkel för affina modeller för att ge explicita uttryck [18],

$$\begin{aligned}\theta^* &= \theta_0 + \text{cov}(\theta^*)K \text{cov}^{-1}(H_C)\Delta_0, \\ \text{cov}^{-1}(\theta^*) &= K \text{cov}(H_C)K^T\end{aligned}\quad (8)$$

där $\Delta_0(r) \equiv [H_C(r) - H(\theta_0, r)]$ är residualen för den okalibrerade modellen och K dess känslighet. Här finns ett par generella svårigheter:

- Standarden för kalibreringsexperiment inom mätteknik är *inte* anpassad till kraven för modellidentifiering. Specifikt betyder det att kalibreringsinformation inte är fullständig, eller har en konsistent approximationsnivå. Etablerade metoder kräver dock fullständighet och därför framtvingar *antaganden*.
- Modellkalibrering är en krävande invers operation som kan kräva drastiska modell-*approximationer*. Det kan helt ta udden av identifieringsmetodens prestanda. T.ex. är det inte meningsfullt med omfattande slumpsampling efter linjärisering, eftersom en enkel känslighetsanalys då är precis lika bra.

Det som kräver extra uppmärksamhet är således giltigheten för antaganden och de approximationer som användandet av surrogatmodeller innebär.

Identifierbarhet

Begreppet identifierbarhet [7] är centralt för modellkalibrering. Det uttrycker om det ens är möjligt att entydigt finna en modell, givet ett modellresultat. Är modellen inte identifierbar betyder det att minst två olika parameteruppsättningar ger samma modellresultat. Traditionellt syftar det endast på bestämningen av bästa skattning och utgår från säkra resultat, utan osäkerhet. Matematisk uttrycks det som,

$$H(r, \theta^{(1)}) = H(r, \theta^{(2)}) \Rightarrow \theta^{(1)} = \theta^{(2)}.\quad (9)$$

Det generaliserar i princip begreppet inverterbarhet för enkla funktioner av en variabel till modeller, något som logiskt krävs för att lösa det inversa kalibreringsproblemet.

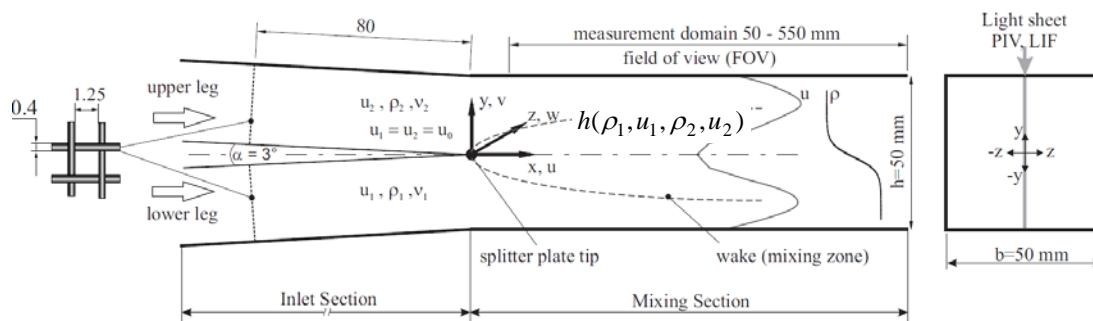
Konsistens

Ett krävande test av olika metoders konsistens och representativitet är att omsluta hela osäkerhetsanalysen i ett cirkelresonemang, för att undersöka om ursprungliga förutsättningar och antaganden återfås. I Fig. 1 illustreras en konsistensbestämning med sekvensen 1-2-3-4: Identifieringsmetoden testas genom att man låter modellen som ska identifieras först generera ett resultat (1) som får bli kalibreringsdata (2). Därefter ändras parametervärden innan modellen identifieras (3), i förhoppning att 'hitta tillbaka' till ursprungliga värden (4). Normalt begränsas sådana konsistenskontroller till bästa skattning och innefattar ej deras osäkerhet. I detta projekt är emellertid konsistensen i representationen av osäkerhet central. Full *representationsförmåga* betyder här att även

osäkerheten återskapas fullt ut i en konsistensbestämning, medan ingen representationsförmåga av osäkerhet motsvaras av att en helt säker modell identifieras i steg (3). Representation av osäkerhet undersöks här med parameter-*ensembler*, snarare än parametrarnas statistik. Rätt hantering av kovarians för kalibreringsdata är centralt för att representationsförmågan ska kunna anses vara relevant.

GEMIX kalibrering

Benchmark-problemet 'GEMIX' inom OECD-projektet "Uncertainty Methods for CFD Application to Nuclear Reactor Thermalhydraulics" [1] beskriver mixning av två flöden. GEMIX är en förkortning av "GEneric MIXing" experiment, som har genomförts vid Paul Scherrer Institut i Schweiz. Det finns en rad situationer inom kärnkraftsektorn där mixning är av intresse; borinblandning kan påverka stabiliteten i reaktorn, blandning vätgas och luft i inneslutning och reaktorbyggnad kan leda till explosion, varmt och kallt vatten i T-stycken och styrtavar kan leda till utmattningssprickor, etc. En schematisk illustration av GEMIX vätskeflödesproblem visas i Fig. 2 nedan.



Inlet section length	1250	Total channel length	3000
Inlet section height x width	25x50 (2x)	Nominal inlet velocities	0.2-1.2 m/s
Honeycomb $d = 2, l = 50$	@ $x = -670$	Re-Numbers in mixing section	10,000 – 60,000
1. Grid $d = 1, w = 4$	@ $x = -520$	Volumetric flow rate	15 – 90 l/min
2. Grid $d = 1, w = 4$	@ $x = -300$	Density difference	0 – 10%
3. Grid $d = 0.4, w = 1.25$	@ $x = -80$	Temperature difference	0 – 50K
Mixing section length	550	Viscosity difference	0 – 100%
Mixing section height x width	50x50	Streamwise turbulence level	$\approx 5\%$

Fig. 2. (Reproducerat [1].) GEMIX benchmark-problem. Modellen $h(\rho_1, u_1, \rho_2, u_2)$ representerar CFD-beräkning som bestäms av inkommande densiteter ρ_1, ρ_2 (ger Froude-talet, samt flödes hastigheter u_1, u_2 från de två tilloppsrören (ger Reynolds tal).

Mixningsexperimentet fokuserar på grundläggande turbulenta blandningsmekanismer. Flödeskanalen är tillverkad av genomskinlig akryl för att kunna observera flödet, med undantag av 80 mm långa delningsplattan ('splitter plate tip'). För att ge tillräcklig mekanisk stabilitet är den gjord av rostfritt stål.

I detta projekt kalibrerades modellen mot flödes hastighetens fält $u(x, y)$. Totalt tre olika fall användes för kalibrering (C), medan ett sparades för validering (V), se Fig. 3. nedan. Det ger ett kalibreringsfält i form av en tensor av ordning 3, $u(C_k, x_k, y_k)$.

Nominell inloppshastighet (m/s)	0.6	1	Upplösning x (mm)	100
Globalt Reynoldstal	30000	50000	Upplösning y (mm)	1
$\Delta\rho = 0\%, \Delta T = 0K$	C ₁ :N339	C ₂ :N337	Omfång x (mm)	[50,450]
$\Delta\rho = 1\%, \Delta T = 2.5K$	C ₃ :N320	V:N318	Omfång y (mm)	[-21,22]
Kalibreringsdata (C)	N320 - N337 - N339		Valideringsdata (V)	N318

Fig. 3. Experimentell design av kalibrerings- (C) och valideringsexperiment (V). Upplösning och omfång gäller för modellresultat, dvs beskriver valda kalibreringspunkter.

Anpassning och konvertering

Modellresultatens omfång har reducerats till att ligga inom de officiellt tillgängliga mätresultatens intervall över samtliga mätningar, för att undvika riskabel extrapolation. Det reducerade tvärkoordinaten, $y \in [-25, 25] \Rightarrow y \in [-21, 22]$ mm. Mätpunkterna sammanföll inte med de jämnt fördelade modellpunkterna. För att bli jämförbara interpolerades därför kalibreringsdata två-dimensionellt med en kubisk spline.

För att enkelt kunna jämföra och illustrera i 2D-grafer konverterades kalibreringspunkterna till en vektor, genom skanning av hela kalibreringsfältets tensor. Variablerna varierades i tur och ordning, för att ge kalibreringsvektorn $u(C_k, x_k, y_k) \in \mathfrak{R}^{3 \times 5 \times 44} \rightarrow u_C \in \mathfrak{R}^{660 \times 1}$, som sekventiellt visar flödes hastigheten tvärs kanalen y , i de olika punkterna x längs blandningsområdet, för vardera fall C . Med andra ord visar de första 220 punkterna av u_C flödes hastigheten för 'N320', varav de första 44 punkterna motsvarar position $x=50$ i blandningsområdet 'mixing section' (Fig. 2). Valideringsdata innefattade endast ett fall, dvs $u(V_k, x_k, y_k) \in \mathfrak{R}^{1 \times 5 \times 44} \rightarrow u_V \in \mathfrak{R}^{220 \times 1}$. Hur fallen arrangerats framgår att sista raden i Fig. 3.

Strålsäkerhetsmyndigheten tillhandahöll beräkningsstöd för motsvarande deterministiska (givna parametrar) k-epsilon CFD-modell, för de parameterkombinationer som beräknats i projektet.

Startensemble

Utgångspunkten för kalibreringen gavs av modellens ursprungliga *parameterensemble*, som bestämts, eller syntetiserats från uppskattade bästa värden och osäkerheter enligt Fig. 4.

$$\Sigma_0 = \begin{matrix} c_\mu : \\ C_{\varepsilon 2} : \\ \sigma_k : \\ K : \\ \sigma_T : \\ u_{IN} : \\ k_{IN} : \end{matrix} \begin{pmatrix} 0.0892 & 0.0844 & 0.0892 & 0.0844 & 0.0892 & 0.0844 & 0.0892 & 0.0844 \\ 2.0990 & 2.0990 & 1.8010 & 1.8010 & 2.0990 & 2.0990 & 1.8010 & 1.8010 \\ 1.0167 & 1.0167 & 1.0167 & 1.0167 & 0.9833 & 0.9833 & 0.9833 & 0.9833 \\ 0.4051 & 0.4149 & 0.4149 & 0.4051 & 0.4051 & 0.4149 & 0.4149 & 0.4051 \\ 0.6630 & 0.6630 & 0.8370 & 0.8370 & 0.8370 & 0.8370 & 0.6630 & 0.6630 \\ 1.0050 & 0.9950 & 1.0050 & 0.9950 & 0.9950 & 1.0050 & 0.9950 & 1.0050 \\ 0.9850 & 1.0150 & 1.0150 & 0.9850 & 1.0150 & 0.9850 & 0.9850 & 1.0150 \end{pmatrix}. \quad (10)$$

För varje sampel, eller parameteruppsättning (kolumn av Σ_0), så erhålls en komplett kalibreringsvektor, dvs ett fullständigt resultat $u_{h,C} \in \mathfrak{R}^{660 \times 1}$, för CFD-modellen $h(\rho_1, u_1, \rho_2, u_2)$. Alla delresultat ger *modellensemblen* $u_{h,C}(\Sigma_0) \in \mathfrak{R}^{660 \times 8}$.

Parameter			Constants
c_μ	Weibull	$g(x) = \frac{\alpha_w}{\beta_w} \left(\frac{x}{\beta_w}\right)^{\alpha_w-1} e^{-(x/\beta_w)^{\alpha_w}}$ As in ref [3]	$\alpha_w = 45.54$ $\beta_w = 8.77 \times 10^{-2}$ $\langle c_\mu \rangle = 0.0868$ $\sigma_{c_\mu} = 2.4 \times 10^{-3}$
$C_{\varepsilon 2}$	Beta	$g(x) = \frac{(x - A_1)^{p-1} (A_2 - x)^{q-1}}{\int_{A_1}^{A_2} (x - A_1)^{p-1} (A_2 - x)^{q-1} dx}$ As in ref [3]	$A_1 = 1.80, A_2 = 2.20$ $p = 1.20, q = 2.0$ $\langle C_{\varepsilon 2} \rangle = 1.95$ $\sigma_{C_{\varepsilon 2}} = 0.149$
σ_k	Normal	$g(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-(\sigma_k))^2}{2\sigma^2}}$ As in ref [3]	$\langle \sigma_k \rangle = 1.0$ $\sigma = 1.67 \times 10^{-2}$
κ	Normal	$g(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-(\kappa))^2}{2\sigma^2}}$ As in ref [3]	$\langle \kappa \rangle = 0.41$ $\sigma = 4.89 \times 10^{-3}$
σ_T		$0.6 < \sigma_T < 0.9$	$\langle \sigma_T \rangle = 0.75$ $\sigma_{\sigma_T} = 0.087$
Factor U_{inlet}	1% flow difference between inlets	Estimated from measured inlet profiles	$1 \pm \sigma$ $\sigma = 0.005$ 1.005 upper inlet 0.995 lower inlet Or vice versa
Factor k_{inlet}	Estimated variation in k-profile	Estimated from measured inlet profiles	$1 \pm \sigma$ $\sigma = 0.015$

Fig. 4. Ursprungliga uppskattningar av parametrar med osäkerheter.

Resultat

Traditionell mot ny modellkalibreringsmetodik

De skattningstekniker som alla etablerade kalibreringsmetoder bygger på kräver en bestämd, eller 'säker' modellhypotes. För modellkalibrering betyder det en specifik beräkningsstruktur, eller modellekvationer, alternativt en uppsättning av sådana. Inom matematisk statistik testas hypotesens riktighet genom fysisk sampling och diverse standardiserade tester. Sådana tester har en direkt motsvarighet för modellkalibrering som benämns *validering*. Resultatet är 'digitalt', antingen förkastas hypotesen fullständigt eller inte alls. Det senare är emellertid inte detsamma som acceptans, bara att vi ännu inte funnit orsaker att diskvalificera den. Det finns således ingen nivå mellan 'diskvalificerad' och 'ännu inte diskvalificerad'. Valideringen jämför modellens kvalitet med kalibreringsdatas kvalitet, vilken inte alls behöver motsvara kraven för tilltänkt *användning*. Kvalitetskrav och relevans är olika aspekter — valideringens experimentella design kan vara relevant för tilltänkt användning men mätosäkerheten kan vara långt bättre än vad användningen kräver. Det senare är mer troligt än otroligt, eftersom valideringen ofta görs med högkvalitativa instrument, under bättre kontrollerade förhållanden än i den verkliga användningen. Ofta sannolikt utan tanke på vad modellen ska användas till. Modeller kan därför av misstag lätt förkastas, trots att de kan vara tillräckligt bra. Metod 1 utgår från sådan traditionell statistisk skattningsteknik.

Representationstekniker torde i princip vara enklare att förstå än traditionella skattningstekniker. Representationer uttrycker vad vi vet i den 'bas' vi har tillgänglig. Det som inte kan beskrivas förloras till ett okänt fel som avgörs av representationens fullständighet. Därför är det viktigt att uppskatta hur stort detta fel rimligen skulle kunna vara, en formulering som stämmer väl överens med den generella iden bakom osäkerhet. En angiven osäkerhet behöver inte vara 'sann' men däremot rimlig och inte alltför överdriven, för att vara praktiskt användbar. Framförallt krävs ingen fullständighet för en representation, varken för informationen den beskriver eller dess bas. Det är pragmatiskt och därför också praktiskt. Exempel på välkända representationer är ortogonala funktionsserier och Fourierserieuppdelning av signaler. Däremot har representationstanken inte ännu använts för att uttrycka statistisk information med ett begränsat antal sampel, förutom i de deterministiska samplingsmetoder som författaren bidragit till.

Istället för att uttrycka en funktion i en bas av harmoniskt oscillerande basfunktioner som i Fourieranalys, uttrycks statistisk information i detta projekt i ett modellsampel bestående av en uppsättning av olika parameterkombinationer. Statistiken för modellsamplet ska sammanfalla med känd statistisk information, på samma sätt som Fourierserien ska beskriva den funktion eller signal vi studerar. För en representation finns ingen validering i normal bemärkelse. Däremot går det att studera *representationsförmåga*, dvs hur mycket av tillgänglig kunskap som kan beskrivas med vår valda bas. Bestämning av denna representationsgrad motsvarar validering. Dessutom bör den göras för data som inte använts för identifiering, precis som normal validering. En stor skillnad mot skattningstekniker är att utfallet vid denna 'validering' är kontinuerligt, t.ex. kan modellens förutsägelse vara konsistent med kalibreringsdata i 93% av alla observationspunkter. Är det inte acceptabelt kan vi revidera vår information för att erhålla en ny representation med lämpligare konsistensnivå. Väsentligt är att vi inte ställt någon ovillkorlig hypotes. Metod 2 bygger på en sådan representationsteknik.

Att använda representationsteknik för kalibrering är också konsekvent med de idéer om representation av statistisk information för enkel osäkerhetskvantifiering som föreslogs i föregående projekt [2]. Att 'representera tillgänglig kunskap som sedan propageras', antingen från modell till resultat, eller tvärtom, blir då en grundprincip som används för både osäkerhetskvantifiering och modellkalibrering. Skillnaden består endast i vilken 'riktning' detta sker, direkt (eng. forward/direct UQ) eller omvänt (eng. inverse UQ) relativt modellbeskrivningen. Kort sagt, "Vi beskriver allt vi *vet* och som är *relevant*, så *bra* som möjligt". Matematiska modeller blir endast hjälpmedel för att uttrycka vår kunskap. Tar vi med den kunskap som modellerna implicit beskriver, måste också dess kvalitet på något sätt vara känd, kvantifierbar och transporterats genom analysen. Ignoreras modellens existens totalt, som för enkel osäkerhetskvantifiering, så undanhålls känd information. Med mindre information borde osäkerheten i allmänhet överskattas. Fig. 11 visar precis detta. Kalibreringen drar ner denna osäkerhet genom att *både* lägga till korrelationer mellan ingående parametrar och justera varianser, se Fig. 14. Kalibreringen blir alltså de facto en metod att finna just de korrelationer som modellen kan generera. Det har vi förtroende för eftersom vi avser använda modellen.

Konsistens

Den etablerade metodiken maximum likelihood skattning, som ligger till grund för metod 1 och som ligger nära statistisk signalbehandling [6] uttrycks inte normalt som en representation. Istället refererar osäkerheten till en bästa skattning. Det gör t.ex. att osäkerheten inte kan innefatta residualinformation, eftersom den bästa skattningen enligt definition inte alls kan beskriva residualen. Osäkerheten beskriver därför inte all osäkerhet vi faktiskt har kännedom om. Den beskriver *skattningens* och inte modellens osäkerhet. Frågan blir då om skattningsmetodikerna och representationsmetodikerna ändå är jämförbara i något avseende. Kan osäkerheten för bästa skattning av modellen härbärgera *all* osäkerhet, förutsatt att den har möjlighet till det? De två metoderna kan i så fall anses ekvivalenta, förutsatt att modellen i båda fall klarar validering. En metods *konsistensgrad* kan definieras av hur stor andel av kvantifierad osäkerhet som återfås vid kalibrering.

Enkel osäkerhetskvantifiering av den affina modellen

$H = H_0 + K_1(r)\delta\theta_1 + K_1(r)\delta\theta_1 + \dots = H_0(r) + K(r)\delta\theta$ ger enligt ekv. 4 kovariansmatrisen,

$$\text{cov}(H_C) = K \text{cov}(\theta) K^T, \quad (11)$$

som här omklassificeras att gälla för kalibreringsdata för att kunna bestämma konsistensgraden. Maximum likelihood metoden ger direkt kovariansen för bästa skattning θ^* av den identifierade modellen enligt ekv. 8,

$$\text{cov}^{-1}(\theta^*) = K^T \text{cov}^{-1}(H_C) K. \quad (12)$$

Sätts ekv. 11 in i ekv. 12 erhålls,

$$\text{cov}^{-1}(\theta^*) = K^T [K \text{cov}(\theta) K^T]^{-1} K = K^T K^{T,-1} \text{cov}^{-1}(\theta) K^{-1} K \equiv A^T \text{cov}^{-1}(\theta) A. \quad (13)$$

Eftersom känslighetsmatrisen K inte är kvadratisk, får 'inversen' K^{-1} utifrån ovanstående multiplikation av $A = K^{-1}K$ från höger, *tolkas* som lösningen för regressionsproblemet,

$$XK = Y \Rightarrow X = YK^T (KK^T)^{-1} \equiv YK^{-1}. \quad (14)$$

Med användning av linjär regression för att bestämma K^{-1} erhålls,

$$A \equiv K^{-1}K = K^T (KK^T)^{-1} K \Rightarrow AA^T = A \Leftrightarrow A = I \Rightarrow \text{cov}(\theta^*) = \text{cov}(\theta). \quad (15)$$

Alla kovariansselement $\text{cov}(\theta)$ för den ursprungliga modellen H representeras till fullo i bästa skattningens osäkerhet $\text{cov}(\theta^*)$. Representationsgraden är således 100%. Om alla matriselement i $\text{cov}(H_C)$ representerar vad vi anses veta, så ger maximum likelihood-metoden en relevant osäkerhet.

Om vi däremot *gissar* $\text{cov}(H_C(r_i), H_C(r_{j \neq i}))$ så är det mycket svårt att säga mer än att $\text{cov}(\theta^*)$ också är en gissning av samma dignitet. Det gäller även om varianserna är väl kända eftersom de är i minoritet av alla bidrag, enligt ekv. 4. Därför kan vi minska

osäkerheten, eller nästan eliminera den helt, genom antaganden som strider mot modellens egenskaper: Vad vi definitivt vet om *alla* modeller, är att de ger starkt korrelerade resultat, eller 'mönster'. Det är nämligen just sådana samband som får oss att formulera 'modeller' överhuvudtaget. Med ett ansenligt mått av 'Willful ignorance' [11] kan vi kanske ändå förtränga detta uppenbara faktum helt och hållet och istället påstå att vi inte vet något alls. Det vore i princip inte fel men torde ge en högre osäkerhet än vad den borde vara, $\text{cov}(\theta^*) > \text{cov}(\theta)$, p.g.a. av att modellens deterministiska förkunskap ignoreras.

Om vi däremot också påstår att observationsfelen är oberoende av varandra *eftersom de är okända* begås ett mycket vanligt misstag. Det *minskar* osäkerheten kraftigt(!), $\text{cov}(\theta^*) \ll \text{cov}(\theta)$. Inkonsekvensen är uppenbar genom att samtidiga påståenden om beroende (modellens existens) och oberoende (vår s.k. okunskap) knappast kan anses förenliga i någon form. Det enda som ger någon återstående osäkerhet vid prediktion med kalibrerad modell är samplingsvariansen för samplet av kalibreringsdata. Enligt konventionell tolkning av kalibrering,

*Med tillräckligt antal punkter för kalibreringsdata kan osäkerheten för bästa skattningen av identifierad modell **alltid** drivas ned till **godtyckligt lågt** värde.*

Vi kan uppenbarligen bli precis hur säkra som helst på vår modell, utan att vi testat dess absoluta riktighet (systematiska fel). En giltig validering avslöjar endast att modellens systematiska fel understiger mätosäkerheten för kalibreringsdata, ett krav som varken har något med $\text{cov}(\theta^*)$ att göra, eller behöver ha någon som helst relation till den tilltänkta användningen.

Det finns inte något förlåtande med gissningar och ogrundade antaganden. Matematiskt kan gissningar inte särskiljas från kunskap. Gissningar blir i analysen ekvivalent med kunskap. Om det vi inte anser oss ha någon kunskap alls, blir vilket som helst antagande lika acceptabelt som något annat, oberoende av vad det leder till.

Vore alla observationsfel helt okorrelerade med varandra, skulle t.ex. all mätapparatur få oändlig bandbredd, vilket knappast kan anses korrekt. Problemet är att mätteknik p.g.a. bristande förståelse för modellering i allmänhet inte inkluderar kovarians i begreppet mätosäkerhet, trots att i princip all modern signalbehandling gör det. Det orsakar svårigheter vid modellering eftersom oberoendeantaganden knappast kan anses korrekta. Den enda logiska konsistenta alternativet blir därför att anse dem okända och inte göra några antaganden alls, som i metod 2. Det gör att vi inte kan omsätta modellekvationernas förkunskap till fullt och att osäkerheten därför konsekvent borde överskattas, snarare än underskattas som i konventionell kalibrering.

I praktiken är problemet allvarligt, eftersom alla existerande metoder för identifiering havererar p.g.a. ofullständig information. Identifiering bör i sådana fall ersättas med enkel kurvanpassning som inte tillåter jämförelser överhuvudtaget, eftersom osäkerhetsintervall då inte anges. Den fundamentala skillnaden mellan naiv kurvanpassning och kalibrering består i hanteringen av osäkerhet. Det vassa instrumentet för kvalificerat beslutsfattande

som en korrekt kalibrerad modell erbjuder har då kollapsat. Detta faktum är den fundamentala orsaken till att metod 2 föreslogs i detta projekt.

Kalibrering av deterministiska ensembler

Utökningen av identifierbarhetsbegreppet i avsnitt Identifierbarhet till *osäkra modeller* beskrivna med deterministiska ensembler är mer komplicerad än vad som kanske är förväntat. Kravet på identifierbarhet kan uttryckas,

$$H(r, \Sigma^{(1)}) = H(r, \Sigma^{(2)}) \Rightarrow M(\Sigma^{(1)}) = M(\Sigma^{(2)}), \quad (16)$$

där operatoren $M(\Sigma)$ genererar all relevant statistik för ensemblen Σ . Två olika ensembler kan emellertid ha samma relevanta statistik. Dvs,

$$M(\Sigma^{(1)}) = M(\Sigma^{(2)}) \not\Rightarrow \Sigma^{(1)} = \Sigma^{(2)}. \quad (17)$$

Med inskränkningen till samma typ av ensemble $\hat{\Sigma}$, dvs unik *syntes* eller regel att bestämma $\hat{\Sigma}$ utifrån $M(\hat{\Sigma})$ så gäller dock för lämpliga pivoteringar $P_{1,2}$ av samplets parameteruppsättningar att,

$$M(\hat{\Sigma}^{(1)}) = M(\hat{\Sigma}^{(2)}) \Rightarrow P_1(\hat{\Sigma}^{(1)}) = P_2(\hat{\Sigma}^{(2)}). \quad (18)$$

Därför är parameterensemblen identifierbar med begränsning till varje unik syntes.

Fixpunktsiteration — successiv förbättring

Motiven för att iterera kalibreringen är desamma för båda metoder 1 och 2: Den statistiska ensemble (Σ) som representerar den kalibrerade modellen bör helst sammanfalla med de parameteruppsättningar, eller ensemble (Ψ) som använts för att undersöka modellen under dess identifiering. Iterationen sammanför dessa två ensembler successivt, med förhoppningsvis allt mindre avvikelse i varje steg. Kalibreringen för den sista iterationen görs därmed för samma information som den kalibrerade modellen representeras med. Symboliskt kan detta formuleras,

$$\Sigma_0 \equiv \Psi_1 \xrightarrow{\text{IDENT}} \Sigma_1 \equiv \Psi_2 \rightarrow \Sigma_2 \equiv \dots \rightarrow \Sigma_n \equiv \Psi_{n+1} \rightarrow \Sigma_{n+1} = \Sigma_n = \Psi_n, \quad (19)$$

där \equiv betecknar tilldelning och \rightarrow motsvarar identifiering/kalibrering med metod 1 eller 2. Om t.ex. en surrogatmodell bestäms med hjälp av linjär regression som i kalibreringsmetod 1, så minimeras approximationsfelet i regressionspunkterna. Samplar vi sedan den kalibrerade modellen i just dessa regressionspunkter där surrogatmodellen är som bäst, minimeras samplingsfelet. Det finns emellertid två viktiga aspekter att beakta:

- Metodens komplexitet (Ψ) har ingen given relation till resultatets representation (Σ) : Ensemblen Σ kan därför i princip vara otillräcklig för att användas som Ψ . Kompletteringar med fler punkter kan därför behövas för att genomföra analysen.

- Iterationen behöver inte nödvändigtvis konvergera till en fix ensemble, s.k. fixpunkt. Speciella metoder kan krävas för att uppnå konvergens.

Den första aspekten är i synnerhet viktig för metod 1 som använder surrogatmodeller. Den andra aspekten är mest kritisk för metod 2, eftersom den flyttar parameterensemblen direkt och inte via parametrarnas statistik, t.ex. medelvärden och kovarians. Direkt förflyttning betyder här att ett helt ‘moln’ av punkter flyttas i metod 2. Det är mycket svårare än att flytta deras tyngdpunkt (bästa skattning) och därefter bestämma ensemblens spridning som i metod 1. Nedan ges en något förenklad sammanfattning av hur metoden fungerar. En detaljerad beskrivning som behövs för att kunna tillämpa metoderna är för omfattande för att ges här och kommer istället att ges i en planerad vetenskaplig artikel [19].

Metod 1: Maximum likelihood skattning

Metoden för itererad maximum likelihood skattning kombinerar den etablerade maximum likelihood metoden beskriven i avsnitt Modellkalibrering med iteration enligt avsnitt Fixpunktsiteration — successiv förbättring, samt användande av en linjär surrogatmodell bestämd med linjär regression enligt avsnitt Linjär regression.

Representation av medelvärde och kovarians ger en tillräckligt stor ensemble för bestämning av en linjär (affin) surrogatmodell. Därför behövs inga ytterligare regressionspunkter (Ψ) nedan, utöver det som parameterensemblen (Σ) ger. Kalibreringen upprepas/itereras enligt följande,

1. Välj startensemble $\Sigma_{[0]}$ indexerad med $k = 0$.
2. Sätt regressionspunkter lika med aktuell parameterensemble, $\Psi_{[k+1]} = \Sigma_{[k]}$.
3. Beräkna modellensemble $H(r, \Psi_{[k+1]})$ för alla regressions- ($\Psi_{[k+1]}$) och observationspunkter r .
4. Bestäm en affin surrogatmodell $H_{\text{SUR}}(r, \theta) = A(r)\theta$. Fältet $A_{[k+1]}(r)$ bestäms då från den linjära ekvationen $A_{[k+1]}(r)R(\Psi_{[k+1]}) = H(r, \Psi_{[k+1]})$ med regressionsmatrix $R(\Psi_{[k+1]})$. Extrahera/beräkna känsligheter $K(r)$ från $A_{[k+1]}(r)$.
5. Propagera kalibreringsdata $H_C(r)$ till ny bästa skattning $H_{\text{SUR}}(r, \theta_{[k+1]}^*) = A_{[k+1]}(r)\theta_{[k+1]}^*$ av surrogatmodellen, under antagandet att den nya avvikelserna $\Delta_{[k+1]}(r) = H_C(r) - H_{\text{SUR}}(r, \theta_{[k+1]}^*)$ är normalfördelad,

$$\begin{aligned} \theta^* &= \theta_0 + \text{cov}(\theta^*)K \text{cov}^{-1}(H_C)\Delta_{[k]} \\ \text{cov}^{-1}(\theta^*) &= K \text{cov}(H_C)K^T \end{aligned}$$

6. Sampla θ^* , $\text{cov}(\theta^*) \equiv U^T S^2 U$ genom att syntetisera en ny parameterensemble Σ_1 med excitationsmatris V , precis som vid enkel osäkerhetskvantifiering med deterministisk sampling [4],

$$\Sigma_{[k+1]} = \theta^* + U^T S U V .$$
7. Låt $k \rightarrow k+1$ och upprepa från steg 2 tills en fixpunkt för parameterensemblen erhålls, $\Sigma_{[n+1]} \approx \Sigma_{[n]}$.

Denna metod är identiskt lika med tidigare publicerad metod [18].

Metod 2: Stratifiering och annealing

[Matematiskt avancerat avsnitt]

Grundidé

Metoden baserad på stratifiering och annealing innehåller också två steg. Istället för ett mellansteg med linjär regression för att bestämma en surrogatmodell som i metod 1, görs en omviktning av parameterensemblen, också med linjär regression. Symboliskt kan metoden illustreras med att omväxlande införa olikformiga vikter (δW), s.k. *stratifiering* och justera ensemblen ($\delta \Sigma$) för att återställa dessa vikter att åter bli lika, s.k. *annealing*,

$$\Sigma_0 \xrightarrow{\text{STRAT}} \delta W_1 \xrightarrow{\text{ANN}} \delta \Sigma_1 \rightarrow \delta W_2 \rightarrow \delta \Sigma_2 \rightarrow \dots \rightarrow \delta W_n = 0 \rightarrow \delta \Sigma_n = 0 \rightarrow \Sigma_{n+1} = \Sigma_n . \quad (20)$$

Ordet stratifiering har sitt ursprung från stratifierad sampling inom matematisk statistik. Det betyder att samplingsområdet delas in i flera delområden som samplas var för sig. Delresultaten *viktas* sedan samman med de sannolikheter de olika områdena motsvarar. Motivet är att få ett bättre och mer representativt sampel i intervall med låg sannolikhet och få sampel. Intervallen viktas sedan efter deras totala inflytande, dvs sannolikhet. Önskan att erhålla bättre representativitet för modellen med olikformig viktning i metod 2, givet den kända fördelningen för kalibreringsdata, stämmer mycket väl överens med motiven bakom normal statistisk stratifiering.

Benämningen annealing motsvarar *glödning* på svenska, vilket beskriver en fysikalisk process som minskar spänningen i ett material genom att molekyler flyttas under inverkan av värme. Olika vikter kan på likartat sätt också anses ge upphov till en 'spänning' inom ensemblen. Denna minskas här genom att flytta dess medlemmar i 'gradientens' riktning, dvs vill säga den riktning som gör vikterna mer lika. Metodiken efterliknar således den naturliga justeringen av atomer i material med inre spänningar som sker under uppvärmning vid 'glödning'. Annealing är för övrigt ett populärt ord som används i många sammanhang, 'laser annealing' och 'stimulated annealing' är två andra exempel.

Fördelen med att justera modellens statistik med vikter är att det matematiska problemet att anpassa till statistiska moment är linjärt. För m samplade värden av en osäker parameter θ ges det $p(\geq 2)$ -te momentet av deterministiska modellensembeln av,

$$M^{(p)} = \frac{1}{m} \sum_{k=1}^m W_k [\delta H(\Sigma^{(k)})]^p \equiv [\delta H(\Sigma)^{\bullet p}]^T W, \quad \delta H(\Sigma) = H(\Sigma) - \langle H(\Sigma) \rangle \otimes 1^{1 \times m}, \quad (21)$$

där $\{X^{\bullet p}\}_{jk} = X_{jk}^p$ och $\langle H(\Sigma) \rangle = H(\Sigma)W$. Här omskrivet på matrisform blir det linjära beroendet i vikter W_k uppenbart. Det gör att matchningen av samtliga statistiska moment av [beroende] parametrar blir linjär. Därför kan de lösas ut direkt från matchingsvillkoren,

$$\begin{aligned} \langle 1 \rangle &= (1 \ 1 \ \dots \ 1) && W \\ \langle H(x, \theta) \rangle_{\theta} \in \mathfrak{R}^{q \times 1} &= (H(x, \Sigma^{(1)}) \ H(x, \Sigma^{(2)}) \ \dots \ H(x, \Sigma^{(m)})) && W \\ \left\langle \underbrace{\delta H(x_j, \theta)}_{\equiv \delta H_j} \underbrace{\delta H(x_k, \theta)}_{\equiv \delta H_k} \right\rangle &= (\delta H(x_j, \Sigma^{(1)}) \delta H(x_k, \Sigma^{(1)}) \ \delta H_j^{(2)} \delta H_k^{(2)} \ \dots \ \delta H_j^{(m)} \delta H_k^{(m)}) && W \\ \langle \delta H_j \delta H_k \delta H_l \rangle &= (\delta H_j^{(1)} \delta H_k^{(1)} \delta H_l^{(1)} \ \delta H_j^{(2)} \delta H_k^{(2)} \delta H_l^{(2)} \ \dots) && W \\ \langle \delta H_j \delta H_k \delta H_l \delta H_p \rangle &= (\delta H_j^{(1)} \delta H_k^{(1)} \delta H_l^{(1)} \delta H_p^{(1)} \ \delta H_j^{(2)} \delta H_k^{(2)} \delta H_l^{(2)} \delta H_p^{(2)} \ \dots) && W \\ \vdots &= \vdots && W \\ \cdot & && (22) \end{aligned}$$

På matrisform kan detta skrivas,

$$M = R(H(\Sigma))W. \quad (23)$$

Med linjär regression enligt avsnitt Linjär regression erhålls direkt lösningen,

$$W = (R^T R)^{-1} R^T M. \quad (24)$$

Den ursprungliga iden var att eftersom viktningen bevaras från parameterensemble till modellensemble vid enkel osäkerhetskvantifiering, så överförs viktningen på samma sätt vid kalibrering, fast i omvänd riktning. En vikt för modellensembeln överförs till samma vikt för parameterensemblen. Det möjliggör en indirekt bestämning av parametrarnas statistik, varifrån en kalibrerad parameterensemble med lika vikter sedan kan bestämmas med annealing, precis som vid enkel osäkerhetskvantifiering.

Komplikationer

I praktiken är det emellertid inte riktigt så enkelt. Frihetsgraderna för ensemblen $H(\Sigma)$ är många fler än för vikterna W . Det medför att omviktning (stratifiering) inte kan ersätta en godtycklig förflyttning av $H(\Sigma)$ (annealing) utan en omfattande utökning av ovanstående ekvationssystem. Utan denna generalisering kommer inte den optimala parameterensemblen att kunna bestämmas via anpassad viktning / stratifiering.

Vikterna för ensemblen bestämmer både centrering ($H(\Sigma)$) och spridning ($\delta H(\Sigma)$) av hela ensemblen. Det ena kan alltså generellt inte justeras utan att samtidigt ändra det

andra. Det blir därför rimligt att använda *olika* vikter för centrerings och normalisering (första ekvationen), samt spridning. Det löser det kanske mest uppenbara problemet med stratifieringen.

Eftersom $\sum_{u=1}^m \delta H_k^{[u]} = 0$ så är variationerna linjärt beroende och en kolumn kan uttryckas som en linjärkombination av de övriga. Därför saknas en frihetsgrad för variationernas $\delta H_k^{[u]}$ vikter W_u .

Även med dubbla uppsättningar vikter, för centrerings/normalisering samt spridning, så representerar ensemblen ändå mer information än dess sampelvikter. I projektet har två principiellt olika lösningar studerats för att expandera antalet vikter så att de kan representera lika mycket relevant information som ensemblen:

1. Utöka kombinationerna av ensemblevariationer från att vara lika, till att innehålla alla möjliga kombinationer, t.ex. $W_u \delta H_j^{[u]} \delta H_k^{[u]} \rightarrow W_{uv} \delta H_j^{[u]} \delta H_k^{[v]}$. Vikterna blir tensorer av samma ordning som aktuellt moment.
2. Tillräcklig översampling av modellen, utgående från parameterensemblen Σ .

Alternativ 1 delar upp vikterna ytterligare i grupper, för de olika momenten. Varje bestämd ordning av moment får en egen deluppsättning vikter. Överföringen av vikter för modellensemble till parameterensemble blir därför svår, eftersom modellens olinjäriteter blandar, eller *sprider* statistiska moment vid propagering av osäkerhet [2]. Endast affina modeller bevarar vikterna: Ett visst moment av modellensemblen beror då endast på moment av parameterensemblen av samma ordning, beskrivna av *samma* vikter. Med olinjäritet i modellen blandas moment av olika ordning, dvs olika uppsättningar vikter, på ett komplicerat svåranalyserat sätt. Därför ger första alternativet en praktisk begränsning till affina modeller, motsvarande metod 1.

För den icke-diagonala viktningen i alternativ 1 ger kravet 'positiv semidefinit kovarians' ett svårt bivillkor i den linjära regressionen. Ignoreras det kan lösningen lätt bli statistiskt otillåten. Det kan i och för sig hanteras i den efterföljande annealingen. En sådan procedur ger dock inte med säkerhet optimal anpassning. Bivillkoret är inte ekvivalent med att trunkera ett resultat som erhållits utan bivillkor! Alternativ 1 kan därför *inte* förväntas ge bäst möjliga representation.

Översampling av modellen i alternativ 2 är lättare att hantera, men även enklare att förstå. Framförallt kan lösningen garanteras ge positivt semidefinit lösning för kovariansmatrisen: Med bibehållen diagonal viktning överförs kravet på semipositiva egenvärden för viktmatrisen till semipositiva diagonala element, dvs vikterna själva. Det gör att regressionsproblemet kan lösas med s.k. *kvadratisk programmering* [20] som garanterar sådana lösningar. Då kan inte en ogiltig stratifierad modell erhållas [som i alternativ 1].

För varje sampelvärde $(\Sigma_1^{[m+1]} \quad \Sigma_2^{[m+1]} \quad \dots \quad \Sigma_n^{[m+1]})^T$ som läggs till *kan som bäst* ytterligare en frihetsgrad erhållas. Det finns dock ingen garanti för det eftersom tillägget kan var linjärt beroende med föregående elements bidrag till regressionsmatrisen rank m i vilket fall representationsförmågan inte ökar alls. För att undvika otillräcklig frihet i stratifieringen så kan det vara bättre att vara pragmatisk än smart. En optimal översamplingen är i regel inte enkel att finna. Däremot går det alltid att slumpmässigt generera ytterligare sampelvärden för att utöka ensemblen. För att med säkerhet ge tillräcklig frihet kan det vara lämpligt att den utökade ensemblen är något större än den måste vara, för att vikterna ska vara tillräckligt många och tillräckligt bra konditionering av normalmatrisen $R^T R$.

Slumpmässig översampling av Σ i stratifieringssteget är *inte* ekvivalent med slumpmässig statistisk sampling. Stratifieringen avser enbart att *indirekt* överföra statistisk information från kalibreringsdata, till information om osäkra parametrar vid modellkalibrering.

Den avgörande fördelen med översampling (alt. 2) är att vikterna universellt gäller för *samtliga* moment. Vikterna relaterar *inte* som i alt. 1 till alla *kombinationer* av sampelvärden, vars antal beror på momentets ordning. Därmed blir vikterna bevarade vid propagering av osäkerhet, trots att olinjära modeller blandar moment av olika ordning. Överföringen av vikter från modellensemble till parameterensemble avgörs därför i detta fall inte på modellens linjäritet, som i alt. 1. Därför valdes slumpmässig översampling av ensemblen (alt. 2), för att ge vikterna tillräckligt stor frihet att representera kalibreringsdata på bästa möjliga sätt. Därmed propageras osäkerheten ‘olinjärt’ i varje kalibreringssteg.

För att förenkla GEMIX-kalibreringen beräknades dock modellensemblen för översamplingen med en linjär surrogatmodell. Att utföra fulla modellberäkningar ansågs i i detta sammanhang onödigt, med tanke på övriga begränsningar för den aktuella CFD-modellen. I praktiken bör dock fulla modellen användas även för översamplingen.

Ensemblekollaps

Annealing flyttar hela ensemblen enligt de vikter som bestämts vid stratifiering. Dessa viktors avvikelser från likformighet är det som avgör hur mycket ensemblens olika medlemmar ska flyttas. Förskjutningen är individuell och resulterande kovarians måste vara positivt semidefinit, vilket är ett annat sätt att uttrycka att variansen alltid måste vara positiv.

Regressionsmatrisen R i ekv. 24 kan i princip diagonaliseras och delas upp i fundamentala bidrag med en egenvärdesuppdelning. I så fall måste alla bidrag ha positiva egenvärden för att kovariansen ska bli positivt semidefinit. Det är fullt möjligt att ett visst bidrag inte förbättrar representationen av osäkerhet. Då utesluts bidraget helt, vilket gör att rangen reduceras. Det betyder att ensemblen helt ‘plattas till’ i de riktningar som ges av egenvektorer för de bidrag som ignoreras. Ensemble spannar sedan upp färre

dimensioner. Förlusten i dimensionalitet kan inte återskapas i följande iteration på annat sätt än att ensemblen påtvingas en expansion i den riktning som *i denna iteration* varit ogynnsam. Denna tillplattning kan ses som en oåterkallelig ensemble-kollaps. Konsekvensen kan bli att den slutgiltigt kalibrerade ensemblen inte får en optimal representation, dock blir den så bra som möjligt m.a.p. kvarvarande frihetsgrader.

Risken för ensemblekollaps under fixpunktsiteration kan i sämsta fall leda till en mycket dålig representation för den slutgiltiga parameterensemblen. Därför kan det vara viktigt att justera parameterensemblen i steg under fixpunktsiterationen. Annars kan den kollapsa av fel orsaker, långt från sin slutgiltiga optimala position. Endast centrerings (väntevärdet) för ensemblen bör justeras i de första iterationerna. Eventuellt tillåts rotation och skalning av ensemblen endast i de sista stegen. Konvergensen för centrerings är inte annorlunda eller mer garanterad än Newton-Raphsons metod [21]. En fullt utvecklad stabil fixpunkt kan ibland vara svår att finna.

Ensemblen har alltid en viss kvalitet på sin representation, oavsett vilka förändringar ytterligare iterationer ger. Det räcker med att finna en tillräckligt bra justering. Iterationen är en systematisk men begränsad metodik att finna en möjlig förbättring av aktuell representation av den osäkra modellen.

Residual

För traditionell skattningsmetodik krävs att residualen är försumbar i relation till osäkerheten för kalibreringsdata, för att kalibrerade modellen ska bli godkänd vid validering. Det finns alltså inget sätt att inkludera eller acceptera ett betydande systematisk modellfel. För komplexa beräkningar kan det vara en alldeles för hög ambition att kräva en försumbar residual, jämfört med osäkerheten för kalibreringsdata. Med metod 2 ges en möjlighet att inkludera residualen i modellosäkerhet.

Residualen gör att referenserna $\{H_C\}, \langle H \rangle$ för varianserna $\text{var}(H_C) = \langle \delta^2 H_C \rangle$ och $\text{var}(H) = \langle \delta^2 H \rangle$, för kalibreringsdata respektive modellprediktion kan bli *väsentligt* olika. Normalt fås endast ett kalibreringsresultat i varje punkt, eller ett sampelvärde $\{H_C\}$ av *statistiska variabeln* H_C beskrivande kalibreringsdata. Variationen δH_C av experimentella kalibreringsdata bör därför istället relateras till det förväntade *modell*-resultatet $\langle H \rangle$ för att bli jämförbart,

$$\delta_M H_C \equiv H_C - \langle H \rangle = H_C - \{H_C\} + \{H_C\} - \langle H \rangle = \delta_C H_C + \underbrace{\{H_C\} - \langle H \rangle}_{\{\rho\}}. \quad (25)$$

Residualen ρ är precis som H_C en slumpvariabel med bl.a. medelvärde och varians, här angiven med ett specifikt utfall $\{\}$. Variansen för kalibreringsdata justeras till andra moment kring $\langle H \rangle$, vilket explicit inkluderar variansen för residualen enligt ekv. 25 ovan,

$$\langle \delta_c^2 H_c \rangle \rightarrow \langle \delta_M^2 H_c \rangle = \langle \delta_c^2 H_c \rangle + \langle \rho^2 \rangle. \quad (26)$$

Här antas kalibreringsdata $\delta_c H_c$ vara statistiskt oberoende av residualen ρ , $\langle \delta_c H_c \rho \rangle \approx 0$ eftersom det ingår som ett krav på identifierad modell [7]. Modellen får inte beskriva ett specifikt slumpmässigt mätfel för att kunna anses vara allmängiltig.

Nu återstår det praktiska problem att räkna ut väntevärdet $\langle \rho^2 \rangle$ av den kvadrerade residualen från endast ett(!) sampelvärde $\{\rho\}^2$. Det motsvarar att bestämma medelvärdet av endast ett tärningskast, vilket verkar omöjligt. Matematiskt blir variansen av en sådan skattning 'oändlig', vilket uttrycker denna omöjlighet. Den avgörande skillnaden som ändå gör det möjligt att ge en bra skattning av $\langle \rho^2 \rangle$ är att $\rho(r)$ är en *vektor* och inte en skalär (som tärningens utfall). Förutsatt att $\rho(r)$ är någorlunda *ergodisk* [22], är det förväntade värdet $\langle \rho^2(r) \rangle$ över *upprepade* mätningar ungefär detsamma som medelvärdet över *närliggande* (ε) punkter, $\tilde{r} : |\tilde{r} - r| \leq \varepsilon$. I så fall kan digital lågpas (LP) -filtrering över r ersätta ensemblemedelvärdet,

$$\langle \rho^2 \rangle(r_k) \approx LP(\{\rho\}^2) = \sum_l h_l \frac{\{\rho\}^2(r_{k-l}) + \{\rho\}^2(r_{k+l})}{2}. \quad (27)$$

Här är h_l filtrets impulsrespons. Med explicit lika filtrering framåt och bakåt ovan har filtrets tidsfördröjning, eller färfel helt eliminerats. Ett lågpas-filter utan rippel som inkluderar runt 20 sampelvärden, dvs $\varepsilon \approx 20 \cdot |r_{k+1} - r_k|$ kan vara lämpligt. I projekt har ett andra ordningens Butterworth-filter med normerad gränshäns $\omega_N \approx 0.05$ använts med gott resultat. Risken med att filtrera för hårt är förstås att $\langle \rho^2 \rangle$ blir alltför utsmetad för att kunna upplösa detaljer korrekt där modellen lokalt har stora fel, typiskt vid hörn, kanter och andra diskontinuiteter. Där kan inte heller $\rho(r)$ vara ergodisk, eftersom det förutsätter stationaritet, eller likhet över r .

Residualen inkluderas således som en expanderad mätsäkerhet för *kalibreringsdata*, för att sedan representeras på bästa sätt med den kalibrerade osäkra modellen. För det ska vara möjligt behövs en hypotes om hur modellen bör utökas, eller generaliseras till att kunna inkludera modellfel.

Mätfel

En förutsättning för att erhålla kalibreringsdata $\{H_c\}$ är att vi utför en faktisk mätning. Den kan aldrig bli helt perfekt utan måste innehålla mätfel. En konsekvent kalibrerad modell bör följdaktligen även innefatta mätningens imperfektion, precis som residualen beskriver modellens fysikaliska brister. Därför behöver den *fysikaliska modellen* kompletteras med en *mätmodell* som representerar möjliga mätfel.

Inom mätteknik används oftast multiplikativa och additiva statistiska variabler (α, β) för att representera en mätosäkerhet som täcker möjliga mätfel. Under det rimliga antagandet att mätfel och modellfel är okorrelerade, $\langle \delta\alpha(x)\delta H(x, \theta) \rangle = 0$, blir utökningen av den fysiska modellen,

$$\begin{aligned} H(x, \theta) &\rightarrow H(x, \theta) + \alpha(x)H(x, \theta) + \beta(x), & \alpha &\sim F_\alpha(0, \sigma_\alpha^2), & \beta &\sim F_\beta(0, \sigma_\beta^2) \\ \langle \delta^2 H \rangle &\rightarrow \langle \delta^2 H \rangle + \sigma_\alpha^2 \langle H \rangle^2 + \sigma_\beta^2, \end{aligned} \quad (28)$$

där fördelningarna $F_\alpha(0, \sigma_\alpha^2), F_\beta(0, \sigma_\beta^2)$ har medelvärden noll under antagande att systematiska mätfel är försumbara. Varianserna $\sigma_\alpha^2, \sigma_\beta^2$ beskriver slumpmässiga mätfel.

Osäker modell

Normalt utgår beräkningsmodellen från en fysikalisk modell som inte är perfekt. För konsistent kalibrering mot kalibreringsdata så bör hypotesen för hela beräkningsmodellen egentligen innefatta ett flertal komponenter:

- Fysikalisk grundmodell(-er), t.ex. CFD-beräkningsprogram
- Mätfelsmodell
- Residualmodell

Sammantaget motsvarar det en *osäker modell* — ett nytt huvudkoncept som introducerats i detta projekt. Den fysikaliska modellen får anses given. Finns det flera alternativa modeller kan de kombineras i en multi-modell kalibrering, där viktningen av de olika modeller är fria osäkra parametrar, vilka kalibreras på motsvarande sätt som övriga parametrar för var och en av modellerna.

En etablerad allmän mätfelsmodellen visas i ekv. 28 ovan. Residualmodellen är svårast att ansätta. Den måste vara lika generell som den fysikaliska modellen för att inte inskränka tillämpbarheten, dvs vara oberoende av aktuellt kalibreringsfall. Den fysikaliska modellen kan inte användas eftersom residualen beskriver precis det som modellen inte kan uttrycka. Någon bättre beskrivning av det observerade finns inte heller, eftersom den i så fall redan skulle inkluderas i den fysikaliska grundmodellen. Sammantaget liknar bestämningen av residualmodell därför det generella problemet med att bestämma en mätosäkerhet. I frånvaro av specifik information om residualen kan därför residualmodellen ansättas på samma *form* som mätmodellen. Det är *inte* samma sak som att de är lika. Med samma form kan de däremot kombineras.

Om de generella mätfels- och residualmodellerna har precis samma form, så blir de inte heller matematiskt särskiljbara, eller *identifierbara* enligt avsnitt Identifierbarhet. Det är

ett mindre problem eftersom det inte finns någon anledning att särskilja dem i den kalibrerade beräkningsmodellen. Med samma form kan de kombineras, vilket gör att (α, β) skalas om, eller *renormaliseras*, till $(\tilde{\alpha}, \tilde{\beta})$.

Stratifiering

Stratifieringen i metod 2 blir därmed sekventiell:

1. Först sök bästa representation av medelvärde, motsvarande bästa skattning för traditionell metodik. Dvs, minimera den samplade residualen $\{\rho\} = \{H_C\} - \langle H \rangle = \{H_C\} - H(\Sigma)W$ genom att variera vikter W för Σ .
2. Lågpass-filtrera (*LP*) den samplade residualen för att skatta dess förväntade värde, $\{\rho\}^2 \rightarrow \langle \rho^2 \rangle (r_k) \approx LP(\{\rho\}^2)$.
3. Matcha högre statistiska moment för modellensembelns spridning $\delta H(\Sigma)$ med vikter \tilde{W} . För andra moment med generella delmodeller enligt avsnitt Mätfel och Osäker modell erhålls, $\langle \delta^2 H \rangle + \sigma_{\tilde{\alpha}}^2 \langle H \rangle^2 + \sigma_{\tilde{\beta}}^2 = \langle \delta^2 H_C \rangle + LP(\{\rho^2\})$, vilket ger,

$$\begin{pmatrix} \sum_{1 \leq u \leq m-1} \delta^2 H & H(\Sigma)W & 1 \end{pmatrix} \begin{pmatrix} \tilde{W} \\ \sigma_{\tilde{\alpha}}^2 \\ \sigma_{\tilde{\beta}}^2 \end{pmatrix} = \langle \delta^2 H_C \rangle + LP(\{\rho^2\}). \quad (29)$$

Modelleras $\sigma_{\tilde{\alpha}}^2, \sigma_{\tilde{\beta}}^2$ som säkra hyperparametrar [17], så kan hela ekvationssystemet i ekv. 29 lösas med linjär regression. [Här har de redundanta variationerna $\delta H^{(m)}$ uteslutits.]

Annealing

I projektet begränsades anpassningen av beräkningsmodellen endast till variansen för kalibreringsdata, eftersom [som vanligt] ingen information om varken kovarianser eller högre moment fanns tillgängliga [för GEMIX]. Stratifieringen justerar sedan vikterna så att modellresultatets medelvärde och varians överensstämmer så bra det är möjligt med kalibreringsdata och given okalibrerad parameterensemble. Därefter antas vikterna för modellensembeln även gälla för parameterensemblen. Bestäms sedan parametrarnas statistik från $W, \Sigma, \tilde{W}, \delta \Sigma$, erhålls en full kovariansmatris.

Annealing, dvs bestämning av parameterensembler utifrån parametrars statistik, är identisk med den typ av annealing som används vid enkel osäkerhetskvantifiering [23]. Begränsat till första och andra statistiska moment kan parameterensemblen direkt beräknas utifrån vald excitationsmatris V [4], se avsnitt Metod 1: Maximum likelihood skattning, punkt 6.

Det innebär att om t.ex. geometrin ändras blir justeringen helt(!) okänd. Den enda sättet att erhålla en allmängiltig justering är att kalibrera en residualmodell med enbart generella konstanter. Exempelvis en residualmodell enligt ekv. 28, fast med *säkra* parametrar α, β .

Utvärdering

Föreslagna metoder har dels utvärderats med enkla testmodeller, dels med kalibrering av CFD-modell mot GEMIX kalibreringsdata. Testmodellerna ger den stora fördelen att det perfekta resultatet av kalibrering kan bestämmas. Det ger den absoluta referens som krävs för att utvärdera kalibreringsmetoderna. Kalibreringen av CFD-modellen är däremot avsedd att vara realistisk och så typisk som möjligt, för att kunna påvisa effekten av kalibrering i ett verkligt fall.

Konsistens

Begreppet *konsistens* beskriver hur stor andel av alla observationspunkter som det beräknade osäkerhetsintervallet innesluter kalibreringsdata (utan osäkerhet). Det motsvarar i viss utsträckning konsistensbegreppet inom identifiering. För en s.k. konsistent modell så är motsvarande konsistens nära 100%. Skillnaden här är att konsistens blir ett kvantitativt mått. Vid validering mot ett resultat av kalibreringsdata ska konsistensen motsvara täckningsgraden för beräknat resultat, för tillräckligt många mätpunkter och förutsatt att mätfel i kalibreringsdata inte har alltför hög korrelation.

Osäkerhetsbidrag

Osäkerhetsbidragen tillhör två kategorier:

- *Epistemisk osäkerhet* beskriver kunskapsbrist om underliggande samband. Det ger ett okänt specifikt modellfel som antas inneslutas inom en generell osäkerhet. Osäkerheten blir därför avsevärt större än modellfelet i nästan alla mätpunkter — det är priset för att inte känna till det specifika felet och samtidigt kräva att nästan alltid ha rätt, dvs hög konsistens.
- *Aleatorisk osäkerhet* beskriver slumpmässiga variationer för sampel av kalibreringsdata. Det motsvarar den normala föreställningen om mätfel, eller mätbrus. Precis som för den epistemiska osäkerheten så antas den aleatoriska osäkerheten innesluta möjliga mätfel i nästan alla punkter, typiskt med 95% konfidens.

För att kunna associera ett perfekt kalibreringsresultat för kalibrering av testmodellerna inkluderas inte specifika epistemiska modellfel. Istället beskrivs de av den typ av osäkerhet som används vid kalibreringen för att representera just dessa modellfel. Epistemisk osäkerhet representeras i kalibreringsdata därför med spridning av en *referensensemble*, för samma modellstruktur som senare används vid identifiering. Ett perfekt kalibreringsresultat motsvarar därmed 100% representation av denna osäkerhet i alla mätpunkter, dvs identifierad osäkerhet sammanfaller med osäkerheten för kalibreringsdata.

Kalibreringsdata inkluderar normalt endast ett sampelvärde, dvs ett uppmätt resultat i varje observationspunkt. Till varje sådant värde associerar vi en mätosäkerhet, vilken

omsluter de flesta tänkbara mätfel. Även om det innebär miljoner värden, så motsvarar kalibreringsdata normalt endast ett 'sampelvärde', så som det definieras inom matematisk statistik. Samplingsvarians ger därför ett slumpmässigt fel i varje mätpunkt. Kalibreringsdata är alltså i regel en slumpvektor/-tensor. Slumpmässiga fel undviks i jämförelser här genom att låta kalibreringsdata innehålla ett stort sampel.

Täckningsintervall

Innebörden av *konfidensintervall* och av *täckningsintervall* är central, i synnerhet vid kalibrering av beräkningsmodeller. Täckningsintervall är en generalisering av konfidensintervall för observationer, till statistiska modeller och kunskap. Beräkningen ger ett täckningsintervall, snarare än ett konfidensintervall. Konfidensintervallet för kalibreringsdata är inget uttalande om ett enskilda sampelvärde. Det uttrycker endast det förväntade mätresultatets variation över alla möjliga (oändligt antal) repetitioner av kalibreringsexperimentet. Det är inte alls ekvivalent med att beräknade och uppmätta resultat är konsistenta, motsvarande konfidensnivån, i alla observationspunkter. Om t.ex. felen för kalibreringsdata är fullständigt korrelerade, så får kalibreringsdata endast en(!) frihetsgrad. Är däremot kalibreringsdata helt okorrelerade, eller oberoende så har kalibreringsdata lika många frihetsgrader som antal observationspunkter. När korrelationen för kalibreringsdata går ner, ökar alltså informationsmängden. Det är det som gör att traditionell identifiering (metod 1) ger säkrare skattningar när kovariansen minskar. Beskrivs okänd kovarians som ingen kovarians, enligt normal praxis, fyller det således på med mycket(!) information som inte existerar.

Sampelstorlek kalibreringsdata

För att kunna bestämma kalibreringsmetodernas prestanda men även vara typisk för kalibrering valdes en mix av sampelstorlekar. Den epistemiska osäkerheten för kalibreringsdata representerar hela dess population [med deterministisk sampling], medan endast ett sampelvärde tas av den aleatoriska. Det gör att 'medelvärdet'¹ för kalibreringsdata får ett mätbrus, liksom att residualen är en slumpad vektor, som normalt vid modellkalibrering.

Testmodeller

Kalibrering av testmodellerna har inte enbart använts för utvärdering, utan har även varit ovärderlig för att verifiera och utveckla metod 2 i allmänhet. Modellerna har följdaktligen främst valts för att kunna illustrera relevanta egenskaper.

Affin modell

Modeller som är linjära-i-parametrar (affina) är ofta en naturlig utgångspunkt för osäkerhetskvantifieringsmetoder. Både studerade metoder 1 och 2 blir för affina modeller helt felfria. Iterationerna *ska* därför konvergera i endast ett steg. Kalibrering av affina modeller ger därför en mycket bra kontroll men påvisar även skillnader mycket tydligt.

¹ Medelvärdet för den epistemiska osäkerheten, men ett sampelvärde för den aleatoriska osäkerheten.

Konvergens under iteration måste dock studeras med icke affina modeller, se nedan. I detta fall valdes den affina modellen,

$$H(x, \theta) = \theta_1 + \theta_2 \sin(x) + \theta_3 x^3 + \beta. \quad (30)$$

Den epistemiska osäkerheten för kalibreringsdata representeras av referensensemblen,

$$\Sigma_{REF} = \begin{pmatrix} 1.1 & 0.5 & 1.1 & 0.5 \\ 0.7 & 0.7 & 0.3 & 0.3 \\ 1.1 & 0.7 & 0.7 & 1.1 \end{pmatrix}, \Rightarrow \langle \theta \rangle = \begin{pmatrix} 0.8 \\ 0.5 \\ 0.9 \end{pmatrix}, \text{cov}(\theta) = \begin{pmatrix} 0.3 & 0 & 0 \\ 0 & 0.2 & 0 \\ 0 & 0 & 0.2 \end{pmatrix}. \quad (31)$$

Båda metoder 1 och 2 klarade det grundläggande kravet på konvergens i ett steg för kalibreringsdata utan mätbrus, $\alpha = \beta = 0$, men med full representation av modellosäkerhet. Konsistensen var alltså 100%, dvs beräknade resultat och kalibreringsdata var helt identiska.

Med relativt starkt additivt mätbrus $\text{var}(\beta) = (0.2)^2$ blev det dock stor skillnad på osäkerheterna, medan medelvärdena alltså var oskiljaktiga, se Fig. 5. Notera att för metod 1 *minskar* osäkerheten för bästa skattning av modellen när mätbrus adderas!

Mer störningar i kalibreringsdata ger säkrare(!) skattning med traditionell identifiering.

För metod 2 är det precis tvärtom. Utöver modellosäkerhet så representeras även mätbrus. Modellosäkerheten blir då högre med mer störningar i kalibreringsdata. Anledningen till att inte allt mätbrus representeras fullt ut, är att residualfiltreringen jämnar ut och minskar de extrema felen, vilket underskattar motsvarande osäkerhet. Anpassningen av den kombinerade mätfels- och residualmodellen till den filtrerade residualen visas i Fig. 6.

Det kan verka orimligt att bli mer säker på sin slutsats (modell) när informationen den grundar sig på blir mer osäker. Det är ändå begripligt utifrån vad som menas. Att bli mer säker på skattningen av modellen med metod 1 betyder inte att den är bättre, inte heller att denna osäkerhet avspeglar osäkerheten för kalibreringsdata. Osäkerheten refererar inte till något annat än just bestämningen av *skattningen*. Konkret så ökar förstås variansen för kalibreringsdata med tillägg av mätbrus. Det medför att osäkerheten för modellen ökar med metod 2, vilken endast tar variansen i beaktande. Det mest väsentliga bidraget för metod 1 är dock att korrelationerna för kalibreringsdata minskar. Modellen får det svårare att reproducera möjliga variationer (osäkerhet) i kalibreringsdata. Det är just denna 'styvhet' som ger lägre osäkerhet. De ökade kraven på modellens anpassning är en konsekvens av att informationsmängden ökar när kovariansen minskar p.g.a. mätbrus.

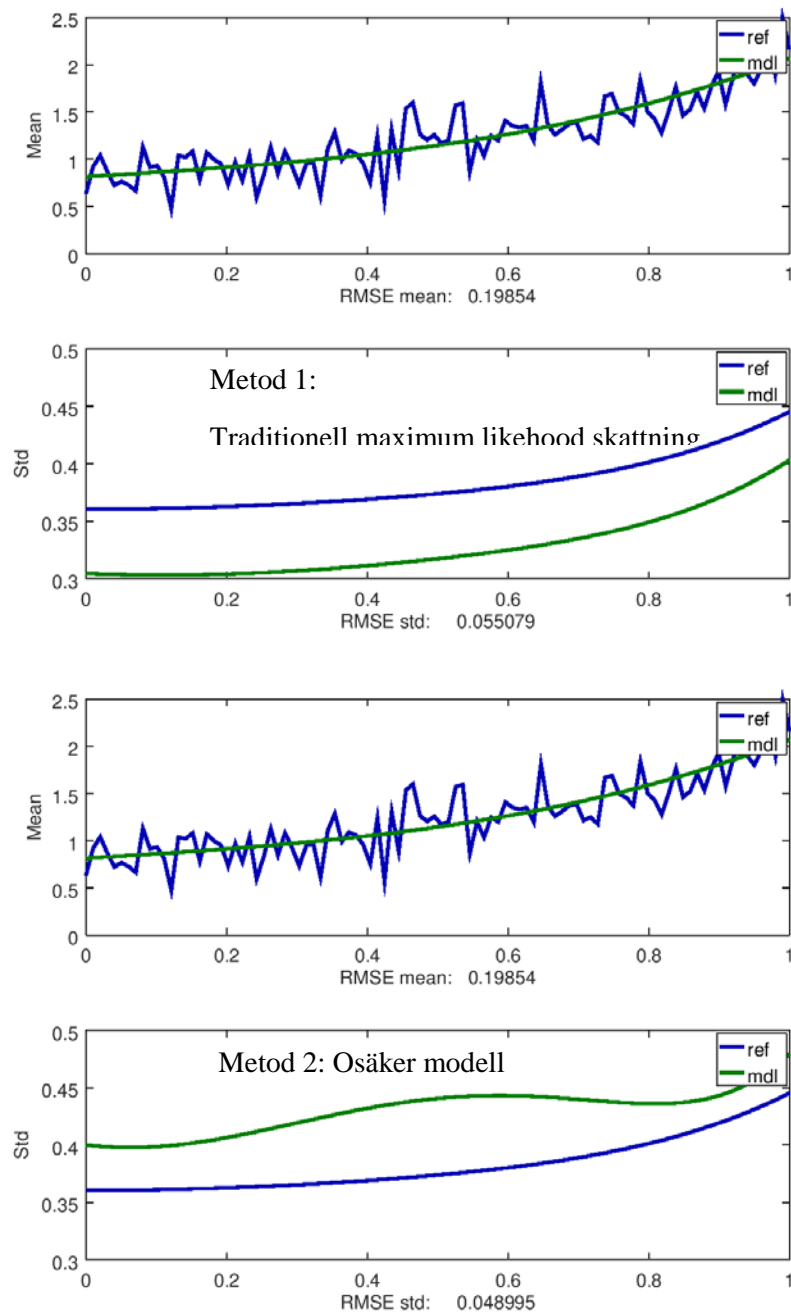


Fig. 5. Identifiering av affin modell med metod 1, maximum likelihood (överst), samt metod 2, stratifiering och annealing (nederst).

Anledningen till att metod 2 ger en *högre* osäkerhet för kalibrerad beräkningsmodell, än för kalibreringsdata är dess samplingsvarians. Det finns ju endast *en* realisering av kalibreringsdata. En hel uppsättning kalibreringsdata skulle inneburit mer information och bättre bestämning, dvs lägre osäkerhet för modellen efter kalibrering. ‘Kostnaden’ för att inte ha en stor uppsättning kalibreringsmätningar är således skillnaden mellan ‘ref’ och ‘mdl’ i Fig. 5 (nederst). Omvägen till ett resultat via modellen bör ju öka osäkerheten, ifall information om populationen av kalibreringsdata går förlorad, som i detta fall.

Residualfiltreringen visas i Fig. 6 nedan. Eftersom modellen i ekv. 30 redan har en additiv parameter blir brusmodellen ej identifierbar: Mätbruset $\text{var}(\alpha)=0$, $\text{var}(\beta)$ som här endast är additivt kommer därför att helt kunna representeras av $\text{var}(\theta_1)$.

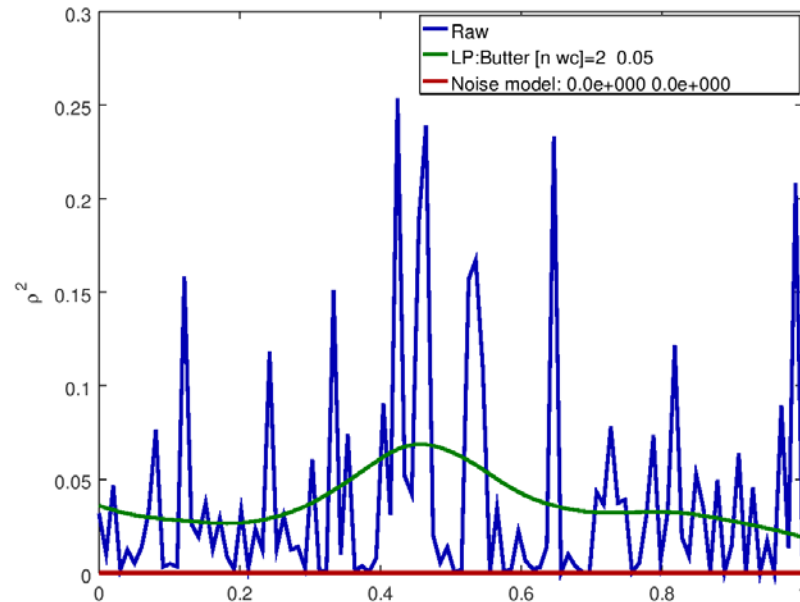


Fig. 6. Residualfiltrering för metod 2, stratifiering och annealing.

En intressant aspekt är hur det totala antalet observationspunkter påverkar kalibreringsresultatet för metod 1, se Fig. 7 där $m = 10, 100, 1000$. För metod 2 bestäms inte modellens förmåga att representera kalibreringsdata av antalet jämförelsepunkter, förutsatt (som här) att korrelationer mellan fel i olika punkter inte beaktas. För metod 1 minskar dock osäkerheten, dvs representationsförmågan med antalet punkter:

Ju mer kalibreringsdata av samma kvalitet, desto säkrare(!) skattning med traditionell identifiering. Genom att endast lägga till mätpunkter kan skattningen för metod 1 således bli obegränsat(!) säker.

Det kan verka logiskt ur ett matematiskt statistiskt perspektiv på slumpmässig sampling att mängden information ökar med antalet observationer [8]. Kvaliteten för kalibreringsdata blir emellertid inte högre *lokalt* i varje enskild punkt, bara för att flera punkter undersöks. Eftersom vår referens(kalibreringsdata) egentligen inte blir bättre, tillkommer inte någon väsentlig information.

Att modellens osäkerhet minskar med antalet punkter för kalibreringsmetod 1 beror på kravet att modellera korrelationer för kalibreringsdata. När antalet punkter ökar, så ökar antalet kovariansbidrag mycket snabbare än antalet variansbidrag. Modellen får därför allt svårare att anpassa sig efter alla krav. Till slut 'fryser' skattningen av modellen helt fast och blir helt säker — oberoende av hur *bra* den faktiskt är. Att vara säker är uppenbarligen inte alls samma sak som att ha rätt.

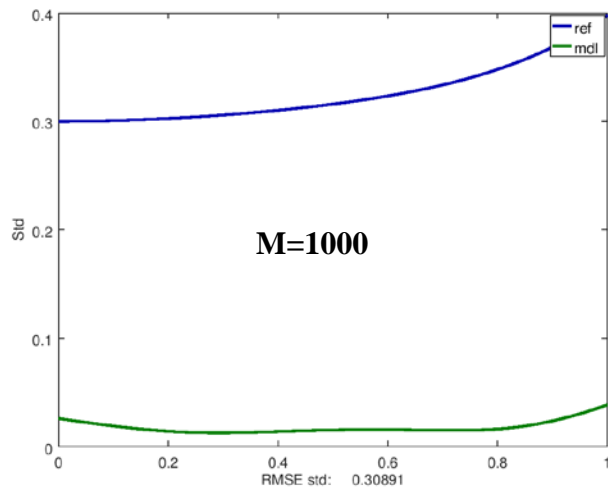
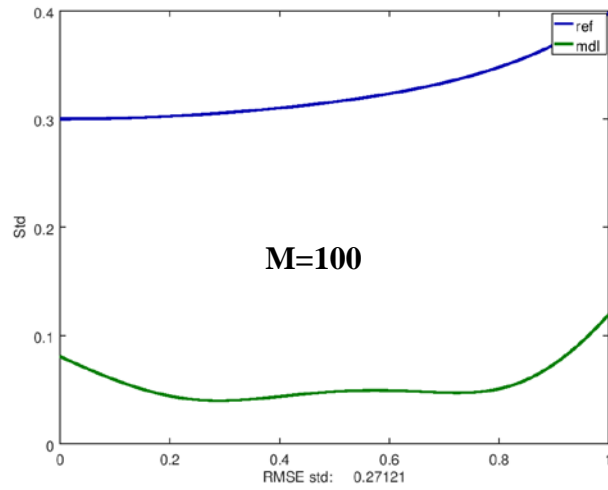
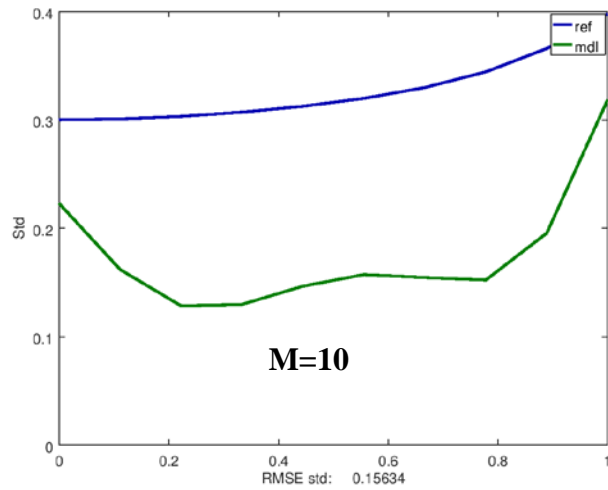


Fig. 7. Identifiering av affin modell med metod 1, för olika antal ($M = 10, 100, 1000$) kalibreringspunkter och $\text{var}(H_C) \neq 0$, vid frånvaro av mätbrus ($\alpha = \beta = 0$) och kovarians $\text{cov}(H_C)_{i \neq j} \rightarrow 0$ för kalibreringsdata. Det senare är ett normalt antagande vid kalibrering.

Chirp modell

Ordet ‘chirp’ syftar på en harmonisk signal med ökande eller minskande frekvens. I detta fall ökas frekvensen med ett kubiskt beroende,

$$H(x, \theta) = \sin(\theta_1 + \theta_2 x^3) + \beta. \quad (32)$$

Kalibreringsdata genererades med en parameterensemble av typ Hadamard [23],

$$\Sigma_{REF} = \begin{pmatrix} 0.43 & -0.17 & 0.43 & -0.17 \\ 1.06 & 1.06 & 0.66 & 0.66 \end{pmatrix}, \Rightarrow \langle \theta \rangle = \begin{pmatrix} 0.13 \\ 0.86 \end{pmatrix}, \quad \text{cov}(\theta) = \begin{pmatrix} 0.3 & 0 \\ 0 & 0.2 \end{pmatrix}. \quad (33)$$

För de studerade kalibreringsmetoderna finns det två mycket väsentliga aspekter med just olinjäritet-i-parametrar:

1. *Konvergens.* Iterationen kan konvergera till en fix ensemble, divergera eller kanske oscillera mellan olika ensembler. Resultatet kan också bli en icke optimal parameterensemble, ifall iterationen fastnar i ett lokalt minimum.
2. *Representation.* Modellens förmåga att beskriva information kan starkt reduceras om den baseras på en approximativ surrogatmodell, t.ex. linjäriserad modell.

Förenklat har vardera studerad metod en av dessa aspekter som sin akilleshäla. Den traditionella metoden 1 använder ofta linjäriserad surrogatmodell. Konvergensen under iteration syftar endast på medelvärdet av parameterensemblen, vilket medför konvergensproblem liknande Newton-Raphsons metod [21]. Frånvaro av konvergens i sig är inte ett problem, däremot kan resultatet bli icke-optimalt. Aspekt nr 2 ovan blir därför det huvudsakliga dilemmat för metod 1.

Metod 2 kan förflytta hela parameterensemblens alla medlemmar individuellt i varje iteration. Justeringar blir därför avsevärt mer avancerade än för metod 1, eftersom det finns många frihetsgrader i ett sådant ‘moln’ av parameteruppsättningar. Skulle molnet av någon anledning bli ‘tillplattat’ genom projicering i ett plan, försvinner frihetsgrader som kanske skulle vara användbara i en senare iteration. Det skulle i så fall försämra den kalibrerade modellens representation. Därför kan det vara väsentligt att inte förändra parameterensemblen för mycket i början, när optimum ligger långt ifrån dess aktuella konfiguration.

För att finna bästa tänkbara och förhoppningsvis globalt optimala ensemblen kan det krävas insikt och tålamod. Skulle den kalibrerade ensemblen inte vara optimal blir representationen bara lite sämre än nödvändigt, men kan likväl vara bättre än före kalibrering. Konvergerar metod 2 så är dess representationsförmåga överlägsen metod 1 eftersom den varken kräver fullständig information eller någon surrogatmodell för att propagera osäkerheten.

Ett mycket enkelt sätt att identifiera olinjäritet är att beräkna modellens s.k. *scent* ζ [4], eller 'lukt' på svenska,

$$\zeta(x) = \langle H(x, \theta) \rangle_{\theta} - H(x, \langle \theta \rangle_{\theta}) = H(x, \Sigma) \cdot W - H(x, \Sigma \cdot W). \quad (34)$$

Här är W viktvektorn som bestämmer det förväntade resultatet. Scent är illustrerad i Fig. 8. Däremot gäller inte det omvända, dvs frånvaro av scent betyder *inte* att modellen är linjär, som $\zeta(1.5) \approx \zeta(1.9) \approx 0$. Skulle $\zeta(x) = 0$ för alla värden på x är det dock mycket osannolikt att modellen ändå skulle vara olinjär-i-parametar θ .

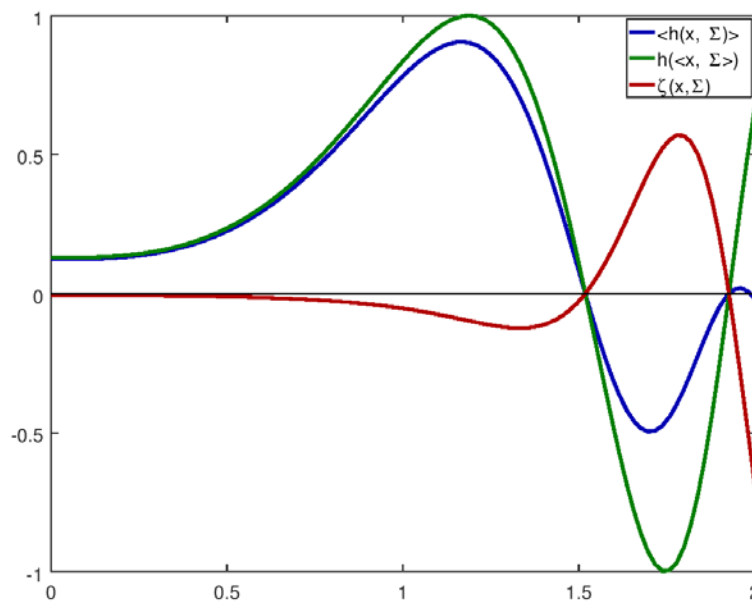


Fig. 8. Scent $\zeta(x)$ för chirp testmodell i ekv. 32.

Identifiering av Chirp-modellen med metod 2 visas i Fig. 9 nedan, från den ursprungliga (PRIOR='Före') till den kalibrerade (POST(-erior)='Efter') beräkningsmodellens resultat. Representationsgraden är som synes hög för både medelvärde (Mean) och standardavvikelse (Std).

Kalibrering med metod 1 gav dock ett annorlunda resultat, se Fig. 10. Även i detta fall minskar osäkerheten för skattningen när kovariansen för kalibreringsdata sätts till noll (nederst). Det bör betonas att dess primära anpassning av medelvärde även kräver en bra fördelning/spridning av ensemblen, vilken återspeglas i visad standardavvikelse. Om den skulle vara dålig så kan inte bästa skattningen bli bra. Förhoppningsvis är spridningen så liten att begränsningen till linjärisering i metod 1 är oväsentlig. Det motsvaras av försumbar scent. Det verkar dock inte vara fallet här eftersom representationen av medelvärdet är tämligen dålig där scenten är hög, jämför Fig. 10 och Fig. 8. Konvergensen var emellertid bättre för metod 1 än metod 2.

Detta illustrerar de förväntade egenskaper som nämnts ovan.

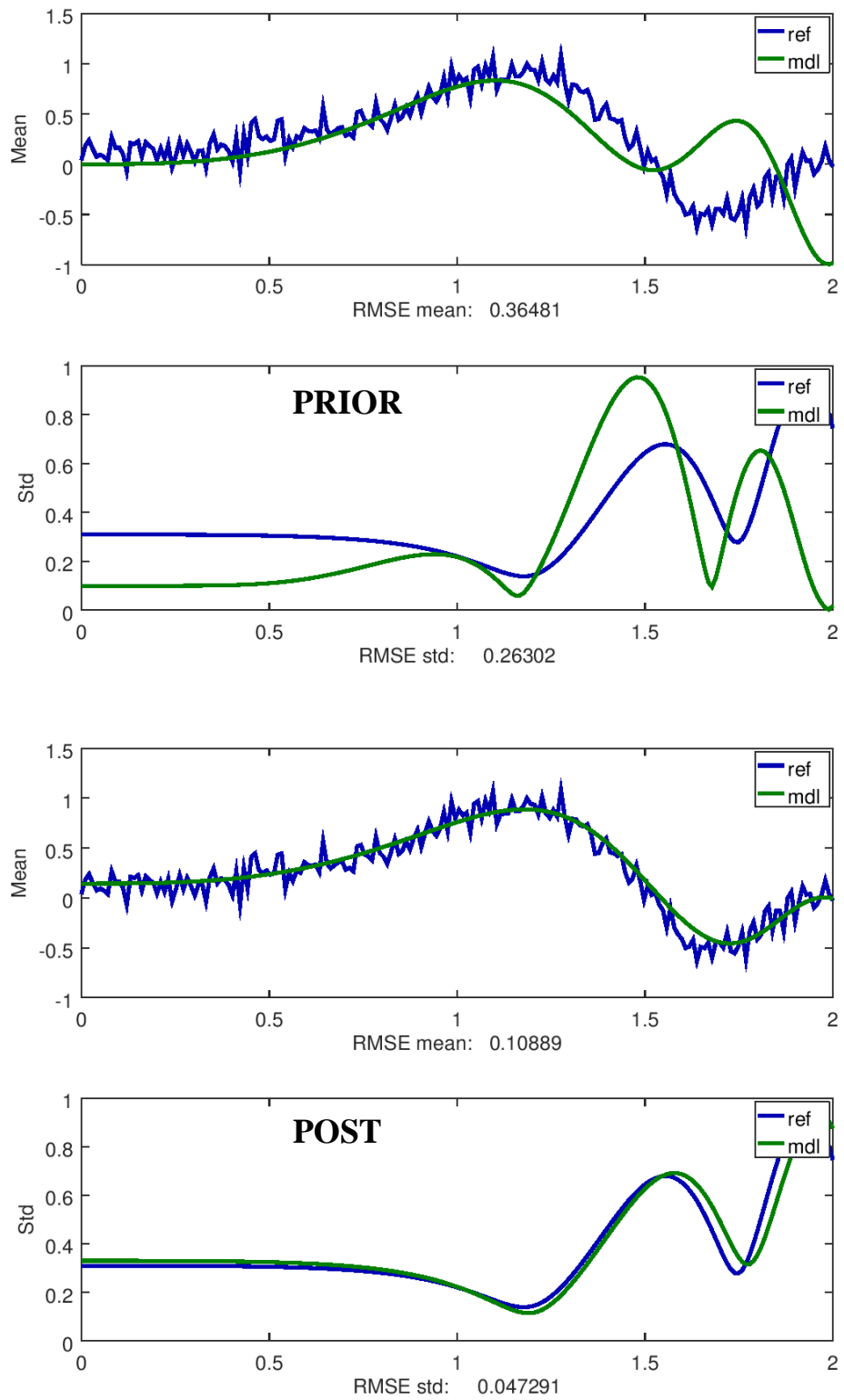


Fig. 9. Identifiering av chirp modell (ekv. 32) med metod 2, start-(överst) och slut-(underst) resultat i 200 mätpunkter.

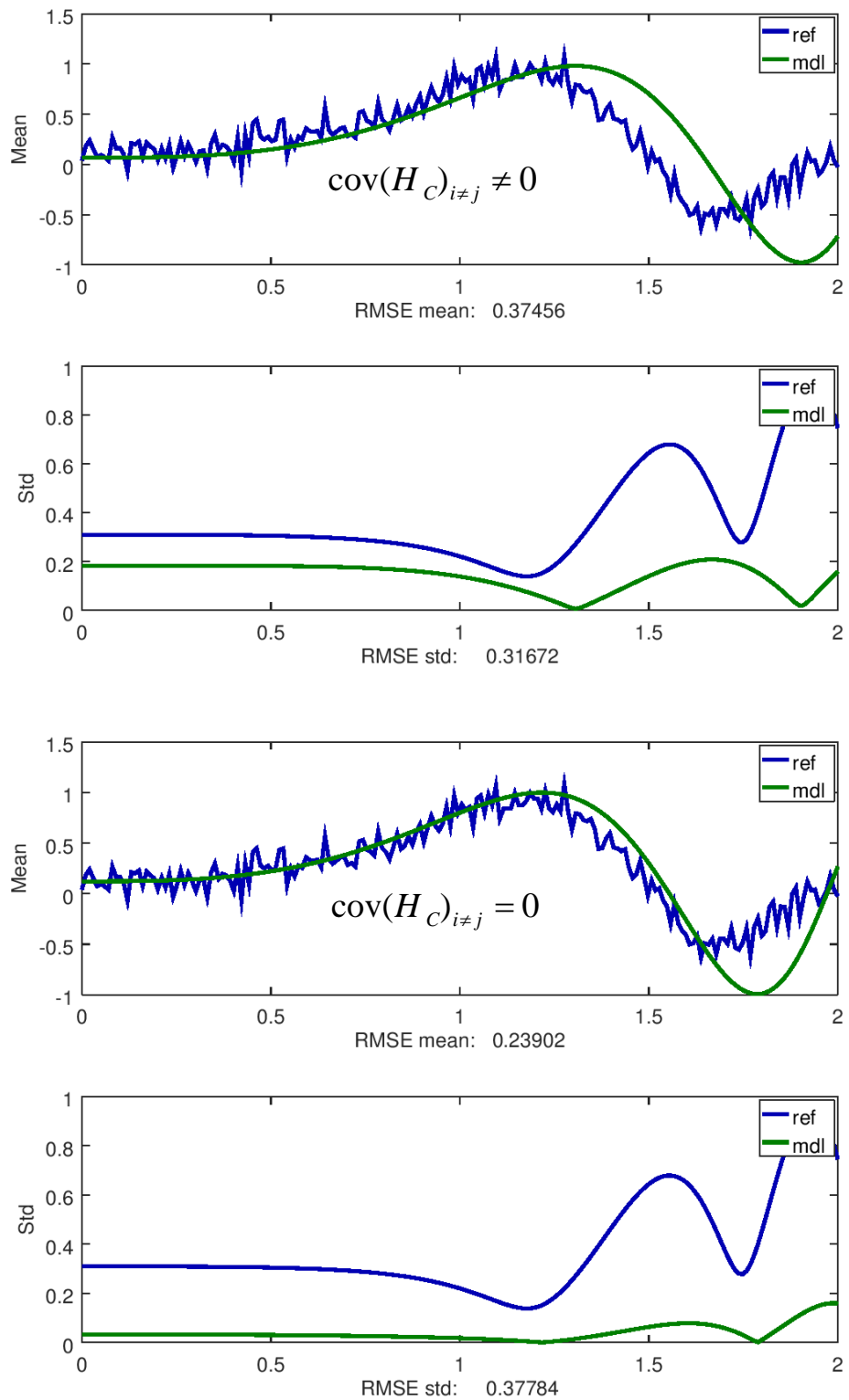


Fig. 10. Identifiering av chirp modell (Ekv. 32) med metod 1, maximum likelihood. Antagande om kalibreringsdata med (överst) och utan (underst) kovarians, för 200 mätpunkter.

GEMIX

En fullskalig kalibrering utfördes även för en relevant CFD-modell. Kalibreringsdata tillhandahölls av laboratoriet GEMIX vid Paul Scherrer Institutet i Schweiz, inom ett jämförelseprojekt organiserat av OECD [1]. För detta fall användes ca 660 observationer för kalibrering, och runt 220 för validering. Notera att kalibreringspunkterna kan begränsas till ett urval av alla tillgängliga observationspunkter. Urvalet bör vara representativt för vad som är mest intressant men även testa modellen på ett adekvat sätt. Är det beräknade fältet mycket stort, kan ett mindre urval vara nödvändigt, dock inte i detta fall.

Alla beräkningar, inklusive 10-talet iterationer utfördes under ett fåtal minuter på en ca 5 år gammal normalpresterande persondator. Det som är mest krävande är hanteringen av regressionsmatriser. Troligtvis kan den effektiviseras mångfalt, liksom hårdvaran. Totalt sett är därför bedömningen att båda kalibreringsmetoder har tillräckligt hög effektivitet för storskalig användning, även för mycket krävande beräkningsmodeller med många parametrar.

För detta realistiska fall finns förstås inget 'facit' för att kontrollera att metoderna fungerar på allra bästa sätt. Därför är det omöjligt att bestämma om representationsgraden är begränsad av modellen eller kalibreringsmetoden. Konsistensen är däremot enklare att utvärdera eftersom den beskriver hur ofta förutsägelsen inte motsägs av kalibreringsdata. Genom att öka osäkerheten kan 100% konsistens alltid uppnås (i varje fall med generell brusmodell), dock till ett förväntat högt pris av en alltför hög osäkerhet. För att undvika detta bör konsistensen ligga nära täckningsgraden. Med expansion till approximativt 95% täckningsgrad, får alltså en perfekt kalibrering konsistensen 95%.

Kalibrering 0 – Start

En rimlig startensemble Σ_0 valdes, se avsnitt Startensemble, och en ursprunglig modellensemble $H(\Sigma_0)$ beräknades. Den inkluderade alla tre testerna, N320-N337-N339, med beräknad flödes hastighet i tvärled (y), nedströms längs röret i vardera mätpunkt (x). Kalibreringsdata inkluderade totalt 660 flödes hastigheter sammanställda i en vektor. Medelvärden och standardavvikelser för modellen H_{PRIOR} , samt kalibreringsdata H_C med standardavvikelse (std) $\sqrt{\text{var}(H_C)}$ bestämd från angiven mätosäkerhet med antagande om täckningsfaktor är sammanställda i Fig. 11. Samma resultat är för tydlighets skull även uppdelat i de fem olika positionerna (kolumner), för de tre olika testfallen (rader) i Fig. 11-M (medelvärde) och Fig. 11-S (std).

Uppenbarligen överskattas osäkerheten ('Std') ganska kraftigt av CFD-modellen för den ursprungliga parameterensemblen, i detta första steg som direkt motsvarar enkel osäkerhetskvantifiering baserad på hypotes snarare än observation. Ofta är Fig. 11. ett *slut*-resultat som ska förutsäga ett rimligt täckningsintervall för de aktuella mätningarna H_C . Det ursprungliga täckningsintervallet skulle emellertid ge en konsistens på endast ca 14% enligt Fig. 12. I genomsnitt skulle endast ett mätresultat av sju vara konsistent med

beräknade resultat. Det kan tyckas förvånande eftersom 'Std' samtidigt överskattas kraftigt (Fig. 11). Problemet är att den relativt sett stora residualen överstiger även den kraftigt överskattade osäkerheten, samt att denna residual normalt helt ignoreras vid osäkerhetskvantifieringen. Det är knappast förvånande eftersom det krävs kalibreringsdata för att bedöma residualens storlek. Kalibrering tycks därför vara nödvändigt för att använda CFD-modellen på ett trovärdigt sätt. Den stora frågan blir då om residualen till en bästa skattning kan minskas så mycket att osäkerheten dominerar. Det är nämligen kravet för att konventionell kalibrering (metod 1) ska lyckas.

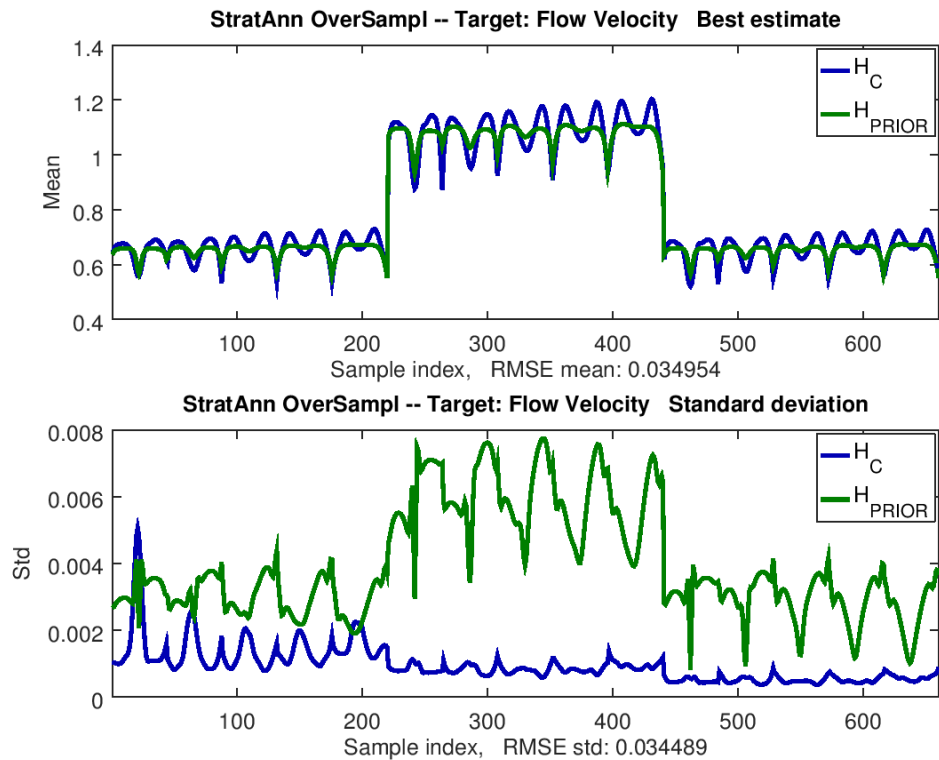


Fig. 11. Ursprungliga beräkningsresultat för CFD-modellen H_{PRIOR} , mot GEMIX H_C .

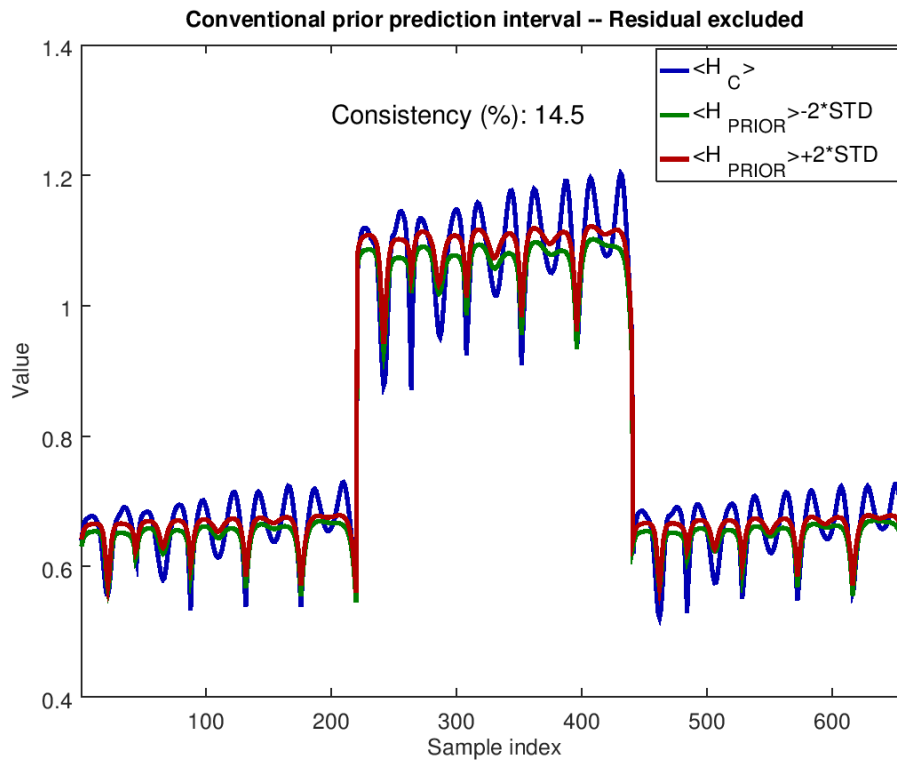


Fig. 12. Ursprunglig konsistens för CFD-modellen H_{PRIOR} , mot GEMIX H_C .

Fig. 11-M. Kalibrering 0 - Start:

Uppmätt (blå) och beräknad (grön) flödes hastighet för olika testfall (rader):

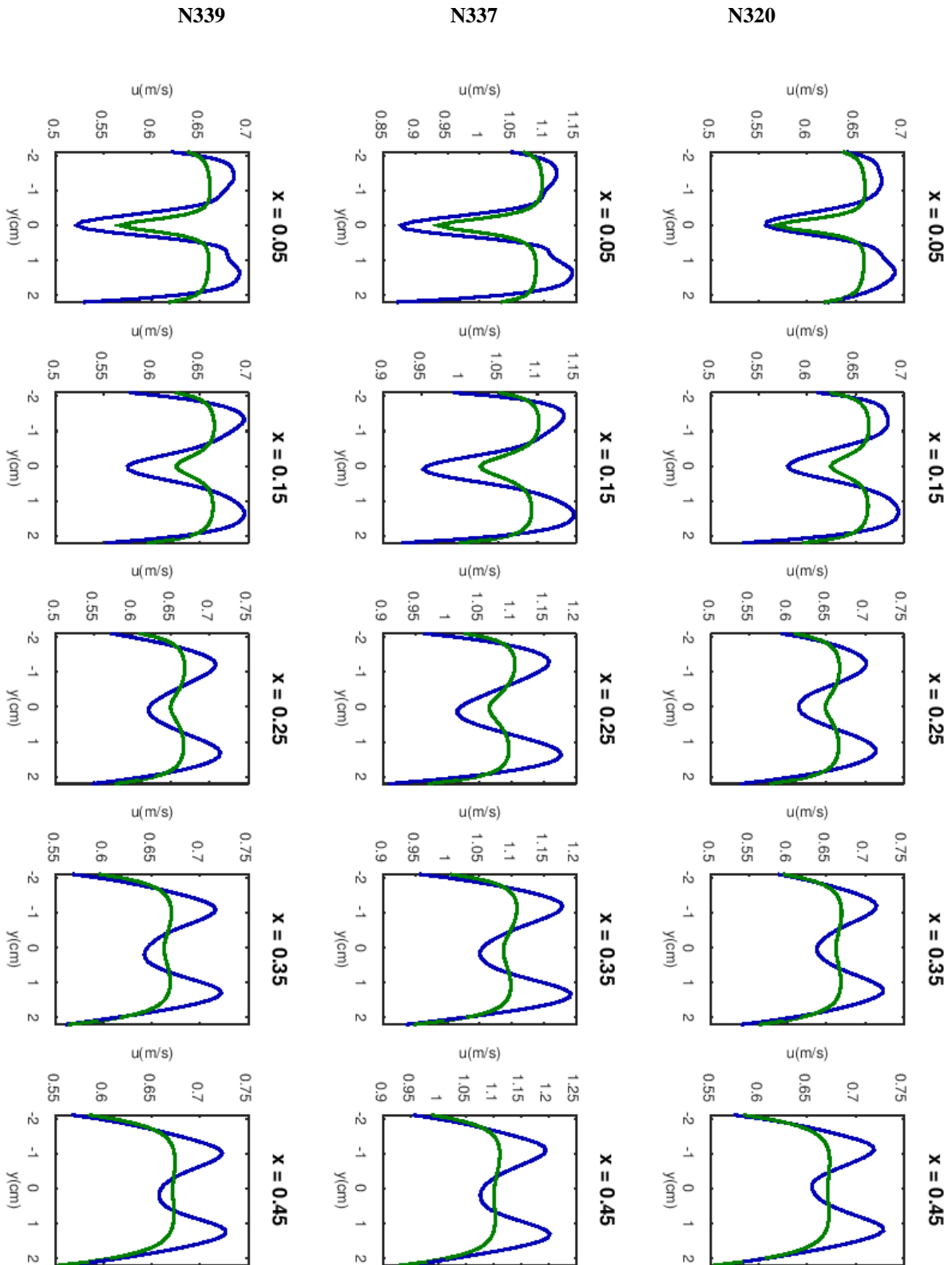
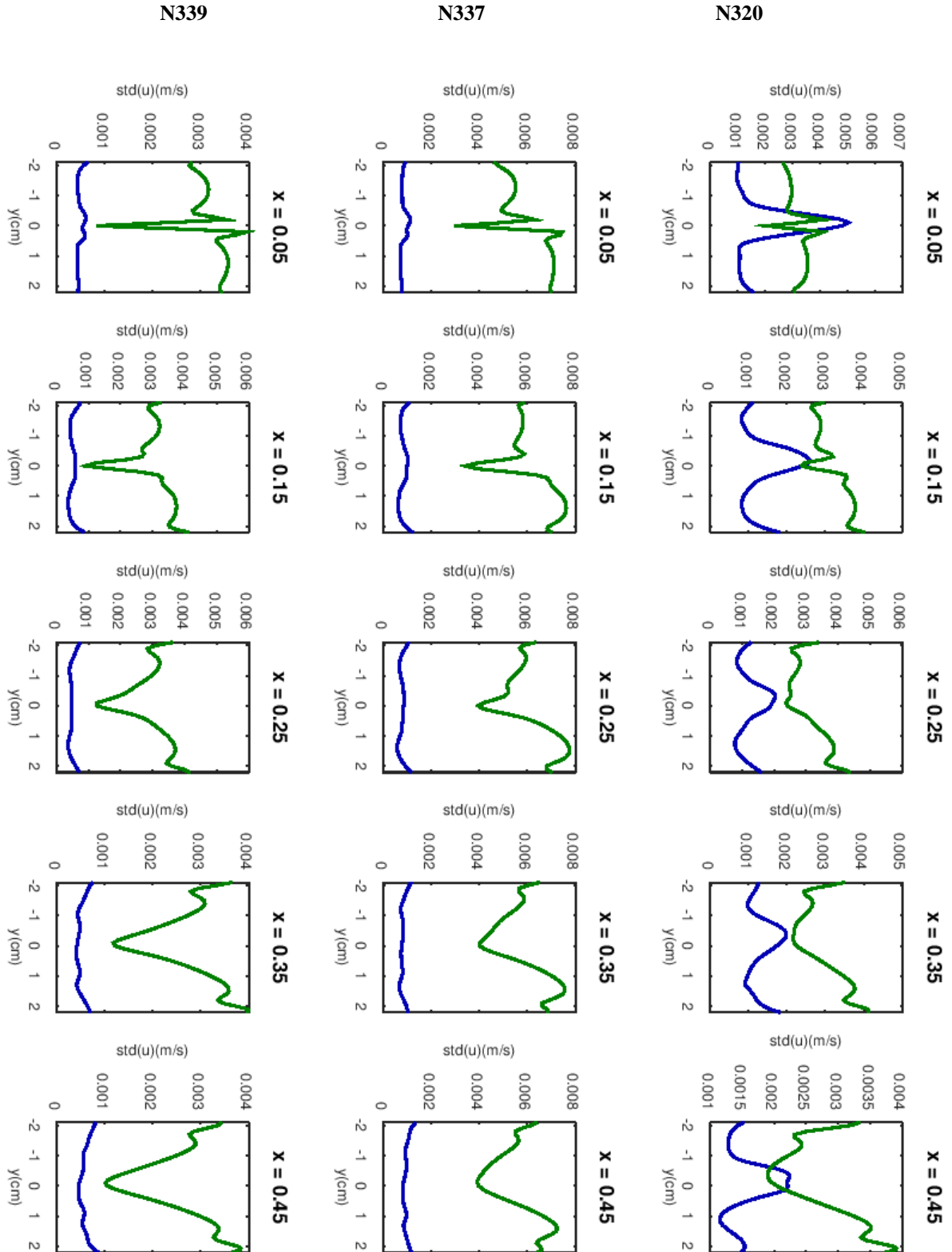


Fig. 11-S. Kalibrering 0 - Start:

Standardavvikelse för uppmätt (blå) och beräknad (grön) flödes hastighet:



Kalibrering 1

Första kalibreringen med metod 1 ger en mycket stor reduktion av prediktionsosäkerheten $\text{std}(H_{\text{POST}})$ enligt Fig. 13. Den hamnar långt under mätosäkerheten $\text{std}(H_C)$, vilket är jämförbart med resultatet i Fig. 5. Eftersom antal mätpunkter är tämligen högt, blir reduktionen med normalt oberoendeantagande om mätfel, dvs $\text{cov}_{i \neq j}(H_C) = 0$, betydande. Det överensstämmer även med resultatet i Fig. 7.

Förbättringen av det bästa skattningen som kalibreringen ger ökar konsistensen, samtidigt som den låga osäkerheten (std) minskar densamma till blygsamma 26.5%, se Fig. 16.

Första kalibreringen med metod 2 ger däremot en mer måttlig reduktion av prediktionsosäkerheten $\text{std}(H_{\text{POST}})$ enligt Fig. 14. Samma resultat är för tydlighets skull även uppdelat i de fem olika positionerna (kolumner), för de tre olika testfallen (rader) i Fig. 14-M (medelvärde) och Fig. 14-S (std). Mät- och residualmodellen är exkluderad i Fig. 14 för att lättare kunna se hur mycket av osäkerheten som representeras. Analysen för att bestämma mät- och residualmodellen som visas i Fig. 15 ger en konsistens på 88% för metod 2 enligt Fig. 17. Något lågt men betydligt bättre än för metod 1 (Fig. 16).

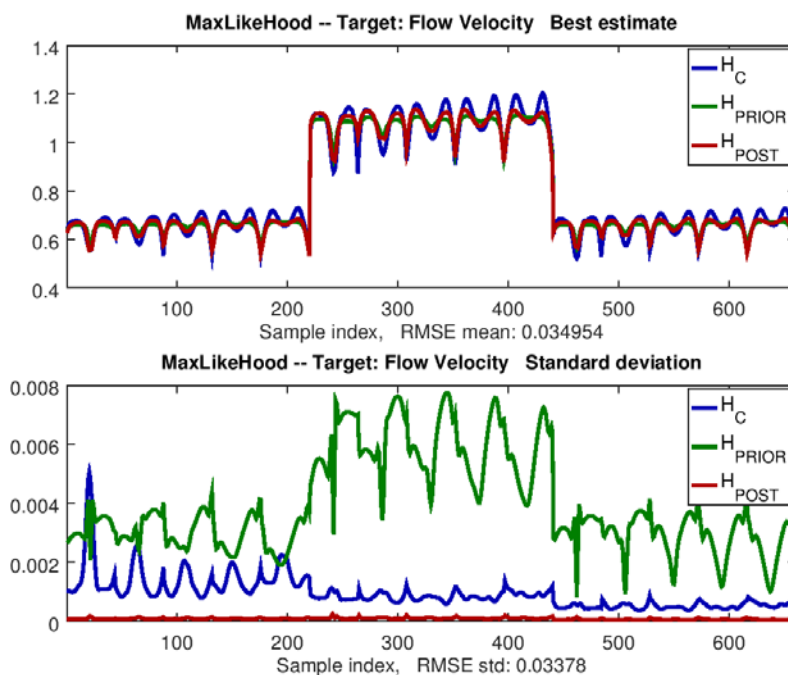


Fig. 13. Kalibrering 1, metod 1. CFD-modellen H mot GEMIX H_C .

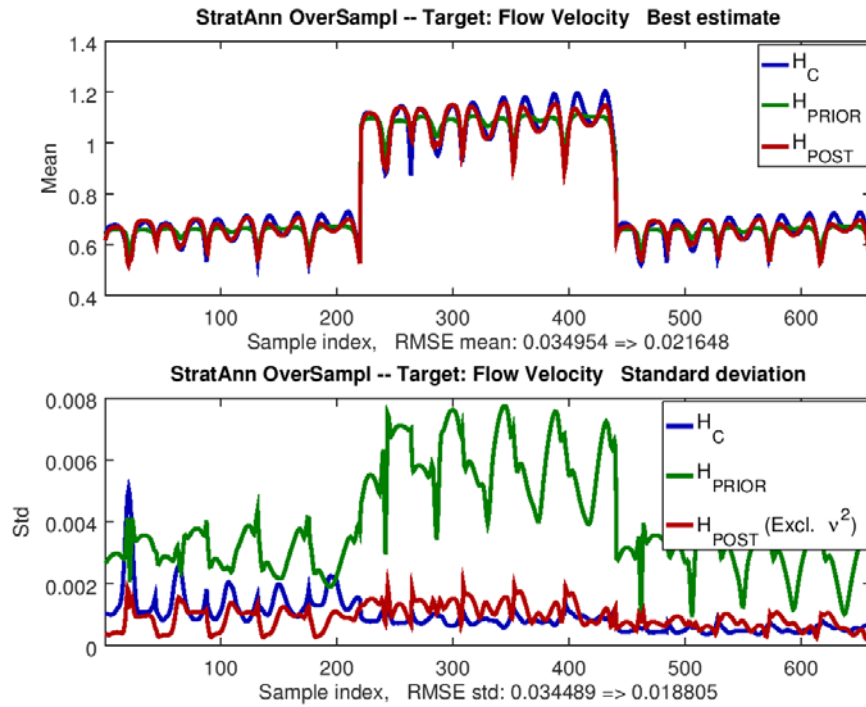


Fig. 14. Kalibrering 1, metod 2. CFD-modellen H mot GEMIX H_C . Mät- och residualmodellen (v^2) är exkluderad för bättre jämförbarhet.

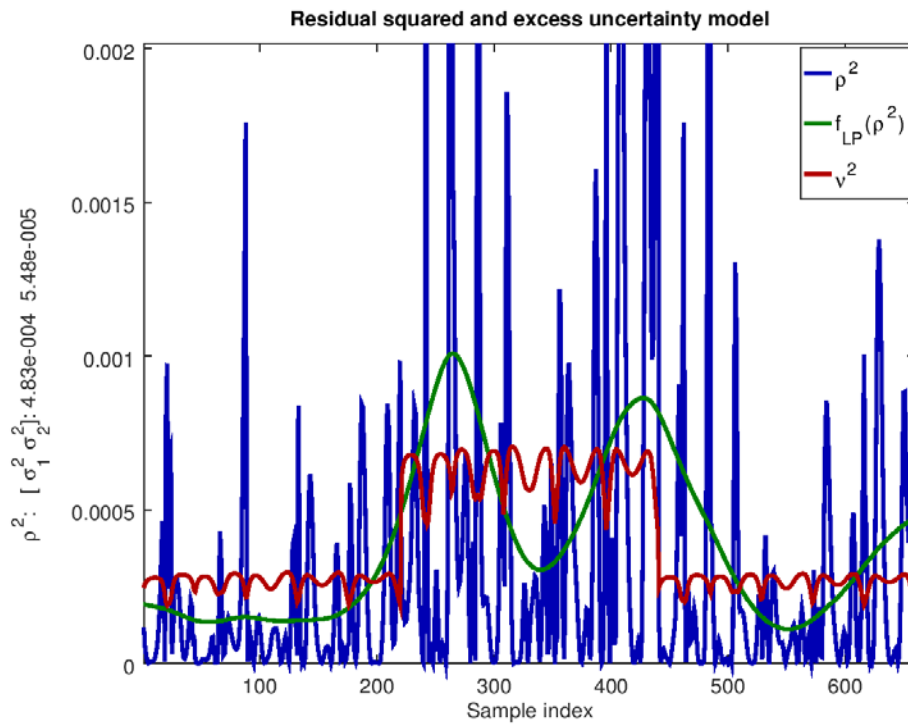


Fig. 15. Kalibrering 1, residualanalys för metod 2: Mät- och residualmodell $v^2 = \sigma_1^2 + \sigma_2^2 H^2(x)$ anpassas mot filtrerad (f_{LP}) kvadrerad residual ρ^2 .

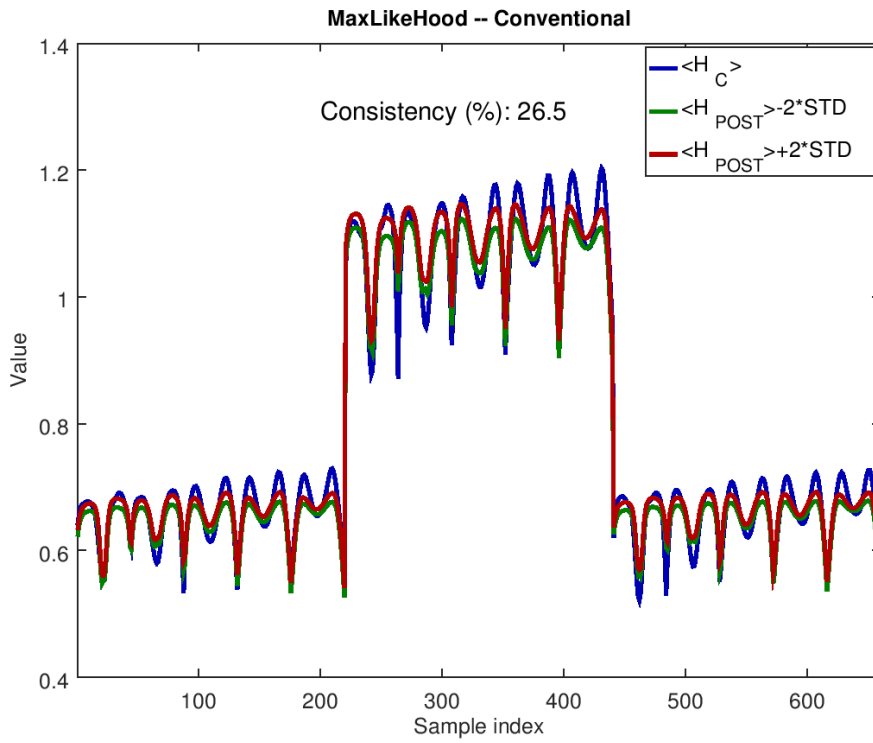


Fig. 16. Kalibrering 1, metod 1. Konsistens för CFD-modellen H_{POST} mot GEMIX H_C .

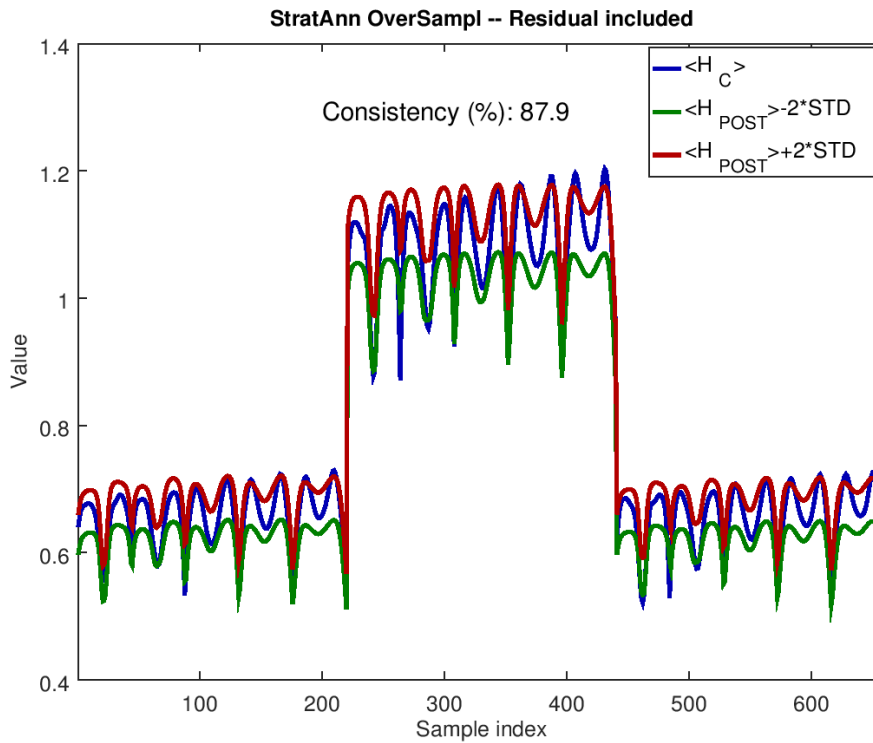


Fig. 17. Kalibrering 1, metod 2. Konsistens för CFD-modellen H_{POST} mot GEMIX H_C .

Fig. 14-M. Kalibrering 1, metod 2:

Uppmätt (blå) och beräknad flödes hastighet före (grön) och efter (röd) kalibrering:

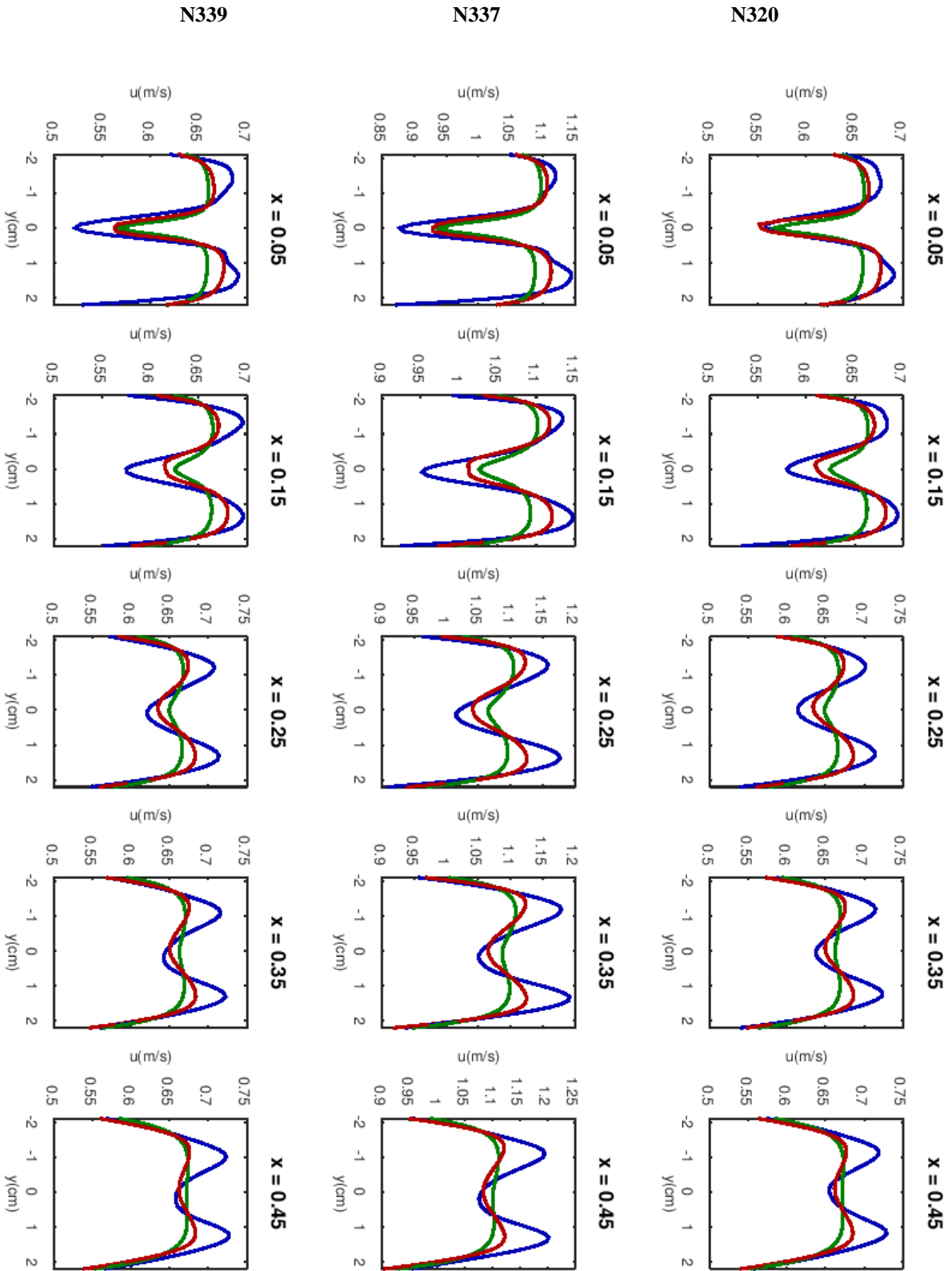
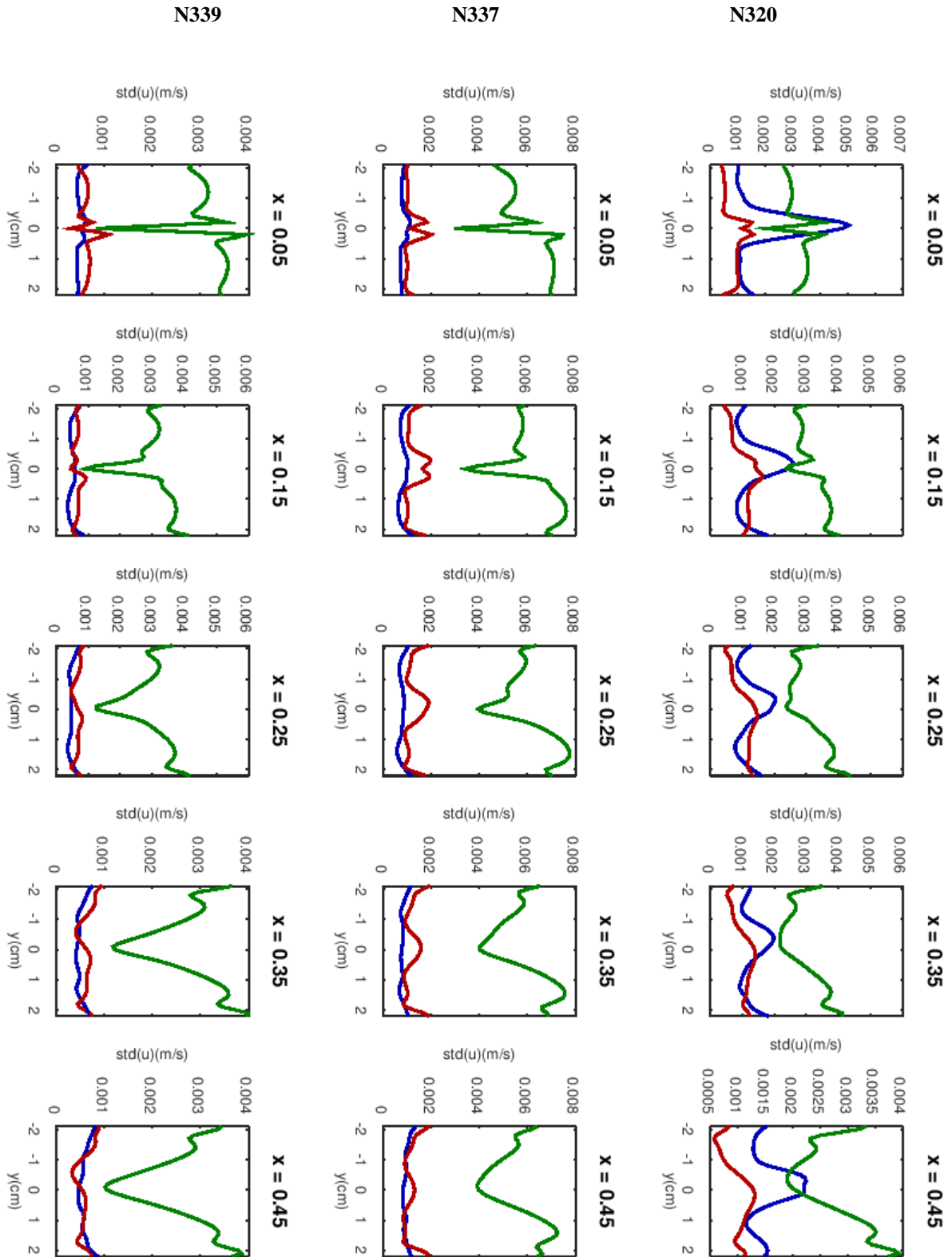


Fig. 14-S. Kalibrering 1, metod 2:
Standardavvikelse för uppmätt (blå) och beräknad flödes hastighet,
före (grön) och efter (röd) kalibrering:



Kalibrering 2 – Slutgiltig

Den andra kalibreringen med metod 2, stratifiering och annealing visade sig ge bäst resultat av alla fyra genomförda upprepningar av kalibreringen. Resultatet redovisas i Fig. 18. Den förväntade strömningshastigheten förändrades knappt alls från kalibrering 1, $H_{\text{PRIOR}} \approx H_{\text{POST}}$. Det betyder att modellensembeln befinner sig nära en fixpunkt, förhoppningsvis den globalt optimala.

Representationsgraden av standardavvikelse (Std) var också förvånansvärt hög (Fig. 18, nederst), med tanke på CFD-modellens relativt stora och komplexa residual. Slutresultatet visas även uppdelat i de fem olika positionerna (kolumner), för de tre olika testfallen (rader) i Fig. 18-M (medelvärde) och Fig. 18-S (std).

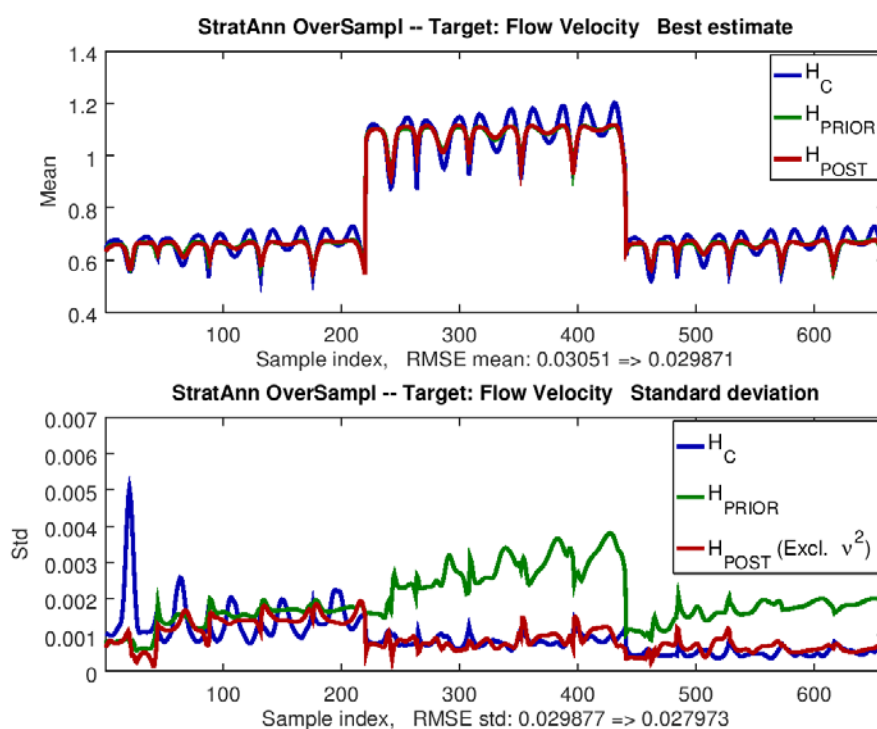


Fig. 18. Kalibrering 2, metod 2. CFD-modellen H mot GEMIX H_C . Mät- och residualmodell (v^2) är exkluderad för bättre jämförbarhet.

När modellensembeln räknas om för den kalibrerade parameterensemblen i första kalibreringen ändras den p.g.a. att metoden inte förutsäger (\tilde{H}) förändringarna helt korrekt eftersom surrogatmodell $\tilde{H}(x, \Sigma)$ används för översamplingen, dvs $\tilde{H}(x, \Sigma_{\text{POST}}^{(1)}) \neq H(x, \Sigma_{\text{POST}}^{(1)}) = H(x, \Sigma_{\text{PRIOR}}^{(2)})$ trots att $\Sigma_{\text{POST}}^{(1)} = \Sigma_{\text{PRIOR}}^{(2)}$. Förskjutningen mellan $H(x, \Sigma_{\text{POST}}^{(1)})$ i Fig. 14 och $H(x, \Sigma_{\text{PRIOR}}^{(2)})$ i Fig. 18 är ett mått på hur bra metoden är på att förutsäga dess föreslagna förflyttning av parameterensemblen. Den andra kalibreringen med metod 2 korrigerar för felet när hela modellensembeln räknas om. Mät- och residualmodellen är exkluderad för att lättare kunna se hur mycket av osäkerheten som representeras. Residualanalysen för att bestämma denna modell visas i Fig. 19. Den ger en konsistens på 96% för metod 2 enligt Fig. 20, mycket nära det uppskattat ideala (95%).

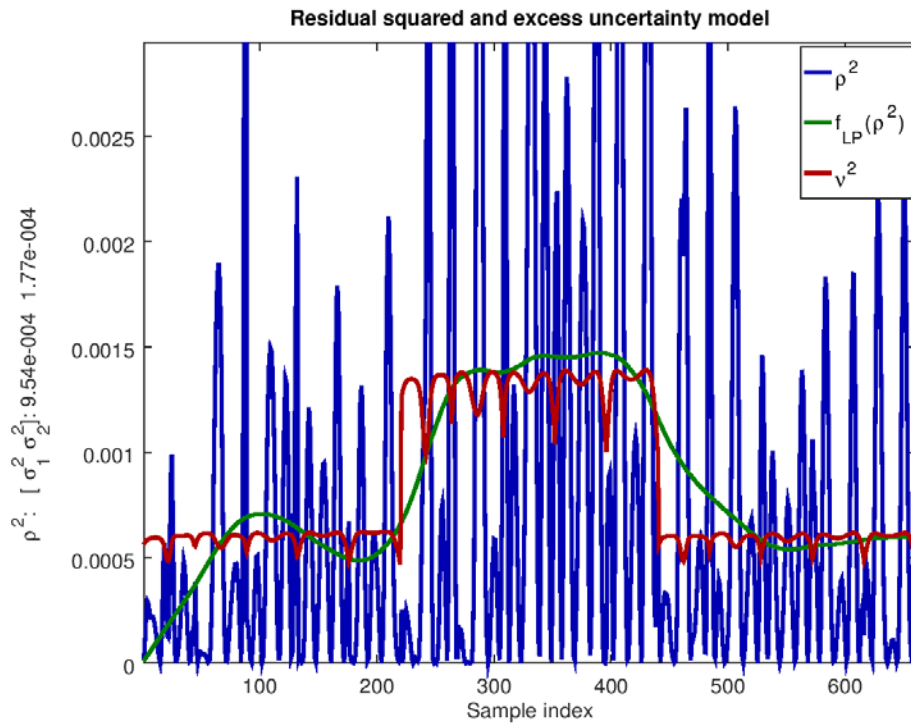


Fig. 19. Kalibrering 2, residualanalys för metod 2: Mät- och residualmodell $v^2 = \sigma_1^2 + \sigma_2^2 H^2(x)$ anpassas mot filtrerad (f_{LP}) kvadrerad residual ρ^2 .

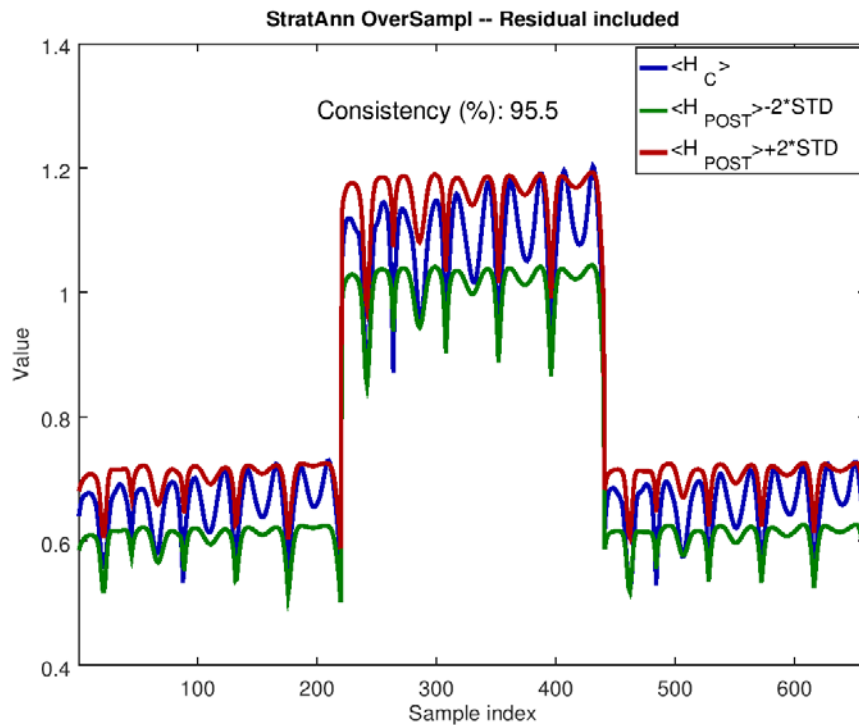


Fig. 20. Kalibrering 2, metod 2. Konsistens för CFD-modellen H_{POST} mot GEMIX H_C .

Fig. 18-M. Kalibrering 2, metod 2:

Uppmätt (blå) och beräknad flöde hastighet före (grön) och efter (röd) kalibrering:

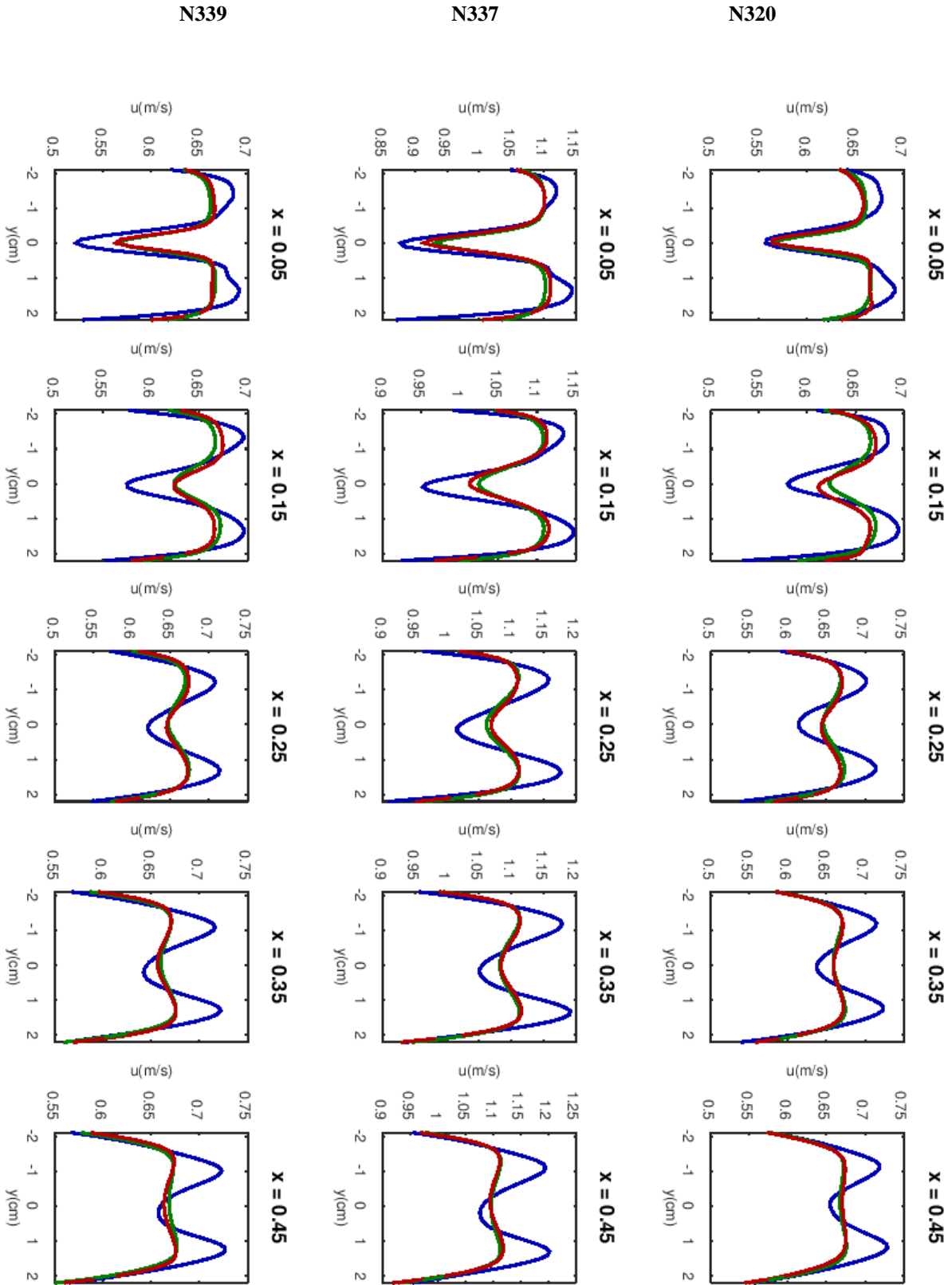
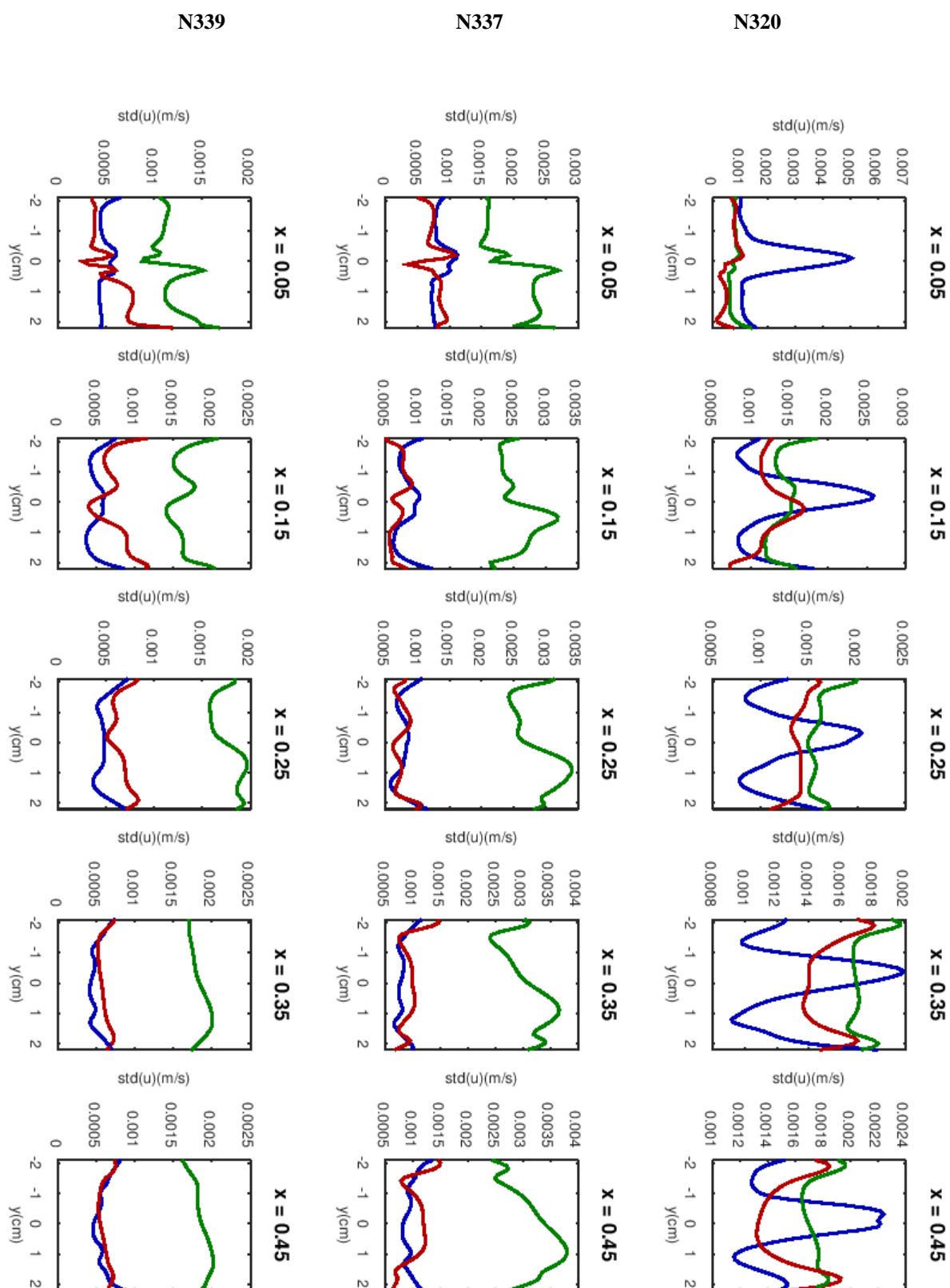


Fig. 18-S. Kalibrering 2, metod 2:

Standardavvikelse för uppmätt (blå) och beräknad flödes hastighet,
 före (grön) och efter (röd) kalibrering:



Parameterensemble (kalibrerad)

Utgångspunkten för kalibreringen var modellens ursprungliga *parameterensemble*,

$$\Sigma_0 = \begin{matrix} c_\mu : \\ C_{\varepsilon_2} : \\ \sigma_k : \\ K : \\ \sigma_T : \\ u_{IN} : \\ k_{IN} : \end{matrix} \begin{pmatrix} 0.0892 & 0.0844 & 0.0892 & 0.0844 & 0.0892 & 0.0844 & 0.0892 & 0.0844 \\ 2.0990 & 2.0990 & 1.8010 & 1.8010 & 2.0990 & 2.0990 & 1.8010 & 1.8010 \\ 1.0167 & 1.0167 & 1.0167 & 1.0167 & 0.9833 & 0.9833 & 0.9833 & 0.9833 \\ 0.4051 & 0.4149 & 0.4149 & 0.4051 & 0.4051 & 0.4149 & 0.4149 & 0.4051 \\ 0.6630 & 0.6630 & 0.8370 & 0.8370 & 0.8370 & 0.8370 & 0.6630 & 0.6630 \\ 1.0050 & 0.9950 & 1.0050 & 0.9950 & 0.9950 & 1.0050 & 0.9950 & 1.0050 \\ 0.9850 & 1.0150 & 1.0150 & 0.9850 & 1.0150 & 0.9850 & 0.9850 & 1.0150 \end{pmatrix}$$

Den slutgiltigt kalibrerade modellen bestäms av parameterensemblen,

$$\Sigma_2 = \begin{matrix} c_\mu : \\ C_{\varepsilon_2} : \\ \sigma_k : \\ K : \\ \sigma_T : \\ u_{IN} : \\ k_{IN} : \end{matrix} \begin{pmatrix} 0.061971 & 0.061665 & 0.062410 & 0.061784 & 0.061855 & 0.062095 & 0.061970 & 0.061601 \\ 2.955756 & 2.961468 & 2.922767 & 2.921832 & 2.961976 & 2.956582 & 2.921370 & 2.922165 \\ 1.042128 & 1.041441 & 1.039501 & 1.043094 & 1.043173 & 1.038942 & 1.040647 & 1.042900 \\ 0.387741 & 0.387698 & 0.387697 & 0.387764 & 0.387469 & 0.387107 & 0.387398 & 0.387428 \\ 1.111747 & 1.097606 & 1.114342 & 1.099275 & 1.096463 & 1.112644 & 1.100998 & 1.113222 \\ 1.013606 & 1.013101 & 1.013034 & 1.013583 & 1.012766 & 1.012537 & 1.013335 & 1.014457 \\ 1.017233 & 1.014410 & 1.018524 & 1.019686 & 1.014076 & 1.017856 & 1.020306 & 1.017897 \end{pmatrix} \quad (35)$$

Parameterstatistik (kalibrerad)

Modellens ursprungliga statistik före kalibrering gavs enligt Fig. 4 av:

$$\langle \theta \rangle = \begin{pmatrix} 0.0868 \\ 1.9500 \\ 1.0000 \\ 0.4100 \\ 0.7500 \\ 1.0000 \\ 1.0000 \end{pmatrix},$$

$$\text{cov}(\theta) = \begin{matrix} c_\mu : \\ C_{\varepsilon_2} : \\ \sigma_k : \\ K : \\ \sigma_T : \\ u_{\text{IN}} : \\ k_{\text{IN}} : \end{matrix} \begin{pmatrix} (2.4 \cdot 10^{-3})^2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.149^2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & (1.67 \cdot 10^{-3})^2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & (4.89 \cdot 10^{-3})^2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & (0.0870)^2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & (0.005)^2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & (0.015)^2 \end{pmatrix} \quad (36)$$

Modellens statistik efter kalibrering ($(X^{\bullet 2})_{uv} \equiv X_{uv}^2, i^2 = -1$):

$$\langle \theta \rangle^* = \begin{pmatrix} 0.0619 \\ 2.9405 \\ 1.0415 \\ 0.3875 \\ 1.1058 \\ 1.0133 \\ 1.0175 \end{pmatrix},$$

$$\text{cov}(\theta)^* = \begin{matrix} c_\mu : \\ C_{\varepsilon_2} : \\ \sigma_k : \\ K : \\ \sigma_T : \\ u_{\text{IN}} : \\ k_{\text{IN}} : \end{matrix} \begin{pmatrix} 1.679 \cdot 10^{-4} & 8.177 \cdot 10^{-4} & i3.848 \cdot 10^{-4} & 1.810 \cdot 10^{-4} & 1.102 \cdot 10^{-3} & i2.169 \cdot 10^{-4} & i4.234 \cdot 10^{-4} \\ 8.177 \cdot 10^{-4} & 1.810 \cdot 10^{-2} & 2.186 \cdot 10^{-3} & 1.415 \cdot 10^{-3} & i6.474 \cdot 10^{-3} & i1.663 \cdot 10^{-3} & i3.753 \cdot 10^{-3} \\ i3.848 \cdot 10^{-4} & 2.186 \cdot 10^{-3} & 1.283 \cdot 10^{-3} & 5.899 \cdot 10^{-5} & i3.476 \cdot 10^{-3} & 5.641 \cdot 10^{-4} & 9.089 \cdot 10^{-4} \\ 1.810 \cdot 10^{-4} & 1.415 \cdot 10^{-3} & 5.899 \cdot 10^{-5} & 5.248 \cdot 10^{-4} & i7.628 \cdot 10^{-4} & 8.220 \cdot 10^{-5} & i8.551 \cdot 10^{-4} \\ 1.102 \cdot 10^{-3} & i6.474 \cdot 10^{-3} & i3.476 \cdot 10^{-3} & i7.628 \cdot 10^{-4} & 1.232 \cdot 10^{-2} & i1.658 \cdot 10^{-3} & i2.260 \cdot 10^{-3} \\ i2.169 \cdot 10^{-4} & i1.663 \cdot 10^{-3} & 5.641 \cdot 10^{-4} & 8.220 \cdot 10^{-5} & i1.658 \cdot 10^{-3} & 3.681 \cdot 10^{-4} & 5.687 \cdot 10^{-4} \\ i4.234 \cdot 10^{-4} & i3.753 \cdot 10^{-3} & 9.089 \cdot 10^{-4} & i8.551 \cdot 10^{-4} & i2.260 \cdot 10^{-3} & 5.687 \cdot 10^{-4} & 1.778 \cdot 10^{-3} \end{pmatrix}^{\bullet 2} \quad (37)$$

Korrelation parametrar (kalibrerade)

Hur starka beroenden är mellan olika kalibrerade parametrar beskrivs av korrelationsmatrisen, här benämnd X ($(X \cdot Y)_{uv} \equiv X_{uv} \cdot Y_{uv}$),

$$\text{cov}(\theta) \equiv \Gamma \Gamma^T \cdot X,$$

där

$$\Gamma = \begin{pmatrix} \sigma_1 \\ \sigma_2 \\ \sigma_3 \\ \sigma_4 \\ \sigma_5 \\ \sigma_6 \\ \sigma_7 \end{pmatrix},$$

$$X = \begin{matrix} c_\mu : \\ C_{\varepsilon 2} : \\ \sigma_k : \\ K : \\ \sigma_T : \\ u_{IN} : \\ k_{IN} : \end{matrix} \begin{pmatrix} 1.000 & 0.220 & -0.687 & 0.372 & 0.588 & -0.761 & -0.601 \\ 0.220 & 1.000 & 0.206 & 0.211 & -0.188 & -0.415 & -0.438 \\ -0.687 & 0.206 & 1.000 & 0.005 & -0.765 & 0.674 & 0.362 \\ 0.372 & 0.211 & 0.005 & 1.000 & -0.090 & 0.035 & -0.784 \\ 0.588 & -0.188 & -0.765 & -0.090 & 1.000 & -0.607 & -0.233 \\ -0.761 & -0.415 & 0.674 & 0.035 & -0.607 & 1.000 & 0.494 \\ -0.601 & -0.438 & 0.362 & -0.784 & -0.233 & 0.494 & 1.000 \end{pmatrix} \quad (38)$$

Dessa [starka] korrelationer motsvarar den förkunskap som CFD-modellens ekvationer beskriver i abstrakt form. Att uppskatta dessa på annat sätt är genom kalibrering är mycket komplicerat. Korrelationer mellan parametrar är emellertid mycket viktiga vid osäkerhetskvantifiering. Därför krävs kalibrering för att kunna ange en trovärdig osäkerhet för modellens resultat som representerar experimentella observationer.

Validering

En slutgiltig kontroll av kalibreringen bör alltid göras med en validering mot kalibreringsdata som inte använts vid identifieringen. I detta fall har testfallet N318 valts för validering i benchmark-projektet för GEMIX experimentet, se avsnitt GEMIX kalibrering och Fig. 3.

Giltighet

Val av valideringsfall kan emellertid knappast anses helt korrekt eftersom det är alltför lika kalibreringsfallen. Valideringsfallet kan nämligen återskapas *exakt*(!), dvs geometri, randvillkor etc. genom interpolation mellan de tre kalibreringsfallen. Därför blir valideringsdata starkt *beroende* av kalibreringsdata, vilket inte kan anses acceptabelt. Med andra ord bryter valideringen mot kravet på att vara avvikande men ha en likartad experimentell design som kalibreringen. Det ska garantera ett så oberoende test som möjligt, för att på ett så fullständigt sätt som möjligt kunna utvärdera kalibreringens effekt.

Med aktuellt valideringsfall blir det dessvärre lönsamt att justera modellens resultat i varje observationspunkt, för att reproducera systematiska modellfel som är specifika för kalibreringsfallen men i övrigt saknar allmän giltighet. Denna justering gjordes också av deltagare inom GEMIX-projektet, som beskrivet i avsnitt Traditionell justering inom GEMIX-projekt. Vid konventionell systemidentifiering krävs det nämligen att residualen inte får ha någon signifikant tvärkorrelation med insignalen [7], vilken i detta fall får översättas med kalibreringsdata. Detta villkor kan dock aldrig vara uppfyllt för modeller med systematiska residualfel, vars auto-korrelation redan gör dem inkonsistenta i konventionell mening. Det inträffade mycket riktigt för kalibreringen av CFD-modellen med metod 1.

Utvärdering

Förutsägelse av förbättringar i varje kalibrering har generellt överskattats när översamplingsmodellensemblen beräknats med surrogatmodeller. Det har upptäckts när den kalibrerade parameterensemblen senare beräknats med den fulla modellen i nästa iteration. Därför har sannolikt förbättringen överskattats även med den slutgiltiga kalibreringen visad i Fig. 18. En direkt jämförelse med ursprunglig och kalibrerad ensemble för valideringsfallet visas i Fig. 21. Det illustrerar den totala effekten, dvs. nyttan av kalibreringen utan modellapproximation i någon form.

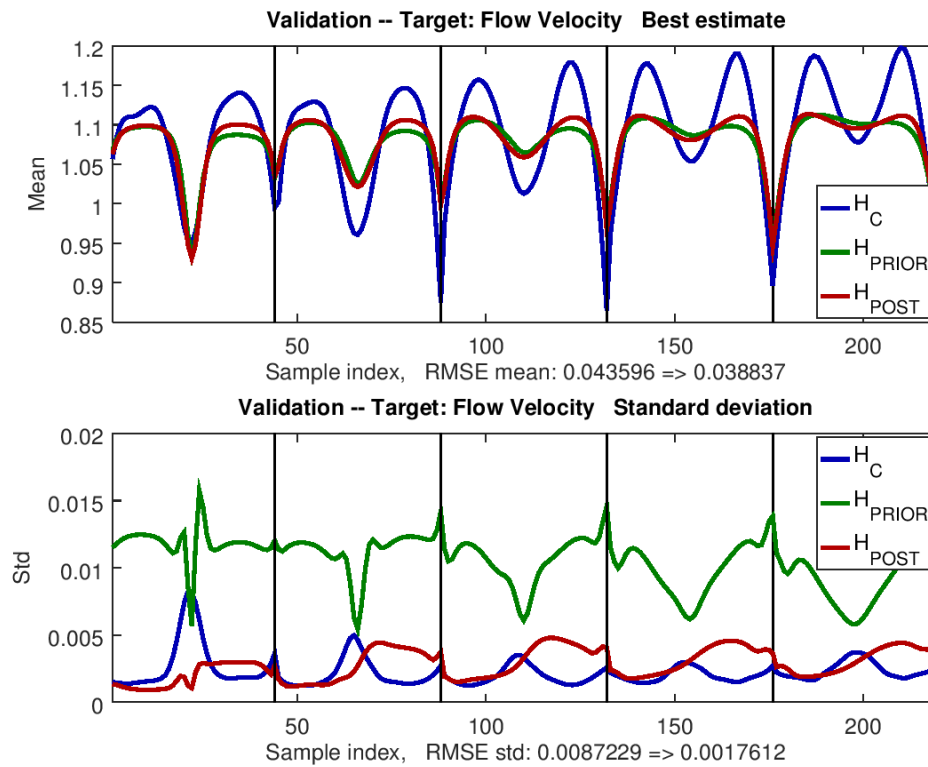


Fig. 21. Validering av ursprunglig H_{PRIOR} och slutkalibrerad H_{POST} CFD-modell mot GEMIX H_C , för testfall N318.

Den bästa skattningens RMSE (Root Mean Square Error) minskar endast från 0.0436 till 0.0388 (Fig. 21 överst), vilket antyder att den ursprungliga fysikaliska deterministiska modellen i stort sett är så bra den kan vara. Att pragmatiskt justera [24] parametrarna med kalibrering ger relativt marginella vinster. Beräkningsmodellens osäkerhet ändras däremot mycket (Fig. 21 nederst). Representationen verkar acceptabel även för valideringen, både för medelvärde och standardavvikelse.

Konsistensen förblir därför mest en fråga om hur residualen hanteras eftersom den är så dominerande i relation till standardosäkerheten (std). Speciellt om den tillåts ingå i beräkningsmodellens osäkerhet som i metod 2, eller inte som i metod 1. Enligt Fig. 22 och Fig. 23 innebär kalibrering i detta fall en mindre justering från 92% till 94% för valideringen (mot 96% för den slutgiltiga kalibreringen enligt Fig. 20).

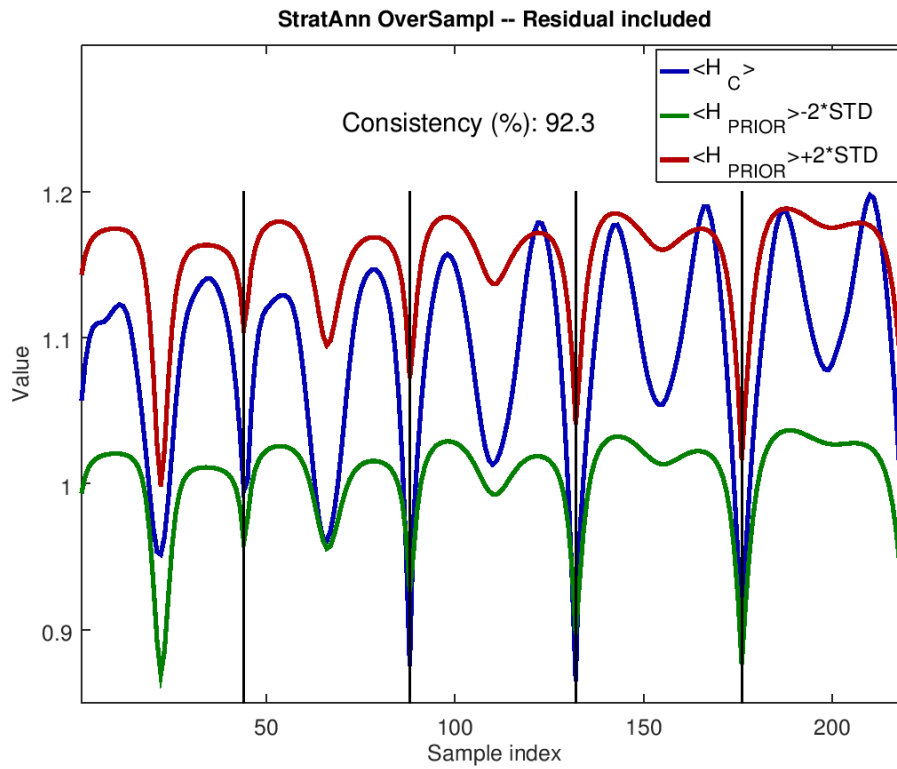


Fig. 22. Konsistens för ursprunglig CFD-modell H_{POST} mot GEMIX H_C , enl. metod 2.

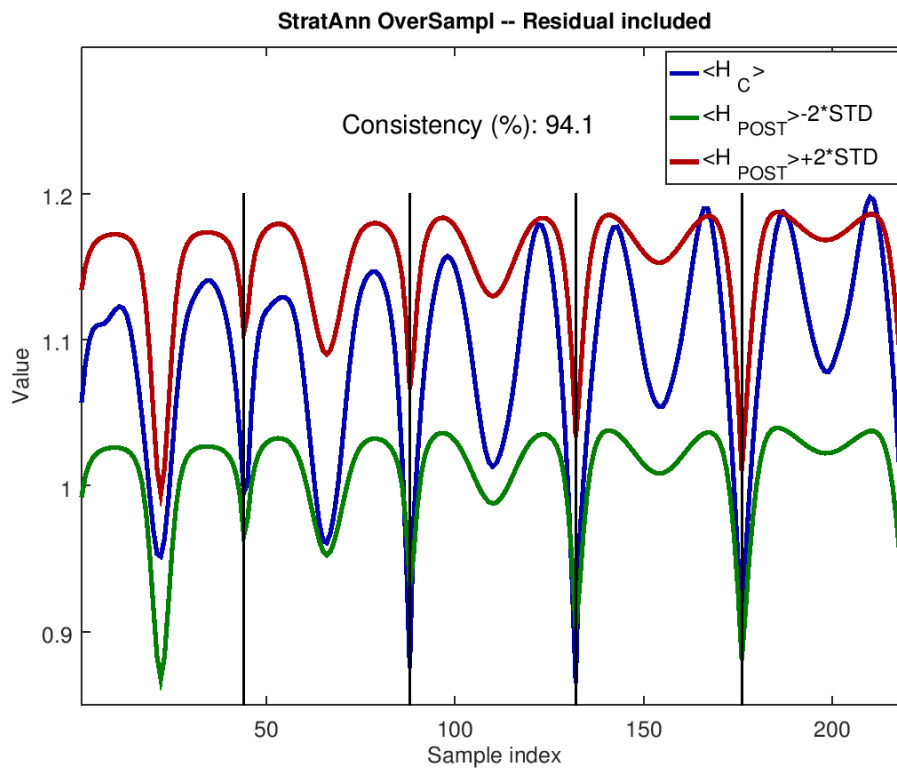


Fig. 23. Konsistens för slutkalibrerad CFD-modell H_{POST} mot GEMIX H_C , enl. metod 2.

Slutsats

Två olika metoder för kalibrering av beräkningsmodeller har studerats. Båda är iterativa med successiva förbättringar av deterministiskt bestämda modellsampel. Den första metoden baserades på traditionell bästa skattning av en modell med maximum likelihood-teknik. En hypotes ställs före kalibrering om modellstrukturens absoluta riktighet, vilken inte testas förrän i valideringen. Den andra metoden bygger på ett helt nytt angreppssätt med s.k. osäkra modeller. De ska på bästa sätt beskriva vad som anses vara känt från ett kalibreringsexperiment, både förväntat resultat och förväntat fel i form av osäkerhet. Till skillnad mot metod 1 ställs inga hypoteser om modellens riktighet. Modellstrukturen är istället endast en bas i vilken kunskap från kalibreringen uttrycks, eller *representeras*. Den begränsning som användningen av modellen innebär utgör deterministisk förkunskap, som reducerar antagna osäkerheter på motsvarande sätt som Bayes metod reducerar osäkerhet med statistisk förkunskap.

Metod 1 föreslogs före projektet och har varit en viktig referens för utveckling av metod 2. Enkla testmodeller har använts för att studera skillnader mellan metoderna och vad dessa innebär: Användbarhet, effektivitet, förväntad grad av konsistens vid prediktion med kalibrerad modell, tolkning av osäkerhet och hur systematiska fel kan inkluderas. Även hur och varför kalibrerade modeller inte blir godkända vid validering har undersökts.

Ett antal nya begrepp och tekniker såsom t.ex. *konsistens* och *representationsgrad* har införts för att beskriva egenskaperna för metod 2. Bestämningen av sampelvikter för metod 2, s.k. *stratifiering*, har justerats i ett flertal steg till att slutligen blivit baserad på översampling av modellen för att finna dess optimala justering. Acceptabel konsistens uppnås även för modeller med stora systematiska fel genom att inkludera residualen i osäkerheten. Unikt är också att inkludera det oundvikliga felet i kalibreringsdata som kalibreringsmätningen bidrar med. En vetenskaplig detaljerad publicering är planerad.

Kalibreringsmetoderna har testats på en CFD-modell med kalibreringsdata från GEMIX experiment som använts för s.k. benchmark i osäkerhetskvantifiering inom ett OECD projekt. Sammanlagt fyra iterationer av kalibreringar genomfördes, samt validering. Kalibreringen av CFD-modellen för GEMIX-experimentet blev dock huvudsakligen begränsad till en större justering av osäkerhet. Orsaken var att modellens systematiska fel var helt dominerade och att justeringar av modellens fysikaliska parametrar inte kunde reducera dessa nämnvärt. Valideringsfallet var alltför lika kalibreringsfallen för att erbjuda en bra kontroll. CFD-modellen underkändes efter kalibrering med metod 1, vilket var förväntat p.g.a. dess stora systematiska fel som strider mot metodens grundantagande. Metod 2 gav däremot en godkänd modell med rimlig konsistens på avsett sätt, genom att inkludera residualen i en stor osäkerhet. Även om CFD-modellen blev godkänd i detta fall, kan nyttan med att använda den vara begränsad med tanke på dess stora osäkerhet.

Sammanfattningsvis fungerade båda kalibreringsmetoderna som avsett, för både testmodeller och GEMIX. Modeller kalibrerade med metod 2 får allmänt mycket högre

osäkerhet, eftersom den *inte* baseras på de starka antaganden och hypoteser om både modellens giltighet och kalibreringsdatas osäkerhet som kännetecknar allmän praxis för metod 1. Metod 2 är tvärtom speciellt konstruerad för att vara förlåtande mot systematiska modellfel som i praktiken är svåra att undvika vid avancerad modellering, simulering och beräkning. Därför bedöms potentialen för metod 2 vara avsevärt större än för metod 1.

Referenser

- [1] OECD/NEA CFD–UQ Benchmark Exercise: CFD prediction and Uncertainty Quantification of a GEMIX mixing layer test (Project description)

OECD/NEA CFD–UQ Benchmark Exercise: CFD prediction and Uncertainty Quantification of a GEMIX mixing layer test (Benchmark project description)
A. Badillo, B. Niceno, R. Kapulla, J. Fokken*, B. Krohn* and H.-M. Prasser, OECD/NEA PSI GEMIX Benchmark Exercise (Presentation)

NEA/CSNI/R(2016)4: REVIEW OF UNCERTAINTY METHODS FOR CFD APPLICATION TO NUCLEAR REACTOR THERMALHYDRAULICS, Organisation for Economic Co-operation and Development, (25-Feb-2016) (Project Report, unclassified)
- [2] SSM-Projekt 2037012-71: Modellosäkerhet i beräkningar, 2014-02-01 – 2015-12-31.
- [3] Wilks, Samuel S., Determination of sample sizes for setting tolerance limits, The Annals of Mathematical Statistics 12.1 (1941): 91-96.
- [4] Hessling, Jan Peter, Deterministic sampling for propagating model covariance, SIAM/ASA Journal on Uncertainty Quantification 1.1 (2013): 297-318.
- [5] Smith R., Uncertainty Quantification – Theory, Implementation and Applications, SIAM (2014).
- [6] Kay, Steven M., Statistical signal processing, Estimation Theory 1 (1993).
- [7] Ljung, Lennart, System Identification: Theory for the User, PTR Prentice Hall Information and System Sciences Series (1999).
- [8] Fisher, Ronald Aylmer. Statistical methods, experimental design, and scientific inference. Oxford Univ. Press (1990).
- [9] Rubenstein Kroese, Simulation and the Monte Carlo method, 2nd ed., John Wiley & Sons (2008).
- [10] Sivia, Devinderjit, and John Skilling. Data analysis: a Bayesian tutorial. OUP Oxford (2006).
- [11] Weisberg, Herbert I., Willful ignorance: The mismeasure of uncertainty, John Wiley & Sons (2014).
- [12] Jaynes, Edwin T., Information theory and statistical mechanics, Physical review 106.4 (1957): 620.
- [13] Hessling, Jan Peter, Subjectivity in application of the principle of maximum entropy, Open Journal of Statistics 2013 (2013).
- [14] Shannon, C. E., A mathematical theory of communication, Bell System Technical Journal 27: 379-423 and 623–656, Mathematical Reviews (MathSciNet): MR10, 133e (1948).

- [15] Kreyszig, Introductory Functional Analysis with Applications, Krieger, Malabar, FL (1989).
- [16] Ellenberg, Jordan, How not to be wrong: The power of mathematical thinking, Penguin (2015).
- [17] Carlin, Bradley P., and Thomas A. Louis, Bayesian methods for data analysis, CRC Press (2008).
- [18] Hessling, Jan Peter, Identification of complex models, SIAM/ASA Journal on Uncertainty Quantification 2.1 (2014): 717-744.
- [19] Hessling Jan Peter, Uncertain models, opublicerad.
- [20] Jönsson Henrik och Migdalas Sakis, Icke-linjär programmering, Optimeringslära, Matematiska Inst., Linköpings Universitet (1989).
- [21] https://en.wikipedia.org/wiki/Newton's_method
- [22] Händel Peter, Ottoson Ragnar and Hjalmarsson Håkan, Signalteori (2002).
- [23] Hessling Jan Peter, Annealing of deterministic ensembles, opublicerad.
- [24] Kennedy, Marc C., and Anthony O'Hagan, Bayesian calibration of computer models, Journal of the Royal Statistical Society: Series B (Statistical Methodology) 63.3 (2001): 425-464.



2017:23

Strålsäkerhetsmyndigheten har ett samlat ansvar för att samhället är strålsäkert. Vi arbetar för att uppnå strålsäkerhet inom en rad områden: kärnkraft, sjukvård samt kommersiella produkter och tjänster. Dessutom arbetar vi med skydd mot naturlig strålning och för att höja strålsäkerheten internationellt.

Myndigheten verkar pådrivande och förebyggande för att skydda människor och miljö från oönskade effekter av strålning, nu och i framtiden. Vi ger ut föreskrifter och kontrollerar genom tillsyn att de efterlevs, vi stödjer forskning, utbildar, informerar och ger råd. Verksamheter med strålning kräver i många fall tillstånd från myndigheten. Vi har krisberedskap dygnet runt för att kunna begränsa effekterna av olyckor med strålning och av avsiktlig spridning av radioaktiva ämnen. Vi deltar i internationella samarbeten för att öka strålsäkerheten och finansierar projekt som syftar till att höja strålsäkerheten i vissa östeuropeiska länder.

Strålsäkerhetsmyndigheten sorterar under Miljödepartementet. Hos oss arbetar drygt 300 personer med kompetens inom teknik, naturvetenskap, beteendevetenskap, juridik, ekonomi och kommunikation. Myndigheten är certifierad inom kvalitet, miljö och arbetsmiljö.

Strålsäkerhetsmyndigheten
Swedish Radiation Safety Authority

SE-17116 Stockholm
Solna strandväg 96

Tel: +46 8 799 40 00
Fax: +46 8 799 40 10

E-mail: registrator@ssm.se
Web: stralsakerhetsmyndigheten.se