

---

# Modellierung und Kalibrierung von Computermodellen

## mit Anwendung auf einen Umformprozess

---

### Dissertation

zur Erlangung des Grades  
eines Doktors der Naturwissenschaften  
der Technischen Universität Dortmund



Der Fakultät Statistik  
der Technischen Universität Dortmund

vorgelegt von  
**Hilke Kracker**

Weil am Rhein, 2011

1. Gutachterin: Prof. Dr. Ursula Gather

2. Gutachterin: PD. Dr. Sonja Kuhnt

Tag der mündlichen Prüfung: 21.06.2011

# INHALTSVERZEICHNIS

<b>1</b>	<b>Einleitung</b>	<b>1</b>
<b>2</b>	<b>Gauß Prozess Interpolation</b>	<b>7</b>
2.1	Grundlagen der Gauß Prozess Interpolation . . . . .	8
2.2	Vorhersagen bei der Gauß Prozess Interpolation . . . . .	11
2.3	Maximum Likelihood Ansatz für die Gauß Prozess Interpolation . . . . .	18
2.4	Gauß Prozess Regression . . . . .	20
2.5	Zusammenfassung und Bemerkungen . . . . .	21
<b>3</b>	<b>Vorhersageintervallüberdeckung bei der Gauß Prozess Interpolation</b>	<b>23</b>
3.1	Objektive a priori Verteilungen für Korrelationsparameter . . . . .	25
3.2	Untersuchungen zur Überdeckungsgüte . . . . .	29
<b>4</b>	<b>Kalibrierung von Computermodellen</b>	<b>41</b>
4.1	Notation und allgemeine Modellformulierung . . . . .	42
4.2	Kalibrierung als nichtlineares Regressionsproblem . . . . .	45
4.3	Bayesianische Kalibrierung von Computermodellen bei Verzerrung . . . . .	51
4.4	Simulationsstudie zur Bayesianischen Kalibrierung . . . . .	65
<b>5</b>	<b>Kalibrierung von Computermodellen unter Modellunsicherheit</b>	<b>83</b>
5.1	Kalibrierungsmodell mit Modellunsicherheitskomponente . . . . .	84
5.2	Simulationsstudie zur Kalibrierung unter Modellunsicherheit . . . . .	93
<b>6</b>	<b>Anwendung: Kalibrierung einer Rückfederungssimulation</b>	<b>103</b>
6.1	Kalibrierung einer Rückfederungssimulation durch numerische Parameter	105
6.2	Kalibrierung einer Rückfederungssimulation durch Materialparameter .	111
<b>7</b>	<b>Fazit und Ausblick</b>	<b>123</b>

<b>A Anhang</b>	<b>127</b>
A.1 Monte Carlo Methoden . . . . .	127
<b>Symbolverzeichnis</b>	<b>131</b>
<b>Literaturverzeichnis</b>	<b>133</b>

# 1 EINLEITUNG

Die vorliegende Arbeit beschäftigt sich mit der statistischen Modellierung, Optimierung und Kalibrierung von Computermodellen in der Blechumformung. Unter einem Computermodell verstehen wir ein mathematisches Modell, das einen realen Prozess nachbildet. Häufig handelt es sich um physikalische Prozesse, die durch Systeme von Differentialgleichungen beschrieben werden. Diese sind oft so komplex, dass sie nur numerisch auf dem Computer gelöst werden können. Typischerweise sind Computermodelle deterministisch, das heißt wiederholtes Ausführen für gleiche Anfangsbedingungen liefert das gleiche Resultat. Häufige Anwendung von Computermodellen sind unter anderem Atmosphärenmodelle zur Wettervorhersage, Klimamodelle zur Prognose des Klimawandels, pharmakokinetische Modelle zur Modellierung von Medikamenteffekten im Menschen oder Modelle zur Beschreibung von Produktionsprozessen im Maschinenbau.

In der vorliegenden Arbeit werden wir Computermodelle zur Beschreibung eines Blechumformungsprozesses betrachten, die mit Hilfe von Finite-Elemente Methoden ausgewertet werden. In diesem Zusammenhang ist es von Interesse den Blechumformungsprozess so zu optimieren, dass das Produkt des Umformprozesses bestimmte gewünschte Eigenschaften hat. Fragestellungen, die sich dem Ingenieur in diesem Zusammenhang stellen, sind zum Beispiel: Welche Blechdicke soll verwendet werden? Mit welchem Druck soll das Werkstück in die Maschine eingespannt werden? Welche Bauteilgeometrie ist geeignet? Wie sollen unbekannte physikalische Konstanten gewählt werden? Das Computermodell für den Blechumformprozess kann dann für diese Einstellgrößen das Ergebnis simulieren, und der Ingenieur kann somit eine optimale Einstellung für den Blechumformprozess finden. Ganz allgemein kann man die Ausgabe eines Computermodells also als eine deterministische Funktion auffassen, die für beliebige Elemente einer Definitionsmenge ausgewertet werden kann, aber keine einfache geschlossene Darstellung besitzt. Eine Funktionsauswertung wird Simulationslauf oder in der Statistik

auch Computerexperiment genannt, da sie echte Experimente zum Teil ersetzen. Die statistische Modellierung von deterministischen Computermodellausgaben ist aus verschiedenen Gründen sinnvoll. Ein wichtiger Punkt ist, dass Computermodelle sehr rechenintensiv sein können. Je nach Anwendung und Problem können einzelne Simulationsläufe eine Stunde oder einen ganzen Tag in Anspruch nehmen. Die Durchführung einer Vielzahl von Simulationsläufen in Trial-and-Error Zyklen zum Beispiel für eine Optimierung ist in diesen Situationen ineffizient. Eine Alternative besteht in einer statistischen Modellierung von Computermodellausgaben. Dabei werden die Zusammenhänge zwischen Einflussgrößen und einer oder mehrerer Zielgrößen aufgrund der begrenzten Anzahl von Computerexperimenten statistisch approximiert. Für weitere Analysen, wie zum Beispiel die Optimierung, dient die Approximation dann als Ersatz für das Computermodell. Neben der reinen Approximation ist bei der Modellierung von Computermodellausgaben auch die Unsicherheit über die Approximation von Interesse. Auch Methoden zur effizienten Versuchsplanung kommen im Zusammenhang von Computermodellen häufig zum Einsatz.

Trotz der Tatsache, dass das Computermodell auf substanzwissenschaftlichen Überlegungen beruht, bleibt es nur ein Modell des wahren physikalischen Prozesses, dessen Anpassungsgüte an die Realität überprüft werden muss. Dies erfolgt häufig durch einen Abgleich der Computermodellausgaben mit realen Experimenten. Typischerweise sind bei den vorliegenden Computermodellen auch eine Reihe von Einflussgrößen unbekannt, die für die Ausführung der Computersimulation jedoch festzulegen sind. Dies können zum Beispiel unbekannte physikalische Konstanten oder Einstellparameter für die numerischen Lösungsverfahren sein. Diese unbekannt Parameter sollten möglichst so kalibriert werden, dass das Computermodell die Realität gut beschreibt. Die Kalibrierung von Computermodellen beschäftigt sich mit diesem Aspekt und verknüpft die Parameterschätzung mit der Validierung des Computermodells. Ziel ist dabei die Entwicklung eines Vorhersagemodells für den realen Prozess auf Basis von Computer- und realen Experimenten.

In dieser Arbeit werde ich mich sowohl mit der Modellierung von Computermo-

---

dellausgaben als auch mit der Kalibrierung von Computermodellen mit Hilfe von realen Experimenten beschäftigen. Die Modellierung von Computermodellausgaben wird von Sacks *et al.* (1989) erstmals in der statistischen Literatur aufgegriffen und zur Kalibrierung von Computermodellen veröffentlichten Cox *et al.* (2001) und Kennedy und O'Hagan (2001) erste Artikel. Mittlerweile hat sich die Analyse von Computerexperimenten zu einem eigenständigen Gebiet der Statistik entwickelt, in dem insbesondere seit der Jahrtausendwende eine Reihe von Artikeln publiziert wurden.

Für die Modellierung von Computermodellausgaben ist die Gauß Prozess Interpolation ein verbreitetes statistisches Verfahren. Sie löst das Approximationsproblem durch eine stochastische Formulierung und liefert neben einer interpolierenden Schätzfunktion auch Informationen über die Unsicherheit in jedem Punkt. Die Gauß Prozess Interpolation wurde zunächst von Sacks *et al.* (1989) und Currin *et al.* (1991) auf die Analyse von Computerexperimenten angewendet. Einen umfassenden Überblick in die Versuchsplanung und Analyse von Computerexperimenten bieten die Monographien von Fang *et al.* (2006) und Santner *et al.* (2003). Heutzutage liegen zahlreiche Artikel vor in denen die Gauß Prozess Interpolation für verschiedene Probleme bei der Auswertung komplexer Computermodelle eingesetzt werden, zum Beispiel in der Sensitivitätsanalyse von Computermodellen (z.B. Marrel *et al.*, 2009, Oakley, 2009, Schonlau und Welch, 2006) oder der sequentiellen Versuchsplanung von Computerexperimenten (z.B. Jones *et al.*, 1998, Williams *et al.*, 2000).

Ein wesentlicher Vorteil des Gauß Prozess Modells gegenüber anderen Interpolationsansätzen besteht darin, dass die Methode nicht nur eine Vorhersage sondern auch Unsicherheitsintervalle liefert. In der Literatur finden sich bisher jedoch noch keine Untersuchungen zur Überdeckungsgüte der Vorhersageintervalle bei der Gauß Prozess Interpolation. In dieser Arbeit wird dies in Kapitel 3 anhand von Simulationsstudien überprüft. Dabei werden unterschiedliche objektive a priori Verteilungen für Gauß Prozesse (siehe Paulo, 2005) und numerische Verfahren verglichen.

Für die Kalibrierung und Validierung von rechenintensiven Computermodellen werden ebenfalls Gauß Prozesse verwendet (Bayarri *et al.*, 2007b, Cox *et al.*, 2001, Higdon

*et al.*, 2008, Kennedy und O’Hagan, 2001). Cox *et al.* (2001) gehen für die Kalibrierung davon aus, dass das vorliegende Computermodell bei einer geeigneten Wahl der unbekannt Parameter die Realität korrekt beschreiben kann. Dieser Ansatz stellt somit die Schätzung der unbekannt Parameter in den Vordergrund und wird auch Tuning des Computermodells genannt. Dahingegen lässt das Kalibrierungsmodell von Kennedy und O’Hagan (2001) eine systematische Verzerrung des Computermodells zu. Die Erwartungswertfunktion des realen Prozesses wird durch die Summe aus Computermodellfunktion und Verzerrungsfunktion beschrieben. Durch die gemeinsame Berücksichtigung von Computerexperimenten und realen Experimenten soll ein möglichst gutes Vorhersagemodell für den realen Prozess erstellt werden. Für die Modellerstellung liegen neben den Computerexperimenten in der Regel nur wenige teure reale Experimente vor. Bayarri *et al.* (2007b) stellen eine Implementierung des Modells vor und illustrieren wie es für die Validierung von Computermodellen genutzt werden kann. Erweiterungen auf die Kalibrierung hochdimensionaler Computermodelle finden sich in Higdon *et al.* (2008) und auf die Kalibrierung bei funktionalen Zielgrößen in Bayarri *et al.* (2007a) und Bayarri *et al.* (2009).

In der vorliegenden Arbeit (Kapitel 5) wird eine Methode zur Kalibrierung von Computermodellen entwickelt, die sowohl die Situation mit als auch ohne Verzerrung berücksichtigt. Die Methodik erlaubt einerseits zu entscheiden, ob eine Verzerrung des Computermodells vorliegt und andererseits eine Kombination beider Modelle im Rahmen einer Modellmittelung.

In dieser Arbeit motiviert sich die Modellierung und Kalibrierung von Computermodellen aus Problemen bei der Rückfederungssimulation für die Blechumformung (siehe Kapitel 6). Rückfederung tritt auf, wenn bei Nachlassen der Kräfte das Bauteil aus seiner Form zurückfedert. Dadurch kommt es zu Formabweichungen bei der resultierenden Bauteilgeometrie. Ziel ist eine Prozessauslegung mit möglichst geringer Abweichung zur gewünschten Bauteilgeometrie. Dazu soll zunächst ein geeignetes Vorhersagemodell für die Rückfederung anhand von Computer- und realen Experimenten erstellt werden. Im Computermodell sind unbekannt Materialparameter und numerische Einstellparameter zu wählen. Dieses praktische Problem wird in

Kapitel 6 mit den vorgestellten Methoden zur Modellierung und Kalibrierung von Computermodellen ausgewertet.

Der Aufbau dieser Dissertation gliedert sich wie folgt. Die Gauß Prozess Interpolation als Methode zur Modellierung von Computermodellausgaben und Grundlage für die Analyse von Computereperimenten wird in Kapitel 2 vorgestellt. Anschließend wird in Kapitel 3 die Überdeckungsgüte der Vorhersageintervalle bei der Gauß Prozess Interpolation analysiert. Methoden zur Kalibrierung von Computermodellen mit und ohne Verzerrungsfunktion werden in Kapitel 4 beschrieben und untersucht. Anschließend wird in Kapitel 5 ein Kalibrierungsmodell entwickelt und diskutiert, das die Situationen mit und ohne Verzerrung kombiniert. In Kapitel 6 werden Computermodelle für die Rückfederungssimulation bei einem Umformprozess mit den vorgestellten Methoden kalibriert und ausgewertet. Abschließend werden die Ergebnisse dieser Arbeit in Kapitel 7 zusammengefasst und diskutiert.



## 2 GAUSS PROZESS INTERPOLATION

Ein Computermodell wird als deterministische Funktion  $y_M(\cdot) : \mathcal{Z} \rightarrow \mathbb{R}$  aufgefasst, die aus dem Wertebereich  $\mathcal{Z} \subset \mathbb{R}^{d_M}$  in die reellen Zahlen abbildet. Der Funktionswert  $y_M(\mathbf{z})$  kann für beliebiges  $\mathbf{z} \in \mathcal{Z}$  berechnet werden, jedoch ist die Durchführung solch eines Computerexperiments zeitintensiv.

Sei  $\mathbf{y}_M = (y_M(\mathbf{z}_1), \dots, y_M(\mathbf{z}_{n_M}))^T$  der Vektor von Computerexperimenten erhoben an Designpunkten  $\mathbf{z}_i$ , die in der Designmatrix  $\mathbf{D}_M = (\mathbf{z}_1, \dots, \mathbf{z}_{n_M})^T$  zusammen gefasst werden. Ziel bei der statistischen Modellierung von Computerexperimenten ist es eine Approximation und Vorhersage der Funktion  $y_M(\cdot)$  zu finden. Für deterministische Computermodelle sind Vorhersagefunktionen, die die vorliegenden Beobachtungen interpolieren gut geeignet. Neben der Vorhersage ist auch die Vorhersageunsicherheit von Interesse. Die Gauß Prozess Interpolation ist eine statistisch basierte Methode zur Interpolation von Beobachtungen, die diese Unsicherheit quantifizieren kann.

Bei der Gauß Prozess Interpolation wird das Konzept der statistischen Abhängigkeit genutzt, um die Annahme einer stetigen und glatten Funktion  $y_M(\cdot)$  zu modellieren: Bei kleinem Abstand der Einflussgrößen sind Funktionswerte ähnlich und können somit als statistisch abhängig modelliert werden. Dieses Konzept ist auch für stochastische Zielgrößen anwendbar (siehe Kapitel 2.4).

Die Gauß Prozess Interpolation selbst wird schon seit langem in der räumlichen Statistik angewendet. Vorgeschlagen von Daniel Krige (Krige, 1951) und weiter entwickelt von Georges Matheron (Matheron, 1963) ist sie auch unter dem Namen Kriging bekannt (siehe z.B. Cressie, 1993). Das Modell wurde in der Bayes Statistik von O'Hagan (1978) wieder aufgegriffen und hat sich in der Analyse von Computerexperimenten in den letzten Jahren durchgesetzt (siehe z.B. Fang *et al.*, 2006, Sacks *et al.*, 1989, Santner *et al.*, 2003).

Die Gauß Prozess Interpolation lässt sich mit verschiedenen Interpolations- und nicht-parametrischen Regressionsmethoden in Zusammenhang bringen. Fang *et al.* (2006)

illustrieren beispielsweise, dass die funktionale Form der Vorhersage in der Gauß Prozess Interpolation der Form der Vorhersagen bei lokalen Polynommodellen und Radial Basis Funktionen für jeweils eine bestimmte Wahl von Erwartungswert- und Kovarianzfunktionen entspricht. Weiterhin können Interpolationssplines, die durch Minimierung eines Rauheitsstrafers unter der Interpolationsbedingung definiert sind, als Spezialfall einer Gauß Prozess Interpolation für eine bestimmte (nicht stationäre) Kovarianzfunktion dargestellt werden (Gu, 2002, Wahba, 1978). Weitere enge Zusammenhänge bestehen zu Regressionsmethoden des maschinellen Lernens, wie zum Beispiel neuronalen Netzen (Seeger, 2004).

## 2.1 Grundlagen der Gauß Prozess Interpolation

Im Bayes Ansatz zur Modellierung von Computereperimenten wird eine a priori Verteilung  $\tilde{y}_M(\cdot)$  für die Funktion  $y_M(\cdot)$  festgelegt, die alle Informationen vor Beobachtung der Daten beschreiben soll. Anschließend wird die bedingte Verteilung von  $\tilde{y}_M$  gegeben den Beobachtungen  $\mathbf{y}_M$ , die a posteriori oder auch Vorhersageverteilung bestimmt. Sie enthält die Informationen aus a priori Verteilung und Daten.

In der Gauß Prozess Interpolation wird als a priori Verteilung für das Computermodell  $y_M(\cdot)$  ein Gauß Prozess

$$\tilde{y}_M(\cdot) \sim GP(\mu_M(\cdot), C_M(\cdot, \cdot)) \quad (2.1)$$

mit Erwartungswertfunktion  $\mu_M(\cdot)$  und Kovarianzfunktion  $C_M(\cdot, \cdot)$  verwendet. Damit wird die Computermodellfunktion als Realisation eines Gauß Prozesses modelliert.

Definiert ist ein Gauß Prozess  $\tilde{y}_M(\cdot)$  als stochastischer Prozess, für den an einer beliebigen Menge von Punkten  $\mathbf{z}_1, \dots, \mathbf{z}_t$ ,  $t = 1, 2, 3, \dots$  in  $\mathcal{Z}$  der Vektor  $(\tilde{y}_M(\mathbf{z}_1), \dots, \tilde{y}_M(\mathbf{z}_t))^T$  multivariat normalverteilt ist mit Erwartungswert  $\boldsymbol{\mu}_M = (\mu_M(\mathbf{z}_1), \dots, \mu_M(\mathbf{z}_t))^T$  und Kovarianzmatrix  $\boldsymbol{\Sigma} \in \mathbb{R}^{t \times t}$ , wobei  $\Sigma_{(i,j)} = C_M(\mathbf{z}_i, \mathbf{z}_j)$ . Ein Gauß Prozess kann somit als Verallgemeinerung der Normalverteilung auf einen unendlich dimensionalen Raum verstanden werden. Seine Realisationen sind Funktionen auf  $\mathcal{Z}$  anstatt Vektoren bei der

multivariaten Normalverteilung. Die a priori Verteilung aus (2.1) entspricht somit der Annahme, dass die Computermodellfunktion an einer beliebigen Anzahl von Punkten einer multivariaten Normalverteilung folgen.

Für die Wahl von  $\mu_M(\cdot)$  und  $C_M(\cdot, \cdot)$  gibt es eine Vielzahl von Möglichkeiten (siehe z.B. Santner *et al.*, 2003). In der Analyse von Computermodellen wird häufig eine lineare Erwartungswertfunktion  $\mu_M(\cdot)$  und eine stationäre Kovarianzfunktion verwendet. Ein Prozess heißt kovarianzstationär, wenn die Kovarianz zwischen zwei Punkten nur von der Differenz der Punkte abhängt, d.h.  $C_M(\mathbf{z}, \mathbf{z} + \mathbf{h}) = C_M(\mathbf{z}', \mathbf{z}' + \mathbf{h})$  für  $\mathbf{z}, \mathbf{z} + \mathbf{h}, \mathbf{z}', \mathbf{z}' + \mathbf{h} \in \mathcal{Z}$ . Damit lässt sich die Kovarianzfunktion in der Form  $C_M(\mathbf{z}, \mathbf{z}') = \sigma_M^2 R_M(\mathbf{z}, \mathbf{z}')$  mit konstanter Varianz  $\sigma_M^2$  und Korrelationsfunktion  $R_M(\mathbf{z}, \mathbf{z}')$  darstellen. Die Korrelationsfunktion  $R_M(\mathbf{z}, \mathbf{z}')$  hängt wieder nur von der Differenz  $\mathbf{z} - \mathbf{z}'$  ab. Die Kovarianzstationarität ist eine recht starke Annahme, da der zugehörige Gauß Prozess homogen im Raum ist, so dass die Glattheit der Realisationen über  $\mathcal{Z}$  ähnlich sind.

Es werden lineare parametrische Erwartungswertfunktionen der Form

$$\mu_M(\cdot) = f_M(\cdot)^T \boldsymbol{\beta}_M \quad (2.2)$$

mit  $p_M$ -dimensionalen Hyperparameter  $\boldsymbol{\beta}_M$  und fester Regressionsfunktion  $f_M(\cdot)$  betrachtet. Für die Analyse von Computermodellen sind oft verhältnismäßig einfache Regressionsfunktionen wie zum Beispiel  $f_M(\mathbf{z}) = 1$  oder  $f_M(\mathbf{z})^T = (1, z^{(1)}, \dots, z^{(p_M)})$  gut geeignet, so lange keine weiteren Informationen über den Erwartungswert vorliegen.

Die Korrelationsfunktion  $R_M(\cdot, \cdot)$  wird ebenfalls in parametrischer Form gewählt. Häufig wird die so genannte „Power-Exponential“ Korrelationsfunktion

$$R_M(\mathbf{z}, \mathbf{z}') = \exp \left( - \sum_{j=1}^d \psi_{M,j} |z^{(j)} - z'^{(j)}|^{a_{M,j}} \right) \quad (2.3)$$

mit Korrelationsparametervektor  $\boldsymbol{\xi}_M = (\boldsymbol{\psi}_M^T, \mathbf{a}_M^T)^T$  verwendet. Dabei ist  $\boldsymbol{\psi}_M = (\psi_{M,1}, \dots, \psi_{M,d})^T$  und  $\mathbf{a} = (a_{M,1}, \dots, a_{M,d})^T$  mit  $\psi_{M,j} > 0$  und  $0 < a_{M,j} \leq 2$ . Ist  $a_{M,j} = 2$

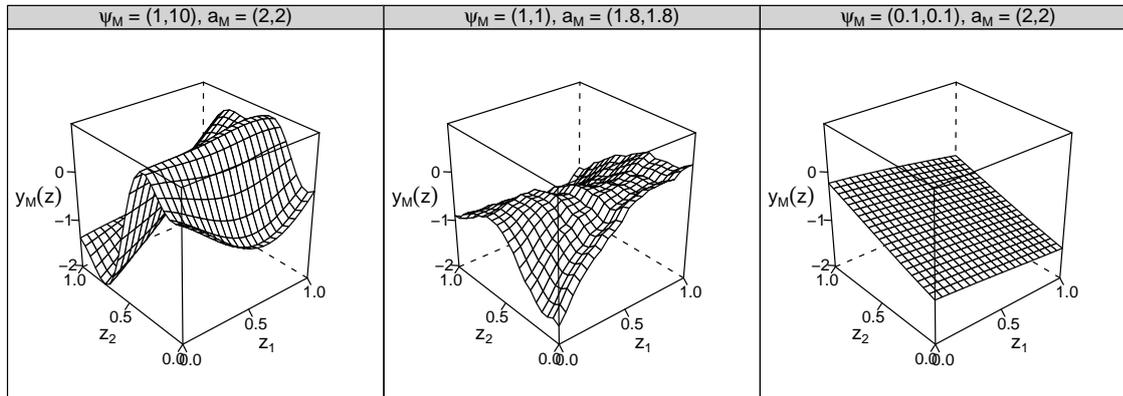


Abbildung 2.1: Gauß Prozess Realisationen mit verschiedenen Korrelationsparametern ( $\mu_M(\mathbf{z}) = 0$  und  $\sigma_M^2 = 1$ )

für alle  $j = 1, \dots, d$  erhalten wir den Spezialfall der Gauß Korrelationsfunktion und wenn  $\psi_{M,j} = \psi_M$  für alle  $j$  gilt, so liegt eine isotropische Korrelationsfunktion vor. Die Korrelation zwischen zwei Punkten ist nach Gleichung (2.3) im multivariaten Sinne monoton fallend in der Differenz der Punkte. Damit wird angenommen, dass die Ähnlichkeit der Computermodellausgaben mit fallendem Abstand der Punkte in  $\mathcal{Z}$  steigt. Für die Power-Exponential Korrelationsfunktion kann gezeigt werden, dass die Realisationen (fast sicher) stetig sind. Ist  $a_{M,j} = 2$ , für alle  $j = 1, \dots, d$ , (Gauß Korrelationsfunktion) sind die Realisationen zusätzlich unendlich oft stetig differenzierbar, für  $a_{M,j} < 2$  jedoch nicht differenzierbar.

Gauß Prozesse können sehr flexibel für die Modellierung von Funktionen eingesetzt werden. Die Vielfalt der Gestalt der Realisationen wird in Abbildung 2.1 illustriert. Dort werden drei zwei dimensionale Realisationen mit Erwartungswert 0 und Power-Exponential Korrelationsfunktion mit variierendem Korrelationsparameter dargestellt. Der Parameter  $\psi_M$  kontrolliert wie stark die Funktion oszilliert, die Anzahl der Maxima und Minima steigt mit steigendem  $\psi_M$ . Der Exponent Parameter  $a_M$  kontrolliert die Glattheit der Funktion im Sinne der Differenzierbarkeit: Werte kleiner 2 liefern nicht differenzierbare Funktionen. Eine alternative Korrelationsfunktion, bei der auch der Grad der Differenzierbarkeit kontrolliert werden kann, ist die Matérn Korrelationsfunktion (siehe z.B. Diggle und Ribeiro, 2007, S.51). Auf diese wird hier aber nicht

weiter eingegangen.

Zusammenfassend folgt der Vektor  $\tilde{\mathbf{y}}_M = \tilde{y}(\mathbf{D}_M) = (\tilde{y}_M(\mathbf{z}_1), \dots, \tilde{y}_M(\mathbf{z}_{n_M}))^T$  a priori im Gauß Prozess Modell damit der multivariaten Normalverteilung

$$\tilde{\mathbf{y}}_M \sim N(\mathbf{F}_M \boldsymbol{\beta}_M, \sigma_M^2 \mathbf{R}_M) \quad (2.4)$$

mit Regressionsmatrix  $\mathbf{F}_M = (f_M^T(\mathbf{z}_1), \dots, f_M^T(\mathbf{z}_{n_M}))^T$  und Korrelationsmatrix  $\mathbf{R}_M$  mit Einträgen  $R_{M,i,i'} = R_M(\mathbf{z}_i, \mathbf{z}_{i'})$  für  $i, i' \in 1, \dots, n_M$  mit Parametern  $\boldsymbol{\beta}_M, \sigma_M^2$  und  $\boldsymbol{\xi}_M$ .

Das Gauß Prozess Modell ähnelt auf dem ersten Blick einem linearen Regressionsmodell. Bei beiden wird eine Normalverteilungsannahme und eine lineare Regressionsfunktion verwendet. Durch die Kovarianzstruktur des Fehlerterms unterscheidet sich die im Folgenden vorgestellte resultierende Vorhersage beim Gauß Prozess Modell jedoch deutlich.

## 2.2 Vorhersagen bei der Gauß Prozess Interpolation

Die Vorhersage- bzw. prädiktive Verteilung einer Größe  $y_{\text{neu}}$  ist in der Bayes Statistik durch die a posteriori Verteilung  $p(y_{\text{neu}}|\mathbf{y})$  gegeben, der bedingten Verteilung von  $y_{\text{neu}}$  gegeben einen beobachteten Datenvektor  $\mathbf{y}$ . Ist diese bedingte Verteilung nur in Abhängigkeit von Parametern  $\boldsymbol{\theta}$  bekannt, wird zusätzlich eine a priori Verteilung  $p(\boldsymbol{\theta})$  angenommen. Dann lässt sich  $p(y_{\text{neu}}|\mathbf{y})$  durch herausintegrieren der Parameter bezüglich ihrer a posteriori Verteilung  $p(\boldsymbol{\theta}|\mathbf{y})$  bestimmen:  $p(y_{\text{neu}}|\mathbf{y}) = \int p(y_{\text{neu}}, \boldsymbol{\theta}|\mathbf{y})p(\boldsymbol{\theta}|\mathbf{y})d\boldsymbol{\theta}$ . Als Vorhersage wird der Erwartungswert der Vorhersageverteilung  $p(y_{\text{neu}}|\mathbf{y})$  genutzt und Grenzen für ein  $1 - \alpha$  Vorhersageintervall resultieren aus den  $\alpha/2$  und  $1 - \alpha/2$  Quantil. In der Gauß Prozess Interpolation ist die Vorhersageverteilung von  $\tilde{y}_M(\mathbf{x})$  somit durch die bedingte Verteilung  $p(\tilde{y}_M(\mathbf{x})|\mathbf{y}_M)$  gegeben. Um diese Verteilung zu bestimmen kann folgendes Resultat über die multivariate Normalverteilung verwendet werden.

Seien  $X_1$  und  $X_2$  zwei gemeinsam normalverteilte Zufallsvektoren

$$\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim N \left( \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \Sigma_1 & \Sigma_{1,2} \\ \Sigma_{1,2}^T & \Sigma_2 \end{pmatrix} \right),$$

so ist die bedingte Verteilung von  $X_1|X_2$  gegeben durch

$$X_1|X_2 \sim N(\mu_1 - \Sigma_{1,2}^T \Sigma_2^{-1}(X_2 - \mu_2), \Sigma_1 - \Sigma_{1,2}^T \Sigma_2^{-1} \Sigma_{1,2}), \quad (2.5)$$

siehe z.B. O'Hagan und Forster (2004).

### *Bekannte Parameter*

Im Fall der Gauß Prozess Interpolation ist die Vorhersageverteilung unter der Annahme, dass alle Parameter  $\boldsymbol{\theta}_M = (\boldsymbol{\beta}_M^T, \sigma_M^2, \boldsymbol{\xi}_M^T)^T$  bekannt sind, also wieder gegeben durch einen Gauß Prozess

$$\tilde{y}_M(\cdot) | \mathbf{y}_M, \boldsymbol{\theta}_M \sim GP(y_M^*(\cdot), \Sigma_M^*(\cdot, \cdot)), \quad (2.6)$$

mit Erwartungswertfunktion

$$y_M^*(\mathbf{z}) = f_M(\mathbf{z})^T \boldsymbol{\beta}_M - \mathbf{r}_M^T(\mathbf{z}) \mathbf{R}_M^{-1}(\mathbf{y}_M - \mathbf{F}_M \boldsymbol{\beta}_M) \quad (2.7)$$

und Kovarianzfunktion

$$\Sigma_M^*(\mathbf{z}, \mathbf{z}') = \sigma_M^2 \cdot [R_M(\mathbf{z}, \mathbf{z}') - \mathbf{r}_M^T(\mathbf{z}) \mathbf{R}_M^{-1} \mathbf{r}_M(\mathbf{z}')] \quad (2.8)$$

Dabei ist  $\mathbf{r}_M^T(\cdot) = (R_M(\cdot, \mathbf{z}_{M,1}), \dots, R_M(\cdot, \mathbf{z}_{M,n_M}))$  der Vektor der Korrelationen zwischen Vorhersagepunkt und Designpunkten. Erwartungswert- und Kovarianzfunktion ergeben sich durch Anwendung von (2.5) für jeden Punkt aus  $\mathcal{Z}$ .

Bei der Gauß Prozess Interpolation gehen in die Vorhersage, das heißt in den Erwartungswert von (2.6), neben der Regressionskomponente  $f_M(\cdot)^T \boldsymbol{\beta}_M$  auch die bereits erhobenen Beobachtungen gewichtet durch die Korrelationsstruktur ein. Bei großem

Abstand zu den Designpunkten und damit schwacher Korrelation überwiegt die Regressionskomponente und bei kleinem Abstand und starker Korrelation wird die Vorhersage in Richtung der Beobachtungen gezogen. Die a posteriori Korrelation besteht aus der a priori Korrelation abzüglich der Korrelation zwischen Vorhersagepunkt und Beobachtungen.

### *Bekannte Korrelationsstruktur*

Im nächsten Schritt werden die Parameter  $\beta_M$  und  $\sigma_M^2$  als unbekannt und nur noch der Korrelationsparameter  $\xi_M$  als bekannt betrachtet. In einem Bayes Ansatz sind für die unbekannt Hyperparameter a priori Verteilungen zu wählen. Da bei der Modellierung von Computereperimenten in der Regel keine a priori Informationen vorliegen, wird eine so genannte nicht-informative a priori Verteilung gewählt. Hier wird  $p(\beta_M, \sigma_M^2) \propto \frac{1}{\sigma_M^2}$ , die Jeffreys a priori Verteilung in diesem Modell (siehe z.B. Paulo, 2005) eingesetzt. Die Schreibweise „ $\propto$ “ bedeutet dabei, dass die a priori Verteilung für beliebige  $\beta_M$  und  $\sigma_M^2$  proportional zu  $\frac{1}{\sigma_M^2}$  ist. Es handelt sich nicht um eine echte Wahrscheinlichkeitsverteilung, da das Integral über die Verteilung nicht 1 sondern unendlich ist. Solche a priori Verteilungen werden auch unechte oder diffuse a priori Verteilungen genannt. Nicht immer resultieren bei unechten a priori Verteilungen echte a posteriori Verteilungen. Bei bekannter Korrelationsstruktur ist dies hier jedoch der Fall (Berger *et al.*, 2001).

Die Vorhersageverteilung ist nun nicht länger durch einen Gauß Prozess sondern einen so genannten Student Prozess

$$\tilde{y}_M(\cdot) | \mathbf{y}_M, \xi_M \sim t_{n_M - p_M} \left( \hat{y}_M(\cdot), \hat{\Sigma}_M^*(\cdot, \cdot) \right) \quad (2.9)$$

mit Erwartungswertfunktion  $\hat{y}_M(\cdot)$ , Skalenfunktion  $\hat{\Sigma}_M^*(\cdot, \cdot)$  und  $n_M - p_M$  Freiheitsgraden gegeben. Der Student Prozess ist eine Verallgemeinerung der multivariaten t-Verteilung bei dem die Zielgröße für eine beliebige Menge an Einflussgrößen multivariat t-verteilt ist (Santner *et al.*, 2003, S.212). Der a posteriori Erwartungswert  $\hat{y}_M(\cdot)$

entspricht dem Erwartungswert aus (2.6), jedoch indem  $\boldsymbol{\beta}_M$  durch

$$\widehat{\boldsymbol{\beta}}_M = (\mathbf{F}_M^T \mathbf{R}_M^{-1} \mathbf{F}_M)^{-1} \mathbf{F}_M^T \mathbf{R}_M^{-1} \mathbf{y}_M \quad (2.10)$$

ersetzt wird.  $\widehat{\boldsymbol{\beta}}_M$  entspricht dem verallgemeinerten kleinste Quadrate Schätzer im linearen Modell. Die Skalenfunktion ist gegeben durch  $\widehat{\Sigma}_M^*(\cdot, \cdot) = \widehat{\sigma}_M^2 c_M^*(\cdot, \cdot)$  mit  $\widehat{\sigma}_M^2 = \frac{1}{n_M - p_M} (\mathbf{y}_M - \mathbf{F}_M \widehat{\boldsymbol{\beta}}_M)^T \mathbf{R}_M^{-1} (\mathbf{y}_M - \mathbf{F}_M \widehat{\boldsymbol{\beta}}_M)$  und a posteriori Korrelation  $c_M^*(\mathbf{z}, \mathbf{z}')$  zwischen beliebigen  $\mathbf{z}, \mathbf{z}'$

$$\begin{aligned} c_M^*(\mathbf{z}, \mathbf{z}') &= R_M(\mathbf{z}, \mathbf{z}') - \mathbf{r}_M^T(\mathbf{z}) \mathbf{R}_M^{-1} \mathbf{r}_M(\mathbf{z}') \\ &+ (\mathbf{f}_M^T(\mathbf{z}) - \mathbf{r}_M^T(\mathbf{z}) \mathbf{R}_M^{-1} \mathbf{F}_M) (\mathbf{F}_M^T \mathbf{R}_M^{-1} \mathbf{F}_M)^{-1} (\mathbf{f}_M^T(\mathbf{z}') - \mathbf{r}_M^T(\mathbf{z}') \mathbf{R}_M^{-1} \mathbf{F}_M)^T. \end{aligned} \quad (2.11)$$

Im Vergleich zur Korrelation bei bekannten Parametern in (2.6) wird der dritte Term in (2.11) addiert, der die Varianz aus der Schätzung von  $\boldsymbol{\beta}$  berücksichtigt.

Ein punktweises  $(1 - \alpha) \cdot 100\%$  Vorhersageintervall für  $\tilde{y}_M(\mathbf{z})$  ist gegeben durch

$$\left[ \hat{y}_M(\mathbf{z}) - t_{n_M - p_M, 1 - \alpha/2} \sqrt{\widehat{\Sigma}_M^*(\mathbf{z}, \mathbf{z})}, \hat{y}_M(\mathbf{z}) + t_{n_M - p_M, 1 - \alpha/2} \sqrt{\widehat{\Sigma}_M^*(\mathbf{z}, \mathbf{z})} \right], \quad (2.12)$$

mit  $t_{\nu, 1 - \alpha/2}$  dem  $1 - \alpha/2$  Quantil der t-Verteilung mit  $\nu$  Freiheitsgraden.

### Unbekannte Korrelationsstruktur

Die Parameter der Korrelationsfunktion  $\boldsymbol{\xi}_M$  bestimmen wesentlich die Glattheit des a posteriori Prozesses (siehe Abbildung 2.1). Im Folgenden wird auch  $\boldsymbol{\xi}_M$  als unbekannt betrachtet und a priori angenommen, dass  $\boldsymbol{\xi}_M$  unabhängig von  $\boldsymbol{\beta}_M$  und  $\sigma_M^2$  ist und Dichte  $p(\boldsymbol{\xi}_M)$  besitzt. Die a posteriori für  $\boldsymbol{\psi}_M$  erhalten wir durch analytisches herausintegrieren der übrigen Parameter aus der gemeinsamen a posteriori Verteilung. Sie ist proportional zu

$$p(\boldsymbol{\xi}_M | \mathbf{y}_M^n) \propto p(\boldsymbol{\xi}_M) \cdot \det(\mathbf{R}_M)^{-1/2} \det(\mathbf{F}_M^T \mathbf{R}_M^{-1} \mathbf{F}_M)^{-1/2} \widehat{\sigma}_M^{-(n_M - p_M)/2} \quad (2.13)$$

(siehe Handcock und Stein, 1993). Die Vorhersageverteilung für  $\tilde{y}_M$  ist damit formal gegeben durch

$$\int p(\tilde{y}_M(\cdot)|\mathbf{y}_M^n, \boldsymbol{\xi}_M)p(\boldsymbol{\xi}_M|\mathbf{y}_M^n)d\boldsymbol{\xi}_M. \quad (2.14)$$

Dieses Integral kann nicht analytisch sondern nur numerisch durch eine Monte Carlo Integration bestimmt werden. Dazu sind numerische Methoden zur Generierung von Zufallszahlen wie zum Beispiel Markov Chain Monte Carlo (MCMC) Methoden erforderlich (siehe Anhang A.1 für eine kurze Einführung). Es wird eine große Anzahl von Zufallszahlen  $\boldsymbol{\xi}_M^{(1)}, \dots, \boldsymbol{\xi}_M^{(T)}$  aus der a posteriori (2.13) von  $\boldsymbol{\xi}_M$  gezogen, um die Vorhersageverteilung (2.14) durch das arithmetische Mittel  $1/T \sum_{t=1}^T p(\tilde{y}_M(\cdot)|\mathbf{y}_M^n, \boldsymbol{\xi}_M^{(t)})$  zu approximieren. Alternativ kann (2.14) durch Einsetzen des a posteriori Modalwerts  $\boldsymbol{\xi}_M^*$  in die Vorhersageverteilung (2.9) bei bekannter Korrelationsstruktur,  $p(\tilde{y}_M(\cdot)|\mathbf{y}_M^n, \boldsymbol{\xi}_M^*)$ , approximiert werden. Der a posteriori Modalwert kann durch Maximieren von (2.13) bestimmt werden. Dieses ist aktuell das Standardvorgehen in der Analyse von Computereperimenten. Dabei wird die Unsicherheit gegenüber dem Korrelationsparameter vernachlässigt und somit die a posteriori Varianz etwas unterschätzt. Jedoch ist der Optimierungsansatz deutlich weniger rechenintensiv als die numerische Integration. Dies macht ihn für die Analyse von Computereperimenten attraktiv, wo eine schnell berechenbare Approximation zu bestimmen ist. In Kapitel 3 werden die Unterschiede zwischen diesen beiden rechentechnischen Ansätzen und ihr Einfluss auf die Schätzung von Unsicherheitsintervallen in einer Simulationsstudie genauer untersucht.

Eine verbreitete a priori Verteilung für  $\boldsymbol{\xi}_M$  ist die flache Verteilung  $p(\boldsymbol{\xi}_M) \propto 1$  auf dem Wertebereich von  $\boldsymbol{\xi}_M$ . In Kapitel 3 werden bei Verwendung einer Gauß Korrelationsfunktion verschiedene nichtinformativ a priori Verteilungen für den Korrelationsparameter untersucht und anhand von Simulationsstudien und Testfunktionen verglichen.

Es wird nun gezeigt, dass die Vorhersagefunktion  $\hat{y}_M(\cdot)$  die Beobachtungen interpoliert und Varianz der Vorhersageverteilung in den Designpunkten  $\mathbf{z}_{M,i}$  Null ist (siehe auch

Santner *et al.*, 2003, S.57). Im folgenden wird die Vorhersageverteilung bei bekannter Korrelation aus (2.9) betrachtet. Die Argumentationskette lässt sich jedoch auch auf die Vorhersage bei bekannten Parametern (2.6) und unbekannter Korrelation übertragen. Betrachte den Designpunkt  $\mathbf{z}_{M,i}$  für  $i \in \{1, \dots, n_M\}$ . Es gilt

$$\mathbf{r}_M^T(\mathbf{z}_{M,i})\mathbf{R}_M^{-1} = \mathbf{e}_i^T, \quad (2.15)$$

wobei  $\mathbf{e}_i$  der  $i$ -te Einheitsvektor ist. Damit ist  $\hat{y}_M(\mathbf{z}_{M,i}) = \mathbf{f}_M^T(\mathbf{z}_{M,i})\hat{\boldsymbol{\beta}}_M - \mathbf{e}_i^T(\mathbf{y}_M - \mathbf{F}_M\hat{\boldsymbol{\beta}}_M)$  und wegen  $\mathbf{e}_i^T\mathbf{y}_M = y_M(\mathbf{z}_{M,i})$  und  $\mathbf{e}_i^T\mathbf{F}_M\hat{\boldsymbol{\beta}}_M = \mathbf{f}_M^T(\mathbf{z}_{M,i})\hat{\boldsymbol{\beta}}_M$  ist  $\hat{y}_M(\mathbf{z}_{M,i}) = y_M(\mathbf{z}_{M,i})$ .

Die Varianz der Vorhersageverteilung in  $\mathbf{z}_{M,i}$  ist proportional zu  $\hat{\Sigma}_M^*(\mathbf{z}_{M,i}) = \hat{\sigma}_M^2 c_M^*(\mathbf{z}_{M,i}, \mathbf{z}_{M,i})$ . Mit (2.15) gilt

$$c_M^*(\mathbf{z}_{M,i}, \mathbf{z}_{M,i}) = 1 - \mathbf{e}_i^T \mathbf{r}_M(\mathbf{z}_{M,i}) + (\mathbf{f}_M^T(\mathbf{z}_{M,i}) - \mathbf{e}_i^T \mathbf{F}_M) (\mathbf{F}_M^T \mathbf{R}_M^{-1} \mathbf{F}_M)^{-1} (\mathbf{f}_M^T(\mathbf{z}_{M,i}) - \mathbf{e}_i^T \mathbf{F}_M)^T$$

und da  $\mathbf{e}_i^T \mathbf{r}_M(\mathbf{z}_{M,i}) = R_M(\mathbf{z}_{M,i}, \mathbf{z}_{M,i})$  sowie  $\mathbf{e}_i^T \mathbf{F}_M = \mathbf{f}_M^T(\mathbf{z}_{M,i})$  ist  $c_M^*(\mathbf{z}_{M,i}, \mathbf{z}_{M,i})$  und damit die Varianz gleich Null.

Für deterministische Computermodelle ist der Funktionswert an den Designpunkten bekannt und somit besteht hier keine Unsicherheit mehr. Interpolation und eine Varianz Null an Designpunkten machen die Gauß Prozess Interpolation deshalb für die Analyse deterministischer Computerexperimente besonders geeignet. Die Varianzfunktion  $\hat{\Sigma}_M^*(\cdot)$  eignet sich zum Beispiel als Designkriterium für die Erhebung weiterer Beobachtungen (Jones *et al.*, 1998, Ranjan und Bingham, 2008, Williams *et al.*, 2000).

### Beispiel

Die Gauß Prozess Interpolation soll nun anhand eines einfachen Beispiels illustriert werden. Die eindimensionale Funktion  $y_M(z) = (z - 5)/10 \sin(-z) + 1$  soll auf dem Intervall  $[0, 10]$  mit 6 äquidistant erhobenen Beobachtungen approximiert werden. Abbildung 2.2 links zeigt die Vorhersage die sich durch  $\hat{y}_M(\cdot)$  bei einer Gauß Prozess

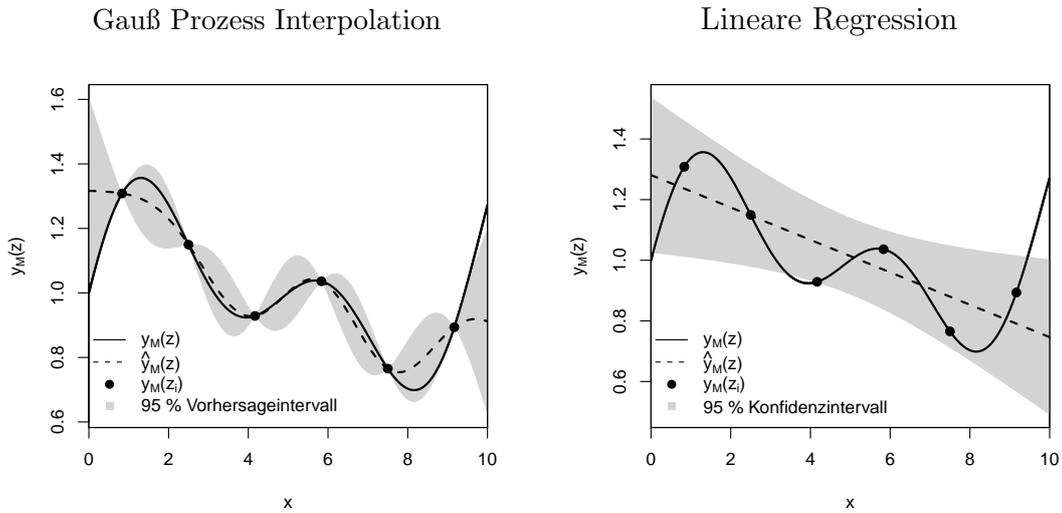


Abbildung 2.2: Approximation der Funktion  $y_M(z) = (z - 5)/10 \sin(-z) + 1$

Interpolation mit linearer Erwartungswertfunktion  $f_M(z) = (1, z)^T$  und Gauß Korrelationsfunktion ergibt. Die punktwisen 95% Vorhersageintervalle sind grau hinterlegt. Für den Korrelationsparameter wurde der a posteriori Modalwert eingesetzt. Zum Vergleich in Abbildung 2.2 rechts die Vorhersage, die sich durch eine einfache lineare Regression – also alleine durch die Erwartungswertfunktion – ergibt. Punktweise 95% Konfidenzintervalle für den Erwartungswert sind auch hier grau hinterlegt.

Die Abbildungen illustrieren deutlich, dass die Korrelationsstruktur die Gestalt der Vorhersagefunktion in der Gauß Prozess Interpolation wesentlich beeinflusst. Die lineare Regression liefert nur eine sehr grobe Trendanpassung und interpoliert die Daten nicht. Dagegen werden die Beobachtungen bei Verwendung eines Gauß Prozesses interpoliert und die vorliegende deutlich variierende Funktion kann sehr gut approximiert werden. Lediglich an den Rändern des betrachteten Intervalls weichen die Vorhersagen deutlich von der zu Grunde liegenden Funktion ab. Eine Extrapolation über den Bereich der beobachteten Daten hinaus ist also nicht möglich und damit ist die Gauß Prozess Interpolation eine Methode zur lokalen Approximation.

Neben den Vorhersagen unterscheiden sich auch die Unsicherheitsintervalle zwischen den beiden Methoden deutlich. Bei der Gauß Prozess Interpolation ist die Unsicherheit in der Nähe der Beobachtungen klein und steigt mit dem Abstand zu ihnen. Maximale

Unsicherheit besteht hier am Rand des Wertebereichs. Dagegen sind die Konfidenzintervalle für den Erwartungswert bei der linearen Regression weniger stark abhängig von den Beobachtungen, da Abweichungen von der gewählten Regressionsfunktion als Zufallsfehler interpretiert werden.

### 2.3 Maximum Likelihood Ansatz für die Gauß Prozess Interpolation

Die Gauß Prozess Interpolation lässt sich auch durch einen Likelihood Ansatz motivieren. Dabei wird die Computermodellfunktion  $y_M(\cdot)$  als Realisation eines Gauß Prozesses modelliert. Erwartungswertfunktion und Kovarianzfunktionen werden analog zum Bayes Ansatz gewählt. An den Beobachtungen  $\mathbf{y}_M$  liegt die Modellannahme  $\mathbf{y}_M \sim N(\mathbf{F}_M \boldsymbol{\beta}_M, \sigma_M^2 \mathbf{R}_M)$  vor. Entsprechend ist die Likelihood gegeben durch

$$L(\mathbf{y}_M, \boldsymbol{\beta}_M, \sigma_M^2, \boldsymbol{\xi}_M) = (2\pi)^{p_M/2} (\sigma_M^2)^{-n_M/2} (\det \mathbf{R}_M)^{-1/2} \exp \left\{ -\frac{1}{2\sigma_M^2} (\mathbf{y}_M - \mathbf{F}_M \boldsymbol{\beta}_M)^T \mathbf{R}_M^{-1} (\mathbf{y}_M - \mathbf{F}_M \boldsymbol{\beta}_M) \right\}. \quad (2.16)$$

Maximieren der Likelihood (2.16) liefert die ML Schätzer  $\hat{\boldsymbol{\beta}}_M$  aus (2.10) für  $\boldsymbol{\beta}_M$  und

$$\hat{\sigma}_{M,ML}^2 = \frac{1}{n_M} \left( \mathbf{y}_M - \mathbf{F}_M \hat{\boldsymbol{\beta}}_M \right)^T \mathbf{R}_M^{-1} \left( \mathbf{y}_M - \mathbf{F}_M \hat{\boldsymbol{\beta}}_M \right) \quad (2.17)$$

für  $\sigma_M^2$ . Durch Einsetzen von  $\hat{\boldsymbol{\beta}}_M$  und  $\hat{\sigma}_{M,ML}^2$  in die Log-Likelihood ergibt sich

$$l(\boldsymbol{\xi}_M) = -\frac{1}{2} \left[ n_M \log(2\pi) + n_M \log(\hat{\sigma}_{M,ML}^2(\boldsymbol{\xi}_M)) + \log(\det(\mathbf{R}_M(\boldsymbol{\xi}_M))) + n_M \right]$$

und damit der ML-Schätzer  $\hat{\boldsymbol{\xi}}_{M,ML}$  für  $\boldsymbol{\xi}_M$  durch Minimieren von

$$n_M \log(\hat{\sigma}_{M,ML}^2(\boldsymbol{\xi}_M)) + \log(\det(\mathbf{R}_M(\boldsymbol{\xi}_M)))$$

über den Definitionsbereich von  $\boldsymbol{\xi}_M$ .

Die ML-Schätzer für  $\sigma_M^2$  und  $\boldsymbol{\xi}_M$  sind nicht erwartungstreu. Weniger verzerrte Schätzungen können durch die so genannte restringierte Maximum-Likelihood (REML) Me-

thode (Patterson und Thompson, 1974). Dieser Ansatz wird insbesondere bei gemischten linearen Modellen angewendet. Dabei wird die Zufallsvariable der Beobachtungen so transformiert, dass ihre Verteilung nicht länger von  $\boldsymbol{\beta}_M$  abhängt. Dadurch reduziert sich die effektive Dimension der Beobachtungen auf  $n_M - p_M$ , wobei  $p_M$  der Dimension von  $\boldsymbol{\beta}_M$  entspricht. Die logarithmierte restringierte Log-Likelihood für  $\mathbf{y}_M$  lautet

$$l_r(\boldsymbol{\beta}_M, \sigma_M^2, \boldsymbol{\xi}_M) = \frac{1}{2} \left[ (n_M - p_M) \ln(\sigma_M^2) - \ln \det(\mathbf{R}(\boldsymbol{\xi}_M)) - \ln \det(\mathbf{F}_M^T \mathbf{R}_M(\boldsymbol{\xi}_M)^{-1} \mathbf{F}_M) - \frac{1}{\sigma_M^2} (\mathbf{y}_M^n - \mathbf{F}_M \boldsymbol{\beta}_M)^T \mathbf{R}_M^{-1} (\mathbf{y}_M^n - \mathbf{F}_M \boldsymbol{\beta}_M) \right]$$

(siehe Diggle und Ribeiro, 2007, S.17).

Die Schätzung für  $\boldsymbol{\beta}_M$  entspricht dem ML-Schätzer. Der REML Schätzer für  $\sigma_M^2$

$$\hat{\sigma}_{\text{REML}}^2 = \frac{1}{n_M - p_M} \left( \mathbf{y}_M - \mathbf{F}_M \hat{\boldsymbol{\beta}}_M \right)^T \mathbf{R}_M^{-1} \left( \mathbf{y}_M - \mathbf{F}_M \hat{\boldsymbol{\beta}}_M \right) \quad (2.18)$$

ist im Gegensatz zum ML Schätzer bei gegebenen  $\boldsymbol{\xi}_M$  erwartungstreu und der Schätzer  $\hat{\boldsymbol{\xi}}_{M,\text{REML}}$  für den Korrelationsparameter ergibt sich durch Minimieren des Ausdrucks

$$(n_M - p_M) \ln(\hat{\sigma}_{M,\text{REML}}^2) + \ln(\det(\mathbf{R}_M(\boldsymbol{\xi}_M))) + \ln(\det(\mathbf{F}_M^T \mathbf{R}_M^{-1}(\boldsymbol{\xi}_M) \mathbf{F}_M)) \quad (2.19)$$

über den Wertebereich von  $\boldsymbol{\xi}_M$ .

Bei bekannter Korrelationsstruktur ist die lineare unverzerrte Vorhersage mit kleinster Varianz im Likelihood Ansatz gleich  $\hat{y}_M(\cdot)$  aus (2.9). Sie besitzt Varianz  $\sigma_M^2 c^*(\mathbf{z}, \mathbf{z})$ . Der REML Schätzer  $\hat{\sigma}_{\text{REML}}^2$  entspricht genau  $\hat{\sigma}_M^2$  aus (2.9) und  $\hat{\boldsymbol{\xi}}_{M,\text{REML}}$  dem a posteriori Modalwert für  $\boldsymbol{\xi}_M$  aus dem Bayes Ansatz. Damit stimmt dieser Ansatz im wesentlichen mit dem Bayes Ansatz überein.

Maximum Likelihood Schätzer werden bei der Gauß Prozess Interpolation in einem Bayes Ansatz häufig in die Vorhersageverteilung bei bekannten Parametern aus (2.6) eingesetzt (siehe z.B. Bayarri *et al.*, 2007b). Dies liefert die Erwartungswertfunktion  $\hat{y}_M(\cdot)$  und Kovarianzfunktion  $\hat{\Sigma}_M(\mathbf{z}, \mathbf{z}') = \hat{\sigma}_{M,\text{ML}}^2 (R_M(\mathbf{z}, \mathbf{z}') - r_M^T(\mathbf{z}) \mathbf{R}_M^{-1} r_M^T(\mathbf{z}'))$ . Dadurch kann der numerische Aufwand für die Gauß Prozess Interpolation im Vergleich zu einem vollständigen Bayes Ansatz deutlich reduziert werden.

## 2.4 Gauß Prozess Regression

Gauß Prozesse können auch für die Modellierung stochastischer Zielgrößen  $Y_E(\mathbf{x})$  eingesetzt werden. Es wird das allgemeine Regressionsmodell  $Y_E(\mathbf{x}) = \mu_E(\mathbf{x}) + e$  mit  $e \sim N(0, \sigma_E^2)$  betrachtet. Als a priori Annahme für  $\mu_E(\mathbf{x})$  wird ein Gauß Prozess wie in (2.1) gewählt. Es gilt

$$\tilde{y}_E(\cdot) \sim GP(f_E(\cdot)^T \boldsymbol{\beta}_E, \tau^2 R_E(\cdot, \cdot) + \sigma_E^2). \quad (2.20)$$

mit Regressionsfunktion  $f_E(\cdot)$ , Regressionsparameter  $\boldsymbol{\beta}_E$  sowie Korrelationsfunktion  $R_E(\cdot, \cdot)$  mit Korrelationsparameter  $\boldsymbol{\xi}_E$  und zugehöriger Prozessvarianz  $\tau^2$ . In der Kovarianz beschreibt  $\sigma_E^2$  die experimentelle Streuung und  $\tau^2 R_E(\cdot, \cdot)$  die Unsicherheit über die zu Grunde liegende Funktion  $\mu_E(\cdot)$ . In der Literatur der Geostatistik wird das Modell aus (2.20) auch als Kriging Modell mit Nugget Effekt bezeichnet.

Sei  $\mathbf{y}_E$  der Datenvektor korrespondierend zu der Designmatrix  $\mathbf{D}_E = (\mathbf{z}_1, \dots, \mathbf{z}_{n_E})^T$ . Die Likelihood ist hier gegeben durch die Dichte der Normalverteilung  $\mathbf{y}_E \sim N(\mathbf{F}_E \boldsymbol{\beta}_E, \tau^2 \mathbf{R}_E + \sigma_E^2 \mathbf{I})$ . Für die Herleitung der Vorhersageverteilung  $p(y_E(\cdot) | \mathbf{y}_E)$  hilft die Reparametrisierung  $\tau^2 \mathbf{R}_E + \sigma_E^2 \mathbf{I} = \tau^2 (\mathbf{R}_E + \nu^2 \mathbf{I})$  mit  $\nu^2 = \frac{\sigma_E^2}{\tau^2}$ . Bei bekannten Parametern ist die Vorhersageverteilung analog zu dem Gauß Prozess in (2.6), wobei  $\mathbf{R}_M$  durch  $\mathbf{R}_E + \nu^2 \mathbf{I}$  zu ersetzen und die Notation anzupassen ist. Bei bekannten  $\nu^2$  und  $\boldsymbol{\xi}_E$  und a priori  $p(\boldsymbol{\beta}_E, \tau^2) \propto \frac{1}{\tau^2}$  ist die Vorhersageverteilung wieder ein Student Prozess wie in (2.9). Auch hier sind in Erwartungswert- und Skalenfunktion jeweils  $\mathbf{R}_M$  durch  $\mathbf{R}_E + \nu^2 \mathbf{I}$  sowie  $\widehat{\sigma}_M^2$  durch  $\widehat{\tau}^2$  zu ersetzen und die Notation anzupassen. Bei unbekanntem  $\boldsymbol{\xi}_E$  und  $\nu^2$  ist eine geschlossene Darstellung der Vorhersageverteilung nicht möglich, sondern muss durch numerische Integration wie in (2.14) bestimmt werden.

Alternativ ist auch hier das Einsetzen des a posteriori Modalwerts oder eines ML-Schätzers verbreitet. Der ML Schätzer  $\widehat{\tau}^2$  entspricht dem Schätzer aus (2.17). Die ML-Schätzer für  $\nu^2$  und  $\boldsymbol{\xi}_E$  ergeben sich durch Minimieren des Ausdrucks

$$n \log(\widehat{\tau}^2) + \log \det(\mathbf{R}_E + \nu^2 \mathbf{I}), \quad (2.21)$$

(siehe Diggle und Ribeiro, 2007, S. 113).

Die Grundstruktur der Vorhersageverteilung unterscheidet sich im Vergleich zur Gauß Prozess Interpolation nicht wesentlich. Allerdings werden die Beobachtungen nicht länger interpoliert und die Varianz an Beobachtungspunkten ist größer Null. Die Unsicherheit über die zu Grunde liegende Funktion bleibt weiterhin realistisch: sie ist in der Nähe der Beobachtungen relativ klein und steigt mit dem Abstand zu ihnen an. Um zwischen der Varianz für den Zufallsfehler  $\sigma_E^2$  und der Prozessvarianz  $\tau^2$  zu differenzieren, sind Experimentwiederholungen erforderlich. Ohne weitere Annahmen sind die beiden Parameter andernfalls nicht identifizierbar.

Gelegentlich werden deterministische Computermodelle mit der Gauß Prozess Regression für stochastische Zielgrößen analysiert. Dabei wird ein kleiner Zufallsfehler zugelassen, um so eine numerisch stabilere Vorhersage zu erhalten. Dieses Vorgehen bietet sich in Situationen an, in denen die Ausgaben eines deterministischen Computermodells durch kleine numerische deterministische Artefakte verrauscht sind.

## 2.5 Zusammenfassung und Bemerkungen

In diesem Kapitel wurde die Modellierung mit Gauß Prozessen als flexible Methode zur Interpolation und Regression glatter Funktionen vorgestellt. Neben einer Approximation liefert sie Informationen über die Unsicherheit unter den getroffenen Modellannahmen. Dabei lässt sich die Vorhersageverteilung bei bekannter Korrelation geschlossen darstellen. Einsetzen von a posteriori Modalwert oder Maximum Likelihood Schätzern für die Hyperparameter liefern schnell berechenbare Vorhersagen und approximative Vorhersageintervalle.

Für die Modellierung deterministischer Computermodelle ist die Gauß Prozess Interpolation gut geeignet, da geeignete Vorhersagen und Vorhersageintervalle im Vergleich zu der Auswertung des tatsächlichen Computermodells vergleichsweise schnell berechnet werden können. Allerdings ist der numerische Aufwand im Vergleich zu einer klassischen linearen Regression deutlich größer. Selbst wenn auf einen vollständigen Bayes Ansatz verzichtet wird, ist die numerische Optimierung der Likelihood bzw. a posteriori bezüg-

lich der Korrelationsparameter aufwändig, da in jedem Schritt die Korrelationsfunktion häufig ausgewertet und Korrelationsmatrizen der Dimension  $n_M \times n_M$  invertiert werden müssen. Das R-Paket `DiceKriging` bietet eine effiziente Implementierung mit REML Schätzern in der Software R (Roustant *et al.*, 2010).

In der Praxis kann es bei der Schätzung der Korrelationsparameter zu numerischen Schwierigkeiten kommen. Kleine Werte für  $\psi_{M,j}$  bei der Gauß Korrelationsfunktion können zu einer numerisch singulären Korrelationsmatrix führen, so dass keine (stabile) Invertierung mehr möglich ist. Dies wird durch kleine Abstände zwischen Designpunkten zusätzlich begünstigt. In anderen Situationen ist die Likelihood für  $\psi_{M,j}$  sehr flach, so dass unrealistisch große ML oder REML Schätzer resultieren. Im Extremfall entsprechen die Vorhersagefunktionen dann größtenteils der Regressionskomponente mit einzelnen Peaks, um die Interpolation der Beobachtungen zu erreichen. Aus diesem Grund schlagen Li und Sudjianto (2005) einen penalisierten Likelihood Ansatz vor, der große Werte für  $\psi_{M,j}$  bestraft. Im Bayes Ansatz lässt sich dieses Problem durch die Verwendung von a priori Verteilungen begegnen. Roustant *et al.* (2010) gehen extremen Werten für die Komponenten von  $\psi_M$  durch Beschränkung des zulässigen Wertebereichs auf ein Intervall aus dem Weg. Für die hier gewählte Parametrisierung ergibt sich mit den Standardwerten des Programms das Intervall  $[\frac{1}{8r_j^2}, 5 \cdot 10^{-19}]$ , wobei  $r_j$  der Spannweite der  $j$ -ten Variable aus der Designmatrix entspricht.

### 3 VORHERSAGEINTERVALLÜBERDECKUNG BEI DER GAUSS PROZESS INTERPOLATION

Die Gauß Prozess Interpolation aus Kapitel 2 liefert neben einer schnell zu berechnenden Approximation auch Informationen über deren Unsicherheit. Dahingegen berücksichtigen die meisten anderen Interpolationsmethoden die Unsicherheit der Approximation nicht und liefern lediglich einen Schätzwert  $\hat{y}_M(\mathbf{z})$  ohne Angaben zu den zu erwartenden Abweichungen von der Vorhersage zum Funktionswert.

Die berechenbare Unsicherheit der Interpolationsfunktion resultiert dabei aus einer stochastischen Formulierung des Interpolationsproblems mit Hilfe von Gauß Prozessen, wie in Kapitel 2 vorgestellt: Es werden a priori Annahmen über die zu Grunde liegende Funktion durch einen Gauß Prozess mit seiner Erwartungswert- und Kovarianzfunktion formuliert. In Kombination mit den vorliegenden Computerexperimenten kann schließlich die punktweise a posteriori Verteilung und damit auch ein Unsicherheitsintervall anhand von (2.14) bestimmt werden. Die Relevanz von Unsicherheitsangaben bei der Analyse von Computermodellen soll zunächst kurz an zwei Beispielen illustriert werden. Bei der Betrachtung von Computermodellen ist eine nahe liegende Frage wie gut die Simulation die Realität widerspiegelt. Dazu sind sowohl reale als auch simulierte Experimente durchzuführen. Für rechenintensive Simulationen, insbesondere wenn wie in der Kalibrierung von Computermodellen (siehe Kapitel 4 und 5) weitere zu schätzende Einflussgrößen in der Simulation vorliegen, ist ein direkter Abgleich der Experimente mit Computermodellauswertungen aus Zeitgründen nicht praktikabel, so dass Interpolationsmethoden unverzichtbar sind. Weicht nun die Approximation der Computersimulation  $\hat{y}_M(\mathbf{z})$  an einem Punkt  $\mathbf{z}$  stark vom Experiment ab, so hängt es von der Unsicherheit über Computermodell und Realität in  $\mathbf{z}$  ab, ob das Computermodell geeignet erscheint oder nicht. Bei großer Unsicherheit lässt sich noch keine Entscheidung über das Modell treffen, wohingegen bei kleiner Unsicherheit geschlossen werden kann,

dass das Computermodell verzerrt ist und wie groß die Abweichung von Simulation und Experiment ist.

Für die statistische Modellierung von Computermodellen ist die Wahl des Designs  $D = (z_1, \dots, z_{n_M})$ , an dem die Funktion  $y_M(\cdot)$  auszuwerten ist, von Bedeutung. Häufig werden raumfüllende Versuchspläne verwendet, die den Wertebereich  $\mathcal{Z}$  mit wenigen Punkten so gut wie möglich abdecken (siehe Fang *et al.*, 2006, Santner *et al.*, 2003). Solche raumfüllende Versuchspläne sind im Allgemeinen gut geeignet wenn die gesamte zu Grunde liegende Funktion approximiert werden soll. Ist jedoch ein bestimmtes Merkmal des Computermodells  $y_M(\cdot)$  von Interesse, wie zum Beispiel die Minimalstelle  $z^*$  von  $y_M(\cdot)$ , so können sequentielle Versuchspläne geeigneter sein. Dabei werden nach Auswertung des Computermodells an einem raumfüllenden Startdesign Designpunkte so hinzugefügt, dass am meisten über die Minimalstelle gelernt werden kann. Die Auswertung des Computermodells an der Minimalstelle von  $\hat{y}_M(\cdot)$  ist dabei nicht unbedingt die geeignetste Stelle. Wurde beispielsweise an dieser Stelle bereits ein Computerexperiment durchgeführt, so erhalten wir bei einem deterministischen Computermodell keinen Erkenntnisgewinn durch eine wiederholte Auswertung. Um am meisten im Hinblick auf eine Minimierung zu lernen, sollte das Computermodell an einer potentiellen Minimalstelle mit möglichst großer Unsicherheit ausgewertet werden. Somit werden intelligente sequentielle Designs eine Balance zwischen der Minimierung der Interpolationsfunktion und der Interpolationsunsicherheit finden. Dazu kann der nächste Punkt beispielsweise so gewählt werden, dass die erwartete Verbesserung im Vergleich zum aktuellen Minimum möglichst groß ist (siehe z.B. Jones *et al.*, 1998).

In den beiden vorgestellten Beispielen helfen Angaben zur Unsicherheit bei der Gauß Prozess Interpolation in Form von Vorhersageverteilungen oder -intervallen praxisrelevante Entscheidungen zu treffen. Die Angabe einer Vorhersageverteilung im Rahmen einer Interpolation stellt dabei ein sehr ambitioniertes Ziel dar und es ist zu überprüfen, wie zuverlässig die Angaben zur Unsicherheit in der Gauß Prozess Interpolation sind. Auch wenn viele Verfahren der Literatur die Vorhersageverteilungen nutzen, wurde diese Zuverlässigkeit im Hinblick auf die Überdeckung der Computermodellfunktion  $y_M(\cdot)$  durch Vorhersageintervalle in der Literatur noch nicht näher untersucht. Die hier

vorgestellten Resultate wurden auch in Kracker *et al.* (2010) veröffentlicht.

Im Folgenden wird die Überdeckungsgüte der Gauß Prozess Interpolation anhand von Simulationsstudien untersucht. Dabei werden unterschiedliche objektive a priori Verteilungen für den Korrelationsparameter und unterschiedliche numerische Ansätze eingesetzt. Kapitel 3.1 gibt zunächst einen Überblick über die objektiven a priori Verteilungen. Anschließend wird die Überdeckungsgüte in Kapitel 3.2 für Simulationen aus einem Gauß Prozess, Testfunktionen und ein Problem aus der Praxis untersucht.

### 3.1 Objektive a priori Verteilungen für Korrelationsparameter

Objektive a priori Verteilungen für den Korrelationsparameter bei der Gauß Prozess Interpolation wurden von Berger *et al.* (2001) und insbesondere Paulo (2005) vorgeschlagen und untersucht. Dabei nehmen wir im Folgenden an, dass  $a_{M,j} = 2$  für  $j = 1, \dots, d_M$  gilt - betrachten also nur den wichtigen Fall der Gauß Korrelationsfunktion, die zu differenzierbaren Realisationen des Gauß Prozesses führt. Die a posteriori Verteilung für den Korrelationsparameter  $\boldsymbol{\psi}$  ist dann gegeben durch

$$p(\boldsymbol{\psi}_M | \mathbf{y}_M^n) \propto p(\boldsymbol{\psi}_M) \cdot \det(\mathbf{R}_M)^{-1/2} \det(\mathbf{F}_M^T \mathbf{R}_M^{-1} \mathbf{F}_M)^{-1/2} \widehat{\sigma}_M^2^{-(n_M - p_M)/2}, \quad (3.1)$$

siehe auch (2.13). Im Folgenden werden wir drei verschiedene a priori Verteilungen für  $\boldsymbol{\psi}_M$  vergleichen: (i) eine nicht integrierbare flache Gleichverteilung auf  $[0, \infty)^{d_M}$ ; dies ist die klassische a priori Verteilung, wenn keine weiteren Informationen vorhanden sind. Auch wenn diese Verteilung nicht unbedingt uninformativ ist, wird sie häufig für die Gauß Prozess Interpolation verwendet, insbesondere da der Modalwert der a posteriori für  $\boldsymbol{\psi}_M$  dem REML Schätzer entspricht (Santner *et al.*, 2003). (ii) Eine a priori Verteilung für  $\boldsymbol{\psi}_M$  basierend auf dem Produkt von Dichten der Exponentialverteilung, deren Erwartungswert jeweils dem ML Schätzer entspricht. Es handelt sich somit um einen so genannten empirischen Bayes Ansatz, da die a priori Verteilung von den Daten abhängt. (iii) Der dritte Ansatz ist eine marginal unabhängige Jeffreys a priori, wie in Paulo (2005) vorgestellt. Eine Jeffreys a priori besitzt die wünschenswer-

te Eigenschaft, dass sie invariant gegenüber Reparametrisierungen des Modells ist und weiterhin kann gezeigt werden, dass resultierende Kreditabilitätsintervalle die nominale frequentistische Überdeckung schneller erreichen als bei anderen a priori Klassen (Kass und Wasserman, 1996). Für einen multivariaten Gauß Prozess existieren verschiedene Varianten dieser objektiven a prioris, von denen einige in Paulo (2005) hinsichtlich ihrer frequentistischen Überdeckungswahrscheinlichkeit in einer Simulationsstudie verglichen werden. Dazu wurden Realisationen aus einem Gauß Prozess mit festen Hyperparametern simuliert und die Überdeckungswahrscheinlichkeit der Kreditabilitätsintervalle für  $\boldsymbol{\psi}_M$  im Hinblick auf das nominale Niveau bewertet. Hierbei kommt Paulo (2005) zu dem Ergebnis, dass die marginal unabhängige Jeffreys a priori im Hinblick auf Überdeckungsgenauigkeit und Rechenaufwand im Vergleich zu weiteren objektiven a priori Verteilungen zu empfehlen ist. Deshalb wird hier diese Jeffreys a priori betrachtet. Zusammenfassend vergleichen wir

$$p^{(1)}(\boldsymbol{\psi}_M) \propto \mathbb{1} \text{ für } \psi_j > 0,$$

$$p^{(2)}(\boldsymbol{\psi}_M) = \prod_{i=1}^{d_M} f_{Exp} \left( \psi_{M,j}, \lambda = \left( \hat{\psi}_{M,j,ML} \right)^{-1} \right),$$

und

$$p^{(3)}(\boldsymbol{\psi}_M) \propto |I_J(\boldsymbol{\psi}_M)|^{-1/2}$$

mit der Matrix

$$I_J(\boldsymbol{\psi}_M) = \begin{pmatrix} n & \text{tr}U_1 & \text{tr}U_2 & \cdots & \text{tr}U_d \\ & \text{tr}U_1^2 & \text{tr}U_1U_2 & \cdots & \text{tr}U_1U_d \\ & & \ddots & \cdots & \vdots \\ & & & & \text{tr}U_d^2 \end{pmatrix},$$

dabei bezeichnet  $f_{Exp}$  die Dichte der Exponentialverteilung mit Rate  $\lambda$  und  $U_j = (\frac{\partial}{\partial \psi_{M,j}} \mathbf{R}_M) \mathbf{R}_M^{-1}$  für  $j = 1, \dots, d$  ist das Produkt der Korrelationsmatrix und deren partiellen Ableitung bezüglich  $\psi_{M,j}$ . Interessanterweise hängt die Jeffreys a priori  $p^{(3)}(\boldsymbol{\xi})$  über die Korrelationsmatrix  $\mathbf{R}_M$  vom Design  $\mathbf{D}_M$  ab. Die gemeinsame a priori Verteilung aller Hyperparameter ist gegeben durch  $p(\boldsymbol{\beta}_M, \sigma_M^2, \boldsymbol{\psi}_M) \propto p^{(m)}(\boldsymbol{\psi}_M)(\sigma_M^2)^{-1}$ , mit  $m = 1, 2, 3$  unter der Annahme dass  $\boldsymbol{\beta}_M$  und  $(\sigma_M^2, \boldsymbol{\psi}_M)$  unabhängig sind (Paulo, 2005). Rechentechnisch ist die Jeffreys a priori am aufwändigsten zu berechnen, da Matrixinverse berechnet werden müssen. Das Produkt der Exponential a prioris ist ebenfalls etwas komplexer, da zunächst der ML Schätzer numerisch bestimmt werden muss. Am schnellsten kann die flache Gleichverteilung ausgewertet werden. Die Gestalt der flachen a priori  $p^{(1)}$  und des Produkts der Exponential a prioris ist gut bekannt, wohingegen sich die Gestalt der Jeffreys a priori nicht ohne weiteres aus der Formel ableiten lässt. Abbildung 3.1 illustriert die Jeffreys a priori bei äquidistantem Design für einen eindimensionalen Gauß Prozess: kleine Werte für  $\psi_M$  besitzen größere a priori Wahrscheinlichkeit. Der Stichprobenumfang beeinflusst wie schnell die a priori Wahrscheinlichkeit für steigendes  $\psi_M$  gegen Null geht. Diese Form lässt sich auch auf höhere Dimensionen übertragen und ähnelt somit eher der Exponential als der flachen a priori Verteilung. Die drei a priori Verteilungen werden alle als so genannte nichtinformativ oder objektive a priori Verteilungen bezeichnet. Dabei stellt sich natürlich die Frage welche a priori in dem vorliegenden Modell als objektiv bezeichnet werden kann. Auf den ersten Blick wirkt die flache a priori Verteilung  $p^{(1)}$ , die proportional zu einer Gleichverteilung auf dem Parameterbereich  $(0, \infty)^{d_M}$  gewählt wird, nichtinformativ.

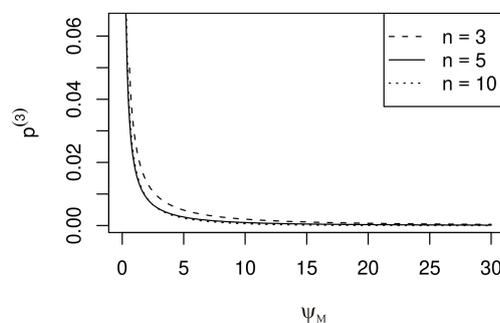


Abbildung 3.1: Jeffreys a priori im eindimensionalen Fall für äquidistantes Design auf  $[0, 1]$  mit verschiedenen Stichprobenumfängen ( $n = 3, 5, 10$ ) und festem  $a_{M,j} = 2$ .

In Abbildung 3.2 wird die Korrelation zwischen zwei Punkten für festen Abstand 0.1 in Abhängigkeit des Korrelationsparameters  $\psi_M$  dargestellt. Es ist deutlich erkennbar, dass sich die Korrelation für große Werte von  $\psi_M$  nur noch wenig ändert und einen flachen Funktionsverlauf besitzt. Die a priori  $p^{(1)}$  berücksichtigt diesen Zusammenhang nicht, dort bekommen somit kleine Korrelationen ein stärkeres Gewicht als große Korrelationen. Dahingegen werden beispielsweise bei der Jeffreys a priori  $p^{(3)}$  die kleinen Parameterwerte stärker gewichtet um diesen Effekt auszugleichen.

In speziellen Situationen kann unter Umständen auch eine informative a priori sinnvoll sein. Higdon *et al.* (2008) parametrisieren beispielsweise bei der Auswertung hochdimensionaler Computermodelle den Korrelationsparameter so, dass sich die umparametrisierten Korrelationsparameter inhaltlich als Korrelation interpretieren lassen und damit nur Werte zwischen 0 und 1 annehmen können. Um Informationen auch für  $\beta$  und  $\sigma^2$  sinnvoll einfließen zu lassen, werden dort neben dem Wertebereich auch die Beobachtungen skaliert.

Werden nicht integrierbare a priori Verteilungen (wie  $p^{(1)}$  und  $p^{(3)}$  von oben) verwendet, ist die a posteriori unter Umständen ebenfalls nicht integrierbar, d.h. das Integral in (2.14) kann unendlich sein. Berger *et al.* (2001) haben für den eindimensionalen Fall in einem sehr ähnlichen Modell gezeigt, dass eine nicht integrierbare a priori für  $\psi_M$  in einer nicht integrierbaren a posteriori resultiert. Bei der Integrierbarkeit spielt die Regressionsfunktion eine Rolle. Paulo (2005) konnte zeigen, dass die Jeffreys a priori im multivariaten Fall beispielsweise bei einem faktoriellen Versuchsplan und einer konstanten Erwartungswertfunktion eine integrierbare a posteriori liefert. In der Praxis

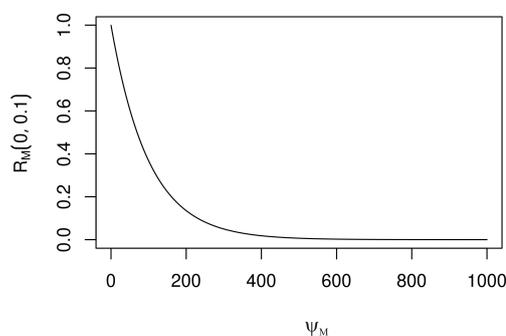


Abbildung 3.2: Gauß Korrelationsfunktion in Abhängigkeit von  $\psi_M$  für Abstand  $x - x' = 0.1$ .

kann die Integrierbarkeit durch einen kompakten Parameterraum für  $\boldsymbol{\psi}_M$  erzwungen werden. Dazu müssen die Korrelationsparameter  $\psi_{M,j}$  sowohl nach oben als auch nach unten auf ein kompaktes Intervall beschränkt werden. In dem a posteriori Modalwert Ansatz verursacht die fehlende Integrierbarkeit theoretisch keine Probleme.

Für die Beschränkung des Korrelationsparameter gibt es neben der theoretischen Probleme mit der Integrierbarkeit auch praktische: Der Grenzwert der Korrelationsmatrix für  $\psi_{M,j} \rightarrow_+ 0$  in allen Komponenten ist eine singuläre Matrix  $\mathbf{J}$  deren Einträge alle 1 sind. Für kleine Werte von  $\psi_{M,j}$  kann es je nach Datensituation deshalb zu numerischen Schwierigkeiten beim Invertieren der Korrelationsmatrix kommen. Bei sehr großen Korrelationsparametern konvergiert die Korrelationsmatrix gegen die Einheitsmatrix. Die Vorhersagen entsprechen dann in großen Teilen des Parameterraums der Schätzung aus dem generalisierten linearen Modell, allerdings kommt es durch die erzwungene Interpolation zu einigen wenigen Peaks innerhalb der Funktion. Diese Formen können zwar häufig a priori ausgeschlossen werden, allerdings kann durch das Erzwingen einer glatten Vorhersage auch die Überdeckungsgüte reduziert werden.

### 3.2 Untersuchungen zur Überdeckungsgüte

In diesem Kapitel werden wir die Überdeckungsgüte der Vorhersageintervalle für die Gauß Prozess Interpolation in einigen Situationen überprüfen. Selbstverständlich spielen dabei insbesondere die getroffenen Modellannahmen und die zu approximierende Funktion eine wesentliche Rolle. Hier soll die Überdeckung einerseits für Simulationen aus einem Gauß Prozess sowie andererseits für einige Testfunktionen und ein praktisches Datenbeispiel aus der Umformung untersucht werden. Dabei werden die drei verschiedenen objektiven a priori Verteilungen  $p^{(1)}, p^{(2)}, p^{(3)}$  für den Korrelationsparameter  $\boldsymbol{\psi}_M$  und die zwei verschiedenen numerische Ansätze a posteriori Modalwert (PM) und Markov Chain Monte Carlo (MCMC), (Kapitel 2.2), hinsichtlich der resultierenden Überdeckungsgüte verglichen.

Für die Auswertung wird die Gauß Prozess Interpolation mit einem konstanten Erwartungswert  $\mu_M(\mathbf{z}) = \beta_0$  und der Gauß Korrelationsfunktion verwendet. Der Kor-

relationsparameter  $\psi_M$  wird als unbekannt betrachtet. In allen Situationen werden raumfüllende Versuchspläne, genauer gesagt Maximin Latin Hypercube Versuchspläne, verwendet. Dieser verbreitete Versuchsplan für Computereperimente maximiert den minimalen Abstand der Punkte in der Klasse der Latin Hypercube Versuchspläne. Dabei ist die Projektion der Punkte in jeder Dimension gleichmäßig verteilt. Hier wurde er mit Hilfe des `lhs` Packets für die Statistiksoftware `R` bestimmt (Carnell, 2009).

In Kapitel 3.2.1 werden zunächst Zufallsfunktionen aus einem Gauß Prozess erzeugt und für die resultierenden Datensätze wird jeweils die Gauß Prozess Interpolation angewandt. Dies stellt den Idealfall für die Gauß Prozess Interpolation dar, da hier das Modell für die Interpolation mit dem Modell zur Generierung der Zufallsfunktionen übereinstimmt. Diese Art von Simulationen wurden auch von Paulo (2005) für unterschiedliche objektive a priori Verteilungen und einen faktoriellen Versuchsplan der Größe  $n_M = 25$  durchgeführt. Dort wird jedoch nur die Überdeckungsgüte für die Unsicherheitsintervalle der Hyperparameter  $\psi_M$  untersucht. Die Unsicherheitsintervalle für die zu Grunde liegende Funktion  $y_M(\cdot)$  sind bei der Analyse von Computereperimenten jedoch von größerem praktischen Interesse. In Kapitel 3.2.2 wird anschließend die Überdeckung für einige realistische Testfunktionen untersucht, die nicht notwendigerweise wie eine typische Realisation eines kovarianzstationären Gauß Prozesses aussehen. In dieser Situation sind also die Annahmen der Gauß Prozess Interpolation nicht zwingend erfüllt. Abschließend werden in Kapitel 3.2.3 die Überdeckungs- und Vorhersageeigenschaften der Gauß Prozess Interpolation für einen echten Datensatz betrachtet. Dabei liegen Daten aus einem Computermodell für die Rückfederung beim Tiefziehen vor.

Im PM Ansatz für die Schätzung des Korrelationsparameters  $\psi_M$  wurde der Modalwert der a posteriori Dichte (2.13) für  $\psi_M$  mit einem Nelder-Mead Optimierer (Funktion `optim` in `R`) bestimmt. Häufig ist die a posteriori Dichte multimodal, so dass verschiedene Startwerte für die Optimierung verwendet werden. Für den MCMC Ansatz wurde ein Random Walk Metropolis Algorithmus basierend auf der a posteriori Dichte für  $\psi_M$  in Gleichung (3.1) verwendet. Die Implementierung erfolgt mit Hilfe des `R` Packets `mcmc` (Geyer, 2009). Um eine effiziente Markovkette zu erhalten wurde der Algorithmus in einer Startphase zunächst eingestellt. Dabei wurde die Schrittweite des Random

Walks so gewählt, dass eine Akzeptanzrate von 0.3 erreicht wurde und die Kovarianzmatrix der Vorschlagsdichte proportional zu der geschätzten Kovarianzmatrix dieser vorläufigen Kette ist (siehe auch Robert und Casella (2004) und Anhang A.1). Mit der gewonnenen Spezifikation wurden anschließend 15000 Iterationen ausgehend vom a posteriori Modalwert durchgeführt. Die ersten 100 Iterationen wurden als Anfangsperiode verworfen und nur jeder 10. Wert wurde für die Auswertung verwendet.

Die Zuverlässigkeit der Unsicherheitsangaben der Gauß Prozess Interpolation werden über Maßzahlen bewertet. Insbesondere wird die funktionsweite Überdeckung betrachtet, bei der die punktweise Überdeckung der wahren Funktion durch Vorhersageintervalle über den Wertebereich gemittelt wird. Dieses Maß ist von einem praktischen Blickwinkel aus sinnvoll und eine Variante hiervon wurde beispielsweise in der Literatur zu Glättungssplines bereits theoretisch untersucht (Gu, 2002, Kapitel 3.3).

Für die Berechnung der Überdeckungsgüte werden zunächst punktweise Vorhersageintervalle  $P_{y_M}(\mathbf{z}) = [L_{y_M}(\mathbf{z}), U_{y_M}(\mathbf{z})]$  berechnet, wobei  $L_{y_M}(\mathbf{z})$  und  $U_{y_M}(\mathbf{z})$  die 0.025 und 0.975 Quantile der a posteriori Verteilung am Punkt  $\mathbf{z}$  bezeichnet. Für den PM Ansatz sind die Grenzen gerade die Quantile der t-Verteilung aus Gleichung (2.9), wohingegen im MCMC Ansatz empirische Quantile aus den Zufallszahlen der Vorhersageverteilung (2.14) verwendet werden. Anschließend wird überprüft, ob der wahre Funktionswert  $y_M(\mathbf{z})$  im Intervall  $P_{y_M}(\mathbf{z})$  enthalten ist. Dies wird in der Größe  $c_{y_M}(\mathbf{z})$ ,  $c_{y_M}(\mathbf{z}) = \mathbb{1}_{[L_{y_M}(\mathbf{z}), U_{y_M}(\mathbf{z})]}(y_M(\mathbf{z}))$ , festgehalten. Durch Integrieren über den Wertebereich  $C_{y_M} = \int_{\mathcal{Z}} c_{y_M}(\mathbf{z}) d\mathbf{z}$  erhalten wir eine Maßzahl für die Überdeckungsgüte der gesamten Funktion. Somit ist  $C_{y_M}$  die durchschnittliche Überdeckung über den Wertebereich. Neben der Überdeckung ist auch die Länge der Vorhersageintervalle von Interesse. Die mittlere Länge der Vorhersageintervalle über den Wertebereich wird als  $H_{y_M} = \int_{\mathcal{Z}} |U_{y_M}(\mathbf{z}) - L_{y_M}(\mathbf{z})| d\mathbf{z}$  definiert. Alle Integrale werden durch numerische Integration auf  $\mathcal{Z}$  berechnet.

Liegen wie bei der Simulation aus dem Gauß Prozess  $n_{\text{sim}}$  unterschiedliche Funktionen vor, so kann die mittlere funktionsweite Überdeckung  $\bar{C}_{y_M} = \frac{1}{n_{\text{sim}}} \sum_{i=1}^{n_{\text{sim}}} C_{y_M}^{(i)}$  und mittlere funktionsweite Vorhersageintervalllänge  $\bar{H}_{y_M} = \frac{1}{n_{\text{sim}}} \sum_{i=1}^{n_{\text{sim}}} H_{y_M}^{(i)}$  durch Mitteln über die einzelnen Simulationen bestimmt werden. Dies entspricht dem Schätzer für

die funktionsweite Überdeckungswahrscheinlichkeit und erwartete Länge.

### 3.2.1 Interpolation von simulierten Gauß Prozessen

In einem ersten Schritt werden 100 Realisationen von einem zweidimensionalen Gauß Prozess mit konstanter Erwartungswertfunktion  $\mu_M(\mathbf{z}) = 1$ ,  $\sigma_M^2 = 1.5$ , Gauß Korrelationsfunktion und  $\boldsymbol{\psi} = (3, 0.5)^T$  auf  $\mathcal{Z} = [0, 1]^2$  erzeugt. Es wird ein Maximin Latin Hypercube als Versuchsplan mit Stichprobenumfang  $n_M = 10, 15, 20, 25, 30$  verwendet, um den Einfluss des Stichprobenumfangs auf die Angaben zur Unsicherheit zu untersuchen.

In Tabelle 3.1 werden die geschätzten Überdeckungswahrscheinlichkeiten  $\bar{C}_{y_M}$  und erwarteten Intervalllängen  $\bar{H}_{y_M}$  dargestellt, jeweils gemittelt über die 100 simulierten Funktionen. Für die MCMC basierte Analyse wurde außerdem die Überdeckung für die Korrelationsparameter  $\bar{C}_{\psi_{M,j}}$  auf Basis der empirischen Quantile der MCMC Simulationen geschätzt. Die funktionsweite Überdeckung für kleine Stichprobenumfänge ( $n_M = 10$ ) liegt für alle a priori Verteilungen unterhalb des Zielwertes von 95%. Dabei schneiden die flache und die datenabhängige Exponential a priori in Kombination mit einer MCMC Auswertung am besten ab. Mit steigendem Stichprobenumfang erreichen alle Methoden approximativ das nominale Niveau 0.95 mit leicht konservativen Ergebnissen für  $n_M = 30$ . Dabei lässt sich der bekannte Zielkonflikt zwischen der Intervalllänge und Überdeckungswahrscheinlichkeit beobachten: Der PM Ansatz liefert kürzere Vorhersageintervalle jedoch auch eine geringere Überdeckungsgüte als der MCMC Ansatz. Dabei kann das nominale Niveau mit dem PM Ansatz in den meisten Situationen nicht erreicht werden. Die Unterschiede zwischen den beiden numerischen Ansätzen werden jedoch für steigenden Stichprobenumfang immer kleiner. Insgesamt dominiert keine der a priori Verteilungen die anderen hinsichtlich Überdeckungswahrscheinlichkeit und Intervalllänge.

Für den Korrelationsparameter  $\boldsymbol{\psi}_M$  sind die Simulationsergebnisse recht ähnlich: für kleine Stichprobenumfänge ist die Überdeckung schlechter, erreicht aber schnell das nominale Niveau. Für die Hyperparameter  $\psi_{M,j}$  kann dies erwartet werden, da Baye-

Tabelle 3.1: Geschätzte Überdeckungswahrscheinlichkeiten der 95% Vorhersage- / Kreditabilitätsintervalle gemittelt über 100 Gauß Prozess Realisationen (in Klammern: 0.1, 0.9 Quantil der funktionsweiten Überdeckung  $C_{y_M}$ ).

$n_M$	$\bar{C}_{y_M}^{(\text{MCMC})}$	$\bar{C}_{y_M}^{(\text{PM})}$	$\bar{H}_y^{(\text{MCMC})}$	$\bar{H}_{y_M}^{(\text{PM})}$	$\bar{C}_{\psi_{M,1}}^{(\text{MCMC})}$	$\bar{C}_{\psi_{M,2}}^{(\text{MCMC})}$
Flache a priori						
10	0.88 (0.64,1)	0.77 (0.34,1)	1.179	0.167	0.78	0.77
15	0.96 (0.84,1)	0.87 (0.64,1)	0.081	0.048	0.94	0.93
20	0.96 (0.88,1)	0.91 (0.72,1)	0.026	0.021	0.92	0.93
25	0.95 (0.84,1)	0.88 (0.69,1)	0.008	0.007	0.99	0.96
30	0.98 (0.95,1)	0.95 (0.88,1)	0.004	0.004	0.92	0.93
Exponential a priori						
10	0.87 (0.66,1)	0.74 (0.41,1)	0.209	0.148	0.80	0.83
15	0.94 (0.78,1)	0.86 (0.61,1)	0.065	0.046	0.93	0.90
20	0.94 (0.76,1)	0.87 (0.60,1)	0.025	0.020	0.94	0.91
25	0.95 (0.84,1)	0.89 (0.64,1)	0.008	0.007	0.96	0.93
30	0.98 (0.95,1)	0.94 (0.87,1)	0.004	0.004	0.97	0.94
Jeffreys a priori						
10	0.73 (0.24,1)	0.63 (0.26,0.98)	0.247	0.128	0.67	0.59
15	0.94 (0.76,1)	0.84 (0.51,1.00)	0.061	0.044	0.95	0.89
20	0.93 (0.74,1)	0.87 (0.60,1.00)	0.025	0.021	0.98	0.95
25	0.97 (0.89,1)	0.92 (0.76,1.00)	0.008	0.007	0.94	0.95
30	0.98 (0.92,1)	0.92 (0.80,1.00)	0.004	0.004	0.98	0.94

sianische Kreditabilitätsintervalle asymptotisch als frequentistische Konfidenzintervalle interpretiert werden können, insbesondere wenn objektive a priori Verteilungen verwendet werden.

Zusammenfassend sind die Ergebnisse recht zufriedenstellend: die Überdeckungsgüte bei der Gauß Prozess Interpolation liefert in der Modellsituation gute Resultate, so lange der Stichprobenumfang nicht zu klein ist.

### 3.2.2 Interpolation von Testfunktionen

Im nächsten Schritt soll die Überdeckung der Gauß Prozess Interpolation untersucht werden, wenn diese auf Testfunktionen angewendet wird, die nicht notwendigerweise

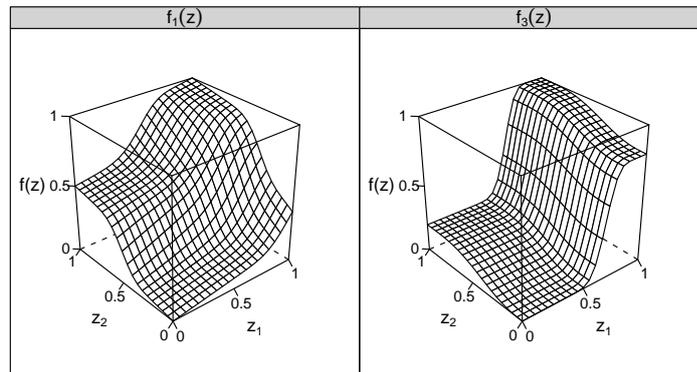


Abbildung 3.3: Testfunktionen  $f_1(\mathbf{x})$  and  $f_3(\mathbf{x})$ .

Realisationen eines Gauß Prozesses sind. Dazu werden hier drei Testfunktionen der Dimension Zwei und Sieben betrachtet, deren Komplexität und die zur Auswertung verwendeten Stichprobenumfänge sind vergleichbar mit der Situation in der Anwendung auf die Blechumformung aus Kapitel 6.

Die erste Testfunktion  $f_1$  ist eine zweidimensionale glatte Konvexkombination von Verteilungsfunktionen  $F_{\text{Beta}}(\mathbf{z}, \alpha, \beta)$  der Beta Verteilung:

$$f_1(\mathbf{z}) = 1/3 \cdot F_{\text{Beta}}(0.5z_1 + 0.5z_2, 1, 1) + 1/3 \cdot F_{\text{Beta}}(0.7z_1 + 0.3z_2, 20, 10) \\ + 1/3 \cdot F_{\text{Beta}}(0.15z_1 + 0.85z_2, 15, 20).$$

Diese Funktion (dargestellt in Abbildung 3.3 links) steigt moderat an und sollte somit gut durch eine kovarianzstationären Gauß Prozess beschrieben werden können.

Die zweite Testfunktion beschreibt einen physikalischen Prozess, der analytisch beschrieben und somit schnell simuliert werden kann. Der Kolben Simulator (engl. piston simulator) von Kenett und Zacks (1998) modelliert die Taktzeit einer vollständigen Umdrehung einer Kolbenachse in Abhängigkeit von sieben Einflussgrößen: Kolbengewicht ( $M$ , 30-60 kg), Kolbenoberfläche ( $S$ , 0.005-0.02  $m^2$ ), Anfangsgasvolumen ( $V_0$ , 0.002 – 0.01  $cm^3$ ), Federkonstante ( $k$ , 1000 – 5000  $N/m$ ), atmosphärischer Druck ( $P_0$ ,  $9 \cdot 10^4$  –  $11 \cdot 10^4$   $N/m^2$ ), Raumtemperatur ( $T$ , 290 – 396  $K$ ) und Füllgastemperatur

$(T_0, 340 - 3602K)$  als

$$\text{Taktzeit} = 2\pi \sqrt{\frac{M}{k + S^2 \frac{P_0 V_0}{T_0} \frac{T}{V^2}}}$$

mit  $V = \frac{S}{2k} \left( \sqrt{A^2 + 4k \frac{P_0 V_0}{T_0} T} - A \right)$  und  $A = P_0 S + 19.62M - \frac{kV_0}{S}$ . Diese Funktion dient als höherdimensionales Beispiel, dass auf einem realen Prozess basiert. Der Einfachheit halber skalieren wir alle Einflussgrößen auf  $[0, 1]$  und nennen die Funktion der Taktzeit  $f_2(\mathbf{z})$ . In diesem höherdimensionalen Beispiel werden im MCMC Algorithmus 20000 Iterationen verwendet, wobei nur jeder 10. Wert nach einer Einschwingphase von 1000 gespeichert wird.

Die dritte Funktion ist wieder eine monotone Kombination von Beta Verteilungsfunktionen:

$$f_3(\mathbf{z}) = 0.8 \cdot F_{\text{Beta}}(z_1, 30, 15) + 0.2 \cdot F_{\text{Beta}}(z_2, 3, 4).$$

Diese Funktion besitzt einen einzelnen steilen Anstieg und ist in den übrigen Bereichen nahezu konstant (siehe Abbildung 3.3 rechts). Dieses Beispiel wurde ausgewählt, um zu untersuchen wie die Gauß Prozess Interpolation bei der Auswertung einer Funktion funktioniert, die nicht einer typischen Realisation eines kovarianzstationären Gauß Prozesses gleicht.

Für jede Testfunktion werden drei Stichprobenumfänge (5, 7 und 10 Beobachtungen je Dimension  $d_M$ ), drei verschiedene a priori Verteilungen ( $p^{(1)}$ ,  $p^{(2)}$  und  $p^{(3)}$ ) sowie zwei numerische Methoden (MCMC and PM) verglichen. Die Erwartungswertfunktion für den Gauß Prozess wird wiederum als konstant angenommen und eine Gauß Korrelationsfunktion verwendet, so dass der Gauß Prozess glatte differenzierbare Realisationen besitzt.

In Tabelle 3.2 wird die funktionsweite Überdeckung  $C_{y_M}$  der Testfunktionen zusammen mit der Vorhersageintervalllänge  $H_{y_M}$  angegeben. Hier resultieren die vorliegenden Werte aus einer einzelnen Funktionsauswertung; da keine Zufallsfunktionen aus einem stochastischen Modell gezogen werden, kann nicht wie in Kapitel

Tabelle 3.2: Funktionsweite Überdeckung für Testfunktionen

n / d	a priori	$C_{y_M}$ (MCMC)	$C_{y_M}$ (PM)	$H_{y_M}$ (MCMC)	$H_{y_M}$ (PM)	RMSE (MCMC)	RMSE (PM)
$f_1(\mathbf{z})$ ( $d_M = 2$ )							
5	flache	0.875	0.512	0.149	0.082	0.056	0.059
5	Exponential	0.578	0.397	0.094	0.070	0.064	0.064
5	Jeffreys	0.748	0.333	0.154	0.061	0.051	0.069
7	flache	1.000	0.862	0.218	0.141	0.037	0.044
7	Exponential	0.891	0.796	0.160	0.129	0.045	0.046
7	Jeffreys	0.989	0.771	0.178	0.122	0.043	0.048
10	flache	0.891	0.760	0.087	0.067	0.022	0.024
10	Exponential	0.844	0.703	0.078	0.063	0.024	0.026
10	Jeffreys	0.837	0.680	0.078	0.062	0.024	0.027
$f_2(\mathbf{z})$ ( $d_M = 7$ )							
5	flache	0.913	0.832	0.048	0.038	0.013	0.014
5	Exponential	0.871	0.833	0.041	0.038	0.013	0.014
5	Jeffreys	1.000	0.811	2.608	0.061	0.059	0.025
7	flache	0.964	0.781	0.043	0.032	0.013	0.016
7	Exponential	0.860	0.703	0.034	0.026	0.015	0.017
7	Jeffreys	0.863	0.771	0.050	0.039	0.020	0.020
10	flache	0.786	0.758	0.020	0.017	0.009	0.009
10	Exponential	0.781	0.727	0.018	0.015	0.009	0.009
10	Jeffreys	0.791	0.789	0.020	0.018	0.009	0.009
$f_3(\mathbf{z})$ ( $d_M = 2$ )							
5	flache	1.000	0.900	1.011	0.210	0.123	0.067
5	Exponential	0.961	0.816	0.237	0.191	0.077	0.076
5	Jeffreys	0.950	0.705	0.247	0.176	0.115	0.092
7	flache	0.986	0.812	0.170	0.115	0.041	0.039
7	Exponential	0.902	0.730	0.148	0.128	0.037	0.035
7	Jeffreys	0.878	0.605	0.144	0.103	0.035	0.034
10	flache	0.506	0.385	0.073	0.061	0.054	0.056
10	Exponential	0.426	0.327	0.066	0.056	0.058	0.059
10	Jeffreys	0.424	0.311	0.067	0.054	0.059	0.060

3.2.2 über Wiederholungen gemittelt werden. Die Wurzel der Fehlerquadratsumme  $\text{RMSE} = \sqrt{\int_{\mathcal{Z}} (\hat{y}_M(\mathbf{z}) - y_M(\mathbf{z}))^2 d\mathbf{z}}$ , mit  $\hat{y}_M(\mathbf{z})$  als a posteriori Erwartungswert im MCMC oder PM Ansatz, misst die Approximationsgüte der Interpolation.

Für alle Testfunktionen liegt die funktionsweite Überdeckung beim MCMC Ansatz in der Regel über 80% und für den a posteriori Modalwertansatz über 70%. Der Unter-

schied zwischen MCMC und PM Ansatz ist hier ein wenig größer als bei den Simulationen im letzten Kapitel. Von den drei verschiedenen a priori Verteilungen scheint die flache a priori  $p^{(1)}$  Resultate am nächsten zum nominalen Überdeckungslevel zu liefern. Insgesamt erreichen die Überdeckungen in vielen Situationen jedoch nicht die 95%. Wie erwartet sind die Resultate für die Funktion  $f_3$  schlechter, insbesondere für 10 Beobachtungen pro Dimension. Im Gegensatz zu den Ergebnissen bei Simulation aus dem Gauß Prozess steigt die mittlere Überdeckung hier interessanterweise nicht mit dem Stichprobenumfang, sondern fällt bei großen Stichprobenumfängen (insbesondere für  $f_2$  und  $f_3$ ) wieder. Ein großer Stichprobenumfang kann anscheinend eine falsche Sicherheit bewirken: Wenn die wahre Funktion stark von den Annahmen für ein stationäres Gauß Prozess Modell abweicht, werden die Unsicherheitsangaben weniger zuverlässig. Die Funktion  $f_3$  ist durch den einzelnen steilen Anstieg keine typische Realisation eines stationären Gauß Prozesses, somit sind hier die Annahmen nicht erfüllt. Insgesamt sind die Resultate für die Überdeckung, insbesondere für den MCMC Ansatz, jedoch annehmbar zufriedenstellend.

### 3.2.3 Überdeckungs- und Vorhersagegüte für die Rückfederungssimulation

Abschließend soll nun die Überdeckungs- und Vorhersagegüte der Gauß Prozess Interpolation anhand realer Computerexperimente für einen Umformprozess untersucht werden. Die vorliegenden Daten sind Rückfederungssimulationen für das Tiefziehen des so genannten Hutprofils mit dem Material DX55. Der physikalische Prozess wird in Kapitel 6 ausführlicher beschrieben. Hier wurde einerseits die Ziehtiefe ( $z_1$ ) und die Niederhalterkraft ( $z_2$ ) variiert. Außerdem wurden vier numerische Einflussgrößen - die Elementkantenlänge ( $z_3$ ), die Anzahl der Elemente über den Stempelradius ( $z_4$ ), die Anzahl der Integrationspunkte über das Blech ( $z_5$ ) sowie die Stempelgeschwindigkeit ( $z_6$ ) - im Computermodell variiert. Insgesamt wurden 79 Beobachtungen erhoben, davon wurde die Ziehtiefe und die Niederhalterkraft nur auf 3 Stufen variiert. Für die übrigen Größen liegen zwischen 5 und 7 unterschiedliche Werte vor. Als Zielgröße dient der Flanschwinkel als Maß der Rückfederung. Ziel bei der Datenerhebung war die Sen-

sitivität der Rückfederungssimulation gegenüber den numerischen Einflussgrößen zu untersuchen. Hier werden wir den Datensatz als reales Beispiel für eine Gauß Prozess Interpolation verwenden.

Die Kreuzvalidierung ist in der vorliegenden Situation, bei der Beobachtungen ohne Zufallsfehler vorliegen, die einzige Möglichkeit zur Gütebewertung. Hier wird eine so genannte Leave-one-out Kreuzvalidierung durchgeführt. Für den  $i$ -ten Simulationslauf werden die Vorhersage  $\hat{y}_M^{(-i)}(\mathbf{z}_i)$  und das korrespondierende Vorhersageintervall  $[L_{y_M}^{(-i)}(\mathbf{z}_i), U_{y_M}^{(-i)}(\mathbf{z}_i)]$  anhand einer Gauß Prozess Interpolation, angepasst auf den übrigen  $n - 1$  Daten, bestimmt. Für eine Auswertung wird einerseits der „root mean square error“,  $\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_{M,i} - \hat{y}_M^{(-i)}(\mathbf{z}_i))^2}$ , und andererseits die mittlere Überdeckungsgüte

$$\bar{c}_{y_M} = \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{[L_{y_M}^{(-i)}(\mathbf{z}_i), U_{y_M}^{(-i)}(\mathbf{z}_i)]}(y_M(\mathbf{z}_i))$$

sowie die mittlere Vorhersageintervalllänge  $\bar{h}_{y_M} = \frac{1}{n} \sum_{i=1}^n U_{y_M}^{(-i)}(\mathbf{z}_i) - L_{y_M}^{(-i)}(\mathbf{z}_i)$  am Punkt  $\mathbf{z}_i$  bestimmt.

Auch hier werden Überdeckungs- und Vorhersagegüte für die drei objektiven a priori Verteilungen  $p^{(1)}$ ,  $p^{(2)}$  und  $p^{(3)}$  und die zwei verschiedenen numerische Ansätze a posteriori Modalwert (PM) und Markov Chain Monte Carlo (MCMC) verglichen. Für die Gauß Prozess Interpolation wurde eine lineare Erwartungswertfunktion  $f_M(\mathbf{z}) = (1, z_1, \dots, z_p)^T$  verwendet, die sich in vergangenen Auswertungen ähnlicher Datensätze als geeignet gezeigt hat. Die Ergebnisse sind in Tabelle 3.3 dargestellt. Sie enthält für eine bessere Einordnung auch die Resultate, die sich bei Betrachtung eines linearen Modells mit Regressionsfunktion analog zu der Erwartungswertfunktion bei der Gauß Prozess Interpolation ergeben. Als Vorhersageintervalle wurde das punktweise Konfidenzintervall für den Erwartungswert im linearen Modell berücksichtigt.

Die Überdeckungsgüte für die vorliegenden Daten aus der Rückfederung liegt bei der Auswertung durch die Kreuzvalidierung zwischen 72% und 81%. Somit ist sie akzeptabel, liegt aber unter dem nominalen Niveau von 95%, jedoch deutlich über der Überdeckungsgüte im linearen Modell von 50%. Insgesamt liefert die flache a priori

Tabelle 3.3: Geschätzte Überdeckungswahrscheinlichkeiten der 95% Vorhersageintervalle aus Leave-one-out Kreuzvalidierung für reales Datenbeispiel.

a priori	$\bar{c}_{y_M}$ (MCMC)	$\bar{c}_{y_M}$ (PM)	$\bar{h}_{y_M}$ (MCMC)	$\bar{h}_{y_M}$ (PM)	RMSE (MCMC)	RMSE (PM)
flache	0.810	0.810	3.70	3.67	2.26	2.51
Exponential	0.797	0.722	3.66	3.50	1.97	2.44
Jeffreys	0.785	0.747	4.01	3.73	2.48	2.60
lineares Modell	-	0.506	-	3.98	-	3.46

Verteilung die besten Resultate hinsichtlich der Überdeckung unabhängig davon, ob ein a posteriori Modalwert oder vollständiger Bayes Ansatz mit MCMC Methoden angewendet wird. Bei Verwendung der Exponential- oder Jeffreys a priori Verteilung mit einem vollständigen Bayes Ansatz steigt die Überdeckungsgüte auf 79 bzw. 80% gegenüber von 72 bzw. 74% beim a posteriori Modalwertansatz. Da die Auswertungen in jedem Kreuzvalidierungsschritt auf einem Stichprobenumfang von 75 Beobachtungen bei 6 Dimensionen beruhen, ist mit steigendem Stichprobenumfang kein wesentlicher Gewinn hinsichtlich der Überdeckungsgüte zu erwarten. Wie wir in Kapitel 3.2.2 gesehen haben, können bei größerem Stichprobenumfang auch Abweichungen von der Stationaritätsannahme die Überdeckungsgüte negativ beeinflussen. Hinsichtlich der Vorhersagegüte schneiden Exponential- oder flache a priori mit MCMC Ansatz am besten ab. Aber auch der PM Ansatz liefert hier bei Verwendung der flachen a priori äquivalente Resultate. Im Vergleich zum linearen Modell kann die Vorhersage- und Überdeckungsgüte für die Gauß Prozess Interpolation unabhängig von der gewählten a priori Verteilung deutlich gesteigert werden.

Insgesamt konnte in diesem Kapitel gezeigt werden, dass die Überdeckungs- und Vorhersagegüte bei der Gauß Prozess Interpolation bei ausreichend großem Stichprobenumfang zufriedenstellend ist. Allerdings wird das nominale Niveau der Vorhersageintervalle häufig nicht ganz erreicht. Dies gilt selbst bei großen Stichprobenumfängen, wenn die Funktion  $y_M(\cdot)$  deutlich von einer typischen Realisation eines kovarianzstationären

Gauß Prozesses abweicht. Die schnell berechenbare flache a priori Verteilung  $p^{(1)}$  für den Korrelationsparameter ist eine geeignete objektive a priori Verteilung für die Gauß Prozess Interpolation. Dagegen konnte die Jeffreys a priori als theoretisch begründete objektive a priori Verteilung in der Praxis keinen Gewinn hinsichtlich Überdeckungs- und Vorhersagegüte bei der Gauß Prozess Interpolation aufweisen. Der MCMC Ansatz ist dem PM Ansatz oft überlegen, geht jedoch mit einem erheblich größeren Rechenaufwand einher. Im betrachteten Praxisbeispiel zur Rückfederung bei einem Umformprozess liefert jedoch auch der PM Ansatz gute Ergebnisse.

## 4 KALIBRIERUNG VON COMPUTERMODELLEN

Komplexe Computermodelle enthalten oft eine Vielzahl von Einflussgrößen. Viele der Einflussgrößen können in der Realität eingestellt oder gemessen werden. Dies können zum Beispiel einstellbare Kräfte in einer Umformsimulation oder aber aktuelle Temperatur- und Luftdruckbedingungen bei einer Wettersimulation sein. Andere Einflussgrößen des Computermodells sind typischerweise in der Realität unbekannt. Hierbei kann es sich zum Beispiel um unbekannte physikalische Größen oder aber um numerische Parameter handeln. Für eine Durchführung der Computersimulation sind jedoch auch diese zu wählen. Bei der Kalibrierung von Computermodellen sollen diese unbekanntes Einflussgrößen, die so genannten Kalibrierparameter, anhand von experimentellen Beobachtungen des realen Prozesses optimal gewählt werden. Neben der Parameterschätzung wird außerdem die Entwicklung eines Vorhersagemodells für den realen Prozess durch die gemeinsame Betrachtung von Computereperimenten und physikalischen Experimenten angestrebt. Wie bei jedem Modell können auch bei einem Computermodell Abweichungen zur Realität vorliegen. Bei der Validierung des Computermodells wird die Simulation mit der Realität abgeglichen, um so zu entscheiden, ob das Computermodell ein geeigneter Ersatz für die Realität ist.

Für die Kalibrierung von Computermodellen durch experimentelle Daten werden unterschiedliche Ansätze vorgeschlagen. Cox *et al.* (2001) fassen das Kalibrierungsproblem als ein nichtlineares Regressionsproblem auf und schätzen mit Hilfe von Maximum Likelihood Methoden den Kalibrierparameter. Bei diesem so genannten Tuning Ansatz soll das Computermodell durch eine Punktschätzung des Kalibrierparameters verbessert werden. Ein Nachteil dieses Tuning Ansatzes ist, dass hierbei die Modellannahmen häufig nicht erfüllt sind und es damit zu einer Modellüberanpassung kommt. Ist dieser Ansatz jedoch geeignet, so ist er sehr effizient und bietet dem Anwender eine leicht interpretierbare Lösung. Einen alternativen Ansatz für die Kalibrierung von Computermodellen schlagen Kennedy und O'Hagan (2001) vor. Hierbei wird eine Verzerrung

zwischen Modell und Realität zugelassen, so dass eine möglichst gute Vorhersage des realen Prozesses unter Unsicherheit bezüglich des Kalibrierparameters möglich ist. Erweiterungen hinsichtlich der Validierung und Hinweise für die praktische Umsetzung finden sich in Bayarri *et al.* (2007b). Dieser Ansatz ist deutlich flexibler, bringt jedoch ein Identifizierbarkeitsproblem für den Kalibrierparameter mit sich. In dieser Arbeit in Kapitel 5 wird außerdem ein Kalibrierungsansatz unter Modellunsicherheit entwickelt, der diese beiden Ansätze verknüpft.

Im Folgenden wird in Kapitel 4.1 zunächst die Notation und das allgemeine Modell für die Kalibrierung eingeführt. In Kapitel 4.2 wird die Kalibrierung durch nichtlineare Regression vorgestellt. Die Bayesianische Kalibrierung für verzerrte Computermodelle wird in Kapitel 4.3 beschrieben und abschließend in Kapitel 4.4 ihre Sensitivität gegenüber a priori Verteilungen untersucht.

#### 4.1 Notation und allgemeine Modellformulierung

Sei  $Y_E(\mathbf{x})$  die Zufallsvariable einer Experimentbeobachtung aus dem interessierenden physikalischen Prozess mit Kovariable  $\mathbf{x}$ . Für den Prozess liegen Beobachtungen  $y_{E,i}$  an Designpunkten  $\mathbf{x}_i$ , für  $i = 1, \dots, n$ , vor. Zusammengefasst ergibt sich der Vektor der Beobachtungen  $\mathbf{y}_E = (y_{E,1}, \dots, y_{E,n})^T$  und die Designmatrix  $\mathbf{D}_E = (\mathbf{x}_1, \dots, \mathbf{x}_n)^T$ . Es wird angenommen, dass die Beobachtungen aus einem datengenerierenden Modell der Form

$$Y_E(\mathbf{x}) = \mu_E(\mathbf{x}) + e \tag{4.1}$$

stammen, mit Erwartungswertfunktion  $\mu_E(\mathbf{x})$  und unabhängig normalverteilten Fehler  $e \sim N(0, \sigma_E^2)$ .

Der Zufallsfehler des physikalischen Prozesses folgt laut Modell (4.1) einer verhältnismäßig einfachen Struktur. Es wird angenommen, dass ein unabhängig normalverteilter additiver Fehler mit konstanter Varianz vorliegt. Für den Fehlerterm sind sicher auch komplexere Modelle zum Beispiel mit anderen Verteilungsannahmen und nicht sta-

tionärer Varianz realistisch. Geeignete Transformationen der Zielgrößen sind eventuell erforderlich, um die getroffene Annahme zu rechtfertigen.

Das deterministische Computermodell  $y_M(\mathbf{z})$  soll den gleichen physikalischen Prozess wie  $Y_E(\mathbf{x})$  beschreiben. Der Vektor der Einflussgrößen  $\mathbf{z} = (\mathbf{x}^T, \mathbf{u}^T)^T$  im Computerexperiment lässt sich in die Einflussgrößen des physikalischen Prozesses  $\mathbf{x} = (x_1, \dots, x_{d_x})^T$  und die unbekanntes Kalibrierparameter  $\mathbf{u} = (u_1, \dots, u_{d_u})^T$  zerlegen. Äquivalent lässt sich das Computermodell somit schreiben als  $y_M(\mathbf{z}) = y_M(\mathbf{x}, \mathbf{u})$ . Im Computermodell ist auch der Kalibrierparameter  $\mathbf{u}$  eine Einflussgröße, die in den Computerexperimenten variiert werden kann. In der Realität wird sie jedoch als unbekannter Parameter betrachtet. Der optimale Wert für die Realität wird mit  $\mathbf{u}^*$  bezeichnet. Handelt es sich bei einem Kalibrierparameter  $u_j$  um eine unbekannte physikalische Konstante, so ist  $u_j^*$  der wahre, in der Realität zu Grunde liegende Wert der Konstante. Unter Umständen existiert der Kalibrierparameter  $u_j$  jedoch in der Realität nicht, wie zum Beispiel bei numerischen Parametern für die Simulation. In dieser Situation wird  $u_j^*$  als Tuningparameter definiert: es handelt sich um den am besten passenden Wert für  $u_j$  gemäß eines festzulegenden Kriteriums. Hier wird angenommen, dass  $\mathbf{u}^*$  ein globaler Kalibrierparameter ist. Das heißt es handelt sich um eine Konstante, die damit insbesondere nicht von den übrigen Einflussgrößen  $\mathbf{x}$  abhängt.

Eine gutes deterministisches Computermodell  $y_M(\cdot)$  entspricht im Idealfall der Erwartungswertfunktion  $\mu_E(\cdot)$  des physikalischen Prozesses. Liegt ein Kalibrierparameter vor, so soll die Realität gerade durch die Computersimulation  $y_M(\mathbf{x}, \mathbf{u}^*)$  mit optimalen Wert  $\mathbf{u}^*$  abgebildet werden, das heißt  $\mu_E(\mathbf{x}) = y_M(\mathbf{x}, \mathbf{u}^*)$  für alle  $\mathbf{x} \in \mathcal{X}$ .

Ist das Computermodell nicht schnell berechenbar, so werden Computerexperimente  $\mathbf{y}_M$  an einem Design  $\mathbf{D}_M$  erhoben. Dabei ist  $\mathbf{D}_M = (\mathbf{z}_1, \dots, \mathbf{z}_{n_M})^T$  und  $\mathbf{z}_j = (\mathbf{x}_{M,j}, \mathbf{u}_j)$  mit  $j = 1, \dots, n_M$ , sowie  $\mathbf{y}_M = (y_M(\mathbf{z}_1), \dots, y_M(\mathbf{z}_{n_M}))^T$ . Hier werden neben den physikalischen Einflussgrößen  $\mathbf{x}$  auch die Einflussgrößen  $\mathbf{u}$  variiert.

Computermodelle beschreiben die Realität häufig nicht perfekt. Kennedy und O'Hagan (2001) formulieren deshalb ein Kalibrierungsmodell, dass die Realität in Abhängigkeit eines verzerrten Computermodells beschreibt. Genauer soll die Erwartungswertfunktion der Realität durch eine "Skalen- und Lagentransformation" des Computermodells der

Form

$$\mu_E(\mathbf{x}) = \rho \cdot y_M(\mathbf{x}, \mathbf{u}) + b_{\mathbf{u}}(\mathbf{x})$$

erklärt werden. In der Literatur (Bayarri *et al.*, 2007b, Higdon *et al.*, 2004, Loepky *et al.*, 2006) hat sich die additive Zerlegung von  $\mu_E(\cdot)$  in Computermodell und Verzerrungsfunktion  $b(\cdot)$  durchgesetzt, die auch in dieser Arbeit verfolgt werden soll. Mit Kalibrierparametern wird also eine Zerlegung der Form

$$\mu_E(\mathbf{x}) = y_M(\mathbf{x}, \mathbf{u}) + b_{\mathbf{u}}(\mathbf{x}) \tag{4.2}$$

betrachtet. Dabei hängt die Verzerrungsfunktion  $b_{\mathbf{u}}(\mathbf{x})$  indirekt auch von dem gewählten  $\mathbf{u}$  ab. Alternativ kann für die Funktion  $b_{\mathbf{u}}(\mathbf{x})$  deshalb auch die Notation  $b(\mathbf{x}, \mathbf{u})$  verwendet werden.

Eine systematische Verzerrung der Simulation ist sicher eine realistische Annahme. Allerdings erhöht sich mit ihr die Komplexität der statistischen Modellbildung deutlich. Insbesondere besteht bei Zulassung eines systematischen Verzerrungsterms das Problem der Nichtidentifizierbarkeit des Kalibrierparameters. Für jedes  $\mathbf{u}$  lässt sich ein  $b_{\mathbf{u}}(\mathbf{x})$  finden, so dass Gleichung (4.2) erfüllt ist.

Ist das Ziel die Vorhersage des physikalischen Prozesses, stellt sich die Frage wieso für ein Vorhersagemodell überhaupt Computerexperimente berücksichtigt werden. Alternativ könnte ein flexibles nichtparametrisches Regressionsmodell alleine für experimentell erhobene Daten eingesetzt werden. Allerdings ist die Erhebung der physikalischen Experimente wesentlich zeit- und kostenintensiver als die Durchführung der Computerexperimente. In vielen Situationen gilt dies insbesondere für die Beobachtung des physikalischen Prozesses an unterschiedlichen Werten einer Einflussgröße  $x_j$ , wenn jede Einstellung beispielsweise mit konstruktiven Arbeiten einhergeht. Dahingegen ist die Durchführung von Computerexperimenten auch bei sehr zeitintensiven Simulationen vergleichsweise günstig. Es ist somit zu erwarten, dass wesentlich mehr Beobachtungen aus dem Computermodell als Beobachtungen des physikalischen Prozesses zur Verfü-

gung stehen. Selbst wenn das Computermodell die Realität nicht vollständig beschreibt, so kann das bereits enthaltene Wissen über den physikalischen Prozess in das Vorhersagemodell mit integriert werden. Die Verzerrungsfunktion  $b_{\mathbf{u}}(\cdot)$  besitzt dann eine wesentlich einfachere Struktur als  $\mu_E(\cdot)$  und für die Schätzung der Verzerrungsfunktion sind nur noch vergleichsweise wenige physikalische Experimente erforderlich.

## 4.2 Kalibrierung als nichtlineares Regressionsproblem

Cox *et al.* (2001) schätzen die unbekanntten Einflussgrößen eines Computermodells durch ein modifiziertes nichtlineares Regressionsverfahren. Dabei gehen sie in ihrem Tuning Ansatz davon aus, dass das Computermodell die Realität bereits vollständig beschreibt, d.h.  $b(\mathbf{x}, \mathbf{u}^*) = 0$  ist, und nur noch der optimale Kalibrierparameter  $\mathbf{u}^*$  zu schätzen ist. Damit ist das Modell für den physikalischen Prozess aus Gleichung (4.1) durch

$$Y_E(\mathbf{x}) = y_M(\mathbf{x}, \mathbf{u}^*) + e \quad (4.3)$$

mit  $e \sim N(0, \sigma_E^2)$  gegeben.

Sei die Computersimulation  $y_M$  zunächst schnell berechenbar. Dann ist (4.3) ein nichtlineares Regressionsmodell mit Regressionsfunktion  $y_M(\mathbf{x}, \mathbf{u})$  und nichtlinearem Parameter  $\mathbf{u}$ . Schätzer für  $\mathbf{u}$  und  $\sigma_E^2$  können durch einen Maximum Likelihood Ansatz bestimmt werden. Für das Aufstellen der Likelihood wird  $\mathbf{u}$  als unbekannter Parameter betrachtet und zunächst die hypothetische Experimentdesignmatrix für das Computermodell  $\mathbf{D}_E^{\mathbf{u}} = ((\mathbf{x}_1, \mathbf{u})^T, \dots, (\mathbf{x}_n, \mathbf{u})^T)^T$  definiert, die jeden Designpunkt  $\mathbf{x}_i$  mit dem gleichen  $\mathbf{u}$  kombiniert. Der Erwartungswertvektor für die Experimentdaten  $\mathbf{y}_E$  ist dann gegeben durch  $y_M[\mathbf{D}_E^{\mathbf{u}}] = (y_M(\mathbf{x}_1, \mathbf{u}), \dots, y_M(\mathbf{x}_n, \mathbf{u}))^T$ . Die Log-Likelihood bei bekanntem Computermodell entspricht dann

$$\begin{aligned} L(\mathbf{y}_E; \mathbf{u}, \sigma_E^2) = & -\frac{1}{2}(n \log(2\pi) + n \log(\sigma_E^2)) \\ & + 1/\sigma_E^2 (\mathbf{y}_E - y_M[\mathbf{D}_E^{\mathbf{u}}])^T (\mathbf{y}_E - y_M[\mathbf{D}_E^{\mathbf{u}}]). \end{aligned}$$

Der ML Schätzer  $\widehat{\sigma}_E^2 = \frac{1}{n} (\mathbf{y}_E - y_M[\mathbf{D}_E^u])^T (\mathbf{y}_E - y_M[\mathbf{D}_E^u])$  für  $\sigma_E^2$  lässt sich in Abhängigkeit von  $\mathbf{u}$  geschlossen darstellen, dagegen wird  $\hat{\mathbf{u}}$  durch numerisches Maximieren der Likelihood bestimmt.

Ist das Computermodell nicht schnell auszuwerten und nur für die Beobachtungen  $\mathbf{y}_M$  am Design  $\mathbf{D}_M$  bekannt, modellieren Cox *et al.* (2001)  $y_M(\cdot)$  durch einen Gauß Prozess  $GP(f_M(\cdot)^T \boldsymbol{\beta}_M, \sigma_M^2 R_M(\cdot, \cdot | \boldsymbol{\xi}_M))$ , wie in Kapitel 2. In den Computerexperimenten wird dabei  $\mathbf{u}$  genau wie  $\mathbf{x}$  als Einflussgröße betrachtet und das Computermodell für die Werte  $\mathbf{u}_j$  mit  $j = 1, \dots, n_M$  aus dem Design  $\mathbf{D}_M$  ausgewertet. Für die Kalibrierung ist  $\mathbf{u}$  hingegen ein unbekannter Parameter. Die gemeinsame Verteilung von Experiment und Simulation lässt sich leicht angeben, da jeweils Normalverteilungen angenommen werden. Die Zufallsvariable  $\mathbf{Y}_{\text{ges}}$  zu dem gemeinsamen Vektor von Beobachtungen  $\mathbf{y}_{\text{ges}} = (\mathbf{y}_E^T, \mathbf{y}_M^T)^T$  aus Experiment und Simulation folgt einer multivariaten Normalverteilung  $\mathbf{Y}_{\text{ges}} \sim N(\boldsymbol{\mu}_{\text{ges}}, \boldsymbol{\Sigma}_{\text{ges}})$  mit Erwartungswert

$$\boldsymbol{\mu}_{\text{ges}} = \begin{pmatrix} f_M[\mathbf{D}_E^u] \boldsymbol{\beta}_M \\ \mathbf{F}_M \boldsymbol{\beta}_M \end{pmatrix}$$

und Kovarianzmatrix

$$\boldsymbol{\Sigma}_{\text{ges}} = \begin{pmatrix} \sigma_E^2 \mathbf{I} + \sigma_M^2 R_M[\mathbf{D}_E^u] & \sigma_M^2 R_M[\mathbf{D}_E^u, \mathbf{D}_M] \\ \sigma_M^2 R_M^T[\mathbf{D}_E^u, \mathbf{D}_M] & \sigma_M^2 \mathbf{R}_M \end{pmatrix}.$$

Der Erwartungswert entspricht der Erwartungswertfunktion des Gauß Prozesses für das Computermodell  $f_M(\cdot)^T \boldsymbol{\beta}_M$ . Der Ausdruck  $f_M[\mathbf{D}_E^u]$  steht dabei für die zeilenweise Anwendung der Regressionsfunktion  $f_M$  auf  $\mathbf{D}_E^u = ((\mathbf{x}_1, \mathbf{u}), \dots, (\mathbf{x}_n, \mathbf{u}))^T$ . Die Kovarianzmatrix besteht im Wesentlichen aus der Kovarianzstruktur des Computermodells. Dabei bezeichnet  $R_M[\mathbf{D}_E^u]$  die Korrelationsmatrix, die mit der Korrelationsfunktion  $R_M(\cdot, \cdot)$  für  $\mathbf{D}_E^u$  resultiert und  $R_M[\mathbf{D}_E^u, \mathbf{D}_M]$  gibt die korrespondierende Kreuzkorrelationsmatrix, d.h. die paarweise Korrelation zwischen Zeilenelementen von  $\mathbf{D}_E^u$  und  $\mathbf{D}_M$ , an. Für die experimentelle Streuung der physikalischen Experimente wird die Kovarianzmatrix  $\sigma_E^2 \mathbf{I}$  addiert.

ML Schätzer für die Parameter ergeben sich durch gemeinsames numerisches Maximie-

ren der Likelihood. Dieser vollständige Likelihood Ansatz wird von Cox *et al.* (2001) als „FMLE“ (Full Maximum Likelihood Estimator) Ansatz bezeichnet. Alternativ stellen sie einen modularen „SMLE“ (Separate Maximum Likelihood Estimator) Ansatz vor, bei dem die Resultate aus Simulation und Experiment getrennt betrachtet werden. Dabei werden zunächst alle Parameter für die Modellierung des Computermodells alleine aus den Computereperimenten  $\mathbf{y}_M$  durch Maximum Likelihood Methoden geschätzt. Anschließend werden die Schätzwerte in den a posteriori Gauß Prozess (2.6) bei bekannten Parametern eingesetzt. Die Likelihood im SMLE Ansatz ergibt sich dann durch Einsetzen der Vorhersage  $\hat{y}_M[\mathbf{D}_E^u]$  für  $y_M[\mathbf{D}_E^u]$  und der Kovarianzmatrix  $\hat{\Sigma}_M[\mathbf{D}_E^u]$  für  $\sigma_M^2 R_M[\mathbf{D}_E^u]$  in die Likelihood für  $\mathbf{y}_E$ :

$$L(\mathbf{y}_E; \mathbf{u}, \sigma_E^2) = -\frac{1}{2} \left\{ (\mathbf{y}_E - \hat{y}_M(\mathbf{D}_E^u))^T \left( \sigma_E^2 \mathbf{I} + \hat{\Sigma}_M[\mathbf{D}_E^u] \right)^{-1} (\mathbf{y}_E - \hat{y}_M[\mathbf{D}_E^u]) + n \log(2\pi) + \log(\det(\sigma_E^2 \mathbf{I} + \hat{\Sigma}_M[\mathbf{D}_E^u])) \right\}. \quad (4.4)$$

Maximieren der Likelihood liefert die ML Schätzer für  $\mathbf{u}$  und  $\sigma_E^2$ . Der SMLE Ansatz ist numerisch günstiger, da hier zwar in zwei Schritten jedoch jeweils auf einem niedrigdimensionaleren Raum maximiert wird. Das modulare Vorgehen lässt sich inhaltlich dadurch rechtfertigen, dass eine Verbesserung der Computermodellapproximation durch Experimentdaten in der Praxis unrealistisch ist, da in der Regel deutlich mehr Computereperimente als teure physikalische Experimente vorliegen.

Neben einer Punktschätzung für den Kalibrierparameter ist auch eine Unsicherheitsregion von Interesse. Cox *et al.* (2001) verwenden eine approximative Konfidenzregion auf Basis einer Taylor Entwicklung zweiten Grades der Residuenfunktion um  $\hat{\mathbf{u}}$ . Die Konfidenzregion wird über die approximative asymptotische Normalverteilung des ML Schätzers  $\hat{\mathbf{u}} \underset{approx}{\sim} N(\mathbf{u}^*, \mathbf{C})$  bestimmt. Dabei ist die Kovarianzmatrix  $\mathbf{C} = \left[ \frac{\partial^2 -L(\mathbf{y}_E; \mathbf{u}, \hat{\sigma}_E^2)}{\partial \mathbf{u} \partial \mathbf{u}^T} \right]_{\mathbf{u}=\hat{\mathbf{u}}}$  die Inverse der Hessematrix für die negative Log-Likelihood an der Stelle des ML Schätzers.

Diese Approximation ist bei multimodaler Likelihood weniger gut geeignet. Dann ist die Konfidenzregion auf Basis eines Likelihood Ratio Tests zu bevorzugen, da diese nicht von der Krümmung der Likelihood in den Parametern abhängt (Seber und Wild, 2003,

S.194ff). Eine asymptotische  $(1 - \alpha) \cdot 100$  % Konfidenzregion für den  $d_u$ -dimensionalen Parameter  $\mathbf{u}$  mit der konzentrierten Likelihood  $L(\mathbf{y}_E; \mathbf{u}, \hat{\sigma}_E^2(\mathbf{u}))$  ist durch

$$K(\mathbf{u}) = \{\mathbf{u} : 2[L(\mathbf{y}_E; \hat{\mathbf{u}}, \hat{\sigma}_E^2(\hat{\mathbf{u}})) - L(\mathbf{y}_E; \mathbf{u}, \hat{\sigma}_E^2(\mathbf{u}))] \leq \chi_{d_u}^2(1 - \alpha)\} \quad (4.5)$$

gegeben. Dabei ist  $\hat{\sigma}_E^2(\mathbf{u})$  der ML Schätzer für  $\sigma_E^2$  bei gegebenem  $\mathbf{u}$ . Diese Konfidenzregion geht jedoch mit einem erheblich größeren Rechenaufwand einher, da die Likelihood für den gesamten Parameterraum auszuwerten ist.

Das weitere praktische Vorgehen bei diesem Kalibrierungsansatz ist wie folgt. Mit  $\hat{\mathbf{u}}$  kann der Anwender eine einzelne am besten passende Simulation auswählen und diese zum Beispiel für eine Optimierung oder Übertragung auf geringfügig modifizierte Probleme nutzen. Mit festem  $\mathbf{u} = \hat{\mathbf{u}}$  reduziert sich die Zahl der erforderlichen Computermodellauswertungen wesentlich. Neben der Punktschätzung sollte jedoch auch die Unsicherheit für  $\mathbf{u}$  berücksichtigt werden. Für eine Vorhersage des physikalischen Prozesses ist dabei weniger die Unsicherheit über  $\mathbf{u}$  als über das Computermodell von Interesse. Ist für ein festes  $\mathbf{x}$  die Variation von  $y_M(\mathbf{x}, \cdot)$  für Punkte  $\mathbf{u}$  aus der Konfidenzregion  $K(\mathbf{u})$  klein, so besteht dort nur wenig Unsicherheit über das Computermodell. Dahingegen kann das Computermodell  $y_M(\mathbf{x}', \cdot)$  für ein anderes  $\mathbf{x}'$  eine größere Funktionsvariation aufweisen und damit liegt dort eine größere Unsicherheit über das Computermodell vor. Dieser Aspekt kann in der Praxis für eine robuste Optimierung im Sinne einer Varianzreduktion genutzt werden, erfordert jedoch einen größeren Rechenaufwand als die alleinige Betrachtung des Punktschätzers  $\hat{\mathbf{u}}$ .

#### 4.2.1 Verhalten der nichtlinearen Regression bei verzerrtem Computermodell

Die zentrale Annahme bei der Kalibrierung von Computermodellen durch nichtlineare Regression ist ein unverzerrtes Computermodell. Das heißt, es soll ein  $\mathbf{u}^* \in \mathcal{U}$  mit  $\mu_E(\mathbf{x}) = y_M(\mathbf{x}, \mathbf{u}^*)$  für alle  $\mathbf{x} \in \mathcal{X}$  existieren. Im Folgenden wird das Verhalten der Methode bei verzerrtem Computermodell untersucht.

Die Erwartungswertfunktion  $\mu_E(\cdot)$  des physikalischen Prozesses lässt sich nun nicht länger alleine durch das Computermodell beschreiben, sondern es wird eine additive Verzerrung wie in (4.2) zugelassen, d.h.  $\mu_E(\mathbf{x}) = y_M(\mathbf{x}, \mathbf{u}^*) + b(\mathbf{x}, \mathbf{u}^*)$ . Ganz allgemein lässt sich  $b(\mathbf{x}, \mathbf{u}) = \mu_E(\mathbf{x}) - y_M(\mathbf{x}, \mathbf{u})$  für beliebige  $\mathbf{x}$  und  $\mathbf{u}$  definieren. Insgesamt wird nun also die Modellgleichung

$$Y_E^*(\mathbf{x}) = y_M(\mathbf{x}, \mathbf{u}) + b(\mathbf{x}, \mathbf{u}) + e^* \quad (4.6)$$

mit  $e^* \sim N(0, \sigma_E^{2*})$  betrachtet. Dabei werden  $Y_E(\mathbf{x})$  und  $\sigma_E^2$  in diesem Kapitel mit einem Stern gekennzeichnet, um deutlich zwischen dem Modell mit und ohne Verzerrung unterscheiden zu können. Der ML Schätzer für die Kalibrierung mit nichtlinearem Regressionsansatz ergibt sich durch Einsetzen der Zufallsvariable  $\mathbf{Y}_E^*$  in die Log-Likelihood aus (4.4).

Bei vollständig bekanntem Computermodell ergibt sich

$$L(\mathbf{y}_E = \mathbf{Y}_E^*; \mathbf{u}, \sigma_E^2) \propto - \left\{ n \log(\sigma_E^2) + \frac{1}{\sigma_E^2} (b(\mathbf{D}_E \mathbf{u}) + \mathbf{e}^*)^T (b(\mathbf{D}_E \mathbf{u}) + \mathbf{e}^*) \right\}.$$

Um Herauszufinden wie sich die Likelihood bei verzerrtem Computermodell im Mittel verhält, wird die erwartete Likelihood, d.h. der Erwartungswert der Log-Likelihood bezüglich  $\mathbf{Y}_E^*$ , betrachtet. Diese ist gegeben durch

$$E(L(\mathbf{y}_E = \mathbf{Y}_E^*; \mathbf{u}, \sigma_E^2)) \propto - \left\{ n \log(\sigma_E^2) + \frac{1}{\sigma_E^2} b(\mathbf{D}_E \mathbf{u})^T b(\mathbf{D}_E \mathbf{u}) + \frac{n \sigma_E^{2*}}{\sigma_E^2} \right\}.$$

Das Maximum der erwarteten Likelihood bezüglich  $\mathbf{u}$  lässt sich durch Minimieren der quadratischen Verzerrung  $b(\mathbf{D}_E \mathbf{u})^T b(\mathbf{D}_E \mathbf{u})$  finden. Es wird also der  $L_2$  Abstand zwischen Computermodell und Realität an den Designpunkten minimiert. Für  $\sigma_E^2$  liegt das Maximum der erwarteten Likelihood für  $\hat{\sigma}_E^2 = \sigma_E^{2*} + \frac{1}{n} (b(\mathbf{D}_E \mathbf{u})^T b(\mathbf{D}_E \mathbf{u}))$  vor. Diese Schätzung entspricht der wahren experimentellen Varianz  $\sigma_E^{2*}$  erweitert um die mittlere quadratische Verzerrung. Liegen mehrere Wiederholungen im Experimentdesign vor, so ist eine Beurteilung der Computermodellgüte somit durch den Vergleich des Fehlervarianzschätzers aus dem Kalibrierungsmodell mit dem auf Wiederholungen basierenden

Varianzschätzer  $s_p^2 = \frac{\sum_{j=1}^k (n_j-1) s_j^2}{\sum_{j=1}^k (n_j-1)}$  mit  $s_j^2 = \frac{1}{(n_j-1)} \sum_{i=1}^{n_j} (y_{E,i}(\mathbf{x}_j) - \bar{y}_E(\mathbf{x}_j))^2$  möglich (Kracker und Kuhnt, 2009, siehe auch).

Bei unbekanntem Computermodell und Anwendung des SMLE Ansatzes ist die erwartete Likelihood proportional zu

$$\begin{aligned} E(L(\mathbf{y}_E = \mathbf{y}_E^*; \mathbf{u}, \sigma_E^2)) \propto & - \left\{ \log(\det(\widehat{\Sigma}_E)) + \sigma_E^{2*} \text{tr}(\widehat{\Sigma}_E^{-1}) \right. \\ & \left. + (\mathbf{b}_M + b(\mathbf{D}_E \mathbf{u}))^T \widehat{\Sigma}_E^{-1} (\mathbf{b}_M + b(\mathbf{D}_E \mathbf{u})) \right\} \end{aligned}$$

mit  $\widehat{\Sigma}_E = \widehat{\Sigma}_M(\mathbf{D}_E \mathbf{u}) + \sigma_E^2 \mathbf{I}$  und  $\mathbf{b}_M = y_M(\mathbf{D}_E \mathbf{u}) - \hat{y}_M(\mathbf{D}_E \mathbf{u})$ . Für eine Maximierung der erwarteten Likelihood nach  $\mathbf{u}$  ist auch hier im wesentlichen die quadratische Abweichung von Computermodell und Realität zu minimieren. Diese wird jedoch mit der Inversen Kovarianzmatrix  $\widehat{\Sigma}_E^{-1}$  gewichtet. Bei großer Unsicherheit über das Computermodell werden somit Abweichungen zwischen Simulation und Realität weniger stark gewichtet. Die Maximalstelle der erwarteten Likelihood bezüglich  $\sigma_E^2$  lässt sich nicht direkt angeben. Sie wird jedoch auch hier durch die Verzerrung eher überschätzt.

Insgesamt wird bei einem verzerrten Computermodell mit diesem Kalibrierungsansatz der Kalibrierparameter so gewählt, dass die mittlere quadratische Verzerrung minimiert und die Fehlervarianz eher überschätzt wird. Ist das Ziel eine Schätzung des Kalibrierparameters, so liefert dieser Ansatz, insbesondere bei vielen gleichmäßig verteilten Beobachtungen, eine sinnvolle Lösung: die Schätzung des Kalibrierparameters mit kleinstem  $L_2$  Abstand von Simulation und Realität. In Kapitel 6 wird jedoch in einem praktischen Beispiel illustriert, dass die Kalibrierung durch die Computermodellunsicherheit gestört werden kann. Hilfreich ist eine möglichst gleichmäßig große Unsicherheit über das Computermodell im Versuchsraum, wie sie mit einem raumfüllenden Versuchsplan erreicht wird. Als Vorhersagemodell ist dieser Kalibrierungsansatz bei verzerrtem Computermodell jedoch nur bedingt geeignet. Selbst mit möglichst geeignetem Kalibrierparameter werden Vorhersagen, die alleine auf dem Computermodell basieren, stets eine systematische Verzerrung aufweisen.

### 4.3 Bayesianische Kalibrierung von Computermodellen bei Verzerrung

Kennedy und O’Hagan (2001) schlagen für die Kalibrierung von Computermodellen einen Bayes Ansatz auf Basis von Gauß Prozessen vor, der explizit eine Computermodellverzerrung zulässt. Dieser Ansatz wird von einer Reihe von Artikeln aufgegriffen, modifiziert und erweitert (Bayarri *et al.*, 2007a,b, Higdon *et al.*, 2004, Loepky *et al.*, 2006). Dabei ist nicht nur die Schätzung eines optimalen Kalibrierparameters  $\mathbf{u}^*$  von Interesse, sondern insbesondere die Erstellung eines guten Vorhersagemodells für den physikalischen Prozess. Weiterhin soll die Verzerrung quantifiziert werden, um so eine Validierung des Computermodells und Interpretation der Verzerrung zu ermöglichen. Dazu werden Vorhersagen und Vorhersageintervalle für Computersimulation, Verzerrung und Experiment benötigt.

Es wird wieder der physikalische Prozess  $Y_E(\mathbf{x}) = \mu_E(\mathbf{x}) + e$  aus Gleichung (4.1) mit unabhängigem Fehler  $e \sim N(0, \sigma_E^2)$  und Varianz  $\sigma_E^2$  betrachtet. Die Erwartungswertfunktion  $\mu_E(\cdot)$  wird hier wie in Gleichung (4.2) als Summe der Computermodellfunktion  $y_M(\mathbf{x}, \mathbf{u})$  mit Kalibrierparameter  $\mathbf{u}$  und einer Verzerrungsfunktion modelliert:  $\mu_E(\mathbf{x}) = y_M(\mathbf{x}, \mathbf{u}) + b_{\mathbf{u}}(\mathbf{x})$ . Dabei ist die Verzerrungsfunktion  $b_{\mathbf{u}}(\mathbf{x})$  eine Funktion von  $\mathbf{x}$ , die indirekt natürlich auch von  $\mathbf{u}$  abhängt. Es wird a priori angenommen, dass sie durch einen Gauß Prozess  $\tilde{b}(\cdot)$  mit konstantem Erwartungswert  $\mu_b$ , Kovarianzfunktion  $C_b(\cdot, \cdot) = \sigma_b^2 R_b(\cdot, \cdot)$ , Prozessvarianz  $\sigma_b^2$  und Gauß Korrelationsfunktion  $R_b(\cdot, \cdot | \boldsymbol{\psi}_b)$  beschrieben werden kann:

$$\tilde{b}(\cdot) \sim GP(\mu_b \mathbf{1}, C_b(\cdot, \cdot | \sigma_b^2, \boldsymbol{\psi}_b)). \quad (4.7)$$

Dies entspricht der a priori Annahme einer glatten Verzerrungsfunktion, die sich als Realisation eines stationären Gauß Prozesses beschreiben lässt und gleichmäßig um  $\mu_b$  variiert. In der Regel wird  $\mu_b = 0$  gesetzt, da für das Computermodell a priori keine systematische Verzerrung zu erwarten ist. Es ist jedoch zu beachten, dass bei der Modellierung durch einen Gauß Prozess der a posteriori Erwartungswert für die Verzerrung natürlich von Null verschieden sein kann.

Der Kalibrierparameter  $\mathbf{u}$  ist in dem Modell mit Verzerrung aus (4.2) im klassischen Sinne nicht identifizierbar, da für beliebiges  $\mathbf{u}$  eine Verzerrungsfunktion  $b_{\mathbf{u}}(\cdot)$  gefunden werden kann, so dass  $\mu_E(\mathbf{x}) = y_M(\mathbf{x}, \mathbf{u}) + b_{\mathbf{u}}(\mathbf{x})$  gilt. Erst die gewählten a priori Annahmen für die Verzerrungsfunktion machen eine Schätzung des Kalibrierparameters möglich. Mit  $\mu_b = 0$  werden also solche Werte für  $\mathbf{u}$  bevorzugt, die in einer glatten um Null schwankenden Verzerrungsfunktion  $b_{\mathbf{u}}(\mathbf{x})$  resultieren.

Ist das Computermodell nicht schnell berechenbar, so wird wie in Kapitel 2 auch für  $y_M(\cdot)$  a priori ein Gauß Prozess  $\tilde{y}_M(\cdot)$  angenommen. Dieser besitzt eine lineare Erwartungswertfunktion  $f_M(\cdot)\boldsymbol{\beta}$  und Kovarianzfunktion  $C_M(\cdot, \cdot) = \sigma_M^2 R_M(\cdot, \cdot | \boldsymbol{\psi}_M)$  mit Power Exponential Korrelationsfunktion  $R_M(\cdot, \cdot | \boldsymbol{\psi}_M)$  und wird als unabhängig von dem Gauß Prozess zur Beschreibung der Verzerrungsfunktion  $b_{\mathbf{u}}(\cdot)$  betrachtet.

Das Kalibrierungsmodell von Kennedy und O'Hagan (2001) und Bayarri *et al.* (2007b) ist insgesamt wesentlich flexibler als die nichtlineare Regression aus Kapitel 4.2 von Cox *et al.* (2001). Neben der Computermodellfunktion wird hier auch die Verzerrungsfunktion flexibel durch einen Gauß Prozess modelliert. Der physikalische Prozess kann dadurch in der Regel gut beschrieben werden.

Das vorgestellte Modell zur Kalibrierung soll nun an die Beobachtungen  $\mathbf{y}_M$  und  $\mathbf{y}_E$  aus Computermodell und physikalischen Prozess angepasst werden. Für die Auswertung orientiere ich mich im wesentlichen an Bayarri *et al.* (2007b), wo die nachfolgenden Formeln zur Herleitung der a posteriori gefunden werden können.

### *Suffiziente Statistiken*

Um sowohl die Prozessvarianz  $\sigma_b^2$  der Verzerrung als auch die experimentelle Streuung  $\sigma_E^2$  schätzen zu können sind Wiederholungen der physikalischen Experimente erforderlich. Für die Analyse können dann statt  $\mathbf{y}_E$  suffiziente Statistiken betrachtet werden. Dadurch wird auch die Designmatrix ohne Wiederholungen  $\boldsymbol{\Delta}_E$  statt der vollständigen Designmatrix  $\mathbf{D}_E$  betrachtet. Durch eine Reduktion der Dimension der Designmatrix reduziert sich die Dimension der Korrelationsmatrix für den Gauß Prozess und damit die Rechenkomplexität (Bayarri *et al.*, 2007b).

Sei  $\Delta_E = (\boldsymbol{\delta}_1, \dots, \boldsymbol{\delta}_N)$  die Matrix mit  $N$  unterschiedlichen Experimentdesignpunkten. Es handelt sich um die Designmatrix  $\mathbf{D}_E$  reduziert um die Zeilenwiederholungen. In dem vollständigen Design  $\mathbf{D}_E$  liegt jedes Element  $\boldsymbol{\delta}_i$  jeweils  $n_i$ ,  $i = 1, \dots, N$ , mal vor. Die suffizienten Statistiken sind für jedes  $\boldsymbol{\delta}_i$  durch den Mittelwert  $\bar{\mathbf{y}}_{E,i} = \frac{1}{n_i} \sum_j^{n_i} y_{E,j}(\boldsymbol{\delta}_i)$  und Summe der quadratischen Abweichungen  $s_{E,i}^2 = \sum_j^{n_i} (y_{E,j}(\boldsymbol{\delta}_i) - \bar{\mathbf{y}}_{E,i})^2$  gegeben und werden in dem Vektor  $\bar{\mathbf{y}}_E = (\bar{y}_{E,1}, \dots, \bar{y}_{E,N})^T$  und der Summe  $\mathbf{s}_E^2 = \sum_{i=1}^N s_{E,i}^2$  zusammengefasst.

Die Verteilung des Vektors der Mittelwerte  $\bar{\mathbf{y}}_E$  ist nun gegeben durch

$$\bar{\mathbf{y}}_E \sim N(y_M(\Delta_E^u) + b_{\mathbf{u}}(\Delta_E), \bar{\mathbf{S}}_E). \quad (4.8)$$

Der Erwartungswert  $y_M(\Delta_E^u) + b_{\mathbf{u}}(\Delta_E)$  entspricht der Summe aus Computermodell und Verzerrung für das Experimentdesign  $\Delta_E$ , bzw. der Kombination mit festem  $\mathbf{u}$ :  $\Delta_E^u = ((\boldsymbol{\delta}_1, \mathbf{u}), \dots, (\boldsymbol{\delta}_N, \mathbf{u}))^T$ , und die Kovarianzmatrix  $\bar{\mathbf{S}}_E$  einer Diagonalmatrix mit  $i$ -tem Diagonalelement  $\frac{\sigma_E^2}{n_i}$ , für  $i = 1, \dots, n$ . Für jede Quadratsumme gilt  $\frac{s_{E,i}^2}{\sigma_E^2} \sim \chi_{n_i-1}^2$ , dabei ist  $\chi_\nu^2$  eine  $\chi^2$ -Verteilung mit  $\nu$  Freiheitsgraden. Entsprechend besitzt die Summe der unabhängigen Quadratsummen eine Gammaverteilung

$$\frac{s_E^2}{\sigma_E^2} \sim \chi_{\sum_i^N n_i - 1}^2 \hat{=} \Gamma(a_1 = \frac{1}{2} \sum_{i=1}^N n_i - 1, a_2 = 2), \quad (4.9)$$

mit Formparameter  $a_1$  und Skalenparameter  $a_2$ .

### *Likelihood*

Sei das Computermodell  $y_M(\cdot)$  zunächst vollständig bekannt. Dann ergibt sich mit dem Gauß Prozess für  $b_{\mathbf{u}}(\cdot)$

$$Y_E(\cdot) \sim GP(y_M(\cdot, \mathbf{u}) + \mu_b \mathbf{1}, \sigma_b^2 R_b(\cdot, \cdot)) \quad (4.10)$$

auch ein Gauß Prozess für  $Y_E(\cdot)$ , bei gegebenem Parametern  $\mathbf{u}$ ,  $\sigma_E^2$  und  $\boldsymbol{\theta}_b = (\mu_b, \sigma_b^2, \boldsymbol{\psi}_b^T)^T$ . Die Likelihood für den Beobachtungsvektor  $\mathbf{y}_E$  entspricht damit

$$l(\mathbf{y}_E | \mathbf{u}, \sigma_E^2, \boldsymbol{\theta}_b) = \varphi(\mathbf{y}_E | y_M(\mathbf{D}_E^{\mathbf{u}}) + b_{\mathbf{u}}(\mathbf{D}_E), \sigma_E^2 \mathbf{I} + \sigma_b^2 R_b(\mathbf{D}_E)). \quad (4.11)$$

Dabei bezeichnet die Schreibweise  $\varphi(\mathbf{y} | \boldsymbol{\mu}, \boldsymbol{\Sigma})$  die Dichte der multivariaten Normalverteilung mit Erwartungswert  $\boldsymbol{\mu}$  und Kovarianzmatrix  $\boldsymbol{\Sigma}$  an der Stelle  $\mathbf{y}$ . Bei  $R_b(\mathbf{D}_E)$  handelt es sich um die aus der Korrelationsfunktion  $R_b(\cdot, \cdot)$  resultierende Korrelationsmatrix für  $\mathbf{D}_E$ . In Abhängigkeit der suffizienten Statistiken ist die Likelihood gegeben durch

$$l(\bar{\mathbf{y}}_E, s_E^2 | \mathbf{u}, \sigma_E^2, \boldsymbol{\theta}_b) = \varphi(\bar{\mathbf{y}}_E | y_M(\mathbf{D}_E^{\mathbf{u}}) + \mu_b \mathbf{1}, \bar{\mathbf{S}}_E + \sigma_b^2 R_b(\Delta_E)) \cdot p(s_E^2 | \sigma_E^2). \quad (4.12)$$

Ist die Computermodellfunktion nur für  $\mathbf{y}_M$  am Design  $\mathbf{D}_M$  bekannt, wird sie durch einen Gauß Prozess mit Hyperparametern  $\boldsymbol{\theta}_M = (\boldsymbol{\beta}_M^T, \sigma_M^2, \boldsymbol{\psi}_M^T)^T$  modelliert. Dann lässt sich die gemeinsame Likelihood für physikalische Beobachtungen und Computerexperimente in faktorisierter Form angeben:

$$l(\bar{\mathbf{y}}_E, s_E^2, \mathbf{y}_M | \mathbf{u}, \sigma_E^2, \boldsymbol{\theta}_b, \boldsymbol{\theta}_M) = p(s_E^2 | \sigma_E^2) \cdot \varphi(\mathbf{y}_M | \mathbf{F}_M \boldsymbol{\beta}_M, \sigma_M^2 \mathbf{R}_M) \cdot \varphi(\bar{\mathbf{y}}_E | \hat{y}_M(\Delta_E^{\mathbf{u}}) + \mu_b \mathbf{1}, \hat{\boldsymbol{\Sigma}}_M(\Delta_E^{\mathbf{u}}) + \sigma_b^2 R_b(\Delta_E) + \bar{\mathbf{S}}_E). \quad (4.13)$$

Regressionsmatrix  $\mathbf{F}_M$  und Korrelationsmatrix  $\mathbf{R}_M$  für das ComputermodeLL am Design  $\mathbf{D}_M$  sowie die Vorhersage  $\hat{y}_M(\cdot)$  mit zugehöriger Kovarianz  $\hat{\boldsymbol{\Sigma}}_M(\cdot)$  sind wie in Kapitel 2 definiert.

### *A priori Verteilungen und ML Schätzer für Hyperparameter*

Für die Hyperparameter des Kalibrierungsmodells sind in einem Bayes Ansatz a priori Verteilungen zu wählen. Im Folgenden werden die a priori Verteilungen aus Bayarri *et al.* (2007b) vorgestellt. Für den Kalibrierparameter  $\mathbf{u}$  wird die multivariate Gleichverteilung auf dem Wertebereich  $\mathcal{U}$  als a priori Verteilung verwendet, wenn keine weiteren

Informationen vorliegen. Besitzt der Parameter eine physikalische Interpretation und liegen zum Beispiel Informationen aus Vorversuchen vor, so ist auch eine informative a priori Verteilung geeignet. Für die Hyperparameter der Verzerrungsfunktion liegen in der Regel keine a priori Informationen vor, insbesondere da die Verzerrungsfunktion nicht beobachtet werden kann oder direkt physikalisch interpretierbar ist. Bayarri *et al.* (2007b) empfehlen die Verwendung von datenabhängigen a priori Verteilungen für  $\sigma_b^2$  und  $\sigma_E^2$ . Dies soll der Vermischung von  $b_{\mathbf{u}}(\cdot)$  und  $\mathbf{u}$  entgegen steuern und so die numerischen Berechnungen stabilisieren. Dazu wird eine Inverse Exponentialverteilung, d.h. eine Exponentialverteilung für die Präzisionsparameter  $\lambda_b = 1/\sigma_b^2$  und  $\lambda_E = 1/\sigma_E^2$  verwendet. Als Erwartungswert wird datenabhängig ein Vielfaches (z.B. das fünffache) des ML Schätzers von  $\lambda_b$  beziehungsweise  $\lambda_E$  eingesetzt. Alternative a priori Verteilungen für  $\sigma_b^2$  und  $\sigma_E^2$  wie die Inverse Gammaverteilung, die quadrierte  $t$ -Verteilung oder auch a priori Verteilungen mit Punktmasse auf Null werden in Kapitel 4.4.2 und 5 vorgestellt und hinsichtlich der Resultate für die Computermodellkalibrierung untersucht. Der Korrelationsparameter  $\boldsymbol{\psi}_b$  für den Gauß Prozess der Verzerrungsfunktion wird fest auf seinen ML Schätzer gesetzt.

Der ML Schätzer wird in Bayarri *et al.* (2007b) wie folgt bestimmt. Es wird ein plausibles  $\tilde{\mathbf{u}}$  gewählt und der Vektor der „beobachteten“ Verzerrung  $\tilde{\mathbf{b}} := \mathbf{y}_E - y_M(\mathbf{D}_E^{\tilde{\mathbf{u}}})$  berechnet. An  $\tilde{\mathbf{b}}$  wird ein Gauß Prozess mit Zufallsfehler durch ML Methoden angepasst (siehe Kapitel 2.4). Es resultieren Schätzer für die Parameter  $\psi_b$ ,  $\sigma_b^2$ ,  $\sigma_E^2$  und gegebenenfalls  $\mu_b$ . In der Kalibrierung wird  $\psi_b$  auf seinen ML Schätzer gesetzt. Für die Varianzparameter gehen die ML Schätzer in die datenabhängigen a priori Verteilungen ein. Der Erwartungswert der Verzerrung  $\mu_b$  wird in der Regel gleich Null gesetzt, wie bereits diskutiert. Andernfalls wird die a priori Verteilung  $p(\mu_b) \propto 1$  gewählt.

Der Einfluss von  $\tilde{\mathbf{u}}$  und dem Festsetzen von  $\boldsymbol{\psi}_b$  ist laut Bayarri *et al.* (2007b) sehr klein. Für den Korrelationsparameter  $\boldsymbol{\psi}_b$  konnte dies in dieser Arbeit bereits für die Gauß Prozess Interpolation in Kapitel 3 zum Teil bestätigt werden. Ein vollständiger Bayes Ansatz für den Parameter  $\boldsymbol{\psi}_b$  geht mit einem großen Rechenaufwand einher. Da im Kalibrierungsmodell numerische Methoden für die Herleitung von a posteriori Verteilungen erforderlich sind, ist eine zusätzliche Erhöhung der numerischen Komplexität

durch Betrachtung eines variablen  $\psi_b$  zu vermeiden.

Wird auch das Computermodell durch einen Gauß Prozess approximiert, liegen mit  $\beta_M, \sigma_M^2$  und  $\psi_M$  weitere Hyperparameter vor. Für die Kalibrierung empfehlen Bayarri *et al.* (2007b) eine modulare Vorgehensweise statt eines vollständigen Bayes Ansatzes. Dabei werden die Parameter  $\beta_M, \sigma_M^2$  und  $\psi_M$  alleine auf Basis der Computerexperimente  $\mathbf{y}_M$  durch ML Methoden geschätzt und anschließend für die Kalibrierung als bekannt angenommen. In die Kalibrierung geht das Computermodell dann nur über den a posteriori Gauß Prozess bei bekannten Parametern aus (2.6) ein. Dieses modulare Vorgehen entspricht dem separaten Likelihood Ansatz bei der Kalibrierung durch nichtlineare Regression aus (4.4). Das Einsetzen von ML Schätzern vernachlässigt natürlich die Unsicherheit über das Computermodell, vereinfacht jedoch die Analyse des Kalibrierungsmodells wesentlich, da es sich bei der resultierenden Vorhersageverteilung wieder um einen Gauß Prozess handelt. Dadurch lassen sich bedingte Verteilungen über das Resultat (2.5) herleiten. Dieses Vorgehen lässt sich wie schon bei der Kalibrierung in Kapitel 4.2 dadurch rechtfertigen, dass in der Praxis mehr Simulations- als Experimentdaten vorliegen werden und somit keine zusätzlichen Informationen über die Simulation aus den physikalischen Experimenten zu erwarten ist. Außerdem wird durch dieses schrittweise Vorgehen die Nichtidentifizierbarkeit der einzelnen Modellkomponenten in (4.2) nicht weiter verschärft, da die Approximation des Computermodells nicht durch Experimentdaten verzerrt wird. Die für die Kalibrierung besonders wichtigen Parameter  $\mathbf{u}, \sigma_b^2$  und  $\sigma_E^2$  werden auch hier als unbekannt betrachtet.

### *A posteriori Verteilung*

Die a posteriori Verteilung für  $Y_E(\cdot)$  ist in einem vollständigen Bayes Ansatz ganz allgemein gegeben durch

$$p(Y_E(\cdot)|\mathbf{y}_E, \mathbf{y}_M) = \int p(Y_E(\cdot)|\mathbf{u}, \sigma_E^2, \boldsymbol{\theta}_b, \boldsymbol{\theta}_M) p(\mathbf{u}, \sigma_E^2, \boldsymbol{\theta}_b, \boldsymbol{\theta}_M | \mathbf{y}_E, \mathbf{y}_M) d(\mathbf{u}, \sigma_E^2, \boldsymbol{\theta}_b, \boldsymbol{\theta}_M),$$

d.h. durch Integrieren bezüglich der a posteriori Verteilung für die Parameter  $p(\mathbf{u}, \sigma_E^2, \boldsymbol{\theta}_b, \boldsymbol{\theta}_M | \mathbf{y}_E) \propto l(\mathbf{y}_E | \mathbf{u}, \sigma_E^2, \boldsymbol{\theta}_b, \boldsymbol{\theta}_M) p(\mathbf{u}, \sigma_E^2, \boldsymbol{\theta}_b, \boldsymbol{\theta}_M)$ . In dem hier gewählten em-

pirischen Bayes Ansatz werden für die Parameter  $\boldsymbol{\theta}_M$  des Gauß Prozesses zur Beschreibung des Computermodells und den Korrelationsparameter  $\boldsymbol{\psi}_b$  des Gauß Prozesses zur Beschreibung der Verzerrungsfunktion jeweils ML Schätzer eingesetzt und anschließend als bekannt betrachtet. Somit ist hier die a posteriori Verteilung

$$p(Y_E(\cdot)|\mathbf{y}_E, \mathbf{y}_M) = \int p(Y_E(\cdot)|\mathbf{u}, \sigma_E^2, \sigma_b^2, \mu_b)p(\mathbf{u}, \sigma_E^2, \sigma_b^2, \mu_b|\mathbf{y}_E, \mathbf{y}_M)d(\mathbf{u}, \sigma_E^2, \sigma_b^2, \mu_b)$$

herzuleiten. Eine geschlossene Darstellung dieser a posteriori Verteilung existiert nicht. Insbesondere für den Kalibrierparameter  $\mathbf{u}$ , der einem nichtlinearen Parameter in einem nichtlinearen Regressionsmodell entspricht (siehe Kapitel 4.2), müssen stets numerische Methoden, wie z.B. MCMC Methoden, zur Bestimmung einer a posteriori Stichprobe angewendet werden. Das allgemeine Vorgehen ist dabei wie folgt. Zunächst werden Zufallszahlen aus der a posteriori Verteilung der Parameter  $\mathbf{u}, \sigma_E^2, \sigma_b^2$  und  $\mu_b$  erzeugt. Anschließend werden für jedes a posteriori Stichprobenelement Zufallszahlen aus  $\tilde{b}(\mathbf{x})$  und  $\tilde{y}_M(\mathbf{x}, \cdot)$  gezogen, um so ein a posteriori Stichprobenelement für  $y_E(\mathbf{x})$  zu erhalten. Die a posteriori für die Parameter ist hier proportional zu

$$p(\mathbf{u}, \sigma_E^2, \sigma_b^2, \mu_b|\mathbf{y}_E, \mathbf{y}_M) \propto l(\bar{\mathbf{y}}_E, s_E^2, \mathbf{y}_M|\mathbf{u}, \sigma_E^2, \sigma_b^2, \mu_b)p(\mathbf{u})p(\sigma_E^2)p(\sigma_b^2)p(\mu_b) \quad (4.14)$$

mit der Likelihood aus (4.13) und den zuvor beschriebenen a priori Verteilungen für die Hyperparameter.

Neben der Modellformulierung ist die Implementierung für solch ein komplexes Modell in der Bayes Statistik entscheidend. Dabei stehen verschiedene Ansätze zur Erzeugung von Zufallszahlen aus der a posteriori Verteilung zur Auswahl. In dieser Arbeit werden einerseits MCMC (Markov Chain Monte Carlo) Methoden und Importance Sampling mit iterativer Laplace Approximation angewendet. Diese numerischen Methoden werden im Anhang A.1 kurz beschrieben.

Für die Analyse durch MCMC Algorithmen, wie durch Gibbs Sampling oder den Metropolis-Hastings Algorithmus, werden nun die verwendeten Vorschlagsdichten angegeben. Für das Importance Sampling mit iterativer Laplace Approximation wird die

Strategie zur Wahl geeigneter Startwerte erläutert.

### *Gibbs Sampling für das Kalibrierungsmodell*

Bayarri *et al.* (2007b) implementieren für das Kalibrierungsmodell einen Gibbs Sampler mit einem Metropolis Schritt. Dabei werden die suffizienten Statistiken  $\bar{\mathbf{y}}_E$  und  $s_E^2$  und die reduzierte Designmatrix  $\Delta_E$  für die Analyse herangezogen, um so die Dimension des MCMC Algorithmus zu reduzieren. Im Gibbs Sampler werden neben den Parametern auch die Vektoren  $\mathbf{y}_M^* = \tilde{y}_M(\Delta_E^u)$  und  $\mathbf{b}^* = \tilde{b}(\Delta_E)$  als Hilfsvariablen aufgenommen. Diese entsprechen Computermodell und Verzerrung am Experimentdesign  $\Delta_E$  und sind für eine geschlossene Darstellung der bedingten Verteilungen erforderlich. Für den Gibbs Sampler ist es hilfreich die Präzisionsparameter  $\lambda_E$  und  $\lambda_b$  statt der Varianzen  $\sigma_E^2$  und  $\sigma_b^2$  zu betrachten.

Die Likelihood lässt sich über die suffizienten Statistiken  $\bar{\mathbf{y}}_E, s_E^2$  und die Hilfsvariablen  $\mathbf{y}_M^*$  und  $\mathbf{b}^*$  faktorisieren zu

$$l(\mathbf{y}_E | \mathbf{u}, \lambda_E, \mu_b, \lambda_b) = p(\bar{\mathbf{y}}_E | \mathbf{b}^*, \mathbf{y}_M^*, \lambda_E) \cdot p(s_E^2 | \lambda_E) \cdot p(\mathbf{b}^* | \mu_b, \lambda_b) \cdot p(\mathbf{y}_M^* | \mathbf{y}_M), \quad (4.15)$$

siehe Bayarri *et al.* (2007b). Die Faktoren der Likelihood lauten dabei

$$p(\bar{\mathbf{y}}_E | \mathbf{b}^*, \mathbf{y}_M^*, \lambda_E) = \varphi(\bar{\mathbf{y}}_E | \mathbf{y}_M^* + \mathbf{b}^*, \bar{\mathbf{S}}_E)$$

für den Mittelwert der physikalischen Experimente,

$$p(s_E^2 | \lambda_E) = \lambda_E \cdot f_{\chi^2}(s_E^2 \lambda_E | \nu = \frac{1}{2} \sum_{i=1}^N n_i - 1)$$

für die Quadratsumme der physikalischen Experimente mit Dichte  $f_{\chi^2}(y | \nu)$  der  $\chi^2$ -Verteilung mit  $\nu$  Freiheitsgraden an der Stelle  $y$ ,

$$p(\mathbf{b}^* | \mu_b, \lambda_b) = \varphi\left(\mathbf{b}^* | \mu_b \mathbf{1}, \frac{1}{\lambda_b} R_b(\Delta_E)\right)$$

für die Verzerrung durch die Gauß Prozess Annahme und

$$p(\mathbf{y}_M^* | \mathbf{y}_M) = \varphi(\mathbf{y}_M^* | \hat{\mathbf{y}}_M[\mathbf{\Delta}_E^u], \hat{\Sigma}_M[\mathbf{\Delta}_E^u])$$

für das Computermodell mit bekannten Parametern wie bei der Gauß Prozess Interpolation in Gleichung (2.6).

Als a priori Verteilungen für die Hyperparameter wird eine multivariate Gleichverteilung für  $\mathbf{u}$ , die diffuse Verteilung  $p(\mu_b) \propto 1$  für  $\mu_b$ , die Exponentialverteilung  $\text{Exp}(\lambda_E | c_E \cdot \lambda_{EML})$  für  $\lambda_E$  und die Exponentialverteilung  $\text{Exp}(\lambda_b | c_b \cdot \lambda_{bML})$  für  $\lambda_b$  verwendet. Dabei bezeichnet  $\text{Exp}(x|a)$  die Exponentialverteilung mit Erwartungswert  $a$ . Für den Gibbs Sampler wird die a posteriori Verteilung über vollständig bedingte Dichten („full conditionals“) faktorisiert. Dabei werden in Bayarri *et al.* (2007b) die folgenden erforderlichen Dichten hergeleitet. Die bedingte Verteilung für  $\lambda_E$  lautet

$$\lambda_E | \bar{\mathbf{y}}_E, s_E^2, \mathbf{y}_M^*, \mathbf{b}^*, \lambda_b \sim \Gamma(a_E^1, a_E^2)$$

mit Formparameter  $a_E^1 = \frac{n-N+2}{2}$  und Skalenparameter  $a_E^2 = r_E + \frac{1}{2}s_E^2 + \frac{1}{2}(\bar{\mathbf{y}}_E - \mathbf{y}_M^* - \mathbf{b}^*)^T \text{diag}\{n_1, \dots, n_N\} (\bar{\mathbf{y}}_E - \mathbf{y}_M^* - \mathbf{b}^*)$ . Diese Dichte ergibt sich aus  $p(\lambda_E | \bar{\mathbf{y}}_E, s_E^2, \mathbf{y}_M^*, \mathbf{b}^*) \propto p(\bar{\mathbf{y}}_E | \mathbf{y}_M^*, \mathbf{b}^*, \lambda_E) p(s_E^2 | \lambda_E) p(\lambda_E)$ . Die  $\chi^2$ -Verteilung mit  $\nu$  Freiheitsgraden lässt sich als  $\Gamma(\frac{\nu}{2}, \frac{1}{2})$  und die Exponentialverteilung  $\text{Exp}(a)$  als  $\Gamma(1, a)$  Verteilung schreiben. Bei unabhängigen gammaverteilten Zufallsvariablen folgt das Produkt wieder einer Gammaverteilung.

Die bedingte Verteilung für  $\lambda_b$  hängt nur über  $\mathbf{b}^*$  von der Stichprobe ab:

$$\lambda_b | \mathbf{b}^* \sim \Gamma(a_b^{(1)}, a_b^{(2)})$$

mit Formparameter  $a_b^{(1)} = \frac{n+2}{2}$  und Skalenparameter  $a_b^{(2)} = r_b + \frac{1}{2}\mathbf{b}^* R_b(\mathbf{\Delta}_E)^{-1}\mathbf{b}^*$ . Diese Dichte ergibt sich aus dem Produkt  $p(\lambda_b | \bar{\mathbf{y}}_E, s_E^2, \lambda_E, \mathbf{y}_M^*, \mathbf{b}^*) \propto p(\mathbf{b}^* | \lambda_b) p(\lambda_b)$ .

Wird für  $\mu_b$  kein fester Wert verwendet wird, hängt auch diese vollständig bedingte

Verteilung wieder nur von  $\mathbf{b}^*$  ab:

$$\mu_b | \mathbf{b}^* \sim N \left( \frac{\mathbf{1}^T R_b(\Delta_E)^{-1} \mathbf{b}^*}{\mathbf{1}^T R_b(\Delta_E)^{-1} \mathbf{1}}, \frac{1}{\lambda_b \mathbf{1}^T R_b(\Delta_E)^{-1} \mathbf{1}} \right).$$

Für die zusätzlichen Hilfsvariablen  $\mathbf{y}_M^*$  und  $\mathbf{b}^*$  wird die gemeinsame bedingte Verteilung  $p(\mathbf{y}_M^*, \mathbf{b}^* | \bar{\mathbf{y}}_E, \mathbf{y}_M, \boldsymbol{\theta})$  mit Parametervektor  $\boldsymbol{\theta} = (\mathbf{u}^T, \lambda_b, \lambda_E, \mu_b)^T$  über bedingte Normalverteilungen nach (2.5) aus der gemeinsamen Verteilung von  $\mathbf{y}_M^*, \mathbf{b}^*, \bar{\mathbf{y}}_E, \mathbf{y}_M | \boldsymbol{\theta}$  hergeleitet. Der Erwartungswert ergibt sich in der gemeinsamen Verteilung komponentenweise aus  $f_M(\cdot) \boldsymbol{\beta}_M$  und  $\mu_b$ . Die Kovarianzmatrix wird blockweise aus den einzelnen Kovarianzfunktionen bestimmt. Da  $y_M(\cdot)$ ,  $\tilde{b}(\cdot)$  und der Zufallsfehler  $e$  a priori als unabhängig angenommen werden, kann die Kovarianzmatrix über Summen bestimmt werden. Sei  $\mathbf{w}_1^T = (\mathbf{y}_M^{*T}, \mathbf{b}^{*T})$  und  $\mathbf{w}_2^T = (\bar{\mathbf{y}}_E^T, \mathbf{y}_M^T)$ . Die Erwartungswerte  $\boldsymbol{\mu}_1 = E(\mathbf{w}_1)$  und  $\boldsymbol{\mu}_2 = E(\mathbf{z}_2)$  sind gegeben durch

$$\boldsymbol{\mu}_1 = \begin{pmatrix} f_M(\Delta_E^u)^T \boldsymbol{\beta}_M + \mu_b \mathbf{1} \\ f_M(\mathbf{D}_M)^T \boldsymbol{\beta}_M \end{pmatrix} \quad \text{und} \quad \boldsymbol{\mu}_2 = \begin{pmatrix} f_M(\Delta_E^u)^T \boldsymbol{\beta}_M \\ \mu_b \mathbf{1} \end{pmatrix},$$

die Kovarianzmatrizen  $\Sigma_1 = \text{Cov}(\mathbf{w}_1)$  und  $\Sigma_2 = \text{Cov}(\mathbf{z}_2)$  durch

$$\Sigma_1 = \begin{pmatrix} R_M(\Delta_E^u) & \mathbf{0} \\ \mathbf{0} & R_b(\Delta_E) \end{pmatrix} \quad \text{und} \quad \Sigma_2 = \begin{pmatrix} R_M(\Delta_E^u) + R_b(\Delta_E) & R_M(\Delta_E^u, \mathbf{D}_M) \\ R_M(\Delta_E^u, \mathbf{D}_M)^T & R_M(\mathbf{D}_M) \end{pmatrix}$$

sowie die Kreuzkorrelationsmatrix  $\Sigma_{1,2} = \text{Cov}(\mathbf{w}_1, \mathbf{z}_2)$  durch

$$\Sigma_{1,2} = \begin{pmatrix} R_M(\Delta_E^u) & R_b(\Delta_E) \\ R_M(\Delta_E^u, \mathbf{D}_M)^T & \mathbf{0} \end{pmatrix}.$$

Die gemeinsame bedingte Verteilung von  $\mathbf{y}_M^*$  und  $\mathbf{b}^*$  ergibt sich durch Bestimmung der bedingten Verteilung von  $\mathbf{w}_1 | \mathbf{w}_2$  durch Einsetzen in Gleichung (2.5). Es resultiert eine Normalverteilung mit Erwartungswert  $\boldsymbol{\mu}_1 - \Sigma_{1,2}^T \Sigma_2^{-1} (\mathbf{X}_2 - \boldsymbol{\mu}_2)$  und Kovarianz  $\Sigma_1 - \Sigma_{1,2}^T \Sigma_2^{-1} \Sigma_{1,2}$ .

Die bedingte Verteilung von  $\mathbf{u}$  lässt sich nicht geschlossen angeben. Für diesen Para-

meter ist im Gibbs Sampler ein Metropolis Schritt erforderlich. Die a posteriori für  $\mathbf{u}$  ist dabei als Produkt von Likelihood und  $p(\mathbf{u})$  gegeben. Geeignet ist hier die marginale Likelihood bei der  $\mathbf{y}_M^*$  herausintegriert wird:

$$p(\mathbf{u}|\mathbf{y}_E, \mathbf{y}_M, \mathbf{b}^*, \boldsymbol{\theta}) \propto \varphi(\bar{\mathbf{y}}_E|\hat{y}_M[\Delta_E^u] + \mathbf{b}^*, \Sigma_M[\Delta_E^u] + \bar{\mathbf{S}}(\Delta_E)) \cdot p(\mathbf{u}).$$

Wird auf das Herausintegrieren von  $\mathbf{y}_M^*$  verzichtet, wie in Bayarri *et al.* (2005), so hängt in der faktorisierten Likelihood aus (4.15) nur der Faktor  $p(\mathbf{y}_M^*|\mathbf{y}_M)$  für das Computermodell von  $\mathbf{u}$  ab. Bei kleiner Unsicherheit bezüglich  $y_M(\cdot)$  bewegt sich die MCMC Kette für  $\mathbf{u}$  nur extrem langsam. Dies sieht man wie folgt: Im  $t$ -ten Schritt des Gibbs Samplers wird bei gegebenem  $\mathbf{u}^{(t)}$  ein  $\mathbf{y}_M^{*(t)}$  generiert, anschließend wird die a posteriori von  $\mathbf{u}^{(t+1)}$  bei gegebenem  $\mathbf{y}_M^{*(t)}$  ausgewertet. Bei kleiner Unsicherheit des Computermodells fällt die Likelihood hier jedoch sehr schnell mit dem Abstand von  $\mathbf{u}^{(t)}$  und  $\mathbf{u}^{(t+1)}$  und neue Vorschläge werden nur bei extrem kleinen Abständen akzeptiert. Durch Herausintegrieren von  $\mathbf{y}_M^*$  wird dieses Problem abgeschwächt (persönliche Mitteilung, Rui Paulo, 2008).

Insgesamt verwenden Bayarri *et al.* (2007b) den folgenden MCMC Algorithmus. Aus den bedingten Verteilungen von  $\lambda_E$ ,  $\lambda_b$ ,  $\mu_b$ ,  $\mathbf{y}_M^*$  und  $\mathbf{b}^*$  werden sukzessive in Gibbs Schritten Zufallszahlen erzeugt. Für  $\mathbf{u}$  wird in einem Metropolis Schritt eine Zufallszahl vorgeschlagen. Als Vorschlagsdichte wird eine Mischung aus einer Gleichverteilung auf  $\mathcal{U}$  und einem modifizierten Random Walk Schritt jeweils mit Gewicht  $p_q = 0.5$  gewählt. Für den Random Walk Schritt wird eine Zahl aus einer lokalen Gleichverteilung um den alten Wert, gestutzt auf den Wertebereich, generiert:  $q_{u,2}(\mathbf{u}_{\text{neu}}|\mathbf{u}_{\text{alt}}) = \prod_{j=1}^{d_u} \frac{1}{c_j} \mathbb{1}(|u_{\text{neu},j} - u_{\text{alt},j}| \leq \epsilon_j) \cdot \mathbb{1}(u_{\text{neu},j} \in \mathcal{U}_j)$  mit komponentenweisen Abstand  $\epsilon_j$  und Wertebereich  $\mathcal{U}_j = [u_{j,u}, u_{j,o}]$  und  $c_j = \min\{u_{\text{alt},j} + \epsilon_j, u_{j,o}\} - \max\{u_{\text{alt},j} - \epsilon_j, u_{j,u}\}$ . Mit diesem Vorgehen werden die Vorschläge sowohl in kleinen Schritten um den alten Wert als auch in neuen Regionen des Wertebereichs generiert. Die Akzeptanzwahrscheinlichkeit für einen Vorschlag wird mit Hilfe des Produkts von Likelihood und a priori berechnet.

Die Implementierung des Gibbs Samplers erfordert zunächst die analytische Herleitung der vollständig bedingten Dichten. Anschließend werden für die Parameter  $\sigma_b^2$ ,  $\sigma_E^2$  und  $\mu_b$  alle Vorschläge akzeptiert, so dass der Algorithmus sehr effizient sein kann. In der Praxis (hier in der Anwendung auf die Rückfederungssimulation aus Kapitel 6) zeigt sich jedoch, dass sich die MCMC Kette beim Gibbs Sampler durch Korrelationen zwischen den Parametern zum Teil nur sehr langsam bewegt. Um eine möglichst unabhängige Stichprobe der a posteriori Verteilung zu erhalten und alle Bereiche des Parameterraums zu erreichen ist dann eine recht lange Markovkette erforderlich, bei der für die a posteriori Stichprobe ausreichend viele aufeinander folgende Werte zur Reduktion der Korrelation entfernt werden. Durch die erforderlichen Hilfsvariablen  $\mathbf{y}_M^*$  und  $\mathbf{b}^*$  ist die Dimension der MCMC Kette im Gibbs Sampler größer als die Parameteranzahl. Bei einer großen Anzahl unterschiedlicher Experimentdesignpunkte  $N$  entspricht die Dimension der Hilfsvariablen  $2 \cdot N$  schnell einem Vielfachen der Anzahl  $d_u + 3$  an Hyperparametern  $\mathbf{u}$ ,  $\lambda_b$ ,  $\lambda_E$ ,  $\mu_b$ . Bei großem Stichprobenumfang für die physikalischen Experimente und hohen Korrelationen kann deshalb ein einfacher Metropolis-Hastings Algorithmus geeigneter sein.

### *Metropolis-Hastings Algorithmus für das Kalibrierungsmodell*

Neben dem Gibbs Sampler betrachte ich auch den Metropolis-Hastings Algorithmus zur Analyse des Kalibrierungsmodells (siehe Anhang A.1 für eine Einführung). Hierbei werden Vorschläge für alle Parameter aus einer Vorschlagsdichte generiert und jeweils nur mit einer zu berechnenden Wahrscheinlichkeit akzeptiert. Entscheidend für die Effizienz des Metropolis-Hastings Algorithmus ist die Wahl einer geeigneten gemeinsamen Vorschlagsdichte für die Parameter  $\lambda_E$ ,  $\lambda_b$ ,  $\mu_b$  und  $\mathbf{u}$ . In dieser Arbeit werden unabhängige Random Walk Vorschlagsdichten verwendet. Für die Komponenten von  $\mathbf{u}$  wird eine lokale Gleichverteilung mit Intervalllänge  $\rho_u$  um den vorhergehenden Wert von  $\mathbf{u}$  analog zum Gibbs Sampler verwendet. Für  $\log(\lambda_E)$ ,  $\log(\lambda_b)$  und  $\mu_b$  werden jeweils normalverteilte Vorschlagsdichten mit Standardabweichungen  $sd_b$ ,  $sd_E$ ,  $sd_\mu$  gewählt. Die Akzeptanzwahrscheinlichkeit für einen Vorschlag berechnet sich über den Metropolis-Hastings Quotienten aus Vorschlagsdichte und a posteriori Dichte aus Glei-

chung (4.14). Der Algorithmus wird über die Tuningparameter der Vorschlagsdichten, dies sind hier die Standardabweichungen  $sd_b$ ,  $sd_E$ ,  $sd_\mu$  und die Intervalllänge  $\rho_u$ , eingestellt. Diese werden anhand von Testläufen so gewählt, dass eine geeignete Akzeptanzrate der MCMC Kette erreicht wird (siehe Anhang A.1). Die Konvergenz der MCMC Kette ist anhand von diagnostischen Grafiken zu überprüfen. Da im Gegensatz zum Gibbs Sampling keine vollständig bedingten Dichten herzuleiten sind, ist dieser Algorithmus sehr universell einsetzbar, bei unkorrelierten Parametern jedoch im Vergleich zum Gibbs Sampler weniger effizient. Für die Auswertung des Kalibrierungsmodells kann im Metropolis-Hastings Algorithmus auf die Hilfsvariablen  $\mathbf{y}_M^*$  und  $\mathbf{b}^*$  verzichtet werden, so dass sich die Dimension des zu erzeugenden Zufallsvektors gegenüber dem Gibbs Sampler reduziert. Insgesamt ist der Metropolis-Hastings Algorithmus jedoch stets sehr rechenintensiv. Eine schneller zu berechnende Implementierung wird im Folgenden vorgestellt.

#### *Importance Sampling mit iterativer Laplace Approximation*

Eine Stichprobe aus der a posteriori Verteilung kann auch mit Hilfe eines Importance Sampling Algorithmus generiert werden. Dabei werden Zufallszahlen aus einer Vorschlagsdichte entsprechend der a posteriori Verteilung gewichtet. Eine geeignete a posteriori Stichprobe kann über gewichtetes Resampling generiert werden. Das hier verwendete Importance Sampling mit iterativer Laplace Approximation (Bornkamp, 2011) wird in Anhang A.1 genauer beschrieben. Dabei wird die Vorschlagsdichte über eine iterative Laplace Approximation der a posteriori Dichte automatisch gewählt. Bei der Laplace Approximation wird eine Dichte durch eine Mischung von Normalverteilungen in den Maximalstellen approximiert. Für das Auffinden dieser Maximalstellen der a posteriori Verteilung sind mehrere geeignet gewählte Startwerte zu empfehlen. Bei der Kalibrierung von Computermodellen kann zum Beispiel ein raumfüllender Versuchsplan für den nichtlinearen Kalibrierparameter  $\mathbf{u}$  verwendet werden, um dessen Wertebereich möglichst gut abzudecken. So erfolgt auch für multimodale Funktionen eine geeignete Optimierung. Die iterative Laplace Approximation kann die a posteriori Dichte in der Regel gut approximieren wenn das Problem nicht zu hochdimensional

ist. Mit einer zunehmenden Anzahl von Parametern wird die Approximation jedoch schwieriger und ihre Güte ist zu überprüfen. Dies kann zum Beispiel über den effektiven Stichprobenumfang NESS erfolgen (siehe Anhang A.1 und Robert und Casella (2004), Kapitel 3).

#### 4.3.1 Vorhersagen für die Bayes Kalibrierung

Mithilfe einer a posteriori Stichprobe für die Parameter lassen sich nun Vorhersagen und Unsicherheitsintervalle für den physikalischen Prozess  $Y_E(\mathbf{x})$ , das Computermodell  $\tilde{y}_M(\mathbf{x}, \cdot)$  und Verzerrung  $\tilde{b}(\mathbf{x})$  herleiten. Dabei werden für das  $i$ -te Element der a posteriori Stichprobe  $(\mathbf{u}^{(i)}, \sigma_E^2{}^{(i)}, \sigma_b^2{}^{(i)}, \mu_b^{(i)})$  Zufallszahlen für  $\tilde{y}_M^{(i)}(\mathbf{x}, \mathbf{u}^{(i)})$ ,  $\tilde{b}^{(i)}(\mathbf{x})$  und  $e^{(i)}$  erzeugt. Kenngrößen der Vorhersageverteilung, wie zum Beispiel Erwartungswert oder Quantile, können durch ihre empirischen Gegenstücke aus der a posteriori Stichprobe geschätzt werden. Der Erwartungswert des physikalischen Prozesses wird zum Beispiel durch  $\hat{\mu}_E(\mathbf{x}) = \frac{1}{N_{\text{MC}}} \sum_{i=1}^{N_{\text{MC}}} \mu_E^{(i)}(\mathbf{x})$  mit  $\mu_E^{(i)}(\mathbf{x}) = \tilde{y}_M^{(i)}(\mathbf{x}, \mathbf{u}^{(i)}) + \tilde{b}_u^{(i)}(\mathbf{x})$  geschätzt. Computermodell und Verzerrung können analog geschätzt werden. Eine schneller berechenbare Alternative setzt den a posteriori Modalwert  $\hat{\mathbf{u}}$  für  $\mathbf{u}$  in das Computermodell ein und nutzt  $\hat{y}_M(\mathbf{x}, \hat{\mathbf{u}})$  als Vorhersage für das Computermodell und  $\hat{b}(\mathbf{x}) = \hat{\mu}_E(\mathbf{x}) - \hat{y}_M(\mathbf{x}, \hat{\mathbf{u}})$  als Vorhersage für die Verzerrung.

Unsicherheitsintervalle für die Kalibrierung von Computermodellen bestimmen Bayarri *et al.* (2007b) ebenfalls aus der a posteriori Stichprobe für  $\mu_E(\mathbf{x})$ . Für die Validierung des Computermodells sind die folgenden Toleranzintervalle geeignet. Ein  $1 - \alpha$  Toleranzintervall für das Computermodell bezüglich des physikalischen Prozesses ist gegeben durch  $[\hat{y}_M(\mathbf{x}, \hat{\mathbf{u}}) - \tau_M(\mathbf{x}), \hat{y}_M(\mathbf{x}, \hat{\mathbf{u}}) + \tau_M(\mathbf{x})]$  mit  $P(|\hat{y}_M(\mathbf{x}, \hat{\mathbf{u}}) - \mu_E(\mathbf{x})| < \tau_M(\mathbf{x})) = 1 - \alpha$ . Dabei ist die Toleranz  $\tau_M(\mathbf{x})$  so zu wählen, dass für  $(1 - \alpha) \cdot 100\%$  der a posteriori Stichprobe

$$|\hat{y}_M(\mathbf{x}, \hat{\mathbf{u}}) - \mu_E^{(i)}(\mathbf{x})| < \tau_M(\mathbf{x})$$

gilt. Interpretieren lässt sich die Toleranz  $\tau_M(\boldsymbol{x})$  wie folgt. Mit Wahrscheinlichkeit  $1 - \alpha$  besitzt das Computermodell in  $\boldsymbol{x}$  einen Abstand kleiner  $\tau_M(\boldsymbol{x})$  zum Erwartungswert des physikalischen Prozesses  $\mu_E(\boldsymbol{x})$ . Dem Anwender kann somit die zu erwartende Abweichung vermittelt werden. Ob das Computermodell als valide bezeichnet werden kann, ergibt sich dann aus der tolerierbaren Abweichung für die Anwendung. Für eine Modellwahl ist dieses Verfahren nur bedingt geeignet. Da die Toleranzen von  $\boldsymbol{x}$  abhängen, kann erst nach Berechnung von Vorhersagen auf ganz  $\mathcal{X}$  entschieden werden ob eine Verzerrung des Computermodells vorliegt.

Insgesamt ist der vorgestellte Ansatz zur Bayesianischen Kalibrierung durch seine große Flexibilität besonders geeignet zur Erstellung eines Vorhersagemodells für den physikalischen Prozess. Neben einer Vorhersage werden dabei auch Unsicherheitsintervalle angegeben. Über Toleranzintervalle lässt sich die erwartete Abweichung vom Computermodell zur Realität quantifizieren und das Computermodell validieren. Außerdem kann die Unsicherheit über den nichtlinearen Parameter  $\boldsymbol{u}$  bestimmt werden. Durch Hinzunahme der Verzerrungsfunktion liegt ein realistisches Modell zur Beschreibung der Daten vor. Allerdings ist das Modell durch diese Formulierung nicht direkt identifizierbar und das Vorliegen der Verzerrung macht die Interpretation von  $\boldsymbol{u}$  schwieriger. Ein erheblicher Nachteil des Modells ist die hohe Rechenkomplexität. Dies gilt nicht nur für die Herleitung der a posteriori Verteilung für die Parameter sondern auch für die Vorhersageverteilungen des physikalischen Prozesses. In jedem Punkt  $\boldsymbol{x}$  sind hier alle a posteriori Stichprobenelemente der Parameter zu berücksichtigen. Allerdings liefert der Ansatz nur so neben Vorhersagen auch Unsicherheitsintervalle. Um den numerischen Aufwand möglichst gering zu halten, wurde hier bereits auf einen vollständigen Bayes Ansatz verzichtet.

#### 4.4 Simulationsstudie zur Bayesianischen Kalibrierung

In diesem Abschnitt wird das Bayesianische Kalibrierungsmodell aus Kapitel 4.3 näher betrachtet und seine Sensitivität gegenüber a priori Verteilungen für die Varianzpara-

meter  $\sigma_b^2$  und  $\sigma_E^2$  anhand einer Simulationsstudie untersucht. In Kapitel 4.4.1 werden zunächst vier Testbeispiele vorgestellt. Anschließend werden das Design der Simulationsstudie und die verwendeten a priori Verteilungen näher beschrieben. Abschließend werden in Kapitel 4.4.3 die Ergebnisse diskutiert.

#### 4.4.1 Testbeispiele für die Kalibrierung von Computermodellen

Um die Kalibrierung von Computermodellen anhand von Simulationsstudien untersuchen zu können, werden hier vier niedrig dimensionale Testbeispiele betrachtet. Ein Testbeispiel für die Kalibrierung besteht immer aus einer Computermodellfunktion und einem Modell für den physikalischen Prozess.

Das erste Testbeispiel (Bayarri *et al.*, 2007b, Loeppky *et al.*, 2006) lässt sich aus der chemischen Kinetik motivieren. Das Computermodell ist hier gegeben durch

$$f_M^1(x, u) = 5 \cdot \exp(-(3 \cdot u) \cdot (3 \cdot x)).$$

Es beschreibt bei einer Anfangskonzentration 5 eines Stoffes und Reaktionsrate  $3 \cdot u$  den noch nicht reagierten Anteil des Stoffes zum Zeitpunkt  $3 \cdot x$ . Der Wertebereich für  $x$  und  $u$  liegt hier jeweils im Intervall  $[0, 1]$ . Es wird angenommen, dass der Restanteil des untersuchten Stoffes in der Realität durch

$$Y_E^1|x = 3.5 \exp(-1.7 \cdot 3 \cdot x) + 1.5 + \epsilon$$

mit unabhängigen Zufallsfehler  $\epsilon \sim N(0, 0.3^2)$  gegeben ist. Somit liegt ein verzerrtes Computermodell mit wahrer Reaktionsrate  $u^* = 1.7/3 = 0.57$  vor. Die Verzerrung  $b^1(x)$  entspricht hier  $b^1(x) = 1.5(1 - \exp(-1.7 \cdot 3 \cdot x))$ . Der theoretische nichtlineare Kleinste Quadrate Schätzer  $u_{KQ} = \operatorname{argmin}_{u \in \mathcal{U}} \int_{\mathcal{X}} (E(Y_E^1|x) - f_M^1(x, u))^2 dx$  liegt in diesem Beispiel bei  $u_{KQ} = 0.21$ .

Als zweites Testbeispiel (Han und Santner, 2009) wird die Computermodellfunktion

$$f_M^2(x, u) = x^2 + u_1(x + 1) + 0.5u_2$$

und für das physikalische Experiment die Modellgleichung

$$Y_E^2|x = x^2 + 0.1x + 0.5 + \epsilon,$$

mit unabhängigem Fehler  $\epsilon \sim N(0, 0.1)$ , betrachtet. Der Wertebereich für den gemeinsamen Vektor von  $x$  und  $u$  ist der dreidimensionale Einheitswürfel. Bei diesem Beispiel handelt es sich um ein unverzerrtes Computermodell, da das Computermodell mit  $u^* = (0.1, 0.8)$  den Erwartungswert der physikalischen Experimente perfekt beschreiben kann. Dieses Beispiel ist so gewählt, dass sich  $u_1$  einfach und  $u_2$  etwas schwieriger aus Daten schätzen lässt (siehe Han und Santner, 2009).

Als drittes Testbeispiel wird ein Kolbensimulator, der so genannte piston simulator aus Kenett und Zacks (1998), als Computermodell verwendet. Dabei wird die vollständige Umlaufzeit der Kolbenwelle in Abhängigkeit von der Kolbenoberfläche ( $S$ , 0.005-0.02  $m^2$ ) sowie der Federkonstante ( $k$ , 1000 – 5000  $N/m$ ) und der Füllgastemperatur ( $T_0$ , 340 – 360  $K$ ) modelliert. Die Kolbenoberfläche wird als Einflussgröße im physikalischen Prozess und die Federkonstante und Füllgastemperatur als nicht vollständig bekannte Größen der Simulation – und somit als Kalibrierparameter – betrachtet. Die Umlaufzeit wird im Computermodell beschrieben durch

$$\text{Umlaufzeit} = 2\pi \sqrt{\frac{M}{k + S^2 \frac{P_0 V_0}{T_0} \frac{T}{V^2}}}$$

mit  $V = \frac{S}{2k} \left( \sqrt{A^2 + 4k \frac{P_0 V_0}{T_0} T} - A \right)$  und  $A = P_0 S + 19.62M - \frac{kV_0}{S}$ . Die weiteren Einflussgrößen im Computermodell werden auf folgende feste Werte gesetzt: Kolbengewicht  $M = 45kg$ , Anfangsgasvolumen  $V_0 = 0.006cm^3$ , atmosphärischer Druck  $P_0 = 100000N/m^2$  und Umgebungstemperatur  $T = 293K$ . Als Testfunktion  $f_M^3(x, u)$  wird nun die Funktion der Umlaufzeit mit auf  $[0, 1]$  skalierten Einflussgrößen definiert. Für den wahren physikalischen Prozess  $Y_E^3$  wird davon ausgegangen, dass das Computermodell den Prozess bereits geeignet beschreibt, jedoch in der Realität die

physikalische Einflussgröße  $\mathbf{x}$  geringfügig verschoben vorliegt:

$$Y_E^3|x = f_M^3(x + 0.01, \mathbf{u}^*) + \epsilon$$

mit  $\mathbf{u}^* = (0.1, 0.3)^T$ . Im wahren Prozess liegt somit eine systematische Abweichung der Kolbenoberfläche von  $0.00015 \text{ m}^2$  zu Grunde. Der Fehler  $\epsilon$  sei wieder unabhängig normalverteilt mit  $\epsilon \sim N(0, 0.01^2)$ . Die Verzerrungsfunktion ist hier  $b^3(x) = 2\pi \left( \sqrt{\frac{45}{1400+(x-0.015+0.015)^2}} - \sqrt{\frac{45}{1400+(x-0.015+0.005)^2}} \right)$ . Der theoretische kleinste Quadrate Schätzer lautet  $\mathbf{u}_{\text{KQ}} = (0.598, 1)^T$ .

Als viertes Testbeispiel wird ebenfalls das Kolbensimulationsproblem betrachtet. Nur wird diesmal davon ausgegangen, dass das Computermodell die Realität vollständig korrekt beschreibt. Somit wird die Computermodellfunktion  $f_M^4(x, \mathbf{u}) = f_M^3(x, \mathbf{u})$  verwendet und der physikalische Prozess wird durch

$$Y_E^4|x = f_M^4(x, \mathbf{u}^*) + \epsilon$$

beschrieben. Dabei ist wieder  $\mathbf{u}^* = (0.1, 0.3)^T$  und  $\epsilon \sim N(0, 0.01^2)$ .

Insgesamt liegen somit vier unterschiedliche Testfunktionen für die Kalibrierung von Computermodellen vor. Die Konstruktion von Testbeispielen für die Kalibrierung ist schwierig, wenn eine Verzerrung zugelassen wird. Aus dem Modell ergibt sich durch die Identifizierbarkeitsproblematik, dass beliebige Funktionen für Computermodell und Prozessmittelwert verwendet werden können. Die ausgewählten Testfunktionen 1 und 3 mit Verzerrung illustrieren, wie eine Abweichung zur Realität zu Stande kommen kann. Hier bilden die Computermodellfunktionen den physikalischen Prozess zum großen Teil richtig ab. Allerdings wurden einige kleinere Effekte im Computermodell nicht korrekt beachtet, die in einer komplexeren additiven Abweichung resultieren können.

#### 4.4.2 *Simulationsstudie: Kalibrierung von Computermodellen mit unterschiedlichen a priori Verteilungen*

In einer Simulationsstudie soll nun das Kalibrierungsmodell aus Kapitel 4.3 anhand der vorgestellten Testbeispiele näher untersucht werden. Für jedes der vier Testbeispiele werden die Computerexperimente nach einem Maximin-LHC Design ausgewählt. Der Stichprobenumfang entspricht dem Zehnfachen der Dimension der Computermodellfunktion, d.h. 20 Beobachtungen für  $f_M^1$  und 30 Beobachtungen für  $f_M^2$ ,  $f_M^3$  und  $f_M^4$ . Die experimentellen Beobachtungen werden nach einem äquidistanten Design auf 5 Stufen mit jeweils 3 Wiederholungen simuliert, so dass insgesamt  $n_E = 15$  Beobachtungen für die eindimensionalen Testprobleme vorliegen. Auf diese Daten wird jeweils die Bayesianische Kalibrierung mit Verzerrung aus Kapitel 4.3 angewendet. Insgesamt wird die Generierung der Datensätze für Simulation und Experiment jeweils 50-mal zufällig wiederholt. So liegen für jedes Testbeispiel insgesamt 50 Kombinationen aus Computerexperimenten  $\mathbf{y}_M$  zu einem Design  $\mathbf{D}_M$  sowie physikalischen Experimenten  $\mathbf{y}_E$  zu einem Design  $\mathbf{D}_E$  vor.

In der Simulationsstudie sollen unterschiedliche a priori Verteilungen für die Varianzparameter  $\sigma_b^2$  und  $\sigma_E^2$  verglichen werden. Neben einer datenabhängigen inversen Exponentialverteilung für die Varianzparameter wie bei Bayarri *et al.* (2007b) werden auch schwächer informative Verteilungen wie eine inverse Gammaverteilung oder die von Gelman (2006) für hierarchische Modelle empfohlene quadrierte  $t$ -Verteilung untersucht. Von Interesse ist dabei einerseits wie sensitiv die Resultate der Kalibrierung gegenüber den unterschiedlichen a priori Verteilungen sind. Andererseits ist zu untersuchen, ob sich die Empfehlungen für die quadrierte  $t$ -Verteilung von Gelman (2006) auch auf das vorliegende Kalibrierungsmodell übertragen lassen. Zunächst werden in einem Exkurs die unterschiedlichen a priori Verteilungen beschrieben.

##### *A priori Verteilungen für Varianzparameter*

Bayarri *et al.* (2007b) verwenden als a priori Verteilungen für die Skalenparameter  $\sigma_b^2$

und  $\sigma_E^2$  datenabhängige inverse Gammaverteilungen, dies entspricht Gammaverteilungen für die Präzisionsparameter  $\lambda_b$  und  $\lambda_E$ . Als konjugierte a priori Verteilung für die Präzision bei Normalverteilung bietet die Gammaverteilung den Vorteil, dass die a posteriori Verteilung der Präzision wieder eine Gammaverteilung besitzt, aus der in einem Gibbs Sampler einfach Zufallszahlen generiert werden können.

Die Dichte der inversen Gammaverteilung mit Formparameter  $\alpha$  und Skalenparameter  $\beta$  für  $\sigma^2$  lautet

$$p(\sigma^2) = \frac{\beta^\alpha}{\Gamma(\alpha)} \left( \frac{1}{\sigma^2} \right)^{\alpha+1} \exp\left(-\frac{\beta}{\sigma^2}\right),$$

mit Erwartungswert  $E(\sigma^2) = \frac{\alpha}{\beta}$ . Für  $\alpha = 1$  entspricht die inverse Gammaverteilung einer inversen Exponentialverteilung, das heißt der Kehrwert  $1/\sigma^2$  besitzt eine Exponentialverteilung mit Parameter  $\beta$ . Abbildung 4.1 zeigt die Dichte der inversen Gamma-Verteilung für verschiedene Parameterwerte. Es handelt sich um eine unimodale Dichte mit Modalwert an der Stelle  $\frac{\beta}{\alpha+1}$ . Für kleine Werte gibt es zunächst einen starken Anstieg und anschließend für größere Werte einen mehr oder weniger starken Abfall der Dichte. Für Varianzparameter ist es in der Regel sinnvoll, dass größere Werte a priori unwahrscheinlicher sind. Diese a priori Verteilung ist von Null weg beschränkt, das heißt die zu der Varianz gehörende Zufallsvariable kann nicht zu einer Einpunktverteilung in Null zusammenfallen.

Für die Kalibrierung von Computermodellen empfehlen Bayarri *et al.* (2007b) die Einschränkung auf die datenabhängige Exponentialverteilung, das heißt  $\sigma_b^2 \sim \text{Inv}\Gamma(1, \beta_b)$  und  $\sigma_E^2 \sim \text{Inv}\Gamma(1, \beta_E)$ . Dabei wird  $\beta_b = \frac{c_b}{\sigma_{b,\text{ML}}^2}$  und  $\beta_E = \frac{c_E}{\sigma_{E,\text{ML}}^2}$  jeweils als Vielfaches des ML Schätzers gewählt. Für die Konstanten  $c_b$  und  $c_E$  werden kleine ganze Zahlen (zum Beispiel  $c_b = c_E = 5$ ) eingesetzt, so dass die a priori Verteilung für  $\sigma_b^2$  und  $\sigma_E^2$  eher kleinere Werte als den ML Schätzer bevorzugt.

Um eine datenunabhängige und wenig informative Wahl innerhalb des Rahmens der inversen Gammaverteilung als a priori zu ermöglichen, können sowohl Skalen- als auch Formparameter auf einen kleinen Wert  $\epsilon$  gesetzt werden. Für  $\epsilon$  werden beispielsweise Werte wie 0.01 oder 0.001 verwendet. So liegt der Modalwert sehr nahe Null und die

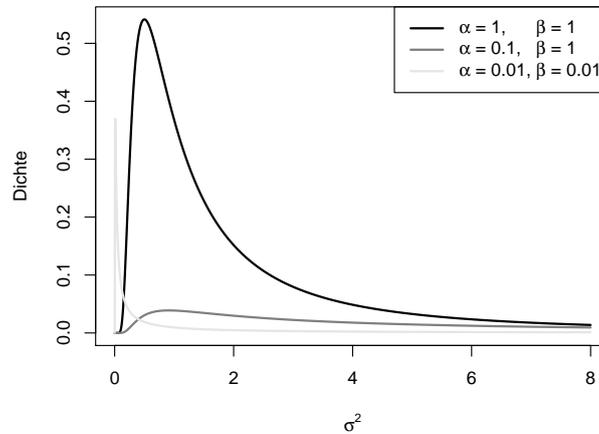
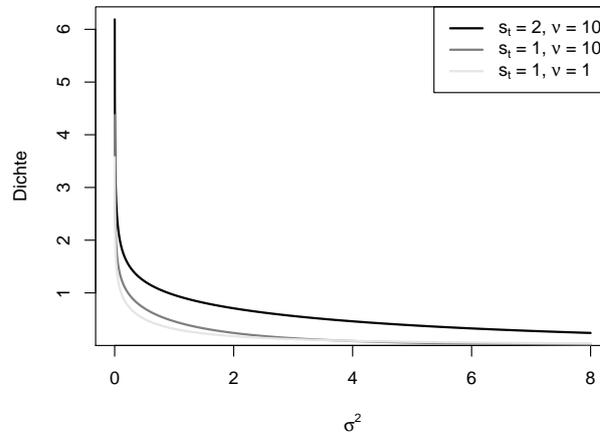


Abbildung 4.1: Dichte der inversen Gammaverteilung

Dichte ist mit Ausnahme eines kleinen Bereichs zu Beginn monoton fallend ist (siehe Abb. 4.1). Gelman (2006) weist jedoch darauf hin, dass eine solche inverse Gammaverteilung  $\text{Inv}\Gamma(\epsilon, \epsilon)$  für die Varianzparameter in einem hierarchischen Modell sehr sensitiv gegenüber der Wahl von  $\epsilon$  ist, wenn tatsächlich eine kleine Varianz vorliegt. In dieser Situation kann sich der monoton steigende Bereich bis zum Modalwert auswirken. Konkret betrachtet Gelman (2006) das allgemeine hierarchische Modell  $y = \mu + \alpha + e$  mit  $\alpha \sim N(0, \tau^2)$  und  $e \sim N(0, \sigma^2)$ . In diesem Modell bestehen Probleme bei Verwendung einer inversen Gammaverteilung für den Varianzparameter  $\tau^2$ . Eine ähnliche hierarchische Modellsituation liegt auch im Kalibrierungsmodell mit Verzerrungsfunktion vor - auch wenn die zu Grunde liegende Modellstruktur hier wesentlich komplizierter ist. Das allgemeine Mittel  $\mu$  erweitert sich auf die Funktion des Computermodells und die Rolle des zufälligen Effekts  $\alpha$  übernimmt die Verzerrungsfunktion  $b(\mathbf{x})$ . Die Verteilungsannahme wird im Vergleich zum einfachen hierarchischen Modell auf den Funktionenraum durch die Gauß Prozess Annahme  $GP(0, \sigma_b^2 \mathbf{C}_b(\cdot, \cdot))$  erweitert. Somit entspricht der Parameter  $\sigma_b^2$  dem Varianzparameter im hierarchischen Modell. Es ist deshalb zu prüfen, ob für dieses Modell eine nichtinformative inverse Gammaverteilung als a priori für  $\sigma_b^2$  geeignet ist oder ob selbst die datenabhängige inverse Exponentialverteilung zu ähnlichen Problemen führen kann. Dies könnte insbesondere für die unverzerrten Testfunktionen 2 und 4 der Fall sein, da hier  $\sigma_b^2 = 0$  geeignet wäre.

Abbildung 4.2: Dichte der  $t^2$ -Verteilung

Als Alternative empfiehlt Gelman (2006) eine so genannte absolute  $t$ -Verteilung für die Standardabweichung  $\sigma$ . Bei der absoluten  $t$ -Verteilung wird eine  $t$ -Verteilung mit  $\nu$  Freiheitsgraden und Skalenparameter  $s_t$  um den Erwartungswert 0 „gespiegelt“, so dass nur die positiven reellen Zahlen eine positive Wahrscheinlichkeitsdichte besitzen. Kurz wird dies als  $|t_\nu|$ -Verteilung bezeichnet und die Dichte ist proportional zu

$$p(\sigma) \propto \left(1 + \frac{1}{\nu} \left(\frac{\sigma}{s_t}\right)^2\right)^{-(\nu+1)/2}. \quad (4.16)$$

Für  $\nu = 1$  entspricht diese der absoluten Cauchyverteilung. Für große aber endliche Werte von  $s_t$  handelt es sich bei der absoluten Cauchyverteilung um eine schwach informative Verteilung, deren Dichte schwere Ränder hat in denen sie monoton fallend in  $\sigma$  ist. Folgt  $\sigma$  einer  $|t|$ -Verteilung, so folgt  $\sigma^2$  einer quadrierten  $t$ -Verteilung, kurz  $t^2$ -Verteilung. Die Dichte der  $t^2$ -Verteilung ist in Abbildung 4.2 dargestellt. Sie ist im Gegensatz zur Inversen Gammaverteilung für alle positiven reellen Zahlen monoton fallend in  $\sigma^2$ . Damit liegt ihr Modalwert in Null und sie ist als a priori Verteilung auch in der zuvor diskutierten Situation eines hierarchischen Modells geeignet. Beschreibt die zugrunde liegende Zufallsvariable zum Beispiel einen zufälligen Effekt, so ist eine Varianz von Null oder nahe Null a priori sinnvoll. Im Kalibrierungsmodell korrespondieren zum Beispiel sehr kleine Werte von  $\sigma_b^2$  zu einem (nahezu) unverzerrten

Szenario	a priori Verteilung	Parameter für $\sigma_b^2$	Parameter für $\sigma_E$
1	$t^2 (\nu = 1)$	$s_t = 100$	$s_t = 100$
2	$t^2 (\nu = 1)$	$s_t = \widehat{\sigma}_{b,ML}$	$s_t = 100$
3	$t^2 (\nu = 1)$	$s_t = 100$	$s_t = \widehat{\sigma}_{E,ML}$
4	$t^2 (\nu = 1)$	$s_t = \widehat{\sigma}_{b,ML}$	$s_t = \widehat{\sigma}_{E,ML}$
5	$t^2 (\nu = 10)$	$s_t = \widehat{\sigma}_{b,ML}$	$s_t = \widehat{\sigma}_{E,ML}$
6	$\text{Inv}\Gamma(\beta, 1)$	$\beta = 5/\widehat{\sigma}_{b,ML}^2$	$\beta = 5/\widehat{\sigma}_{E,ML}^2$
7	$\text{Inv}\Gamma(\beta, 0.1)$	$\beta = 5/\widehat{\sigma}_{b,ML}^2$	$\beta = 5/\widehat{\sigma}_{E,ML}^2$
8	$\text{Inv}\Gamma(\beta, \beta)$	$\beta = 0.01$	$\beta = 0.01$

Tabelle 4.1: A priori Verteilungen für Varianzparameter  $\sigma_b^2$  und  $\sigma_E^2$  in der Simulationsstudie

Computermodell. Diese sind somit nicht auszuschließen und im Sinne eines möglichst einfachen Modells falls geeignet auch größeren Werten von  $\sigma_b^2$  vorzuziehen.

Gelman (2006) empfiehlt als schwach informative a priori für Varianzparameter in hierarchischen Modellen die quadrierte Cauchyverteilung mit großem Skalenparameter  $s_t$ , der jedoch nicht vollkommen außerhalb des erwarteten Wertebereichs liegen sollte. Insbesondere bei Modellen mit mehreren Varianzparametern schlägt er vor  $s_t$  datenabhängig zu wählen.

Tabelle 4.1 gibt eine Übersicht über die in der Simulationsstudie verwendeten a priori Verteilungen. Insgesamt werden 8 verschiedene Simulationsszenarien verwendet. Szenarien 1-5 vergleichen datenabhängige und datenunabhängige Wahl der Parameter für die  $t^2$ -Verteilung. In Szenario 1 wird sowohl für  $\sigma_b^2$  als auch  $\sigma_E^2$  eine datenunabhängige quadrierte Cauchyverteilung gewählt und der Skalenparameter  $s_t$  jeweils fest auf einen großen Wert, hier auf 100, gesetzt. Um den Einsatz von datenabhängigen und datenunabhängigen a priori Verteilungen für  $\sigma_b^2$  und  $\sigma_E^2$  getrennt zu untersuchen, wird in den Szenarien 2 und 3 für jeweils einen der beiden Parameter eine datenunabhängige quadrierte Cauchyverteilung wie in Szenario 1 und für den anderen Parameter eine datenabhängige quadrierte Cauchyverteilung gewählt. In der datenabhängigen Situation wird der Skalenparameter  $s_t$  aus (4.16) auf den ML Schätzer für die Standardabweichung  $\sigma_b$  bzw.  $\sigma_E$  gesetzt. In Szenario 4 und 5 werden für beide Varianzparameter datenabhängige  $t^2$ -Verteilungen verwendet: eine quadrierte Cauchyverteilung (Szena-

rio 4) sowie eine informativere quadrierte  $t$ -Verteilung mit  $\nu = 10$  Freiheitsgraden. Damit liegt bei der datenabhängigen quadrierten Cauchyverteilung 99 % der Wahrscheinlichkeitsmasse etwa unter dem 4000-fachen und bei der quadrierten  $t$ -Verteilung mit 10 Freiheitsgraden 99 % etwa unter dem 10-fachen des ML Schätzers für  $\sigma_b^2$  bzw.  $\sigma_E^2$ .

In den Szenarien 6-8 werden a priori Verteilungen aus der Klasse der inversen Gamma-verteilung betrachtet. Dabei wird in den Szenarien 6 und 7 der Skalenparameter der Gammaverteilung datenabhängig auf das fünffache des ML Schätzers gesetzt. Kombiniert wird dies mit einem Formparameter  $\alpha = 1$  in Szenario 6, dies entspricht der Exponentialverteilung, und  $\alpha = 0.1$  in Szenario 7 als etwas schwächer informative Alternative. In Szenario 6 liegen somit 99 % der Wahrscheinlichkeitsmasse etwa unter dem 8-fachen sowie in Szenario 7 unter dem 23-fachen des ML Schätzers für  $\sigma_b^2$  bzw.  $\sigma_E^2$ . In Szenario 8 wird die datenunabhängige  $\text{Inv}\Gamma(0.01, 0.01)$  Verteilung verwendet.

#### 4.4.3 Ergebnisse der Simulationsstudie zur Bayes Kalibrierung

Die Ergebnisse der Simulationen werden nun nachfolgend beschrieben. Zielgrößen sind dabei der Vorhersagefehler für  $\mu_E(\cdot)$ , der Schätzfehler für  $\sigma_E^2$  sowie die a posteriori Verteilungen für  $\sigma_b^2$  und  $\mathbf{u}$ .

Für eine Bewertung der Kalibrierung wurde zunächst der Vorhersagefehler (RMSE) auf einem Gitter für die vier Testszenarien ermittelt. Ergebnisse für die Szenarien 1-5 sind in Abbildung 4.3 dargestellt. Dabei wurde aus den 50 Werten der Simulationsläufe jeweils ein Kerndichteschätzer mit Hilfe der R Funktion `density` angepasst. Es zeigt sich, dass die Spezifikation der  $t^2$ -Verteilung für alle Simulationsszenarien einen geringen aber erkennbaren Einfluss auf den Vorhersagefehler besitzt. Bei den betrachteten Testfunktionen ist der Vorhersagefehler bei einer datenabhängigen Wahl für  $\sigma_b^2$  kleiner als bei einer datenunabhängigen Wahl. Besonders deutlich ist dieser Unterschied für die Testfunktion  $f_M^2$  zu sehen. Der Vorhersagefehler unterscheidet sich zwischen einer datenabhängigen oder unabhängigen a priori Verteilung für  $\sigma_E^2$  hingegen nicht wesentlich.

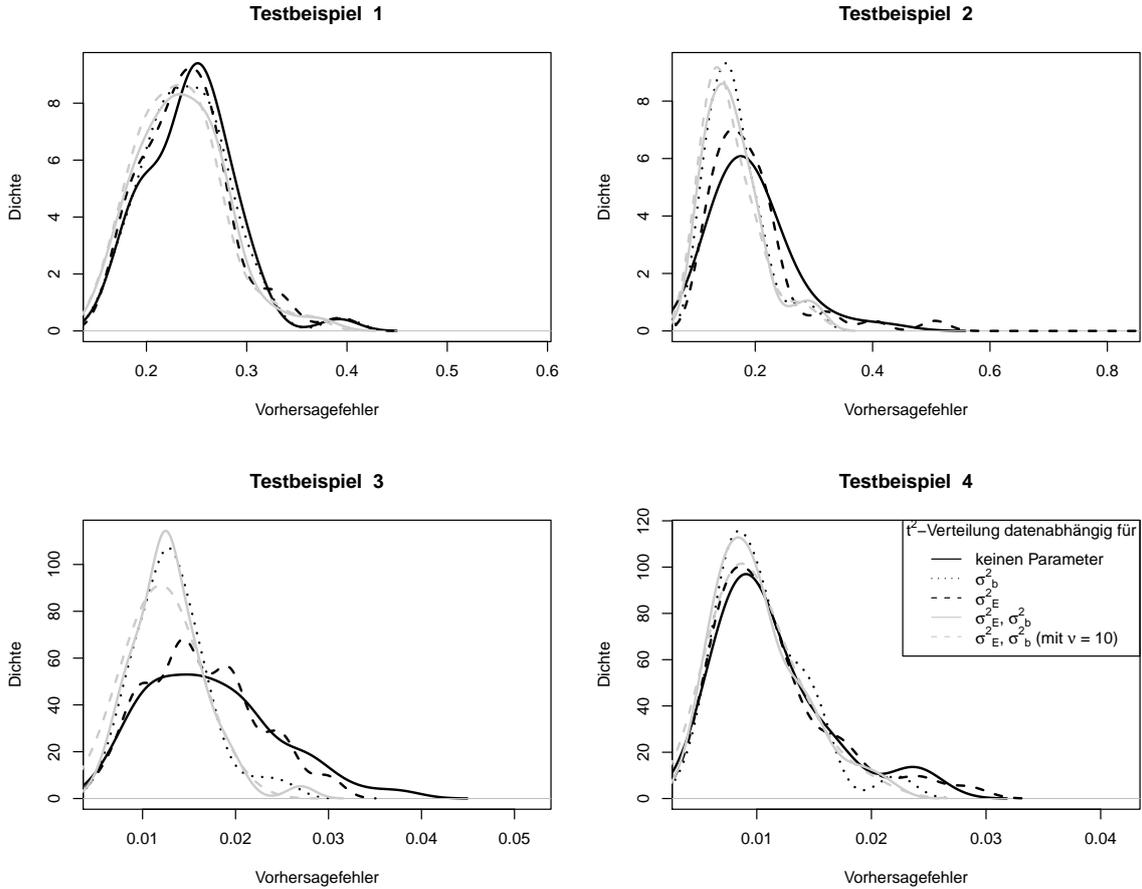


Abbildung 4.3: Vorhersagefehler mit  $t^2$  a priori

Abbildung 4.4 stellt den Vorhersagefehler für die a priori Verteilungen aus der Klasse der inversen Gammaverteilung und als Referenz die datenabhängige  $t^2$ -Verteilung mit 10 Freiheitsgraden gegenüber (Szenarien 5-8). Für die untersuchten Testfunktionen  $f_M^2(\cdot)$  und  $f_M^3(\cdot)$  liefert die Kalibrierung mit der informativeren Inversen Exponentialverteilung einen etwas kleineren Vorhersagefehler. Ansonsten unterscheiden sich die Resultate zwischen den datenabhängigen a priori Verteilungen nur wenig. Die datenunabhängige  $\text{Inv}\Gamma(0.01, 0.01)$ -Verteilung führt hingegen für die Testfunktion  $f_M^3(\cdot)$  und  $f_M^4(\cdot)$  zu deutlich schlechteren Ergebnissen. Insgesamt lässt sich aus den Simulationen schließen, dass für die untersuchten Testfunktionen der Vorhersagefehler nicht extrem sensitiv gegenüber den unterschiedlichen a priori Verteilungen ist. Nur die datenunabhängige inverse Gammaverteilung ist für die Kalibrierung der Testfunktionen  $f_M^3(\cdot)$  und  $f_M^4(\cdot)$  nicht geeignet. Weiterhin gilt in der Tendenz, je informativer die a priori

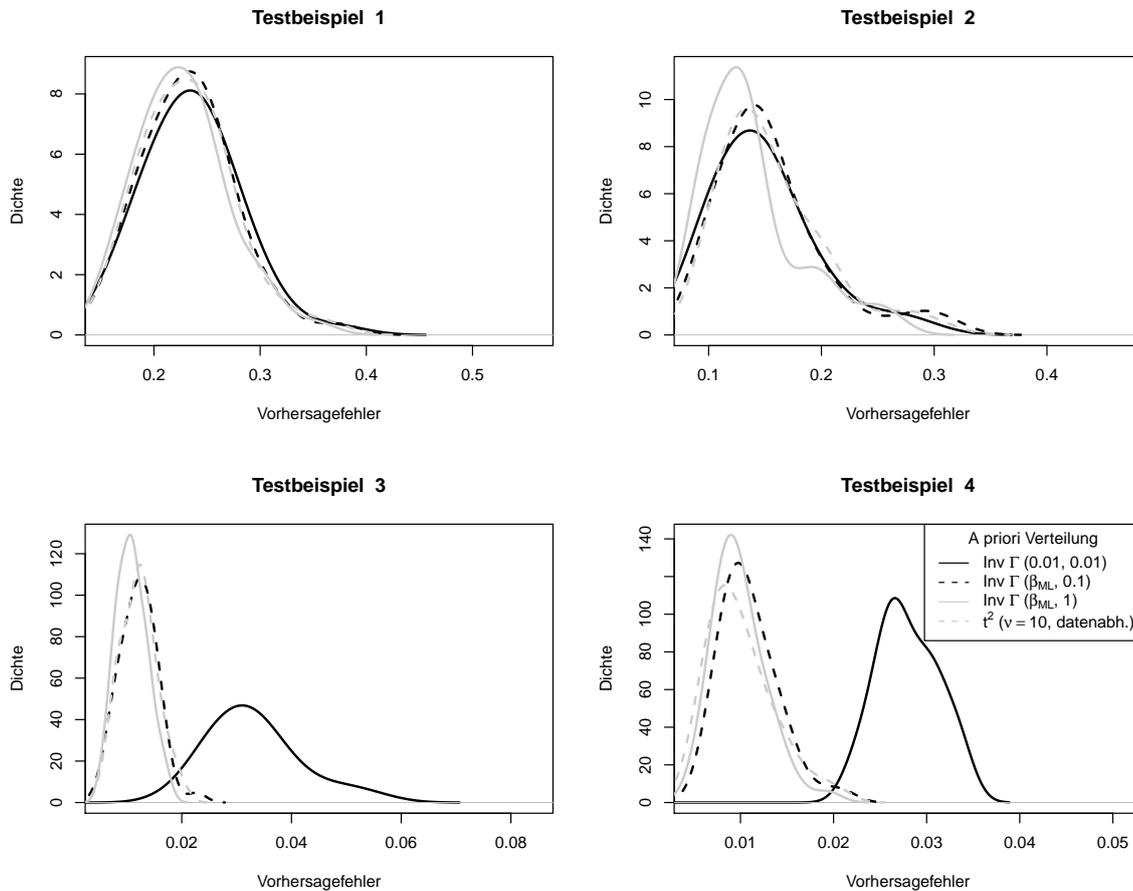


Abbildung 4.4: Vorhersagefehler für unterschiedliche a priori Verteilungen

Verteilungen für  $\sigma_b^2$  und  $\sigma_E^2$  gewählt wurden, desto kleiner wurde der Vorhersagefehler. Nun wird als zweite Zielgröße in der Simulationsstudie der Schätzfehler der angepassten Kalibrierungsmodelle bezüglich  $\sigma_E^2$  betrachtet. Dieser ist definiert als die Wurzel der mittleren quadratischen Abweichung der a posteriori Stichprobe vom wahren Wert für  $\sigma_E^2$ . Die Resultate der unterschiedlichen Szenarien werden in Abbildung 4.5 verglichen. Es zeigen sich auch hier für viele der a priori Verteilungen nur kleine Unterschiede bei den unterschiedlichen a priori Verteilungen. Ein deutlich größerer Schätzfehler resultiert nur für die  $\text{Inv}\Gamma(0.01, 0.01)$  Verteilung. Da der wahre Wert  $\sigma_E^2 = 0.0001$  in Testbeispiel 3 sehr klein ist, wirkt sich hierbei der geringe monotone Anstieg für kleine Werte bei der inversen Gamma-Verteilung aus. Dieses Problem besitzt die  $t^2$ -Verteilung nicht und kann deshalb bessere Resultate liefern. Analog zum Vorhersagefehler gilt auch hier wieder, dass der Schätzfehler für  $\sigma_E^2$  im Vergleich zu der datenunabhängigen quadrierten

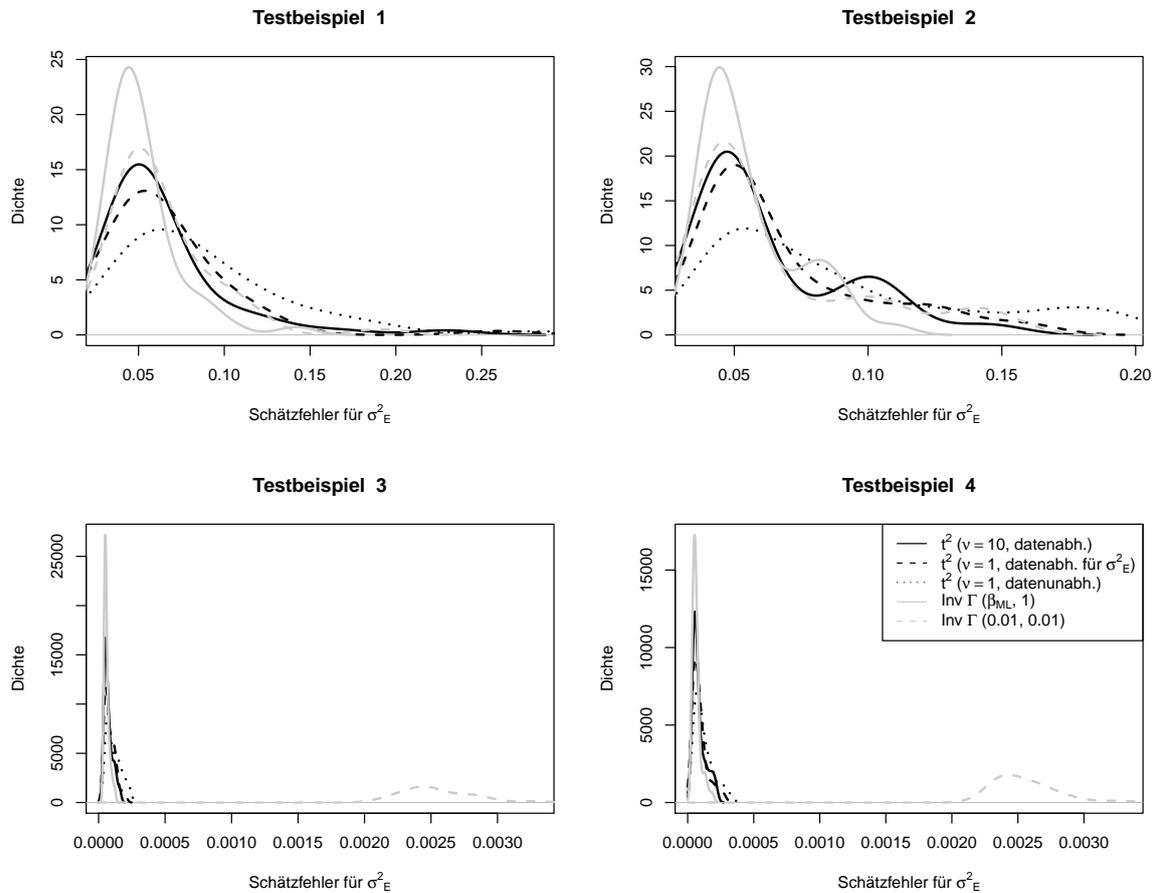


Abbildung 4.5: Schätzfehler für  $\sigma_E^2$  bei unterschiedlichen a priori Verteilungen

Cauchyverteilung kleiner wird, wenn für  $\sigma_E^2$  eine datenabhängige a priori gewählt wird. Der Schätzfehler wird noch kleiner, wenn neben  $\sigma_E^2$  auch die a priori für  $\sigma_b^2$  datenabhängig gewählt wird, insbesondere bei Verwendung der inversen Exponentialverteilung.

Nun soll die a posteriori Verteilung von  $\sigma_b^2$  im Kalibrierungsmodell bei unterschiedlichen a priori Verteilungen verglichen werden. Diese ist kein direktes Gütemaß, hilft aber das Kalibrierungsmodell besser zu verstehen. Für jedes Szenario wurden dazu die a posteriori Stichproben für  $\sigma_b^2$  aus den 50 Simulationsläufen zu einer gemeinsamen Stichprobe zusammengefasst. Abbildung 4.6 zeigt die resultierenden Kerndichteschätzer für die Szenarien 1, 5, 6 und 7. Die a posteriori Verteilung reagiert sehr sensitiv auf die unterschiedlichen a priori Verteilungen. Bei Verwendung der datenabhängigen inversen Exponentialverteilung werden kleine Werte für  $\sigma_b^2$  bevorzugt. Im Gegensatz

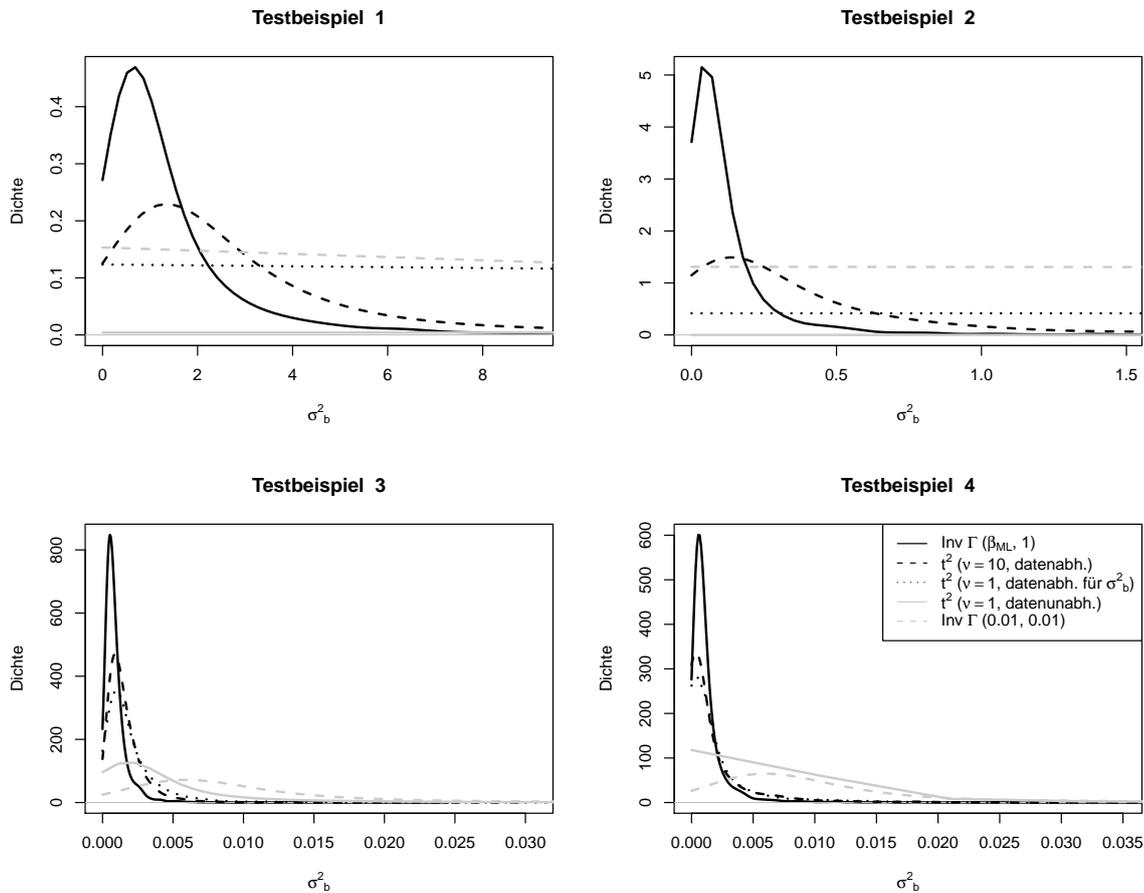


Abbildung 4.6: A posteriori Verteilung für  $\sigma_b^2$  bei unterschiedlichen a priori Verteilungen

dazu ist die a posteriori Verteilung für  $\sigma_b^2$  bei Verwendung einer datenunabhängigen a priori Verteilung für  $\sigma_b^2$  sehr flach und es liegen viele große Werte in der a posteriori Stichprobe vor. Interessanterweise trifft dies bei den Testfunktionen  $Y_E^1$  und  $Y_E^2$  auch für das Szenario zu bei dem nur für  $\sigma_b^2$  eine datenabhängige quadrierte Cauchyverteilung verwendet wurde. Große Werte von  $\sigma_b^2$  stehen im Kalibrierungsmodell für große Verzerrungen. Selbst bei den Testfunktionen ohne Verzerrung  $Y_E^2$  und  $Y_E^4$ , bei denen ein  $\sigma_b^2$  von Null geeignet wäre, schlägt sich dies insbesondere bei  $f_M^2$  nicht in engeren a posteriori Verteilungen nieder. Der Grund für die Sensitivität von  $\sigma_b^2$  gegenüber den unterschiedlichen a priori Verteilungen liegt sicher in den Identifizierbarkeitsproblemen von  $\sigma_b^2$  im Kalibrierungsmodell. Die Wahl der a priori Verteilung für  $\sigma_b^2$  scheint somit eine wichtige Modellspezifikation. Für die betrachteten Testbeispiele liefern datenunabhängige a priori Verteilungen häufig keine realistischen a posteriori Verteilungen.

Abschließend werden nun die marginalen a posteriori Verteilungen für die Kalibrierparameter  $\mathbf{u}$  betrachtet. Auch hier wird für eine bessere Übersichtlichkeit eine gemeinsame a posteriori Stichprobe aus den 50 Simulationsläufen für die Schätzung von  $u_j$ ,  $j = 1, \dots, p_u$  für jede Testfunktion und jede a priori Verteilung generiert. Abbildung 4.7 zeigt die Kerndichteschätzungen. Durch eine vertikale Linie wird dort jeweils der wahre Wert  $u_j^*$  gekennzeichnet. Bei allen betrachteten Testfunktionen gibt es zwischen den marginalen a posteriori Verteilungen für  $u_j$  erstaunlich wenige Unterschiede. Lediglich die a posteriori Verteilungen für  $u_1$  bei  $Y_E^2$ ,  $Y_E^3$  und  $Y_E^4$  variieren etwas. Die Unterschiede bei der a posteriori Verteilung von  $\sigma_b^2$  spiegeln sich somit nicht in unterschiedlichen a posteriori Verteilung der  $u_j$  wieder. Zum Teil sind die a posteriori Verteilungen recht flach und nicht immer können die wahren Werte identifiziert werden. Für die Testfunktionen 2, 3 und 4 gleicht die a posteriori Verteilung für  $u_2$  jeweils nahezu einer Gleichverteilung und somit der gewählten a priori Verteilung. Es konnten also nur wenige Informationen für diese Parameter aus den Daten gewonnen werden. Dies gilt selbst für die eigentlich unverzerrten Testfunktionen 2 und 4. Für die Testfunktion 3 (Kolbensimulator mit Verzerrung) konzentriert sich die a posteriori Verteilung für  $u_1$  nicht um den wahren Wert  $u_1^*$ , sondern um 0.4. Hier kann offensichtlich die Verzerrung besser durch Verwendung eines anderen Kalibrierparameter beschrieben werden. Dies entspricht jedoch nicht dem theoretischen Kleinste Quadrate Schätzer.

Insgesamt zeigt sich, dass mit dem gewählten Kalibrierungsansatz nicht immer eine gute Schätzung für den Kalibrierparameter  $\mathbf{u}$ , das heißt wenig verzerrt und mit kleiner Varianz, erreicht werden kann. Dies ist für verzerrte Computermodelle nicht verwunderlich, da hierbei die Definition eines wahren Wertes  $\mathbf{u}^*$  für  $\mathbf{u}$  immer etwas beliebig ist, so dass eine perfekte Schätzung von  $\mathbf{u}^*$  nicht zu erwarten ist. Allerdings kann selbst bei unverzerrtem Computermodell nicht immer eine gute Schätzung erreicht werden. In diesen Situationen sind weitere Experimente erforderlich, mit denen der Parameter im vorliegenden nichtlinearen Modell gut geschätzt werden kann.

Interessanterweise sind der Vorhersagefehler für den Gesamtprozess und die a posteriori Verteilung für  $\mathbf{u}$  vergleichsweise robust gegenüber den a priori Verteilungen für die Prozessvarianzen, trotz der a posteriori Sensitivität von  $\sigma_b^2$ . Ist das alleinige Ziel der

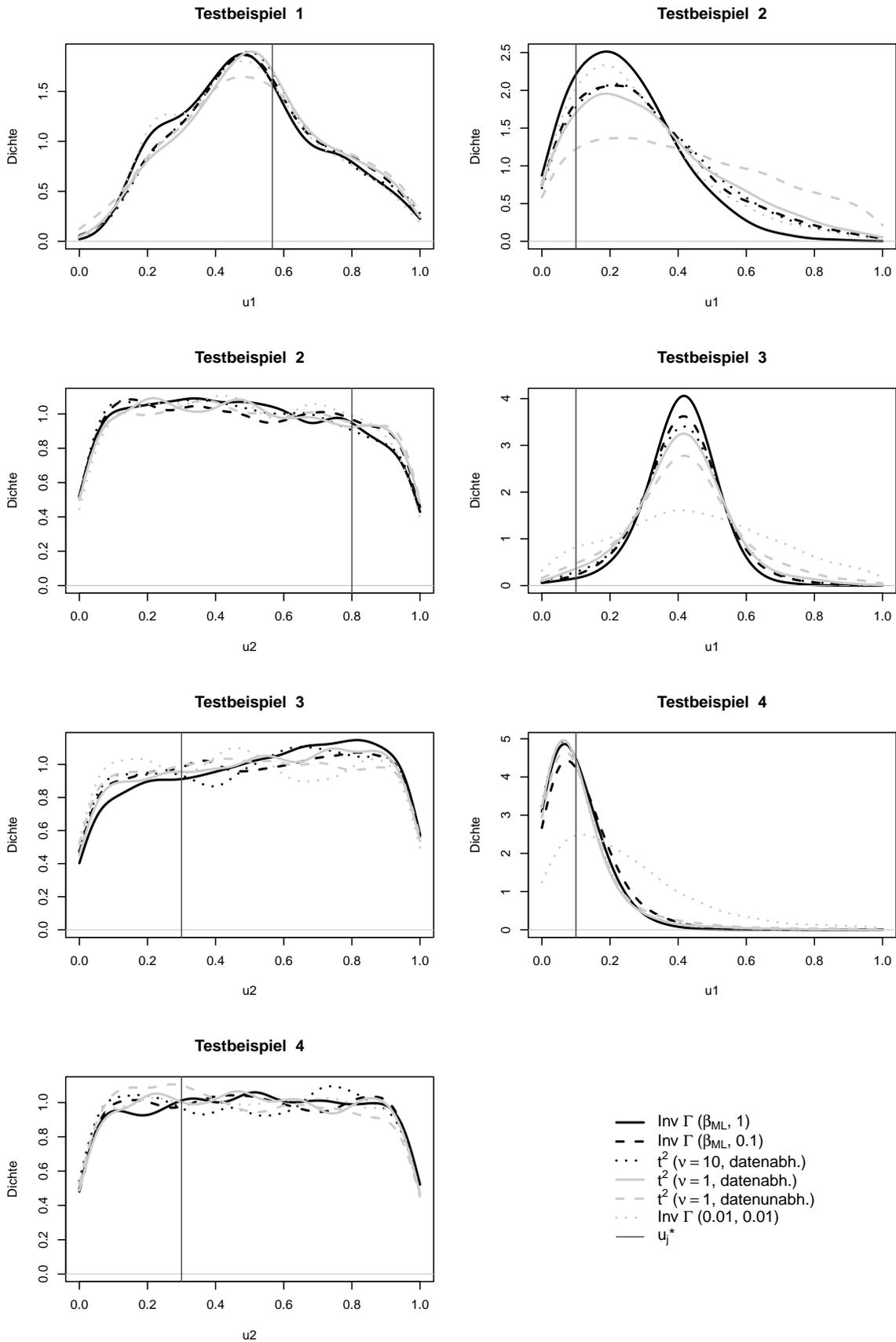


Abbildung 4.7: A posteriori Verteilung für  $u_j$  bei unterschiedlichen a priori Verteilungen

Kalibrierung die Vorhersage des Gesamtprozesses  $Y_E$ , so liefern alle datenabhängigen a priori Verteilungen für  $\sigma_b^2$  ähnlich gute Vorhersagen mit einem kleinen Vorteil für die datenabhängige inverse Exponentialverteilung. Geringfügig schlechter sind die datenunabhängigen  $t^2$ -Verteilungen für  $\sigma_b^2$ . Auch für den Schätzfehler bezüglich  $\sigma_E^2$  wiederholt sich dieses Verhalten. Lediglich die datenunabhängige inverse Gammaverteilung liefert bei einigen der vorliegenden Testfunktionen schlechtere Resultate.

Dahingegen ist die a posteriori Verteilung von  $\sigma_b^2$  deutlich sensitiver gegenüber den unterschiedlichen a priori Verteilungen. Bei extrem schwach informativen a priori Verteilungen resultieren flache a posteriori Verteilungen für  $\sigma_b^2$  einhergehend mit großer Verzerrung. Aus dem Kalibrierungsmodell lässt sich jedoch nicht ohne weiteres beantworten, ob eine große Verzerrung prinzipiell a priori auszuschließen ist. Wenn geeignet sind jedoch möglichst kleine Verzerrungen wünschenswert und tragen zur Interpretierbarkeit des Modells bei.

Insgesamt können informative datenabhängige a priori Verteilungen im Kalibrierungsmodell empfohlen werden. Diese resultieren in Modellen mit vergleichsweise kleiner Verzerrung und liefern so besser interpretierbare Modelle. Bei den betrachteten Testfunktionen konnte so außerdem eine bessere Prozessvorhersage und zuverlässigere Schätzung der Fehlervarianz erreicht werden. Soll eine datenunabhängige a priori Verteilung gewählt werden, so ist die quadrierte  $t$ -Verteilung der inversen Gammaverteilung vorzuziehen. Die Schätzung für den Kalibrierparameter ist hingegen wenig sensitiv gegenüber den unterschiedlichen a priori Verteilungen. Sie liefert jedoch selbst in Situationen in denen ein unverzerrtes Computermodell vorliegt nicht immer besonders gute Resultate. Auch wenn insgesamt informative, datenabhängige a priori Verteilungen empfohlen werden können, ist zu beachten, dass bei einer datenabhängigen Wahl die a priori Verteilungen wesentlich von den dafür verwendeten ML Schätzern abhängen. Eine schlechte Schätzung kann unter Umständen in einer weniger geeigneten a priori Verteilung resultieren.



## 5 KALIBRIERUNG VON COMPUTERMODELLEN UNTER MODELLUNSICHERHEIT

In Kapitel 4 wurden bereits Methoden für die Kalibrierung von Computermodellen vorgestellt. Dabei ist im Wesentlichen zwischen Verfahren zu unterscheiden, die von einem vollständig korrekten Computermodell ausgehen und solchen, die im Modell einen Term für systematische Verzerrungen zwischen Computermodell und Realität enthalten. In diesem Kapitel wird eine alternative Methode vorgestellt, die beide Ansätze zu einem Modell kombiniert oder alternativ einen der beiden Ansätze auswählt.

In Situationen in denen das Computermodell deutlich von der Realität abweicht, ist das Modell ohne Verzerrungsfunktion aus Kapitel 4.2 nicht für eine Kalibrierung geeignet. Kennedy und O’Hagan (2001) und Bayarri *et al.* (2007b) verwenden deshalb für die Kalibrierung ein Modell mit Verzerrungsfunktion. Es resultiert ein flexibles Modell, das den realen Prozess mit Unterstützung des Computermodells vorhersagt. Durch die zusätzliche Verzerrungsfunktion ist das Modell sehr komplex: In der allgemeinen Modellgleichung  $\mu_E(\mathbf{x}) = y_M(\mathbf{x}, \mathbf{u}) + b_{\mathbf{u}}(\mathbf{x})$  aus (4.2) lassen sich systematische Abweichungen zwischen  $\mu_E(\mathbf{x})$  und  $y_M(\mathbf{x}, \mathbf{u})$  immer durch  $b_{\mathbf{u}}(\mathbf{x})$  ausgleichen – und zwar für beliebiges  $\mathbf{u}$ . Eine Identifizierbarkeit der Verzerrungsfunktion wird durch a priori Annahmen erreicht (siehe Kapitel 4.3).

Ein großer Nachteil des Modells mit Verzerrung ist die schwierige Interpretierbarkeit der Ergebnisse. Die Vorhersagen lassen sich nicht alleine aus dem Computermodell und der a posteriori Verteilung für den Kalibrierparameter  $\mathbf{u}$  ableiten, sondern es muss zusätzlich die Verzerrungsfunktion berücksichtigt werden. Durch die flexible Modellierung der Verzerrung mit einem Gauß Prozess ist die vorliegende Verzerrung dem Anwender in höheren Dimensionen nur schwer zu beschreiben und damit nur schwer interpretierbar. Bestehen in der Realität keine (oder nur sehr kleine) Abweichungen von Computermodell und Prozessmittelwert, so ist das wesentlich einfachere Modell ohne Verzerrungs-

funktion zu bevorzugen. Hierbei sind dann auch effizientere Vorhersagen zu erwarten. Für den Anwender bietet dieses Modell den Vorteil, dass Vorhersagen allein mit dem bestehenden Computermodell durchgeführt werden. Damit ist es prinzipiell möglich, dass die Ergebnisse der Kalibrierung für das Computermodell auch auf Bereiche über den betrachteten Designraum  $\mathcal{X}$  hinaus extrapoliert werden könnten. Ein Modell mit Verzerrungsfunktion kann hingegen ohne zusätzliche physikalische Experimente nicht extrapoliert werden. Zusammengefasst ist also das Modell ohne Verzerrungsfunktion zu bevorzugen, wenn genügend Evidenz vorliegt, dass dieses eine geeignete Approximation der Realität bietet. Im Gegenzug führt jedoch bei einer wesentlichen Verzerrung die Kalibrierung im Sinne einer nichtlinearen Regression nicht zu einem geeigneten Modell und es ist das Modell mit Verzerrung zu bevorzugen.

In diesem Kapitel wird eine Erweiterung des Kalibrierungsmodells aus Kapitel 4.3 um eine Modellunsicherheitskomponente vorgeschlagen. Dazu wird eine zusätzliche Variable eingeführt, die es erlaubt sich zwischen dem Modell mit Verzerrung und dem Modell ohne Verzerrung hin- und her zu bewegen. Auf diesem Weg soll einerseits mit Hilfe von Daten entschieden werden, ob ein unverzerrtes Computermodell vorliegt und außerdem durch Regularisierung des Verzerrungsterms die effizienteren Schätzmöglichkeiten bei einem unverzerrten Computermodell genutzt werden.

## 5.1 Kalibrierungsmodell mit Modellunsicherheitskomponente

Im Folgenden wird für die Kalibrierung das Modell

$$Y_E(\mathbf{x}) = y_M(\mathbf{x}, \mathbf{u}) + \gamma \cdot b_u(\mathbf{x}) + e \quad (5.1)$$

mit unabhängig identisch verteiltem Fehler  $e \sim N(0, \sigma_E^2)$  betrachtet. Dabei bezeichnet  $\gamma$  eine Indikatorvariable, die entweder den Wert 0 oder 1 annimmt. So entspricht das Modell (5.1) für  $\gamma = 0$  dem erweiterten nichtlinearen Regressionsmodell in (4.3) und für  $\gamma = 1$  dem Kalibrierungsmodell mit Verzerrung aus Kapitel 4.3 mit Erwartungswert aus (4.2). Insgesamt wird damit der Erwartungswert  $E[Y_E(\mathbf{x})]$  in (5.1) entweder allei-

ne durch das Computermodell  $y_M(\mathbf{x}, \mathbf{u}^*)$  oder durch Addition von Computermodell  $y_M(\mathbf{x}, \mathbf{u}^*)$  und Verzerrungsfunktion, jeweils mit geeignetem  $\mathbf{u}^*$ , beschrieben.

Für  $y_M(\cdot)$  und  $b_{\mathbf{u}}(\cdot)$  werden analog zu Kapitel 4.3 unabhängige Gauß Prozesse der Form  $\tilde{y}_M(\cdot) \sim GP(f_M(\cdot)\beta_M, \sigma_M^2 R_M(\cdot, \cdot))$  beziehungsweise  $\tilde{b}(\cdot) \sim GP(0, \sigma_b^2 R_b(\cdot, \cdot))$  mit Gaußschen Korrelationsfunktionen  $R_M(\cdot, \cdot)$  und  $R_b(\cdot, \cdot)$  angenommen. Dabei wird für die Verzerrung ein Gauß Prozess mit Erwartungswert 0 berücksichtigt. Für die Korrelationsparameter  $\boldsymbol{\psi}_M$  und  $\boldsymbol{\psi}_b$  werden ihre ML Schätzer eingesetzt.

Die Likelihood im neuen Modell aus Gleichung (5.1) ändert sich im Vergleich zu dem Modell aus Kapitel 4.3 nur wenig. Die gemeinsame Likelihood für  $\mathbf{y}_E, \mathbf{y}_M$  lässt sich wieder als Produkt von Dichten der suffizienten Statistiken  $\bar{\mathbf{y}}_E, s_E^2$  und  $\mathbf{y}_M$  darstellen:

$$l(\bar{\mathbf{y}}_E, s_E^2, \mathbf{y}_M | \mathbf{u}, \sigma_b^2, \sigma_E^2, \gamma, \boldsymbol{\psi}_b, \boldsymbol{\theta}_M) = \varphi(\mathbf{y}_M | f_M(\mathbf{D}_M)\beta, C_M(\mathbf{D}_M)) \cdot p(s_E^2 | \sigma_E^2) \cdot \varphi(\bar{\mathbf{y}}_E | \hat{y}_M(\boldsymbol{\Delta}_E^u), \hat{\Sigma}_M^*(\boldsymbol{\Delta}_E^u) + \gamma \cdot \sigma_b^2 \cdot R_b[\boldsymbol{\Delta}_E] + \overline{\mathbf{S}}_E). \quad (5.2)$$

In der Likelihood Komponente für  $\bar{\mathbf{y}}_E$  wird hier in der Kovarianzmatrix die Indikatorvariable  $\gamma$  vor  $\sigma_b^2$  multipliziert. Ist  $\gamma = 0$ , so entfällt der additive Term, der zu der Kovarianzstruktur von  $\tilde{b}(\cdot)$  korrespondiert. Die resultierende Kovarianzstruktur entspricht der Struktur, die ohne Verzerrungsfunktion vorliegt. Der Erwartungswert verändert sich nicht, da für  $\tilde{b}(\cdot)$  stets ein Erwartungswert von 0 angenommen wird.

Das neue Modell in (5.1) ist nun noch mit geeigneten a priori Verteilungen zu versehen. Die a priori Verteilungen für  $\sigma_b^2$ ,  $\sigma_E^2$  und  $\mathbf{u}$  werden analog zu Kapitel 4.3 gewählt. Für die Prozessvarianzen kann zum Beispiel eine der in Kapitel 4.4 diskutierten a priori Verteilungen wie die Inverse Exponentialverteilung oder quadrierte t-Verteilung, verwendet werden und für den Kalibrierparameter beispielsweise eine Gleichverteilung. Zusätzlich ist in diesem Modell eine a priori Verteilung für die Modellindikatorvariable  $\gamma$  zu wählen.

Da  $\gamma$  nur die Werte 0 und 1 annimmt, wird a priori eine Bernoulliverteilung mit Erfolgswahrscheinlichkeit  $\pi$  angenommen:  $\gamma | \pi \sim \text{Bin}(1, \pi)$ . Für  $\pi$  kann in einem hierarchischen Modellansatz ebenfalls eine a priori Verteilung gewählt werden. Da in der Regel keine a priori Informationen über diesen Parameter vorliegen, sollte eine möglichst nichtinfor-

mative a priori Verteilung für  $\gamma$ , wie eine Gleichverteilung auf  $[0, 1]$ , d.h.  $\pi \sim \mathcal{U}(0, 1)$ , gewählt werden. Soll auf eine hierarchische a priori Verteilung verzichtet werden, so bietet sich für  $\pi$  die Einpunktverteilung in 0.5 an. Die resultierende a priori für  $\gamma$ ,  $\gamma \sim \text{Bin}(1, 0.5)$ , entspricht der marginalen a priori, die sich durch Herausintegrieren von  $\pi$  unter der Gleichverteilungsannahme im hierarchischen Modell ergibt.

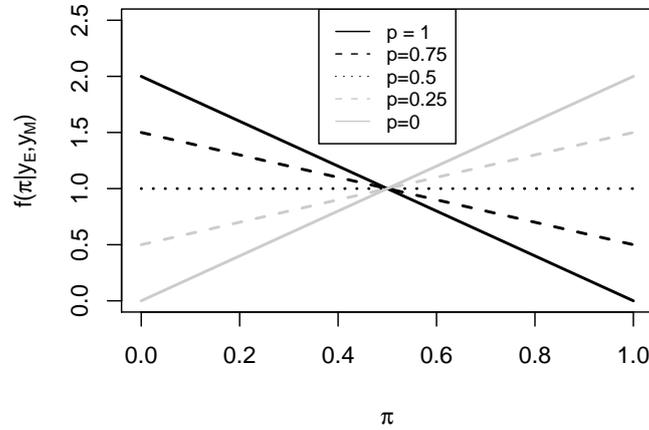
Um den hierarchischen und nichthierarchischen a priori Ansatz für  $\gamma$  zu vergleichen, wird im nächsten Schritt die a posteriori Verteilung für  $\pi$  bei Verwendung einer Gleichverteilung für  $\pi$  hergeleitet und untersucht. Die a posteriori Verteilung für  $\pi$  lässt sich durch

$$p(\pi|\mathbf{y}_E, \mathbf{y}_M) = \sum_{\gamma=0}^1 p(\pi|\gamma, \mathbf{y}_E, \mathbf{y}_M)p(\gamma|\mathbf{y}_E, \mathbf{y}_M)$$

über Integration von  $p(\pi|\gamma, \mathbf{y}_E, \mathbf{y}_M)$  bezüglich des Zählmaßes  $\gamma|\mathbf{y}_E, \mathbf{y}_M$  bestimmen. Durch den hierarchischen Modellansatz hängt  $\pi$  alleine über  $\gamma$  von  $\mathbf{y}_E$  und  $\mathbf{y}_M$  ab (vergleiche auch O'Hagan und Forster, 2004, S. 151). Somit gilt  $p(\pi|\gamma, \mathbf{y}_E, \mathbf{y}_M) = p(\pi|\gamma)$ . Unter der a priori Annahme  $\pi \sim \mathcal{U}(0, 1)$  folgt  $\pi|\gamma \sim \text{Beta}(1 + \gamma, 2 - \gamma)$  einer Beta-Verteilung mit Parametern  $1 + \gamma$  und  $2 - \gamma$ . Die a posteriori Verteilung für  $\pi$  ist damit insgesamt gegeben durch

$$\begin{aligned} p(\pi|\mathbf{y}_E, \mathbf{y}_M) &= \sum_{\gamma=0}^1 f_{\text{Beta}}(\pi|1 + \gamma, 2 - \gamma)p(\gamma|\mathbf{y}_E, \mathbf{y}_M) \\ &= p \cdot f_{\text{Beta}}(\pi|1, 2) + (1 - p) \cdot \text{Beta}(\pi|2, 1), \end{aligned} \quad (5.3)$$

einer gewichteten Mischung der Dichten  $f_{\text{Beta}}$  einer  $\text{Beta}(1, 2)$  und  $\text{Beta}(2, 1)$  Verteilung mit Gewichten  $p$  und  $1 - p$ . Dabei ist  $p = P(\gamma = 0|\mathbf{y}_E, \mathbf{y}_M)$  die a posteriori Modellwahrscheinlichkeit für das Modell ohne Verzerrungsfunktion. Die a posteriori für  $\pi$  lässt sich also geschlossen in Abhängigkeit von  $p$  darstellen. Abbildung 5.1 zeigt sie für verschiedene Werte von  $p$ . Selbst für extreme Werte wie  $p = 0$  oder  $p = 1$  ist die a posteriori Verteilung wenig informativ. Für  $p = 0.5$  ergibt sich als a posteriori die Gleichverteilung und damit genau die gewählte a priori. Durch diese Betrachtung wird deutlich, dass in dem vorliegenden Modell nur sehr wenig Information über den

Abbildung 5.1: A posteriori Verteilung für  $\pi$ 

Hyperparameter  $\pi$  zu gewinnen ist. Es liegt jeweils nur eine Beobachtung  $\gamma$  für die Herleitung der a posteriori Verteilung von  $\pi$  vor. Anders als in hierarchischen Modellen häufig üblich, wird hier kein Parameter einer Population beschrieben.

Die hierarchische a priori Verteilung für  $\gamma$  mit Hyperparameter  $\pi$  lässt somit keinen wesentlichen Erkenntnisgewinn zu. Von diesem Standpunkt ist es ausreichend  $p(\gamma) \sim \text{Bin}(1, 0.5)$ , also eine Einpunktverteilung für  $\pi$  in 0.5 zu wählen. Auf diesen Fall werde ich mich im Folgenden beschränken.

Im nächsten Schritt soll eine alternative Interpretation des Kalibrierungsmodell unter Modellunsicherheit vorgestellt werden. Das beschriebene Kalibrierungsmodell mit einer zusätzlichen Modellindikatorvariable  $\gamma$  lässt sich alternativ auch als Kalibrierungsmodell aus Kapitel 4.3 mit einer speziellen a priori Verteilung für die Prozessvarianz des Verzerrungsterms auffassen. Das Produkt  $b^*(\cdot) = \gamma \cdot \tilde{b}(\cdot)$  von Modellindikatorvariable und Gauß Prozess für die Verzerrungsfunktion ist wieder ein Gauß Prozess  $\tilde{b}^*(\cdot) \sim GP(0, \sigma_b^{2*} R_b(\cdot, \cdot))$ , mit Prozessvarianz  $\sigma_b^{2*} = \gamma \cdot \sigma_b^2$ . Für  $\gamma = 0$  gilt  $\sigma_b^{2*} = 0$  und somit fällt der Gaußprozess für  $\tilde{b}^*(\cdot)$  auf einen Einpunktprozess in 0 zusammen. In (5.1) resultiert das Kalibrierungsmodell ohne Verzerrungsfunktion. Die Modellindikatorvariable lässt sich durch  $\gamma = \mathbb{1}(\sigma_b^{2*} = 0)$  als Funktion von  $\sigma_b^{2*}$  darstellen, so dass auf diesen Parameter verzichtet werden kann. Insgesamt ergibt sich wieder das ursprüngli-

che Kalibrierungsmodell diesmal mit einem Gauß Prozess  $b^*(\cdot)$  als a priori Annahme für die Verzerrungsfunktion mit Prozessvarianz  $\sigma_b^{2*}$ . Statt die Likelihood durch Hinzufügen von  $\gamma$  anzupassen, wird bei dieser Parametrisierung lediglich für  $\sigma_b^{2*}$  eine veränderte a priori Verteilung verwendet. Diese besitzt die Form einer Mischverteilung

$$p(\sigma_b^{2*}|\pi) = \begin{cases} 1 - \pi & \sigma_b^{2*} = 0 \\ \pi \cdot p(\sigma_b^2) & \sigma_b^{2*} > 0, \end{cases} \quad (5.4)$$

mit der a priori Verteilung  $p(\sigma_b^2)$  für die ursprüngliche Prozessvarianz.

Das Modell unter Modellunsicherheit lässt sich somit als Regularisierung des Verzerrungsterms durch eine a priori Verteilung für  $\sigma_b^{2*}$  verstehen. Dabei wird eine Varianz  $\sigma_b^{2*} = 0$  und damit ein unverzerrtes Computermodell bevorzugt.

### 5.1.1 Vorhersagen für die Kalibrierung unter Modellunsicherheit

Das betrachtete Modell mit Unsicherheit über die Verzerrungsfunktion aus (5.1) lässt sich für Vorhersagen auf zwei unterschiedliche Arten und Weisen nutzen. Klassischerweise wird im Bayes Ansatz die a posteriori Verteilung aller Parameter für die Erstellung der Vorhersageverteilung eingesetzt. Im vorliegenden Modell resultieren die Vorhersagen dann durch ein gewichtetes Mittel der Kalibrierungsmodelle mit und ohne Verzerrungsfunktion. Alternativ lässt sich die Unsicherheit über die Verzerrungsfunktion auch als Modellwahlproblem auffassen. Dabei wird zunächst eine Entscheidung für oder gegen das Vorliegen einer Verzerrung getroffen und die Vorhersage nur auf Basis des ausgewählten Modells erstellt. Betrachten wir zunächst die Vorhersage durch Modellmittelung. Formal wird die Vorhersage im Bayes Ansatz durch Integration bezüglich der a posteriori Verteilung erstellt:

$$\begin{aligned} p(Y_E(\cdot)|\mathbf{y}_E, \mathbf{y}_M) &= \int p(Y_E(\cdot)|\gamma, \boldsymbol{\theta}) \cdot p(\gamma, \boldsymbol{\theta}|\mathbf{y}_E, \mathbf{y}_M) d(\boldsymbol{\theta}, \gamma) \\ &= \sum_{\gamma=0}^1 p(\gamma|\mathbf{y}_E, \mathbf{y}_M) \int p(Y_E(\cdot)|\boldsymbol{\theta}) \cdot p(\boldsymbol{\theta}|\mathbf{y}_E, \mathbf{y}_M, \gamma) d\boldsymbol{\theta}, \end{aligned} \quad (5.5)$$

dabei fasst der Vektor  $\boldsymbol{\theta}$  alle übrigen Modellparameter  $(\sigma_b^2, \sigma_E^2, \mathbf{u}, \pi)$  zusammen. Die Vorhersage in (5.5) ist ein gewichtetes Mittel für die zwei möglichen Realisationen von  $\gamma$ , wobei die Gewichte der a posteriori Wahrscheinlichkeit für  $\gamma = 0$  beziehungsweise  $\gamma = 1$  entsprechen. Auf die Verzerrungsfunktion wird bei dieser Betrachtung mit Ausnahme des Extremfalls  $p(\gamma = 0 | \mathbf{y}_E, \mathbf{y}_M) = 1$  nicht vollständig verzichtet. Somit wird die Modellgleichung gegenüber dem Bayesianischen Kalibrierungsmodell aus Kapitel 4.3 nicht reduziert. Allerdings kann hier durch die Mischung der Modelle eine sinnvolle Regularisierung der Verzerrungsfunktion erreicht werden. Das lässt bei weniger deutlichen Ergebnissen für die a posteriori Modellwahrscheinlichkeit  $p$  bessere Vorhersageeigenschaften erhoffen (siehe die Literatur zum Bayesian Model Averaging, z.B. Hoeting *et al.*, 1999).

Bei der Modellwahl wird zunächst das Modell – bzw. genauer  $\gamma^*$  – mit maximaler a posteriori Wahrscheinlichkeit ausgewählt. Als Vorhersageverteilung wird anschließend  $\int p(Y_E(\cdot) | \boldsymbol{\theta}) \cdot p(\boldsymbol{\theta} | \mathbf{y}_E, \mathbf{y}_M, \gamma = \gamma^*) d\boldsymbol{\theta}$ , die Vorhersageverteilung bedingt auf das Modell mit  $\gamma = \gamma^*$  verwendet. Bei der Ermittlung der Vorhersagen mit Modellwahl handelt es sich somit nicht um einen vollständigen Bayes Ansatz, da die Unsicherheit über die Modellwahl nicht in der Vorhersageverteilung berücksichtigt wird. Damit ist zu erwarten, dass Unsicherheitsintervalle zu kurz sind und sich eventuell auch die Vorhersagegüte verschlechtert. Der große Vorteil der Modellwahl gegenüber dem vollständigen Bayes Ansatz liegt jedoch in der besseren Interpretierbarkeit der Resultate. Gilt  $p(\gamma = 0 | \mathbf{y}_E, \mathbf{y}_M) \geq 0.5$  so wird bei der Kalibrierung von Computermodellen das weniger komplexe Modell ohne Verzerrungsfunktion ausgewählt. Die Vorhersagen basieren dann im Wesentlichen auf dem Computermodell und der Unsicherheit bezüglich des Kalibrierparameters  $\mathbf{u}$ . Andernfalls, wenn  $p(\gamma = 0 | \mathbf{y}_E, \mathbf{y}_M) < 0.5$ , wird das Modell mit Verzerrung ausgewählt und die Vorhersagen entsprechen denen aus dem Bayesianischen Kalibrierungsmodell aus Kapitel 4.3. Als Grenze für die Modellwahl wurde hier 0.5 gewählt. Es ist jedoch zu berücksichtigen, dass eine starke Evidenz für eines der Modelle erst bei einer deutlich von 0.5 verschiedenen a posteriori Modellwahrscheinlichkeit vorliegt. Kass und Raftery (1995) liefern eine Hilfestellung für die Interpretation von Bayes Faktoren beziehungsweise posterior odds, die hier  $p/(1 - p)$  entsprechen.

Für das Kalibrierungsmodell besitzt das Vorgehen mit Modellwahl den Vorteil, dass wirklich eine Entscheidung für ein Modell ohne Verzerrungsfunktion getroffen werden kann. Dieses ist anschließend für den Anwender wesentlich einfacher zu interpretieren und Vorhersagen sind schneller berechenbar, da nur Computermodellvorhersagen unter Berücksichtigung der Unsicherheit über  $\mathbf{u}$  erforderlich sind. Allerdings kann diese Reduktion der Modellkomplexität nur erreicht werden, wenn auch tatsächlich das Modell ohne Verzerrungsfunktion ausgewählt wird.

Im Folgenden werden beide Ansätze parallel betrachtet. Für die Herleitung der a posteriori Verteilung ist der gewählte Ansatz zur Vorhersage zunächst unerheblich.

### 5.1.2 A posteriori Verteilung für Kalibrierung unter Modellunsicherheit

Für das beschriebene Modell zur Kalibrierung unter Modellunsicherheit aus Gleichung (5.1) kann bei Herleitung der a posteriori Verteilung im Prinzip analog zu dem Kalibrierungsmodell aus Kapitel 4.3 vorgegangen werden. Neben den bisher betrachteten Parametern  $\sigma_E^2$ ,  $\sigma_b^2$  und  $\mathbf{u}$  kommt nun noch der Parameter  $\gamma$  für die Modellunsicherheit hinzu.

Beginnen wir mit der Herleitung der a posteriori Modellwahrscheinlichkeit  $p = P(\gamma = 0 | \mathbf{y}_E, \mathbf{y}_M)$ . Sie lässt sich mit Hilfe des Bayes Theorems durch

$$p = \frac{p(\mathbf{y}_E, \mathbf{y}_M | \gamma = 0) \cdot P(\gamma = 0)}{p(\mathbf{y}_E, \mathbf{y}_M | \gamma = 0) \cdot P(\gamma = 0) + p(\mathbf{y}_E, \mathbf{y}_M | \gamma = 1) \cdot P(\gamma = 1)} \quad (5.6)$$

bestimmen, wobei  $p(\mathbf{y}_E, \mathbf{y}_M | \gamma)$  die marginale Likelihood bezeichnet und für die a priori Wahrscheinlichkeiten  $P(\gamma = 0) = P(\gamma = 1) = 0.5$  angenommen wird. Die marginale Likelihood ergibt sich durch Herausintegrieren aller übrigen Parameter  $\boldsymbol{\theta} = (\sigma_b^2, \sigma_E^2, \mathbf{u})^T$  aus der vollständigen Likelihood:  $p(\mathbf{y}_E, \mathbf{y}_M | \gamma) = \int p(\mathbf{y}_E, \mathbf{y}_M | \boldsymbol{\theta}, \gamma) \cdot p(\boldsymbol{\theta} | \gamma) d\boldsymbol{\theta}$ . Dies entspricht der Normalisierungskonstanten der a posteriori Verteilung von  $\boldsymbol{\theta}$  gegeben der Modellindikatorvariable  $\gamma$ .

Stichproben aus der a posteriori Verteilung der übrigen Parameter  $\sigma_b^2$ ,  $\sigma_E^2$  und  $\mathbf{u}$  sind durch numerische Methoden aus dem Produkt von a priori und Likelihood herzuleiten.

Geeignete Implementierungsansätze hierfür werden im Folgenden betrachtet.

### 5.1.3 Implementierung der Kalibrierung unter Modellunsicherheit

Für die Implementierung der Kalibrierung unter Modellunsicherheit werden hier zwei unterschiedliche Strategien verfolgt. Einerseits wird ein Random Walk Metropolis-Hastings Algorithmus und andererseits ein Importance Sampling Algorithmus basierend auf der iterativen Laplace Approximation verwendet. Ein Gibbs Sampler lässt sich mit der zusätzlichen Modellindikatorvariable auch bei Verwendung konjugierter a priori Verteilungen nicht anwenden, da keine irreduzible Markovkette erzeugt werden kann.

Zunächst wird kurz auf die Probleme beim Gibbs Sampler eingegangen. Durch die zusätzliche Modellindikatorvariable  $\gamma$  ist die erforderliche so genannte Positivitätsbedingung nicht länger erfüllt. Dies lässt sich leicht durch folgende Situation nachvollziehen. Betrachte  $\mathbf{b}^* = \gamma \mathbf{b}$ , den Verzerrungsvektor am Experimentdesign mit  $\mathbf{b} = (\tilde{b}(\mathbf{x}_1), \dots, \tilde{b}(\mathbf{x}_n))^T$ , der im Gibbs Algorithmus als zusätzliche Variable benötigt wird. Sei nun  $\gamma^{(i)} = 1$  und  $\mathbf{b}^{*(i)}$ , das  $i$ -te Markovkettenelement des Verzerrungsvektors, enthalte Werte ungleich 0. Die bedingte Wahrscheinlichkeit für  $\gamma^{(i+1)} = 0$  gegeben  $\mathbf{b}^* = \mathbf{b}^{*(i)}$  ist damit stets 0. Umgekehrt ist die bedingte Verteilung für  $\mathbf{b}^*$  gegeben  $\gamma^{(i)} = 0$  stets die Einpunktverteilung in dem Nullvektor. Die Übergangswahrscheinlichkeit zwischen den beiden unterschiedlichen Modellen ist somit nicht positiv. Damit ist die resultierende Markovkette nicht irreduzibel, so dass der Gibbs Sampler hier nicht zu der gewünschten a posteriori Verteilung konvergiert.

Im nächsten Schritt wird die Implementierung mit dem Metropolis-Hastings Algorithmus beschrieben. Das Kalibrierungsmodell mit Modellwahl lässt sich über  $\sigma_b^{2*} = \gamma \sigma_b^2$  wie oben beschrieben zu dem ursprünglichen Kalibrierungsmodell ohne Modellwahl mit einer Mischungsverteilung als a priori für  $\sigma_b^{2*}$  umparametrisieren. Deshalb kann weiterhin der Metropolis-Hastings Algorithmus angewendet werden und es sind nur wenige Modifikationen im Vergleich zu dem Kalibrierungsmodell aus Kapitel 4.3 erforderlich. Insbesondere kann hier durch die Reparametrisierung des Modells auf den Reversible-

Jump MCMC Algorithmus (Green, 1995) verzichtet werden, der bei einer variablen Dimension des Parametervektors angewendet werden muss, wie sie sonst in der Regel bei Modellwahlproblemen auftritt.

In der vorliegenden Situation ist für den Metropolis-Hastings Algorithmus lediglich sicherzustellen, dass die Markovkette für  $\sigma_b^{2*}$  auch genau den Wert 0 annehmen kann. Die Vorschlagsdichte für  $\sigma_b^{2*}$  wird deshalb als Mischung aus einer Einpunktverteilung bei 0 und einer modifizierten Random Walk Vorschlagsdichte  $q_+(\sigma_b^{2*(t)}, \sigma_b^{2*(t-1)})$  gewählt

$$q(\sigma_b^{2*(t)}, \sigma_b^{2*(t-1)}) = \begin{cases} \rho_b & \sigma_b^{2*(t)} = 0 \\ (1 - \rho_b) \cdot q_+(\sigma_b^{2*(t)}, \sigma_b^{2*(t-1)}) & \sigma_b^{2*(t)} > 0 \end{cases}$$

mit dem Mischungsfaktor  $\rho_b$  als Tuningparameter. Für  $q_+(\cdot, \cdot)$  wird hier der Betrag einer normalverteilten Zufallsvariable mit altem Wert als Erwartungswert und Standardabweichung  $s_b$  als Tuningparameter:

$$q_+(\sigma_b^{2*(t)}, \sigma_b^{2*(t-1)}) = \varphi(\sigma_b^{2*(t)} | \sigma_b^{2*(t-1)}, s_b^2) \cdot (1 - \Phi(-\sigma_b^{2*(t-1)} / s_b)) + \\ \varphi(-\sigma_b^{2*(t)} | \sigma_b^{2*(t-1)}, s_b^2) \cdot (\Phi(-\sigma_b^{2*(t-1)} / s_b))$$

verwendet. Um  $\sigma_b^{2*} \geq 0$  zu erhalten, wird hierbei die ursprüngliche Dichte der Normalverteilung für negative Werte durch Betragsbildung dem positiven Bereich hinzugefügt. Gewichtung durch die Verteilungsfunktion stellt sicher, dass wieder eine Wahrscheinlichkeitsdichte resultiert. Eine Transformation von  $\sigma_b^{2*}$  auf die log-Skala, wie im Metropolis-Hastings Algorithmus in Kapitel 4.3, ist hier ungeeignet, da der Logarithmus in Null divergiert. Für die übrigen Parameter  $\sigma_E^2$  und  $\mathbf{u}$  werden die Vorschlagsdichten analog zur Implementierung ohne Modellwahlkomponente eingesetzt.

Voraussetzung für eine schnelle Konvergenz des Metropolis-Hastings Algorithmus ist, dass sich die Parameter  $\sigma_E^2$  und  $\mathbf{u}$  zwischen den zwei Modellen nur wenig unterscheiden. Ist dies der Fall kann leicht zwischen den beiden Modellen gewechselt werden. Im vorliegenden Kalibrierungsmodell sollte diese Übertragbarkeit zumindest für kleine  $\sigma_b^2$  gewährleistet sein. Dennoch ist eine gute Einstellung des Algorithmuses über die Tuningparameter erforderlich, um zu verhindern, dass die Markovkette im Metropolis-

Hastings Algorithmus in einem der beiden Modelle stecken bleibt.

Als alternative Implementierungsmethode bietet sich bei Modellwahlproblemen die Verwendung eines Importance Sampling Algorithmus an (siehe Robert, 2007, S. 356 ff). Die Vorschlagsdichten werden hier wieder mit der iterativen Laplace Approximation bestimmt (Bornkamp, 2011, siehe Anhang A.1). Beim Importance Sampling für Modellwahlprobleme wird der Fall  $\gamma = 0$  unabhängig von  $\gamma = 1$  betrachtet und jeweils ein Importance Sample generiert. Anschließend werden die zwei Stichproben entsprechend der a posteriori Modellwahrscheinlichkeit, das heißt  $p = P(\gamma = 0 | \mathbf{y}_E, \mathbf{y}_M)$  und  $1 - p$  gewichtet. Diese ergibt sich wie in Gleichung (5.6) dargestellt in Abhängigkeit von  $\int p(\mathbf{y}_E, \mathbf{y}_M | \boldsymbol{\theta}, \gamma) \cdot p(\boldsymbol{\theta} | \gamma) d\boldsymbol{\theta}$ . Damit muss das Integral über das Produkt von Likelihood und a priori gegeben  $\gamma$  bestimmt werden. Dieses Integral entspricht genau der Normalisierungskonstante des Produkts aus Likelihood und a priori, dass für jede Realisation von  $\gamma$  zur Generierung einer a posteriori Stichprobe verwendet wird. Normalisierungskonstanten können bei Verwendung des Importance Sampling leicht geschätzt werden (siehe Kapitel A.1). Deshalb lässt sich in diesem Fall auch die gemeinsame a posteriori Verteilung aus den einzelnen a posteriori Verteilungen leicht bestimmen. Dieses Vorgehen bietet den Vorteil, dass bei der Herleitung der a posteriori Verteilung auf jeden Fall sowohl die Modellsituation mit und ohne Verzerrung berücksichtigt wird. Analog könnte prinzipiell auch bei Verwendung eines Metropolis-Hastings Algorithmus vorgegangen werden. Allerdings ist dort die Berechnung der Normalisierungskonstanten über Integration anhand der a posteriori Stichprobe wesentlich rechenintensiver, so dass dort eine gemeinsame Markovkette für beide Modellsituationen geeigneter ist.

## 5.2 Simulationsstudie zur Kalibrierung unter Modellunsicherheit

In diesem Kapitel wird die vorgestellte Methode zur Kalibrierung unter Modellunsicherheit in einer Simulationsstudie auf die vier Testbeispiele aus Kapitel 4.4.1 angewendet und mit der Bayesianischen Kalibrierung aus Kapitel 4.3 verglichen. Die Daten werden analog zu der Simulationsstudie in Kapitel 4.4.2 erzeugt. In 50 Simulationsläufen werden Computerexperimente jeweils nach einem Maximin-LHC Design mit einem Stich-

Szenario	Modell	Verzerrung	a priori für $\sigma_b^2$ und $\sigma_E^2$
1	1	$b(\mathbf{x})$	$t^2$
2	2	$\gamma \cdot b(\mathbf{x})$	$t^2$
3	3	0 oder $b(\mathbf{x})$	$t^2$
4	1	$b(\mathbf{x})$	inv. Exp.
5	2	$\gamma \cdot b(\mathbf{x})$	inv. Exp.
6	3	0 oder $b(\mathbf{x})$	inv. Exp.

Tabelle 5.1: Simulationsszenarien zur Kalibrierung unter Modellunsicherheit

probenumfang  $n_M = 10 \cdot d_M$  erhoben, wobei  $d_M$  der Dimension des Computermodells entspricht. Die experimentellen Beobachtungen werden anhand eines äquidistanten Designs auf 5 Stufen mit jeweils 3 Wiederholungen generiert. Auf jeden Datensatz wird einerseits die Bayesianische Kalibrierung mit Verzerrung (Modell 1) und andererseits die hier vorgestellte Kalibrierung unter Modellunsicherheit mit  $\pi = 0.5$  (Modell 2) angewendet. Neben Vorhersagen durch Modellmittelung werden dabei auch Vorhersagen nach Modellwahl (Modell 3) betrachtet. Als a priori Verteilungen für die Prozessvarianzen  $\sigma_b^2$  und  $\sigma_E^2$  werden die datenabhängige inverse Exponentialverteilung sowie die datenabhängige quadrierte  $t$ -Verteilung mit 10 Freiheitsgraden verwendet. Somit werden insgesamt 6 Szenarien (siehe Tabelle 5.1) betrachtet.

Es wird folgende Implementierungsstrategie angewendet. Zunächst wird eine a posteriori Stichprobe der Größe 1000 mit Hilfe des Importance Sampling Algorithmus auf Basis von 5000 Zufallszahlen erhoben. Bei einer äußerst niedrigen effektive Stichprobengröße und damit geringer Approximationsgüte sowie bei numerischen Problemen der iterativen Laplace Approximation wird auf den Metropolis-Hastings Algorithmus zurückgegriffen. Dabei werden zunächst die Tuningparameter in Testläufen so eingestellt, dass eine geeignete Akzeptanzrate erreicht wird, wobei ein Wert von 23 % angestrebt wird (Roberts und Rosenthal, 2001). Es wird eine Markovkette der Länge 11000 erzeugt. Entfernen der ersten 1000 Elemente und Ausdünnen der Kette liefert eine a posteriori Stichprobe vom Umfang 1000.

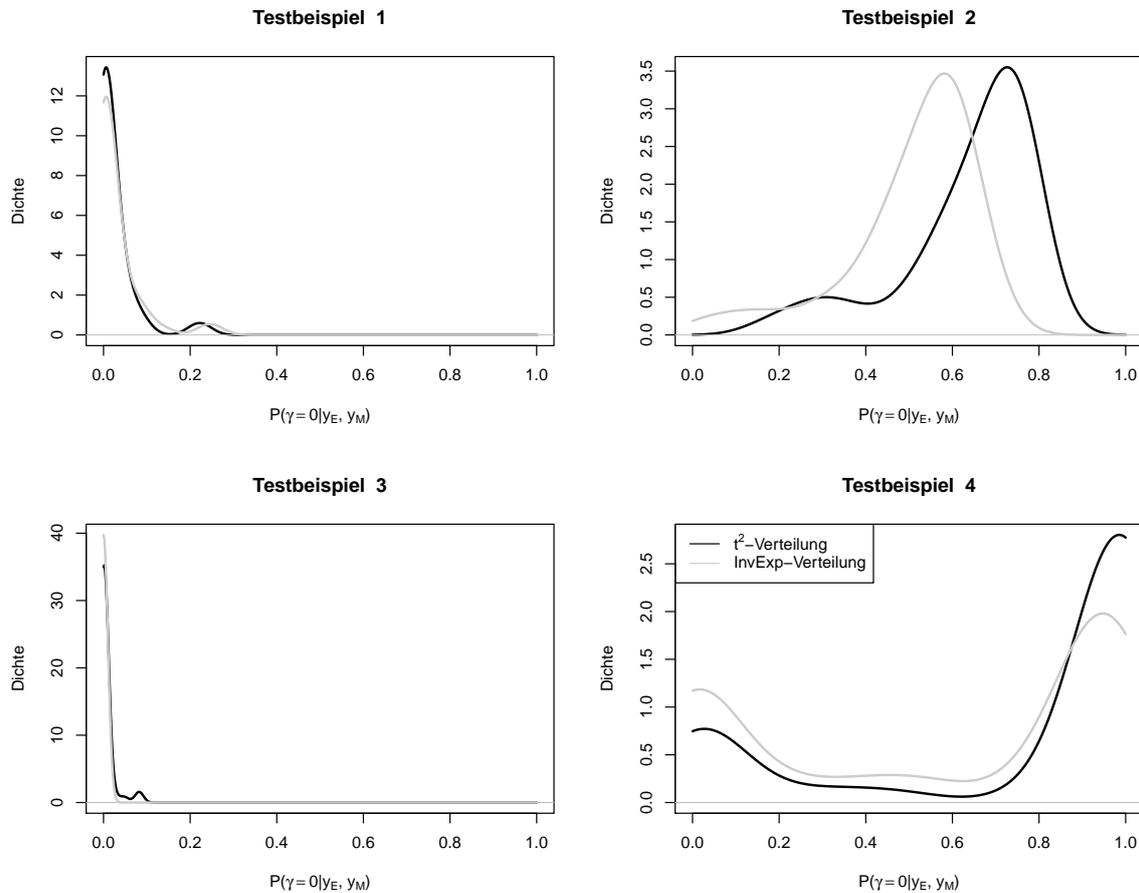


Abbildung 5.2: A posteriori Modellwahrscheinlichkeit  $p = P(\gamma = 0 | \mathbf{y}_M, \mathbf{y}_E)$

### 5.2.1 Ergebnisse

Zunächst werden die Schätzungen der a posteriori Wahrscheinlichkeit für das Modell ohne Verzerrung  $p = P(\gamma = 0 | \mathbf{y}_E, \mathbf{y}_M)$  bei Kalibrierung unter Modellunsicherheit (Modell 2), dargestellt in Abbildung 5.2, betrachtet. Die Ergebnisse aus den 50 Simulationsläufen werden durch Kerndichteschätzer zusammengefasst. Für die Testbeispiele 1 und 3 mit Verzerrung wäre  $p = 0$  und für die Testbeispiele 2 und 4 ohne Verzerrung  $p = 1$  optimal. Die Simulationsergebnisse für die Testbeispiele ohne Verzerrung liegen jeweils in der Nähe des optimalen Wertes 0. Die Ergebnisse sind unabhängig von der gewählten a priori Verteilung und die meisten Werte sind kleiner 0.2 bzw. 0.1, jedoch alle kleiner 0.5. Damit wird für diese beiden Testbeispiele bei Anwendung der vorgestellten Modellwahlstrategie stets das korrekte Modell mit Verzerrungsfunktion ausgewählt. Für

die Testbeispiele 2 und 4 ohne Verzerrung werden wie gewünscht eher größere Werte angenommen. Allerdings sind hier die Realisationen weniger stark um den optimalen Wert 1 konzentriert. Zum Beispiel unterscheidet sich für  $Y_E^2$  die Häufigkeitsverteilung zwischen den unterschiedlichen a priori Verteilungen. Im Modellwahlansatz wird in 84% ( $t^2$ -Verteilung) bzw. 62% (Inverse Exponentialverteilung) der Fälle das korrekte Modell ohne Verzerrungsfunktion ausgewählt. Für  $Y_E^4$  liegt für beide a priori Verteilungen eine bimodale Häufigkeitsverteilung vor: es werden Werte an den beiden Rändern des Parameterbereichs für  $p$  angenommen, wobei große Modellwahrscheinlichkeiten häufiger vorkommen. Auch hier sind die Resultate bei Verwendung der  $t^2$ -Verteilung als a priori geringfügig näher dem optimalen Wert 1. In 74 % ( $t^2$ -Verteilung) bzw. 60% (Inverse Exponentialverteilung) der 50 Simulationsläufe wird im Modellwahlansatz das korrekte Modell ohne Verzerrungsfunktion ausgewählt. In beiden Situationen in denen in der Realität keine Verzerrung vorliegt, werden in einigen Simulationsläufen auch a posteriori Modellwahrscheinlichkeiten in der Nähe von 0 für das richtige Modell realisiert. Diese sind somit weiter entfernt von dem wahren Wert 1 als die gewählte a priori Wahrscheinlichkeit von 0.5. Anhand dieser Ergebnisse wird deutlich, dass es wesentlich leichter ist eine Verzerrung aufzudecken als ein unverzerrtes Computermodell zu identifizieren. Für die Entscheidung, dass eine Verzerrung vorliegt, genügt es einen nicht passenden Datensatz zu finden. Im Gegenzug muss wesentlich mehr Information über  $Y_E$  und  $y_M$  gewonnen werden, um genügend Evidenz zu sammeln, dass keine Verzerrung vorliegt. Dieses Phänomen untersuchen Johnson und Rossell (2010) theoretisch in einer ähnlichen Modellsituation.

Im nächsten Schritt werden die Modelle hinsichtlich der Vorhersagegüte verglichen. Abbildung 5.3 zeigt den Vorhersagefehler für die 4 Testbeispiele und unterschiedliche Simulationsszenarien. Für die Testbeispiele 1 und 3 unterscheidet sich der Vorhersagefehler für das Bayesianische Kalibrierungsmodell mit Verzerrung (Modell 1) und das Modell mit Indikatorvariable (Modell 2) kaum. Da in beiden Testbeispielen für die Kalibrierung unter Modellunsicherheit stets eine a posteriori Modellwahrscheinlichkeit  $p < 0.5$  für die Situation ohne Verzerrung realisiert wurde, entsprechen die Vorhersagen unter Modellwahl (Modell 3) genau denen aus der ursprünglichen Bayesianischen

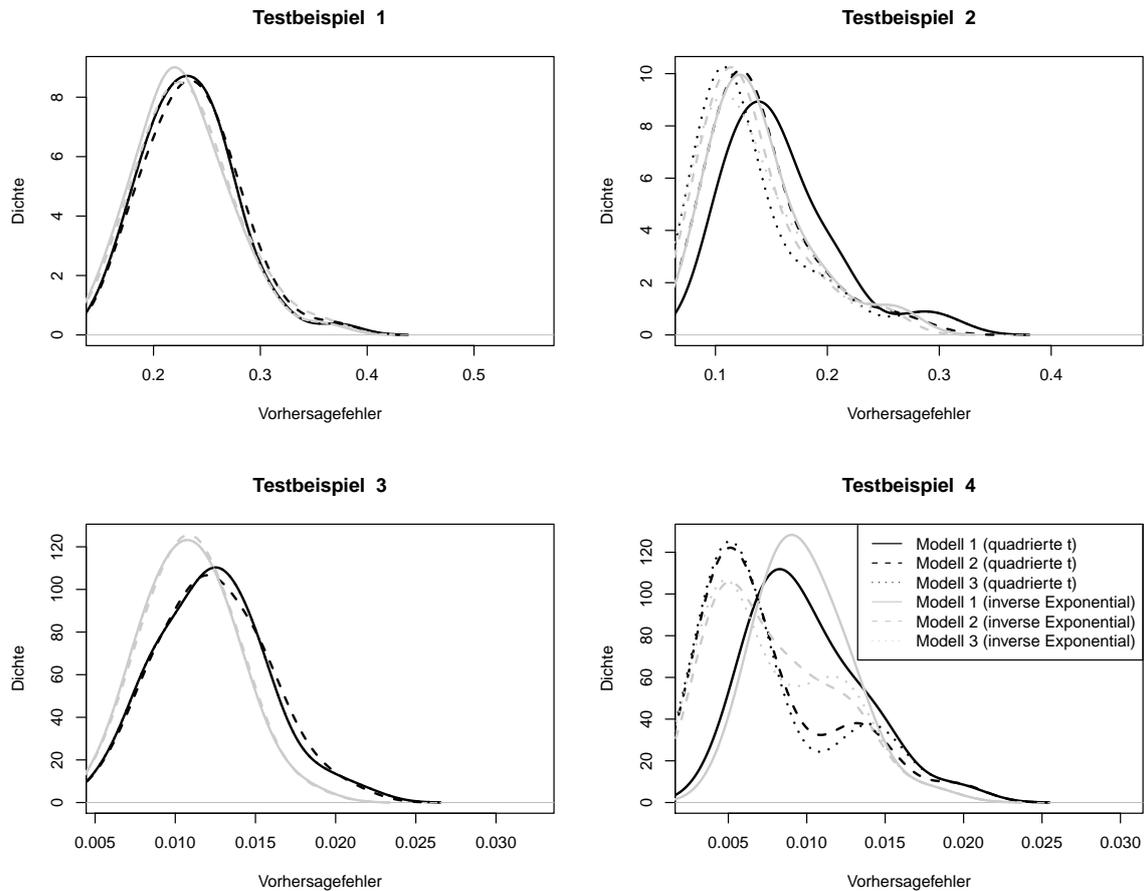
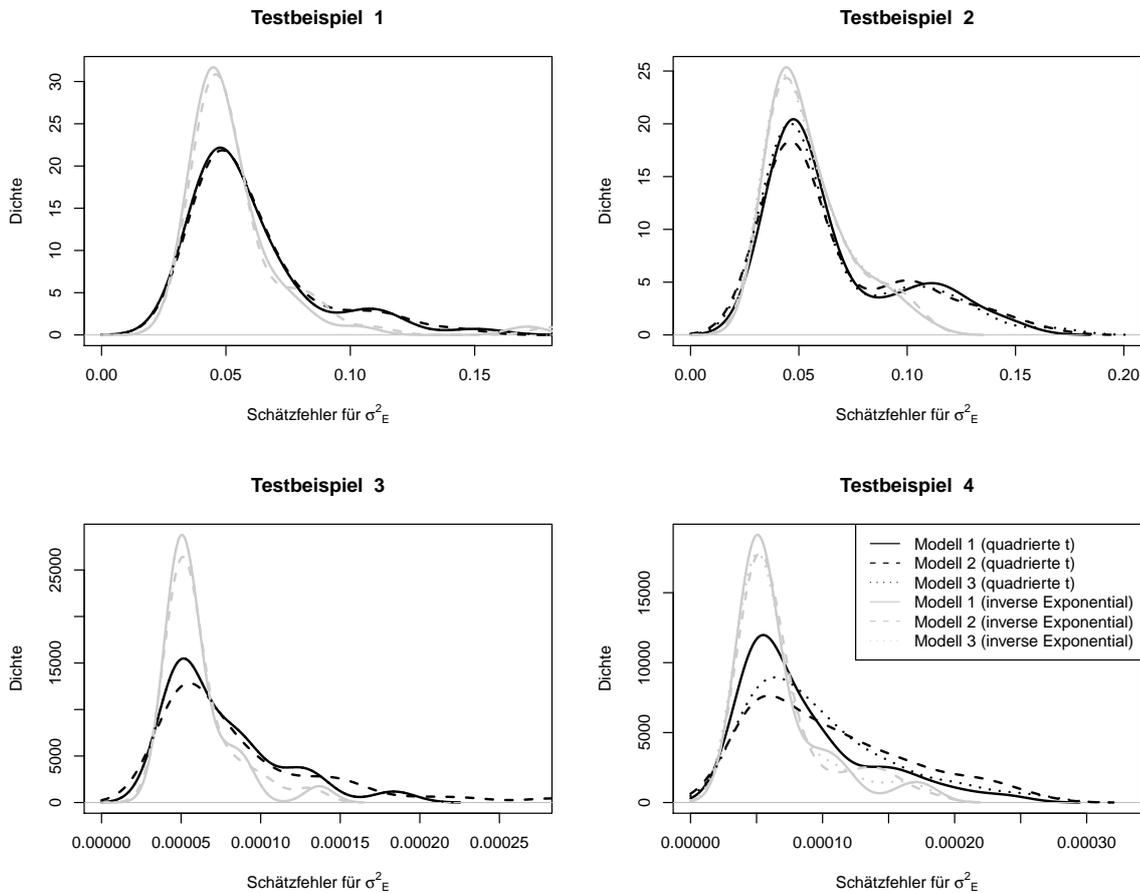


Abbildung 5.3: Vorhersagefehler für den Gesamtprozess

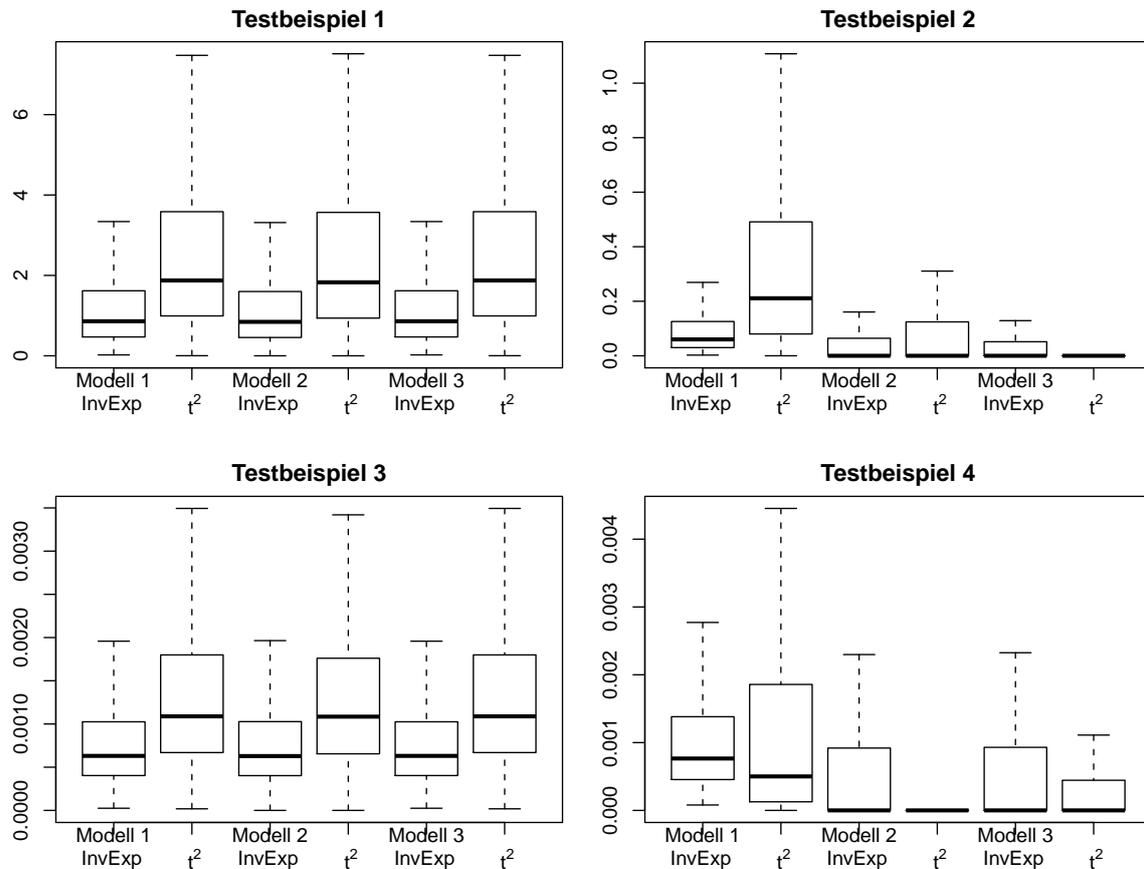
Kalibrierung mit Verzerrung (Modell 1). Der Vorhersagefehler für das Modell mit Verzerrung (Modell 1 und 3) ist hier etwa gleich groß oder geringfügig kleiner als bei Verwendung des Modells 2. Die Berücksichtigung der überflüssigen zusätzlichen Modellsituation ohne Verzerrungsfunktion in Modell 2 und 3 geht somit nicht mit einem wesentlichen Verlust in der Vorhersagegüte einher.

Für die Testbeispiele ohne Verzerrung verbessert sich hingegen die Vorhersagegüte bei Verwendung des Kalibrierungsmodells unter Modellunsicherheit (Modell 2 und 3). Insbesondere bei Testbeispiel 4 ist eine deutliche Verbesserung zu beobachten. Dabei unterscheidet sich die Vorhersagegüte nach Modellwahl oder Modellmittelung nur wenig und insbesondere ist keine Vorgehensweise gleichmäßig besser. Hinsichtlich der zwei betrachteten a priori Verteilungen ist die Vorhersagegüte für die inverse Exponentialverteilung in einigen Situationen geringfügig besser. Nur für die Kalibrierung unter

Abbildung 5.4: Schätzfehler für  $\sigma_E^2$ 

Modellunsicherheit im Testbeispiel 4 ist die  $t^2$ -Verteilung überlegen. Insgesamt ist die Kalibrierung unter Modellunsicherheit – sowohl bei Verwendung der Modellwahl als auch der Modellmittelung – der Kalibrierung mit Verzerrung nicht unterlegen und schneidet für unverzerrte Computermodelle besser ab.

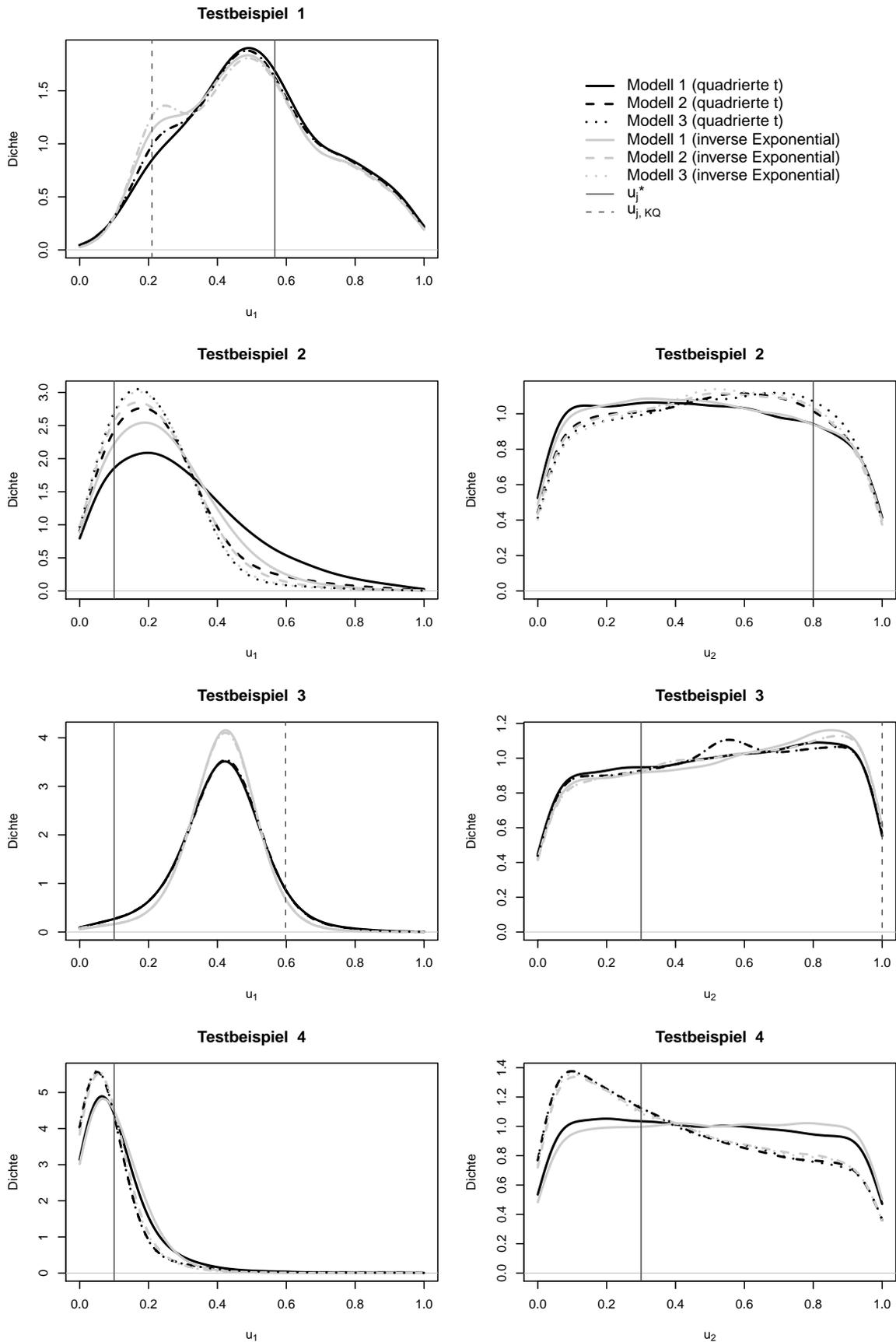
Als dritte Zielgröße wird der Schätzfehler für  $\sigma_E^2$  betrachtet, das heißt die Wurzel der mittleren quadratischen Abweichung der a posteriori Stichprobenelemente zum wahren Wert. Abbildung 5.4 zeigt die Resultate der Simulationsstudie. Im Gegensatz zum Vorhersagefehler steigt der Schätzfehler für  $\sigma_E^2$  zum Teil geringfügig bei Verwendung der Kalibrierung unter Modellunsicherheit (Modell 2 und 3) im Vergleich zu dem Kalibrierungsmodell aus Kapitel 4.3 (Modell 1) an. Durch die vorliegenden Wiederholungen im Experimentdesign ist davon auszugehen, dass die experimentelle Streuung bei der Kalibrierung mit Verzerrung sehr genau allein über Wiederholungen geschätzt wer-

Abbildung 5.5: A posteriori Verteilung für  $\sigma_b^{2*}$ 

den kann. Bei der Kalibrierung ohne Verzerrung kann die Schätzung der Fehlervarianz durch Abweichungen von Simulation und Realität verzerrt sein, wie in Kapitel 4.2.1 illustriert. Dies bewirkt insgesamt eine geringfügig schlechtere Schätzung der Fehlervarianz. Unabhängig von der gewählten Methode werden bei Verwendung der inversen Exponentialverteilung als a priori Verteilung etwas bessere Schätzungen von  $\sigma_E^2$  erzielt. Nun wird die a posteriori Verteilung für  $\sigma_b^{2*}$  verglichen. Da es sich hierbei in Modell 2 und 3 um eine Mischverteilung mit Punktmasse auf Null handelt, werden Boxplots statt Kerndichteschätzer verwendet. Abbildung 5.5 stellt die Boxplots der Vereinigung der 50 a posteriori Stichproben für  $\sigma_b^{2*}$  für die unterschiedlichen Szenarien ohne Ausreißer dar. Für die Testbeispiele 1 und 3 mit Verzerrung erhielt bei der Kalibrierung unter Modellunsicherheit das Modell ohne Verzerrung nur ein äußerst geringes Gewicht. Somit unterscheiden sich a posteriori Verteilungen von  $\sigma_b^{2*}$  nicht zwischen den

drei unterschiedlichen betrachteten Modellen. Die a priori Punktmasse auf Null für  $\sigma_b^{2*}$  wirkt sich also bei verzerrtem Computermodell nicht auf die a posteriori Verteilung von  $\sigma_b^{2*}$  aus. Liegt jedoch ein unverzerrtes Computermodell vor, wie in den Testbeispielen 2 und 4, so verändert sich auch die a posteriori Verteilung zwischen den verschiedenen Modellen. Wie erwartet und erwünscht wird die a posteriori Verteilung für  $\sigma_b^{2*}$  bei Kalibrierung unter Modellunsicherheit (Modell 2 und 3) im Vergleich zu Modell 1 stärker in Richtung der Null gezogen. Man erhält in diesen Fällen also die gewünschte Regularisierung von  $\sigma_b^{2*}$ .

Abschließend wird nun die a posteriori Verteilung für den Kalibrierparameter  $\mathbf{u}$  betrachtet. Dazu wird die gemeinsame a posteriori Stichprobe der 50 Simulationsläufe für jede Komponente  $u_j$ ,  $j \in 1, \dots, p_u$ , von  $\mathbf{u}$  betrachtet. Abbildung 5.6 zeigt die Kern-dichteschätzer. Durch vertikale Linien werden die wahren Werte  $\mathbf{u}_j^*$  und in den Testbeispielen mit Verzerrung auch die theoretischen kleinsten Quadrate Schätzer  $u_{j,KQ}$  (liefert kleinste  $L_2$ -Norm der Verzerrungsfunktion) visualisiert. Die marginalen a posteriori Verteilungen für  $u_j$  unterscheiden sich zwischen den unterschiedlichen Kalibrierungsmodellen nicht systematisch und nur geringfügig. Die wesentliche Gestalt der a posteriori aus Modell 1, der Bayesianischen Kalibrierungsmodell ohne Modellunsicherheit, bleibt auch bei Betrachtung der Kalibrierung mit Modellunsicherheit erhalten, unabhängig ob die Vorhersagen mit Modellmittelung oder Modellwahl durchgeführt werden (Modell 2 und 3). Bei den Testbeispielen mit Verzerrung werden weder die wahren Kalibrierparameter noch die theoretischen Kleinsten Quadrate Schätzer durch die Modelle 2 und 3 im Vergleich zu Modell 1 stärker bevorzugt. Selbst bei den Testbeispielen ohne Verzerrung verbessert sich die Schätzung des zu Grunde liegenden wahren Kalibrierparameters für Modell 2 und 3 nur leicht. Zum Beispiel wird für  $u_1$  im Testbeispiel 2 die a posteriori Verteilung durch Verwendung der Kalibrierung unter Modellunsicherheit etwas weniger flach – Werte größer als 0.5 besitzen nun nahezu keine Wahrscheinlichkeitsmasse mehr. In anderen Situationen zum Beispiel für  $u_2$  im Testbeispiel 2 ändert sich die a posteriori Verteilung nicht wesentlich, obwohl bei Verwendung der Modellwahlstrategie in 78 % der Fälle das Modell ohne Verzerrungsfunktion ausgewählt wird. Selbst wenn stets das korrekte Modell ohne Verzerrung vorgegeben wird (ohne grafische Darstellung),

Abbildung 5.6: A posteriori Verteilung für  $u_j$

konzentriert sich die a posteriori Verteilung nicht stärker um den wahren Kalibrierparameter  $u_2^*$ . Die vorliegenden Beobachtungen aus Experiment und Computermodell reichen hier somit nicht für eine genauere Schätzung des Kalibrierparameters aus.

Die Simulationsstudie bestätigt insgesamt die erwarteten Ergebnisse für das hier vorgestellte Kalibrierungsmodell mit Modellindikatorvariable. Wird die Kalibrierung mit Modellunsicherheit (Modell 2 und 3) bei verzerrten Computermodellen eingesetzt, ergeben sich nur geringfügige Änderungen der a posteriori Verteilungen und insbesondere keine nennenswerte Verschlechterung der Vorhersagegüte. Dahingegen kann bei Verwendung der Kalibrierung mit Modellindikatorvariable bei einem Computermodell ohne Verzerrung die Vorhersagegüte geringfügig verbessert werden. Allerdings geht dies nicht mit einer wesentlich verbesserten Schätzung der Kalibrierungsparameter einher. In einem Modellwahlansatz kann die vorgestellte Methode verzerrte Computermodelle zuverlässig erkennen, wohingegen unverzerrte Computermodelle prinzipiell – jedoch wesentlich schwerer – identifizierbar sind.

Das hier vorgestellte Kalibrierungsmodell kann auf Basis der durchgeführten Simulationsstudien empfohlen werden, wenn nicht bekannt ist ob ein verzerrtes Computermodell vorliegt. Dann erlaubt es über die Notwendigkeit einer Verzerrungsfunktion zu entscheiden. Im Vergleich zu der Bayesianischen Kalibrierung aus Kapitel 4.3 kann bei der Kalibrierung unter Modellunsicherheit eine bessere Vorhersagegüte, niedrigere Modellkomplexität sowie bessere Interpretierbarkeit bei einem unverzerrtem Computermodell ermöglicht werden, ohne mit einem wesentlichen Effizienzverlust bei der Analyse eines verzerrten Computermodells einherzugehen.

## 6 ANWENDUNG: KALIBRIERUNG EINER RÜCKFEDE- RUNGSSIMULATION

In diesem Kapitel werden zwei Probleme aus der Umformtechnik mit Hilfe der vorgestellten Ansätze zur Kalibrierung von Computermodellen analysiert. Genauer wird die Rückfederung im Tiefziehprozess untersucht. Beim Tiefziehen wird ein Blech durch einen Stempel in eine formgebende Matrize gedrückt. Abbildung 6.1 zeigt das Grundprinzip (Gösling, 2010). Niederhalterkräfte  $F_{NH}$  spannen das Blech an den Rändern in der Tiefziehpresse ein und der Stempel bewegt sich mit einer Geschwindigkeit  $v_{St}$  nach unten. Durch diesen Vorgang passt sich das Blech mit Blechdicke  $d_B$  an die Form der Matrize an. Die Geometrie von Matrize und Stempel wird hier unter anderem durch den Werkzeugradius  $r_{Wz}$  und die Ziehtiefe  $T_Z$  beschrieben. Beim Nachlassen der Kräfte an Stempel und Niederhalter kann das Bauteil aus der Form zurückfedern und die resultierende Ist-Geometrie weicht von der gewünschten Sollgeometrie ab. Eine adäquate Simulation der Rückfederung beim Tiefziehen ist erforderlich, um anschließend eine auf Computermodellen basierende Kompensation oder Minimierung der Rückfederung ohne zahlreiche experimentelle Trial-and-Error Zyklen zu ermöglichen (Gösling *et al.*, 2011).

Abbildung 6.2 zeigt die hier betrachtete Bauteilgeometrie, das so genannte Hutprofil. Links wird die Istgeometrie, das heißt die tatsächlich realisierte Geometrie, dreidimen-

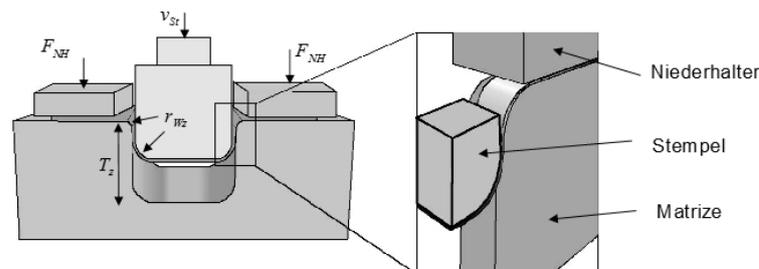


Abbildung 6.1: Tiefziehen Grundprinzip

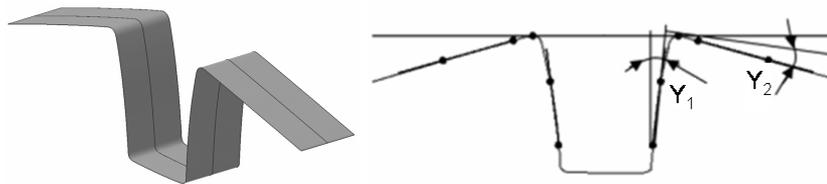


Abbildung 6.2: Werkzeuggeometrie

sional dargestellt. Rechts wird neben der Istgeometrie auch die Sollgeometrie zweidimensional veranschaulicht. Die Rückfederung wird definiert als Abweichung von Soll- und Istgeometrie und kann durch den Zargenwinkel  $Y_1$  und Flanschwinkel  $Y_2$  beschrieben werden. Umformprozesse werden in der Regel durch Computermodelle simuliert, die so genannte Finite-Elemente Methoden zur numerischen Berechnung einsetzen. Es liegen zahlreiche Softwarepakete für die Simulation von Umformprozessen vor. In dieser Arbeit werden die Softwarepakete Abaqus (Dassault Systèmes, 2010) und Autoform (AutoForm Engineering GmbH, 2010) für die Durchführung der Computerexperimente eingesetzt. Der Umformprozess selbst lässt sich in der Regel gut nachbilden. Die Simulation der Rückfederung von Bauteilen ist hingegen verhältnismäßig schwer. Für die Simulation von Umformung und Rückfederung ist die geeignete Wahl unbekannter Parameter im Computermodell erforderlich. Diese sollen hier mit Hilfe von Rückfederungsexperimenten für verschiedene Geometriemodifikationen und Prozessbedingungen geschätzt werden.

In diesem Kapitel liegen zu zwei verschiedenen Rückfederungsprozessen numerische und experimentelle Daten vor. Die beiden Kalibrierungsprobleme unterscheiden sich in dem untersuchten Blechwerkstoff, sowie den betrachteten physikalischen Größen und Kalibrierparametern. In der ersten Anwendung in Kapitel 6.1 liegen als Kalibrierparameter numerische Parameter vor, die in der Computersimulation die Diskretisierung steuern. In der zweiten Anwendung in Kapitel 6.2 handelt es sich um physikalische Parameter, die in der Simulation zur Beschreibung des Werkstoffes eingesetzt werden. Mit Hilfe des kalibrierten Computermodells ist anschließend die Rückfederung des Umformprozesses zu minimieren.

Methodisch werde ich bei der Auswertung wie folgt vorgehen. In Kapitel 6.1 werden die beiden in Kapitel 4 vorgestellten Ansätze angewendet. In Kapitel 6.2 wird die in Ka-

pitel 5 entwickelte Kalibrierung unter Modellunsicherheit eingesetzt und anschließend Ansätze zur Optimierung des Umformprozesses vorgestellt.

## 6.1 Kalibrierung einer Rückfederungssimulation durch numerische Parameter

Es wird die Rückfederung in Simulation und Experiment für den Blechwerkstoff DP600, einen Dualphasenstahl mit einer Mindestzugfestigkeit von 600 MPa, betrachtet. In Experimenten wurde der Werkzeugradius  $r_{Wz}$  (3-5 mm) und die Niederhalterkraft  $F_{NH}$  (500-2500 kP) variiert. Diese beiden physikalischen Größen wurden ebenfalls in einer Computersimulation berücksichtigt (mit geringfügig größeren Wertebereichen  $F_{NH} \in [300, 3000]$  kP und  $r_{Wz} \in [2.2, 7.8]$  mm). Weiterhin wurden in der Computersimulation zwei numerische Einflussgrößen, die das Finite-Elemente Netz beschreiben, variiert. Es wird das Verhältnis der Elementkantenlängen  $V_{EKL}$  (1-6) sowie die Anzahl der Elemente über das Blech  $A_{EB}$  (2-14 Elemente) betrachtet. Beide Größen beschreiben die Diskretisierung des Bleches durch Finite Elemente. Ziel ist es, geeignete numerische Parameterwerte für zukünftige Simulationen zu finden. Es ist zu erwarten, dass das Computermodell mit einer feineren Diskretisierung eine bessere Approximation liefert. Ist jedoch eine gröbere Diskretisierung gleichwertig, so ist diese wegen einer geringeren Rechenzeit zu bevorzugen. Dazu werden die Ansätze zur Kalibrierung von Computermodellen aus Kapitel 4 angewendet.

Die Experimente wurden am Institut für Umformtechnik und Leichtbau der TU Dortmund durchgeführt. Es liegen insgesamt 22 Beobachtungen aus physikalischen Experimenten vor. Dabei wurde der Werkzeugradius und die Niederhalterkraft auf je 3 Stufen,  $r_{Wz} \in \{3, 5, 7\}$  mm und  $F_{NH} \in \{500, 1250, 2500\}$  kP, variiert. Für die Niederhalterkräfte 500 und 1250 kP wurden alle möglichen Kombinationen mit den Werkzeugradien beobachtet, wohingegen für die Niederhalterkraft von 2500 kP nur Beobachtungen mit einem Werkzeugradius von 5 mm vorliegen. Es liegen bis zu 6 Wiederholungen für einige der Designpunkte vor. Die 45 Computerexperimente wurden nach einem zentral zusammen gesetzten Versuchsplan, ergänzt um einige weitere Punkte, erhoben.

Als Zielgrößen werden sowohl Zargen- als auch Flanschwinkel betrachtet (siehe Abbildung 6.2). Hierfür sind die Methoden zur Kalibrierung aus Kapitel 4 auf multivariate Zielgrößen  $\mathbf{Y} = (Y_1, \dots, Y_p)^T$  zu erweitern. Dabei ist sicher zu stellen, dass jede der experimentellen Zielgrößen  $Y_{E,j}$  möglichst gut durch das Computermodell beschrieben werden kann. Damit sind nur solche Kalibrierparameter geeignet, die für alle Zielgrößen eine Übereinstimmung von Computermodell und Experiment liefern. Es wird angenommen, dass die Zielgrößen untereinander stochastisch unabhängig sind, um eine einfache Erweiterung der Kalibrierung zu ermöglichen. Für jede Zielgröße wird die Modellgleichung

$$Y_{E,j}(\mathbf{x}) = y_{M,j}(\mathbf{x}, \mathbf{u}) + b_{\mathbf{u},j}(\mathbf{x}) + e_j$$

mit  $e_j \sim N(0, \sigma_{E,j}^2)$  betrachtet, wobei die Verzerrungsfunktion  $b_{\mathbf{u},j}(\mathbf{x})$  für die nichtlineare Regression konstant Null ist. In jeder Dimension werden Computermodell und Verzerrung durch unabhängige Gauß Prozesse beschrieben. Für die Modellanpassung wird wieder ein modulares Vorgehen gewählt. Die Parameter für den Gauß Prozess zur Beschreibung des Computermodells werden für alle Zielgrößen unabhängig voneinander aus Computerexperimenten bestimmt. Die gemeinsame Likelihood aller Zielgrößen ergibt sich durch die Unabhängigkeitsannahme als Produkt der univariaten Likelihoodfunktionen. Im Gibbs-Algorithmus für die Bayesianische Kalibrierung können alle Komponenten bis auf  $\mathbf{u}$  wie im eindimensionalen Fall generiert werden. Ein neues Element der Markovkette von  $\mathbf{u}$  wird im Metropolis-Schritt mit Hilfe von Produktlikelihood und dem Produkt der a priori Verteilungen erzeugt.

Bei Anwendung der beiden Kalibrierungsmethoden aus Kapitel 4 mit multivariater Erweiterung auf die vorliegenden Daten werden alle Einflussgrößen jeweils auf das Intervall  $[0, 1]$  skaliert. Kurz bezeichne ich wieder mit  $\mathbf{x} = (r_{Wz}, F_{NH})^T$  den Vektor der physikalischen Einflussgrößen und mit  $\mathbf{u} = (V_{EKL}, A_{EB})^T$  den Vektor der Kalibrierparameter. Für beide Zielgrößen wird das Computermodell durch einen Gauß Prozess mit konstanter Erwartungswertfunktion und Gauß Korrelationsfunktion modelliert.

Zunächst wird der vergleichsweise einfach zu implementierende nichtlineare Re-

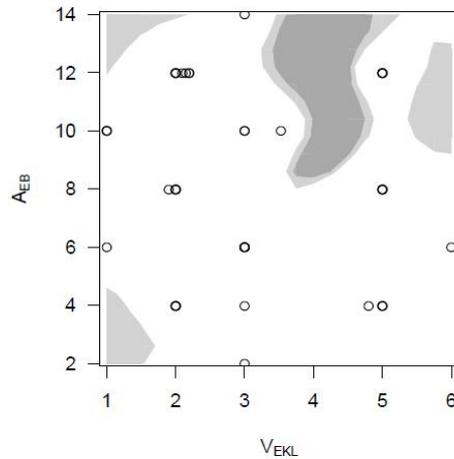


Abbildung 6.3: Ergebnisse des nichtlinearen Regressionsansatzes: 95 (99)% Konfidenzregion für  $V_{EKL}$  und  $A_{EB}$  in dunkelgrau (hellgrau) sowie Projektion der Designpunkte

gressionsansatz angewendet. Der Maximum-Likelihood Schätzer ist gegeben durch  $\hat{\mathbf{u}}_{ML} = (4.25, 12.80)^T$ . Abbildung 6.3 zeigt die 95% und 99% Konfidenzregionen für  $\mathbf{u} = (V_{EKL}, A_{EB})^T$  mit höchster Profil-Likelihood. Die 95% Konfidenzregion konzentriert sich auf den Bereich  $V_{EKL} \approx 4$  und  $A_{EB} > 8$ ; die 99% Konfidenzregion enthält zusätzlich einige Regionen am Rand des Wertebereichs, jedoch keine für großes  $V_{EKL}$  und kleines  $A_{EB}$ . Es fällt auf, dass die Konfidenzregionen keinen der Designpunkte für das Computerexperiment, projiziert auf  $\mathbf{u}$ , enthalten. In den Konfidenzregionen besteht ein recht großer Abstand zum nächsten Designpunkt und somit ist dort die Unsicherheit über das Computermodell verhältnismäßig groß. In Kapitel 4.2.1 wurde illustriert, dass der nichtlineare Regressionsansatz bei verzerrtem Computermodell die quadratischen Abweichungen zwischen Computermodell und Realität minimiert, wobei Abweichungen in Regionen mit großer Varianz der Computermodellvorhersage weniger stark gewichtet werden. Die große Unsicherheit über das Computermodell in den Konfidenzregionen kann hier also eine Verzerrung maskieren. Damit spricht die Form der Konfidenzregionen für ein verzerrtes Computermodell. Zusätzliche raumfüllend ausgewählte Computerexperimente können die Unsicherheit über das Computermodell reduzieren und damit die Schätzung der Kalibrierparameter verbessern. Für die Kalibrierung von Computermodellen ist somit ein raumfüllender Versuchsplan, wie zum Beispiel ein Latin Hypercube Design, für die Erhebung der Computerexperimente

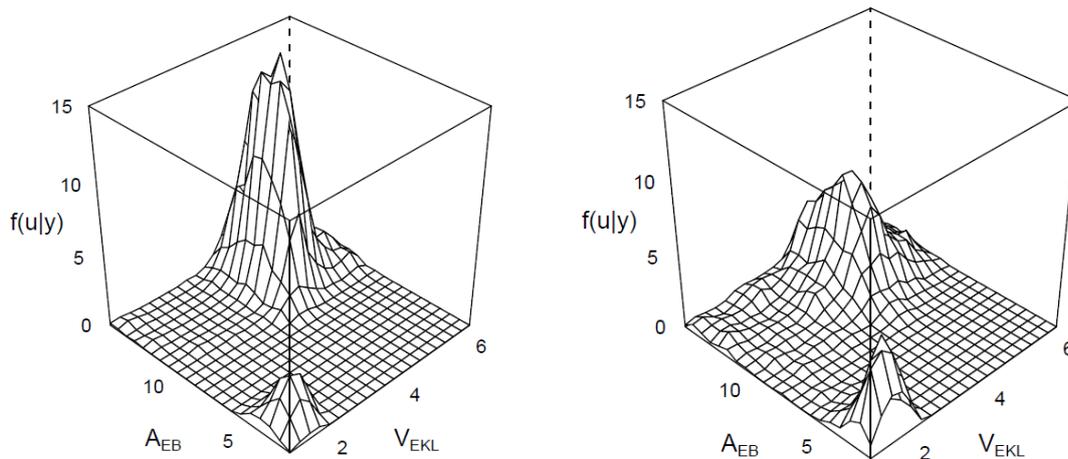


Abbildung 6.4: Ergebnisse der Bayes Kalibrierung: A posteriori für  $V_{EKL}$  und  $A_{EB}$  mit  $\mu_b = 0$  (links) und  $\mu_b = const$  (rechts).

besser geeignet als der hier verwendete zentral zusammengesetzte Versuchsplan.

Diese Beobachtungen motivieren die Verwendung eines Kalibrierungsmodells mit Verzerrungsfunktion und im nächsten Schritt wird der Bayesianische Kalibrierungsansatz auf die vorliegenden Daten angewendet. Dazu wird die multivariate Erweiterung des Gibbs-Sampling Algorithmus von Bayarri *et al.* (2007b) implementiert. Der Implementations- und rechentechnische Aufwand für diese Methode ist vergleichsweise hoch, aber für die Dimension des vorliegenden Problems gut realisierbar. In Abbildung 6.4 ist ein Kerndichteschätzer für die a posteriori von  $\mathbf{u} = (V_{EKL}, A_{EB})^T$  dargestellt – links wird für die Verzerrung ein a priori Erwartungswert  $\mu_b = 0$  und rechts ein konstanter unbekannter Erwartungswert  $\mu_b$  verwendet. Die a posteriori Dichte mit  $\mu_b = 0$  ähnelt der Profil-Likelihood im nichtlinearen Regressionsansatz; die Maximalstellen befinden sich an ähnlichen Stellen, allerdings sind diese weniger stark konzentriert und bleiben nicht strikt von den Designpunkten entfernt. Der a posteriori Modalwert liegt bei  $\hat{\mathbf{u}} = (4.5, 11)^T$ . Auch wenn für die Verzerrungsfunktion ein unbekannter konstanter Erwartungswert zugelassen wird, besitzt die a posteriori Verteilung eine ähnliche Gestalt. Allerdings sind die Maxima der a posteriori Dichte weniger deutlich. Unter beiden Modellannahmen sind Werte im Bereich von  $A_{EB} < 8$  und  $V_{EKL} > 2.5$  – also eine grobe Diskretisierung – sehr unwahrscheinlich.

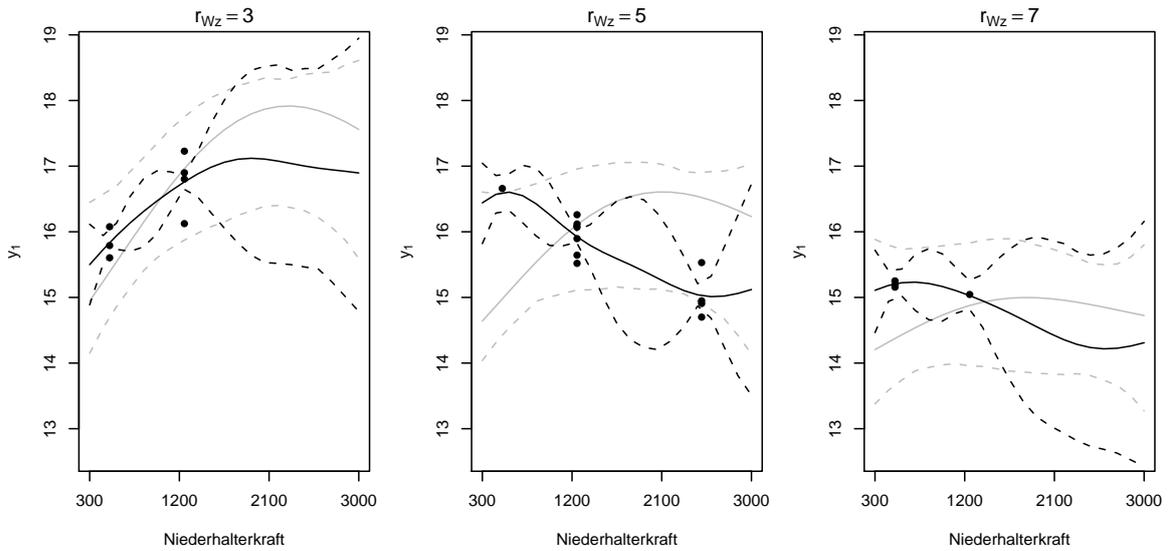


Abbildung 6.5: Vorhersagen für den Zargenwinkel mit  $\mu_b = 0$ .

In grau: Computermodellvorhersagen  $\hat{y}_M(\hat{u})$  mit a posteriori Modalwert  $\hat{u}$  (durchgezogene Linie) und 95% Unsicherheitsintervalle (gestrichelte Linie). In schwarz: Vorhersage für Prozessmittelwert (durchgezogene Linie), 95% Unsicherheitsintervalle (gestrichelte Linie) und experimentelle Beobachtungen (Punkte).

In Abbildung 6.5 sind die Vorhersagen für realen Prozess und Computermodell des Zargenwinkels  $Y_1$  in Abhängigkeit der Niederhalterkraft dargestellt. In den drei Grafiken werden unterschiedliche Werkzeugradien  $r_{Wz}$  von 3, 5 oder 7 mm betrachtet. Die Rückfederungsvorhersage für den Prozessmittelwert  $\mu_{E,1}(\cdot)$  steigt bei einem Werkzeugradius von 3 mm mit der Niederhalterkraft an, wohingegen sie bei den beiden größeren Werkzeugradien fällt. Die Unsicherheit über den physikalischen Prozess ist in vielen Bereichen verhältnismäßig groß und lediglich in der Nähe der experimentellen Beobachtungen klein. Neben der Prozessvorhersage wird in der Grafik auch die Computermodellvorhersage  $\hat{y}_M(\cdot, \hat{u})$ , bei der  $\mathbf{u}$  fest auf den a posteriori Modalwert  $\hat{u}$  gesetzt wird, dargestellt. Sie weicht für alle Werkzeugradien deutlich von der Prozessvorhersage ab. Es werden zwar prinzipiell Werte im richtigen Wertebereich vorhergesagt, allerdings kann der funktionale Zusammenhang zwischen Niederhalterkraft und Rückfederung nicht korrekt beschrieben werden. Anders als im realen Prozess beobachtet, sagt das Computermodell stets eine mit der Niederhalterkraft steigende Rückfederung vorher. Es fällt auf, dass die Unsicherheitsintervalle des Computermodells überall sehr breit sind. Neben der Unsicherheit über die Compu-

termodellfunktion enthalten diese auch die Unsicherheit über  $\mathbf{u}$ . Die betrachteten numerischen Kalibrierparameter können also eine große Variation von Rückfederungswinkeln vorhersagen. Insgesamt ermöglicht das Computermodell somit alleine keine präzise Vorhersage der Realität. In Regionen mit ausreichend vielen experimentellen Beobachtungen kann das Computermodell keine zusätzlichen Informationen liefern. Allerdings kann es in Regionen, in denen keine physikalischen Experimente vorliegen, sinnvolle Vorhersagen und Unsicherheitsintervalle unterstützen.

Beide Ansätze zur Kalibrierung von Computermodellen kommen zu dem Ergebnis, dass ein verzerrtes Computermodell vorliegt. Insbesondere mit einer groben Diskretisierung kann das Computermodell die Realität nicht geeignet approximieren. Ab einer ausreichend feinen Diskretisierung kann zwar die Größenordnung der Rückfederung, jedoch nicht der Einfluss der Prozessparameter auf die Rückfederung vorhergesagt werden. Entweder besteht ein systematischer Fehler bei der Formulierung des Computermodells oder weitere Parameter im Computermodell wurden hier ungünstig gewählt. Die betrachteten numerischen Parameter sind somit nur bedingt für die Kalibrierung des Computermodells geeignet.

Kommen wir nun zu einem Vergleich der beiden angewendeten Methoden. In diesem Beispiel ist die Bayesianische Kalibrierung besser geeignet als die Kalibrierung durch nichtlineare Regression, da von einem verzerrten Computermodell auszugehen ist. Insbesondere die Vorhersage von  $\mu_E(\mathbf{x})$  mit Angabe von Unsicherheitsintervallen ist eine große Stärke des Ansatzes, die die Kalibrierung durch nichtlineare Regression nicht bietet. Für das vorliegende niedrigdimensionale Beispiel kann die Verzerrung dem Anwender gut durch eine grafische Darstellung vermittelt werden.

Bei der Kalibrierung durch nichtlineare Regression zeigte sich, dass ein raumfüllender Versuchsplan für die Computereperimente besser geeignet ist als der hier verwendete zentral zusammengesetzte Versuchsplan. Dies gilt auch für die Bayesianische Kalibrierung. Für beide Methoden erweisen sich die vorliegenden Wiederholungen von physikalischen Experimenten an einem Designpunkt als hilfreich um zwischen Modellabweichung und zufälliger Streuung unterscheiden zu können. Wird jedoch von einer

konstanten Varianz ausgegangen, sind für das vorliegende Modell mehr Designpunkte gegenüber einer größeren Anzahl von Wiederholungen zu bevorzugen.

## 6.2 Kalibrierung einer Rückfederungssimulation durch Materialparameter

In diesem Kapitel soll die Rückfederungssimulation eines Hutprofils für den Werkstoff H240LA, ein mikrolegierter Stahl mit einer Mindeststreckgrenze von 240 MPa, durch Material- und Reibungsparameter kalibriert werden. Anschließend soll der Prozess hinsichtlich der Rückfederung minimiert werden. Im Computermodell wird das Verhalten von Material und Reibung durch nichtlineare Reibungs- und Materialmodelle beschrieben, die unbekannte Parameter enthalten. In dieser Anwendung wird als Materialmodell ein kombiniert isotrop-kinematisches Verfestigungsmodell verwendet. In Gösling (2010) konnte für die vorliegende Werkzeuggeometrie gezeigt werden, dass dieses Modell eine adäquate Beschreibung der Realität ermöglicht. In der verwendeten Simulationssoftware Autoform ist ein solches Materialmodell implementiert (siehe Kubli *et al.*, 2008).

Es werden insgesamt 7 solcher Reibungs- und Materialparameter in einem festgelegten Parameterbereich als unbekannt betrachtet und sollen durch einen Kalibrierungsansatz anhand experimenteller Daten für die Rückfederung geschätzt werden. Der Wertebereich der Parameter wurde anhand von Vorversuchen – so genannten Zug- und Reibungsversuchen – festgelegt.

Für das Reibungsmodell werden die dimensionslosen Parameter Reibungszahl  $u_1$  (0.1-0.2) und Reibungsexponent  $u_2$  (0.5-1) betrachtet. Die Reibungszahl gibt die maximale Reibung bei einem festgelegten Druck an und der Reibungsexponent steuert die Abnahme der Reibung bei weiter ansteigendem Druck. Das Materialverhalten wird in Abhängigkeit der Anfangsfließspannung  $u_3$  (-13 bis 13 MPa), der Zugfestigkeit  $u_4$  (-15 bis 15 MPa), dem Extrapolationsanteil  $u_5$  (0 - 1), der frühzeitigen Plastifizierung  $u_6$  (0.001-0.004) und der Verfestigungsstagnation  $u_7$  (0.1-0.9) beschrieben. Die Anfangsfließspannung gibt Spannung bei kleinen Dehnungen zu Beginn der plastischen Phase (hier als Abweichung zu 238 MPa) an, analog dazu gibt die Zugfestigkeit die maximale Spannung (hier als Abweichung zu 516 MPa) an. Die Fließkurve wird hier aus einer

Konvexkombination zweier Extrapolationsmodelle bestimmt. Der Extrapolationsanteil bezeichnet das Gewicht für die Konvexkombination. Die frühzeitige Plastifizierung und die Verfestigungsstagnation beschreiben das kinematische Verfestigungsverhalten nach einer Lastumkehr.

Die Anzahl unbekannter Kalibrierparameter ist mit sieben verhältnismäßig groß. Die Anwender gehen davon aus, dass diese Material- und Reibeigenschaften die wesentliche Unsicherheit bei der Beschreibung der Rückfederung enthalten. Entsprechend wird vermutet, dass die Computersimulation unter Umständen alleine durch die Wahl der Kalibrierparameter in der Lage ist die Realität gut zu approximieren.

Für die Kalibrierung des Computermodells werden vier physikalische Parameter betrachtet. Dabei handelt es sich um den Werkzeugradius (3-7 mm), die Ziehtiefe (40-60 mm), die Niederhalterkraft (2-10 MPa) und die Blechdicke (0.75-1.2 mm). Zielgröße  $Y$  ist der Rückfederungswinkel im Flanschbereich ( $Y_2$  in Abbildung 6.2). Der Rückfederungswinkel im Zargenbereich ( $Y_1$  in Abbildung 6.2) wird nicht berücksichtigt, da dieser hier nahezu linear von dem Flanschwinkel abhängt und somit keine zusätzlichen Informationen enthält.

Es liegen insgesamt 113 reale Experimente und 443 Computerexperimente, erhoben am Institut für Umformtechnik und Leichtbau der TU Dortmund, vor. Die physikalischen Experimente wurden nach einem faktoriellen Versuchsplan auf je 3 Stufen erhoben. 33 unterschiedliche Kombinationen wurden dabei jeweils mindestens 3 mal wiederholt. Bei den Computerexperimenten wurde der schwerer zu variierende Werkzeugradius ebenfalls nur auf 3 Stufen betrachtet. Für alle übrigen Parameter wurde jeweils ein raumfüllender Versuchsplan (Latin-Hypercube Design) verwendet – mit je 101 Simulationen für die Werkzeugradien 3 und 7 mm sowie 241 Simulationen für einen Werkzeugradius von 5 mm.

Für die Analyse werden alle Einflussgrößen aus  $\mathbf{x}$  und  $\mathbf{u}$  auf das Intervall  $[0, 1]$  skaliert. In einem ersten Schritt wird ein geeignetes Gauß Prozess Modell zur Beschreibung der Computersimulation ausgewählt. Dazu werden die Resultate der Gauß Prozess Interpolation für unterschiedliche Erwartungswertfunktionen (konstante Funktion, lineare Funktion sowie quadratische Funktion mit 2-fach Interaktionen) und Korrelati-

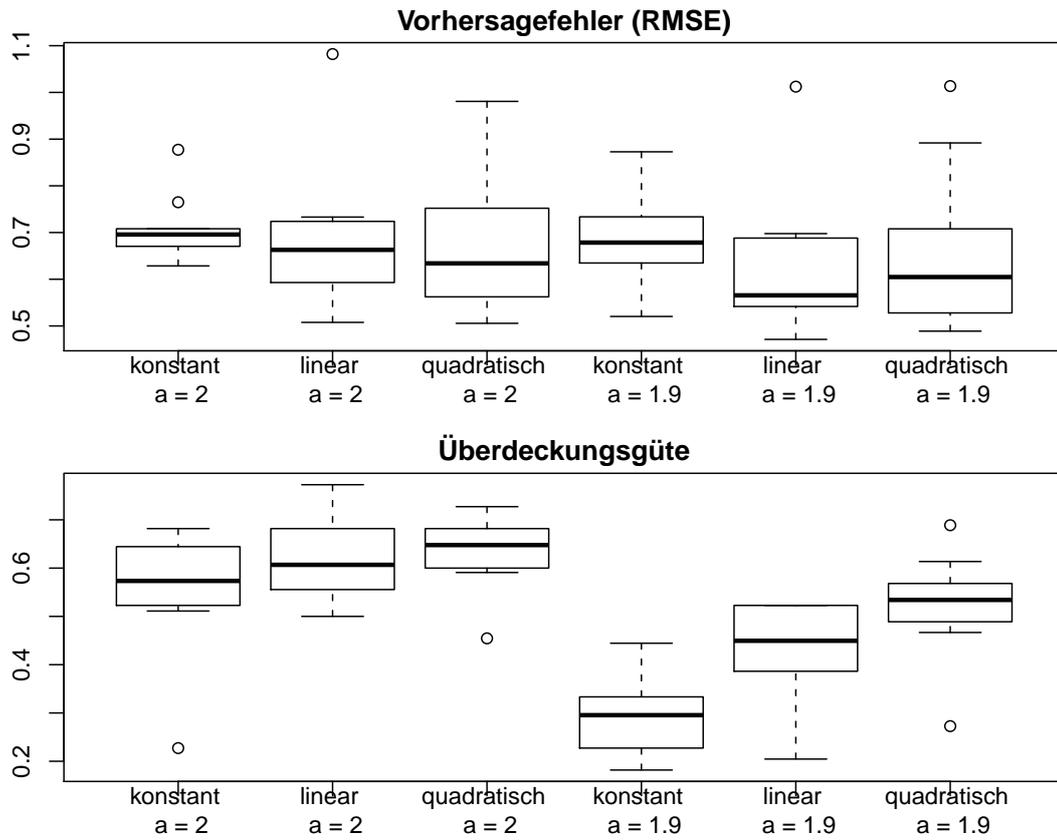


Abbildung 6.6: Vorhersagefehler und Überdeckungsgüte bei 10-facher Kreuzvalidierung

onsfunktionen („Power-Exponential“ Korrelationsfunktion mit  $a = 2$  und  $a = 1.9$  und  $a_j = a$  für alle  $j = 1, \dots, d_M$ ) mit Hilfe einer 10-fachen Kreuzvalidierung verglichen. Als Zielgrößen werden die Vorhersagegüte, genauer die Wurzel des mittleren quadratischen Fehlers (RMSE), sowie die Überdeckungsgüte, das heißt der Anteil der Testbeobachtungen, die innerhalb des punktweisen Vorhersageintervalls liegen, betrachtet. Abbildung 6.6 stellt die Resultate grafisch dar. Die Gauß Prozess Interpolation mit quadratischer Erwartungswertfunktion und Gauß Korrelationsfunktion ( $a = 2$ ) liefert die besten Resultate der Überdeckungsgüte bei gleichzeitig guter Vorhersagegüte. Die mittleren 50 % der Resultate für die Überdeckungsgüte liegen zwischen 60 und 68 %. Das nominale Niveau von 95% wird damit nicht erreicht. Die mittleren Abweichungen der Vorhersagegüte für den Flanschwinkel liegen jedoch stets unter einem Grad - bei einem Wertebereich von 5 bis 26 Grad und einer Standardabweichung von 3.6 Grad des Datenvektors  $\mathbf{y}_M$ . Insgesamt wird durch die Gauß Prozess Interpolation eine

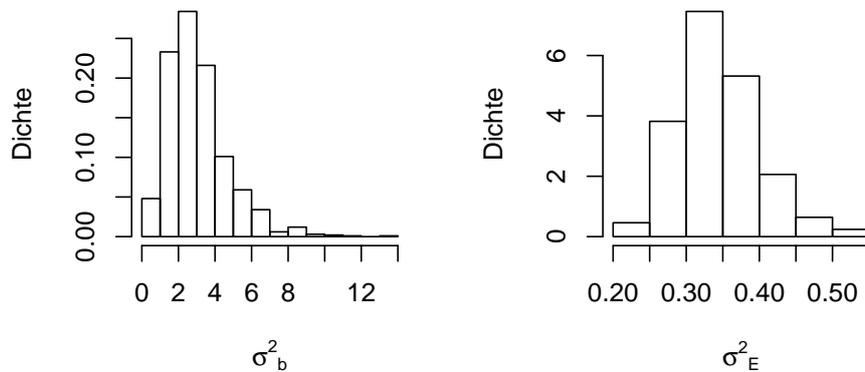
zufrieden stellende Modellanpassung erreicht.

Ob bei der Kalibrierung eine Verzerrungsfunktion erforderlich ist, soll im Folgenden untersucht werden. Entsprechend wird die Kalibrierung unter Modellunsicherheit aus Kapitel 5 angewendet. Dabei wird das Computermodell durch einen Gauß Prozess mit quadratischer Erwartungswertfunktion und Gauß Korrelationsfunktion approximiert. Die Verzerrungsfunktion wird als Gauß Prozess mit Erwartungswert  $\mu_b = 0$  und Gauß Korrelationsfunktion modelliert. Für die Varianzparameter  $\sigma_b^2$  und  $\sigma_E^2$  wird die datenabhängige quadrierte  $t$ -Verteilung mit 10 Freiheitsgraden als a priori verwendet. Die Kalibrierparameter werden a priori als unabhängig gleichverteilt auf  $[0, 1]$  angenommen. Die Modellindikatorvariable  $\gamma$  folgt a priori einer  $Bin(1, 0.5)$  Verteilung.

Die a posteriori Stichprobe für die Parameter wird durch den Importance-Sampling Algorithmus basierend auf der iterativen Laplace-Approximation erstellt. Der geschätzte effektive Stichprobenumfang basierend auf 10000 Zufallsvektoren beträgt dabei je nach Modellsituation zwischen 2000 und 2500. Für die nachfolgenden Vorhersagen wird hieraus eine Stichprobe vom Umfang 1000 ausgewählt.

Im Folgenden werden die Ergebnisse der Kalibrierung analysiert. Die a posteriori Wahrscheinlichkeit  $p = P(\gamma = 0 | \mathbf{y}_M, \mathbf{y}_E)$  für das Kalibrierungsmodell ohne Verzerrungsfunktion beträgt  $p = 0.0006$  und ist damit nahezu Null. Dies ist nach der Interpretationsskala von Kass und Raftery (1995) eine starke Evidenz für ein verzerrtes Computermodell. Im vorliegenden Fallbeispiel kann daher trotz der großen Zahl der Kalibrierparameter der Erwartungswert der Rückfederung nicht alleine durch das Computermodell geeignet beschrieben werden. In der folgenden Analyse wird nur das Modell mit Verzerrungsfunktion betrachtet. Die Resultate werden sich durch das extrem kleine Gewicht  $p$  für das Modell ohne Verzerrungsfunktion nicht wesentlich von den Resultaten bei Modellmittelung unterscheiden.

Abbildung 6.7 zeigt Histogramme für die marginalen a posteriori Stichproben von  $\sigma_b^2$  und  $\sigma_E^2$ . Die Prozessvarianz  $\sigma_b^2$  für die Verzerrungsfunktion besitzt eine rechtsschiefe Verteilung, deren Wahrscheinlichkeitsmasse größtenteils zwischen 0 und 8 liegt und ihren Modalwert bei etwa 2 besitzt. Somit sind große Verzerrungen möglich und es

Abbildung 6.7: A posteriori Verteilungen für  $\sigma_b^2$  und  $\sigma_E^2$ 

besteht eine recht große Unsicherheit über  $\sigma_b^2$ . Die a posteriori Verteilung für die experimentelle Streuung  $\sigma_E^2$  ist nahezu symmetrisch um 0.35 und konzentriert sich auf den Bereich zwischen 0.2 und 0.55. Damit ist die Unsicherheit über die experimentelle Streuung verhältnismäßig klein. Die experimentelle Streuung kann durch die vorliegenden Wiederholungen bereits recht präzise geschätzt werden und ist deutlich kleiner als die Prozessvarianz für den Gauß Prozess des Verzerrungsterms bei verhältnismäßig kleiner Unsicherheit.

Für die Kalibrierparameter  $u_j$  werden die marginalen und paarweisen a posteriori Verteilungen in Abbildung 6.8 dargestellt. Die Histogramme der marginalen a posteriori Stichprobe finden sich auf der Hauptdiagonalen. Die zweidimensionalen Histogramme der paarweisen a posteriori Stichproben sind in Matrixform angeordnet. Dabei steigt die Graustufenintensität mit der absoluten Häufigkeit der Beobachtungen in einem Quadrat. Der Wertebereich der Histogramme entspricht jeweils den festgelegten Parameterbereichen. Die marginalen a posteriori Verteilungen konzentrieren sich für einige Parameter auf einen Bereich des Intervalls, wohingegen sie sich für andere Parameter nur wenig von der gleichverteilten a priori unterscheiden. Kleines a posteriori Gewicht besitzen Reibungszahlen  $u_1 < 0.15$ , Reibungsexponenten  $u_2 > 0.8$  sowie eine frühzeitige Plastifizierung  $u_6$  am Rand des betrachteten Parameterraums. Für die Zugfestigkeit  $u_4$  und den Extrapolationsanteil  $u_5$  werden eher große Werte bevorzugt. Dahingegen sind für die Anfangsfließspannung  $u_3$  und Verfestigungsstagnation  $u_7$  a

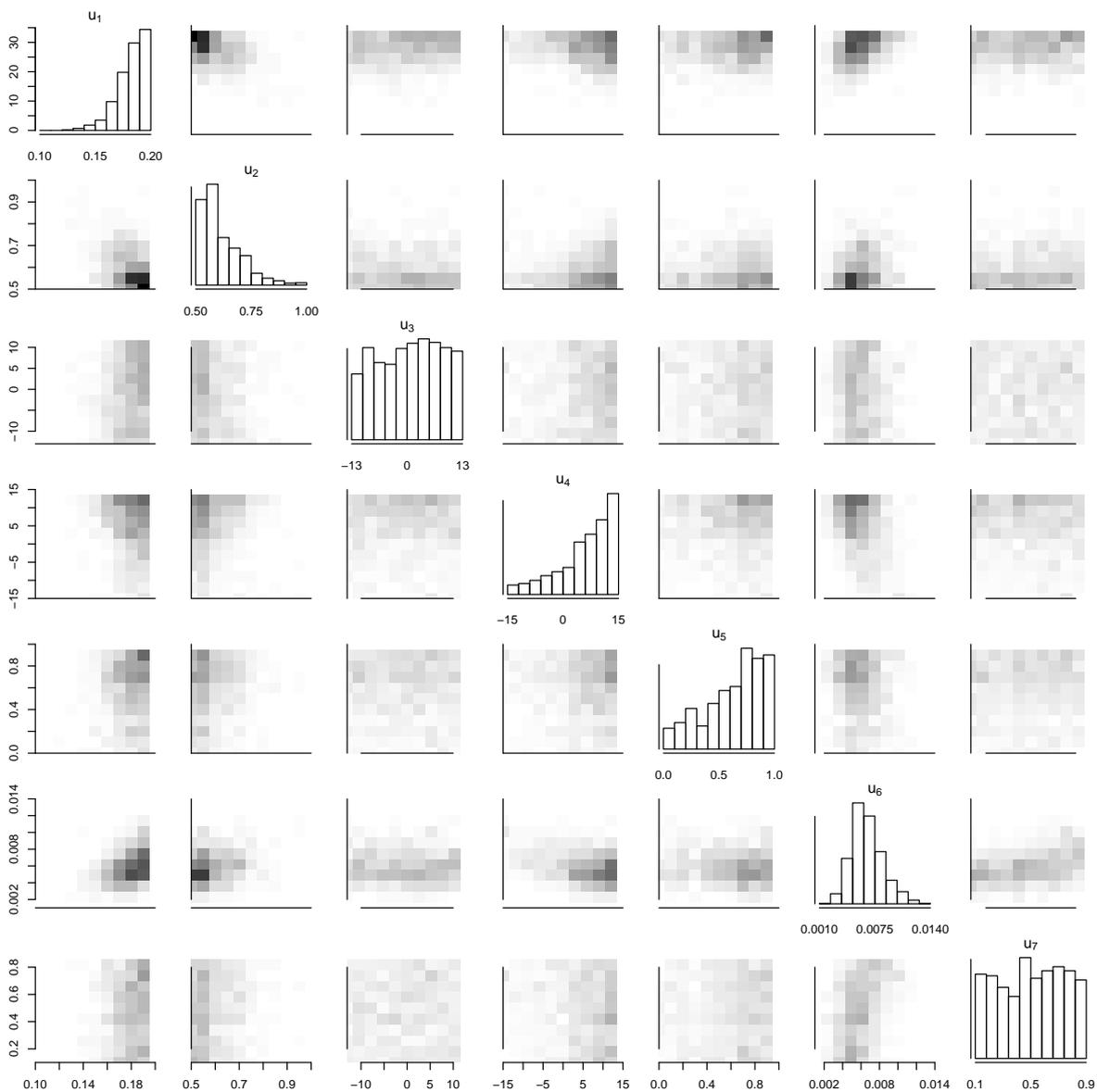


Abbildung 6.8: A posteriori Verteilungen der Kalibrierparameter

posteriori alle Bereiche des Parameterraums etwa gleich wahrscheinlich. Für diese Parameter konnte durch die Kalibrierung keine wesentliche Information gewonnen werden.

Bei Betrachtung der paarweisen a posteriori Verteilungen werden kleinere, jedoch keine extremen, Abhängigkeiten sichtbar. Es besteht zum Beispiel ein paarweiser Zusammenhang zwischen Reibungszahl ( $u_1$ ), Reibungsexponent ( $u_2$ ) und frühzeitiger Plastifizierung ( $u_6$ ). Kleinere Abhängigkeiten dieser Gruppe bestehen ebenfalls zu der Zugfestigkeit ( $u_4$ ).

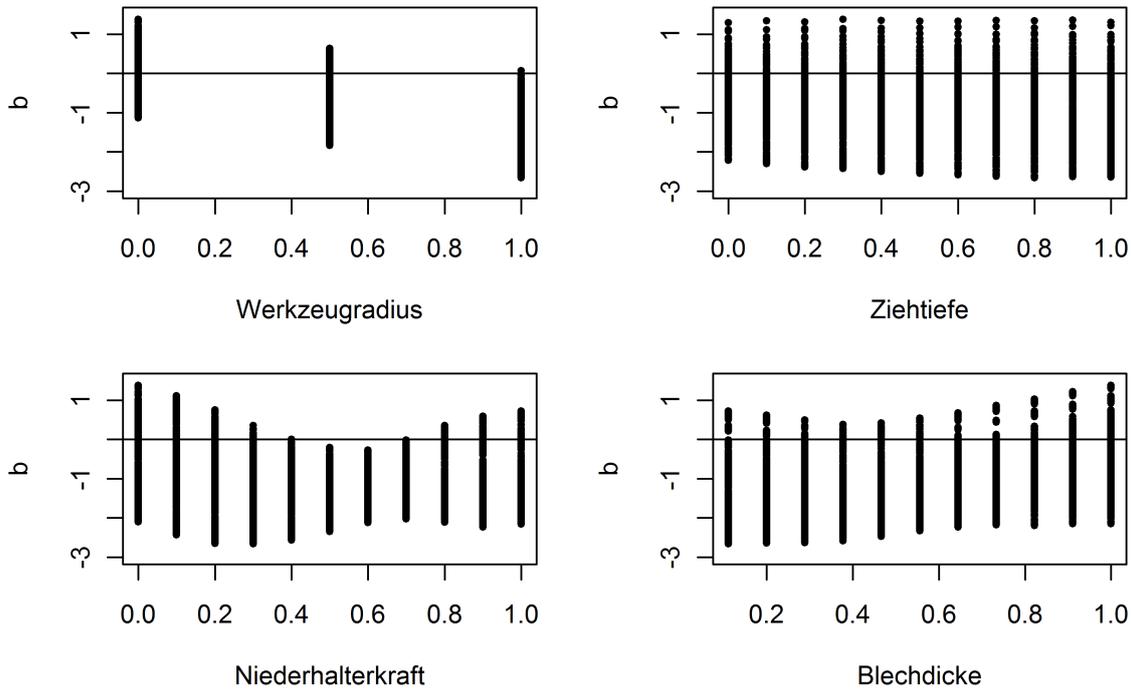


Abbildung 6.9: Vorhersage der Verzerrungsfunktion

Durch die Kalibrierung konnten Informationen über geeignete Parameter im Reibungsmodell gewonnen werden. Für das Materialmodell hingegen gilt dies nur für einige Parameter. Eine präzisere Punktschätzung der Kalibrierparameter erfordert weitere Beobachtungen aus physikalischen Experimenten. Diese sollten jedoch an anderen Experimentdesignpunkten erhoben werden. Zusätzliche Wiederholungen können die Schätzung nur wenig verbessern, da die verhältnismäßig kleine experimentelle Streuung  $\sigma_E^2$  bereits recht präzise geschätzt werden kann.

Im nächsten Schritt werden nun Vorhersageverteilungen für Computermodell, Verzerrung und Gesamtprozess erstellt. Dazu wird ein Gitter der 4 physikalischen Einflussgrößen  $\mathbf{x}$  betrachtet mit Werkzeugradius 3, 5 und 7 mm und jeweils 11 äquidistanten Werte auf den Parameterbereichen im Computerexperiment für Ziehtiefe, Niederhalterkraft und Blechdicke. Eine Stichprobe aus der Vorhersageverteilung wird generiert indem für jedes a posteriori Stichprobenelement  $\boldsymbol{\theta}^{(i)} = (\sigma_E^{2(i)}, \sigma_b^{2(i)}, \mathbf{u}^{(i)})$ ,  $i = 1, \dots, N$ , Zufallszahlen für Computermodell  $y_M^{(i)}(\mathbf{x}, \mathbf{u}^{(i)})$  und Verzerrung  $b_{\mathbf{u}^{(i)}}(\mathbf{x})$ , sowie durch Addition für den physikalischen Prozess  $\mu_E^{(i)}(\mathbf{x}) = y_M^{(i)}(\mathbf{x}, \mathbf{u}^{(i)}) + b_{\mathbf{u}^{(i)}}(\mathbf{x})$

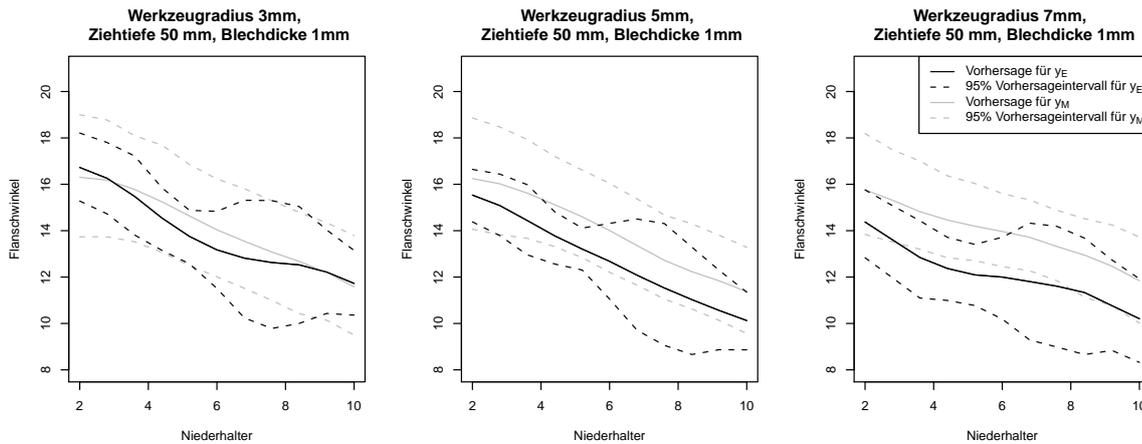


Abbildung 6.10: Vorhersagen für Computermodell und Realität

generiert werden. Als schnell berechenbarer Punktschätzer für das Computermodell wird außerdem die Vorhersage  $\hat{y}_M(\mathbf{x}, \hat{\mathbf{u}})$  mit a posteriori Modalwert  $\hat{\mathbf{u}}$  bestimmt. Das Erstellen der Stichproben aus den Vorhersageverteilungen ist durch die große Anzahl an Gitterpunkten und a posteriori Stichprobenelementen für  $\boldsymbol{\theta}$  rechenintensiv.

Abbildung 6.9 zeigt die geschätzte Verzerrungsfunktion  $\bar{b}(\mathbf{x}) = \frac{1}{N} \sum_{i=1}^N b^{(i)}(\mathbf{x})$ . Sie wird gegen jede Komponente  $x_j$  abgetragen, um so die marginalen Effekte darzustellen. Da im Gitter jeder Wert von  $x_j$  häufiger auftritt, liegen jeweils mehrere Punkte vor. Die Verzerrungen für den Flanschwinkel liegen zwischen -3 und 1.5 Grad, das heißt die Rückfederung wird durch das Computermodell eher überschätzt. Es besteht ein deutlicher, linear aussehender, Zusammenhang zwischen Verzerrung und Werkzeugradius: je größer der Werkzeugradius umso kleiner und weiter von Null entfernt ist  $\bar{b}(\cdot)$ . Auch Niederhalterkraft und Blechdicke besitzen einen Einfluss auf die Verzerrung. Dagegen lässt sich zwischen Verzerrung und Ziehtiefe kein Zusammenhang beobachten. Neben den hier betrachteten Haupteffekten bestehen auch Interaktionseffekte der Einflussgrößen auf die Verzerrung. Dies gilt insbesondere für die Interaktion zwischen Niederhalterkraft und Blechdicke.

Nun werden Vorhersagen und zugehörige Unsicherheitsintervalle für die Rückfederungskalibrierung beispielhaft untersucht und dargestellt. Dazu werden Ziehtiefe und Blechdicke konstant auf 50 mm bzw. 1 mm, das heißt auf der Mitte des Parameterbereichs, festgehalten. Abbildung 6.10 zeigt die Vorhersagen für Computermodell und

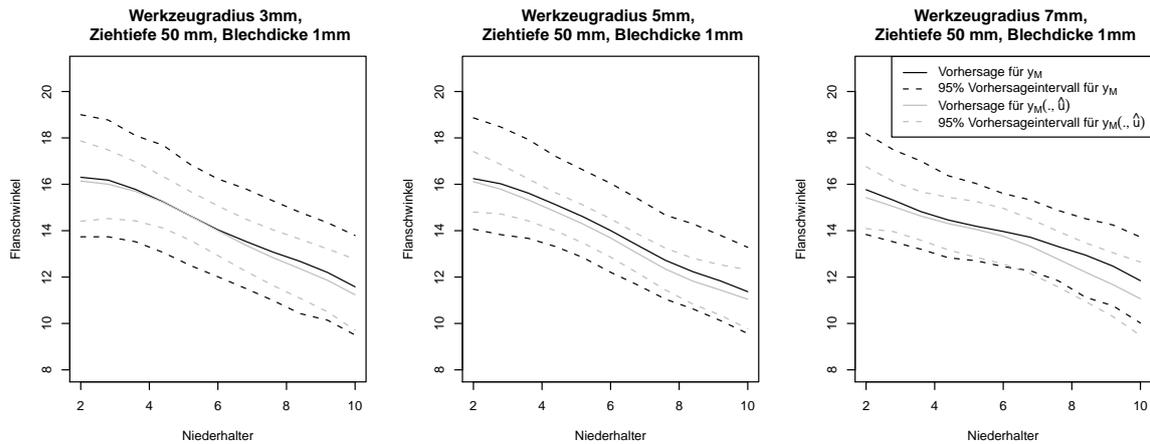


Abbildung 6.11: Computermodellvorhersagen mit Unsicherheitsintervallen

Realität für drei verschiedene Werkzeugradien (3, 5 und 7 mm) in Abhängigkeit der Niederhalterkraft.

Die Computermodellvorhersagen  $\hat{y}_M(\mathbf{x})$  (Vorhersage für  $y_M$ ) weichen hier systematisch von den Vorhersagen für den realen Prozess nach oben ab, wie bereits bei Betrachtung der Verzerrungsfunktion in Abbildung 6.9 festgestellt. Allerdings kann das Computermodell den Trend des realen Prozesses in Abhängigkeit der Niederhalterkraft geeignet beschreiben: die Rückfederung sinkt mit steigender Niederhalterkraft. Für den realen Prozess zeigt sich weiterhin, dass die Rückfederung auch mit steigendem Werkzeugradius sinkt. Dieser Aspekt wird im Computermodell nicht abgebildet: es ist nahezu unabhängig vom Werkzeugradius. Die Unsicherheit über das Computermodell beträgt etwa 4 bis 5 Grad. Für den realen Prozess  $\mu_E(\cdot)$  sind die Unsicherheitsintervalle zwischen 2 und 4 Grad lang und entsprechend dem Experimentdesign auf 3 Stufen am Rand und in der Mitte kürzer als im übrigen Parameterbereich. Eine Reduktion der recht großen Unsicherheit könnte durch zusätzliche Experimente gewonnen werden.

Betrachten wir nun wie sich die Unsicherheit über die Kalibrierparameter auf die Computermodellvorhersage auswirkt. Abbildung 6.11 zeigt einerseits die Computermodellvorhersage  $\hat{y}_M(\mathbf{x})$ , mit Berücksichtigung der Unsicherheit über  $\mathbf{u}$ , sowie die schneller berechenbare Computermodellvorhersage  $\hat{y}_M(\mathbf{x}, \hat{\mathbf{u}})$ , mit  $\mathbf{u}$  fest auf dem a posteriori Modalwert  $\hat{\mathbf{u}}$  und jeweils korrespondierende Unsicherheitsintervalle. Die Unsicherheitsintervalle beschreiben bei festem  $\mathbf{u} = \hat{\mathbf{u}}$  die reine Computermodellunsi-

cherheit und andernfalls zusätzlich die Unsicherheit über die Kalibrierparameter  $\mathbf{u}$ . Die Unsicherheitsintervalle für das Computermodell mit festem  $\mathbf{u} = \hat{\mathbf{u}}$  sind etwa 2 bis 4 Grad lang. Zusätzliche Computorexperimente können diese reine Computermodellunsicherheit reduzieren. Allerdings entsteht auch in diesem Anwendungsproblem ein großer Teil der Unsicherheit über das Computermodell durch die Unsicherheit bezüglich  $\mathbf{u}$ . Diese lässt sich nur durch zusätzliche reale Experimente und zwar mit einer höheren Anzahl von Faktorstufen und -kombinationen im Experimentdesign verbessern.

Im nächsten Schritt soll mit Hilfe des kalibrierten Computermodells die Rückfederung im Prozess minimiert werden. Die Rückfederung bei der Blechumformung ist eine unerwünschte Abweichung von der Sollgeometrie. Ziel ist es, die physikalischen Einflussgrößen so zu wählen, dass die resultierende Rückfederung minimal ist. Dazu werden die vorliegenden Vorhersageverteilungen für den physikalischen Prozess für eine Optimierung genutzt. Um eine Schätzung  $\hat{\mathbf{x}}_{\min}$  für die Minimalstelle des physikalischen Prozesses  $\mu_E(\mathbf{x})$  zu bestimmen, bestehen verschiedene Ansätze. Die einfachste Möglichkeit ist den Mittelwert der a posteriori Verteilung  $\bar{\mu}_E(\mathbf{x})$  zu minimieren. Dieses Vorgehen berücksichtigt jedoch nicht die Unsicherheit über den physikalischen Prozess. Bei großer Varianz wird unter Umständen ein  $\mathbf{x}$  gewählt, dessen wahre erwartete Rückfederung  $\mu_E(\mathbf{x})$  weit vom Minimum entfernt liegt. Um Einstellungen mit kleiner Unsicherheit zu bevorzugen, kann alternativ ein oberes Quantil der Vorhersageverteilung (z.B. das 0.975 Quantil  $\tilde{\mu}_{E(0.975)}(\mathbf{x})$ ) bezüglich  $\mathbf{x}$  minimiert werden.

Im vorliegenden Beispiel dient das für die Vorhersage beschriebene Gitter der Einstellungen für  $\mathbf{x}$  als Optimierungsgrundlage. Die optimale Einstellung bei Betrachtung des Mittelwertes ist gegeben durch einen Radius von 7 mm, einer Ziehtiefe von 40 mm, einer Niederhalterkraft von 10 MPa sowie einer Blechdicke von 1.2 mm. Damit liegt das Optimum am Rand des vorgegebenen Wertebereichs. Wird die alternative Strategie mit Minimierung des 0.975 Quantils der Unsicherheitsverteilung gewählt, ergibt sich eine ähnliche Minimalstelle. Radius, Ziehtiefe und Niederhalterkraft unterscheiden sich nicht von der Minimalstelle bei Betrachtung des Mittelwertes. Nur die Blechdicke ist mit 0.8

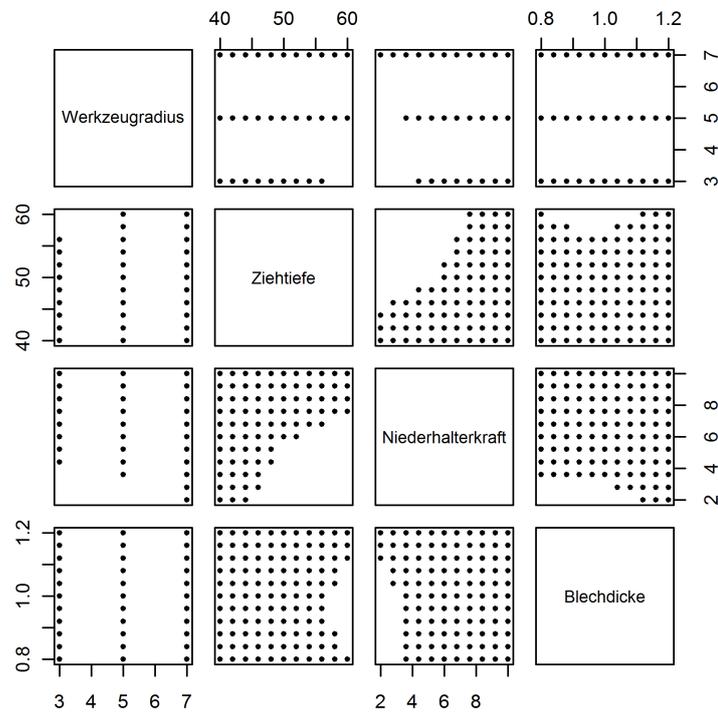


Abbildung 6.12: Unsicherheit über Einstellung mit minimaler Rückfederung

mm nun am anderen Ende des Wertebereichs. Die Unsicherheit über die Minimalstelle  $\hat{\boldsymbol{x}}_{\min}$  wird hier durch die Region  $R = \{\boldsymbol{x} : \tilde{\mu}_{E(0.025)}(\boldsymbol{x}) \leq \tilde{\mu}_{E(0.975)}(\hat{\boldsymbol{x}}_{\min})\}$  veranschaulicht. Diese Region enthält alle Einstellungen  $\boldsymbol{x}$  deren untere Unsicherheitsintervallgrenze kleiner der oberen Grenze an der geschätzten Minimalstelle ist. Abbildung 6.12 zeigt die Gitterpunkte, die innerhalb der Region  $R$  liegen in zweidimensionaler Darstellung. Diese potentiellen Minimalstellen decken den Wertebereich nahezu ab. Nur wenige der betrachteten Gitterpunkte können über die Kalibrierung als nicht geeignet identifiziert werden. Dies gilt lediglich für kleine Niederhalterkräfte insbesondere in Kombination mit großen Ziehtiefen, kleinen Blechdicken oder kleinen Werkzeugradien. Alle übrigen Bereiche des untersuchten Raums  $\mathcal{X}$  können unter Umständen zu einem Minimum führen. Somit besteht eine recht große Unsicherheit über die Minimalstelle der Funktion  $\mu_E(\cdot)$ .



## 7 FAZIT UND AUSBLICK

Computermodelle werden heute in vielen Anwendungsgebieten eingesetzt, um physikalische Prozesse schnell und kostengünstig vorherzusagen. Dabei ist zu prüfen in wie weit das Computermodell die Realität korrekt beschreiben kann. Außerdem sind unbekannte Parameter des Computermodells geeignet zu wählen. Diese Fragestellungen werden unter dem Begriff Kalibrierung von Computermodellen zusammengefasst. Das Endprodukt einer Kalibrierung ist die Erstellung eines Vorhersagemodells für den physikalischen Prozess unter Berücksichtigung von Beobachtungen aus Simulation und Experiment. In dieser Arbeit wurden dazu statistische Methoden zur Modellierung und Kalibrierung von Computermodellen diskutiert und auf ein praktisches Problem aus den Ingenieurwissenschaften angewendet.

In Kapitel 2 wurde zunächst die Gauß Prozess Interpolation als verbreitete Methode zur Modellierung deterministischer Computermodellausgaben vorgestellt. Es wurde illustriert, dass Gauß Prozesse für eine flexible Modellierung von Funktionen eingesetzt werden können. Neben einer Approximation des Computermodells liefert die Gauß Prozess Interpolation auch Unsicherheitsintervalle für die Vorhersagen auf Basis der Modellannahmen. Deren Überdeckungsgüte wurde in dieser Arbeit in Kapitel 3 anhand von Simulationsstudien erstmals untersucht. Dabei zeigte sich, dass die Gauß Prozess Interpolation geeignete Unsicherheitsintervalle liefert, wenn sich die zu approximierende Funktion nicht zu stark von der Realisation eines Gauß Prozesses unterscheidet. Dies gilt bei ausreichend großem Stichprobenumfang unabhängig von den unterschiedlichen betrachteten objektiven a priori Verteilungen für den Korrelationsparameter und auch bei Verwendung des a posteriori Modalwert Ansatzes.

Zentrales Thema dieser Dissertation ist die Kalibrierung von Computermodellen. In Kapitel 4 wurden dazu zwei Ansätze vorgestellt. Der erste Ansatz setzt ein modifiziertes nichtlineares Regressionsverfahren ein, um die Kalibrierung eines Computermodells zu erreichen. Dabei wird der quadratische Abstand zwischen Computermodell und

Realität minimiert. Kann die Realität nicht alleine über das Computermodell beschrieben werden, das heißt liegt eine Verzerrung des Computermodells vor, so liefert dieses Verfahren nur bedingt geeignete Vorhersagen. Der zweite Ansatz modelliert die Realität nicht alleine durch das Computermodell, sondern berücksichtigt bei der Kalibrierung eine additive Verzerrung. Diese Methode zur Kalibrierung wurde in dieser Arbeit anhand von Testbeispielen illustriert und ihre Sensitivität gegenüber datenabhängigen und datenunabhängigen objektiven a priori Verteilungen für die Varianzparameter verglichen. Dabei wird neben der Inversen Gammaverteilung die von Gelman (2006) in einer ähnlichen Modellsituation vorgeschlagene quadrierte  $t$ -Verteilung angewendet. Für die betrachteten Testbeispiele erwiesen sich dabei unabhängig von der Verteilungsfamilie datenabhängige a priori Verteilungen für die Varianzparameter als geeignet.

In der Praxis besteht bei der Kalibrierung von Computermodellen in der Regel eine Modellunsicherheit, das heißt es ist nicht bekannt ob ein verzerrtes oder ein unverzerrtes Computermodell vorliegt. Eine statistische Methode zur Kalibrierung sollte in beiden Modellsituationen angewendet werden können und es erlauben datenbasiert das geeignete Modell auszuwählen. In Kapitel 5 wurde eine Methode entwickelt, die eine Kalibrierung unter einer solchen Modellunsicherheit erlaubt. In den untersuchten Testbeispielen konnte gezeigt werden, dass dieser Ansatz bei unverzerrtem Computermodell gegenüber der Kalibrierungsmethode aus Kapitel 4.3 bessere Ergebnisse liefert, ohne mit einem wesentlichen Effizienzverlust für verzerrte Computermodelle einher zu gehen.

Die vorgestellten Methoden zur Kalibrierung wurden abschließend in Kapitel 6 auf die Kalibrierung einer Rückfederungssimulation für einen Umformprozess angewendet. Dabei wurden zwei unterschiedliche Werkstoffe betrachtet. Als Kalibrierparameter wurden einerseits numerische Parameter und andererseits Materialparameter berücksichtigt. Für den ersten Werkstoff lag trotz der Kalibrierung eine deutliche Verzerrung des Computermodells vor. Insgesamt konnte gezeigt werden, dass das Computermodell nicht in der Lage ist, die Realität alleine über die Wahl der numerischen Parameter zu beschreiben. Für eine geeignete Vorhersage des physikalischen Prozesses ist somit

neben dem Computermodell auch eine Verzerrungsfunktion erforderlich. Für den zweiten untersuchten Werkstoff wurden Materialparameter als Kalibrierparameter berücksichtigt. Vom Anwender wird erwartet, dass diese Parameter die wesentliche Unsicherheit über die Rückfederungssimulation enthalten. Durch Anwendung des statistischen Modells aus Kapitel 5 konnte jedoch gezeigt werden, dass auch hier die Realität nicht alleine über das Computermodell mit geeignet gewählten Materialparametern beschrieben werden kann. Vielmehr ist für die Beschreibung der physikalischen Experimente eine Verzerrungsfunktion erforderlich. Diese Verzerrung fällt jedoch etwas kleiner aus als beim vorherigen Werkstoff und der funktionale Zusammenhang zwischen Rückfederung und physikalischen Parametern wird bereits grob durch das Computermodell approximiert. Die vorliegende Verzerrung des Computermodells konnte mit Hilfe der angewendeten statistischen Methoden analysiert werden. Dabei zeigte sich, dass der Einfluss des Werkzeugradius auf die Rückfederung im Computermodell nicht korrekt vorhergesagt wird. Unter Umständen wurde hier ein physikalischer Effekt für die Rückfederung noch nicht berücksichtigt. Dies ist anhand anderer Beispiele zu verifizieren und liefert dann einem physikalischen Modellierer einen konkreten Hinweis für die Verbesserung des Computermodells. Aufgrund der Verzerrung sollten aktuell zur Vorhersage der Rückfederung neben Simulationen auch stets physikalische Experimente berücksichtigt werden.

Bei der Kalibrierung von Computermodellen gibt es noch eine Vielzahl von offenen Forschungsfragen. Hier möchte ich kurz einige Fragestellungen skizzieren. Bei der Kalibrierung von Computermodellen wurde in dieser Arbeit stets von konstanten globalen Kalibrierungsparametern mit wahren Wert  $\mathbf{u}^*$  ausgegangen. Alternativ lässt sich der Kalibrierparameter auch als Funktion der physikalischen Einflussgrößen  $\mathbf{u}^*(\mathbf{x})$  auffassen, so dass veränderliche Werte für die Beschreibung der Realität vorliegen können. Die Realität soll dann durch  $y_M(\mathbf{x}, \mathbf{u}^*(\mathbf{x}))$  beschrieben werden. Allerdings wird das Modell für die Kalibrierung mit Verzerrung bei dieser Betrachtung schnell sehr komplex und kann nur schwer durch Daten angepasst werden.

Die Versuchsplanung für die Kalibrierung von Computermodellen ist ein weiterer wich-

tiger unberücksichtigter Aspekt. In dieser Arbeit konnten an einigen Stellen Empfehlungen hinsichtlich des Designs ausgesprochen werden, eine systematische Untersuchung zu geeigneten Versuchsplänen steht jedoch noch aus. Bei den physikalischen Experimenten sind Wiederholungen erforderlich, um zwischen Zufallsfehler und Verzerrung unterscheiden zu können. Jedoch wird die Unsicherheit über den funktionalen Zusammenhang durch eine große Anzahl unterschiedlicher Designpunkte reduziert. Bei begrenztem Budget für physikalische Experimente besteht somit ein Trade-Off zwischen der Anzahl der Wiederholungen und der Raumfüllung der Designpunkte. In der betrachteten praktischen Anwendung konnte die Fehlervarianz unter den getroffenen Annahmen bereits durch eine geringe Anzahl vorliegender Wiederholungen sehr präzise geschätzt werden. Bei gleichem Stichprobenumfang wären hier unter Umständen Versuchspläne mit einer größeren Anzahl von Designpunkten bei weniger Wiederholungen für die Kalibrierung geeigneter gewesen. Für die Modellierung von Computermodellausgaben sind raumfüllende Versuchspläne geeignet. Dies bestätigte sich im Beispiel der Rückfederungssimulation auch für die Kalibrierung, da so Probleme durch unterschiedlich große Unsicherheiten bei der Gauß Prozess Interpolation vermieden werden können. Eine weitere interessante Fragestellung ist, ob die physikalischen Parameter  $\mathbf{x}$  in Simulation und Experiment möglichst gleich oder unterschiedlich gewählt werden sollten und ob sequentielle Versuchspläne für die Kalibrierung geeignet sind. Bei unverzerrtem Computermodell sind physikalische Experimente zum Beispiel an Punkten  $\mathbf{x}$  zu empfehlen, wo das Computermodell  $y_M(\mathbf{x}, \mathbf{u})$  als Funktion von  $\mathbf{u}$  eine große Variation aufweist und damit eine große Unsicherheit über die Zielgröße besteht.

Schlussendlich können bei der Kalibrierung von verzerrten Computermodellen auch alternative Konzepte für die Wahl geeigneter Verzerrungsfunktionen eingesetzt werden. Das hier verwendete Kriterium glatter Verzerrungsfunktionen ließ sich über einen Gauß Prozess probabilistisch formulieren. Alternativ können beliebige Abstandsmaße zwischen Computermodell und Realität minimiert werden. Zum Beispiel wählen Han und Santner (2009) Tuningparameter aus  $\mathbf{u}$  so, dass der  $L_2$  Abstand von Simulation und Realität minimiert wird. Da hierbei nicht länger ein probabilistisches Konzept vorliegt, ist die Erstellung von Unsicherheitsintervallen für die Vorhersage jedoch schwierig.

# A ANHANG

## A.1 Monte Carlo Methoden

Zur Generierung von Zufallszahlen aus einer Verteilung werden in dieser Arbeit Markov Chain Monte Carlo und Importance Sampling Methoden verwendet. Diese werden im Folgenden kurz beschrieben.

### A.1.1 Markov Chain Monte Carlo

Markov Chain Monte Carlo (MCMC) ist eine Methode zur Erzeugung von Zufallszahlen aus einer Verteilung mit Dichte  $f$ . Sie kann angewendet werden, selbst wenn die Dichte nur bis auf Proportionalität bekannt ist. Dies ist der entscheidende Grund warum sie in der Bayes Statistik verbreitet ist. Hier ist die Dichte der a posteriori Verteilung oft nicht vollständig bekannt, jedoch proportional zu dem Produkt aus Likelihood und a priori. Die Zufallszahlen werden aus einer (ergodischen) Markovkette generiert, deren stationäre Verteilung der gewünschten Verteilung mit Dichte  $f$  entspricht. Nach einer so genannten Einschwingphase oder Burn-In-Phase erzeugt die Kette abhängige Zufallszahlen aus dieser stationären Verteilung. Für eine genauere Beschreibung von MCMC Methoden und der zugrunde liegenden Theorie siehe Robert und Casella (2004).

Der Metropolis-Hastings Algorithmus ist ein spezieller MCMC Algorithmus bei dem eine neue Zufallszahl  $x^{(t)}$  aus einer Vorschlagsdichte  $q(x^{(t)}|x^{(t-1)})$  vorgeschlagen wird, die von der aktuellen Zufallszahl  $x^{(t-1)}$  abhängen kann.  $x^{(t)}$  wird mit Wahrscheinlichkeit  $\min\left\{\frac{f(x^{(t)})}{f(x^{(t-1)})} \frac{q(x^{(t-1)}|x^{(t)})}{q(x^{(t)}|x^{(t-1)})}, 1\right\}$  angenommen. Andernfalls wird  $x^{(t)} = x^{(t-1)}$  gesetzt. Der Startwert  $x^{(1)}$  ist zu wählen. Es kann gezeigt werden, dass die stationäre Verteilung der so generierten Markovkette der zu  $f$  korrespondierenden Verteilung entspricht (siehe Robert und Casella, 2004). Geeignete Vorschlagsdichten  $q$  sind zum Beispiel Dichten die unabhängig vom vorherigen Zustand sind; dies entspricht dem Independent Metropolis-

Hastings Algorithmus. Aber auch die Normalverteilung zentriert um den vorherigen Wert  $x^{(t)}$  ist eine geeignete Vorschlagsdichte. Sie wird im Random Walk Metropolis Algorithmus verwendet. Die Standardabweichung der Vorschlagsdichte steuert dabei die Akzeptanzrate. Eine hohe Akzeptanzrate ist jedoch kein Qualitätsmerkmal eines Random Walk Metropolis Algorithmus. Die optimale Akzeptanzrate für den Random Walk Metropolis Algorithmus beträgt 0.234 (bei multivariater Normalverteilung) wenn die Dimension  $p$  gegen unendlich geht. Bei kleineren Dimensionen ist eine Akzeptanzrate zwischen 0.234 und 0.5 optimal (Roberts und Rosenthal, 2001). Random Walk Algorithmen werden deshalb häufig zunächst in einem Anpassungsschritt eingestellt, bevor die eigentliche Simulation der Zufallszahlen startet.

Der Gibbs Sampler ist ein alternativer MCMC Algorithmus. Er ist insbesondere in hierarchischen Modellen von großer Bedeutung, da hier Zufallszahlen schrittweise aus vollständig bedingten Verteilungen („full conditionals“) generiert werden. Um eine Stichprobe aus einer gemeinsamen Verteilung von  $\mathbf{X} = (X_1, \dots, X_p)$  mit Dichte  $f$  zu generieren, wird im  $t$ -ten Schritt wie folgt vorgegangen. Generiere  $x_1^{(t+1)}$  aus  $p(x_1|x_2^{(t)}, \dots, x_p^{(t)})$  und anschließend sukzessive  $x_j^{(t+1)}$  aus  $p(x_j|x_1^{(t+1)}, \dots, x_{j-1}^{(t+1)}, x_{j+1}^{(t)}, \dots, x_p^{(t)})$  für  $j = 2, \dots, p$ . Im ersten Schritt sind Startwerte  $x_1^{(1)}, \dots, x_p^{(1)}$  zu wählen. Sind die bedingten Dichten für die Parameter nicht direkt herzuleiten, ist das Hinzufügen von Hilfsvariablen  $\mathbf{Z}$  in die zu erzeugende Zufallsstichprobe sinnvoll. Um eine Stichprobe aus der Verteilung von  $\mathbf{X}$  zu erhalten, wird zunächst eine Stichprobe aus der gemeinsamen Verteilung von  $\mathbf{X}$  und  $\mathbf{Z}$  mit Dichte  $f(\mathbf{x}, \mathbf{z})$  erzeugt. Ignorieren der Stichprobenelemente für die Hilfsvariable  $\mathbf{Z}$  liefert eine marginale Stichprobe für  $\mathbf{X}$ . Beim Gibbs Sampler werden keine Vorschläge abgelehnt. Lassen sich aus den vollständig bedingten Verteilungen leicht Zufallszahlen generieren, kann der Gibbs Sampler sehr effizient sein. Bedingung für eine irreduzible Markovkette ist hier jedoch, dass die gemeinsame Dichte die Positivitätsbedingung erfüllen muss (Robert und Casella, 2004, S. 345). Insbesondere muss der Bereich mit positiver Wahrscheinlichkeitsmasse von  $f$  zusammenhängend sein, da der Algorithmus sonst in einer Region stecken bleibt. Bei starken Korrelationen zwischen den Zufallsvariablen ist das Gibbs Sampling nicht geeignet. Diese können in einer nahezu unzusammenhängenden Dichte resultieren und so eine sehr langsame Bewegung

der Markovkette bewirken.

Um eine möglichst unkorrelierte Stichprobe aus der Zielverteilung  $f$  zu erhalten, werden die ersten  $n_{\text{start}}$  Zufallszahlen entfernt und anschließend nur jeder  $k$ -te Wert verwendet. Geeignete Einstellungen werden in der Regel über Konvergenzdiagnostiken gefunden (siehe Robert und Casella, 2004, Kapitel 12). Die insgesamt erforderliche Kettenlänge  $n_{\text{MCMC}}$  hängt wesentlich von der Dauer der Einschwingphase und der Mischungsstruktur, d.h. der Abhängigkeit aufeinander folgender Zufallszahlen ab.

### A.1.2 Importance Sampling

Neben einem MCMC Ansatz kann für die Generierung von Zufallszahlen aus der a posteriori Verteilung auch ein Importance Sampling Algorithmus eingesetzt werden (siehe Robert und Casella, 2004, Kapitel 3 und 14). Grundidee ist dabei das Integral  $\int f(\mathbf{x})dx = \int \frac{f(\mathbf{x})}{g(\mathbf{x})} \cdot g(\mathbf{x})dx$  über Monte Carlo Methoden durch  $\frac{1}{n} \sum_{i=1}^n \frac{f(\mathbf{x}_i)}{g(\mathbf{x}_i)}$  mit Hilfe einer Vorschlagsdichte  $g$  zu approximieren. Es werden Zufallszahlen  $\mathbf{x}_1, \dots, \mathbf{x}_n$  aus der Vorschlagsdichte  $g$  generiert. Um eine Stichprobe aus der Verteilung mit Dichte  $f$  zu erhalten, wird eine gewichtete Stichprobe mit Gewichten  $w(\mathbf{x}) = f(\mathbf{x})/g(\mathbf{x})$  durch Resampling Methoden erzeugt. Beim Importance Sampling ist die Schätzung eines Integrals besonders leicht: Das Integral  $\int f(\mathbf{x})dx$  wird zum Beispiel durch  $\frac{1}{n} \sum_{i=1}^n w(\mathbf{x}_i)$  geschätzt. Dieses konvergiert nach dem Gesetz der großen Zahlen bei geeignet gewählter Vorschlagsdichte gegen den wahren Wert. Ein Gütemaß für den Abstand von  $g$  und  $f$  und damit für die Stichprobengüte ist der effektive Stichprobenumfang  $\text{NESS} = \frac{1}{\sum_{i=1}^n w(\mathbf{x}_i)^2}$ .

Entscheidend für den Erfolg des Importance Samplings ist die Wahl der Vorschlagsdichte  $g$ . Je näher die Vorschlagsdichte  $g$  an der Dichte  $f$ , desto eher ergibt sich eine geeignete Stichprobe. Bornkamp (2011) schlägt zur Konstruktion der Vorschlagsdichte eine Approximation von  $f$  durch iterative Anwendung der Laplace Approximation vor. In einem ersten Schritt des Algorithmus wird dabei eine Laplace Approximation der Dichte  $f$  vorgenommen. Eine Laplace Approximation ist eine Taylor Approximation zweiten Grades von  $\log(f)$  im Modalwert  $\tilde{\mathbf{x}}$  und entspricht einer Normalverteilung

mit Erwartungswert  $\tilde{\boldsymbol{x}}$  und Kovarianzmatrix  $\boldsymbol{H}$ , der negativ inversen Hessematrix von  $\log(f)$  in  $\tilde{\boldsymbol{x}}$ . Bei mehreren Modalwerten  $\tilde{\boldsymbol{x}}_j, j = 1, \dots, m$  kann in jedem Modalwert eine Laplace Approximation mit Erwartungswert  $\tilde{\boldsymbol{x}}_j$  und Kovarianzmatrix  $\boldsymbol{H}_j$  vorgenommen werden. Die Dichte  $f$  wird dann durch eine Mischung der Normalverteilungen approximiert:  $\tilde{f}(\boldsymbol{x}) = \sum_{j=1}^m k_j \cdot N(\boldsymbol{x}|\tilde{\boldsymbol{x}}_j, \boldsymbol{H}_j)$  mit Gewichten  $k_j = (2\pi)^{p/2} \cdot |\boldsymbol{H}_j|^{1/2} f(\tilde{\boldsymbol{x}}_j)$  (siehe Gelman *et al.*, 2003, Kapitel 12). Für das Auffinden der Modalwerte sind geeignete Startwerte erforderlich. Die iterative Laplace Approximation von Bornkamp (2011) beginnt mit einer multimodalen Laplace Approximation  $\tilde{f}_0$  und approximiert im  $j$ -ten Iterationsschritt die Residuenfunktion  $r_j(\boldsymbol{x}) = f(\boldsymbol{x}) - \tilde{f}_{j-1}(\boldsymbol{x})$ , das heißt die Abweichung von aktueller Approximation  $\tilde{f}_{j-1}$  und  $f$ , wieder durch eine Laplace Approximation. Die neu gewonnenen Komponenten gehen ebenfalls in die Mischung aus Normalverteilungen ein und es resultiert eine verfeinerte Approximation von  $\tilde{f}_j$ . Die Mischungsgewichte werden dabei so gewählt, dass der quadratische Abstand zwischen  $f$  und  $\tilde{f}_j$  minimiert wird. Durch dieses Vorgehen kann auch für komplexe Dichtefunktionen eine geeignete Approximation auf Basis einfacher Komponenten konstruiert werden, die im Importance Sampling als Vorschlagsdichte verwendet wird. Im Vergleich zu MCMC Methoden kann durch dieses Vorgehen die Rechenzeit zur Generierung einer a posteriori Stichprobe deutlich reduziert werden.

## SYMBOLVERZEICHNIS

$\mathbf{X} \sim P$	Zufallsvariable $\mathbf{X}$ folgt Verteilung $P$
$p(\mathbf{X}) \propto f(\mathbf{x})$	Dichte der Zufallsvariable $\mathbf{X}$ ist proportional zu $f(\mathbf{x})$
$N(\boldsymbol{\mu}, \Sigma)$	multivariate Normalverteilung mit Erwartungswert $\boldsymbol{\mu}$ und Kovarianzmatrix $\Sigma$
$GP(\mu(\cdot), C(\cdot, \cdot))$	Gauß Prozess mit Erwartungswertfunktion $\mu(\mathbf{x})$ und Kovarianzfunktion $C(\mathbf{x}, \mathbf{x}')$
$t_\nu(\mu(\cdot), S(\cdot, \cdot))$	Student Prozess mit $\nu$ Freiheitsgraden, Erwartungswertfunktion $\mu(\cdot)$ und Skalenfunktion $S(\cdot, \cdot)$
$y_M(\cdot)$	deterministisches Computermodell
$\mathbf{z}$	Vektor der Einflussgrößen im Computermodell
$\mathcal{Z}$	Definitionsmenge der Computermodellfunktion mit $\mathbf{z} \in \mathcal{Z}$
$d_M$	Dimension des Wertebereichs $\mathcal{Z}$
$\mathbf{y}_M$	Beobachtungsvektor der Computerexperimente $\mathbf{y}_M = (y_M(\mathbf{z}_1), \dots, y_M(\mathbf{z}_{n_M}))^T$
$n_M$	Anzahl Computerexperimente
$\mathbf{D}_M$	Design für Computerexperimente $\mathbf{D}_M = (\mathbf{z}_1, \dots, \mathbf{z}_{n_M})^T$
$Y_E(\mathbf{x})$	Zufallsvariable für physikalischen Prozess in Abhängigkeit von $\mathbf{x}$
$e$	Zufallsfehler des physikalischen Prozesses
$\sigma_E^2$	Varianz des Zufallsfehlers $\text{Var}(e)$
$\mathbf{x}$	Vektor der Einflussgrößen im physikalischen Experiment
$\mathcal{X}$	Definitionsmenge des physikalischen Prozesses mit $\mathbf{x} \in \mathcal{X}$
$d_x$	Dimension des Wertebereichs $\mathcal{X}$
$\mathbf{u}$	Vektor der Kalibrierparameter des Computermodells. Es gilt $\mathbf{z} = (\mathbf{x}^T, \mathbf{u}^T)^T$ .
$\mathcal{U}$	Definitionsmenge der Kalibrierparameter mit $\mathbf{u} \in \mathcal{U}$ . Es gilt $\mathcal{Z} = \mathcal{X} \times \mathcal{U}$

---

$d_u$	Dimension des Wertebereichs $\mathcal{U}$
$\tilde{y}_M(\cdot)$	stochastischer Prozess für das Computermodell
$\mu_M(\cdot)$	Erwartungswertfunktion von $\tilde{y}_M(\cdot)$
$f_M(\cdot)$	Regressionsfunktion im Computermodell
$\boldsymbol{\beta}_M$	Vektor der Regressionsparameter zu $f_M(\cdot)$
$p_M$	Dimension von $\boldsymbol{\beta}_M$
$\sigma_M^2$	Varianz des Gauß Prozesses $\tilde{y}_M(\cdot)$
$C_M(\cdot, \cdot)$	Kovarianzfunktion von $\tilde{y}_M(\cdot)$
$R_M(\cdot, \cdot)$	Korrelationsfunktion von $\tilde{y}_M(\cdot)$
$\boldsymbol{\xi}_M$	Korrelationsparameter für $R_M$
$\boldsymbol{\psi}_M$	Vektor der Gewichte für die Exponential Korrelationsfunktion
$\boldsymbol{a}_M$	Vektor der Potenzen für die Exponential Korrelationsfunktion
$\tilde{b}(\cdot)$	stochastischer Prozess für die Verzerrung
$\mu_b$	Erwartungswert für Verzerrung $\tilde{b}(\boldsymbol{x})$ mit $\boldsymbol{x} \in \mathcal{X}$
$\sigma_b^2$	Varianz des Gauß Prozesses $\tilde{b}(\cdot)$
$R_b(\cdot, \cdot)$	Korrelationsfunktion von $\tilde{b}(\cdot)$
$\boldsymbol{\xi}_b$	Korrelationsparameter für $R_b$

## LITERATURVERZEICHNIS

- AutoForm Engineering GmbH (2002-2010). *AutoForm*, [www.autoform.com](http://www.autoform.com).
- Bayarri, M., Berger, J., Paulo, R., Sacks, J., Cafeo, J., Cavendish, J., Lin, C., und Tu, J. (2005). A framework for validation of computer models. Technical report, National Institute of Statistical Sciences.
- Bayarri, M. J., Walsh, D., Berger, J. O., Cafeo, J., Garcia-Donato, G., Liu, F., Palomo, J., Parthasarathy, R. J., Paulo, R., und Sacks, J. (2007a). Computer model validation with functional output. *Annals of Statistics*, **35**(5), 1874–1906.
- Bayarri, M. J., Berger, J. O., Paulo, R., Sacks, J., Cafeo, J. A., Cavendish, J., Lin, C.-H., und Tu, J. (2007b). A framework for validation of computer models. *Technometrics*, **49**, 138–154.
- Bayarri, M. J., Berger, J. O., Kennedy, M. C., Kottas, A., Paulo, R., Sacks, J., Cafeo, J. A., H., L. C., und Tu, J. (2009). Predicting vehicle crashworthiness: Validation of computer models for functional and hierarchical data. *Journal of the American Statistical Association*, **104**, 929–943.
- Berger, J., de Oliveira, V., und Sanso, B. (2001). Objective bayesian analysis of spatially correlated data. *Journal of the American Statistical Association*, **96**, 1361–1374.
- Bornkamp, B. (2011). Approximating probability densities by iterated laplace approximations. *Journal of Computational and Graphical Statistics*. Erscheint.
- Carnell, R. (2009). *lhs: Latin Hypercube Samples*. R package version 0.5.
- Cox, D., Park, J., und Singer, C. (2001). A statistical method for tuning a computer code to a data base. *Computational Statistics & Data Analysis*, **37**, 77–92.
- Cressie, N. (1993). *Statistics for spatial data*. Wiley, New York, Revised edition.

- Currin, C., Mitchell, T., Morris, M., und Yilvisaker, D. (1991). Bayesian prediction of deterministic functions with applications to the design and analysis of computer experiments. *Journal of the American Statistical Association*, **86**, 953–963.
- Dassault Systèmes (2010). *Abaqus FEA*, [www.simulia.com](http://www.simulia.com).
- Diggle, P. J. und Ribeiro, P. (2007). *Model-based Geostatistics*. Springer, New York.
- Fang, K., Li, R., und Sudjianto, A. (2006). *Design and Modeling for Computer Experiments*. Chapman & Hall/CRC Computer Science & Data Analysis.
- Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis*, **1**, 515–533.
- Gelman, A., Carlin, J. B., Stern, H. S., und Rubin, D. B. (2003). *Bayesian Data Analysis*. Chapman and Hall, 2nd edition.
- Geyer, C. J. (2009). *mcmc: Markov Chain Monte Carlo*. R package version 0.6.
- Green, P. J. (1995). Reversible jump markov chain monte carlo computation and bayesian model determination. *Biometrika*, **82**, 711–732.
- Gösling, M. (2010). *Metamodell unterstützte Simulation und Kompensation von Rückfederungen in der Blechumformung*. Ph.D. thesis, TU Dortmund.
- Gösling, M., Kracker, H., Brosius, A., Kuhnt, S., und Tekkaya, A. (2011). Strategies for springback compensation regarding process robustness. *Production Engineering - Research and Development*, **5**, 49–57.
- Gu, C. (2002). *Smoothing Spline ANOVA models*. Springer, Berlin.
- Han, G. und Santner, T. J. (2009). Simultaneous determination of tuning and calibration parameters for computer experiments. *Technometrics*, **51**, 464–474.
- Handcock, M. S. und Stein, M. L. (1993). A bayesian analysis of kriging. *Technometrics*, **35**, 403–410.

- Higdon, D., Kennedy, M., Cavendish, J., Cafeo, J., und Ryne, R. (2004). Combining field data and computer simulations for calibration and prediction. *SIAM Journal of Scientific Computing*, **26**, 448–466.
- Higdon, D., Gattiker, J., Williams, B., und Rightley, M. (2008). Computer model calibration using high-dimensional output. *Journal of the American Statistical Association*, **103**, 570–583.
- Hoeting, J., Madigan, D., Raftery, A., und Volinsky, C. (1999). Bayesian model averaging. *Statistical Science*, **14**, 382–401.
- Johnson, V. und Rossell, D. (2010). On the use of non-local prior densities in bayesian hypothesis tests. *Journal of the Royal Statistical Society B*, **72**, 143–170.
- Jones, D., Schonlau, M., und Welch, W. (1998). Global optimization of expensive black-box functions. *Journal of Global Optimization*, **13**, 455–492.
- Kass, R. und Raftery, A. (1995). Bayes factors. *Journal of the American Statistical Association*, **90**, 773–795.
- Kass, R. und Wasserman, L. (1996). The selection of prior distributions by formal rules. *Journal of the American Statistical Association*, **91**, 1343–1370.
- Kenett, R. und Zacks, S. (1998). *Modern Industrial Statistics: Design and Control of Quality and Reliability*. Duxbury Press, San Francisco.
- Kennedy, M. und O’Hagan, A. (2001). Bayesian calibration of computer models. *Journal of the Royal Statistical Society B*, **63**, 425–464.
- Kracker, H. und Kuhnt, S. (2009). Computer model calibration for springback prediction in sheet metal forming. In *Proceedings of Enbis EMSE 2009*.
- Kracker, H., Bornkamp, B., Kuhnt, S., Gather, U., und Ickstadt, K. (2010). Uncertainty in gaussian process interpolation. In R. Korn, L. Devroye, M. Kohler, und B. Karasözen, editors, *Recent Developments in Applied Probability and Statistics*, 79–102. Springer, Heidelberg.

- Krige, D. G. (1951). A statistical approach to some basic mine valuation problems on the witwatersrand. *Journal of the Chemical Mining and Metallurgy Society of South Africa*, **52**, 119–139.
- Kubli, W., Krasovsky, A., und Sester, M. (2008). Advanced modeling of reverse loading effects for sheet metal forming processes. In *Proceedings of NUMISHEET Conference*, 479–484.
- Li, R. und Sudjianto, A. (2005). Analysis of computer experiments using penalized likelihood in gaussian kriging models. *Technometrics*, **47**, 111–120.
- Loeppky, J., Bingham, D., und Welch, W. (2006). Computer model calibration or tuning in practice. Technical report, University of British Columbia, Department of Statistics.
- Marrel, A., Iooss, B., Laurent, B., und Roustant, O. (2009). Calculations of sobol indices for the gaussian process metamodel. *Reliability Engineering & System Safety*, **94**, 742–751.
- Matheron, G. (1963). Principles of geostatistics. *Economic Geology*, **58**, 1246–1266.
- Oakley, J. E. (2009). Decision-theoretic sensitivity analysis for complex computer models. *Technometrics*, **51**, 121–129.
- O’Hagan, A. (1978). Curve fitting and optimal design for prediction (with discussion). *Journal of the Royal Statistical Society B*, **40**, 1–42.
- O’Hagan, A. und Forster, J. (2004). *Kendall’s Advanced Theory of Statistics, Volume 2B: Bayesian Inference*. Arnold, London, 2nd edition.
- Patterson, H. und Thompson, R. (1974). Recovery of interblock information when block sizes are unequal. *Biometrika*, **58**, 545–554.
- Paulo, R. (2005). Default priors for gaussian processes. *The Annals of Statistics*, **33**, 556–582.

- Ranjan, P. und Bingham, D. (2008). Sequential experiment design for contour estimation from complex computer codes. *Technometrics*, **50**, 527–541.
- Robert, C. P. (2007). *The Bayesian Choice*. Springer, New York.
- Robert, C. P. und Casella, G. (2004). *Monte Carlo Statistical Methods*. Springer, New York.
- Roberts, G. O. und Rosenthal, J. S. (2001). Optimal scaling for various Metropolis-Hastings algorithms. *Statistical Science*, **16**, 351–367.
- Roustant, O., Ginsbourger, D., und Deville, Y. (2010). *DiceKriging: Kriging methods for computer experiments*. R package version 1.0.
- Sacks, J., Welch, W., Mitchell, T., und Wynn, H. (1989). Design and analysis of computer experiments. *Statistical Science*, **4**, 409–435.
- Santner, T., Williams, B., und Notz, W. (2003). *Design & Analysis of Computer Experiments*. Springer, New York.
- Schonlau, M. und Welch, W. (2006). Screening the input variables to a computer model via analysis of variance and visualization. In *Screening: methods for experimentation in industry, drug discovery, and genetics*, 308–327. Springer, New York.
- Seber, G. und Wild, C. (2003). *Nonlinear Regression*. Wiley, Hoboken.
- Seeger, M. (2004). Gaussian processes for machine learning. *International Journal of Neural Systems*, **14**, 69–104.
- Wahba, G. (1978). Improper priors, spline smoothing and the problem of guarding against model errors in regression. *Journal Royal Statistical Society B*, **40**, 364–372.
- Williams, B., Santner, T., und Notz, W. (2000). Sequential design of computer experiments to minimize integrated response functions. *Statistica Sinica*, **10**, 1133–1152.