4 Das Overlap-Konzept

Das Overlap-Konzept wurde in [31] ausführlich beschrieben. Im folgenden werden wir die Aspekte darlegen, die nötig sind, um das Konzept zu implementieren, und werden uns dabei eng an die in [31] vorgegebenen Definitionen und Begriffe halten. Neben den Begriffen Datenvariabilität und Modellvariabilität muss die lineare Propagation beschrieben werden, um die Überlappung zwischen Modell und Daten überhaupt auswerten zu können. Die Einbeziehung des Overlap in die Parameterschätzung wird im nachfolgenden Kapitel dargestellt.

Das Konzept entspringt der folgenden Idee: Messdaten, die bei einem Messprozess gewonnen werden, liegen in der Regel nicht als feste Zahlen vor, sondern werden als statistische Verteilungen (Datenverteilung D) angegeben. Fasst man nun diese Verteilung der Messgröße als wichtige Information auf, die uns sagt, welche Messpunkte besonders bzw. in welchem Maße anfällig für Störungen sind, bzw. welche Messpunkte völlig insensitiv gegenüber Störungen sind, und damit strukturelle Aussagen über den Prozess macht, so ist es sicher sinnvoll, zu untersuchen, ob ein Modell, welches gerade diesen Prozess abbilden soll, in ebendieser Weise auf Störungen in seinen Parametern reagiert. Die Idee ist also, die Verteilung eines Modells in jedem Messpunkt M_t mit der Verteilung des zugehörigen Datums D_t zu vergleichen und dadurch ein zusätzliches Kriterium für die Bewertung eines Modells zu erhalten.

Grundlegend dazu ist die Definition der „Modellvariabilität" oder Verteilung eines Modells in jedem Punkt (entnommen aus [30]):

Definition 1 *Modellvariabilität M(t) ist eine (statistische) Verteilung, die für jede im Modell enthaltene Zustandsgröße in einem Zeitpunkt t angibt, mit welcher statistischen Häufigkeit bestimmte Werte der Zustandsgröße angenommen werden, wenn man das Modell statistisch verteilten Störungen unterwirft.*

Störungen des Modells entstehen z.Bsp. durch Störungen in den Parametern des Modells. Fasst man die Parameter $\theta = (\theta_1, \theta_2..\theta_p)$ des Modells als statistisch verteilte Größen $\pi(\theta)$ auf, die z.B. um einen Mittelwert schwanken und deren Verteilung auch tatsächlich bekannt ist, so kann man durch Bestimmung aller Trajektorien, die durch diese Parameterverteilungen erzeugt werden, die Verteilungen des Modells $M(t)$ in jedem Zeitpunkt ermitteln. Die Parameter behalten bei jeder Trajektorieberechnung ihren Wert bei, dieser wird nicht zeitabhängig geändert.

Die Idee, nach der die Auswirkung von Parameteränderungen auf die Trajektorien eines Modells eine wichtige Rolle bei der Auswahl eines Modells spielen sollte, wurde auch schon im Zusammenhang mit Modellbewertung in der Literatur ([38],[39],[40],[41]) genutzt.

Bemerkung 2 *Diese Beschreibung, nach der die Parameter zwar nicht genau bestimmt werden können, aber ihren Wert über die Zeit behalten, entspricht dem naturwissenschaftlichen Verständnis, nach dem ein Parameter eine physikalische bzw. chemische Bedeutung haben muss und damit nicht von der Zeit abhängen kann.*

Die Modellvariabilität lässt sich formal (siehe [29]) schreiben als

$$\begin{aligned} M &: \Gamma \times \mathbb{R} - > [0, a], a \in \mathbb{R}^+ \setminus \{0\} \\ M_t(A) &= \frac{1}{C_t} \int_{\Theta} 1_A(\Phi_\theta^t y_0) \pi(\theta) d\theta \end{aligned} \tag{10}$$

für jedes $A \subset \Gamma$, wobei Γ den gesamten Raum der Zustandsgrößen bezeichnet, mit einer Konstanten C_t, die so gewählt wird, dass $\|M_t\|_2 = 1$ für jedes t.

Die Funktion 1_A ist definiert als

$$1_A(x) := \left\{ \begin{matrix} 1 \text{ falls } x \in A \\ 0 \text{ sonst} \end{matrix} \right\} \tag{11}$$

4.1 Der Overlap

Sobald die Datenverteilung und damit die Datenvariabilität D_t sowie die Modellvariabilität M_t definiert sind, kann der Overlap dieser Verteilungen definiert werden.

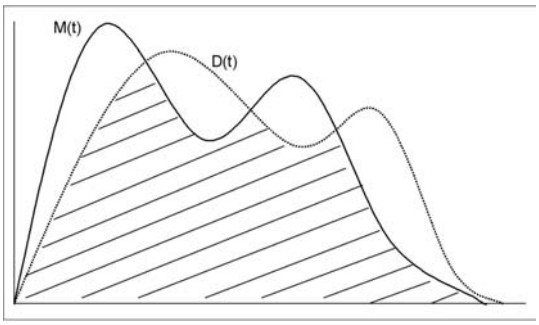

Abbildung 1: M_t und D_t in einem Zeitpunkt t, Overlap schraffiert

Definition 3 *Der Overlap in einem Punkt t ist definiert als Skalarprodukt aus Modellvariabilität M_t und Datenvariabilität D_t.*

$$F_O(t) = \langle M_t, D_t \rangle_2 \leq \|D_t\|_2 \|M_t\|_2 \leq 1$$

Der Overlap kann als eine Wahrscheinlichkeit interpretiert werden, wenn sowohl D_t als auch M_t in normierter Darstellung vorliegen.

4.2 Bestimmung der Datenvariabilität

In den meisten Fällen aus der Praxis kann für jedes einzelne Messdatum eine Normalverteilung angenommen werden $D_t \sim N(d(t), \sigma(t)^2)$.. Ist die Varianz-Kovarianz der Daten gegeben über $\Sigma_D(t)$, so kann die Datenverteilung, die wir betrachten wollen, angenommen werden als

$$D_t(x) \sim \frac{1}{\sqrt[4]{\pi}} \Sigma_D(t)^{-\frac{1}{2}} \exp^{-\frac{(x-d(t))^2}{2\sigma(t)^2}} \tag{12}$$

Der Umgang mit einer solchen Verteilung wird natürlich einfacher, wenn $\Sigma_D(t)$ in Diagonalform vorliegt. Da $\Sigma_D(t)$ symmetrisch ist, kann man dies über einen Wechsel der Basis des Zustandsraumes in eine orthogonale Basis erreichen (N sei die Anzahl der Daten).

$$\Sigma_D(t) = \begin{pmatrix} \sigma_1(t) & 0 & .. & 0 \\ 0 & & & \\ & & & \\ 0 & & & \sigma_N(t) \end{pmatrix} \tag{13}$$

Die Datenverteilung läßt sich dann schreiben als

$$D_t(x) \sim \frac{1}{\sqrt[4]{\pi}\sqrt{\sigma(t)}} \exp^{-\frac{(x-d(t))^2}{2\sigma(t)^2}} \tag{14}$$

und jedes Messdatum wird als unabhängig von den anderen Messdaten betrachtet. Da Varianzen invariant unter orthogonalen Transformationen sind, bedeutet diese Transformation keine Einschränkung bzgl. der Overlap-Auswertung. Im folgenden werden wir von $\Sigma_D(t)$ in Diagonalform ausgehen.

In der Praxis besteht die Gesamtheit der vorliegenden Messdaten aus Daten (über die Zeit) für verschiedene Zustandsgrößen, und wird oft automatisch über angeschlossene Prozessbeobachtungssysteme ermittelt. Die Autorin kann aus Erfahrung sagen, dass eine Transformation der gemessenen Daten auf eine unabhängige Form nur in den seltensten Fällen durchgeführt wird. Es ist daher Aufgabe desjenigen, der eine Messung veranlasst, die zu messenden Zustandsgrößen nach stochastischer Unabhängigkeit auszusuchen. Hier liegt tatsächlich eine Gefahr, die aber Modellierung und Parameterschätzung im allgemeinen betrifft und nicht nur in unserem Zusammenhang zum Tragen kommt.

Ebenfalls in der Praxis gibt es die Option, in Ermangelung echter festgestellter Messdatenvarianzen eine Messdatenvarianz für alle Daten anzugeben, z. Bsp. abgelesen aus der Gerätegenauigkeit. Anschaulich gesprochen wird damit um die Messdaten herum ein gleichmäßiger Schlauch gelegt. Eine solche Angabe enthält keinerlei Information über die Struktur des Prozesses und muss vermieden werden, will man den Overlap als Kriterium einsetzen.

4.3 Bestimmung der Modellvariabilität

Die Modellvariabilität M_t hängt insbesondere von der Verteilung $\pi(\theta)$ der Parameter ab. Wie aber sieht diese Verteilung aus?

Zunächst wollen wir annehmen, dass jeder einzelne Parameter normalverteilt ist: $\pi(\theta_i) \sim N(\theta_i, \Delta\theta_i^2)$, d.h. jeder Parameter verteilt sich mit einer Varianz $\Delta\theta_i^2$ um einen Erwartungswert. Die Modellvariabilität beschreibt sich zunächst durch die Frage: was passiert, wenn jeder Parameter einzeln entsprechend seiner Verteilung geändert wird. Diese Fragestellung geht von einer vollständigen stochastischen Unabhängigkeit der Parameter aus und entspricht der Annahme, dass die Verteilung $\pi(\theta) \sim N(\theta, \Delta\theta^2)$ sich mit diagonaler Kovarianzmatrix

$$\Sigma_\theta(t) = \begin{pmatrix} \Delta\theta_1 & 0 & ... & & 0 \\ 0 & \Delta\theta_2 & 0 & ... & 0 \\ & & & & \\ & & & \Delta\theta_{p-1} & \\ 0 & ... & & & \Delta\theta p \end{pmatrix} \tag{15}$$

darstellen läßt.

Bemerkung 4 *Nach der obigen Definition 3 hängt der Overlap demnach sowohl von den Daten $d(t)$ und den berechneten Trajektorien $\Phi(\theta)$ als auch von den Datenvarianzen $\sigma(t)$ und den Parametervarianzen $\Delta\theta$ ab:*

$$F_O(t) = F_O(\theta, \Delta\theta, d(t), \sigma(t))$$

Bemerkung 5 *Unter der Voraussetzung, dass die Kovarianzmatrix der Parameter Diagonalform besitzt, ist der Overlap abhängig von $2 \cdot p$ Parametern.*

Die Variabilität des Modells bestimmt sich aber grundsätzlich zusätzlich durch die Auswirkungen, die (stochastisch) voneinander abhängige Parameter bei Änderung erzeugen. Der Grad der Abhängigkeit zweier Parameter innerhalb eines Modells wird in der Kovarianzmatrix beschrieben, die wir in (15) als diagonale Matrix präsentiert haben. Angenommen, die Kovarianzmatrix sei vollbesetzt.

$$\Sigma_\theta(t) = \begin{pmatrix} \Delta\theta_{11} & \Delta\theta_{12} & & \Delta\theta_{1p} \\ & & & \\ \Delta\theta_{p1} & & & \Delta\theta_{pp} \end{pmatrix} \tag{16}$$

Als Folge davon würde der Overlap von p Parametern und p^2 Parametervarianzen abhängen. An dieser Stelle müssen wir betrachten, was eigentlich unser Ziel ist: wir wollen nicht nur den Overlap eines Modells auswerten (dazu müssten Parameter und Parametervarianzen vorliegen), sondern wir wollen eine Parameterschätzung über sämtliche Parameter (dann Parameter + Parametervarianzen) durchführen, die uns erst

die Werte für die Varianzen liefert, mit denen der Overlap ausgewertet werden soll. Eine Parameterschätzung über $p + p^2$ Parameter ist aber für komplexe Modelle mit großem p absehbar nicht möglich und nicht zuverlässig (siehe dazu auch Kap. 4.4). In der Konsequenz beschränken wir uns auf die Diagonalgestalt der Kovarianzmatrix, wohl wissend, dass im Grunde die Modellvariabilität nicht vollständig präsentiert wird.

4.3.1 Auswertung der Modellvariabilität

Für gegebene/gewählte Parameter und Parametervarianzen ließe sich jetzt durch Berechnung sämtlicher Trajektorien der Overlap des Modells mit den Daten berechnen. Dieses Vorgehen ist praktisch nicht möglich, weil bereits die Auswertung für ein Modell (rechen-)zeitaufwändig ist, selbst wenn wir uns (wie oben diskutiert) auf die Einführung von p Parametervarianzen beschränken. Sollen viele Modelle untersucht werden bzw. eine Parameterschätzung über das Modell durchgeführt werden, so gerät die Rechenzeit und der gesamte Aufwand in nicht-vertretbare Dimensionen. Damit stellt sich die Frage: Kann man die Modellvariabilität in einem Punkt in Abhängigkeit von gegebenen Parametervarianzen auf eine adäquate Weise bestimmen? Und eine weitere Frage ist aufgeworfen: Wie wählt man die Parameter und Parametervarianzen, mit denen ein Modell ausgewertet wird?

Wir haben uns entschieden, die Modellvariabilität mit Hilfe einer linearen Propagation zu bestimmen, die wir im folgenden beschreiben werden. Untersuchungen an einfachen Beispielen (siehe [31]) haben gezeigt, dass durch die Durchführung einer linearen Propagation im Vergleich zur Durchführung einer vollständigen Propagation Unterschiede in der erzeugten Modellvariabilität auftreten. Diese erwiesen sich aber modellabhängig in manchen Fällen als verschwindend klein, in anderen als deutlich größer. In der Konsequenz, nämlich der Modellbewertung, führte die unterschiedliche Propagation zu keinen Unterschieden. Der oben schon angesprochene vermutete Aufwand zur Durchführung einer vollständigen Propagation konnte schon bei diesen einfachen Beispielen nachgewiesen werden; auf komplexe Modelle hochgerechnet ist er nicht vertretbar.

4.3.2 Lineare Propagation der Parametersensitivität

Sei $\pi(\theta_0)$ als Normalverteilung eines Parametersatzes θ_0 mit einer Varianz $\Delta\theta_0$ gegeben. Die Auswirkungen der Varianz als Störung der Parameter $\theta_0 \rightarrow \theta_0 + \delta\theta_0$ auf die Modelltrajektorien $y(t) \rightarrow y(t) + \delta y(t)$ kann (aus Aufwandsgründen) nicht als vollständige, exakte Propagation der Störung bestimmt werden. Daher gehen wir auf eine lineare Propagation über (siehe dazu auch [8]) und hoffen, dass diese ausreichend gut ist.

Die lineare Propagation kann mit Hilfe der Sensitivitätsmatrix P durchgeführt werden:

$$\delta y(t) = P(t;\theta_0)\delta\theta \tag{17}$$

wobei P mit der Jacobi-Matrix des Problems (siehe (2), $\Phi_\theta^t y_0 = y(t)$) übereinstimmt

$$P(t;\theta_0) = D_\theta \Phi_\theta^t y_0|_{\theta=\theta_0} = J(\theta,t) \tag{18}$$

und die Sensitivitätsgleichung für Anfangswertprobleme (2) erfüllt

$$P^{'}(t;\theta_0) = \frac{\partial}{\partial y} f(y(t),\theta_0) P(t;\theta_0) + \frac{\partial}{\partial \theta} f(y(t),\theta_0)|_{\theta=\theta_0} \tag{19}$$

mit $P(t_0;\theta_0) = 0$ (siehe [9]). Da die Parameterverteilung normal ist und linear propagiert wird, ist auch die erhaltene Modellvariabilität normal ([1]). Daher genügt es, nur den Mittelwert und die Standardabweichung zu propagieren ([4]). Die Propagation des Mittelwerts θ_0 wird gerade von der Trajektorie $\Phi_\theta^t y_0$ geliefert. Die Varianz-Covarianzmatrix der Modellvariabilität ist gegeben durch

$$\Sigma_M(\theta, \Delta\theta, t) = J(\theta,t)\Sigma_\theta J(\theta,t)^T. \tag{20}$$

Die Varianz der i-ten Dimension der Modellvariabilität im Zeitpunkt t ist der i-te Diagonaleintrag der Varianz-Covarianzmatrix aus (20) und wird mit $\Sigma_i(\theta,\Delta\theta,t)^2$ bezeichnet. Sie berechnet sich leicht aus (20) und der Annahme einer diagonalen Kovarianzmatrix Σ_θ (15) zu

$$\Sigma_i(\theta,\Delta\theta,t)^2 = \sum_{j=1}^{p} (\frac{\partial}{\partial\theta_j}(\Phi_\theta^t y_0)_i)^2 \Delta\theta_j^2|_{\theta_j=\theta_{jo}} \tag{21}$$

Bei der Bestimmung des Overlap werden nur diese Diagonal-Einträge benötigt, die gemischten, nicht-diagonalen Einträge von $\Sigma_M(\theta,\Delta\theta,t)$ werden ignoriert, da wir für die Daten (ebenso wie für die Parameter) eine diagonale Kovarianzmatrix voraussetzen.

Insgesamt kann das Overlap-Funktional unter Berücksichtigung einer linearisierten Propagation und (unter Berücksichtigung von diagonalen Daten- und Parameterkovarianzmatrizen) daraus resultierender Modellvariabilität, die wir jetzt mit $M_{t,L}$ bezeichnen, bestimmt werden zu

$$\begin{aligned} F_L(\Phi_\theta^t y_0, \Sigma_M(\theta,\Delta\theta,t), d(t), \sigma(t)) &= \langle M_{t,L}, D_t\rangle_2 \qquad (22)\\ &= \text{(Overlap-Funktional)} \end{aligned}$$

$$\sum_{i=1}^{N} \sqrt{\frac{2\sigma_i(t)\Sigma_i(\theta,\Delta\theta,t)}{\sigma_i(t)^2+\Sigma_i(\theta,\Delta\theta,t)^2}}\, \exp(-\frac{1}{2}\frac{(\Phi_\theta^t y_0 - d_i(t))^2}{\sigma_i(t)^2+\Sigma_i(\theta,\Delta\theta,t)^2}) \tag{23}$$

wobei N die Anzahl der Daten angibt.

4.3.3 Zusammenfassung der einschränkenden Bedingungen

Folgende einschränkende Bedingungen haben wir nun in den Overlap (23) eingehen lassen:

1. wir betrachten die Messdaten als normalverteilte Größen
2. wir gehen von einer Diagonalgestalt der Kovarianzmatrix der Messdaten aus

3. wir beschreiben die Verteilung der Parameter eines Modells als Normalverteilung
4. wir betrachten in der Kovarianzmatrix der Parameter nur die Diagonalgestalt
5. wir propagieren die Varianz der Parameter linear und nicht vollständig nichtlinear.

Wir sind uns bewusst, dass die Punkte 3,4 und 5 echte Einschränkungen darstellen. Im Hinblick auf das eigentliche Ziel der Durchführung einer Parameterschätzung über den Overlap erscheinen sie aber aus Aufwandsgründen auf dem jetzigen Stand der Entwicklung notwendig.

4.4 Wahl von Parametern und Parametervarianzen

Um das Overlapfunktional tatsächlich auswerten zu können, müssen wir uns mit der Frage beschäftigen, mit welchen Parametern bzw. Parametervarianzen es ausgewertet werden soll. Messdaten und Messdatenvarianzen seien gegeben. Die Parameter können natürlich beliebig gewählt werden; für die Parametervarianzen kann die Frage nicht so leicht beantwortet werden, denn a priori gibt es keine sinnvollen Werte für die Varianzen der Parameter.

Im Zusammenhang mit der Modelldiskriminierung liegen Modelle vor, die bereits einer Parameterschätzung bezüglich der Parameter θ mit dem Ziel der Minimierung des Residuums F_R unterworfen wurden: das Residuum als Kriterium verlangt eine Parameterschätzung. Äquivalent dazu macht es wenig Sinn, den Overlap mit beliebigen Parameterwerten und (geschätzten) Varianzwerten auszuwerten. Statt dessen muss eine Parameterschätzung durchgeführt werden, die optimale Parameter UND Parametervarianzen so bestimmt, dass sowohl der Abstand der Daten vom Modell möglichst klein ist, als auch die Überlappung von Datenverteilung und Modellverteilung möglichst groß wird.

Diese Vorgehensweise definiert einen großen Unterschied zu anderen Modelldiskriminierungsverfahren, wie etwa die Berechnung des Residuums mit anschließender Auswertung der Konfidenzintervalle. Die Betrachtung der Datenvarianz findet dort erst a posteriori, also im Anschluss an die PE statt, während beim Overlap-Funktional die Datenvarianz innerhalb der PE berücksichtigt wird. Diesen Vorteil erkauft man sich natürlich durch größeren Aufwand, den wir kurz skizzieren, ohne an dieser Stelle eine genaue Aufwandsbetrachtung machen zu wollen:

Eine Optimierung zum Zwecke der Residuumsminimierung berücksichtigt p Parameter; je nach Verfahren gestaltet sich der Aufwand (Auswertung des Zielfunktionals, Bestimmung einer Jacobi-Matrix, Abarbeiten einer Anzahl Schritte usw.), der aber im Wesentlichen abhängig ist von der Anzahl Parameter.

Eine Optimierung, die das Overlap-Funktional, so wie wir es in (22) formuliert haben, minimiert, berücksichtigt wegen der Annahme einer diagonalen Kovarianzmatrix der Parameter $2 \cdot p$ Parameter - der Aufwand zur Optimierung erhöht sich stark: er verdoppelt sich schon alleine durch die doppelte Anzahl an Parametern (die Dimension der Jacobi-Matrix verdoppelt sich, die Anzahl Schritte wird sich nicht verringern). Zusätzlich muss die Propagation durchgeführt werden, um den Overlap in

allen Messpunkten überhaupt auswerten zu können. Im Vorgriff auf die tatsächlich definierte Zielfunktion (siehe (5.2.1)) erwähnen wir hier schon die Notwendigkeit, Hesse-Matrizen der Dimension $N \cdot 2p$ auszuwerten. Korrelationen zwischen Parametern und Parametervarianzen sind natürlicherweise zu erwarten, so dass dieses Problem noch stärker als bei dem normalen Residuum in den Vordergrund treten wird.

Bei der Optimierung des Overlap-Funktionals mit voller Kovarianzmatrix der Parameter, die $p + p^2$ Parameter unterstützen müsste, ist eine entsprechende Vervielfältigung des Aufwandes zu erwarten. Um das Verfahren rechenbar zu gestalten, müssen wir uns also auf die Diagonalform der Parametervarianzmatrix beschränken.