

5 Modellvalidierung und -dokumentation

Ob ein mathematisches Modell "richtig" ist, läßt sich im allgemeinen nicht beweisen. Bestenfalls kann man das Modell einer Reihe von Prüfungen unterziehen, die das Vertrauen in die Gültigkeit des Modells im Rahmen bestimmter Problemstellungen erhöhen. In diesem Kapitel wird zunächst aufgezeigt, weshalb eine Gültigkeitsprüfung bei Simulationsmodellen besonders wichtig, zugleich aber auch besonders schwierig ist (5.1). Anschließend werden Grundsätze für ein Validierungskonzept aufgestellt (5.2). Auf dieser Basis wird ein Ansatz zur Modellvalidierung in mehreren Stufen – Validierung des konzeptuellen Modells, Verifikation, Sensitivitätsanalyse, Outputvergleich/Kalibrierung, prognostische und dynamische Gültigkeitsprüfung – eingeführt (5.3). Der letzte Abschnitt gibt Hinweise für eine angemessene Dokumentation von Simulationsmodellen (5.4).

5.1 Zur Notwendigkeit einer Gültigkeitsprüfung

Die Entwicklung von mathematischen Modellen ist im allgemeinen ein komplexer Vorgang, der aus einer größeren Anzahl von Einzelschritten besteht. Diese verlaufen in der Regel nicht streng sequentiell, sondern teilweise parallel oder auch in Zyklen (Modellbildungszyklus, vgl. Kap. 1, Abb. 1-8). Jede Phase des Modellbildungsprozesses birgt zahlreiche **Fehlerquellen**. Es können Meßfehler bei der Datenerfassung auftreten, subjektive Einschätzungen des Modellentwicklers können zu Verfälschungen führen (*human bias*), beim Modellentwurf können wichtige Variablen und Relationen übersehen, falsche Parameterwerte und Variablenverknüpfungen gewählt werden, bei der Implementation auf dem Rechner Programmfehler auftreten – um nur einige Fehlerquellen zu nennen. Fehler können sich kumulieren oder sich gelegentlich auch zufällig kompensieren. Auf jeden Fall führen sie zu einer teilweisen oder vollständigen Ungültigkeit eines mathematischen Modells. Sollen Modellexperimente einen brauchbaren Ersatz für Experimente am Realsystem darstellen, muß daher die Gültigkeit (Validität) des Modells untersucht werden. Andernfalls wäre der Aussagewert solcher Experimente "weder hoch noch niedrig zu veranschlagen – er wäre überhaupt nicht abschätzbar" [Har74, S. 155]. G. S. Fishman's Argumentation zielt in die gleiche Richtung:

"Before an investigator claims that his simulation model is a useful tool for studying behavior under new hypothetical conditions, he is well advised to check its consistency with the true system, as it exists before any change is made. The success of this validation procedure establishes a basis for confidence in the results that the model generates under new conditions. After all, if a model cannot reproduce system behavior without change, we hardly expect it to produce truly representative results with change." [Fis73, S. 328 f.]

Obwohl die mathematische Modellbildung und Simulation in vielen Anwendungsgebieten inzwischen große Verbreitung gefunden hat, wird der Modellvalidierung von den Modellentwicklern selten die ihr zustehende Bedeutung beigemessen. In vielen Modellstudien bleibt die Gültigkeitsprüfung sogar gänzlich unerwähnt. Hinzu kommt das Problem der uneinheitlichen Begriffsbildung auf dem Gebiet der Validierung von Simulationsmodellen (vgl. z. B. [Nay67], [Fis73], [Har74], [Kle74], [Stü75], [Pag83], [Sar84], [Bos89]). Ein wichtiges Ziel dieses Kapitels ist es, einen Ansatz zur Durchführung von Gültigkeitsprüfungen für komplexe Simulationsmodelle zu skizzieren, der einem Modellentwickler als Leitlinie dienen kann.

Die Validierungsproblematik stellt sich grundsätzlich in allen Modellstudien – unabhängig vom mathematischen Modelltyp. Es kann sich dabei um Optimierungsmodelle aus dem *Operations Research* (z. B. der Linearen Programmierung), um Wahrscheinlichkeitstheoretische Modelle (z. B. der Warteschlangentheorie) oder um Modelle der multivariaten Statistik (z. B. der Faktorenanalyse) genauso handeln wie um ökonometrische Modelle. Bei **Simulationsmodellen** ist die Gültigkeitsprüfung jedoch besonders komplex. Während beispielsweise bei einem Optimierungsmodell der Linearen Programmierung oder von einem analytischen Warteschlangenmodell bestimmte Annahmen – z. B. der Linearität oder des statistischen Verteilungstyps – und damit Einschränkungen des Modells bekannt sind, an denen die Gültigkeitsprüfung ansetzen kann, können bei Simulationsmodellen mit ihren vielfältigen Erscheinungsformen höchst unterschiedliche Annahmen und Einschränkungen vorliegen. Die Validierung kann daher eine völlig andere Vorgehensweise erfordern. Die Gültigkeitsprüfung ist in der Simulation also besonders problematisch.

5.2 Grundsätze für ein Validierungskonzept

Die grundlegenden Überlegungen, auf denen der im folgenden Abschnitt eingeführte Validierungsansatz für Simulationsmodelle basiert, lassen sich in zehn Punkten zusammenfassen:

1. Die Validierung ist eine **entscheidende Voraussetzung** für die praktische Anwendung eines mathematischen Modells. Modelle sollten weniger nach ihrem Detail-

lierungsgrad, nach Komplexität oder Umfang beurteilt werden, sondern in erster Linie danach, wie sorgfältig und mit welchem Ergebnis ihre Validität geprüft wurde.

2. Die Modellvalidierung ist als ein Prozeß zu verstehen, dessen Ziel die Schaffung eines **angemessenen Vertrauensgrades** in ein mathematisches Modell darstellt. Dabei kann es nicht um einen exakten "Test" gehen, mit dem sich die Validität des Modells endgültig beweisen oder widerlegen ließe. Vielmehr geht es bei der Validierung vor allem darum, größtmögliche Transparenz für die Beurteilung von Modellen zu schaffen und alle Möglichkeiten auszuschöpfen, inkorrekte Modelle zu falsifizieren.
3. Modellexperimente sollen als Ersatz für Experimente mit realen Systemen dienen. Aus den Ergebnissen werden Zielaussagen gewonnen, die in der Regel in einem Entscheidungsprozeß verwendet werden. Folglich ist das eigentliche Ziel der Validierung, sicherzustellen, daß ein Modell eingesetzt wird, welches den Modellanwender zu den **gleichen Entscheidungen** führt, wie er sie unter realen Bedingungen treffen würde. Dies setzt *keine* umfassende Äquivalenz zwischen Modell und Realsystem voraus.
4. Ein mathematisches Modell sollte immer auf ein **explizit definiertes Untersuchungsziel** ausgerichtet sein. Die Gültigkeit kann nur in bezug auf diese Problemstellung nachgewiesen werden. Ändert sich das Untersuchungsziel der Modellstudie, so muß die Validität neu geprüft werden. Modelle mit universeller Gültigkeit (im Sinne einer Äquivalenz mit dem Realsystem) kann es nicht geben, da der Modellbildungsprozeß durch Abstraktion, Idealisierung und Komplexitätsreduzierung gekennzeichnet ist. Ein Modell ist grundsätzlich nur eine Approximation an die jeweils problemrelevanten Aspekte des Realsystems.
5. Die Validierung sollte auf der Basis der gleichen **Genauigkeitsanforderungen** durchgeführt werden, die auch bei realen Experimenten und Entscheidungssituationen gestellt werden.
6. Die Anwendung mathematisch-statistischer Verfahren stößt in der Modellvalidierung an Grenzen. Daher müssen neben statistischen Verfahren auch **qualitative Methoden** (wie die Plausibilitätsprüfung durch Experten, die Erfahrungswissen über das Realsystem besitzen) zur Prüfung eines Modells herangezogen werden.
7. Die Vorgehensweise bei der Validierung ist **spezifisch für bestimmte Modelltypen und Problemstellungen**. Es kann kein einheitliches, für alle Modelltypen und Untersuchungsziele verbindliches Validierungsverfahren geben.
8. Alle Schritte des Validierungsprozesses sind genau zu **dokumentieren**. Nur durch die so hergestellte Transparenz der gesamten Modellstudie wird ein Modell vertrauenswürdig.

9. Ein wichtiger Bestandteil der Modellvalidierung ist die enge Zusammenarbeit mit den späteren **Modellanwendern**.
10. Validierung darf nicht als eine Art "Restschuld" betrachtet werden, die nach Abschluß der Modellentwicklung nur dann eingelöst wird, wenn dies Zeit und Finanzmittel zulassen. Vielmehr sollte sie den **gesamten Modellbildungsprozeß** begleiten, wobei jeweils zu dokumentieren ist, welche Schritte der Modellbildung mit welchen Maßnahmen überprüft wurden.

5.3 Ein mehrstufiger Ansatz zur Modellvalidierung

Der hier eingeführte Ansatz zum Gültigkeitsnachweis von Simulationsmodellen orientiert sich an den Abbildungsebenen zwischen den Objekten der Modellbildung (Realsystem – konzeptuelles Modell – Computer-Modell), wie sie in 1.4 vorgestellt wurden. Daraus leiten sich drei prinzipielle Stufen der Modellvalidierung ab, die in Abbildung 5-1 schematisch dargestellt sind. In jeder Stufe wird eine der Beziehungen überprüft, die zwischen den drei Objekten bestehen können.

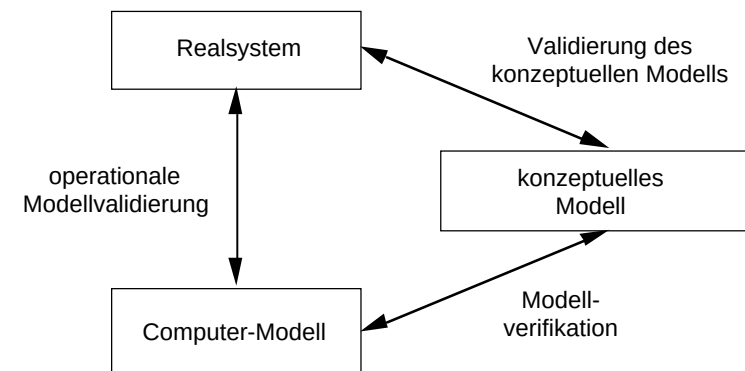


Abb. 5-1: Schematische Darstellung der Modellvalidierung (in Anlehnung an [SCS79])

5.3.1 Validierung des konzeptuellen Modells

Die erste Stufe des Gültigkeitsnachweises dient zur Prüfung der Abbildung zwischen Realsystem und konzeptuellem Modell. Dabei sind die folgenden Punkte zu prüfen.

Überprüfung der Hypothesen und vereinfachenden Annahmen

Typischerweise wird man beim Modellentwurf von bestimmten *Hypothesen* ausgehen. Diese sollten mit bekannten Theorien des Anwendungsgebietes vereinbar sein und einer Plausibilitätsprüfung durch Fachleute standhalten, die entsprechendes Erfahrungswissen über das Realsystem besitzen. Daneben sind bei der Modellbildung stets *vereinfachende Annahmen* notwendig (z. B. Verteilungsannahmen über Ankunftsprozesse). Hier kann zum Teil durch Vergleich mit empirischen Daten statistisch überprüft werden, ob die Vereinfachung zu große Abweichungen zur Folge hat und daher nicht mehr vertretbar erscheint. Außerdem können auch hierzu mehrere Experten befragt werden.

Methoden: Expertenbefragungen, statistische Anpassungstests

Datenverifikation

Sowohl für den Modellentwurf als auch zur Modellvalidierung sind in der Regel empirische Daten erforderlich. Daher müßten aus dem verfügbaren Datenbestand eigentlich zwei Teilmengen gebildet werden, die jeweils unabhängig voneinander für Modellentwurf und -validierung eingesetzt werden. Häufig scheitert jedoch diese an sich notwendige Vorgehensweise daran, daß die Datenbasis schon für den Modellentwurf kaum ausreichend ist. In jedem Fall ist eine Qualitätsprüfung ("Verifikation") aller in einer Modellstudie verwendeten Daten erforderlich. Zur Prüfung der Datenqualität können Plausibilitätskontrollen unter Mitwirkung der Modellanwender oder anderer Fachexperten herangezogen werden.

Methoden: Datenanalysemethoden, Glättungsmethoden, Expertenbefragungen

Strukturprüfung

Hier wird der innere Aufbau des Modells auf seine Plausibilität geprüft. Da es keine exakten Verfahren gibt, um die "Strukturähnlichkeit" zwischen Modell und Original zu überprüfen, ist es hier besonders wichtig, daß auch Personen an der Prüfung teilnehmen werden, die am Modellentwurf nicht beteiligt waren. Gegenüber diesen Experten muß die Struktur des Modells möglichst transparent dargestellt werden. Dabei ist zu diskutieren, ob für das Modell – gemessen am jeweiligen Untersuchungsziel – das richtige Aggregationsniveau gewählt wurde und ob die problemrelevanten Objekte, Attribute und Beziehungen korrekt erfaßt wurden.

Methoden: Begutachtung von Struktur- und Ablaufdiagrammen durch Experten

5.3.2 Modellverifikation

Die Überprüfung der Abbildung des konzeptuellen Modells auf das Computer-Modell wird als Modellverifikation bezeichnet. Es wird geprüft, ob das mit Hilfe der ausgewählten Implementationssprache erstellte Programm das konzeptuelle Modell korrekt wiedergibt. In der Praxis kommt es nicht selten vor, daß ein Simulationsprogramm völlig andere Werte liefert, als man aufgrund des konzeptuellen Modells erwarten würde. Problematischer sind jedoch Fehler, die nicht auf den ersten Blick zu erkennen sind. Auch plausibel erscheinende Ergebnisse sollten daher in systematisch ausgewählten Fällen nachgerechnet werden, um mögliche Programmfehler aufzudecken.

Der Korrektheitsnachweis, der sich angesichts der hohen Komplexität von Simulationssoftware formal nicht führen lassen wird, muß durch eine saubere Anwendung bewährter Methoden des Software-Engineering, wie die schrittweise Programmentwicklung und -verfeinerung, Modularisierung, *Information Hiding* und systematische Programmdokumentation wesentlich gestützt werden. Programmtestmethoden und die Verwendung leistungsfähiger Testhilfen sind ebenfalls hilfreich.

Methoden: Beachtung von Software-Engineering-Prinzipien, Programmtestmethoden, Kontrollrechnungen

5.3.3 Operationale Modellvalidierung

Über die *Strukturgültigkeit* eines Modells hinaus ist auch die *Verhaltensgültigkeit*, also die Ähnlichkeit der Prozeßabläufe in Realsystem und Modell, zu prüfen. Dazu müssen zahlreiche Simulationsläufe durchgeführt werden, um das **dynamische Modellverhalten** zu explorieren. Diese aufwendigste Phase der Modellvalidierung sollte die nachfolgend erläuterten Schritte beinhalten.

Plausibilitätsprüfung

Dieser Schritt zielt auf die Untersuchung des plausiblen Verhaltens des Modells und seiner Komponenten. Hierbei erweisen sich **analytische Alternativmodelle** geringerer Komplexität häufig als nützlich, die einen Vergleich mit dem Simulationsmodell zumindest für einige Spezialfälle ermöglichen. Außerdem können bekannte Theorien (z. B. über Ankunftsprozesse) oder Modellanwender und andere Kenner des realen Systems herangezogen werden, um das Modellverhalten zu interpretieren und zu beurteilen (*face validity*).

Methoden: analytische Vergleichsrechnungen, Expertenbefragungen

Sensitivitätsanalyse

In diesem Schritt der operationalen Modellvalidierung wird untersucht, wie die **Modellausgabe** auf Veränderungen der **Modelleingabe** oder der **Modellstruktur** reagiert, und von welchen exogenen Variablen, Relationen oder Parametern die Modellergebnisse am stärksten abhängen. Dadurch ist es möglich eventuelle Strukturfehler im Modell aufzudecken, kritische Größen zu erkennen, die im Modell wie in der Realität besonders beachtet werden müssen, wesentliche Einflußgrößen von unwesentlichen – und eventuell zu vernachlässigenden – Variablen zu trennen und Aussagen über die Prognosequalität des Modells zu gewinnen. Außerdem läßt sich herausfinden, ob bestimmte Modellresultate – beispielsweise in Form von Entscheidungsalternativen – nur für eingeschränkte Wertebereiche von Parametern gültig sind. Schließlich lassen sich Erkenntnisse über die Stabilität des Modells gewinnen.

Die Ergebnisse der Sensitivitätsanalyse geben somit Hinweise darauf, welche Variablen und Parameter (möglicherweise durch zusätzlichen empirischen Forschungsaufwand) besonders genau bestimmt werden müssen. Umgekehrt kann man sich bei weniger sensitiven Modellgrößen mit einer geringeren Daten- bzw. Abbildungsgenauigkeit begnügen.

Als formales Maß für die **Sensitivität** können die relativen Veränderungen einer Ergebnisvariablen Y_i und einer Eingabevariablen X_j in Beziehung gesetzt werden. Ein *Sensitivitätskoeffizient* läßt sich dann definieren als

$$S_{ij} = (dY_i / Y_i) / (dX_j / X_j) \quad (5.3-1)$$

Werte von S_{ij} nahe 0 zeugen von geringer Sensitivität; die Eingabevariablen sind folglich unkritisch für das Modellergebnis. Sehr große Werte ($|S_{ij}| \rightarrow \infty$) weisen dagegen auf die Instabilität des Modells hin. In diesem Fall sollte abgeklärt werden, ob nicht ein **Strukturfehler** des Modells die Ursache ist. Bei mittleren Werten ist die Interpretation jedoch von der Problemstellung abhängig: Unter dem Aspekt der modellgestützten Entscheidungsfindung ist letztlich von Bedeutung, wann die durch Eingabeänderungen hervorgerufenen Modellresultate zu anderen Entscheidungen führen würden.

Outputvergleich und Kalibrierung

Der Outputvergleich nimmt eine zentrale Stellung im Modellvalidierungsprozeß ein. Der einfachste Ansatz besteht darin, einen oder mehrere statistische **Kennwerte** der Beobachtungen aus dem Realsystem zu berechnen und diese mit den entsprechenden Größen des Modells zu vergleichen. Dazu eignen sich Maßzahlen wie der Mittelwert, die Stichprobenvarianz, der Median, das Minimum oder Maximum. Klassische statistische Verfahren wie der t-Test oder verteilungsfreie Verfahren wie der Rangtest bieten sich für diesen Vergleich an.

Eine andere Möglichkeit besteht in der Bestimmung eines **Konfidenzintervalls** (vgl. 4.4.3) für die Ergebnisdifferenz zwischen Simulationsdaten und empirischen Daten. Befindet sich der Nullwert innerhalb des Konfidenzintervalls, so unterscheiden sich Modell und Realsystem hinsichtlich der betrachteten Ergebnisvariablen nicht signifikant. Doch selbst wenn die Differenz statistisch signifikant sein sollte, muß das Modell für praktische Anwendungen noch nicht unbedingt als ungültig angesehen werden. Subjektiv können durchaus größere Abweichungen toleriert werden, sofern diese nicht zu anderen Schlußfolgerungen (und damit Entscheidungen) führen.

Problematisch an diesem Ansatz sind die methodischen Schwierigkeiten, auf die man unmittelbar stößt, wenn man den Outputvergleich auf mehrere Ergebnisvariablen ausdehnen will (*simultane Konfidenzintervalle*, vgl. [Kle74], S. 237).

Ein anderer Ansatz besteht darin, Verteilungen aus Realsystem und Modell mit Hilfe von **statistischen Testverfahren** zu vergleichen. Zu den verfügbaren Testverfahren gehört der (Zweistichproben-) χ^2 -Test, bei dem man analog zum χ^2 -Anpassungstest (vgl. 4.3) Abweichungen zwischen zwei Verteilungen durch Berechnung einer Prüfgröße bestimmt. Während beim Anpassungstest eine empirische Verteilung mit einem theoretischen Verteilungstyp verglichen wird, so werden hier die Abweichungen zwischen empirischen und simulierten Verteilungen gemessen. Der Wert der Prüfgröße gibt Auskunft darüber, ob die Abweichungen noch als zufällig oder schon als signifikant anzusehen sind. Signifikante Abweichungen führen zu einem negativen Ergebnis des Outputvergleichs und damit zur Ablehnung der Hypothese der Modellvalidität.

Es ist zu beachten, daß die genannten Testverfahren nicht ohne weiteres anwendbar sind, da sie unabhängige, identisch verteilte Beobachtungen erfordern. Die meisten realen und simulierten Prozesse sind jedoch durch Autokorrelation und teilweise Instationarität gekennzeichnet (vgl. 4.4). Man muß über bestimmte Maßnahmen wie die Abtrennung der transienten Anlaufphase (vgl. 4.4.2) und das Batchmittelwertverfahren (vgl. 4.4.3.3) sicherstellen, daß man stationäre und (quasi-) unabhängige Stichprobenwerte erhält, übrigens für das Modell *und* für das Realsystem.

Ein sehr datenintensiver Ansatz für den Outputvergleich besteht darin, die Eingabevariablen des Modells mit **historischen Daten** aus dem Realsystem zu beschicken und zu prüfen, ob ein Modellergebnis ähnlich den bekannten Ergebniswerten des Realsystems erzeugt wird. Hier werden also empirische Rohdaten anstelle von Zufallszahlen verwendet. Voraussetzung für diesen Ansatz ist die Verfügbarkeit sehr detaillierter empirischer Beobachtungswerte und der resultierenden Reaktionen. Nachteilig an dieser an sich plausiblen Vorgehensweise ist der hohe Datenbedarf und die geringe Allgemeingültigkeit, die auf eine mögliche Abhängigkeit von den speziellen empirischen Daten zurückzuführen ist.

Es sind vorab verschiedene statistische Verfahren für den exakten quantitativen Outputvergleich angesprochen worden. Neben großen methodischen Schwierigkeiten, auf die man unmittelbar stößt, wenn mehrdimensionale Problemstellungen vorliegen, d. h.

mehrere Ergebnisvariablen gleichzeitig zu analysieren sind, setzt auch der erwähnte Datenmangel einem quantitativen Outputvergleich Grenzen.

Daher wird man für den Outputvergleich neben exakten statistischen Testverfahren typischerweise auch **qualitative** Ansätze heranziehen müssen. Dies ist beispielsweise dann der Fall, wenn aus Gründen der Datenlage nur für die wichtigste Ergebnisvariable ein exakter Test durchgeführt werden kann, die anderen Ergebnisse jedoch einer subjektiven Beurteilung unterzogen werden, die auch für die Modellanwender akzeptabel ist. Es kann auch sinnvoll sein, nach dem Vorbild des *Turing-Tests* den folgenden Blindversuch durchzuführen: Fachleute, die mit dem modellierten Realsystem gut vertraut sind, werden vor die Aufgabe gestellt, den Modelloutput und die realen Daten zu unterscheiden. Sind sie dazu nicht in der Lage, hat das Modell diesen Test bestanden (vgl. [Law82], S. 341).

Während beim Outputvergleich Modellergebnisse und empirische Meßwerte verglichen werden, geht es bei der *Kalibrierung* um eine Einregulierung der Modellparameter mit dem Ziel, eine Angleichung der Outputs zu erreichen. Genauer versteht man unter Kalibrierung die **Anpassung eines Modells an das Realsystem** (*model fitting*) durch Veränderungen solcher Parameter, die nur ungenau oder überhaupt nicht in der Realität gemessen werden können.

Obwohl die Kalibrierung zuweilen als eigenständiger Arbeitsschritt der Validierung angesehen wird (vgl. [Stü75], S. 239-245), wird sie hier nur als untergeordneter, in sehr enger Verbindung mit dem Outputvergleich stehender Teilschritt behandelt. Denn zum einen beinhaltet die Kalibrierung einen wiederholten Outputvergleich, um die Modellanpassung nach jeder Parameteränderung zu überprüfen; somit sind prinzipiell die gleichen statistischen Methoden anwendbar. Zum anderen sollte der Kalibrierung bewußt die Bedeutung einer eigenständigen Phase der Modellvalidierung vorenthalten werden, weil sie aufgrund der doch recht willkürlichen Parameterveränderungen zu **subjektiv bestimmten Manipulationen** verführen kann. Im Extremfall könnte man nämlich an allen Parametern solange "drehen", bis Modellergebnisse und reale Meßwerte übereinstimmen. Das Problem bei dieser Vorgehensweise ist der Verlust einer Kontrollmöglichkeit: Sind die Parameter eines Modells empirisch fundiert, so ist mit dem Outputvergleich, in den ja andere empirische Größen eingehen, eine Möglichkeit zur Überprüfung des Modells gegeben. Erzeugt es mit den abgesicherten Parameterwerten ein falsches Ergebnis, muß es fehlerhaft sein; ein erfolgreicher Outputvergleich dagegen wird als Bestätigung des Modells ausgelegt. Werden die Parameterwerte jedoch durch Kalibrierung bestimmt, so wird dies das Vertrauen in ein Modell kaum stärken. Folglich ist die Kalibrierung als ein wenig exaktes, eher pragmatisches Instrument im Validierungsprozeß anzusehen, das es sehr "sparsam" einzusetzen gilt. Aufgrund der meist schlechten Datenlage in der Praxis ist dieses Instrument jedoch selten verzichtbar.

Zum kontrollierten Einsatz der Kalibrierung sollte man sich tunlichst auf Parameter beschränken, die sich im Rahmen der Sensitivitätsanalyse als *wenig sensitiv* erwiesen

haben, insbesondere auf diejenigen, die keinen bestimmenden Einfluß auf die aus den Modellergebnissen abzuleitenden Entscheidungsaussagen haben. Außerdem sollte die Parameteränderung nur innerhalb solcher Wertebereiche vorgenommen werden, die entweder empirisch abgesichert sind oder von den Modellanwendern bzw. anderen Experten als realistisch eingeschätzt werden. Schließlich ist es für die Glaubwürdigkeit der Modellstudie vorteilhafter, wenn man die Kalibrierung auf sehr wenige Parameter beschränkt und dabei die Ergebnisse der Kalibrierungsläufe genauestens offenlegt, um sich nicht einem Manipulationsvorwurf auszusetzen.

Prognostische und dynamische Gültigkeitsprüfung

Die letzte Phase der operationalen Modellvalidierung besteht in dem Nachweis der prognostischen und dynamischen Validität eines Modells. Voraussetzung für diesen Test ist jedoch, daß ein Modell tatsächlich in der Praxis eingesetzt wird, und zwar nicht nur für einmalige Analysen, sondern routinemäßig als Planungs- und Entscheidungsinstrument. Dieser **Praxistest** zeigt, ob sich das Modell unter realen Bedingungen auf Dauer bewährt. Bei der *prognostischen Gültigkeitsprüfung* muß sich erweisen, ob die Prognosen des Modells mit den eintretenden Ereignissen übereinstimmen (Prognosefähigkeit). Die prognostische Validierung kann auch als Variante des Outputvergleichs betrachtet werden, bei der die Daten des Realsystems zum Zeitpunkt der Simulation noch nicht vorliegen. Auftretende Abweichungen zwischen Modell und Realsystem können Hinweise für eine Modellrevision geben.

Doch selbst wenn ein Modell verlässliche Prognosen liefert, muß man bei einem Routineinsatz über einen längeren Zeitraum dennoch die *dynamische Modellvalidität* im Auge haben, denn die Systemdynamik mit raschen Veränderungen der Realität kann schon bald zu einer Ungültigkeit des Modells führen. Zur Sicherstellung der dynamischen Modellvalidität muß folglich ein Verfahren zur laufenden Erfassung von Informationen entwickelt werden, die anzeigen, wann Modellparameter und -struktur angepaßt werden müssen.

Methodisch unterscheiden sich die prognostische und dynamische Gültigkeitsprüfung nicht vom Outputvergleich; es kommen die gleichen statistischen und qualitativen Verfahren zur Anwendung.

Die hier aufgeführten Arbeitsschritte in den drei Stufen der Modellvalidierung sind als Maximalcatalog zu verstehen, von dem für den überwiegenden Teil der Modellstudien nur eine bestimmte Auswahl anwendbar ist. Beispielsweise ist ein historischer Outputvergleich nur für Modelle von bereits existierenden Systemen möglich – in vielen Modellstudien geht es jedoch um fiktive, hypothetische Systeme. Die prognostische und dynamische Gültigkeitsprüfung ist nur für solche Modelle relevant, die über einen längeren Zeitraum im praktischen Einsatz stehen.

Zusammenfassend läßt sich festhalten, daß der Gültigkeitsnachweis einen entscheidenden Bestandteil einer Simulationsstudie darstellt, der bisher zu wenig Beachtung

find. Validierungsaspekte sind im gesamten Modellbildungszyklus von Bedeutung. Die Modellvalidierung sollte es den Modellanwendern ermöglichen, sich ein eigenes Urteil über die Gültigkeit des Modells zu bilden. Grundvoraussetzung dafür ist eine hohe Transparenz der Modellstudie, die nur durch eine ausführliche Dokumentation aller Phasen der Modellbildung erreicht werden kann. Ein Gültigkeitsnachweis ist kein vollständig formaler Test im Sinne der Prüfstatistik sein; er wird neben quantitativen auch immer qualitative Methoden beanspruchen, die zusätzliche Möglichkeiten zur Erkennung unzulänglicher Modelle – also zur Falsifikation von Modellen – bieten. Welche Maßnahmen der Validierung im Einzelfall anwendbar sind, ist vom Modelltyp und von den verfügbaren Daten abhängig.

5.4 Modelltransparenz durch Dokumentation

Die Dokumentation einer Modellstudie hat einen hohen Stellenwert für ihre **Transparenz** und damit auch für die **Vertrauenswürdigkeit** des Simulationsmodells. S. I. Gass bemerkt dazu:

"Although no studies that demonstrate precisely which documentation is necessary or sufficient to ensure appropriate model use exist, it is clear that the lack of documentation hinders both the dissemination and use of models." [Gas81, S. 728]

Es erscheint notwendig, daß selbst dann, wenn keine allgemein akzeptierten Dokumentationsstandards für Modelle verfügbar sind, einige Grundregeln eingehalten werden, um die Kommunikation zwischen Modellersteller und -anwender zu verbessern.

In Hinblick auf die wachsende Bedeutung der Simulation in der Entscheidungsfindung ist eine Dokumentation zu fordern, die die Prämissen eines Modells offenlegt und seine Anwendung einschließlich daraus resultierender Entscheidungen nachvollziehbar macht. Eine derartige **Dokumentation für Simulationsmodelle** muß weit über eine herkömmliche Programmdokumentation hinausgehen. Sie sollte so vollständig sein, daß sie einen Anwender in die Lage versetzt,

- die Qualität des Modells selbst zu beurteilen,
- die dokumentierten Simulationsexperimente selbständig zu wiederholen und eigene Experimente durchzuführen,
- bei Bedarf eigene Arbeiten auf dem vorliegenden Modell möglichst nahtlos aufzubauen und fortzuführen, also das Modell selbst weiterzuentwickeln.

Die Dokumentation eines Modells sollte mindestens die folgenden Bestandteile enthalten (vgl. auch [Gas81]):

1. den *Modellnamen*, der als Referenz und Suchbegriff dient und auch erkennen lassen sollte, welches System das Modell beschreibt,
2. den *Namen des verantwortlichen Modellkonstruktors*, damit bei späteren Rückfragen ein Ansprechpartner vorhanden ist,
3. das *Modellbildungsziel*, das der (ebenfalls zu benennende) Auftraggeber verfolgt,
4. die *Daten*, die zur Modellbildung herangezogen wurden, sowie ihre Quellen,
5. die *Prämissen*, die der Modellbildung zugrundeliegen,
6. die *Methoden* die bei der Modellerstellung zur Anwendung kommen, z. B. Simulationskonzepte, Differentialgleichungstypen, numerische Lösungsmethoden, statistische Auswertungsmethoden,
7. die durchgeführten Maßnahmen zur *Modellvalidierung* mit ausführlicher Darstellung der Resultate,
8. die *schematische Modelldarstellung*, möglichst in Form übersichtlicher, verständlicher Modelldiagramme,
9. die *formale Modellbeschreibung*, d. h. mathematische Modellgleichungen oder Algorithmen,
10. die *kommentierten Programmlistings* einschließlich der Flußdiagramme oder Struktogramme und weiterer Programmdokumentation,
11. die *technischen Voraussetzungen* für einen Simulationslauf wie Hardwareanforderungen, Betriebssystem, Compiler, Kompatibilitäts- und Portabilitätsprobleme,
12. die *Ergebnisse* mindestens eines exemplarischen Simulationslaufs mit den zugehörigen Eingabedaten,
13. die *Analyse und Bewertung* dieser exemplarischen Ergebnisdaten bezüglich der Prämissen und der Zielsetzung, der Hypothesen und vereinfachenden Annahmen,
14. die *Anwendungsbreite* des Modells (die Klasse der Fragestellungen, auf die es anwendbar ist) sowie den Vergleich mit anderen (verwandten) Modellen,
15. die *Literaturliste*, die dem Anwender weitere Hintergrundinformationen zur Modellstudie erschließt.

Man erkennt leicht, daß die Dokumentation eines Modells beschwerlich, zeitaufwendig und damit teuer ist. Dennoch muß sie fester Bestandteil einer jeden Modellstudie sein, wenn diese nicht nur für die Modellkonstruktoren, sondern auch für andere Anwender von Nutzen sein soll. Nur durch eine gute Dokumentation kann letztlich die Kommunikation zwischen Modellkonstruktoren, Anwendern und anderen Fachleuten über das Modell ermöglicht und damit seine Qualität sichergestellt werden.