

Veröffentlichung

Qualitätssicherung von Verkehrsnachfragemodellen

Quality Control of Travel Demand Models

Eric Pestel, Markus Friedrich, Udo Heidl, Juliane Pillat, Christian Schiller, Martin Schimpf

Verfasseranschriften: Dipl.-Ing. Eric Pestel,
Univ.-Prof. Dr.-Ing. Markus Friedrich,
Universität Stuttgart,
Lehrstuhl für Verkehrsplanung und Verkehrsleittechnik,
Pfaffenwaldring 7,
70569 Stuttgart

Dipl.-Wirt.-Ing. Udo Heidl,
Dr.-Ing. Juliane Pillat,
PTV AG,
Haid-und-Neu-Straße 15,
76131 Karlsruhe

PD Dr.-Ing. habil. Christian Schiller,
Dipl.-Ing. Martin Schimpf,
Institut für Verkehrsplanung und Straßenverkehr,
Fachbereich Verkehrsnachfragemodellierung,
Hettnerstraße 1,
01069 Dresden

Veröffentlicht in: Straßenverkehrstechnik 10.2016

Pestel, E., Friedrich, M., Heidl, U., Pillat, J., Schiller, C., Schimpf, M., (2016):
Qualitätssicherung von Verkehrsnachfragemodellen. Straßenverkehrstechnik, 60, 10,
Kirschbaum Verlag, Bonn, 658-670.

Universität Stuttgart

Institut für Straßen- und Verkehrswesen

Lehrstuhl für Verkehrsplanung und Verkehrsleittechnik



Qualitätssicherung von Verkehrsnachfragemodellen

Quality Control of Travel Demand Models

Eric Pestel, Universität Stuttgart, Lehrstuhl für Verkehrsplanung und Verkehrsleittechnik

Markus Friedrich, Universität Stuttgart, Lehrstuhl für Verkehrsplanung und Verkehrsleittechnik

Udo Heidl, PTV AG, Karlsruhe

Juliane Pillat, PTV AG, Karlsruhe

Christian Schiller, TU Dresden, Institut für Verkehrsplanung und Straßenverkehr

Martin Schimpf, TU Dresden, Institut für Verkehrsplanung und Straßenverkehr

Kurzfassung

Verkehrsnachfragemodelle sind Grundlage für weitreichende verkehrspolitische und verkehrsplanerische Entscheidungen. Aus diesem Grund sollte ein Qualitätssicherungsprozess für Verkehrsnachfragemodelle von den Modellanwendern gefordert und von den Modellerstellern während des Erstellungsprozesses durchgeführt werden. In deutschsprachigen Ländern gibt es viel Erfahrung mit der Erstellung und Anwendung von Verkehrsnachfragemodellen. Diese Erfahrung ist bisher jedoch kaum formal dokumentiert und hat nicht zu Richtlinien für die Nachfragemodellierung geführt. Ziel dieses Beitrags ist es, einen Überblick über das Thema Qualität von Verkehrsnachfragemodellen im Personenverkehr zu geben. Es wird eine Vorgehensweise bei der Qualitätssicherung beschrieben und Gütemaße für die Überprüfung der Modellergebnisse diskutiert. Anschließend werden die Schritte Verifizierung der Modellspezifikation und der Modellimplementierung, Überprüfung der Eingangsdaten, Kalibrierung und Validierung beschrieben.

Abstract

Travel demand models provide the basis for extensive decisions in transport politics and transport planning. Therefore, model users should ask for a defined quality control process conducted by the model builders during the model building process. German speaking countries have a decent amount of experience with building and applying travel demand models. This experience is not yet well documented and has not led to modelling guidelines. The paper aims at providing an overview on the topic of quality of travel demand models for person transport. It describes an approach for a quality control process and discusses quality measures for checking the model results. Then the steps verification of the model specification and the model implementation, the examination of input data, the calibration and the validation are described.

Keywords

Verkehrsnachfragemodelle, Qualitätssicherung, Kalibrierung, Validierung

Keywords

Stand 17.08.2016

1 Einleitung

Verkehrsplaner nutzen Verkehrsnachfragemodelle, um die Wirkungen von zukünftigen Entwicklungen und geplanten verkehrlichen oder raumplanerischen Maßnahmen zu quantifizieren. Dabei würden sie gerne davon ausgehen, dass das eingesetzte Modell valide Ergebnisse liefert, auf deren Grundlage planerische Entscheidungen getroffen werden können. Ein Verkehrsnachfragemodell wird dann als gut bezeichnet, wenn für den heutigen Zustand beobachtete und berechnete Werte möglichst gut übereinstimmen und wenn es Wirkungszusammenhänge zwischen Verkehrsangebot und Verkehrsnachfrage realitätsnah abbilden kann. Auch wenn Verkehrsplaner immer wieder punktgenaue Prognosen veröffentlichen (z.B. +12% Personenkilometer zwischen 2010 und 2030 in der Verflechtungsprognose für den Bundesverkehrswegeplan), ist den meisten Verkehrsplanern und Modellierern sicher klar, dass selbst die besten Modellergebnisse von den tatsächlichen Werten abweichen werden. Bei der Modellerstellung und bei der Modellanwendung stellt sich aber immer wieder die Frage, welche Abweichungen sind vertretbar und wie lässt sich prüfen, ob ein Nachfragemodell die Wirkungszusammenhänge hinreichend gut wiedergibt.

In deutschsprachigen Ländern existieren bisher, anders als beispielsweise in UK (WebTAG, [1]) und in den USA (Travel Model Validation and Reasonableness Checking Manual [2]), keine etablierten Modellierungsrichtlinien zur Ermittlung der Qualität von Verkehrsnachfragemodellen. Deshalb gibt es seit einiger Zeit im deutschsprachigen Raum Bestrebungen, die Anforderungen an die Qualität von Verkehrsnachfragemodellen in Empfehlungen oder Richtlinien - ähnlich zu den Hinweisen zur mikroskopischen Verkehrsflusssimulation [3] – zu formalisieren. Hierbei kann neben den Erkenntnissen aus den USA und UK auf abgeschlossene Forschungsprojekte in Deutschland (QUALIMOD [4]) und Österreich (QUALIVERMO [5]) aufgebaut werden. Zu diesem Thema laufen derzeit zudem Forschungsprojekte in Deutschland (Projekt der Deutschen Forschungsgemeinschaft DFG: Einflussgrößen auf die Qualität von makroskopischen Nachfragemodellen im Personenverkehr [6]) und in der Schweiz (Projekt der Schweizerischen Vereinigung der Verkehrsingenieure und Verkehrsexperten SVI: Qualitätssicherung von Verkehrsmodellberechnungen [7]).

Ziel dieses Beitrags ist es, einen Überblick über das Thema Qualität von Verkehrsnachfragemodellen im Personenverkehr zu geben. Der Beitrag dient auch der Diskussion über die Inhalte einer FGSV-Empfehlung zum Einsatz von Verkehrsnachfragemodellen.

2 Vorgehensweise bei einer Qualitätssicherung

Die Qualitätssicherung bei der Nachfragemodellierung ist kein einmaliger Verfahrensschritt, sondern ein Prozess, der in alle Stufen der Modellerstellung eingebunden werden sollte. Bild 1 gibt einen Überblick über den Qualitätssicherungsprozess beim Aufbau eines Verkehrsnachfragemodells, dessen Arbeitsschritte in den folgenden Abschnitten beschrieben werden.

Bei der Modellüberprüfung ist es wichtig, zwischen den Parametern und den Variablen des Modells zu unterscheiden. **Parameter** sind die Stellschrauben des Modells, mit deren Hilfe das Modell bei der Modellerstellung an die Realität angepasst wird. Nach Abschluss der Modellerstellung sollen die Parameter nicht mehr verändert werden. Ein Verkehrsnachfragemodell umfasst mehrere Arten von Parametern:

- Geschätzte Parameter, die mit statistischen Methoden aus Erhebungen abgeleitet werden, z.B. die Parameter der Nutzenfunktion für Ziel- und Verkehrsmittelwahl.
- Gesetzte Parameter, die aufgrund mangelnder empirischer Grundlagen aus der Literatur oder aufgrund von Erfahrungen angenommen werden. Beispiele hierfür sind die Parameter der CR-Funktion oder der Widerstandsfunktion bei der Routenwahl.
- Eigenschaften des Netzmodells, z.B. Zellengröße, Zahl und Lage von Anbindungsknoten und Streckenkapazitäten.
- Eigenschaften der Nachfragemodells, z.B. Zahl der Personengruppen und Wegezwecke.

Variablen umfassen Größen, die sich im Laufe der Zeit verändern (z.B. Bevölkerungsstruktur oder Energiepreise) oder die durch planerische Entscheidungen gesetzt werden (z.B. Arbeitsplätze in einer Zelle, Anzahl der Fahrstreifen einer Strecke, Lage von Haltepunkten, Preise für Parken oder Fahrkarten).

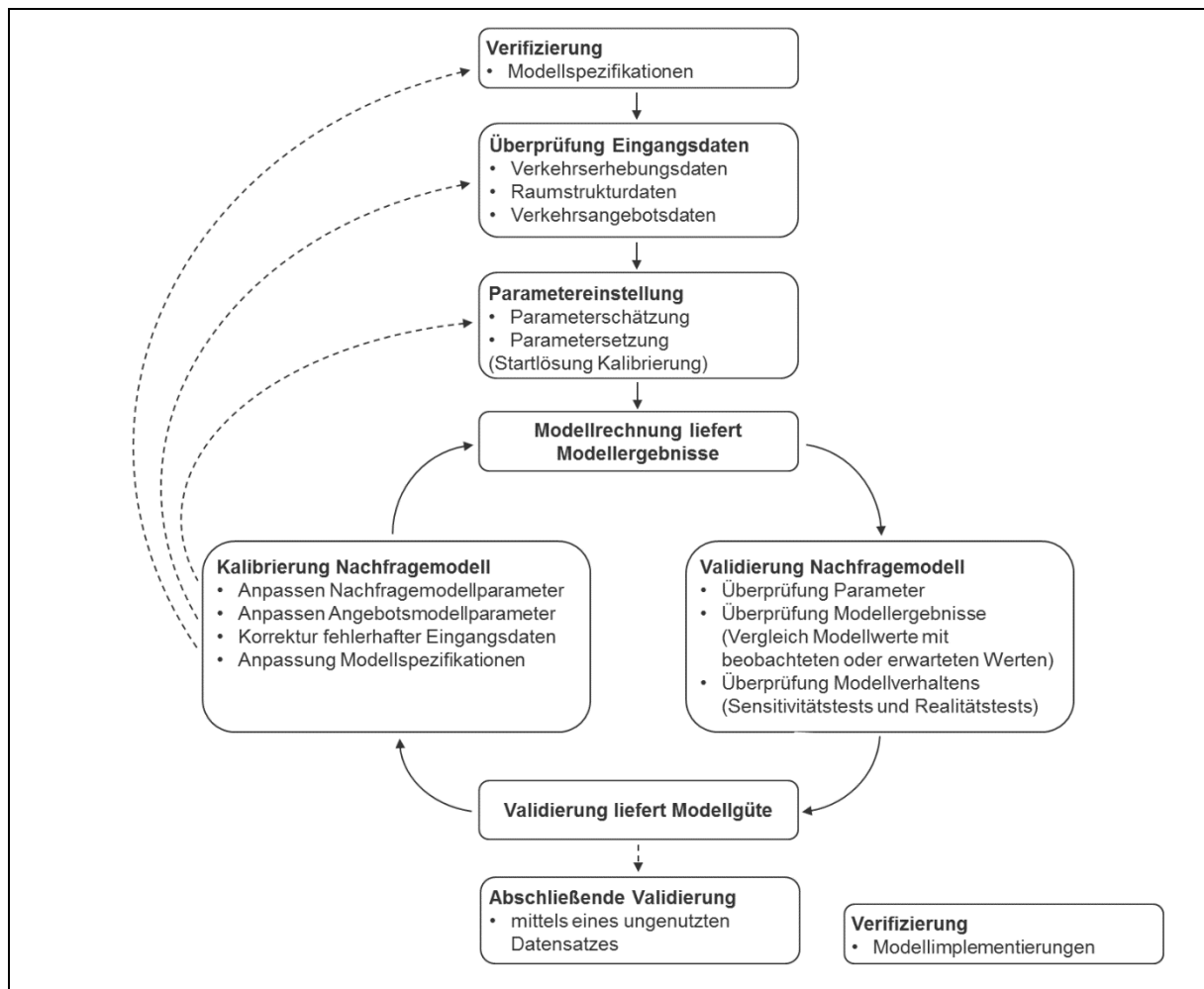


Bild 1: Qualitätssicherungsprozess beim Aufbau eines Verkehrsnachfragemodells

3 Gütemaße für die Überprüfung der Modellergebnisse

Damit der Nachweis der Validität eines Modells erbracht werden kann, müssen die Abweichungen zwischen Modellwerten m ($m = \text{Modell}$) und Erhebungswerten c ($c = \text{Count}$) durch geeignete statistische Gütemaße quantifiziert und bewertet werden. Der Vergleich von Modell- und Erhebungswerten kann sich auf Einzelwerte, Mengen von Einzelwerten und Verteilungen von Werten beziehen. Typische Werte, die bei der Validierung eines Verkehrsnachfragemodells untersucht werden, sind Zahl der Wege, Modal-Split Anteile, Reisezeiten, Reiseweiten und Verkehrsstärken. Abweichungen zwischen Modellwerten und Erhebungswerten können verschiedene Ursachen haben:

- **Messfehler:** Jede Messung kann Fehler aufweisen. Das beginnt bei der Erfassung der Wege (Werden alle Aktivitäten berichtet?) mit den zugehörigen Reiseweiten (ohne/mit Geocodierung) und endet bei der Zählung von Verkehrsmengen. Messmethoden (webbasierte, persönliche oder schriftliche Befragung, ohne/mit GPS, manuelle /automatisierte Zählung), Messgeräte und Auswertemethoden beeinflussen die resultierenden Erhebungswerte.
- **Zufällige Abweichungen bei den Erhebungswerten:** Die meisten Erhebungswerte beziehen sich nicht auf eine Erfassung der Grundgesamtheit (z.B. alle Einwohner eines Modellraums), sondern auf eine Zufallsstichprobe. Diese Zufallsstichprobe weist Abweichungen gegenüber dem wahren, unbekanntem Wert der Grundgesamtheit auf. Das Konfidenzintervall einer Messgröße definiert einen Bereich, in dem der wahre Wert für ein vorgegebenes Konfidenzniveau (z.B. 95%) liegt. Die relativen Abweichungen eines gemessenen Mittelwerts vom wahren, unbekanntem Mittelwert sind bei kleinen Stichproben größer als bei großen Stichproben.
- **Unbekannte Fehler oder Abweichungen bei den Eingangsdaten:** Ein Verkehrsnachfragemodell für den Personenverkehr bildet den Zusammenhang zwischen der Nachfrage von Einwohnern nach Aktivitäten und dem Verkehrsangebot ab. Als Ergebnis liefert es Ortsveränderungen und damit Verkehrsstärken. Die Verkehrsstärken hängen damit unmittelbar von den Raumstrukturdaten ab,

die aber immer unbekannte Ungenauigkeiten aufweisen werden (z.B. aufgrund von Fehlern in der amtlichen Statistik, von Zweitwohnsitzen und Wochenendpendler).

- **Modellabstraktion:** Ein Nachfragemodell abstrahiert die realen Werte. Beispiele für diese Abstraktion sind Verkehrszellen, Anbindungsknoten oder die Abgrenzung des Modellraums. Diese Abstraktionen führen dazu, dass der Zellbinnenverkehr keinen Beitrag zu Verkehrsstärken liefert, dass Straßen im Bereich von Anbindungen unrealistische Verkehrsstärken aufweisen können, dass Reisezeiten zwischen nahegelegenen Verkehrszellen von den Anbindungspunkten beeinflusst werden und dass Teile der Nachfrage nicht modelliert, sondern aus anderen Datenquellen ergänzt werden.
- **Fehler in der Modellspezifikation oder in der Modellimplementierung:** Ein ungeeignetes Modell wird die Erhebungswerte nicht richtig abbilden können.

Aus den genannten Ursachen lassen sich für die Quantifizierung der Modellgüte eines Verkehrsnachfragemodells folgende Aussagen ableiten:

- Ursache für Abweichungen zwischen Modellwerten und Erhebungswerten können Fehler auf der Modellseite, Messfehler und zufällige Schwankungen in der Stichprobe sein.
- Die relativen Abweichungen zwischen Modellwerten und Erhebungswerten sollten bei Werten, die sich aus vielen Ortsveränderungen ergeben, geringer sein als bei Werten, die nur aus wenigen Ortsveränderungen resultieren. Eine hohe Verkehrsstärke muss eine geringere relative Abweichung als eine niedrige Verkehrsstärke aufweisen. Der Modal-Split Anteil für den Modus Rad darf größere relative Abweichungen aufweisen als der Modal-Split für den Modus Pkw. Die mittlere Reisezeit der Personen einer Personengruppe oder eines Teilraumes kann größere Abweichungen ergeben als die mittlere Reisezeit über alle Einwohner des Modellraums.
- Vergleiche, die nur auf einer kleinen Zahl von Ortsveränderungen beruhen, sollten nicht zur Überprüfung der Modellgüte herangezogen werden.
- Abweichungen, die ein definiertes Gütemaß überschreiten, müssen interpretiert und zum Beispiel mit den Modellabstraktionen erklärt werden.

3.1 Einzelwerte

Bei der Überprüfung von Einzelwerten wird ein einzelner Modellwert m mit einem Erhebungswert c verglichen. Für jedes Wertepaar wird eine Aussage über die Güte der Übereinstimmung gemacht. Beispiele für Einzelwerte sind Verkehrsstärken an einer Zählstelle, Fahrzeiten auf einer Route oder mittlere Reisezeiten für einen Wegezweck. Um die Übereinstimmung zu quantifizieren, kann die absolute oder die relative Abweichung des betrachteten Wertepaares herangezogen werden. Es ist leicht zu erkennen, dass alleine mit der absoluten oder mit der relativen Abweichung keine angemessene Bewertung möglich ist. Bei hohen Verkehrsstärken (z.B. 10.000 Fz/d) sind – wie oben erläutert – höhere Anforderungen an die relative Abweichung zu stellen als bei geringen Verkehrsstärken (z.B. 1.000 Fz/d). Für die absolute Abweichung ist es umgekehrt. Hier sind bei hohen Verkehrsstärken größere Differenzen akzeptabel als bei geringen Verkehrsstärken. Deshalb sind Gütemaße wünschenswert, die sowohl die relative als auch die absolute Abweichung berücksichtigen. Ein verbreitetes Gütemaß, das diese Anforderungen für die Bewertung der Verkehrsstärke erfüllt, ist der sog. *GEH*-Wert:

$$GEH = \sqrt{\frac{2 \cdot (m - c)^2}{m + c}}$$

mit

m Verkehrsstärke des Modells ($m = \text{Model}$)

c Verkehrsstärke der Erhebung ($c = \text{Count}$)

Dieser Wert ist nach Geoffrey E. Havers benannt, der ihn für die Zwecke der Verkehrsplanung einführte. Der *GEH*-Wert wird u.a. in den englischen Richtlinien (WebTAG Unit 3.1 [1]) als Gütemaß herangezogen. Hier wird für stündliche Verkehrsstärken ein *GEH*-Sollwert von 5 gefordert. Das nachstehende Bild 2 zeigt für drei Sollwerte des *GEH* (5, 10, 15) den Verlauf der unteren und oberen Modellwerte m , die sich bei einer gemessenen Verkehrsstärke c ergeben. Der *GEH* erfüllt die planerische Vorstellung, dass bei höheren Erhebungswerten geringere relative Abweichungen eingehalten werden sollen. Bezogen auf einen Sollwert des *GEH* von 5 ist bei einer Verkehrsstärke von $c = 1.000$ Fahrzeugen für die nach unten abweichenden Modellwerte m eine relative Abweichung von

16% zulässig. Die zulässige relative Abweichung reduziert sich bei einer Verkehrsstärke von $c = 10.000$ Fahrzeugen auf 5%.

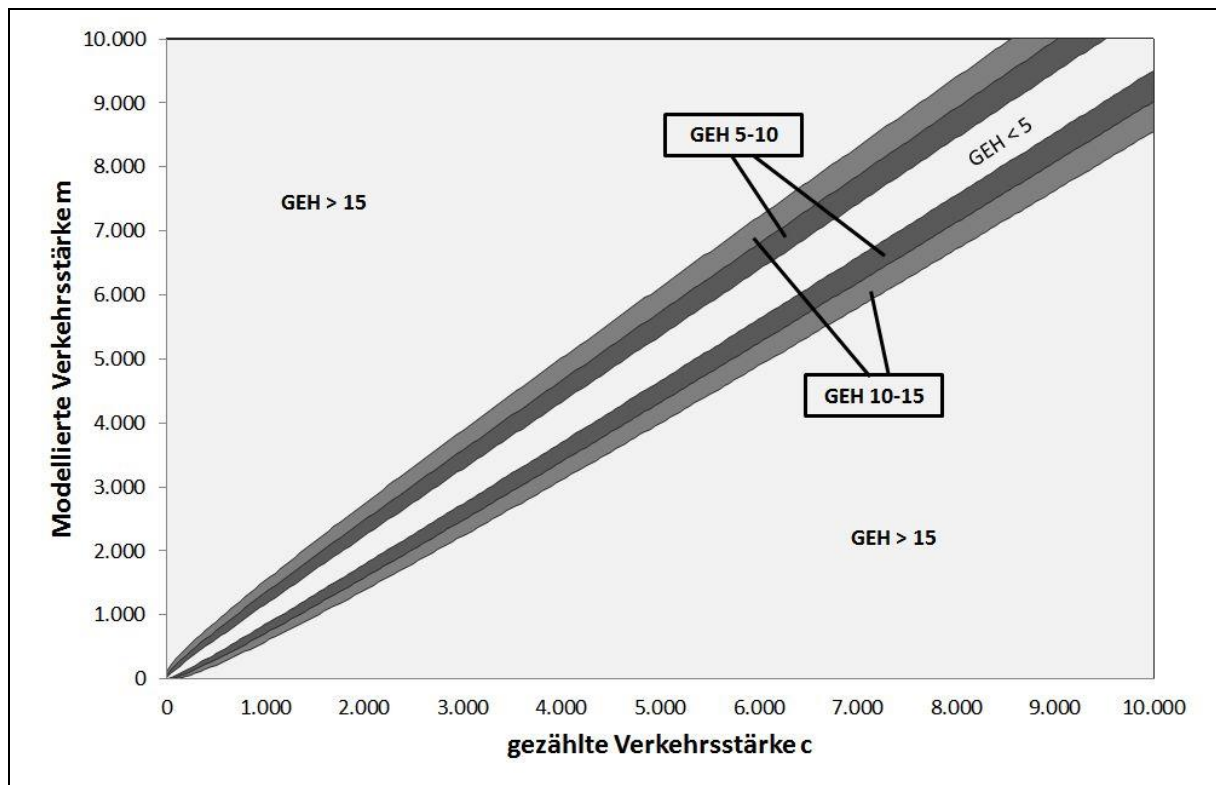
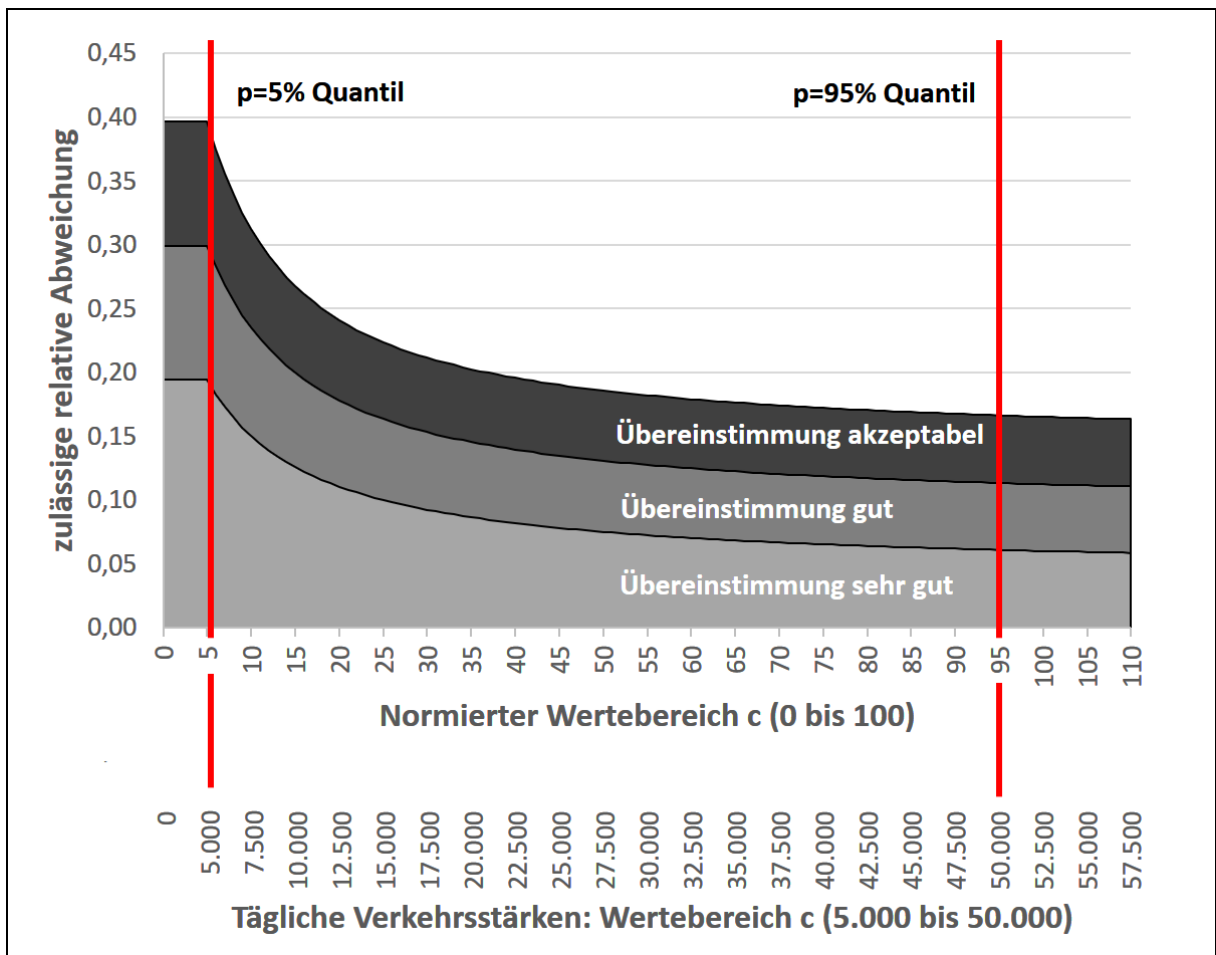


Bild 2: Verlauf der unteren und oberen Grenzen für die berechneten Werte m für einen GEH -Wert von 5, 10 und 15

Neben wünschenswerten Eigenschaften weist der GEH aber auch problematische Eigenschaften auf. Der GEH ist nicht selbstskalierend, d.h. die Sollwerte müssen abhängig vom Wertebereich der zu beurteilenden Größe angepasst werden. Damit beispielsweise die Zählwerte einer Stundenzählung und einer Tageszählung an einer Zählstelle vergleichbare Ergebnisse ergeben, muss der Sollwert für die Tageszählung erhöht werden. Deshalb ist der GEH auch nicht für die Bewertung anderer Kenngrößen, z.B. von Reisezeiten, geeignet. Außerdem ist der GEH nicht spiegelsymmetrisch, d.h. eine gleiche absolute Abweichung nach oben und unten wird unterschiedlich bewertet.

- Aus der Diskussion zum GEH lassen sich folgende Anforderungen an ein Gütemaß für Einzelwerte ableiten: Das Gütemaß sollte zwischen 0 (keine Übereinstimmung) und 1 (perfekte Übereinstimmung) liegen.
- Die zulässige absolute Abweichung soll bei großen Werten größer als bei kleinen Werten sein.
- Die zulässige relative Abweichung soll bei kleinen Werten größer als bei großen Werten sein.
- Das Gütemaß sollte auf Werte mit unterschiedlichem Wertebereich (z.B. Tageswerte und Stundenwerte für Verkehrsstärken oder Reisezeiten in Stunden und Minuten) anwendbar sein.
- Das Gütemaß soll spiegelsymmetrisch bzgl. der Abweichung vom Erhebungswert c sein.
- Die Berechnungsvorschrift des Gütemaßes sollte mit möglichst wenigen Parametern auskommen.

Bild 3 zeigt einen Vorschlag für eine Bewertung von Einzelwerten für Kenngrößen mit unterschiedlichen Wertebereichen. In einem ersten Schritt wird eine geeignete Bewertungsfunktion $f(c)$ gewählt, die die zulässige relative Abweichung für Erhebungswerte c im Wertebereich 0 bis 100 definiert. Diese Bewertungsfunktion ist aber auch für Werte über 100 gültig und konvergiert für große Werte von c gegen eine maximale zulässige relative Abweichung (z.B. 5%). Die Skalierung für eine Kenngröße erfolgt dann über die Quantile des Wertebereichs, die typische untere und obere Grenzen der Kenngröße definieren. Hierfür kann das 5% und das 95%-Quantil herangezogen werden. Über diese beiden Stützstellen werden dann die Werte der c -Achse angepasst, d.h. proportional verteilt. Im Diagramm ist das beispielhaft für den Wertebereich von täglichen Verkehrsstärken dargestellt. In der Tabelle unter dem Diagramm sind für ausgewählte Kenngrößen typische Werte des 5% und des 95%-Quantils eingetragen und die daraus resultierenden zulässigen absoluten Abweichungen ausgewiesen. Die Anforderungen an die absoluten Abweichungen einer Verkehrsstärke von 5.000 Fahrzeugen sind dabei von der betrachteten Zeitdauer abhängig. Stundenwerte müssen höhere Anforderungen erfüllen als Tageswerte.



Kenngröße	Einheit	unterer Grenze p=5%	obere Grenze p=95%	zul. absolute Abweichung bei p=5%		zul. absolute Abweichung bei p=95%	
				sehr gut	akzeptabel	sehr gut	akzeptabel
Verkehrsstärke Tag	[Fz/d]	5.000	50.0000	±1.000	±2.000	±3.000	±8.300
Verkehrsstärke Stunde	[Fz/h]	500	5.000	±100	±200	±300	±830
Anz. Wege pro Person	[Wege/d]	1	5	±0,2	±0,4	±0,3	±0,8
Anz. Wege pro Person x Zweck	[Wege/d]	0,2	1	±0,04	±0,08	±0,06	±0,17
mittlere Reiseweite	[km]	2	50	±0,4	±0,8	±3,0	±8,3
mittlere Reisezeit	[min]	10	60	±2,0	±4,0	±3,6	±10,0

Bild 3: Möglichkeit zur Beurteilung von Einzelwerten mithilfe einer zulässigen relativen Abweichung, die über die Quantilwerte der zu bewertenden Kenngröße skaliert wird.

In die Klasse der Einzelwerte fallen auch Aussagen, für die mit einem Ja/Nein-Test überprüft wird, ob eine Anforderung erfüllt wird. Tabelle 1 zeigt Beispiele für Ja/Nein Tests.

Test	Beispiel
Liegt ein Wert im erwarteten Wertebereich?	<ul style="list-style-type: none"> Liegt das Reisezeitverhältnis ÖV/Pkw auf allen Relationen im zu erwartenden Bereich (z.B. zwischen 0,5 und 10)? Liegen die Parameter für die Mobilitätsrate in einem Bereich, der aus anderen Erhebungen bekannt ist?
Erfüllt das Modell die geforderten Konvergenzbedingungen?	<ul style="list-style-type: none"> Werden bei einer Gleichgewichtsumlegung und bei einer Rückkopplung zwischen Umlegung und Nachfrageberechnung die vorgegebenen Konvergenzmaße eingehalten?
Erfüllt das Modell die geforderten Konsistenzbedingungen?	<ul style="list-style-type: none"> Ist für alle Zellen erfüllt, dass die Zahl der Wege, die innerhalb eines Tages in der Zelle beginnen und enden, identisch ist? Verteilen sich die Wege zur Arbeit gleichmäßig auf die angebotenen Arbeitsplätze?

Tabelle 1: Beispiele für Ja/Nein Tests

3.2 Mengen von Einzelwerten

Der Vergleich von Mengen von Einzelwerten vergleicht eine Menge von modellierten und gemessenen Wertepaaren und führt zu Aussagen über die Güte der Menge. Die betrachteten Wertepaare sollten dabei (weitgehend) unabhängig voneinander sein. Beispiele für Wertemengen sind Verkehrsstärken an mehreren Zählstellen oder Fahrzeiten über mehrere Relationen. Um statistische Gütemaße für Mengen von Einzelwerten anhand vorgegebener Sollwerte bewerten zu können, müssen die Einzelwerte in Werteklassen gruppiert werden. Sammer et al. [5] geben einige Hinweise für die Klassenskalierung:

- Es sollen mindestens 3 (besser 5) Klassen vorhanden sein.
- Jeder Klasse sollen mindestens 10 Elemente (z.B. Messquerschnitte) zugeordnet werden können, um einen Mindeststichprobenumfang sicherzustellen und dadurch eine Normalverteilung annehmen zu können. Befinden sich weniger als 10 Elemente in einer Klasse, so ist diese mit den benachbarten Klassen zusammenzufassen.
- Das arithmetische Mittel der Elemente einer Klasse sollte ungefähr der Klassenmitte entsprechen.
- Um mehrere Planfälle und Modelle besser vergleichen zu können, schlagen Sammer et al. [5] eine einheitliche Klasseneinteilung für Verkehrsstärken vor, die in Bild 4 als dargestellt ist (schwarz dargestellte Klassen). Das bereits beschriebene Problem der „Selbstskalierung“ bleibt bei diesem Vorgehen allerdings bestehen.

Um die Übereinstimmung der modellierten und gemessenen Wertemengen – insbesondere Verkehrsstärken – zu bewerten, werden häufig die nachstehend erläuterten Gütemaße herangezogen. Für diese Gütemaße gilt, dass sie nur die Menge der Werte beurteilen. Fehlerhafte Einzelwerte können dabei übersehen werden, obwohl sie z.B. aufgrund ihrer Lage von großer Bedeutung sind. Aus Sicht der Autoren sollten diese Gütemaße deshalb nur ergänzend zu den Gütemaßen für die Einzelwerte herangezogen werden.

Korrelationskoeffizient und Bestimmtheitsmaß

Der Korrelationskoeffizient R und das Bestimmtheitsmaß R^2 stellen Qualitätsmaße dar, die die Abhängigkeit von zwei Datensätzen überprüfen. R ist eine dimensionslose Größe zwischen -1,0 und 1,0. Sie spiegelt das Ausmaß der linearen Abhängigkeit zwischen zwei Datensätzen wieder. Das Bestimmtheitsmaß drückt aus, wie hoch der Anteil der Variation ist, der von den unabhängigen Variablen des Modells erklärt wird. Dabei gilt, dass mit einem größer werdenden Bestimmtheitsmaß der Anteil der Daten höher ist, die durch die Regressionsfunktion erklärt werden. Weiterhin wird R^2 als Verhältnis der Varianz der abhängigen Variable zur Varianz der unabhängigen Variablen interpretiert. Diese Interpretation kann jedoch nur teilweise auf die Validierung einer Verkehrsumlegung übertragen werden, da die Umlegung bedingt durch verschiedene Kalibrierungs- und Validierungsdatensätze nicht vollständig unabhängig von den Verkehrszählungen ist. Weiterhin ist kritisch zu betrachten, dass das Bestimmtheitsmaß einerseits nichts über die Kausalität des betrachteten Zusammenhangs aussagt und andererseits mit zunehmender Anzahl an untersuchten Einflussfaktoren zwangsläufig steigt [8; 2]. R^2 ist daher nur mit Einschränkungen als Gütemaß geeignet. In der Literatur genannte Grenzen für akzeptable R^2 -Werte schwanken zwischen 0,88 [2], 0,95 [2] und 0,98 [1].

$$R = \frac{\left(N \cdot \left(\sum_{n=1}^N (c_n \cdot m_n) \right) - \sum_{n=1}^N c_n \cdot \sum_{n=1}^N m_n \right)^2}{\left(N \cdot \left(\sum_{n=1}^N c_n^2 \right) - \left(\sum_{n=1}^N c_n \right)^2 \right) \cdot \left(N \cdot \left(\sum_{n=1}^N m_n^2 \right) - \left(\sum_{n=1}^N m_n \right)^2 \right)}$$

mit

N Zahl der betrachteten Strecken

m_n n. Modellwert

c_n n. Erhebungswert

Da das Bestimmtheitsmaß nichts über die Streuung der Daten aussagt, bietet es sich zusätzlich an, eine lineare Regressionsfunktion zu bestimmen. Die Steigung der Regressionsgeraden, die zwar auch keine Aussage zur Streuung liefert, aber ein eigenes Qualitätsmaß für die Übereinstimmung der modellierten und erhobenen Daten darstellt, sollte laut WebTAG (Unit 3.1) [1] für Verkehrszellenwerte zwischen 0,98 und 1,02 liegen.

Für die Geradensteigung ist es allerdings entscheidend, ob der Schnittpunkt zwischen Gerade und Koordinatenursprung erzwungen wird. Die Steigung und das Bestimmtheitsmaß weisen bei einem erzwungenen Schnittpunkt auf eventuelle systematische Fehler hin (z.B. die Modellwerte sind immer zu groß), allerdings geht durch dieses Erzwingen auch ein zusätzliches Qualitätsmaß, der Ordinaten Schnittpunkt, verloren. Idealerweise sollte der Schnittpunkt von Regressionsgerade und Ordinate nahe dem Koordinatenursprung liegen.

Wurzel des mittleren quadratischen Fehlers

Dieses Qualitätsmaß, das u.a. im Travel Model Validation and Reasonableness Checking Manual [2] vorgeschlagen wird, findet im Bereich der Verkehrsnachfragemodellierung häufig Anwendung beim Vergleich von Verkehrsstärken. Voraussetzung ist, dass die Strecken nach Straßenklasse (funktionale Klassierung) oder Verkehrsstärke (qualitative Klassierung) klassifiziert sind. Zusätzlich ist es möglich, die Strecken gemäß ihrer geographischen Lage zusammenzufassen, was die Validierungsinformation weiter erhöht. Eine allzu sehr aggregierte Wurzel des mittleren quadratischen Fehlers sagt allerdings wenig aus und lässt somit keine Rückschlüsse auf das Gesamtmodell zu. Neben der Beurteilung der Umlegungsqualität dient die Wurzel des mittleren quadratischen Fehlers auch zur Einschätzung der Modellstufenqualität [5; 2].

Der Wert der Wurzel des mittleren quadratischen Fehlers ist unter verschiedenen Bezeichnungen und Akronymen geläufig, die an dieser Stelle kurz aufgeführt werden sollen, um auf die unterschiedlichen Definitionen aufmerksam zu machen. Sammer et al. [5] verwenden den Begriff der Wurzel der mittleren Abweichungsquadrate, wobei sie hierbei noch in die „absolute Wurzel der mittleren Abweichungsquadrate“ (*AWA*) und die „prozentuale Wurzel der mittleren Abweichungsquadrate“ (*PWA*) unterscheiden. Um zu berücksichtigen, dass verschiedene Zählstellen eine unterschiedliche Genauigkeit aufweisen, stellen die Autoren des QUALIVERMO eine zusätzliche Variante der Wurzel des mittleren quadratischen Fehlers vor, die eine Genauigkeitsgewichtung vornimmt.

In den Hinweisen zur mikroskopischen Verkehrsflusssimulation [3] wird die prozentuale Wurzel des mittleren quadratischen Fehlers als „Root Mean Squared Percent Error“ (*RMSPE*) verwendet. Sie entspricht der Berechnungsformel des *PWA* und stellt die Größe des Fehlers relativ zum Mittelwert dar.

Die englischen und amerikanischen Richtlinien (z.B. Travel Model Validation and Reasonableness Checking Manual [2]) sprechen vom „Root Mean Squared Error“ (*RMSE*) bzw. von dessen prozentualer Form, dem „Percent Root Mean Squared Error“ (*%RMSE*). Während der *RMSE* und die österreichischen *AWA* eine identische Berechnungsvorschrift aufweisen, weichen die Formel des *%RMSE* und des österreichischen *PWA* (bzw. des deutschen *RMSPE*) voneinander ab. Der *RMSE* und der *%RMSE* berechnen sich auf folgende Weise:

$$RMSE = \sqrt{\frac{\sum_{n=1}^N ((m_n - c_n)^2)}{N}}$$

$$\%RMSE = \frac{RMSE}{\frac{\sum_{n=1}^N c_n}{N}} \cdot 100\% = \frac{\sqrt{\frac{\sum_{n=1}^N ((m_n - c_n)^2)}{N}}}{\frac{\sum_{n=1}^N c_n}{N}} \times 100\%$$

mit

N Zahl der Strecken in der Klasse

m_n n. Modellwert in der betrachteten Klasse

c_n n. Erhebungswert in der betrachteten Klasse

Für den $\%RMSE$ finden sich in den Regelwerken amerikanischer Bundesstaaten Vorschläge für Richtwerte, die in Bild 4 dargestellt sind. Dabei ist zu beachten, dass die Klassenbreiten zwischen den einzelnen Richtlinien variieren.

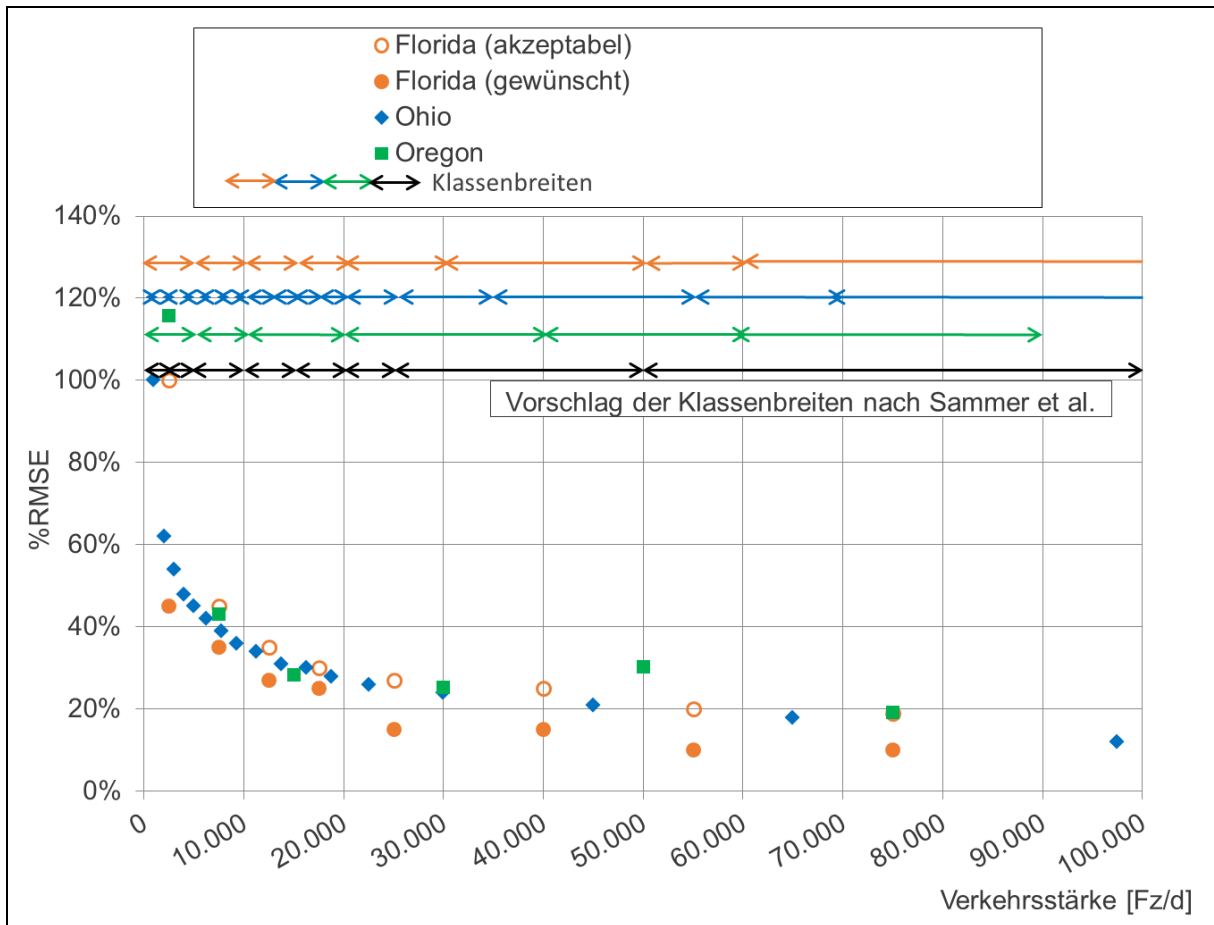


Bild 4: Zulässige $\%RMSE$ -Werte in Abhängigkeit von der Verkehrsstärkenklasse [2] und Klasseneinteilung nach Sammer et al [5].

Häufigkeitsverteilungen von Einzelwerten

Für Mengen von Einzelwerten lassen sich neben ihrer Betrachtung als Einzelwerte, Häufigkeitsverteilungen der statistischen Gütemaße für Einzelwerte ausweisen. Sie geben den Anteil der Wertepaare an, deren Gütemaße in bestimmte Gütemaßklassen (z.B. *GEH* 0 bis 5, 5 bis 10, > 10) entfallen.

3.3 Verteilungen

Die Übereinstimmung von modellierten und gemessenen Verteilungen (z.B. Reiseweiten- oder Reisezeitenverteilungen) kann beispielsweise durch folgende statistische Gütemaße erfasst werden:

- Lagemaße der Verteilung (Mittelwert einer Verteilung, Perzentile einer Verteilung), die dann wie Einzelwerte beurteilt werden können.
- der Korrelationskoeffizient R bzw. das Bestimmtheitsmaß R^2 .
- das Koinzidenz-Verhältnis (Coincidence Ratio CR), das wie folgt definiert ist:

$$CR = \frac{\sum_{n=1}^N \min\{m_n, c_n\}}{\sum_{n=1}^N \max\{m_n, c_n\}}$$

mit

N Zahl der Intervalle in der Häufigkeitsverteilung

m_n Modellwert im n . Intervall

c_n Erhebungswert im n . Intervall

Das Coincidence Ratio CR liegt zwischen 0 und 1. Ein Wert von 1,0 ergibt sich bei einer vollständigen Übereinstimmung der beiden Verteilungen. Das Travel Model Validation and Reasonableness Checking Manual [2] schlägt einen Sollwert von mindestens 0,7 vor. Bild 5 zeigt zwei Reisezeitenverteilungen und die zugehörigen Gütemaße.

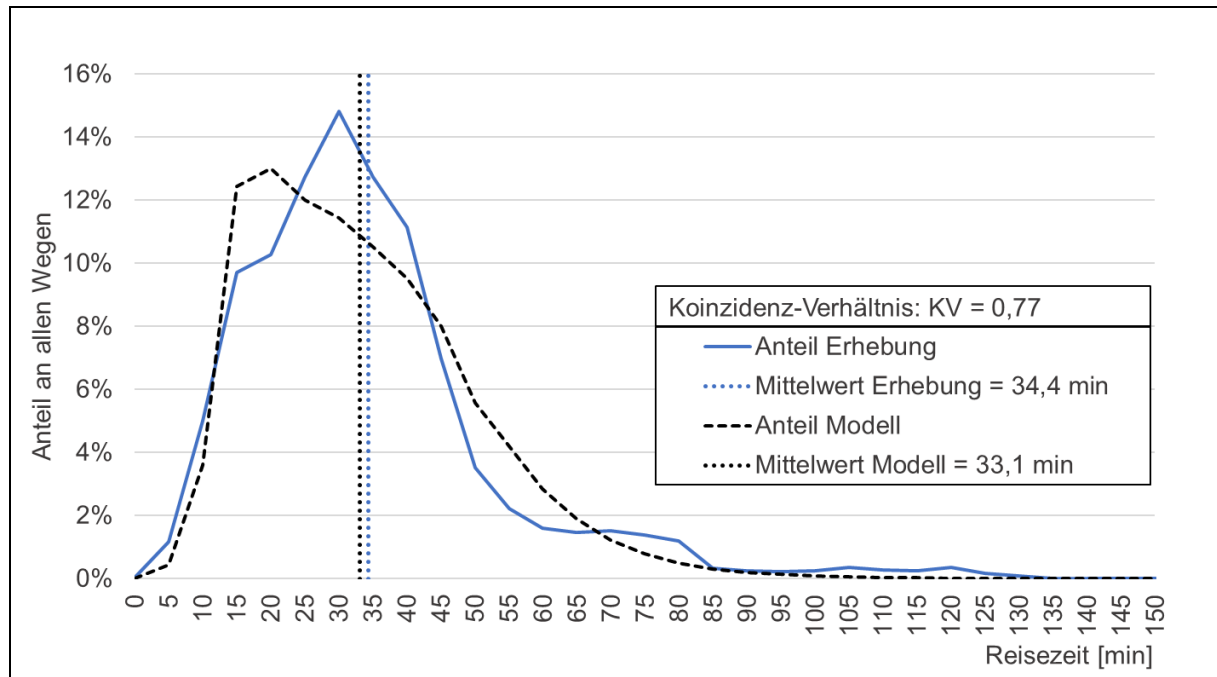


Bild 5: Vergleich einer gemessenen und einer modellierten Reisezeitverteilung.

4 Verifizierung der Modellspezifikation und der Modellimplementierung

Versteht man Verifizieren als das Beglaubigen oder Bestätigen eines Sachverhalts, dann können im Rahmen der Modellerstellung die Prüfung der Modellspezifikation und der Modellimplementierung als Verifizierung bezeichnet werden. Bei der **Verifizierung der Modellspezifikationen** prüft eine unabhängige dritte Person, ob die in der Modellspezifikation beschriebene Vorgehensweise angemessen ist, um die Anforderungen der Modellnutzer zu erfüllen. Dabei ist besonders auf folgende Aspekte zu achten:

- Sind die Modellanforderungen und die benötigten Modellergebnisse ausreichend beschrieben?
- Ist der Modellraum ausreichend groß und die räumliche Segmentierung (Zellengröße, abgebildetes Verkehrsnetz) angemessen?
- Ist die Vorgehensweise zum Aufbau des Netzmodells angemessen?
- Ist die Segmentierung der Verkehrsnachfrage (Personengruppen, Wegezwecke) und der Verkehrsmittel bzw. Verkehrsmodi angemessen?
- Ist die zeitliche Segmentierung (Tagesmodell, Hauptverkehrszeiten, Stundenmodell) angemessen?
- Wird der Wirtschaftsverkehr auf geeignete Weise abgebildet?
- Sind die zugrundeliegende Modelltheorie (z.B. Aktivitätenkettenmodell oder Einzelwegmodell, synthetisches oder inkrementelles Modell) und die Modellannahmen benannt?
- Sind die Entscheidungsprozesse benannt, die mit dem Modell abgebildet werden können (Aktivitätenwahl, Zielwahl, Moduswahl, Abfahrtszeitwahl, Routenwahl)?
- Welche Rückkopplungen (z.B. zwischen Nachfrage und Angebot) sind vorgesehen?

- Sind die Kenngrößen der Angebotsqualität beschrieben, die in den Entscheidungsmodellen berücksichtigt werden? Können damit alle geforderten Maßnahmen abgebildet werden?
- Ist die Vorgehensweise zur Kalibrierung der Modellparameter beschrieben und sinnvoll?
- Sind Maßnahmen und Tests zur Qualitätssicherung beschrieben?
- Sind die Datenquellen für Verkehrserhebungsdaten, Raumstrukturdaten, Angebotsdaten und für externe Verkehre angegeben?
- Ist die technische Umsetzung beschrieben, d.h. mit welcher Software die Daten verwaltet und die Berechnungen ausgeführt werden? Sind bei der Berechnung manuelle Eingriffe erforderlich?

Die **Verifizierung der Modellimplementierung** prüft, inwieweit die in der Modellspezifikation beschriebenen Daten und Methoden tatsächlich umgesetzt wurden. Dazu begibt sich die prüfende Person in die Rolle des Modellanwenders, installiert das Verkehrsnachfragemodell auf einem neuen Rechner und führt eigene, anwendungsorientierte Testrechnungen durch, die sich an den Realitätstests orientieren.

5 Überprüfung der Eingangsdaten

Die Qualität eines Verkehrsnachfragemodells wird maßgeblich von der Qualität der Eingangsdaten bestimmt. Wesentliche Eingangsdaten sind Verkehrserhebungsdaten, Raumstrukturdaten und Verkehrsangebotsdaten. Diese Daten sollten beim Aufbau des Verkehrsnachfragemodells und vor der Kalibrierung der Modellparameter auf ihre Qualität hin untersucht werden.

5.1 Verkehrserhebungsdaten

Verkehrserhebungsdaten umfassen personenbezogene Daten zum Mobilitätsverhalten aus Befragungen und netzbezogene Daten zu Verkehrsstärken an Zählstellen und Fahrzeiten zwischen ausgewählten Punkten im Netz. Mobilitätsverhaltensdaten beinhalten u.a. die Kenngrößen Mobilitätsrate, Reiseweite und Reisezeit differenziert nach Wegezweck und Verkehrsmodus, die auf folgende Weise überprüft werden sollten:

- Einzelwertprüfungen mit erwarteten Werten aus anderen Verkehrsbefragungen (z.B. MiD [9]).
- Angabe der Konfidenzintervalle für die Mobilitätskenngrößen für ein Konfidenzniveau von 95%.

Die Daten sollten für die spätere Validierung so aufbereitet werden, dass die Mobilitätskenngrößen für alle Wege und für die Teilmenge der Wege im Modellraum bekannt sind.

Bei Zählwerten sollte jeder Zählwert mit einem Gütemaß bewertet werden. Zählwerte aus einer Dauerzählstelle haben eine höhere Aussagekraft als Einmalzählungen, neuere Daten haben eine höhere Aussagekraft als ältere Daten. Für Dauerzählstellen sollte das Konfidenzintervall der Verkehrsstärke für ein Konfidenzniveau von 95% bestimmt werden.

5.2 Raumstrukturdaten

Raumstrukturdaten umfassen die Siedlungsflächen, die Einwohnerdaten differenziert nach Altersklassen und die Standortdaten von Aktivitätenorten mit zugehörigen Eigenschaften (Arbeitsstandorte/Arbeitsplätze, Schulstandorte/Schulplätze, Einkaufsstandorte/Größenklasse). Raumstrukturdaten werden auf Verkehrszellen bezogen, die als Quellen und Ziele von Ortsveränderung die Raumstruktur, das Verkehrsangebot und die Verkehrsnachfrage miteinander verknüpfen. Folgende Kenngrößen der Raumstrukturdaten sollten ausgewiesen und überprüft werden:

- Siedlungsdichte (Einwohner bezogen auf die Siedlungsfläche) auf Zellenebene.
- Arbeitsplatzdichte (Arbeitsplatzdichte bezogen auf die Siedlungsfläche) auf Zellenebene.
- Verhältnis Arbeitsplätze pro Einwohner und Erwerbstätige pro Arbeitsplatz auf Gemeindeebene.
- Verhältnis Schüler pro Schulplatz auf Gemeindeebene.
- Vergleich der aggregierten Strukturdaten (Einwohner, Beschäftigte am Arbeitsplatz) auf Gemeindeebene mit Daten der statistischen Ämter.

Liegen Standortdaten als Einzelpunkte (POI) vor, sollte die Qualität der Daten stichprobenhaft überprüft werden. Außerdem ist eine visuelle Prüfung mit Hilfe von Dichtekarten empfehlenswert.

Für eine standardisierte Überprüfung wäre ein Leitfaden mit typischen Wertebereichen wünschenswert, so wie es beispielsweise das Dokument FSUTMS-Cube Framework Phase I und II [10; 11] für die USA bietet.

5.3 Verkehrsangebotsdaten

Verkehrsangebotsdaten beinhalten sämtliche Daten zur Infrastruktur des Modells (Strecken, Knoten, Haltestellen, etc.) und alle Daten, die für den Betrieb der Infrastruktur notwendig sind (Fahrpläne, Preissysteme, etc.). Verkehrsangebotsdaten lassen sich auf der Ebene der einzelnen Netzelemente und auf der Relationsebene überprüfen. Die Überprüfung der Angebotsdaten erfolgt zu einem Zeitpunkt, zu dem keine oder nur abgeschätzte Nachfragematrizen und Verkehrsstärken im Netz vorliegen. Deshalb sind nur Aussagen zu Fahrzeiten im unbelasteten Netz oder zu abgeschätzten Fahrzeiten möglich.

Auf der Ebene von Netzelementen können Netztopologie und Netzattribute überprüft werden:

- Anzahl der Arme pro Knotenpunkt (z.B. maximal 5 Zufahrten pro Knotenpunkt).
- Vergleich der Netzlängen nach Straßenklasse und Vergleich mit Werten aus der amtlichen Statistik.
- Vergleich der mittleren Geschwindigkeit je Straßenklasse und Ortslage mit Erwartungswerten.
- Anzahl der Haltestellen je Verkehrszelle (z.B. maximal eine Schienenhaltestelle je Zelle).

Auch für Angebotsdaten bietet sich eine visuelle Prüfung mittels Kartendarstellung an. Das Straßennetz kann z.B. differenziert nach Geschwindigkeitsniveaus oder Fahrstreifenanzahlen dargestellt werden. Für das Netz im öffentlichen Verkehr kann eine streckenfeine Darstellung der angebotenen Fahrtenanzahl je Tag sinnvoll sein.

Sobald das Netzmodell erstellt ist, können Kenngrößenmatrizen berechnet und Testumlegungen durchgeführt werden. Dabei sollten folgende Werte überprüft werden:

- Vergleich der Kenngrößen (Reisezeiten, ggf. Fahrpreise) und der Fahrtrouten für ausgewählte Relationen mit Daten aus unabhängigen Quellen, z.B. einem Routenplanungsprogramm oder Floating Car Daten.
- Berechnung und Überprüfung relationsbezogener Kenngrößen:
 - Ist das Netz zusammenhängend, d.h. kann für jede Relation die Reisezeit berechnet werden?
 - Luftliniengeschwindigkeit: Werte, die nach den Richtlinien für die integrierte Netzgestaltung (RIN, 2008 [12]) als sehr gut oder sehr schlecht eingestuft werden, sollten überprüft werden,
 - Umwegfaktor: Werte über 1,5 sollten überprüft werden,
 - Reisezeitverhältnis ÖV/Pkw: dieser Wert sollte üblicherweise zwischen 0,5 (ÖV ist doppelt so schnell wie Pkw) und 5 liegen (ÖV ist 5-mal langsamer als der Pkw).
 - Reisezeitverhältnis Hin- und Gegenrichtung: In der Normalverkehrszeit sollten beide Richtungen ähnliche Werte aufweisen.
 - Anteil der Wartezeiten an Knotenpunkten bezogen auf die gesamte Fahrtzeit.
 - Anteil der Anbindungszeiten (Zu- und Abgang) an der gesamten Reisezeit.
- Durchführung und Analysen von Testumlegungen z.B. mit einer 1er Matrix
 - Die Verkehrsstärke im Netz sollte bei einer Bestwegumlegung in beiden Richtungen ähnlich sein.
 - Die Verkehrsstärken sollten in einem unabhängigen Netzmodell (z.B. Open Street Map) im Hauptstraßennetz ähnlich sein.

6 Kalibrierung und Validierung

Im Mittelpunkt einer Modellerstellung stehen die Kalibrierung und die Validierung, bei der in mehreren Durchläufen die Modellparameter gesetzt und ihre Wirkungen auf die Modellergebnisse geprüft werden:

- Kalibrierung:
Anpassen der Modellparameter, so dass die mit dem Modell berechneten Werte möglichst gut mit der beobachteten Realität übereinstimmen.
- Validierung:
Überprüfung der Modellparameter, der Modellergebnisse und des Modelverhaltens mit geeigneten Gütemaßen und Tests.

6.1 Einstellen der Parameter

Ausgangspunkt für den Prozess der Kalibrierung und Validierung ist eine Startlösung, bei der alle Modellparameter mit geschätzten oder gesetzten Parametern belegt werden.

- **Parameterschätzung:**

Bei der Parameterschätzung werden die Parameter entweder mit Hilfe eines Optimierungsverfahrens bestimmt oder iterativ solange verändert, bis die Abweichungen zwischen den aus dem Modell errechneten Werten und den gemessenen Werten minimal sind. Für die Parameterschätzung von Entscheidungsmodellen eignet sich die sogenannte Maximum-Likelihood Methode. Die Methode nutzt beobachtete Wahlentscheidungen von Verkehrsteilnehmern und die zugehörigen Eigenschaften der Menge an Alternativen. Wichtige Eigenschaften einer Alternative sind die Kenngrößen der Angebotsqualität: Zu- und Abgangszeit, Wartezeit, Fahrzeit im Verkehrsmittel, Kosten und Umsteigehäufigkeit. Diese Kenngrößen sollten nicht aus den Angaben der Probanden abgeleitet werden, sondern für alle Probanden und alle Alternativen aus den Kenngrößen, die mit dem Verkehrsnachfragemodell ermittelt wurden. Grundlage für die Kenngrößen sollten dabei nicht die Verkehrszellen sein, in denen die Probanden ihre Aktivitäten durchführen, sondern die genauen geocodierten Aktivitätenorte. Nur so kann der Einfluss der Zu- und Abgangszeit – diese Werte können für verschiedene Orte innerhalb einer Zelle deutlich variieren - bei der Parameterschätzung berücksichtigt werden. Die Güte einer Maximum-Likelihood-Schätzung sollte mittels des Likelihood-Ratio-Test und des rho-square-Test (siehe z.B. Huelsenbeck & Crandall [13]) dokumentiert werden.

- **Parametersetzung:**

Parameter, die nicht empirisch erhoben werden können, müssen durch den Modellierer festgesetzt werden. Sinnvolle Wertebereiche der Parameter können der Literatur (z.B. Travel Model Validation and Reasonableness Checking Manual [2], FSUTMS-Cube Framework Phase II [10], BVWP Axhausen et al. [14], Nationales Personenverkehrsmodell des UVEK [15]) oder anderen Modellen entnommen werden.

6.2 Anpassen der Modellparameter, Eingangsdaten und Modellspezifikation

Mit den initialen Parametern wird ein erster Validierungsprozess gestartet. Aufbauend auf den Prüfungsergebnissen beginnt der eigentliche Kalibrierungsprozess, bei dem die Modellparameter angepasst, ggf. die Eingangsdaten korrigiert und die Modellspezifikation ergänzt wird. Anhand der beobachteten Abweichungen der Modellergebnisse müssen mögliche Ursachen identifiziert und Anpassungen im Modell vorgenommen werden. Dabei ist zu beachten, dass eine Anpassung der Modellparameter der jeweiligen Modellstufe die Ergebnisse der nachfolgenden Stufe beeinflusst. Entsprechend ist es sinnvoll, mit der Kalibrierung der Verkehrserzeugung zu beginnen, daraufhin die Zielwahl- und Moduswahl zu kalibrieren und mit der Umlegung abzuschließen.

Wie in Bild 1 dargestellt, kann die Überprüfung der Modellergebnisse im Rahmen der Validierung wiederholt Rückkopplungen zu vorangegangenen Stufen erforderlich machen, z.B.

- Überarbeitung Modellspezifikation: z.B. Einführen einer zusätzlichen Personengruppe.
- Korrektur Verkehrsangebotsdaten: z.B. Änderungen der Anbindungsknoten oder der Streckenkapazitäten.
- Parameterschätzung: erneute Schätzung der Parameter aufgrund veränderter Reisezeiten.

Auch die Überprüfung des Modellverhaltens, die i.d.R. erst nach Erreichen eines gewissen Gütemaßes der Modellergebnisse erfolgt, kann zu den beschriebenen Rückkopplungen führen.

6.3 Überprüfung der Parameter

Als Ergebnis einer Parameterschätzung für Entscheidungsmodelle erhält man die Parameterwerte für die in einer Nutzenfunktion verwendeten Kenngrößen. Diese Parameter geben die Bedeutung der jeweiligen Kenngröße auf die Wahlentscheidung an. Diese Werte sollten auf folgende Weise geprüft werden:

- Vorzeichen der Parameter: Die Parameterwerte zur Bewertung von Widerständen sollten in einer Nutzenfunktion ein negatives Vorzeichen haben.

- Verhältnis der Parameter: Die Parameterwerte für die Gehzeit, das Umsteigen oder die Kosten werden in Bezug gesetzt zum Parameterwert der Fahrzeit. Diese resultierenden Verhältniswerte geben den Value of Time für eine Zeiteinheit Gehzeit, einen Umsteigevorgang oder eine Geldeinheit an, z.B. BVWP Axhausen et al. [14]. Diese Vorgehensweise kann nur für Modelle mit linearen Nutzenfunktionen gewählt werden.
- Wertebereich der Parameter: Die Parameterwerte können mit Werten aus der Literatur und mit Werten aus anderen Modellen verglichen werden. Das Travel Model Validation and Reasonableness Checking Manual [2] enthält beispielsweise für die Verkehrsmittelwahl mit einem Logit-Modell Tabellen mit Wertebereichen für die Parameter Fahrzeit, Gehzeit und Kosten. Eine Übertragbarkeit auf europäische Modelle ist allerdings nur eingeschränkt möglich. Deshalb könnte ein Regelwerk hilfreich sein, dass Wertebereiche für Modelle im deutschsprachigen Raum ausweist. Aber auch dann muss berücksichtigt werden, dass der Wert eines Parameters von der Modellarchitektur, von der verwendeten Bewertungsfunktion (Exponentialfunktion, Potenzfunktion, Halbglockenkurven) und von der Gesamtanzahl der Parameter in der Nutzenfunktion abhängt.

6.4 Überprüfung der Modellergebnisse

In diesem Schritt werden mit dem Modell errechnete Werte und gemessene Werten mit geeigneten Gütemaßen (siehe Abschnitt 3) überprüft. Als Kontrolldaten dienen personenbezogene Daten zum Mobilitätsverhalten aus Befragungen und netzbezogene Daten zu Verkehrsstärken an Zählstellen und Fahrzeiten zwischen ausgewählten Punkten im Netz.

Für die Überprüfung der Modellstufen Verkehrserzeugung, Ziel- und Moduswahl werden Wegedaten von Einzelpersonen aggregiert und mit den berechneten Wegen verglichen:

- berechnete Zahl der Wege mit befragter Zahl der Wege,
- zellenfein berechnetes Quell- und Zielverkehrsaufkommen der Zielwahl mit Quell- und Zielverkehrsaufkommen aus Strukturgrößen
- berechnete Modal-Split Anteile der Verkehrsmoduswahl mit Modal-Split Anteilen aus Befragungen differenziert nach Entfernungsklassen,
- berechnete Reiseweiten- oder Reisezeitverteilung der Verkehrsverteilung mit Reiseweiten- oder Reisezeitverteilung, die aus Befragungen abgeleitet werden,

Die Wegedaten sollten für verschiedene Segmentierungen überprüft werden:

- räumlich: gesamter Modellraum und ggf. für Teilräume (Stadt, Umland)
- inhaltlich: alle Einwohner und differenziert nach Personengruppen und Wegezwecken

Für die Überprüfung der Verkehrsstärken und Reisezeiten werden folgende Werte verglichen:

- berechnete Quelle-Ziel Reisezeiten mit gemessenen Reisezeiten für ausgewählte Relationen,
- berechnete Verkehrsstärken der Umlegung mit gezählten Verkehrsstärken an ausgewählten Zählstellen oder an Screenlines. Eine Screenline umfasst eine Menge von Strecken, die durch eine Polygonlinie definiert werden (z.B. alle Brücken über einen Fluss). Alle Strecken, die durch die Polygonlinie geschnitten werden, gehören zur Screenline. Die Verkehrsstärken aller Strecken einer Screenline werden zu einer Screenline-Verkehrsstärke aufsummiert.

Tabelle 2 zeigt beispielhaft und nicht vollständig einen Vorschlag, welche Kenngrößen überprüft werden sollten, welche Differenzierung räumlich und inhaltlich wünschenswert ist und welches Gütemaß in Frage kommt. Neben Haushaltsbefragungen und Zählungen können Befragungen im ÖV (z.B. Linienbeförderungsfälle) und Reisezeitmessungen im Netz als Kontrolldaten herangezogen werden.

Bei Verkehrsnachfragemodellen, die im Rahmen der Modellerstellung für einen oder mehrere Modi eine Korrektur der Nachfragematrix durchführen, sollten die Kenngrößen vor und nach der Matrixkorrektur überprüft und die Änderungen dokumentiert werden. Nach der Matrixkorrektur sollte u.a. geprüft werden, ob die Konsistenz der erzeugten und angezogenen Wege eingehalten ist und ob sich die Reisezeitverteilung geändert hat.

	Kenngröße	Differenzierung		Kontroll daten	statistisches Gütemaß
		inhaltlich	räumlich		
Verkehrsaufkommen	Anzahl Wege	<ul style="list-style-type: none"> • pro Einwohner • pro Person einer PG x Zweck • Gesamtanzahl pro PG, pro Zweck, pro Modus 	<ul style="list-style-type: none"> • gesamter MR • Teilräume 	HH-Bef	<ul style="list-style-type: none"> • Mittelwert: größenabhängige rel. Abweichung • Konsistenz erzeugte und angezogene Wege
	wegebezogener Modal-Split	<ul style="list-style-type: none"> • Grundgesamtheit 	<ul style="list-style-type: none"> • gesamter MR • Teilräume 	HH-Bef	<ul style="list-style-type: none"> • Prozentwert je Modus: größenabhängige rel. Abweichung
		<ul style="list-style-type: none"> • pro PG • pro Zweck • pro PG x Zweck 	<ul style="list-style-type: none"> • gesamter MR 	HH-Bef	<ul style="list-style-type: none"> • Prozentwert je Modus: größenabhängige rel. Abweichung
		<ul style="list-style-type: none"> • pro Zweck 	<ul style="list-style-type: none"> • pro Entfernungs-klasse 	HH-Bef	<ul style="list-style-type: none"> • Prozentwert je Modus: größenabhängige rel. Abweichung
Verkehrsleistung	Verteilung der Reiseweite der Einzelwege	<ul style="list-style-type: none"> • Grundgesamtheit • pro Modus • pro Zweck 	<ul style="list-style-type: none"> • gesamter MR 	HH-Bef	<ul style="list-style-type: none"> • Mittelwert: größenabhängige rel. Abweichung • Koinzidenz-Verhältnis
	verkehrsleistungs-bezogener Modal-Split	<ul style="list-style-type: none"> • Grundgesamtheit 	<ul style="list-style-type: none"> • gesamter MR 	HH-Bef	<ul style="list-style-type: none"> • Prozentwert je Modus: größenabhängige rel. Abweichung
	Verkehrsleistung Personenkilometer [Pkm]	<ul style="list-style-type: none"> • pro Einwohner • pro Person einer PG x Zweck 	<ul style="list-style-type: none"> • gesamter MR 	HH-Bef	<ul style="list-style-type: none"> • Mittelwert: größenabhängige rel. Abweichung
	Verkehrsleistung Fahrzeugkilometer [Fgzk]	<ul style="list-style-type: none"> • pro Einwohner • pro Person einer PG x Zweck 	<ul style="list-style-type: none"> • gesamter MR 	HH-Bef	<ul style="list-style-type: none"> • Mittelwert: größenabhängige rel. Abweichung
Verkehrszeitaufwand	Verteilung der Reisezeit der Einzelwege	<ul style="list-style-type: none"> • Grundgesamtheit • pro Modus • pro Zweck 	<ul style="list-style-type: none"> • gesamter MR 	HH-Bef	<ul style="list-style-type: none"> • Mittelwert: größenabhängige rel. Abweichung • Koinzidenz-Verhältnis
	Verkehrszeit-aufwand Personenstunden [Ph]	<ul style="list-style-type: none"> • pro Einwohner • pro Person einer PG x Zweck 	<ul style="list-style-type: none"> • gesamter MR 	HH-Bef	<ul style="list-style-type: none"> • Mittelwert: größenabhängige rel. Abweichung
Verkehrsstärke	Verkehrsstärke	<ul style="list-style-type: none"> • Pkw • Lkw • Rad • Bus • Bahn 	<ul style="list-style-type: none"> • Strecke • Abbieger • Strecke einer Linienroute • Screenline 	Zählung	<ul style="list-style-type: none"> • Verkehrsstärke Einzelwert: größenabhängige rel. Abweichung • %RMSE abhängig von der Verkehrsstärken-lassen
HH-Bef Haushaltsbefragung mit Wegetagebüchern. Die Datensätze müssen gewichtet und vollständig dokumentiert vorliegen. Es werden nur solche Wege als Kontrolldaten berücksichtigt, die im Modelraum beginnen und enden. MR Modellraum, in dem die Verkehrsnachfrage modelliert wird. PG Personengruppe					

Tabelle 2: Überprüfung von Kenngrößen der Nachfrage bei der Modellvalidierung

6.5 Überprüfung des Modellverhaltens

Sensitivitätstests - Überprüfung des Modells mit modifizierten Parametern

Mit Sensitivitätstests wird der Einfluss der Modellparameter auf das Ergebnis geprüft. Dabei werden die Modellvariablen – in der Regel die Modellvariablen des Analysezustands – unverändert gelassen, so dass Ergebnisänderungen auf einzelne Parameter zurückgeführt werden können. Eine Variation der Parameterwerte sollte besonders für folgende Parameter angestrebt werden:

- Empirisch nicht abgesicherte Parameter der Nutzenfunktion: Parameterwerte von Kenngrößen, die nicht oder nur eingeschränkt geschätzt werden können, sollten einmal reduziert und einmal erhöht werden. WebTAG [16] schlägt eine Reduktion bzw. Erhöhung um 25% bis 50% vor. Das gilt häufig für Parameter der Zahlungsbereitschaft und andere Parameter, die „nur“ mit einer Stated-Choice Befragung ermittelt werden konnten. Die Ergebnisse dieser Variationen können als Elastizitäten der untersuchten Parameter ausgedrückt werden, z.B. die Erhöhung der Parameters Zeitwert um x% reduziert die Pkw-Fahrleistung um y%.
- Konstanten der Nutzenfunktion: Nutzenfunktionen enthalten in der Regel eine sogenannte alternativenspezifische Konstante, die den Nutzen aller nicht abgebildeten, unbekanntem Einflussgrößen beschreibt. Um die Bedeutung dieser Einflussgrößen zu ermitteln, sollten in einem Sensitivitätstest alle alternativenspezifischen Konstanten auf den Wert 0 gesetzt werden. Analog sollte auch einen Sensitivitätstest durchführen werden, bei dem mit Ausnahme der Konstanten alle Parameter auf 0 gesetzt werden. Dieser Prozess hat zur Folge, dass sehr gut sichtbar wird, welche Modellergebnisse nur durch die alternativenspezifischen Konstanten und welche Ergebnisse nur durch die im Modell genutzten Aufwände inkl. ihrer Parameter innerhalb der Nutzenfunktion generiert werden. Ziel muss sein, dass der Einfluss der alternativenspezifischen Konstanten minimiert wird und somit das Modell mehr Aussagekraft, im Besonderen für die Prognose, erhält.
- Abbruchbedingungen für Gleichgewichtsberechnungen: Strengere Abbruchbedingungen bei einer Gleichgewichtsumlegung (siehe Kenngröße Duality Gap in Bild 6) oder bei einer Rückkopplung zwischen Nachfrageberechnung und Umlegung erhöhen die Modellkonvergenz auf Kosten der Rechenzeit. Um Aussagen über den Einfluss der Abbruchbedingungen auf das Ergebnis machen zu können, sollten die geplanten Abbruchbedingungen um den Faktor 10 verschärft werden (z.B. maximaler zulässiger Gap wird von 0,0001 auf 0,00001 reduziert).

Im Prognosefall sollten die Modellparameter nicht verändert werden. Daher ist es ausreichend, die Sensitivitätstests im Analysezustand durchzuführen.

$$G = \frac{V_{Ist} - V_{Hypo}}{V_{Hypo}}$$

mit

V_{Ist} tatsächlicher Verkehrszeitaufwand (Fahrzeugstunden), der sich aus dem Produkt der aktuellen Fahrzeiten und Streckenbelastungen ergibt

V_{Hypo} hypothetischer Verkehrszeitaufwand (Fahrzeugstunden), der sich ergibt, wenn alle Fahrzeuge die Fahrzeit auf der aktuell zeitkürzesten Route nutzen könnten

Bild 6: Duality Gap als Konvergenzkenngroße bei der Gleichgewichtsumlegung

Realitätstests - Überprüfung des Modells mit modifizierten Variablen

Mit Realitätstests wird geprüft, ob die Wirkungen eines Modells bei Änderungen der Variablen in der erwarteten Größenordnung liegen. Dazu werden die Variablen Verkehrsangebot oder die Strukturdaten in kontrollierter Weise verändert (z.B. Erhöhung der Fahrzeit oder der Preise oder der Einwohnerzahl um 10 %). Dann werden die Nachfrageänderungen ermittelt. Aus dem Vergleich der Angebotsänderungen und der Nachfrageänderungen (Personenkilometer) lassen sich Elastizitäten der Zeit oder des Preises berechnen und mit erwarteten Elastizitäten vergleichen. Dabei ist zu berücksichtigen, dass sich Elastizitäten in Abhängigkeit der Aufwandskenngroße ändern können, was insbesondere in großräumigen Modellen auftritt.

Außerdem sollte ein Realitätstest durchgeführt werden, bei dem die Wirkung der Kapazitätsänderung eines Netzelements (z.B. Neubau einer Straße) einmal ohne Rückkopplung (nur Umlegung) und einmal mit Rückkopplung (Nachfrageberechnung und Umlegung) untersucht wird. Auf welchen Relationen

erhöhen bzw. reduzieren sich die Reisezeiten. Sind Änderungen der Verkehrsstärken – differenziert nach Pkw und Lkw – auch in größerer Entfernung vom untersuchten Netzelement erklärbar oder eher auf eine unzureichende Konvergenz zurückzuführen?

6.6 Abschließende Validierung der Modellergebnisse

Nach dem Abschluss des Kalibrierungs- und Validierungsprozesses liegt ein geprüftes Modell vor. Um die Qualität dieses Modells nachzuweisen, können die Modellergebnisse nun mit einem bis zu diesem Zeitpunkt unbenutzten Datensatz überprüft werden. Bei der Überprüfung mikroskopischer Verkehrsflussmodelle [3] lässt sich ein Datensatz für einen weiteren Kontrollzustand zum Beispiel durch Reisezeitmessungen zu einer anderen Tageszeit relativ leicht erzeugen. Bei makroskopischen Modellen, die meist die mittlere Verkehrsnachfrage an einem Werktag außerhalb der Ferien abbilden, ist es schwer, einen weiteren, unbenutzten Datensatz zu erhalten. Im einfachsten Fall kann man Verkehrsstärken anderer Zählstellen heranziehen. Eine andere, aufwändigere Vorgehensweise besteht in einem Backcasting, bei dem die Modellvariablen in einen definierten Zustand der Vergangenheit gesetzt werden und die Modellergebnisse mit Verkehrsstärken oder Mobilitätskenngrößen zugehöriger Verkehrserhebungen verglichen werden.

Die Durchführung einer abschließenden Validierung ist sicher wünschenswert, es stellt sich aber unweigerlich die Frage nach der Konsequenz, wenn der abschließende Validierungstest negativ ausfällt. Ein erneutes Kalibrieren verbietet sich, da dies erstens nicht Zweck dieses Tests ist und zweitens der Datensatz in diesem Fall von Anfang an zu Kalibrierungszwecken hätte verwendet werden können. Ein simples Ignorieren des Testergebnisses rechtfertigt seine Anwendung ebenso wenig. Konsequenterweise müsste bei einem negativen Ausgang das komplette Verkehrsnachfragemodell verworfen werden, um seine Modellspezifikationen grundlegend neu zu definieren. In der Praxis gibt es diese Fälle durchaus. Sie treten immer dann ein, wenn ein Modellanwender in seiner Rolle als Auftraggeber das von einem Modellersteller angefertigte Modell nicht abnimmt. Meist sind sie die Folge einer ungenauen Modellspezifikation und einer nachlässigen Modellüberprüfung.

7 Fazit

Verkehrsnachfragemodelle sind Grundlage für weitreichende verkehrspolitische und verkehrsplanerische Entscheidungen. Aus diesem Grund sollte ein Qualitätssicherungsprozess für Verkehrsnachfragemodelle von den Modellanwendern gefordert und von den Modellerstellern während des Erstellungsprozesses durchgeführt werden. In deutschsprachigen Ländern gibt es Erfahrung mit der Erstellung und Anwendung von Verkehrsnachfragemodellen. Diese Erfahrung ist bisher jedoch kaum formal dokumentiert und hat nicht zu Richtlinien für die Nachfragemodellierung geführt. Deshalb sollte es ein Ziel sein, Empfehlungen für die Erstellung und Überprüfung von Verkehrsnachfragemodellen zu entwickeln, die idealerweise in den deutschsprachigen Ländern abgestimmt sind, sich aber auch an internationalen Erkenntnissen orientieren.

Das in diesem Artikel beschriebene Vorgehen zeigt, wie ein Qualitätssicherungsprozess aussehen könnte. Es wird aber deutlich, dass eine standardisierte Überprüfung die Definition geeigneter Kenngrößen und Benchmarks erfordert. Dieser Schritt steht bisher noch aus.

Modellanwender und die Nutzer der Modellergebnisse, die ein geprüftes Modell für die Zwecke einer Verkehrsprognose einsetzen, können folgende Ergebnisqualität erwarten:

- Die Prognosewerte der Verkehrsstärken können in etwa mit der Genauigkeit berechnet werden, die bei der Validierung von Einzelwerten als Grenzwerte definiert wurden.
- Die Prognosewerte basieren auf der Annahme, dass die Inputdaten (Bevölkerungsprognose, Entwicklung der Siedlungsstruktur, zukünftiges Verkehrsangebot, zukünftige Preise) weitgehend richtig sind. Außerdem kann das Modell nur Wirkungen von Entwicklungen und Maßnahmen prognostizieren, die dem vorher definierten Einsatzbereich entsprechen.
- Entscheidungen bei denen eine hohe Prognosesicherheit gefordert ist, sollten mit Sicherheitszuschlägen arbeiten. Für verkehrstechnische Leistungsfähigkeitsuntersuchungen sollte eher von höheren Verkehrsstärken ausgegangen werden, bei der Berechnung von Erlösen (z.B. aus Fahrkarten) oder bei volkswirtschaftlichen Untersuchungen sollte eher von einer geringeren Nachfrage ausgegangen werden. Welchen Sicherheitszuschlag man annimmt, ist eine politische und bei Kostenrechnungen eine unternehmerische Entscheidung.

- Ein Verkehrsnachfragemodell ist ein kontrolliertes Experiment in einer virtuellen Welt, in der viele Einflussfaktoren als gegeben angenommen werden. Die Ergebnisse müssen deshalb interpretiert und die Annahmen mit den Ergebnissen benannt werden.

8 Literaturverzeichnis

- [1] *Department for Transport* (Hrsg.): TAG UNIT M3.1. Highway Assignment Modelling. Transport analysis guidance: WebTAG. M3.1 2014
- [2] *Cambridge Systematics, I.*: Travel Model Validation and Reasonableness Checking Manual. Second Edition. 2nd ed. Hrsg.: Federal Highway Administration, Cambridge, Massachusetts 2014. <http://www.camsys.com/pubs/FHWA-HEP-10-042.pdf>, abgerufen am 04. November 2015
- [3] Hinweise zur mikroskopischen Verkehrsflusssimulation. Grundlagen und Anwendungen. FGSV. 388. Ausg. 2006. FGSV-Verl., Köln 2006
- [4] *Herkt, S.; Leerkamp, B.; Althoff, T.; Goebel, D.; Janßen, T.; Meißner, A.*: QUALIMOD. Qualitätsanforderungen und -standards für Verkehrsmodellrechnungen 2008, abgerufen am 01. April 2016
- [5] *Sammer, G.; Röschel, G.; Gruber, C.*: Qualitätssicherung für die Anwendung von Verkehrsnachfragemodellen und Verkehrsprognosen. Entwurf eines Merkblattes Projekt QUALIVERMO, Forschungsbericht. Hrsg.: Autobahnen- und Schnellstraßen-Finanzierungs-Aktiengesellschaft; Bundesministerium für Verkehr, Innovation und Technologie, Graz
- [6] *Friedrich, M.; Pestel, E.; Schimpf, M.; Schiller, C.*: Einflussgrößen auf die Qualität von makroskopischen Nachfragemodellen im Personenverkehr. DFG-Projekt. <http://www.isv.uni-stuttgart.de/vuv/projekte/QTDM/index.html>
- [7] *Ernst Basler + Partner*. Qualitätssicherung von Verkehrsmodellberechnungen. SVI-Projekt 2015/001. <http://www.transport-research.info/project/quality-assurance-transport-model-calculations>
- [8] *Backhaus, K.; Erichson, B.; Plinke, W.; Weiber, R.*: Multivariate Analysemethoden. Eine anwendungsorientierte Einführung ; [Extras im Web]. Springer-Lehrbuch. 13., überarbeitete Auflage. Springer, Berlin, Dordrecht, London, New York 2011
- [9] *infas, D. L.*: Mobilität in Deutschland 2008. Ergebnisbericht, Bonn und Berlin 2010. www.mobilitaet-in-deutschland.de, abgerufen am 27. Juli 2016
- [10] FSUTMS-Cube Framework Phase II. Model Calibration and Validation Standards, final report. Hrsg.: Cambridge Systematics, I., Tallahassee, Florida 2008. http://www.fsutmsonline.net/images/uploads/reports/FR2_FDOT_Model_CalVal_Standards_Final_Report_10.2.08.pdf, abgerufen am 04. November 2015
- [11] FSUTMS-Cube Framework Phase I. Default Model Parameters, final report. Hrsg.: Cambridge Systematics, I., Tallahassee, Florida 2006. http://www.fsutmsonline.net/images/uploads/reports/FR1_FSUTMS-Cube_Parameters_All-in-One_Version_10-31-06%5B1%5D.pdf, abgerufen am 04. November 2015
- [12] Richtlinien für integrierte Netzgestaltung. RIN. FGSV. 121. FGSV Verlag, Köln 2008
- [13] *Huelsenbeck, J. P.; Crandall, K. A.*: Phylogeny Estimation and Hypothesis Testing Using Maximum Likelihood. Annual Review of Ecology and Systematics 28 (1997), S. 437–466
- [14] *Axhausen, K. W.; Ehreke, I.; Glemser, A.; Hess, S.; Jödden, C.; Nagel, K.; Sauer, A.; Weis, C.*: Ermittlung von Bewertungsansätzen für Reisezeiten und Zuverlässigkeit auf der Basis eines Modells für modale Verlagerungen im nicht-gewerblichen und gewerblichen Personenverkehr für die Bundesverkehrswegeplanung. Entwurf Schlussbericht, FE-Projekt-Nr. 96.996/2011 2014. https://www.bmvi.de/SharedDocs/DE/Anlage/VerkehrUndMobilitaet/BVWP/bvwp-2015-zeitkosten-pv.pdf?__blob=publicationFile, abgerufen am 25. Juli 2016
- [15] Nationales Personenverkehrsmodell des UVEK. Aktualisierung auf den Basiszustand 2010, Endbericht. Hrsg.: Bundesamt für Raumentwicklung (ARE); Eidg. Departement für Umwelt, Verkehr, Energie und Kommunikation (UVEK) 2014. http://www.are.admin.ch/themen/verkehr/00256/00513/index.html?lang=de&download=NHZLpZeg7t,lnp6l0NTU042l2Z6ln1acy4Zn4Z2qZpnO2YUq2Z6gpJCEenx9gmym162epYbg2c_JjKbNoKSn6A--, abgerufen am 04. August 2016
- [16] *Department for Transport* (Hrsg.): TAG UNIT M2. Variable Demand Modelling. Transport analysis guidance: WebTAG. M2 2014