

86. Jahrgang – Heft 2 – 2015

ZEITSCHRIFT FÜR VERKEHRSWISSENSCHAFT

INHALT DES HEFTES:

- Implementierung der Pkw-Telematik in die Kfz-Versicherungstarifizierung – Seite 91
Ein Analyse-Ansatz für Fahrprofile
Ein Überblick
Von Wiltrud Weidner, Hannover, Robert Weidner, Hamburg,
Fabian W.G. Transchel, Hannover
- Leistungsmaßzahlen für die Straßenerneuerung Seite 122
Von Craig Richmond, Clemens Kielhauser, Bryan T. Adey, Zürich
- Delivery Time Coordination at Customer Sites with Implicit Time Windows – Seite 155
A Computational Evaluation
Von Jörn Schönberger, Dresden

Manuskripte sind zu senden an die Herausgeber:
Prof. Dr. Bernhard Wieland
Institut für Wirtschaft und Verkehr,
Fakultät Verkehrswissenschaften an der Technischen Universität Dresden
01062 Dresden
Prof. Dr. Thorsten Beckers
Fachgebiet Wirtschafts- und Infrastrukturpolitik (WIP)
an der Technischen Universität Berlin
Straße des 17. Juni 135
10623 Berlin

Verlag – Herstellung – Vertrieb – Anzeigen:
Verkehrs-Verlag J. Fischer, Corneliusstraße 49, 40215 Düsseldorf
Telefon: (0211) 9 91 93-0, Telefax (0211) 6 80 15 44
www.verkehrsverlag-fischer.de
Einzelheft EUR 25,50 – Jahresabonnement EUR 71,00
zuzüglich MwSt und Versandkosten
Für Anzeigen gilt Preisliste Nr. 25 vom 1.1.2009
Erscheinungsweise: drei Hefte pro Jahr

Es ist ohne ausdrückliche Genehmigung des Verlages nicht gestattet, photographische Vervielfältigungen, Mikrofilme, Mikrophotos u.ä. von den Zeitschriftenheften, von einzelnen Beiträgen oder von Teilen daraus herzustellen.

Herausgeber

Prof. Dr. Bernhard Wieland (Technische Universität Dresden, federführender Herausgeber)
Prof. Dr. Thorsten Beckers (Technische Universität Berlin, federführender Herausgeber)
Prof. Dr. Herbert Baum (Universität zu Köln)
Prof. Dr. Karl-Hans Hartwig (Universität Münster)
Prof. Dr. Kay Mitusch (Karlsruher Institut für Technologie - KIT)
Prof. Dr. Kai Nagel (Technische Universität Berlin)

Schriftleitung

Prof. Dr. Bernhard Wieland (Technische Universität Dresden)
Prof. Dr. Thorsten Beckers (Technische Universität Berlin)
Dr. Christos Evangelinos (IUBH Internationale Hochschule Bad Honnef · Bonn)
Dr. Martin Winter (Technische Universität Berlin)

Herausgeberbeirat

Prof. Dr. Gerd Aberle (Universität Gießen)
Prof. Dr. Kay W. Axhausen (Eidgenössische Technische Hochschule - ETH, Zürich)
Prof. Dr. Johannes Bröcker (Universität zu Kiel)
Dr. Astrid Gühnemann (Institut for Transport Studies - ITS, Universität Leeds)
Dr. Hendrik Haßheider (Bundesministerium für Verkehr, Bau und Stadtentwicklung, Berlin)
Prof. Dr. Georg Hauger (Technische Universität Wien)
Prof. Dr. Christian von Hirschhausen (Technische Universität Berlin)
Prof. Dr. Christian Kirchner † (Humboldt-Universität zu Berlin)
Prof. Dr. Günter Knieps (Universität Freiburg)
Prof. Dr. Jürgen Kühling (Universität Regensburg)
Prof. Dr. Gernot Liedtke (Technische Universität Berlin/ DLR Berlin)
Dr. Heike Link (Deutsches Institut für Wirtschaftsforschung - DIW, Berlin)
Dr. Robert Malina (Universität Münster)
Prof. Dr. Hans-Martin Niemeier (Hochschule Bremen)
Prof. Dr. Werner Rothengatter (Karlsruher Institut für Technologie - KIT)
Prof. Dr. Bernhard Schlag (Technische Universität Dresden)

© Verkehrs-Verlag J. Fischer, Corneliusstraße 49, 40215 Düsseldorf
Telefon: (0211) 9 91 93-0, Telefax (0211) 6 80 15 44
www.verkehrsverlag-fischer.de
Hinweise für Autoren: www.z-f-v.de
ISSN: 0044-3670

Implementierung der Pkw-Telematik in die Kfz-Versicherungstarifizierung – Ein Analyse-Ansatz für Fahrprofile

VON WILTRUD WEIDNER, HANNOVER, ROBERT WEIDNER, HAMBURG,
FABIAN W.G. TRANSCHEL, HANNOVER

1. Einleitung

Die Konsequenzen des technischen Fortschritts wie Assistenzsysteme und die technische Voraussetzung für das autonome Fahren werden in den nächsten Jahren starken Einfluss auf die Produktlandschaft und das Geschäftsmodell der Kraftfahrzeugversicherung haben. Neben dem zunehmenden, deregulierungsbedingten Preiswettbewerb im beitragsmäßig größten Versicherungszweig der deutschen Schaden- und Unfallversicherung werden neue Technologien die Pricingstrategien und Produktentwicklung maßgeblich bestimmen. Im Zusammenhang mit einem risikoadäquaten Pricing sowie einer differenzierten Produktgestaltung stehen bereits die technischen Entwicklungen im Bereich der Telematik im Fokus. Die Pkw-Telematik basiert üblicherweise auf einer mit einem GPS-Modul kombinierten Blackbox zur Aufzeichnung spezifischer Informationen zur Fahrzeugnutzung und zum individuellen Fahrverhalten. Die auf Basis der Telematik exakt erstellten, individuellen Risikoprofile bilden die Grundlage für neue Versicherungsmodelle in der Kfz-Versicherung wie die Pay-How-You-Drive (PHYD)-Tarifizierung (vgl. z.B. Gerpott und Berg 2012; Fried-

Anschriften der Verfasser:

Wiltrud Weidner
Leibniz Universität Hannover
Institut für Versicherungsbe-
triebslehre
Otto-Brenner-Straße 1,
30159 Hannover

ww@ivbl.uni-hannover.de

Dr. Robert Weidner
Helmut-Schmidt-Universität/
Universität der Bundeswehr Hamburg
Institut für Konstruktions- und Ferti-
gungstechnik
Holstenhofweg 85,
22043 Hamburg

robert.weidner@hsu-hh.de

Fabian W.G. Transchel
Leibniz Universität Hannover
Institut für Theoretische Physik
Appelstraße 2,
30167 Hannover

fabian.transchel@itp.uni-hannover.de

richs 2013).¹ Sie erlauben es Versicherern, an bestimmten Kundensegmenten ausgerichtete Produkte zu gestalten, welche dann spezifisch bepreist werden können. Telematiktarife bzw. versicherungstelematische Produkte können in einem zunehmend gesättigten Markt, in dem Wachstum kaum noch möglich ist, als Kundenbindungsinstrumente eingesetzt werden.

Versicherungskonzepte aus dem Ausland zeigen, dass die Technik und ihre Anwendung bereits weit fortgeschritten sind (vgl. z.B. Ippisch et al. 2007; Gerpott und Berg 2012). In den USA und Großbritannien sind Pay-As-You-Drive (PAYD)-Modelle bei den Versicherern wie Progressive, State Farm, Insurethebox und Co-operative Insurance fest etabliert. Versicherer wie die S-Direkt und die VHV sammeln derzeit erste Erfahrungen auf dem deutschen Kfz-Markt.

Ausführungen zu telematischen Versicherungslösungen finden sich seit dem letzten Jahrzehnt vermehrt in der wissenschaftlichen Literatur, sind allerdings aufgrund der neuartigen Thematik immer noch selten und allgemein gehalten. Oberholzer (2003), Ippisch et al. (2007), Lee (2008), Gerpott und Berg (2012) und Friedrichs (2013) behandeln überwiegend die Grundzüge und Voraussetzungen der PHYD-Tarifierung sowie deren Auswirkungen auf das Geschäftsmodell eines Versicherers. Vonseiten der technischen Literatur stellt Coroma (2006) einen Prototyp zur individualisierten Messung von Kosten im Zusammenhang mit dem Verkehr vor. Jun et al. (2005) beschreiben Glättungsverfahren, die darauf abzielen, die Auswirkungen von GPS-Zufallsfehlern auf die Wegstrecke, Geschwindigkeit und geschätzten Beschleunigungsprofile zu begrenzen. In Weidner und Weidner (2014) wird der Aspekt der Gestaltbarkeit der Datenqualität vor dem Hintergrund von systematischen Fehlern aufgegriffen. Iqbal und Lim (2006) präsentieren einen Ansatz eines Board-Zahlungssystems, welcher die Privatsphäre des Fahrzeugführers über eine anonyme Zahlungsweise der Versicherungsprämie schützt. Des Weiteren führen die Autoren die Erzeugung individueller Fahrprofile basierend auf GPS-Daten beispielhaft vor (vgl. Iqbal und

¹ Die in der Literatur meist genannten Versicherungsvarianten sind die Pay-As-You-Drive (PAYD)-Tarifizierung und Pay-How-You-Drive (PHYD)-Tarifizierung. Erstere beruht auf der Betrachtung der Fahrzeugnutzung, zweite bezieht darüber hinaus das Fahrverhalten mit ein (vgl. z.B. Iqbal und Lim 2006; Lee 2008; Pletziger 2012). Der vorliegende Beitrag bezieht sich überwiegend auf das PHYD-Versicherungskonzept.

Lim, 2010). Empirische Studien zu konkreten Tarifierungsansätzen unter Einbindung von Telematikdaten sind indessen sehr begrenzt. Jun et al. (2007, 2011) sowie Kremslehner und Muermann (2013) weisen auf die Relevanz des Fahrverhaltens für die Vertragsgestaltung und das Risiko hin. In Toledo et al. (2008) werden Risikokennzahlen anhand riskanter Fahrmanöver ermittelt. Im Gegensatz dazu führen Paefgen et al. (2013) einen Klassifikationsansatz für aggregierte PAYD-Tarifierungsmerkmale ein. Ansätze zur Gestaltbarkeit der Auswertung von Fahrprofilen auf Basis der zur Verfügung stehenden begrenzten Daten, und insbesondere ohne Rückgriff auf eine in Verbindung mit Telematikdaten in der Praxis bisher nicht aufgebauten Schadenhistorie, wurden bislang überhaupt nicht thematisiert.

In dieser Arbeit werden Fahrprofile anhand des Fahrverhaltens analysiert, um im weiteren Forschungsverlauf risikobasierte Prämien für Telematik-Tarife ohne Einschluss einer Schadenhistorie in Bezug auf Telematikdaten herleiten zu können. Die Analyse basiert auf einer umfangreichen Datenbasis über 10.000 Fahrprofile, die mithilfe eines stochastischen Simulationsmodells erstellt wurde. Bei simultaner Betrachtung von Geschwindigkeitsabläufen und Beschleunigungsverhalten werden dabei sechs Fahrstile differenziert, die die Grundlage für eine allgemeine Risikokategorisierung bilden.

Der Rest des Beitrags ist wie folgt aufgebaut: In Kapitel 2 wird die gegenwärtige Situation der Kfz-Versicherung sowie ihrer Tarifizierung dargestellt. In Kapitel 3 wird die Methodik der zugrundeliegenden Generierung der Fahrprofile sowie deren Auswertung erläutert. Kapitel 4 veranschaulicht die Differenzierung von Fahrstilen in einer Simulationsstudie, wobei abschließend die Konsequenzen für die Tarifizierung der Kfz-Versicherung diskutiert werden.

2. Hintergrund

Die Prämienkalkulation in der deutschen Kraftfahrzeugversicherung basiert auf zahlreichen Faktoren – so vielen wie in keinem anderen Versicherungszweig der deutschen Schaden- und Unfallversicherung. Die Komplexität der Tarife ist stark durch das Spannungsfeld aus Wettbewerb und Profitabilität getrieben. Die Tarifizierung erfolgt anhand von subjektiven und objektiven Tarifmerkmalen wie der Kilometer-, Schadenfreiheits- oder Typklasse unter

Anwendung aktueller Methoden basierend auf Statistiken und Wahrscheinlichkeiten (vgl. z.B. Heep-Altiner und Klemmstein 2001). Typischerweise sind dies Merkmale, die leicht messbar sind. Einige, wie die jährliche Fahrleistung, sind jedoch im Einzelfall nur schwer nachprüfbar.

Über die festgelegten Tarifkriterien wird versucht, die Risikoprofile annähernd abzubilden. Allerdings können Versicherer auf diese Weise nicht alle Details zu den Risiken erfassen. Insbesondere sind zur Beobachtung von Fahrverhaltensweisen und Fahrzeugnutzungsarten weitere Kriterien, z.B. das Geschwindigkeits- und Beschleunigungsverhalten, die Tageszeit, in der das Fahrzeug gefahren wird, oder der Straßentyp, der am meisten genutzt wird, erforderlich.

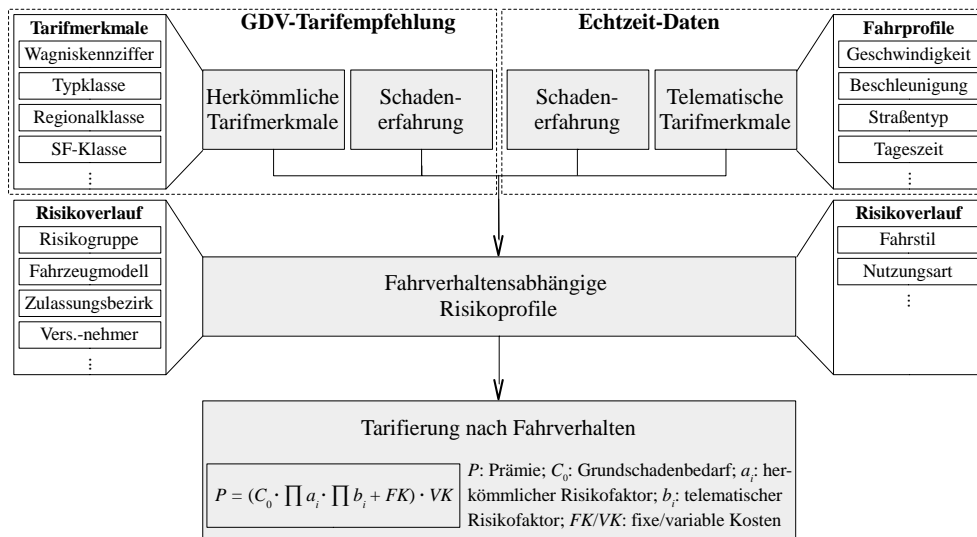
Vollkommene Informationen liegen höchstens dem Versicherungsnehmer vor, nicht aber dem Versicherungsunternehmen. Diese unvollkommenen, ungleich verteilten Informationen charakterisieren das Verhältnis zwischen Versicherten und Versicherer und führen zum bekannten Problem der Informationsasymmetrie. Daraus resultiert die Gefahr der adversen Selektion und des Moral Hazard (vgl. z.B. Akerlof 1970; Alger und Ma 2003): Die Informationsasymmetrie vor Zeitpunkt des Vertragsabschlusses hat zur Folge, dass der Versicherer die fehlenden Daten nicht in der Prämie abbilden kann und somit günstigen sowie ungünstigen Risiken hinsichtlich der fehlenden Informationen dieselben Vertragskonditionen anbieten muss. Nach Vertragsabschluss kann die Nichtberücksichtigung der Fahrweise und der Fahrzeugnutzung für den Versicherungsnehmer hingegen ein Grund für leichtfertiges Verhalten im Straßenverkehr sein.

Versicherer müssen versuchen, die Informationen zu den übernommenen Risiken laufend aktuell und umfassend zu halten, um Veränderungen von Verhaltensweisen und Zuständen reflektieren zu können. Eine aktuarielle Weiterentwicklung der Tarifierungsverfahren, begleitet von einer Steigerung der Anzahl und Qualität tarifrelevanter Merkmale, steht daher im Interesse der Versicherer (vgl. z.B. Laas et al. 2014). Dies kann im Rahmen eines PHVD-Konzeptes erfolgen, welches Informationen über die Fahrzeugnutzer und -nutzung erfasst sowie auswertet, und damit eine im Vergleich zu herkömmlichen Tarifmerkmalen ergänzende Datenbasis für eine individuellere sowie risikoadäquatere Prämiengestaltung

liefert, wie sie in Abbildung 1 dargestellt ist (vgl. z.B. Coroma und Höckl 2004; Ippisch et al. 2007; Weidner und Weidner 2014). Erforderlich ist dazu ein im Fahrzeug fest installiertes oder ein mobil mitzuführendes sensorik- und standortverarbeitendes technisches System (vgl. z.B. Ippisch et al. 2007; Gerpott und Berg 2012). Durch diese Telematikeinheit gesammelte Informationen betreffen signifikante Einflussfaktoren auf die Unfallzahlen wie z.B. Fahrtzeiten, das Beschleunigungsverhalten, Geschwindigkeitsabläufe, Witterungsverhältnisse und den Straßentyp (vgl. z.B. Oberholzer 2003; Zantema et al. 2008; Gerpott und Berg 2012).

Die Berücksichtigung von Variablen des Fahrverhaltens und der Fahrsituation ermöglicht es Versicherern, individuelle Risikoprofile besser widerzuspiegeln und erlaubt eine nutzenbasierte sowie verhaltensabhängige Kalkulation, die zu einer aus versicherungsmathematischer Sicht ausgeprägteren Prämiendifferenzierung führt. PHYD-Tarife bieten den Vorteil einer günstigeren Versicherungsprämie bei situationsangepasstem Fahrverhalten, wohingegen die Prämie bei einer riskanten Fahrweise steigt.

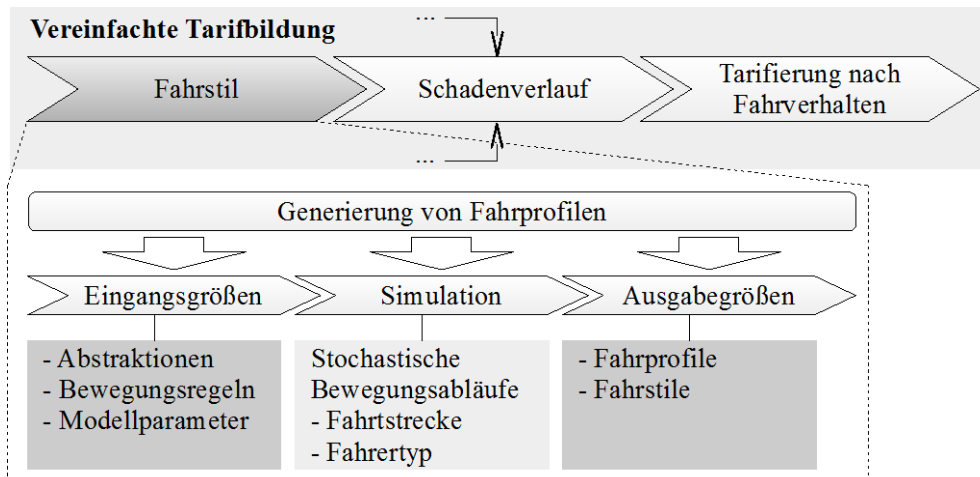
Abbildung 1: Darstellung der fahrverhaltensabhängigen Tarifierungsmethodik



3. Methodik

Im Fokus steht die Bewertung von Fahrprofilen zur Schaffung der Voraussetzungen zur Implementierung der Pkw-Telematik in die Tarifikalkulation der Kfz-Versicherung. Zur Erzielung einer hohen Präzision und großen Variationsbreite an Fahrprofilen wird die Datenbasis über eine umfangreiche stochastische Simulation gebildet; dieses Vorgehen schließt technische Fehlerquellen, z.B. aufgrund von Betriebs- und Wartungsproblemen oder einer geforderten hohen Sensitivität, aus. In einem Simulationslauf werden simultan zufällige Fahrtstrecken und Fahrertypen generiert, aus denen sich spezifische Fahrprofile ableiten lassen. Die allgemeine Vorgehensweise ist in Abbildung 2 dargestellt und wird im Anschluss näher erläutert. Vorab werden zunächst Modellanforderungen und Bewegungsregeln aufgestellt.

Abbildung 2: Modellstruktur zur Generierung von Fahrprofilen



Abstraktionen

Zunächst sind einige Abstraktionen erforderlich, um die Simulation von realitätsnahen Bewegungsmustern zu ermöglichen:

- Zeitlich diskretes System: Die Bewegung lässt sich in Bewegungs- und Ruhephasen zerlegen. Das Fahrzeug bewegt sich entlang einer Folge von

Wegpunkten fort, an denen es jeweils eine bestimmte Zeit verharret. Die Strecke zwischen zwei Wegpunkten ist durch eine spezifische Geschwindigkeit charakterisiert. Zustandsübergänge, die zu bestimmten Zeitpunkten erfolgen, werden mit einer zeitlich diskreten Simulation abgebildet; zwischen den einzelnen Zeitpunkten bleiben die Zustandsgrößen konstant. Diese Verallgemeinerung ist insbesondere vor dem Hintergrund sinnvoll, dass GPS-basierte Telemetrie typischerweise ein solches Format aufweist.

- **Mikroskopisches Modell:** Im Vordergrund steht das Bewegungsmuster einzelner Fahrzeuge basierend auf dem individuellen Fahrverhalten des Fahrzeugführers sowie auf den technischen Eigenschaften des Fahrzeugs. Der Detaillierungsgrad ist daher so zu wählen, dass die Erfassung einer Fahrer-Fahrzeug-Einheit möglich ist; die Verkehrsströme und der Verkehrsablauf spielen dabei eine untergeordnete Rolle. Das Verhalten der einzelnen Fahrer-Fahrzeug-Einheit mit ihren individuellen Eigenschaften kann über eine mikroskopische Simulation nachgebildet werden (vgl. z.B. FGSV 2006).
- **Stochastischer Prozess:** In der Verkehrsmobilität sind die Verkehrsteilnehmer (Fahrer-Fahrzeug-Einheiten) sowie Rahmenbedingungen (Wetter, Verkehrsstaus und Verweildauern) stochastischer Natur. Die Geschwindigkeit, Fahrtrichtung und Wartezeit sind nicht-deterministische Variablen, zu deren Abbildung stochastische Prozesse eingesetzt werden müssen.

Bewegungsregeln

Des Weiteren werden folgende Bewegungsregeln getroffen, um die Gegebenheiten des Straßenverkehrs realitätsnah abzubilden:

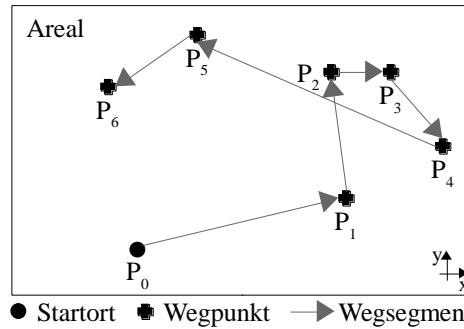
- **Höchstgeschwindigkeit:** Das Fahrzeug passt seine Geschwindigkeit in Abhängigkeit von seiner Wunschgeschwindigkeit an die angegebene Geschwindigkeitsbegrenzung bzw. Richtgeschwindigkeit an. Limitierende Faktoren wie die Verkehrsdichte oder der Straßentyp können über die Richtgeschwindigkeit widerspiegelt werden.

- Beschleunigung/Bremsverzögerung: Fahrzeuge stoppen und bewegen sich nicht übergangslos. Ein Fahrzeug beschleunigt zu Beginn bzw. verzögert die Bewegung gegen Ende eines Wegsegments (Streckenabschnitt zwischen zwei Wegpunkten) linear um eine individuelle Beschleunigungsrate bis zur Erreichung der segmentspezifischen Höchstgeschwindigkeit bzw. bis zum Stillstand.
- Verweildauer: An Hindernissen wie z.B. (Ampel-)Kreuzungen, Zebrastreifen oder Staus verharret das Fahrzeug eine spezifische Wartezeit.
- Distanz: Die Länge des zurückgelegten Weges ist definiert als die Summe der Längen der Wegsegmente.

Stochastische Bewegungsabläufe

Zur Simulation der stochastischen Bewegungsabläufe (siehe exemplarisch Abbildung 3) wird in diesem Beitrag ein Fahrzeug-Mobilitäts-Modell eingesetzt; in der Literatur wird eine Vielzahl von Mobilitäts-Modellen diskutiert. In der Physik und Finanzmathematik sind der Random Walk und die Brownsche Bewegung als beschreibende Approximation an die Realität vorherrschend (vgl. z.B. Hänggi und Marchesoni 2005). Im Bereich der drahtlosen Kommunikationsnetzwerke (vgl. Guérin 1987; Johnson und Malz 1996; Zonoozi und Dassanayake 1997; Markoulidakis et al. 1997; Bettstetter 2001) und des Verkehrsmanagements (vgl. Heidemann 1989; Nagel und Schreckenberg 1992; Briesemeister und Hommel 2010; Helbing 2001) wurden zahlreiche Modellvarianten entwickelt. In diesem Beitrag erfolgt die Modellbildung nach dem Random Waypoint (RWP) Prinzip², eingeführt durch Johnson und Malz (1996), welches den genannten Abstraktionen und Bewegungsregeln in leicht modifizierter Form adäquat gerecht wird.

² Bettstetter et al. (2004) geben einen Überblick über das RWP-Modell und zeigen wesentliche stochastische Eigenschaften auf.

Abbildung 3: Veranschaulichung der Random-Waypoint-Bewegung

Das RWP Modell beschreibt die Bewegung einer Fahrer-Fahrzeug-Einheit in einem bestimmten Areal. Als Areal wird ein begrenztes, ebenerdiges Rechteck $[0, a] \times [0, b] \subset \mathbb{R}^2$ betrachtet.³ Die räumliche Anordnung der zweidimensionalen Wegpunkte $\mathbf{P} = (P_x, P_y)$ ist durch eine Gleichverteilung gegeben. Zu Simulationsbeginn wird für einen RWP-Knoten j , $j = 1, 2, 3, \dots, m$, stellvertretend für eine Fahrer-Fahrzeug-Einheit, ein zufälliger Startpunkt $\mathbf{P}_0^{(j)}$ gewählt. Anschließend wird eine Folge von Wegpunkten $\mathbf{P}_i^{(j)}$, $i \in \mathbb{N}$, aus dem Areal bestimmt. Die Bewegungsabfolge eines RWP-Knotens j kann formal als zeitlich diskreter Prozess beschrieben werden (vgl. z.B. Bettstetter et al. 2004):

$$\{P_i^{(j)}\}_{i \in \mathbb{N}_0} = P_0^{(j)}, P_1^{(j)}, P_2^{(j)}, \dots$$

An den Wegpunkten $\mathbf{P}_i^{(j)}$ verharret der Knoten j jeweils eine zufällige Zeit (Wartezeit) $T_{p,i}$. Hierzu wird zu Beginn eines jeden Schrittes normalverteilt eine Wartezeit $t_{p,i} \in [0, t_{p,\max}]$ mit $t_{p,\max} < \infty$ bestimmt. Die Bewegung von \mathbf{P}_{i-1} zu \mathbf{P}_i erfolgt mit einer zufälligen Geschwindigkeit V_i . Die Geschwindigkeit am Punkt $r_{i,k}$, $k = 0, 1, 2, \dots, n_i$, ergibt sich aus einer segment-spezifischen nominalen Geschwindigkeit v_i und einer persönlichen Komponente $v(r_{i,k})$:

$$v_i(r_{i,k}) = v_i \cdot v(r_{i,k}).$$

Die nominale Geschwindigkeit v_i wird gleichverteilt aus dem Intervall $[v_{\min}, v_{\max}]$ bestimmt, wobei $v_{\min} > 0$ und $v_{\max} < \infty$ gewählt sind. Sie ist durch externe Faktoren wie die angegebene Geschwindigkeitsbeschränkung bzw. Richtgeschwindigkeit in Abhängigkeit vom Stra-

³ Technisch lässt sich ein zweidimensionaler Random Walk einfach umsetzen und später unproblematisch auf tatsächliche GPS-Daten adaptieren. Die dritte Dimension wird hier ohne Beschränkung der Allgemeinheit ausgelassen.

ßentyp und der Verkehrsdichte vorgegeben. Die persönliche Komponente $v(r_{i,k})$ basiert auf spezifischem Fahrer- sowie Fahrzeugtyp. Durch Einbezug dieser internen Einflussfaktoren kann die Geschwindigkeit in Abhängigkeit von der individuellen Fahrerwunschgeschwindigkeit und der spezifischen Beschleunigungsrate der Fahrer-Fahrzeug-Einheit angepasst werden. Mit dieser Erweiterung des RWP-Modells kann die Bewegung auf einer Teilstrecke verzögert bzw. beschleunigt werden. Eine Bewegungsperiode i wird damit komplett über den Vektor $(\mathbf{p}_{i-1}, \mathbf{p}_i, v_i(r_{i,k}), t_{p,i})$ beschrieben.

Fahrprofile und charakterisierende Kennzahlen

Auf Grundlage der simulierten Bewegungsabläufe werden in einem letzten Schritt spezifische Fahrprofile erstellt. In Analogie zum stochastischen Prozess als Abbild der Bewegungsabfolge eines RWP-Knotens j ist der stochastische Prozess der Abstände zwischen zwei aufeinanderfolgenden Wegpunkten gegeben durch (vgl. z.B. Bettstetter et al. 2004):

$$\{L_i^{(j)}\}_{i \in \mathbb{N}} = L_1^{(j)}, L_2^{(j)}, L_3^{(j)}, \dots,$$

mit $L_i^{(j)} = \|P_i^{(j)} - P_{i-1}^{(j)}\|_2 = \left((P_{x,i}^{(j)} - P_{x,i-1}^{(j)})^2 + (P_{y,i}^{(j)} - P_{y,i-1}^{(j)})^2 \right)^{1/2}$. Eine Realisierung dieses Prozesses wird geschrieben als $\{L_i^{(j)}\}$, wobei sich $L_i^{(j)}$ weiter zerlegen lässt in Teilstreckenabschnitte $\{s_{i,k}^{(j)}, k = 0, 1, 2, \dots, n_i\}$ in Abhängigkeit von der gewählten Abtastfrequenz.⁴ Aus den zeitlich äquidistanten Tupeln $(s_{i,k}, v_i(r_{i,k}))$, $i \in \mathbb{N}_0$, $k = 0, 1, 2, \dots, n_i$, kann ein Geschwindigkeits-Weg-Profil erstellt werden, wobei für die Wartezeiten $t_{p,i}$ die Tupel $(0, 0)$ entsprechend einzufügen sind. Die Geschwindigkeit des Fahrzeugs zwischen zwei Wegpunkten ist nicht konstant, sodass die Abtastfrequenz f_A einen entscheidenden Einfluss auf die Güte des Geschwindigkeitsprofils hat.

Im Allgemeinen gilt, dass Details nur dann (ausreichend) genau erfasst werden, wenn die Abtastfrequenz entsprechend hoch ist. Ist die Abtastfrequenz klein, kann es sein, dass Geschwindigkeitsveränderungen nicht abgebildet werden (vgl. Weidner und Weidner, 2014). Zur Ermittlung eines Geschwindigkeitsverlaufes können die Abtastungen ggf. als Stützpunkte für eine lineare Interpolation gewählt werden.

⁴ Die Abtastfrequenz f_A beschreibt die Zahl der vorhandenen Abtastungen pro Sekunde (vgl. Werner 2012). Insofern stellt sie ein Qualitätsmerkmal für die Auflösung der GPS-Signale dar.

Für die mittlere Geschwindigkeit zwischen den Wegpunkten \mathbf{P}_{i-1} zu \mathbf{P}_i gilt (vgl. z.B. Meschede 2010):

$$\bar{v}_i = \frac{\sum_{k=1}^{n_i} s_{i,k}}{\sum_{k=1}^{n_i} t_{i,k}}, \quad i \in \mathbb{N},$$

mit der Zeit $t_{i,k}$ für den Teilstreckenabschnitt $s_{i,k}$. Bei einem ausreichend kleinem Abtastintervall t_A [s] kann $t_{i,k} = 1/t_A$ gesetzt werden, sodass sich die Formel vereinfacht zu:

$$\bar{v}_i = \frac{1}{n_i} \cdot t_A \cdot \sum_{k=1}^{n_i} s_{i,k}, \quad i \in \mathbb{N}$$

Das Beschleunigungsprofil ergibt sich aus der ersten Ableitung der Geschwindigkeits-Zeit-Funktion (vgl. z.B. Meschede 2010):

$$a_i(r_{i,k}) = \dot{v}_i(t, r_{i,k}), \quad i \in \mathbb{N}_0, \quad k = 0, 1, 2, \dots, n_i.$$

Die mittlere positive bzw. negative Beschleunigung pro Beschleunigungsvorgang zwischen den Wegpunkten \mathbf{P}_{i-1} zu \mathbf{P}_i lässt sich dann wie folgt ermitteln:

$$\bar{a}_i = \frac{v_{i,m_i} - v_{i,l_i}}{\sum_{k=l_i}^{m_i} t_{i,k}}, \quad i \in \mathbb{N},$$

wobei $v_{i,m_i-1} < v_{i,m_i} \geq v_{i,m_i+1}$ und $v_{i,l_i-1} \geq v_{i,l_i} < v_{i,l_i+1}$ für $a > 0$ bzw. $v_{i,m_i-1} > v_{i,m_i} \leq v_{i,m_i+1}$ und $v_{i,l_i-1} \leq v_{i,l_i} > v_{i,l_i+1}$ für $a < 0$ zu wählen sind.

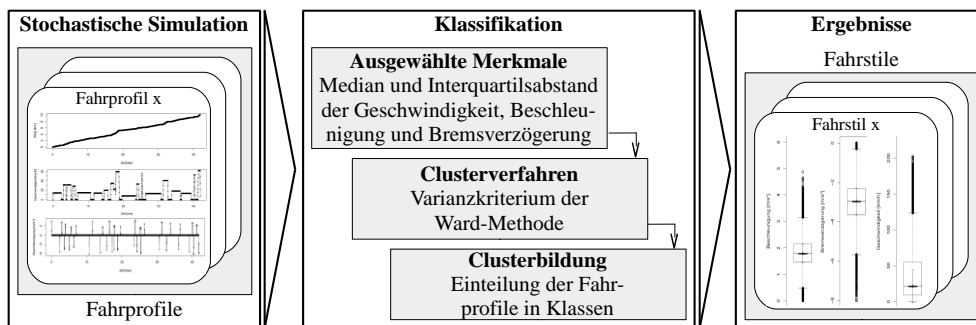
Klassifizierung der Fahrprofile

Die Geschwindigkeitsabläufe und Beschleunigungsverhalten der verschiedenen Fahrer-Fahrzeug-Einheiten zeigen eine weite Bandbreite von Variationen. Zur Identifikation spezifischer Fahrstile ist es sinnvoll, die generierten Fahrprofile über eine Clusteranalyse in Klassen zu unterteilen; Abbildung 4 gibt einen Überblick. Die einem Cluster zugeordneten Fahrprofile sollen sich dabei hinsichtlich der Variablenausprägungen möglichst ähnlich sein, sich aber von anderen Clustern zugeordneten Fahrprofilen möglichst stark unterscheiden (vgl. z.B. Han und Kamber 2006).

Das Fahrverhalten einer spezifischen Fahrer-Fahrzeug-Einheit ist originär gekennzeichnet über die Fahrzeuggeschwindigkeit und die jeweilige Beschleunigung bzw. Bremsverzögerung. Zur Clusterbildung wird, vor dem Hintergrund der Informationsaggregation, die Ähnlichkeit der Fahrprofile hinsichtlich der Mediane (als robustes Lagemaß) und Interquartils-

abstände (als Streuungsmaß) dieser drei Variablen untereinander betrachtet. In der Simulationsstudie erfolgt die Ermittlung der Cluster unter Anwendung des Varianzkriteriums der Ward-Methode (vgl. Ward 1963), rekursiv implementiert unter Verwendung des Lance-Williams-Algorithmus (vgl. Lance und Williams 1967).⁵ In der Praxis stellt diese Methode ein stabiles Verfahren dar, das zu sehr homogenen, kompakten Clustern führt (vgl. z.B. Mangiameli et al. 1996). Schließlich bilden die Cluster die zu charakterisierenden, spezifischen Fahrstile; auf diese Weise wird die Bandbreite an individuellen Fahrstilen zu einem Fahrstil im jeweiligen Cluster reduziert.

Abbildung 4: Identifikation von Fahrstilen auf Grundlage generierter Fahrprofile



Entscheidend bei der Festlegung der Clusteranzahl ist, dass sich die zu ermittelnden Cluster differenzieren lassen. Hierzu lässt sich die statistische Sicherheit der Mediane über ein approximatives 95%-Konfidenzintervall bestimmen.⁶ Gibt es zwischen den Konfidenzintervallen zweier Cluster keine Überschneidung, ist dies ein starkes Anzeichen dafür, dass sich die Verteilungszentren signifikant unterscheiden (Chambers et al. 1983) und der Unterschied zwischen den Medianen nicht nur zufallsbedingt, sondern statistisch begründbar ist (vgl. McGill et al. 1978). Dies bedeutet folglich, dass sich die spezifischen Fahrstile

⁵ Die Anwendung der Ward-Methode erscheint geeignet aufgrund der unmittelbaren Verbindung zur Euklidischen Distanz. Fahrprofile mit z.B. ähnlichen Beschleunigungswerten weisen eine Euklidische Distanz nahe Null auf. Sie werden in der gleichen Gruppe zusammengefasst und die Quadratsumme innerhalb der Gruppe ist klein.

⁶ Obere und untere Schranke sind gegeben als $Q_{0,5} \pm 1,58 \cdot IQR / \sqrt{n}$ (vgl. Chambers et al. 1983).

tatsächlich über die Mediane der Cluster charakterisieren lassen und durch die Datenaggregation genügend Informationen erhalten bleibt.

4. Ergebnisse und Diskussion

4.1 Erzeugung von Fahrprofilen

Anhand einer Simulationsstudie sollen die Charakteristika verschiedener Fahrprofile herausgearbeitet und veranschaulicht werden, die dann im Rahmen einer innovativen Prämienkalkulation Anwendung finden können. Zur Erstellung beispielhafter Fahrprofile wird das in Abschnitt 3 eingeführte modifizierte RWP-Modell eingesetzt. Tabelle 1 fasst die Modellannahmen und -parameter zusammen, die als Variationsgrößen zur Profilbildung die Ausgangsbasis darstellen. Die Parameterwahl erfolgte anhand von Fachliteratur vor dem Hintergrund, ein möglichst weites Spektrum an individuellen Eigenschaften der Fahrer-Fahrzeug-Einheiten sowie spezifischen örtlichen Gegebenheiten, mit dem Schwerpunkt auf durchschnittlichen Stadt- und Rush-Hour-Fahrten, abzubilden. Die im Rahmen dieser Analyse gewählten Parameterbereiche sind im weiteren Forschungsverlauf an spezifische zu untersuchende Situationen adäquat anzupassen.

In der Simulationsstudie werden gleichverteilte Wegpunkte ohne Beschränkung der Allgemeinheit aus einem Areal von 2 x 5 km gewählt. Weitere spezifische Örtlichkeiten und deren Einflüsse auf den Fahrstil und Geschwindigkeiten werden sowohl in Bewegungs-, als auch in Ruhephasen berücksichtigt. Segmentspezifisch wird gleichverteilt auf dem Intervall [5 km/h, 200 km/h] jeweils eine nominale Geschwindigkeit festgelegt. Wartezeiten an Wegpunkten werden pauschal für alle Hindernisse über eine Normalverteilung mit Mittelwert 40 [s] und einer Standardabweichung von 20 [s] modelliert.⁷

⁷ Die Parameter sind in Anlehnung an in Deutschland gültige Richtlinien der Forschungsgesellschaft für Straßen- und Verkehrswesen gewählt; z.B. liegt die normale Umlaufzeit einer Ampel nach der Richtlinie zur Anlage von Lichtsignalanlagen (RiLSA) bei 50-75 s, minimal 30 s und maximal 90-120 s.

Variierter Parameter	Symbol	Variationsmethode	Parametrisierung
Wegpunkte im Areal $A = [0,a] \times [0,b] \subset \mathbb{R}^2$	$p = (p_x, p_y)$	$P = (P_x, P_y) \sim U(A)$, d.h. $P_x \sim U[0,a]$, $P_y \sim U[0,b]$	$a = 2$ [km] $b = 5$ [km]
Wartezeiten an Wegpunkt	t_p	$T_p \sim N(\mu, \sigma^2)$ $t_p \in [0, t_{p,max}]$ mit $t_{p,max} < \infty$	$\mu = 40$ [s] $\sigma = 20$ [s] $t_{p,max} = 120$ [s]
Segmentspezifische nominale Geschwindigkeit	v	$V \sim U[v_{min}, v_{max}]$ mit $v_{min} > 0$ und $v_{max} < \infty$	$v_{min} = 5$ [km/h] $v_{max} = 200$ [km/h]
Geschwindigkeitsskalierung und -begrenzung pro Fahrer- Fahrzeug-Einheit	$sf_{v^{(i)}}$ $v^{(i)}_{max}$	$SF_{v^{(i)}} \sim N(\mu_{sf}, \sigma_{sf}^2)$ $V^{(i)}_{max} \sim N(\mu_v, \sigma_v^2)$ $v^{(i)}_{max} > v_{minmax}$, mit $v_{minmax} < \infty$	$\mu_{sf} = 1,02$ $\sigma_{sf} = 0,05$ $\mu_v = 140$ [km/h] $\sigma_v = 25$ [km/h] $v_{minmax} = 70$ [km/h]
Segmentspezifische Beschleunigung pro Fahrer-Fahrzeug-Einheit	a	$A^{(i)} \sim N(\mu^{(i)}, (\sigma^{(i)})^2)$ $a^{(i)} \in [a_{min}, a_{max}]$ mit $a_{min} > 0$ und $a_{max} < \infty$	$M^{(i)} \sim N(2.5, 0.4^2)$ mit $\mu^{(i)}$ $\in [a_{min}, a_{max}]$ und $\rho(\mu^{(i)}, sf_{v^{(i)}}) = 0,90$ $\sigma^{(i)}$ $= \min \left\{ \frac{\mu^{(i)} - a_{min}}{2,576}; \frac{a_{max} - \mu^{(i)}}{2,576} \right\}$ [m/s ²] $a_{min} = 1,0$ [m/s ²] $a_{max} = 6,0$ [m/s ²]
Segmentspezifische Bremsverzögerung pro Fahrer-Fahrzeug-Einheit	d	$D^{(i)} \sim N(\mu^{(i)}, (\sigma^{(i)})^2)$ $d^{(i)} \in [d_{min}, d_{max}]$ mit $d_{min} > 0$ und $d_{max} < \infty$	$M^{(i)} \sim N(4.75, 1.0^2)$ mit $\mu^{(i)} \in [d_{min}, d_{max}]$ und $\rho(\mu^{(i)}, sf_{v^{(i)}}) = 0,90$ $\sigma^{(i)}$ $= \min \left\{ \frac{\mu^{(i)} - d_{min}}{2,576}; \frac{d_{max} - \mu^{(i)}}{2,576} \right\}$ [m/s ²] $d_{min} = 1,5$ [m/s ²] $d_{max} = 8,0$ [m/s ²]

Tabelle 1: Zentrale Eingangsparameter der Simulationsstudie

Neben der Örtlichkeit wirken sich individuelle Fahrereigenschaften (Alter, Fahrerfahrung, Fahrzeugnutzungsart, etc.) und die Leistungscharakteristik des Fahrzeuges auf die Geschwindigkeitsabläufe aus.⁸ Einerseits erfolgt dazu eine Skalierung und Begrenzung der nominalen Geschwindigkeit, um aus ihr eine Fahrerwunschgeschwindigkeit abzuleiten. Der Faktor zur Geschwindigkeitsskalierung wird aus einer Normalverteilung mit Mittelwert 1,02 und einer Standardabweichung von 0,05 ermittelt.⁹ Positiv korreliert mit dem Skalierungsfaktor ($\rho = 0,9$) wird pro Fahrer-Fahrzeug-Einheit die Normalverteilung für die linearen Beschleunigungs- bzw. Bremsverzögerungsfaktoren parametrisiert. Es wird angenommen, dass die segmentspezifische Beschleunigung¹⁰ zwischen 1 und 6 m/s² liegt und die segmentspezifische Verzögerung¹¹ aus dem Intervall 1,5 und 8,0 m/s² stammt. Abweichungen in Richtung 0 m/s² können aufgrund eines gewöhnlichen Auslaufprozesses am Ende eines Beschleunigungs- bzw. Bremsverzögerungsvorgangs punktuell auftreten, sind aber in den Simulationsergebnissen angesichts der festgelegten unteren Intervallgrenzen nicht gehäuft anzutreffen.

Durch Ausschluss von sogenanntem weißen Rauschen fokussiert das Modell auf tatsächlichen Geschwindigkeitsänderungen, d.h. auf den eigentlich interessierenden Wertebereich zur Beschreibung von Beschleunigungsprofilen. Zudem werden mit der nach oben gewählten Intervallbeschränkung sogenannte Hochwerttrisiken, wie insbesondere Sportfahrzeuge und Vollbremsungen, bewusst ausgeschlossen und unterliegen einer gesonderten Bewertung.

⁸ Individuellen Fahreigenschaften und Leistungscharakteristiken der Fahrzeuge haben bekanntermaßen Einfluss auf den Schadenverlauf (vgl. Statistisches Bundesamt 2014), sodass eine Kausalität zwischen statistischen sowie personenbezogenen Merkmalen und GPS-basierten Fahrzeugdaten zu Grunde gelegt werden kann.

⁹ Die Parameter wurden auf Basis einer Bestandsanalyse des Kraftfahrt-Bundesamts festgelegt (vgl. KBA 2014). Hierzu wurden die Fahrzeughalter erfahrungsgemäß verschiedenen Personengruppen (risikoavers, risikoaffin, risikoneutral, indifferent) zugeordnet und die spezifischen Fahrstile charakterisiert.

¹⁰ Das Intervall ist in Anlehnung an die Studie von Lange (2006) gewählt. Lange führt für geradeaus fahrende Pkws ein normales Beschleunigungsintervall von 1,5-2,5 m/s² und ein maximales Beschleunigungsintervall von 3,0-6,0 m/s² sowie für ein-/abbiegende Pkws ein Intervall von 1,0-2,0 m/s² auf.

¹¹ Die Wahl des Intervalls erfolgte in Anlehnung an Becke (2005). Das Bremsverzögerungsniveau hängt neben den individuellen Faktoren wie z.B. den Reifen, Bremsen und spezifischen Bremsverhalten entscheidend von den Fahrbahnverhältnissen wie z.B. dem Fahrbahnbelag, den Witterungsverhältnissen und der Steigung ab. Ein Pkw hat unter idealen Bedingungen eine Bremsverzögerung von 7-8 m/s², vereinzelt bis 10 m/s². Die Bremsverzögerung sinkt auf feuchter/nasser Fahrbahn auf 6-7 m/s², bei Schnee- und Eisglätte auf bis zu 1 m/s².

4.2 Analyse des Fahrverhaltens

Die Fahrprofilanalyse basiert auf 10.000 Fahrer-Fahrzeug-Einheiten und einer Folge von je 20 Wegpunkten. Dieses Vorgehen stellt sicher, dass die statistische Analyse der Fahrprofile alle relevanten Verhaltens-Stereotypen – Stadtfahrt, Überlandfahrt und Autobahn, jeweils mit sicherem, erfahrener, defensivem, risikoreichem und aggressivem Fahrstil – für verschiedene auf dem Markt befindlichen Fahrzeugklassen ausreichend genau abbildet. Um detaillierte Fahrprofile ohne Informationsverlust zu erhalten, wird die Abtastfrequenz f_A auf 1 Hz gesetzt.¹² Die Clusterbildung selbst erfolgt unter Anwendung der Ward-Methode (vgl. Ward 1963) anhand der Mediane sowie der Interquartilsabstände der Geschwindigkeits-, Beschleunigungs- und Bremsverzögerungswerte der generierten Fahrprofile.

Die Fahrprofile der Simulationsstudie lassen sich in sechs Fahrstile einteilen (siehe Abbildung 5), die als statistisch signifikant verschieden, zu einem (approximativen) Konfidenzintervall von 95 %, hinsichtlich ihrer zentralen Lage betrachtet werden können. Die Beschleunigungswerte schwanken – ohne Berücksichtigung von Ausreißern – bis maximal 5 m/s². Werte, die weiter als das 1,5-fache des Interquartilsabstandes außerhalb des unteren und oberen Quartils liegen, werden in den Box-Plots nicht abgebildet, kommen aber in jedem der sechs Fahrstile vor. Etwa 2,6 % der Stichprobe entfallen auf den Wertebereich, innerhalb dessen Werte als Ausreißer definiert werden. Die Mediane der sechs Untergruppen liegen zwischen 1,8 und 2,5 m/s². Weiter zeigen die Boxplots eine Erhöhung des Interquartilsabstands der Daten mit zunehmendem Median, was wiederum Rückschlüsse über die Streuung der Daten erlaubt. Der Interquartilsabstand schwankt zwischen 0,6 und 1,2 m/s².

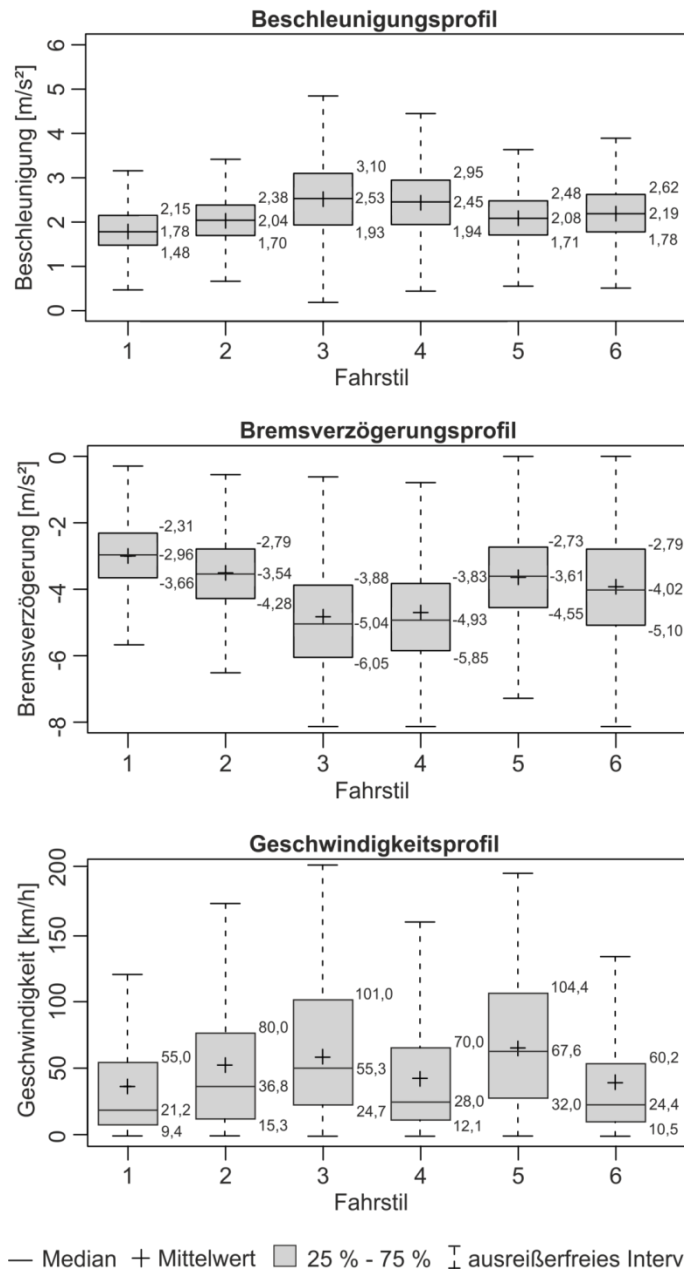
Aus den Fahrstilen für die Bremsverzögerungswerte wird ersichtlich, dass die Spannweite bis zu 8 m/s² beträgt. Ausreißer werden nur bei 1,9 % der Stichprobenwerte beobachtet. Bei drei der sechs Fahrstile kommen niedrige bzw. hohe Werte so häufig in der Stichprobe vor, dass sie gar nicht erst als Ausreißer klassifiziert werden. Die Mediane der sechs Fahrstile nehmen Werte von -3,0 bis -5,0 m/s² an. Der Interquartilsabstand steigt auch hier tendenzi-

¹² Eine Analyse des systematischen Fehlers, dessen Ausprägung mit der gewählten Abtastfrequenz verknüpft ist, kann in Weidner und Weidner (2014) nachgelesen werden.

ell – unter Ausschluss von Fahrstil 6 – mit zunehmendem absoluten Betrag des Medians an und nimmt Werte zwischen -1,3 und -2,3 m/s³ an.

Beim Vergleich der Fahrstile bezüglich der Bremsverzögerung mit den Stilen bezüglich der Beschleunigung fällt eine Symmetrie auf, so beträgt etwa der Korrelationskoeffizient bezüglich der Mediane -0,99. Dies bedeutet insbesondere, dass übereinstimmende Fahrstile mit hoher Beschleunigung ebenfalls eine hohe Bremsverzögerung – und umgekehrt – aufweisen. Dies deutet darauf hin, dass der Fahrstil sowohl auf der Beschleunigung als auch auf der Bremsverzögerung beruht und nicht aus einem dieser beiden Parameter allein abgeleitet werden kann.

Abbildung 5: Box-Plot-Darstellung der Beschleunigungs-, Bremsverzögerungs- und Geschwindigkeitswerte der Simulationsergebnisse



Die Boxplot-Darstellung der Geschwindigkeitswerte zeigt signifikante Mediane zwischen 21,2 und 67,6 km/h. Weiter wird ersichtlich, dass diejenigen Werte aller Fahrstile häufiger beobachtet werden, die kleiner als der Mittelwert sind. Folglich sind die Verteilungen (leicht) rechtsschief. Hierfür gibt es drei Gründe: Zunächst das Anfahren und Abbremsen aufgrund von Hindernissen wie z.B. Verkehrsampeln. Hinzu kommen die vorgegebenen Geschwindigkeitsbegrenzungen und schließlich auch das Ausmaß der Abweichung der Fahrerwunschgeschwindigkeit von der vorgeschriebenen Geschwindigkeit (vgl. Schüller, 2010). Darüber hinaus gibt es Ausreißer nach oben, die etwa 2,0 % der Stichprobe ausmachen; signifikante Abweichungen nach unten können aufgrund der wiederholt auftretenden Mindestgeschwindigkeit von 0 km/h an (Ampel-)Kreuzungen, Zebrastreifen oder Staus nicht vorkommen. Wiederum steigt der Interquartilsabstand mit zunehmendem Median und schwankt zwischen 45,7 und 76,3 km/h.

Neben dem Beschleunigungs- und Bremsverhalten hängen auch die Geschwindigkeitsabläufe von den technischen Fahrzeugeigenschaften und der Haltung des Fahrers ab. Deshalb ist auch die Geschwindigkeit verknüpft mit dem Beschleunigungs- und Bremsverhalten.¹³ Allerdings ist die Wechselbeziehung nur schwach, weil einige Faktoren die naturgemäße Tendenz ausgleichen: Einerseits können Fahrer mit identischen Beschleunigungs- und Bremsprozessen zwischen verschiedenen Straßentypen, d.h. Straßen innerhalb geschlossener Ortschaften, Straßen außerhalb geschlossener Ortschaften und Autobahnen, mit unterschiedlichen Geschwindigkeitsbegrenzungen bzw. Richtgeschwindigkeiten wählen. Abbildung 6 und 7 verdeutlichen, dass sich die Fahrprofile eines spezifischen Fahrzeugführers in Abhängigkeit von der Fahrzeugnutzung, d.h. der Fahrtstrecke, unterscheiden. Neben der Geschwindigkeit und der Standzeit variiert insbesondere die Länge der Beschleunigungs- und Bremsverzögerungsprozesse. Andererseits führt eine defensive Fahrweise, als gleichmäßige Fahrweise ohne störende Beschleunigungs- und Bremsvorgänge, zu einer höheren Durchschnittsgeschwindigkeit.

In Summe lässt sich festhalten, dass der Fahrstil nur unter Einbezug der Geschwindigkeit als dritten Parameter umfassend differenziert wird. Dabei ist anzumerken, dass nicht an die

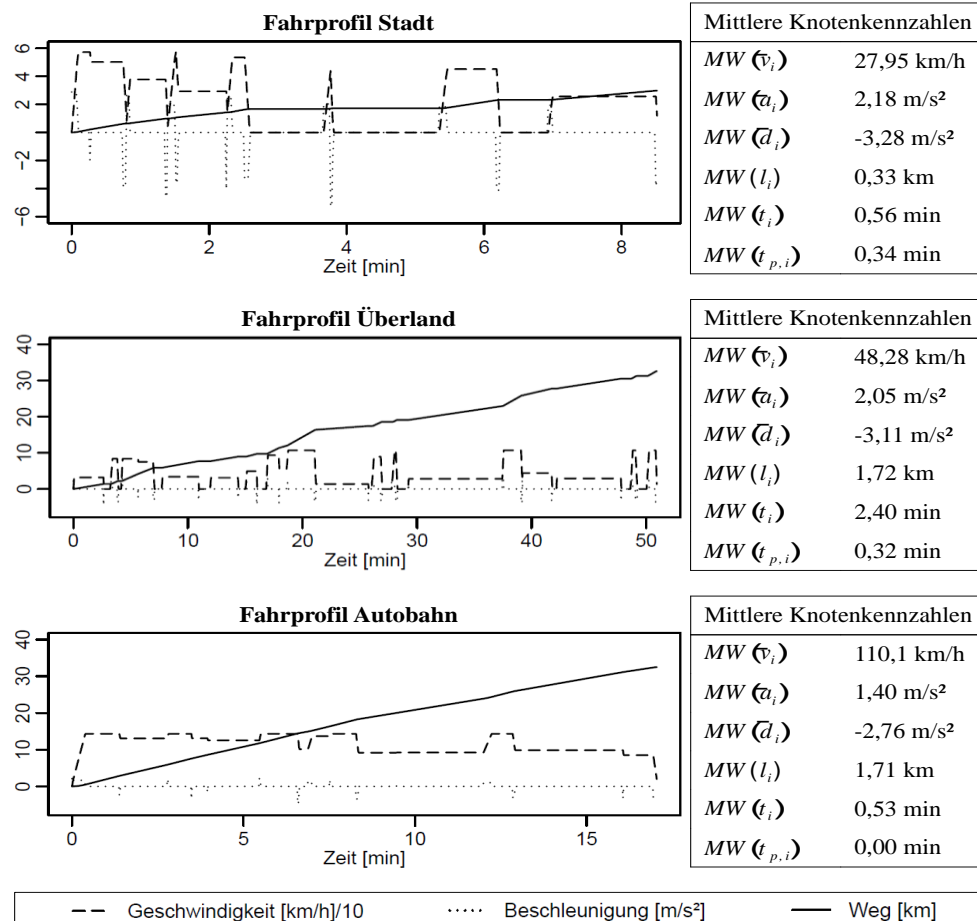
¹³ Dies liegt nicht zuletzt daran, dass die Ableitung der Geschwindigkeit nach der Zeit die Beschleunigung ist.

Situation angepasste Geschwindigkeiten¹⁴ im Rahmen der vorliegenden Analyse nicht näher betrachtet werden und einer gesonderten Wertung unterliegen.

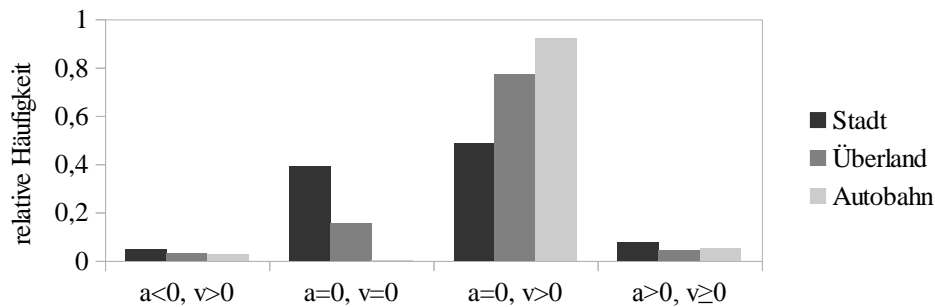
Der Fokus der Modellbildung lag auf der Generierung realistischer Fahrprofile. Ein Abgleich der Simulationsergebnisse mit der Fachliteratur zeigt, dass sich die Ergebnisse des physikalischen Modells mit der Realität decken, d.h. dass die gewählten Parameterbereiche und Verteilungsannahmen eine Übertragung der generierten Fahrprofile sowie der daraus resultierenden Fahrstile auf die Realität zulassen. Die Ergebnisse zeigen neben dem Differenzierungspotential zwischen defensiver und adverser Fahrweise ebenfalls, dass verschiedene Abstufungen in der Fahrweise möglich sind. Über die Häufigkeit des Auftretens sowie die Risikobewertung einzelner Fahrstile lässt sich allein anhand der Simulationsergebnisse keine Aussagen machen. Dazu gilt es im nächsten Schritt die identifizierten simulationsbasierten Fahrstile mit Telematik-Daten aus einem empirischen Versicherungsportfolio zu koppeln.

¹⁴ Auch an dieser Stelle gilt, dass ein Risikoprofil nur unter Einbezug von Höhe und Dauer der Geschwindigkeitsabweichungen im Allgemeinen, d.h. bei Betrachtung von Geschwindigkeitsüberschreitungen sowie -unterschreitungen im Sinne eines Verstoßes gegen § 3 Abs. 2 StVO, umfassend differenziert wird.

Abbildung 6: Gegenüberstellung nach Straßentyp differenzierter Fahrprofile eines beispielhaften Fahrzeugführers



$MW(\bar{v}_i)$: Mittelwert über die mittleren Geschwindigkeiten zwischen den Wegpunkten P_{i-1} und P_i ; $MW(\bar{a}_i)$: Mittelwert über die mittleren Beschleunigungen pro Beschleunigungsvorgang zwischen den Wegpunkten P_{i-1} und P_i ; $MW(\bar{d}_i)$: Mittelwert über die mittleren Bremsverzögerungen zwischen den Wegpunkten P_{i-1} und P_i ; $MW(l_i)$: Mittelwert über die mittleren Streckenlängen zwischen den Wegpunkten P_{i-1} und P_i ; $MW(t_i)$: Mittelwert über die Fahrtzeiten zwischen den Wegpunkten P_{i-1} und P_i ; $MW(t_{p,i})$: Mittelwert über die mittleren Standzeiten am Wegpunkt P_i ; $i = 1, 2, \dots, 10$ bei Fahrprofil Stadt und $i = 1, 2, \dots, 20$ bei Fahrprofil Überland bzw. Autobahn

Abbildung 7: Zerlegung der Tangentialbeschleunigung

$a < 0, v > 0$: negative Beschleunigung; $a = 0, v = 0$: Stillstand; $a = 0, v > 0$: geradlinige Bewegung; $a > 0, v \geq 0$: positive Beschleunigung

4.3 Konsequenzen für die Kfz-Versicherungstarifizierung

Die Ergebnisse der Simulationsstudie zeigen, dass die sogenannten Telematik-Technologien in der Kfz-Versicherung auf die bislang vorherrschende Tarifierungsmethodik disruptiv wirken könnten. Die Telematik bietet neue Informationen durch eine Verknüpfung herkömmlicher mit komplementären telematischen Tarifmerkmalen und dadurch die actuarielle Möglichkeit, Prämien differenzierter und individueller zu gestalten. Insbesondere erlaubt sie eine Tariffdifferenzierung durch Berücksichtigung von Fahrstilen innerhalb bis dato vermeintlich gleicher Risikogruppen; die Variablen des Fahrverhaltens können zwischen verschiedenen Fahrprofilen innerhalb mit herkömmlichen Tarifmerkmalen nicht differenzierbarer Risikogruppen stark variieren.

Die Prämienhöhe hängt bei einer Telematik-Versicherung weniger von langfristigen Schadenerfahrungen ab, sondern stärker vom kurz- bis mittelfristigen Fahrverhalten und Fahrzeugnutzung. Bewertet wird dazu, ob das Fahrverhalten und die Fahrzeugnutzung auf Einzelfahrtenniveau Unfall vermeidend oder eventuell risikoreich ist.

Ein Mehrwert ergibt sich bei Betrachtung des Merkmals Alter des Versicherungsnehmers. Für die herkömmliche Tariffkalkulation liegen den Versicherungsunternehmen altersgruppenabhängige Schadenbedarfe vor, die dem jeweiligen Versicherungsvertrag entsprechend zugeordnet werden. Beispielsweise verursacht das Teil-Risikokollektiv der Fahranfänger statistisch deutlich mehr Unfälle als erfahrene Fahrer (vgl. Statistisches Bundesamt 2014),

sodass die hohen Risikoaufschläge für Fahranfänger somit dem hohen (mittleren) Schadenbedarf dieses Kollektivs entsprechen. Der Einbezug des Fahrverhaltens als neuen multiplikativen Risikofaktor ermöglicht es dem Versicherer jedoch, innerhalb dieser Risikogruppe vereinzelte Abweichungen vom statistischen Mittel zu bestimmen und in die Tarifizierung entsprechend aufzunehmen. Kann z.B. ein 18-Jähriger Fahranfänger über sein telematisches Fahrprofil zeigen, dass sein (zu erwartender) tatsächlicher Schadenbedarf nicht annahm gemäß dem seiner Altersgruppe, sondern dem Schadenbedarf einer höheren Altersgruppe entspricht, fällt sein altersabhängiger Risikoaufschlag bei einem aktuariell auskalkuliertem Tarif geringer aus.

Hohe Risikoaufschläge sind ebenfalls mit den Merkmalen der Schadenfreiheitsklasse und Typklasse verbunden. Ergibt sich die Einstufung in eine niedrige Schadenfreiheitsklasse nicht aufgrund einer regelmäßigen Rückstufung einhergehend mit Schadenfällen, sondern aufgrund einer fehlenden Schadenhistorie (bei erst kürzlich abgeschlossen Verträgen), kann Telematik zu einem unverzüglichen Abbau der Informationsasymmetrie verknüpft mit einer deutlichen Prämienreduktion zugunsten des Versicherungsnehmers beitragen. Herkömmlich erfolgt etwa die Einstufung bei Vertragsbeginn für komplette Neuverträge üblicherweise in die Schadenfreiheitsklasse 0. Für diese Klasse sehen die unternehmensindividuellen Schadenfreiheitsklassensysteme einen Beitragssatz von rund 100 % vor, wohingegen der Beitragssatz bei einer Dauer des ununterbrochenen, schadenfreien Verlaufs von 35 Jahren in der Regel auf 20 % abfällt. Lässt sich nun die im Falle eines Neuvertrags noch nicht aufgebaute langfristige Schadenerfahrung durch ein auf Basis von Telematik-Daten ermitteltes individuelles Unfallrisiko anreichern, wären deutliche Reduktionen des Beitragssatzes aktuariell theoretisch begründbar. Nämlich dann, wenn etwa ein Neuvertrag durch einen erfahrenen, jahrelang unfallfreien Fahrer als alleiniger Nutzer abgeschlossen wird. Ferner kann die langfristige Schadenhistorie durch die individuelle Fahrverhaltensbewertung de facto überflüssig werden und Telematik somit zur Rationalisierung der Tarifkriterien beitragen.

Des Weiteren bestehen vermehrte Informationsasymmetrien, sofern das Fahrzeug von verschiedenen Nutzern verwendet wird. Wenngleich der Versicherungsnehmer in der Regel

zur Angabe des Nutzerkreises verpflichtet ist, kann die personenbezogene Verteilung der Fahrzeugnutzung von Seiten des Versicherers für den Tarifizierungszweck lediglich abgeschätzt werden. Insofern ermöglicht das Telematik-Konzept dem mit großen Nutzerkreisen, im Speziellen mit neuen Mobilitätskonzepten wie z.B. dem CarSharing, einhergehenden Wegfall von eindeutigen Tarifmerkmalen zu begegnen.

Ferner kann die Preisgestaltung durch Telematik aufgrund der (unmittelbaren) Belohnung einer guten und sicheren Fahrweise das Fahrverhalten positiv beeinflussen (vgl. Towers Watson 2013). Verbesserungen des Fahrverhaltens können, neben der üblichen Risikoselektion zu Vertragsabschluss, langfristig gesehen zu einer Verbesserung der Risikostruktur im Versichertenkollektiv eines Versicherers führen.

Eine Telematik-Versicherung wirkt sich jedoch nicht notwendigerweise bei allen Fahrern oder allen Fehlerursachen gleichmäßig positiv aus. Auf der einen Seite ist der Sachverhalt zu beachten, dass sich ein Übermaß an Kontrolle und Bevormundung nicht gleichzeitig auf alle möglichen Fehlerquellen schadensmindernd auswirken muss. Sei es aus Nervosität, aus Ablenkung oder Überflutung, hinter jeder Ecke lauert ein Fehler. Andererseits kann die Situation eintreten, dass ein übervorsichtiger Fahrer selbst keinen Fehler verursacht, dafür aber selbst die tiefere Ursache fremder Fehler wird.

An dieser Stelle sei jedoch darauf hingewiesen, dass den deutschen Versicherern derzeit kein aktuariell ausdifferenziertes Telematikprodukt vorliegt (vgl. Morawetz 2015). Eine grundlegende Voraussetzung dafür ist zunächst der Aufbau einer adäquaten Datengranularität, signifikanten Schadenhistorie in Bezug auf Telematik-Daten sowie systematischer Auswertungsmethoden (vgl. Nießen 2015). Die vorherrschenden Rabattsystematiken werden einstweilen ex post auf traditionell kalkulierte Prämien angerechnet. Die Rabatte von bis zu 30 % ergeben sich vornehmlich aus der univariaten Bewertung von Fahrverhaltens- und Fahrsituationsvariablen (vgl. z.B. Sparkassen Direktversicherung 2014).

Es gibt verschiedene Blickwinkel für das Angebot von Telematik-Produkten. Zum Aspekt der Prämienersparnis kommen Sicherheits- und Servicewünsche des Kunden hinzu. Ein integriertes Servicepaket zielt auf technische Interessen und eine Verbesserung der Fahr-

leistung ab, wohingegen das Sicherheitspaket Aspekte wie ein Notruf-System und einen Diebstahl-Schutz aufgreift (vgl. Hartmann und Nutzenadel 2015).

Angesichts der Vielschichtigkeit hinsichtlich der Produktgestaltung – abgesehen von der Tatsache, dass Test- und Experimentierfeldern zur Erzielung von Anreizen lediglich Prämienschläge vorsehen – können Auswirkungen auf den Versicherungsbestand, insbesondere Selektions- und Wanderungseffekte, zum Zeitpunkt der Markteinführungen telematischer Produkte in Deutschland für die unmittelbare Zukunft nicht abgeschätzt werden. Im Experimentierstadium werden günstige Prämien eingesetzt, danach muss es aus Wettbewerbsgründen gelingen, die mit der genutzten Telematik-Technologie einhergehenden höheren Akquisitions-, Produkt- und Betriebskosten wesentlich niedriger zu halten als die Ersparnis im Schadenbedarf. Erst dann stellt sich heraus, ob sich das Konsumentenverhalten für das neue Produkt entscheidet. Langfristig ist davon auszugehen, dass das System der Integration von Big Data in die Prämienberechnung Eingang findet. Aller Voraussicht nach ist die Telematik lediglich eine Vorbereitung der Versicherer auf weitere technologische Entwicklungen im Hinblick auf das autonome Fahren.

Über dies hinaus veranschaulicht der Modellansatz, dass die Messung und Bewertung von Fahrverhalten auf Basis aggregierter Daten, d.h. Lage- und Streuungsmaßen, unter Heranziehung einiger weniger Parameter einschließlich ihrer Interdependenzen erfolgen kann. Dieses Ergebnis ist von besonderer Bedeutung vor dem Hintergrund der personenbezogenen Datensammlung. Gemäß dem in Deutschland geltenden Grundsatz der Datensparsamkeit und Datenvermeidung (§ 3a Bundesdatenschutzgesetz) und dem bundesweiten Datenkodex der Versicherungswirtschaft (Code of Conduct) sind so wenig personenbezogene Daten wie möglich zu sammeln, auszuwerten und zu verarbeiten. Diese Zielausrichtung ist bereits im Pflichtenheft der Techniker beim Aufsatz der GPS-Technologie zu berücksichtigen, da nur zwingend notwendige Telematikdaten an Dritte zu übermitteln sind. Weiter kann durch die Ermittlung von Fahrstilen mittels der eingeführten minimalen Datenanforderung sichergestellt werden, dass sich jeder erhobene Datenpunkt auf die Tarifgestaltung auswirkt. Damit wird der Grundsatz der Zweckbindung (§ 31 Bundesdatenschutzgesetz) erfüllt, dem zufolge jeder Datenverarbeitung ein bestimmter Zweck zugrunde liegen muss.

5. Fazit und Ausblick

Anders als im europäischen und außereuropäischen Ausland, in dem telematische Kfz-Versicherungsprodukte bereits fest etabliert sind, werden diese in Deutschland noch erprobt. Ursächlich für die Zurückhaltung ist, dass die Prämienberechnung mit vielen Tarifierungsmerkmalen bereits sehr differenziert und das zusätzliche Potential erst bei niedriger Granularität der Daten und ausgereiften Auswertungsmethoden ausgeschöpft werden kann (vgl. Nießen 2015). Maßgeblich sind ebenfalls datenschutzrechtliche Bedenken und Kosten der technologischen Implementierung. Allerdings veranlasst die verpflichtende Einführung des automatischen Notrufsystems eCall für Neuwagen in der Europäischen Union ab voraussichtlich 2018 die deutsche Versicherungswirtschaft dazu, sich in den nächsten Jahren vermehrt mit der Verkehrstelematik auseinanderzusetzen.

Im Rahmen der Simulationsstudie wurde eine Klassifizierung von Fahrprofilen anhand des Fahrverhaltens vorgestellt. Basierend auf stochastisch simulierten Fahrprofilen lassen sich unter simultaner Betrachtung von Geschwindigkeits- und Beschleunigungsparametern sechs Fahrstile unterscheiden. Die präsentierten Fahrstile stehen beispielhaft für die untersuchten Parameterbereiche – ein echtes Versicherungsportfolio kann, auch in der durch die Clusteranalyse bestimmten Anzahl der Stile, basierend auf verschiedenen Abwägungen des Pricing und der technischen Realisierung davon erheblich abweichen.

Der Mehrwert der vorgestellten Methode besteht insbesondere darin, die Bewegungsvektoren differenziert auszuwerten. Die Clusteranalyse der differenziellen Betrachtung von Geschwindigkeit und Beschleunigung (inkl. Bremsen) ergibt ein genaues Scoring des Fahrverhaltens. Gegenüber der isolierten Betrachtung jedes Faktors lassen sich auf diese Weise falsch-negative Bewertungen des Fahrverhaltens vermeiden.

Die Ergebnisse sind für die Versicherungspraxis von besonderer Bedeutung, da hier die Integration von Telematikdaten in den aktuariellen Pricing-Prozess exemplarisch ersichtlich wird. Vor diesem Hintergrund wird im weiteren Forschungsverlauf angestrebt, die identifizierten modellbasierten Fahrstile mit empirischen Telematikdaten zu überprüfen. Darüber hinaus sind die Ergebnisse auch relevant für Techniker, weil sie für die technische Realisierbarkeit gemäß dem Pflichtenheft zuständig sind.

6. Abstract:

Telematics tariffs are one of the future issues in German car insurance. This paper presents a systematic approach to the evaluation of driving profiles, which allows to calculate risk-based premiums for telematics tariffs. We propose an allocation of the driving profiles based on velocity, acceleration and deceleration processes to specific driving styles for assessing the driving behavior. A vehicle mobility model forms the high quality and broad database into the analysis. The result is highly relevant for insurers, who calculate the tariffs, but also for engineers, since they have to prepare the specifications for the technologies.

7. Literaturverzeichnis

- Akerlof, G.A. (1970), The market for lemons: Quality uncertainty and the market mechanism, *The Quarterly Journal of Economics*, 84(3), 488-500.
- Alger, I. und Ma, C.A. (2003), Moral hazard, insurance, and some collusion, *Journal of Economic Behavior and Organization*, 50(2), 225-247.
- Becke, M. (2005), Unfallrekonstruktion – Geschwindigkeitsermittlung vor Bremsbeginn, *VerkehrsrechtsReport*, 1/2005, 20-24.
- Bettstetter, C. (2001), Smooth is Better than Sharp: A Random Mobility Model for Simulation of Wireless Networks, *Proc. ACM International Workshop on Modeling, Analysis, and Simulation of Wireless and Mobile Systems (MSWiM)*, 19-27, New York.
- Bettstetter, C., Hartenstein, H. und Pérez-Costa, X. (2004), Stochastic Properties of the Random Waypoint Mobility Model, *Wireless Networks*, 10(5), 555-567.
- Briesemeister, L. und Hommel, G. (2000), Role-based multicast in highly mobile but sparsely connected ad hoc networks, *Proc. Workshop on Mobile Ad Hoc Networking and Computing*, 45-50.
- Chambers, J.M., Cleveland, W.S., Kleiner, B. und Tukey, P.A. (1983), *Graphical Methods of Data Analysis*, Duxbury Press.

- Coroma, V. (2006), *The Smart Tachograph – Individual Accounting of Traffic Costs and its Implications*, Zürich.
- Coroma, V. und Höckl, N. (2004), *Pervasive Insurance Markets and their Consequences*, Zürich.
- FGSV (2006), *Hinweise zur mikroskopischen Verkehrsflusssimulation: Grundlagen und Anwendung*, Forschungsgesellschaft für Straßen- und Verkehrswesen, FGSV Verlag.
- Friedrichs, F. (2013), Black Box mit Durchblick, *Topics Magazin*, 1/2013, 40-45.
- Gerpott, T.J. und Berg, S. (2012), Pay-As-You-Drive Angebote von Erstversicherern für Privatkunden: Eine betriebswirtschaftliche Analyse, *ZVersWiss*, 101, 3-29.
- Guérin, R.A. (1987), Channel occupancy time distribution in a cellular radio system, *Vehicle Technol.*, 36(3), 89-99.
- Hartmann, M. und Nutzenadel, C. (2015), Trends in der Technologie sowie Erkenntnisse des Behavioural Pricings vereinbaren, *I · VW Management Information – St. Galler Trendmonitor für Risiko- und Finanzmärkte*, 37(1), 3-9.
- Hänggi, P. und Marchesoni, F. (2005), 100 Years of Brownian motion, *Chaos* 5, 026101.
- Han, J. und Kamber, M. (2006), *Data Mining: Concepts and Techniques*, Morgan Kaufmann, San Francisco.
- Heep-Altiner, M. und Klemmstein, M. (2001), *Versicherungsmathematische Anwendungen in der Praxis: Mit Schwerpunkt Kraftfahrt und Allgemeine Haftpflicht*, Verlag Versicherungswirtschaft GmbH, Karlsruhe.
- Heidemann, D. (1989), Ein mathematisches Modell des Verkehrsflusses, *Schriftenreihe des Instituts für Verkehrswesen der Universität Karlsruhe*, (42), Karlsruhe.
- Helbing, D. (2001) Traffic and related self-driven many-particle systems, *Rev. Modern Phys*, 73(4), 1067-1141.
- Ippisch, T., Thiesse, F. und Fleisch, E. (2007), *An Analysis of Sensor-Based Premium Schemes in the Car Insurance Industry*, St. Gallen.
- Iqbal, M.U. und Lim, S. (2006) A Privacy Preserving GPS-based Pay-as-You-Drive Insurance Scheme, *Symposium on GPS/GNSS (IGNSS)*, Sydney.

- Iqbal, M.U. und Lim, S. (2010), Privacy Implications of Automated GPS Tracking and Profiling. *Technology and Society Magazine*, 29(2), 39-46.
- Johnson, D.B. und Malz, D.A. (1996), Dynamic Source Routing in Ad Hoc Wireless Networks, In: Imielinkis, T., Korth, H.F. (Hrsg.) *Mobile Computing*, 153-181, Kluwer Academic Publishers.
- Jun, J., Guensler, R. und Ogle, J.H. (2005), Smoothing Methods Designed to Minimize the Impact of GPS Random Error on Travel Distance, Speed, and Acceleration Profile Estimates, *Transportation research record*, 1972, 141-150.
- Jun, J., Ogle, J. und Guensler, R. (2007), Relationships Between Crash Involvement and Temporal Spatial Driving Behaviour Activity Patterns: Use of Data for Vehicles with Global Positioning Systems, *Transportation Research Record*, 2019, 246-255.
- Jun, J., Ogle, J. und Guensler, R. (2011), Differences in observed speed patterns between crash-involved and crash-not-involved drivers: Application of in-vehicle monitoring technology, *Transportation Research Part C* 19(4), 569-578.
- KBA (2014), *Fahrzeugzulassungen (FZ) – Bestand an Kraftfahrzeugen und Kraftfahrzeuganhängern*, Kraftfahrt-Bundesamt, Flensburg.
- Kremslehner, D. und Muermann, A. (2013), *Asymmetric Information in Automobile Insurance: Evidence from Driving Behavior*, Wien.
- Laas, D., Hartmann, M., Nützenadel, C., Schmeiser, H. und Wagner, J. (2014), *Pricing-Strategien in der KFZ-Versicherung*, St. Gallen/Zürich
- Lance, G.N. und Williams, W.T. (1967), A General Theory of Classificatory Sorting Strategies, *Computer Journal* 9(4), 373-380.
- Lange, F. (2006), Unfallrekonstruktion – Anfahrbeschleunigungen, *VerkehrsrechtsReport*, 10/2006, 377-382.
- Lee, Y. (2008), Applications of Sensing Technologies for the Insurance Industry, In: Michahelles, F. (Hrsg.) *Business Aspects of the Internet of Things*, 8-13. http://www.inf.ethz.ch/personal/cagri.balkesen/pdf/iot_seminar_2008_proceedings.pdf [Stand 09.04.2014].

- Mangiameli, P., Chen, S.K. und West, D. (1996), A comparison of SOM neural network and hierarchical clustering methods, *European Journal of Operational Research*, 93(2), 402-417.
- McGill, R., Tuckey, J.W. und Larsen, W.A. (1978), Variations of Box Plots. *The American Statistician*, 32(1), 12-16.
- Meschede, D. (2010), *Gerthsen Physik*, Springer-Verlag, Berlin.
- Markoulidakis, J.G., Lyberopoulos, G.L., Tsirkas, D.F. und Sykas, E.D. (1997), Mobility modeling in third-generation mobile telecommunication systems, *IEEE Personal Communications*, 4(4), 41-56.
- Morawetz, M. (2015), "Keine aktuariell auskalkulierten Telematikprodukte", <http://versicherungswirtschaft-heute.de/koepfe/keine-aktuariell-auskalkulierten-telematikprodukte/> [Stand 12.08.2015].
- Nagel, K. und Schreckenber, M. (1992), A cellular automaton model for freeway traffic, *J. Phys. I France*, 2(12), 2221-2229.
- Nießen, G. (2015), Towers Watson: "Telematik-Tarife sind in fünf Jahren etabliert", <http://versicherungswirtschaft-heute.de/maerkte/towers-watson-telematik-tarife-sind-in-funf-jahren-etabliert/> [Stand 12.08.2015].
- Oberholzer, M. (2003), *Strategische Implikationen des Ubiquitous Computing für das Nichtleben-Geschäft im Privatkundensegment der Assekuranz*, VVW, Karlsruhe.
- Paefgen, J., Staake, T. und Thiesse, F. (2013), Evaluation and aggregation of pay-as-you-drive insurance rate factors: A classification analysis approach, *Decision Support Systems*, 56, 192-201.
- Pletziger, R. (2012), Komposit: Technologieschub als Bedrohung, *Versicherungswirtschaft*, 67(12), 870-872.
- Schüller, H. (2010): Modelle zur Beschreibung des Geschwindigkeitsverhaltens auf Stadtstraßen und dessen Auswirkungen auf die Verkehrssicherheit auf Grundlage der Straßengestaltung, <http://nbn-resolving.de/urn:nbn:de:bsz:14-qucosa-61493> [Stand 27.01.2015].

- Sparkassen Direktversicherung (2014), Berechnungsregeln, https://www.sparkassen-direkt.de/fileadmin/images/divers/Telematik/2014-08-08_score-logik.jpg [Stand 12.08.2015].
- Statistisches Bundesamt (2014): *Verkehr – Verkehrsunfälle 2013*, Fachserie 8 Reihe 7, Wiesbaden.
- Toledo, T., Musicant, O. und Lotan, T. (2008), In-vehicle data recorders for monitoring and feedback on driver's behavior, *Transportation Research, Part C* 16(3), 320-331.
- Towers Watson (2013), Towers Watson-Studie zur Kfz-Versicherung – Mehrheit der Europäer findet Telematik-Tarife attraktiv, <http://www.towerswatson.com/de-DE/Press/2013/09/Towers-Watson-Studie-zur-Kfz-Versicherung> [Stand 27.03.2015].
- Ward, J.H. (1963), Hierarchical Grouping to Optimize an Objective Function, *Journal of the American Statistical Association*, 58(301), 236-244.
- Weidner, W. und Weidner, R. (2014), Identifikation neuer Ansätze zur individuellen Kfz-Tarifierung, *ZVersWiss*, 103(2), 167-193.
- Werner, M. (2012), *Digitale Signalverarbeitung mit MATLAB®*, Vieweg+Teubner Verlag, Wiesbaden, 5. Auflage.
- Zanema, J., van Amelsfort, D., Bliemer, M. und Bovy, P. (2008), Pay-as-you-drive strategies: Case study of safety and accessibility effects, *J. Trans. Res. Board*, 2078, 8-16.
- Zonoozi, M.M. und Dassanayake, P. (1997), User Mobility Modeling and Characterization of Mobility Patterns, *IEEE Journal on Selected Areas in Communications*, 15(7), 1239-1252.

Leistungsmaßzahlen für die Straßenerneuerung

VON CRAIG RICHMOND; CLEMENS KIELHAUSER UND BRYAN T. ADEY,
ZÜRICH

1. EINLEITUNG

Im folgenden Beitrag werden Methoden untersucht, wie Leistungen des Erneuerungsmanagements von Straßennetzen so gemessen werden können, dass die resultierenden Maßzahlen zwischen Verwaltungseinheiten vergleichbar werden. Im Unterschied zu anderen Ansätzen mit ähnlicher Zielsetzung (Cook, Kazakov, & Roll, 1993; Deller, Chicoine, & Walzer, 1988; Kalb, 2009; Rouse, Putterill, & Ryan, 1997) wird hier ein „bottom-up“ Ansatz gewählt. Unter Zuhilfenahme physischer Verhaltensfunktionen werden die Leistungsmaßzahlen der Straßennetzwerkverwaltung (SNV) basierend auf extensiven Datensätzen von Merkmalen der einzelnen Straßenabschnitte errechnet. Der Ansatz ist in vielerlei Hinsicht mit der Erneuerungsplanung in modernen Pavement Management Systemen (PMS) verwandt. Soweit den Verfassern bekannt, ist die Anwendung dieser Berechnungsmöglichkeiten für einen Leistungsvergleich neu.

Insbesondere werden Maßzahlen gesucht, die Antworten auf zwei Fragen liefern: ob das Netzwerk ausreichend unterhalten wird und ob das Netzwerk effizient erhalten wird. Das Ziel ist eher pragmatisch, da Methoden gesucht werden, die mit heute zur Verfügung stehenden Daten gerechnet werden können. Die Schweizerischen Kantonsstraßennetze, die den zwischenörtlichen Verkehr sicherstellen, werden als Kontext verwendet. Dafür werden hier mögliche Maßzahlen vorgeschlagen und beispielhaft an Hand einer Fallstudie mit fünf Kantonen berechnet. Darüber hinaus wird eine Reihe zusätzlicher Maßzahlen vorgeschlagen, die ebenfalls berechnet werden sollten, um ungewollte Anreize zu bekämpfen, die dann entstehen, wenn nur einzelne Aspekte einer komplexen Leistung gemessen werden.

Leistungsmaßzahlen werden insbesondere benötigt, da die Bereitstellung der meisten Straßinfrastrukturdienste ein öffentliches Monopol ist und somit in Bezug auf die Effizienz

Anschriften der Verfasser:

Dr. Craig Richmond
ETH Zürich
Insitut für Bau- und Infra-
strukturmanagement (IBI)
Stefano-Frascini-Platz 5
8093 Zürich

Clemens Kielhauser
ETH Zürich
Insitut für Bau- und Infra-
strukturmanagement (IBI)
Stefano-Frascini-Platz 5
8093 Zürich

Prof. Dr. Bryan T. Adey
ETH Zürich
Insitut für Bau- und Infra-
strukturmanagement (IBI)
Stefano-Frascini-Platz 5
8093 Zürich

richmond@ibi.baug.ethz.ch

kielhauser@ibi.baug.ethz.ch

adey@ibi.baug.ethz.ch

die disziplinierende Wirkung des Marktes fehlt. Die Öffentlichkeit bestimmt über das Mengengerüst und Qualität der Dienste, was entsprechende Maßzahlen erfordert, um festzustellen, ob die Ansprüche an Menge und Qualität ausreichend erfüllt werden. Des Weiteren sind vergleichbare Maßzahlen auch für die Straßeninfrastrukturverwaltung selbst von großem Nutzen, um die eigene Leistung und/oder Organisation besser zu kennen und zu steuern.

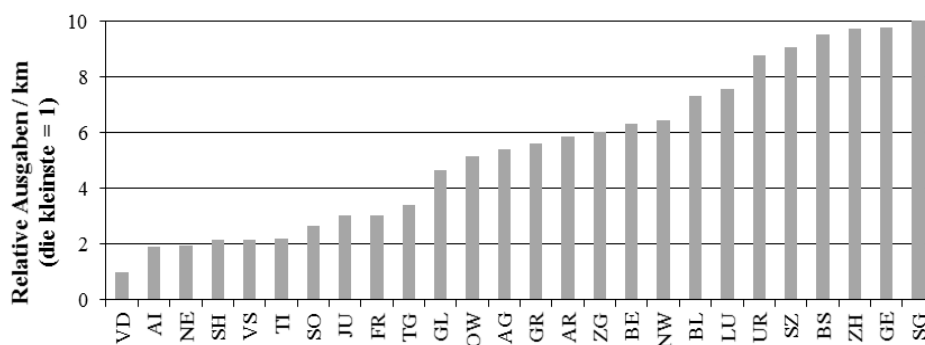
Obwohl eine SNV für viele verschiedene Aufgaben verantwortlich ist, liegt der Fokus dieses Beitrags jedoch fast ausschließlich auf Leistungen im Bereich Erneuerung, womit hier die periodische gänzliche Ersetzung der Deckschicht, oder je nachdem auch die darunterliegenden Schichten, über einen längeren Straßenabschnitt gemeint ist. Kleinere Maßnahmen, die zu einer Verlängerung der Dienstleistungszeit führen, werden je nach Systematik als Reparaturen, Instandhaltungs- oder Instandsetzungsmaßnahmen bezeichnet. In diesem Beitrag werden diese im Allgemeinen der einfacheren Lesbarkeit wegen unter dem Begriff „Instandsetzung“ zusammengefasst. Auf einzelne spezifische Kategorien wird nur dann eingegangen, wenn ein Bezug zu spezifischen externen Datenreihen herzustellen ist. Empirisch lässt sich bei den Ausgabendaten keine kategorische Trennung zwischen Erneuerungsmaßnahmen und Instandsetzungsmaßnahmen durchführen. In den schweizerischen Statistiken, sowie in den dort gültigen Normen (VSS Norm 640900a) wird zwischen den Begriffen „Betrieblicher Unterhalt“, „Baulicher Unterhalt“, „Ausbau und Verbesserungen“, sowie „Neubau“ unterschieden. Diese Begriffstrennung ist für die vorliegende Arbeit von erheblicher Bedeutung, da die empirische Modellierung, der die Kosten-Effizienzmaßzahl – eine der hier vorgestellten Maßzahlen – unterliegt, sich dieser Begriffstrennung anzupassen hat. Für die anderen vorgestellten Maßzahlen, die ohne Kostendaten auskommen, besteht diese Einschränkung nicht und dort können Maßzahlen vorgestellt werden, die tatsächlich nur einzelne Maßnahmentypen betreffen, wie zum Beispiel die Erneuerungsmaßnahmen. Als empirische Annäherung an die Erneuerung wird für die Kosten-Effizienzmaßzahl die Kategorie Baulicher Unterhalt in den Ausgabendaten verwendet.

Aus theoretischer Sicht bietet die Einschränkung auf Erneuerungsmaßnahmen gewisse Vorteile. Zum einen ist die empirische Identifikation der Erneuerungsmaßnahmen relativ einfach. Die Erneuerung findet in geographisch klar abgegrenzten Projekten statt, bewirkt eine eindeutige Verbesserung des Straßenzustands und wird in der Schweiz fast ausschließlich durch Drittfirmen durchgeführt, was klar ersichtliche Einträge in den Finanzdatenstämmen zur Folge hat. Deshalb sind diese Maßnahmen sowohl in den physischen Daten als auch in den Finanzdaten eher erkennbar, was bei betrieblichem Unterhalt, Instandhaltung oder Instandsetzung nicht der Fall ist. Zum anderen ist bei Erneuerungsmaßnahmen eine relative endogene Bestimmung des Mengengerüsts möglich, was bei Verbesserungen, Ausbau oder Neubau nicht gegeben ist. Dort gibt es viele externe beeinflussende Faktoren, wie z.B. das wirtschaftliche Wachstum. Das Mengengerüst an benötigten Erneuerungsmaßnahmen wird in erster Linie durch das vorhandene Netzwerk bestimmt. Die Modellierbarkeit der geforderten Menge an Erneuerung stellt für die Konstruktion von vergleichbaren Maßzahlen einen großen Vorteil dar.

Dass die Vergleichbarkeit eine nicht zu unterschätzende Herausforderung darstellt, zeigt sich eindeutig an Hand der Ausgabendaten der Schweizerische Straßenrechnung in Abbildung 1. Die Schweiz sammelt diese Daten systematisch seit 1913 auf Gemeinde- und Kantonebene, und – etwas verspätet – auch auf nationaler Ebene (Swiss Federal Statistical Office, 2014). Ein zwar einfaches, aber doch naheliegendes Konzept für eine vergleichende Maßzahl sind die Ausgaben pro Kilometer Straße. Beispiele von Anwendungen dieser Maßzahl findet man in BAKBASEL (2013) bezüglich der Effizienz oder Koch & Forster (2010) bezüglich der Hinlänglichkeit. Da beide Verfasser gute Kenner der schweizerischen Verhältnisse sind, darf man daraus schließen, dass bessere Maßzahlen für die Schweiz nicht zu Verfügung stehen. Damit tritt das Ausmaß an fehlender Vergleichbarkeit deutlich zu Tage (wie in Abbildung 1 gezeigt) und kann als Ausgangslage angesehen werden.

Abbildung 1 zeigt die Ausgaben pro Kilometer Länge als 15-jährigen Durchschnitt für die Kategorie des baulichen Unterhalts der Straßen und Kunstbauten für die 26 Kantone. Überraschenderweise scheint es, dass St. Gallen (SG) etwa zehnmal so viel pro Kilometer ausgibt wie der Kanton Waadt (VD). Nicht weniger überraschend sind die Vergleiche zwischen den benachbarten Kantonen Solothurn (SO) und Basel-Landschaft (BL) mit einem Verhältnis von 1:4 oder zwischen Solothurn (SO) und Aargau (AG) mit fast 1:3. Natürlich ist es kaum zu glauben, dass die Effizienz der Kantone derart stark voneinander abweichen könnte, und das Gegenteil wird von den Verfassern vermutet. Dennoch belegt die Grafik, dass große Differenzen in der Effizienz unbemerkt existieren können, woraus abgeleitet wird, dass bessere Messmethoden entwickelt werden sollten.

Abbildung 1: Die Vergleichbarkeit von bloßen Ausgaben pro Kilometer scheint nicht gegeben zu sein: Index der jährlichen kantonalen Ausgaben für baulichen Unterhalt pro km; 1994 - 2011¹



Quelle: Eigene Berechnung aus der Schweizerischen Straßenrechnung des BFS, inflationsangepasst mit dem LIK

Eine Liste von möglichen Erklärungen für die Unterschiede in Abbildung 1 würde Einflussfaktoren aufführen, die in dieser Maßzahl nicht berücksichtigt werden: Kunstbauten müssen von Straßen getrennt behandelt werden; Erneuerungskosten fallen nicht proportional zur Länge sondern zur Fläche an; Innerortsstraßen haben Entwässerungsanlagen, Beleuchtungen, Randsteine und Gehwege; auf manchen Untergründen ist das Bauen schwieriger als auf anderen; steile Gegenden verlangen viele Kunstbauten; usw. Netzwerke, die von diesen Schwierigkeiten besonders betroffen sind, müssten daher höhere Ausgaben haben. Des Weiteren müsste auf die vielen Detailfragen hingewiesen werden, die hinter den Ausgabenaggregaten verborgen sind, zum Beispiel: Was ist genau mit baulichem Unterhalt gemeint? Wie wird mit Straßenverbreiterungen umgegangen? Wird die Teilung von Ausgaben zwischen Gemeinden und Kantonen überall gleich behandelt?

Etwas tiefer verborgen sind verzehrende zeitliche Aspekte, die zu Mehr- oder Minderausgaben führen können. Straßen sind langlebig. Die erwartete Lebensdauer der Deckschicht kann durchaus 25 Jahre betragen, und die der darunter liegenden Schichten auch einiges mehr (Gunsteren, Kieliger, Traber, Nellen, & Zahnd, 2005). Daher sind sogar bei einer Betrachtungslänge von 15 Jahren die Ausgangs- und Endbedingungen nicht vernachlässigbar. Übernimmt eine SNV beim Antritt ein im Schnitt älteres Netzwerk, wird dies über mehrere Jahre höhere Ausgaben erfordern, um das erforderliche Dienstleistungsniveau zu gewährleisten, als dies bei einem im Schnitt jüngeren Netzwerk der Fall wäre. Dies trifft umso mehr bei kürzeren Betrachtungsperioden zu, die für eine typische Leistungsbesmes-

¹ Eine sehr breite Streuung in Ausgaben pro Kilometer lässt sich mit österreichischen oder US-amerikanischen Daten ebenfalls darlegen. Beide Länder publizieren die benötigten Grunddaten.

sung relevant sind. Je nach politischen oder finanziellen Umständen kann eine Behörde mehr oder weniger für die Straßenerneuerung budgetieren. Es fragt sich dann, welche Anfangs- und Endbedingungen zu erfassen sind und welche kompensierenden Berechnungen angestellt werden sollen, damit am Ende die Maßzahlen vergleichbar werden.

Die vorhergehenden Ausführungen belegen, dass die Vergleichbarkeit von Maßzahlen über die Erneuerung nicht einfach herbeizuführen ist. Es fehlt Literatur darüber, wie vergleichbare Maßzahlen für die mengenmäßig ausreichende und kostenmäßig effiziente Erneuerung aus detaillierten Straßendaten eben „bottom-up“ gerechnet werden können, wie dies für Benchmarking benötigt wird. Das Ziel dieses Beitrags ist daher, Wege aufzuzeigen, wie diese Lücke ansatzweise gefüllt werden kann.

Vorab sollten zwei Begriffe geklärt werden, und zwar was unter „ausreichend“ und was unter „Leistung“ zu verstehen ist. „Ausreichend“ ist ein relativer Begriff und es muss als erstes die Zielsetzung, auf die er bezogen ist, bestimmt werden. Anstatt eine fixierte Zielsetzung vorzuschreiben, soll die Maßzahl so formuliert sein, dass die Zielsetzung offen bleibt und als unabhängige Variable in der Formel auftritt. Vergleiche zwischen mehreren SNVen lassen sich dann – je nach Zweck – relativ zu individuellen Zielen oder relativ zu einer postulierten gemeinsamen Zielsetzung durchführen.

Der Begriff ‚Leistung‘ ist in Bezug auf die Straßenerneuerung zweideutig. Sie kann entweder als physischer Output oder als Outcome/Wirkung verstanden werden, ungefähr wie in der Literatur zur Performance Messung zwischen „Output“ und „Outcome“ unterschieden wird (Rouse & Putterill, 2008) oder bei der Balanced Scorecard Literatur zwischen „Performance Driver“ und „Outcome Measure“ (Kaplan & Norton, 1996) unterschieden wird. Wenn es um Leistung als physischen Output geht, muss die Leistung in physischen Maßeinheiten gemessen werden. Zum Beispiel kann die Leistung einer Maßnahme mit der Anzahl Quadratmeter von Maßnahmentyp k auf Straßenaufbautyp j bemessen werden. Ein Nachteil dieser Definition zeigt sich deutlich in dem Fall, wenn der besagte Belagseinbau bei einer intakten Oberfläche gemacht wurde, nur um an den darunterliegenden Leitungen zu arbeiten. Ein Erneuerungs-Output wurde geleistet, aber kaum eine Erneuerungswirkung. Dieselbe Problematik zeigt sich weniger deutlich, wenn Erneuerungsmaßnahmen bei unterschiedlichen Belagszuständen ausgelöst werden. Interessiert man sich hingegen für die Wirkung, also die bewirkte Verbesserung der Straßensubstanz, dann muss die Leistung in der Menge der bewirkten Zustandsverbesserung gemessen werden². Die Unterscheidung zwischen physischem Output und Wirkung ist von besonderer Wichtigkeit, falls die Maßzahlen als Teil eines Anreizsystems fungieren (siehe Kerr, 1975).

² Genauer gesagt, die Erneuerung bewirkt das zur Verfügung stellen eines Vektors an potentiellen Dienstleistungen, die je nach Gebrauch abgenutzt werden. Überlässt man die Strasse allein dem Wetter und den Fussgängern, hält ihre Dienstleistung vielleicht 80 Jahre an. Überlässt man die Strasse jedoch schweren Lastwagen, dauert die Dienstleistung vielleicht nur 12 Jahre. Idealerweise würde man dieses „zur Verfügung-stellen“ messen. Leider ist das nicht direkt möglich. Der Strassenzustand eines bestimmten Strassentyps ist ein approximatives Mass für die potentielle Benutzung.

Der verbleibende Teil dieses Beitrag ist folgendermaßen strukturiert: In Kapitel 2 werden Methoden diskutiert, die für die Bemessung von ausreichender Erneuerung angewendet werden können. Kapitel 3 behandelt die Frage nach effizienter Erneuerung. Schlussfolgerungen sind in Kapitel 4 aufgeführt.

2. Maßzahlen für die Feststellung, ob die Strassen ausreichend erhalten werden

Grundsätzlich gibt es zwei Vorgehensweisen, um zu prüfen, ob eine dauerhafte Infrastruktur ausreichend erhalten wird. Die Unterscheidung folgt im Allgemeinen der Abgrenzung zwischen Output und Wirkung. Die zwei in Frage kommenden Maßzahlen sind:

- i. Der durchschnittliche Netzwerkzustand oder dessen Veränderungsrate
- ii. Die Maßnahmendurchführungsrate

Neben einer kurzen Diskussion der Vorteile und Nachteile in Bezug auf die Vergleichbarkeit dieser zwei Haupt-Maßzahlen werden in diesem Kapitel auch zusätzliche Maßzahlen vorgestellt, die ungewollte Anreize bekämpfen sollen, welche dann entstehen, wenn die oben genannten Maßzahlen im Betrieb als Leistungsmaß benutzt werden. Wenn zum Beispiel die Anzahl eingebauter Quadratmeter als Maß der Zielerreichung bestimmt würde, entstähe ein Anreiz, neuen Belag auch dort einzubauen, wo es nicht unbedingt angezeigt wäre.

2.1 Der durchschnittliche Zustand

Der durchschnittliche Netzwerkzustand fasst die Wirkung der vergangenen Maßnahmen gegenüber den vorhandenen Einflussfaktoren zusammen und ist deshalb ein guter Indikator dafür, ob in der Vergangenheit ausreichend viel für die Erhaltung unternommen wurde, sofern unter „Vergangenheit“ eine genügend lange Periode verstanden wird, wie z.B. 5 Jahre oder länger. Sind kürzere Perioden von Interesse, spiegelt der Durchschnittswert des heutigen Tages die jüngsten Leistungen der SNV nur bedingt wider, da Straßen langlebig sind. Obgleich die Berechnung selbst trivial und die Aussage recht eindeutig ist wird diese Maßzahl selten publiziert. Eine Ausnahme findet sich z.B. in den Berichten des Österreichischen Rechnungshofes über die österreichischen Autobahnen, in denen der durchschnittliche Zustand im historischen Verlauf gezeigt wird (siehe Der Rechnungshof, 2012, Abbildung 4). Statt eines Durchschnittswerts sieht man dagegen öfter eine Einteilung der Netzwerkanteile in Zustandskategorien, wie bei Koch & Forster (2010). Diese Darstellungsweise ist jedoch weniger eindeutig.

2.1.1 Vor- und Nachteile

Sofern es eine Zielvorgabe für den durchschnittlichen Zustand gibt, ist die Schlussfolgerung bezüglich einer ausreichenden Erhaltung bei einem Vergleich eindeutig. Von Vorteil

sind die einfache Berechnungsart und die vielen Möglichkeiten, über diverse Gewichtungsschemen spezielle Merkmale eines Erhaltungsprogramms hervorzuheben. Eine Gewichtung nach Verkehrsaufkommen würde beispielsweise eine Benutzersicht betonen; eine Gewichtung nach unfallbezogenen Risikofaktoren würde eine gesellschaftliche Perspektive betonen. Identische Zielwerte über verschiedene SNVen sind keine zwingende Voraussetzung für eine Vergleichbarkeit, da auch die Differenz zum jeweiligen Zielwert verglichen werden kann. Ebenfalls muss als Vorteil gewertet werden, dass es sich beim Straßenzustand bereits vom Konzept her um einen vergleichbaren Parameter handelt. Ein Wert von 3,0 sollte für eine Strasse durch ein Wohngebiet dasselbe bedeuten wie für eine Autobahn.

Hingegen ist das Fehlen eines objektiven Kriteriums, wie zum Beispiel ein Zielwert für den durchschnittlichen Zustand ausgewählt werden soll, als Nachteil zu werten. Je nach Wohlstand, Beschwerlichkeit der lokalen Bedingungen bezüglich des Straßenunterhalts, Risiko oder der Lage der öffentlichen Finanzen sind unterschiedliche Zielwerte zu erwarten. Wenn alle anderen Einflussfaktoren gleich gehalten werden, hängt der durchschnittliche Zustand von dem „Auslösezustand“ ab, nach dem gemäß der Erhaltungsstrategie eine Erneuerung ausgelöst wird. Ist der Auslösezustand niedriger (d.h. in unserem Fall besser), dann werden die Straßen öfter erneuert, was die Gesamterneuerungskosten asymptotisch ansteigen lässt³. Da es auch unter optimalen Bedingungen relativ lang dauern kann, bis sich ein neuer Durchschnittszustand eingependelt hat, ist es ohne Prozessmodelle nicht einfach, einen stetig zu haltenden Durchschnittszustand mit einem Ausgabeniveau zu verknüpfen. Anhaltspunkte bieten Erfahrungen aus dem eigenen oder aus benachbarten Netzwerken.

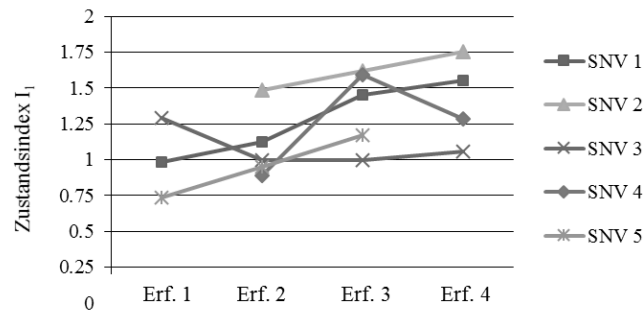
Abbildung 2 zeigt ein vergleichendes Beispiel dieser Maßzahl aus der Fallstudie. Aufgrund der Skala von I_1 ⁴, bedeutet ein Aufwärtstrend eine Verschlechterung, was in vier von fünf SNVen der Fall ist. Die Skala bezeichnet 0-1 als „gut“ und 1-2 als „mittel“, was zu dem Schluss führt, dass diese Netzwerke im Schnitt einen „guten“ bis „mittleren“ Zustand aufweisen, wobei bedeutende Unterschiede bestehen. Eine wichtige Schlussfolgerung ist, dass für das Beibehalten eines stetigen Zustands diese Netze in diesem Betrachtungszeitraum im Durchschnitt nicht ausreichend unterhalten wurden. Bemerkenswert ist, wie aussagekräftig diese einfache Maßzahl ist, wenn sie über die Zeit und über mehrere SNVen vergleichend

³ Wenn die Erneuerung nach T Jahren erfolgt, ist die erwartete Fläche, die jedes Jahr bei einem stetigen Zustand erneuert werden muss, proportional zu $1/T$.

⁴ In der Schweiz werden für die Messung des Straßenzustands sechs Indizes verwendet. Diese werden als I_0 bis I_5 bezeichnet. I_2 bis I_5 berücksichtigen spezifische physikalische Merkmale wie zum Beispiel die Quer- und Längsebenheit. Der Index I_1 charakterisiert die Gesamtheit der Oberflächenschäden unter Einschluss der visuell geschätzten Spurrinnentiefe, während I_0 dieses Schadensmerkmal nicht enthält. (VSS 640925b). Für alle Indizes beträgt der beste Wert 0 und der schlechteste 5. Die meisten der in diesem Artikel präsentierten Maßzahlen können von jedem Index aus berechnet werden. I_1 wurde hier primär verwendet, da dafür die größte Anzahl Beobachtungen im Datensatz vorlagen.

dargestellt ist - selbst, wenn bekannt ist, dass bei den Erhebungen nicht durchgehend die gleichen Methoden benutzt wurden⁵.

Abbildung 2 Der durchschnittliche Netzwerkzustand in Zeitverlauf, Fallbeispiel mit Fünf Schweizerischen Kantonen



Quelle: Richmond, Kielhauser, and Adey (2013); zwischen den einzelnen Zustandserfassungen liegen 2 bis 3 Jahre.

2.1.2 Zusatzmasszahlen aufbauend auf Gewichtungsschemen

Ein gegebener durchschnittlicher Straßenzustand kann von vielen verschiedenen Verteilungen stammen. Zum Beispiel kann sich hinter einem akzeptablen Durchschnitt eine nicht akzeptable bimodale Verteilung verstecken. Es kann auch nicht allgemeingültig bestimmt werden, welche Verteilung der Zustände eines Straßenabschnitts die richtige ist. Die Verteilung variiert je nach der verfolgten Erhaltungsstrategie. Abhilfe gegenüber ungewollten Verteilungen bei einem gleichbleibenden Durchschnitt können Gewichtungsschemen, wie in Gleichung (1) schematisch dargestellt, liefern. Zum Beispiel kann man die Abschnittsflächen mit nichtlinearen Nutzerkosten gewichten, damit sehr schlechte Zustände überproportional in den Durchschnittswert eingehen. Auf diese Weise vermittelt man einen Anreiz, die schlechtesten Straßen zuerst zu erneuern. Dasselbe Prinzip lässt sich für andere Zwecke analog anwenden.

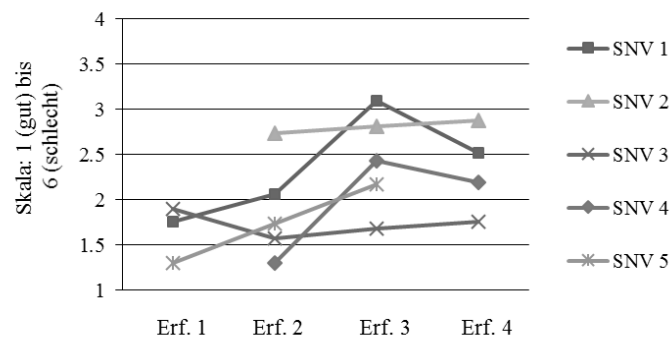
$$Masszahl = \frac{\sum_i w_i A_i}{\sum_i A_i} \quad (1)$$

Abbildung 3 wurde mit der gleichen Datengrundlage wie Abbildung 2 erstellt. Aber um ein an den „Fahrentutzen“ angelehntes Schema zu erzeugen, wird der Straßenzustand in dritter

⁵ SNV 4 benutzte vor und nach Erfassung eine visuelle Erhebungsmethode, aber bei Erfassung eine maschinelle Methode, was den Ausreißer in jenem Jahr vermutlich erklärt.

statt erster Potenz in den Gewichten eingesetzt: $w_i = 7 - (27 - \text{Zustand}^3) / 4.5$. Unter dieser Gewichtung ist es von Vorteil, Straßen mit schlechtem Zustand prioritär zu erneuern. Noch näher am tatsächlichen Fahrernutzen wäre eine zusätzliche Gewichtung nach Verkehrsaufkommen. Im konkreten Fall bleibt der Gesamteindruck wie in Abbildung 2 unverändert, außer dass die Werte auseinanderzogen wurden: Der durchschnittliche Zustand wird in vier von fünf Fällen schlechter; es gibt bedeutende Unterschiede zwischen den Teilnehmern; aber die Straßenzustände sind im Allgemeinen gut. Im Detail sieht man einzelne Unterschiede, etwa bei der SNV 1 im dritten Jahr oder bei der SNV 3, die unter dieser Berechnungsmethode eine Verschlechterung des Zustands nach Erfassung 2 erlebt.

Abbildung 3: Nach Dem Fahrernutzen Gewichteter Durchschnitt der Zustände (I_1); Gewichtung des Zustands in dritter Potenz; Skaliert auf 1 bis 6⁶



Quelle: Eigene Berechnungen; zwischen den einzelnen Zustandserfassungen liegen 2 bis 3 Jahre.

2.2 Die Maßnahmendurchführungsrate als Maßzahl

Der zweite Typus von Leistungsmaßzahlen für die ausreichende Erneuerung zielt nicht auf die geschaffene Wirkung der vergangenen Maßnahmen ab, sondern geht von der Menge der ausgeführten Maßnahmen während der laufenden Beobachtungsperiode aus: die Maßnahmendurchführungsrate. Es ist ein Maß des „Outputs“ statt des „Outcomes“. Auch hier stellt sich die Frage, welche Menge als ausreichend angesehen werden sollte.

Diese Maßzahlen sind mit den Prognosen des Erneuerungsbedarfs methodisch stark verwandt, welche z.B. mit Pavement Management Systemen (PMS) erstellt werden. In beiden

⁶ Die Gewichtungformel lautet $7 - (27 - I_1^3) / 4.5$ und wurde aus Praktikabilitätsgründen gewählt.

Fällen wird eine Quantität an benötigten Maßnahmen aufgrund von erwarteten Verfallsraten und einer Erhaltungsstrategie, d.h. „bottom-up“, gerechnet. Um die benötigten Finanzmittel zu berechnen, müssen die Kosten der möglichen Maßnahmen geschätzt und diese dann mit der Menge an Maßnahmen multipliziert werden (siehe Maerschalk, 2014 für eine methodische Übersicht). Danach lässt sich der so bestimmte Bedarf mit dem tatsächlichen Budget vergleichen, um die Frage zu beantworten, ob das Budget ausreichend ist. Im Unterschied dazu wird hier nur aufgrund der Menge sanierter Fläche gerechnet, damit alle empirischen Schwierigkeiten hinsichtlich der Kosten entfallen. Außerdem gehen Prognosen des Erneuerungsbedarfs eines PMS von den heutigen Netzzuständen aus, oder allenfalls vom heutigen Netzalter. Hier wird hingegen ein Erwartungswert berechnet, wie viel Quadratmeter im Schnitt pro Jahr erneuert werden müssten, um einem bestimmten Netzzustand stetig beizubehalten. Der heutige Zustand des Netzwerks wird dabei nicht berücksichtigt. Eine prinzipielle Verwandtschaft mit den sogenannten Duration-Berechnungen bei beispielsweise Newbery (1988) oder Small and Winston (1986) ist zu erkennen, obwohl anstelle von deren Annahme eines einheitlichen Netzwerks hier der Erwartungswert verwendet wird.

Anders als bei Newbery oder Small & Winston werden hier die erwarteten Verfallsraten abschnittsspezifisch gerechnet, d.h. in Anbetracht der lokalen Einflussfaktoren. Um auf Abschnitt j die Länge der erwarteten Erneuerungsperiode T_j bis zu einer Erneuerung zu berechnen, wird neben den Verfallsraten auch eine Erhaltungsstrategie inklusive kleineren Maßnahmen und deren Reset-Werten benötigt – so wie dies in einem PMS ebenfalls gemacht wird (vgl. hierzu Hinsch, Krause, Maerschalk, & Rübstein, 2005). Die erwartete lokale Erneuerungsrate ist dann $1/T_j$. Eine flächengewichtete Summe dieser Werte über das ganze Netzwerk dient als netzwerkspezifischer Vergleichswert (Benchmark) der zu erwartenden Quadratmeter erneuerter Fläche pro Jahr. Setzt man diese Summe wiederum ins Verhältnis zu der gesamten Netzwerkfläche, ergibt sich eine erwartete netzwerkweite Erneuerungsrate. Für das Netzwerk i lässt sich diese wie folgt anschreiben:

$$\text{Erwartete Erneuerungsrate}_i = \frac{\sum_j \frac{A_{ij}}{T_{ij}}}{\sum_j A_{ij}} = \sum_j \frac{A_{ij}}{A_i} \frac{1}{T_{ij}} \quad (2)$$

Der gleiche Quotient kann ebenfalls aus der empirisch geleisteten Menge an Erneuerungen pro Jahr gebildet werden. Bezeichnet man einen tatsächlich erneuerten Quadratmeter von Abschnitt j mit A'_j , ist die geleistete Erneuerungsrate für das Netzwerk i

$$\text{Geleistete Erneuerungsrate}_i = \frac{\sum_j A'_{ij}}{A_i} \quad (3)$$

Diese beiden Quotienten können miteinander verglichen werden. Die resultierende Differenz ist ein über Netzwerke hinweg vergleichbares Maß der langfristig ausreichenden Erneuerung. Formell lautet diese

$$\text{Maß der langfristig ausreichenden Erneuerung} = \sum_j \frac{A'_{ij}}{A_i} - \sum_j \frac{A_{ij}}{A_i} \frac{1}{T_{ij}} \quad (4)$$

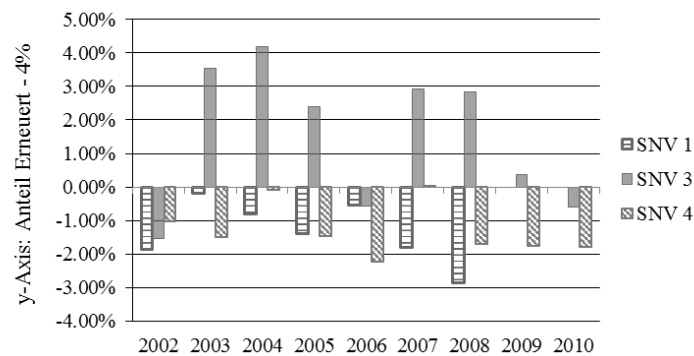
2.2.1 Vorteile und Nachteile

Ein erster Vorteil kann in der Genauigkeit der Messung gesehen werden. Die Menge an Maßnahmen nach Quadratmetern anstatt nach dem ausgegebenen Geld zu bemessen, repräsentiert das Objekt des Interesses deutlich besser: den Anteil des Netzwerks, der tatsächlich erneuert wurde. Bei einer Liste aller Erneuerungsmaßnahmen, die innerhalb eines gegebenen Zeitraums durchgeführt wurden, kann der erneuerte Anteil des Netzwerks relativ genau durch Aufsummierung der Quadratmeter an denen eine Maßnahme durchgeführt wurde bestimmt werden. Wenn man stattdessen die Kosten aller Maßnahmen addieren würde, und versuchen würde, davon den erneuerten Netzwerkanteil abzuleiten, würden die Aussagen deutlich ungenauer sein. Zum Beispiel würden sich sofort viele Detailfragen stellen, etwa wie belagsbezogene Kosten innerhalb der Projekte von anderen Kostentypen abgegrenzt wurden. Des Weiteren müssen weder die Preise noch die Qualitätsstandards von Ort zu Ort vergleichbar sein, wenn die Menge an Erneuerung in Quadratmetern gemessen wird. Schließlich will auch niemand die Leistung einer SNV anhand der ausgegebenen Mittel messen! Dies würde genau in die falsche Richtung führen.

Ebenfalls vorteilhaft ist die zeitnahe Messung. Straßenzustände verändern sich langsam und die Zustandsmessungen sind Unsicherheiten unterlegen. Dazu kommt, dass die Zustandserfassung selten jährlich getätigt wird. Es dauert deshalb mehrere Jahre, bis eine gesicherte Aussage über die ungenügende Erneuerung aufgrund des durchschnittlichen Straßenzustands gemacht werden kann. Demgegenüber ist die geleistete Erneuerungsrate relativ präzise und zeitnah. Unsicherheiten gibt es vor allem in der Berechnung der erwarteten Erneuerungsrate, die als Vergleichswert dient. Es gibt in der Literatur einige Ansätze, wie diese erwartete Erneuerungsrate zu berechnen ist, auf die hier verwiesen wird (Aguiar-Moya, Prozzi, & de Fortier Smit, 2011; Bardaka, Labi, & Haddock, 2014; Hong & Prozzi, 2010; Khattak, Nur, Bhuyan, & Gaspard, 2014; Mishalani & Madanat, 2002; Oeser, 2014; Scanziga, 2008; Weninger-Vycudil, Simanek, Rohringer, & Haberl, 2009). Letztendlich ist die Unsicherheit einer geschätzten erwarteten netzweiten Erneuerungsrate mit der Unsicherheit eines PMS-Maßnahmenprognosemodells vergleichbar.

Ein weiterer Vorteil des Maßzahlkonzepts ist, dass Differenzen zwischen Netzwerken hinsichtlich ihres Ausgesetztseins gegenüber Einflussfaktoren aufgefangen werden können. Aber auch unter der Annahme einer gemeinsamen erwarteten Erneuerungsrate, wie in Abbildung 4 gezeigt, gibt dieses schon einen deutlichen Eindruck darüber, ob zu wenig für die Erhaltung unternommen wird.

Abbildung 4: Der Anteil des Netzwerks, an dem im betrachteten Jahr die Deckschicht Erneuert wurde, abzüglich des Vergleichswerts von 4%



Quelle: Richmond et al. (2013)

In Abbildung 4 wurden anstatt netzwerkspezifischen erwarteten Erneuerungsrate für jede SNV eine erwartete Erneuerungsrate von 4% abgezogen, passend zu einer durchschnittlichen Lebensdauer der Deckschicht von 25 Jahren. Übereinstimmend mit den Verläufen der Zeitreihen in Abbildung 2 ist die Erneuerungsrate von SNV 3 eindeutig höher als die der SNV 1 oder 4. Auch mit dieser Maßzahl ist die Aussage bezüglich ausreichender Erneuerung eindeutig. Bei zwei von drei SNV müssten die Straßen im Schnitt mehr als 25 Jahre überdauern, damit bei den beobachteten Erneuerungsrate ein stetiger Zustand gehalten werden kann. Bei Einbeziehung der lokalen Verfallsraten sowie des extern bestimmten Auslösezustands für Erneuerungsmaßnahmen wäre die Aussagekraft noch höher.

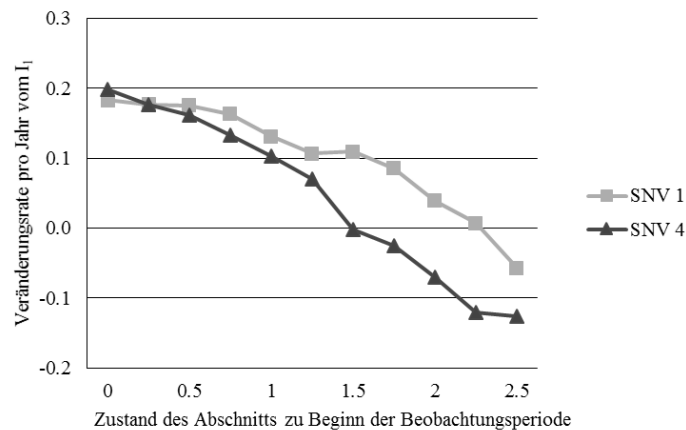
Es muss nicht zwingend optimal sein, dass in jedem Jahr die langjährig erwartete Erneuerungsrate erreicht wird. Budgetengpässe, außerordentlich schlechte Witterungsperioden oder vergangene Perioden der Unterinvestition können zu einem über- oder unterdurchschnittlichen Bedarf an Erneuerung führen. Aber über mehrere Jahre hinweg muss ein Zusammenhang zwischen der erfolgten Menge an erneuerten Quadratmetern und dem durchschnittlichen Zustand bestehen. Die Kombination von Abbildung 2 und Abbildung 4 bestätigt dies.

2.2.2 Zusatzmaßzahlen aufbauend auf altersspezifischen Verfallsraten

Die Verwendung der Erneuerungsrate als Maßzahl kann jedoch zu unerwünschten Anreizen führen. Zum Beispiel können Instandsetzungsmaßnahmen vernachlässigt werden, um mit den dadurch gesparten Ressourcen Erneuerungsmaßnahmen zu forcieren. Die erste von zwei Gruppen Zusatzmaßzahlen baut auf den unterschiedlichen durchschnittlichen Verfallsraten auf, die in verschiedenen Altersphasen der Straßen vorkommen. Die zweite basiert auf einem Vergleich der Verteilung der empirischen Zustände mit einer Modellverteilung derselben.

Das Ausmaß von Instandsetzungsarbeiten wie das Füllen von Rissen lässt sich nicht einfach quadrateinheitlich erfassen. Je nach Ausmaß des Schadens, muss mehr oder weniger intensiv gearbeitet werden, um den dahinterstehenden Schadensprozess einzudämmen. Als alternative Messstrategie kann die durchschnittliche Verfallsrate bei Abschnitten, die einen bestimmten Zustand bereits erreicht haben, einen Hinweis auf die bessere oder schlechtere Ausführung von Instandsetzungsarbeiten liefern. Ein Beispiel dieser Maßzahlen wird in Abbildung 5 gezeigt. Es wurden nur Straßenabschnitte in die Rechnung einbezogen, auf denen keine Erneuerungsmaßnahme stattfand. Die Verlangsamung der gemessenen jährlichen Zustandsveränderungsrate bei höheren (schlechteren) Ausgangszuständen zeigt die Wirkung von Instandsetzungsmaßnahmen. Die eigenen historischen Daten oder Daten benachbarter Netzwerke können als Vergleichswerte herangezogen werden um festzustellen, ob die Verfallsraten auffallend hoch sind, nachdem ein Zustand erreicht worden ist, bei dem vermehrte Instandsetzungen zu erwarten wären.

Abbildung 5: Durchschnittliche jährliche Veränderungsrate, anhand von Paaren aufeinanderfolgender ZustandsErfassungen (Bezeichnet nach dem Zustand im Ausgangsjahr)



Quelle: Eigene Berechnungen; dargestellt sind die Mittelwerte über mehrere Paare von Zustandserfassungen

2.2.3 Zusatzmaßzahlen aufbauend auf einer Modellverteilung der Zustände

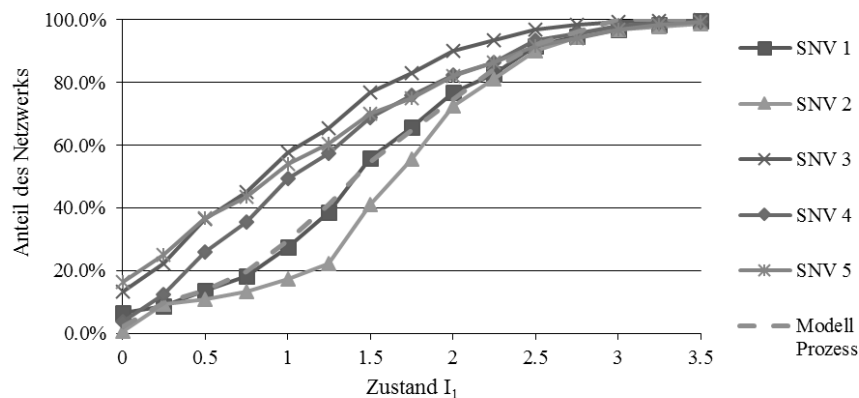
Die Summe der erneuerten Fläche sagt nichts darüber aus, ob jene Straßenabschnitte erneuert wurden, die dies am dringendsten benötigten. Die eine Hälfte der Straßen doppelt so oft zu erneuern und die andere Hälfte überhaupt nicht, resultiert zwar in der gleichen Summe, aber nicht in der gleichen Verteilung der Zustandswerte. Eine Reihe von Maßzahlen lassen sich mit Statistiken über die Verteilung der Zustände berechnen, um Probleme in der Erhaltungspraxis sichtbar zu machen. Obwohl Maßzahlen in manchen Fällen direkt zwischen einzelnen Netzwerken sinnvoll verglichen werden können, müssen jedoch - um breitere Vergleiche anzustellen - Differenzen zu lokal angepassten Vergleichswerten benutzt werden, die aus Modellverteilungen der Abschnittszustände bestimmt werden. Im Folgenden wird auf zwei Beispiele solcher Zusatzmaßzahlen eingegangen, es sind jedoch noch viele andere vorstellbar.

Die Modellverteilung wird wie folgt bestimmt: Würde eine SNV eine Erhaltungsstrategie definieren, könnte dies genau wie in einem PMS oder Life-Cycle Kostenmodell (siehe beispielsweise Ressel, Tejkl, & Klöpfer, 2013; Santos & Ferreira, 2013) verwendet werden, um eine zu erwartende Verteilung der Zustände zu berechnen. Für jeden Abschnitt wird ein Lebenszyklus in Anbetracht der Einflussfaktoren und Erhaltungsstrategie berechnet. Die Anzahl Jahre, in denen sich der Abschnitt in einem bestimmten Zustand befindet, definiert

– als Anteil der gesamten Lebenszykluszeit – eine Verteilung. Solche Verteilungen können flächengewichtet summiert und wie in Abbildung 6 als Modellprozess dargestellt werden.

Abbildung 6 vergleicht den prozentuellen Anteil der Gesamtfläche eines Netzwerks zwischen SNVn, der einen gegebenen Zustand nicht übersteigt. Neben den tatsächlichen Werten aus den fünf teilnehmenden SNVn beinhaltet die Abbildung auch eine Modellverteilung. Letztere entstammt einem idealisierten Netzwerk mit gleichmäßig verteilten Abschnittsaltern, angenommenen Verfallsraten und einer vordefinierten Reihe von Maßnahmen. Die gute Übereinstimmung mit SNV 1 ist zufällig. Wie die Abbildung zeigt, scheint es zwischen den Erhaltungsstrategien der dargestellten SNVn wichtige Unterschiede zu geben. Im Unterschied zu dem Durchschnittszustand oder der Erneuerungsrate ist hier eine Art Zeitprofil zu sehen. Aufgestauter Investitionsbedarf, wie vermutlich bei SNV 2 ersichtlich, oder zyklische Erneuerungen, wie vermutlich bei SNV 3 und 5, werden erkennbar.

Abbildung 6: Kumulativer Anteil des Netzwerks über einem bestimmten Zustandswert



Quelle: Richmond et al. (2013)

Abbildung 7 zeigt in zwei Matrizen, dass die Verteilungen der Zustände sowie die Verteilung ihrer Veränderung deutlich unterschiedlich zwischen den SNVn sein können. Die Zelle in Reihe i und Spalte j zeigt den Prozentsatz aller Abschnitte, die einen Ausgangszustand i und einem Endzustand j hatten. Größere Zahlen sind dunkler gefärbt. SNV 3 weist ein weniger breit gestreutes Bild auf, bei dem fast keine Abschnitte mit schlechterem Zustand als 3 vorliegen. SNV 1 hat Abschnitte bis zum Zustand 5 in der ersten Zustandserfassung. Es scheint allerdings, dass die meisten dieser Abschnitte im Folgenden eine Maßnahme erfuhren, sodass bei der zweiten Zustandsaufnahme nur noch sehr wenige Abschnitte einen Zustand von 5 aufwiesen. Gemäß Lehrbuch müsste die Erhaltungsstrategie durch Häufungen klar zu erkennen sein. Die horizontale Streuung der Werte rechts der Diagonalen erlaubt die Vermutung, dass Messfehler im Spiel sind, die Streuung nach links könnte

2.3 Vorläufige Zusammenfassung

Sowohl der durchschnittliche Zustand als auch die jährliche Erneuerungsrate sind vielversprechende Maßzahlen, um die Frage zu beantworten, ob ein Netzwerk ausreichend unterhalten wird. Um jedoch ungewollten Anreizen entgegenzuwirken, sind zusätzliche Maßzahlen empfehlenswert. Davon gibt es viele. Alle diese Maßzahlen haben die Gemeinsamkeit, auf rein physikalischen Quantitäten aufgebaut zu sein. Keine beinhaltet geschätzte finanzielle Angaben. Somit werden ein Hindernis der Vergleichbarkeit und eine große Quelle für Messunsicherheiten umgangen.

3. Maßzahlen für die Kosteneffizienz der Unterhaltsausgaben

Bei der Frage der Kosteneffizienz (KE) ist es nicht mehr möglich, auf Finanzdaten und die damit verbundenen Schwierigkeiten zu verzichten. Der Schlüssel zum Erfolg liegt, wie bei den schon vorgestellten Maßzahlen auch, in der Fähigkeit, die Kostenwirkung der lokalen Einflussfaktoren zu kompensieren, damit die Vergleiche als fair – oder zumindest fair genug – angesehen werden können. 100% Fairness ist bei so vielen verschiedenen die Kosten beeinflussenden Faktoren illusorisch. Der Einsatz einer KE-Maßzahl bedarf daher einer Kosten-Nutzen-Überlegung, bei der die Verfügbarkeit und Qualität der Daten eine entscheidende Rolle zu spielen haben. In diesem Sinne wird hier untersucht, inwiefern es in einem spezifischen Kontext möglich ist, eine Maßzahl der Kosteneffizienz aus einem gegebenen Datenstamm zu konstruieren. Die Schlussfolgerung im hier untersuchten spezifischen Fall fiel zwar eher ernüchternd aus, was jedoch nicht heißt, dass sie nicht anders sein könnte, wenn die Datenqualität verbessert werden könnte.

3.1 Definition der KE-Maßzahl

Wie eingangs erwähnt, ist der spezifische Kontext dieser Arbeit eine Fallstudie bestehend aus 5 Schweizerischen Kantonen. Einheitlich strukturierte jährliche Ausgaben mit einer Granularität bis zum baulichen Unterhalt der Straßen mit entsprechenden Abgrenzungen gegenüber Kunstbauten, Ausbau, und betrieblichem Unterhalt bildeten die Daten zu den Ausgaben. Die das Netzwerk beschreibenden Daten kamen aus den jeweiligen PMS-Systemen in einer Granularität von maximal 100 m langen Abschnitten. In jedem Fall lagen mehrere Zustandserfassungen vor, die aber nicht gleichzeitig oder in gleichen Intervallen stattfanden. Nicht überall waren Datenbestände über erfolgte Erneuerungsmaßnahmen vorhanden, was eine direkte Bemessung der Wirkung der Erneuerungen verhinderte. Deshalb ist das unten vorgeschlagene Messkonzept auf der netzwerkweiten Zustandsveränderung aufgebaut, was impliziert, dass die Wirkung von Instandsetzungsmaßnahmen mit denjenigen der Erneuerungsmaßnahmen vermischt sind. Anders ausgedrückt, die anfangs angestrebte Fokussierung auf reine Erneuerungsmaßnahmen, lässt sich hier empirisch nicht durchsetzen. Dies ist nicht kritisch sofern die Ausgabendaten ebenfalls entsprechend ausgeweitet werden, was mit dem Begriff des baulichen Unterhalts sowieso teilweise zutrifft.

Die Einflussfaktoren können mit einem Vektor \mathbf{z} angeschrieben werden. Um zu unterscheiden, wann eine Variable als netzwerkweiter Durchschnittswert zu verstehen ist, wird ein Großbuchstabe \mathbf{Z} benutzt. Kleine Buchstaben stehen für Variablen, die abschnittsweise variieren. Im Fallbeispiel fehlten zum Teil abschnittsspezifische Daten über die Einflussfaktoren. Insbesondere fehlten diese für die Regressionsanalyse, die in Sektion 3.2.2 beschrieben ist. Deshalb wurde in Übereinstimmung mit dieser Beschränkung eine lineare Form der Kosteneffizienz-Maßzahl (KE) wie in Gleichung 5 gewählt, die die Kosteneinwirkungen der Einflussfaktoren als gebietsweiter Durchschnittswert erlaubt.

$$KE_{i,\bar{t}} = \left(\mathbf{p}^T \bar{\mathbf{Z}} + \mathbf{p}^T \Delta \mathbf{Z}_{i,\bar{t}} \right) - \frac{TC_{i,\bar{t}}}{\Delta \hat{C}_{i,\bar{t}}} \quad (5)$$

Es gelten dabei folgende Definitionen:

- $TC_{i,\bar{t}}$: die Gesamtausgaben von SNV i in der Zeitperiode \bar{t} , die zwischen zwei Zustandserhebungen bei t_1 und t_2 getätigt wurden
- $\bar{\mathbf{Z}}$: der Durchschnittswert der Einflussfaktoren über alle teilnehmenden SNVen.
- $\Delta \mathbf{Z}_{i,\bar{t}}$: die Differenz zum $\bar{\mathbf{Z}}$ für SNV i
- \mathbf{p} : ein Vektor aus den erwarteten Einheitskosten pro m^2 für jeden Einflussfaktor
- $\Delta \hat{C}_{i,\bar{t}}$: die Anzahl Einheiten der Zustandsverbesserung in $\Delta C \times m^2$; das Caret (^) bezeichnet eine Schätzung

Der erste Term auf der rechten Seite von Gleichung (5) entspricht einem Vergleichswert für das Kostenniveau pro m^2 , der im jährlichen Durchschnitt für das gesamte Netzwerk von SNV i erwartet wird. Mit ausreichenden Daten über die geographische Verteilung der Einflussfaktoren kann ein Vergleichswert auch abschnittsweise berechnet werden. Darin können dann interaktive Kosteneffekte berücksichtigt werden, wie zum Beispiel aufwendige Bautypen bei starkem Verkehrsaufkommen und gleichzeitiger Innerortslage. In dem verwendeten Modell wird angenommen, dass die Auswirkungen voneinander unabhängig sind und in einem linearen Zusammenhang stehen.

Es wäre prinzipiell möglich, die Menge an erfolgter Zustandsverbesserung einer Erneuerungsmaßnahme durch eine direkte Messung vor und nach der Maßnahme zu ermitteln. Im Fallbeispiel standen solche Daten jedoch nicht zur Verfügung, und ein anderes Konzept für die Berechnung der Zustandsverbesserung musste gefunden werden. Die gemessene Zustandsveränderung auf Abschnitt j , $C_{j,t2} - C_{j,t1}$, ist das Nettoergebnis aus zwei Prozessen:

der Erneuerung und das Verfalls. Die gesuchte Bruttozustandsverbesserung ist die netto gemessene Verbesserung unter Einbeziehung des Verfalls. Das heißt, mit der Einrechnung einer geschätzten Verfallsrate, die – wie in einem PMS – die lokalen Einflussfaktoren berücksichtigt, kann von einer gemessenen Nettoverbesserung auf eine Bruttoverbesserung geschlossen werden. Da die Abschnitte, an denen Maßnahmen stattfanden, ebenfalls nicht eindeutig klar waren, musste die Berechnung der Bruttozustandsverbesserung für das ganze Netz berechnet werden.

Wenn die Menge an Zustandsverbesserung auf diese Art berechnet wird, dann beinhaltet das Resultat auch die bewirkte Zustandsverbesserung von Instandsetzungsmaßnahmen zusätzlich zu derjenigen der Erneuerungsmaßnahmen. Dies ist zwar an sich unerwünscht, passt aber erfreulicherweise zu der Kategorie Baulicher Unterhalt, die in den Ausgabendaten verwendet wurde.

Die geschätzte Zustandsverbesserung $\Delta \hat{C}_{i,t}$ ist deshalb wie folgt definiert:

$$\begin{aligned} \Delta \hat{C}_{i,t} &= -1 \sum_j A_{i,j,\bar{i}} \left((C_{j,t_2} - C_{j,t_1}) - \lambda_{j,\bar{i}}(\mathbf{z}_{j,\bar{i}})(t_2 - t_1) \right) \\ &= -A_{i,\bar{i}}(t_2 - t_1) \sum_j \frac{A_{i,j,\bar{i}}}{A_{i,\bar{i}}} \left(\frac{(C_{j,t_2} - C_{j,t_1})}{(t_2 - t_1)} - \lambda_{j,\bar{i}}(\mathbf{z}_{j,\bar{i}}) \right) \end{aligned} \quad (6)$$

Es gelten folgende Definitionen:

$A_{j,\bar{i}}$: die Fläche des Abschnitts j

$A_i = \sum_j A_{i,j}$: die Fläche von SNV i . Ansonsten wird auf das gemeinsame Anschreiben von i und j verzichtet.

C_{j,t_2} : der gemessene Zustand von Abschnitt j zum Zeitpunkt t_2

$\lambda_{j,\bar{i}}(\mathbf{z}_{j,\bar{i}})$: die durchschnittliche Verfallsrate (bis zur Erneuerung) als Funktion der Einflussfaktoren

$t_2 - t_1$: die Länge der Betrachtungsperiode zwischen t_1 und t_2

Die Einheiten an Verbesserung des Zustandes nehmen die Form einer flächengewichteten Summe der Zustandsverbesserungen an. Die geschätzte Zustandsverbesserung auf einem bestimmten Abschnitt besteht aus zwei Elementen, einem gemessenen Unterschied $C_{j,t_2} - C_{j,t_1}$ und einem erwarteten Verfall $\lambda_{j,\bar{i}}(\mathbf{z}_{j,\bar{i}})(t_2 - t_1)$.

Für die Funktion $\lambda(\mathbf{z})$ eignet sich jede Form der Verfallsfunktionen, die auch in PMS-Systemen eingesetzt werden (siehe beispielsweise Oeser, 2014; Scazziga, 2008; Weninger-Vycudil et al., 2009). Konkret ist λ die durchschnittliche jährliche Verfallsrate bis zu dem Zustandswert, bei dem gemäß der Erhaltungsstrategie die Erneuerungsmaßnahme ausgelöst wird. Im Fallbeispiel wurden in Anlehnung an die VSS NORM 640986 Lebenserwartungen bis zur Erneuerung je nach Straßentyp berechnet. Im Prinzip sind jedoch auch viel komplexere Vorgehen möglich. Weil bei einem Indexverlauf von 0 bis 5⁷ die Verbesserung als negative Veränderung herauskommt, wird die ganze Summe in (6) mit -1 multipliziert.

Es wird nicht behauptet, dass diese KE-Maßzahl die einzig mögliche ist oder dass sie bereits optimal wäre, wenn keine Datenbeschränkungen vorliegen. Die Maßzahl kann drei Mechanismen des Einwirkens von Einflussfaktoren widerspiegeln: solche, die über die Fläche wirken; solche, die über die Verfallsrate wirken; und solche, die über die Baukosten pro Meter wirken. Jede dieser Anpassungen kann auf der Granularität einzelner Straßenabschnitte stattfinden. Im Fallbeispiel wurden nur die Flächen und die erwarteten Verfallsraten auf dieser Granularitätsstufe umgesetzt.

Als weiterer Vorteil kann die Möglichkeit angesehen werden, unterschiedliche Periodenlängen zwischen den Zustandsaufnahmen korrekt in die Maßzahl einfließen zu lassen. Der zu erwartende Verfall wird dementsprechend angepasst. Außerdem setzen Vergleiche zwischen SNVen keine Gleichheit in Erneuerungsstrategien oder Budgetsituation voraus. Vergleichend wird die auf der Stufe der Ausgaben pro Einheit bewirkte Bruttoverbesserung. Dadurch wird die Abgrenzung gegenüber den Maßzahlen aus Kapitel 2 hervorgehoben. Die KE-Maßzahl gibt nur Auskunft über die Kosteneffizienz des Mitteleinsatzes, nicht aber über deren Höhe. Bei den Maßzahlen für die ausreichende Erhaltung war es umgekehrt.

3.2 Empirische Umsetzung der KE-Maßzahl

Die Variablen TC , A , \mathbf{z} , und C sind alle Beobachtungen. Die Verfallsrate λ resultiert aus \mathbf{z} bei einer angenommenen Verfallsfunktion und festgelegtem Auslösezustand. Nur der Vektor \mathbf{p} muss empirisch bestimmt werden, um die Maßzahl zu berechnen.

Es ist anzunehmen, dass je nach Anwendungsfall unterschiedliche Einflussfaktoren berücksichtigt werden. Im Fallbeispiel wurde neben der Literatur über Performance-Messung (Cook et al., 1993; Deller et al., 1988; Fallah-Fini, Triantis, de la Garza, & Seaver, 2012; Kalb, 2009; Rouse et al., 1997) auch eine Gruppe von Kantonsingenieuren konsultiert und über die wichtigsten Kosteneinflussfaktoren befragt. Neben Faktoren wie Verkehrsaufkommen und Klima, die gemäß der Bauingenieurliteratur mit Verfallsprozessen in Verbindung zu bringen sind, (Bardaka et al., 2014; Khattak et al., 2014; Oeser, 2014; Weninger-Vycudil et al., 2009), gab es Vorschläge, die wenig oder gar nichts mit Verfall zu tun ha-

⁷ Siehe Fußnote 5

ben, wie Einkommen, „Komplexe Urbane Umgebungen“ oder die Verminderung von Stau bei Baumaßnahmen. Es wurden allerdings auch Einflussfaktoren vorgeschlagen, bei denen eine Kostenrelevanz über mehrere Mechanismen vermutet werden kann, wie z.B. die Kriterien „innerorts“ oder „Steilheit“⁸. Die Einflussfaktoren, die in der Fallstudie zur Anwendung kamen, sind in Tabelle 1 aufgelistet. Es kamen nur solche Faktoren zur Anwendung, für die auch Daten für eine Regression gefunden werden konnten. Bei einer zukünftigen Arbeit wird davon ausgegangen, dass diese Liste erweitert werden kann.

Tabelle 1: Beschreibende Statistiken der Regressionsvariablen

Einflussfaktor	Empirische Um-	Einheiten	Durch-	Stand.	Max	Min
	setzung		schnitt	Abw.		
Erneuerungs-	Neubau und Er-	1000 CHF	34.21	25.81	250	3.07
kosten	neuerung / km					
Innerorts Umge-	Wohnfläche / km	ha	9.67	4.11	37.7	2.45
bung						
Komplexe Urbane	Einwohner /	Personen /	39.35	14.81	124	13.5
Umgebung	Wohnfläche	ha				
Lastwagen Ver-	Registrierte	Anzahl	13.48	9.08	77.5	2.27
kehr	Lastwagen / km	/km				
Einkommen	Bundessteuer /	CHF	1144	963	9474	421
	Einwohner					
Steile Umgebung	Durchschnittliche	Grad	5.72	4.94	27.7	0.14
	Geländeneigung					
Geologische	Index der Flächen	(0, 1)	0.029	0.087	0.50	0.00
Schwierigkeiten	mit schwierigen					
	Bodentypen					

Quelle: Richmond et al. (2013)

⁸ Als Beispiel: Innerorts. Der Straßenraum wird innerorts mit Leitungen geteilt, welche gelegentlich Eingriffe durch die Straßensubstanz hindurch benötigen, was zur beschleunigten Verfall führen kann (Chow & Troyan, 1999). Andererseits müssen großflächige Arbeiten an Leitungen in der optimalen Planung von Straßenerneuerungen miteinbezogen werden, was zu Verschiebungen des optimalen Zeitpunkts führen kann. Die daraus resultierenden Mehrkosten gegenüber dem optimalen Zeitpunkt gäbe es auch bei einer gleichbleibenden Verfallsrate.

3.2.1 Das Schätzen der durchschnittlichen Kostenbeeinflussung durch Einflussfaktoren

Mit entsprechenden Daten könnte man den Einheitskostenvektor \mathbf{p} als Parameter einer Kostenfunktion schätzen, wobei $KE_{i,\bar{t}}$ als Teil der Zufallsvariable $v_{i,\bar{t}}$ auftreten würde und $TC_{i,\bar{t}}$ zur abhängigen Variable wird.

$$\begin{aligned} TC_{i,\bar{t}} &= \left(\mathbf{p}^T \bar{\mathbf{Z}} + \mathbf{p}^T \Delta \mathbf{Z}_{i,\bar{t}} \right) \Delta \hat{C}_{i,\bar{t}} - \Delta \hat{C}_{i,\bar{t}} \left(\mathbf{z}_{i,\bar{t}} \right) KE_{i,\bar{t}} + \varepsilon_{i,\bar{t}} \\ &= \left(\mathbf{p}^T \bar{\mathbf{Z}} + \mathbf{p}^T \Delta \mathbf{Z}_{i,\bar{t}} \right) \Delta \hat{C}_{i,\bar{t}} \left(\mathbf{z}_{i,\bar{t}} \right) + v_{i,\bar{t}} \end{aligned} \quad (7)$$

Top-Down Leistungsvergleiche in der Stochastic Frontier Analysis (SFA) Gattung werden in dieser allgemeinen Form vorgenommen (Farsi, Filippini, & Greene, 2005; Greene, 2005; Kalb, 2009). Dafür standen jedoch in dem Fallbeispiel nur fünf Beobachtungen mit kompletten Datensätzen zur Verfügung und es musste daher auf einen größeren Datensatz ausgewichen werden. Ein weitaus größerer Datensatz von Ausgabendaten wurde für größere Gemeinden gefunden⁹. Hier fehlten jedoch die Netzwerk- und Zustandsdaten, aus denen $\Delta \hat{C}_{i,\bar{t}}$ berechnet wird.

Um diese Beschränkung zu umgehen, sind zwei vereinfachende Annahmen nötig. Erstens ist aus Gleichung (6) ersichtlich, dass bei einer ausreichend langen Zeitperiode die Unterschiede zwischen der Anfangs- und Endzustandsaufnahme, $(C_{j,t_2} - C_{j,t_1})$, eine untergeordnete Rolle gegenüber dem linear ansteigenden erwarteten Gesamtverfall $\lambda_{j,\bar{t}} \left(\mathbf{z}_{j,\bar{t}} \right) (t_2 - t_1)$ einnimmt. Dies ist umso mehr der Fall, wenn von dem durchschnittlichen Zustand für das Gesamtnetzwerk gesprochen wird und das Netzwerk als Ganzes in einem brauchbaren Zustand gehalten wird. Zweitens wird für die Regression vereinfachend angenommen, dass jede Gemeinde die gleiche durchschnittliche Verfallsrate aufweist¹⁰.

Gleichung (6) kann nach Anwendung der oben genannten Annahmen wie folgt geschrieben werden

$$\Delta \hat{C}_{i,t} = A_{i,\bar{t}} (t_2 - t_1) \bar{\lambda} \quad (8)$$

Somit kann Gleichung (7) umformuliert werden:

⁹ Schweizerische Gemeinden/Städte mit mehr als 5000 Einwohnern

¹⁰ Aufgrund der großen Variation in den geographischen, klimatischen und verkehrsbezogenen Einflussfaktoren in der Schweiz ist diese Annahme vermutlich nicht realistisch, aber - wie oben bereits erwähnt - fehlten für eine empirische Schätzung der unterschiedlichen Verfallsparameter die Netzwerkdaten.

$$\frac{TC_{i,\bar{t}}}{A_{i,\bar{t}}(t_2 - t_1)} = (\mathbf{p}^T \bar{\mathbf{Z}} + \mathbf{p}^T \Delta \mathbf{Z}_{i,\bar{t}}) \bar{\lambda} + v_{i,\bar{t}} \quad (9)$$

Es muss erwartet werden, dass die Zufallsvariable $v_{i,\bar{t}} = -\Delta \hat{C}_{i,\bar{t}}(\mathbf{z}_{i,\bar{t}}) KE_{i,\bar{t}} + \varepsilon_{i,\bar{t}}$ wegen des Faktors $\Delta \hat{C}_{i,\bar{t}}(\mathbf{z}_{i,\bar{t}})$ heteroskedastisch verteilt ist.

3.2.2 Beschreibung der Regressionsdaten

Die Daten über die Straßenausgaben der etwa 130 größten Gemeinden der Schweiz werden durch das Schweizerische Bundesamt für Statistik erstellt und durch den Schweizerischen Städteverband (SSV) jährlich publiziert. Drei achtjährige Perioden, anfangend im Jahr 1986, wurden gewählt, um die Problematik der fehlenden Zustandsveränderungsdaten zu reduzieren. Daraus ergaben sich 391 Beobachtungen. Zwei Dummy-Variablen erfassen eine zeitabhängige Veränderung für die Zeitperioden 1994-2001 und 2002-2009. Alle Daten in Schweizer Franken wurden mit dem Verbraucherpreisindex Basis 2010 für Inflation korrigiert.

Spezifische Bemerkungen

1. „Ausgaben pro Km_Jahr“: Der SSV publiziert die Straßenrechnung des BFS für die Kantone, allerdings nicht so detailliert wie das BFS selbst. Es gibt z.B. keine Unterscheidung zwischen Neubau, Verbesserung und baulichem Unterhalt. Diese Ausgaben pro km und Jahr stellen daher die abhängige Variable der Regressionsgleichung dar. Es kann deshalb davon ausgegangen werden, dass mit Neubau und Verbesserung korrelierte Einflussfaktoren einen positiven Bias in der geschätzten Wirkung erhalten. Zum Beispiel erscheint es plausibel, dass Einkommen mit Neubau positiv korreliert..
2. „Wohnfläche pro km“: Diese Variable repräsentiert den Anteil von innerorts/außerorts liegenden Straßen in der Gemeinde. Die Variable benutzt das Verhältnis von Hektar Wohngebiet zu Kilometern Straßennetzwerk. Angaben über das Wohngebiet sind durch das Bundesamt für Landestopographie (swisstopo) tabellarisch verfügbar und werden ca. alle 12 Jahre neu erfasst. Die Länge des Straßennetzwerks wird durch den SSV publiziert.
3. „Einwohnerdichte im Wohngebiet“: Unter Verwendung des gleichen Maßes für das Wohngebiet wie oben genannt, wurde die durchschnittliche Bevölkerungsdichte errechnet. Diese Variable soll die Effekte der steigenden Komplexität der Infrastruktur erfassen, die durch eine höhere Bevölkerungsdichte hervorgerufen wird.
4. „LKW pro km“: Da kein direktes Maß für die Ausgesetztheit gegenüber dem Verkehr im Allgemeinen oder Schwerverkehr im Besonderen *auf Netzwerkebene* – weder auf Gemeinde- noch auf Kantonsebene – gefunden werden konnte, wurde diese Näherung benutzt. Es wird der Quotient aus der Anzahl registrierter LKWs durch die Länge des

Straßennetzwerkes dieser Stadt berechnet. Die Daten werden vom SSV publiziert und reichen zurück bis in das Jahr 1980.

5. „Bundessteuer pro EW“: Es wird angenommen, dass einer der Einflussfaktoren auf die Kostenunterschiede die politische Erwartungshaltung über die durchschnittliche Straßenqualität ist und diese positiv mit dem Einkommen der Einwohner korreliert. Die durchschnittliche direkte Bundessteuer pro Kopf dient als Maß für das durchschnittliche Einkommen, da diese Steuer über alle Kantone gleich ist. Gemeindesteuersätze sind nicht überall gleich, weshalb die Bundessteuer pro Kopf nicht mit dem zur Verfügung stehenden Budget gleichzusetzen ist und noch weniger mit dem zur Verfügung stehenden Budget pro Km. Also wird hier etwas anderes als das Budget selbst gemessen. Neben dem Einkommen müsste eine zweite ähnliche Wirkung auf die politische Erwartungshaltung über die durchschnittliche Straßenqualität hinsichtlich der Bevölkerungsdichte bei gleichbleibendem Einkommen vermutet werden¹¹. Da die Bevölkerungsdichte bereits als Maß der komplexen urbanen Umgebung in der Regression fungiert, kann vermutet werden, dass diese Variable einen Teil der Effekte der durchschnittlichen Straßenqualität aufnimmt.
6. „Durchschn. Geländeneigung“: Es wurden Zufallsstichproben aus der GeoVITe¹² Datenbank extrahiert. Die dahinterliegenden Daten stammen von swisstopo. Der Wert der Gemeinde ist der Durchschnitt der Stichprobenpunkte.
7. „Geol. Schwierigkeit“: Es wurden Zufallsstichproben aus der GeoVITe Datenbank extrahiert. Die dahinterliegenden Daten stammen von swisstopo. Die 144 Unterkategorieen wurden einem der drei Schwierigkeitsgrade zugeordnet: Schwierig = +1, Durchschnitt = 0, Einfach = -1. Der Wert der Gemeinde ist der Durchschnitt der Stichprobenpunkte.

3.2.3 Regressionsgleichung und Parameterwerte

Mit den Variablen in Tabelle 1 wurde Gleichung (10) mit einem gewichteten Kleinst-Quadrat-Schätzer ermittelt, wobei die gesamten Straßenausgaben ohne Normierung nach Kilometern als Gewichtungsvektor verwendet wurden, weil diese am stärksten mit der Streuung der Schätzfehler korrelierten.

$$\begin{aligned}
 \text{Ausgaben pro Km}_{i,\bar{t}} = & \beta_0 + \beta_1 \text{Dummy94}_{i,\bar{t}} + \beta_2 \text{Dummy02}_{i,\bar{t}} \\
 & + \beta_3 \text{Wohnfläche pro km}_{i,\bar{t}} + \beta_4 \text{Einwohnerdichte im Wohngebiet}_{i,\bar{t}} \\
 & + \beta_5 \text{LKW pro Km}_{i,\bar{t}} + \beta_6 \text{Bundessteuer pro EW}_{i,\bar{t}} \\
 & + \beta_7 \text{Durchschn. Geländeneigung}_{i,\bar{t}} + \beta_8 \text{Geol. Schwierigkeit}_{i,\bar{t}} + v_{i,\bar{t}}
 \end{aligned}
 \tag{10}$$

¹¹ Ein anonymer Gutachter hat auf diesen Effekt hingewiesen.

¹² Quelle: Bundesamt für Landestopografie swisstopo (Art. Nr. 30 GeoIV): 5704 000 000.

Tabelle 2 zeigt die Ergebnisse der Regression. Mit Ausnahme der geologischen Schwierigkeit weisen alle Einflussfaktoren sowohl statistisch als auch numerisch relevante Zusammenhänge zu den Gemeindeausgaben für Straßenneubau und Straßenerneuerung auf, wobei die oben erwähnten statistischen Schwierigkeiten zur Vorsicht mahnen. Die Vorzeichen sind wie erwartet. Die jährlichen durchschnittlichen Ausgaben pro km in 1000 CHF₂₀₁₀ aus Tabelle 1 liegen bei 34.21¹³. Demgegenüber ist der mittlere Einfluss der Wohnfläche pro km in Tabelle 2 bei 23.74 numerisch bedeutsam und mit einem t-Wert von 6,3 statistisch relativ präzise geschätzt. Der Geländeneigung ist im Durchschnitt weniger bedeutsam, aber erhöht man die Neigung einer durchschnittlichen Straße um eine Standardabweichung (etwa 5 Grad), werden $5.82/34.21 = 17\%$ mehr Kosten generiert. Eine Umfrage unter den an der Fallstudie beteiligten Kantonsingenieuren ergab erwartete Mehrkosten bei einer „steilen und kurvenreichen Straße“ von 15%, bezogen auf die durchschnittlichen Erneuerungskosten.

¹³ Um diese Werte auf die Kosten pro m² bei einer Massnahme umzurechnen, müssen sie mit der erwarteten Lebensdauer multipliziert und durch die durchschnittlichen Fläche pro km Länge dividiert werden. Bei 8000 m² und 25 Jahren sind das ca. CHF 110.- pro m². Bei 7000m² und 30 Jahren sind es ca. CHF 150.-.

Tabelle 2: Regressionsschätzung der Einheitskosten der Einflussfaktoren

	R	R ²	korrigiertes R ²	Standard- schätzfehler		
	0.613	0.3755	0.3625	328.7		
	Variation	Freiheitsgrad	Quadr. Mittel	F	Sig.	
Regresswerte	24893773	8	3111721	28.79	0.000	
Residuen	41385662	383	108056	0	0	
Gesamt	66279435	391	0	0	0	
Variable	Koeff.	Std.Abw.	t-Wert	Signif.- Niveau	Mittl. Einfluss ¹⁴	Std.Abw.- Einfluss ¹⁵
(Konstante)	-8.444	6.503	-1.298	0.195		
Dummy_94_01	-24.439	3.776	-6.473	0.000		
Dummy_02_09	-19.856	3.699	-5.367	0.000		
Wohnfläche pro km	2.455	0.392	6.267	0.000	23.74	10.08
Einwohnerdichte im Wohngebiet	0.455	0.091	4.98	0.000	17.9	6.74
LKW pro km	0.581	0.185	3.138	0.002	7.83	5.28
Bundessteuer pro EW	0.00486	0.001	3.471	0.001	5.57	4.69
Durchschn. Gelän- deneigung	1.178	0.331	3.561	0.000	6.73	5.82
Geol. Schwierigkeit	29.409	17.568	1.674	0.095	0.86	2.55

Quelle: Richmond et al. (2013)

3.2.4 Berechnung der KE-Maßzahl für die Kantone

Mit den geschätzten Parameterwerten aus Tabelle 2 lässt sich die KE-Maßzahl gemäß Gleichung (11) berechnen¹⁶. Der Durchschnitt der teilnehmenden Kantone wurde für $\mathbf{p}^T \bar{\mathbf{Z}}$

¹⁴ Der Durchschnittswert aus Tabelle multipliziert mit dem geschätzten Parameter.

¹⁵ Die Standardabweichung aus Tabelle multipliziert mit dem geschätzten Parameter.

¹⁶ Die Parameter \mathbf{p} sind in CHF pro Einheit Zustandsverbesserung pro m². Die geschätzten Parameter β sind in 1000 CHF pro Jahr pro km Länge, unter den drei Annahmen: 6000m² Fläche pro km Länge; $\bar{\lambda} = \frac{1}{T} = \frac{1}{24}$; und 3 Einheiten Zustandsverbesserung pro Erneuerungsmaßnahme. Zusammen ergibt dies einen Umrechnungsfaktor von $\mathbf{p} = \frac{4}{3} \beta$.

substituiert, damit die Kosteneffizienz einen Durchschnitt von Null hat. Die flächengewichtete Summe $\Delta\hat{C}_{i,\bar{i}}$ wurde aus den Abschnittsdaten der Teilnehmer an zwei nacheinander folgenden Zustandsaufnahmen berechnet, wie in Gleichung (6) dargestellt.

$$KE_{i,\bar{i}} = \left(\frac{1}{5} \sum_i \frac{TC_{i,\bar{i}}}{\Delta\hat{C}_{i,\bar{i}}} + \mathbf{p}^T \Delta\mathbf{Z}_{i,\bar{i}} \right) - \frac{TC_{i,\bar{i}}}{\Delta\hat{C}_{i,\bar{i}}} \quad (11)$$

Die Resultate dieser Berechnung sind in Tabelle 3 aufgeführt, wobei (a) und (b) die zwei Terme der rechten Seite von (11) sind. Gemäß der I_1 Skala und nach Beobachtungen aus der Fallstudie treten Erneuerungsmaßnahmen selten später als bei Zustand 3,5 ein. Nur etwa 0.3% der Abschnitte wiesen einen schlechteren Zustand auf. Um Anhalt dafür zu geben, wie die Größe der Zahlen in Tabelle 3 zu interpretieren ist, kann man eine durchschnittliche Zustandszunahme bei einer Maßnahme von 3 annehmen. Dann würde der durchschnittlichen Quadratmeterpreis einer Maßnahme CHF 243.- (3 x 81.02) betragen.

Tabelle 3: Beispiel KE-Maßzahlen für fünf Kantone

	(a)	(b)	(a)-(b)
	Vergleichswert	Angepasste	KE-
SNV	Ausgaben / ΔC	Ausgaben / ΔC	Maßzahl
1	76.54	116.48	-39.93
2	84.19	66.09	18.10
3	75.88	57.77	18.11
4	73.00	77.77	-4.77
5	95.49	86.99	8.50
∞	81.02	81.02	0.00

Quelle: Richmond et al. (2013)

Gemäß Tabelle 3, Spalte (b) ist SNV 1 etwa 50% teurer im Bewirken von Zustandsverbesserungen als der Durchschnitt – auch nach allen Anpassungen für die Einflussfaktoren. Dieser sehr große Unterschied wird in Tabelle 4 in einzelnen Anpassungsschritten auseinandergenommen. Die jeweilige Standardabweichung gibt Auskunft darüber, ob die Streuung nach dem Anpassungsschritt kleiner wird oder nicht. Da die tatsächliche Kosteneffizienz der SNVen nicht bekannt ist, lässt sich die Maßzahl nur über die Streuung evaluieren, unter der Annahme, dass alle SNVen relativ ähnlich in ihrer Kosteneffizienz sind. Die Indi-

zes sind so konstruiert, daß der Durchschnitt immer bei 100 liegt. Somit können die Standardabweichungen direkt miteinander verglichen werden.

Die Spalte (e) in Tabelle 4 bildet die Ausgangslage und gibt die Ausgaben pro km an, so wie dies in Abbildung 1 gezeigt wird. Berücksichtigt man zunächst nur die unterschiedlichen Flächen (Spalte (e) zu (f)), sinkt die Standardabweichung von 28.3 auf 23.3 markant. Spalte (d) zeigt die unterschiedlichen Mengen an Zustandsverbesserung, die in der Periode bewirkt wurden. SNV 3 sticht mit beinahe dem Doppelten an Zustandsverbesserung im Schnitt pro m^2 heraus. Scheinbar gab es eine Erneuerungskampagne im betrachteten Zeitraum. Wird dieser Mengenunterschied ebenfalls berücksichtigt (Spalte (f) zu (g)), wechselt SNV 3 von ineffizient (116) zu effizient (66). Die Streuung nimmt aber gleichzeitig wieder zu, was zwar unerwünscht ist, aber keinen Beweis dafür liefert, dass diese Anpassung nicht erfolgen sollte. Spalte (h) zeigt die Verhältnisse unter den SNVen bezüglich der relativen Betroffenheit gegenüber den teuer zu erneuernden Straßentypen. SNV 1 hat nicht nur breitere Straßen, die Straßen sind auch vermehrt in einer höheren Verkehrslastklasse (VLK). Rechnet man diesen Unterschied mit ein (Spalte (g) zu (i)), rücken sowohl SNV 1 wie auch 3 zur Mitte und die Streuung fällt erneut. Die letzte Anpassung von Spalte (i) zu (k) berücksichtigt die mit der Regression geschätzten Kosten der Einflussfaktoren. Diese sind in Spalte (j) angegeben. Da $\mathbf{p}^T \Delta \mathbf{Z}$ zu den Vergleichswertausgaben gehört, wird es von den empirischen Ausgaben subtrahiert. Die Größenordnung dieser Anpassung beträgt im größten Fall fast 17% (107/17.86) und ist somit nicht vernachlässigbar. Wenn dies ebenfalls berücksichtigt wird, wechselt SNV 5 von ineffizient zu effizient.

Tabelle 4: Einzelne Anpassungsschritte in der Maßzahlberechnung

	(d)	(e)	(f)	(g)	(h)	(i)	(j) ¹⁷	(k)
	∞	Index: Ausg./	Index: Ausg./	Index: Ausg./	∞	Index: Ausg. /		Index: Ausg. /
SNV	$\Delta C / m^2$	km	m ²	ΔC	VLK Faktor	ΔC	$\mathbf{p}^T \Delta \mathbf{Z}$	ΔC
1	-0.094	145	137	152	1.07	143	-5.53	149
2	-0.100	72	77	81	1.00	81	3.91	77
3	-0.182	119	116	66	0.95	71	-6.34	77
4	-0.091	73	82	94	0.99	96	-9.90	105
5	-0.086	89	86	104	0.99	107	17.86	89
St. Abw.		28.3	23.3	29.3		25.1		26.7

Quelle: Richmond et al. (2013); Alle Indizes sind gegenüber dem Gruppendurchschnitt als Verhältnis und mal 100 gerechnet

Zusammenfassend sind die diversen Anpassungen zwar numerisch bedeutend, aber insgesamt können sie die Unterschiede zwischen den SNVen nicht erklären oder vermindern. Störend bleibt die Diskrepanz zwischen dem besten und dem schlechtesten Resultat von fast 100% - auch nach allen Anpassungen. Als Erklärung bleibt die Vermutung der fehlenden Vergleichbarkeit in den Grunddaten über Ausgaben¹⁸, wie dies bereits bei Abbildung 1 gezeigt wurde. Eine Maßzahl für Kosteneffizienz kann nur dann zu überzeugenden Vergleichen führen, wenn bei den Ausgabendaten von allen Teilnehmern in konsistenter Weise zwischen Erneuerungsmaßnahmen, Ausbau, Neubau und betrieblicher Unterhalt unterschieden wird.

Wie in der vorgestellten KE-Maßzahl gezeigt, ist eine Abgrenzung zwischen Erneuerung und Instandsetzung als Ursache der Zustandsverbesserung nicht zwingend nötig, obwohl präzisere Masszahlen konstruierbar wären, falls diese Unterscheidung durch die Daten konsistent nachzuvollziehen wäre. Sofern Zustandsdaten vor und nach den Erneuerungsmaßnahme nicht erhoben worden sind, oder falls die genauen Abschnitte, auf denen die Maßnahmen erfolgten, nicht in den Datensätzen erfasst sind, muss man die Ausgaben für

¹⁷ Die eigentlichen Werte wurden mit dem Faktor 100/81.02 skaliert, um proportional zu den anderen Werten zu bleiben.

¹⁸ Das genaue Buchungsprozedere wurde in der Fallstudie ebenfalls untersucht. Obwohl es nicht möglich war, zu eruieren, um wie viel jeder Teilnehmer im Schnitt voneinander abweichen müsste, ließ sich schematisch zeigen, dass Unterschiede in einer Größenordnung von bis zu 4:1 möglich sind.

Erneuerung und Instandsetzung ohnehin aggregieren. Nur so erhält man die Gesamtausgaben, die zu der Menge an Zustandsverbesserungen passen, die für Gleichung (6) benötigt werden.

4. Schlussfolgerungen

In diesem Bericht wurden Maßzahlen präsentiert, die über die mengenmäßig ausreichende sowie kostenmäßig effiziente Erneuerung Auskunft geben. Ob die Erneuerung ausreichend ist, lässt sich am besten mit physikalischen Quantitäten beantworten, wie zum Beispiel dem durchschnittlichen Zustand oder dem Anteil des Netzwerks, der jedes Jahr erneuert wird. Solche Maßzahlen sind relativ einfach zu berechnen und – im Vergleich zu der Schätzung der Kosten von Maßnahmentypen – unterliegen sie weniger Messunsicherheiten. Diese Unsicherheiten treten unausweichlich in jenen Berechnungen auf, in denen die Frage nach einer ausreichenden Erneuerung in der Form einer Frage nach ausreichenden Ausgaben gestellt wird.

Entscheidet sich die Aufsichtsinstanz für die wiederholte Leistungsbemessung nach einer bestimmten Maßzahl, ist davon auszugehen, daß sich das Verhalten der Beaufsichtigten auf diese Tatsache einstellt. Ist die Maßzahl einseitig – und das ist fast immer der Fall – entsteht ein einseitiges Anreizsystem. Es ist daher ratsam für eine Aufsichtsinstanz möglichst vielseitige Maßzahlen gleichzeitig einzusetzen. Mit diesem Gedanken im Sinne wurde eine Reihe zusätzlicher Maßzahlen vorgestellt, die ebenfalls aus den Daten der Pavement Management Systeme berechnet werden können, um die unerwünschten einseitigen Anreize zu bekämpfen. Diese Maßzahlen liefern auch wichtige Einsichten in die Erneuerungsstrategien und wie sie implementiert werden.

Die vorgeschlagene Maßzahl zur Kosteneffizienz gestaltet sich als eher kompliziert und benötigt auch Daten, bei denen die Messunsicherheit groß ist. Es wurden Methoden gefunden, um die Kosteneinwirkung von einigen wichtigen Einflussfaktoren auszugleichen, obwohl die Datenlage nicht ausreicht, um den Ausgleich in allen Fällen in ausreichender Qualität zu parametrisieren. Entsprechend nüchtern fällt auch ein Rechenbeispiel bezogen auf fünf Schweizerische Kantone aus. Die Größenordnung der Streuung vor und nach vier Anpassungen blieb so hoch, dass der Ineffizienteste doppelt so viel aufwendet wie der Effizienteste, was eher auf Probleme in den Ausgabendaten hindeutet. Dennoch konnte eine theoretisch konsistente Messmethode für die Leistung einer SNV gefunden werden. Im Unterschied zu den auf physischen Daten basierenden Maßzahlen, wird von einem allgemeinen Einsatz der KE-Maßzahl im konkreten Fallbeispiel abgeraten, bis grundlegende Datenprobleme angemessen gelöst werden können.

5. Abstract

Road networks are nearly exclusively managed by public bodies and hence there is a need for comparable measures to determine whether the networks are being maintained suffi-

ciently and efficiently. Generally, both types of measures are missing. The primary cause of that absence is likely to be found in the difficulty of achieving comparability across networks. In this article theoretical and practical methods for achieving comparability are derived and tested on the data from a group of Swiss cantons. The presented measures focus primarily on renewal maintenance. In addition to measures concerning sufficiency and efficiency, a set of further measures is introduced, as this is done in a Balanced Scorecard approach, to combat potential skewed incentives that will likely result from putting too much management attention on single quantitative measures. Collectively, the presented measures should enable a reasonable benchmarking of renewal management activity for road networks.

6. Literaturverzeichnis

- Aguiar-Moya, J. P., Prozzi, J. A., & de Fortier Smit, A. (2011), Mechanistic-empirical IRI model accounting for potential bias, *Journal of Transportation Engineering*, 137(5), 297-304.
- Bardaka, E., Labi, S., & Haddock, J. (2014), Using Enhanced Econometric Techniques to Verify the Service Life of Asset Interventions, *Transportation Research Record: Journal of the Transportation Research Board*, 2431(1), 16-23.
- Chow, C., & Troyan, V. (1999), Quantifying Damage from Utility Cuts in Asphalt Pavement by Using San Francisco's Pavement Management Data, *Transportation Research Record: Journal of the Transportation Research Board*, 1655(-1), 1-7.
- Cook, W. W., Kazakov, A., & Roll, Y. (1993), On the Measurement and Monitoring of Relative Efficiency of Highway Maintenance Patrols. In A. Charnes, W. W. Cooper, A. Y. Lewin & L. M. Seiford (Eds.), *Data Envelopment Analysis: Theory, Methodology and Application*. London: Kluwer Academic Publishers.
- Deller, S. C., Chicoine, D. L., & Walzer, N. (1988), Economies of Size and Scope in Rural Low-Volume Roads. *The Review of Economics and Statistics*, 70(3), 459-465.
- Der Rechnungshof. (2012), Erhaltungsmaßnahmen bei Autobahnen und Schnellstraßen *Bericht des Rechnungshofes*. Wien.
- Fallah-Fini, S., Triantis, K., de la Garza, J. M., & Seaver, W. L. (2012), Measuring the efficiency of highway maintenance contracting strategies: A bootstrapped non-parametric meta-frontier approach, *European Journal of Operational Research*, 219(1), 134-145.
- Farsi, M., Filippini, M., & Greene, W. (2005), Efficiency Measurement in Network Industries: Application to the Swiss Railway Companies, *Journal of Regulatory Economics*, 28(1), 69-90.

- Greene, W. (2005), Fixed and Random Effects in Stochastic Frontier Models, *Journal of Productivity Analysis*, 23(1), 7-32.
- Gunsteren, E. v., Kieliger, T., Traber, M., Nellen, A., & Zahnd, U. (2005), Abschätzung des durchschnittlichen jährlichen Wertverlustes von kommunalen Strassennetzen (pp. 110): VSS.
- Hinsch, K., Krause, G., Maerschalk, G., & Rübenstein, J. (2005), Kategorisierung von beschreibenden Größen für das Gebrauchsverhalten von Fahrbahnbefestigungen und die Wirkung von Erhaltungsmaßnahmen, *Forschung Strassenbau und Strassenverkehrstechnik* (Vol. 915). Bonn: Bundesministerium für Verkehr und digitale Infrastruktur, Abteilung Straßenbau.
- Hong, F., & Prozzi, J. A. (2010), Roughness model accounting for heterogeneity based on in-service pavement performance data, *Journal of Transportation Engineering*, 136(3), 205-213.
- Kalb, A. (2009), What Determines Local Governments' Technical Efficiency? The Case of Road Maintenance, *Discussion Paper* (Vol. 09, pp. 34).
- Kaplan, R. S., & Norton, D. P. (1996), *The Balanced Scorecard: Translating Strategy into Action*. Boston: Harvard Business School Press.
- Kerr, S. (1975), On the Folly of Rewarding A while Hoping for B., *Academy of Management Journal*, 18(4), 14.
- Khattak, M. J., Nur, M. A., Bhuyan, M. R. U. K., & Gaspard, K. (2014), International roughness index models for HMA overlay treatment of flexible and composite pavements, *International Journal of Pavement Engineering*, 15(4), 334-344.
- Koch, B., & Forster, M. (2010), Zustandsanalyse und Werterhaltung bei den Kantonsstrassen in der Schweiz (pp. 18). Zürich: Infrastruktur Strasse.
- Maerschalk, G. (2014), *Prognose des Erhaltungsbedarfs aller Anlageaggregate der Fernstrasseninfrastruktur*, Paper presented at the Verkehrsökonomik und -politik, Berlin. http://www.sep-maerschalk.de/cms/files/erhaltungsbedarfsprognose_fernstrassen.pdf, Zugriff am 24.06.2014.
- Mishalani, R. G., & Madanat, S. M. (2002), Computation of infrastructure transition probabilities using stochastic duration models, *Journal of Infrastructure Systems*, 8(4), 139-148.
- Newbery, D. M. (1988). Road Damage Externalities and Road User Charges, *Econometrica*, 56(2), 295-316.

- Oeser, M. (2014). Entwicklung von Prognosefunktionen für den Straßenzustand kommunaler Straßen, *Forschung Strassenbau und Strassenverkehrstechnik* (Vol. 1107). Bonn: Bundesministerium für Verkehr und digitale Infrastruktur, Abteilung Straßenbau.
- Ressel, W., Tejkl, K., & Klöpfer, C. (2013). Methodenstudie zur Life-Cycle-Bewertung von Strassenbefestigungen, *Forschung in Strassenbau und Strassenverkehrstechnik* (Vol. 1096). Bonn: Bundesministerium für Verkehr und digitale Infrastruktur, Abteilung Straßenbau.
- Richmond, C., Kielhauser, C., & Adey, B. T. (2013). KPI - Key Performance Indicators for Cantonal Road Management (pp. 160). Zurich, Switzerland: Eidgenössischer Technischer Hochschule, IBI.
- Rouse, P., & Putterill, M. (2008). Highway Performance. In D. A. Hensher & K. J. Button (Eds.), *Handbook of Transport Modelling* (pp. 743-759). London: Elsevier.
- Rouse, P., Putterill, M., & Ryan, D. (1997). Towards a General Managerial Framework for Performance Measurement: A Comprehensive Highway Maintenance Application, *Journal of Productivity Analysis*, 8(2), 127-149.
- Santos, J., & Ferreira, A. (2013). Life-cycle cost analysis system for pavement management at project level, *International Journal of Pavement Engineering*, 14(1), 71-84.
- Scazziga, I. (2008). Forschungspaket Massnahmenplanung im Erhaltungsmanagement, Forschungsauftrag VSS 2004/712; 1304 Eidgenössisches Departement für Umwelt, Verkehr, Energie und Kommunikation UVEK, Bundesamt für Strassen; 1304.
- Small, K. A., & Winston, C. (1986). Efficient pricing and investment solutions to highway infrastructure needs, *American Economic Review*, 76(2), 165-169.
- Weninger-Vycudil, Simanek, Rohringer, & Haberl. (2009). Handbuch Pavement Management in Österreich 2009 (Vol. 584). Wien: Bundesministerium für Verkehr, Innovation und Technologie.

Delivery Time Coordination at Customer Sites with Implicit Time Windows – A Computational Evaluation

BY JÖRN SCHÖNBERGER, DRESDEN

1. Introduction

Transport services connect the different locations in a value chain and realizes the physical flow of materials. As a consequence of the high level of labor share among partners in (global) value chains, transport services become more and more important in the business-to-business (B2B) sector. Due to the significant grow of online retail businesses and related internet-based retailing ideas transport services become also more important in the business-to-consumer (B2C) branch. Both the B2B sector as well as the B2C sector offer multi sourcing options where (different) products are bought from different sources (Agrawal et al., 2002). This results in a fragmented flow of goods since different lots and/or parcels are shipped independently from other lots and/or parcels from different supply site to a common customer site. This delivery site is visited several times per period. In the B2B setup, a coordination of the visiting operations becomes necessary in order to ensure that internal downstream processes fit with the (external) transport processes. Furthermore, setup as well as cleaning costs associated with the used handling tools increase if the multiple delivery operations remain uncoordinated here. In a B2C sector it is necessary to coordinate multiple visits at one customer site in order to increase the chance to realize the handover of shipments in a first attempt but to avoid costly and environmental critical re-visits.

The aforementioned challenge of coordination of delivery times belongs to the class of operational fleet management dispatching problems (Crainic and Laporte, 1998) in which the portfolio of requests to be fulfilled is known. Here, the decision about the short term routing of vehicles forming a fleet is addressed. Vehicle routing problems (Golden et al., 2008) like the capacitated vehicle routing problem (CVRP) or the pickup and delivery problem (PDP) represent the core decision tasks in the field of operational fleet management.

Anschrift des Verfassers:

Prof. Dr. Jörn Schönberger
Dresden University of Technology
"Friedrich List" Faculty of Transportation and Traffic Sciences
01062 Dresden

joern.schoenberger@tu-dresden.de

Approaches to achieve a time-coordination between (several) delivery operations and downstream internal processes at a common customer site are based on time windows. An explicit time window determines a period in which an unloading operation at a customer side may start. However, a time window represents a very strict delivery time restriction that is often unnecessarily restrictive. In this article we investigate so-called *moveable* or *implicit time windows* to achieve the aforementioned coordination between several inbound transport processes and downstream processes at a customer. Instead of specifying an absolute position for the delivery operation on the time axis as done by defining an explicit time window, the operation starting times associated with the necessary customer visits are groups together in an a priori unspecified period of the time axis. Only the length of this time window is given as part of the dispatching scenario but not its absolute position on the time axis. Practically, an implicit time window specifies an upper bound on the difference of the starting times of different operations at a commonly served customer site.

Offering implicit time windows to customers can be a competitive advantage of a distribution company that must deliver customer locations twice or even more often during a planning period. However, the consideration of implicit time windows represents a sophisticated challenge for the fleet dispatcher since several unloading times (of different vehicles) require a coordination during the dispatching of the necessary vehicle operations. Waiting times as well as deviations from least distance vehicle routes might become necessary to meet the implicit time windows. In order to contribute to a better understanding of the benefits and challenges related to the consideration of implicit time windows in operational fleet dispatching we are going to find answers to the following two research questions in this article:

1. How can the consideration of implicit time windows into state-of-the-art fleet disposition tools be realized?
2. What are the impacts of the consideration of implicit time windows on the formation of vehicle routes with respect to additional travel distances and additional vehicles needed?

With the goal to find answers to the aforementioned questions we evaluate results from computational experiments. In these experiments we simulate the disposition of a vehicle fleet. The disposition situation is similar to the CVRP scenario but each customer must be visited twice by two different vehicles (each delivering only one specific commodity). A maximal time difference between the necessary two visits must be respected at each customer site. We introduce a dispatching procedure that generates high quality solutions for this dispatching problem by constructing and improving solutions of a model of the disposition problem proposed in Schönberger (2015).

We present the investigated fleet disposition scenario in Section 2. Section 3 describes central aspects of the disposition system used in the simulation experiments. Experimental results are presented and analyzed in Section 4.

2. A Fleet Dispatching Scenario with Multiple Customer Site Visits Coordinated by Implicit Time Windows

The consideration of implicit time windows in fleet management and vehicle routing comes along with additional challenges compared to the consideration of explicit time windows. Therefore, it is necessary to carefully introduce into the concept of this coordination scheme. Subsection 2.1 contains a survey of research related to the consideration of multiple visits at a customer site and their coordination. Subsection 2.2 introduces the fleet management scenario with implicit time windows discussed in the remainder of this article. Subsection 2.3 compares the coordination of multiple visits by implicit time windows with other time window based coordination approaches. Subsection 2.4 summarizes opportunities to achieve and ensure feasibility with respect to implicit time windows.

2.1 Literature

Golden et al. (2008) survey contributions and research directions on the vehicle routing problem class. Problems in which individual pickup and/or delivery locations are considered are investigated and classified by Parragh et al. (2008).

Vidal et al. (2012) write about models and algorithmic approaches on a broader class of vehicle routing problems with several depots. Crevier et al. (2007) report about a variant of the vehicle routing problem in which vehicles are replenished at different re-filling depots while being on route. Cattanzuza et al. (2014) investigate the multi trip vehicle routing problem in which only small vehicles are allowed to enter a downtown area so that a regular return to a depot to refill a vehicle becomes necessary. Goel and Meisel (2013) propose a decision support tool for a vehicle deployment problem in which operations of different vehicles at different locations are coupled by a scheduling constraint.

Differences between routing problems with only one and with several commodities are revealed by Archetti et al. (2014). The split of demand associated with a request in the context of vehicle routing is addressed by Archetti and Speranza (2008).

A general discussion of different types of synchronization constraints in the context of vehicle routing is contributed by Drexl (2012).

The usage of time windows for the coordination of internal processes at a customer site with the delivery times decided by the fleet dispatcher is investigated by Solomon (1987), de Jong et al. (1996), Favaretto et al. (2007) as well as Breier (2015) discuss situations in which the dispatcher has the choice among two or even more alternative time windows at a customer location.

Doerner et al. (2008) report a vehicle routing problem in which sensitive blood products must be collected. In the reported application, a blood collection site must be re-visited

several times in order to ensure that the meanwhile collected blood can be processed without deterioration. The concept is quite similar to an implicit time window since the first visit of a location sets limits for subsequent visits. The authors refer to this scenario as a vehicle routing problem with interdependent time windows.

2.2 Dispatching Problem Description

We consider the fleet management scenario introduced in Schönberger (2015). In this scenario the distribution of two types of commodities, A and B, is addressed from the perspective of a trucking company. The trucking company must fulfill transport requests with its own fleet. A transport order comprises the fulfillment of demand of both commodities. Each commodity requires a specially fitted vehicle. Therefore, two types of vehicles are maintained and compose the fleet that is deployed in order to cover the supply demand.

Each of the n_A type-A vehicles can be used to transport only one or several pieces of commodity A and each of the n_B type-B vehicles is able to carry one or several pieces of commodity B only. There are no vehicles that can carry both types of commodities.

All commodities to be distributed are stored at the warehouses WH-A as well as WH-B. The first mentioned warehouse stocks only commodity A but commodity B is exclusively stored at WH-B. Each customer order comprises two transport requests r^A as well as r^B (origin-to-destination requests, for short: od-requests). Request r^A expresses the need to transport a certain quantity of commodity A from warehouse WH-A to a customer location. Similarly, request r^B expresses the need to transport a certain quantity of commodity B from warehouse WH-B to the same customer location specified in r^B .

There is a scheduling requirement that prevents the decomposition of the vehicle routing problem into a type-A routing problem as well as into a type-B routing problem. The delivery operations associated with the two requests of an order must not differ by more than δ time units. A decomposition of the outlined vehicle routing problem into two individual one-commodity-routing problem is therefore inappropriate.

The trucking company has to determine routes for the vehicles so that the total sum of travel distances across the fleet is minimized while the aforementioned implicit time window condition is met. Furthermore, no vehicle route is allowed to last longer than MS^{\max} time units (maximal allowed makespan) This decision problem is called the two-commodity capacitated vehicle routing problem with synchronization (2-CVRP-S).

2.3 Operation Starting Time Coordination

It is necessary to ensure that both unloading operations associated with an order are scheduled so that their starting times do not differ more than δ time units. This condition has to be formalized in order to prepare the computer-supported fleet disposition.

2.3.1 Explicit Time Windows

An explicit operation time window is an interval $T := [e_O; l_O]$ of the time axis limited by an earliest allowed starting time e_O of an operation O and by a latest allowed operation starting time l_O for this operation. Explicit time windows are often used in the context of vehicle routing, i.e. in the Vehicle Routing Problem with Time Windows (VRPTW) investigated for example by Solomon (1987) or in the Pickup and Delivery Problem with Time Windows (PDPTW) survey by Parragh et al. (2008). Time windows are typically defined in order to enforce that a certain unloading happens within a certain time period.

Consider now a situation with only one unloading operation O and let st_O denote the determined starting time of operation O so that $st_O \in T$. Here, the consideration of the time window enforces the operation starting time into the specific interval on the time axis so that internal upstream or downstream process comply with the unloading operation. Figuratively spoken, the time window consideration enforces the operation starting time to be positioned within a pre-specified part of the time axis (*position restricting property of a time window*).

Assume now two different unloading operations P as well as Q . These two operations are contained in two different routes executed by different vehicles. In case that both operations are scheduled with the interval T then it is sure that the operation starting times of the two aforementioned operations do not differ more than $l_O - e_O$ time units (*difference restricting property of a time window*).

In order to ensure that the two unloading operations are started with a time difference of not more than δ time units it is possible to define a time window $[t_0; t_0 + \delta]$ and to ensure that both associated operation starting times fall into this time window. Immediately, the question about an adequate value for t_0 arises. Actually, every t_0 -value leads to a time window that implies the fulfillment of the maximal operation starting time difference restriction. The absolute positions of the two operation starting times on the time axis are irrelevant here. It is sufficient to exploit the difference property of the time window but it is not necessary to fulfill the position property of this time window. In conclusion, an explicit time window is inappropriate to take care that the starting time difference between two operations do not differ more than a given threshold value. An explicit time window is too restrictive since we do not need the position restricting property.

In order to overcome this shortcoming and with the goal to relax the position restricting property of an explicit time window two related time window concepts are proposed. These concepts will be introduced and compared now.

2.3.2 Alternative Time Windows

de Jong et al. (1996), Favaretto et al. (2007) as well as Breier (2015) propose to define several time windows for each operation (customer site) and to give these time windows to

the dispatcher of the fleet. A dispatcher has to select one of the proposed explicit time windows and all associated operation starting times must fall into the selected explicit time window. The main motivation of specifying several time windows for one customer site is to give the fleet planner more freedom to compile profitable vehicle routes. If the same time window must be selected for both operations associated with an order then the maximal allowed operation starting time difference condition is fulfilled. Alternative time windows enable the dispatcher to exploit the difference restricting property of explicitly formulated time windows. The position restricting property of an explicit time window, which is not requested here, is (partly) deactivated.

However, beside the determination of the operation starting times of the considered operations it is necessary for a fleet planner to select the appropriate time window from the set of available time windows. This is a quite complicated task since a decision problem of enlarged complexity has to be solved (de Jong et al., 1996). Two different kinds of interrelated decisions (routing as well as time window selection) must be made during the fleet dispatching process.

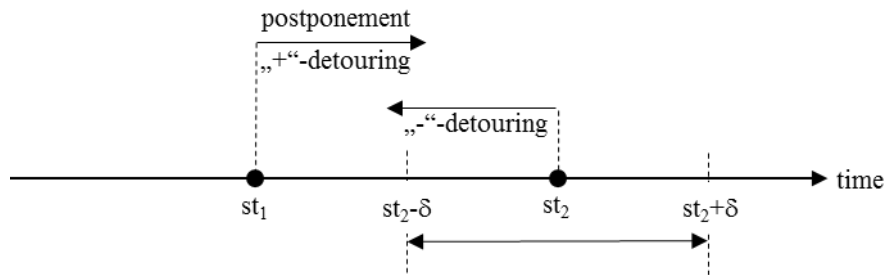
2.3.3 Implicit Time Windows

Schönberger (2015) proposes a non-stationary (or shiftable) time window approach to coordinate the operation starting times of several coupled operations. Here, it is not necessary to specify an explicit time window as part of the problem data. Instead, the maximal arrival time difference between pairs of operation starting times to be respected is added as problem data to the dispatching problem formulated directly as a constraint to be respected.

Practically, the determination of a first operation starting time at the corresponding customer location defines an explicit time window in which the second operation starting time must be determined in a feasible way. This time window reduces whenever additional operation starting times at this customer location are defined. Clausen (2011) as well as Stieber et al. (2015) introduce the term *implicit time window* for such a variable time window. Here, it is not necessary to select a time window. Only operation starting times have to be decided.

Implicit time windows provide the here required difference restricting property but not the here redundant and too restrictive position restricting property of an explicit time window. However, since the position of the implicit time window is not part of the fixed problem data it is necessary to determine this position during the solving of the fleet dispatching decision task. Therefore, implicit time windows increase the already quite high complexity of the fleet dispatching task by adding a further decision problem component.

2.4 Dispatching Strategies to Meet Implicit Time Windows

Figure 1: Options to achieve implicit time window feasibility

The common agreement of an implicit time window between a carrier and a customer, e.g. the specification of a maximal allowed time difference between two visits at the customer's location is assumed. Now, it is necessary to ensure that the agreed time window is considered during the composition of the vehicle routes.

Route composition comprises the solving of two interdependent decision tasks (the vehicle routing problem). First, it is necessary to partition the available set of od-requests into clusters (tours) and to assign each cluster to a vehicle that is able to serve the contained od-requests. Second, the loading and unloading operations associated with the od-requests in a cluster must be arranged in a sequence in order to build the executable vehicle routes (routing). Obviously, both decisions (clustering and routing) are interdependent. The need to keep the total travel distances resulting from the route execution as low as possible, makes the vehicle routing problem quite challenging and requires regularly an iterative revision of the tentatively made clustering and the tentatively made routing decisions.

Beside clustering and sequencing loading and unloading operations of the vehicles it is necessary to determine the starting time of each individual operation (schedule building). The typical scheduling approach in the context of vehicle routing is left-to-right shifting, where each operation is scheduled to start as early as possible. Here, a vehicle starts from a depot at time 0 and the arrival time of the vehicle at the first visited operation site is set by adding the travel time between the depot and this site. In case that a vehicle arrives before a given explicit time window has opened, a waiting time is inserted. The operation starting time is shifted to a later time until the earliest allowed operation time is reached. Adding the operation duration (service time) to the operation starting time determines the operation finishing time which is equal to the vehicle leaving time from the associated site. Operation starting times of subsequently visited sites are consecutively determined in the same way.

In case that the schedules for the vehicle routes are setup independently for each individual vehicle it may happen that the two determined starting times st_1 as well as st_2 ($st_1 < st_2$) of the two unloading operations associated with an order fail to fulfill the implicit time win-

dow. Such a situation is given in Figure 1. It is necessary to revise the so far made scheduling and or routing decisions. Two general revising options are available. One approach is to shift the secondly scheduled earlier operation starting time st_1 to the right until it falls in the interval $[st_2-\delta;st_2+\delta]$. This right shifting means to postpone the starting time of st_1 . Postponing this operation establishes feasibility with respect to the implicit time window associated with this order. However, the right shifting of an operation requires the right shifting of all subsequently visited operations in this route so that the updated starting times are endangered to cause (additional) infeasibilities with respect to (implicit or explicit) time windows to be considered. In addition the latter scheduled operation cannot be postponed with the goal to establish implicit time window feasibility.

Another approach to vary the scheduled operation starting times st_1 or st_2 is to modify the sequence in which earlier visited operations in the two considered routes are processed. This route modification leads, in general, to additional travel distances for processing the updated route(s) (detouring). However, detouring can be applied to both unloading operations of the considered order. Modifying the sequence before operation 1 means to shift st_1 to the right („+“-detouring) but modifying the sequence of the operations before operation 2 by re-position some earlier processes operations in the route that serves operation 2 after operation 2 realizes a left shifting of st_2 („-“-detouring). Similar to postponing a request the application of detouring can lead to time window infeasibilities of other operations served by the affected route.

Taking care of the fulfillment of implicit time windows is therefore a very challenging task especially since scheduling decisions and, in several situations, routing decisions in two (or more different routes) have to be coordinated. In contrast, feasibility with respect to an explicit time window can be achieved by updating routes independently, since the explicit time window does not rely on any scheduling decision made before (for another vehicle). The position restricting property of an explicit time window takes care that after the independent shifting of st_1 and st_2 these two times still fulfill the time difference restricting property.

3. Fleet Disposition with Implicit Time Windows

Ensuring feasibility of an implicit time window associated with a customer location requires the coordination of operation starting times within two different vehicle routes. This involves the confirmation and/or revision of scheduling and/or routing decisions in at least two different routes. Scheduling concepts developed for vehicle routing problems with explicit time windows are insufficient to cope with this challenge (Solomon, 1987). They are able to compare a variable operation starting time with a fixed time window opening (or closing) time but they cannot take care about ensuring that the two starting times of operations in two different routes differ by at most δ time units. It is therefore necessary to equip such an approach with additional capabilities to take care about implicit time windows. This section contains the presentation of such an enriched combined routing and scheduling

fleet management tool. In Subsection 3.1 the framework of such a combined dispatching system is presented. Subsection 3.2 contains the description of the mechanism to control the consideration of implicit time windows.

3.1 Metaheuristic Route Construction

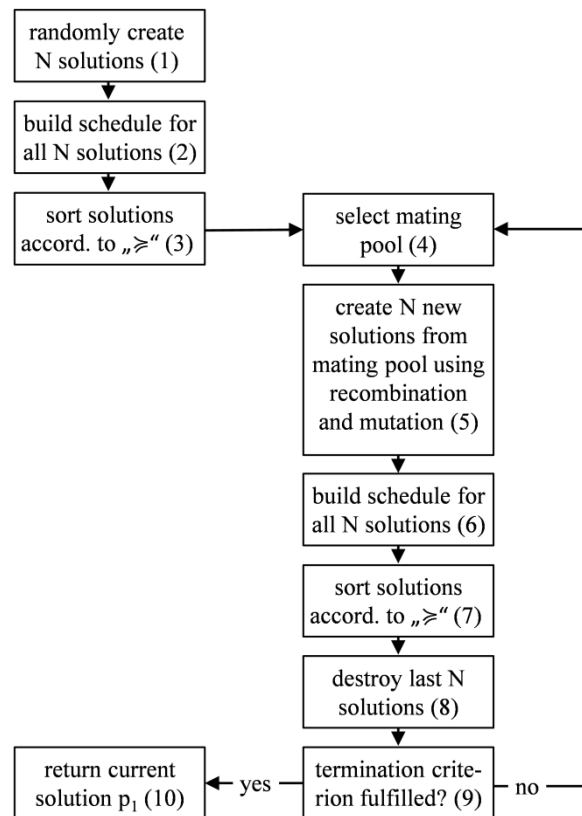
The simultaneous route generation and schedule determination in a situation of a vehicle routing problem with implicit time windows is a very complicated and complex decision task. In Schönberger (2015) a mixed-integer linear optimization model for the situation outlined in Subsection 2.2 has been developed. It has been demonstrated that the application of a commercial model solver like IBM CPLEX is possible only for very small problem instances (with 6 orders and 12 requests). The application of a heuristic model processing algorithm seems appropriate and promising. In the remainder of this section, the development of a genetic algorithm based heuristic for searching and determining high quality solutions of the considered vehicle routing problem with implicit time windows is addressed. This genetic algorithm is an extension of the framework developed and evaluated in Schönberger (2011).

A genetic algorithm is a population-based heuristic. Here, a solution is a route set generated for the considered vehicle routing problem with implicit time windows. Given a set of N solutions (forming the „parental population“) it generates N additional solutions (forming the „offspring population“) using the three mechanisms of „selection“, „recombination“ as well as „mutation“.

Each solution in a population is evaluated and the fitness value represents a quantification of the evaluation result. We use a three-facetted evaluation scheme. The evaluation of a solution p starts with some preparatory steps. Initially, the MS^{\max} -exceeding (expressed in time units) is collected from all routes („sum of exceeding of MS^{\max} “). Next, the exceeding of δ is collected from all orders („sum of exceeding of the implicit time windows“). Using these two sums as well as the sum of travel distances associated with the routes stored in p enables a three-dimensional evaluation of the solution p . First, $F^1(p)$ gives the sum of exceeding of MS^{\max} observed in solution p . Second, $F^2(p)$ carries the sum of exceeding of the implicit time windows observed in p . Finally, $F^3(p)$ represents the travel distances resulting from the route in p . Based on these three evaluation criterions, we are able to define a sorting scheme to order the members of the population („ranking“). We write „ $p_1 \succcurlyeq p_2$ “ (solution p_1 is ranked higher than solution p_2) according to the following definition of „ \succcurlyeq “:

$$p_1 \succcurlyeq p_2 \Leftrightarrow \begin{cases} F_1(p_2) \geq F_1(p_1) \\ F_2(p_2) \geq F_2(p_1) \\ F_3(p_2) \geq F_3(p_1) \end{cases} \quad \begin{array}{l} \text{if } F_1(p_2) = F_1(p_1) \\ \text{if } F_2(p_2) = F_2(p_1) \wedge F_1(p_2) = F_1(p_1) \end{array}$$

We can use „ \succcurlyeq “ to sort the entire population, so that the highest positioned solutions comply with the restrictions and objectives to a higher extend then those solutions positioned at the end of the implied ranking.

Figure 2: Flow chart of the genetic algorithm fleet dispatching tool

The basic idea of a genetic algorithm based optimization procedure is to replace an existing parental population of solutions by a child population so that the average fitness observed in the child population is larger than the average fitness observed for the parental population. The replacement process is as follows. First, „selection“ selects preferentially some individuals that are positioned at the beginning of the sorted solution list according to „>“. These selected individuals form the so-called „mating-pool“. Next, N new solutions are generated in two steps. First „recombination“ randomly selects two solutions from the mating pool and re-combines properties from both parental solutions to create a new offspring solution. After the offspring solution has been created „mutation“ implements some random variations of the offspring. After N offspring solutions have been created they are inserted into the parental population. The resulting population consists of $2 \cdot N$ solutions. It is sorted using „>“. Only the solutions in the first N positions are kept and the remaining N solutions are destroyed. Now, an iteration is completed and a new iteration is initiated if a given

termination criterion is not yet fulfilled (e.g. maximal number of iterations reached). Otherwise, the procedure stops.

The genetic algorithm procedure used to find and to improve feasible solutions for the aforementioned fleet management problem with implicit time windows is represented in Fig. 2. The dispatching procedure starts with the creation of randomly generated route sets (1). Next, operation starting times are determined (2). Now, all members of the population are sorted according to „ \succ “ (3). The initial population is iteratively replaced by another population by the genetic algorithm steps (4)-(8). First, the mating pool is selected (4). Second, offspring solutions are created (5). Operation starting times are fixed (6) followed by the determination of the ranking of solutions (7). The worse half of the mixed parental and child population is destroyed (8). After these five steps it is checked whether an a priori defined termination criterion is reached (9). If the genetic algorithm has been stopped the currently best ranked solution is returned as solution of the fleet dispatching problem (10).

3.2 Achieving Implicit Time Window Feasibility

Sets of vehicle routes are generated and evaluated during the execution of the genetic algorithm. After having ensured that the maximal makespan constraints is fulfilled, the genetic algorithm primarily searches for solutions that fulfill all implicit time windows using both options detouring as well as operation starting time postponement.

The sorting of the members of a population prefers those solution with few or even no violations of implicit time windows. In case that a solution exhibits violations of implicit time windows it is ranked quite low compared to a solution without any time window violation even if the feasible solution requires to travel longer (detours).

Usually, the scheduling procedure used in steps (2) and (6) tries to determine the earliest starting time for an unloading operation at a customer site. However, postponing the starting time of an operation can help to fulfill the implicit time window at a customer site in case that the earliest possible starting time st_1 of the currently considered operation is more than δ time units earlier than the starting time st_2 of the second operation associated with this customer site. If the starting time of the currently considered operation 1 is postponed from st_1 until $st_2 - \delta$ then the maximal allowed time difference of the two operation starting times at this customer site is respected. This postponement option is implemented in the procedure responsible for the schedule determination in steps (2) as well as (6). Whenever it has to determine the starting time st_1 of a delivery / unloading operation it checks first whether starting time st_2 of the corresponding unloading operation associated with the second request in the same order has already been determined. If this is true then the earliest possible starting time is postponed as long as this postponement prevents an infeasibility with respect to the implicit time window. If the associated second operation has not yet been scheduled then the earliest possible operation starting time is assigned to the currently considered operation. If necessary, postponement is applied later to the so far unscheduled delivery operation of the currently considered order.

The schedule determining procedure employs a route-based schedule building strategy. It determines the starting times of all operations for a first route. Then it determines the starting times for all operations belonging to a second route and so on. With the goal to exploit as much postponement opportunities as possible, we sort the routes by decreasing route duration before we start the determination of the schedules for the individual vehicles, i.e. operations in the longest lasting route are scheduled prior to operations of shorter routes.

4. Computational Experiments

This section reports about the specification, preparation, execution and evaluation of computational experiments in which the aforementioned dispatching system is assessed. The achieved results enable an estimation of additional costs as well as of other impacts of considering implicit time windows in vehicle route construction.

Subsection 4.1 addresses the specification of simulation scenarios. In Subsection 4.2 we summarize the experimental setup, i.e. we outline the conducted experiments and introduce the observed performance indicators. The presentation of the observed results as well as their analysis are found in Subsection 4.3.

4.1 Specification of Test Scenarios

In order to conduct the announced computational simulation experiments we setup a collection of test scenarios. The subsequently outlined general setup applies for all these test cases and comprises a fleet of 10 vehicles. This fleet comprises five type-A vehicles and five type-B vehicles. The fleet operates in the area represented by the square $[-300; 300] \times [-300; 300]$. Initially, all vehicles are positioned at the trucking company's depot at point $(0;0)$. It is necessary that each vehicle finally returns to this location after it has completed all assigned operations.

There are two warehouses available. Each warehouse stores one commodity exclusively. Commodity A is stocked at warehouse WH-A which is located at $(-150; 150)$. Commodity B is stocked at warehouse WH-B located at $(200; -50)$.

There are 25 orders. Each order comprises two origin-to-destination requests (od-requests). Each od-request requires the pickup of a commodity quantity at the corresponding warehouse and the delivery of this commodity to an individual customer site. The two od-requests contained in an order fulfill the following two properties. First, their delivery locations coincide. Second, the first od-request requires the delivery of a type-A commodity but the second od-request is associated with a type-B commodity. Consequently, each of the 25 customer sites must be visited twice: once by a type-A vehicle and once by a type-B vehicle. We randomly draw five different sets of 25 customer locations using five different random number generator seeding values $\omega \in \Omega = \{1, \dots, 5\}$.

The fleet dispatcher has to setup a set of routes with a least possible sum of total travelled distance units for the fleet so that all 50 od-requests contained in the 25 orders are served. We are going to analyze the impacts of varying the maximal allowed difference δ between the two visits to be scheduled by the fleet dispatcher at each customer site (common length of the implicit time window for all customer sites). Three different situations are distinguished. If $\delta=\infty$ then there is no coordination of the two visits necessary. Preliminary experiments have revealed that arrival time difference larger than 500 time units are to be implemented if the travel distance sum over all vehicle routes is minimal. In a second experiment we therefore limit the maximal allowed unloading starting time difference at each customer site to $\delta=500$ time units. This enforces the fleet dispatcher to revise the least distance route set in order to fulfill the requirements of the implicit time windows of length $\delta=500$ time units at each customer site. Finally, in a third experiment, we want to analyze the impacts of enforcing the two unloading operations to the same starting time ($\delta=0$). In total, we analyze the three implicit time window lengths $\delta \in \Delta := \{\infty; 500; 0\}$.

With the intention to keep the total travel distances as short as possible a fleet dispatcher would preferentially apply a waiting strategy to achieve the implicit time window feasibility of the generated vehicle operation schedules. The insertion of waiting times at customer sites contributes to the prolongation of the makespan, which is the time span between the leaving of the first vehicle from the depot and the return of the last vehicle to the depot. The specification of a maximal allowed makespan MS^{\max} implies that the fleet dispatcher has to revise some routes in order to avoid any exceeding of MS^{\max} . In the aforementioned preliminary experiments we have seen that the maximal makespan without any time-related operation starting restriction is larger than 3000 time units. The reduction of MS^{\max} from (the referential value) of ∞ time units to 3000 time units makes a route set revision become necessary in order to comply with this the maximal allowed makespan. We investigate scenarios with the maximal allowed makespan values $MS^{\max} \in \Pi := \{\infty; 3000; 2000\}$.

In summary, following the aforementioned ideas, we setup $|\Omega| \cdot |\Delta| \cdot |\Pi| = 5 \cdot 3 \cdot 3 = 45$ different fleet dispatching (vehicle routing) scenarios. For each of the 5 customer location sets we can have 9 different time-oriented limitation sets by combining different maximal makespan values with different values for the length of the implicit time windows at the customer sites.

4.2 Setup of Experiments

In each individual simulation experiment, the genetic algorithm outlined in Section 3 is used to mimic a dispatcher that determines a feasible schedule with least travel distances for a given test case $(\omega; \delta; MS^{\max})$. The genetic algorithm is a randomized procedure so that it is necessary to apply it with several random number seeding values to the same instance in order to get a reliable average solution quality estimation. Here, each of the 45 test cases is solved by the genetic algorithm with five different seeding values. This leads to $5 \cdot 45 = 225$ individual simulation experiments.

In all experiments, we observe several performance indicator values, store them and calculate the average values for each combination $(\delta; MS^{\max})$ of the length δ of the implicit time window and the maximal allowed makespan MS^{\max} . The averagely observed travel distance is stored in $D(\delta; MS^{\max})$. Since we are interested in getting insights into the impacts of varying the length of the implicit time window, we calculate and store the relative increase $D^{\text{var}}(\delta; MS^{\max}) := D(\delta; MS^{\max}) / D(\infty; MS^{\max})$ of the travel distance that results from the reduction of δ . Similarly, we store the average number of deployed vehicles in $V(\delta; MS^{\max})$ and its relative variation in $V^{\text{var}}(\delta; MS^{\max})$. Furthermore, we save the average contribution of waiting (idle) times to the total “away from the depot”-time of the deployed vehicles and store this percentage in $W(\delta; MS^{\max})$.

In order to identify structural changes of a route set implied by the reduction of the implicit time window length it is necessary to compare routing as well as clustering decisions with and without consideration of the implicit time windows at customer sites. We can quantify the percentage of revised decisions using the H^2 -route set comparison measure proposed in Schönberger (2015a). Let $K^{\text{clust}}(\delta; MS^{\max})$ denote the percentage of revised clustering decisions (varied assignments of requests among vehicles) resulting from sharpening the length of the implicit time window from ∞ to δ . In the same way, we define $K^{\text{seq}}(\delta; MS^{\max})$ to represent the percentage of the implied sequencing decision variations.

4.3 Presentation and Discussion of Results

Figure 3 shows the solution of an example instance with unlimited makespan and an implicit time window of length $\delta = \infty$ generated by the genetic algorithm. Two vehicles are deployed, one vehicle is of type A but the other one vehicle is of type B. The type B – vehicle (route printed in gray) follows the route depot \rightarrow WH-B \rightarrow X \rightarrow W \rightarrow V \rightarrow U \rightarrow T \rightarrow S \rightarrow P \rightarrow Q \rightarrow R \rightarrow Y \rightarrow F \rightarrow E \rightarrow D \rightarrow A \rightarrow G \rightarrow H \rightarrow I \rightarrow J \rightarrow K \rightarrow L \rightarrow M \rightarrow N \rightarrow O \rightarrow B \rightarrow C \rightarrow depot. The type A – vehicle (route printed in black) travels along the route depot \rightarrow WH-A \rightarrow A \rightarrow B \rightarrow C \rightarrow D \rightarrow E \rightarrow F \rightarrow G \rightarrow H \rightarrow I \rightarrow J \rightarrow K \rightarrow L \rightarrow M \rightarrow N \rightarrow O \rightarrow P \rightarrow Q \rightarrow R \rightarrow S \rightarrow T \rightarrow U \rightarrow V \rightarrow W \rightarrow X \rightarrow Y \rightarrow depot.

Figure 3: Routes generated for an instance with unlimited makespan and implicit time window length $\delta=\infty$

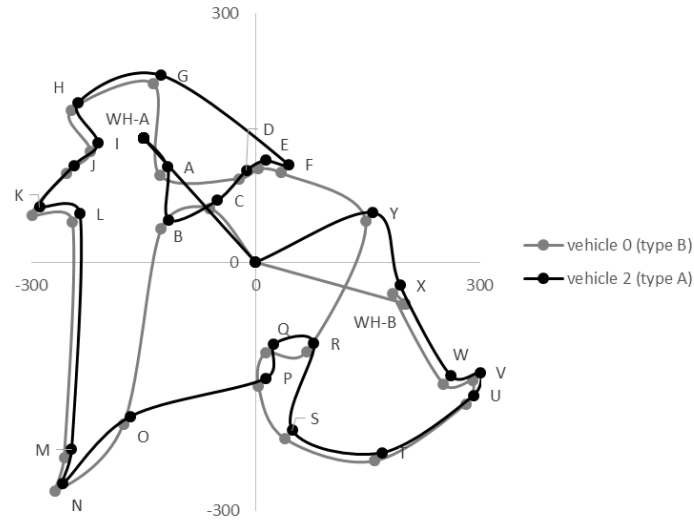


Figure 4: Routes generated for the instance from Figure 3 with unlimited makespan and implicit time window length $\delta=500$

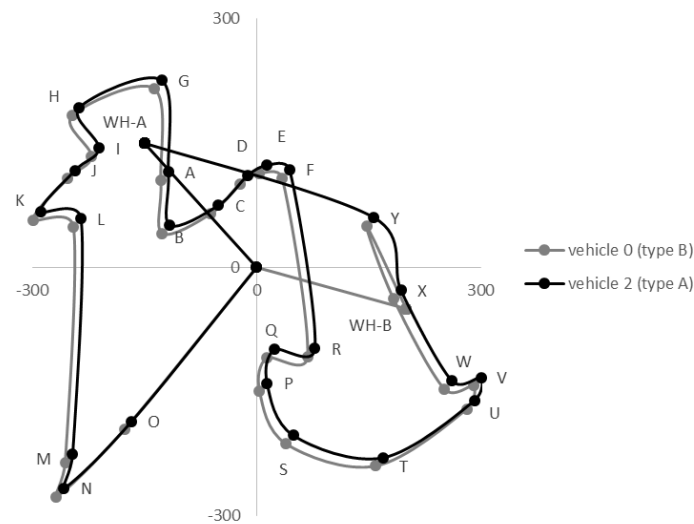


Table 1: Observed travel distances $D(\delta; MS^{\max})$ and implied travel distance increase $D^{\text{var}}(\delta; MS^{\max})$

δ	MS^{\max}		
	∞	3000	2000
∞	5696 (0%)	6317 (0%)	6565 (0%)
500	5875 (3%)	6597 (4%)	7183 (9%)
0	5940 (4%)	7370 (17%)	8269 (26%)

Table 1 summarizes the averagely observed travel distance sum as well as the relative increase if the implicit time window length δ is reduced. In the situation that the maximal makespan of a route is not restricted ($MS^{\max}=\infty$) we observe only a slight travel distance increase of 4% at most. Here, the insertion of waiting times often remedies implicit time window constraint violations. However, the insertion of waiting times does not fix the violation of an implicit time window constraint in every case. In case that a pair of customer locations is visited in two different sequences in the two associated vehicle routes, then the insertion of waiting periods does not solve the implicit time window infeasibility. For example, consider the customer locations B and W in Figure 3. In the route of the type-A-vehicle B is visited before W but in the route of the type-B-vehicle W is visited earlier than B. In addition, the time difference between the two visits at B is larger than $\delta=500$ time units. Letting the type-A vehicle wait at B would solve the conflict associated with the implicit time windows appended to B but the time window conflict at W cannot be solved. There, the arrival time difference is enlarged. To solve both conflicts it is necessary to modify the visiting sequences in at least one of the routes. This causes detours. Figure 4 shows the corresponding solution proposed for the case that $\delta=500$. The route of the type-B-vehicle is adjusted. Both vehicles follow the same visiting sequence now. Due to the need to travel first to the corresponding warehouse, the arrival times of the vehicles at the first visited customer location are different and, sometimes, the earlier arrived vehicle has to wait until the implicit time window opens to met the implicit time window requirement.

In case that the maximal allowed makespan is rather tight, then the opportunity to insert waiting times to achieve feasibility with respect to the implicit time window decreases. Now, excessive re-routing and even the incorporation of (an) additional vehicle(s) might become necessary (Figure 5 and Figure 6). Both the deviation from the shortest possible routes as well as the incorporation of additional vehicles result in additional travel distances. As it can be seen in Table 1 the total needed travel distance increases by 17% (for $MS^{\max}=3000$) resp. by 26% (for $MS^{\max}=2000$).

In summary, the consideration of implicit time windows leads to additional travel distances compared to the situation without time windows. Even in the situation without effective

maximal makespan, small travel distance increases are observed in order to meet the requirements of an implicit time window since inserted waiting periods are insufficient to achieve time window feasibility. If the maximal allowed makespan is rather short then more than 25% of additional travel distances must be realized in order to fulfill the implicit time windows.

Figure 5: Routes generated for an instance with maximal allowed makespan $MS^{\max}=2000$ and unlimited implicit time window length $\delta=\infty$

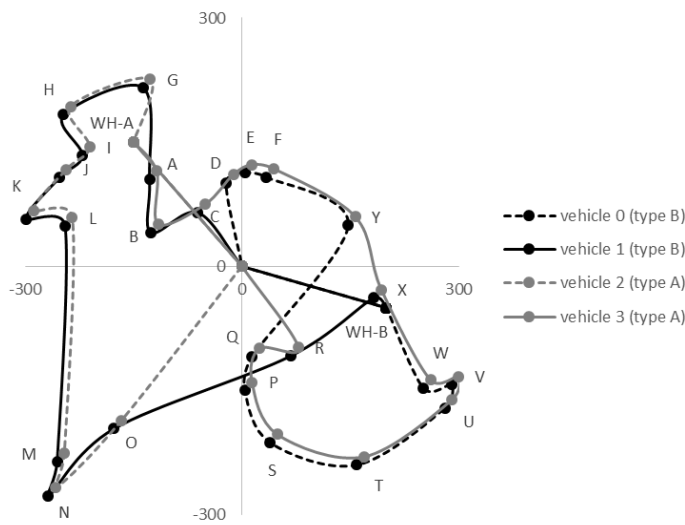


Table 2: Averagely observed number of deployed vehicles $V(\delta;MS^{\max})$ and implied increase of needed vehicles $V^{\text{var}}(\delta;MS^{\max})$

δ	MS^{\max}		
	∞	3000	2000
∞	2 (0%)	3.52 (0%)	4.16 (0%)
500	2 (0%)	3.68 (5%)	4.64 (12%)
0	2 (0%)	3.92 (11%)	5.8 (39%)

A major driver of the additional travel distances is the increase of the number of additionally needed vehicles to be deployed in order to meet the requirements of the implicit time

windows at the customer sites. Whenever a vehicle is deployed then additional travel distances for travelling back from the last customer occur. The values summarized in Table 2 show that the number of deployed vehicles can be kept stable only if the maximal allowed makespan is large ($MS^{\max}=\infty$). But as soon as the route duration is limited ($MS^{\max}=3000$) then the sharpening of the time window length implies an increase of the number of averagely deployed vehicles by up to 11%. If the maximal allowed makespan is quite tight ($MS^{\max}=2000$) then even an increase of 39% of the number of incorporated vehicles is reported.

Figure 6: Routes generated for an instance with maximal allowed makespan $MS^{\max}=2000$ and unlimited implicit time window length $\delta=\infty$

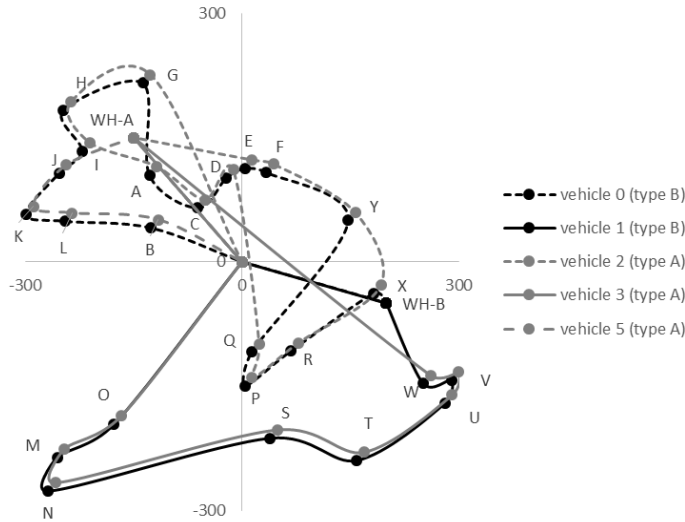


Table 3: Contribution of waiting time $W(\delta;MS^{\max})$

δ	MS^{\max}		
	∞	3000	2000
∞	0%	0%	0%
500	0%	12%	5%
0	3%	9%	12%

The usage of inserting waiting times to synchronize the arrival times of both vehicles at customer sites can be analyzed by means of the values compiled in Table 3. If the route duration is unlimited then the routes are quite long and waiting periods are relatively short so that these idle periods contribution to the total time until the return to the depot by 3% at most. In case that $MS^{\max}=3000$ then the reduction of the length of the implicit time window down to 500 time units requires additional waiting times and the contribution of these idle times summarizes up to 12% contribution to the total operational time. A further reduction of the implicit time window length implies the deployment of additional vehicles. Now, the sum of waiting period lengths is reduced but detours (in form of additional travel distances caused by the additionally incorporated vehicles) have to be implemented to meet the implicit time window requirements. Only 9% of the total operational time is spent with waiting. In case that $MS^{\max}=2000$ then the incorporation of additional vehicle does not avoid the increase of idle times. Here, up to 12% of the time vehicles are away from the depot is spent with waiting.

Table 4: Implied revision of clustering decisions $K^{\text{clust}}(\delta; MS^{\max})$

δ	MS^{\max}		
	∞	3000	2000
∞	0%	0%	0%
500	0%	39%	56%
0	0%	44%	59%

With the goal to get additional insights and a better understanding of the impacts of sharpened implicit time windows we compare (for a fixed maximal makespan value) the route sets proposed with and without ($\delta=\infty$) implicit time windows. In particular, we analyze the percentage of those pairs of requests that are served together in a common route without implicit time window consideration but which are not served in one route in the route set proposed for the situation with implicit time windows. Table 4 contains these percentage values. If the route duration is unlimited ($MS^{\max}=\infty$) then no revision of the clustering decision is necessary to comply with the implicit time windows. In case that an effective route duration limitation is stated ($MS^{\max}\leq 3000$) then more than 39% of the pairs of clustered requests are split as soon as $\delta\leq 500$. At maximum, 59% of these request pairs are split up.

Table 5: Implied revision of sequencing decisions $K^{\text{seq}}(\delta; MS^{\text{max}})$

δ	MS^{max}		
	∞	3000	2000
∞	0%	0%	0%
500	21%	48%	52%
0	21%	45%	52%

Quantifying revisions of sequencing decisions is a sophisticated task. The basic idea to determine the percentage of revised sequencing decisions is to compare again pairs of operations. However, the major challenge is that it is necessary to ensure that a pair of operations must be served by the same vehicle either in both situations for the $\delta=\infty$ -situation and for $\delta=3000$ ($\delta=2000$). Otherwise, the base for a comparison would be too small to get a meaningful statement. In Schönberger (2015a) a so-called giant-route that is defined by a consecutive arrangement of the routes of all vehicles is used to overcome this challenge. This giant route contains all operations and therefore enables the comparison of the sequence among all pairs of operations. We count the percentage of pairs of operations in which both associated operations are processed in the same sequence (but not necessarily as direct successors). From the values shown in Table 5, we observe that sequencing revisions are established also in the case of an unlimited maximal route duration. Here, 21% of sequencing decisions are updated after the length of the implicit time window at the customer sites has been sharpened. Approximately 50% of the sequencing decisions are revised if the maximal allowed makespan is so small that the insertion of waiting times is not sufficient to achieve feasibility with respect to the implicit time windows.

In summary, between 10% and more than 50% of all route generating decisions (clustering as well as sequencing decisions) are revised in order to fulfill the implicit time window constraint. The incorporation of detours is necessary and causes up to 26% additional travel distances caused by the deployment of up to 39% additionally deployed vehicles.

5. Summary, Conclusions & Outlook

In this article we addressed a sophisticated extension of the well-known and broadly-investigated vehicle routing problem. The consideration of maximal allowed delivery time differences (implicit time windows) at customer sites requiring multiple visits was in the focus of the reported research. Implicit time windows are a tool to achieve time-coordination among different delivery activities at customer sites without the need to specify the absolute position of operations on the time axis as necessary when using explicit time windows. Implicit time windows enable the exploitation of more freedom during fleet dis-

position. However, the management of implicit time windows in computer-supported automatic fleet disposition comes along with a lot of challenges.

The contribution of this article to the understanding and promotion of implicit time windows is two folded. First, we have proposed a prototypic automatic scheduling system that is able to handle implicit time windows by inserting waiting times as well as detours to enforce coupled operation starting times in the implicit time window. Second, we have used this tool to conduct simulation experiments in which we observed the impacts of considering implicit time windows during fleet disposition. These contributions answer the research questions stated in the introduction of this research report.

The introduction and/or the sharpening of implicit time windows require the revision of routing decisions. Implicit time windows help to offer logistics service providers a better and more appropriate service level. The interface performance between (external) transport processes and (internal) downstream processes in a value chain benefits from implicit time windows. However, we have seen that a significant increase of travel distances comes along with the consideration of implicit time windows. Furthermore, more vehicles are necessary to meet implicit time window requirements in time sensitive distribution scenarios.

Although we have already achieved first insights into the benefits and drawbacks of using implicit time windows in distribution logistics further research is necessary to get a deeper understanding of this coordination tool. Especially, it is necessary to investigate setups in which implicit time windows are customized, i.e. each customer defines the length of its implicit time window individually. Second, limited vehicle as well as limited handling capacities must be integrated into the fleet disposition system. Third, other objectives guiding the vehicle scheduling process require an assessment. In this context, a total cost calculation is necessary in order to estimate the surcharge a customer has to pay for being served in a shorter implicit time window. In addition, the optimization of emissions resulting from the transport operations should be assessed in the context of implicit time windows.

6. Abstract

Enterprises as well as private households exploit multi-sourcing strategies that lead to fragmented material flows and multiple deliveries. Often, customer (unloading) sites are visited several times by different vehicles of a transport service provider. In order to avoid costly setup at business customer sites and with the goal to increase the probability to meet a private customer in a first delivery attempt it is necessary to coordinate the multiple visits. In this context moveable time windows (also called implicit time windows) are proposed as coordination tool. In contrast to explicit time windows, a moveable time window does not restrict the customer visiting time to a certain part of the time axis but it only limits the time differences among the necessary unloading operations at a certain customer site. The primary goal of the here reported research comprises the evaluation of impacts on least distance vehicle routes and on the number of required vehicles implied by the consideration of im-

licit time windows from the perspective of a carrier company. We report about simulation experiments.

Multisourcing-Strategien von Unternehmen und Privathaushalten führen zu einer Fragmentierung von Transporten und Anlieferungsvorgängen im Wareneingang bzw. bei der Sendungsaushändigung. Häufig werden dadurch mehrfache Besuche eines Kunden durch verschiedene Fahrzeuge (eines Transportdienstleisters) notwendig. Zur Vermeidung von Umrüstaufwendungen bzw. zur Erhöhung der Wahrscheinlichkeit einer erfolgreichen Sendungsaushändigung ist eine zeitliche Abstimmung der Anlieferungen an Auslieferungsorten notwendig. In dieser Arbeit werden variable bzw. bewegliche Zeitfenster (sog. implizite Zeitfenster) als Koordinationsmechanismus vorgeschlagen und aus der Sicht eines Straßengüterverkehrsunternehmens evaluiert. Im Gegensatz zu explizit definierten Zeitfenstern wird der Belieferungszeitpunkt nicht vorab eingeschränkt sondern lediglich der Abstand zwischen den einzelnen Anlieferungen nach oben beschränkt. Das übergeordnete Ziel der hier berichteten Forschungsarbeiten umfasst die Abschätzung von Auswirkungen verschiedener Konfigurationen von impliziten Zeitfenstern. Anhand der in Simulationsexperimenten beobachteten streckenminimalen Transportprozesse werden die Auswirkungen verschiedener Konfigurationen von impliziten Zeitfenstern auf Fahrtrouten und benötigte Fahrzeuge abgeschätzt.

7. References

- Agrawal, N., Smith, S.A., Tsay, A.A., 2002. Multi-Vendor Sourcing in A Retail Supply Chain. *Production and Operations Management* 11(2), 157-182
- Archetti, C., Campbell, A., Speranza, M. G., 2014. Multicommodity vs. Single-Commodity Routing. *Transportation Science* (to appear).
- Archetti, C., Speranza, M. G., 2008. The split delivery vehicle routing problem: A survey. In: Golden, B., Raghavan, S., Wasil, E. (Eds.), *The Vehicle Routing Problem: Latest Advances and New Challenges*. Springer, New York, 103– 122.
- Breier, H., 2015. Tourenplanung mit alternativen Lieferperioden und Teillieferungen. KIT Scientific Publishing, Karlsruhe.
- Cattanzuza, D., Absi, N., Feillet, D., Vidal, T., 2014. A memetic algorithm for the multi trip vehicle routing problem. *European Journal of Operational Research* 236, 833–848.
- Clausen, T., 2011. Airport Ground Staff Scheduling. DTU Management Engineering, PhD Thesis.
- Crainic, T.G., Laporte, G. (Eds.), 1998. *Fleet Management and Logistics*. Springer, New York

- Crevier, B., Cordeau, J.-F., Laporte, G., 2007. The multi-depot vehicle routing problem with inter-depot routes. *European Journal of Operational research* 176, 756–773.
- de Jong, C., Kant, G., van Vliet, A., 1996. On Finding Minimal Route Duration in the Vehicle Routing Problem with Multiple Time Windows. Utrecht University, Technical Report TR-120-96.
- Doerner, K.F., Gronalt, M., Hartl, R.F., Kiechle, G., Reimann, M., 2008. Exact and heuristic algorithms for the vehicle routing problem with multiple interdependent time windows. *Computers & Operations Research* 35, 3034 – 3048.
- Drexl, M., 2012. Synchronization in vehicle routing—a survey of vrps with multiple synchronization constraints. *Transportation Science* 46(3), 297–316.
- Favaretto, D., Moretti, E., Pellegrini, P., 2007. Ant colony system for a vrp with multiple time windows and multiple visits. *Journal of Interdisciplinary Mathematics* 10 (2), 263–284.
- Goel, A., Meisel, F., 2013. Workforce routing and scheduling for electricity network maintenance with downtime minimization. *European Journal of Operational Research* 231, 210–228.
- Golden, B., Raghavan, S., Wasil, E. (Eds.), 2008. *The Vehicle Routing Problem: Latest Advances and New Challenges*. Springer, New York.
- Parragh, S., Doerner, K., Hartl, R., 2008. A survey on pickup and delivery problems. *Journal für Betriebswirtschaft* 58, 21–51.
- Schönberger, J., 2015. The two-commodity capacitated vehicle routing problem with synchronization. *IFAC-PapersOnLine* 48 (3), 168 – 173, 15th IFAC Symposium on Information Control Problems in Manufacturing (INCOM 2015)
- Schönberger, J., 2015a. Computing Indicators for Differences and Similarities among Sets of Vehicle Routes. Technical Report. Chair of Transport Services and Logistics, TU Dresden.
- Schönberger, J., 2011. *Model-Based Control of Logistics Processes in Volatile Environments*. Springer New York.
- Solomon, M. M., 1987. Algorithms for the vehicle routing and scheduling problems with time window constraints. *Operations Research* 35 (2), 254–265.
- Stieber, A., Yuan, Z., Fügenschuh, A., 2015. School Taxi Routing for Children with Special Needs. Universität der Bundeswehr Hamburg, Technical Report AMOS#19(2015).

Vidal, T., Crainic, T., Gendreau, M., Lahrichi, N., Rei, W., 2012. A hybrid genetic algorithm for multidepot and periodic vehicle routing problems. *Operations Research* 60(3), 611–624.