Justus-Liebig-Universität
Fachbereich 07 - Mathematik und Informatik, Physik und Geographie
Institut für Geographie

Methodenvergleich zur Schätzung von räumlichen Interaktionsprozessen und Prognosen von Mobilitätsströmen

Vom Fachbereich 07 der Justus-Liebig-Universität Gießen zur Erlangung des akademischen Grades Doktorin der Naturwissenschaften (Dr. rer. nat.) genehmigte Dissertation

vorgelegt von
Marie-Louise Litmeyer

Gießen, September 2022

Wissenschaftliche Betreuung:

Prof. Dr. Stefan Hennemann, Dipl. Geogr.
Professur für Wirtschaftsgeographie mit dem Schwerpunkt wissensbasierter Regionalentwicklung
Justus-Liebig-Universität Gießen

Erstgutachter: Prof. Dr. Stefan Hennemann
Zweitgutachter: Prof. Dr. Christian Diller

Danksagung

Zunächst gilt mein Dank meinem Betreuer Prof. Dr. Stefan Hennemann und meinem Zweitgutachter Prof. Dr. Christian Diller für den fachlichen Austausch, die konstruktive Unterstützung bei der Themenfindung und Umsetzung der Arbeit sowie für das Vertrauen in den letzten vier Jahren. Außerdem möchte ich mich bei Prof. Dr. Lea Schneider und Prof. Dr. Sebastian Schorcht für die Bereitschaft bedanken, Teil der Prüfungskommission zu sein.

Zudem möchte ich mich bei Michael Hollenhorst bedanken, der mit der Vergabe des Referatsthemas „Das Gravitationsmodell und seine Anwendung in der Verkehrsplanung und -prognose“ im vierten Semester mein grundlegendes Interesse an räumlichen Interaktionsprozessen geweckt hat. Mein besonderer Dank für die vielen konstruktiven Vorschläge inhaltlicher und visueller Art sowie für die Rechtschreibkorrekturen gilt Annette Briesemeister, Lisett Diehl, Sarah Hermsen, Jessica Ikerna-Müller und Dr. Jörn Profe. Ferner möchte ich mich bei meinen KollegInnen Stefanie Alsfeld, Julian Adler, Dr. Geritt Eichner, Kerstin Fischer, Carsten Klaholz, Charlotte Hanewinkel, Julia Kollotzek, Julia Jeretzky, David Pehlke, Johannes Damster, Melanie Geier, Dr. Manuel Gardt, Dr. Philipp Gareis, Jonas Goebel, Dr. Sarah Karic, Lukas Kriesch, Martin Kohl, Dr. Thomas Kolb, Christin Müller, Sergej Nesterenko, Dr. Natalie Schmiede, Mareike Schnorr, Moritz Schütz, Saskia Thorbecke, Barbara und Katharina Thörner, Niklas van der Au, Dr. Frank Volker, Gabriele Weiß, Dr. Daniela Wilden und Lisa Zirbes für den fachlichen und privaten Austausch sowie die technische Unterstützung in den letzten Jahren bedanken. Mein Leben ohne euch wäre wie ein metrischer Raum, in dem mindestens eine Cauchy-Folge nicht konvergiert. Dank gilt auch meinen Studentischen Hilfskräften Lilly, Jule K., Fabian, Jan, Jule R., Julian, Tim, Yeliz für die Unterstützung in unterschiedlichen Lehrveranstaltungen und Projekten.

Bedanken möchte ich mich ferner bei Egon Rücker, Edgar Hackspiel, dem Team der Landesstelle für Statistik Vorarlberg und der Abteilung Raumplanung und Baurecht sowie Julia Salvatori für die Möglichkeit „das Ländle”kennen und lieben zulernen.

Abschließend möchte ich mich bei meinen Eltern und meiner Schwester bedanken. Ohne eure liebevolle Unterstützung, wäre es nicht möglich gewesen die Schule, das Studium und diese Arbeit erfolgreich abzuschließen. Es ist gut, dass ihr um die Bedeutung von Besenheide und Kiefernspitzen wisst.

Zusammenfassung

Die vorliegende Dissertation hat das Ziel, einen Methodenvergleich zur Schätzung von räumlichen Interaktionsprozesse und Prognosen von Mobilitätsbewegungen durchzuführen. Als Untersuchungsgegenstand dienen dabei Studienanfänger und Pendler.

Klassischerweise wurden Schwerkraftmodelle zur Schätzung der Mobilitätsströme verwendet. In den letzten Jahren wurden jedoch eine Reihe neuer Modelle, z. B. Strahlungsmodelle zur Schätzung der menschlichen Mobilität eingeführt. Der Schwerpunkt lag dabei im Allgemeinen auf Modellen, die sich mit Pendlerbewegungen befassen. Die Mobilitätsströme wurden auf Grundlage der Einwohnerzahl am Ziel- und Quellort sowie zwischen den Regionen berechnet. Ein systematischer Vergleich für unterschiedliche Länder existiert jedoch nicht. Ähnlich verhält es sich mit Mobilitätsströmen von Studienanfängern. Diese wurden bisher nicht mit den neu entwickelten Modellen prognostiziert. Daher werden mittels der wenigen öffentlich zugänglichen Daten anhand von Teilchenmodellen Vorhersagen auf unterschiedlichen räumlichen Ebenen und für ausgewählte europäische Länder gemacht. Weiter werden die verschiedenen Modelle miteinander verglichen. Die besten Resulate erzielen dabei die Modelle, welche eine Kombination aus Strahlungs- und Gravitationsmodell darstellen. Insgesamt weisen alle Teilchenmodelle jedoch Schwächem auf, da sie die individuellen Mobilitätsmotive nicht berücksichtigen können. Eine Möglichkeit, diese unterschiedlichen Motive besser abzubilden, besteht darin, Vorhersagen mithilfe von Algorithmen des maschinellen Lernens zu treffen. Es können so Aspekte wie z. B. die Beschäftigungsquote oder Aspekte der Lebensqualität berücksichtigt werden. Die Ergebnisse machen deutlich, dass insbesondere gegenüber des Gravitations- und Strahlungsmodells deutliche Verbesserungen bei der Prognoseleistung erzielt werden können.

Insgesamt bietet die Anwendung der Modelle auf die studentische Mobilität trotz vorhandener Limitationen (z. B. Datenverfügbarkeit) die Möglichkeit, zuverlässige Prognosen der studentischen Mobilitätsströme auf der Mikroebene zu berechnen, mittel- bis langfristige Entscheidungen auf Hochschulebene zu treffen und eine nachhaltige strategische Ausrichtung umzusetzen. Ähnlich verhält es sich mit den Pendlerflüssen. Diese können genutzt werden, um auf kleinräumiger Ebene Schätzungen vorzunehmen und somit langfristige Angebote, zum Beispiel im ÖPNV-Bereich, zu schaffen.

Abstract

Gravity models have always been used as a standard to estimate mobility flows. In recent years however, a number of new models, such as radiation models, have been introduced to estimate human mobility.

The aim of this dissertation is to compare the methods for the estimation of spatial interaction processes and forecasts of mobility movements. First-semester students and commuters serve as the objects of this study.

Generally, the scientific community has focused on modeling commuter movements. Mobility flows have been calculated based on population at the destination, the source, and between regions. However, a systematic comparison for different countries does not exist. Mobility flows of first-year students show many similarities. These have not yet been forecast using the newly developed models. Therefore, using the few publicly available data based on the particle models, predictions are made on different spatial levels and for selected European countries. In addition the different models are compared with each other.

The best results were achieved by the models representing a combination of radiation and gravitation models. As a whole, all particle models show weaknesses, since they cannot take the individual mobility motives into account. One way to better represent these different motives is to make predictions using machine learning algorithms. In this way, aspects such as the employment rate or quality of life can be anyed. The results clearly show that significant improvements in forecasting performance can be analysed, especially when compared to gravity and radiation models.

Overall, despite existing limitations (e.g. data availability), the application of the models to student mobility offers the possibility to calculate reliable forecasts of student mobility flows at the micro level, to make medium- to long-term decisions at the university level, and to implement a sustainable strategic orientation. The same goes for commuter flows. These can be used to make estimates at the small-scale level and thus create long-term offers, for example in the public transport sector.

Inhaltsverzeichnis

Danksagung **I**

Zusammenfassung **II**

Abstract **III**

Inhaltsverzeichnis **IV**

Abbildungsverzeichnis **VIII**

Tabellenverzeichnis **IX**

Abkürzungsverzeichnis **X**

1 Einleitung **1**

2 Theoretischer Rahmen **5**
2.1 Begriffserklärung 5
2.2 Interaktionen in Städtesystemen 6
2.3 Klassische Migrationsmodelle 17
2.4 Motive für räumliche Mobilität 21
2.5 Instrumente zur Prognose von räumlichen Interaktionen 26

3 Hypothesen **37**

4 Datengrundlage **41**

5 Mobilitätsprognosen für Studienanfänger **53**

6 Mobilitätsprognosen für Pendlerbewegungen **67**

7 Zusammenfassung der empirischen Ergebnisse und Schlussfolgerungen **157**
7.1 Diskussion der Prognoseergebnisse 157
7.2 Anwendungen der Prognoseinstrumente in der Praxis 164
7.3 Limitationen 165
7.4 Zukünftiger Forschungsbedarf 165

8 Fazit **168**

9 Literaturverzeichnis **169**

Abbildungsverzeichnis

Abb. 1 Mobilitätstypen von Studierenden 6
Abb. 2 Entwicklung eines monozentralen hin zum polyzentralen Städtesystem 9
Abb. 3 Lokales Rauschen und trans-lokales Pfeifen 16
Abb. 4 Vergleich von s_{ij} und S_{ij} am Beispiel der Pendlerstrecke Fulda-Gießen (Kreisebene) 33
Abb. 5 Übersicht der untersuchten Länder in Europa 47
Abb. 6 Vergleich von Schulabgängern und Studienanfängern 53
Abb. 7 Der Sörensenindex im Zeitverlauf 55
Abb. 8 Der Pearson-Korrelationskoeffizient im Zeitverlauf 55
Abb. 9 Der adjustierte R^2-Wert im Zeitverlauf 55
Abb. 10 Der mittlere absolute Fehler im Zeitverlauf 56
Abb. 11 Die Wurzel der mittleren quadratischen Abweichung im Zeitverlauf 56
Abb. 12 Abweichung zwischen den empirischen Daten und den Modellen (Studienanfänger) 57
Abb. 13 P(n) im Verhältnis zur Größe des Hochschulstandorts . . . 58
Abb. 14 P(Distanz) im Verhältnis zur Entfernung (Mobilität von Studienanfängern) 59
Abb. 15 Zeitliche Entwicklung der Parameter 61
Abb. 16 SHAP-Werte für den XGBoost-Regressor zur Prognose von studentischer Mobilität (Grund- und Erweiterungsmodell) 63
Abb. 17 SHAP-Werte für den XGBoost-Regressor zur Prognose von studentischer Mobilität (Individualmodell) 64
Abb. 18 Ergebnisse der Simulation für die Mobilität von Studienanfängern 65
Abb. 19 Ergebnisse der Evaluationsindikatoren für studentische Mobilitätsbewegungen mittels XGBoost-Algorithmus 66
Abb. 20 Abweichung zwischen den empirischen Daten und den Modellvorhersagen bezüglich der Pendlerbewegungen auf Gemeindeverbandsebenen 68
Abb. 21 P(Distanz) im Verhältnis zur Entfernung (Pendlerbewegung auf Gemeindeverbandsebene) 69
Abb. 22 Ergebnisse der Evaluationsindikatoren für verschiedene Länder in Europa 72
Abb. 23 Ergebnisse der Simulation für Deutschland (Gemeindeverbandsebenen) 73
Abb. 24 Ergebnisse der Evaluationsindikatoren mittels XGBoost-Algorithmus für verschiedene Länder in Europa 75
Abb. 25 SHAP-Werte für den XGBoost-Regressor bei Pendlerbewegungen auf Gemeindeverbandsebenen (Grund- und Erweiterungsmodell) 77
Abb. 26 SHAP-Werte für den XGBoost-Regressor bei Pendlerbewegungen auf Gemeindeebenen (Individualmodell) 78

Abb. 27 Abweichung zwischen den empirischen Daten und den Modellvorhersagen bezüglich der Pendlerbewegungen auf Kreisebene . . . 80
Abb. 28 P(Distanz) im Verhältnis zur Entfernung (Pendlerbewegung auf Kreisebene) . . . 82
Abb. 29 Ergebnisse der Simulation Deutschland (Kreisebene) . . . 83
Abb. 30 SHAP-Werte für den XGBoost-Regressor bei Penderbewegungen auf Kreisebenen (Grund- und Erweiterungsmodell) . . . 85
Abb. 31 SHAP-Werte für den XGBoost-Regressor bei Penderbewegungen auf Kreisebenen (Individualmodell) . . . 86
Abb. 32 Abweichung zwischen den empirischen Daten und den Modellvorhersagen bezüglich der Pendlerbewegungen für England und Wales . . . 88
Abb. 33 P(Distanz) im Verhältnis zur Entfernung (Pendlerbewegung in England und Wales) . . . 89
Abb. 34 Ergebnisse der Simulation für England und Wales . . . 90
Abb. 35 SHAP-Werte für den XGBoost-Regressor bei Pendlerbewegungen in England und Wales (Grund- und Erweiterungsmodell) . . . 92
Abb. 36 Abweichung zwischen den empirischen Daten und den Modellvorhersagen bezüglich der Pendlerbewegungen für Frankreich . . . 94
Abb. 37 P(Distanz) im Verhältnis zur Entfernung (Pendlerbewegung in Frankreich) . . . 95
Abb. 38 Ergebnisse der Simulation für Frankreich . . . 96
Abb. 39 SHAP-Werte für den XGBoost-Regressor bei Pendlerbewegungen in Frankreich (Grund- und Erweiterungsmodell) . . . 98
Abb. 40 Abweichung zwischen den empirischen Daten und den Modellvorhersagen bezüglich der Pendlerbewegungen für Irland . . . 100
Abb. 41 P(Distanz) im Verhältnis zur Entfernung (Pendlerbewegung in Irland) . . . 101
Abb. 42 Ergebnisse der Simulation für Irland . . . 103
Abb. 43 SHAP-Werte für den XGBoost-Regressor bei Pendlerbewegungen in Irland (Grund- und Erweiterungsmodell) . . . 104
Abb. 44 Abweichung zwischen den empirischen Daten und den Modellvorhersagen bezüglich der Pendlerbewegungen für Österreich . . . 106
Abb. 45 P(Distanz) im Verhältnis zur Entfernung (Pendlerbewegung in Österreich) . . . 107
Abb. 46 Ergebnisse der Simulation für Österreich . . . 109
Abb. 47 SHAP-Werte für den XGBoost-Regressor bei Pendlerbewegungen in Österreich (Grund- und Erweiterungsmodell) . . . 110

Abb. 48 Abweichung zwischen den empirischen Daten und den Modellvorhersagen bezüglich der Pendlerbewegungen für die Schweiz . . . 112
Abb. 49 P(Distanz) im Verhältnis zur Entfernung (Pendlerbewegung in der Schweiz) . . . 113
Abb. 50 Ergebnisse der Simulation für die Schweiz . . . 115
Abb. 51 SHAP-Werte für den XGBoost-Regressor bei Pendlerbewegungen in der Schweiz (Grund- und Erweiterungsmodell) . . . 116
Abb. 52 Abweichung zwischen den empirischen Daten und den Modellvorhersagen bezüglich der Pendlerbewegungen für Dänemark . . . 118
Abb. 53 P(Distanz) im Verhältnis zur Entfernung (Pendlerbewegung in Dänemark) . . . 119
Abb. 54 Ergebnisse der Simulation für Dänemark . . . 121
Abb. 55 SHAP-Werte für den XGBoost-Regressor bei Pendlerbewegungen in Dänemark (Grund- und Erweiterungsmodell) . . . 122
Abb. 56 Abweichung zwischen den empirischen Daten und den Modellvorhersagen bezüglich der Pendlerbewegungen für Finnland . . . 124
Abb. 57 P(Distanz) im Verhältnis zur Entfernung (Pendlerbewegung in Finnland) . . . 125
Abb. 58 Ergebnisse der Simulation für Finnland . . . 126
Abb. 59 SHAP-Werte für den XGBoost-Regressor bei Pendlerbewegungen in Finnland (Grund- und Erweiterungsmodell) . . . 127
Abb. 60 Abweichung zwischen den empirischen Daten und den Modellvorhersagen bezüglich der Pendlerbewegungen für Norwegen . . . 129
Abb. 61 P(Distanz) im Verhältnis zur Entfernung (Pendlerbewegung in Norwegen) . . . 130
Abb. 62 Ergebnisse der Simulation für Norwegen . . . 132
Abb. 63 SHAP-Werte für den XGBoost-Regressor bei Pendlerbewegungen in Norwegen (Grund- und Erweiterungsmodell) . . . 133
Abb. 64 Abweichung zwischen den empirischen Daten und den Modellvorhersagen bezüglich der Pendlerbewegungen für Schweden . . . 135
Abb. 65 P(Distanz) im Verhältnis zur Entfernung (Pendlerbewegung in Schweden) . . . 136
Abb. 66 Ergebnisse der Simulation für Schweden . . . 137
Abb. 67 SHAP-Werte für den XGBoost-Regressor bei Pendlerbewegungen in Schweden (Grund- und Erweiterungsmodell) . . . 138

Abb. 68 Abweichung zwischen den empirischen Daten und den Modellvorhersagen bezüglich der Pendlerbewegungen für Kroatien 140
Abb. 69 P(Distanz) im Verhältnis zur Entfernung (Pendlerbewegung in Kroatien) 141
Abb. 70 Ergebnisse der Simulation für Kroatien 143
Abb. 71 SHAP-Werte für den XGBoost-Regressor bei Pendlerbewegungen in Kroatien (Grund- und Erweiterungsmodell) 144
Abb. 72 Abweichung zwischen den empirischen Daten und den Modellvorhersagen bezüglich der Pendlerbewegungen für Italien 146
Abb. 73 P(Distanz) im Verhältnis zur Entfernung (Pendlerbewegung in Italien) 147
Abb. 74 Ergebnisse der Simulation für Italien 149
Abb. 75 SHAP-Werte für den XGBoost-Regressor bei Pendlerbewegungen in Italien (Grund- und Erweiterungsmodell) 150
Abb. 76 Abweichung zwischen den empirischen Daten und den Modellvorhersagen bezüglich der Pendlerbewegungen für Spanien 152
Abb. 77 P(Distanz) im Verhältnis zur Entfernung (Pendlerbewegung in Spanien) 153
Abb. 78 Ergebnisse der Simulation für Spanien 155
Abb. 79 SHAP-Werte für den XGBoost-Regressor bei Pendlerbewegungen in Spanien (Grund- und Erweiterungsmodell) . . . 156

Tabellenverzeichnis

Tab. 1 Variablenübersicht 29
Tab. 2 Erwarteter Einfluss der Variablen auf die studentische Mobilität aus Perspektive i 44
Tab. 3 Erwarteter Einfluss der Variablen auf das Pendlerverhalten aus Perspektive i 46
Tab. 4 Übersicht über die Hypothesen 163

Abkürzungsverzeichnis

Abb.	Abbildung
bzw.	beziehungsweise
ca.	circa
d. h.	das heißt
EC-Bahnhof	EuroCity-Bahnhof
et al.	et alii
etc.	et cetera
f.	eine folgende Seite
ff.	zwei folgende Seiten
IC-Bahnhof	InterCity-Bahnhof
ICE-Bahnhof	InterCityExpress-Bahnhof
km	Kilometer
km^2	Quadratkilometer
KMK	Kultusministerkonferenz
m	Meter
ÖPNV	öffentlicher Personennahverkehr
S.	Seite
Tab.	Tabelle
u. a.	unter anderem
vgl.	vergleiche
z. B.	zum Beispiel

1 Einleitung

Räumliche Mobilität und Interaktion sind ein wesentlicher Bestandteil unserer heutigen Gesellschaft. Besonders mobil sind die Gruppen der Bildungswanderer und Berufstätigen im Alter zwischen 18 und 30 Jahren (Gans 2015, S. 15; Simons/Weiden 2016, S. 265). Hauptursache für die Wanderung dieser Altersgruppe ist der Beginn einer Ausbildung bzw. eines Studiums oder der Einstieg in die Berufswelt. Im Fokus stehen dabei häufig Agglomerationsräume. Diese sind in den letzten zwei Jahrzehnten in Deutschland durch ein starkes Wachstum gekennzeichnet (Volgmann et al. 2021), da urbane Wohnstandorte zunehmend attraktiv für Haushalte geworden sind. Zudem existieren in Ballungsräumen besonders viele Arbeitsplätze in wissensintensiven Branchen (Moos et al. 2019, S. 1076). Das heißt, eine zunehmende Akademisierung verstärkt die Zuwanderung in Agglomerationsräumen (Jessen/Siedentop 2018, S. 2075; Florida et al. 2017, S. 86). Daher lohnt sich zunächst ein Blick auf die Gruppe der Studierenden.

Für Deutschland ist festzustellen, dass seit Anfang der 2000er Jahre die Studierendenzahlen (Gruppe der Bildungswanderer) stark gestiegen sind (Multrus et al. 2017, S. 43). Ursache hierfür sind unter anderem die massive Expansion des Hochschulangebotes, der steigende Anteil an Abiturientinnen und Abiturienten, die Einführung des Bachelor- / Mastersystems, die Abschaffung der Bundeswehrpflicht sowie doppelte Abiturjahrgänge (Deutscher Bundestag 2011; KMK 2011; KMK 2012). Das heißt, um das zukünftige Wachstum von städtischen Räumen besser abschätzen zu können, sind vor allem zuverlässige Prognosen von studentischer Mobilität notwendig. Zwar existieren Studien, die ein Wachstum an Studierenden prognostizierten, dennoch wurde der Anstieg auf nationaler Ebene deutlich unterschätzt (Nutz 1991; KMK 2000; Gösta/von Stuckrad 2007; Wissenschaftliche Dienste des Deutschen Bundestages 2006). Ähnliche Vorhersagefehler sind auf Ebene der Bundesländer, z. B. für Mecklenburg-Vorpommern (vgl. Kramer 2005) und Sachsen-Anhalt (Lischka/Kreckel 2006), gemacht worden.

Neben Auswirkungen auf das Wachstum in Agglomerationsräumen hat diese Unterschätzung der Studierendenzahlen insbesondere für Hochschulen eine große Relevanz. Ursache ist, dass die Grundfinanzierung und somit auch Arbeitsplätze im Bereich der Forschung und Lehre an die Anzahl der Studierenden geknüpft ist (HMWK 2015, S. 10 ff.). Zuverlässige Prognosen der Mobilitätsströme auf Mikroebene eröffnen daher die Möglichkeit, mittel- bis

langfristige Entscheidungen zu treffen und eine nachhaltige strategische Ausrichtung durchzuführen (z. B. Schaffung von neuem Wohnraum, Einrichtung von neuen Professuren oder Schwerpunkten an Hochschulen).

Doch neben den zunehmenden Urbanisierungsprozessen sind auf Grund von hohen Bodenpreisen in zentralen Lagen auch Suburbanisierungsprozesse zu beobachten (Adam 2019, S. 37). Insbesondere in der Phase der Familiengründung wird ein Umzug in suburbane oder ländliche Gebiete angestrebt (Growe/Münter 2010; Albrecht/Scheiner 2022). Diese Beobachtung führt zu einer steigenden Anzahl an Pendlern, da der Wohnortwechsel nicht zwangsläufig zu einer Verlagerung des Arbeitsplatzes führt.

Grundsätzlich ist die Pendlerbereitschaft in den letzten Jahren europaweit stark gestiegen (Aguilera 2005; Einig/Pütz 2007; Eurostat 2022). Für Deutschland konnte festgestellt werden, dass im Jahr 2020 bereits 49,9 % der Erwerbstätigen 10 bis 30 Minuten zum Arbeitsplatz pendeln und für weitere 22,2 % betrug der Fahrweg zwischen 30 und 60 Minuten (Statistisches Bundesamt 2022). Diese Beobachtung bietet Chancen und Risiken: Während zunehmendes Pendeln aktuell Probleme wie Fachkräftemangel und Landflucht entgegenwirkt (Epifani/Gancia 2005; Kersting et al. 2020), birgt die erhöhte Pendlerbereitschaft allerdings auch neue Probleme. So muss die tägliche Wanderung der Beschäftigten auch vor dem Hintergrund des Klimawandels diskutiert werden. Insbesondere in urbanen Gebieten ist die Luftverschmutzung durch den motorisierten Individualverkehr ein besonders großes Problem (Mansfield et al. 2015). Daher müssen nachhaltige Alternativen geschaffen werden, um die Umweltbelastungen zu senken und die Luftqualität zu verbessern. Darüber hinaus birgt die zunehmende Pendlerbereitschaft auch auf individueller Ebene Gefahren. Staus in Verbindung mit dem motorisierten Individualverkehr führen häufig zu Stress bzw. Zeitdruck und können mittelfristig medizinische Probleme wie Bluthochdruck und psychische Probleme erzeugen (Hansson 2011; Kersting et al. 2020). Das heißt, es müssen Angebote und Lösungen für Pendler entwickelt werden, um Emissionen zu senken und die individuellen Gefahren zu minimieren. Hauptaugenmerk wird in diesem Zusammenhang auf der Verringerung der Fahrzeit und einer Verbesserung des ÖPNV-Angebots liegen. Insbesondere vor diesem Hintergrund ist es wichtig, zuverlässige Prognosen auf möglichst kleinräumiger Ebene vorliegen zu haben, um langfristige politische Entscheidungen treffen zu können.

Als Instrument zur Vorhersage dieser Mobilitätsströme werden seit mehr als

100 Jahren Gravitationsmodelle eingesetzt. Die Grundannahme der Gravitationsmodelle basiert auf dem Newtonschen Gesetz. Die Interaktion zwischen zwei Standorten ist proportional zu der Größe des Standortes (z. B. Einwohnerzahl etc.) und die Anziehung nimmt mit zunehmender Entfernung ab (Carey 1858; Zipf 1946; Wilson 1970; Erlander/Stewart 1990). Diese universelle Gravitation kann auf viele Bereiche, z. B. auf das Verkehrswesen (Jung et al. 2008; Kaluza et al. 2010; Odlyzko 2015; Hong/Jung 2016), auf Bevölkerungsströme und Pendelbewegungen (Griffith 2009; Murat 2010; Lenormand et al. 2012; Liang et al. 2013; Masucci et al. 2013; Thomas/Tutert 2013; Lenormand et al. 2016; Liu et al. 2014), auf den Gütertransport (Kaluza et al. 2010), auf den internationalen Handel (Bergstrand 1985; Fagiolo 2010), auf die Telekommunikation (Krings et al. 2009), auf die wissenschaftliche Zusammenarbeit (Pan et al. 2012) und auf die Betrachtung von menschlicher Mobilität im Zusammenhang mit infektiösen Krankheitsverbreitungen (Viboud et al. 2006; Balcan et al. 2009; Tizzoni et al. 2014; Sallah 2017; Marshall et al. 2018; Tuite et al. 2018) übertragen werden. Auch im Zusammenhang mit Migrationsströmen von Studenten wurden Gravitationsmodelle zur Schätzung verwendet (SÁ et al. 2004; Alm/Winters 2009; Cooke/Boyle 2011; Faggian/Franklin 2014).

Die bisher in Studien eingesetzten Gravitationsmodelle weisen jedoch einige Schwächen auf. In der Praxis werden unterschiedliche Funktionen und bis zu neun Parameter zur Anpassung an die empirischen Daten eingesetzt (Viboud et al. 2006; Balcan et al. 2009; Kaluza et al. 2010; Krings et al. 2009; Simini et al. 2012). Hinzu kommt, dass empirische Daten zum Vergleich bzw. zur Schätzung vorliegen müssen. Daher können nur Vorhersagen für Regionen durchgeführt werden, für die bereits systematische Vergleichsdaten existieren. Ferner weist das Gravitationsmodell durch fehlende Berücksichtigung von Standorten zwischen Ziel- und Quellgebiet eine systematische Vorhersage-Diskrepanz auf. Das heißt, für Interaktionen zwischen zwei Standorten mit ähnlichen Herkunfts- und Zielmaßen und vergleichbaren Entfernungen werden ähnlich große Flüsse prognostiziert, ohne den Zwischenraum zu berücksichtigen. Ferner steigt die prognostizierte Wanderungsbewegung bei Erhöhung der Standortmaße der Zielregion unbegrenzt an. Außerdem kann das klassische Gravitationsmodell keine Schwankungen bei Interaktionen zwischen zwei Standorten berücksichtigen (Simini et al. 2012). Aus diesem Grund entwickelten Simini et al. (2012) das klassische Strahlungsmodell als alternativen Ansatz zur Schätzung der Mobilität zwischen zwei Standorten. Vorteil dieses

Ansatzes ist zum einen die geringe Menge benötigter Daten und zum anderen die Parameterfreiheit. Auf Grundlage der Überlegungen von Simini et al. (2012) wurden in den vergangenen Jahren weitere Modellalternativen entwickelt (vgl. u. a. Masucci et al. 2013; Yan et al. 2014; Ren et al. 2014; Kang et al. 2015; Lenormand et al. 2016; Liu/Yan 2019).

Außerdem gelang es Morton et al. (2018) und Spadon et al. (2019) mit Algorithmen des maschinellen Lernens (XGBoost-Algorithmus) Pendlerinteraktionen in den USA und Brasilien zu schätzen. Morton et al. (2018) prognostizierten jedoch lediglich in wie weit eine Verbindung exisitert. Spadon et al. (2019) berechneten zusätzlich noch die Gewichte der Verbindung und konnten somit die Stärke des Stroms vorhersagen. Eine vergleichende Untersuchung von den Varianten des Strahlungsmodells und den Prognosen mittels XGBoost-Algorithmus auf kleinräumiger Ebene in Europa existiert allerdings bisher nicht.

Insgesamt hat daher die vorliegende Dissertation das Ziel, die unterschiedlichen Modelle miteinander zu vergleichen und das beste Modell für die Prognose studentischer Migration und Pendlerbewegungen in ausgewählten Ländern Europas zu identifizieren.

Zunächst wird der theoretische Rahmen der Dissertation vorgestellt. Hierzu wird eine Definition der Begriffe Migration, studentische Mobilität und Pendeln gegeben, um die verwendeten Begriffe klar abgrenzen zu können. Im Anschluss folgt ein Kapitel zum Thema Interaktion im Städtesystem. Diese Kapitel hat das Ziel, die Verteilung von Städten im Städtesystem zu erläutern und die Bedeutung von Agglomerationsräumen für das Wirtschaftswachstum herzuleiten. Im weiteren Verlauf werden die klassischen Migrationsmodelle kurz vorgestellt, um die grundsätzliche Mikro- und Makroperspektive von menschlicher Interaktion zu verstehen. Zudem werden aus empirischen Studien die Motive für die Mobilität von Studierenden und das Pendlen abgeleitet. Es folgt in Kapitel 3 und 4 die Herleitung der Hypothesen und die Vorstellung der Datengrundlage. Anschließend wird in Kapitel 5 zunächst eine Schätzung der Mobilitätsflüsse von Studierenden auf Kreisebene durchgeführt. Es folgt in Kapitel 6 eine Prognose von Pendlerbewegungen für Deutschland und weitere ausgewählte Länder in Europa mit dem Ziel, die Mobiltätsbewegungen mithilfe der Modelle bestmöglich zu schätzen. Die Diskussion der Ergebnisse, Handlungsempfehlungen, Limitationen, der weitere Forschungsbedarf sowie das Fazit wird in Kapitel 7 und 8 dargestellt.

2 Theoretischer Rahmen

2.1 Begriffserklärung

Beschäftigt man sich mit Mobilität von Pendlern und Studierenden, so muss zunächst der Begriff Migration definiert werden. Migration wird in der Mobilitätsliteratur sehr unterschiedlich diskutiert. Grundsätzlich wird in inter- un intraregionale Mobilität unterschieden (Zax 1994). Bei Greenwood (1975, 1985) und Bartel (1979) wird implizit davon ausgegangen, dass Migration sowohl auf inter- als auch auf intraregionaler Mobilität basiert, während Siegel (1975) dies explizit annimmt. Linneman/Graves (1983) betrachten hingegen sowohl inter- als auch intraregionale Mobilität als Migration (Zax 1994).

Grundsätzlich stammt das Wort Migration vom lateinischen Wort *migrare* ab und kann mit „seinen Wohnort wechseln“ übersetzt werden (Zax 1994). Das heißt, Migration wird in dieser Dissertation wie im engeren Sinne als Wechsel (egal, ob auf inter- oder intraregionalen Ebene) des Wohnortes verwendet.

Demgegenüber steht der Begriff „pendeln“ bzw. die Mobilität von Arbeitskräften. Haas/Osland (2014) definieren pendeln als regelmäßige wiederkehrende Reise zwischen Wohn- und Arbeitsort (Haas/Osland 2014). Das heißt, Personen, die täglich zu ihrem Arbeitsplatz fahren, werden als Pendler bezeichnet.

Eine Spezialform der räumlichen Interaktion ist die studentische Mobilität. In der internationalen Literatur ist bis ca. 1990 häufig der Begriff Migration verwendet worden. Heute wird die Bezeichnung studentische Mobilität genutzt. Unter studentischer Mobilität wird in jüngeren empirischen Arbeiten die Wanderung vom Standort, an dem die Hochschulzugangsberechtigung erworben wurde, zum Hochschulstandort verstanden (Gareis 2019).

Allerdings existieren weitere Differenzierungsmöglichkeiten. Bailey (2015, S. 73 ff.) unterscheidet in seiner Arbeit zu Mobilitätsprozessen von Studierenden im Vereinigten Königreich acht verschiedene Mobilitätstypen (vgl. Abbildung 1). Er differenziert hierzu in Heimat- und Wohnort sowie Hochschulstandort. Es existieren zunächst interne Studierende bzw. Zuwanderer (vgl. Abbildung 1 (a)). Sie stammen ursprünglich aus einer anderen Region, der Wohn- und Hochschulstandort ist jedoch gleich. Sofern die Heimatregion außerhalb des Hochschullandes liegt, spricht man auch von international migrierenden Studierenden (vgl. Abbildung 1 (d)). Zudem gibt es lokale Studierende. Für sie

gilt, dass der Heimatort dem Wohnort und Hochschulstandort entspricht (vgl. Abbildung 1 (b)). Außerdem gibt es pendelnde Studierende (vgl. Abbildung 1 (c)). Das heißt, Heimat- und Wohnort liegen außerhalb der Hochschulregion. Wenn der Wohnort zudem ebenfalls nicht in der Heimatregion liegt, spricht Bailey (2015) von Binnenmigranten und Pendlern (vgl. Abbildung 1 (e)). Für international migrierende Pendler gilt, dass der Wohnort außerhalb der Hochschulregion und der Heimatort nicht im Hochschulland liegt (vgl. Abbildung 1 (f)). Darüber hinaus existieren migrierende Pendler, sofern der Wohnort zwar außerhalb der Hochschulregion angegeben wird, der Heimatort jedoch innerhalb dieser liegt (vgl. Abbildung 1 (g)).

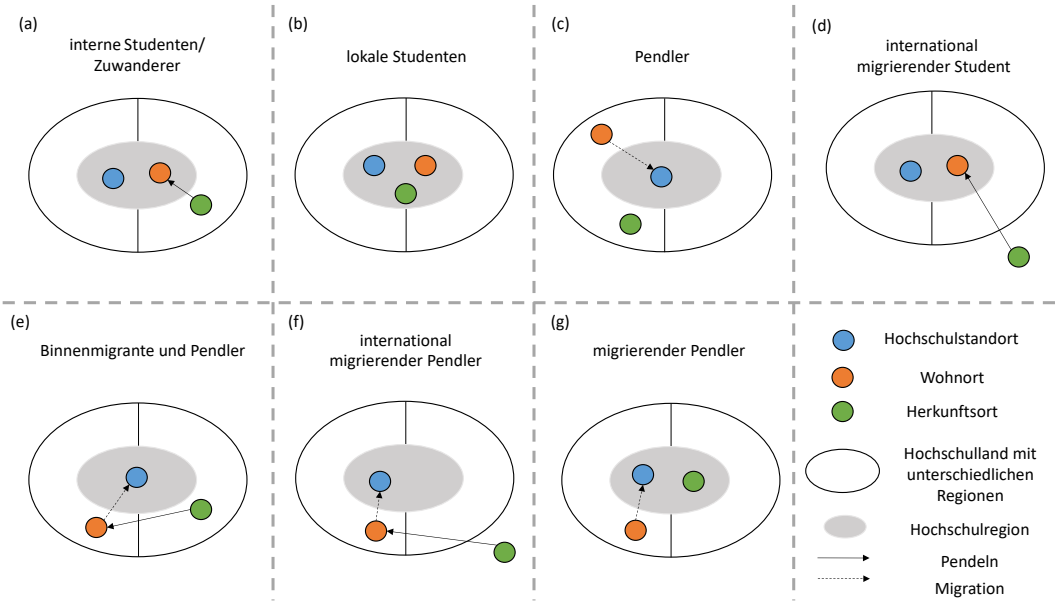


Abbildung 1: Mobilitätstypen von Studierenden. Quelle: Eigene Darstellung angelehnt an Bailey 2015.

2.2 Interaktionen in Städtesystemen

Pendeln bzw. Migration wird im Wesentlichen durch drei Determinanten (Arbeit, Lohn, Wohnraum) eingegrenzt. In den Industrieländern macht der Anteil der Bevölkerung im erwerbsfähigen Alter den größten Teil der Gesamtbevölkerung aus. Zudem ist das Einkommen, das durch Arbeit erwirtschaftet wird, ein wichtiger Bestandteil des Gesamteinkommens einer Volkswirtschaft (Cahuc/Zylberg 2004). Dieses Einkommen wird zum großen Teil dazu genutzt, Mieten für Wohnraum zu bezahlen und Kreditrückzahlungen für Eigenheime zu leisten (Haas/Osland 2014). Das heißt, aus der wirtschaftlichen Perspektive ist es am sinnvollsten, sich dort niederzulassen, wo die entspre-

chenden Arbeitsplätze existieren, um die Pendlerkosten zu minimieren. Die Arbeitsplätze sind oftmals in Agglomerationsräumen zu verorten. Diese sind zwar häufig durch gute Verdienstmöglichkeiten gekennzeichnet, weisen allerdings hohe Bodenpreise auf. Das bedeutet, die räumlichen Unterschiede der Wohnungspreise und Einkommensaussichten beeinflussen die Pendlerentfernung entscheidend.

Das Modell des Stadtökonomen Alonso (1964) greift diese Überlegungen auf und berücksichtigt Wohn- und Arbeitsmarkt sowie die Entscheidung über die Dauer des Pendelns in gleicher Weise (Rouwendal/Meijer 2001). Alonso (1964) geht davon aus, dass lediglich ein Beschäftigungszentrum (CBD) existiert (monozentrische Struktur). Der jeweilige Haushalt sucht seinen Wohnstandort so aus, dass der möglichst größte Nutzen zwischen Pendlerkosten, Grundstückspreisen und Lohnniveau besteht. Das heißt, Arbeitnehmer wählen die Entfernung zwischen Wohn- und Arbeitsort möglichst gering. Da allerdings in den Beschäftigungszentren die Grundstückspreise bzw. die Kosten für Wohnraum bedingt durch eine große Nachfrage relativ hoch sind, nehmen Haushalte weitere Entfernungen und hohe Pendlerkosten in Kauf. Diese werden jedoch durch niedrigere Grundstückspreise im Umland und höhere Löhne im Stadtzentrum ausgeglichen (Haas/Osland 2014, S. 466).

Ein Konzept, das diese monozentrischen Modelle in Frage stellt, stammt von Hamilton (1982). Er konnte für Städte in den USA und Japan feststellen, dass das tatsächliche Pendleraufkommen deutlich höher ist als das mit den monozentrischen Modellen prognostizierte. Dieses Phänomen wird auch als das Problem des „übermäßigen Pendelns“ bezeichnet. Nimmt man zudem an, dass Personen ihren Wohn- und Arbeitsort nach dem Zufallsprinzip wählen, beträgt das durchschnittliche tatsächliche Pendelaufkommen in etwa drei Viertel des zufälligen Pendelns. Ferner führt eine Dezentralisierung der Beschäftigung zu keiner Verringerung der Pendlerzahl, sondern im Durchschnitt eher zu einer leichten Zunahme (Hamilton 1982). Doch obwohl das „übermäßige Pendeln“ die Annahmen des stadtökonomischen Standardmodells grundsätzlich in Frage stellt, bleibt dennoch die grundlegende Idee in Kraft: Arbeitnehmer verteilen sich unter Abwägung von hohen Wohnungspreisen und Pendlerkosten um die Beschäftigungsschwerpunkte herum (Rouwendal/Meijer 2001).

Räumliche Mobilität von Arbeitskräften kann somit nicht losgelöst von der räumlichen Verteilung von Städten und dem Städtesystem (z. B. polyzentrische Städtesysteme) diskutiert werden (Haas/Osland 2014).

Der funktionale Begriff „Städtesystem“ wurde ursprünglich von Berry (1964) eingeführt und Pred (1977, S. 13) definierte *system of cities* als eine interagierende nationale oder regionale Gruppe von Städten.

Die Analysemöglichkeit dieser Städte ist vielfältig. Ein Versuch das Städtesystem eines Landes oder einer Region mit mathematischen Gesetzmäßigkeiten zu beschreiben, ist die sogenannte *Rank-Size-Rule* (Auerbach 1913; Zipf 1949). Die Idee ist, dass die Größenverteilung der Städte eine gewisse Regelmäßigkeit aufweist und oftmals das System aus vielen kleinen und nur wenigen großen Städten besteht. Der deutsche Geograph Auerbach konnte 1913 zeigen, dass die Einwohnerzahl einer Stadt unter Berücksichtigung der Größe proportional zu ihrem Rang ist. Voraussetzung ist allerdings, dass alle Städte eines Landes nach der Größe geordnet werden. Das bedeutet, die fünftgrößte Stadt eines Landes hat eine fünfmal kleinere Einwohnerzahl als die größte Stadt. Benannt wurde dieses Gesetz nach dem amerikanischen Linguisten Zipf (1949).

Zudem versuchten Wissenschaftler wie Christaller Mitte des 20. Jahrhunderts mit der Theorie zu den zentralen Orten regelmäßige, wiederkehrende räumliche Muster von Städtewachstum und -verteilung zu identifizieren und zu beschreiben. Idee des Modells ist es, eine hierarchische Raumstruktur für die Bedeutung von Orten bzw. Städten zu entwickeln. Gemessen wird der Bedeutungsüberschuss (auch Zentralität) einer Stadt, indem das Angebot an zentralen Gütern und Dienstleistungen ins Verhältnis zum Umland gesetzt wird (Christaller 1968, S. 26). Als optimale räumliche Verteilung zentraler Orte und ihrer Einzugsgebiete wird ein hexagonales, wabenförmiges Muster angesehen. Allerdings unterlagen diese noch starken Einschränkungen. Christaller nahm zum Beispiel an, dass eine homogene Raumstruktur ohne natürliche Barrieren vorliegt und die Bevölkerung gleich über den Raum verteilt ist (Christaller 1968, S. 71).

Grundsätzlich reichen die Ausprägung des städtischen Systems von einer völlig monozentrischen Struktur bis zu einem vollständig polyzentrischen System (Batten 1995).

In monozentralen Städtesystemen dominiert eine Stadt das gesamte System. In der Regel liegt dies bei einer historisch gewachsenen Stadt vor, die von Wohngebieten umgeben ist. Die Funktionen Wohnen und Arbeiten sind eindeutig getrennt und die Arbeitsplätze befinden sich aus ökonomischen Gründen im Zentrum. Das Umland wird vor allem für die Bereiche Wohnen und Freizeit ge-

nutzt (Growe 2012). Das heißt, monozentral ausgerichtete Städte weisen die zentralörtliche Funktion im „klassischen“ Sinne auf und verfügen über einen eindeutigen Verflechtungsraum (Limtanakool et al. 2007; Winkel/Greivinger 2008; Flex et al. 2016).

Für Polyzentralität als eines der Schlüsselkonzepte der Stadt- und Regionalforschung gibt es bis heute keine einheitliche Definition. Als minimale Anforderung für eine polyzentrische Region kann ein räumlich zusammenhängendes Gebiet mit mindestens zwei Zentren angesehen werden (Münter/Volgmann 2021, S. 678). Zudem werden weitere Begriffe für polyzentrale Regionen verwendet, unter anderem *„network cities“* (Batten 1995), *„city networks“* (Camagni/Salone 1993).

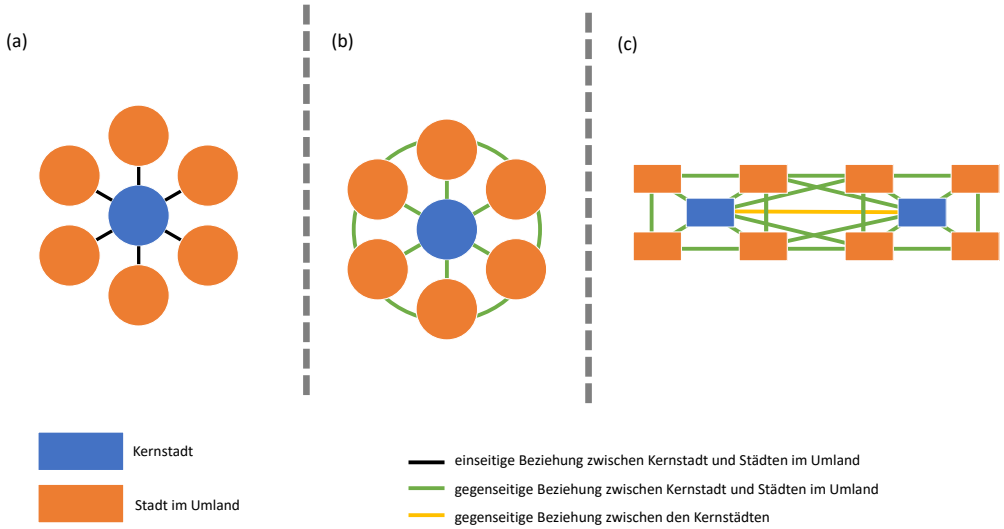


Abbildung 2: Entwicklung eines monozentralen hin zum polyzentralen Städtesystem. Quelle: Eigene Darstellung angelehnt an Growe 2012.

Ausgangspunkt für die Entwicklung von polyzentrischen Regionen ist häufig eine monozentrische Struktur (vgl. Abbildung 2 (a)) des Städtesystems (De Goei et al. 2010, S. 1152; Growe 2012, S. 297). De Goei et al. (2010) beschreiben diesen Prozess von einer monozentrischen hin zu einer polyzentrischen Struktur als Regionalisierung. Er wird unter anderem durch günstige Bodenpreise und eine hohe Lebensqualität angetrieben (van der Laan 1998; Growe 2012, S. 297). Die Idee ist, dass sich ehemalige suburbane Standorte mit der ursprünglichen Funktion Wohnen zu sekundären Beschäftigungszentren entwickeln (vgl. Abbildung 2 (b)). Das heißt, die ehemaligen Vororte gewinnen an Bedeutung. Die kleinen Zentren werden zum

einen unabhängiger von der Kernstadt und zum anderen sind sie immer anziehender für Einpendler aus anderen suburbanen Regionen oder dem Zentrum (Growe 2012, S. 297). Eine der Folgen von diesem Prozess ist die Entstehung von interurbanen polyzentrischen Regionen.

Typischerweise sind interurbane polyzentrische Regionen (vgl. Abbildung 2 (c)) durch mehrere historisch und administrativ getrennte, größere und kleinere Städte gekennzeichnet. Die Entfernung zwischen diesen Stadtzentren beträgt ungefähr die tägliche Pendlerdistanz. Zudem weisen die größeren Städte dieser Regionen eine vergleichbare Einwohnerzahl sowie ähnliche wirtschaftliche und politische Funktionen auf (Kloosterman/Musterd 2001; Münter/Volgmann 2021, S. 678). Außerdem sind Umland und Kernstadt ein verschmolzenes Städtesystem (Champion 2001; Parr 2014). Das heißt, in interurbanen polyzentrischen Regionen existieren keine räumlichen Strukturhierachien. Die Agglomerationseffekte treten nicht zentriert auf (Meijer 2005 & 2016; Wiechmann et al. 2016). Ein typisches Beispiel ist die Rhein-Ruhr-Region oder die Industrieregion in Oberschlesien. Aber auch Stadtregionen mit einem dominierenden Kern und weiteren großen Zentren (z. B. Rhein-Main-Gebiet oder der Großraum um Mailand) können als interurbane polyzentrische Regionen bezeichnet werden (Münter/Volgmann 2021, S. 680).

Ein Konzept, welches versucht, den Bedeutungsüberschuss von kleinen Städten innerhalb einer Metropolregion zu erklären, stammt von dem Ökonomen William Alonso (Growe 2012, S. 297). Er entwickelte das *„borrowed size“* Konzept in den 1970er Jahren. Dieses versucht innerhalb eines metropolitanen Stadtkomplexes die Diskrepanz zwischen Größe und Funktion kleinerer Städte zu erklären. Die Grundidee ist, dass kleine Städte innerhalb eines metropolitanen Städtesystems deutlich besser dastehen als es sich durch ihre Größe erklären lässt. Als Ursache nennt Alonso den Zugang zu Agglomerationsvorteilen von größeren Nachbarstädten (Alonso 1973). Allerdings profitieren nicht alle kleineren Städte innerhalb des Systems. Um nennenswerte Vorteile aufzuweisen, ist vor allem die räumliche Nähe und die Zugehörigkeit zum funktionalen Verflechtungsraum besonders wichtig. Das bedeutet, sofern Städte in der Nähe des jeweiligen Zentrums liegen und/oder ein Teil des funktionalen Verflechtungsraumes sind, profitieren sie von Nachbarschaftseffekten. Daher müssen Netzwerkbeziehungen analysiert werden, um die Größe und das Wachstum einer Stadt erklären zu können. Während jedoch für Alonso noch die Agglomerationsvorteile und die geographische Nähe im Vordergrund standen, betonen Meijers/Burger (2017) die Bedeutung der Interaktion zwischen

den Akteuren des Netzwerkes.

Die Netzwerk-Metapher tritt im planerischen Konzept bzw. in der regionalwissenschaftlichen und geographischen Literatur relativ häufig im Zusammenhang mit polyzentrischen Stadtregionen auf (Meijers 2005, S. 766; Reggiania et al. 2011). Im Vordergrund stehen die Synergieeffekte innerhalb der Netze. Das heißt, es müssen wirtschaftliche Netzwerktheorien in Bezug auf das Prinzip der Synergie betrachtet werden.

Das griechische Wort *„Synergie“* beschreibt das Phänomen, dass der Effekt von zwei oder mehr kombinierten Einrichtungen oder Funktionen größer ist als die Summe der jeweiligen einzelnen Betrachtung (Meijers 2005, S. 766). Bei Capello/Rietveld (1998, S. 64) werden drei verschiedene Bedeutungen von Synergie herausgearbeitet. Zunächst kann von Synergie gesprochen werden, wenn zwei oder mehr Akteure zusammenarbeiten und ein positives Ergebnis für beide erzielt werden kann. Außerdem liegen Synergien vor, wenn eine Zusammenarbeit sinnvoll ist, um die Komplementarität bei der Produktion eines bestimmten Gutes zu nutzen und Vorteile zwischen Wirtschaftsakteuren zu erzielen. Die dritte Bedeutung von Synergie ergibt sich im Zusammenhang von Gruppen. Sofern Einzelpersonen oder Unternehmen Teil einer Gruppe sind (freiwillig oder unfreiwillig), können Externalitäten vorhanden sein, die ebenfalls einen positiven Einfluss haben (Capello/Rietveld 1998, S. 64; Meijers 2005, S. 767).

Ein weiterer Aspekt ist die Art des Netzwerkes. In Club-Netzwerken fokussieren die Akteure ein gemeinsames Ziel. Sie sind zum einen durch Aktivität oder Dienstleistung miteinander verbunden. Zum anderen haben sie auch parallele Interessen und Transaktionsketten (z. B. Mitglieder eines Tennisklubs). In diesem Fall spricht man von sogenannten „horizontalen Synergien“. Die Skaleneffekte und positiven Netzwerkexternalitäten ergeben sich aus der Zusammenarbeit. Typischerweise gelten die Größenvorteile nur für die Teilnehmer des Netzes. Sofern die Kosten der Beteiligung am Netzwerk geringer sind als der Nutzen der Kooperation, liegen externe Effekte vor (Capineri/Kamann 1998; Meijers 2005, S. 768 f.).

Web-Netzwerke sind hingegen durch unterschiedliche Aktivitäten der Akteure geprägt und es treten sogenannte „vertikale Synergien“ auf. Typischerweise sind die Aktivitäten komplementär zueinander (z. B. Kette von Unternehmen, die am Herstellungsprozess eines Produktes beteiligtet sind). Bei den „vertikalen Synergien“ basiert der Vorteil vor allem auf den Agglomerations- oder Spe-

zialisierungseffekten. Durch die Umverteilung von Ressourcen und Aktivitäten besteht die Möglichkeit, dass jeder Beteiligte entsprechend seiner Kernkompetenz am Prozess mitwirken kann. Somit wird eine Verbesserung der individuellen Leistung gewährleistet (Capineri/Kamann 1998; Meijers 2005, S. 768 f.). Zusammenfassend lässt sich feststellen, dass Synergien einen wichtigen Beitrag für die wirtschaftlichen Vorteile von Netzwerken leisten.

Überträgt man diese Überlegungen auf städtische Systeme, muss zunächst die Definition eines Netzwerkes vorgenommen werden. Die Städte des Systems sind die Knotenpunkte des Netzes, wobei jede einzelne Stadt erneut aus einer Vielzahl an Knotenpunkten besteht (z. B. Haushalte, Unternehmen etc.). Die Kanten beschreiben beispielsweise die Infrastruktur oder Mobilitätsströme. Das bedeutet, in polyzentrischen Stadtregionen sind eine Reihe an unterschiedlichen Netzwerken auf verschiedenen Maßstabsebenen vorhanden, die alle von Agglomerationsvorteilen profitieren (Meijers 2005, S. 768).

Zudem konnte gezeigt werden, dass polyzentrische Stadtregionen sowohl clubartige Aspekte als auch webartige Dimensionen aufweisen. Camagn/Salone (1993, S. 1059) nutzen als Synonyme für Club- und Webnetzwerke in Städtesystemen auch Synergie- und Komplementärnetzwerke.

Ein Clubnetzwerk liegt vor, wenn eine Bündelung von Kräften aus vergleichbaren Städten stattfindet. Das Ziel ist, die Skaleneffekte und Größenvorteile zu nutzen, um gemeinsame Ziele und Interessen zu erreichen. Typisches Beispiel sind Städte, die eine vergleichbare wirtschaftliche Rolle im System annehmen und daraus resultierende ähnliche städtische Probleme oder Herausforderungen aufweisen (z. B. Hafenstädte, Tourismusstädte). Die webartige Dimension von Netzwerken ist darüber hinaus ebenfalls gegeben, da es möglich ist, dass einzelne Städte unterschiedliche wirtschaftliche Funktionen ausüben und komplementäre Infrastruktureinrichtungen vorhalten. Doch obwohl beide Arten an Netzwerken auftreten, scheinen vor allem webartige Netzwerke eine besondere Rolle im polyzentrischen Städtesystem zu spielen. Die räumliche Nähe, die zu einer Überschneidung von Marktgebieten führt, wird als Hauptgrund angesehen (Meijers 2005).

Im Vordergrund stehen neben der morphologischen Perspektive von Polyzentralität auch die Verflechtung (z. B. im Zeitverlauf) und Beziehung zwischen den Funktionen Arbeit, Freizeit und Wohnen (Münter/Volgmann 2021, S. 683), wobei sowohl die Richtung der Ströme zwischen den Zentren als auch die Stärke ihrer Interaktion gemessen werden muss. Als Messinstrument für

die Untersuchung der funktionalen räumlichen Organisation einer Region dienen typischerweise die Pendlerströme. Ursache ist, dass andere Daten, die die soziale Interaktion der Einwohner einer Region widerspiegeln, oftmals nicht vorliegen (Burger et al. 2013). Alternativ werden auch unternehmensinterne Kommunikations- und Interaktionsströme oder auch Wertschöpfungsketten zwischen Unternehmen (z. B. Burger/Meijers 2012) eingesetzt. In jüngerer Zeit bieten zudem neue Datenquellen (z. B. anonymisierte Mobilfunkdaten) Möglichkeiten, die komplexe soziale Strukturen in der Gesellschaft zu analysieren (Schläpfer et al. 2014).

Ein weiteres Konzept, welches in Bezug auf interurbane Beziehungen und Vernetzung betrachtet werden muss, ist die Komplementarität (vgl. u. a. Ullmann 1956; Camagni/Salone 1993). Die Idee ist, dass sich Städte (auf der Makroebene) gegenseitig ergänzen und komplementäre Aufgaben übernehmen. Flex et al. (2016) bezeichnen dieses Phänomen auch als sogenannte Mehrfachzentren oder Städteverbünde. Die Idee basiert auf der gemeinsamen Funktionsnutzung zweier oder mehrerer Gemeinden. Besonders großen Vorteil bietet dieser Ansatz, wenn jede der beteiligten Kommunen Versorgungsleistungen für die jeweils andere übernimmt (Flex et al. 2016, S. 111). So ist es beispielsweise möglich, dass eine Stadt ein Hallenbad vorhält, während in der Nachbarstadt eine Eisbahn existiert (z. B. Dornbirn und Feldkirch in Vorarlberg (Österreich)). Daraus folgt, die Bewohner beider Städte können beide Angebote nutzen, sofern sie bereit sind, die jeweils andere Stadt aufzusuchen. Demzufolge bedingt Komplementarität räumliche Interaktion. Ullmann (1956) betrachtet die Komplementarität sogar als die wichtigste Erklärung für die Entwicklung räumlicher Interaktion. Zudem stellt Batten (1995) klar, dass vor allem die komplementären Funktionen und nicht die Entfernung oder Nachfrageschwellen die Verbindungen zwischen den Städten in einer polyzentrischen Stadtregion erklären.

Dijst/Cortie (1988) gehen davon aus, dass Interaktionen in städtischen Systemen durch drei S-Dimensionen (*strength, symetry, structure*) charakterisiert werden. Laut Simmons (1986, S. 26) sind Städte mit starker Interaktion durch einen großen Austausch gekennzeichnet. Das heißt, eine Übertragung von Wissen ist leichter möglich. Die Existenz von starken Verflechtungen ist somit das wichtigste Element des Städtesystems (Friedmann 1978; Bourne/Simmon 1978; Growe 2012, S. 297).

Zudem können Interaktionen zwischen Städten symmetrisch (Ströme in bei-

de Richtungen sind gleich stark) und asymmetrisch (es existiert lediglich der Strom in eine Richtung) vorliegen. Typischerweise treten asymmetrische Interaktionen vor allem in monozentrischen Systemen auf. Die Kernstadt, die über die meisten und spezialisiertesten Funktionen verfügt, ist Anziehungspunkt für weniger wichtige Städte. Umgekehrt existiert diese Beziehung jedoch nicht (Growe 2012, S. 297).

Polyzentrische Regionen sind hingegen durch symmetrische Interaktionen geprägt (Limtanakool et al. 2007). Der dritte Aspekt ist die Struktur des Systems. Sinclair (1983, S. 108) unterscheidet in hierarchische (monozentrische Systeme) und nicht-hierarchische Systeme (polyzentrische Systeme). Typisches Element der nicht-hierarchischen Systeme ist die Diffusion von Impulsen im gesamten System.

Generell ist davon auszugehen, dass einwohnerstarke Städte bedingt durch ihre Konnektivitätsstruktur stärker als einwohnerschwache Städte von Interaktionen zwischen Variablen wie zum Beispiel Beschäftigung, Bildung, Innovation, Löhne, Einkommen und Kriminalität profitieren (Bettencourt et al. 2007). Das heißt, laut Bettencourt et al. (2007) steigt die Produktivität und Innovationsfähigkeit einer Stadt überproportional mit der Stadtgröße. Schläpfer et al. (2014) zeigt mithilfe von anonymisierten Mobilfunkdaten für Portugal und dem United Kingdom, dass größere Städte überproportional dichte Kommunikationsnetze aufweisen. Zudem konnten auch Glaeser (1992) und Krugman (1996) die Dichtevorteile von Ballungszentren nachweisen. Russo et al. (2007, S. 791) betont, dass Unternehmen in der wissensbasierten Branche neben räumlicher Nähe zu anderen Unternehmen eine Vielzahl weiterer Angebote wie zum Beispiel Finanzdienstleistungen, hochqualifizierte Arbeitskräfte, Wissensquellen (Universitäten und Forschungszentren), logistische Dienstleistungen etc. benötigen. Das bedeutet, Agglomerationsvorteile spielen eine entscheidende Rolle bei der Wahl des Unternehmensstandortes.

Grundsätzlich wird in Urbanisations- und Lokalisationsvorteile unterschieden. Diese beiden Ansätze versuchen Agglomerationsvorteile von Städten und Ballungsregionen zu erklären. Jacobs (1969) sieht „Vielfalt“ als den wichtigsten Motor für Innovation an (Urbanisationsvorteil). Unternehmen unterschiedlicher Branchen beeinflussen sich gegenseitig positiv. Sie profitieren unter anderem von Wissens- und Informationsflüssen, einem diversifizierten Arbeitsmarkt sowie der kollektiven Nutzung von Infrastruktur.

Marschall (1890), Arrow (1962) und Romer (1986) später zusammenge-

fasst von Glaser et al. (1992) und heute bekannt als MAR-Konzept leiten die Agglomerationsvorteile von Städten hingegen aus der Spezialisierung vom Markt ab (Lokalisationsvorteil). Die Spezialisierung fördert den Austausch von Wissen, Ideen und Informationen und steigert so die Produktivität von Unternehmen. Das heißt, Unternehmen derselben Branche können bei regionaler Konzentration gegenseitig von Wissensspillover gewinnen (Reggiani et al. 2011, S. 528).

Ein weiterer Aspekt, der Berücksichtigung findet, ist die Verfügbarkeit von Humankapital. Es wird generell davon ausgegangen, dass in Großstädten bzw. Ballungsgebieten eine große Qualifikationsheterogenität vorherrscht und eine Vielzahl von Arbeitsplätzen in wissensintensiven Branchen existiert (O'Clery et al. 2019). O'Clery et al. 2019 gehen sogar davon aus, dass die Vielzahl an verfügbaren Qualifikationen der Hauptgrund für die Agglomerationsvorteile sind.

Ferner wird bei Fritsch und Slavtchev (2007) deutlich, dass wissenschaftliche Erkenntnisse eine entscheidende Rolle für den Innovationsprozess und die wirtschaftliche Entwicklung spielen. Zudem konnten auch Schubert und Kroll (2014) zeigen, dass Hochschulen einen großen Beitrag zu regionalen Wirtschaftsaktivitäten leisten. Die Exsistenz einer Hochschule in einer Region führt zu einem Anstieg des BIPs pro Kopf und einer Senkung der lokalen Arbeitslosigkeit (Schubert/Kroll 2014, S. 484 f.).

Daraus folgt, Universitäten und Hochschulen als Zentren für Forschung und Entwicklung sowie die Lehre und Ausbildung von Studierenden nehmen dabei eine besondere Position im (regionalen) Innovationssystem ein. Während sie zum einen Wissen generieren und anderen Akteuren zur Verfügung stellen, ist die Ausbildung von Studierenden zum anderen ein wichtiger Aspekt für den Arbeitsmarkt. Die Privatwirtschaft profitiert durch die gut ausgebildeten Absolventen sowie von dem entsprechenden Wissen der Universitäten und kann somit ihre Innovationsfähigkeit verbessern (Fritsch/Slavtchev 2007, S. 201 ff.).

Ein weiterer Aspekt ist, dass Kooperation von Unternehmen der Privatwirtschaft mit Universitäten, wissenschaftliche Veröffentlichungen, Seminare, Workshops und informelle Beziehungen den Transfer von akadamischen Wissen in die Wirtschaft positiv beeinflussen (Fritsch/Slavtchev 2007). Akademisches Wissen ist dabei jedoch relativ immobil, daher spielt die räumliche Nähe eine besondere Rolle. Absolventen und Wissenschaftler arbeiten häufig

in der Nähe ihrer Heimatuniversität und neu gegründete innovative Firmen suchen ebenfalls aktiv die räumliche Nähe zu Universitäten (Bade/Nerlinger 2000; Audretsch et. al. 2004; Audretsch/Lehmann 2005; Fritsch/Slavtchev 2007). Zudem verbleiben Absolventen, die während des Studiums Arbeits- und Berufserfahrung in der Hochschulregion sammeln konnten, häufiger in der Region und die Abwanderung ist in dieser Gruppe geringer (vgl. Teichert et al. 2020).

Doch trotz der Bedeutung von räumlicher Nähe reichen interne Strukturen und Beziehungen des Netzes langfristig nicht aus, um die Wettbewerbsfähigkeit einer Region bzw. eines Clusters zu steigern. Daher greift die Theorie des lokalen Rauschens und globalen Pfeifens (vgl. Abbildung 3) neben den *local buzz*-Effekten auch externe Impulse, sogenannte trans-lokale Beziehungen auf. Ziel ist, dass die Akteure versuchen, gezielte trans-lokale oder globale Beziehungen (*global pipelines*) aufzubauen, um neue Verbindungen und Kontakte zu anderen Wissenssystemen zu erhalten und langfristig dadurch Wirtschaftswachstum zu generieren (Bathelt et al. 2004).

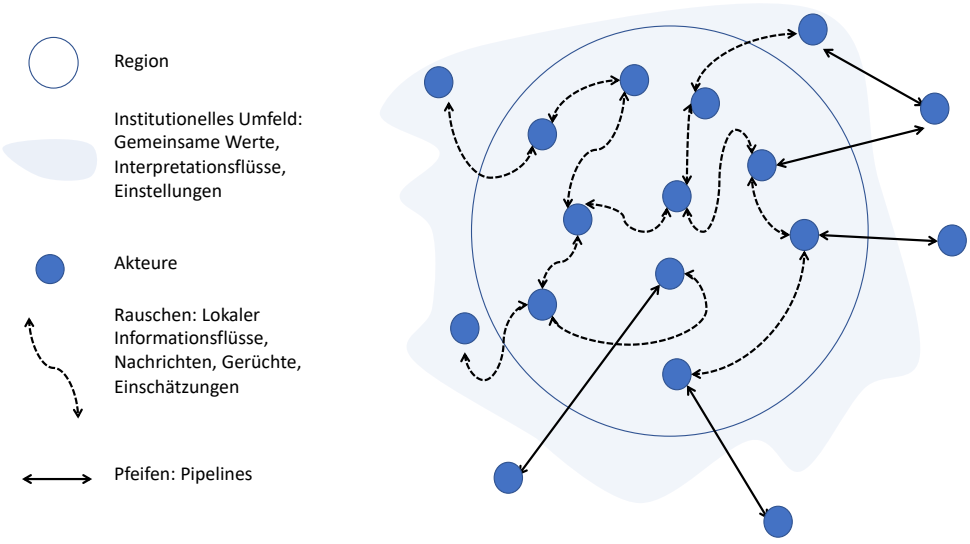


Abbildung 3: Lokales Rauschen und trans-lokales Pfeifen. Quelle: Eigene Darstellung nach Bathelt et al. 2004.

Zusammengefasst lässt sich feststellen, dass die Interaktion zwischen Städten bzw. Städtesystemen und deren Akteuren bei der Wirtschaftswachstumsanalyse eine entscheidende Rollen spielt und sogar als „Motor des Wachstums“ bezeichnet werden kann (Reggiani et al. 2011; Nijkamp 2008). Außerdem ist davon auszugehen, dass insbesondere Großstädte auch in Zukunft eine wich-

tige Quelle des Wirtschaftswachstums sein werden (Quigley 1998, S. 137).

Insgesamt sind neben dem Verständnis von Städtesystemen, dem Erklärungsansatz der Komplementarität sowie der Bedeutung von Interaktion für die wirtschaftliche Entwicklung eines Landes weitere Aspekte heranzuziehen, damit Interaktion zwischen Städten und somit räumliche Mobilität erklärt werden kann. Deshalb lohnt sich ein Blick auf die gängigen theoretischen Beschreibungen von Migrationsprozessen.

2.3 Klassische Migrationsmodelle

Makroökonomische Perspektive

Ausgangspunkt bildet die Arbeit *„The Laws of Migration“* von Ravenstein (1885, 1889). Er stellte insgesamt zehn Gesetze zur Wanderung auf, um Migrationsbewegungen zu erklären. Zunächst sind laut Ravenstein (1885) Wanderungen zwischen Standorten mit kleineren Entfernungen zu erkennen. Lange Distanzen werden oftmals überwunden, um große, wirtschaftlich starke Zentren aufzusuchen (Gesetz 1). Außerdem erfolgt die Migration schrittweise (Gesetz 2). Das bedeutet, die Zuzüge speisen sich zunächst aus nahegelegenen Ortschaften. Zudem bedingt die Wanderung in der Regel eine Gegenwanderung (Gesetz 3). Ferner weist die ländliche Bevölkerung im Vergleich zur städtischen Bevölkerung eine höhere Bereitschaft auf, ihren Wohnort zu wechseln (Gesetz 4). Darüber hinaus sind geschlechtsspezifische Unterschiede festzustellen. Männer migrieren im internationalen Kontext deutlich häufiger als Frauen, während auf nationaler Ebene die Wanderbereitschaft von Frauen höher ist (Gesetz 5). Außerdem handelt es sich um alleinstehende Erwachsene (Gesetz 6). Große Städte wachsen in der Regel durch Zuwanderung und nicht nur durch das natürliche Bevölkerungswachstum (Gesetz 7). Diese zunächst sieben Grundregeln wurden im Jahr 1889 um drei weitere ergänzt (Ravenstein 1889). Typischerweise steigt mit einer positiven wirtschaftlichen Entwicklung und gut ausgebauten Transport- und Infrastruktursystemen die Zuwanderung (Gesetz 8). Zudem ist auf Grund ökonomischer Bedingungen eine Wanderung aus agrarwirtschaftlich geprägten Regionen in die wirtschaftlich starken Zentren zu erkennen (Gesetz 9 & 10).

Hauptkritikpunkt an diesen Wanderungsgesetzen von Ravenstein ist, dass sie sich aus statistischen Beobachtungen ableiten. Das heißt, Hintergründe und Mechanismen, die die Migrationsbewegungen beschreiben, finden keine Berücksichtigung (Kalter 2000, S. 442).

Ein weiterer Ansatz ist die Anwendung von Gravitationsgesetzen auf das Wanderungsverhalten. Die Grundidee basiert auf den physikalischen Gravitationsgesetzen Newtons. Nach Zipf (1946) gilt: Je mehr Einwohner eine Stadt aufweist, desto anziehender ist sie. In Bezug zur Distanz kann festgestellt werden, dass mit zunehmender Entfernung zwischen zwei Städten die Wahrscheinlichkeit für Interaktion abnimmt. Insgesamt ist das Wanderungsverhalten propotional zu $\frac{mn}{r}$, wobei m und n die Einwohner in der Ziel- und Quellregion repräsentieren, während r für die Distanz steht. Da dieses Basismodell jedoch empirisch an seine Grenzen stößt, entwickelte Dodd (1950) eine Variante, in der zusätzlich eine Migrationskonstante, das Aktivitätsniveau pro Kopf in der Bevölkerung sowie die Länge des betrachteten Zeitintervalls mit berücksichtigt werden (Dodd 1950, S. 245). Das Aktivitätsniveau entspricht dabei dem Dichtekoeffizienten im physikalischen Gravitationsgesetz (Kalter 2000, S. 441). Die Integration dieses Niveaus bietet vor allem den Vorteil, dass das Wanderungsverhalten von verschiedenen Gruppen bzw. Subpopulationen genauer untersucht werden kann.

Kritisch zu betrachten ist jedoch, dass der Wanderungsstrom von Region 1 zu Region 2 symmetrisch zu dem Wanderungsstrom von Region 2 zu Region 1 ist. In der Realität liegen in der Regel unterschiedlich starke Interaktionen vor. Sie werden mit diesem Ansatz allerdings nicht berücksichtigt. Zudem werden ähnlich wie bei den Gesetzen von Ravenstein nur die statistischen Zusammenhänge analysiert, nicht aber die dahinter liegenden Mechanismen (Kalter 2000, S. 442).

Außerdem scheinen intervenierende Möglichkeiten sowie die Kosten der Interaktion eine Rolle bei der Entstehung von räumlicher Interaktion zu spielen (Stouffer 1940; Ullmann 1956). Die Migrationstheorie von Stouffer (1940) basiert auf der Überlegung, dass die Wahrscheinlichkeit, eine Entfernung zurückzulegen, proportional zu der Zahl der angebotenen Möglichkeiten in der Zielregion ist. Die Ansiedlungsmöglichkeiten zwischen Ziel- und Quellregion wirken hingegen antiproportional. Das heißt, mit zunehmender Anzahl an Möglichkeiten zwischen beiden Orten sinkt die Wahrscheinlichkeit, den eigentlichen Zielort aufzusuchen. Die Entfernung und die Größe spielen hingegen eine untergeordnete Rolle bei der Wahl des Standortes (Kalter 2000, S. 443 f.).

Dennoch exisiteren laut Kalter (2000, S. 444) auch Schwächen. Zwar wurden die intervenierenden Möglichkeiten über eine zirkuläre Erfassung der Stellen-

angebote abgebildet, jedoch erfolgt die theoretische Auseinandersetzung dieses empirischen Ansatzes nicht ausreichend.

Ein weiteres Modell,welches die Verbindung zwischen dem gravitationstheoretischen Grundmodell und ökonomischen Faktoren schafft, stammt von Lowry (1966). Die theoretischen Überlegungen beruhen auf der Tatsache, dass ein statistischer Zusammenhang zwischen dem Auftreten von hohen Arbeitslosenquoten, einem niedrigen Lohnniveau und Migrationsbewegungen existiert. Es sind grundsätzlich Wanderungsbewegungen von Gebieten mit hohen Arbeitslosenquoten zu Regionen mit einer niedrigen Arbeitslosigkeit zu erkennen. Gleichzeitig steigt die Anzahl an Arbeitnehmern, die von Standorten mit einem niedrigen zu einem hohen Lohnniveau ziehen. Bedingt durch diese eindeutige Wirkrichtung kann auch von Push- und Pull-Faktoren gesprochen werden. Allerdings ist an dieser Stelle darauf hinzuweisen, dass das sogenannte „Push-Pull-Modell“ im eigentlichen Sinne kein eigenständiger theoretischer Ansatz ist, sondern viel mehr als Prinzip bezeichnet werden kann. Im Wesentlichen beeinflussen die „abstoßenden“ Merkmale des Herkunftsortes (Push-Faktoren) und „anziehenden“ Merkmale der Zielregion (Pull-Faktoren) die Migrationsbewegung (Kalter 2000, S. 447 f.).

Alle bisher vorgestellten theoretischen Modelle stellen allerdings nur die makroökonomische Perspektive dar. Somit besteht lediglich die Möglichkeit auf übergeordneter Ebene Pendler- bzw. Studierendenbewegungen zu erklären. Daher ist die Betrachtung von mikroökomischen Theorien von Nöten, um auf individueller Ebene das Handeln der Akteure zu erklären (Kalter 2000, S. 450 f.).

Mikroökonomische Perspektive

Eine Erweiterung des bisherigen Push- und Pull-Ansatzes ist die „Theorie der Wanderung“ von Lee (1966). Die Idee ist, neben den bekannten strukturellen Merkmalen (vgl. Lowry 1966: Arbeitslosigkeit; Lohnniveau), individuelle ökonomische, psychologische und soziale Merkmale der Wanderung mit in das Modell zu integrieren. Die Entfernung zwischen Quell- und Zielregion spielt eine untergeordnete Rolle (Lee 1966, S. 49). Lee unterscheidet in vier Wirkungsfaktoren (Kalter 2000, S. 452):

1. Faktoren in Verbindung mit dem Quellgebiet
2. Faktoren in Verbindung mit dem Zielgebiet

3. Intervenierende Hindernisse
4. Individuelle Aspekte

Besonderer Vorteil dieses Ansatzes ist, dass neben den Push- und Pullfaktoren intervenierende Hindernisse integriert werden können (z. B. Einwanderungsgesetze). Zudem ist es mit dem Ansatz möglich, das Verhältnis zwischen Strom und Gegenstrom der Migrationsflüsse deutlich besser abzubilden. Da jedoch die Entscheidung über die Wanderung auf Basis eines Vergleichs von Faktoren in der Ziel- und Quellregion erfolgt, ist das Verhalten nicht komplett rational und damit nur bedingt objektiv zu überprüfen (Kalter 2000, S. 453).

Einen weiteren Ansatz bietet die Humankapitaltheorie von Sjaastad (1962). Er betrachtet die Wanderung als individuelle Investition in Humankapital (Kalter 2000, S. 454 f.). In diesem Theorieansatz werden monetäre und nicht monetäre Kosten dem Nutzen von Migration gegenüber gestellt (Sjaastad 1962, S. 82 ff.). Die Grundidee ist, dass typische Wanderungsmotive wie ein höheres Lohnniveau in der Zielregion als Nutzen der Migration identifiziert werden. Demgegebenüber stehen Ausgaben für einen Umzug sowie nicht monetäre Kosten. Typische Beispiele hierfür sind emotionale Kosten durch das Verlassen von Familie und Freunden oder Informationsnachteile, die durch das Verlassen eines sozialen Netzwerkes entstehen (Sjaastad 1962, S. 85). Auf der Seite der nicht monetären Erträge stehen hingegen zum Beispiel die freie Wahl des „neuen“ Standortes.

Problematisch an diesem Ansatz ist, dass die Selektion auf Basis der Einkommenssituation geschieht. Zwar werden nicht monetäre Kosten ebenfalls genannt, allerdings nicht bei den eigentlichen Entscheidungsregeln integriert (Kalter 2000, S. 455).

Ein weiterer Ansatz, der versucht die Probleme der makrotheoretischen Modelle zu minimieren, stammt von Wolpert (1965). Basis dieser Idee ist, dass Migration als Reaktion veränderter Rahmenbedingungen im Umfeld entsteht (Wolpert 1965, S. 161). Das heißt, ähnlich wie bei Lee stehen die subjektiven Faktoren bei der Wanderung im Vordergrund (Kalter 2000, S. 455 f.). Ausgangspunkt für die Entscheidung ist die sogenannte „*place utility*“ am Wohnort, die den Nettonutzwert am Wohnort angibt. Zudem existiert für jedes Individuum das sogenannte „*aspiration level*“. Solange die „*place utility*“ über dem „*aspiration level*“ liegt, besteht keine Notwendigkeit für eine Migration. Ursache ist, dass nicht das Prinzip der Nutzenmaximierung im Vordergrund

steht, sondern viel mehr das Prinzip der Zufriedenheit. Sofern der Wert der „*place utility*“ unterschritten wird, ergibt sich beispielsweise durch Unzufriedenheit ein Anpassungsdruck. Allerdings ist die Migration nicht notwendigerweise die Lösung des Problems.

Das Hauptproblem bei der Theorie von Wolpert ist, dass der Mechanismus zwischen den Motiven und dem Migrationsverhalten nicht klar definiert wird und zudem die Zusammensetzung der Aspirationsschwellwerte sowie der „*place utility*“ ebenfalls nicht angegeben wird (Kalter 2000, S. 456).

Eine Synthese der Wanderungstheorie von Lee (1966) mit allgemeinen Einflussfaktoren der Humankapitaltheorie von Sjaastad (1962) und mit Festlegungen von Entscheidungsmechanismen und sozialen psychologische Aspekten wie bei Wolpert (1965) ist das „*Subjective Expected Utility*“-Modell (SEU-Modell). Ausgangspunkt bei dieser Theorie ist, dass die Wahl des Individuums nach dem Vergleich aller Alternativen (in diesem Fall „*move*“ oder „*stay*“) auf diejenige mit dem größten Nutzen fällt. Die Wahl des Zielortes wird auf Grund von subjektiven Bewertungen (z. B. Einkommensmöglichkeiten) vorgenommen. Die Evaluation dieses Nutzens wird um die Wahrscheinlichkeit ergänzt, dass die vorzunehmende Veränderung die angestrebte Verbesserung erzielt. Das heißt, es wird die Variante gewählt, die den SEU-Wert maximiert (Kalter 2000, S. 456 ff.).

Problematisch an diesem Theorieansatz ist allerdings, dass die Wahl des Individuums im Vordergrund steht, Migrationsentscheidungen aber häufig auf Ebene von Haushalten und Familien getroffen werden. Dies findet bisher jedoch keine Berücksichtigung.

2.4 Motive für räumliche Mobilität

Motive für die Mobilität von Studienanfängern

In der internationalen Literatur ist einer der wichtigsten Faktoren in Bezug auf studentische Mobilität die räumliche Nähe zwischen dem Herkunfts- und Zielort. Grundsätzlich nimmt die Wanderungsintensität mit zunehmender Entfernung exponentiell ab (z. B. Dotti et al. 2013; Gibbons/Vignoles 2012; Alm/Winters 2009; Montgomery 2002). Diese Beobachtung konnte auch unter der Berücksichtigung von wirtschaftlichen, sozialen und hochschulspezifischen Faktoren sowie Qualität und Umfang der Lehr- und Forschungsaktivitäten (z. B. Sá et al. 2004) nachgewiesen werden. Einhergehend mit zuneh-

mender Entfernung sind neben höheren Umzugs- und Transportkosten auch emotionale Kosten ein Hemmnis für die Mobilität der Studierenden. Mit einem Standortwechsel sind häufig die Aufgabe sozialer Bindungen und steigende Transaktionskosten verbunden (Denzler/Wolter 2010; Dotti et al. 2013 & 2014; Winters 2011). Studienanfänger geben bei einem Wohnortwechsel ihre sozialen Netze in der Heimatregion auf bzw. verlassen diese. Das heißt, es müssen in der Hochschulregion neue Kontaktnetzwerke aufgebaut werden. Es konnte daher gezeigt werden, dass die Wahrscheinlichkeit, einen bestimmten Hochschulstandort aufzusuchen, mit zunehmender Entfernung sinkt (Frenette 2004 & 2006; Spiess/Wrohlich 2010; Sá et al. 2004). Für Deutschland zeigen Gareis/Brökel (2022) zudem, dass eine große Zahl an Studierenden aus der jeweiligen Hochschulregion stammt. Die Distanz zwischen Heimat- und Hochschulstandort steigt jedoch noch einmal mit dem Übergang zum Masterstudium an.

Allerdings konnte für Irland nachgewiesen werden, dass eine starke Zusammenarbeit zwischen Hochschulen und Schulen zu einer Verringerung der negativen Auswirkung von geographischer Distanz führen kann (Raab et al. 2018). Darüber hinaus bietet die Migration auch neue Möglichkeiten: Es können beispielsweise interessantere und bessere Bildungsangebote genutzt werden. Ferner können Studierende neue soziokulturelle Erfahrungen machen (Holdsworth 2009). Außerdem hat die Lage des Hochschul- und Herkunftsstandorts einen entscheidenden Einfluss auf die studentische Mobilität (Falk/Kratz 2009). So sind Regionen mit einem höheren Urbanisierungsgrad und höherer Einwohnerdichte am Hochschulstandort attraktiver und oftmals anziehender für Studierende (Cullinan/Duggan 2016; Sá et al. 2004; Weisser 2019). Allerdings zeigen Gareis/Diller (2020), dass die Studierendenzahlen in Deutschland zwischen 1999 und 2013 besonders in ländlichen Kreisen stark gestiegen sind. Diese Beobachtung ist auch bei Gareis/Brökel (2022) festzustellen. Generell werden laut dieser Studie Studierende weniger stark von Ballungsräumen angezogen. Ausnahmen bilden Ingenieurswissenschaftler an Universitäten und Masterstudierende in den Wirtschaftswissenschaften.

Neben der Entfernung als wichtigstem räumlichen Faktor spielt auch die Beschäftigungsquote eine Rolle. Regionen mit hoher Beschäftigungsquote sind als Studienort deutlich attraktiver als Regionen mit hohen Arbeitslosenquoten, welche tendenziell eher durch Abwanderung gekennzeichnet sind (Cooke/Boyle 2011; Dotti et al. 2013). Ferner spielen insbesondere in Agglomerationsräumen finanzielle Aspekte, wie das Mietniveau (Dotti et al.

2013) und zukünftige Verdienstmöglichkeiten in der Heimat- und Zielregion, eine Rolle bei der Studienwahl (SÁ et al. 2004). Die Wirkrichtung von Studiengebühren ist im internationalen Kontext hingegen nicht einheitlich (Dwenger et al. 2012; Dotti et al. 2013). Zum einen sind Studierende mit sehr guten Noten – eventuell durch höhere Wahrscheinlichkeiten für Stipendien – weniger von den Gebühren beeinflusst, zum anderen ziehen Universitäten mit Studiengebühren durch eine verbesserte finanzielle Ausstattung mehr potenzielle Studierende an (Dwenger et al. 2012, S. 162). Der Einfluss der Qualität der Lehre in den Hochschuleinrichtungen variiert ebenfalls. So konnte in Italien (Ciriaci 2014), den USA (Cook/Boyle 2011) und Irland (Walsh et al. 2018) ein signifikanter Einfluss gemessen werden, während hingegen bei SÁ et al. (2004) für die Niederlande keine Auswirkungen nachweisbar waren. Als Messindikator dienen beispielsweise die Beschäftigungsquote der Absolventen (SÁ et al. 2011), Dozenten- und Studierendenanteile (SÁ et al. 2004), Ausgaben pro Studierenden (Cullinan/Duggan 2016), die Forschungsintensität (Adkisson/Peach 2008) oder der Platz in internationalen Rankings (Ciriaci 2014). Neben ökonomischen und räumlichen Faktoren spielen auch soziale Aspekte eine wichtige Rolle bei der Wahl des Hochschulstandortes. So beeinflusst zum Beispiel der Bildungshintergrund der Eltern das studentische Wanderungsverhalten (Lörz 2008, S. 430 f.). Studierende, die aus Haushalten mit geringen Einkommen stammen, sind immobiler als Studierende aus einem Elternhaus mit hohem Haushaltseinkommen. In Bezug auf das Geschlecht wurden unterschiedliche empirische Ergebnisse sichtbar. Für italienische Studentinnen konnte gezeigt werden, dass sie immobiler sind als männliche Studierende (Ciriaci 2014; Genova et al. 2021). In Großbritannien konnten Belfileld/Morris (1999) hingegen keine signifikanten Unterschiede zwischen den Geschlechtern feststellen (Belfileld/Morris 1999).

Außerdem deuten unterschiedliche Studien darauf hin, dass potenzielle Studierende aufgrund ähnlicher Lebensstile häufig in studentisch geprägte Regionen bzw. in Regionen mit einem hohen Anteil an Hochqualifizierten wandern (Buenstorf et al. 2016; Haussen/Uebelmesser 2018).

Ein Aspekt, der in den letzten Jahren immer wichtiger geworden ist bei der Wahl des Hochschulstandortes, ist der Suchprozess über hochschulspezifische Informationen. In entwickelten Ländern ist die Informationsbeschaffung über Hochschulen und die zugehörigen Studiengänge bedingt durch die Digitalisierung und das Internet deutlich einfacher geworden. Darüber hinaus spielen heute vor allem die stärkere Beeinflussung durch Medien, Bekannte und der

wachsende Wettbewerb zwischen den Hochschulen eine wesentliche Rolle beim Entscheidungsprozess (Vrontis et al. 2007, S. 985 f.). Somit erlangt das Hochschulmarketing eine ganz neue Bedeutung bei der Wahl des Hochschulstandortes (Vrontis et al. 2007, S. 987).

Insgesamt zeigt diese Vielzahl an Indikatoren zur Erklärung studentischer Mobilität, dass eine sehr große Anzahl an höchst individuellen Faktoren beim Entscheidungsprozess für und gegen eine bestimmte Hochschule eine Rolle spielen können.

Motive für die Mobilität von Pendlern

Die Motivation zum Pendeln wird in wirtschaftliche, demographische und soziale Einflüsse unterteilt (Kersting et al. 2020). Zunächst stehen aus wirtschaftlicher Perspektive das Angebot an Arbeitsplätzen und die Verdienstmöglichkeiten im Vordergrund (Pfaff 2012; Kersting et al. 2020). Das heißt, Beschäftigte pendeln verstärkt an Standorte mit hohen Beschäftigungsquoten und einem hohen Bruttoinlandsprodukt. Hohe Arbeitslosigkeit und niedrige Haushaltseinkommen haben hingegen einen negativen Einfluss auf die Pendlerbewegungen (Guirao et al. 2018, S. 7).

Zudem ist aus dem klassischen Gravitationsmodell ableitbar, dass die Pendlerintensität von Arbeitnehmern mit zunehmender Entfernung sinkt (z. B. Eckey et al. 2007, S. 9; Kersting et al. 2020, S. 8; Simini et al. 2012, S. 96). Diese Beobachtung muss allerdings differenziert betrachtet werden, da mit zunehmendem Einkommen bedingt durch die daraus reduzierende ökonomische Verbesserung die Pendlerdistanz steigt (Pfaff 2012; Kersting et al. 2020). Außerdem erhöht sich mit steigendem Qualifikationsniveau die Pendlerbereitschaft (Kalter 1994, S. 472; Pfaff 2012, S. 462 ff.; Sandow 2014, S. 529 ff.; Broersma et al. 2020, S. 6), da mit zunehmender Spezialisierung die Wahrscheinlichkeit im direkten Umfeld eine Stelle zu finden sinkt (Kersting et al. 2020).

Ferner sind Agglomerationsräume häufig Schwerpunkt von Pendlerbewegungen (Eckey et al. 2007; O'Clery et al. 2019). Diese Beobachtung wird auch in dem klassischen Gravitationsmodell aufgegriffen. Die Einwohnerzahl eines Standortes ist proportional zur Pendlerwahrscheinlichkeit. Außerdem führt laut Heuermann/Schmieder (2019, S. 355) und Martinus et al. (2020) eine hohe Bevölkerungsdichte am Arbeitsort zu steigenden Pendlerzahlen. Für Deutschland konnte gezeigt werden, dass die Pendlerbereitschaft

in Metropolregionen und deren Umland höher ist als in anderen Gebieten (Kauffmann 2011; Eckey et al. 2007). Außerdem sind Oberzentren auf Grund der hohen Konzentration von Beschäftigten die wichtigsten Arbeitsplatzzentren in Deutschland (Einig/Pütz 2007). Das bedeutet, die Erreichbarkeit des nächsten Oberzentrums beeinflusst ebenfalls das Pendlerverhalten (Kauffmann 2011, S. 70; Schulze 2009, S. 6). Ursachen für das hohe Pendleraufkommen in urbanisierten Räumen und dem Umland sind unter anderem eine große Anzahl von Unternehmen, hohe Boden- und Mietpreise sowie Aspekte wie Schul- bzw. Wohnqualität (Tharakan/Tropeano 2009; Dauth/Haller 2018; Krebs/Pflüger 2019, S. 3; Kersting et al. 2020, S. 5).

Zudem spielt ein gut ausgebautes Infrastruktursystem in urbanisierten Räumen eine wichtige Rolle. Die Pendlerbereitschaft wird unter anderem durch die Distanz zur Autobahnauffahrt (Broersma et al. 2020, S. 3), die Anzahl an Linienverbindungen des ICE und die Entfernung des Bahnhofes zum Stadtzentrum (Kauffmann 2011, S. 76; Heuermann/Schmieder 2019, S. 337; Guirao et al. 2018, S. 7) positiv beeinflusst. Die Nähe zur nächsten ÖPNV-Haltestelle hat hingegen einen negativen Einfluss auf die Pendlerbewegungen (Heuermann/Schmieder 2019, S. 336; Krebs/Pflüger 2019, S. 39; Broersma et al. 2020, S. 6).

Ein weiterer Einflussfaktor auf die Pendleraktivität ist die familiäre Situation. Generell konnte gezeigt werden, dass bei Beschäftigten mit Lebenspartner und Familie die Pendlerbereitschaft zunimmt. Bei Mehrpersonenhaushalten wird die Pendlerstrecke größer, da die Migrationskosten mit jeder weiteren Person steigen (Kalter 1994; Pfaff 2012; Huber/Nowotny 2013, S. 1470). Allerdings sind geschlechtsspezifische Unterschiede erkennbar. Frauen mit Kindern pendeln weniger häufig und über geringere Distanzen (Häfner 2011).

Diese Beobachtung steht in Zusammenhang zum Einkommen und dem Qualifikationsniveau. Auspurg/Schönholzer (2013) konnten feststellen, dass die Pendlerdistanz bei Frauen mit niedrigem Einkommen geringer ist als bei Frauen mit hohem Einkommen, während die Pendlerwahrscheinlichkeit bei Männern mit Kindern steigt (McQuid/Chen 2012). Zudem nimmt die durchschnittliche tägliche Pendlerdistanz und -zeit bei Frauen ohne Hochschulabschluss und Kinder ab, während sie bei Frauen mit Hochschulabschluss unabhängig ist. In diesem Zusammenhang muss auch der große Anteil an Frauen in Teilzeitbeschäftigung und daraus resultierenden niedrigeren Einkünften dis-

kutiert werden (McQuid/Chen 2012), da die Bereitschaft, lange Strecken zu pendelen, in dieser Gruppe sinkt (Kersting et al. 2020).

Außerdem existieren regionale Unterschiede: Die Pendlerbereitschaft in Ostdeutschland ist geringer als in Westdeutschland. Hinzu kommt, dass seit der Wiedervereinigung eine große Anzahl von Personen zwischen Ost- und Westdeutschland pendeln. Ursache hierfür ist unter anderem die höhere Arbeitslosigkeit in den neuen Bundesländern und das niedrigere Lohnniveau (Eckey et al. 2007; Granato et al. 2009; Niebuhr et al. 2012; Kersting et al. 2020).

Ferner spielen individuelle Aspekte, wie die persönlichen Lebenszufriedenheit und die Attraktivität des Wohnortes eine wichtige Rolle (Ingenfeld et al. 2018, S. 2; Stenpass 2020, S. 65; Chatterjee et al. 2019, S. 6). Laut Ingenfeld et al. (2018, S. 4) und Chatterjee et al. (2019, S. 24) ist ein negativer Zusammenhang zwischen Straftaten innerhalb eines Kreises und den Pendlerbewegungen vorhanden. Eine flächendeckende Breitbandverbindung hat nach Falck/Schüller (2016, S. 612) einen positiven Einfluss am Arbeitsort und einen negativen Einfluss am Wohnort, da die Breitbandversorgung zu positivem Beschäftigungseffekten am Wohnort der Arbeitnehmer führt und somit die Notwendigkeit zu pendeln geringer wird.

2.5 Instrumente zur Prognose von räumlichen Interaktionen

Modellübersicht

Klassischerweise sind in der Vergangenheit Gravitationsmodelle zur Prognose von Mobilitätsflüssen eingesetzt worden. Grundsätzlich wird die Einwohneranzahl am Wohnort i mit $m_i \in \mathbb{N}$ und am Arbeitsort j mit $n_j \in \mathbb{N}$ bezeichnet. Außerdem repräsentiert $r \in \mathbb{R}$ die Distanz zwischen Quell- und Zielgebiet und $\alpha, \beta, \gamma \in \mathbb{N}$ stehen für frei wählbare Parameter, sodass das Gravitationsmodell wie folgt definiert ist (Simini et al. 2012):

$$\frac{m_i^{\alpha} n_j^{\beta}}{r^{\gamma}} \tag{1}$$

Die Gravitationsmodelle weisen jedoch einige Schwächen auf. In den bisherigen Studien sind unterschiedliche Funktionen und bis zu neun Parameter zur Näherung an die empirischen Daten verwendet worden (Viboud et al. 2006; Balcan et al. 2009; Kaluza et al. 2010; Krings et al. 2009; Si-

MINI et al. 2012). Zudem müssen die empirische Daten zur Schätzung vorliegen, um Vergleiche mit den Ergebnissen der Prognose durchführen zu können. Darüberhinaus führt die fehlende Berücksichtigung von Standorten zwischen Ziel- und Quellgebiet beim Gravitationsmodell zu einer systematische Vorhersage-Diskrepanz. Das heißt, für Interaktionen zwischen zwei Standorten mit ähnlichen Herkunfts- und Zielmaßen sowie vergleichbaren Entfernungen werden ähnlich große Flüsse vorhergesagt, ohne den Zwischenraum zu berücksichtigen. Außerdem steigt die prognostizierte Wanderungsbewegung bei Erhöhung der Standortmaße der Zielregion unbegrenzt an. Ferner berücksichtigt das klassische Gravitationsmodell keine Schwankungen bei Interaktionen zwischen zwei Standorten (SIMINI et al. 2012). Aus diesem Grund entwickelten SIMINI et al. (2012) das klassische Strahlungsmodell als alternativen Ansatz zur Schätzung von Mobilitätsflüssen. Als Untersuchungsgebiet diente dabei zunächst die USA. Es wurde deutlich, dass das neu entwickelte Modell deutliche Verbesserungen der Schätzung lieferte. So wurde beispielsweise bei einer Pendlerstrecke (Entferung ca. 447 km) von Davis County (UT) nach Washington County (UT) (Empirische Daten: 44 Interaktionen) mit dem Gravitationsmodell eine Pendlerbewegung prognostiziert, während das Strahlungsmodell 66 Interaktionen vorhersagte. Ähnlich gute Ergebnisse konnten für Migrationsbewegungen und Warenströme in den USA sowie für die Anzahl der Telefongespräche zwischen Nutzern unterschiedlicher Gemeinden (Zeitraum: 4 Wochen) in Europa erzielt werden (SIMINI et al. 2012). Vorteil dieses Ansatzes ist zum einen die Datensparsamkeit und zum anderen die Parameterfreiheit.

Das klassische Strahlungsmodell von SIMINI et al. (2012) basiert auf der Annahme, dass Beschäftigte unter Berücksichtigung ihrer Erwartungen (z. B. Einkommen, Arbeitszeiten, Bedingungen etc.) den nächstgelegenen Arbeitsplatz aufsuchen. Die Arbeitnehmer einer Region setzen die Beschäftigungsmöglichkeiten zwischen zwei Standorten ins Verhältnis zueinander und entscheiden sich schließlich für den nächstgelegenen Ort, an dem mehr Leistungen angeboten werden und die Nutzen für sie persönlich höher sind als an ihrem Wohnort. Das bedeutet, dieses Modell integriert die theoretischen Überlegungen von STOUFFER (1940) (vgl. Kapitel 2.3).

Ausgangspunkt für die Schätzung der Pendlerzahl $T_{ij} \in \mathbb{N}$ von Ort i nach j ist, ähnlich wie im Gravitationsmodell, die Einwohneranzahl am Wohnort m_i und am Arbeitsort n_j. SIMINI et al. (2012) berücksichtigen zusätzlich die Einwohnerzahlen zwischen Wohn- und Arbeitsort ($s_{ij} \in \mathbb{N}$). Dies bietet den Vorteil, dass somit alle potenziellen Arbeitsorte in der Berechnung integriert

sind. Zudem ist die Summe der Pendler T_i, die ihren Wohnort i zum Arbeiten verlassen, proportional zu $m_i(v/M)$. $v \in \mathbb{N}$ steht für die Gesamtzahl aller Pendler (ohne Binnenpendler) und $M \in \mathbb{N}$ für die Gesamteinwohnerzahl in Deutschland. Die Definition des Strahlungsmodells lautet:

$$T_{ij}^{orginal} = T_i \cdot \frac{m_i n_j}{(m_i + s_{ij})\,(m_i + n_j + s_{ij})}$$

$$= \frac{v}{M} \cdot \frac{m_i^2 n_j}{(m_i + s_{ij})\,(m_i + n_j + s_{ij})} \tag{2}$$

Angelehnt an Studien mit Gravitationsmodellen (vgl. u. a. Cullinan/Duggan 2016) und den Überlegungen von Simini et al. (2012) wird im Rahmen dieser Dissertation für die Mobilität von Studierenden $m_i \in \mathbb{N}$ als Anzahl an Abiturienten eines Jahrgangs $t \in \mathbb{N}$ am Heimatstandort definiert. Für den zukünftigen Hochschulstandort j wird der Indikator $n_j \in \mathbb{N}$ zum Zeitpunkt $t-1$ gewählt.

Er steht für die Anzahl an Studierenden im ersten Semester. Grundsätzlich gilt die Annahme, dass zukünftige Studierende unter Berücksichtigung unterschiedlicher Bedingungen Hochschulstandorte miteinander vergleichen (z. B. Studienangebot, Nähe zum Heimatstandort etc.). Das heißt, es kann angenommen werden, dass nach Abwägung aller Angebote es am wahrscheinlichsten ist, dass ein Studienanfänger die Entscheidung für den nächstgelegenen Hochschulstandort trifft. Aus diesen Überlegungen folgt, dass $s_{ij} \in \mathbb{N}$ die Anzahl an Studierenden im ersten Semester zum Zeitpunkt $t-1$ zwischen dem Heimatkreis und dem zukünftigen Hochschulstandort beschreibt und damit alle anderen potenziellen Hochschulstandorte in der Berechnung berücksichtigt werden. Zur weiteren Berechnung der durchschnittlichen Studierendenwanderung von Erstsemestern $T_{ij} \in \mathbb{N}$ von Ort i nach j wird zusätzlich noch die durchschnittliche Wanderungsrate von Studienanfängern zum Zeitpunkt $t-1$ für das gesamte Bundesgebiet berechnet. Ferner ist die Summe der Studierenden $T_i \in \mathbb{N}$, die ihre Heimat i zum Studieren verlassen, proportional zu $m_i(v/M)$. Die Gesamtzahl aller mobilen Studierenden (ohne Binnenpendler) wird mit $v \in \mathbb{N}$ bezeichnet und $M \in \mathbb{N}$ ist die Anzahl aller Studienanfänger in Deutschland. Eine Zusammenfassung aller Variablen findet sich in Tabelle 1.

Variablen-name	Studentische Mobiliät	Arbeitskräfte Mobilität
α, β, γ, λ	frei wählbare Parameter	frei wählbare Parameter
r	Distanz zwischen i und j	Distanz zwischen i und j
i	Heimatstandort	Wohnort
j	Hochschulstandort	Arbeitsort
m_i	Anzahl an Abiturienten eines Jahrgangs t in i	Einwohnerzahl in i
n_j	Anzahl an Studierenden im ersten Semester zum Zeitpunkt $t-1$ in j	Einwohnerzahl in j
T_i	Summe der migrierenden Abiturienten in i	Summe der Auspendler in i
s_{ij}	Anzahl an Studierenden im ersten Semester zum Zeitpunkt $t-1$ zwischen i und j (exclusive i und j)	Einwohnerzahlen zwischen i und j (exclusive i und j)
S_{ij}	Anzahl an Studierenden im ersten Semester zum Zeitpunkt $t-1$ zwischen i und j (inclusive i und j)	Einwohnerzahlen zwischen i und j (inclusive i und j)
T_{ij}	Studierendenströme	Pendlerströme
M	Anzahl der Studienanfänger in Deutschland	Gesamteinwohnerzahl in Deutschland
V	Gesamtzahl aller mobilen Studierenden (ohne Studierende mit Heimatstandort = Hochschulstandort)	Gesamtzahl aller Pendler (ohne Binnenpendler)
G	Gesamtzahl aller mobilen Studierenden (mit Studierende mit Heimatstandort = Hochschulstandort)	Gesamteinwohnerzahl in Deutschland (=M)

Tabelle 1: Variablenübersicht. Quelle: Eigene Darstellung auf Basis von Simini et al. (2012), Cullinan/Duggan (2016) sowie eigenen Überlegungen.

Der Hauptvorteil des klassischen Strahlungsmodells (Formel (2)) besteht dabei in der Distanzunabhängigkeit (Masucci et al. 2013). Darüber hinaus ist das

Strahlungsmodell frei von Parametern, daher ist keine Kalibrierung erforderlich (Lenormand et al. 2016). Insgesamt zeigt sich bei der Berechnung von Pendlerströmen mit dem klassischen Strahlungsmodell, dass für sehr große Systeme mit gleichmäßig verteilter Bevölkerung sehr gute Schätzergebnisse erzielt werden können. Dies ist für das Städtesystem in Deutschland gewährleistet (Limtanakool et al. 2005, S. 2124). Für das studentische Wanderungsverhalten in Deutschland ist auf Grund der polyzentralen Struktur des Hochschulsystems und der Regionalisierung in den 1970er Jahren durch Erweiterung der Anzahl der Hochschulstandorte (Framheim 1983, S. 11) davon auszugehen, dass diese Annahme erfüllt ist.

Allerdings ist die Originalversion des Modells durch die Ableitung mithilfe des thermodynamischen Grenzfalls (thermodynamischen Grenzannahme) für ein infinites System vorgesehen. Masucci et al. (2013) kamen bei ihrer Untersuchung für großräumige Pendler- und Verkehrsnetze in England und Wales sowie für die innerstädtischen Pendler- und Verkehrsnetze in der Region London zu dem Schluss, dass eine Normierung der Pendlerströme notwendig ist, um zuverlässige Schätzungen zu erzielen. Für ein endliches System wird daher das Modell um einen Normierungsfaktor ergänzt. Es konnte gezeigt werden, dass die Datenströme systematisch um den Faktor $(1/(1-(m_i/M)))$ unterschätzt werden. Für finite Systeme wird folglich eine leicht modifizierte Variante des Strahlungsmodells verwendet (Masucci et al. 2013):

$$T_{ij}^{finit} = \frac{T_i}{1-\frac{m_i}{M}} \cdot \frac{m_i n_j}{(m_i+s_{ij})\,(m_i+n_j+s_{ij})}$$

$$= \frac{1}{1-\frac{m_i}{M}} \cdot \frac{v}{M} \cdot \frac{m_i^2 n_j}{(m_i+s_{ij})\,(m_i+n_j+s_{ij})} \tag{3}$$

In Kang et al. (2015) wurde zur verbesserten Anwendbarkeit des Strahlungsmodells auf verschiedene Mobilitätssysteme und unterschiedliche räumliche Skalen - angelehnt an das klassische Gravitationsmodell (vgl. Formel (1)) - die Wahrscheinlichkeit des ursprünglichen Strahlungsmodells (vgl. Formel (2)) durch einen Skalierungsexponenten $\lambda \in \mathbb{N}$ erweitert. Dieser Skalierungsparameter ist abhängig vom Mobilitätssystem. Er kann außerdem bei der Modellierung von Pendlerströmen als Indikator für die wirtschaftliche Situation in Agglomerationsräumen angesehen werden und erfasst die maßstabsfreien Eigenschaften. Die Wahl des Parameters in der Dissertation erfolgt über einen

Suchalgorithmus, der den Sörensenindex maximiert.

Zusätzlich wurden vier Varianten des Modells basierend auf Kombinationen von Such- und Auslöserichtung vorgestellt. Damit können die unterschiedlichen Push- und Pullfaktoren gut mit in das Modell integriert werden. Für alle vier Varianten wurde, angelehnt an Masucci et al. (2013), ein Normierungsfaktor für endliche Systeme eingeführt. Während das ursprüngliche Strahlungsmodell (vgl. Formel (2)) nutzenspezifisch ist, handelt es sich bei den verallgemeinerten Modellen von Kang et al. (2015) um wettbewerbsorientierte Strahlungsmodelle (vgl. Formel (4-7)). Das heißt, die Anforderungen des Zielortes spielen eine entscheidende Rolle bei der Reiseentscheidung. Je mehr Angebote (m_i) ein Ort i aufweist und je weniger Wettbewerber in s_{ji} existieren, desto höher ist die Chance, dass der Ort i ausgewählt wird. Ferner ist die Dichte des Mobilitätsnetzwerkes häufig gering und es findet eine Unterschätzung der Pendelbewegungen statt. Als Untersuchungsräume dienten dabei die USA, das Vereinigte Königreich und die Städte Peking, Shenzhen, Abidjan und Chicago. Insgesamt präsentieren die Autoren vier nach Angebots- und Nachfrageaspekten differenzierte Varianten des generalisierten Strahlungsmodells: *Production-constrained Intervention-based Radiation*-Modell (PIR-Modell), *Production-constrained Competition-based Radiation*-Modell (PCR-Modell), *Attraction-constrained Intervention-based Radiation*-Modell (AIR-Modell), *Attraction-constrained Competition-based Radiation*-Modell (ACR-Modell). Sie sind wie folgt definiert:

$$T_{ij}^{PIR} = \frac{T_i}{\sum_{k\neq i}^{N}\left[\frac{m_i n_k}{(m_i+s_{ik})(m_i+n_k+s_{ik})}\right]^{\lambda}}\left[\frac{m_i n_j}{(m_i+s_{ij})\,(m_i+n_j+s_{ij})}\right]^{\lambda} \tag{4}$$

$$T_{ij}^{PCR} = \frac{T_i}{\sum_{k\neq i}^{N}\left[\frac{m_i n_k}{(n_k+s_{ki})(m_i+n_k+s_{ki})}\right]^{\lambda}}\left[\frac{m_i n_j}{(n_j+s_{ji})\,(m_i+n_j+s_{ji})}\right]^{\lambda} \tag{5}$$

$$T_{ij}^{AIR} = \frac{T_j}{\sum_{k\neq j}^{N}\left[\frac{m_k n_j}{(m_k+s_{kj})(m_k+n_j+s_{kj})}\right]^{\lambda}}\left[\frac{m_i n_j}{(m_i+s_{ij})\,(m_i+n_j+s_{ij})}\right]^{\lambda} \tag{6}$$

$$T_{ij}^{ACR} = \frac{T_j}{\sum_{k \neq j}^{N} \left[\frac{m_k n_j}{(n_j + s_{jk})(m_k + n_j + s_{jk})} \right]^{\lambda}} \left[\frac{m_i n_j}{(n_j + s_{ji})(m_i + n_j + s_{ji})} \right]^{\lambda} \quad (7)$$

Ein weiteres typisches parameterfreies Mobilitätsmodell ist das Modell der bevölkerungsgewichteten Chancen (*Population-weighted opportunities*-Modell (PWO-Modell)). Es wurde entwickelt, um die Unterschätzungen im Stadtmaßstab auszugleichen. Besonderheit des Modells ist, dass sowohl Fahrten innerhalb einer Stadt als auch der Intercity-Verkehr auf verschiedenen räumlichen Skalen genau vorhergesagt werden können. Der Hauptunterschied zum Gravitationsmodell ist, dass nicht die intervenierenden Gelegenheiten und der Parameter Berücksichtigung bei der Modellierung findet, sondern vielmehr, ähnlich wie beim Strahlungsmodell, die Bevölkerungsverteilung die Abnahme an Interaktionen bestimmt. Yan et al. (2014) konnten beim Vergleich dieses Modells mit dem klassischen Strahlungsmodell von Simini et al. (2012) für zahlreiche weltweit verteilte Städte und Metropolregionen (Beijing, Shenzhen, Abidjan, Chicago, New York, Seattle, Detroit, Twin Cities, London, Berlin, Prag, Oslo, Kopenhagen, Göteborg) signifikante Verbesserungen der Pendlerprognose feststellen. Die Annahme des Strahlungsmodells, dass Studierende bzw. Pendler den nächstgelegenen Hochschulstandort oder Arbeitsort auswählen, wird in diesem Modell um die Ziel- und Ausgangsregion erweitert. Das heißt, S_{ij} ist eine Erweiterung von s_{ij} um die Einwohnerzahl am Wohn- und Arbeitsort bzw. die Anzahl an Studierenden am Heimat- und zukünftigen Hochschulstandort.

In Abbildung 4 wird der Unterschied bei der Berechnung von s_{ij} und S_{ij} anhand der Pendlerstrecke Fulda-Gießen (Kreisebene) veranschaulicht. Die Entfernung zwischen den beiden Zentroiden der Landkreise beträgt 105 Kilometer. Insgesamt pendelten im Jahr 2017 174 Personen von Fulda nach Gießen und 127 von Gießen nach Fulda (Bundesagentur für Arbeit 2017). Sofern s_{ij} in das jeweilige Modell integriert werden soll, liegt die Einwohnerzahl zwischen den Standorten bei 3.510.488 Einwohner (INKAR 2022). Bei S_{ij} hingegen werden durch die Integration von Gießen und Fulda insgesamt 3.993.125 Einwohner (INKAR 2022) berücksichtigt.

Es erfolgt weiterhin eine Korrektur mit $1/G$ wobei $G \in \mathbb{N}$ bei der Schätzung der Pendlerbewegungen gleichgesetzt ist mit M. Für die Schätzung der Studierenden ist G definiert als die Gesamtanzahl aller Studierenden in Deutschland. Grundsätzlich gilt für dieses Modell die Annahme, dass die Wahrscheinlichkeit an einen Hochschulstandort zu ziehen proportional zu seiner Attraktivität ist.

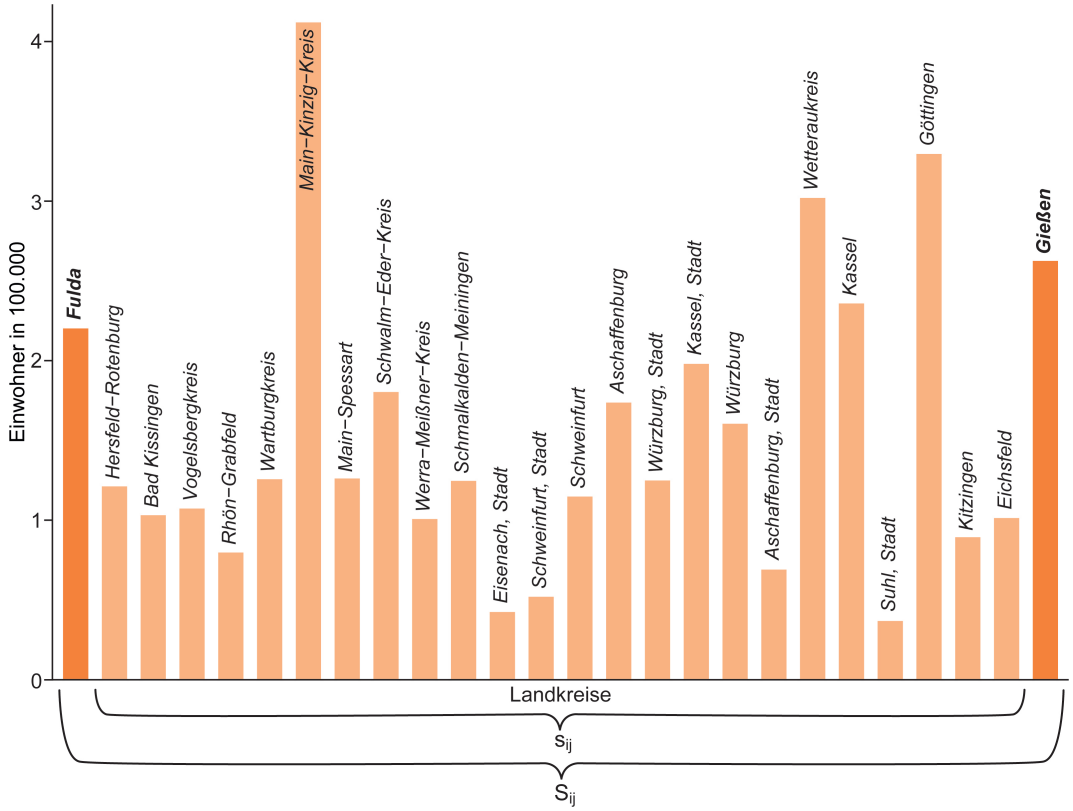


Abbildung 4: Vergleich von s_{ij} und S_{ij} am Beispiel der Pendlerstrecke Fulda-Gießen (Kreisebene). Quelle: Eigene Darstellung auf der Grundlage von INKAR 2022.

Daher ist die Anziehungskraft einer Region mit Universität umgekehrt proportional zur Bevölkerung bzw. zur Anzahl an Studierenden (Yan et al. 2014).

Insgesamt folgt für die Berechnung des PWO-Modells :

$$T_{ij}^{PWO} = T_i \cdot \frac{n_j \left(\frac{1}{S_{ji}} - \frac{1}{G}\right)}{\sum_{k \neq i}^{N} n_k \left(\frac{1}{S_{ki}} - \frac{1}{G}\right)} \tag{8}$$

Eine weitere Alternative zur Vorhersage der studentischen Mobilität und des Pendlerverhaltens ist das *Opportunity-Priority-Selection*-Modell (OPS-Modell). Die Hauptidee ist, dass die Übergangswahrscheinlichkeit zu einem Ziel proportional zur Wahrscheinlichkeit ist, dass das Ziel höheren Nutzen hat als Regionen mit kleinerer Entfernung zum Ausgangsort. Der Hauptunterschied zwischen dem PWO- und dem OPS-Modell ist die Regel der Zielauswahl. Das OPS-Modell wird von einer Reihe von grundlegenden Prinzipien abgeleitet. Das heißt, das Ziel jedes Studierenden ist es, einen Ort auszuwählen, dessen Vorteil höher ist als die Opportunitätsvorteile seines Herkunftslandkreises.

Außerdem sind die Vorteile des Zielortes höher als bei allen anderen dazwischenliegenden Orten (Liu/Yan 2019).

$$T_{ij}^{OPS} = T_i \cdot \frac{\frac{n_j}{S_{ij}}}{\sum_{k \neq i}^{N} \frac{n_k}{S_{ik}}} \tag{9}$$

Die bisher vorgestellten Teilchenmodelle berücksichtigen lediglich die Einwohnerzahl in der Ziel- und Quellregion sowie zwischen den Standorten, individuelle Aspekte finden jedoch keine Integration in die Modelle. Diese sind - wie in Kapitel 2.4 beschrieben - ebenfalls von Bedeutung bei der Wahl eines Arbeits- bzw. Hochschulstandortes. Spadon et al. (2019) integrierten die indiviuellen Aspekte, indem sie mithilfe des XGBoost-Algorithmus die Gewichtung der jeweiligen Verbindungen prognostizieren.

Im Rahmen dieser Dissertation wird daher im Fall der studentischen Mobilität das Grundmodell aus der Anzahl an Abiturienten eines Jahrgangs t m_i und der Anzahl an Studierenden im ersten Semester zum Zeitpunkt $t-1$ (n_j) sowie die Distanz r_{ij} zwischen den Standorten berechnet. Im Erweiterungsmodell wird zusätzlich die Summe aller Studienanfänger zwischen Herkunfts- und Zielregion s_{ij} integriert. Zudem wird ein Individualmodell berechnet, welches weitere sieben Motive (vgl. Tabelle 2) für studentische $U_k := \{u_k^1, \ldots, u_k^7\}$, für $k \in i, j$ und $u_k^f \in \mathbb{R}$, $f \in \mathbb{N}$ Mobilität jeweils für Heimat- und den Hochschulstandort berücksichtigt.

Im Fall der Prognose der Pendlerbewegungen werden als Merkmale, wie bereits bei den Gravitations- bzw. Strahlungsmodellen, die Einwohnerzahl am Wohn- und Arbeitsort ($m_i, n_j \in \mathbb{N}$) sowie die Summe aller Einwohner zwischen den Standorten (s_{ij}) und die Distanz berücksichtigt (vgl. Grund- und Erweiterungsmodell). Außerdem werden im Individualmodell 13 Motive (vgl. Tabelle 3) als $U_k := \{u_k^1, \ldots, u_k^{13}\}$, für $k \in i, j$ und $u_k^f \in \mathbb{R}$, $f \in \mathbb{N}$ integriert.

Für jede Interaktion und Menge: $\mathbb{R}^{|S|} \ni S_i := \{r_{ij}, s_{ij}, U_i, U_j\}$ wird also folgende Funktion gesucht (Spadon et al. 2019):

$$Gewicht : \mathbb{R}^{|S|} \longrightarrow \mathbb{N} \tag{10}$$

Um den Regressor berechnen zu können, werden die Daten zunächst in einen Trainingsdatensatz (70 %) und in einen Testdatensatz (30 %) aufgeteilt. Es folgt eine 5-fache Kreuzvalidierung und Hyperparameter-Abstimmung. Im An-

schluss wird der abgestimmte Regressor auf den Testdatensatz angewendet und mit unterschiedlichen Indikatoren die Güte evaluiert (Spadon et al. 2019). Zudem wird dieser Prozess 1.000 mal im Rahmen einer Monte-Carlos-Simulation wiederholt und auf die Ergebnisse die Evaluationsindikatoren angewendet. Ziel ist es, die Stabilität der Ergebnisse zu überprüfen, da je nach Wahl des Trainings- und Testdatensatzes die Vorhersageergebnisse variieren können.

Evaluationsindikatoren

Zur Bestimmung des bestmöglichen Modells wird der Sörensenindex angewandt. Der Sörensenindex ist ein beliebtes Instrument zur Messung von Fluktuation, daher wird er exemplarisch für die Darstellung der Simulationsergebnisse verwendet (z. B. Abbildung 18). Dieses Ähnlichkeitsmaß gibt den korrekt wiedergegebenen Anteil an Pendelströmen in simulierten Netzwerken an und variiert zwischen 0 und 1. Ein Wert von 0 bedeutet keine Übereinstimmung mit den ursprünglichen Pendelbewegungen, bei einem Wert von 1 sind das empirische und das simulierte Netzwerk identisch. Im Vergleich zu anderen Ähnlichkeitsmaßen, wie zum Beispiel der euklidischen Distanz, behält der Sörensenindex die Empfindlichkeit in heterogeneren Datensätzen bei und ist unempfindlicher gegenüber Ausreißern (McCune/Grace 2002). Der Messindikator wird in nachfolgender Formel berechnet (Soerensen 1948):

$$SI = \frac{2 \sum_{i=1}^{N} \sum_{j=1}^{N} min\left(T_{ij}, \hat{T}_{ij}\right)}{\sum_{i=1}^{N} T_{ij} + \sum_{j=1}^{N} \hat{T}_{ij}} \tag{11}$$

Ergänzt wird die Evaluation durch den mittleren absoluten Fehler, die Wurzel der mittleren quadratischen Abweichung, den Korrelationskoeffizienten von Pearson sowie den adjustierten R^2-Wert (u. a. Masucci et al. 2013; Lenormand et al. 2016). Für die Evaluation des XGBoost-Regressors wird zusätzlich eine Rekonstruktion der gewichteten Struktur des Pendlernetzes vorgenommen. Hierzu wird der Einfluss der Merkmale auf die Prognose mithilfe des SHAP-Wertes bestimmt (Lundberg/Lee 2017, Lundberg et al. 2018 a/b). Bereits 1953 hat Lloyd Shapley im Kontext der Spieltheorie Shapley-Werte eingeführt (Shapley 1953). Die Grundidee ist, dass die Vorhersage des Modells mit und ohne das jeweilige Merkmal vorgenommen wird. Außerdem ist zu beachten, dass die Reihenfolge des Hinzufügens neuer Funktionen einen Einfluss auf das Modell hat, sodass sämtliche Permutationen der Merkmalsreihenfolgen

berechnet werden müssen.

Mathematisch bedeutet dies (Lundberg/Lee 2017; Lundberg et al. 2018 a/b):

$$\phi_i = \sum_{S \subseteq N\{i\}} \frac{|S!\,(M - |S| - 1)!|}{M!} \left[f_x\left(S \cup \{i\}\right) - f_x\left(S\right) \right] \tag{12}$$

Vorteil dieses Ansatzes ist, dass die Auswirkungen der Merkmale auf die Vorhersage einzelner Daten möglich werden. Hierzu wird auf Basis der SHAP-Werte mithilfe des R-Packets SHAPforxgboost eine Grafik (z. B. Abbildung 16) zur besseren Interpretation der Ergebnisse erstellt (Liu et al. 2021). Dabei werden die Variablen unter Berücksichtigung der Summe der SHAP-Werte in absteigender Reihenfolge geordnet. Das wichtigste Merkmal steht an oberster Stelle. Für jedes Merkmal wird zusätzlich ein Punkt, der die prognostizierte Verbindung repäsentiert, gezeichnet. Somit ist es möglich, die Verteilung der Auswirkungen jedes Merkmals auf die jeweilige Interaktion zu bestimmen. Punkte, die im negativen Bereich liegen, weisen darauf hin, dass diese prognostizierte Verbindung einen negativen Einfluss auf die Prognoseleistung des Modells haben. Umgekehrt ist ein Punkt im positiven Bereich ein Hinweis auf die Verbesserung der Vorhersage. Die Farben stellen zudem den SHAP-Wert dar und variieren von niedrigen Werten (gelb) bis zu hohen Werten (lila).

3 Hypothesen

In Kapitel 2.5 werden die Instrumente zur Prognose von räumlichen Interaktionen vorgestellt. Zunächst wurde deutlich, dass das parameterfreie Strahlungsmodell von Simini et al. (2012) dem klassischen Gravitationsmodell bedingt durch die Integration des Einwohnerpotenzials zwischen zwei Standorten überlegen ist und bessere Vorhersagen liefert. Zudem ist erkennbar, dass die Weiterentwicklung für finite Raumsysteme (vgl. Formel (3)) bessere Ergebnisse erzielt. Außerdem ist davon auszugehen, dass die verallgemeinerte Strahlungsmodelle (vgl. Formel (4) bis (7)) aufgrund der Integration eines Parameters besser gesteuert werden können und die Anpassungen zwischen empirischen Daten und Prognosen deutlich verbessert werden. Daher lassen sich folgende drei Hypothesen ableiten (Hypothesen werden mit M gekennzeichnet):

M1: Das Strahlungsmodell liefert bessere Vorhersagen von Mobilitätsströmen als das Gravitationsmodell.

M2: Das finite Strahlungsmodell liefert bessere Vorhersagen von Mobilitätsströmen als das Gravitations- und das Strahlungsmodell.

M3: Die verallgemeinerten Strahlungsmodelle liefern bessere Vorhersagen von Mobilitätsströmen als das Gravitations- und Strahlungsmodell.

Zudem ist aus Kapitel 2.5 zu entnehmen, dass insbesondere die PWO- und OPS-Modelle als weitere parameterfreie Modelle auf unterschiedlichen Maßstabsebenen dem parameterfreien Strahlungsmodell überlegen sind. Das heißt, es kann folgende Hypothese formuliert werden:

M4: Die PWO- und OPS-Modelle liefern bessere Vorhersagen von Mobilitätsströmen als das Strahlungsmodell.

Ein weiterer Aspekt, der für die bisher existierenden Prognosen mittels XGBoost-Algorithmus deutlich wird, ist, dass durch die Integration der individuellen Motive, ähnlich wie in der Thereorie von Lowry (1966) gefordert, die Schätzung der Pendlerströme deutlich verbessert werden können. Daher ist auch davon auszugehen, dass bei der Übertragung dieses Ansatzes auf

die Mobiliät von Studienanfängern und die Pendlerbewegungen in Deutschland bzw. Europa deutliche Verbesserungen gegenüber den „klassischen“ Teilchenmodellen erzielt werden können. Zudem ist auf Grund der theoretischen Überlegungen von Stouffer (1940) davon auszugehen, dass das Erweiterungsmodell (r, m, n, s) durch die Integration der Möglichkeiten zwischen Ziel- und Quellregion bessere Ergebnisse liefert als das Grundmodell (r, m, n). Um die individuellen Aspekte mit zu berücksichtigen, wird zusätzlich das Individualmodell entwickelt. Dadurch werden weitere Verbesserungen erwartet. Es lassen sich daher die folgenden Hypothesen formulieren:

M5: Das mittels XGBoost-Algorithmus berechnet Grundmodell, liefert bessere Vorhersagen von Mobilitätsströmen als das Gravitations- und Strahlungsmodell.

M6: Das mittels XGBoost-Algorithmus berechnet Erweiterungsmodell liefert bessere Vorhersagen von Mobilitätsströmen als das Grundmodell.

M7: Das mittels XGBoost-Algorithmus berechnet Individualmodell liefert bessere Vorhersagen von Mobilitätsströmen als das Grund- und Erweiterungsmodell.

Bei der Integration der unterschiedlichen Aspekte von Mobilität finden zunächst die Gesetze von Ravenstein Berücksichtigung. Aus den ersten drei Gesetzen sowie dem siebten Gesetz von Ravenstein ist die Grundidee von Newton ableitbar. Sie wird auch in klassischen Gravitationsgesetzen berücksichtigt. Die Interaktion zwischen zwei Orten ist proportional zur Masse und antiproportional zur Distanz. Das bedeutet, übertragen auf die Studierenden und Pendler, je mehr Einwohner bzw. Studierende ein Standort hat, desto anziehender ist er, während die Interaktion bei zunehmenden Entfernungen abnimmt. Zudem wird für die Mobilität von Studierenden deutlich, dass die Entfernung u. a. durch die sozialen Netzwerke in der Heimat von entscheidender Bedeutung für die Wahl des Studienortes ist (Frenette 2004 & 2006; Spiess/Wrohlich 2010; Sá et al. 2004). Daher sind folgende Hypothesen für das Mobilitätsverhalten von Studierenden (Hypothese wird mit S gekennzeichnet) und Pendlern (Hypothese wird mit P gekennzeichnet) ableitbar:

S1: Je größer die Entfernung zwischen Heimat- und Hochschulregion ist, desto unwahrscheinlicher ist die studentische Mobilität zwischen den Standorten.

P1: Je größer die Entfernung zwischen Wohn- und Arbeitsort ist, desto unwahrscheinlicher ist die Pendlerwahrscheinlichkeit zwischen den Standorten.

S2: Je mehr Studienanfänger ein Hochschulstandort im Vorjahr hatte, desto anziehender ist er für Studienanfänger.

P2: Je mehr Einwohner ein Arbeitsort hat, desto anziehender ist er für Pendler.

Ein weiterer Aspekt aus den Gesetzen von Ravenstein, der allerdings etwas differenzierter betrachtet werden muss, sind die Mobilitätsströme zwischen Stadt und Land. In der internationalen Literatur wird deutlich, dass für Studierende und Pendler besonders Standorte mit einem hohen Urbanisierungsgrad und hohen Einwohnerdichten anziehend sind. Daher werden folgende Hypothesen abgeleitet:

S3: Je höher der Urbanisierungsgrad ist, desto stärker ist der Mobilitätsstrom von Studienanfängern.

P3: Je höher der Urbanisierungsgrad ist, desto stärker ist der Mobilitätsstrom von Pendlern.

Zudem ist bei der Übertragung der Gesetze sieben bis zehn von Ravenstein zu erkennen, dass es eine ökonomische Komponente bei der Mobilität von Studierenden und Arbeitnehmern gibt. Beide Gruppen wählen häufig Orte mit einer geringen Arbeitslosigkeit und einer hohen Bruttowertschöpfung als Zielregion aus. Es lassen sich daher folgende Hypothesen formulieren:

S4: Je stärker eine Region aus ökonomischer Perspektive ist, desto stärker ist der Mobilitätsstrom von Studienanfängern.

P4: Je stärker eine Region aus ökonomischer Perspektive ist, desto stärker ist der Mobilitätsstrom von Pendlern.

Ein Aspekt, den Lee (1966) aufgreift ist, dass neben harten Push- und Pullfaktoren auch individuelle Aspekte eine entscheidene Rolle bei der Wahl des Wohn- und Arbeitsortes bzw. des Hochschulortes spielen. Sowohl Studierende als auch Pendler wählen ihren Hochschulstandort und Wohnort unter Berücksichtigung von weichen Faktoren, wie der Lebensqualität, aus. Daher spielen trotz intervenierender Hindernisse (z. B. Fahrtkosten, familiäre Verbundenheit, existierende soziale Netzwerke) vorallem auch Erholungsflächen, Angebot an Freizeiteinrichtungen, kulturelle Angebote und das Angebot an Infrastruktureinrichtungen eine Rolle bei der Wahl des Standortes. Ferner greift auch Wolpert (1956) diese Überlegungen auf. In seinem Theorieansatz werden vor allem die Rahmenbedingungen im Umfeld bei der Wahl des Arbeits- und Wohnortes bzw. des Hochschulstandortes berücksichtigt. Daher können folgende Hypothesen abgeleitet werden:

S5: Je höher die Lebensqualität in der Hochschulregion ist, desto stärker ist der Mobilitätsstrom von Studienanfängern.

P5: Je höher die Lebensqualität in der Wohngemeinde ist, desto stärker ist der Mobilitätsstrom von Pendlern.

Darüber hinaus wird bei Stouffer (1940) thematisiert, dass die Wahrscheinlichkeit, eine Entfernung zurückzulegen, proportional zu der Zahl der angebotenen Möglichkeiten in der Zielregion ist. Daher können folgende Hypothesen abgeleitet werden:

S6: Je mehr mögliche Hochschulstandorte zwischen Quell- und Zielregion liegen, desto unwahrscheinlicher ist, dass Studienanfänger den Zielort auswählen.

P6: Je mehr mögliche Arbeitsorte zwischen Quell- und Zielregion liegen, desto unwahrscheinlicher ist, dass Pendler den Zielort auswählen.

4 Datengrundlage

Studentische Mobilität

Der Wissensstandort Deutschland ist für die Schätzung von studierender Mobilität besonders geeignet, da es sich um ein Land mit beträchtlicher Größe und international wettbewerbsfähigen Programmen handelt. Ferner sind Universitäten und Fachhochschulen gleichmäßig über das gesamte Land verteilt. In Deutschland existieren, anders als in den USA, keine „Bildungswüsten" (vgl. Hillman 2016). Dies gilt auch für die neuen Bundesländer. Nach der deutschen Wiedervereinigung sind in den 1990er Jahren vor allem in den ehemaligen Gebieten der DDR neue Hochschulen gegründet worden (insgesamt 64). Dies ist insbesondere für ländliche Gebiete, wie beispielsweise der Lausitz und Ost-Brandenburg, von hoher wirtschaftlicher Bedeutung (Erhart 2002, S. 72 f.). Seit 2000 sind insgesamt 91 Universitäten und Fachhochschulen entstanden. Bei den Neugründungen handelt es sich häufig um Privathochschulen oder zusätzliche Standorte bestehender Hochschulen. Ferner haben sich in Agglomerationsräumen auch zahlreiche regionale Ableger bestehender Berufsakademien oder Ausgründungen aus Universitäten und Forschungseinrichtungen etabliert. Diese Ausgründungen weisen oftmals hohe vierstellige bzw. teilweise sogar mittlere fünfstellige Studierendenzahlen auf, wie beispielsweise die Duale Hochschule Baden-Württemberg oder das Karlsruher Institut für Technologie (HRK 2019/KIT 2018).

Hinzu kommt, dass im Vergleich zum angloamerikanischen Raum, keine Studiengebühren erhoben werden. Zwar sind zwischen 2006 und 2007 in allen westdeutschen Bundesländer, mit Ausnahme von Bremen, Rheinland-Pfalz und Schleswig-Holstein, allgemeine Studiengebühren in unterschiedlicher Höhe eingeführt worden, durch politische Veränderungen und Regierungswechsel wurden diese bis 2014 in allen Bundesländern allerdings wieder abgeschafft (Kauder/Potrafke 2013). Somit entstehen auf Grund der fehlenden finanziellen Möglichkeiten keine Verzerrungseffekte (Weisser 2019). Darüber hinaus dienen die Studiengebühren in den USA zum Teil zur Finanzierung der Hochschulen. Daher existiert eine ausgeprägte Marketingaktivität im Bereich der Rekrutierung von Studenten, um die Einnahmen aus Studiengebühren zu maximieren. Dies führt ebenfalls zu schwer messbaren Effekten, die in Deutschland nicht existieren (Weisser 2019).

Doch trotz dieser Vorteile als Untersuchungsraum muss bei der Analyse

Berücksichtigung finden, dass auch das deutsche Hochschulsystem durch die Angleichung der Studienstrukturen mit Bachelor- und Masterabschlüssen von Fachhochschulen und Universitäten in den letzten 20 Jahren erheblichen Veränderungen unterlegen war. In der Bologna-Erklärung vom 13. Juni 1999 hielten die Bildungsminister der heutigen Europäischen Union sowie von Island, Norwegen und der Schweiz die Grundsätze fest, nach welchen es zu mehr Kompatibilität und Vergleichbarkeit im europäischen Hochschulraum kommen sollte (EHEA 2016c). Zwölf Jahre nach der Bologna-Erklärung waren in der Bestandsaufnahme bereits 82 % der Studiengänge auf das neue Bachelor- und Mastersystem umgestellt (KMK 2011, S. 1). Neben den Änderungen der Abschlüsse kommen weitere hochschulinterne und -externe Einflüsse hinzu. So hat beispielsweise die Umstellung vom neun- zum achtjährigen Abitur in den verschiedenen Bundesländern (KMK 2012, S. 2) und die Aussetzung der Wehrpflicht im Jahr 2011 (Deutscher Bundestag 2011) zu erheblichen Steigerungen der Studierendenzahlen geführt.

Als Untersuchungsebene für den Vergleich der vorgestellten Prognoseinstrumente dienen die 401 Landkreise und kreisfreien Städte. Die Daten liegen im Zeitraum von 1996 bis 2016 vor, wobei die Anzahl an Abiturienten aus der Datenbank INKAR (2022) stammen. Zudem sind die Daten mithilfe des Umrechnungsschlüssels des BBSR (2022) um die Effekte der Kreisreformen bereinigt worden.

Die Anzahl der Studierenden, die Anzahl der Erstsemester sowie die jeweiligen innerdeutschen Wanderungsflüsse der zukünftigen Studierenden vom Heimatkreis zum Hochschulstandort stammen aus der Forschungsdatenbank in Frankfurt (FDZ 1996 - 2016). Es werden aus Datenschutzgründen lediglich Wanderungsbewegungen mit mehr als drei Studierenden berücksichtigt. Eine Einschränkung dieses Datensatzes ist, dass lediglich die studentischen Mobiltätsströme vom Ort der Hochschulzugangangsberechtigung zum jeweiligen Hochschulstandort bestimmt werden können. Inwieweit der Landkreis, in dem die Hochschulzugangsberechtigung erworben wurde, auch dem Wohnort der (potenziellen) Studienanfänger entspricht, lässt sich in diesem Zusammenhang nicht klären.

Zusätzlich werden die euklidische Distanzen zwischen allen Kreisen mithilfe der Haversine-Formel auf Basis der geometrischen Zentren berechnet. Die Distanz r zwischen Ziel- und Quellregion wird somit wie folgt bestimmt:

$$r := 2d \arcsin\left(\sqrt{\sin^2\left(\frac{\varphi_2 - \varphi_1}{2}\right) + \cos\varphi_1 \cdot \cos\varphi_2 \cdot \sin^2\left(\frac{\lambda_2 - \lambda_1}{2}\right)}\right) \quad (13)$$

wobei d der Radius der Sphäre, φ_1, φ_2 die geografische Breite (*latitude*) sowie λ_1, λ_2 die geografische Länge (*longitude*) von Quell- und Zielregion ist.

Zur weiteren Analyse der Motive von studentischer Mobilität werden die 401 Kreise in Deutschland für das Jahr 2016 untersucht. Als Indikatoren zur Bestimmung des Einflusses (vgl. Tabelle 2 und Kapitel 2.4) studentischer Mobilität wird unter anderem für die Herkunft- und Zielregion die Einwohnerdichte herangezogen. Sie bildet den Urbanisierungsgrad des Kreises ab. Die Attraktivität von Hochschulstandorten wird über die Anzahl an Studienanfänger im Vorjahr und dem Durchschnittsalter abgebildet. Außerdem finden die Mietpreise Berücksichtigung. Des Weiteren konnte gezeigt werden, dass Studierende häufig in wirtschaftsstarke Regionen mit einer niedrigen Arbeitslosigkeit ziehen. Daher werden die Beschäftigungsquote am Wohnort sowie das Haushaltseinkommen mit in das Modell integriert. Der Anteil an Naherholungsflächen dient als Indikator für die Lebensqualität. Alle Daten stammen aus der Datenbank INKAR (2022). Eine Übertragung auf weitere europäische Länder ist auf Grund von fehlenden Daten nicht möglich.

Variablen	Definition	Literatur	Erwarteter Trend in i	Erwarteter Trend in j
Distanz	km	Frenette 2004 & 2006; Spiess/Wrohlich 2008; SÁ et al. 2013	+	-
Einwohner-dichte	Einwohner je km^2	SÁ et al. 2004; Cullinan/Duggan 2016; Weisser 2019	-	+
Beschäftigungs-quote	Beschäftigungsquote am Wohnort	Cooke/Boyle 2011; Dotti et al. 2013	-	+
Studienanfänger im Vorjahr		Buenstorf et al. 2016; Haussen/Uebelmesser 2018	-	+
Durchschnitts-alter	Durchschnittsalter der Bevölkerung in Jahren	Buenstorf et al. 2016; Haussen/Uebelmesser 2018	-	+
Mietpreise	Klasse der durchschnittlichen Angebotsmiete	Dotti et al. 2013	+	-
Haushalts-einkommen	Durchschnittliches Haushaltseinkommen in € je Einwohner	SÁ et al. 2004	-	+
Naherholungs-flächen	Erholungsfläche in m^2 je Einwohner	Buch et al. 2014	-	+

Tabelle 2: Erwarteter Einfluss der Variablen auf die studentische Mobilität aus Perspektive i. Quelle: Eigene Darstellung.

Arbeitskräfte Mobiliät

Als Untersuchungsgebiet für den Vergleich der vorgestellten Prognoseinstrumente dienen zunächst die 4.528 Gemeindeverbände sowie die 401 Kreise Deutschlands im Jahr 2017. Als Stichtag dient der 30. Juni 2017. Die Wanderungsflüsse der Pendler von Wohn- zum Arbeitsort stammen von der Bundesagentur für Arbeit (Bundesagentur für Arbeit 2017).

Grundsätzlich sind nach der Beschäftigungsstatistik alle sozialversicherungspflichtig Beschäftigten mit unterschiedlichem Arbeits- und Wohnort Pendler. Es werden aus Datenschutzgründen lediglich Wanderungsbewegungen mit mehr als zehn Pendlern erfasst. Ferner kann es zu Verzerrungseffekten kommen, sofern alle Beschäftigten eines Betriebs mit mehreren Niederlassungen in der Hauptniederlassung gemeldet werden. Zudem ist in den Meldevorschriften nicht klar geregelt, ob der Arbeitnehmer den Erst- oder Zweitwohnsitz angibt. Somit ist es möglich, dass die Angaben des Erstwohnsitzes zu einem Nachweis (wenn auch in geringer Zahl) von Fernpendlern führt. Typisches Beispiel ist, dass der Arbeitnehmer unter der Woche an seinem Zweitwohnsitz lebt (z. B. Pension, Firmenunterkunft) und keine tägliche Pendlerbewegung auftritt (Bundesagentur für Arbeit 2017). Da die Daten nicht verhanden sind, werden die Ein- und Auspendler aus dem Ausland nicht weiter berücksichtigt.

Zusätzlich werden die euklidischen Distanzen zwischen allen Gemeindeverbänden und Kreisen mithilfe der Haversine-Formel auf Basis der geometrischen Zentroide berechnet (vgl. Formel (14)).

Die Indikatoren zur Beschreibung der Pendlermotive, die aus Kapitel 2.4 bzw. Tabelle 3 abgeleitet werden können, wurden in der Datenbank INKAR (2022) abgefragt. Als Indikatoren werden die Einwohnerzahl des Wohn- und Arbeitsortes sowie die Distanzen zwischen den Gemeinden herangezogen. Zusätzlich wird die Urbanität mit der Einwohnerdichte abgebildet. Darüber hinaus wird in Kapitel 2.4 deutlich, dass Pendlermobilität häufig bedingt ist durch ökonomische Aspekte, sodass die Beschäftigungsquote am Wohnort und die Gewerbesteuereinahmen an den jeweiligen Standorten integriert werden. Die Beschäftigungsquote von Frauen erfasst die geschlechtsspezifischen Unterschiede. Da die Anzahl von Haushaltsmitgliedern und Kindern nicht auf Gemeindeverbandsebene zur Verfügung steht, findet eine Integration von dem Anteil an Einwohnern unter 6 Jahre statt. Auf Gemeindeverbandsebene sind die Miet- und Bodenpreise ebenfalls nicht verfügbar, daher wurde alternativ auf die Neubauwohnungen zurückgegriffen. Ferner wurde die Erreichbarkeit

Variablen	Definition	Literatur	Erwarteter Trend in i	Erwarteter Trend in j
Distanz	km	Eckey et al. 2007; Kersting et al. 2020; Simini et al. 2012	+	-
Einwohner-dichte	Einwohner je km²	Heuermann/Schmieder 2019 Martinus et al. 2020	-	+
Gewerbesteuereinnahmen	Gewerbesteuer in € je Einwohner	Pfaff 2012; Kersting et al. 2020	-	+
Beschäftigungs-quote	Beschäftigungsquote am Wohnort	Kersting et al. 2020, Broersma et al. 2020; Eckey et al. 2007; Schulze 2009 Chatterjee et al. 2019	-	+
Beschäftigungsquote von Frauen	Beschäftigungsquote von Frauen am Wohnort	Häfner 2011	-	+
Anteil an Einwohnern unter 6 Jahren	Anteil der Einwohner von unter 6 Jahren an den Einwohnern in %	Kalter 1994; Pfaff 2012; Huber/Nowotny 2013	-	+
Neubauwohnungen	Fertiggestellte Wohnungen in neuen Wohngebäuden je 1.000 Einwohner	Guirao et al. 2018; Krebs/Pflüger 2019; Schulze 2009	+	-
Naherholungs-flächen	Erholungsfläche in m^2 je Einwohner	Chatterjee et al. 2019; Stenpass 2020	-	+
Breitbandversorgung	Anteil der Haushalte mit Breitbandversorgung mit 100 Mbit/s in %	Falck/Schüller 2016	-	+
Erreichbarkeit Oberzentrum	Durchschnittliche Pkw-Fahrzeit zum nächsten Oberzentrum in Minuten	Einig/Pütz 2007; Schulze 2009; Kauffmann 2011	+	-
Durchschnittsdistanz ÖV-Haltestelle	Einwohnergewichtete Luftliniendistanz zur nächsten Haltestelle des ÖV mind. 10 Abfahrten am Tag	Heuermann/Schmieder 2019; Krebs/Pflüger 2019; Broersma et al. 2020	+	-
Erreichbarkeit der Autobahnauffahrt	Pkw-Fahrzeit zur nächsten Autobahnanschlussstelle in Minuten	Broersma et al. 2020	+	-
Erreichbarkeit von IC/EC/ICE-Bahnhöfen	Durchschn. Pkw-Fahrzeit zum nächsten IC/EC/ICE-Bahnhof in Minuten	Kauffmann 2011; Heuermann/Schmieder 2019; Guirao et al. 2018	+	-

Tabelle 3: Erwarteter Einfluss der Variablen auf das Pendlerverhalten aus Perspektive i. Quelle: Eigene Darstellung.

des nächsten Oberzentrums als Globalindikator für eine Fülle von höherwertigen Infrastruktureinrichtungen berücksichtigt. Zudem werden als weitere Variablen folgende Entfernungen zu Infrastrukturen mit einbezogen: Durchschnittsdistanz zur nächsten ÖPNV-Haltestelle und Autobahnauffahrt sowie zum nächsten IC/EC/ICE-Bahnhof. Als ein weiterer Indikator für die Lebensqualität wird stellvertretend der Anteil an Naherholungsflächen und der Ausbaugrad des Breitbandes integriert (vgl. Tabelle 3).



Abbildung 5: Übersicht der untersuchten Länder in Europa. Quelle: Eigene Darstellung. Kartographie: Lisett Diehl.

Im Anschluss der Betrachtung für Deutschland werden die Modelle auf weitere europäische Länder übertragen, um die Vorhersagen noch auf unterschiedliche Städtesyteme anzuwenden (vgl. Abbildung 5). Hierzu wird neben Deutschland, Spanien, England und Wales mit einem relativ ausgeglichenen und homogen verteilten Städtesystem sowie Frankreich als Flächenland mit einer starken Zentralisierung auf Paris herangezogen. Zusätzlich werden Österreich und die Schweiz als alpine Länder mit berücksichtigt. Außerdem werden anhand von Dänemark, Finnland, Schweden, Norwegen und Irland Länder mit wenigen Untersuchungseinheiten die Auswirkungen von unterschiedlichen Bevölkerungsverteilungen getestet. Während in Dänemark die Bevölkerung trotz der großen Bedeutung der Hauptstadt Kopenhagen relativ homogen über das Land verteilt ist, kann in Irland ein starkes Ost-West-Gefälle festgestellt werden. In Italien, Norwegen, Schweden und Finnland dominieren neben den weiten Distanzen die Süd-Nord-Disparitäten. Zudem wurde Kroatien als ehemaliger Ostblockstaat und jüngstes Mitglied der Europäischen Union gewählt.

Als räumliche Ebene werden die NUTS3-Regionen (in Deutschland Landkreise und kreisfreie Städte) der jeweiligen Länder betrachtet. Um eine Vergleichbarkeit zu erzielen, wird für die Pendlerdaten und die Einwohnerzahl das Jahr 2011 als Untersuchungszeitpunkt gewählt. Ausnahme bildet die Schweiz. Hier wird das Jahr 2014 verwendet, da die Daten für 2011 nicht vorliegen (vgl. Bundesamt für Statistik 2014). Die Einwohnerzahlen stammen von der Statistikplattform der Europäischen Union (Eurostat). Die individuellen Motive für das Pendlerverhalten finden jedoch keine Berücksichtigung, da die hergeleiteten Motive für Deutschland (vgl. Tabelle 3) nicht flächendeckend zur Verfügung gestellt werden.

Die Pendlerdaten für England und Wales stammen aus der Zensusbefragung 2011. Es werden die erwerbstätigen Einwohner von England und Wales im Alter von 16 Jahren und älter berücksichtigt. Als Untersuchungszeitraum dient dabei die Woche vor der Zählung (Stichtag: 27. März 2011). Limitation des Datensatzes ist, dass die Arbeitsplatzdaten für Schottland und Nordirland nicht angegeben werden, da die Daten außerhalb von England und Wales nicht weiter untergliedert sind. Auf Grund von Datenschutz wurden Datensätze in der Volkszählungsdatenbank zwischen verschiedenen geografischen Gebieten ausgetauscht, so dass einige Zählungen, insbesondere auf kleinräumiger Ebene, davon betroffen sind und es zu Verzerrungen kommen kann (Office for National Statistics 2011).

Die Daten für Frankreich stammen aus den ergänzenden Untersuchungen der Volkszählung. Es werden alle Erwerbstätige ab 15 Jahren berücksichtigt. Als Erwerbstätige zählen Personen, die einen Beruf (mit oder ohne Gehalt, sowie Teilzeit) ausüben, Lehrlinge oder bezahlte Praktikanten, Arbeitslose bei Ausübung einer eingeschränkten Tätigkeit, Studenten oder Rentner mit einem Job sowie Soldaten. Zudem finden nur Personen mit Wohnsitz in Frankreich Berücksichtigung. Personen mit Zweitwohnsitz werden an ihrem Erstwohnsitz registriert, somit können die Entfernungen zwischen Wohn- und Arbeitsort erhebliche Schwankungen aufweisen. Die Berechnung der Pendlerströme basiert auf folgenden Überlegungen: Für Gemeinden mit weniger als 10.000 Einwohner werden alle Auspendler in dem Jahr erfasst, in dem diese Gemeinde befragt wurde. Sofern die Gemeinde 10.000 oder mehr Einwohner aufweist, wurden die Auspendler über den Zeitraum von fünf Jahren erfasst. Das heißt, die Messung der Volkszählung spiegelt eine durchschnittliche Situation wider. Für die Einpendler gilt, dass die Personen an ihrem Wohnort (bzw. diese Adressen bei Gemeinden ab 10.000 Einwohnern) und damit zu unterschiedlichen

Zeitpunkten (abhängig vom Befragungszeitraum) gezählt werden. Das bedeutet, durch die zeitliche Staffelung der Datenerhebung findet die Erhebung der Ein- und Auspendler nicht zum selben Zeitpunkt statt. Die Abweichungen sind allerdings in der Regel marginal. Ausnahme bilden Gemeinden, in denen während des Fünfjahreszeitraums große Betriebe entstanden sind oder geschlossen wurden. Generell weisen geringe Durchflüsse (weniger als 200) auf Grund der Erhebungsmethode große Unsicherheiten auf (Institut national de la statistique et études économiques 2011).

Die Pendlerdaten für Irland stammen aus der Volkszählung 2011. Der Stichtag ist der 10. April 2011. Für alle Arbeitnehmer mit Wohnsitz in Irland am Stichtag wurde der Arbeitsort geocodiert. Die Daten enthalten neben den Beschäftigten, die tagsüber arbeiten, Nachtschichtarbeiter sowie Personen im Home Office. Personen, die keine Angaben zum Ort ihrer Arbeitsstätte oder keinen festen Arbeitsplatz gemacht haben, sind ausgeschlossen (Central Statistics Office 2011).

Die Daten für Österreich stammen aus der Registerzählung 2011 und stellen eine Vollerhebung zu Merkmalen der österreichischen Bevölkerung dar. Der Stichtag ist der 31. Oktober 2011. Die Auswertung basiert auf Verwaltungs- und Registerdaten. Der Wohnort der Erwerbstätigen liegt dabei auf österreichischem Staatsgebiet. Als Pendler wird nicht die Gesamtzahl der Erwerbspersonen betrachtet, sondern lediglich eine Teilmenge der „Erwerbstätigen“. Das heißt, Personen in Elternzeit (Karenz) werden zum Beispiel nicht berücksichtigt. Aus Datenschutzgründen findet für Daten mit wengier als fünf Pendlerbewegungen eine Bearbeitung mit der Methode *„Target Swapping“* statt. *„Target Swapping“* ist ein Verfahren, das die Daten so verschmutzt, dass diese bei Aggregatauswertungen möglichst unverzerrt sind. Zudem werden gleichzeitig die individuellen personenbezogenen Merkmale optimal geschützt (Statistik Austria 2011).

Für die Schweiz (vgl. Bundesamt für Statistik 2014) werden die Pendler auf Basis der Erwerbstätigen bestimmt, die eine oder auch mehrere Beschäftigungen mit einem alters- und hinterlassenversicherungspflichtigen Lohn (d. h. mindestens 2.300 Franken pro Beschäftigung und Jahr) aufweisen. Als Erhebungszeitraum wird der Dezember des jeweiligen Jahres gewählt. Sofern eine Person für mehrere Unternehmen arbeitet, wird jenes Unternehmen als Arbeitgeber festgelegt, bei dem die erwerbstätige Person den höchsten Lohn bezieht. Sofern die Arbeitnehmer einen Zweitwohnsitz haben und die

Fahrzeit zur Arbeitsstätte geringer ist als vom Erstwohnsitz, wird der Zweitwohnsitz als Wohngemeinde verwendet. Zudem werden Erwerbstätige, die in einem Mehrbetriebs-Unternehmen (2 Millionen Personen im Jahr 2014) arbeiten, mithilfe eines Optimierungsalgorithmus einer bestimmten Arbeitsstätte des Unternehmens zugeteilt. Aus Datenschutzgründen werden Pendlerbewegungen mit weniger als fünf Erwerbstätigen gruppiert.

Die Pendlerstatistik für Dänemark basiert auf der Grundlage der *registerbaserede arbejdsstyrkestatistik* (RAS-Statistik). Als Stichtag dient der letzte Arbeitstag im November und es handelt sich um eine jährliche, personenbezogene Zählung der Erwerbstätigen. Die RAS-Statistik enthält Informationen über die Wohn- und Arbeitsadresse der Beschäftigten. Es werden lediglich Personen mit Wohnsitz in Dänemark berücksichtigt. Unter Erwerbstätigen versteht man eine Person mit einem Arbeitsplatz, die im Bezugszeitraum mindestens eine Stunde einer Erwerbstätigkeit nachgeht. Berücksichtigung finden auch Personen, die im Untersuchungszeitraum abwesend sind (z. B. wegen Krankheit, Elternzeit), aber eine dauerhafte Bindung an einen Arbeitgeber haben. Zudem fallen Selbständige, mitarbeitende Ehegatten und Personen, die einer unterstützten Beschäftigung nachgehen (z. B. Rehabilitation, Flex-Jobs, Lohnzuschüsse etc.) in die Gruppe der Erwerbstätigen. Grundsätzlich wird in der dänischen Statistik davon ausgegangen, dass ein Pendler ein Erwerbstätiger mit einer von Null verschiedenen Pendelstrecke ist. Sofern Arbeits- und Wohnadresse gleich sind, fallen die Erwerbstätigen in die Kategorie „kein Arbeitsweg“. Problematisch ist die Zuordnung zu einem Unternehmen, sofern mehr als eine Beschäftigung vorliegt. Es wird dann der Job mit den meisten Stunden als Hauptarbeitsplatz definiert (Danmarks Statistik 2011).

Die Pendlerdaten für Finnland stammen aus der registerbasierten Beschäftigungsstatistik und berücksichtigen die ständige Wohnbevölkerung des Landes am letzten Tag des Jahres. Die Daten über die Tätigkeit und den Arbeitsplatz der Bevölkerung stammen aus der letzten Woche des Jahres. Zudem werden alle Erwerbstätige von 18 bis 67 Jahre berücksichtigt. Der Gebietsstand der Gemeinden basiert auf dem Jahr 2013 (Statistics Finland 2011).

Die Pendlerdaten für Schweden stammen aus der jährlichen registergestützten Arbeitsmarktstatistiken (*registerbaserade arbetsmarknadsstatistiken* / RAMS) und berücksichtigen die Erwerbstätigen zwischen 16 und 74 Jahren, die in einer Gemeinde gemeldet sind, aber nach Angaben der Unternehmen in einer ande-

ren Gemeinde arbeiten. Als Erwerbstätig gelten in Schweden Person, die über ein Lohneinkommen verfügen, welches einen steuerpflichtigen Schwellenwert übersteigt. Als Zeitraum wird in der Regel der November gewählt. Zudem gilt, dass jede Personen nur mit einem Unternehmen und somit mit einem Arbeitsort verbunden sein kann. Sofern der Erwerbstätige mehr als einen Arbeitsplatz hat, wird das Unternehmen gewählt, an dem das höchste Einkommen erzielt wird (Statistics Sweden 2011).

In Norwegen (Statistic Norway 2011) werden die Erwerbstätigen im Alter zwischen 15 und 74 Jahren differenziert nach Arbeits- und Wohngemeinde im vierten Quartal des Jahres angegeben. Grundsätzlich versteht man unter Erwerbstätigen Personen, die in der Berichtswoche mindestens eine Stunde lang gegen Entgelt oder zur Erzielung eines Gewinns gearbeitet haben oder vorübergehend von der Arbeit abwesend sind (z. B. wegen Krankheit, Urlaub etc.). Wehrpflichtige werden ebenfalls als Erwerbstätige eingestuft sowie Personen, die im Rahmen staatlicher Beschäftigungsmaßnahmen beschäftigt sind. Sofern Erwerbstätige in der Berichtswoche mehr als eine Beschäftigung haben, werden sie dem Unternehmen mit der höchsten Stundenzahl zugeordnet. Der Wohnort wird in der dritten Novemberwoche bestimmt.

Die Pendlerbewegungen stammen aus der Volkszählung 2011 und beziehen sich auf einen einwöchigen Referenzzeitraum (die Referenzwoche lag zwischen dem 25. und 31. März 2011). Es werden Personen ab 15 Jahren erhoben. Arbeitnehmer sind grundsätzlich als Personen definiert, die auf der Grundlage eines formellen oder informellen Vertrags arbeiten und für ihre Arbeit mit Geld- oder Sachleistungen entlohnt werden. Der Wohnort der Personen wird auf Grundlage des Ortes bestimmt, an dem die meiste Zeit verbracht wird (Croatian Bureau of Statistics 2011).

Die Daten für Italien umfassen die Herkunfts-Ziel-Matrix für Arbeits- und Studienpendler. Sie ist im Rahmen der Volkszählung entstanden (Stichtag: 9. Oktober 2011). Als Grundgesamtheit aller Befragten dient die in Haushalten oder Lebensgemeinschaften lebende Bevölkerung in Italien. Insgesamt werden daher 28.871.447 Personen berücksichtigt, die angaben, täglich zu ihrem üblichen Studien- oder Arbeitsort zu fahren, wobei sie ihren Wohnort verließen und zu ihm zurückkehrten (Istituto Nazionale di Statistica 2011).

Die Pendlerdaten für Spanien stammen aus der Volkszählung 2011 (Stichtag: 1. November 2011). Berücksichtigt werden die Personen im Alter von über 15 Jahren, die in der Berichtswoche erwerbstätig gewesen sind. Als Wohnort wird

der Ort definiert, an dem die befragten Personen die meiste Zeit des Jahres leben. Die Angaben über den Arbeitsort werden mithilfe eines Fragebogens erhoben. Aus Datenschutzgründen werden lediglich Interaktionen mit mehr als fünf Beschäftigten berücksichtigt (Instituto nacional de Estadística 2011).

5 Mobilitätsprognosen für Studienanfänger

Betrachtet man zunächst die Entwicklungen der Studienanfänger und Schulabgänger deskriptiv, so fällt auf, dass im Untersuchungszeitraum beide Personengruppen stark gewachsen sind. Wie in Abbildung 6 deutlich wird, ist die Anzahl an Studierenden im ersten Semester von 1995 bis 2016 von 262.407 auf 509.760 gestiegen. Das heißt, die durchschnittliche jährliche Wachstumsrate der Zahl der Erstsemester (deutsche und ausländische Erstsemester an allen Hochschultypen) betrug von 1995 bis 2016 2,82 Prozent. Im Jahr 2011 wurde mit 518.748 Studierenden die höchste Zahl an Erstsemestern gemessen (Statistisches Bundesamt 2008 - 2017).

Der Anteil der Abiturienten, die ein Studium beginnen, und auch die Anzahl an Schulabgängern, die eine Hochschulzugangsberechtigung erworben haben, ist kontinuierlich gewachsen. Die durchschnittliche jährliche Wachstumsrate bei Abiturienten lag bei 1,73 Prozent. Insgesamt sind die Abiturienten von 207.161 im Jahr 1995 auf 296.736 im Jahr 2016 gestiegen und die höchste Anzahl an Schulabgängern mit Hochschulzugangsberechtigung konnte im Jahr 2013 mit 318.900 verzeichnet werden (INKAR 2022).

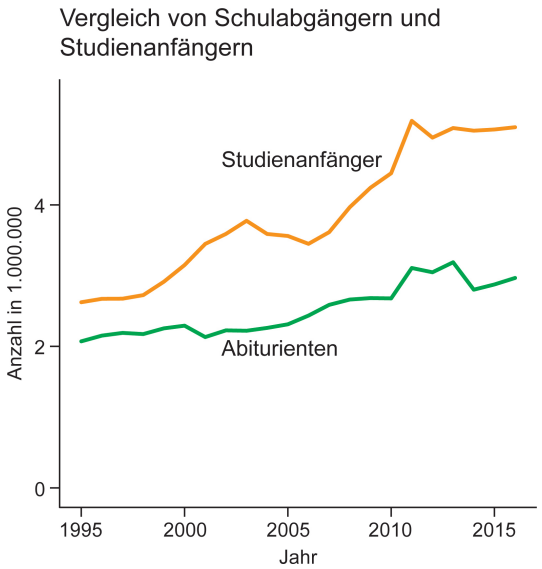


Abbildung 6: Vergleich von Schulabgängern und Studienanfängern. Quelle: Eigene Berechnung auf Grundlage von Statistischem Bundesamt 2008-2017a-b und INKAR 2022.

Die Mobilitätsflüsse der Studienanfänger zeigen, dass die Mobilität zwischen 1995 und 2016 stark gestiegen ist. Während im Jahr 1995 die Mobilitätsströme der Studienanfänger (mit Binnenpendlern) zwischen 3 und 10.930 Personen schwankten, variierten sie im Jahr 2016 zwischen 3 und 15.003 Personen, wobei

der jeweils größte Mobilitätsstrom die Binnenbewegung innerhalb von Berlin repräsentiert. Darüber hinaus liegt das erste Quantil im Jahr 1995 bei 4, der Median und das arithmetische Mittel bei 6 bzw. 23,63 sowie das dritte Quantil bei 17. Für das Jahr 2016 betragen das erste und dritte Quantil ebenfalls 4 bzw. 17. Der Median und das arithmetische Mittel nehmen Werte von 7 und 29,73 an (FDZ 1996-2016).

Die Analyse des Fluktuationsindex von Sörensen (Sörensenindex) macht im Vergleich der unterschiedlichen Modelle im Zeitverlauf (vgl. Abbildung 7) deutlich, dass für die jüngeren Jahrgänge eine bessere Übereinstimmung zwischen empirischer und simulierter studentischer Mobilität auftritt. Zudem wurden neben dem Sörensenindex die Korrelationen berechnet (vgl. Abbildung 8). Auffallend ist in diesem Zusammenhang, dass die Korrelationskoeffizienten am Anfang des Untersuchungszeitraums deutlich niedriger ausfallen als die Berechnungen der Zusammenhangsmaße zwischen 2009 und 2016. Diese Beobachtung ist auch für den adjustierten R^2-Wert (vgl. Abbildung 9) festzustellen. Ferner zeigen Abbildung 10 und 11, dass der mittlere absolute Fehler und die Wurzel der mittleren quadratischen Abweichung im Zeitverlauf geringere Werte annehmen.

Darüber hinaus werden die Boxplots in Abbildung 12 (a) grün gefärbt, sofern die Linie y=x zwischen dem 9. und dem 91. Perzentil in der jeweiligen Klasse liegt. Andernfalls wird die Farbe rot gewählt (vgl. Simini et al. 2012). Das heißt, dass das Gravitationsmodell (vgl. Abbildung 12 (a)) im Bereich von 2.000 Studierenden eine gute Anpassung liefert.

Ein Vergleich mit den Wahrscheinlichkeiten in Bezug zur Anzahl an Studienanfängern aus dem Vorjahr veranschaulicht (vgl. Abbildung 13 (a)), dass für Studienstandorte mit geringen Einschreibezahlen die Prognose mittels Gravitationsmodell eine Überschätzung aufweist. Insbesondere für Standorte mit 1.000 bzw. 2.000 Studierenden liegt die Wahrscheinlichkeit für die Modelle deutlich über den Ergebnissen aus den empirischen Daten. Ferner konnte für diese Modelle festgestellt werden, dass die Wahrscheinlichkeiten für Standorte mit sehr großen Einschreibezahlen leicht unterschätzt werden. Betrachtet man die Wahrscheinlichkeiten, einen Hochschulstandort in Abhängigkeit zur Distanz aufzusuchen (vgl. Abbildung 14 (a)), zeigt sich, dass im direkten Umfeld bis zu zehn Kilometern der Mobilitätsfluss deutlich überschätzt wird.

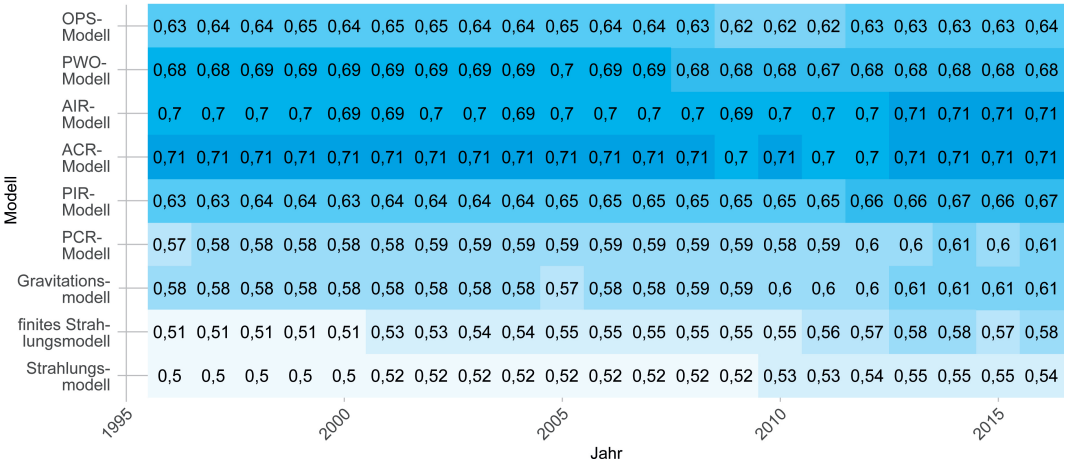


Abbildung 7: Der Sörensenindex im Zeitverlauf. Quelle: Eigene Berechnung auf Grundlage von FDZ 1996-2016 und INKAR 2022.

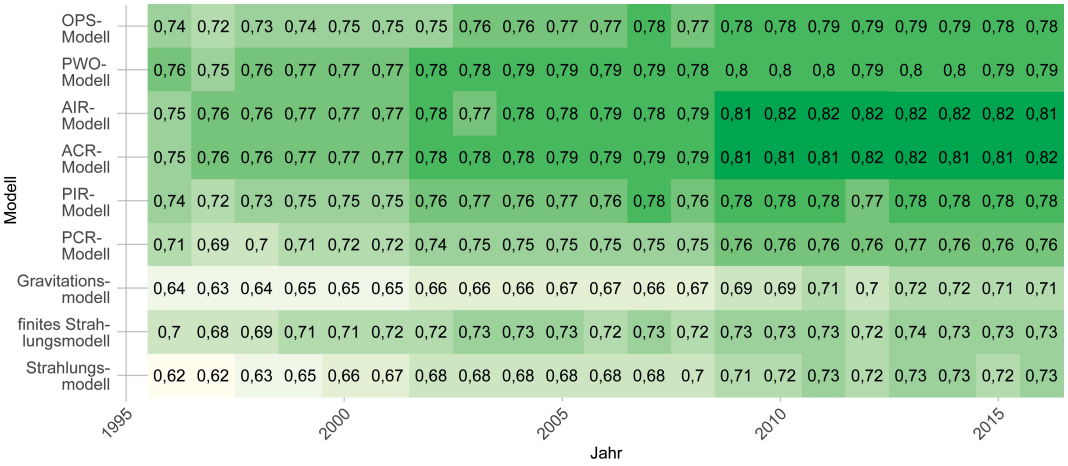


Abbildung 8: Der Pearson-Korrelationskoeffizient im Zeitverlauf. Quelle: Eigene Berechnung auf Grundlage von FDZ 1996-2016 und INKAR 2022.

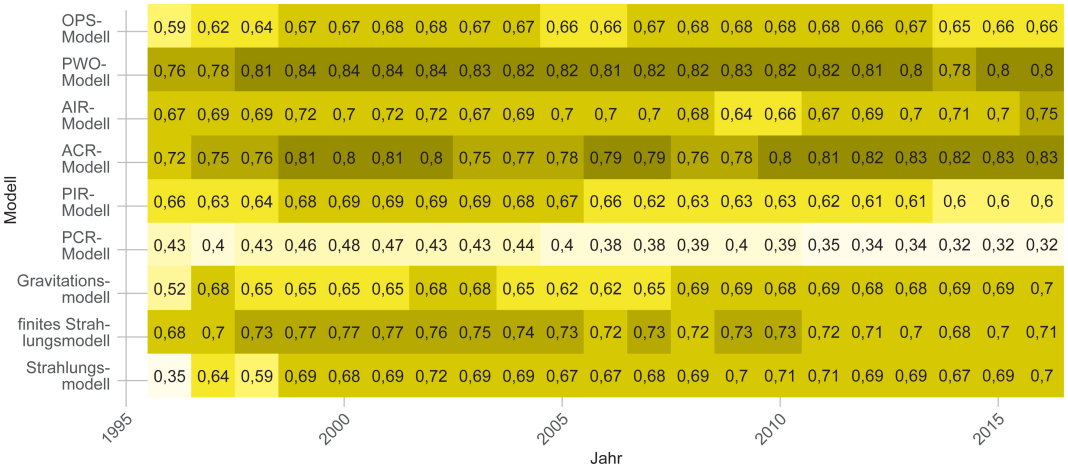


Abbildung 9: Der adjustierte R^2-Wert im Zeitverlauf. Quelle: Eigene Berechnung auf Grundlage von FDZ 1996-2016 und INKAR 2022.

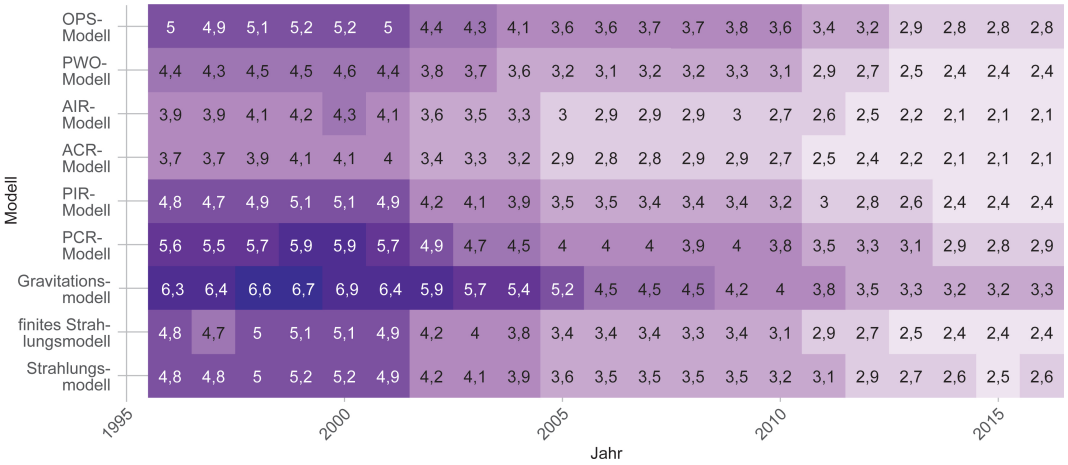


Abbildung 10: Der mittlere absolute Fehler im Zeitverlauf. Quelle: Eigene Berechnung auf Grundlage von FDZ 1996-2016 und INKAR 2022.

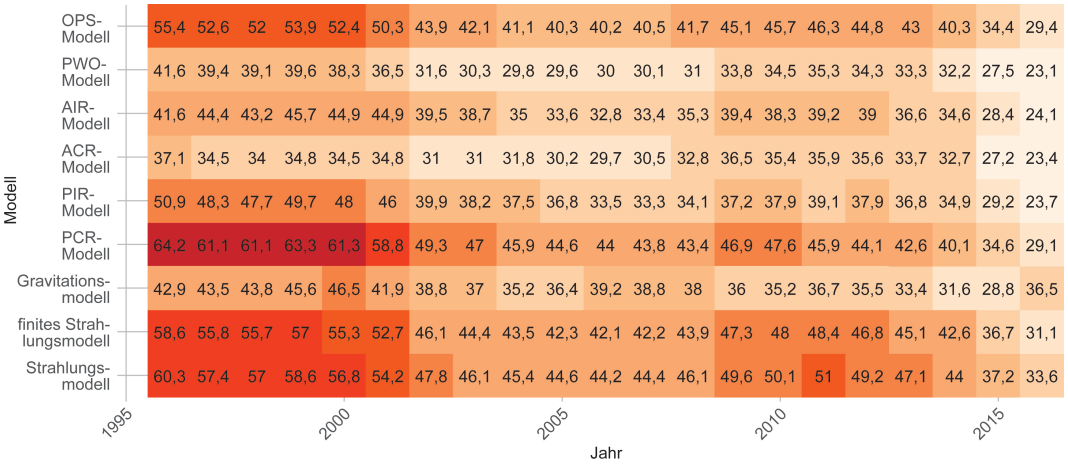


Abbildung 11: Die Wurzel der mittleren quadratischen Abweichung im Zeitverlauf. Quelle: Eigene Berechnung auf Grundlage von FDZ 1996-2016 und INKAR 2022.

Im Gegensatz dazu liegt mit dem Gravitationsmodell die geschätzte Wahrscheinlichkeit zwischen ca. 30 und 130 Kilometern deutlich unter den Ergebnissen aus den empirischen Daten, während ab 170 Kilometern das Modell erneut höhere Wahrscheinlichkeiten erzielt.

Die Betrachtung der Evaluationsindikatoren zeigt zudem, dass ein durchschnittlicher Sörensenindex von 0,59 auftritt (vgl. Abbildung 7). Der Korrelationskoeffizient schwankt zwischen 0,64 und 0,72 (vgl. Abbildung 8) und der adjustierte R^2-Wert zwischen 0,52 und 0,70 (vgl. Abbildung 9). Zudem variiert der mittlere absolute Fehler zwischen 3,3 und 6,6 (vgl. Abbildung 10). Die Schwankungen bei der Wurzel der mittleren quadratischen Abweichung liegen zwischen 31,5 und 46,6 (vgl. Abbildung 11).

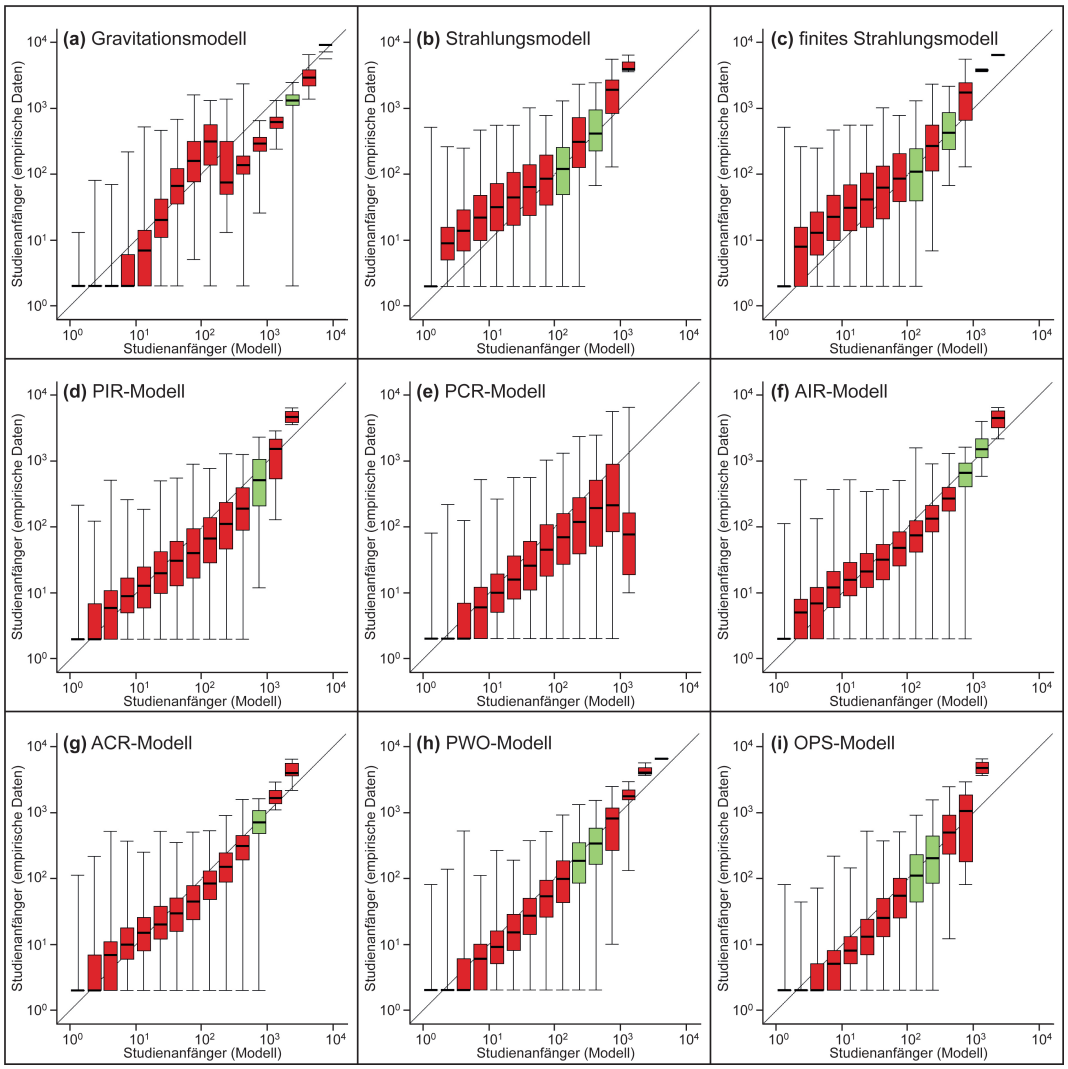


Abbildung 12: Abweichung zwischen den empirischen Daten und den Modellen (Studienanfänger). Quelle: Eigene Berechnung auf Grundlage von FDZ 2016 und INKAR 2022.

Generell werden für das Strahlungsmodell und das finite Strahlungsmodell geringe Wanderungsbewegungen von Studierenden überschätzt. Besonders gute Anpassungen sind für Mobilitätsströme bei ca. 100 bzw. 1.000 Studierende erkennbar (vgl. Abbildung 12 (b) und (c)). Bezüglich der Standortgröße kommt es sowohl beim Strahlungmodell als auch beim finiten Modell zu Überschätzungen im Bereich zwischen 1.000 und 2.000 Studierenden, während große Standorte systematisch unterschätzt werden (vgl. Abbildung 13 (b) und (c)). Die Wahrscheinlichkeiten für das Strahlungsmodell und das finite Strahlungsmodell, dass Studierende einen Hochschulstandort in Abhängigkeit der Distanz aufsuchen, weisen insgesamt nur marginale Unterschiede auf. Die Modellwahrscheinlichkeiten entsprechen bis zu ca. 60 Kilometern Radius den empirischen Daten beider Modelle. Ab ca. 70 Kilometern werden die Wahrschein-

lichkeiten der Modelle leicht überschätzt (vgl. Abbildung 14 (b) und (c)). Der Sörensenindex für das Strahlungsmodell weist mit durchschnittlich 0,52 die geringste Anpassung an die wirkliche Verteilung auf. Der Durchschnitt für das finite Strahlungsmodell liegt bei 0,54 (vgl. Abbildung 7). Das Zusammenhangsmaß von Pearson liefert für das Strahlungsmodell Werte zwischen 0,62 und 0,73 (vgl. Abbildung 8). Die Ergebnisse für das finite Strahlungsmodell schwanken zwischen 0,63 und 0,74. Beim adjustierten R^2-Wert (vgl. Abbildung 9) sind für das Strahlungsmodell und das finite Strahlungsmodell Schwankungen zwischen 0,35 und 0,71 bzw. zwischen 0,68 und 0,77 festzustellen. Zudem variiert der mittlere absolute Fehler beim Strahlungsmodell zwischen 2,5 und 5,2 sowie beim finiten Strahlungsmodell zwischen 2,4 und 5,1 (vgl. Abbildung 10). Die Schwankungen der Wurzel der mittleren quadratischen Abweichung liegen bei 33,6 und 60,3 (Strahlungsmodell) sowie 31,1 und 58,6 (vgl. Abbildung 11).

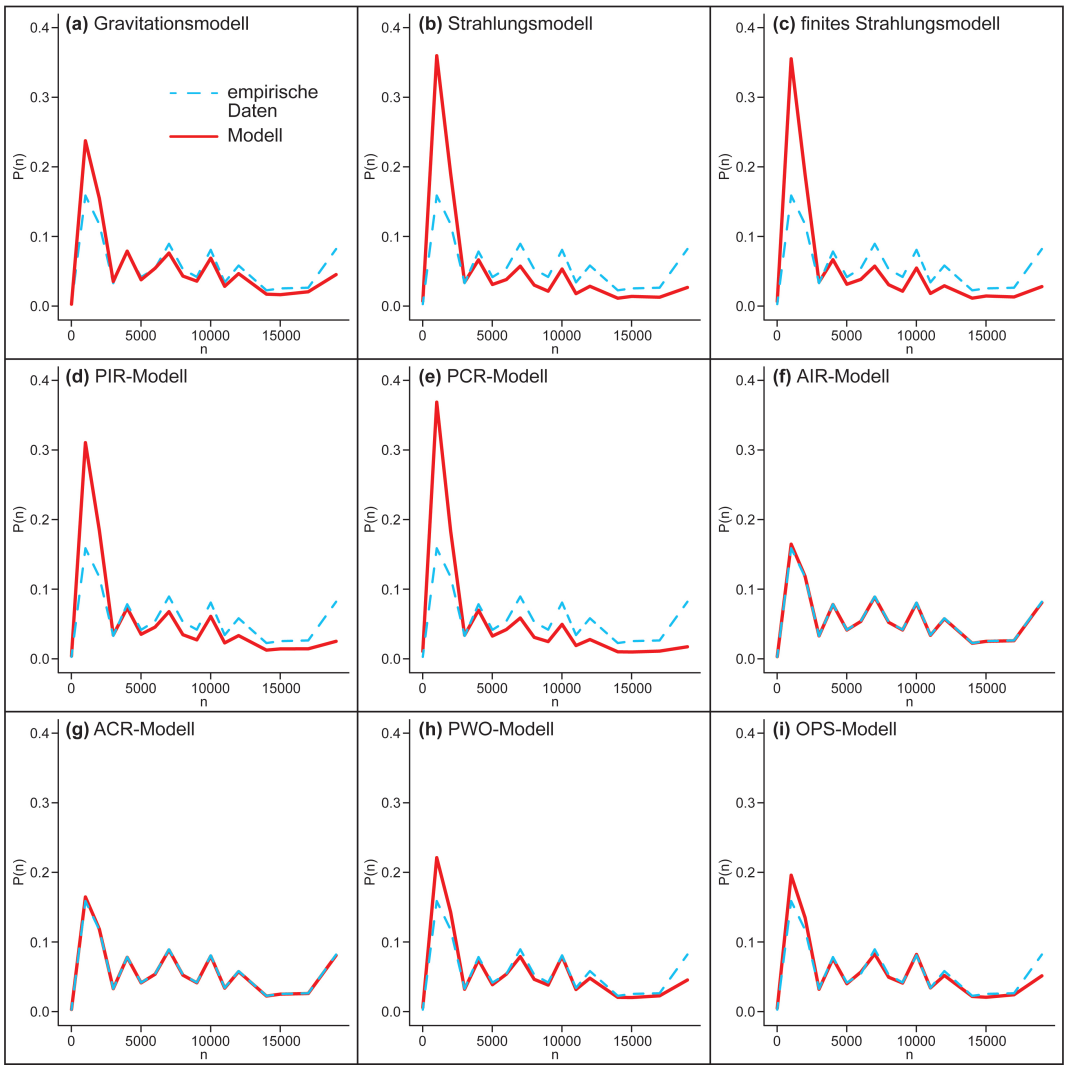


Abbildung 13: P(n) im Verhältnis zur Größe des Hochschulstandorts. Quelle: Eigene Berechnung auf Grundlage von FDZ 2016 und INKAR 2022.

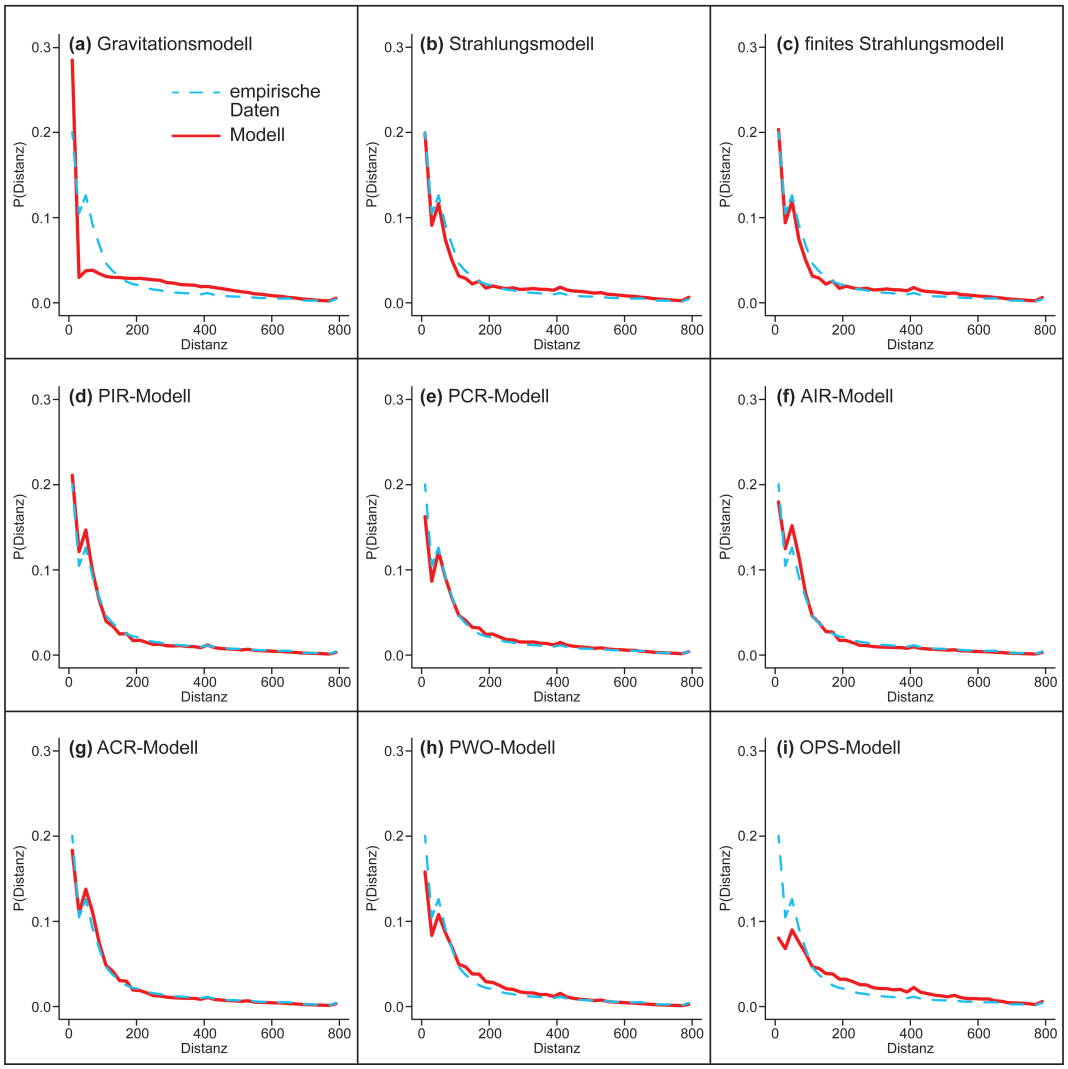


Abbildung 14: P(Distanz) im Verhältnis zur Entfernung (Mobilität von Studienanfängern). Quelle: Eigene Berechnung auf Grundlage von FDZ 2016 und INKAR 2022.

Eine deutlich höhere Übereinstimmung zwischen Modell und empirischen Daten erbrachten die generalisierten Modelle. Betrachtet man die Studierendenflüsse für die verallgemeinerten Strahlungsmodelle, so fällt auf, dass geringe Mobilitätsströme in allen vier Modellen unterschätzt werden. Bei ca. 1.000 Studierenden pro Mobilitätsfluss ist die Anpassung optimal (vgl. Abbildung 12. Ausnahme bildet das PCR-Modell, hier liegen keine optimalen Anpassungen vor (vgl. Abbildung (d) bis (g)). Bei der Wahrscheinlichkeit, einen Hochschulstandort aufzusuchen, sind, in Bezug auf die Studierendenzahl, insbesondere für das AIR- und ACR-Modell sehr geringe Abweichungen festzustellen. Beim PIR- bzw. PCR-Modell wird hingegen für kleine Standorte die Wahrscheinlichkeit überschätzt und für große Standorte unterschätzt (vgl. Abbildung 13 (d) bis (g)). Bezüglich der Distanz sind besonders gute Ergebnisse

für das PIR-Modell erzielt worden. Lediglich im Umkreis von 30 bis 50 Kilometern wird die Wahrscheinlichkeit der Auswahl des Hochschulstandortes leicht überschätzt. Ähnliche Ergebnisse sind auch für das ACR-Modell zu erkennen. Für das AIR-Modell ist die Wahrscheinlichkeit, die sich aus dem Modell ergeben hat, im Umkreis der ersten zehn Kilometer leicht unterschätzt. Das Modell weist weiterhin zwischen 30 und 70 Kilometern eine leichte Überschätzung auf. Beim PCR-Modell wird die Wahrscheinlichkeit, einen Hochschulstandort im Umkreis von 30 Kilometern aufzusuchen, unterschätzt. Ab ca. 50 Kilometern sind lediglich marginale Unterschiede zwischen Modell und den realen Daten zu erkennen (vgl. Abbildung 14 (d) bis (g)). Unter diesen vier Modellvarianten erweisen sich das ACR- und das AIR-Modell zwischen 1996 bis 2016 mit einem durchschnittlichem Sörensenindex von 0,71 und 0,70 als die besten Modelle (vgl. Abbildung 7). Der Durchschnitt für das PIR- und PCR-Modell liegt bei 0,59 bzw. 0,65. Außerdem konnten für das ACR- und das AIR-Modell eine hohe Korrelation zwischen 0,75 und 0,82 festgestellt werden. Diese Beobachtungen (vgl. Abbildung 9) sind auch für den adjustierten R^2-Wert zu machen (ACR-Modell: 0,72-0,83 / AIR-Modell: 0,64-0,75). Der Korrelationskoeffizient für das PIR- und PCR-Modell liegt zwischen 0,72 und 0,78 bzw. 0,69 bis 0,76 (vgl. Abbildung 8) und der adjustierte R^2-Wert zwischen 0,6 und 0,69 (PIR-Modell) sowie 0,32 und 0,48 (PCR-Modell) (vgl. Abbildung 9). Zudem untermauern die Ergebnisse des mittleren absoluten Fehlers (vgl. Abbildung 10) und der Wurzel der mittleren quadratischen Abweichung (vgl. Abbildung 11) diese Beobachtungen (AIR-Modell: mittlerer absoluter Fehler: 2,1-4,3; Wurzel der mittleren quadratischen Abweichung: 24,1-45,7 / ACR-Modell: mittlerer absoluter Fehler: 2,1-4,1; Wurzel der mittleren quadratischen Abweichung: 23,4-37,1 / PIR-Modell: mittlerer absoluter Fehler: 2,4-5,1; Wurzel der mittleren quadratischen Abweichung: 23,7-50,9 / PCR-Modell: mittlerer absoluter Fehler: 2,9-5,9; Wurzel der mittleren quadratischen Abweichung: 29,1-64,2).

Zusätzlich werden noch die Entwicklung der Parameter im Zeitverlauf für die generalisierten Modelle untersucht (vgl. Abbildung 15). Der Parameter für das ACR-Modell ist von 1996 bis 2016 stabil bei 1. Beim PIR- und AIR-Modell wird das Modell bis 2006 bzw. 2008 mit einem Faktor von 1 gewichtet. Es folgt eine Gewichtung mit 0,9. Beim PIR-Modell liegt der maximierende Parameter bis ins Jahr 2001 bei 1.

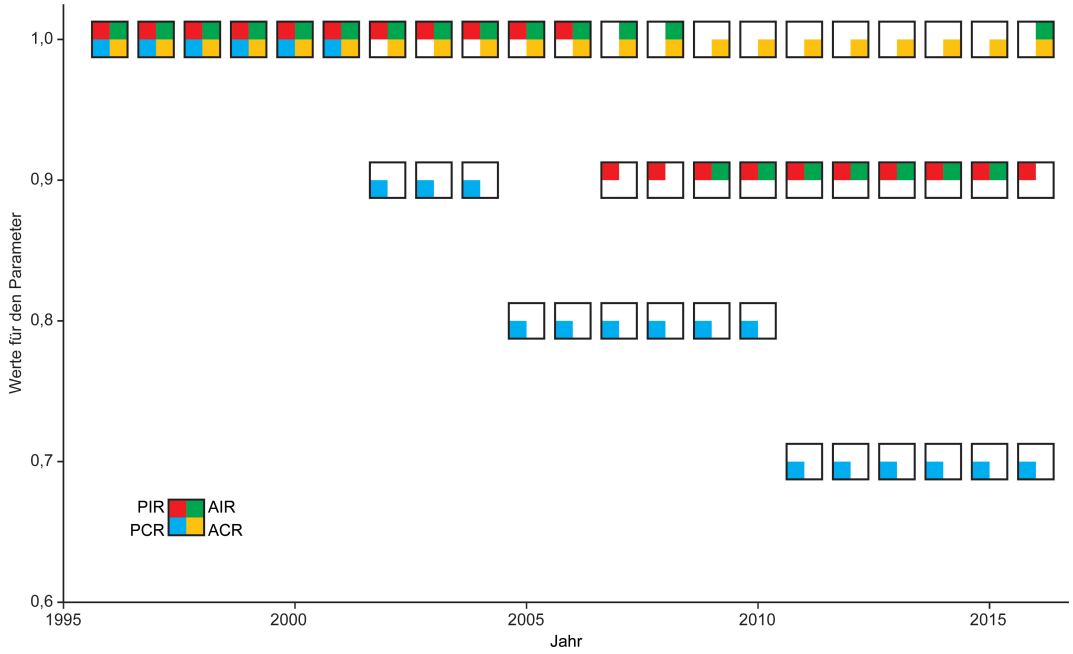


Abbildung 15: Zeitliche Entwicklung der Parameter. Quelle: Eigene Berechnung auf Grundlage von FDZ 1996-2016 und INKAR 2022.

Es folgen drei Jahre mit einem Koeffizienten von 0,9 und weitere fünf Jahre mit einem Parameter von 0,8. Ab 2011 liegt der optimierende Parameter bei 0,7.

Des Weiteren veranschaulicht Abbildung 12 (h), dass geringe studentische Mobilitätsflüsse unterschätzt werden. Auch in Bezug zur Größe des Hochschulstandortes sind die Abweichungen geringer als bei den nicht parametrisierten Modellen (vgl. Abbildung 13 (h)). Darüber hinaus wird in Abbildung 14 (h) deutlich, dass für das PWO-Modell kleine Distanzen bis ca. 30 Kilometer unterschätzt werden. Beim parameterfreien PWO-Modell variiert der Sörensenindex zwischen 0,67 und 0,70 (vgl. Abbildung 7). Durchschnittlich lag er bei 0,68. Der Korrelationskoeffizient nach Pearson bewegt sich zwischen 0,76 und 0,8 (vgl. Abbildung 8). Der adjustierte R^2-Wert nimmt Messwerte zwischen 0,76 und 0,83 an (vgl. Abbildung 9), der mittlere absolute Fehler Werte zwischen 2,4 und 4,6 (vgl. Abbildung 10) und die Wurzel der mittleren quadratischen Abweichung zwischen 23,1 und 41,6 (vgl. Abbildung 11).

Eine ähnlich gute Übereinstimmung kann für das OPS-Modell erzielt werden. Abbildung 12 (i) veranschaulicht, dass es für das OPS-Modell bei geringen Studierendenflüssen zu Unterschätzungen des Modells kommt, während Mobilitätsflüsse mit großen Studienanfängern systematisch überschätzt werden. Besonders gute Anpassungen sind für Mobilitätsströme von ca. 300 bzw. 500

Studienanfängern erkennbar. Die Wahrscheinlichkeit, einen Hochschulstandort aufzusuchen in Bezug zur Hochschulgröße, wird ähnlich wie in den nichtparametrisierten Modellen ebenfalls überschätzt. Allerdings sind die Abweichungen geringer (vgl. Abbildung 13 (i)). In Abbildung 14 (i) wird deutlich, dass für das OPS-Modell kleine Distanzen bis ca. 90 Kilometer unterschätzt werden. Im OPS-Modell ist die Unterschätzung stärker als im PWO-Modell. Im Bereich zwischen 90 und 400 Kilometern werden die Wahrscheinlichkeiten leicht überschätzt. Der Sörensenindex beträgt im Durchschnitt 0,64 (vgl. Abbildung 7). Der Korrelationskoeffizient nach Person variiert zwischen 0,62 und 0,65, der adjustierte R^2-Wert zwischen 0,59 und 0,68 (vgl. Abbildung 9), der mittlere absolute Fehler zwischen 2,8 und 5,2 (vgl. Abbildung 10) sowie die Wurzel der mittleren quadratischen Abweichung zwischen 29,4 und 55,4 (vgl. Abbildung 11).

Betrachtet man die SHAP-Werte für den XGBoost-Regressor (vgl. Abbildung 17 und 16) wird deutlich, dass die Anzahl an Studienanfängern am Hochschulstandort das wichtigste Merkmal in allen drei Modellen ist. Insbesondere Hochschulstandorte mit einer hohen Anzahl an Erstsemestern wirken sich positiv auf die Prognose der studentischen Mobilitätsflüsse aus. Standorte mit geringen Einschreibungen haben hingegen einen negativen Einfluss.

Außerdem zeigt sich, dass der von Simini et al. (2012) eingeführte Parameter s – der die Summe an Studierenden im ersten Semester zwischen dem Heimatkreis und dem zukünftigen Hochschulstandort repräsentiert – für das Erweiterungs- und Individualmodell (vgl. Abbildung 17 und 16a) als zweit wichtigstes Merkmal identifiziert wird. Darüber hinaus spielt in allen drei Modellen die Anzahl an Studienanfängern am Wohnort eine Rolle. Ein weiterer relevanter Aspekt ist die Distanz. Besonders kleine Distanzen haben einen hohen Einfluss auf die Vorhersageleistung der Modelle, während große Distanzen die Prognoseleistung verringern.

Zudem beeinflussen hohe Einwohnerdichten am Hochschulstandort die Berechnung der Studierendenströme des Individualmodells negativ (vgl. Abbildung 17). Hingegen bewirken geringe Einwohnerdichten das Gegenteil. Ähnliche Beobachtungen sind auch für die Einwohnerdichten im Heimatkreis festzustellen. Allerdings hat das durchschnittliche Alter am Hochschulstandort ein etwas höheres Gewicht. Vor allem Kreise mit einem hohen Durchschnittsalter führen zu einer Verschlechterung der vorhergesagten Mobilitätsflüsse.

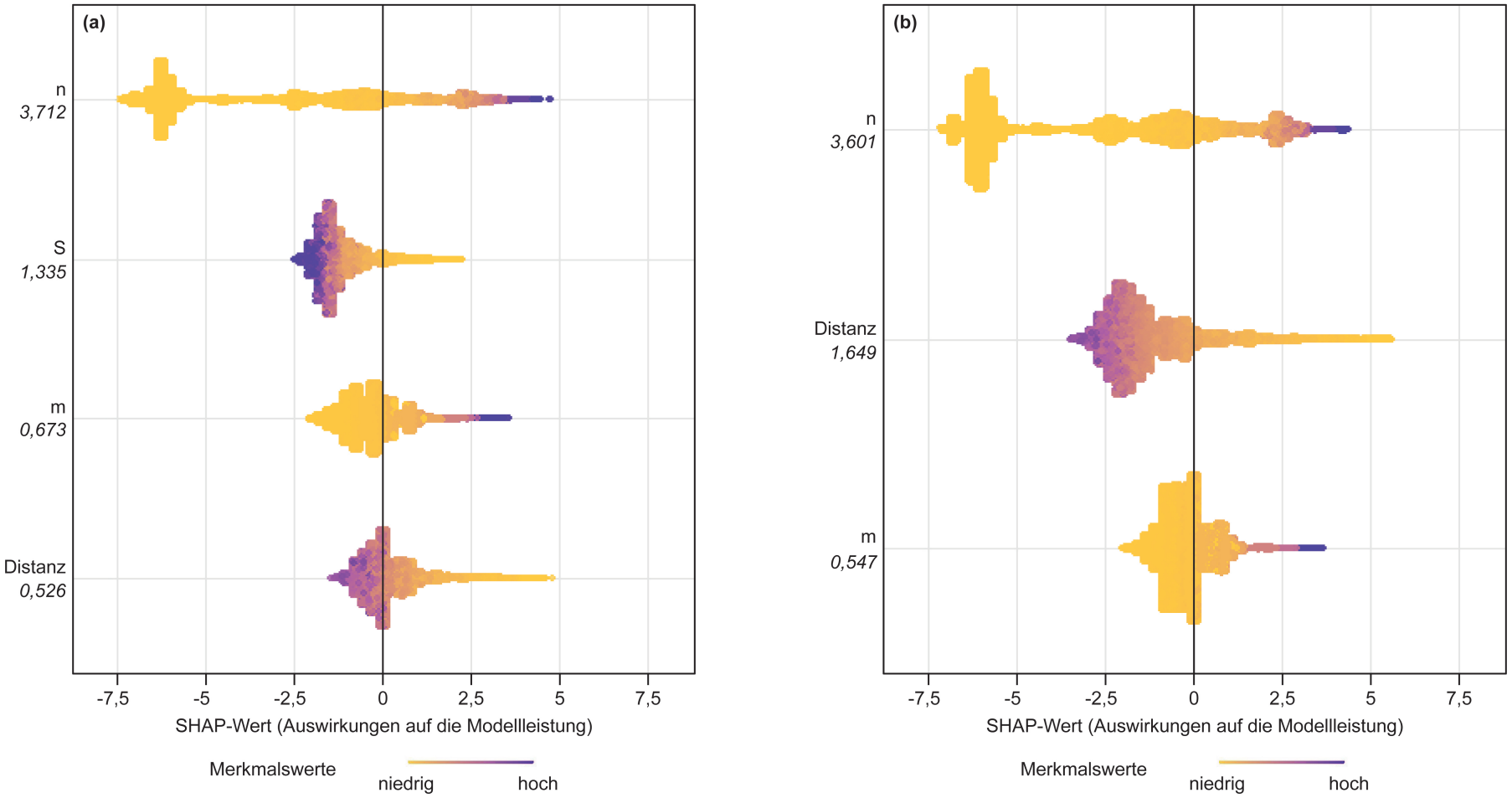


Abbildung 16: SHAP-Werte für den XGBoost-Regressor zur Prognose von studentischer Mobilität (Grund- und Erweiterungsmodell). Quelle: Eigene Darstellung auf Grundlage von eigenen Berechnungen von FDZ 2016 und INKAR 2022.

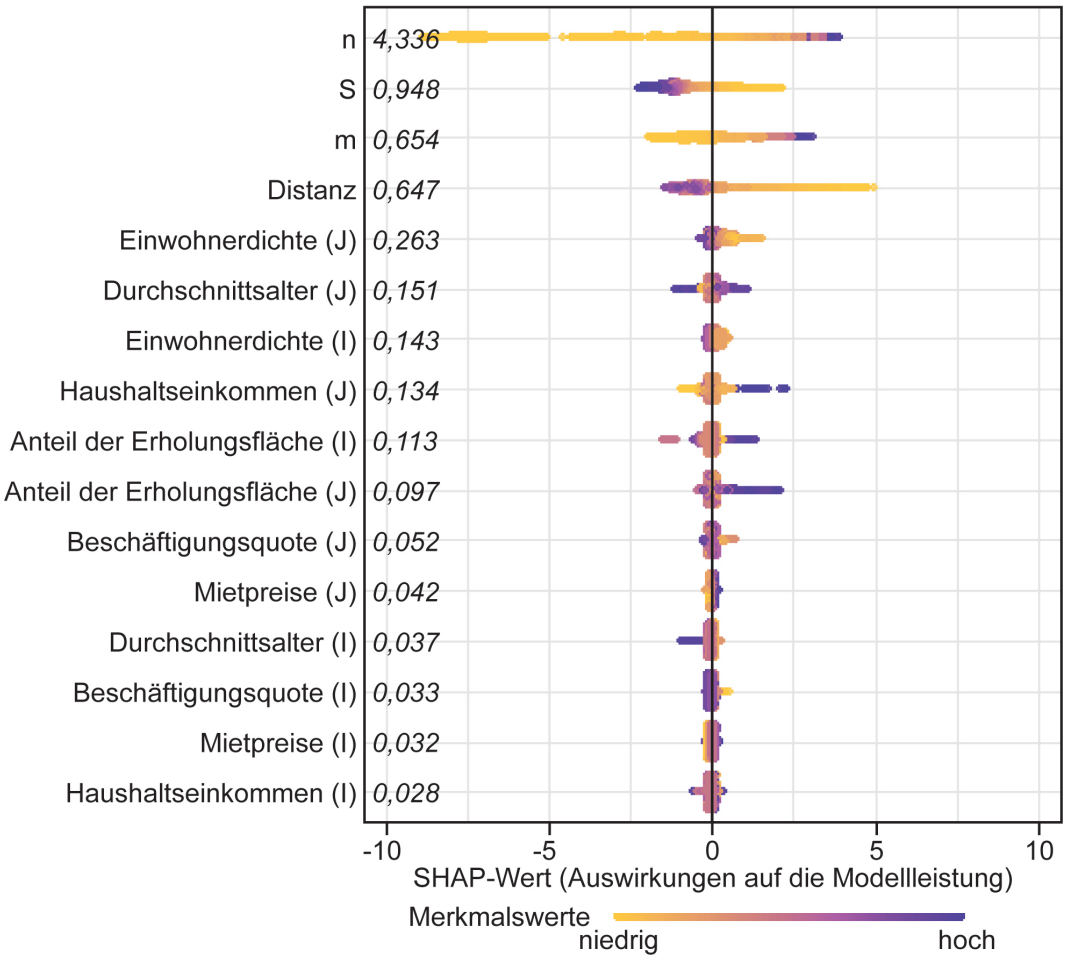


Abbildung 17: SHAP-Werte für den XGBoost-Regressor zur Prognose von studentischer Mobilität (Individualmodell). Quelle: Eigene Darstellung auf Grundlage von eigenen Berechnungen von FDZ 2016 und INKAR 2022.

Dies konnte auch für die Einwohnerstruktur im Heimatkreis festgestellt werden. Ferner sind hohe Haushaltseinkommen und ein hoher Anteil an Naherholungsflächen in der Ziel- und Herkunftsregion von positiver Bedeutung bei der Prognose. Für alle weiteren Merkmale, wie die Mietpreise, die Beschäftigungsquote und das Haushaltseinkommen in der Herkunftsregion, sind keine eindeutigen Aussagen zu treffen.

Zusätzlich zur Auswertung mittels SHAP-Werte werden noch Messindikatoren berechnet. Abbildung 18 (a) veranschaulicht für das Grundmodell, dass der Sörensenindex in 38,8 % der Simulationen bei 0,77 liegt. Der Korrelationskoeffizient nach Pearson beträgt 0,89, der adjustierte R^2-Wert 0,8. Außerdem nehmen der mittlere absolute Fehler und die Wurzel der mittleren quadratischen Abweichung Werte von 1,8 bzw. 14,1 an.

Zudem ist für das Erweiterungsmodell festzustellen, dass der Sörensenindex in

47,2 % der Simulationen 0,78 beträgt (vgl. Abbildung 18 (b)). Der Korrelationskoeffizient nach Pearson nimmt Werte von 0,9 und der adjustierte R^2-Wert von 0,82 an. Zudem liegt der mittlere absolute Fehler bei 1,6 und die Wurzel der mittleren quadratischen Abweichung bei 12,88.

Außerdem kann für das Individualmodell festgestellt werden, dass der Sörensenindex in 45,2 % der Simulationen bei 0,80 (vgl. Abbildung 18 (c)), der Korrelationskoeffizient nach Pearson bei 0,96, der adjustierte R^2-Wert bei 0,81, der mittlere absolute Fehler bei 1,55 und die Wurzel der mittleren quadratischen Abweichung bei 11,3 liegt.

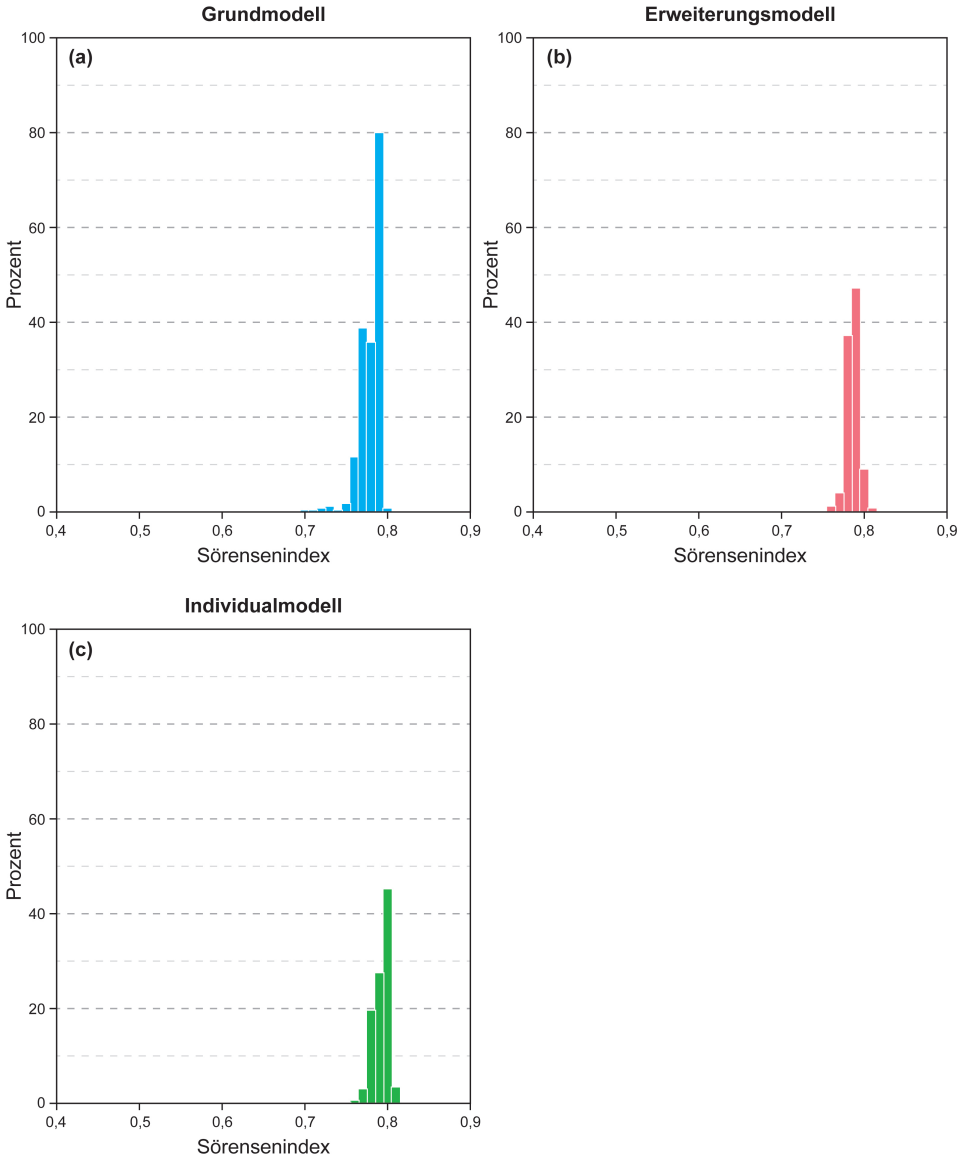


Abbildung 18: Ergebnisse der Simulation für die Mobilität von Studienanfängern. Quelle: Eigene Darstellung auf Grundlage von eigenen Berechnungen von FDZ 2016 und INKAR 2022.

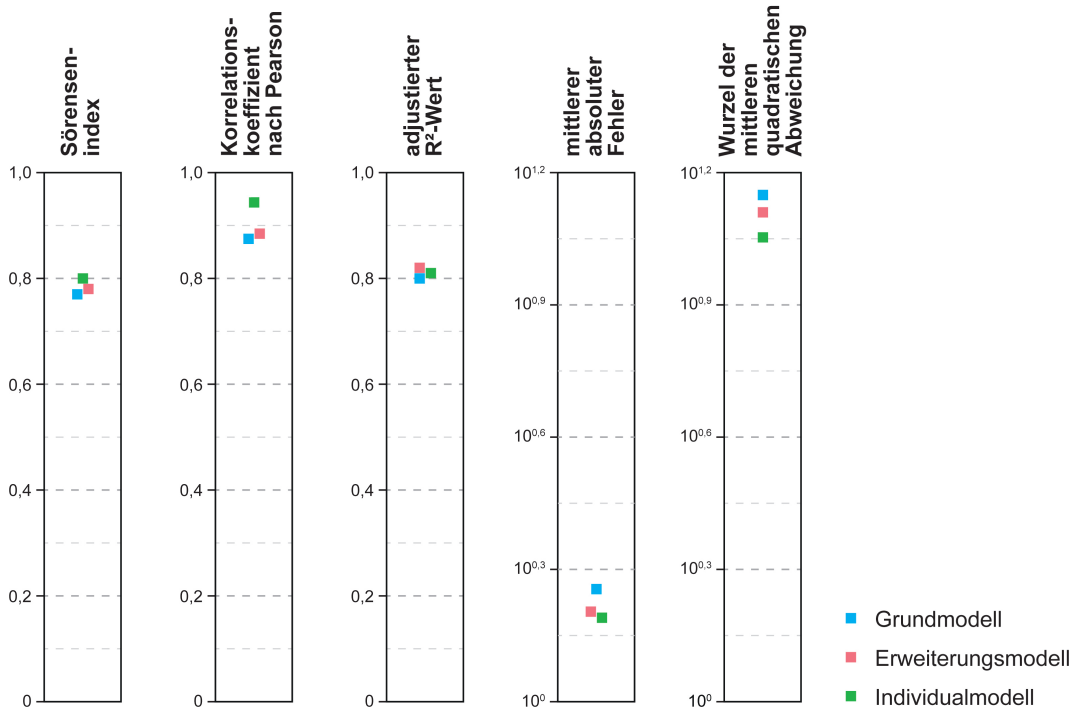


Abbildung 19: Ergebnisse der Evaluationsindikatoren für studentische Mobilitätsbewegungen mittels XGBoost-Algorithmus. Quelle: Eigene Darstellung auf Grundlage von eigenen Berechnungen von FDZ 2016 und INKAR 2022.

6 Mobilitätsprognosen für Pendlerbewegungen

Deutschland (Gemeindeverbandsebene)

Betrachtet man die Pendlerströme in Deutschland auf Gemeindeverbandsebene, wird deutlich, dass die Zahl der Pendler zwischen 10 und 22.513 schwankt. Die Verbindung mit den meisten Pendlern ist die Strecke zwischen Fürth und Nürnberg. Durchschnittlich fahren ca. 125 Personen täglich über die Gemeindeverbandsgrenzen. Das erste Quantil liegt bei 15, das zweite Quantil bei 30 und das dritte Quantil bei 81 Interaktionen (Bundesagentur für Arbeit 2017).

In Abbildung 20 (a) ist erkennbar, dass im Fall des Gravitationsmodells keine optimalen Anpassungen zwischen Modell und empirischen Daten vorliegen, da sich die Linie y=x nicht zwischen dem 9. und 91. Perzentil in der jeweiligen Klasse befindet. Zudem werden besonders die kleinen Distanzen bis ca. 50 Kilometer unterschätzt. Im Bereich von ca. 70 bis 250 Kilometern liegen Überschätzungen vor (vgl. Abbildung 21 (a)). Der Sörensenindex liegt für dieses Modell bei 0,51. Ferner beträgt der Korrelationskoeffizient nach Pearson und der adjustierte R^2-Wert 0,70 bzw. 0,49 (vgl. Abbildung 22). Darüber hinaus zeigen die Berechnungen in Abbildung 22, dass der mittlere absolute Fehler 2,52 und die Wurzel der mittleren quadratischen Abweichung Werte von 27,11 annimmt.

Der Vergleich der empirischen Daten mit den vorhergesagten Mobilitätsflüssen für das Strahlungsmodell zeigt, dass für Pendlerströme im Bereich von ca. 100 und über 10.000 Personen die Linie y=x zwischen dem 9. und 91. Perzentil liegt (vgl. Abbildung 20 (b)). Die Abweichungen bezüglich der Distanz (vgl. Abbildung 21 (b)) sind marginal. Der Sörensenindex liegt bei 0,79, der Korrelationskoeffizient nach Pearson bei 0,68, der adjustierte R^2-Wert bei 0,47, der mittlere absolute Fehler bei 0,66 und die Wurzel der mittleren quadratischen Abweichung bei 26,96 (vgl. Abbildung 22).

Eine ähnliche Beobachtung ist für das finite Strahlungsmodell zu machen. Allerdings können keine optimalen Anpassungen zwischen empirischen Daten und dem Modell festgestellt werden, da die Linie y=x nicht zwischen dem 9. und 91. Perzentil in der jeweiligen Klasse liegt (Abbildung

20 (c)). Zudem ist aus Abbildung 21 (c) zu entnehmen, dass im direkten Umfeld die Pendlerströme überschätzt werden. Außerdem existieren bei den Evaluationsindikatoren (vgl. Abbildung 22) für das finite Strahlungsmodell lediglich geringe Abweichungen (Sörensenindex: 0,8; Korrelationskoeffizient nach Pearson: 0,70; adjustierter R^2-Wert: 0,48; mittlerer absoluter Fehler: 0,66; Wurzel der mittleren quadratischen Abweichung: 26,97).

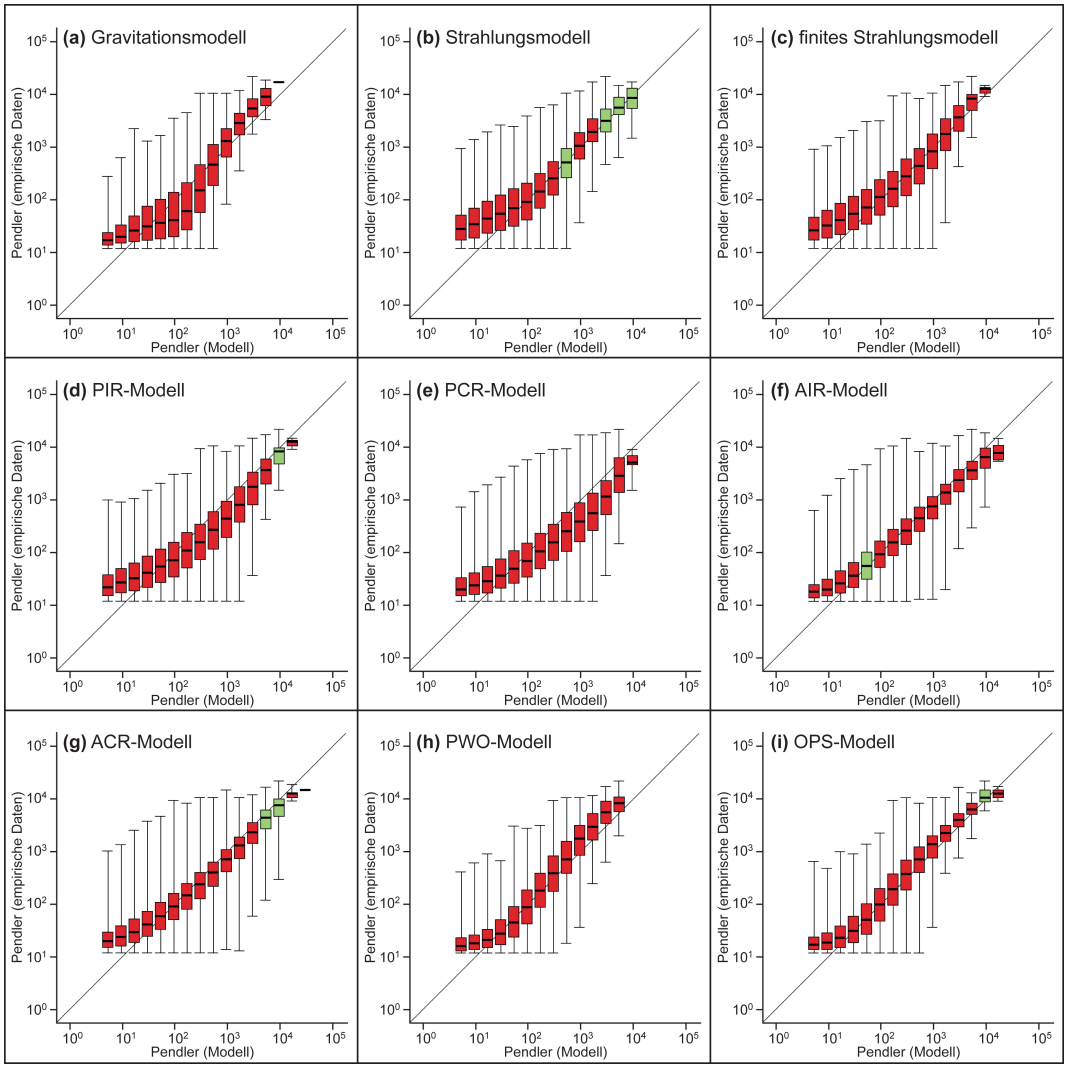


Abbildung 20: Abweichung zwischen den empirischen Daten und den Modellvorhersagen bezüglich der Pendlerbewegungen auf Gemeindeverbandsebenen. Quelle: Eigene Darstellung auf Basis von Bundesagentur für Arbeit (2017).

Darüber hinaus zeigt der Vergleich mit den verallgemeinerten Strahlungsmodellen, dass vor allem das ACR-Modell gute Anpassungen liefert. In Abbildung 20 (g) ist zu erkennen, dass Pendlerströme von etwa 5.000 bis ca. 10.000 Personen besonders gute Anpassungen erzielen. Außerdem ist in Abbildung 21 (g) ersichtlich, dass vor allem kleine Distanzen leicht überschätzt werden.

Die Betrachtung der Evaluationsindikatoren zeigt zudem, dass der Sörensenindex bei 0,85, der Korrelationskoeffizient nach Pearson bei 0,82, der adjustierte R^2-Wert bei 0,68, der mittlere absolute Fehler bei 20,37 und die Wurzel der mittleren quadratischen Abweichung bei 0,53 liegt (vgl. Abbildung 22). Ähnlich gute Ergebnisse liefert das AIR-Modell (Sörensenindex: 0,84; Korrelationskoeffizient nach Pearson: 0,77; adjustierter R^2-Wert: 0,60; mittlerer absoluter Fehler: 0,57; Wurzel der mittleren quadratischen Abweichung: 22,53).

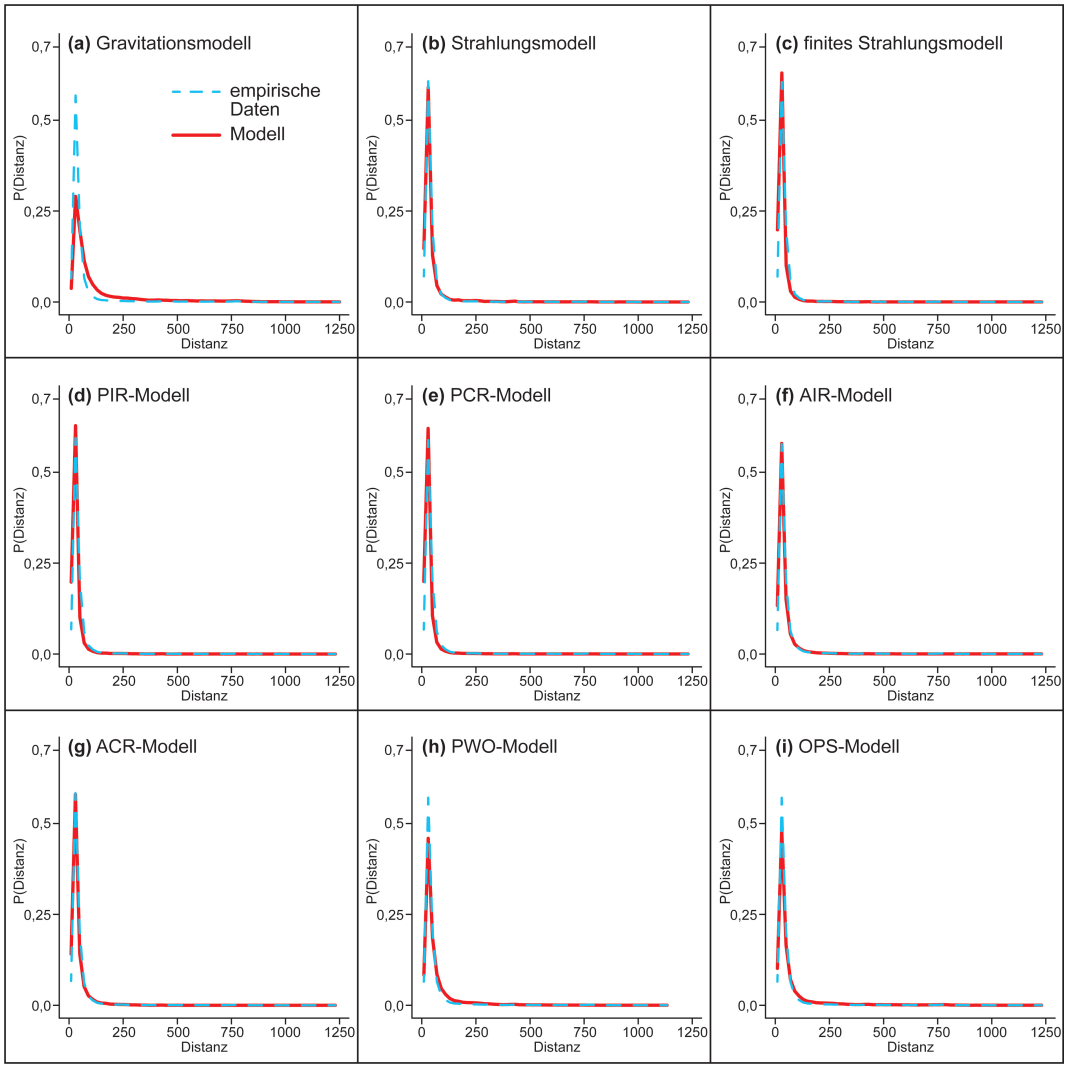


Abbildung 21: P(Distanz) im Verhältnis zur Entfernung (Pendlerbewegung auf Gemeindeverbandsebene). Quelle: Eigene Darstellung auf Grundlage von eigenen Berechnungen auf Basis von Bundesagentur für Arbeit (2017).

Die Anpassungen für das PIR- und das PCR-Modell sind etwas schwächer (PIR-Modell: Sörensenindex: 0,79; Korrelationskoeffizient nach Pearson: 0,70; adjustiert R^2-Wert: 0,48; mittlerer absoluter Fehler: 0,73; Wurzel der mittleren quadratischen Abweichung: 25,50 / PCR-Modell: Sörensenindex: 0,77; Korrelationskoeffizient nach Pearson: 0,56; adjustierter R^2-Wert: 0,31; mittlerer absoluter Fehler: 0,81; Wurzel der mittleren quadratischen Abweichung: 29,54).

Abbildung 20 (h) zeigt für das PWO-Modell, dass die Linie y=x nicht zwischen dem 9. und 91. Perzentil in der jeweiligen Klasse liegt und somit die Vorhersage nicht gut ist. Darüber hinaus werden die Distanzen im Bereich von ca. 30 Kilometern leicht unterschätzt (vgl. Abbildung 21 (h)). Die Ergebnisse des nicht parametrisierten Modells (vgl. Abbildung 22) veranschaulichen ferner, dass in Bezug auf den Sörensenindex, den Korrelationskoeffizienten nach Pearson und den adjustierten R^2-Wert deutliche Verbesserungen im Verhältnis zum klassischen Strahlungsmodell vorliegen (Sörensenindex: 0,75; Korrelationskoeffizient nach Pearson: 0,80; adjustierter R^2-Wert: 0,64; mittlerer absoluter Fehler: 0,91; Wurzel der mittleren quadratischen Abweichung: 25,68).

In Abbildung 20 (i) ist bezüglich des OPS-Modell zu erkennen, dass Pendlerbewegungen in einer Größenordnung von ca. 10.000 Personen besonders gute Anpassungen liefern. Zudem exisiteren leichte Unterschätzungen im Bereich von ca. 50 bis 80 Kilometern (vgl. Abbildung 21). Der Sörensenindex beträgt 0,77, der Korrelationskoeffizient nach Pearson 0,84, der adjustierte R^2-Wert 0,71, der mittlere absolute Fehler 0,82 und die Wurzel der mittleren quadratischen Abweichung 21,14 (vgl. Abbildung 22).

Aus der Betrachtung der SHAP-Werte für den XGBoost-Regressor vom Grundmodell (vgl. Abbildung 25 (b)) ergibt sich, dass die Distanz zwischen Wohn- und Arbeitsort das wichtigste Merkmal ist. Einen besonders hohen Einfluss auf die Vorhersage der Modelle haben insbesondere kleine Distanzen, während weite Entfernungen die Prognoseleistung verringern. Zudem zeigt sich, dass sich große Einwohnerzahlen (sowohl als Arbeits- also auch als Wohnort) positiv auf die Prognose der Mobilitätsflüsse von Pendlern auswirken. Standorte mit wenig Einwohnern haben hingegen einen negativen Einfluss auf die Vorhersage.


1,0
0,8
0,6
0,4
0,2
0
Sörensenindex
Korrelationskoeffizient nach Pearson
adjustierter R²-Wert
Dänemark
Finnland
Norwegen
Schweden
Deutschland (Landkreise)
Deutschland (Gemeinde-verbände)
England & Wales
Frankreich
Irland
Österreich
Schweiz
Kroatien
Italien
Spanien
Nordeuropa
West- und Zentraleuropa
Osteuropa
Südeuropa

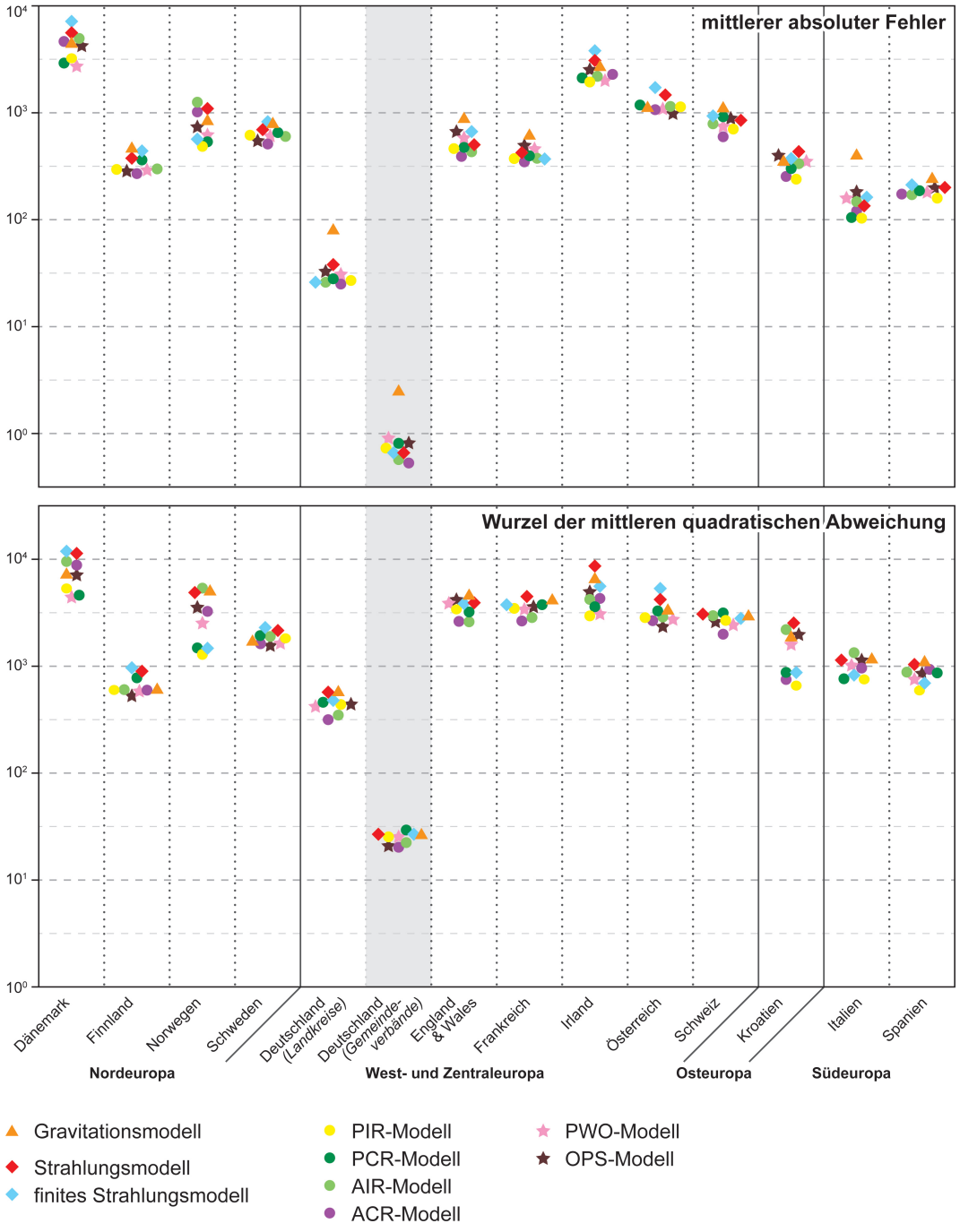


Abbildung 22: Ergebnisse der Evaluationsindikatoren für verschiedene Länder in Europa. Quelle: Eigene Darstellung auf Grundlage von eigenen Berechnungen auf Basis von Bundesagentur für Arbeit (2017), Bundesamt für Statistik (2014), Central Statistics Office (2011), Croatian Bureau of Statistics (2011),Danmarks Statistik (2011), Institut national de la statistique et des études économiques (2011), Instituto Nacionale de Estadística (2011), Instituto Nazionale di Statistica (2011), Office for National Statistic (2011), Statistik Austria (2011), Statistics Finland (2011), Statistics Norway (2011), Statistics Sweden (2011).

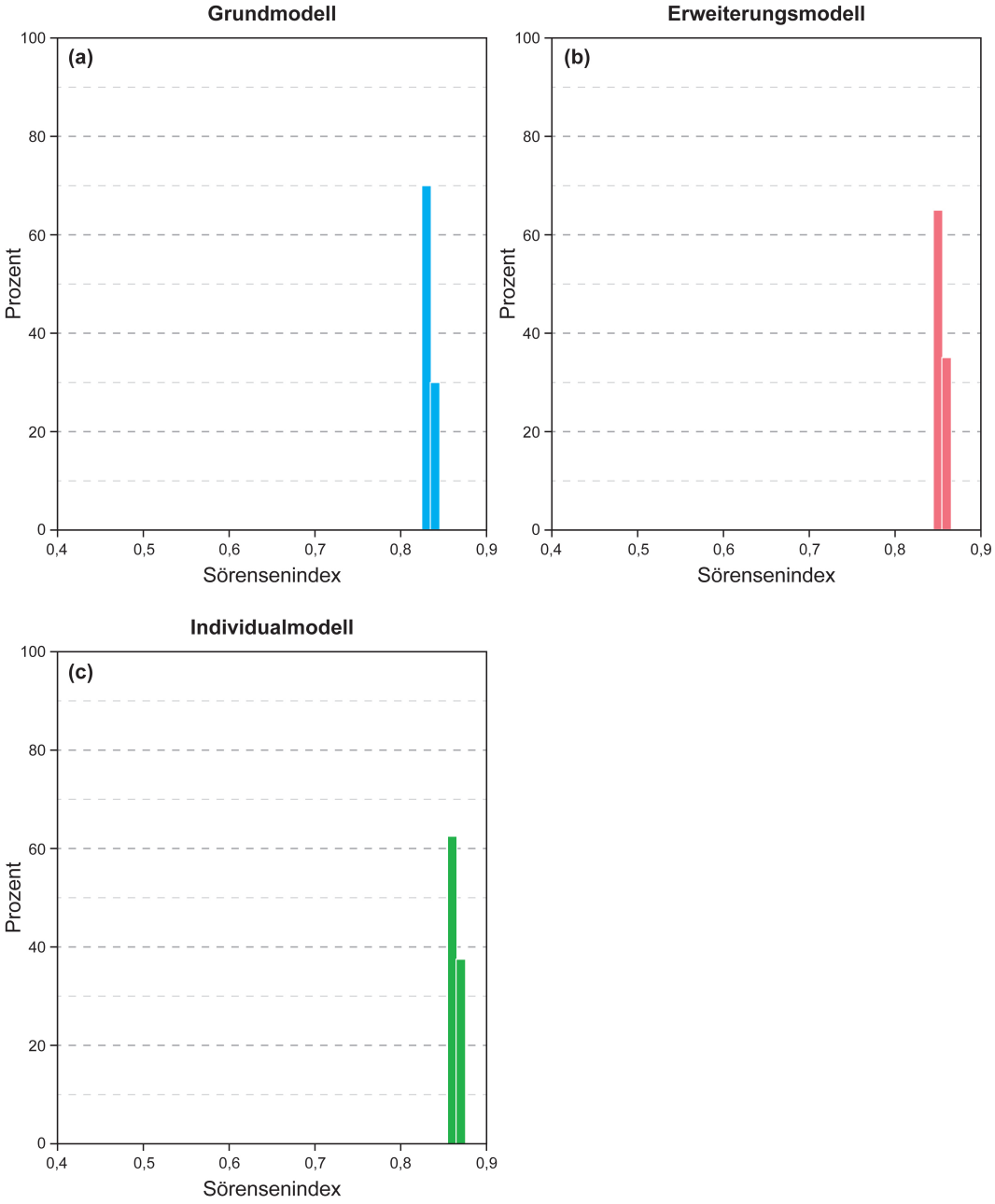
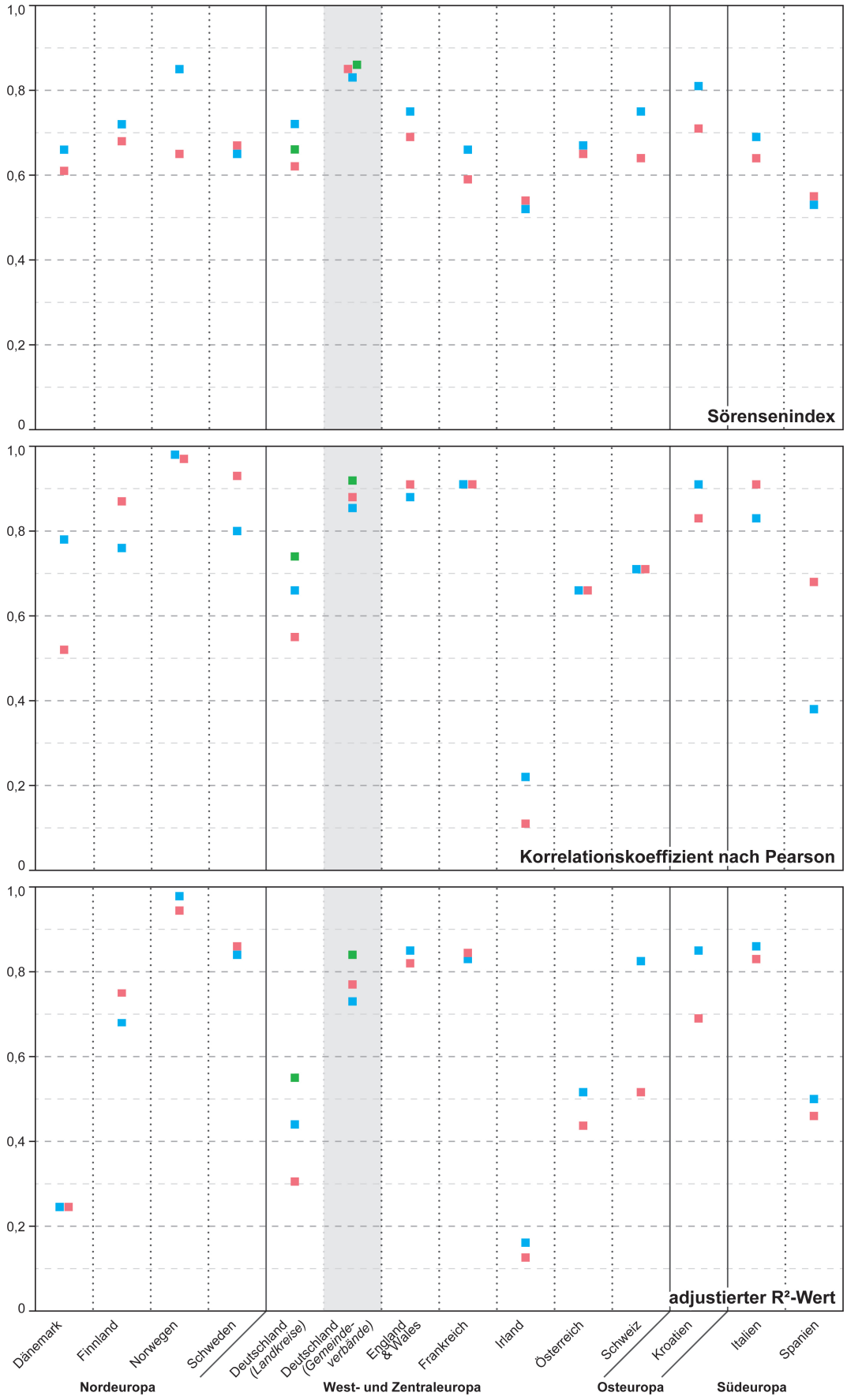


Abbildung 23: Ergebnisse der Simulation für Deutschland (Gemeindeverbandsebenen). Quelle: Eigene Darstellung auf Grundlage von eigenen Berechnungen auf Basis von Bundesagentur für Arbeit (2017).

Der Sörensenindex liegt in 70 % der Simulationen bei 0,83 (vgl. Abbildung 23 (a)), der Korrelationskoeffizient nach Pearson bei 0,85, der adjustierte R^2-Wert bei 0,73, der mittlere absolute Fehler liegt bei 0,59 und die Wurzel der mittleren quadratischen Abweichung bei 17,4 (vgl. Abbildung 24).


1,0
0,8
0,6
0,4
0,2
0
Sörensenindex
Korrelationskoeffizient nach Pearson
adjustierter R²-Wert
Dänemark
Finnland
Norwegen
Schweden
Deutschland (Landkreise)
Deutschland (Gemeinde-verbände)
England & Wales
Frankreich
Irland
Österreich
Schweiz
Kroatien
Italien
Spanien
Nordeuropa
West- und Zentraleuropa
Osteuropa
Südeuropa

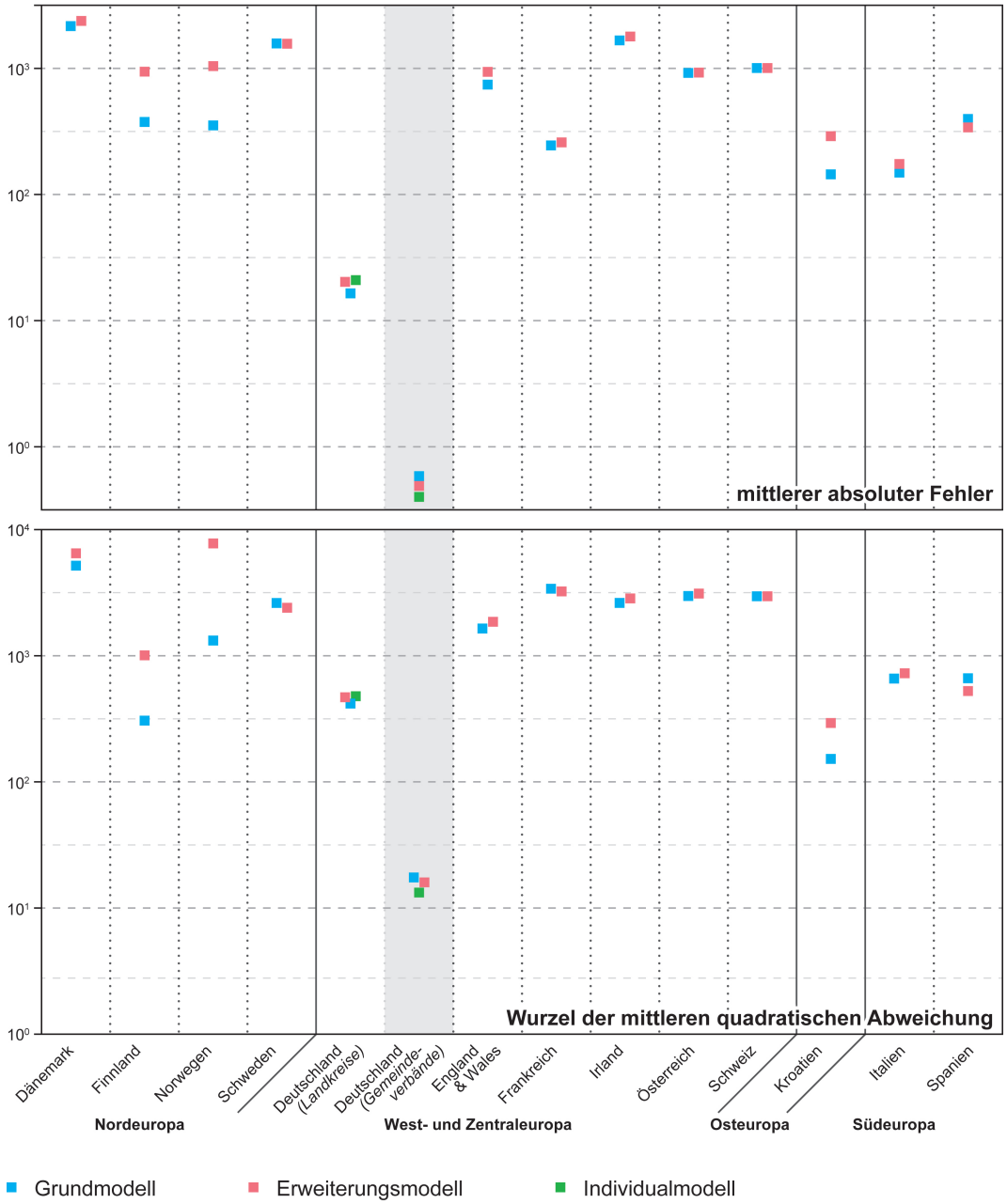


Abbildung 24: Ergebnisse der Evaluationsindikatoren mittels XGBoost-Algorithmus für verschiedene Länder in Europa. Quelle: Eigene Darstellung auf Grundlage von eigenen Berechnungen auf Basis von Bundesagentur für Arbeit (2017), Bundesamt für Statistik (2014), Central Statistics Office (2011), Croatian Bureau of Statistics (2011),Danmarks Statistik (2011), Institut national de la statistique et des études économiques (2011), Instituto Nacionale de Estadística (2011), Instituto Nazionale di Statistica (2011), Office for National Statistic (2011), Statistik Austria (2011), Statistics Finland (2011), Statistics Norway (2011), Statistics Sweden (2011).

Im Erweiterungsmodell (vgl. Abbildung 25 (a)) nimmt s als Repräsentant der Einwohnerzahlen zwischen Wohn- und Arbeitsort das größte Gewicht für die Prognose von Mobilitätsflüssen an. Es ist hierbei festzustellen, dass die Fälle, in denen zwischen Wohn- und Arbeitsort eine Vielzahl an einwohnerstarken Standorten liegen, einen negativen Einfluss auf die Prognose haben. Ferner ist aus Abbildung 25 (a) ableitbar, dass insbesondere hohe Einwohnerzahlen an Wohn- und Arbeitsort zu einer Verbesserung der vorhergesagten Mobilitätsflüsse beitragen. Die Distanz hat auch in diesem Modell den geringsten Einfluss. Allerdings zeigt sich ähnlich wie beim Grundmodell, dass weite Entfernungen zwischen Wohn- und Arbeitsort einen negativen Einfluss auf die Prognose haben. Der Sörensenindex beträgt in 65 % der Simulationen 0,85 (vgl. Abbildung 23 (b)), der Korrelationskoeffizient nach Pearson 0,88, der adjustierte R^2-Werte 0,77, der mittlere absolute Fehler 0,49 und die Wurzel der mittleren quadratischen Abweichung 16 (vgl. Abbildung 24).

Im Individualmodell (vgl. Abbildung 26) werden neben der Distanz, den Einwohnerzahlen am Wohn- und Arbeitsort und im Zwischenraum individuelle Motive für das Pendeln integriert. Ähnlich wie beim Erweiterungsmodell hat auch im Individualmodell s den größten Einfluss auf die Vorhersage. Hohe Einwohnerzahlen zwischen Wohn- und Arbeitsort haben einen negativen Einfluss auf die Prognose. Für die Gewerbesteuereinnahmen am Arbeitsort kann keine abschließende Bewertung vorgenommen werden. Außerdem hat ein hoher Anteil an Naherholungsflächen am Arbeitsort eine positive Bedeutung bei der Prognose. Abbildung 26 ist ebenfalls zu entnehmen, dass hohe Einwohnerdichten am Arbeitsort einen negativen Einfluss auf die Prognose haben. Ähnliche Beobachtungen sind auch für die Entfernung zur nächsten Autobahnauffahrt festzustellen, mit zunehmender Distanz verschlechtert sich die Vorhersage. Eine gut ausgebaute Breitbandversorgung am Arbeitsort beeinflusst die Berechnung der Pendlerströme hingegen positiv. Für alle weiteren Merkmale sind keine eindeutigen Aussagen zu treffen. Der Sörensenindex nimmt in 62,5 % der Simulationen einen Wert von 0,86 an (vgl. Abbildung 23 (c)). Zudem liegt der Korrelationskoeffizient nach Pearson bei 0,91, der adjustierte R^2-Wert bei 0,84, der mittlere absolute Fehler bei 0,40 und die Wurzel der mittleren quadratischen Abweichung bei 13,2.

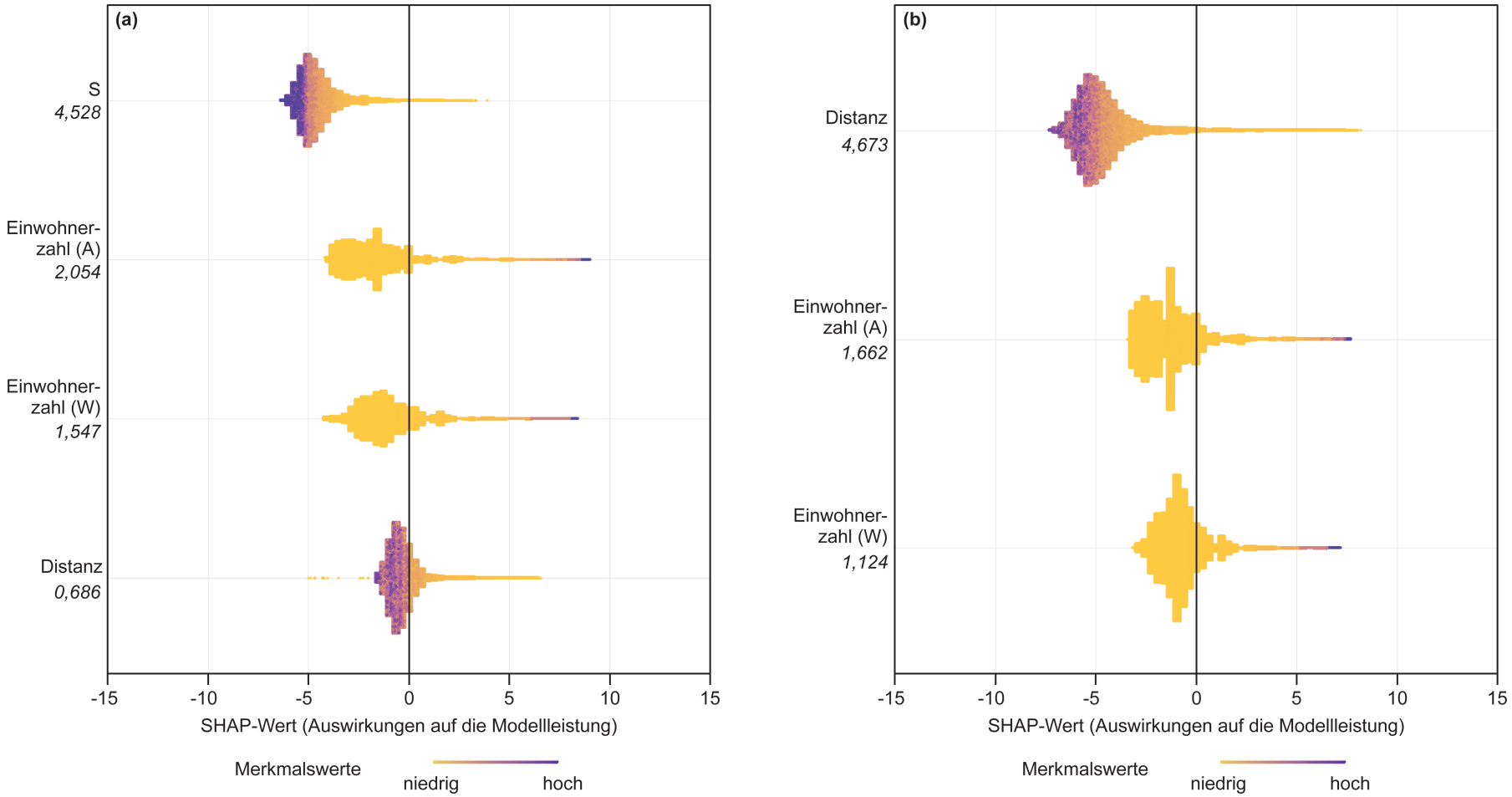


Abbildung 25: SHAP-Werte für den XGBoost-Regressor bei Pendlerbewegungen auf Gemeindeverbandsebenen (Grund- und Erweiterungsmodell). Quelle: Eigene Darstellung auf Grundlage von eigenen Berechnungen auf Basis von Bundesagentur für Arbeit (2017) und INKAR (2022).

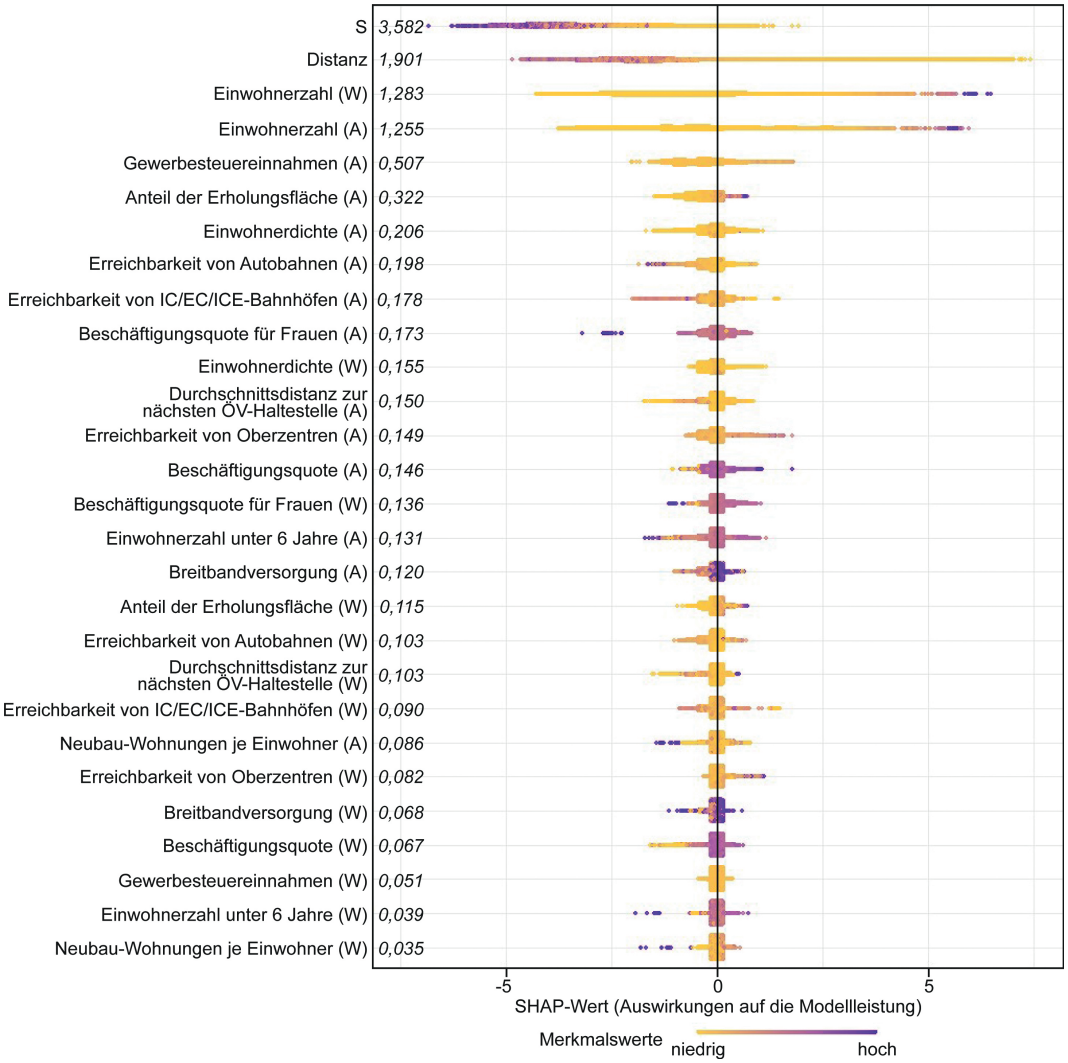


Abbildung 26: SHAP-Werte für den XGBoost-Regressor bei Pendlerbewegungen auf Gemeindeebenen (Individualmodell). Quelle: Eigene Darstellung auf Grundlage von eigenen Berechnungen auf Basis von Bundesagentur für Arbeit (2017) und INKAR (2022).

Deutschland (Kreisebene)

Bei der Betrachtung der Pendlerbewegungen auf Kreisebene in Deutschland fällt auf, dass die Pendleranzahl zwischen 10 und 76.110 Personen schwankt, wobei die Verbindung mit den meisten Interaktionen zwischen München Stadt und dem Landkreis München vorliegt. Im Mittel fahren etwa 327 Personen zum Arbeiten über die Kreisgrenze. Der Median liegt bei 27 Interaktionen und das erste bzw. dritte Quantil beträgt 27 und 77 Personen (vgl. Bundesagentur für Arbeit 2017).

Vergleicht man die Schätzungen des Gravitationsmodells mit den empirischen Daten (vgl. Abbildung 27 (a)), dann wird deutlich, dass insbesondere sehr kleine Pendlerbewegungen (ca. 10 Pendler) und sehr große Pendlerbewegungen (ca. 10.000 Pendler) gute Schätzungen liefern. Die Betrachtung der Pendlerwahrscheinlichkeit in Abhängigkeit der Distanz veranschaulicht, dass das Gravitationsmodell bei kleinen Distanzen zur Unterschätzung der Mobilitätsflüsse neigt (vgl. Abbildung 28(a)). Es folgt eine starke Überschätzung mittels des Modells. In Abbildung 22 ist ablesbar, dass der Sörensenindex bei 0,19 liegt, während für den Korrelationskoeffizient nach Pearson und den adjustierten R^2-Wert 0,45 und 0,20 berechnet wurden. Der mittlere absolute Fehler beträgt 81 und die Wurzel der mittleren quadratischen Abweichung 593.

Für das Strahlungsmodell ist ersichtlich, dass große Pendlerbewegungen (mehr als 10.000 Pendler) gute Schätzungen liefern (vgl. Abbildung 27 (b)). Für sehr kleine Distanzen (10 bis 30 Kilometer) liegt eine Unterschätzung der Vorhersage vor, während die Mobilitätsflüsse zwischen 90 und 150 Kilometern leicht überschätzt werden (vgl. Abbildung 28 (b)). Der Sörensenindex, der Korrelationskoeffizient nach Pearson und der adjustierte R^2-Wert liefern Werte von 0,27, 0,17 und 0,41. Zudem beträgt der mittlere absolute Fehler 38,0 und die Wurzel der mittleren quadratischen Abweichung 578,7 (vgl. Abbildung 22).

Der Vergleich zwischen den empirischen Daten und den vorhergesagten Prognosen mittels finitem Strahlungsmodell (vgl. Abbildung 27 (c)) veranschaulicht sehr gute Anpassungen im Bereich zwischen 1.000 und 10.000 Pendlern. Außerdem wird deutlich, dass kleine Distanzen (30 bis 50 Kilometer) leicht überschätzt werden (vgl. Abbildung 28 (c)). Es lässt sich ein Sörensenindex von 0,52 messen. Der Korrelationskoeffizient nach Pearson beträgt 0,71 und der adjustierte R^2-Wert 0,50 (vgl. Abbildung 22). Ferner liegen der mittlere absolute Fehler und die Wurzel der mittleren quadratischen Abweichung bei 26 und 485.

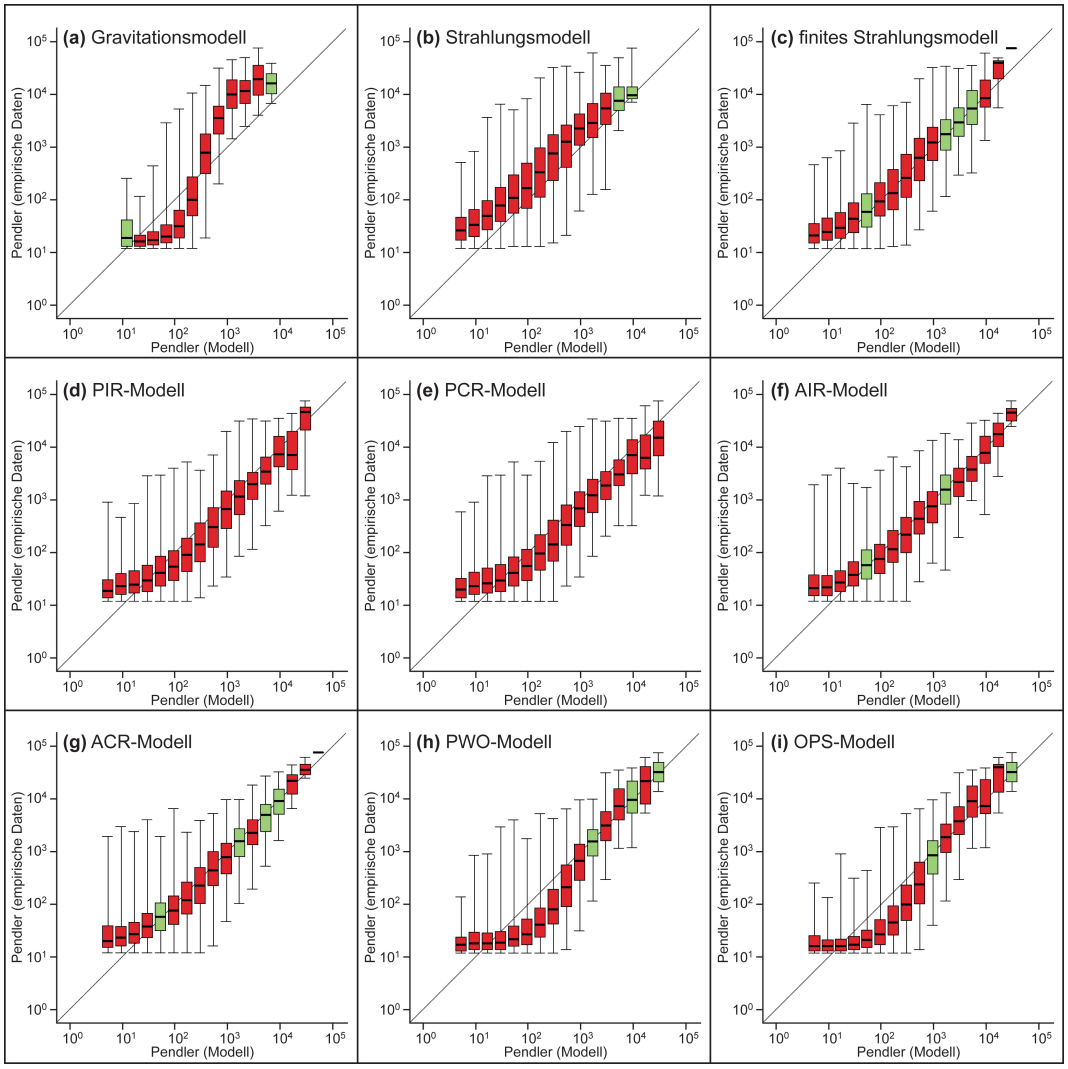


Abbildung 27: Abweichung zwischen den empirischen Daten und den Modellvorhersagen bezüglich der Pendlerbewegungen auf Kreisebene. Quelle: Eigene Darstellung auf Grundlage von eigenen Berechnungen auf Basis von Bundesagentur für Arbeit (2017).

Der Vergleich zwischen den empirischen Daten und der Prognose mithilfe der verallgemeinerten Strahlungsmodelle zeigen zudem, dass vor allem das ACR-Modell besonders geeignet ist, unterschiedlich große Pendelbewegungen zu schätzen. So zeigt Abbildung 27 (g), dass für Mobilitätsströme im Bereich von ca. 80,1000 und zwischen 5.600 und 10.000 Interaktionen die Vorhersage gute Anpassungen liefert. Abbildung 22 bestätigt diese Beobachtung. Für das ACR-Modell weist der Sörensenindex mit 0,61 den höchsten Wert auf. Ähnliche Beobachtungen sind auch für den Korrelationskoeffizienten nach Pearson (0,87) und den adjustierten R^2-Wert (0,76) zu erkennen. Des Weiteren liefern die Wurzel der mittleren quadratischen Abweichung (320,0) und der mittlere absolute Fehler (24,7) die geringsten Werte und weisen insgesamt darauf hin, dass auf Kreisebene dieses Modell die besten Prognosen liefert. Auch das AIR-

Modell erbringt vergleichbare Ergebnisse (Sörensenindex: 0,61; Korrelationskoeffizient nach Pearson: 0,84; adjustierter R^2-Wert: 0,71; mittlerer absoluter Fehler: 26; Wurzel der mittleren quadratischen Abweichung: 353). Abbildung 28 (d) bis (g) veranschaulicht außerdem, dass für das PCR- und das PIR-Modell vor allem die Distanzen zwischen 30 und 50 Kilometern überschätzt werden. Zwischen 110 und 150 Kilometer liefern das AIR- und das ACR-Modell jedoch leichte Überschätzungen. Des Weiteren ist in Abbildung 22 zu erkennen, dass die Evaluationsindikatoren auf deutlich schlechtere Ergebnisse hindeuten. Der Sörensenindex für das PCR-Modell liegt bei 0,58, während er beim PIR-Modell 0,59 aufweist. Der Korrelationskoeffizient nach Pearson nimmt für das PCR-Modell und das PIR-Modell Werte von 0,68 und 0,71 an. Zudem weisen auch der mittlere absolute Fehler (PCR-Modell: 28,0; PIR-Modell: 27,0) und die Wurzel der mittleren quadratischen Abweichung (PCR-Modell: 466,0; PIR-Modell: 444) auf eine schlechtere Anpassung zwischen den Modellen und den empirischen Daten hin.

Für das PWO-Modell wird sichtbar, dass vor allem größere Pendlerströme (ab ca. 1.000 Pendler) besonders gute Anpassungen liefern (vgl. Abbildung 27 (h)). Abbildung 28 (h) veranschaulicht leichte Unterschätzungen im Bereich von 80 bis 100 Kilometern. Zwischen 110 und 300 Kilometern wird die Wahrscheinlichkeit zu pendeln überschätzt. Der Sörensenindex für dieses Modell liegt bei 0,55, während der Korrelationskoeffizient nach Pearson einen Wert von 0,78 und der adjustierte R^2-Wert von 0,61 annimmt (vgl. Abbildung 22). Der mittlere absolute Fehler liegt bei 31 und die Wurzel der mittleren quadratischen Abweichung bei 428.

Der Vergleich der empirischen Daten mit der Prognose der Mobilitätsflüsse mittels OPS-Modell in Abbildung 27 (i) zeigt, dass Pendlerflüsse mit ca. 1.000 Personen und mit mehr als 10.000 Personen besonders gute Ergebnisse liefern. Bezüglich der Distanz sind bei ca. 50 Kilometern deutliche Unterschätzungen festzustellen. Ab etwa 90 Kilometern werden die Mobilitätsströme mit dem Modell überschätzt (vgl. Abbildung 28 (i)). Der Sörensenindex, der Korrelationskoeffizent nach Pearson sowie der adjustierte R^2-Wert liegen bei 0,52, 0,78 und 0,61. Darüber hinaus ergeben die Berechungen des mittleren absoluten Fehlers und der Wurzel aus der mittleren quadratischen Abweichung Werte von 33 und 450 (vgl. Abbildung 22).

In Abbildung 30 (b) wird zunächst auf Kreisebene das Grundmodell mit Distanz und der jeweiligen Einwohnerzahl präsentiert. Es wird deutlich, dass

der Einwohnerzahl am Wohnort eine hohe Bedeutung zukommt. Insbesondere Orte mit vielen Einwohnern führen zu einer Verbesserung des Modells. Der zweitwichtigste Indikator im Grundmodell ist die Distanz. Auch hier kann festgestellt werden, dass kleine Distanzen einen signifikanten positiven Einfluss auf die Prognose haben, während große Entfernungen sich negativ auswirken. Zusätzlich wird deutlich, dass die Einwohnerzahl am Arbeitsort ebenfalls einen positiven Einfluss hat. Der Sörensenindex beträgt in 70 % der Simulationen 0,72 (vgl. Abbildung 29 (a)). Zudem liegt der Korrelationskoeffizient nach Pearson bei 0,66, der adjustierte R^2-Wert bei 0,44, der mittlere absolute Fehler bei 18 und die Wurzel der mittleren quadratischen Abweichung bei 559 (vgl. Abbildung 24).

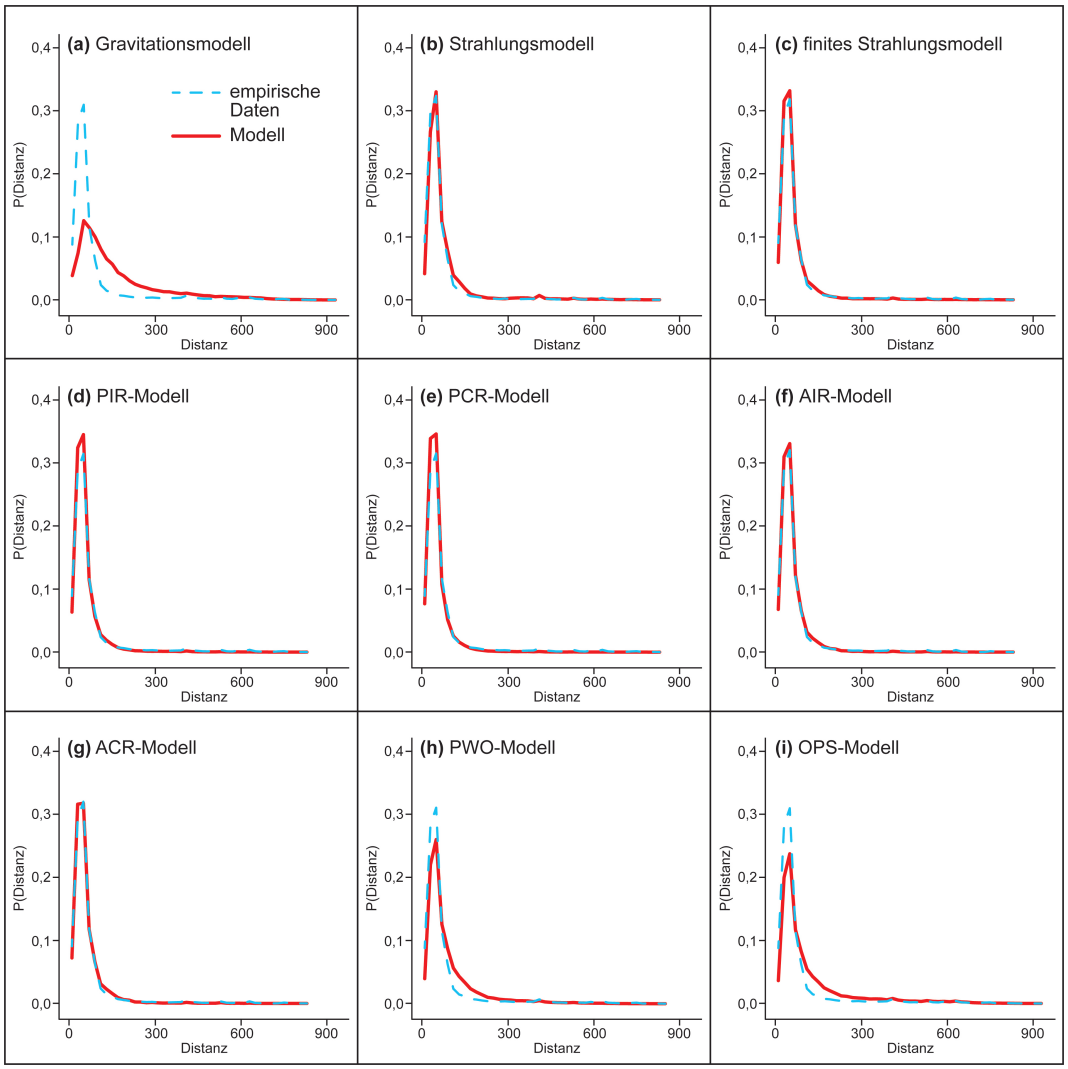


Abbildung 28: P(Distanz) im Verhältnis zur Entfernung (Pendlerbewegung auf Kreisebene). Quelle: Eigene Darstellung auf Grundlage von eigenen Berechnungen auf Basis von Bundesagentur für Arbeit (2017).

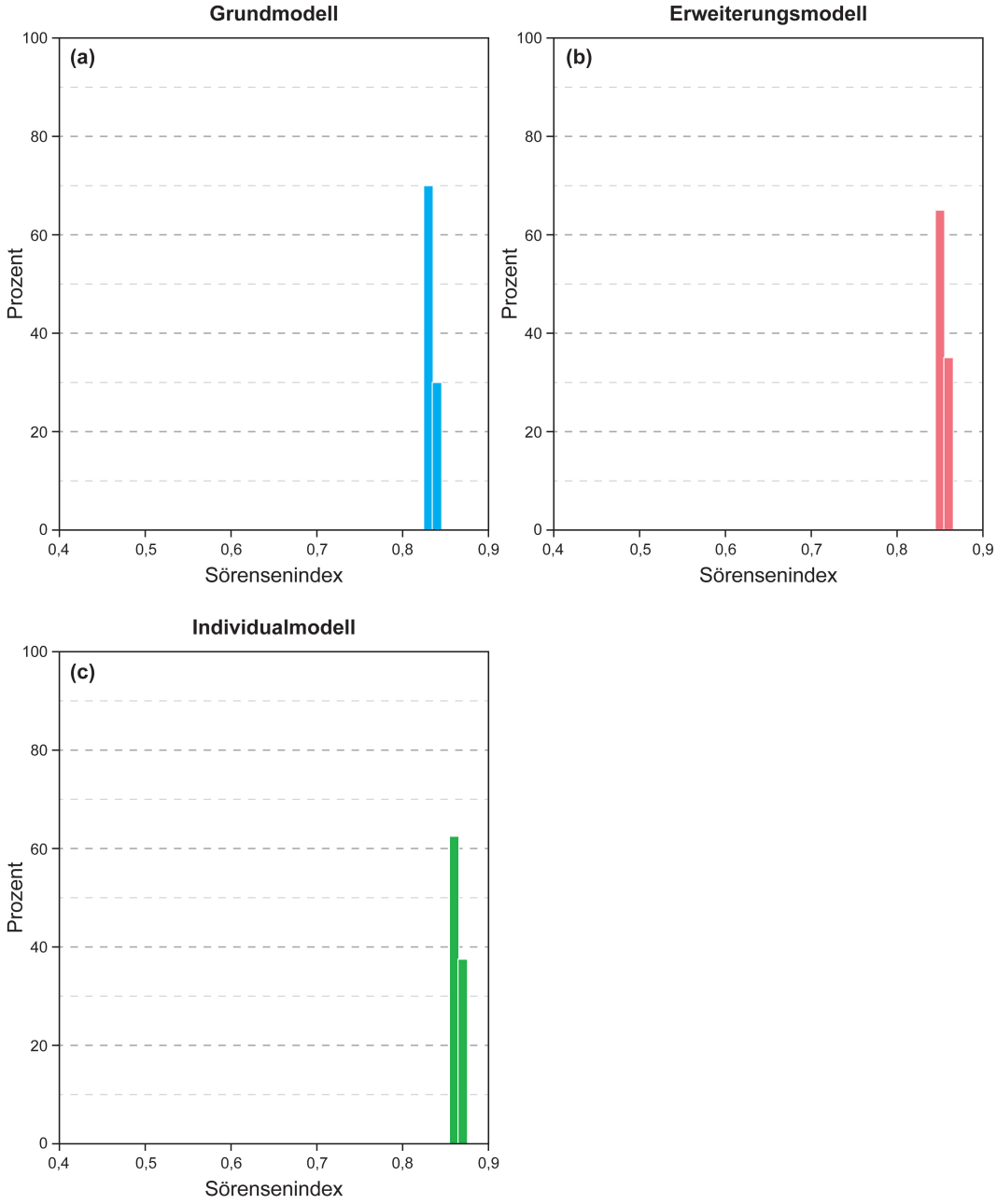


Abbildung 29: Ergebnisse der Simulation Deutschland (Kreisebene). Quelle: Eigene Darstellung auf Grundlage von eigenen Berechnungen auf Basis von Bundesagentur für Arbeit (2017).

Des Weiteren wird in Abbildung 30 (a) die Erweiterung des Grundmodells um s (Erweiterungsmodell) präsentiert. Die Einwohnerzahl zwischen Wohn- und Arbeitsort (s) ist für dieses Modell das wichtigste Merkmal. Hohe Einwohnerzahlen zwischen Ziel- und Quellregion führen zu einer Verschlechterung der Prognoseleistung. Außerdem kann festgestellt werden, dass die räumliche Nähe (anders als auf Gemeindeverbandsebene) ein deutlich größeres Gewicht annimmt, wobei auf Kreisebene besonders große Entfernungen einen positi-

ven Einfluss auf die Vorhersage haben. Die Einwohnerzahlen am Wohn- und am Arbeitsort verhalten sich dagegen ähnlich wie im Erweiterungsmodell auf Gemeindeverbandsebene: Große Städte mit einer hohen Einwohnerzahl wirken sich positiv auf die Schätzung der Pendlerflüsse aus. Der Sörensenindex liegt in 70 % der Simulationen bei 0,62 (vgl. Abbildung 29 (b)), der Korrelationskoeffizient nach Pearson bei 0,55, der adjustierte R^2-Wert bei 30,5, der mittlere absolute Fehler bei 22, 72 und die Wurzel der mittleren quadratischen Abweichung 610 (vgl. Abbildung 24).

Während im Grund- und Erweiterungsmodell auf Kreisebene ähnliche Beobachtungen gemacht wurden wie auf der Gemeindeverbandsebene, ist für das Individualmodell ein deutlicher Unterschied zu erkennen (vgl. Abbildung 31). Auf der Kreisebene ist festzustellen, dass eine hohe Beschäftigungsquote von Frauen am Wohnort die Berechnung der Pendlerflüsse am stärksten positiv beeinflusst, während auf Gemeindeverbandsebene mit einem niedrigen Anteil an beschäftigten Frauen ein negativer Einfluss zu verzeichnen ist. Das zweite wichtige individuelle Merkmal ist die Erreichbarkeit von Oberzentren. Eine geringe Entfernung zum Oberzentrum weist eine positive Bedeutung bei der Prognose auf. Kreise, die durch hohe durchschnittliche Fahrzeiten zum nächsten Oberzentrum gekennzeichnet sind, beeinflussen die Vorhersage hingegen negativ. Ferner zeigt sich, dass Kreise mit einem hohen Anteil an Erholungsfläche (Wohnort) einen negativen Einfluss auf die Prognose haben. Diese Beobachtung ist einhergehend mit dem fünftwichtigsten Indikator (Einwohnerdichte am Wohnort). Hohe Einwohnerdichten verbessern die Prognoseleistung. Es ist auch zu erkennen, dass Regionen oder Kreise, die durch eine Vielzahl weiterer Standorte mit hohen Einwohnerzahlen (s) im Umfeld gekennzeichnet sind, die Prognose negativ beeinflussen. Außerdem wirkt sich eine hohe Breitbandversorgung positiv auf die Vorhersage aus, während hohe Beschäftigungsquoten am Wohnort eine negative Auswirkung aufzeigen. Hohe Beschäftigungsquoten am Arbeitsort wirken sich hingegen positiv aus. Darüber hinaus ist zu erkennen, dass in diesem Modell kleine Distanzen zwischen Wohn- und Arbeitsort vorteilhafter für die Prognose sind. Ähnliche Beobachtungen sind für Wohnorte festzustellen, an denen der Anteil von Neubauwohnungen pro Einwohner besonders hoch ist. Ferner haben geringe Entfernungen zur nächsten Autobahnauffahrt einen signifikant positiven Einfluss. Hohe Gewerbesteuereinnahmen am Wohn- und Arbeitsort sowie einwohnerstarke Standorte beeinflussen die Prognose positiv. Für alle weiteren Indikatoren konnte keine klare Aussage getroffen werden.

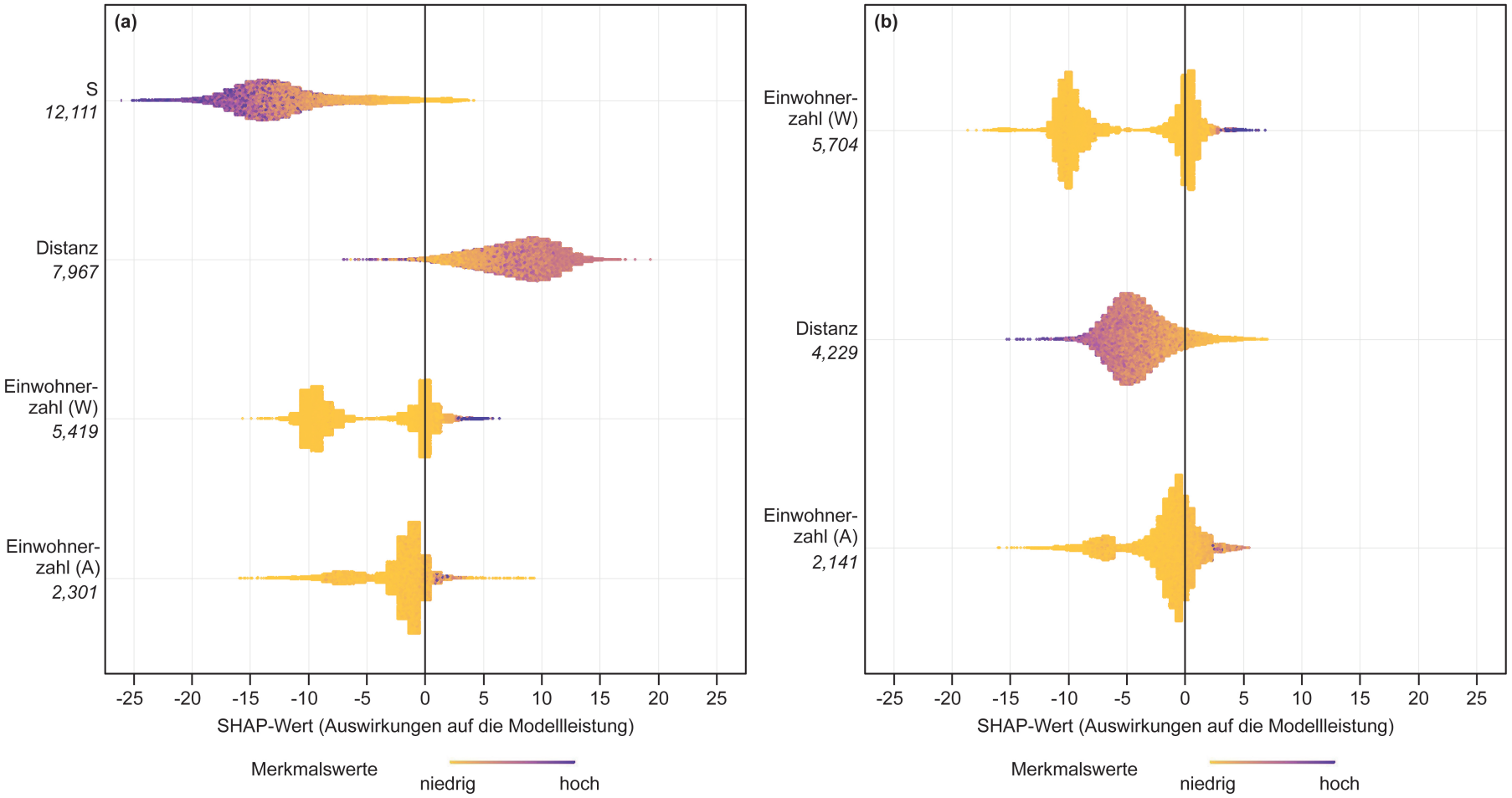


Abbildung 30: SHAP-Werte für den XGBoost-Regressor bei Pendlerbewegungen auf Kreisebenen (Grund- und Erweiterungsmodell). Quelle: Eigene Darstellung auf Grundlage von eigenen Berechnungen auf Basis von Bundesagentur für Arbeit (2017).

Der Sörensenindex liegt in 0,68 % der Simulationen bei 0,66 (vgl. Abbildung 29 (c)). Daneben beträgt der Korrelationskoeffizient nach Pearson 0,74, der adjustierte R^2-Wert 0,55, der mittlere absolute Fehler 23,19 und die Wurzel der mittleren quadratischen Abweichung 495 (vgl. Abbildung 24) .

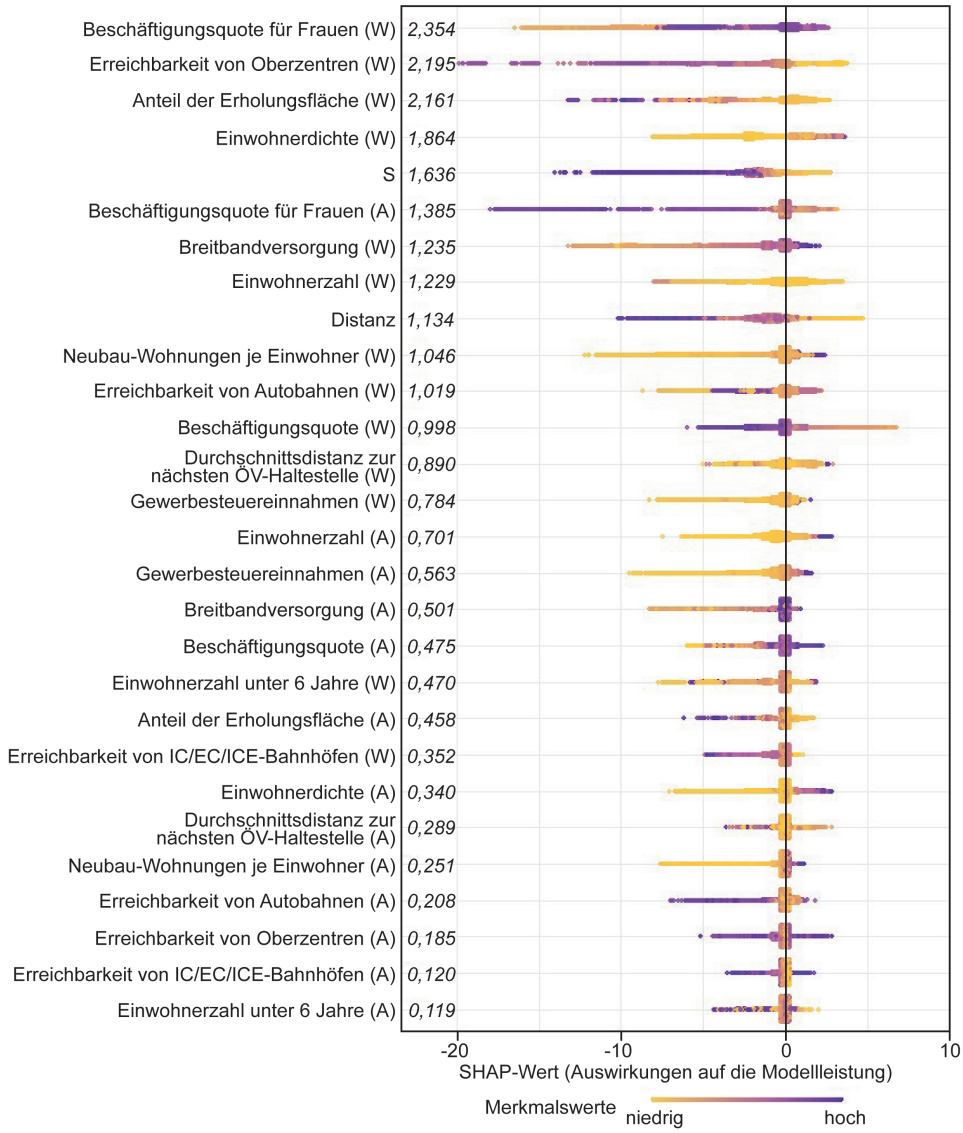


Abbildung 31: SHAP-Werte für den XGBoost-Regressor bei Pendlerbewegungen auf Kreisebenen (Individualmodell). Quelle: Eigene Darstellung auf Grundlage von eigenen Berechnungen auf Basis von Bundesagentur für Arbeit (2017) und INKAR (2022).

England und Wales

Betrachtet man die Mobilitätsströme von Arbeitskräften in England, so zeigt sich, dass sie, sofern sie auftreten, zwischen 1 (z. B. Darlington - Leicester) und 341.837 Personen (Inner London East - Inner London West) schwanken. Im Mittel pendeln 670 Personen täglich auf NUTS3-Ebene. Das erste Quantil liegt bei 10, das zweite Quantil bei 29 und das dritte Quantil bei 92 Interaktionen (Office for National Statistic 2011).

Für England und Wales ist in Abbildung 32 (a) für die Vorhersage mittels Gravitationsmodell erkennbar, dass für die Linie y=x nicht zwischen dem 9. und 91. Perzentil in der jeweiligen Klasse liegt und somit die Pendlerbewegungen nicht optimal geschätzt werden können. Des Weiteren wird in Abbildung 33 (a) deutlich, dass das Modell die empirischen Daten bis ca. 80 Kilometer unterschätzt. Im Anschluss folgt eine Überschätzung der Daten. Außerdem liegt der Sörensenindex bei 0,38, der Korrelationskoeffizient nach Pearson bei 0,67, der adjustierte R^2-Wert bei 0,42, der mittlere absolute Fehler bei 906 und die Wurzel der mittleren quadratischen Abweichung bei 4.725 (vgl. Abbildung 22).

Die Ergebnisse für das Strahlungsmodell zeigen, dass Pendlerbewegungen in einer Größenordnung von ca. 30, 1.800, 18.000 und 100.000 Personen besonders gute Anpassungen liefern (vgl. Abbildung 32 (b)). Zudem werden beim finiten Strahlungsmodell besonders Interaktionen von ca. 180.000 Personen sehr gut geschätzt (vgl. Abbildung 32 (c)). Des Weiteren weisen beide Modelle bezüglich der Distanz lediglich marginale Abweichungen auf (vgl. Abbildung 33 (b) und (c)). Darüber hinaus liefern die Evaluationsindikatoren folgende Werte (vgl. Abbildung 22): Strahlungsmodel: Sörensenindex: 0,61; Korrelationskoeffizient nach Pearson: 0,67; adjustierter R^2-Wert: 0,49; mittlerer absoluter Fehler: 502; Wurzel der mittleren quadratischen Abweichung: 3.919 / finites Strahlungsmodell: Sörensenindex: 0,61; Korrelationskoeffizient nach Pearson: 0,80; adjustierter R^2-Wert: 0,63; mittlerer absoluter Fehler: 669; Wurzel der mittleren quadratischen Abweichung: 3.847.

Beim Vergleich der verallgemeinerten Strahlungsmodelle wird deutlich, dass die Prognose der Mobilitätsflüsse für das ACR-Modell besonders gut sind. Abbildung 32 (g) veranschaulicht geringe Abweichungen zwischen Prognose und empirischen Daten bei Mobilitätsströmen in einer Größenordnung von ca. 5.700 bis 30.000 Personen sowie bei 55.000 Interaktionen. Die Abweichungen bezüglich der Distanz sind marginal (vgl. Abbildung 33 (g)). Diese Beobachtung wird auch durch die Evaluationskoeffizienten unterstrichen (vgl. Abbil-

dung 22). Der Sörensenindex liegt bei 0,70, der Korrelationskoeffizient nach Pearson bei 0,90 und der adjustierte R^2-Wert bei 0,81. Der mittlere absolute Fehler beträgt 391 und die Wurzel der mittleren quadratischen Abweichung 2.625. Die Ergebnisse für die anderen drei Modelle fallen etwas schlechter aus (AIR-Modell: Sörensenindex: 0,67; Korrelationskoeffizient nach Pearson: 0,89; adjustierter R^2-Wert: 0,79; mittlerer absoluter Fehler: 432; Wurzel der mittleren quadratischen Abweichung: 2.601 / PCR-Modell: Sörensenindex: 0,64; Korrelationskoeffizient nach Pearson: 0,81; adjustierter R^2-Wert: 0,66; mittlerer absoluter Fehler: 474; Wurzel der mittleren quadratischen Abweichung: 3.215 / PIR-Modell: Sörensenindex: 0,65; Korrelationskoeffizient nach Pearson: 0,78; adjustierter R^2-Wert: 0,62; mittlerer absoluter Fehler: 462; Wurzel der mittleren quadratischen Abweichung: 3.401).

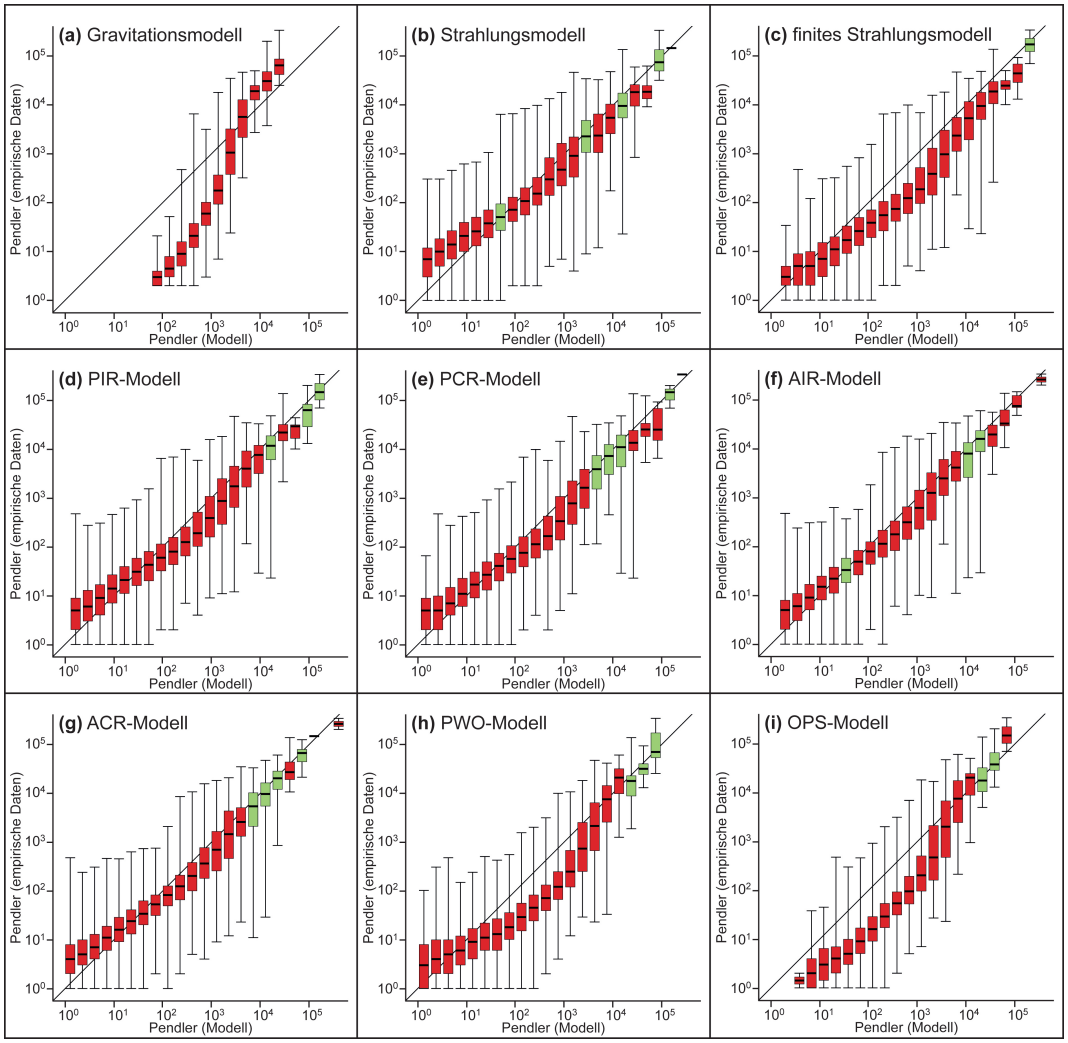


Abbildung 32: Abweichung zwischen den empirischen Daten und den Modellvorhersagen bezüglich der Pendlerbewegungen für England und Wales. Quelle: Eigene Darstellung auf Grundlage von eigenen Berechnungen auf der Basis von Office for National Statistic (2011).

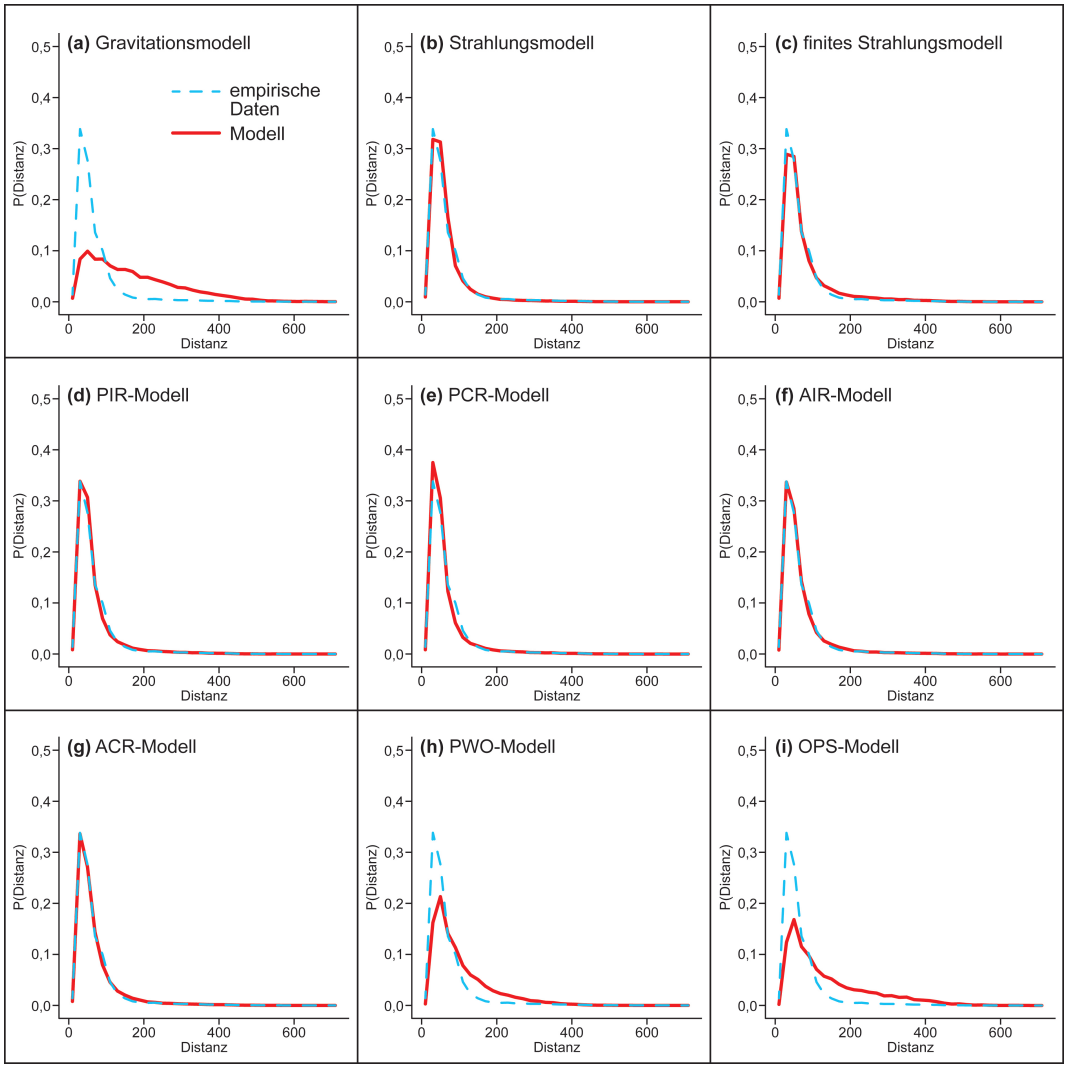


Abbildung 33: P(Distanz) im Verhältnis zur Entfernung (Pendlerbewegung in England und Wales). Quelle: Eigene Darstellung auf Grundlage von eigenen Berechnungen auf Basis von Office for National Statistic (2011).

Für das PWO-Modell wird sichtbar, dass vor allem größere Pendlerströme (ab ca. 30.000 Pendler) besonders gute Anpassungen liefern (vgl. Abbildung 32 (h)). Ferner ist aus Abbildung 33 (h) zu entnehmen, dass Pendlerbewegungen in den ersten ca. 80 Kilometern unterschätzt werden. Es folgt eine Überschätzung bis ca. 380 Kilometer. Der Sörensenindex für dieses Modell liegt bei 0,55, während der Korrelationskoeffizient nach Pearson einen Wert von 0,75 und der adjustierte R^2-Wert von 0,56 annimmt. Der mittlere absolute Fehler des PWO-Modells beträgt 588 und die Wurzel der mittleren quadratischen Abweichung 3.921 (vgl. Abbildung 22).

Darüber hinaus zeigt der Vergleich der empirischen Daten mit der Prognose der Mobilitätsflüsse mittels OPS-Modell in Abbildung 32 (i), dass Pendler-

flüsse mit ca. 30.000 bis 56.000 Personen besonders gute Ergebnisse liefern. Bezüglich der Distanz sind bis zu 100 Kilometern deutliche Unterschätzungen festzustellen. Ab etwa 110 bis ca. 500 Kilometern werden die Mobilitätsströme mit dem Modell überschätzt (vgl. Abbildung 33 (i)). Der Sörensenindex (vgl. Abbildung 22), der Korrelationskoeffizient nach Pearson sowie der adjustierte R^2-Wert liegen bei 0,49, 0,72 und 0,52. Darüber hinaus ergeben die Berechnungen des mittleren absoluten Fehlers und der Wurzel der mittleren quadratischen Abweichung Werte von 674 und 4.240.

Für das Grundmodell des XGBoost-Algorithmus wird in Abbildung 35 (b) deutlich, dass mit zunehmender Entfernung zwischen Arbeits- und Wohnort die Vorhersageleistungen des Modells abnimmt. Zudem wirken sich große Einwohnerzahlen positiv aus. Der Sörensenindex beträgt in 70 % der Simulationen 0,75 (vgl. Abbildung 34 (a)), der Korrelationskoeffizient 0,88, der adjustierte R^2-Wert 0,85, der mittlere absolute Fehler 289 und die Wurzel der mittleren quadratischen Abweichung 1.642 (vgl. Abbildung 24).

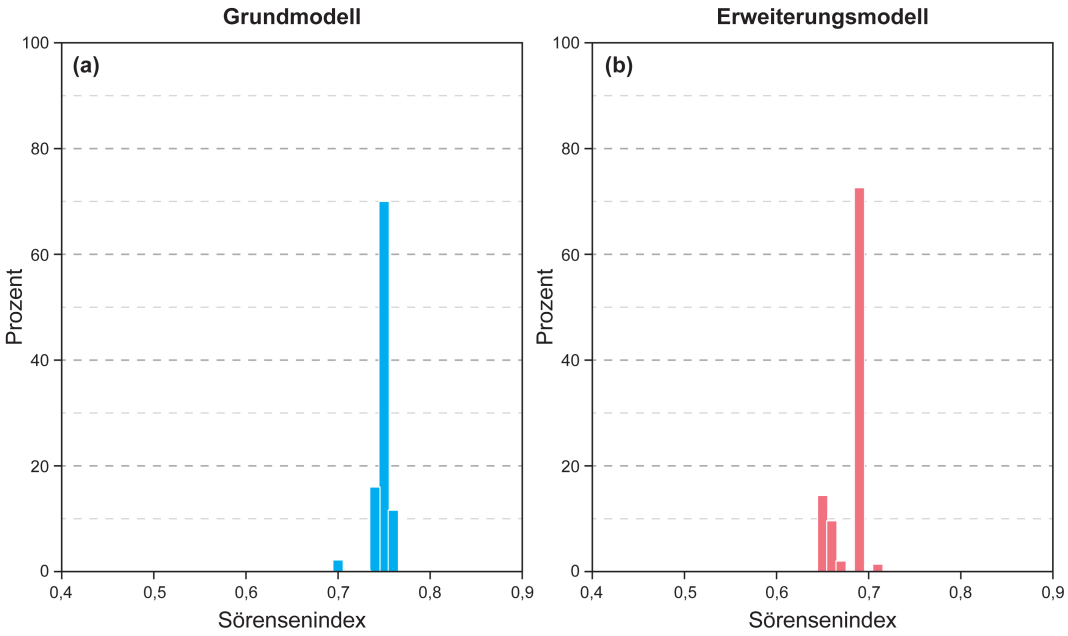


Abbildung 34: Ergebnisse der Simulation für England und Wales. Quelle: Eigene Darstellung auf Grundlage von eigenen Berechnungen auf Basis von Office for National Statistic (2011).

Für die Erweiterung des Grundmodells sind vergleichbare Ergebnisse zu erkennen. Wie in Abbildung 35 (a) führen auch in diesem Fall kleine Entfernungen zwischen den Standorten sowie große Einwohnerzahlen zur Verbesserung des Modells. Zudem ist für hohe Bevölkerungszahlen zwischen den Standorten ein negativer Einfluss auf die Schätzung festzustellen. Der Sörensenindex liegt in 73 % aller Simulationen bei 0,69 (vgl. Abbildung 34 (b)), der Korrelationskoeffizient nach Pearson bei 0,91, der adjustierte R^2-Wert bei 0,82, der mittlere absolute Fehler bei 353 und die Wurzel der mittleren quadratischen Abweichung bei 1.859 (vgl. Abbildung 24).

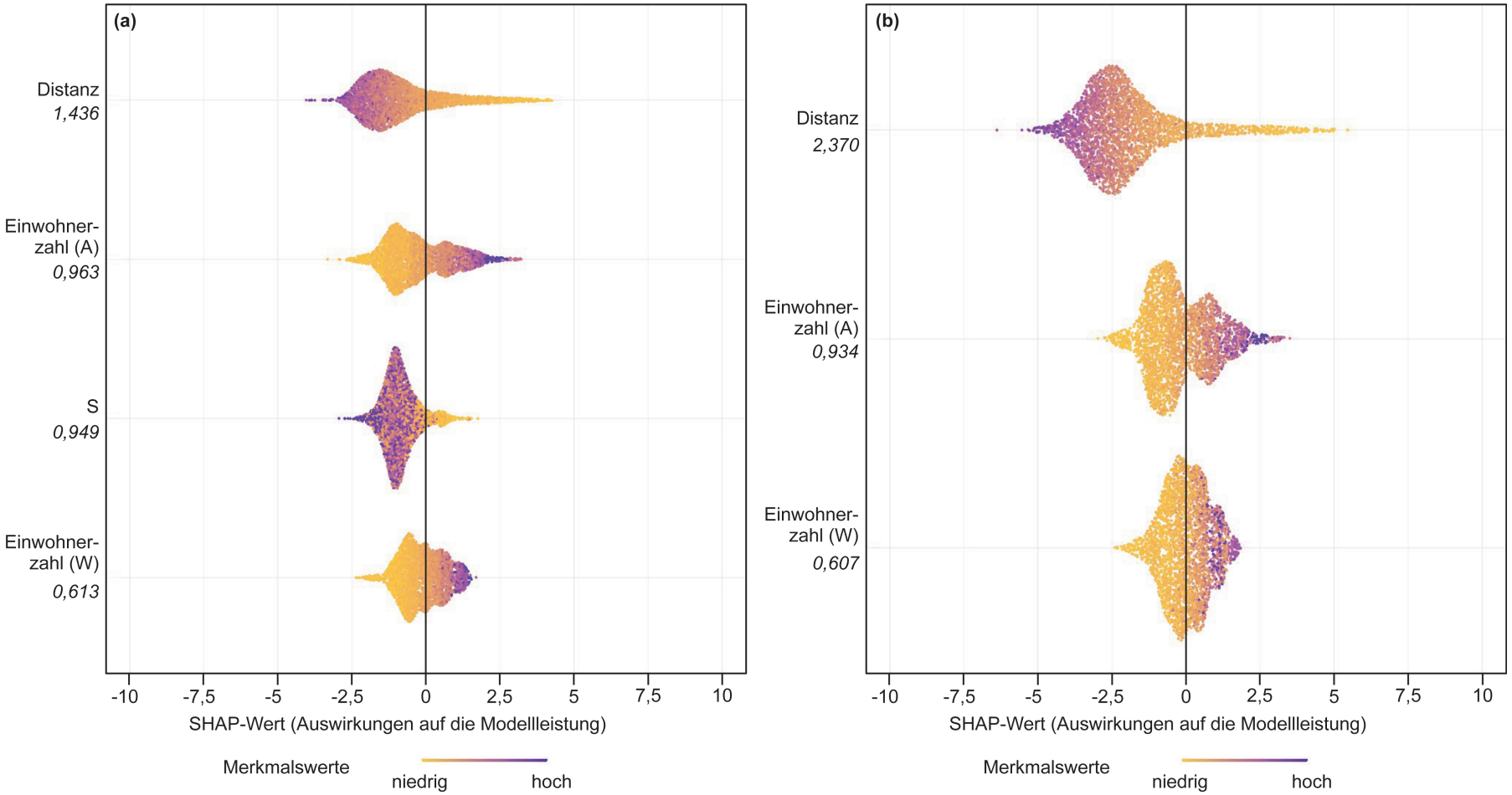


Abbildung 35: SHAP-Werte für den XGBoost-Regressor bei Pendlerbewegungen in England und Wales (Grund- und Erweiterungsmodell). Quelle: Eigene Darstellung auf Grundlage von eigenen Berechnungen auf Basis von Office for National Statistic (2011).

Frankreich

Bei der deskriptiven Betrachtung der Pendlerbewegungen für Frankreich fällt auf, dass die Interaktionen zwischen 0 und 208.549 Personen schwanken. Die größten Mobilitätsflüsse existieren zwischen Hauts-de-Seine und Paris. Das erste Quantil liegt bei 5 Personen, während der Median 17 und das dritte Quantil 57 beträgt. Im Durchschnitt pendeln 493 Personen täglich (Institut national de la statistque et des études économiques 2011).

Für das Gravitationsmodell in Frankreich wird in Abbildung 36 (a) deutlich, dass die Pendlerbewegungen nicht optimal angepasst werden. Die Linie y=x befindet sich nicht zwischen dem 9. und 91. Perzentil in den jeweiligen Klassen. Zudem ist in Abbildung 37 (a) zu erkennen, dass im Umkreis von 180 Kilometern die Wahrscheinlichkeit zu pendeln unterschätzt wird. Im Anschluss folgt eine leichte Überschätzung. Ferner nimmt der Sörensenindex einen Wert von 0,37, der Korrelationskoeffizient nach Pearson einen Wert von 0,84 und der adjustierte R^2-Wert einen Wert von 0,36 an. Der mittlere absolute Fehler liegt bei 633 und die Wurzel der mittleren quadratischen Abweichung bei 4.274 (vgl. Abbildung 22).

Ähnlich wie im Gravitationsmodell werden auch für das Strahlungsmodell die Pendlerbewegungen nicht optimal geschätzt (vgl. Abbildung 36 (b)). Darüber hinaus macht Abbildung 37 (b) deutlich, dass die Wahrscheinlichkeit zu pendeln bis ca. 80 Kilometer unterschätzt wird, während im Anschluss leichte Überschätzungen zu erkennen sind. Der Sörensenindex liefert für das Strahlungsmodell einen Wert von 0,44 und für das finite Strahlungsmodell einen Wert von 0,59. Des Weiteren liegt der Korrelationskoeffizient nach Pearson bei 0,58 bzw. 0,72 und der adjustierte R^2-Wert nimmt Werte von 0,32 bzw. 0,52 an. Außerdem beträgt der mittlere absolute Fehler 424 (Strahlungsmodell) bzw. 370 (finites Strahlungsmodell) und die Wurzel der mittleren quadratischen Abweichung 4.487 (Strahlungsmodell) und 3.758 (finites Strahlungsmodell) (vgl. Abbildung 22).

Der Vergleich der verallgemeinerten Strahlungsmodelle zeigt, dass das ACR-Modell die besten Ergebnisse liefert (Sörensenindex: 0,65; Korrelationskoeffizient nach Pearson: 0,87; adjustierter R^2-Wert: 0,75; mittlerer absoluter Fehler: 347, Wurzel der mittleren quadratischen Abweichung: 2.648). Ferner ist der Abbildung 36 (g) zu entnehmen, dass sich für Pendlerbewegungen in einer Größenordnung von ca. 30 sowie ab ca. 10.000 Personen besonders gute Ergebnisse erzielen lassen. Außerdem sind in Abbildung 37 (g) die Abweichung

zwischen der vorhergesagten und der empirischen Wahrscheinlichkeit marginal. Die anderen drei Modelle schätzen die Pendlerflüsse etwas weniger gut (AIR-Modell: Sörensenindex: 0,65; Korrelationskoeffizient nach Pearson: 0,84; adjustierter R^2-Wert: 0,70; mittlerer absoluter Fehler: 395; Wurzel der mittleren quadratischen Abweichung: 2.850 / PCR-Modell: Sörensenindex: 0,60; Korrelationskoeffizient nach Pearson: 0,71; adjustierter R^2-Wert: 0,51; mittlerer absoluter Fehler: 347; Wurzel der mittleren quadratischen Abweichung: 3.760 / PIR-Modell: Sörensenindex: 0,62; Korrelationskoeffizient nach Pearson: 0,76; adjustierter R^2-Wert: 0,57; mittlerer absoluter Fehler: 373; Wurzel der mittleren quadratischen Abweichung: 3.474) (vgl. Abbildung 22).

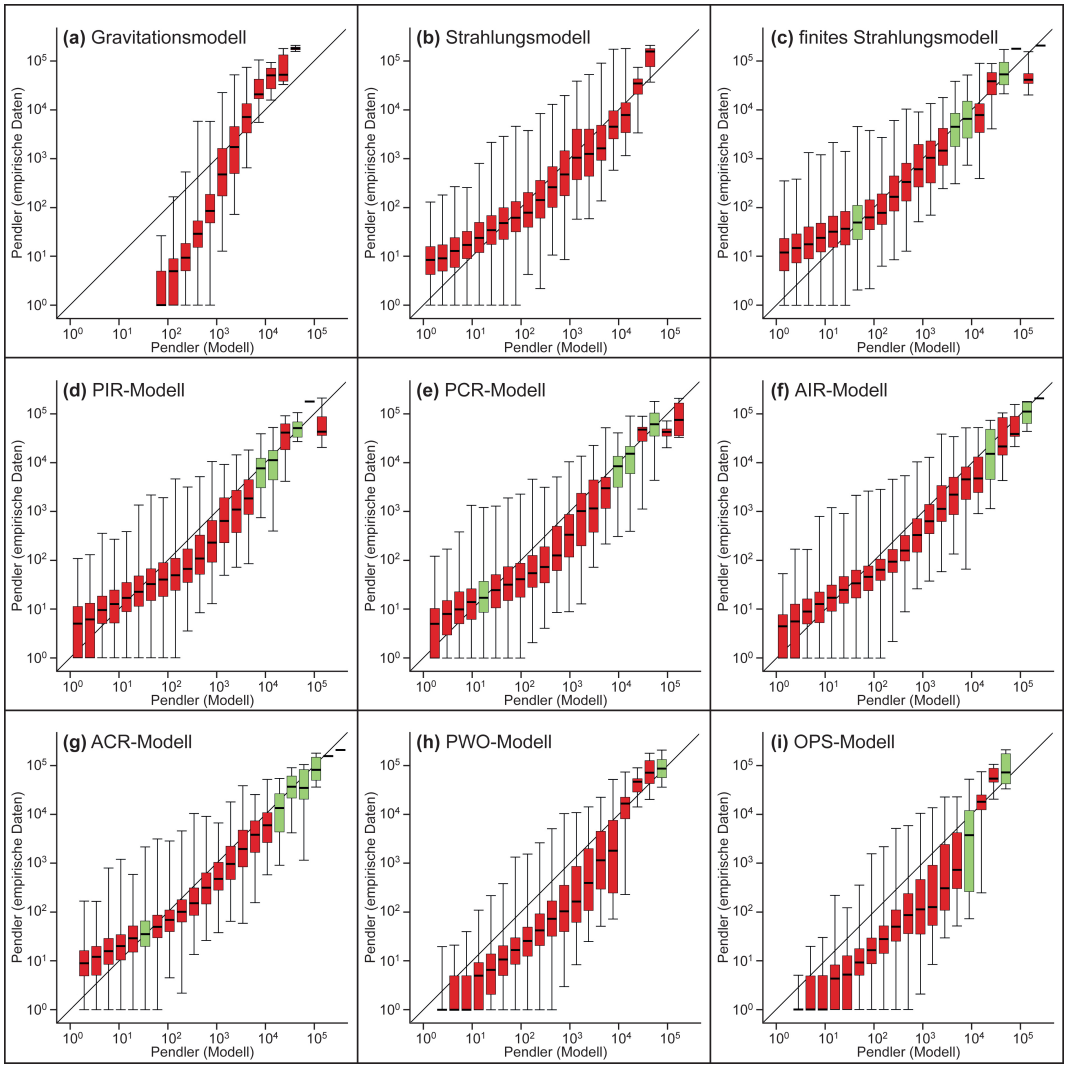


Abbildung 36: Abweichung zwischen den empirischen Daten und den Modellvorhersagen bezüglich der Pendlerbewegungen für Frankreich. Quelle: Eigene Darstellung auf Grundlage von eigenen Berechnungen auf Basis von Institut national de la statistique et des études économiques (2011).

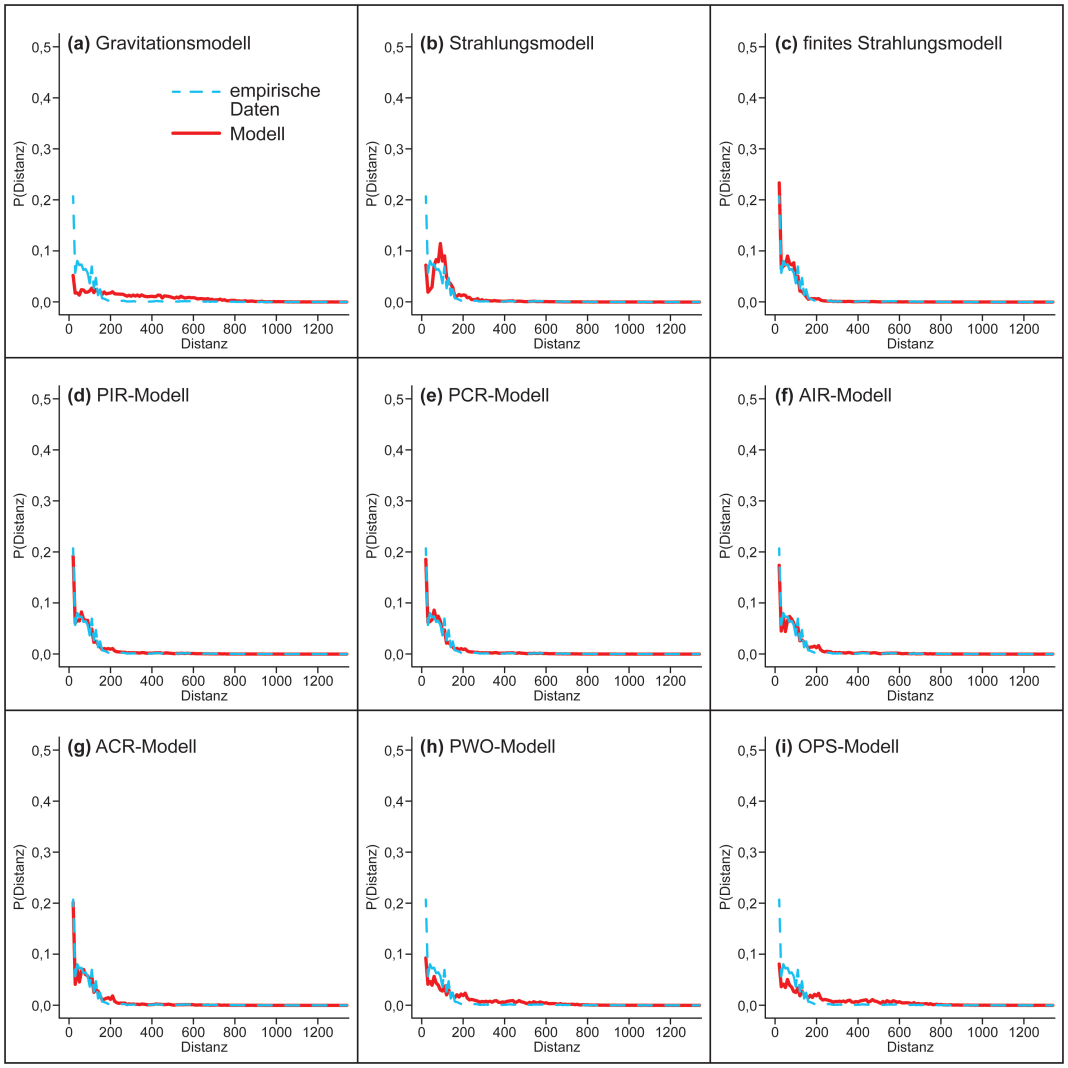


Abbildung 37: P(Distanz) im Verhältnis zur Entfernung (Pendlerbewegung in Frankreich). Quelle: Eigene Darstellung auf Grundlage von eigenen Berechnungen auf Basis von Institut national de la statistique et des études économiques (2011).

Für die Prognose mit dem PWO-Modell werden sehr große Pendlerströme (100.000 Personen) besonders gut geschätzt (vgl. Abbildung 36 (h)). Außerdem wird in Abbildung 37 (h) deutlich, dass bis ca. 150 Kilometer die Wahrscheinlichkeit zu pendeln mit dem Modell unterschätzt wird. Ab ca. 180 Kilometer sind Überschätzungen festzustellen. Die Ergebnisse des nicht parametrisierten PWO-Modells in Abbildung 22 veranschaulichen ferner, dass in Bezug auf den Sörensenindex, den Korrelationskoeffizienten nach Pearson und den adjustieren R^2-Wert leichte Verschlechterungen im Verhältnis zum ACR-Modell vorliegen (Sörensenindex: 0,53; Korrelationskoeffizient nach Pearson: 0,83; adjustierter R^2-Wert: 0,69). Zudem liegt auch der mittlere absolute Fehler mit 465 und die Wurzel der mittleren quadratischen Abweichung mit 3.447 unter den Ergebnissen des ACR-Modells (vgl. Abbildung 22).

Abbildung 36 (i) zeigt zudem für das OPS-Modell, dass für die gewählten Klassenabstände Pendlerströme mit ca. 10.000 und 100.000 Personen sehr gute Vorhersagen gemacht werden können. Die Distanzen bis ca. 100 Kilometer weisen leichte Unterschätzungen auf, es folgt ab ca. 200 Kilometer eine Überschätzung (vgl. Abbildung 36 (i)). Bezüglich der Evaluationsindikatoren sind ähnliche Beobachtungen wie beim PWO-Modell zu machen (vgl. Abbildung 22) (Sörensenindex: 0,49; Korrelationskoeffizient nach Pearson: 0,82; adjustierter R^2-Wert 0,67; mittlerer absoluter Fehler: 501; Wurzel der mittleren quadratischen Abweichung: 3.636).

Für das Grundmodell des XGBoost-Algorithmus wird in Abbildung 39 (b) deutlich, dass vor allem große Distanzen zwischen Ziel- und Quellregion einen negativen Einfluss auf die Vorhersageergebnisse haben. Außerdem wirken sich hohe Einwohnerzahlen an Wohn- und Arbeitsort positiv auf die Schätzung aus. Der Sörensenindex beträgt in 68 % aller Simulationen 0,66 (vgl. Abbildung 38 (a)), der Korrelationskoeffizient nach Pearson 0,91, der adjustierte R^2-Wert 0,83, der mittlere absolute Fehler 245 und die Wurzel der mittleren quadratischen Abweichung 3.398 (vgl. Abbildung 24).

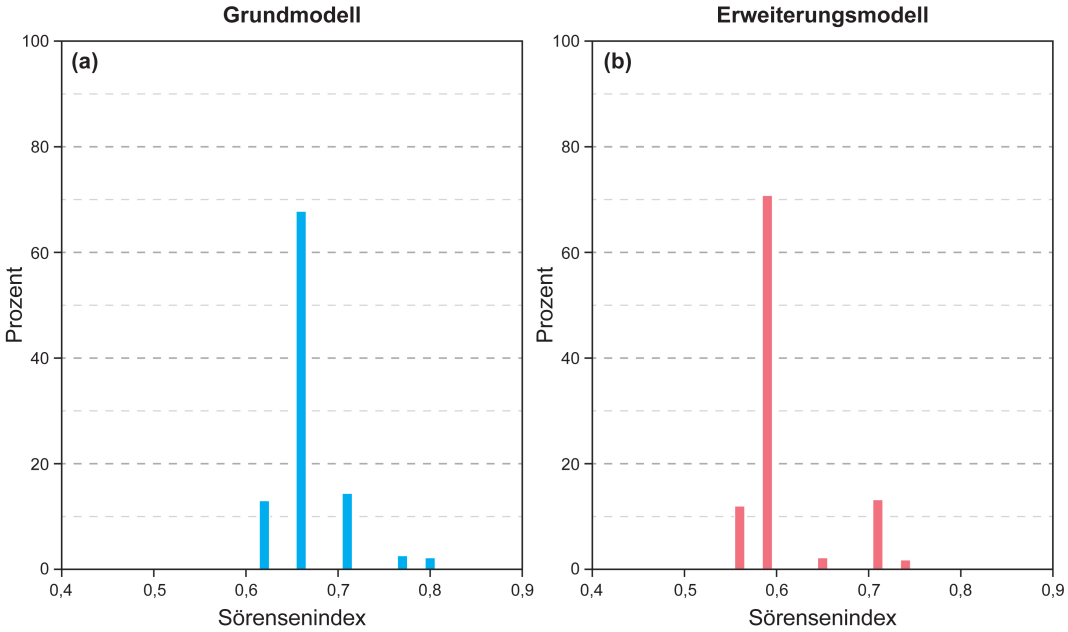


Abbildung 38: Ergebnisse der Simulation für Frankreich. Quelle: Eigene Darstellung auf Grundlage von eigenen Berechnungen auf Basis von Institut national de la statistique et des études économiques (2011).

Für die Erweiterung des Grundmodells des XGBoost-Algorithmus (vgl. Abbildung 39 (a)) ist festzustellen, dass die Einwohnerzahl zwischen Wohn- und Arbeitsort das größte Gewicht bei der Prognose einnimmt. Sofern diese Bevölkerungszahl besonders groß ist, wird ein negativer Einfluss auf die Vorhersage gezeigt. Bei der Distanz (als zweitwichtigster Indikator) kann keine eindeutige Wirkrichtung ausgemacht werden. Außerdem beeinflussen hohe Bevölkerungszahlen in Ziel- und Quellregion die Prognose positiv. Der Sörensenindex liefert in 71 % aller Simulationen einen Wert von 0,59 (vgl. Abbildung 38 (b)), der Korrelationskoeffizient nach Pearson einen Wert von 0,91, der adjustierte R^2-Wert einen Wert von 0,85. Zudem liegt der mittlere absolute Fehler bei 259 und die Wurzel der mittleren quadratischen Abweichung bei 3.238 (vgl. Abbildung 24).

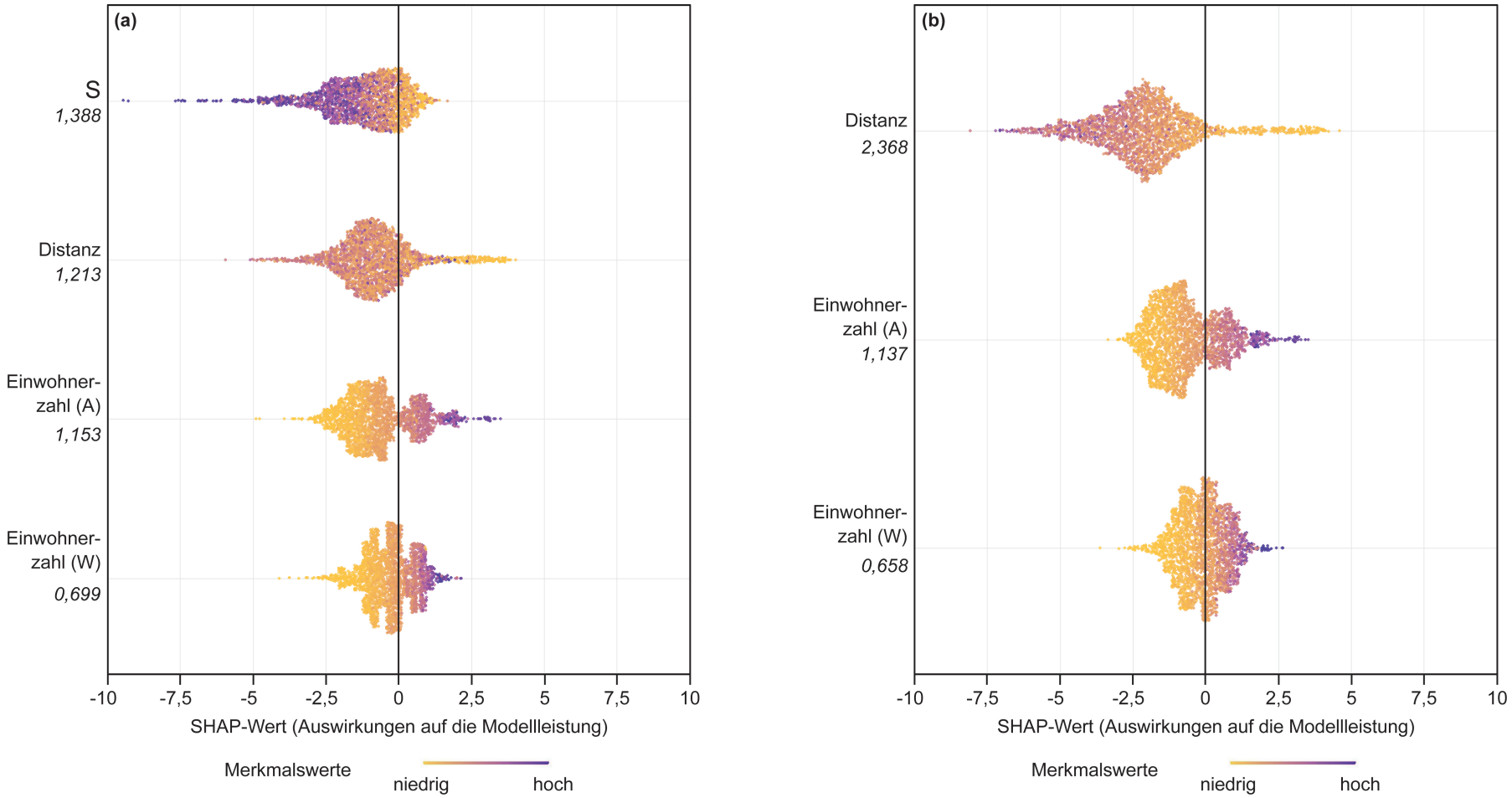


Abbildung 39: SHAP-Werte für den XGBoost-Regressor bei Pendlerbewegungen in Frankreich (Grund- und Erweiterungsmodell). Quelle: Eigene Darstellung auf Grundlage von eigenen Berechnungen auf Basis von Institut national de la statistique et des études économiques (2011).

Irland

Die Pendlerbewegungen (ohne Binnenpendler) in Irland auf NUTS3-Ebene schwanken zwischen 71 und 74.022 Interaktionen, wobei der schwächste Mobilitätsfluss zwischen Mid-West und Border auftritt. Die Regionen Mid-East und Dublin weisen zudem die stärksten Pendlerbewegungen auf. Durchschnittlich hat der Pendlerstrom eine Größe von 3.630 Personen und der Median liegt bei 1.140, das erste Quantil bei 252 und das dritte Quantil bei 3.040 Interaktionen (Central Statistics Office 2011).

Die Ergebnisse des Gravitationsmodells für Irland zeigen, dass keine optimalen Anpassungen an die empirischen Daten vorliegen (vgl. Abbildung 40 (a)). Die Linie y=x ist nicht zwischen dem 9. und 91. Perzentil in der jeweiligen Klasse. Darüber hinaus ist in Abbildung 41 (a) zu erkennen, dass die Wahrscheinlichkeit zu pendeln im Umkreis der ersten 100 Kilometer deutlich unterschätzt wird und ab ca. 200 Kilometern eine Überschätzung vorliegt. Der Sörensenindex beträgt 0,62, der Korrelationskoeffizient nach Pearson 0,91, der adjustierte R^2-Wert 0,82, der mittlere absolute Fehler 2.772 und die Wurzel der mittleren quadratischen Abweichung 6.747 (vgl. Abbildung 22).

Aus Abbildung 40 (b) ist ferner ersichtlich, dass die Vorhersagen mittels des Strahlungsmodells für die Pendlerflüsse von ca. 100 Personen besonders gute Schätzungen liefert. Außerdem wird in Abbildung 41 (b) gezeigt, dass die Wahrscheinlichkeit einer täglichen Fahrt zum Arbeitsplatz bis ca. 100 Kilometer mit dem Modell unterschätzt wird. Im Anschluss folgt bis ca. 180 Kilometer eine Überschätzung. Die Unterschiede zwischen Modell und empirischen Daten werden ab etwa 200 Kilometern deutlich geringer. Für das Strahlungsmodell liefert der Vergleich mit den empirischen Daten einen Sörensenindex von 0,48, einen Korrelationskoeffizient nach Pearson von 0,55 und einen adjustierten R^2-Wert von 0,29. Der mittlere absolute Fehler liegt bei 3.492 und die Wurzel der mittleren quadratischen Abweichung bei 8.642 (vgl. Abbildung 22).

Betrachtet man die Ergebnisse des finiten Strahlungsmodells, so wird deutlich, dass die Linie y=x nicht zwischen dem 9. und 91. Perzentil in der jeweiligen Klasse liegt und somit die Pendlerbewegungen nicht optimal geschätzt werden können (vgl. Abbildung 40 (c)). Des Weiteren wird, ähnlich wie beim Strahlungsmodell, die Wahrscheinlichkeit bis zu einer Distanz von ca. 100 Kilometern unterschätzt. Die Abweichungen sind jedoch geringer. Es folgt eine Überschätzung der Daten bis ca. 140 Kilometern. Außerdem treten ab 230 Kilometern Überschätzungen auf (vgl. Abbildung 41 (c)). Für

das finite Strahlungsmodell konnte außerdem festgestellt werden, dass der Sörensenindex bei 0,63, der Korrelationskoeffizient nach Pearson bei 0,94 und der adjustierte R^2-Wert bei 0,88 liegt. Der mittlere absolute Fehler nimmt einen Wert von 3.804 und die Wurzel der mittleren quadratischen Abweichung einen Wert von 5.593 an (vgl. Abbildung 22).

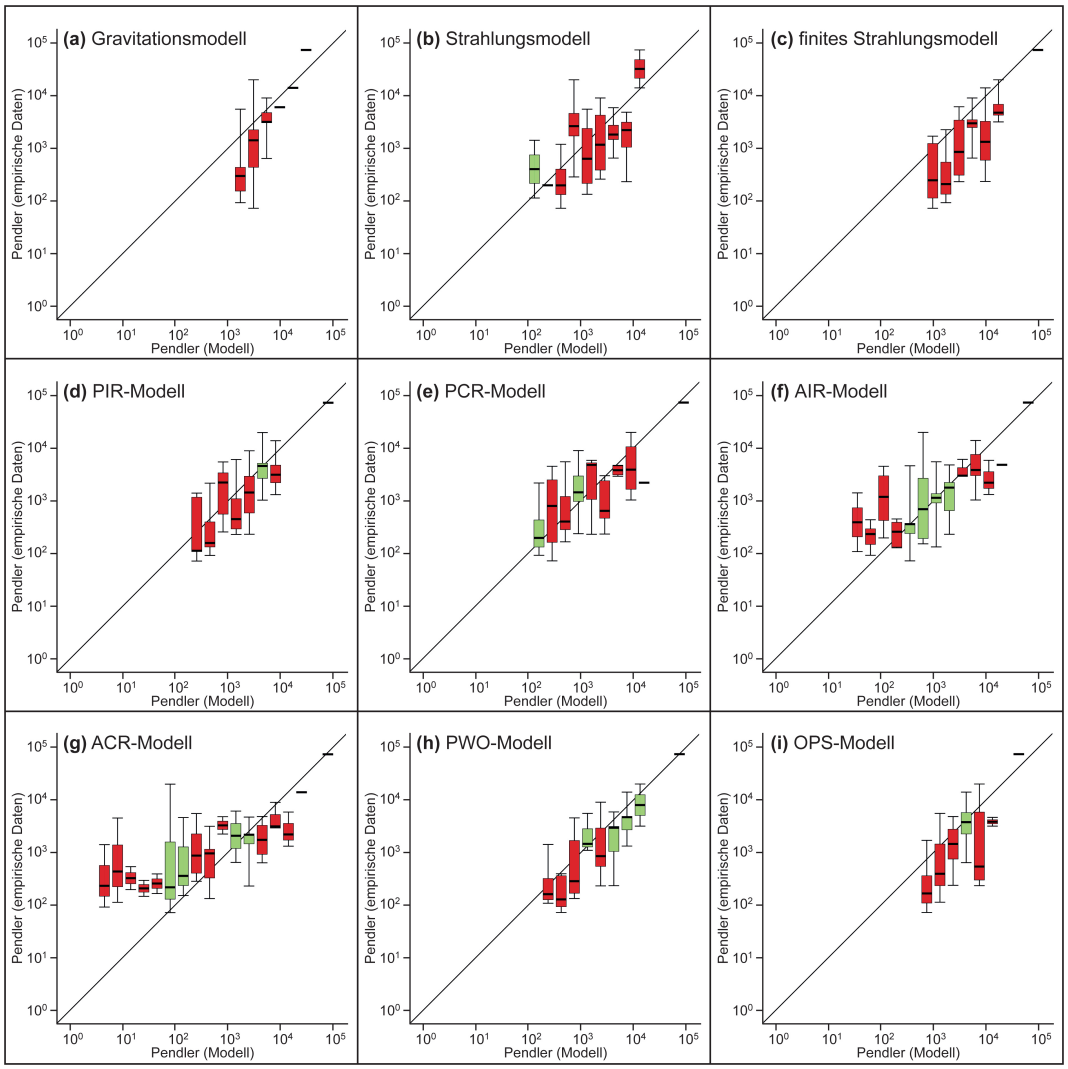


Abbildung 40: Abweichung zwischen den empirischen Daten und den Modellvorhersagen bezüglich der Pendlerbewegungen für Irland. Quelle: Eigene Darstellung auf Grundlage von eigenen Berechnungen auf Basis von Central Statistics Office (2011).

Unter den verallgemeinerten Strahlungsmodellen liefern die Evaluationsindikatoren für das PIR-Modell besonders gute Ergebnisse. Der Sörensenindex beträgt 0,73, der Korrelationskoeffizient nach Pearson 0,95 und der adjustierte R^2-Wert 0,90. Zudem ist der mittlere absolute Fehler mit 1.939 und die Wurzel der mittleren quadratischen Abweichung mit 3.155 am geringsten. Dennoch werden in diesem Modell lediglich Pendlerflüsse mit einer Größe

von ca. 6.000 Pendlern optimal geschätzt (vgl. Abbildung 40 (d)). Zudem ist in Abbildung 41 (d) zu erkennen, dass bis zu ca. 100 Kilometer eine Unterschätzung der Wahrscheinlichkeit zu pendeln vorliegt. Allerdings sind diese Abweichungen deutlich geringer als beim Gravitations-, Strahlungs- und finiten Strahlungsmodell.

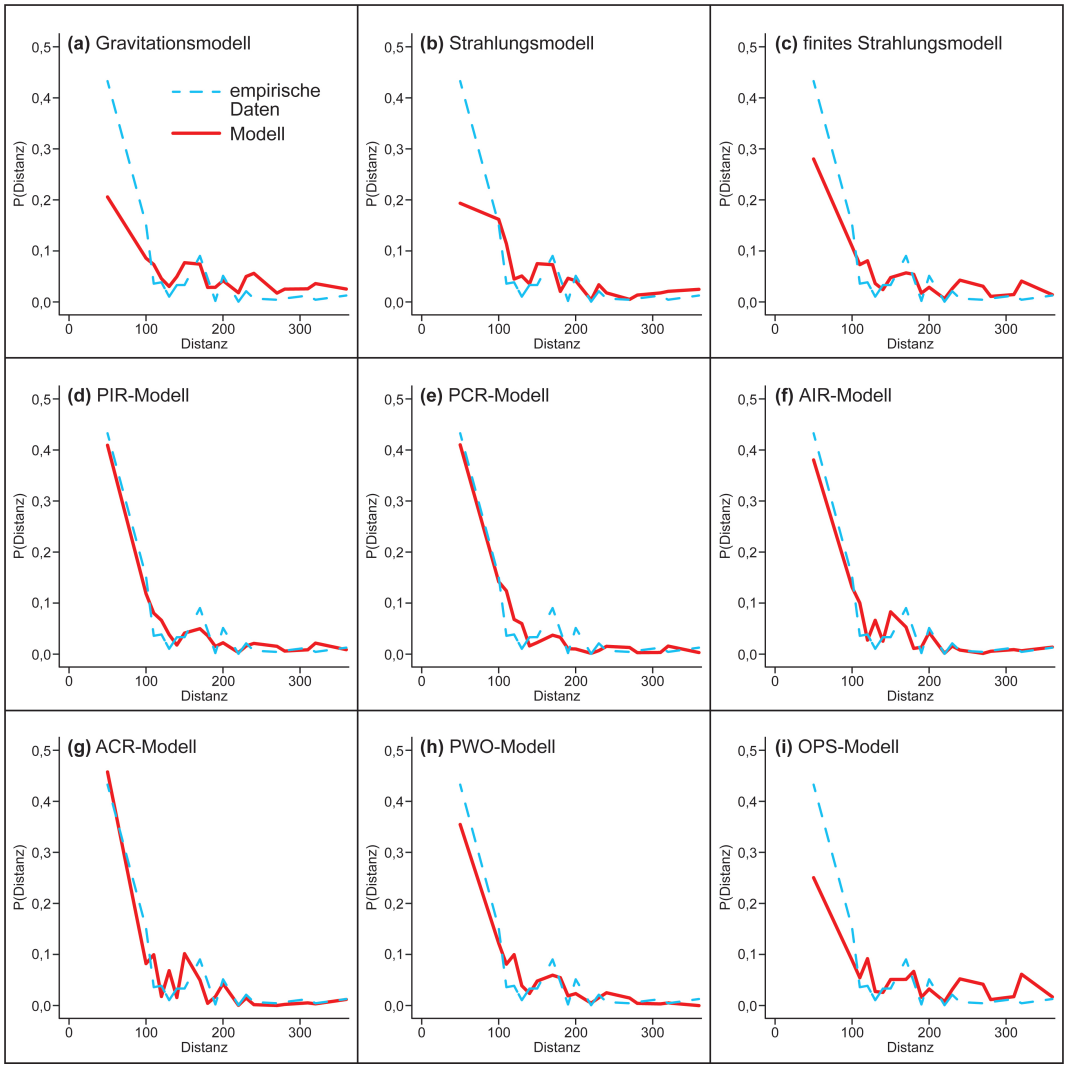


Abbildung 41: P(Distanz) im Verhältnis zur Entfernung (Pendlerbewegung in Irland). Quelle: Eigene Darstellung auf Grundlage von eigenen Berechnungen auf Basis von Central Statistics Office (2011).

Ähnliche Beobachtungen sind auch für die anderen drei verallgemeinerten Strahlungsmodelle zu machen. Die Abweichungen (vgl. Abbildung 22) zwischen empirischen Daten und Schätzungen sind jedoch etwas größer (AIR-Modell: Sörensenindex: 0,7; Korrelationskoeffizient nach Pearson 0,91; adjustierter R^2-Wert: 0,83; mittlerer absoluter Fehler: 2.189; Wurzel der mittleren

quadratischen Abweichung: 4.214 / PCR-Modell: Sörensenindex: 0,71; Korrelationskoeffizient nach Pearson: 0,94; adjustierter R^2-Wert: 0,87; mittlerer absoluter Fehler: 2.118; Wurzel der mittleren quadratischen Abweichung: 3.598 / ACR-Modell: Sörensenindex: 0,69; Korrelationskoeffizient nach Pearson: 0,91; adjustierter R^2-Wert: 0,82; mittlerer absoluter Fehler: 2.287; Wurzel der mittleren quadratischen Abweichung: 4.303).

Für das PWO-Modell wird sichtbar, dass vor allem größere Pendlerströme (ab ca. 1.000 Pendler) besonders gute Anpassungen liefern (vgl. Abbildung 40 (h)). Die Abweichungen bezüglich der Pendlerwahrscheinlichkeit in Abhängigkeit zur Distanz sind deutlich geringer als in den übrigen parameterfreien Modellen (vgl. Abbildung 41 (h)). Der Sörensenindex für dieses Modell liegt bei 0,72, während der Korrelationskoeffizient nach Pearson einen Wert von 0,96 und der adjustierte R^2-Wert von 0,91 annimmt. Der mittlere absolute Fehler des PWO-Modells beträgt 2.023 und die Wurzel der mittleren quadratischen Abweichung 3.119 (vgl. Abbildung 22).

Darüber hinaus zeigt der Vergleich der empirischen Daten mit der Prognose der Mobilitätsflüsse mittels OPS-Modell in Abbildung 40 (i), dass Pendlerflüsse mit ca. 6.000 Personen besonders gute Ergebnisse liefern. Bezüglich der Distanz sind bis zu 100 Kilometer deutliche Unterschätzungen festzustellen. Ab etwa 110 bzw. 230 Kilometern werden die Mobilitätsströme mit dem Modell überschätzt (vgl. Abbildung 41 (i)). Der Sörensenindex (vgl. Abbildung 22), der Korrelationskoeffizient nach Pearson sowie der adjustierte R^2-Wert liegen bei 0,65, 0,91 und 0,83. Ferner ergeben die Berechnungen des mittleren absoluten Fehlers und die Wurzel der mittleren quadratischen Abweichung Werte von 2.559 und 5.024.

Aus Abbildung 43 (b) ist für das Grundmodell des XGBoost-Algorithmus zu entnehmen, dass weitere Entfernungen zwischen Arbeits- und Wohnort die Vorhersageleistungen des Modells verschlechtern. Zudem wirken sich große Einwohnerzahlen positiv aus. Der Sörensenindex beträgt in 65 % der Simulationen 0,52 (vgl. Abbildung 42 (a)), der Korrelationskoeffizient 0,22, der adjustierte R^2-Wert 0,16, der mittlere absolute Fehler 1.666 und die Wurzel der mittleren quadratischen Abweichung 2.624 (vgl. Abbildung 24).

Ferner ist in Abbildung 43 (a) für die Erweiterung des XGBoost-Algorithmus erkennbar, dass auch in diesem Fall große Entfernungen zwischen der Ziel- und Quellregion zur Verschlechterung der Prognose führen. Für die Einwohnerzahl zwischen den Standorten als zweitwichtigster Indikator lässt sich keine eindeutige Wirkrichtung festlegen, wohingegen Standorte mit vielen Einwohnern die Vorhersage verbessern. Der Sörensenindex nimmt in 66 % der Simulation einen Wert von 0,54 (vgl. Abbildung 42 (b)), der Korrelationskoeffizient nach Pearson einen Wert von 0,11 und der adjustierte R^2-Wert einen Wert von 0,13 an. Außerdem liegt der mittlere absolute Fehler bei 1.789 und die Wurzel der mittleren quadratischen Abweichung bei 2.846 (vgl. Abbildung 24).

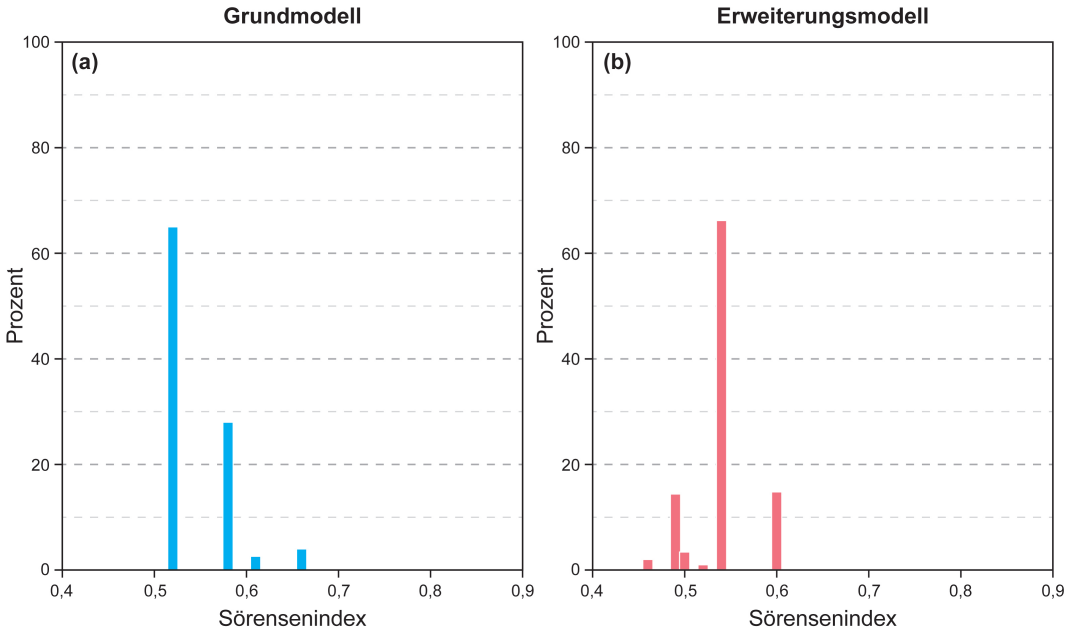


Abbildung 42: Ergebnisse der Simulation für Irland. Quelle: Eigene Darstellung auf Grundlage von eigenen Berechnungen auf Basis von Central Statistics Office (2011).

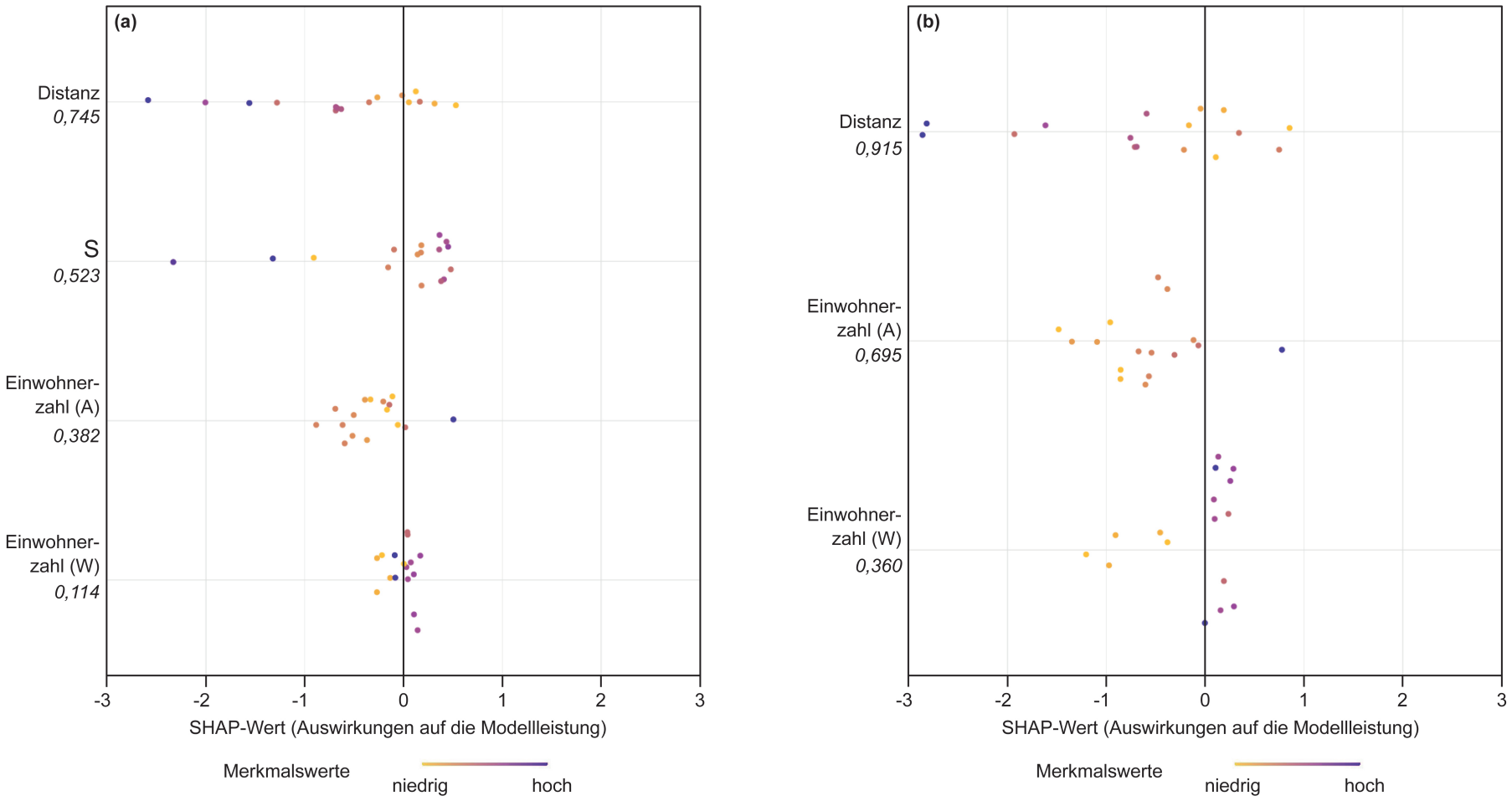


Abbildung 43: SHAP-Werte für den XGBoost-Regressor bei Pendlerbewegungen in Irland (Grund- und Erweiterungsmodell). Quelle: Eigene Darstellung auf Grundlage von eigenen Berechnungen auf Basis von Central Statistics Office (2011).

Österreich

Betrachtet man die Mobilitätsflüsse von Arbeitskräften in Österreich, so wird deutlich, dass sie zwischen 5 und 41.829 Personen schwanken. Die stärkste Interaktion liegt für die NUTS3-Regionen Oststeiermark und Wien vor. Im Mittel pendeln 1.436 Personen täglich auf NUTS3-Ebene. Das erste Quantil liegt bei 59, das zweite Quantil bei 285 und das dritte Quantil bei 1.093 Interaktionen (Statistic Austria 2011).

In Abbildung 44 (a) werden die Ergebnisse für Österreich zusammengefasst. Es ist zu erkennen, dass das Gravitationsmodell insbesondere für Pendlerströme bei einer Größenordnung von ca. 10.000 Personen besonders gute Ergebnisse liefert. Die Übersicht in Abbildung 45 (a) zeigt zudem, dass die Wahrscheinlichkeit zu pendeln bis ca. 30 Kilometer leicht überschätzt wird und zwischen 70 und 100 Kilometern Unterschätzungen vorliegen. Der Sörensenindex liegt bei 0,59, der Korrelationskoeffizient nach Pearson bei 0,5, der adjustierte R^2-Wert bei 0,25, der mittlere absolute Fehler bei 1.149 und die Wurzel der mittleren quadratischen Abweichung bei 3.434 (vgl. Abbildung 22).

In Abbildung 44 (b) ist zu erkennen, dass mithilfe des Strahlungsmodells bei Pendlerbewegungen von ca. 1.000 Personen eine gute Anpassung zwischen Prognose und empirischen Daten erzielt werden kann. Zudem wird die Wahrscheinlichkeit in einen Ort zu pendeln im Umkreis der ersten 90 Kilometer überschätzt (vgl. Abbildung 45 (b)). Außerdem liegt der Sörensenindex bei 0,47, der Korrelationskoeffizient nach Pearson bei 0,5, der adjustierte R^2-Wert bei 0,25, der mittlere absolute Fehler bei 1.468 und die Wurzel der mittleren quadratischen Abweichung bei 4.205 (vgl. Abbildung 22).

Betrachtet man die Ergebnisse des finiten Strahlungsmodells, so wird deutlich, dass kleine Pendlerbewegungen (5, 20, 80-100, 300 Personen) gute Anpassungen liefern (vgl. Abbildung 44 (c)). Außerdem ist aus Abbildung 45 (c) zu entnehmen, dass die Wahrscheinlichkeit eine Strecke zu pendeln bis ca. 90 Kilometer überschätzt wird. Zudem sind die Evaluationsindikatoren (vgl. Abbildung 22) für das finite Strahlungsmodell mit den Ergebnissen des Strahlungsmodells vergleichbar (Sörensenindex: 0,52; Korrelationskoeffizient nach Pearson: 0,6; adjustierter R^2-Wert: 0,35; mittlerer absoluter Fehler: 1.722; Wurzel der mittleren quadratischen Abweichung: 5.330).

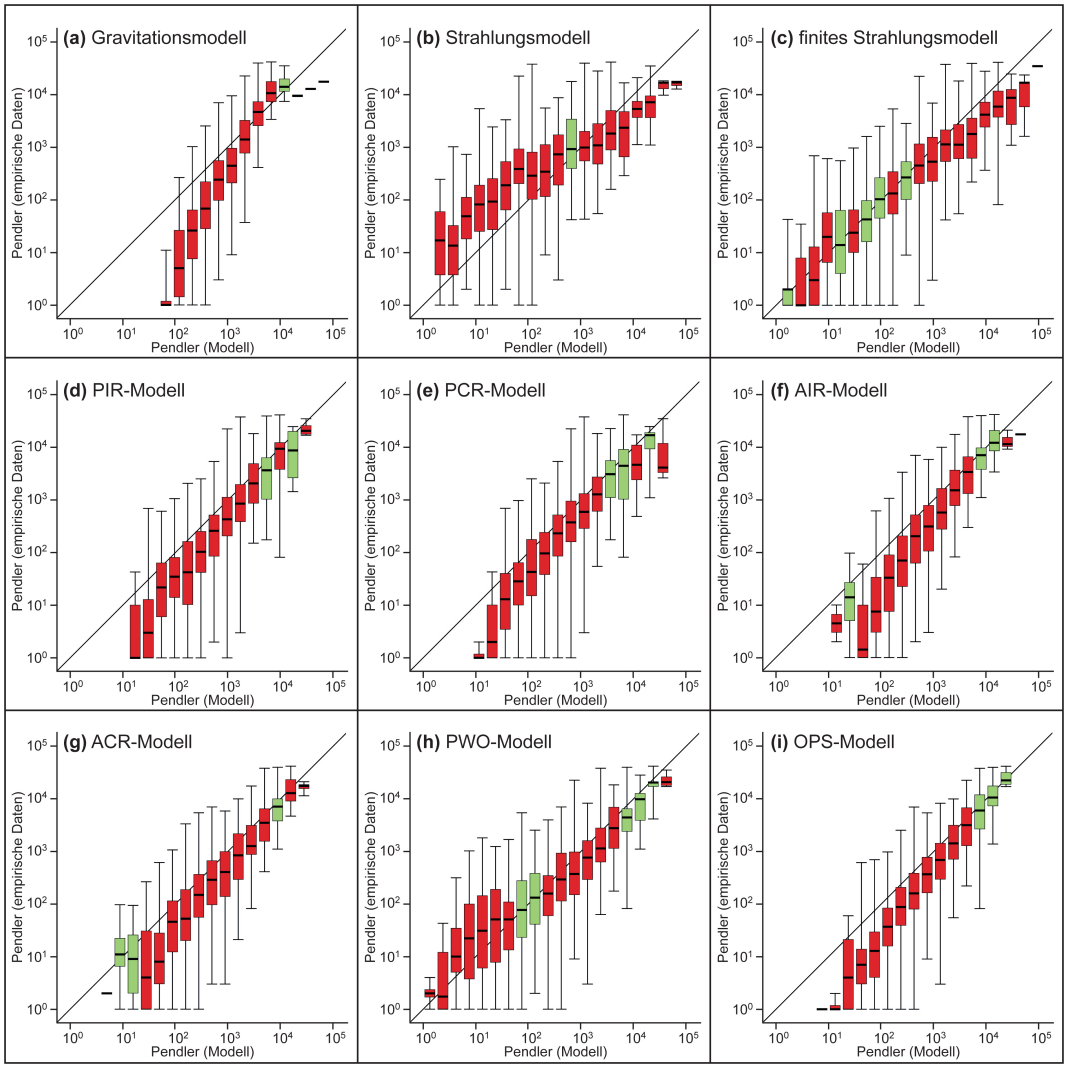


Abbildung 44: Abweichung zwischen den empirischen Daten und den Modellvorhersagen bezüglich der Pendlerbewegungen für Österreich. Quelle: Eigene Darstellung auf Grundlage von eigenen Berechnungen auf Basis von Statistic Austria (2011)

.

Beim Vergleich der verallgemeinerten Strahlungsmodelle wird deutlich, dass die Prognose der Mobilitätsflüsse für das ACR-Modell besonders gut sind. Abbildung 44 (g) veranschaulicht geringe Abweichungen zwischen Prognose und empirischen Daten bei Mobilitätsströmen in einer Größenordnung von ca. 10 bis 20 Personen sowie bei 10.000 Interaktionen. Kleine Unterschätzungen sind jedoch für das direkte Umfeld (bis ca. 40 Kilometer) zu erkennen (vgl. Abbildung 45 (g)). Diese Beobachtung wird auch durch die Evaluationskoeffizienten unterstrichen (vgl. Abbildung 22). Der Sörensenindex liegt bei 0,63, der Korrelationskoeffizient nach Pearson bei 0,7 und der adjustierte R^2-Wert bei 0,49. Der mittlere absolute Fehler beträgt 1.069 und die Wurzel der mittleren quadratischen Abweichung 2.666. Die Ergebnisse für die anderen drei Modelle fal-

len etwas schlechter aus (AIR-Modell: Sörensenindex: 0,6; Korrelationskoeffizient nach Pearson: 0,64; adjustierter R^2-Wert: 0,41; mittlerer absoluter Fehler: 1.146; Wurzel der mittleren quadratischen Abweichung: 2.885 / PCR-Modell: Sörensenindex: 0,59; Korrelationskoeffizient nach Pearson: 0,55; adjustierter R^2-Wert: 0,3; mittlerer absoluter Fehler: 1.182; Wurzel der mittleren quadratischen Abweichung: 3.289 / PIR-Modell: Sörensenindex: 0,6; Korrelationskoeffizient nach Pearson: 0,64; adjustierter R^2-Wert: 0,41; mittlerer absoluter Fehler: 1.136; Wurzel der mittleren quadratischen Abweichung: 2.848).

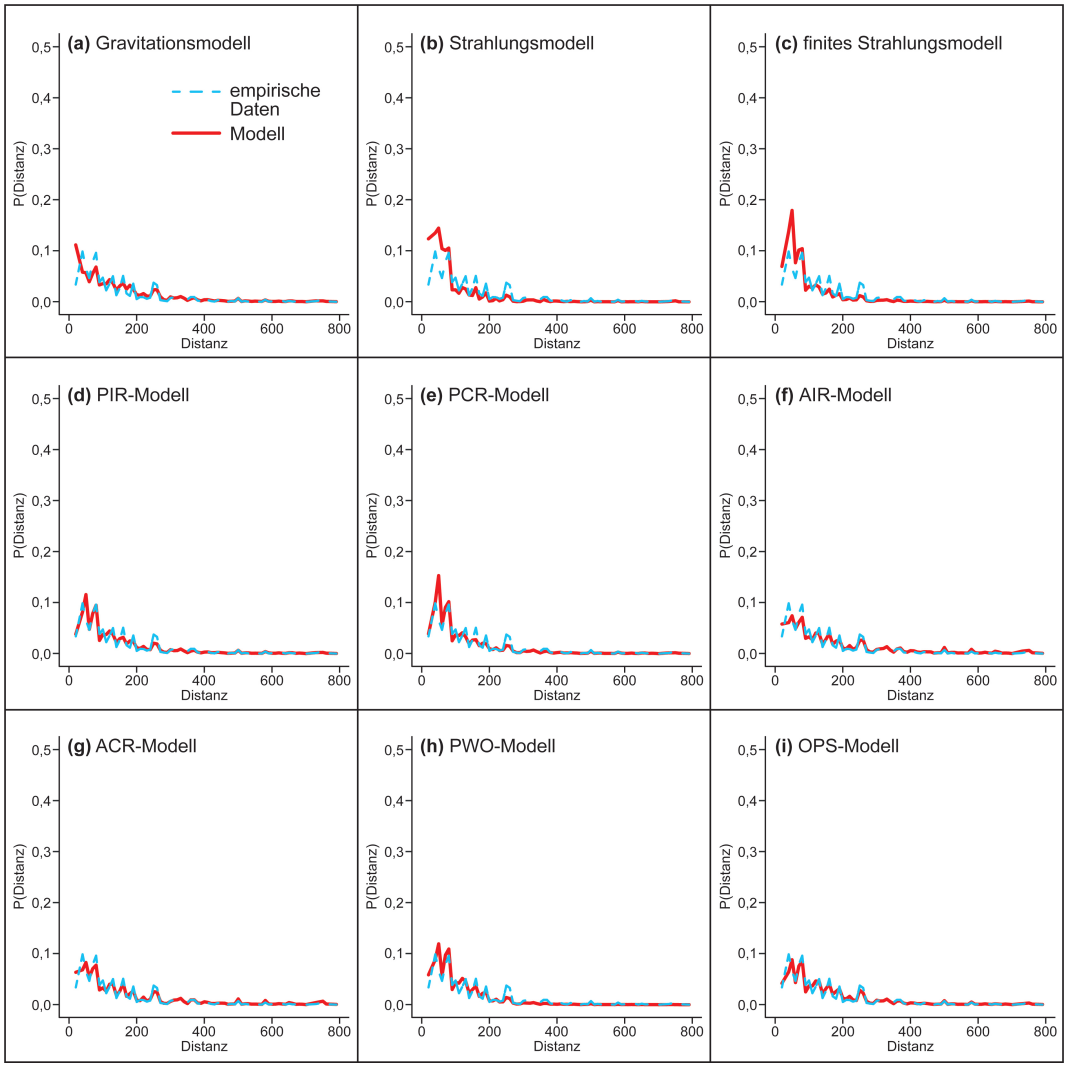


Abbildung 45: P(Distanz) im Verhältnis zur Entfernung (Pendlerbewegung in Österreich). Quelle: Eigene Darstellung auf Grundlage von eigenen Berechnungen auf Basis von Statistik Austria (2011).

Darüber hinaus zeigt Abbildung 44 (h) für das PWO-Modell, dass prognostizierte Mobilitätsflüsse von ca. 80 bis 100 sowie 10.000 bis 30.000 Pendlern geringe Abweichungen aufweisen. Es sind vor allem Unterschätzungen für Distanzen von ca. 250 bis 280 Kilometer zu erkennen (vgl. Abbildung 45 (h)). Der

Sörensenindex nimmt einen Wert von 0,62, der Korrelationskoeffizient nach Pearson einen Wert von 0,69 und der adjustierte R^2-Wert von 0,48 an. Zudem konnte ein mittlerer absoluter Fehler von 1.099 und eine Wurzel der mittleren quadratischen Abweichung von 2.770 festgestellt werden (vgl. Abbildung 22).

Für Österreich wird ebenfalls deutlich, dass das OPS-Modell die besten Schätzungen liefert. Grundsätzlich werden vor allem Mobilitätsströme in einer Größenordnung von mehr als 10.000 Personen besonders gut angenähert (vgl. Abbildung 44 (i)). Aus Abbildung 45 (i) ist darüber hinaus zu entnehmen, dass die Wahrscheinlichkeit zu pendeln in Abhängigkeit zur Distanz im direkten Umfeld marginale Unterschätzungen aufweist. Außerdem beträgt der Sörensenindex 0,65, der Korrelationskoeffizient nach Pearson 0,77 und der adjustierte R^2-Wert 0,59. Der mittlere absolute Fehler nimmt einen Wert von 995 und die Wurzel der mittleren quadratischen Abweichung von 2.373 an (vgl. Abbildung 22).

Untersucht man zusätzlich die Ergebnisse des XGBoost-Algorithmus ist für das Grundmodell zu erkennen (vgl. Abbildung 47 (b)), dass große Entfernungen die Prognose negativ beeinflussen und hohe Einwohnerzahlen am Wohn- und Arbeitsort einen positiven Einfluss haben. Der Sörensenindex liegt in 71 % der Simulationen bei 0,67 (vgl. Abbildung 46 (a)), der Korrelationskoeffizient nach Pearson bei 0,66, der adjustierte R^2-Wert bei 0,52, der mittlere absolute Fehler bei 920 und die Wurzel der mittleren quadratischen Abweichung bei 2.962 (vgl. Abbildung 24).

In Abbildung 47 (a) wird deutlich, dass auch für die Erweiterung des Grundmodells des XGBoost-Algorithmus festzustellen ist, dass die Anzahl an Einwohnern am Arbeitsort den stärksten Einfluss auf die Vorhersage hat. Eine große Einwohnerzahl beeinflusst die Prognose positiv. Der zweitwichtigste Aspekt ist die Distanz. Es ist erkennbar, dass kleine Entfernungen ebenfalls einen positiven Einfluss erbringen. Zudem veringern hohe Einwohnerzahlen zwischen Standorten die Prognoseleistung, während eine hohe Bevölkerungszahl am Wohnort positive Auswirkungen hat. Der Sörensenindex nimmt in 71 % der Simulationen einen Wert von 0,67 (vgl. Abbildung 46 (b)), der Korrelationskoeffizient nach Pearson 0,66 und der adjustierte R^2-Wert einem Wert von 0,44 an. Der mittlere absolute Fehler liegt bei 926 und die Wurzel der mittleren quadratischen Abweichung bei 3.111 (vgl. Abbildung 24).

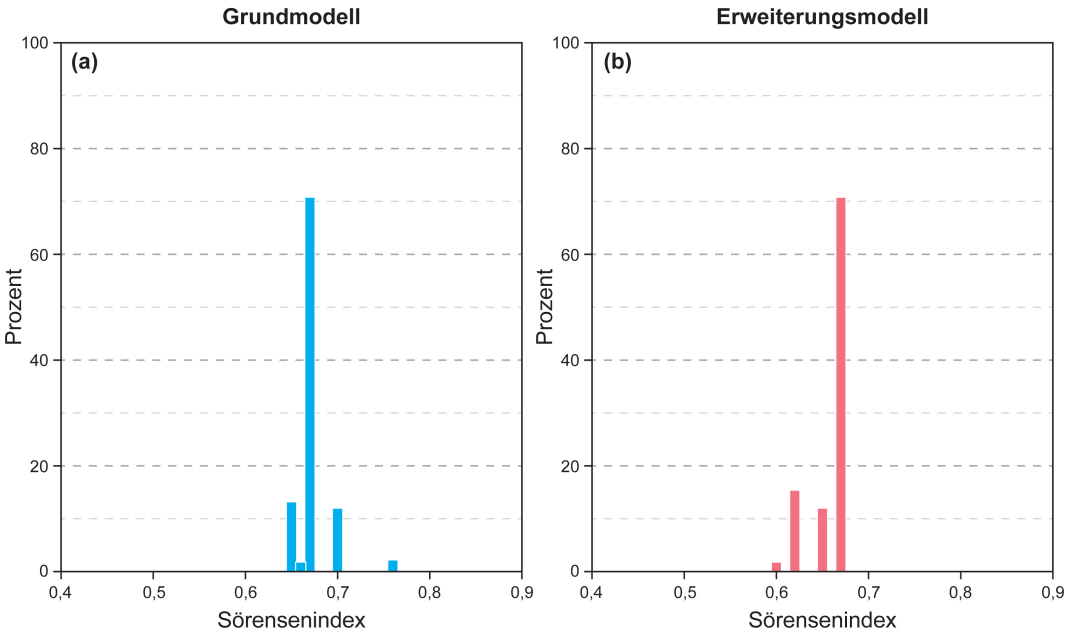


Abbildung 46: Ergebnisse der Simulation für Österreich. Quelle: Eigene Darstellung auf Grundlage von eigenen Berechnungen auf Basis von Statistik Austria (2011).

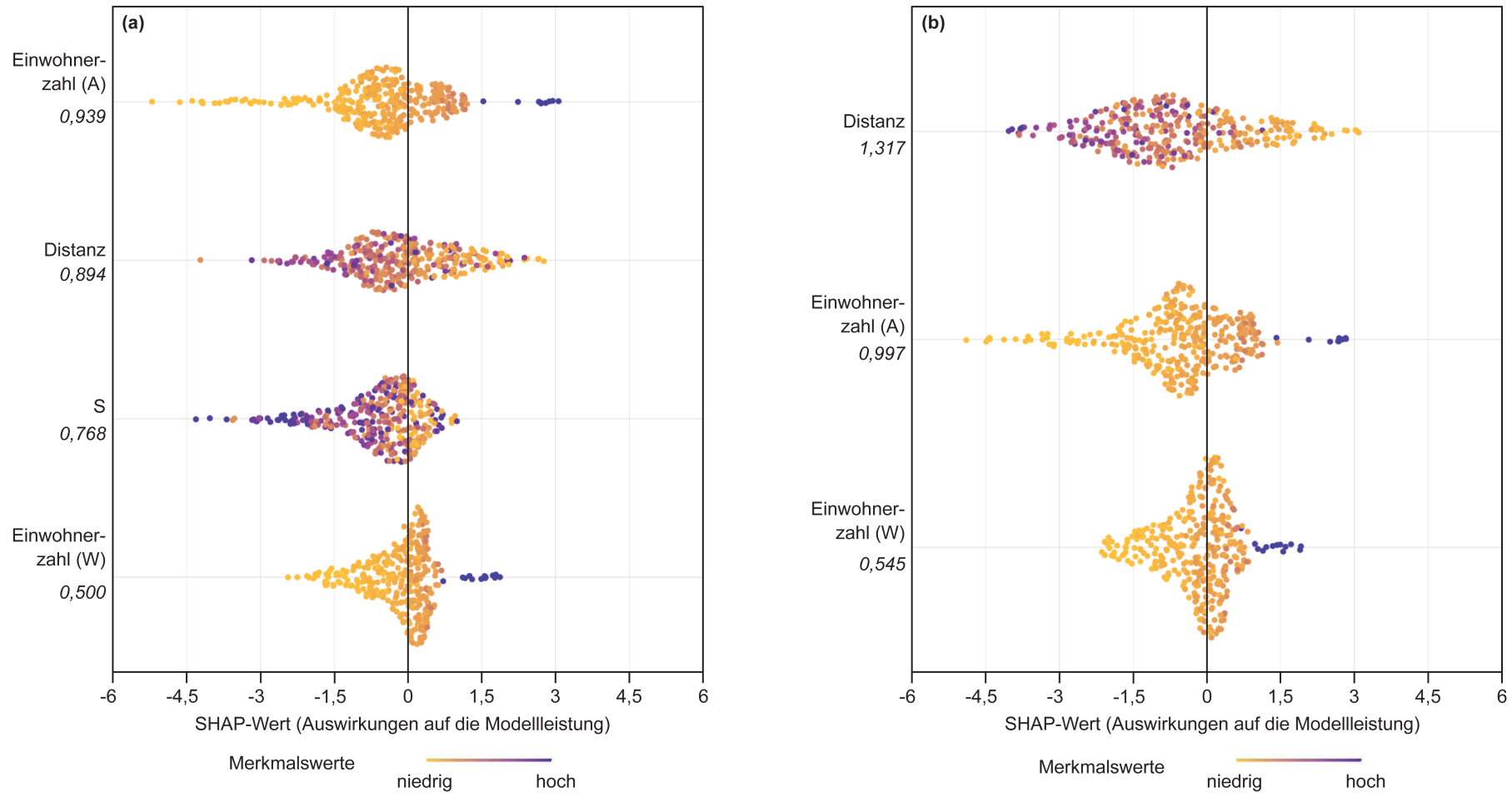


Abbildung 47: SHAP-Werte für den XGBoost-Regressor bei Pendlerbewegungen in Österreich (Grund- und Erweiterungsmodell). Quelle: Eigene Darstellung auf Grundlage von eigenen Berechnungen auf Basis von Statistik Austria (2011).

Schweiz

Die deskriptive Analyse der Pendlerströme zeigt, dass die Mobilitätsbewegungen in der Schweiz auf NUTS3-Ebene zwischen 5 und 51.374 Interaktionen schwanken. Die Verbindung mit den meisten Vernetzungen liegt zwischen dem Kanton Aargau und dem Kanton Zürich. Im Mittel pendeln ca. 1.746 Personen täglich in der Schweiz. Zudem liegt das erste Quantil bei 17, das zweite Quantil bei 87 und das dritte Quantil bei 1.118 Pendlern (Bundesamt für Statistik 2014).

Die Ergebnisse für das Gravitationsmodell in der Schweiz zeigen, dass Pendlerbewegungen in einer Größenordnung von ca. 3.000 bis 5.000 Personen besonders gute Anpassungen liefern (vgl. Abbildung 48 (a)). Ferner ist aus Abbildung 49 (a) zu entnehmen, dass Pendlerbewegungen in den ersten 90 Kilometern unterschätzt werden. Es folgt eine Überschätzung bis ca. 280 Kilometer. Darüber hinaus liefern die Evaluationsindikatoren folgende Werte (vgl. Abbildung 22): Sörensenindex: 0,43; Korrelationskoeffizient nach Pearson: 0,67; adjustierter R^2-Wert: 0,45; mittlerer absoluter Fehler: 1.140; Wurzel der mittleren quadratischen Abweichung: 3.030.

In Abbildung 48 (b) ist zu erkennen, dass das Strahlungsmodell gute Anpassungen im Bereich von 1.000 bis 10.000 Pendlern pro Tag liefert. Zudem zeigt sich in Abbildung 49 (b), dass die Wahrscheinlichkeit zu pendeln im direkten Umfeld von 80 Kilometern deutlich unterschätzt wird. Außerdem liefern die Evaluationsindikatoren (vgl. Abbildung 22) einen Sörensenindex von 0,51, einen Korrelationskoeffizient nach Pearson von 0,55 und einen adjustierten R^2-Wert von 0,3. Der mittlere absolute Fehler nimmt einen Wert von 850 und die Wurzel der mittleren quadratischen Abweichung einen Wert von 3.070 an.

Betrachtet man die Ergebnisse für das finite Strahlungsmodell ist zu erkennen, dass vor allem Pendlerströme von ca. 2.000, 10.000 und 20.000 besonders gut prognostiziert werden können (vgl. Abbildung 48 (c)). Darüber hinaus wird in Abbildung 49 (c) sichtbar, dass die Wahrscheinlichkeit zu pendeln zwischen 50 bis 90 Kilometern unterschätzt wird. Zudem beträgt der Sörensenindex 0,59, der Korrelationskoeffizient nach Pearson 0,72, der adjustierte R^2-Wert 0,51, der mittlere absolute Fehler 934 und die Wurzel der mittleren quadratischen Abweichung 2.801 (vgl. Abbildung 22).

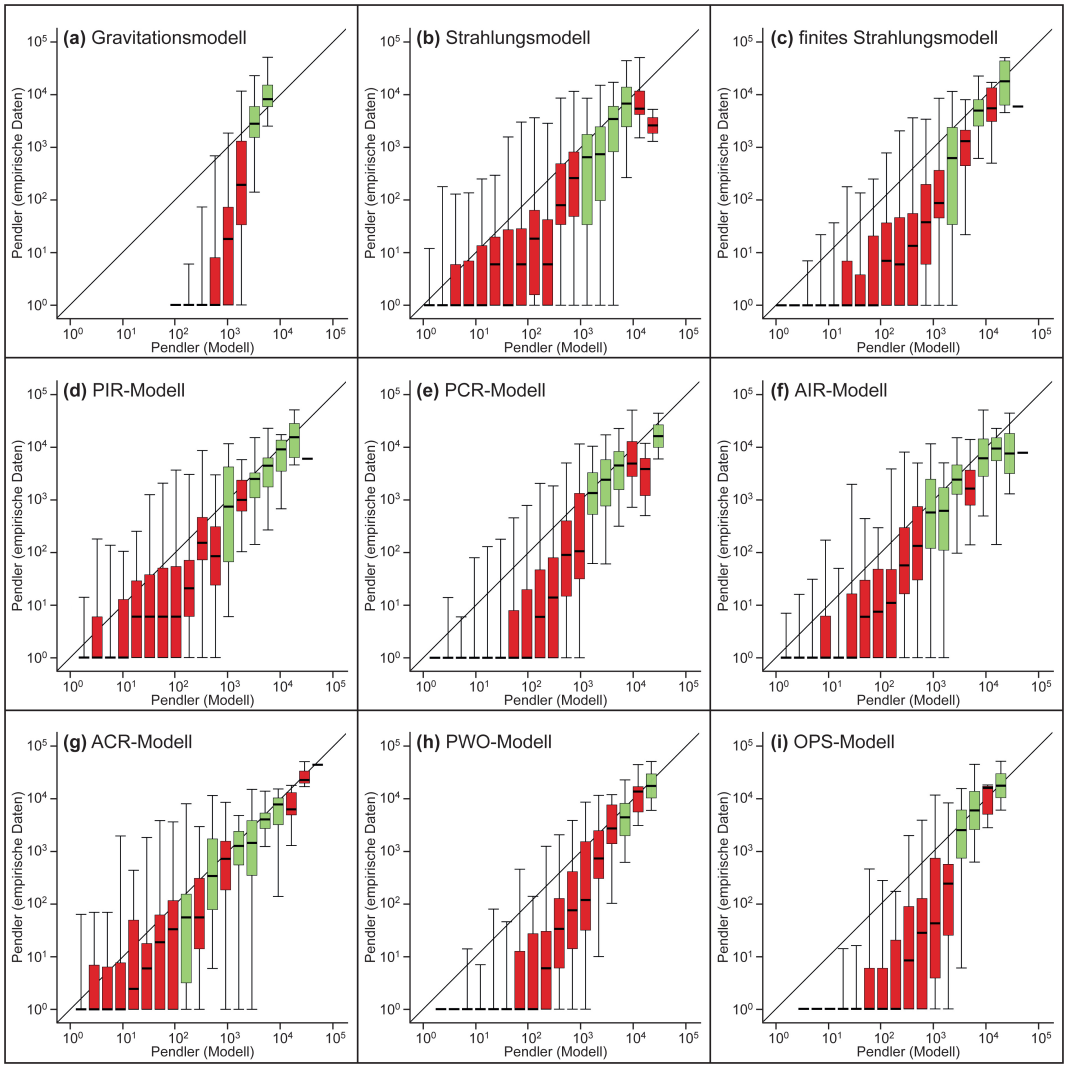


Abbildung 48: Abweichung zwischen den empirischen Daten und den Modellvorhersagen bezüglich der Pendlerbewegungen für die Schweiz. Quelle: Eigene Darstellung auf Grundlage von eigenen Berechnungen auf Basis von Bundesamt für Statistik (2014).

Unter den verallgemeinerten Strahlungsmodellen liefert das ACR-Modell die beste Anpassung. Besonders gute Schätzungen werden für etwa 200, 600 und zwischen 2.000 bis 10.000 Pendlern pro Tag erzielt (vgl. Abbildung 48 (g)). Betrachtet man die Wahrscheinlichkeit zu pendeln in Abbildung 49 (g), ist erkennbar, dass es im Bereich von 50 bis 90 Kilometern zu leichten Unterschätzungen kommt. Der Sörensenindex liegt bei 0,67, der Korrelationskoeffizient nach Pearson bei 0,84 und der adjustierte R^2-Wert bei 0,71. Zudem nimmt der mittlere absolute Fehler einen Wert von 596 und die Wurzel der mittleren quadratischen Abweichung einen Wert von 1.992 an (vgl. Abbildung 22). Die anderen drei Modelle liefern deutlich schlechtere Ergebnisse (AIR-Modell: Sörensenindex: 0,57; Korrelationskoeffizient nach Pearson: 0,64; adjustierter R^2-Wert: 0,41; mittlerer absolute Fehler: 788;

Wurzel der mittleren quadratischen Abweichung: 2.967 / PCR-Modell: Sörensenindex: 0,49; Korrelationskoeffizient nach Pearson: 0,55; adjustierter R^2-Wert: 0,3; mittlerer absoluter Fehler: 918; Wurzel der mittleren quadratischen Abweichung: 3.148 / PIR-Modell: Sörensenindex: 0,61; Korrelationskoeffizient nach Pearson: 0,7; adjustierter R^2-Wert: 0,48; mittlerer absoluter Fehler: 704; Wurzel der mittleren quadratischen Abweichung: 2.680).

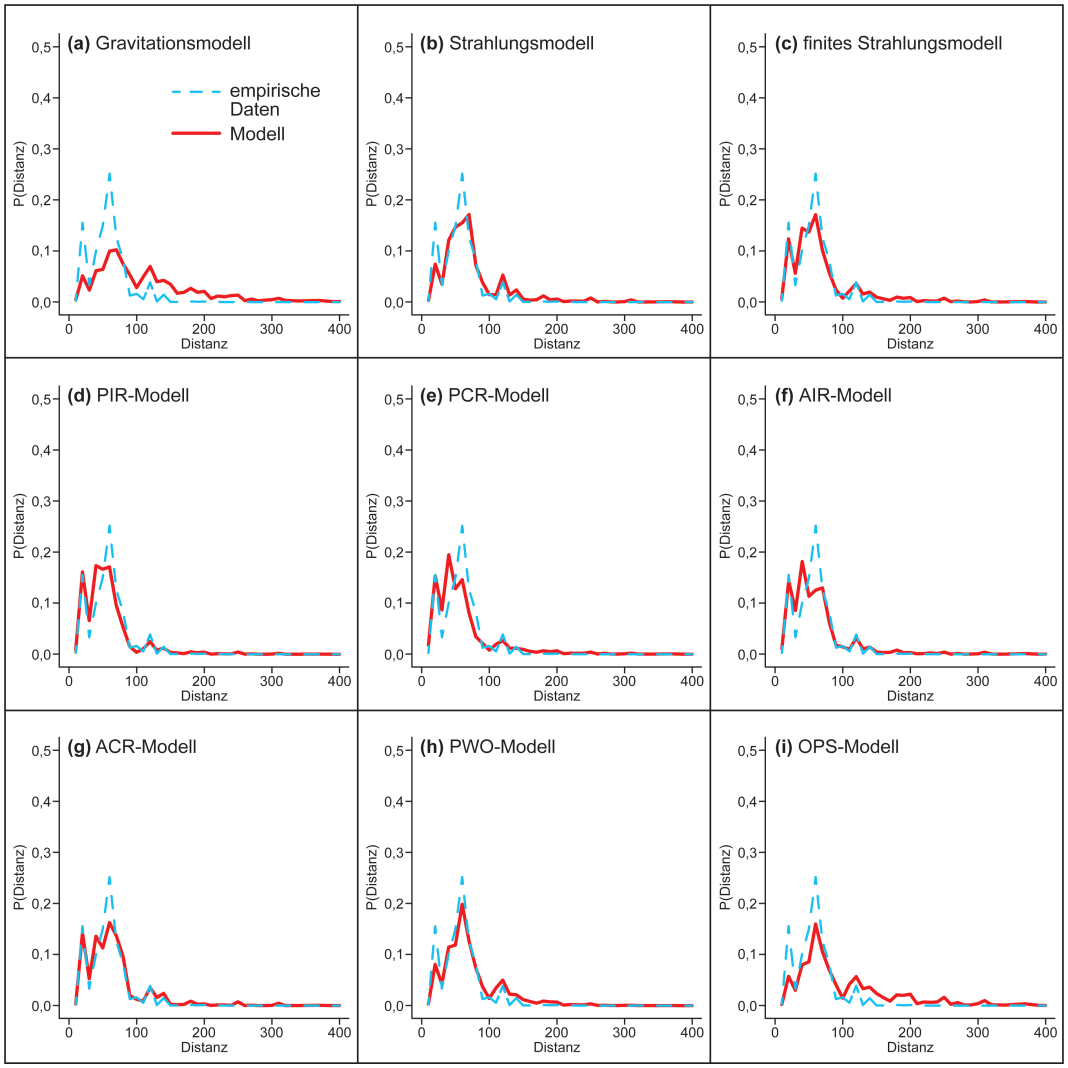


Abbildung 49: P(Distanz) im Verhältnis zur Entfernung (Pendlerbewegung in der Schweiz). Quelle: Eigene Darstellung auf Grundlage von eigenen Berechnungen auf Basis von Bundesamt für Statistik (2014).

Betrachtet man die Ergebnisse für das parameterfreie PWO-Modell wird deutlich, dass Pendlerbewegungen in der Größenordnung von ca. 10.000 und 30.000 Personen besonders gute Schätzungen liefern (vgl. Abbildung 48 (h)). Außerdem ist aus Abbildung 49 (h) zu entnehmen, dass die Abweichungen zwischen Modell und empirischen Daten marginal sind. Der Sörensenindex liefert einen

Wert von 0,58, der Korrelationskoeffizient nach Pearson einen Wert von 0,75 und der adjustierte R^2-Wert beträgt 0,56. Zudem ist erkennbar, dass der mittlere absolute Fehler bei 755 und die Wurzel der mittleren quadratischen Abweichung bei 2.459 liegen (vgl. Abbildung 22).

Für das OPS-Modell ist in Abbildung 48 (i) sichtbar, dass Mobilitätsflüsse von ca. 5.000, 10.000 und 30.000 Personen besonders gut prognostiziert werden können. Außerdem wird im direkten Umkreis von ca. 30 Kilometern sowie zwischen 50 bis 70 Kilometern die Wahrscheinlichkeit für eine Pendlerfahrt unterschätzt. Für größere Distanzen liegt eine leichte Überschätzung vor (vgl. Abbildung 49 (i)). Zudem nimmt der Sörensenindex einen Wert von 0,5, der Korrelationskoeffizient nach Pearson einen Wert von 0,73 und der adjustierte R^2-Wert einen Wert von 0,53 an. Der mittlere absolute Fehler liegt bei 898 und die Wurzel der mittleren quadratischen Abweichung bei 2.632 (vgl. Abbildung 22).

Für das Grundmodell (vgl. Abbildung 51 (b)) des XGBoost-Algorithmus ist festzustellen, dass große Distanzen einen negativen Einfluss auf die Prognose haben. Ferner wirken sich große Einwohnerzahlen (sowohl am Arbeits- als auch Wohnort) positiv auf die Vorhersage aus. Die Evaluationsindikatoren zeigen (vgl. Abbildung 24), dass der Sörensenindex in 64 % der Simulationen bei 0,75 einzuordnen ist (vgl. Abbildung 50 (a)). Ferner liegt der Korrelationskoeffizient nach Pearson bei 0,71 und der adjustierte R^2-Wert bei 0,83. Zudem beträgt der mittlere absolute Fehler 1.007 und die Wurzel der mittleren quadratischen Abweichung 2.961.

Abbildung 51 (a) ist zu entnehmen, dass bei der Erweiterung des Grundmodells des XGBoost-Algorithmus besonders große Distanzen einen negativen Einfluss haben. Zudem wirken sich Standorte, bei denen eine hohe Anzahl an Einwohnern im Zwischenraum leben, negativ auf die Prognose aus, während hohe Einwohnerzahlen am Wohn- und Arbeitsort einen positiven Einfluss haben. Der Sörensenindex beträgt in 67,2 % der Simulationen 0,64 (vgl. Abbildung 50 (b)), der Korrelationskoeffizient nach Pearson 0,71 und der adjustierte R^2-Wert 0,51. Zudem liegt der mittlere absolute Fehler bei 1.006 und die Wurzel der mittleren quadratischen Abweichung bei 2.962 (vgl. Abbildung 24).

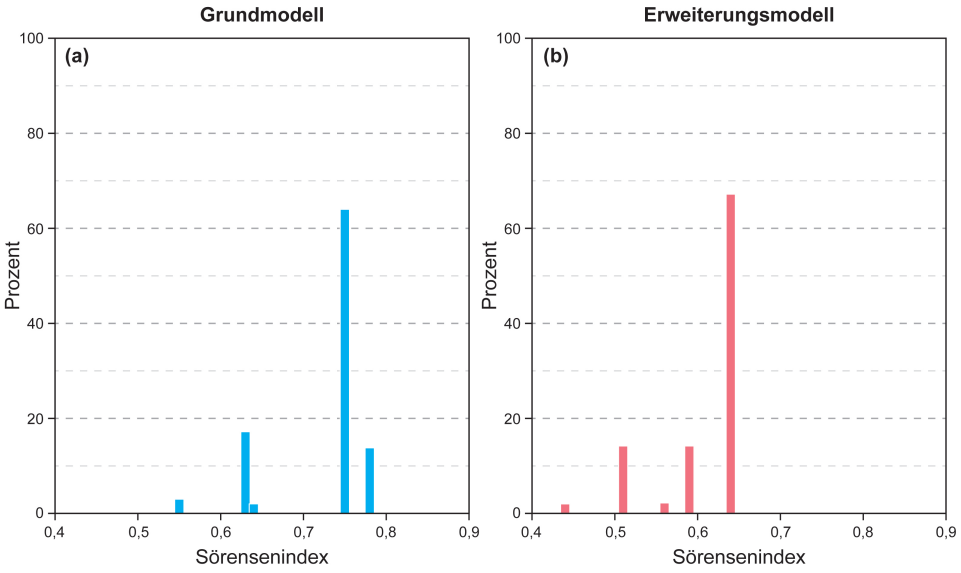


Abbildung 50: Ergebnisse der Simulation für die Schweiz. Quelle: Eigene Darstellung auf Grundlage von eigenen Berechnungen auf Basis von Bundesamt für Statistik (2014).

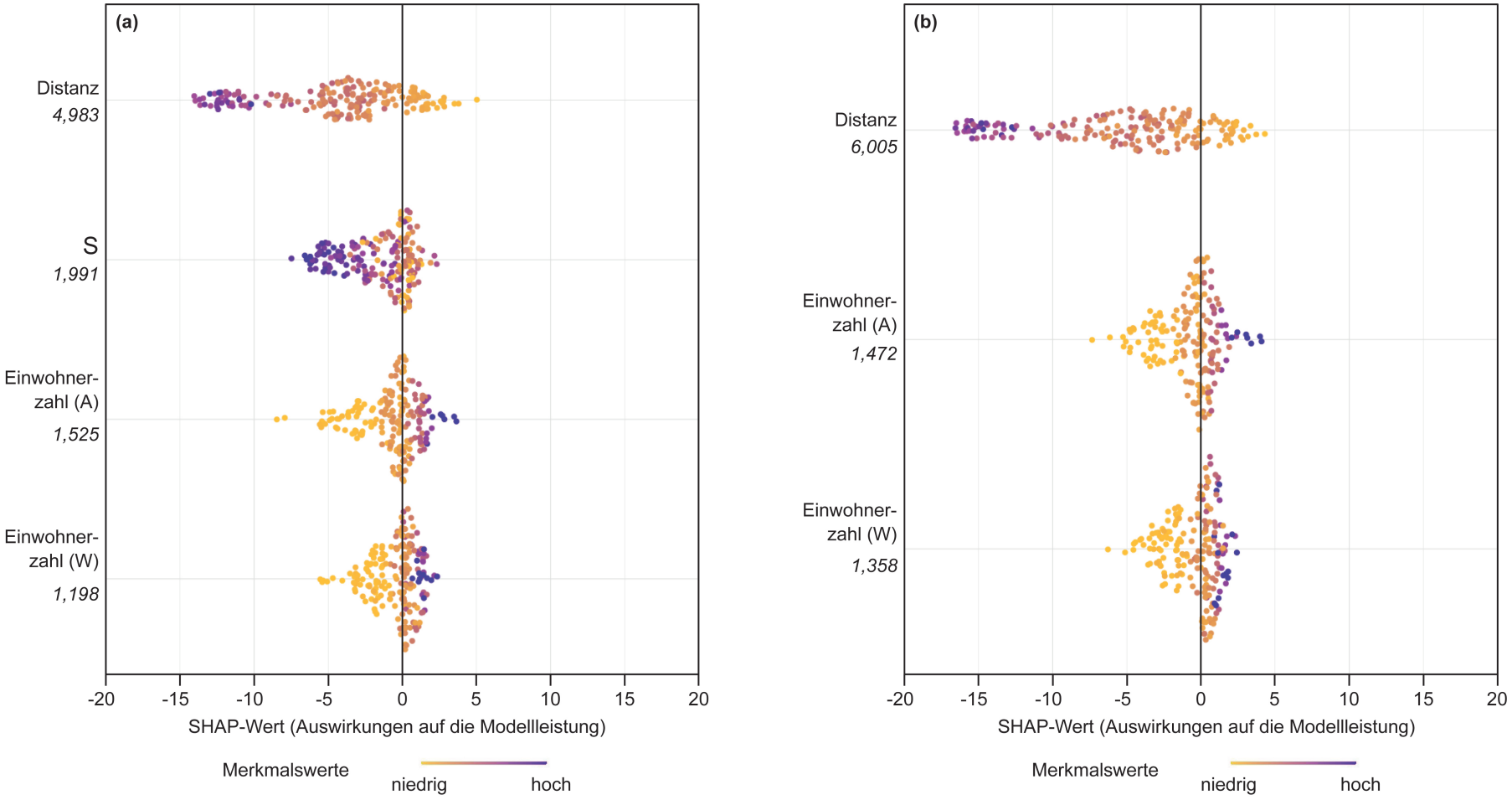


Abbildung 51: SHAP-Werte für den XGBoost-Regressor bei Pendlerbewegungen in der Schweiz (Grund- und Erweiterungsmodell). Quelle: Eigene Darstellung auf Grundlage von eigenen Berechnungen auf Basis von Bundesamt für Statistik (2014).

Dänemark

Betrachtet man die Pendlerbewegungen (ohne Binnenpendler) für Dänemark, so fällt auf, dass die Interaktionen auf NUTS3-Ebene zwischen 152 und 73.851 schwanken. Die stärkste Pendlerbewegung liegt für die Provinzen Byen København und Københavns omegn vor, während die schwächste Verbindung zwischen Province Østsjælland und Vestjylland existiert. Im Mittel beträgt der Penderfluss 6.033 Personen und das erste, zweite und dritte Quantil nimmt Werte von 455, 1.314 und 5.256 an (Danmarks Statistik 2011).

In Abbildung 52 (a) wird für das Gravitationsmodell deutlich, dass vor allem Pendlerbewegungen von ca. 20.000 Pendler pro Tag besonders gute Schätzungen liefern. Bezüglich der Distanz ist zu erkennen, dass für das direkte Umfeld (bis zu 20 Kilometer) leichte Unterschätzungen vorliegen (vgl. Abbildung 53 (a)). Der Sörensenindex liegt bei 0,64, der Korrelationskoeffizient nach Pearson bei 0,90 und der adjustierte R^2-Wert bei 0,82. Zudem beträgt der mittlere absolute Fehler 4.570 und die Wurzel der mittleren quadratischen Abweichung 7.451.

Betrachtet man die Ergebnisse des Strahlungsmodells, so wird deutlich, dass die Pendlerströme zwischen ca. 300 und 600 sowie zwischen 2.000 und 10.000 Interaktionen besonders gute Schätzungen liefern (vgl. Abbildung 52 (a)). Zudem ist in Abbildung 53 (b) erkennbar, dass die Wahrscheinlichkeit zu pendeln bis ca. 90 Kilometer unterschätzt wird. Zwischen 90 und 120 Kilometern tritt eine starke Überschätzung auf. Für größere Distanzen sind hingegen kaum Abweichnungen zu erkennen. Der Sörensenindex beträgt 0,51, der Korrelationskoeffizient nach Pearson 0,47, der adjustierte R^2-Wert 0,21, der mittlere absolute Fehler 5.598 und die Wurzel der mittleren quadratischen Abweichung 11.346.

Für die Pendlerströme des finiten Strahlungsmodells ist nicht festzustellen, dass die Linie y=x zwischen dem 9. und 91. Perzentil in der jeweiligen Klasse liegt (vgl. Abbildung 52 (c)). Die Wahrscheinlichkeit zu pendeln in Abhängigkeit von der Distanz zeigt zudem Unterschätzungen im direkten Umfeld und zwischen 150 und 200 Kilometern (vgl. Abbildung 53 (c)). Der Sörensenindex liegt bei 0,61, der Korrelationskoeffizient nach Pearson bei 0,91 und der adjustierte R^2-Werte bei 0,83. Der mittlere absolute Fehler beträgt 7.159 und die Wurzel der mittleren quadratischen Abweichung 11.885 (vgl. Abbildung 22).

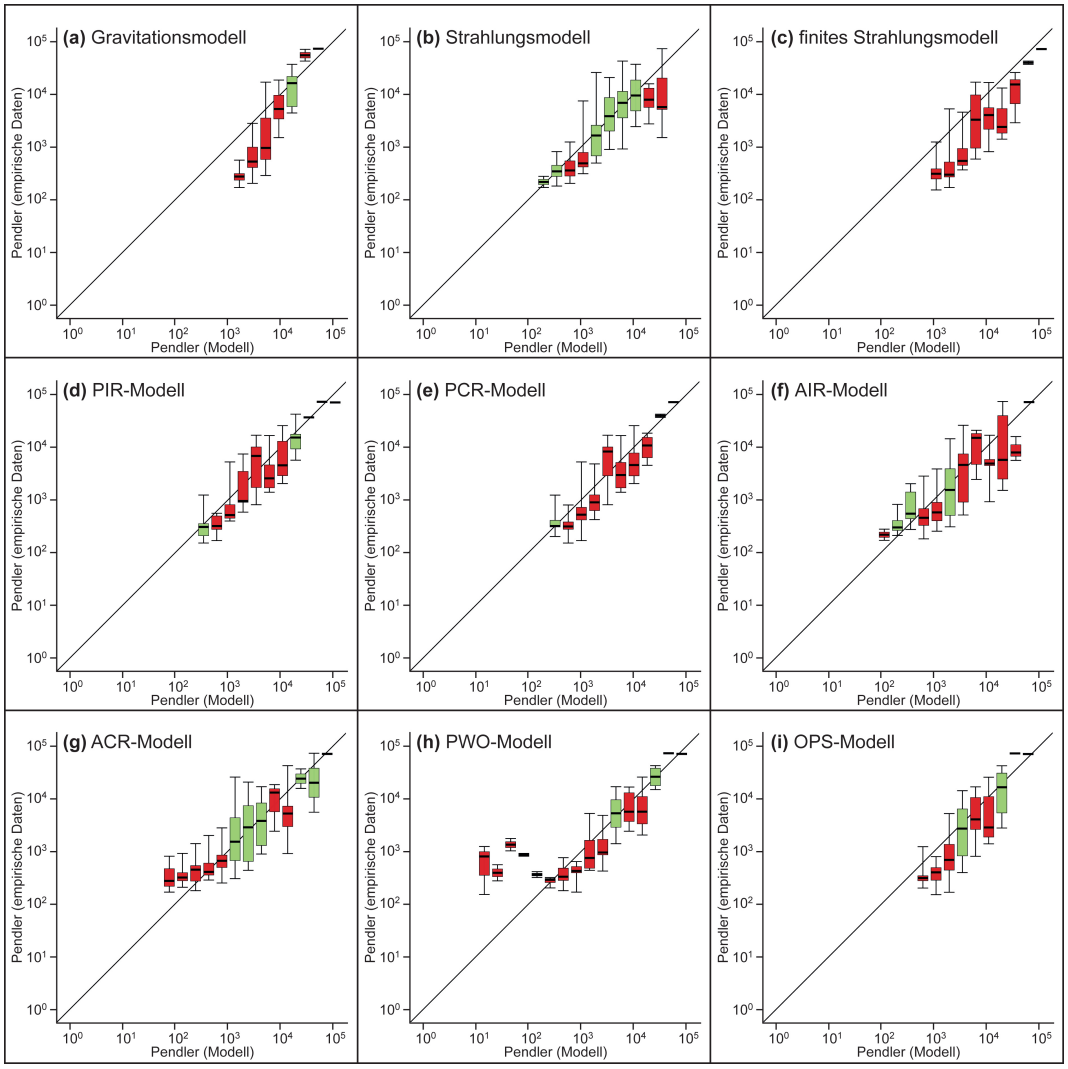


Abbildung 52: Abweichung zwischen den empirischen Daten und den Modellvorhersagen bezüglich der Pendlerbewegungen für Dänemark. Quelle: Eigene Darstellung auf Grundlage von eigenen Berechnungen auf Basis von Danmarks Statistik (2011).

Bei den verallgemeinerten Strahlungsmodellen liefert das PCR-Modell besonders gute Anpassungen. Es wird deutlich, dass vor allem Pendlerbewegungen in der Größenordnung von ca. 300 Personen besonders gut geschätzt werden (vgl. Abbildung 52 (e)). Ferner ist Abbildung 53 (e) zu entnehmen, dass zwischen 50 und 100 sowie zwischen 130 und 200 Kilometern die empirischen Daten unterschätzt werden. Zudem liefert der Sörensenindex mit einem Wert von 0,76 einen weiteren Hinweis, dass wenige Abweichungen zwischen empirischen und prognostizierten Daten existieren. Ähnliche Beobachtungen (vgl. Abbildung 22) können auch für den Korrelationskoeffizienten nach Pearson (0,93) und den adjustierten R^2-Wert (0,86) gemacht werden. Der mittlere absolute Fehler liegt bei 2.916 und die Wurzel der mittleren quadratischen Abweichung bei 4.616. Auch das PIR-Modell erbringt vergleichbare Ergebnisse (Sörensenindex: 0,73;

Korrelationskoeffizient nach Pearson: 0,91; adjustierter R^2-Wert: 0,82; mittlerer absoluter Fehler: 3.233; Wurzel der mittleren quadratischen Abweichung: 5.335). Das AIR- und das ACR-Modell liefern hingegen eine deutlich schlechtere Anpassung an die empirischen Daten (AIR-Modell: Sörensenindex: 0,59; Korrelationskoeffizient nach Pearson 0,67; adjustierter R^2-Wert: 0,44; mittlerer absoluter Fehler: 4.954; Wurzel der mittleren quadratischen Abweichung: 9.540 / ACR-Modell: Sörensenindex: 0,62; Korrelationskoeffizient nach Pearson: 0,73; adjustierter R^2-Wert: 0,52; mittlerer absoluter Fehler: 4.644; Wurzel der mittleren quadratischen Abweichung: 8.824).

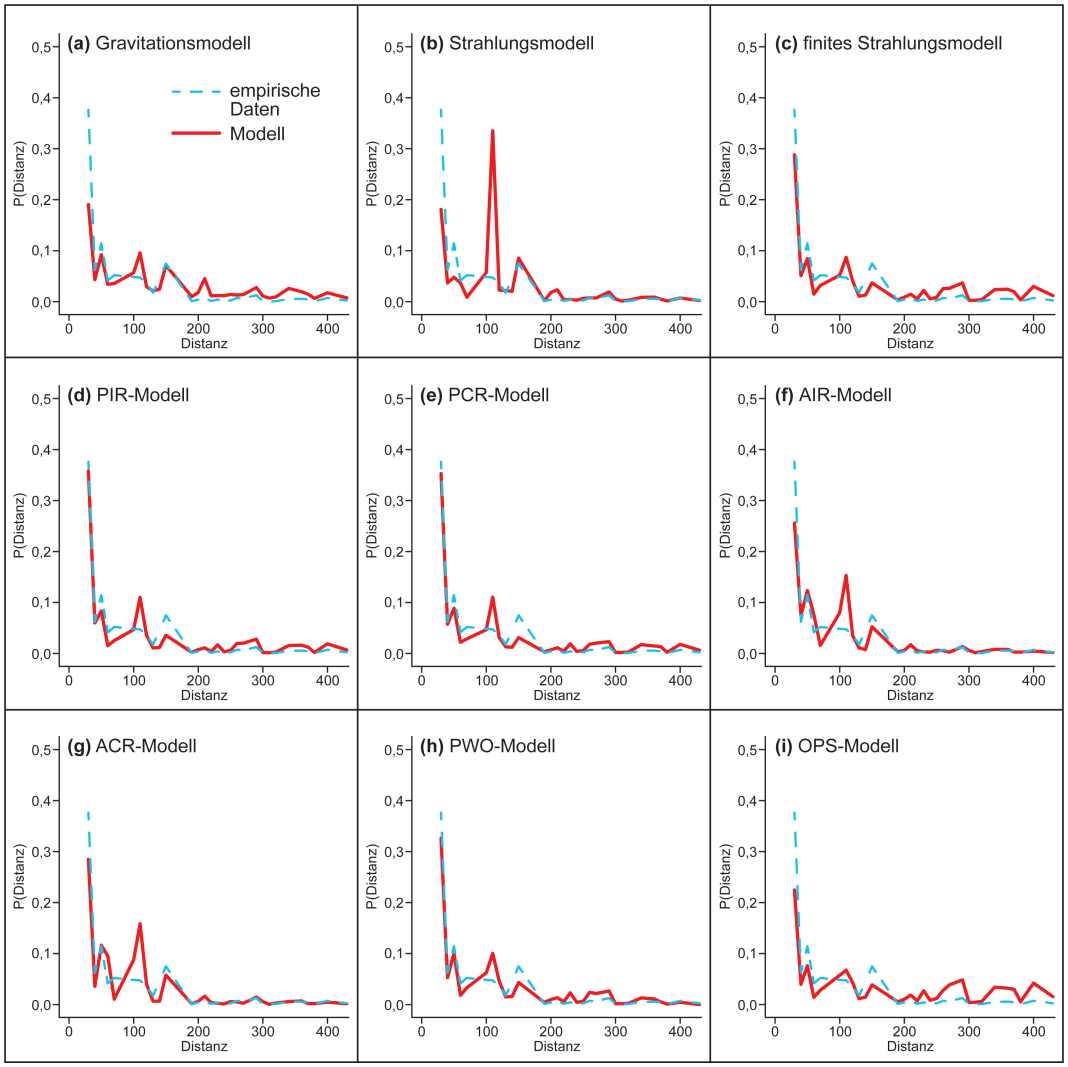


Abbildung 53: P(Distanz) im Verhältnis zur Entfernung (Pendlerbewegung in Dänemark). Quelle: Eigene Darstellung auf Grundlage von eigenen Berechnungen auf Basis von Danmarks Statistik (2011).

Für das PWO- und OPS-Modell liefern Pendlerströme von ca. 6.000 und 30.000 Personen besonders gute Vorhersagen (vgl. Abbildung 52 (h) und (i)). Des Weiteren ist in Abbildung 53 (h) und (i) zu erkennen, dass die Abweichungen für sehr kleine Distanzen beim OPS-Modell deutlich stärker sind. Außerdem treten Unterschätzungen für Distanzen von ca. 60 bis 80 und 130 bis 200 Kilometern auf. Der Sörensenindex nimmt Werte von 0,77 (PWO-Modell) und 0,65 (OPS-Modell) an. Ferner konnte festgestellt werden, dass beide Modelle einen starken Zusammenhang (PWO-Modell: 0,94; OPS-Modell: 0,85) aufweisen und die adjustierten R^2-Werte bei 0,88 und 0,71 liegen. Außerdem wurde für das PWO-Modell mit 2.757 Pendlern der kleinste mittlere absolute Fehler gemessen. Für das OPS-Modell liegt dieser hingegen bei 4.279. Die Wurzel der mittleren quadratischen Abweichung beträgt für das PWO-Modell 4.476 und für das OPS-Modell 7.211 (vgl. Abbildung 22).

In Abbildung 55 (b) ist für das Grundmodell des XGBoost-Algorithmus zu erkennen, dass besonders große Distanzen sowie NUTS3-Regionen mit wenig Einwohnern am Wohn- und Arbeitsort einen negativen Einfluss auf die Prognose haben. Der Sörensenindex liegt in 68 % aller Simulationen bei 0,66 (vgl. Abbildung 54 (a)), der Korrelationskoeffizient nach Pearson bei 0,78 und der adjustierte R^2-Wert bei 0,25. Zudem beträgt der mittlere absolute Fehler 2.164 und die Wurzel der mittleren quadratischen Abweichung 5.179 (vgl. Abbildung 24).

Für die Erweiterung dieses Grundmodells sind ähnliche Ergebnisse zu erkennen. Wie in Abbildung 55 (a) führen auch in diesem Fall große Entfernungen zwischen den Standorten sowie kleine Einwohnzahlen zur Verschlechterung des Modells. Zudem ist keine eindeutige Wirkrichtung für die Einwohnerzahl zwischen den Standorten zu erkennnen. Der Sörensenindex liegt in 74 % aller Simulationen bei 0,61 (vgl. Abbildung 54 (b)), der Korrelationskoeffizient nach Pearson bei 0,52, der adjustierte R^2-Wert bei 0,24, der mittlere absolute Fehler bei 2.378 und die Wurzel der mittleren quadratischen Abweichung bei 6.491 (vgl. Abbildung 24).

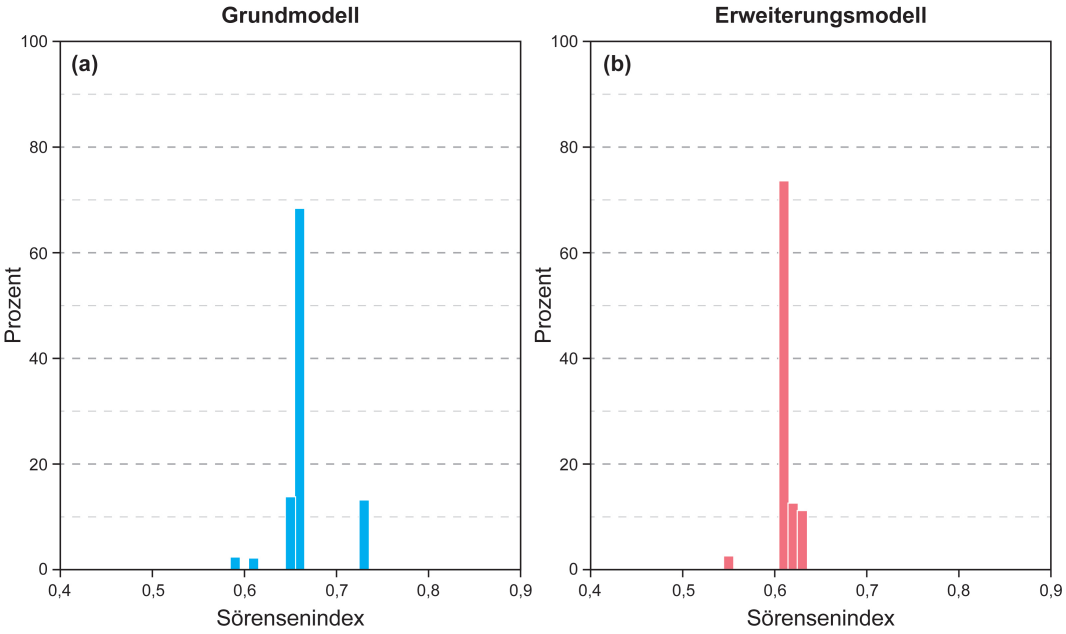


Abbildung 54: Ergebnisse der Simulation für Dänemark. Quelle: Eigene Darstellung auf Grundlage von eigenen Berechnungen auf Basis von Danmarks Statistik (2011).

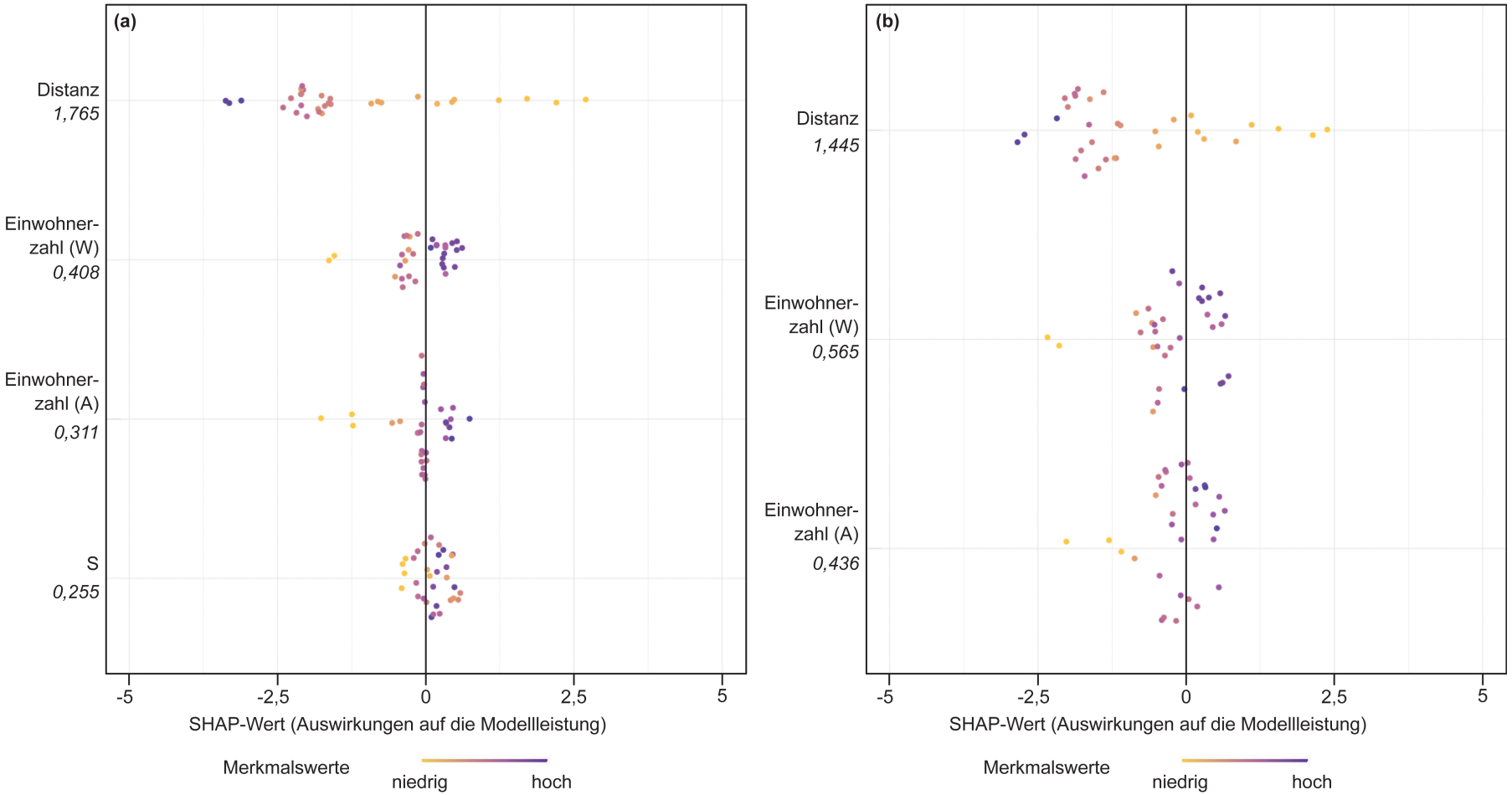


Abbildung 55: SHAP-Werte für den XGBoost-Regressor bei Pendlerbewegungen in Dänemark (Grund- und Erweiterungsmodell). Quelle: Eigene Darstellung auf Grundlage von eigenen Berechnungen auf der Basis von Danmarks Statistik (2011).

Finnland

Für Finnland wird bei der deskriptiven Analyse deutlich, dass die Strecke Satakunta - Kainuu mit 2 Interaktionen den geringsten Mobilitätsfluss aufweist. Der größte Pendlerstrom ist für Kanta-Häme - Helsinki-Uusimaa mit 11.636 Personen festzustellen. Im Mittel pendeln 491 Erwerbstätige. Das erste und dritte Quantil liegt bei 39 und 382. Zudem beträgt der Median 112 Pendler (Statistics Finland 2011).

In Abbildung 56 (a) ist für Finnland zu erkennen, dass das Gravitationsmodell insbesondere für Pendlerströme in einer Größenordnung von ca. 10.000 Personen besonders gute Ergebnisse liefert. Die Übersicht in Abbildung 57 (a) zeigt zudem, dass die Wahrscheinlichkeit zu pendeln bis ca. 250 Kilometer leicht unterschätzt wird und ab ca. 380 Kilometer leichte Überschätzungen vorliegen. Der Sörensenindex liegt bei 0,64, der Korrelationskoeffizient nach Pearson bei 0,89, der adjustierte R^2-Wert bei 0,79, der mittlere absolute Fehler bei 478 und die Wurzel der mittleren quadratischen Abweichung bei 623 (vgl. Abbildung 22).

Für das Strahlungsmodell bzw. finite Strahlungsmodell werden die Pendlerbewegungen bei 300 bzw. zwischen 30 bis 60 Interaktionen optimal geschätzt (vgl. Abbildung 56 (b) und (c)). Darüber hinaus macht Abbildung 57 (b) und (c) deutlich, dass die Wahrscheinlichkeit zu pendeln bis ca. 200 Kilometer überschätzt wird. Der Sörensenindex liegt beim Strahlungsmodell bei 0,59 und beim finiten Strahlungsmodell bei 0,63. Des Weiteren nimmt der Korrelationskoeffizient nach Pearson Werte von 0,89 (Strahlungsmodell) bzw. 0,82 (finites Strahlungsmodell) an. Der adjustierte R^2-Wert beträgt 0,41 bzw. 0,68. Außerdem liefert die Auswertung für den mittleren absoluten Fehler 376 (Strahlungsmodell) bzw. 440 (finites Strahlungsmodell) und für die Wurzel der mittleren quadratischen Abweichung 897 (Strahlungsmodell) und 966 (finites Strahlunsmodell) (vgl. Abbildung 22).

Ferner zeigt der Vergleich mit den verallgemeinerten Strahlungsmodellen, dass vor allem das ACR-Modell gute Anpassungen liefert. In Abbildung 57 (g) ist zu erkennen, dass Pendlerströme von etwa 20 bis 30, 1.800 sowie ab 5.700 Personen besonders gute Anpassungen erzielen. Außerdem ist in Abbildung 57 (g) ersichtlich, dass vor allem kleine Distanzen leicht unterschätzt werden, die Unterschiede jedoch sehr gering sind. Zudem zeigt die Analyse der Evaluationsindikatoren, dass der Sörensenindex bei 0,73, der Korrelationskoeffizient nach Pearson bei 0,85, der adjustierte R^2-Wert bei 0,72, der

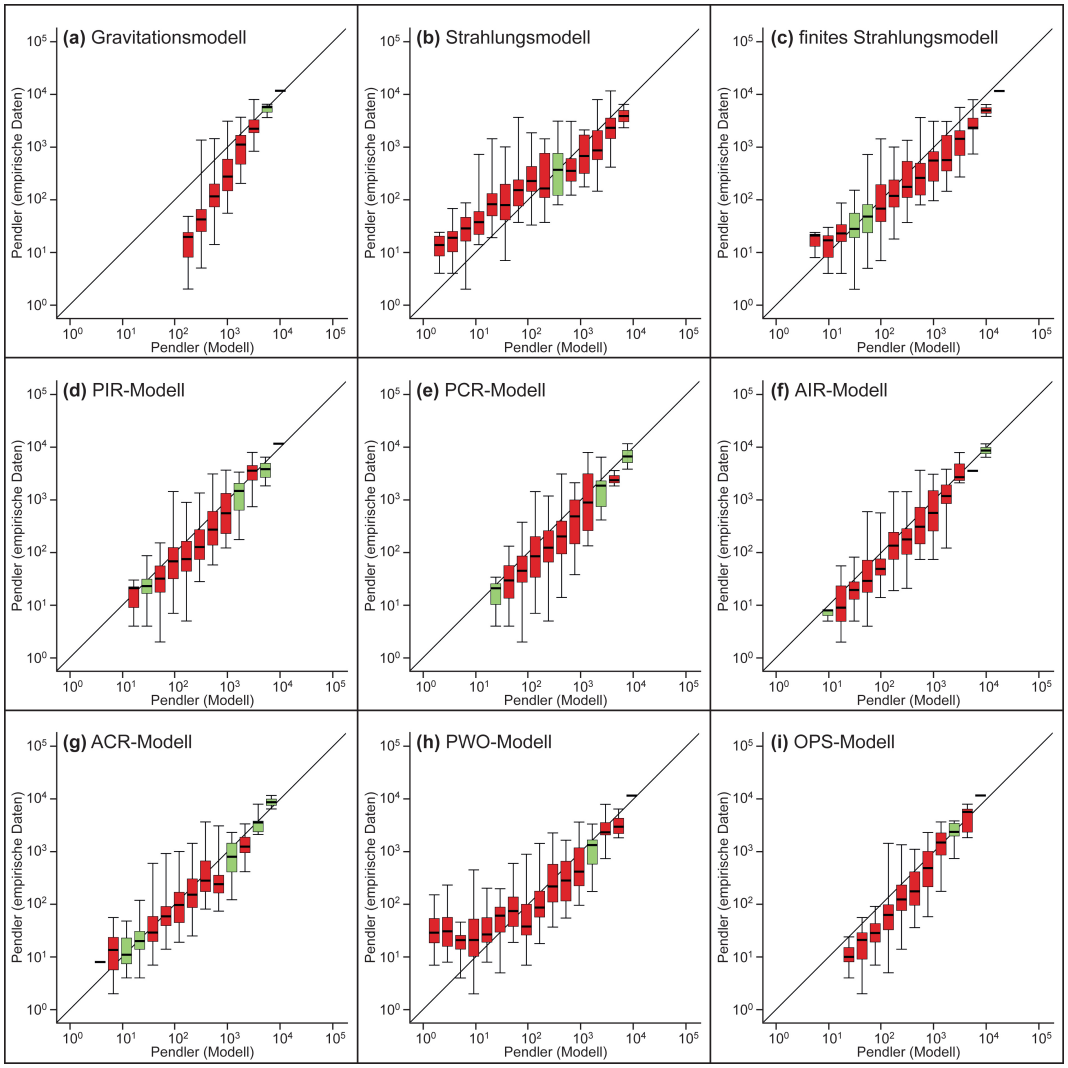


Abbildung 56: Abweichung zwischen den empirischen Daten und den Modellvorhersagen bezüglich der Pendlerbewegungen für Finnland. Quelle: Eigene Darstellung auf Grundlage von eigenen Berechnungen auf Basis von Statistics Finland (2011).

mittlere absolute Fehler bei 270 und die Wurzel der mittleren quadratischen Abweichung g bei 597 liegen (vgl. Abbildung 22). Ähnlich gute Ergebnisse liefert das AIR-Modell (Sörensenindex: 0,70; Korrelationskoeffizient nach Pearson: 0,85; adjustierter R^2-Wert: 0,72; mittlerer absoluter Fehler: 298; Wurzel der mittleren quadratischen Abweichung: 603). Die Unterschiede zwischen den empirischen Daten und den Vorhersagen mittels des PIR- und das PCR-Modells sind etwas größer (PIR-Modell: Sörensenindex: 0,70; Korrelationskoeffizient nach Pearson: 0,85; adjustiert R^2-Wert: 0,72; mittlerer absoluter Fehler: 295; Wurzel der mittleren quadratischen Abweichung: 599 / PCR-Modell: Sörensenindex: 0,63; Korrelationskoeffizient nach Pearson: 0,72; adjustierter R^2-Wert: 0,52; mittlerer absoluter Fehler: 362; Wurzel der mittleren quadratischen Abweichung: 779).

Abbildung 56 (h) zeigt für das PWO-Modell, dass Pendelbewegungen in einer Größenordnung von ca. 1.800 Personen besonders gute Ergebnisse liefern. Darüber hinaus werden kleine Distanzen leicht überschätzt (vgl. Abbildung 57 (h)). Aus den Ergebnisse des nicht parametrisierten Modells (vgl. Abbildung 22) ist ferner zu entnehmen, dass der Sörensenindex bei 0,70, der Korrelationskoeffizient nach Pearson bei 0,85, der adjustierte R^2-Wert bei 0,73, der mittlere absolute Fehler bei 294 sowie die Wurzel der mittleren quadratischen Abweichung bei 590 liegt.

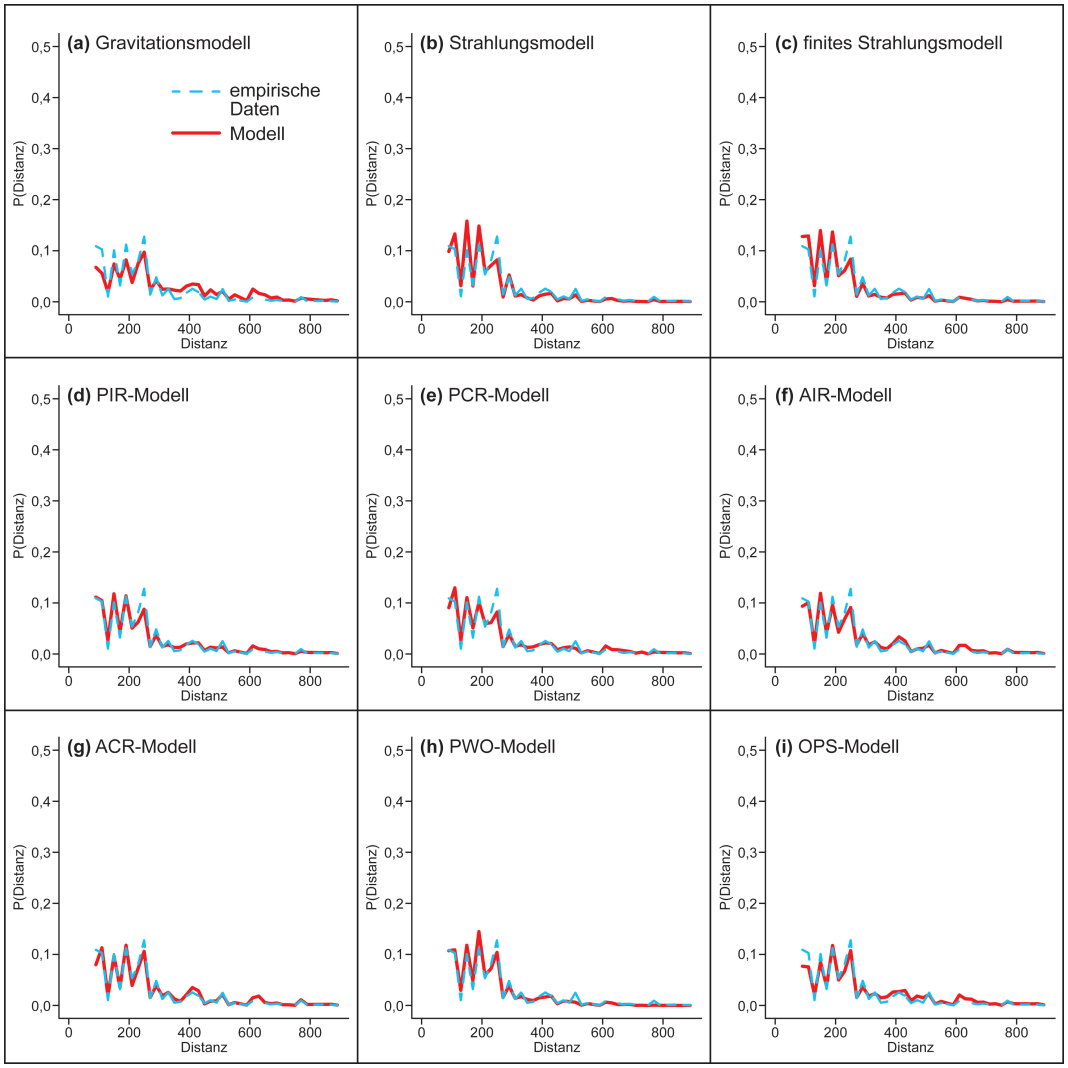


Abbildung 57: P(Distanz) im Verhältnis zur Entfernung (Pendlerbewegung in Finnland). Quelle: Eigene Darstellung auf Grundlage von eigenen Berechnungen auf Basis von Statistics Finland (2011).

In Abbildung 56 (i) ist bezüglich des OPS-Modells zu erkennen, dass Pendlerbewegungen in einer Größenordnung von ca. 3.000 Interaktionen besonders gute Anpassungen liefern. Zudem existieren für kleine Distanzen lediglich leichte Differenzen zwischen den empirischen Daten und den Modellvorhersagen (vgl.

Abbildung 57 (i)). Der Sörensenindex beträgt 0,71, der Korrelationskoeffizient nach Pearson 0,91, der adjustierte R^2-Wert 0,82, der mittlere absolute Fehler 288 und die Wurzel der mittleren quadratischen Abweichung 534 (vgl. Abbildung 22).

Für das Grundmodell (vgl. Abbildung 59 (b)) des XGBoost-Algorithmus wird deutlich, dass große Entfernungen die Prognose negativ beeinflussen. Ferner sind positive Enflüsse bei großen Einwohnerzahlen (sowohl am Arbeits- als auch Wohnort) festzustellen. Aus Abbildung 58 (a) ist zu entnehmen, dass der Sörensenindex in 68 % der Simulationen bei 0,72 liegt. Des Weiteren beträgt der Korrelationskoeffizient nach Pearson 0,76, der adjustierte R^2-Wert 0,68, der mittlere absolute Fehler 161 und die Wurzel der mittleren quadratischen Abweichung 306 (vgl. Abbildung 24).

Außerdem ist für die Erweiterung des Grundmodells des XGBoost-Algorithmus (vgl. Abbildung 59 (a)) festzustellen, dass auch in diesem Fall die Distanz einen sehr hohen Stellenwert einnimmt. Große Entfernungen beeinfussen die Prognose negativ. Zudem wirken sich große Einwohnerzahl am Wohn- und Arbeitsort auf die Vorhersage positiv aus. Bei der Einwohnerzahl zwischen Ziel- und Quellregionen kann jedoch keine eindeutige Wirkrichtung ausgemacht werden. Der Sörensenindex liefert in 0,95 % aller Simulationen einen Wert von 0,68 (vgl. Abbildung 58 (b)). Darüber hinaus liegt der Korrelationskoeffizient nach Pearson bei 0,87, der adjustierte R^2-Wert bei 0,75, der mittlere absolute Fehler bei 354 und die Wurzel der mittleren quadratischen Abweichung bei 1.007 (vgl. Abbildung 24).

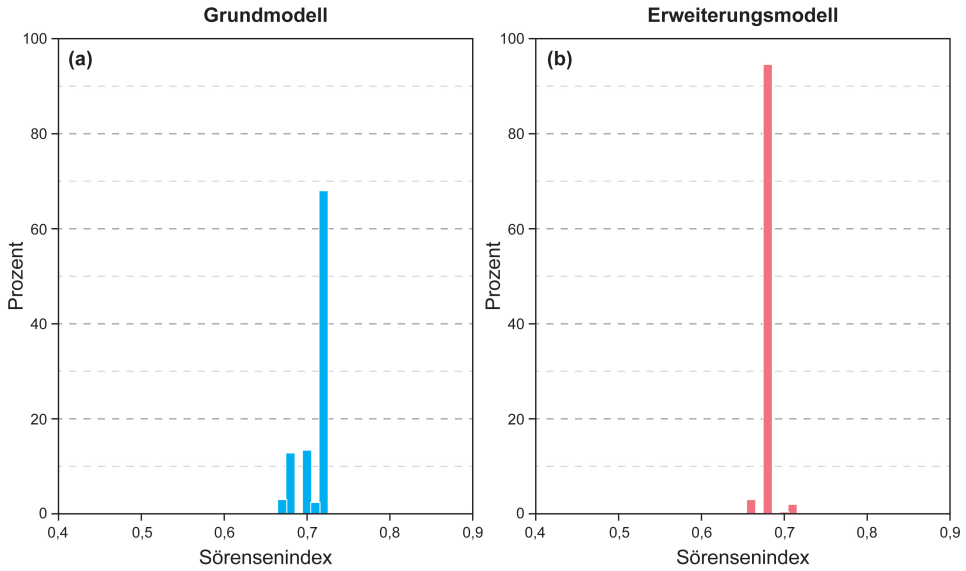


Abbildung 58: Ergebnisse der Simulation für Finnland. Quelle: Eigene Darstellung auf Grundlage von eigenen Berechnungen auf Basis von Statistics Finland (2011).

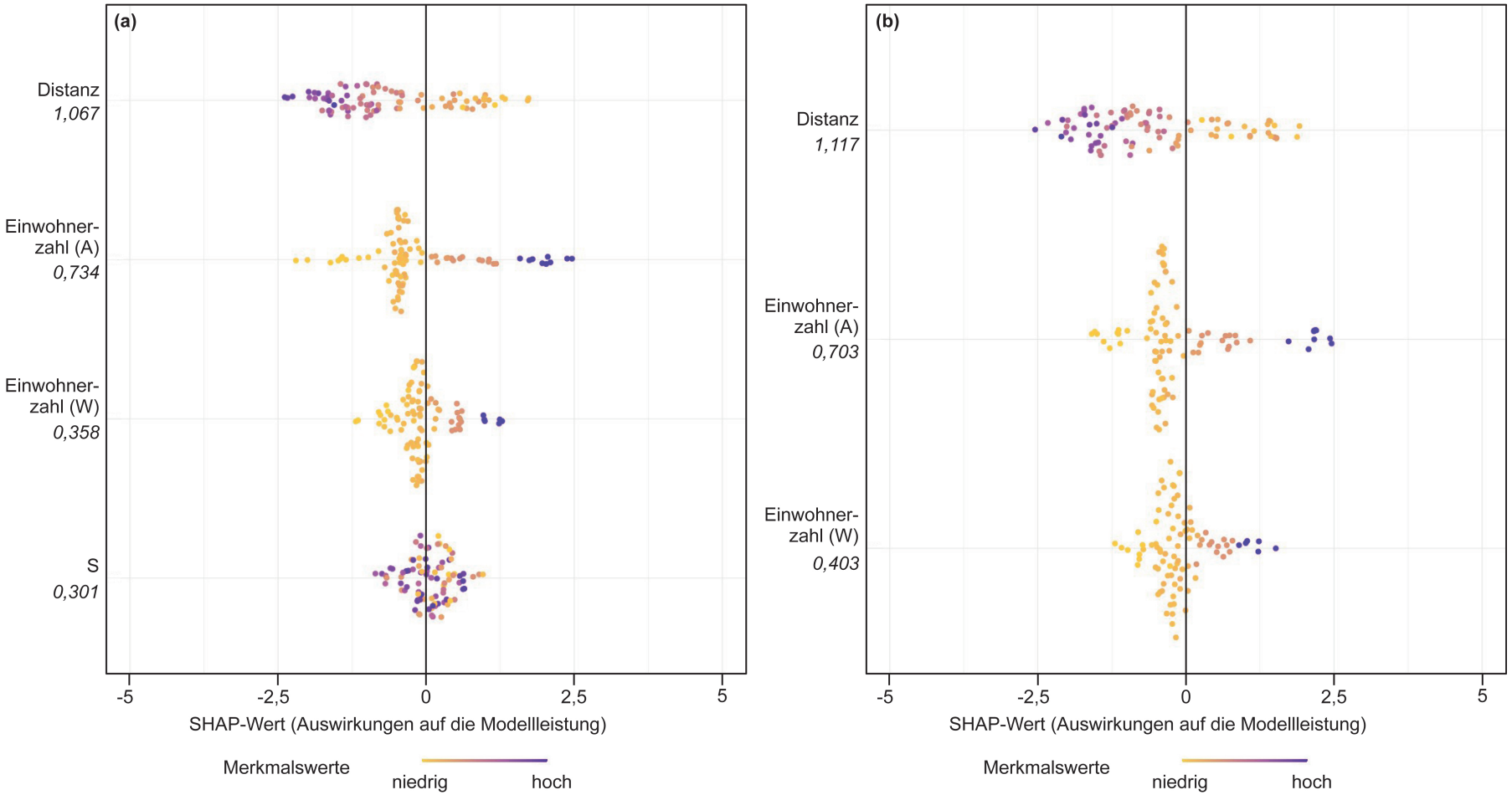


Abbildung 59: SHAP-Werte für den XGBoost-Regressor bei Pendlerbewegungen in Finnland (Grund- und Erweiterungsmodell). Quelle: Eigene Darstellung auf Grundlage von eigenen Berechnungen auf Basis von Statistics Finland (2011).

Norwegen

Die deskriptive Betrachtung der Pendlerströme in Norwegen zeigt, dass sie zwischen 6 und 100.718 Personen schwanken. Die Pendlerstrecke mit der größten Nachfrage ist die Verbindung zwischen Oslo und Akershus. Die Pendlerverbindung mit den wenigsten Interaktionen ist die Strecke zwischen Nord-Trøndelag und Aust-Agder. Die durchschnittliche Interaktion zwischen zwei NUTS3-Regionen beträgt 1.084 Pendler. Der Median liegt bei 66 Personen, das erste und dritte Quantil bei 66 bzw. 489 Pendlern (Statistics Norway 2011).

Die Ergebnisse des Gravitationsmodells liefern für Norwegen besonders gute Anpassungen bei Pendlerströmen von ca. 30 Personen (vgl. Abbildung 60 (a)). Im direkten Umfeld wird die Wahrscheinlichkeit zu pendeln deutlich unterschätzt (vgl. Abbildung 61 (a)). Der Sörensenindex beträgt 0,54, der Korrelationskoeffizient nach Pearson 0,77, der adjustierte R^2-Wert 0,6, der mittlere absolute Fehler 865 und die Wurzel der mittleren quadratischen Abweichung 5.179 (vgl. Abbildung 22).

In Abbildung 60 (b) wird deutlich, dass in Norwegen mithilfe des Strahlungsmodelles gute Schätzungen für Pendlerströme von einer Größe von 100 bis 200 Personen sowie von 3.000 Personen erzielt werden können. Die Betrachtung der Wahrscheinlichkeit in einen Ort zu pendeln in Abhängigkeit von der Distanz zeigt, dass kleine Distanzen bis ca. 30 Kilometer unterschätzt werden. Im Anschluss folgt bis ca. 90 Kilometer eine leichte Überschätzung (vgl. Abbildung 61 (b)). Zudem wird in Abbildung 22 deutlich, dass der Sörensenindex bei 0,48, der Korrelationskoeffizient nach Pearson bei 0,63 und der adjustierte R^2-Wert bei 0,39 liegt. Der mittlere absolute Fehler beträgt 1.094 und die Wurzel der mittleren quadratischen Abweichung 4.884 (vgl. Abbildung 22).

Überdies ist in Abbildung 60 (c) für das finite Strahlungsmodell erkennbar, dass keine optimalen Schätzungen der Pendlerströme vorliegen. Ursache ist, dass die Linie y=x nicht zwischen dem 9. und 91. Perzentil in der jeweiligen Klasse liegt. Zudem macht Abbildung 61 (c) deutlich, dass die Wahrscheinlichkeit des Auftretens einer Pendlerbewegung lediglich marginale Abweichungen zwischen empirischen Daten und Modell aufweist. Der Sörensenindex, der Korrelationskoeffizient nach Pearson und der adjustierte R^2-Wert nehmen Werte von 0,76, 0,97 und 0,94 an. Außerdem liegt der mittlere absolute Fehler bei 566 und die Wurzel der mittleren quadratischen Abweichung bei 1.471 (vgl. Abbildung 22).

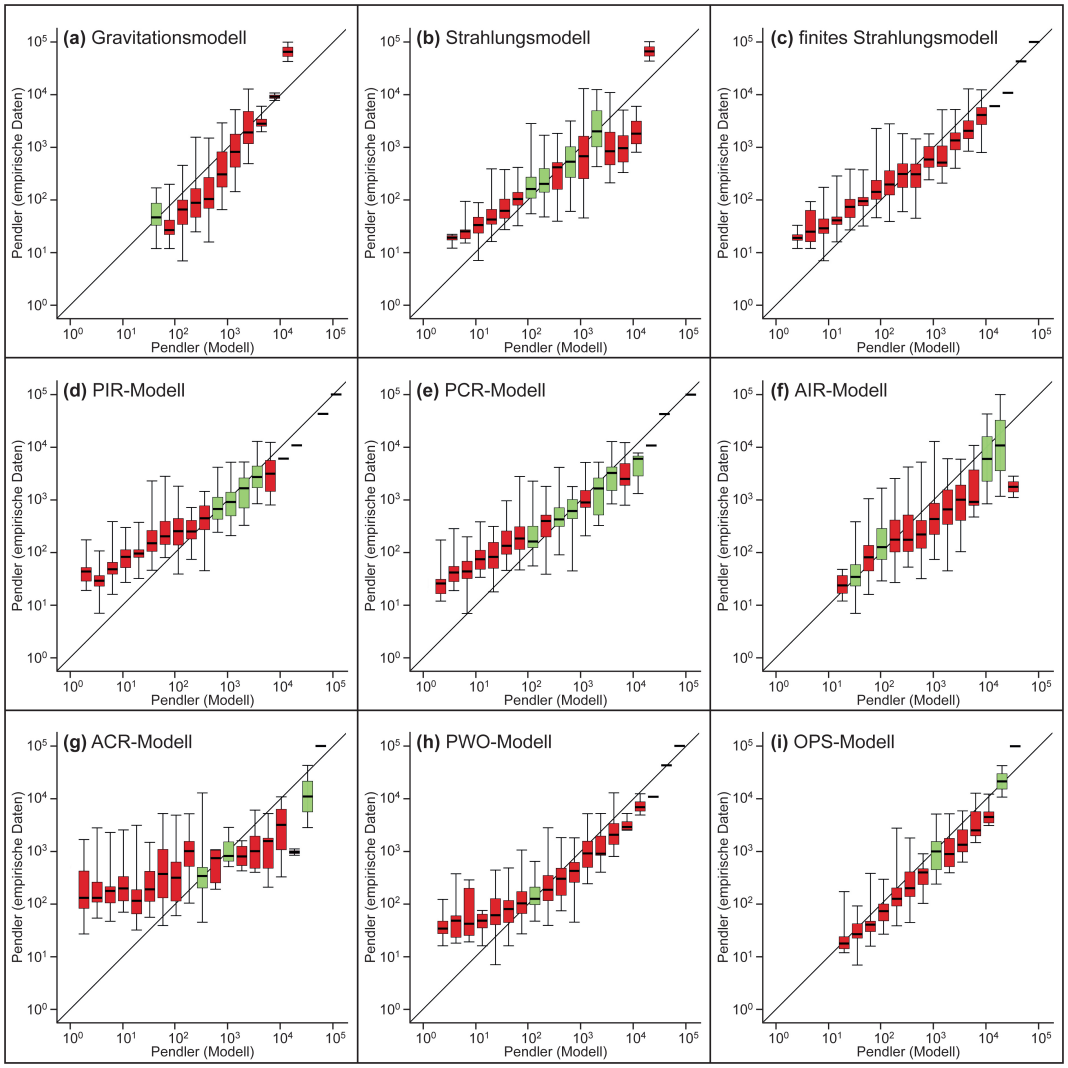


Abbildung 60: Abweichung zwischen den empirischen Daten und den Modellvorhersagen bezüglich der Pendlerbewegungen für Norwegen. Quelle: Eigene Darstellung auf Grundlage von eigenen Berechnungen auf Basis von Statistics Norway (2011).

Beim Vergleich der verallgemeinerten Strahlungsmodelle wird deutlich, dass insbesondere das PIR-Modell sehr gute Ergebnisse liefert. Aus Abbildung 60 (d) ist erkennbar, dass Mobilitätsströme zwischen ca. 1.000 und 5.600 Pendlern sehr gut geschätzt werden. Die größten Abweichungen (Überschätzungen) treten zwischen 30 und 90 Kilometern auf (vgl. Abbildung 61). Zusätzlich liefern die Evaluationsindikatoren den Hinweis, dass die besten Schätzungen der Pendlerflüsse vorliegen (Sörensenindex: 0,78; Korrelationskoeffizient nach Pearson: 0,98; adjustierter R^2-Wert: 0,96; mittlerer absoluter Fehler: 484; Wurzel der mittleren quadratischen Abweichung: 1.289). Ähnlich gute Ergebnisse (vgl. Abbildung 22) sind für das PCR-Modell zu beobachten (Sörensenindex: 0,75; Korrelationskoeffizient nach Pearson: 0,97; adjustierter R^2-Wert: 0,94; mittlerer absoluter Fehler: 535; Wurzel der mittleren quadratischen Abweichung:

1.483). Die Evaluationsindikatoren für das AIR-Modell (Sörensenindex: 0,42; Korrelationskoeffizient nach Pearson: 0,46; adjustierter R^2-Wert: 0,21; mittlerer absoluter Fehler: 1.256; Wurzel der mittleren quadratischen Abweichung: 5.371) und das ACR-Modell (Sörensenindex: 0,53; Korrelationskoeffizient nach Pearson: 0,84; adjustierter R^2-Wert: 0,71; mittlerer absoluter Fehler: 1.018; Wurzel der mittleren quadratischen Abweichung: 3.258) weisen hingegen auf eine deutlich schlechtere Prognoseleistung hin.

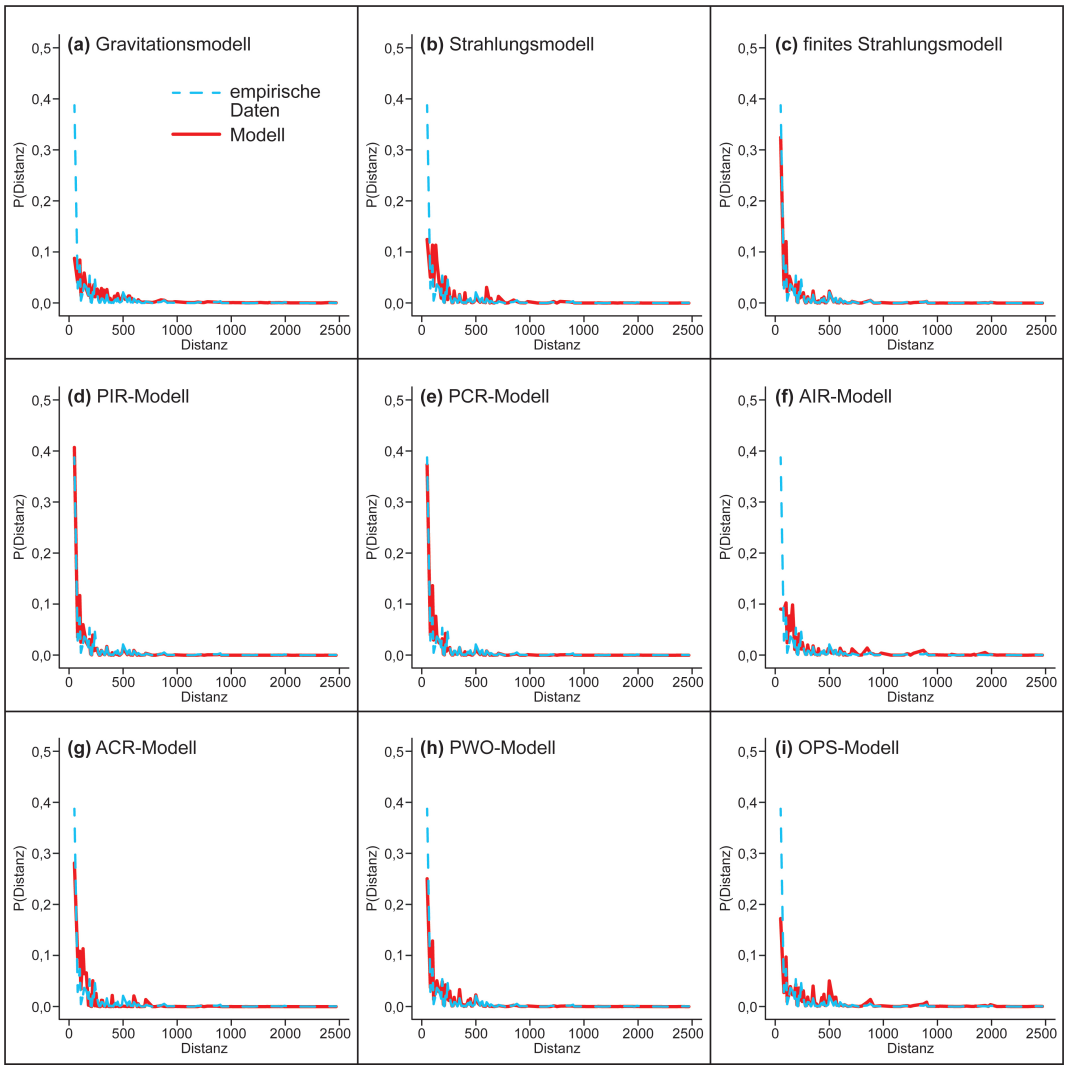


Abbildung 61: P(Distanz) im Verhältnis zur Entfernung (Pendlerbewegung in Norwegen). Quelle: Eigene Darstellung auf Grundlage von eigenen Berechnungen auf Basis von Statistics Norway (2011).

Für das parameterfreie PWO-Modell wird deutlich, dass Pendlerströme in der Größenordnung von ca. 100 Personen besonders gute Anpassungen liefern (vgl. Abbildung 60 (h)). Daneben existieren marginale Abweichungen zwischen den empirischen Daten und den Schätzungen bezüglich der Pendlerwahrscheinlichkeit und der Distanz (vgl. Abbildung 61 (h)). Die Evaluationsindikatoren lie-

fern einen Sörensenindex von 0,71, einen Korrelationskoeffizient nach Pearson von 0,94 und einen adjustierten R^2-Wert von 0,88. Der mittlere absolute Fehler liegt bei 629 und die Wurzel der mittleren quadratischen Abweichung bei 2.557 (vgl. Abbildung 22).

Außerdem wird in Abbildung 60 (i) für das OPS-Modell deutlich, dass Pendlerbewegungen von ca. 2.000 sowie 30.000 Personen besonders gute Prognoseergebnisse liefern. Zudem sind für Pendlerdistanzen im Umkreis von ca. 10 Kilometern Unterschätzungen festzustellen (vgl. Abbildung 61 (i)). Die Evaluationsindikatoren (vgl. Abbildung 22) ergeben für die Schätzungen des OPS-Modells Werte von 0,66 (Sörensenindex), 0,88 (Korrelationskoeffizient nach Pearson), 0,78 (adjustierter R^2-Wert), 746 (mittlerer absoluter Fehler) und 3.589 (Wurzel der mittleren quadratischen Abweichung).

Betrachtet man die Ergebnisse für das Grundmodell des XGBoost-Algorithmus (vgl. Abbildung 63 (b)) ist erkennbar, dass für Norwegen große Entfernungen zwischen Wohn- und Arbeitsort einen negativen Einfluss auf die Prognose haben. Zudem wirken sich vor allem einwohnerstarke Regionen positiv auf die Schätzung aus. Die Evaluationsergebnisse der Indikatoren zeigen, dass der Sörensenindex in 66 % der Simulationen bei 0,85 (vgl. Abbildung 62 (a)), der Korrelationskoeffizient nach Pearson bei 0,98 und der adjustierte R^2-Wert bei 0,97 liegt. Außerdem beträgt der mittlere absolute Fehler 353 und die Wurzel der mittleren quadratischen Abweichung 1.321 (vgl. Abbildung 24).

Aus Abbildung 63 (a) ist für das Erweiterungsmodell zu erkennen, dass Regionen mit vielen Einwohnern am Wohnort den stärksten positiven Einfluss auf die Prognose haben. Zudem liefern hohe Entfernungen einen negativen Einfluss auf die Prognoseleistung, während die Einwohnerzahl am Arbeitsplatz sich positiv auswirkt. Der Einfluss der Einwohnerzahl zwischen den beiden Standorten kann nicht abschließend geklärt werden. Des Weiteren ist in Abbildung 24 sichtbar, dass der Sörensenindex bei 70 % der Simulationen 0,65 (vgl. Abbildung 62 (b)), der Korrelationskoeffizient nach Pearson bei 0,97, der adjustierte R^2-Wert bei 0,94, der mittlere absolute Fehler bei 1.042 und die Wurzel der mittleren quadratischen Abweichung bei 7.771 liegt.

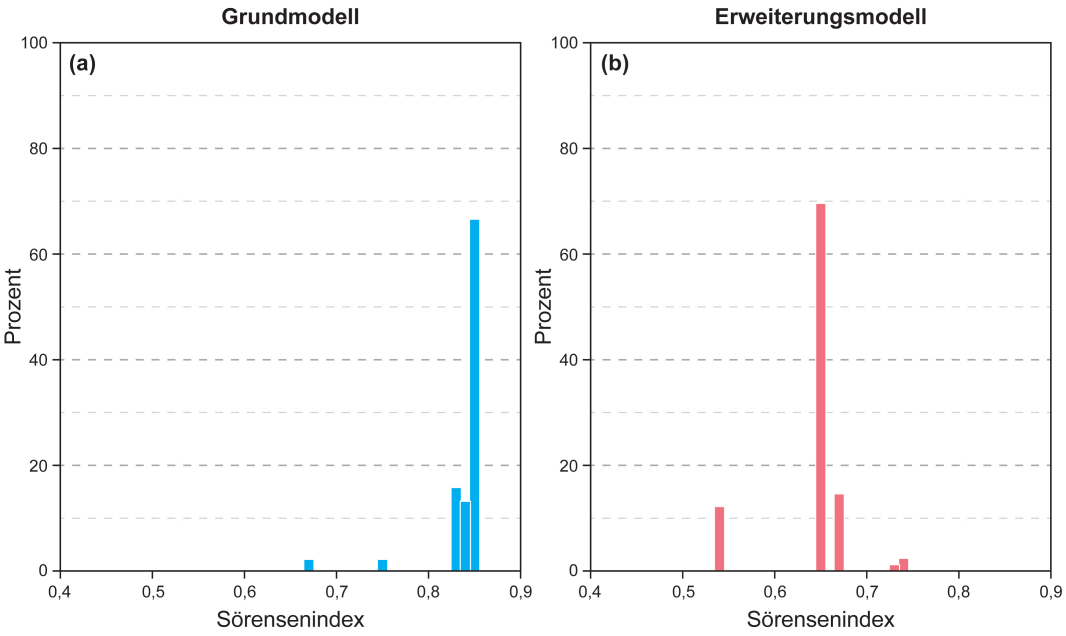


Abbildung 62: Ergebnisse der Simulation für Norwegen. Quelle: Eigene Darstellung auf Grundlage von eigenen Berechnungen auf Basis von Statistics Norway (2011).

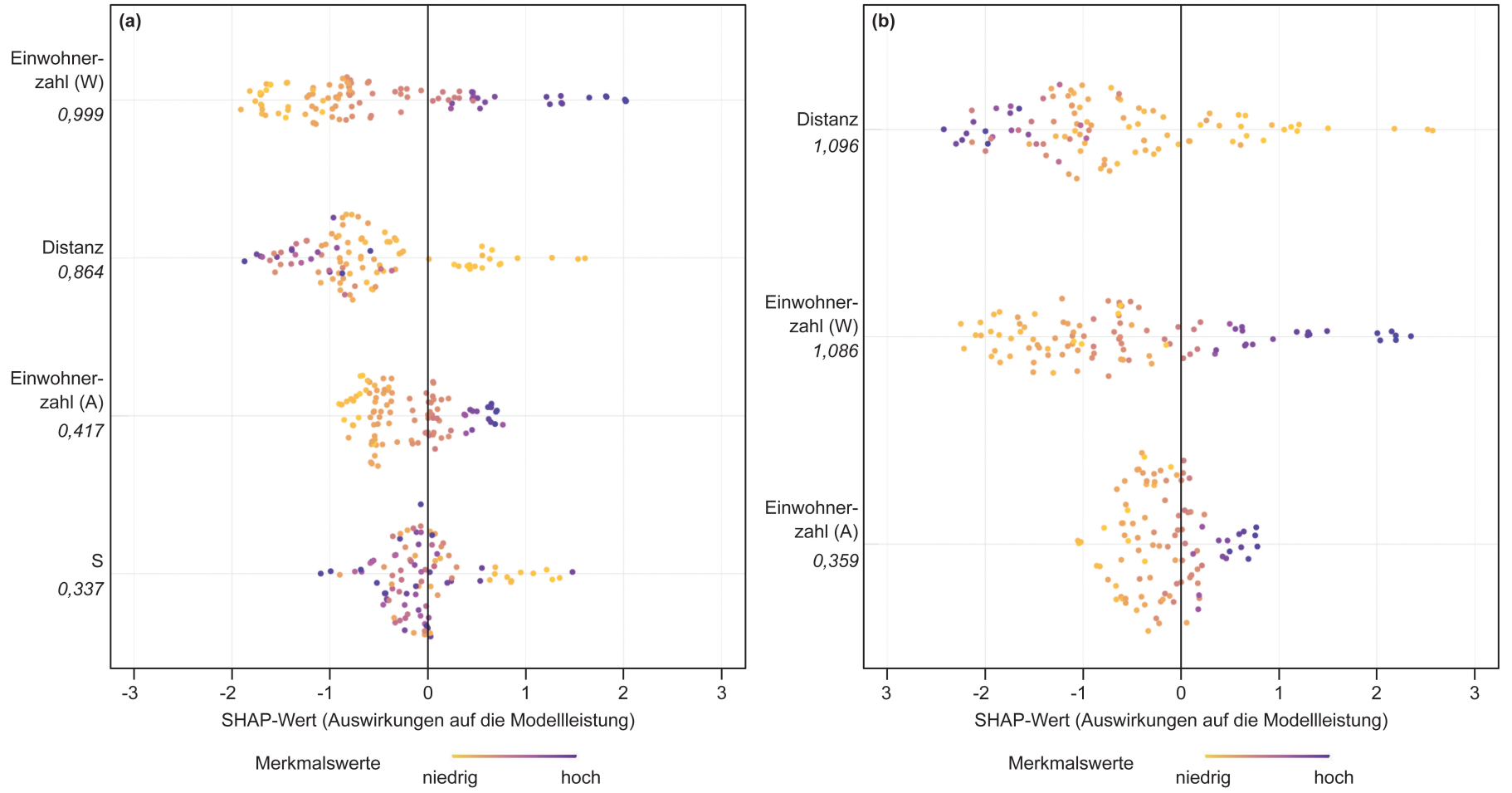


Abbildung 63: SHAP-Werte für den XGBoost-Regressor bei Pendlerbewegungen in Norwegen (Grund- und Erweiterungsmodell). Quelle: Eigene Darstellung auf Grundlage von eigenen Berechnungen auf Basis von Statistics Norway (2011).

Schweden

Die deskriptive Analyse der Pendlerbewegungen in Schweden auf NUTS3-Ebene macht deutlich, dass die Interaktionen zwischen 6 und 31.681 schwanken. Die Pendlerstrecke mit dem geringsten Mobilitätsfluss ist Jämtlands län nach Blekinge, während die Stecke Uppsala nach Stockholm von den meisten Pendlern genutzt wird. Durchschnittlich pendlen 779 Personen täglich. Die Quantile liegen bei 50 (1. Quantil), 121 (2. Quantil) und 602 (3. Quantil) (Statistics Sweden 2011).

Die Betrachtung des Gravitationsmodells in Abbildung 64 (a) macht für Schweden deutlich, dass die Pendlerbewegungen nicht optimal angepasst werden. Die Linie y=x befindet sich nicht zwischen dem 9. und 91. Perzentil in den jeweiligen Klassen. Darüber hinaus ist aus Abbildung 65 (a) ersichtlich, dass im Umkreis bis etwa 200 Kilometer sowie zwischen 600 und 700 Kilometern die Wahrscheinlichkeit zu pendeln unterschätzt wird. Es folgt bis ca. 500 Kilometer eine leichte Überschätzung. Ferner nimmt der Sörensenindex einen Wert von 0,58, der Korrelationskoeffizient nach Pearson einen Wert von 0,76 und der adjustierte R^2-Wert einen Wert von 0,58 an. Der mittlere absolute Fehler liegt bei 813 und die Wurzel der mittleren quadratischen Abweichung bei 1.756 (vgl. Abbildung 22).

Zudem ist für das Strahlungs- und finite Strahlungsmodell zu erkennen, dass die Mobilitätsflüsse bei 200, 500 und 18.000 Interaktionen (Strahlungsmodell) bzw. bei ca. 60 Personen optimal geschätzt werden (vgl. Abbildung 64 (b) und (c)). Darüber hinaus macht Abbildung 65 (b) und (c) deutlich, dass die Wahrscheinlichkeit zu pendeln marginal Abweichungen aufweist. Der Sörensenindex beträgt für das Strahlungsmodell 0,53 und für das finite Strahlungsmodell 0,57. Des Weiteren liefert der Korrelationskoeffizient nach Pearson Werte von 0,76 (Strahlungsmodell) bzw. 0,66 (finites Strahlungsmodell) und der adjustierte R^2-Wert nimmt Werte von 0,25 bzw. 0,44 an. Außerdem beträgt der mittlere absolute Fehler 695 (Strahlungsmodell) bzw. 825 (finites Strahlungsmodell) und die Wurzel der mittleren quadratischen Abweichung 2.158 (Strahlungsmodell) und 2.300 (finites Strahlungsmodell) (vgl. Abbildung 22).

Der Vergleich der verallgemeinerten Strahlungsmodelle macht deutlich, dass die Vorhersage der Mobilitätsflüsse für das ACR-Modell besonders gut sind. Es sind geringe Abweichungen zwischen Prognose und empirischen Daten bei Mobilitätsströmen in einer Größenordnung von 10 bis 60 sowie 300 Interaktionen festzustellen (vgl. Abbildung 64 (g)). Kleine Unterschätzungen bezüglich

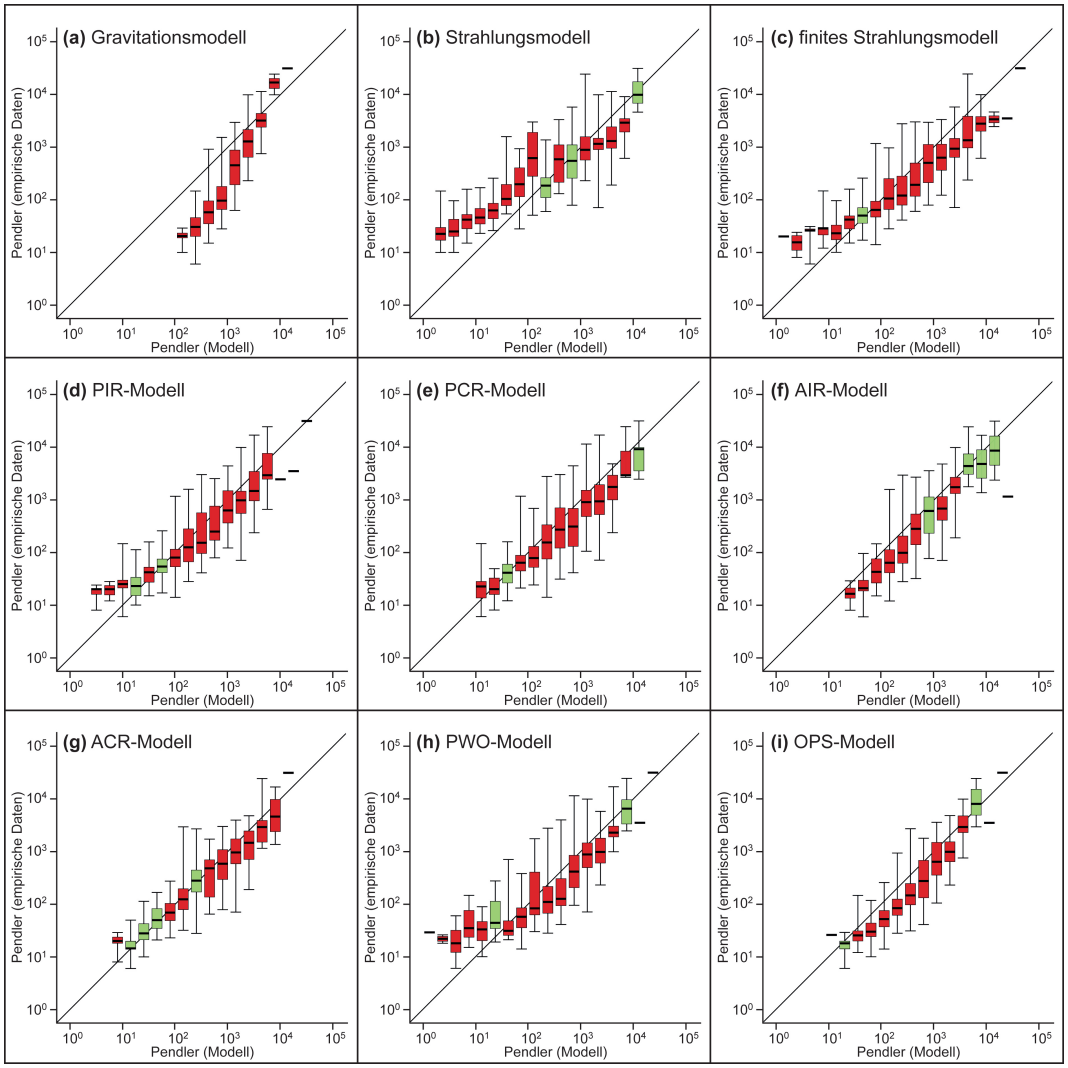


Abbildung 64: Abweichung zwischen den empirischen Daten und den Modellvorhersagen bezüglich der Pendlerbewegungen für Schweden. Quelle: Eigene Darstellung auf Grundlage von eigenen Berechnungen auf Basis von (Statistics Sweden 2011.)

der Distanz sind zwischen 130 und 170 Kilometern zu erkennen (vgl. Abbildung 65 (g)). Diese Beobachtung wird auch durch die Evaluationskoeffizienten unterstrichen (vgl. Abbildung 22). Der Sörensenindex liegt bei 0,67, der Korrelationskoeffizient nach Pearson bei 0,75 und der adjustierte R^2-Wert bei 0,56. Der mittlere absolute Fehler beträgt 510 und die Wurzel der mittleren quadratischen Abweichung 1.624. Die Analyse der Evaluationsindikatoren für die anderen drei Modelle liefert etwas schlechtere Werte (AIR-Modell: Sörensenindex: 0,62; Korrelationskoeffizient nach Pearson: 0,63; adjustierter R^2-Wert: 0,40; mittlerer absoluter Fehler: 600; Wurzel der mittleren quadratischen Abweichung: 1.897 / PCR-Modell: Sörensenindex: 0,58; Korrelationskoeffizient nach Pearson: 0,61; adjustierter R^2-Wert: 0,46; mittlerer absoluter Fehler: 653; Wurzel der mittleren quadratischen Abweichung: 1.932 / PIR-

Modell: Sörensenindex: 0,6; Korrelationskoeffizient nach Pearson: 0,68; adjustierter R^2-Wert: 0,46; mittlerer absoluter Fehler: 617; Wurzel der mittleren quadratischen Abweichung: 1.821).

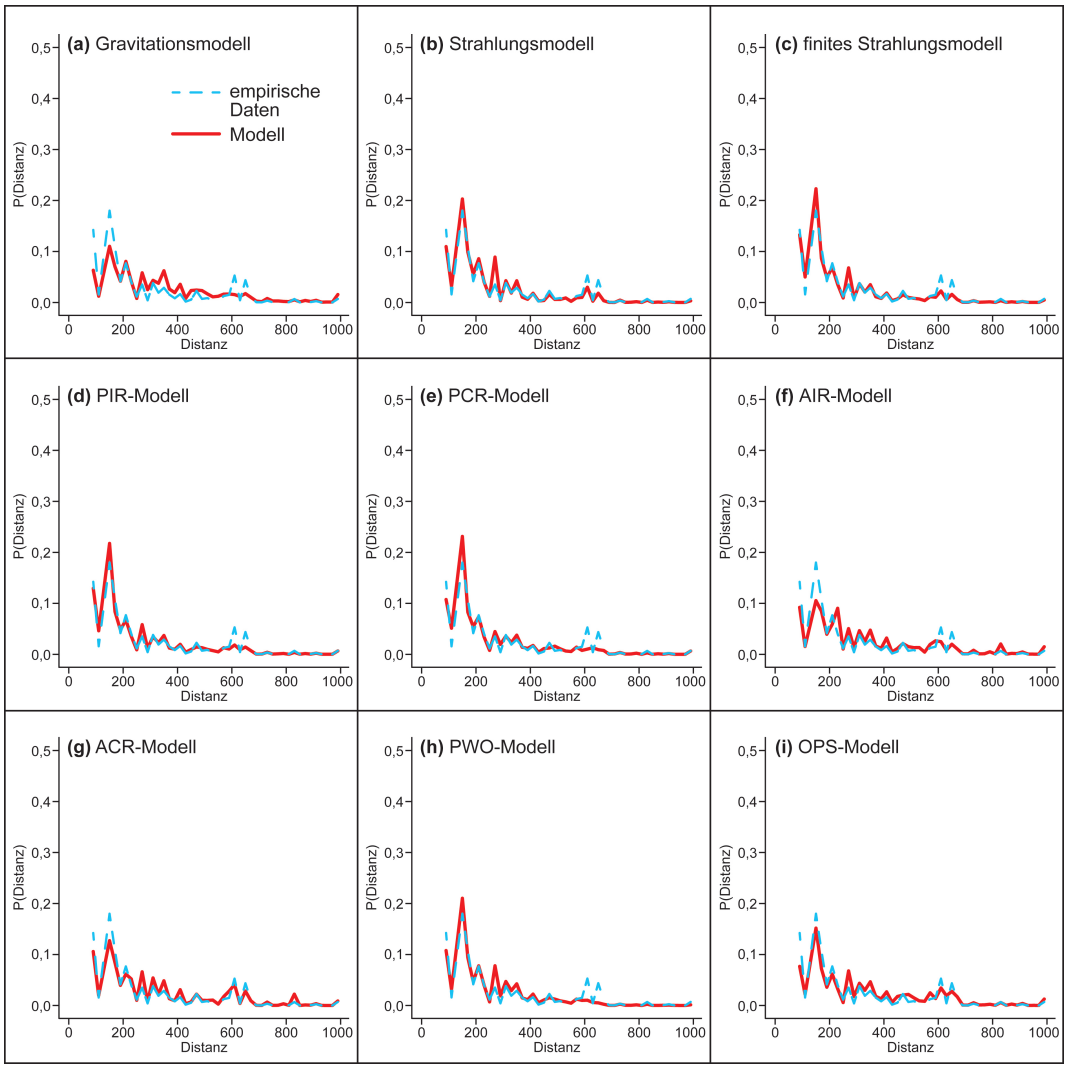


Abbildung 65: P(Distanz) im Verhältnis zur Entfernung (Pendlerbewegung in Schweden). Quelle: Eigene Darstellung auf Grundlage von eigenen Berechnungen auf Basis von Statistics Sweden (2011).

Für das PWO- und OPS-Modell liefern Mobilitätsflüsse zwischen ca. 30 und 10.000 Personen besonders gute Vorhersagen (vgl. Abbildung 64 (h) und (i)). Darüber hinaus ist in Abbildung 65 (h) und (i) zu entnehmen, dass zwischen 250 und 270 Kilometern Überschätzungen auftreten. Zudem treten für beide Modelle zwischen 600 und 700 Kilometer Unterschätzungen auf. Der Sörensenindex liefert Werte von 0,61 (PWO-Modell) und 0,65 (OPS-Modell). Außerdem wird deutlich, dass beide Modelle einen starken Zusammenhang (PWO-Modell: 0,73; OPS-Modell: 0,80) aufweisen. Die adjustierten R^2-Werte

betragen bei 0,54 und 0,63. Außerdem wurde für den mittleren absoluten Fehler für das PWO-Modell 614 gemessen. Für das OPS-Modell liegt dieser hingegen bei 553. Die Wurzel der mittleren quadratischen Abweichung nimmt Werte von 1.659 für das PWO-Modell und 1.579 für das OPS-Modell an (vgl. Abbildung 22).

Aus Abbildung 67 (b) ist für das Grundmodell des XGBoost-Algorithmus ersichtlich, dass kleine Distanzen sowie NUTS3-Regionen mit vielen Einwohnern am Wohn- und Arbeitsort einen positiven Einfluss auf die Prognose haben. Der Sörensenindex beträgt in 65 % aller Simulationen 0,65 (vgl. Abbildung 66 (a)), der Korrelationskoeffizient nach Pearson 0,80 und der adjustierte R^2-Wert 0,84. Zudem liegt der mittlere absolute Fehler bei 551 und die Wurzel der mittleren quadratischen Abweichung bei 2.619 (vgl. Abbildung 24).

Ferner ist für das Erweiterungsmodell des XGBoost-Algorithmus (vgl. Abbildung 67 (a)) erkennbar, dass für Schweden ebenfalls kleine Entfernungen zwischen Wohn- und Arbeitsort einen positiven Einfluss auf die Prognose haben. Zudem wirken sich vor allem einwohnerstarke Regionen positiv auf die Schätzung aus, während die hohe Einwohnerzahlen zwischen Arbeits- und Wohnort zu einer Verschlechterung der Vorhersage führen. Die Analyse der Evaluationsindikatoren zeigen, dass der Sörensenindex in 69 % der Simulationen bei 0,67 (vgl. Abbildung 66 (a)) liegt. Zudem nehmen der Korrelationskoeffizient nach Pearson Werte von 0,93, der adjustierte R^2-Wert Werte von 0,86, der mittlere absolute Fehler Werte von 548 und die Wurzel der mittleren quadratischen Abweichung Werte von 2.402 an (vgl. Abbildung 24).

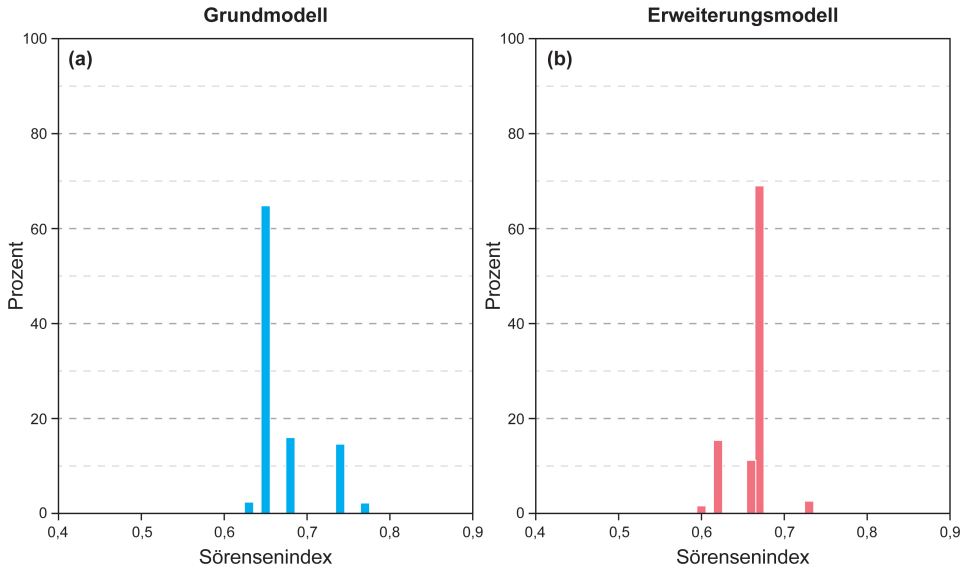


Abbildung 66: Ergebnisse der Simulation für Schweden. Quelle: Eigene Darstellung auf Grundlage von eigenen Berechnungen auf Basis von Statistics Sweden (2011).

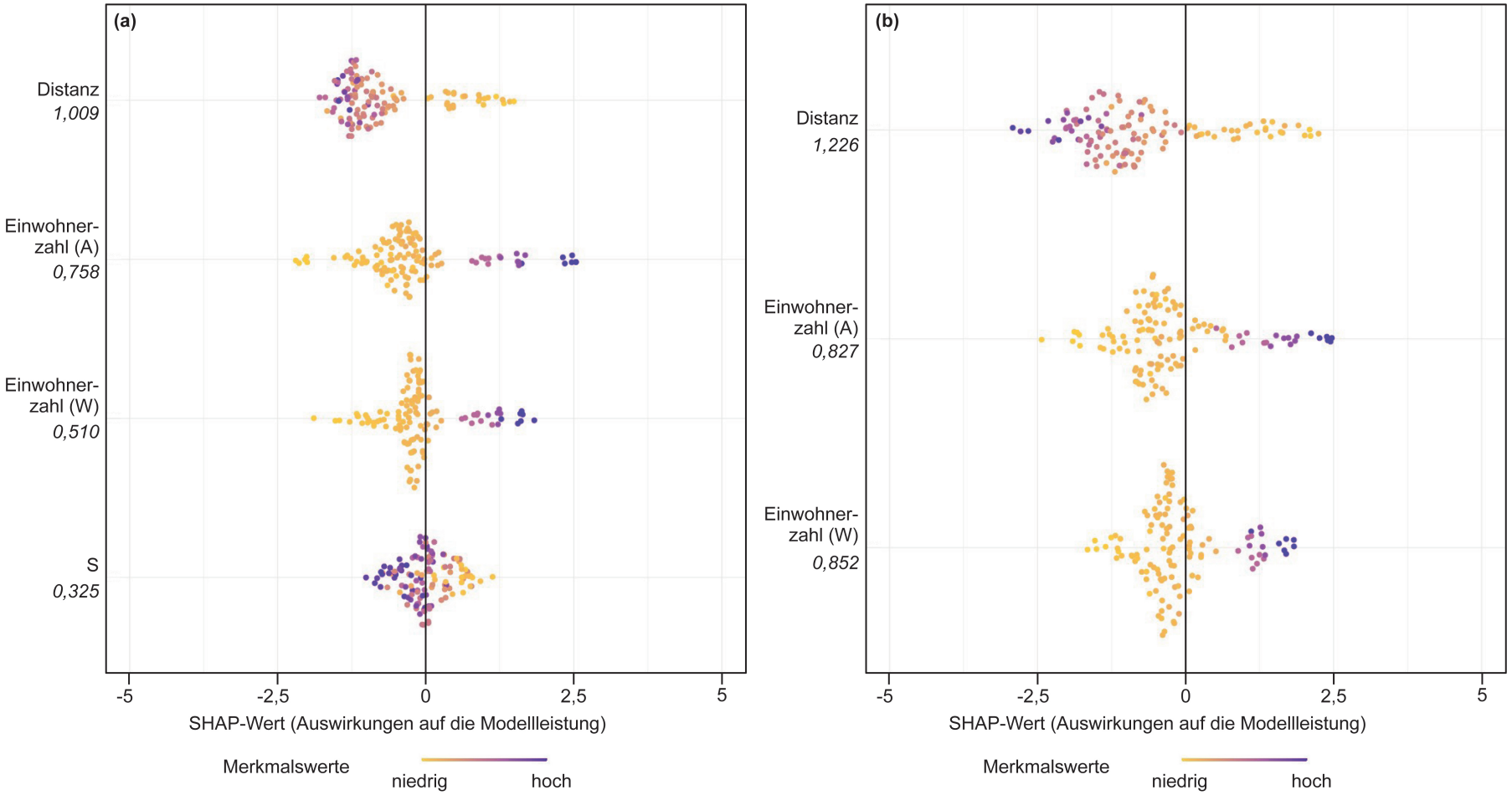


Abbildung 67: SHAP-Werte für den XGBoost-Regressor bei Pendlerbewegungen in Schweden (Grund- und Erweiterungsmodell). Quelle: Eigene Darstellung auf Grundlage von eigenen Berechnungen auf Basis von Statistics Sweden (2011).

Kroatien

Die Analyse der Mobilitätsflüsse mittels Lage- und Streuparamter macht deutlich, dass, sofern überhaupt Interaktionen auftreten, das Minimum bei 1 liegt (z. B. Dubrovnik-Neretva nach Virovitica-Podravina). Zudem weist die Strecke Zagreb-Land - Zagreb-Stadt mit 58.113 Personen den meisten Pendlern auf. Das arithmetische Mittel liegt bei 417 und der Median bei 47. Außerdem nehmen das erste und dritte Quantil Werte von 15 und 169 an (Croatian Bureau of Statistics 2011).

In Abbildung 68 (a) ist zu erkennen, dass das Gravitationsmodell insbesondere für Pendlerströme in einer Größenordnung von ca. 1.800 Personen besonders gute Ergebnisse liefert. Die Übersicht in Abbildung 69 (a) zeigt zudem, dass die Wahrscheinlichkeit zu pendeln bis ca. 30 Kilometer stark unterschätzt wird und zwischen 90 und 400 Kilometern Überschätzungen vorliegen. Der Sörensenindex liegt bei 0,59, der Korrelationskoeffizient nach Pearson bei 0,94, der adjustierte R^2-Wert bei 0,88, der mittlere absolute Fehler bei 361 und die Wurzel der mittleren quadratischen Abweichung bei 1.921 (vgl. Abbildung 22).

Aus Abbildung 68 (b) ist ferner ersichtlich, dass die Vorhersagen mittels des Strahlungsmodells für die Pendlerflüsse von ca. 30 bis 60 und 1.800 Personen besonders gute Schätzungen liefert. Des Weiteren ist in Abbildung 69 (b) erkennbar, dass die Wahrscheinlichkeit einer täglichen Fahrt zum Arbeitsplatz zunächst stark unterschätzt wird. Es folgt ab ca. 30 bis 250 Kilometern eine Überschätzung mittels des Modells. Für das Strahlungsmodell konnte außerdem bei der Betrachtung der Evaluationsindikatoren festgestellt werden, dass der Sörensenindex bei 0,46, der Korrelationskoeffizient nach Pearson bei 0,94 und der adjustierte R^2-Wert bei 0,34 liegt. Der mittlere absolute Fehler nimmt einen Wert von 433 und die Wurzel der mittleren quadratischen Abweichung einen Wert von 2.536 an (vgl. Abbildung 22).

Betrachtet man die Ergebnisse des finiten Strahlungsmodells, so wird deutlich, dass für Mobilitätsströme von 5, 30 und 10.000 Interaktionen besonders gute Anpassungen erzielt werden (vgl. Abbildung 68 (c)). Außerdem wird, ähnlich wie beim Strahlungsmodell, die Wahrscheinlichkeit, einen Ort als Arbeitsplatz zu wählen zunächst unterschätzt. Die Abweichungen sind jedoch geringer. Es folgt eine Überschätzung bis ca. 110 und zwischen 130 und 250 Kilometern (vgl. Abbildung 69 (c)). Für das Strahlungsmodell liefert der Vergleich mit den empirischen Daten einen Sörensenindex von 0,63, einen Korrelationskoeffizient nach Pearson von 0,96 und einen adjustierten R^2-Wert von 0,92. Der mittlere

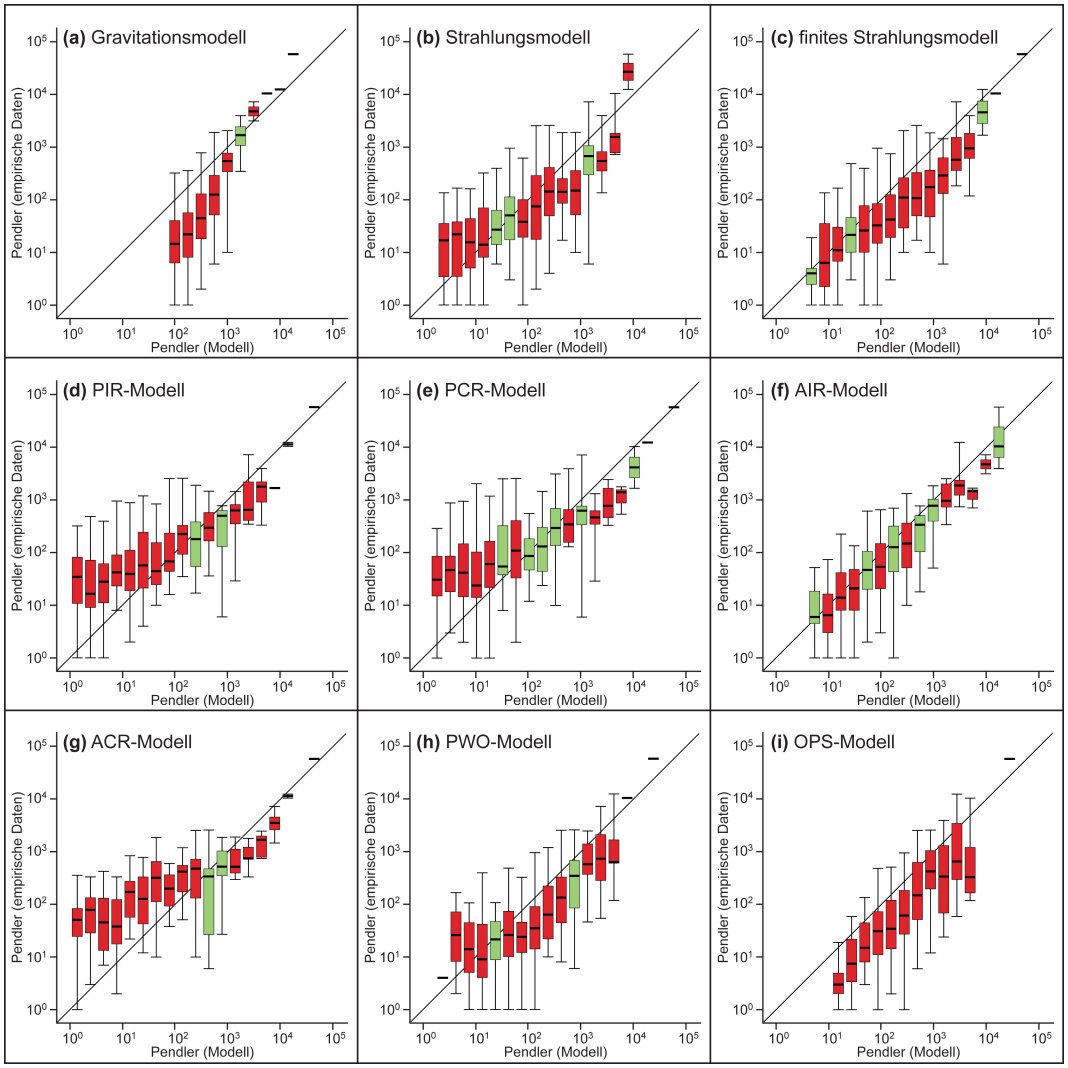


Abbildung 68: Abweichung zwischen den empirischen Daten und den Modellvorhersagen bezüglich der Pendlerbewegungen für Kroatien. Quelle: Eigene Darstellung auf Grundlage von eigenen Berechnungen auf Basis von Croatian Bureau of Statistics (2011).

absolute Fehler liegt bei 373 und die Wurzel der mittleren quadratischen Abweichung bei 873 (vgl. Abbildung 22).

Unter den verallgemeinerten Strahlungsmodellen liefert das PIR-Modell die beste Anpassung. Besonders gute Schätzungen werden für etwa 200 sowie 1.000 Pendler pro Tag erzielt (vgl. Abbildung 68 (d)). Betrachtet man die Wahrscheinlichkeit zu pendeln in Abbildung 69 (d), so ist erkennbar, dass es im Bereich von 30 bis 90 Kilometern zu Überschätzungen kommt. Es folgen kleinere Unterschätzungen. Der Sörensenindex liegt bei 0,71, der Korrelationskoeffizient nach Pearson bei 0,98 und der adjustierte R^2-Wert bei 0,95. Überdies nimmt der mittlere absolute Fehler einen Wert von 239 und die Wurzel der mittleren quadratischen Abweichung einen Wert von 658 an (vgl.

Abbildung 22). Die anderen drei Modelle weisen leichte Verschlechterungen der Ergebnisse auf (ACR-Modell: Sörensenindex: 0,70; Korrelationskoeffizient nach Pearson: 0,97; adjustierter R^2-Wert: 0,93; mittlerer absoluter Fehler: 254; Wurzel der mittleren quadratischen Abweichung: 751 / AIR-Modell: Sörensenindex: 0,60; Korrelationskoeffizient nach Pearson: 0,69; adjustierter R^2-Wert: 0,47; mittlerer absoluter Fehler: 334; Wurzel der mittleren quadratischen Abweichung: 2.199 / PCR-Modell: Sörensenindex: 0,64; Korrelationskoeffizient nach Pearson: 0,96; adjustierter R^2-Wert: 0,91; mittlerer absoluter Fehler: 301; Wurzel der mittleren quadratischen Abweichung: 875).

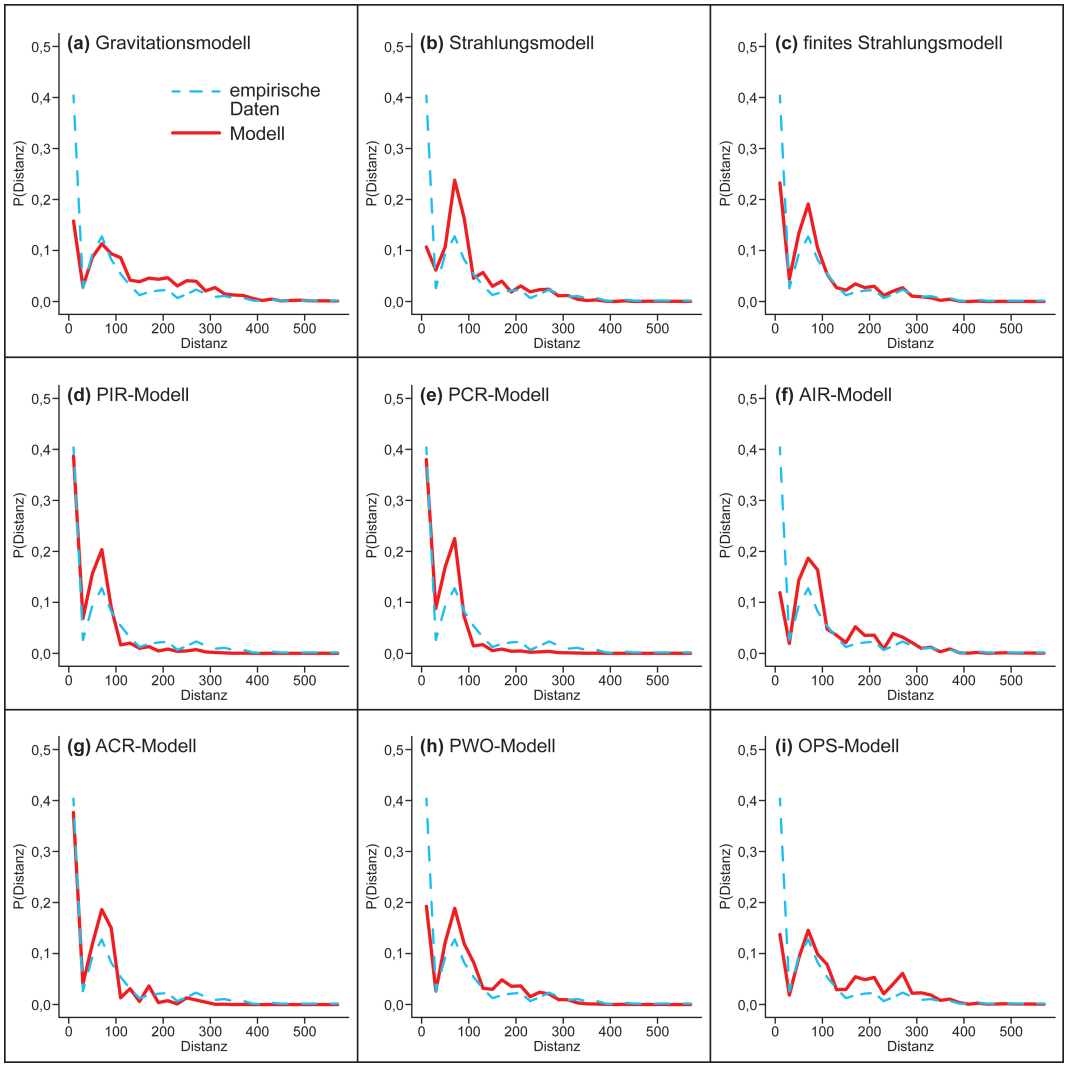


Abbildung 69: P(Distanz) im Verhältnis zur Entfernung (Pendlerbewegung in Kroatien). Quelle: Eigene Darstellung auf Grundlage von eigenen Berechnungen auf Basis von Croatian Bureau of Statistics (2011).

Darüber hinaus zeigt Abbildung 68 (h) für das PWO-Modell, dass prognostizierte Mobilitätsflüsse von ca. 30 sowie 1.000 Pendlern geringe Abweichungen aufweisen. Es treten zunächst Unterschätzungen auf. Ab ca. 50 bis 250 Kilometer sind vor allem Überschätzungen erkennbar (vgl. Abbildung 69 (h)). Der Sörensenindex nimmt einen Wert von 0,57, der Korrelationskoeffizient nach Pearson einen Wert von 0,94 und der adjustierte R^2-Wert einen Wert von 0,88 an. Darüber hinaus konnte ein mittlerer absoluter Fehler von 355 und eine Wurzel der mittleren quadratischen Abweichung von 1.615 gemessen werden (vgl. Abbildung 22).

Für Kroatien wird deutlich, dass für das OPS-Modell die Linie y=x nicht zwischen dem 9. und 91. Perzentil in der jeweiligen Klasse liegt (vgl. Abbildung 68 (i)). Aus Abbildung 69 (i) ist darüber hinaus zu entnehmen, dass die Wahrscheinlichkeit zu pendeln in Abhängigkeit zur Distanz im direkten Umfeld starke Unterschätzungen aufweist. Ab ca. 90 bis 300 Kilometer treten Überschätzungen auf. Außerdem beträgt der Sörensenindex 0,52, der Korrelationskoeffizient nach Pearson 0,90 und der adjustierte R^2-Wert 0,81. Der mittlere absolute Fehler nimmt einen Wert von 403 und die Wurzel der mittleren quadratischen Abweichung von 1.999 an (vgl. Abbildung 22).

Betrachtet man die Ergebnisse für das Grundmodell des XGBoost-Algorithmus (vgl. Abbildung 71 (b)), so ist erkennbar, dass für Kroatien große Entfernungen zwischen Wohn- und Arbeitsort sowie niedrige Einwohnerzahlen einen negativen Einfluss auf die Prognose haben. Die Evaluationsergebnisse der Indikatoren zeigen, dass der Sörensenindex in 71 % der Simulationen bei 0,81 (vgl. Abbildung 70 (a)), der Korrelationskoeffizient nach Pearson bei 0,9 und der adjustierte R^2-Wert bei 0,85 liegen. Außerdem beträgt der mittlere absolute Fehler 71 und die Wurzel der mittleren quadratischen Abweichung 152 (vgl. Abbildung 24).

Für die Erweiterung des Grundmodells des XGBoost-Algorithmus (vgl. Abbildung 71 (a)) ist festzustellen, dass Regionen mit einer hohen Bevölkerung am Wohn- und Arbeitsort einen positiven Einfluss auf die Prognose haben. Zudem liefern hohe Entfernungen einen negativen Einfluss auf die Prognoseleistung. Der Einfluss der Einwohnerzahl zwischen den beiden Standorten kann nicht abschließend geklärt werden.

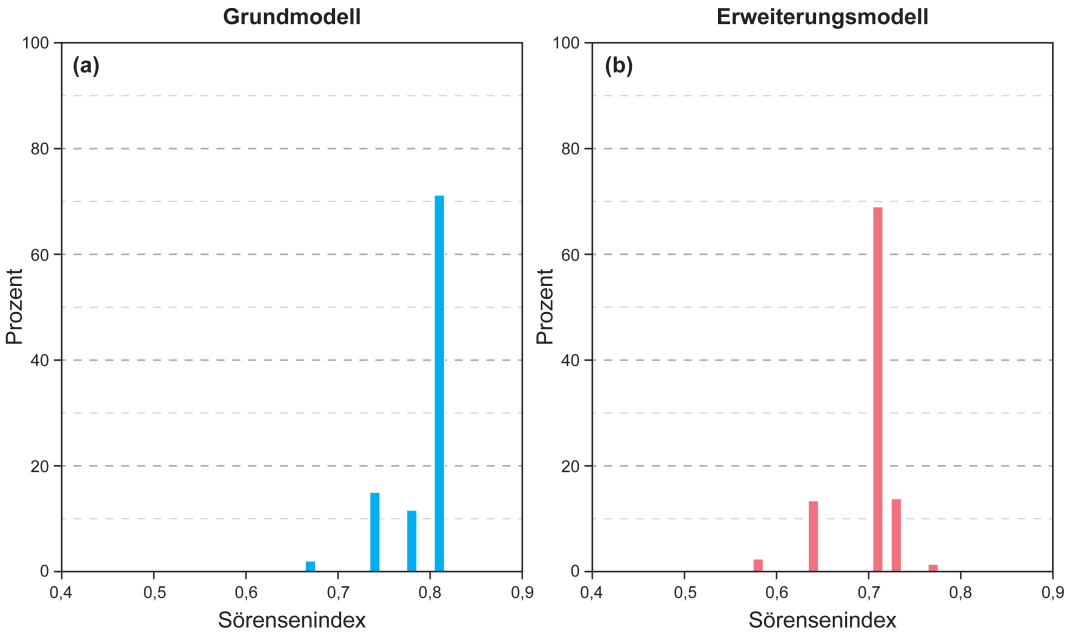


Abbildung 70: Ergebnisse der Simulation für Kroatien. Quelle: Eigene Darstellung auf Grundlage von eigenen Berechnungen auf Basis von Croatian Bureau of Statistics (2011).

Darüber hinaus ist in Abbildung 70 (b) sichtbar, dass der Sörensenindex in 69 % der Simulationen bei 0,71 liegt. Der Korrelationskoeffizient nach Pearson liefert einen Wert von 0,83, der adjustierte R^2-Wert von 0,69, der mittlere absolute Fehler von 129 und die Wurzel der mittleren quadratischen Abweichung von 293 (vgl. Abbildung 24).

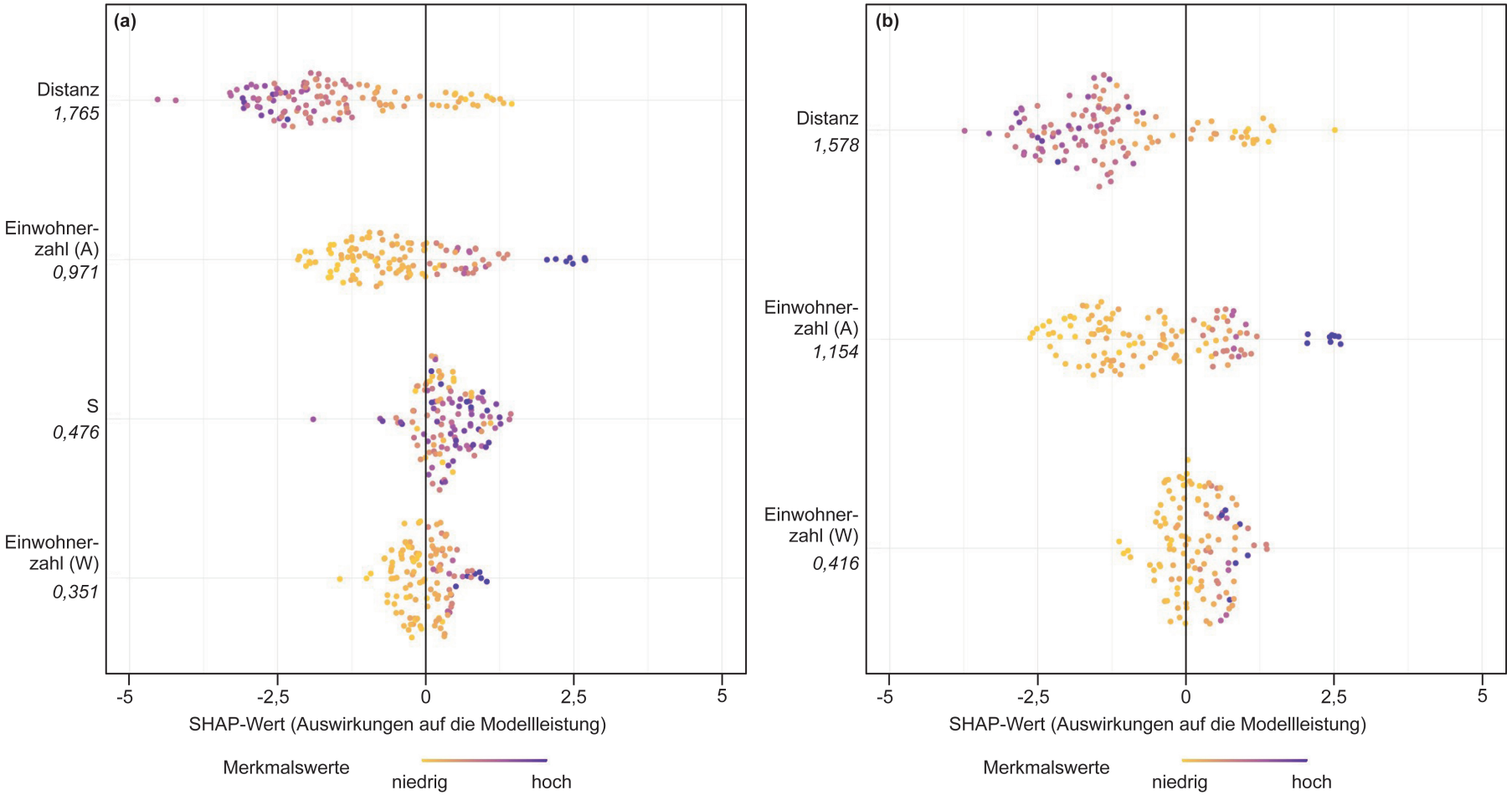


Abbildung 71: SHAP-Werte für den XGBoost-Regressor bei Pendlerbewegungen in Kroatien (Grund- und Erweiterungsmodell). Quelle: Eigene Darstellung auf Grundlage von eigenen Berechnungen auf Basis von Croatian Bureau of Statistics (2011).

Italien

Die deskriptive Untersuchung der Pendlerbewegungen in Italien macht deutlich, dass, die auftretenden Mobilitätsströme zwischen 1 (z. B. Torino - Siracusa) und 98.543 (Milano - Monza e della Brianza) schwanken. Das arithmetische Mittel liegt bei 450 und der Median bei 8. Zudem betragen das erste und dritte Quantil 2 und 46 Interaktionen (Instituto Nazionale di Statistica 2011).

Aus den Ergebnissen für das Gravitationsmodell in Italien ist zunächst abzuleiten, dass die Mobilitätsströme nicht optimal angepasst werden können (Linie y=x nicht zwischen dem 9. und 91. Perzentil in der jeweiligen Klasse) (vgl. Abbildung 72 (a)). Außerdem ist aus Abbildung 73 (a) zu entnehmen, dass Pendlerbewegungen in den ersten 100 Kilometern unterschätzt werden. Es folgt eine Überschätzung. Darüber hinaus werden folgende Evaluationsindikatoren erzielt (vgl. Abbildung 22): Sörensenindex: 0,28; Korrelationskoeffizient nach Pearson: 0,69; adjustierter R^2-Wert: 0,48; mittlerer absoluter Fehler: 412; Wurzel der mittleren quadratischen Abweichung: 1.198.

Für das Strahlungsmodell erzielen die Mobilitätsflüsse ab 10.000 Pendlern besonders gute Anpassungen (vgl. Abbildung 72 (b)). Aus Abbildung 73 (b) ist ferner ersichtlich, dass die Wahrscheinlichkeit einer tägliche Fahrt zum Arbeitsplatz zwischen 140 und 250 Kilometern mit dem Modell überschätzt wird. Für das Strahlungsmodell konnte zudem festgestellt werden, dass der Sörensenindex 0,54, der Korrelationskoeffizient nach Pearson 0,69 und der adjustierte R^2-Wert 0,43 beträgt. Der mittlere absolute Fehler nimmt einen Wert von 135 und die Wurzel der mittleren quadratischen Abweichung einen Wert von 1.139 an (vgl. Abbildung 22).

Betrachtet man die Ergebnisse des finiten Strahlungsmodells, so zeigt sich, dass die Prognose für die Pendlerflüsse von ca. 3.000, 10.000 und 30.000 Personen besonders gute Schätzungen liefert (vgl. Abbildung 72 (c)). Des Weiteren wird, ähnlich wie beim Gravitationsmodell, die Wahrscheinlichkeit bis zu einer Distanz von ca. 100 Kilometern unterschätzt. Es folgt eine Überschätzung der Daten bis ca. 300 Kilometer (vgl. Abbildung 73 (c)). Für das finite Strahlungsmodell macht der Vergleich mit den empirischen Daten ferner deutlich, dass der Sörensenindex Werte von 0,59, der Korrelationskoeffizient nach Pearson 0,88 und der adjustierte R^2-Wert 0,77 erzielt. Der mittlere absolute Fehler liegt bei 162 und die Wurzel der mittleren quadratischen Abweichung bei 825 (vgl. Abbildung 22).

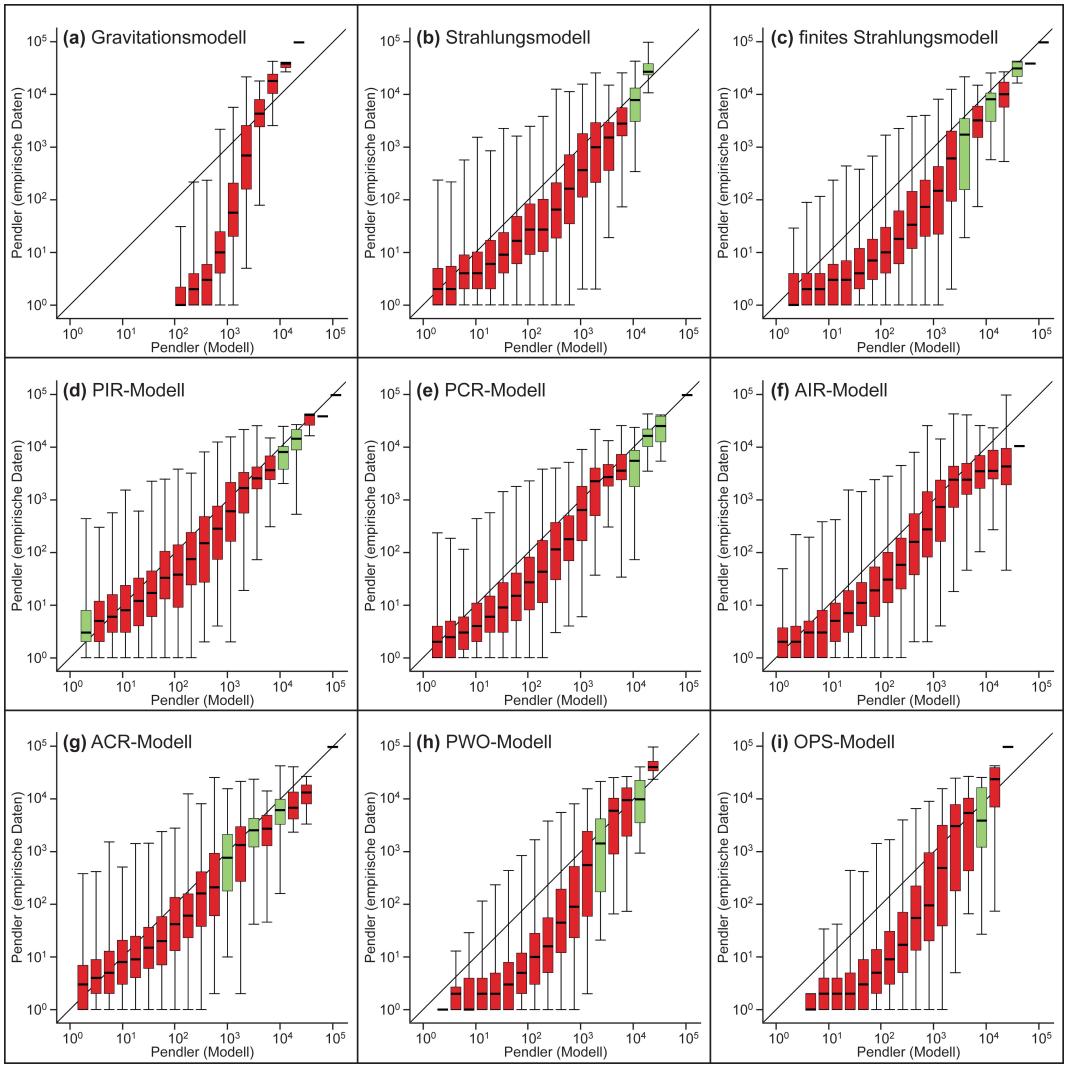


Abbildung 72: Abweichung zwischen den empirischen Daten und den Modellvorhersagen bezüglich der Pendlerbewegungen für Italien. Quelle: Eigene Darstellung auf Grundlage von eigenen Berechnungen auf Basis von Instituto Nazionale di Statistica (2011).

Eine deutliche Verbesserung zwischen Modell und empirischen Daten erbringen die verallgemeinerten Strahlungsmodelle. Es fällt auf, dass für das PIR- und PCR-Modell deutliche Verbesserungen erzielt werden können. Besonders gute Anpassungen sind beim PIR-Modell für Mobilitätsströme von ca. 5, 10.000 und 18.000 Interaktionen festzustellen. Beim PCR-Modell werden Pendlerbewegungen ab ca. 10.000 Personen besonders gut geschätzt und für das ACR-Modell werden Mobilitätsflüsse von 1.000, 3.000 und 10.000 optimal geschätzt. Beim AIR-Modell liegt die Linie y=x nicht zwischen dem 9. und 91. Perzentil in der jeweiligen Klasse (vgl. Abbildung 72 (d) bis (g)). Bezüglich der Distanz sind für die Modelle sehr geringe Abweichungen festzustellen (vgl. Abbildung 73). Insgesamt bestätigen die Evaluationsindikatoren diese Beobachtung (PCR-Model: Sörensenindex:

0,64; Korrelationskoeffizient nach Pearson: 0,87; adjustierter R^2-Wert: 0,76; mittlerer absoluter Fehler: 105; Wurzel der mittleren quadratischen Abweichung: 762 / PIK-Modell: Sörensenindex: 0,65; Korrelationskoeffizient nach Pearson: 0,87; adjustierter R^2-Wert: 0,26; mittlerer absoluter Fehler: 103; Wurzel der mittleren quadratischen Abweichung: 753 / ACR-Modell: Sörensenindex: 0,59; Korrelationskoeffizient nach Pearson: 0,77; adjustierter R^2-Wert: 0,60; mittlerer absoluter Fehler: 121; Wurzel der mittleren quadratischen Abweichung: 961 / AIR-Modell: Sörensenindex: 0,50; Korrelationskoeffizient nach Pearson: 0,51; adjustierter R^2-Wert: 0,28; mittlerer absoluter Fehler: 147; Wurzel der mittleren quadratischen Abweichung: 1.338).

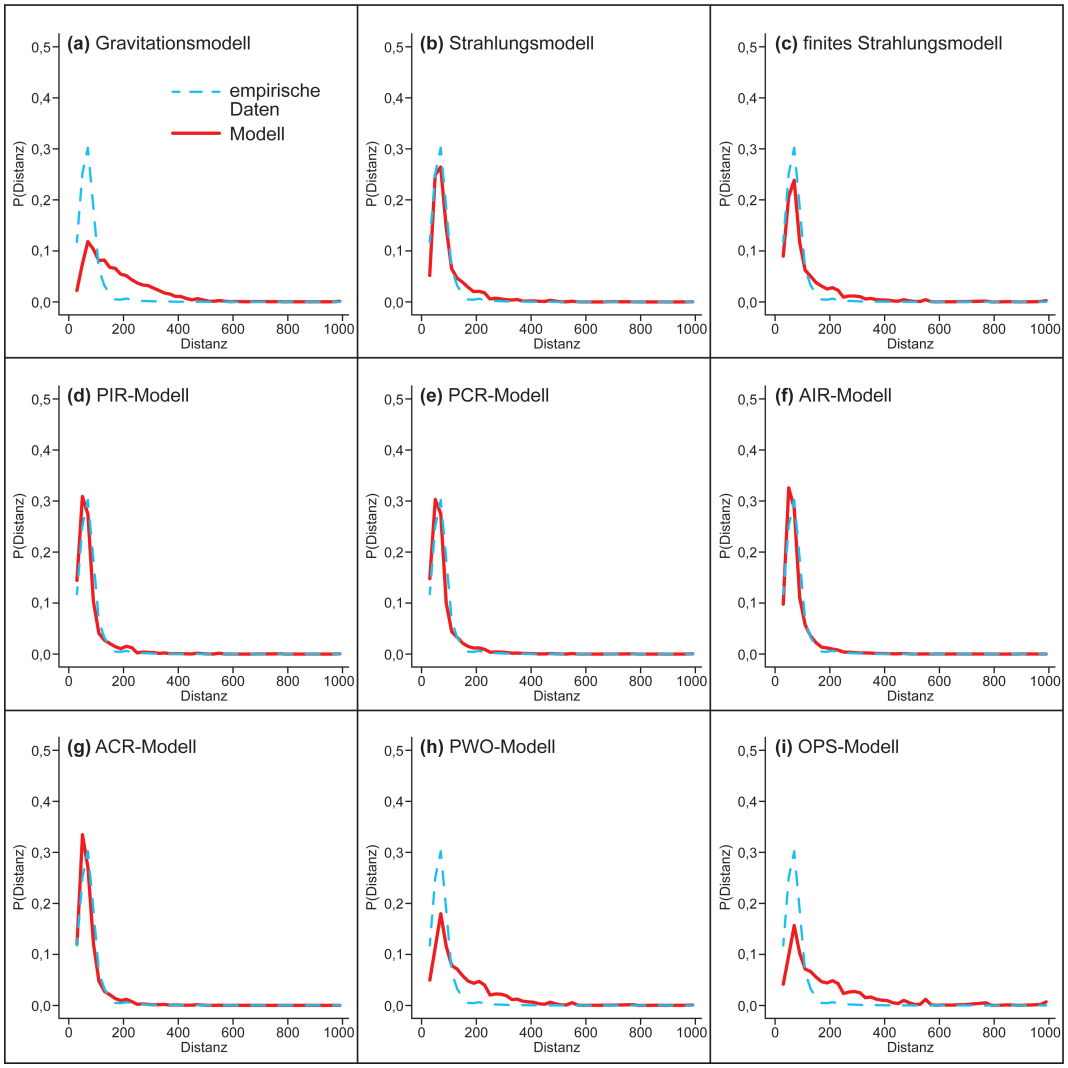


Abbildung 73: P(Distanz) im Verhältnis zur Entfernung (Pendlerbewegung Italien). Quelle: Eigene Darstellung auf Grundlage von eigenen Berechnungen auf Basis von Instituto Nazionale di Statistica (2011).

Für das parameterfreie PWO-Modell wird deutlich, dass Pendlerströme in der Größenordnung von ca. 5.600 und 18.000 Personen besonders gute Anpassungen liefern (vgl. Abbildung 72 (h)). Daneben exisiteren Unterschätzungen bis ca. 100 Kilometer. Danach folgt eine Überschätzung der Wahrscheinlichkeit (vgl. Abbildung 73 (h)). Die Evaluationsindikatoren (vgl. Abbildung 22) ergeben für die Schätzungen des PWO-Modells Werte von 0,45 (Sörensenindex), 0,80 (Korrelationskoeffizient nach Pearson), 0,64 (adjustierter R^2-Wert), 161 (mittlerer absoluter Fehler) und 1.031 (Wurzel der mittleren quadratischen Abweichung).

Außerdem zeigt sich in Abbildung 72 für das OPS-Modell, dass Pendlerbewegungen von ca. 10.000 Interaktionen besonders gute Prognoseergebnisse liefern. Zudem sind für Pendlerdistanzen im Umkreis bis ca. 100 Kilometern Unterschätzungen festzustellen. Im Anschluss folgt eine Überschätzung (vgl. Abbildung 73 (i)). Die Evaluationsindikatoren liefern einen Sörensenindex von 0,37, einen Korrelationskoeffizient nach Pearson von 0,74 und einen adjustierten R^2-Wert von 0,54. Der mittlere absolute Fehler liegt bei 185 und die Wurzel der mittleren quadratischen Abweichung bei 1.153 (vgl. Abbildung 22).

In Abbildung 75 (b) wird für das Grundmodell des XGBoost-Algorithmus deutlich, dass mit zunehmender Entfernung zwischen Arbeits- und Wohnort die Prognoseleistungen des Modells abnimmt. Zudem wirken sich kleine Einwohnerzahlen negativ aus. Der Sörensenindex beträgt in 83 % der Simulationen 0,69 (vgl. Abbildung 74 (a)), der Korrelationskoeffizient 0,83, der adjustierte R^2-Wert 0,86, der mittlere absolute Fehler 73 und die Wurzel der mittleren quadratischen Abweichung 659 (vgl. Abbildung 24).

Aus Abbildung 75 (a) ist ersichtlich, dass beim Erweiterungsmodells des XGBoost-Algorithmus große Entfernungen einen negativen Einfluss haben. Für Standorte, bei denen eine hohe Anzahl an Einwohnern im Zwischenraum leben, kann keine klare Wirkrichtung festgestellt werden. Hohe Bevölkerungszahlen am Wohn- und Arbeitsort haben hingegen einen positiven Einfluss. Die Analyse der Evaluationsindikatoren zeigt, dass in 71 % der Fälle ein Sörensenindex von 0,64 simuliert wird (vgl. Abbildung 74 (b)). Der Korrelationskoeffizient nach Pearson beträgt 0,91, der adjustierte R^2-Wert 0,83, der mittlere absolute Fehler 84 und die Wurzel der mittleren quadratischen Abweichung 726 (vgl. Abbildung 24).

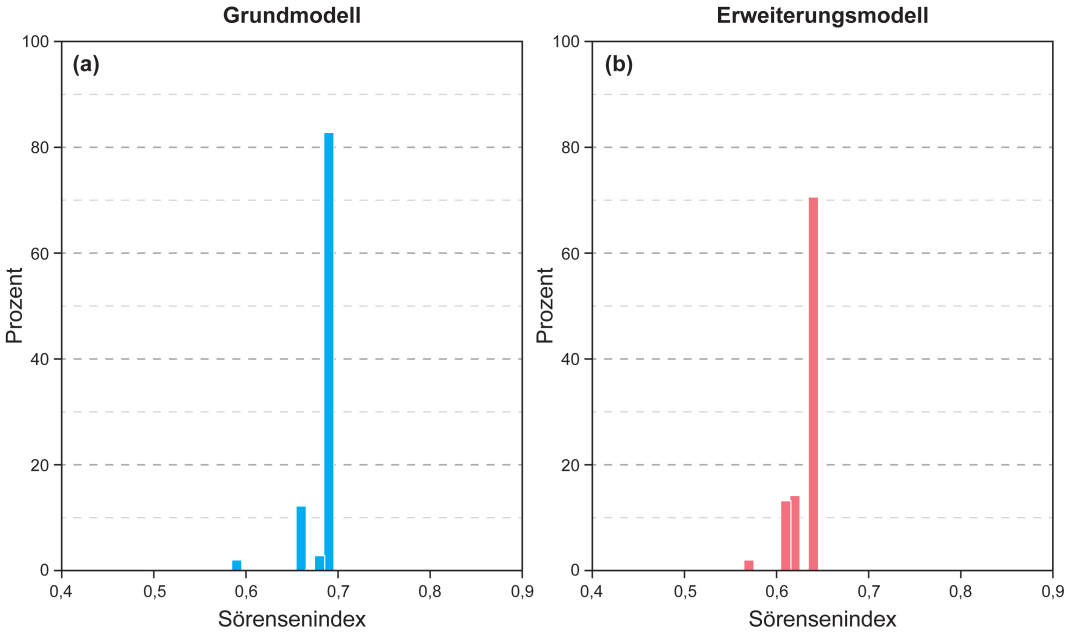


Abbildung 74: Ergebnisse der Simulation für Italien. Quelle: Eigene Darstellung auf Grundlage von eigenen Berechnungen auf Basis von Instituto Nazionale di Statistica (2011).

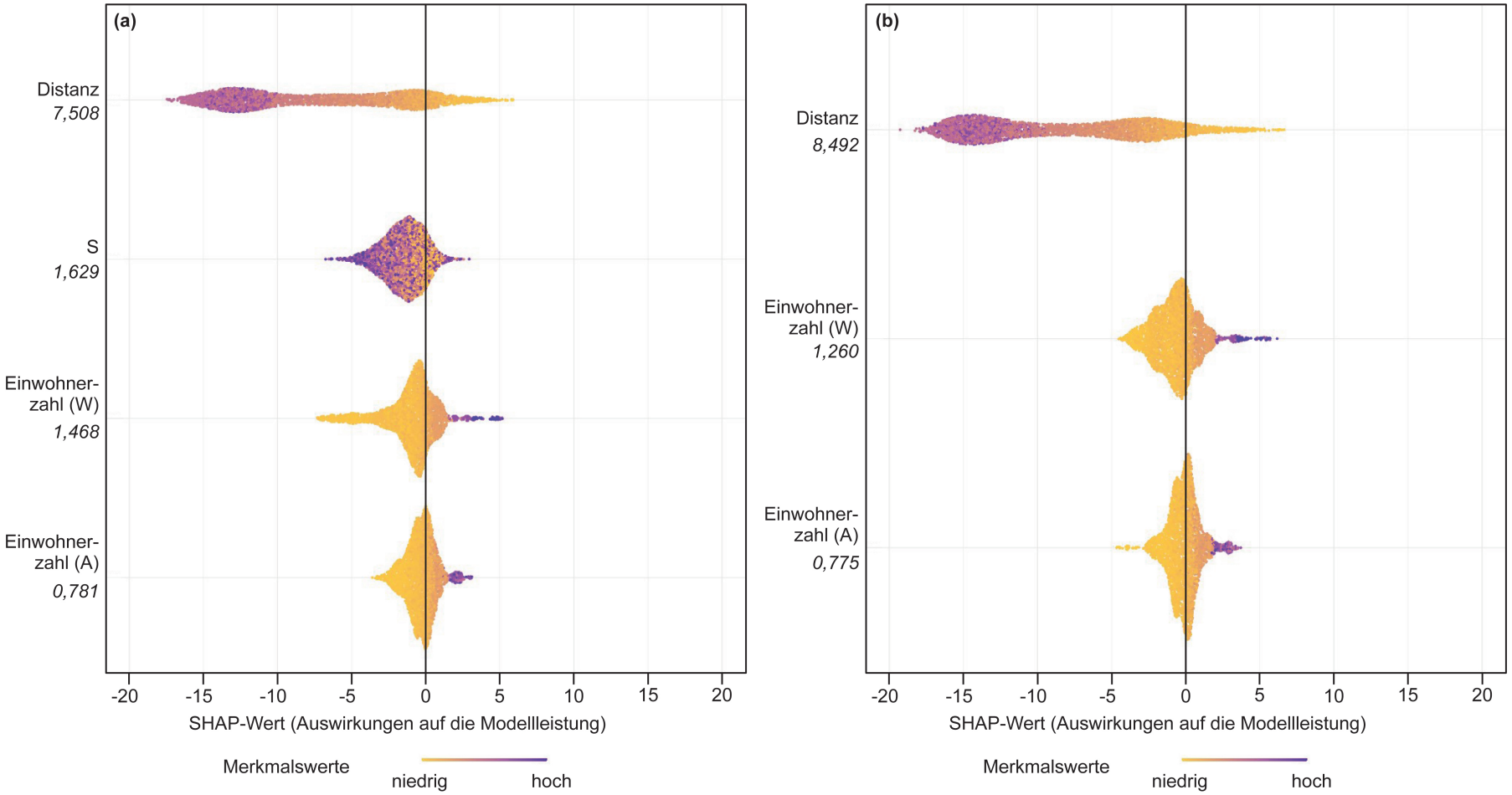


Abbildung 75: SHAP-Werte für den XGBoost-Regressor bei Pendlerbewegungen in Italien (Grund- und Erweiterungsmodell). Quelle: Eigene Darstellung auf Grundlage von eigenen Berechnungen auf der Basis von Instituto Nazionale di Statistica (2011).

Spanien

Betrachtet man die deskriptive Auswertung der Pendlerdaten für Spanien, so wird deutlich, dass das Minimum an Interaktionen bei 5 liegt (z. B. Burgos-Lugo). Die Strecke mit den meisten Pendlern ist Toledo - Madrid mit 40.725. Durchschnittlich pendelten 281 Personen im Jahr 2011. Der Median beträgt 45 und das erste und dritte Quantil 15 und 130 Personen (Instituto Nacionale de Estadística 2011).

Abbildung 76 (a) zeigt zunächst für das Gravitationsmodell, dass die Linie y=x nicht zwischen dem 9. und 91. Perzentil in der jeweiligen Klasse liegt und daher die Prognose der Mobilitätsströme nicht optimal ist. Zudem treten bis etwa 160 Kilometer große Unterschätzungen auf. Während ab ca. 210 Kilometer das Modell die Wahrscheinlichkeit, einen Ort als Arbeitsort zu wählen, überschätzt (vgl. Abbildung 77 (a)). Die Analyse der Evaluationsindikatoren (vgl. Abbildung 22) erbringt ähnliche Beobachtungen. Es wird ein Sörensenindex von 0,46, ein Korrelationskoeffizient nach Pearson von 0,58 und ein adjustierter R^2-Wert von 0,33 gemessen. Außerdem liegt der mittlere absolute Fehler bei 249 und die Wurzel der mittleren quadratischen Abweichung bei 1.129.

Für das Strahlungsmodell ist ersichtlich, dass Pendlerbewegungen in der Größenordnung von ca. 1.800 Pendler gute Schätzungen liefern (vgl. Abbildung 76 (b)). Die Abweichungen bezüglich der Distanz sind gering. Es treten lediglich zwischen 270 und 490 Kilometer vereinzelt Überschätzungen auf (vgl. Abbildung 77 (b)). Der Sörensenindex, der Korrelationskoeffizient nach Pearson und der adjustierte R^2-Wert betragen 0,52, 0,58 und 0,30. Zudem betragen der mittlere absolute Fehler 200 und die Wurzel der mittleren quadratischen Abweichung 1.038 (vgl. Abbildung 22).

Der Vergleich zwischen den empirischen Daten und den vorhergesagten Prognosen mittels finitem Strahlungsmodell (vgl. Abbildung 76 (c)) erzielt sehr gute Anpassungen für ca. 30 und 10.000 Interaktionen. Außerdem ist bei der Abhängigkeit von der Distanz aus Abbildung 77 (c) zu entnehmen, dass kaum Abweichungen zwischen empirischen Daten und Modell existieren. Der Sörensenindex nimmt zudem einen Wert von 0,62 an. Der Korrelationskoeffizient nach Pearson beträgt 0,86 und der adjustierte R^2-Wert 0,75 (vgl. Abbildung 22). Ferner können für den mittleren absoluten Fehler und die Wurzel der mittleren quadratischen Abweichung Werte von 212 und 695 gemessen werden.

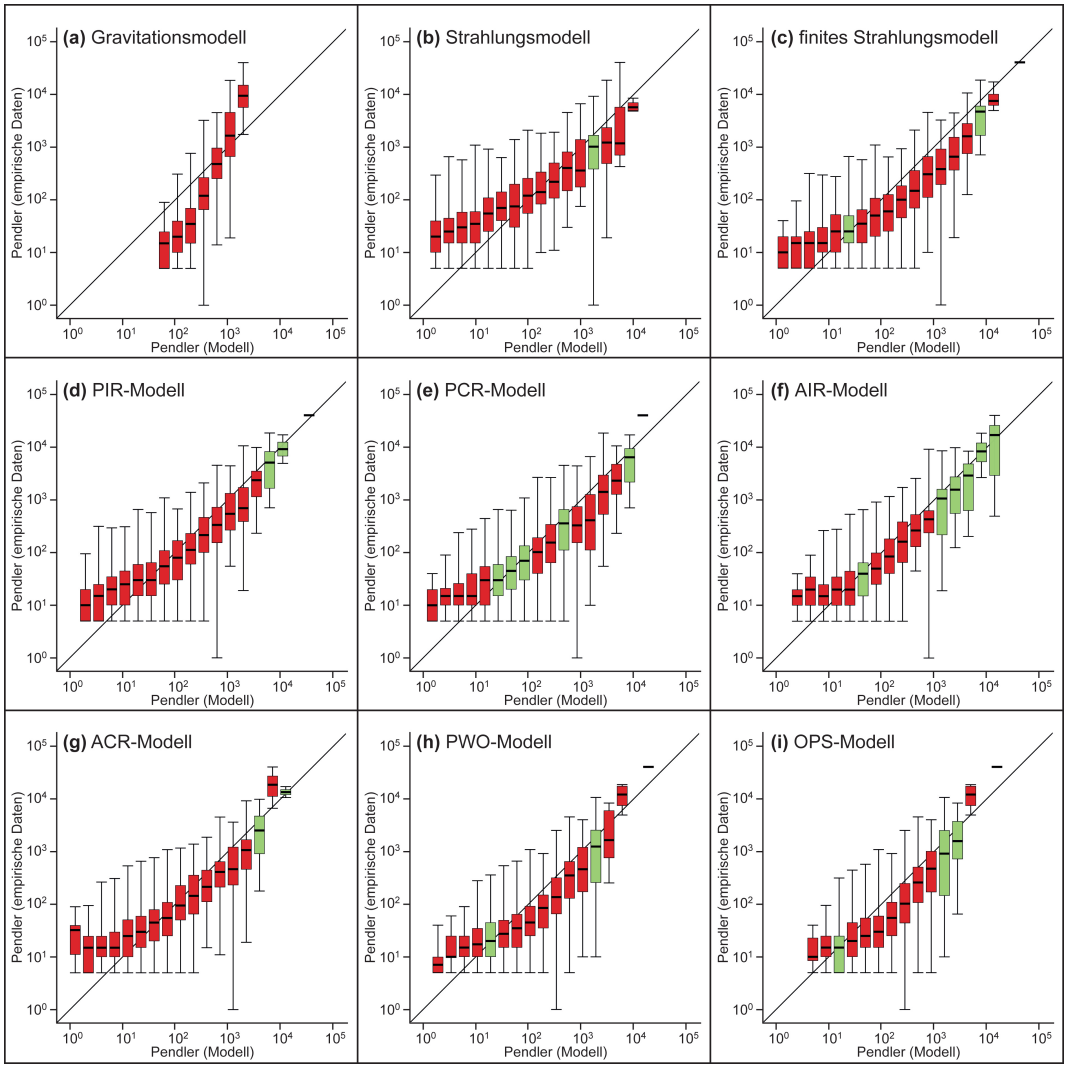


Abbildung 76: Abweichung zwischen den empirischen Daten und den Modellvorhersagen bezüglich der Pendlerbewegungen für Spanien. Quelle: Eigene Darstellung auf Grundlage von eigenen Berechnungen auf Basis von Instituto Nacionale de Estadística (2011).

Unter den verallgemeinerten Strahlungsmodellen erzielt vor allem das PIR-Modell gute Anpassungen. Pendlerbewegungen in der Größenordnung von etwa 5.600 bis 10.000 Interaktionen können besonders gut geschätzt werden (vgl. Abbildung 76 (d)). Zudem ist die Abweichung bezüglich der Distanz gering (vgl. Abbildung 77 (d)). Außerdem lieferen der Sörensenindex mit 0,63, der Korrelationskoeffizient nach Pearson mit 0,88, der adjustierte R^2-Wert von 0,77, der mittlere absolute Fehler von 159 und die Wurzel der mittleren quadratischen Abweichung von 593 einen weiteren Hinweis, dass wenige Abweichungen zwischen empirischen und prognostizierten Daten existieren. Ähnliche Beobachtungen (vgl. Abbildung 22) können auch für das PCR-Modell gemacht werden. Der Sörensenindex beträgt 0,58, der Korrelationskoeffizient nach Pearson 0,72 und der adjustierte R^2-Wert (0,52).

Der mittlere absolute Fehler liegt bei 187 und die Wurzel der mittleren quadratischen Abweichung bei 865. Außerdem erzielen auch die ACR- und AIR-Modelle vergleichbare Ergebnisse (ACR-Modell: Sörensenindex: 0,61; Korrelationskoeffizient nach Pearson 0,66; adjustierter R^2-Wert: 0,44; mittlerer absoluter Fehler: 174; Wurzel der mittleren quadratischen Abweichung: 934 / AIR-Modell: Sörensenindex: 0,61; Korrelationskoeffizient nach Pearson: 0,71; adjustierter R^2-Wert: 0,50; mittlerer absoluter Fehler: 171; Wurzel der mittleren quadratischen Abweichung: 880).

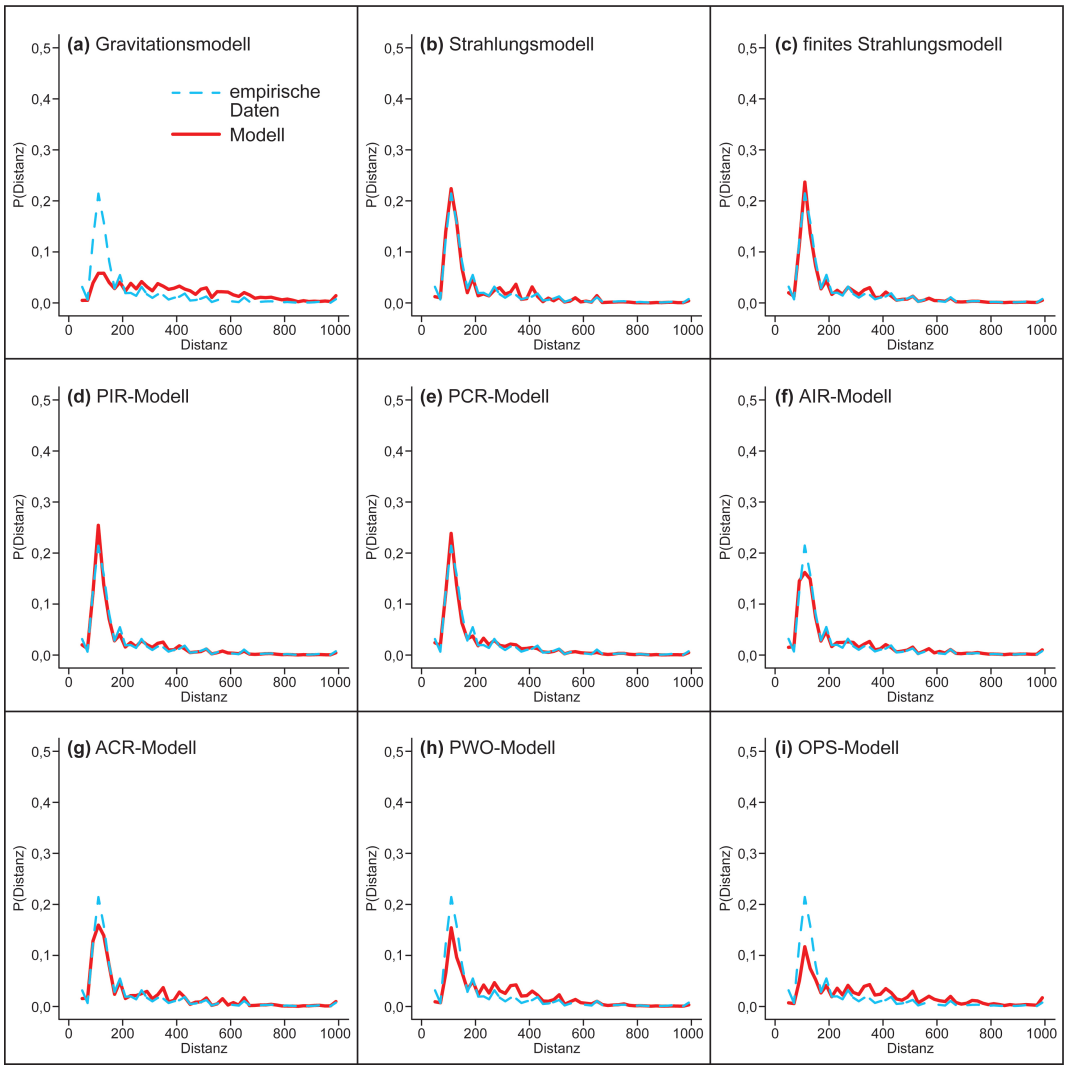


Abbildung 77: P(Distanz) im Verhältnis zur Entfernung (Pendlerbewegung in Spanien). Quelle: Eigene Darstellung auf Grundlage von eigenen Berechnungen auf Basis von Instituto Nacionale de Estadística (2011).

Für das PWO- und OPS-Modell liefern Pendlerströme von ca. 20 und 6.000 Personen besonders gute Vorhersagen (vgl. Abbildung 76 (h) und (i)). Des Weiteren ist in Abbildung 76 (h) und (i) zu entnehmen, dass bis ca. 180 Kilometer Unterschätzungen auftreten. Ab ca. 230 Kilometern liegen Überschätzungen vor. Die Differenz zwischen Modell und empirischen Daten ist jedoch für das OPS-Modell etwas stärker ausgeprägt. Der Sörensenindex nimmt Werte von 0,59 (PWO-Modell) und 0,55 (OPS-Modell) an. Darüber hinaus konnte festgestellt werden, dass beide Modelle einen starken Zusammenhang (PWO-Modell: 0,85; OPS-Modell: 0,82) aufweisen. Die adjustierten R^2-Werte betragen bei 0,73 und 0,67. Außerdem wurde für den mittleren absoluten Fehler bei dem PWO-Modell 183 gemessen. Für das OPS-Modell liegt dieser hingegen bei 201. Die Wurzel der mittleren quadratischen Abweichung nimmt Werte von 763 für das PWO-Modell und 864 für das OPS-Modell an (vgl. Abbildung 22).

Für das Grundmodell des XGBoost-Algorithmus wird in Abbildung 79 (b) deutlich, dass vor allem große Entfernungen zwischen Ziel- und Quellregion einen negativen Einfluss auf die Prognoseleistung haben. Ferner wirken sich hohe Einwohnerzahlen an Wohn- und Arbeitsort positiv auf die Vorhersage aus. In Abbildung 78 (a) ist ersichtlich, dass der Sörensenindex in 71 % aller Simulationen 0,53 beträgt. Außerdem ist für den Korrelationskoeffizient nach Pearson ein Wert von 0,38, für den adjustierten R^2-Wert ein Wert von 0,5, für den mittleren absoluten Fehler ein Wert von 169 und für die Wurzel der mittleren quadratischen Abweichung ein Wert von 663 festzustellen.

Abbildung 79 (a) ist zu entnehmen, dass bei der Erweiterung des Grundmodells für die Distanz sowie die Einwohnerzahlen am Wohn- und Arbeitsort ähnliche Ergebnisse wie im Grundmodell geliefert werden. Zudem kann nicht eindeutig geklärt werden, welchen Einfluss die Einwohnerzahlen zwischen den Standorten haben. Der Sörensenindex beträgt in 74 % der Simulationen 0,55 (vgl. Abbildung 78 (b)), der Korrelationskoeffizient nach Pearson 0,68 und der adjustierte R^2-Wert 0,46. Des Weiteren liegt der mittlere absolute Fehler bei 148 und die Wurzel der mittleren quadratischen Abweichung bei 525 (vgl. Abbildung 24).

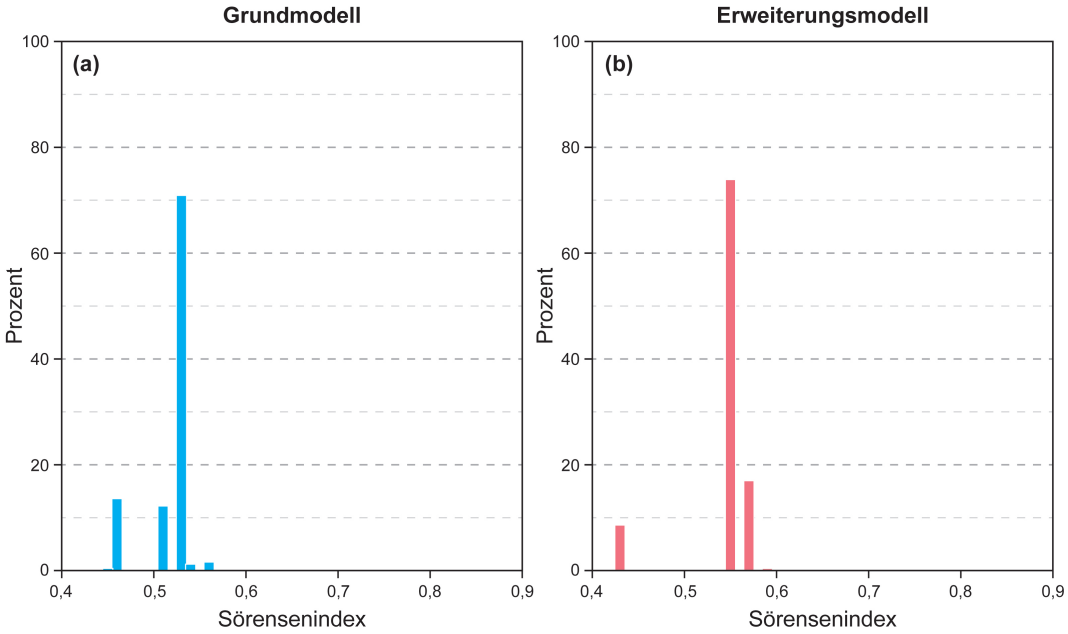


Abbildung 78: Ergebnisse der Simulation für Spanien. Quelle: Eigene Darstellung auf Grundlage von eigenen Berechnungen auf Basis von Instituto Nacionale de Estadística (2011).

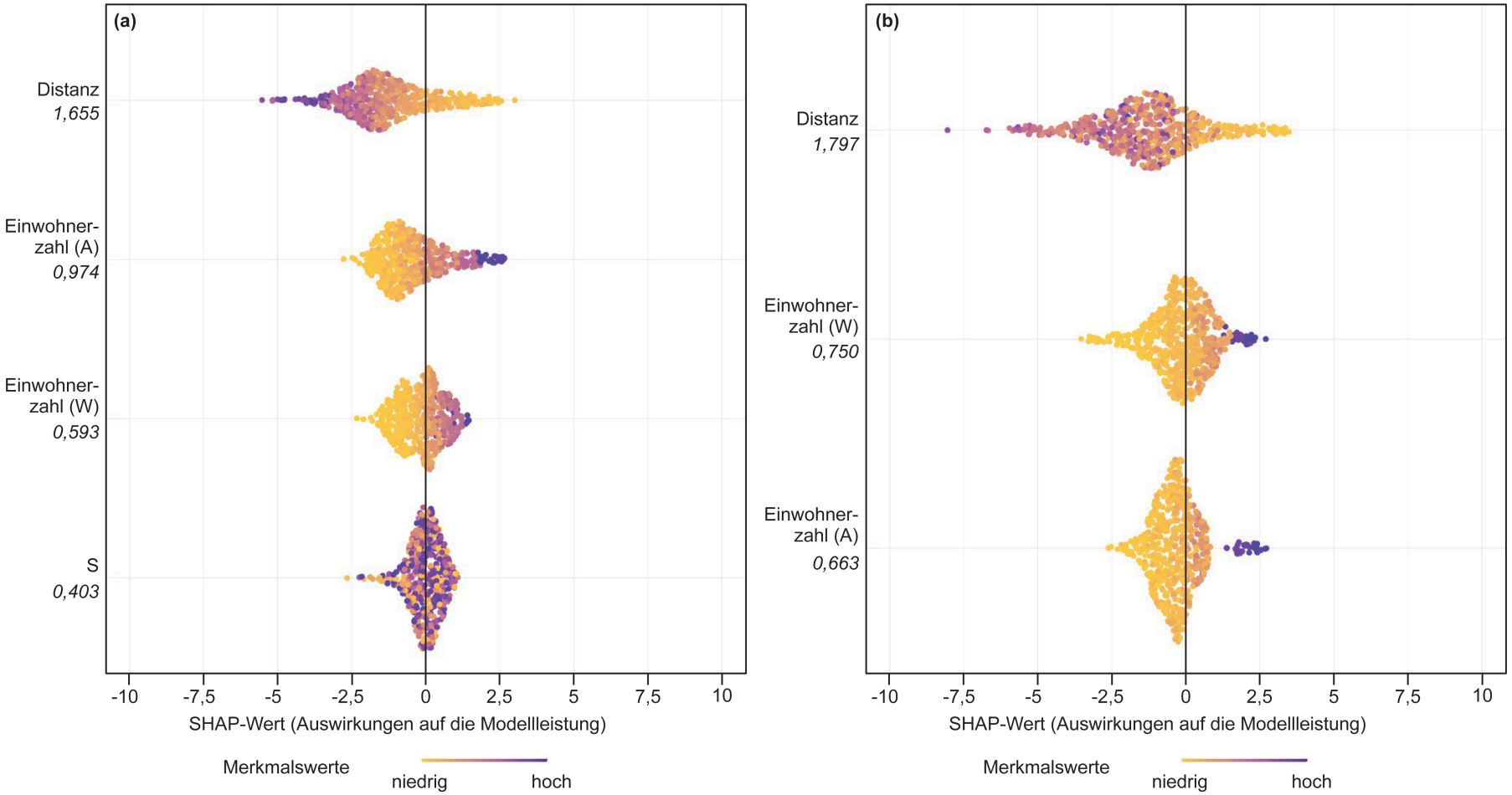


Abbildung 79: SHAP-Werte für den XGBoost-Regressor bei Pendlerbewegungen in Spanien (Grund- und Erweiterungsmodell). Quelle: Eigene Darstellung auf Grundlage von eigenen Berechnungen auf Basis von Instituto Nacionale de Estadística (2011).

7 Zusammenfassung der empirischen Ergebnisse und Schlussfolgerungen

7.1 Diskussion der Prognoseergebnisse

Zusammengefasst lässt sich für die Prognose von studentischen Mobilitätsflüssen feststellen, dass die Analyse mithilfe der Evaluationsindikatoren für das Gravitationsmodell bessere Ergebnisse liefert als das Strahlungsmodell (M1). Zudem konnte gezeigt werden, dass das finite Strahlungsmodell zwar bessere Anpassungen erzielt als das Strahlungsmodell, allerdings die Ergebnisse des Gravitationsmodells nicht erreicht werden (M2). Die verallgemeinerten Strahlungsmodelle liefern im Vergleich aller Teilchenmodelle die besten Befunde (M3), wobei vor allem das AIR- und das ACR-Modell besonders gute Ergebnisse erzielen. Unter den parameterfreien Modellen weist das PWO-Modell deutliche Verbesserungen gegenüber dem Strahlungsmodell auf (M4).

Die Ergebnisse der Prognose von Pendlerbewegungen auf Gemeindeverbandsebenen in Deutschland zeigen, dass das Strahlungsmodell dem Gravitationsmodell überlegen ist. Das finite Strahlungsmodell verbessert die Prognose nocheinmal geringfügig (M1 und M2). Im Gesamtvergleich liefern die verallgemeinerten Modelle (ACR- und AIR-Modell) die besten Anpassungen an die empirischen Daten (M3), während die parameterfreien Modelle lediglich dem Gravitationsmodell überlegen sind (M4).

Eine ähnliche Beobachtung kann auch auf Kreisebene gemacht werden. Das Gravitationsmodell ist dem Strahlungsmodell überlegen (M1). Zusätzlich liefert das finite Strahlungsmodell noch einmal deutliche Verbesserungen (M2). Darüber hinaus erbringen das ACR- und das AIR-Modell die besten Ergebnisse unter den klassischen Teilchenmodellen (M3). Zudem sind auch das PWO- und das OPS-Modell als parameterfreie Modelle dem Strahlungsmodell überlegen (M4), wobei die Anpassungen etwas schlechter sind als die Ergebnisse der verallgemeinerten Strahlungsmodelle.

Für die ausgewählten Länder Europas zeigt sich jedoch ein differenzierteres Bild. Zwar sind in diesem Fall das Gravitations- und das Strahlungsmodell den anderen Modellen unterlegen, allerdings zeigt sich auch, dass insbesondere die verallgemeinerten Strahlungsmodelle gute Anpassungen zwischen empirischen Daten und Modellen liefern. Während in Finnland, Schweden, England und Wales, Frankreich, Schweiz und Österreich das ACR-Modell besonders gu-

te Vorhersagen erzielt, sind für Irland, Spanien, Italien und Kroatien das PIR- und das PCR-Modell besonders gut geeignet. Die Prognose für Norwegen und Dänemark zeigt, dass diese beiden Modelle ebenfalls optimale Anpassungen erzielen (M3). Zudem wird deutlich, dass die parameterfreien Modelle in allen ausgewählten Ländern deutlich bessere Ergebnisse erzielen als das Strahlungsmodell.

Das heißt, insgesamt scheinen sowohl für die Mobilität von Studienanfängern als auch für die Pendler das finite Strahlungsmodell und die verallgemeinerten Strahlungsmodelle, die sich auf Stouffers Theorie der intervenierenden Gelegenheiten beziehen und über einen Skalierungsexponenten verfügen, auf mikro- und makroperspektivischer Ebene besonders gute Anpassungen zu erbringen. Diese Beobachtung ist besonders interessant, da Lenormand et al. (2016, S. 168) auf Gemeindeebene für ausgewählte Länder in Europa und Nord- und Mittelamerika zeigen konnten, dass das Gravitationsgesetz mit der exponentiellen Entfernungsabnahmefunktion den Strahlungsmodellen überlegen ist. Als Hauptgrund wird bei Lenormand et al. (2016, S. 168) angegeben, dass die meisten Fahrten Kurzstrecken sind und der exponentielle Entfernungsabfall mithilfe des Gravitationsgesetzes besser abgebildet werden kann.

Problematisch ist jedoch, die in diesem Zusammenhang bereits von Simini et al. (2012) diskutierte fehlende Parameterfreiheit der verallgemeinerten Strahlungsmodelle. Kang et al. (2015) und Lenormand et al. (2016) gehen davon aus, dass die Einführung des Parameters jedoch zu einer höheren Flexibilität und daraus resultierend zu einer besseren Prognose führt.

Die Betrachtung der Ergebnisse des Grund-, Erweiterungs- und Individualmodells für die Mobilität von Studienanfängern veranschaulicht außerdem, dass alle drei Modelle dem Strahlungsmodell überlegen sind (M5, M6, M7). Das Erweiterungsmodell liefert hingegen bessere Anpassungen als das Grundmodell. Das Individualmodell erzielt noch einmal leichte Verbesserungen.

Außerdem macht die Vorhersage mittels XGBoost-Algorithmus deutlich, dass die Anzahl an Erstsemestern im Vorjahr und die Distanz die wichtigsten Indikatoren bei der Wahl des Hochschulstandortes darstellen (S1, S2, S3).

Diese Beobachtung kann auch für die klassischen Teilchenmodelle gemacht werden. Betrachtet man die Wahrscheinlichkeit einen Hochschulstandort aufzusuchen bezüglich der Distanz, so wird deutlich, dass auch in diesem Fall die verallgemeinerten Strahlungsmodelle besonders geringe Abweichungen zu den

empirischen Daten liefern. Außerdem ergab sich bei der Betrachtung der Resultate, dass auch in Bezug auf die Größe der Hochschulstandorte besonders gute Schätzungen für das AIR- und das ACR-Modell erzielt werden konnten. Das heißt, Modelle, die besonders auf die Anziehungskraft des Hochschulstandortes setzen, liefern vergleichsweise gute Ergebnisse.

Einen Erklärungsansatz liefert die flächendeckende historisch gewachsene Versorgung von Hochschulen. Sie bietet für eine Vielzahl an Studierenden die Möglichkeit, an Standorten in Heimatnähe ein Studium beginnen zu können. In Kapitel 3 wurde bereits diskutiert, dass die Nähe und das bestehende Netzwerk aus Familie und Freunden ein häufiges Motiv für die Wahl des Standortes ist (vgl. u.a. Faggian/Franklin 2014). Darüber hinaus ist die Wahl einer lokalen Hochschule eine kostensparende Alternative, da Lebens- und Umzugskosten vermieden werden (Frenette 2006). Diese Beobachtung scheint auch über die Zeit stabil zu sein, da in den parametrisierten Modellen der Faktor zur Gewichtung der Mobilitätsflüsse im Gesamtzeitraum lediglich für das PCR-Modell kleiner wird.

Zudem ist die Abhängigkeit von der Größe des Standortes vergleichbar mit den Resulaten aus unterschiedlichen Studien. Potenzielle Studierende wandern auf Grund ähnlicher Lebensstile häufig in studentisch geprägte Regionen bzw. in Regionen mit einem hohen Anteil an Hochqualifizierten (Buenstorf et al. 2016; Haussen/Uebelmesser 2018). Zudem ist davon auszugehen, dass Regionen mit einem hohen Urbanisierungsgrad, wie zum Beispiel Berlin, Hamburg oder München mit vielen unterschiedlichen Angeboten sehr attraktiv und anziehend für Studierende sind (Cullinan/Duggan 2016; Sá et al. 2004; Weisser 2019).

Ferner zeigte sich, dass der Entscheidungsprozess ebenfalls stark abhängig von der Anzahl weiterer Möglichkeiten ist (S6). In Regionen, in denen es eine Vielzahl an Universitäten in unmittelbarer räumlicher Nähe gibt, führt dies zu einer Verschlechterung der Prognose. Außerdem wird deutlich, dass in Herkunftsregionen mit vielen Studienanfängern die Wahrscheinlichkeit, einen anderen Hochschulstandort aufzusuchen, sinkt. Hingegen ist in Orten, in denen lediglich eine Universität in unmittelbarer Nähe liegt, davon auszugehen, dass diese eine große Anzahl an Abiturienten absorbiert. Zudem studieren Abiturienten deutlich eher in ihrer Heimatregion, wenn es sich um einen großen Hochschulstandort handelt.

Des Weiteren war die Prognose mit dem Individualmodell besonders vorteil-

haft. Dieser Ansatz ermöglicht, die verwendeten Variablen um individuelle, soziale, hochschulpolitische und ökonomische Motive für studentische Mobilität (z. B. Mietniveau, Arbeitslosigkeit) zu erweitern. Unter den individuellen Merkmalen sind vor allem hohe Haushaltseinkommen und ein hoher Anteil an Naherholungsflächen in Ziel- und Herkunftsregion von positiver Bedeutung bei der Prognose (S4 und S5). Beide Beobachtungen entsprechen den Erkenntnissen aus der Literatur, dass Studierende Standorte mit guten Verdienstmöglichkeiten (SÁ et al. 2004) und einer hohen Lebensqualität aufsuchen (Weisser 2019).

Ein weiterer Vorteil aller vorgestellten Modelle ist die Möglichkeit, die Vorhersagen auf Kreisebene durchzuführen. Dies eröffnet die Chance, dass Universitäten eine ungefähre Vorstellung erlangen, mit wie vielen Studierenden sie in den kommenden Semestern zu rechnen haben. Dies liefert Planungssicherheit und ist für die finanzielle Grundsicherung der Universität von besonderer Bedeutung (HMWK 2015, S. 10 ff.).

Darüber hinaus wird aus den Befunden für die Pendlerbewegungen deutlich, dass alle drei Prognosen mit dem XGBoost-Algorithmus sehr gute Anpassungen an die empirischen Daten erzielen. Die Unterschiede zwischen dem Grund- und Erweiterungsmodell sowie den klassischen Teilchenmodellen auf Gemeindeverbandsebene in Deutschand sind marginal. Auf Kreisebene in Deutschland ist zu erkennen, dass mittels des Ansatzes im Bereich des maschinellen Lernens lediglich für das Grundmodell Verbesserungen der Prognose erzielt werden können. Zudem zeigen die Ergebnisse für Dänemark, Finnland, Norwegen, England und Wales, Frankreich, Österreich, die Schweiz, Kroatien und Italien, dass das Grundmodell des XGBoost-Algorithmus dem Erweiterungsmodell überlegen ist. Für Irland, Schweden und Spanien liefert hingegen das Erweiterungsmodell bessere Anpassungen. Allerdings sind die Verbesserungen gegenüber den Teilchenmodellen, sofern überhaupt vorhanden, gering.

Zudem ist aus den Ergebnissen des Grund- und Erweiterungsmodells erkennbar, dass die Distanz in fast allen Fällen das stärkste Gewicht bei der Prognose annimmt (P1). Dieses Phänomen wird bereits in den traditionellen Migrationsmodellen sowie bei den klassischen Teilchenmodellen (vgl. Kapitel 2.3 und 2.5) aufgegriffen. Große Entfernungen von Städten machen tägliche Pendlerfahrten unmöglich. Somit ist die zurückgelegte Pendlerstrecke nach oben beschränkt und die Pendlerintensität nimmt mit zunehmender Entfernung ab.

Die Anzahl der Interventionsmöglichkeiten und die Einwohnerzahl spielen hin-

gegen eine schwächere Rolle bei der Prognose (P2), wobei die abgeleiteten Tendenzen aus Kapitel 2.4 dennoch angenommen werden können. Große Städte gelten durch ihre hohe Einwohnerzahl und die Agglomerationsvorteile als besonders anziehend für Arbeitnehmer (Heuermann/Schmieder 2019; Martinus et al. 2020). Zudem ist den Ergebnissen zu entnehmen, dass, wie bereits im theoretischen Ansatz von Stouffer beschrieben, die Wahrscheinlichkeit einen Standort auszuwählen, sinkt, wenn zunehmende Angebote für Arbeitsplätze entstehen (P6).

Insgesamt stellt sich dennoch die Frage, warum das Erweiterungsmodell nicht für alle betrachteten Länder Verbesserungen liefert. Ein möglicher Erklärungsansatz ist die geringe Fallzahl an NUTS3-Regionen. So existieren in Irland im Jahr 2011 acht Regionen, das heißt es liegen 56 Interaktionen vor. Ferner ist denkbar, dass die Schwächen des XGBoost-Algorithmus bedingt durch den Zuschnitt der NUTS3-Regionen entstanden sind. In NUTS3-Regionen liegt die Bevölkerungszahl generell zwischen 150.000 und 800.000 Einwohnern (Statistisches Bundesamt 2022b), das heißt, bei Ländern mit geringen Bevölkerungsdichten wie Irland (67 Einwohner je Quadratkilometer) weisen die Regionen eine große Fläche auf. Marcińczaka/Bartosz (2018) konnten diese bereits zeigen, so scheint die Fläche einer Region Auswirkung auf die Entfernung und somit auf den Umfang des Pendelns zu haben. Zudem wurde deutlich, dass ein höheres Maß an Polyzentralität kürzere Pendelfahrten fördert (Marcińczaka/Bartosz 2018) und somit nicht ausgeschlossen werden kann, dass unterschiedliche Bevölkerungsverteilung zu unvermeidbaren Verzerrungseffekten führen.

Die Erweiterung der Prognose mit dem XGBoost-Algorithmus um die individuellen Aspekte für das Pendlerverhalten auf Gemeindeverbandsebene liefert zudem Hinweise für eine Verbesserung der Vorhersage. Die Integration der weiteren individuellen Merkmale steigert die Prognoseleistung geringfügig. Neben den klassischen Merkmalen (Distanz, intervenierende Möglichkeiten, Einwohnerzahl am Wohn- und Arbeitsort) ist die Gewerbesteuereinnahme das wichtigste Merkmal (P4). Gemeindeverbände mit hohen Gewerbesteuereinnahmen sind oftmals durch eine Vielzahl an Unternehmen geprägt und bieten somit gute Berufsmöglichkeiten. Als Konsequenz ziehen sie mehr Pendler an.

Des Weiteren liefert die Prognose mittels des Individualmodells des XGBoost-Algorithmus auf Ebene der Kreise keine nennenswerten Verbesserungen. Dennoch können interessante Hinweise aus dem Individualmodell abgeleitet wer-

den. Auf individueller Ebene scheinen insbesondere die Erwerbstätigkeit von Frauen am Wohnort, die Erreichbarkeit von Oberzentren, der Anteil an Erholungsflächen sowie die Einwohnerdichte am Wohnort ein großes Gewicht bei der Vorhersage anzunehmen (P3, P5). Erst nach Abwägung dieser Aspekte berücksichtigen Pendler in dem vorliegenden Modell die möglichen Alternativen zwischen Ziel- und Quellregion. Während auf kleinräumiger Ebene vor allem die Distanz und die Anzahl an Möglichkeiten eine Rolle bei der Wahl des Arbeitsplatzes spielt, so scheint auf der Ebene der Kreise eine deutlich differenziertere Abwägung stattzufinden. Mögliche Ursache für diese Beobachtung ist, dass insbesondere gut qualifizierte Arbeitnehmer mit einem hohen Einkommen große Pendlerstrecken auf sich nehmen, um einen geeigneten Arbeitsplatz zu finden (Kalter 1994; Pfaff 2012; Sandow 2014; Broersma et al. 2020). Diese Arbeitsplätze liegen oftmals in Agglomerationsräumen bzw. den nächstgelegenen Oberzentren (vgl. Einig/Pütz 2007), da die Ballungsräume, bedingt durch ihre Innovationsfähigkeit, besonders viele Arbeitsmöglichkeiten und Spezialisierungen anbieten.

Trotzdem muss diskutiert werden, warum das Individualmodell vor allem auf Kreisebene nur bedingt gute Anpassungen liefert. Möglicher Erklärungsansatz ist, dass die vorliegende Analyse auf Grund von fehlenden Daten keine individuellen Faktoren, wie beispielsweise den Bildungshintergrund und die Anzahl der Kinder im Haushalt, beim Entscheidungsprozess berücksichtigt. Zwar wird versucht, diese Aspekte durch andere Variablen abzubilden, dennoch ist nicht ausgeschlossen, dass sie die Realität nur bedingt widerspiegeln und das Erklärungspotenzial des Modells nicht voll ausgeschöpft werden kann. Außerdem sind Variablen, die das Bruttoinlandsprodukt, das Mietniveau und die Wohnqualität abbilden, auf Gemeindeverbandsebene nicht vorhanden, sodass auf die Gewerbesteuereinnahmen, die Anzahl der neu gebauten Wohnungen und der Anteil der Naherholungsflächen zurückgegriffen werden musste. Das heißt, es ist davon auszugehen, dass das Hinzufügen dieser Indikatoren die Modelle noch einmal verbessern würde. Eine Übersicht der gesamten untersuchten Hypothesen findet sich in Tabelle 4.

Region	M1	M2	M3	M4	M5	M6	M7	S1	S2	S3	S4	S5	S6	P1	P2	P3	P4	P5	P6
Deutschland (S)	×	×	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	-	-	-	-	-	-
Deutschland (Gemeindeverband,P)	✓	✓	✓	×	✓	✓	×	-	-	-	-	-	-	✓	✓	✓	✓	✓	✓
Deutschland (Kreise, P)	✓	✓	✓	✓	✓	×	✓	-	-	-	-	-	-	✓	✓	✓	✓	×	✓
England und Wales (P)	✓	×	✓	×	✓	×	-	-	-	-	-	-	-	✓	✓	-	-	-	✓
Frankreich (P)	✓	✓	✓	✓	✓	×	-	-	-	-	-	-	-	✓	✓	-	-	-	✓
Irland (P)	×	✓	✓	✓	✓	✓	-	-	-	-	-	-	-	✓	✓	-	-	-	×
Österreich (P)	×	×	✓	✓	✓	×	-	-	-	-	-	-	-	✓	✓	-	-	-	✓
Schweiz (P)	✓	✓	✓	✓	✓	×	-	-	-	-	-	-	-	✓	✓	-	-	-	✓
Dänemark (P)	×	×	✓	✓	✓	×	-	-	-	-	-	-	-	✓	✓	-	-	-	×
Finnland (P)	×	×	✓	✓	✓	×	-	-	-	-	-	-	-	✓	✓	-	-	-	×
Norwegen (P)	×	✓	✓	✓	✓	×	-	-	-	-	-	-	-	✓	✓	-	-	-	×
Schweden (P)	×	✓	✓	✓	✓	✓	-	-	-	-	-	-	-	✓	✓	-	-	-	✓
Kroatien (P)	×	✓	✓	✓	✓	×	-	-	-	-	-	-	-	✓	✓	-	-	-	×
Italien (P)	✓	✓	✓	×	✓	×	-	-	-	-	-	-	-	✓	✓	-	-	-	×
Spanien (P)	✓	✓	✓	✓	✓	✓	-	-	-	-	-	-	-	✓	✓	-	-	-	×

✓wird nicht verworfen; × wird verworfen; - wird nicht untersucht; S Studienanfänger; P Pendler

Tabelle 4: Übersicht über die Hypothesen. Quelle: Eigene Darstellung.

7.2 Anwendungen der Prognoseinstrumente in der Praxis

Aus den Ergebnissen für die Prognosse von Mobilitätsströmen lassen sich folgende Anwendungen der Prognoseinstrumente in der Praxis ableiten:

Bedingt durch die große Bedeutung von räumlicher Nähe beim Entscheidungsprozess für den Hochschulstandort kann angenommen werden, dass Hochschulen ihre Studierenden in Numerus Clausus freien Fächern aus dem direkten Umfeld generieren. Da die Distanz ein nicht veränderliches Merkmal ist und außerdem die Anzahl an Studienanfängern eine überdurchschnittliche Rolle bei der Entscheidung von Studierenden spielt, sollte versucht werden, die Immatrikulationen an den jeweiligen Universitäten zu erhöhen. In diesem Zusammenhang sollte der Fokus - bedingt durch die Bedeutung von räumlicher Nähe - auf gezielten Marketingmaßnahmen im direkten Umfeld liegen. Denkbar sind hier neben klassischen Publikationsformen, wie Broschüren oder Informationsblätter, auch interaktive Kommunikationsangebote (vgl. Brune et al. 2009), wie zum Beispiel Diskussionsforen, Podcasts, Vorträge in Schulen oder Exkursionen für Schulklassen. Diese Maßnahmen fördern schon vor Beginn des Studiums den Aufbau einer Verbindung zur jeweiligen Hochschule und können sogar negative Auswirkungen von geographischer Distanz reduzieren (Raab et al. 2018).

Die Ergebnisse der Pendlerprognosen lassen sich darüber hinaus unter anderem für Planungsvorhaben einsetzen. So ist - bedingt durch die Stabilität der Modelle - denkbar, zukünftige Mobilitätsflüsse zum Beispiel auf Basis der Bevölkerungsprognose zu berechnen. Dies ermöglicht Kommunen zum einen bei der Planung von neuen Wohn- oder Gewerbegebieten das zukünftige Pendleraufkommen zu berechnen. Zum anderen wird es durch die prognostizierten Einzugsgebiete möglich, gezielt mit Verkehrsträgern über mittel- bis langfristige ÖPNV-Verbindungen zu verhandeln bzw. ÖPNV-Angebote einzurichten, um langfristig eine nachhaltige Mobilitätswende zu erzielen.

Allerdings muss den Kommunen und Universitäten ein einfaches Tool zur Verfügung gestellt werden, dass ihnen ermöglicht, die Mobiliätsströme durch eine Eingabe von Bevölkerungsprognosen (z. B. vom Statistischen Landes- und Bundesamt) und Studienanfängerzahlen aus den Vorjahren zu berechnen.

7.3 Limitationen

Insgesamt weist die Prognose der Mobilitätsströme für Studienanfänger und Pendler trotz aller Vorteile auch Schwächen auf. Im Fokus sind dabei die Limitationen bei der methodischen Vorgehensweise.

Problematisch ist in diesem Zusammenhang vor allem, dass durch die Normierung (vgl. Kang et al. 2015; Masucci et al. 2013; Yang et al. 2014; Liu/Yan 2019) insbesondere die Größe an Studierenden bzw. Einwohnern geschätzt wird, die bereits im Vorjahr ein Studium aufgenommen haben oder in einer Region wohnen.

Kritisch anzumerken bleibt zudem, dass die vorliegende Analyse auf Grund von fehlenden Daten keine individuellen Faktoren, wie beispielweise Geschlecht, Bildungshintergrund und familiäre Bindung, berücksichtigt. Zudem werden auch bei den sozioökomischen Push- und Pull-Faktoren nicht alle Apekte abschließend berücksichtigt. So sind zum Beispiel insbesondere für Großstädte steigende Immobilienpreise von besonderer Bedeutung (u. a. Adam 2019). Allerdings wurden die Miet- und Bodenpreise nicht in der Analyse berücksichtigt, da sie auf Gemeindeverbandsebene nicht zur Verfügung gestellt werden.

Bei der Betrachtung der Ergebnisse ist ferner kritisch anzumerken, dass mit den vorgestellten Instrumenten nicht auf politische Entscheidungen reagiert werden kann. So können beispielsweise Beschlüsse, wie die Aussetzung der Wehrpflicht im Jahr 2012, nicht in die Prognosen der Studierendenmobilität integriert werden. Damit unterliegen die Vorhersagen immer einem nicht einkalkulierten Fehler.

Abschließend ist, wie bereits in Kapitel 4 thematisiert, noch einmal auf die Problematik der unterschiedlichen Erhebungsformen von Pendlerdaten in Europa hinzuweisen. Die Resultate aus den verschiedenen europäischen Ländern sind daher auf Grund der fehlenden Harmonisierung nicht abschließend vergleichbar. Zwar wurden die Modelle für das jeweilige Land extra geschätzt, dennoch können Verzerrungseffekte nicht komplett ausgeschlossen werden.

7.4 Zukünftiger Forschungsbedarf

Aus den Befunden dieser Dissertation sowie den aufgezeigten Limitationen lässt sich folgender Forschungsbedarf ableiten: Zunächst ist in Bezug auf die Mobilitätsprognosen anzumerken, dass weiterführende Analysen auf kleinräumiger Ebene gemacht werden müssen. Da bereits für Deutschland

gezeigt wurde, dass die Modelle auf Gemeindeverbandsebene besonders gute Anpassungen erzielen. Daher sollte untersucht werden inwieweit diese Beobachtung auch auf andere Länder in Europa zutrifft. Zudem ist denkbar, mit Hilfe des Bevölkerungsrasters (Censusdaten), innerhalb einer Stadt bis auf die Quartiersebene Untersuchungen durchzuführen und Mobilitätsströme auf Mikroebene zu berechnen.

Ferner ist in der Literaturauswertung in Kapitel 2.4 deutlich geworden, dass das Mobilitätsverhalten häufig vom Geschlecht abhängig ist. Frauen mit Kindern pendeln weniger häufig und die zurückgelegte Strecke ist oftmals kürzer (Häfner 2011), sie haben aber dennoch in weiten Teilen Europas einen signifikanten positiven Einfluss auf die wirtschaftliche Leistung eines Landes (Litmeyer et al. 2022). In Bezug auf die Mobilität von Studierenden liefern empirische Studien hingegen ein differenziertes Bild. In Italien konnte gezeigt werden, dass Studentinnen immobiler sind als männliche Studierende (Ciriaci 2014; Genova et al. 2021). Hingegen wurden bei Belfield/Moriris (1999) in Großbritannien keine signifikanten Unterschiede zwischen den Geschlechtern festgestellt. Die bisherigen Studien beschreiben und erklären allerdings lediglich die Unterschiede zwischen männlicher und weiblicher Mobilität. Geschlechtsbezogene Vorhersagen von Mobilitätsflüssen werden hingegen nicht vorgenommen. Daher wäre denkbar, mit den vorgestellten Methoden (vgl. Kapitel 2.5) jeweils nach Geschlecht differenzierte Mobilitätsflüsse für Pendler und Studierende zu prognostizieren und die Eignung dieser Modelle zu testen.

Ein weiterer Aspekt, auf den im Rahmen dieser Arbeit aber nicht genauer eingegangen wird, ist, dass menschliche Mobilität in Zusammenhang mit infektiösen Krankheitsverbreitungen steht (Viboud et al. 2006; Balcan et al. 2009; Tizzoni et al. 2014; Sallah 2017; Marshall et al. 2018; Tuite et al. 2018). Allerdings nutzen die bisher exisitierenden Studien vor allem das klassische Gravitations- bzw. das Strahlungsmodell von Simini et al. (2012). Diese Beobachtung ist insbesondere vor dem Hintergrund der Corona-Pandemie interessant. Da die Auswertungen in Kapitel 5 und 6 dieser Dissertation zeigten, dass insbesondere die verallgemeinerten Strahlungsmodelle besonders gut für die Vorhersage von studentischer Mobilität und Pendlerbewegungen geeignet sind, sollte versucht werden, die Migrationsbewegungen bzw. Pendlerflüsse mit allen vorgestellten Modellen zu schätzen und mit diesen Vorhersagen das Infektionsgeschehen und die Ausbreitung des Corona-Virus zu modellieren.

Zudem sollte vor dem Hintergrund der aktuellen Corona-Pandemie und den daraus resultierenden digitalen Lehrangeboten von Hochschulen mit aktuellen Studierendenzahlen überprüft werden, inwieweit die räumlichen Mobilitätsmuster von Studierenden auch in Zukunft noch gültig sind.

Darüber hinaus sollte langfristig untersucht werden, inwieweit sich die räumlichen Mobilitätsmuster von Pendlern bedingt durch die Corona-Pandemie verändern. Insbesondere sind viele hochqualifizierte Beschäftigte während der Pandemie auf Telearbeit umgestiegen. Florida et al. (2020) gehen davon aus, dass in diesen Arbeitsbereichen auch in Zukunft eine Mischform von Arbeiten im Home-Office und im Büro existieren wird. Dies hätte einen erheblichen Einfluss auf das Mobilitäts- und Verkehrsgeschehen (Florida et al. 2020, S. 12 f.).

Ein weiterer Aspekt, der aus methodischer Perspektive interessant ist, ist der Einsatz von Persistenten Homologien zur Identifikation von Mobilitätsmustern. Dieser Ansatz ist eine relativ neue Methode in der Datenanalyse und ist auf Überlegungen aus der Graphentheorie in den 1990er Jahren zurückzuführen. Die Grundidee basiert auf der Konstruktion von impliziten Komplexen in einem gerichteten Netzwerk. Mit Hilfe von Werkzeugen aus der algebraischen Topologie können dann topologische Merkmale extrahiert werden. Diese ermöglichen die Identifikation von Flussmustern im gerichteten Netzwerk. Besonders vorteilhaft an diesem Ansatz ist es, dass die Beschreibung höherer Ordnung der gesamten Netzwerkflüsse leicht möglich ist (Ignacio/Darcy 2019). Ignacio/Darcy (2019) untersuchten mit diesem Verfahren der Mustererkennung die asiatischen Nettomigrations- und Überweisungsnetzwerke aus dem Jahr 2015 und konnten klare Migrationsströme und Cluster identifizieren. Denkbar ist daher die Übertragung auf die Pendler- und Studierendenströme. Im Idealfall wäre es so möglich, „neue“ Arbeitsmarktregionen oder auch Einzugsgebiete von Universitäten zu identifizieren und ökonomische und gesellschaftliche Auswirkungen sowohl in regionalen als auch in globalen Umgebungen zu verstehen.

8 Fazit

Insgesamt wurde bei der Prognose der Mobilitätsflüsse von Studierenden deutlich, dass vor allem die verallgemeinerten Strahlungsmodelle sowie die parameterfreien Modelle besonders gute Anpassungen an die empirischen Daten liefern. Darüber hinaus erzielen die Modelle basierend auf dem XGBoost-Algorithmus noch einmal deutliche Verbesserungen. Bereits für das Grund- und Erweiterungsmodell konnten geringe Abweichungen zwischen empirische Daten und Modellvorhersage gemessen werden. Zudem verbessert die Berücksichigung von individuellen Variablen auf Kreisebene nocheinmal die Prognose der Mobilitätsströme.

Bei der Prognose der Pendlerverflechtungen zeigt sich jedoch ein etwas differenzierteres Bild. Auf Gemeindeverbands- und Kreisebene in Deutschland liefert das Strahlungsmodell bessere Vorhersagen von Mobilitätsströmen als das Gravitationsmodell. Im europäischen Vergleich kann dieser Befund lediglich für Frankreich, England und Walse, die Schweiz, Spanien sowie Italien sowie festgestellt werden. In allen anderen untersuchen Ländern ist das Gravitationsmodell dem Strahlungsmodell überlegen. Allerdings sind, ähnlich wie bei den Mobiliätsströmen der Studierenden, bei den verallgemeinerten Strahlungsmodellen sowie den parameterfreien Modellen deutliche Verbesserungen festzustellen. Die Ergebnisse des XGBoost-Algorithmus machen deutlich, dass zwar auf Gemeindeverbandsebene mittels Individualmodell Verbesserungen erzielt werden können, europaweit erbringt jedoch das Grundmodell bis auf einzelne Ausnahmen die besten Vorhersagen.

Zusammengefasst bietet das Thema dieser Dissertation trotz der Limitationen von Datenverfügbarkeit und modellspezifischen Einschränkungen auch in Zukunft Anknüpfungspunkte. Im Vordergrund steht neben weiteren Untersuchungen auf unterschiedlichen räumlichen Ebenen die Anwendung der vorgestellten Teilchenmodelle sowie der XGBoost-Agorithmus auf geschlechtsspezifische Mobilität von Studienanfängern und Pendlern. Denkbar ist auch die Anwendung der Vorhersagen auf die Verbreitung von Infektionskrankheiten. Zudem bietet der Einsatz von neuen Verfahren aus der Topologie Möglichkeiten, räumliche Interaktionen besser beschreiben zu können.

9 Literaturverzeichnis

Adam, B. (2019): Vom Siedlungsbrei zum Städtischen? Eine mehrdimensionale Bestandsaufnahme der Suburbanisierung. In: Raumforschung und Raumordnung — Spatial Research and Planning 77 (1), 35–55. doi: 10.2478/rara-2019-0001.

Adkisson, R. V. / Peach, J. (2008): Non-resident enrolment and non-resident tuition at land grant colleges and universities. In: Educ Econ 16 (1), 75–88. doi: 10.1080/09645290701563156.

Aguilera, A. (2005): Growth in commuting distances in French polycentric metropolitan areas: Paris, Lyon and Marseille. In: Urban Studies 42 (9), 1537–1547. doi: 10.1080/00420980500185389.

Albrecht, J. / Scheiner, J. (2022): Gehen, Bleiben und Zurückkommen - der Ort der Kindheit und Jugend: Wohnstandortwahl von Paaren zur Zeit der Familiengründung. In: Raumforschung und Raumordnung — Spatial Research and Planning. doi: 10.14512/rur.132.

Alm, J. / Winters, J. V. (2009): Distance and intrastate college student migration. In: Economics of Education Review 28 (6), 728–738. doi: 10.1016/j.econedurev.2009.06.008.

Alonso, W. (1964): Location and Land Use: Toward a General Theory of Land Rent. Cambridge, MA: Harvard University Press.

Alonso, W. (1973): Urban zero population growth. In: Daedalus 102 (4), 191–206.

Arrow, K. (1962): The economic implications of learning by doing. In: Review of Economic Studies 29, 155–172.

Audretsch, D. / Lehmann, E. / Warning, S. (2004): University Spillovers: Does the Kind of Science Matter. In: Industry and Innovation 11, 193-206. doi: 10.1080/1366271042000265375.

Audretsch, D. / Lehmann, E. (2005): Does the Knowledge Spillover Theory of Entrepreneurship Hold for Regions. In: Research Policy 34, 1191-1202. doi: 10.1016/j.respol.2005.03.012.

Auerbach, F. (1913): Das Gesetz der Bevölkerungskonzentration. In: Petermann's Geographische Mitteilungen 59, 74–76.

Auspurg, K. / Schönholzer, T. (2013): An Heim und Herd gebunden? Zum Einfluss von Pendelstrecken auf geschlechtsspezifische Lohnunterschiede. In: Zeitschrift für Soziologie 42 (3), 138–156. doi: 10.1515/zfsoz-2013-0205.

Bade, F. / Nerlinger, E. (2000): The spatial distribution of new technology- based firms: empirical results for West Germany. In: Papers in Regional Science 79, 155-176. doi: 10.1007/s101100050041.

Bailey, N. (2015): The migration processes of students into higher educational institutions in the United Kingdom. Dissertation, University of Southampton.

Balcan, D. / Colizza, V. / Goncalves, B. / Hud, H. / Ramasco, J. / Vespignani, A. (2009): Multiscale mobility networks and the spatial spreading of infectious diseases. In: Proceedings of the National Academy of Sciences 106 (51), 21484-21489. doi: 10.1073/pnas.0906910106.

Batten, D. F. (1995): Network cities: creative urban agglomerations for the 21st century. In: Urban Studies 32, 313–327. doi: 10.1080/00420989550013103.

Bartel, A. P. (1979): The migration decision: What role does job mobility play? In: American Economic Review 69, 775-786.

Bathelt, H. / Malmberg, A. / Maskell, P. (2004): Clusters and Knowledge: Local Buzz. Global Pipelines and the Process of Knowledge Creation. In: Progress in Human Geography 28, 31-56. doi: 10.1191/0309132504ph469oa.

Belfield, C. / Morris, Z. (1999): Regional migration to and from higher education institutions: scale, determinants and outcomes. In: Higher Education Quarterly 53 (3), 240-263. doi: 10.1111/1468-2273.00129.

Bettencourt, L. M. / Lobo, J. / Helbing, D. / Kühnert, C. / West, G. B. (2007): Growth, innovation, scaling, and the pace of life in cities. In: Proceedings of the National Academy of Sciences 104 (17), 7301–7306. doi: 10.1073/pnas.0610172104.

Bergstrand, J. (1985): The gravity equation in international trade: some microeconomic foundations and empirical evidence. In: The Review of Economics and Statistics 67, 474–481. doi: 10.2307/1925976.

Berry, B. J. L. (1964): Cities as systems within systems of cities. Papers of the Regional. In: Science Association 13, 147–163.

Bourne L. S. / Simmons, J. W. (1978): Introduction: The urban system as a unit of analysis, In: Bourne L. S. / Simmons, J.W. (Hrsg.): Systems of Cities: Reading on Structure, Growth, and Policy. New York: Oxford University Press. 1–18.

Buch, T. / Hamann, S. / Niebuhr A. / Rossen, A. (2014): What Makes Cities Attractive? The Determinants of Urban Labour Migration in Germany. In: Urban Studies 51 (9), 1960–1978. doi: 10.1177/0042098013499796.

Buenstorf, G. / Geissler, M. / Krabel, S. (2016): Locations of labor market entry by German university graduates: is (regional) beauty in the eye of the beholder?. In: Review of Regional Research 36 (1), 29–49. doi: 10.1007/s10037-015-0102-z.

Bundesinstitut für Bau-, Stadt- und Raumforschung (BBSR) (2022): Umstiegsschlüssel für konsistente Zeitreihen. https://www.bbsr.bund.de/BBSR/DE/forschung/raumbeobachtung/umstiegsschluessel/umstiegsschluessel.html?nn=2544954 (24. März 2022).

Bundesagentur für Arbeit (2017): Sozialversicherungspflichtig Beschäftigte - Einpendler nach Gemeinden.

Bundesamt für Statistik (2014): Erwerbstätige nach Wohn- und Arbeitsgemeinde, 2014. https://www.bfs.admin.ch/asset/de/ts-x-11.04.04.05-2014 (23. März 2022).

Broersma, L. / Edzes, A. / van Dijk, J. (2020): Commuting Between Border Regions in The Netherlands, Germany and Belgium: An Explanatory Model. In: Journal of Borderland Studies 37 (3), 551–573. doi: 10.1080/08865655.2020.1810590.

Brune, G. / Seyfarth, F. / Spoun, S. (2009): Hochschulmarketing in Zeiten von Blogs, Wikipedia und YouTube: Zur prozessanstoßenden Kommunikation für öffentliche Universitäten. In: Virale Kommunikation: Möglichkeiten und Grenzen des prozessanstoßenden Marketings, Sascha Spoun, 215–244.

Burger, M. / Meijers, E. (2012): Form follows function? Linking morphological and functional polycentricity. In: Urban Studies 49 (5), 1127–1149. doi: 10.1177/0042098011407095.

Burger, M. / Van der Knaap, B. / Wall, R. S. (2013): Polycentricity and the multiplexity of urban networks. In: European Planning Studies 22 (4), 816–840. doi: 10.1080/09654313.2013.771619.

Cahuc, P. / Zylberberg, A. (2004): Labor Economics. Cambridge, MA: The MIT Press.

Camagni, R. / Salone, C. (1993): Network urban structures in northern Italy: elements for a theoretical framework. In: Urban Studies 30, 1053–1064. doi: 10.1080%2F00420989320080941.

Capineri, C. / Kamann, D. J. F. (1998): Synergy in networks: concepts. In: Button, K. / Nijkamp, P. / Priemus, H. (Hrsg.): Transport Networks in Europe. Cheltenham: Edward Elgar. 35–56.

Cappello, R. / Rietveld, P. (1998): The concept of network synergies in economic theory: policy implications. In: Button, K. / Nijkamp, P. / Priemus, H. (Hrsg.): Transport Networks in Europe. Cheltenham: Edward Elgar, 57–83.

Carey, H. C. (1858): Principles of Social Science. Lippincott.

Central Statistics Office (2011): Population Aged 15 Years And Over At Work, Usually Resident And Present In The State By Regional Authority Of Usual Residence And Regional Authority Of Work.https://www.cso.ie/en/census/census2011reports/census2011profile10doortodoorcommutinginireland/ (23. März 2022).

Champion, A. G. (2001): A changing demographic regime and evolving polycentric urban regions: Consequences for the size, composition and distribution of city populations. In: Urban Studies 38 (4), 657–677. doi: 10.1080/00420980120035277.

Chatterjee, K. / Chng, S. / Clark, B. / Davis, A. / Vos, J. / Ettema, D. / Handy, S. / Martin, A. / Reardon, L. (2020): Commuting and Wellbeing: A Critical Overview of the Literature with Implications for Policy and Future Research. In: Transport Reviews 40 (1), 5–34. doi: 1080/01441647.2019.1649317.

Christaller, W. (1968): Die Zentralen Orte in Süddeutschland (2. Auflage). Darmstadt: Wissenschaftliche Buchgesellschaft.

Ciriaci, D. (2014): Does university quality influence the interregional mobility of students and graduates? The case of Italy. In: Reg Stud 48 (10), 1592–1608. doi: 10.1080/00343404.2013.821569.

Croatian Bureau of Statistics (2011): Census of Population, Households and Dwellings 2011. https://web.dzs.hr/arhiva_e.htm (19. September 2019).

Cooke, T. J. / Boyle, P. (2011): The migration of high school graduates to college. In: Educational Evaluation and Policy Analysis 33 (2), 202–213. doi: 10.3102/0162373711399092.

Cullinan, J. / Duggan, J. (2016): A School-Level Gravity Model of Student Migration Flows to Higher Education Institutions. In: Spatial Economic Analysis 11 (3), 294–314. doi: 10.1080/17421772.2016.1177195.

Dauth, W. / Haller, P. (2018): Berufliches Pendeln zwischen Wohn- und Arbeitsort: Klarer Trend zu längeren Pendlerdistanzen. IAB-Kurzbericht. Nürnberg: Institut für Arbeitsmarkt- und Berufsforschung (IAB), No. 10.

Danmarks Statistik (2011): PEND100: Beskæftigede (ultimo november) efter bopælskommune, arbejdsstedsområde og køn. https://www.statistikbanken.dk/statbank5a/SelectVarVal/Define.asp?MainTable=PEND100&PLanguage=0&PXSId=0&wsid=cflist (23. März 2022).

De Goei, B. / Burger, M. / van Oort, F. / Kitson, M. (2010): Functional polycentrism and urban network development in the Greater South East, United Kingdom. Evidence from commuting patterns, 1981–2001. In: Regional Studies 44, 1149–1170. doi: 10.1080/00343400903365102.

Deutscher Bundestag (2011): Aussetzung der allgemeinen Wehrpflicht beschlossen. https://www.bundestag.de /dokumente/textarchiv/2011/33831649_kw12_de_wehrdienst-204958 (10. Juni 2019).

Dijst, M. J. / Cortie, C. (1988): Van 'low' naar'high': een voorstudie naar de ontwikkeling en toekomstmogelijkheden van polycentrische stedelijke systemen in Nederland, Belgie en Noord-Rijn Westfalen. Amsterdamse Sociaal-Geografische Studies No. 15, Instituut voor Sociale Geografie, Universiteit van Amsterdam.

Dodd, S. C. (1950): The Interactance Hypothesis: A Gravity Model Fitting Physical Masses and Human Groups. In: American Sociological Review 15 (2), 245–256. doi: 10.2307/2086789.

Dotti, N. F. / Fratesi, U. / Lenzi, C. / Percoco, M. (2013): Local labour markets and the interregional mobility of Italian university students. In: Spatial Economic Analysis 8, 443–468. doi: 10.1080/17421772.2013.833342.

Dotti, N. F. / Fratesi, U. / Lenzi, C. / Percoco, M. (2014): Local labour market conditions and the spatial mobility of science and technology university students: evidence from Italy. In: Review of Regional Research 34 (2), 119–137. doi: 10.1007/s10037-014-0088-y.

Dwenger, N. / Storck, J. / Wrohlich, K. (2012): Do tuition fees affect the mobility of university applicants? Evidence from a natural experiment. In: Economics of Education Review 31 (1), 155–167. doi: 10.1016/j.econedurev.2011.10.004.

Eckey, H. / Kossfeld, R. / Türck, M. (2007): Pendlerbereitschaft von Arbeitnehmern in Deutschland In: Raumforschung und Raumordnung — Spatial Research and Planning 65 (1), 5–14.

Einig, K. / Pütz, T. (2007): Regionale Dynamik der Pendlergesellschaft: Entwicklung von Verflechtungsmustern und Pendeldistanzen. In: Informationen zur Raumentwicklung 3 (2), 73–91.

EHEA (2016): The Bologna Declaration of 19 June 1999. http://ehea.info/me-dia.ehea.info/file/Ministerial_conferences/02/8/1999_Bologna_Declaration_English_553028.pdf (02. Mai 2019).

Epifani, P. / Gancia, G. A. (2005): Trade, migration and regional unemployment. In: Regional Science and Urban Economics 35 (6), 625–644. doi: 10.1016/j.regsciurbeco.2004.09.003.

Erhart, M. (2002): Gemeinsame Strukturen finden – Der „Masterplan“ zu Beginn der 90er Jahre. In: 10 Jahre danach – Zur Entwicklung der Hochschulen und Forschungseinrichtungen in den neuen Ländern und Berlin. Stifterverband für die Deutsche Wissenschaft, ed. Stifterverband für die Deutsche Wissenschaft e.V., 70–73.

Eurostat (2022): Employment and commuting by NUTS 2 regions (1000). lfst_r_lfe2ecomm http://data.europa.eu/88u/dataset/lqGmNpkz8gDTIprMfP34DA. (25. Februar 2022).

Erlander, S. / Stewart, N. F. (1990): The gravity model in transportation analysis: theory and extensions. Topics in transportation. VSP.

Falck, O. / Schüller, S. (2016): Querschnittstechnologie Internet - Universallösung für den Arbeitsmarkt der Zukunft? In: Wirtschaftsdienst 96 (8), 609–613. doi: 10.1007/s10273-016-2023-z.

Faggian, A. / McCann, P. / Sheppard, S. (2007): Some evidence that women are more mobile than men: Gender differences in UK graduate migration behavior. In: Journal of Regional Science 47 (3), 517–539. doi: 10.1111/j.1467-9787.2007.00518.x.

Faggian, A. / McCann, P. (2009): Human capital, graduate migration and innovation in British regions. In: Cambridge Journal of Economics 33 (2), 317–333. doi: doi: 10.1093/cje/ben042

Faggian, A. / Franklin, R. S. (2014): Human capital redistribution in the USA: the migration of the college-bound. In: Spatial Economic Analysis 9 (4), 376–395. doi: 10.1080/17421772.2014.961536.

Fagiolo, G. (2010): The international-trade network: gravity equations and topological properties. In: Journal of Economic Interaction and Coordination 5, 1–25. doi: 10.1007/s11403-010-0061-y.

FDZ (1996-2016): Forschungsdatenzentrum: Statistik der Studenten. https://www.forschungsdaten-zentrum.de/de/bildung/studenten (06. Juni 2019).

Flex, F. / Greiving, S. / Terfrüchte, T. (2016): Strukturen eines (fortentwickelten) Zentrale-Orte-Konzepts. In: Greiving, S. / Flex, F. (Hrsg.) (2016): Neuaufstellung des Zentrale-Orte-Konzepts in Nordrhein-Westfalen, Hannover: Verlag der Akademie für Raumforschung und Landesplanung, 106–122.

Florida, R. / Adler, P. / Mellander, C. (2017): The city as innovation machine. In: Regional Studies 51 (1), 86–96. doi: 10.1080/00343404.2016.1255324.

Florida, R. / Rodriguez-Pose, A. / Storper, M. (2020): Cities in a Post-COVID World. Utrecht University: Human Geography and Planning, 1–28.

Framhein, G. (1983): Alte und neue Universitäten. Einzugsbereiche und Ortswahl der Studenten, Motive und Verhalten. In: Der Bundesminister für Bildung und Wissenschaft (Hrsg.), Schriftenreihe Hochschule (44). Bock.

Frenette, M. (2004): Access to college and university: Does distance to school matter?. In: Canadian Public Policy, 427–43. doi: 10.2307/3552523.

Frenette, M. (2006): Too far to go on? Distance to school and university participation. In: Education Economics 14 (1), 31–58. doi: 10.1080/09645290500481865.

Fritsch, M. / Slavtchev, V. (2007): Universities and Innovation in Space. In: Industry & Innovation 14 (2), 201–218.

Friedmann, J. (1978): The urban field as human habitat. In: Bourne, L. S. / Simmons, J. W. (Hrsg.): Systems of Cities: Reading on Structure, Growth and Policy. New York: Oxford University Presspp, 42–52.

Gans, P. (2015): Reurbanisierungstypen in Deutschland. Wissensökonomie und Komponenten der Bevölkerungsentwicklung (2004-2010). In: Fricke, A. / Siedentop, S. / Zakrzewski, P. (Hrsg.) (2015): Reurbanisierung in baden-württembergischen Stadtregionen. Hannover: Arbeitsberichte der ARL 14, 11–31.

Gareis, P. / Diller, C. / Huchthausen, H. (2018): Braindrain und Bologna-Drain – Räumliche Implikationen der Bologna-Reform und Auswirkungen auf ländliche Räume. Eine Analyse am Beispiel der Justus-Liebig-Universität Gießen. In: Raumforschung und Raumordnung — Spatial Research and Planning 76 (4), 281–296. doi: 10.1007/s13147-018-0529-0.

Gareis, P. (2019): Räumliche Auswirkungen der Einführung gestufter Studiengänge in Deutschland. Dissertation.

Gareis, P. / Diller, C. (2020): Räumliche Aspekte der Studierendenmobilität. Stand der Forschung, eigene regionalstatistische Untersuchungen und die These vom „Bologna-Drain“ und möglichen Auswirkungen auf eine nachhaltige Hochschul- und Regionalentwicklung. In: Postlep, R. / Blume, L. / Hülz, M. (Hrsg.) (2020): Hochschulen und ihr Beitrag für eine nachhaltige Regionalentwicklung. Hannover: Forschungsberichte der ARL 11, 260–286.

Gareis, P. / Brökel, T. (2022): The spatial patterns of student mobility before, during, and after the Bologna process in Germany. In: Tijdschrift voor economische en sociale geografie — Journal of Economic and Human Geography. doi: 10.1111/tesg.12507.

Genova, V.G. / Tumminello, M. / Aiello, F. / Attanasio, M. (2021): A network analysis of student mobility patterns from high school to master's. In: Statistical Methods & Applications 30, 1445–1464. doi: 10.1007/s10260-021-00592-4.

Gibbons, S. / Vignoles, A. (2012): Geography, choice and participation in higher education in England. In: Regional Science and Urban Economics 42 (1–2), 98–113. doi: 10.1016/j.regsciurbeco.2011.07.004.

Glaeser, E.L. / Kallal, H. D. / Scheinkman, J. A. / Shleifer, A. (1992): Growth in Cities. In: Journal of Political Economy 100 (6), 1126–1152.

Gösta, G. / von Stuckrad, T. (2007): Die Zukunft vor den Toren. Aktualisierte Berechnungen zur Entwicklung der Studienanfängerzahlen bis 2020. https://www.che.de/download/che_prognose_studienanfaengerzahlen_ap100-pdf/?wpdmdl=11190&ind=5d1a0a27b4528 (13. November 2020).

Granato, N. / Haas, A. / Hamann, S. / Niebuhr, A. (2009): Arbeitskräftemobilität in Deutschland - Qualifikationsspezifische Befunde regionaler Wanderungs- und Pendlerströme. In: Raumforschung und Raumordnung — Spatial Research and Planning 67 (1), 21–33. doi: 10.1007/BF031 83140.

Greenwood, M. (1975): Research on internal migration in the United States: A survey. In: Journal of Economic Literature 13, 397433.

Greenwood, M. (1985): Human migration: Theory, models and empirical studies. In: Journal of Regional Science 25, 521–544. doi: 10.1111/j.1467-9787.1985.tb00321.x.

Griffith, D. A. (2009): Modeling spatial autocorrelation in spatial interaction data: empirical evidence from 2002 Germany journey-to-work flows. In: Journal of Geographical Systems 11, 117–140. doi: 10.1007/s10109-009-0082-z.

Growe, A. / Münter, A. (2010): Die Renaissance der großen Städte. In: Geographische Rundschau 62 (11), 54–59.

Growe, A. (2012): Emerging polycentric city-regions in Germany. Regionalisation of economic activities in metropolitan regions. In: Erdkunde 66 (4), 295–311. doi: 10.3112/erdkunde.2012.04.02.

Guirao, B. / Campa, J. / Casado-Sanz, N. (2018): Labour mobility between cities and metropolitan integration: The role of high speed rail communting in Spain. In: Cities 78,140–154. doi: 10.1016/j.cities.2018.02.008.

Haas, A. / Osland, L. (2014): Commuting, Migration, Housing and Labour Markets: Complex Interactions. In: Urban Studies 51 (3), 463–476. doi: 10.1177/0042098013498285.

Häfner, S. (2011): Auswirkungen beruflicher Mobilität auf Partnerschaften und Familienleben. Ein Literaturüberblick über den gegenwärtigen Forschungsstand unter besonderer Berücksichtigung der Commuter-Ehe. In: Zeitschrift für Psychosomatische Medizin und Psychotherapie 57 (2), 185–201.

Hansson, E. / Mattisson, K. / Björk, J. / Östergren, P. / Jakobsson, K. (2011): Relationship between commuting and health outcomes in a cross-sectional population survey in southern Sweden. In: BMC Public Health 11. doi: 10.1186/1471-2458-11-834.

Hamilton, B. W. (1982): Wasteful commuting. In: Journal of Political Economy 90, 1035–1053.

Haussen, T. / Uebelmesser, S. (2018): No place like home? Graduate migration in Germany. In: Growth and Change 49 (3), 442–472. doi: 10.1111/grow.12249.

Heuermann, D. / Schmieder, J. (2019): The effect of infrastructure on workers mobility: evidence from high-speed rail expansion in Germany. In: Journal of Economic Geography 19 (2), 335–372. doi: 10.1093/jeg/lby019.

Hillman, N. W. (2016): Geography of college opportunity: The case of education deserts. In: American Educational Research Journal 53 (4), 987–1021. doi: 10.3102/0002831216653204.

Holdsworth, C. (2009): 'Going away to uni': mobility, modernity, and independence of english higher education students. In: Environment and Planning A 41 (8), 1849-1864. doi: 10.1068/a41177.

Hong, I. / Jung, W. S. (2016): Application of gravity model on the Korean urban bus network. In: Physica A: Statistical Mechanics and its Applications 46, 48–55. doi: 10.1016/j.physa.2016.06.055.

HMWK (2015): Hochschulpakt 2016 – 2020. https://wissenschaft.hessen.de/si-tes/default/files/media/hmwk/hsp_2016-2020.pdf (12. Juli 2019).

HRK (2019): Download von Hochschulliste. https://www.hochschulkom-pass.de/hochschulen/downloads.html (28. Januar 2022).

Huber, P. / Nowotny, K. (2013): Moving across Borders: Who is Willing to Migrate or to Commute? In: Regional Studies 47 (9), 1462–1481. doi:10.1080/00343404.2011.624509.

Ignacio, P. / Darcy, K. (2019): Tracing patterns and shapes in remittance and migration networks via persistent homology. In: EPJ DataScience 8 (1). doi: 10.1140/epjds/s13688-018-0179-z.

Ingenfeld, J. / Wolbring, T. / Bless, H. (2019): Commuting and Life Satisfaction Revisited: Evidence on a Non-linear Relationship. In: Journal of Happiness Studies 20, 2677–2709. doi: 10.1007/s10902-018-0064-2.

INKAR (2022): Indikatoren und Karten zur Raum- und Stadtentwicklung. https://www.inkar.de (11. Mai 2022).

Institut national de la statistique et des études économiques (2011): Mobilités professionnelles en 2011: déplacements domicile - lieu de travail Recensement de la population - Base flux de mobilité. https://www.insee.fr/fr/statistiques/2022113#consulter (23. März 2022).

Instituto nacional de Estadística (2011): MB038_Ocupados. https://www.ine.es/en/censos2011 (19. Oktober 2019).

Istituto Nazionale di Statistica (2011): Matrici del pendolarismo 2011. https://www.istat.it/it/archivio/139381 (23. September 2019).

Jacobs, J. (1969): The Economies of Cities. Random House, New York.

Jessen, J. / Siedentop, S. (2018): Reurbanisierung. In: ARL – Akademie für Raumforschung und Landesplanung (Hrsg.): Handwörterbuch der Stadt- und Raumentwicklung, Hannover, 2073–2084.

Jung, W. S. / Wang, F. / Stanley, H. (2008): Gravity model in the Korean highway. In: Europhysics Letters Association 81 (4). doi: 10.1209/0295-5075/81/48005.

Kaluza, P. / Kölzsch, A. / Gastner, M. T. / Blasius, B. (2010): The complex network of global cargo ship movements. In: Journal of the Royal Society Interface 7, 1097–1103. doi: 10.1098/rsif.2009.0495.

Kalter, F. (2000): Theorien der Migration. In: Müller, U. / Nauck, B. / Diekmann, A. (2000): Handbuch der Demographie 1. Berlin: Springer Verlag, 438–475.

Kang, C. / Liu, Y. / Guo, D. / Qin, K. (2015): A Generalized Radiation Model for Human Mobility: Spatial Scale, Searching Direction and Trip Constraint. In: PLoS ONE 10, Article e0143500. doi: 10.1371/journal.pone.0143500.g003.

Kauder, B. / Potrafke, N. (2013): Government Ideology and Tuition Fee Policy: Evidence from the German States. In: CESifo Working Paper, No. 4205, Center for Economic Studies and ifo Institute (CESifo), Munich.

Kauffmann (2011): Metropolregion Mitteldeutschland: Wie stark sind die Pendlerverflechtungen zwischen den Städten. In: Wirtschaft im Wandel 17 (2), 70–78.

Kersting, M. / Matthies, E. / Lahner, J. / Schlüter, J. (2020): A socioeconomic analysis of commuting professionals. In: Transportation 48, 2127–2158. doi: 10.1007/s11116-020-10124-w.

KIT (2018): Geschichte - Forschungszentrum und Universität: Pioniere in Forschung und Lehre. http://www.kit.edu/kit/geschichte.php (28. Januar 2022).

KMK (2005): Prognose der Studienanfänger, Studierenden und Hochschulabsolventen bis 2020. https://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/2005/2005_10_01-Studienanfaenger-Absolventen-2020.pdf (23. Oktober 2020).

KMK (2011): Bestandsaufnahme und Perspektiven der Umsetzung des Bologna-Prozesses. https://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/2011/2011_03_10-Bestandsaufnahme-Bologna-Prozess.pdf (06. Mai 2019).

KMK (2012): Vorausberechnung der Studienanfängerzahlen 2012-2025. Fortschreibung. https://www.kmk.org/fileadmin/Dateien/pdf/Statistik/Vorausberechnung_der_Studienanfaengerzahlen_2012-2025_01.pdf (1. April 2019).

Kloosterman, R. C. / Musterd, S. (2001): The polycentric urban region: Towards a research agenda. In: Urban Studies 38 (4), 623–633. doi: 10.1080/00420980120035259.

Krabel, S. / Flöther, C. (2014): Here Today, Gone Tomorrow? Regional Labour Mobility of German University Graduates. In: Regional Studies 48 (10), 1609–1627. doi: 10.1080/00343404.2012.739282.

Kramer, J. W. (2005): Zur Prognose der Studierendenzahlen in Mecklenburg-Vorpommern bis 2020. In: Wismarer Diskussionspapiere, No. 12/2005, Hochschule Wismar, Fachbereich Wirtschaft, Wismar.

Krebs, O. / Pflüger, M. (2019): On the Road (Again): Commuting and Local Emplayment Elasticities in Germany. In: CESifo Working Paper No. 9190.

Krings, G. / Calabrese, F. / Ratti, C. / Blondel, V. D. (2009): Urban gravity: a model for inter-city telecommunication flows. In: Journal of Statistical Mechanics Theory and Experiment, 2009 (7). doi: 10.1088/1742-5468/2009/07/L07003.

Lee, E. S. (1966): A theory of migration. In: Demography 3 (1), 47–57.

Lenormand, M. / Huet, S. / Gargiulo, F. / Deuant, G. (2012): A universal model of commuting networks. In: PLoS One 7, Article, e45985. doi: 10.1371/journal.pone.0045985.

Lenormand, M. / Bassolas, A. / Ramasco, J. J. (2016): Systematic comparison of trip distributionlaws and models. In: Journal of Transport Geography 51, 158–169. doi: 10.1016/j.jtrangeo.2015.12.008.

Liang, X. / Zhao, J. / Dong, L. / Xu, K. (2013): Unraveling the origin of exponential law in intraurban human mobility. In: Scientific Reports 3, Article 2983. doi: 10.1038/srep02983.

Limtanakool, N. / Dijst, M. / Schwanen, T. (2007): A Theoretical Framework and Methodology for Characterising National Urban Systems on the Basis of Flows of People: Empirical Evidence for France and Germany. In: Urban Studis 44 (11), 2123–2145. doi: 10.1080/00420980701518990.

Linneman, P. / Graves, P.E. (1983): Migration and job change: A multinomial logit approach. In: Journal of Urban Economics 14, 263–279. doi: 10.1016/0094-1190(83)90009-8.

Lischka, I. / Kreckel, R. (2006): Zur künftigen Entwicklung der Studierendenzahlen in Sachsen-Anhalt – Prognosen und Handlungsoptionen. Expertise im Auftrag der Landesrektorenkonferenz von Sachsen-Anhalt. (Arbeitsberichte 2'06). Hrsg. von HoF Wittenberg – Institut für Hochschulforschung an der Martin-Luther-Universität Halle-Wittenberg.

Litmeyer, M.-L. / Bender, L. / Happel, S. / Spory, N. / Hennemann, S. (2022): The effects of gender equality on economic development in Europe. In: Erdkunde 76 (1), 21–40. doi: 10.3112/erdkunde.2022.01.02.

Liu, Y. / Sui, Z. / Kang, C. / Gao, Y. (2014): Uncovering patterns of inter-urban trip and spatial interaction from social media check-in data. In: PLoS One 9 Article e86026. doi: 10.1371/journal.pone.0086026.

Liu, E. / Yan, X. (2019): New parameter-free mobility model: Opportunity priority selection model. In: Physica A 526. doi: 10.1016/j.physa.2019.04.259.

Liu, Y. / Just, A. / Mayer, M. (2021): SHAP Plot for 'XG-Boost'. Package 'SHAPforxgboost'.

Lörz, M. (2008): Räumliche Mobilität beim Übergang ins Studium und im Studienverlauf: Herkunftsspezifische Unterschiede in der Wahl und Nachhaltigkeit des Studienortes. In: Bildung und Erziehung 61, 413–436.

Lowry, I.S. (1966): Migration and Metropolitan Growth: Two Analytical Models. San Francisco.

Lundberg, S. M. / Lee S-I. (2017): A unified approach to interpreting model predictions. Proceedings of the 31st International Conference on Neural Information Processing Systems. NIPS'17. USA: Curran Associates Inc, 4768–4777.

Lundberg, S. M. / Erion, G. / Lee S-I. (2018): Consistent individualized feature attribution for tree ensembles. In: arXiv preprintarXiv:1802.03888.

Lundberg, S. M. / Nair, B. / Vavilala, M. / Horibe, M. / Eisses, M. / Adams, T. / Liston, D. / King-Wai Low, D. / Newman, S-F. / Kim, J. / Lee S-I. (2018): Explainable machine-learning predictions for the prevention of hypoxaemia during surgery. In: Nature Biomedical Engineering 2, 749–760. doi: 10.1038/s41551-018-0304-0.

McCune, B. / Grace, J. (2002): Analysis of Ecological Communities. MjM Software Design. Gleneden Beach.

McQuaid, R. W. / Chen, T. (2012): Commuting times: the role of gender, children and part-time work. In: Res. Transp. Econ. 34 (1), 66–73.

Mansfield, T. J. / Rodriguez, D. / Huegy, J. / MacDonald Gibson, J. (2015): The Effects of Urban Form on Ambient Air Pollution and Public Health Risk: A Case Study in Raleigh, North Carolina. In: Risk Analysis: An International Journal 35 (5), 901–918. doi: 10.1111/risa.12317.

Marcińczaka, S. / Bartosz B. (2018): Commuting patterns and urban form: Evidence from Poland. In: Journal of Transport Geography 70, 31–39.

Marshall, A. (1890): Principles of Economics. MacMillan, London.

Marshall, J. M. / Wu, S. L. / Sanchez, C. / Kiware, S. S. / Ndhlovu, M. / Ouédraogo, A. L. / Touré, M. B. / Sturrock, H. J. / Ghani, A. C. / Ferguson, N. M. (2018): Mathematical models of human mobility of relevance to malaria transmission in Africa. In: Scientific Reports 8, Article. doi: 10.1038/s41598-018-26023-1.

Martinus, K. / Suzuki, J. / Bossaghzadeh, S. (2020): Agglomeration economics, interregional commuting and innovation in the peripheries. In: Regional Studies 54 (6), 776–788. doi: 10.1080/00343404.2019.1641592.

Masucci, A. / Serras, J. / Johansson, A. / Batty, M. (2013): Gravity versus radiation models: on the importance of scale and heterogeneity in commuting fows. In: Physical Review 88 (2). doi: 10.1103/PhysRevE.88.022812.

Meijers, E. (2005): Polycentric urban regions and the quest for synergy: Is a network of cities more than the sum of the parts? In: Urban Studies 42 (4), 765–781. doi: 10.1080=00420980500060384.

Meijers, E. / Burger, M. (2017): Stretching the concept of 'borrowed size'. In: Urban Studies 54 (1), 269–291. doi: 10.1177/0042098015597642.

Meister, M. / Stiller, J. / Peters, J. C. / Birkeneder, A. (2019): Die Binnenwanderung von Arbeitskräften in Deutschland: Eine deskriptive Analyse für ländliche Räume auf Basis der Integrierten Erwerbsbiografien (IEB) des IAB. In: Thünen Working Paper 132. doi: 10.3220/WP1574415954000.

Montgomery, M. (2002): A nested logit model of the choice of a graduate business school. In: Economics of Education Review 21 (5), 471–480. doi: 10.1016/S0272-7757(01)00032-2.

Moos, M. / Revington, N. / Wilkin, T. / Andrey, J. (2019): The knowledge economy city: Gentrification, studentification and youthification, and their connections to universities. In: Urban Studies 56 (6), 1075–1092. doi: 10.1177/0042098017745235.

Morton, A. / Piburn, J. / Nagle, N. (2018): Need A Boost: A Comparison of Traditional Commuting Models with the XGBoost Model for Predicting Commuting Flows (Short Paper). In: GIScience. doi: 10.4230/LIPIcs.GISCIENCE.2018.51.

Multrus, F. / Majer, S. / Bargel, T. / Schmidt, M. (2017): Studiensituation und studentische Orientierungen. https://www.bmbf.de/ SharedDocs/Publikationen/de/bmbf/pdf/ studiensituation-und-studentische-orientierungen-zusammen-fassung.pdf;jsessionid= E11D4ABAF32E7A355FBFBCE3D0AD4 AD4.live471? __blob=publicationFile&v=2 (28. Januar 2022).

Murat, C. H. (2010): Sample size needed for calibrating trip distribution and behavior of the gravity model. In: Journal of Transport Geography 18 (1), 183–190. doi: 10.1016/j.jtrangeo.2009.05.013.

Münter, A. / Volgmann, K. (2021): Polycentric regions: Proposals for a new typology and terminology. In: Urban Studies 58 (4), 677–695. doi: 10.1177/0042098020931695.

Niebuhr, A. / Granato, N. / Haas, A. / Hamann, S. (2012): Does Labour Mobility Reduce Disparities between Regional Labour Markets in Germany? In: Regional Studies 46 (7), 841–858. doi: 10.1080/00343404.2010.532118.

Nijkamp, P. (2008): XXQ Factors for Sustainable Urban Development: A Systems Economics View. In: Romanian Journal of Regional Science 2 (1), 1–33.

Nutz, M. (1991): Räumliche Mobilität der Studierenden und Struktur des Hochschulwesens in der Bundesrepublik Deutschland. Köln.

O'Clery, N. / Curiel, R. P. / Lora, E. (2019): Commuting times and the mobilisation of skills in emergent cities. In: Applied Network Science 118 (4). doi: 10.1007/s41109-019-0235-z.

Odlyzko, A. (2015): The forgotten discovery of gravity models and the inefficiency of early railway networks. In: OEconomia 5, 157–192. doi: 10.4000/oeconomia.1684.

Office for National Statistics (2011): Location of usual residence and place of work (OA level). https://www.nomisweb.co.uk/census/2011/wf01bew (23. September 2019).

Pan, R. K. / Kaski, K. / Fortunato, S. (2012): World citation and collaboration networks: uncovering the role of geography in science. In: Scientific Reports 2, Article 902. doi: 10.1038/srep00902.

Parr, J. B. (2014): The regional economy, spatial structure and regional urban systems. In: Regional Studies 48 (12), 1926–1938. doi: 10.1080/00343404.2013.799759.

Pfaff, S. (2012): Pendeln oder umziehen? Mobilitätsentscheidungen in Deutschland zwischen 2000 und 2009. In: Zeitschrift für Soziologie 41 (6), 458–477.

Phelps, N. A. (2019): In what sense a post-suburban era? In: Hanlon B. / Vicino T. J. (Hrgs.): The Routledge Companion to the Suburbs. London and New York: Routledge, Taylor & Francis Group, 39–47.

Pred, A. (1977): City-systems in Advanced Economies: Past Growth, Present Processes and Future Development Options. London: Hutchison.

Quigley, J. M. (1998): Urban Diversity and Economic Growth. In: Journal of Economic Perspectives 12, 127–138.

Raab, J. / Knobem, J. / Aufurth, L. / Kaashoek, B. (2018): Going the distance: the effects of university - secondary school collaboration on student migration. In: Papers in Regional Science 97 (4), 1131–1149. doi: 10.1111/pirs.12288.

Ravenstein, E. G. (1885): The laws of migration. In: Journal of the statistical society of London 48 (2), 167–235.

Ravenstein, E. G. (1889): The laws of migration: Second Paper. In: Journal of the statistical society of London 52 (2), 241–305.

Reggiania, A. / Buccib, P. / Russoc, G. / Haas, A. / Nijkampc, P. (2011): Regional labour markets and job accessibility in City Network systems in Germany. In: Journal of Transport Geography 19 (4), 528–536. doi: 10.1016/j.jtrangeo.2010.05.008.

Ren, Y. / Ercsey-Ravasz, M. / Wang, P. / González, M. C. / Toroczkai, Z. (2014): Predicting commuter flows in spatial networks using a radiation model based on temporal ranges. In: Nature Communication 5, Article 5347. doi: 10.1038/ncomms6347.

Romer, P. (1986): Increasing returns and long-run growth. In: Journal of Political Economy 94, 1002–1037.

Rouwendal, J. / Meijer, E. (2001): Preferences for housing, jobs and commuting: a mixed logit analysis. In: Journal of Regional Science 41 (3), 475–505. doi: 10.1111/0022-4146.00227.

Sá, C. / Florax, R. J. / Rietveld, P. (2004): Determinants of the regional demand for higher education in the Netherlands: a gravity model approach. In: Regional Studies 38 (4), 375–392. doi: 10.1080/03434002000213905.

Sá, C. / Dias, D. / Justino, E. / Amarale, A. (2011): Higher education (related) choices in Portugal: joint decisions on institution type and leaving home. In: Studies in Higher Education 36 (6), 687–703. doi: 10.1080/0307507100372534.

Sallah, K. / Giorgi, R. / Bengtsson, L. / Lu, X. / Wetter, E. / Adrien, P. / Rebaudet, S. / Piarroux, R. / Gaudart, J. (2017): Mathematical models for predicting human mobility in the context of infectious disease spread: introducing the impedance model. In: International Journal of Health Geographics 16, Article 42. doi: 10.1186/s12942-017-0115-7.

Sandow, E. (2014): Til Work Do Us Part: The Social Fallacy of Long-distance Commuting. In: Urban Studies 51 (3), 526–543. doi: 10.1177/0042098013498280.

Schubert, T. / Kroll, H. (2014): Universities' effects on regional GDP and unemployment: The case of Germany. In: Regional Science 95 (3), 467–489. doi: 10.1111/pirs.12150.

Schulze, S. (2009): Einige Beobachtungen zum Pendlerverhalten in Deutschland. In: HWWI Policy Paper 1 (19).

Schläpfer, M. / Bettencourt, L. M. / Grauwin, S. / Raschke, M. / Claxton, R. / Smoreda, Z. / Ratti, C. (2014): The scaling of human interactions with city size. In: Journal of the Royal Society Interface 11 (98), Article 20130789. doi: 10.1098/rsif.2013.0789.

Shapley, L. S. (1953): A value for n-person games. In: Contributions to the Theory of Games 2, 307–317.

Siegel, J. (1975): Intrametropolitan migration: A simultaneous model of employment and residential location of white and black households. In: Journal of Urban Economics 2, 2947.

Simini, F. / Gonzalez, M.C. / Maritan, A. / Barabasi, A. L. (2012): A universal model for mobility and migration patterns. In: Nature 484, 96–100. doi: 10.1038/nature10856.

Simons, H. / Weiden, L. (2016): Schwarmverhalten, Reurbanisierung und Suburbanisierung. In: Informationen zur Raumentwicklung 3, 263–273.

Sinclair, R. (1983): Changing spatial trends at the national, regional and metropolitan levels of the settlement system in developed western countries. In: Geographia Polonica 47, 101–113.

Sjaastad, L. A. (1962): The costs and returns of human migration. In: Journal of Political Economy 70 (5/2), 80–93.

Soerensen, T. (1948): A method of establishing groups of equal amplitude in plant sociology based on similarity of species and its application to analyses of the vegetation on danish commons. In: Biologiske Skriftler 5, 1–34.

Spadon, G. / de Carvalho, A. / Rodrigues-Jr, J. / Alves, L. (2019): Reconstructing commuters network using machine learning and urban indicators. In: Scientific Reports 9. doi: 10.1038/s41598-019-48295-x.

Spiess, C. K. / Wrohlich, K. (2010): Does distance determine who attends a university in Germany?. In: Economics of Education Review 29(3), 470–479. doi: 10.1016/j.econedurev.2009.10.009.

Statistics Finland (2011): Employed by municipality of residence, area of workplace and time. https://stat.fi/tup/tilastotietokannat/index_en.html (23. September 2019).

Statistics Norway (2011): Employed persons (aged 15-74), by municipality of work and municipality of residence. 4th quarter (M) 2011. https://www.ssb.no/en/statbank/table/03321 (23. März 2022).

Statistics Sweden (2011): Gainfully employed commuters by municipality 16+ years by municipality of residence, municipality of work and sex. Year 2004 – 2018. https://www.statistikdatabasen.scb.se /pxweb/en/ssd/START__AM__AM0207__AM0207L/AM0207 PendlKomA04/table/tableViewLayout1/ (30. August 2022).

Statistik Austria (2011): Registerzählung 2011.

Statistisches Bundesamt (2008-2017a): Bevölkerung nach Geschlecht und Altersgruppen (17) - Stichtag 31.12. - regionale Tiefe: Gemeinden. Code: 12411-02-03-5.

Statistisches Bundesamt (2008-2017b): Gestorbene nach Altersgruppen - Jahressumme – regionale Ebenen. Code: 12613.

Statistisches Bundesamt (2008-2017c): Beherbergungsbetriebe, Gästebetten, Gästeübernachtungen, Gästeankünfte - Jahressumme - regionale Tiefe: Gemeinden (bis 2017). Code: 45412-1-2-5.

Statistisches Bundesamt (2008-2017d): Fachserie 11 Reihe 4.1. Bildung und Kultur. Studierende an Hochschulen. Wintersemester 2007/2008 – Wintersemester 2016/2017, Wiesbaden.

Statistisches Bundesamt (2022a): Erwerbstätigkeit nach Stellung im Beruf, Entfernung, Zeitaufwand und benutztem Verkehrsmittel für den Hinweg zur Arbeitsstätte 2020 in %. https://www.destatis.de/DE/Themen/Arbeit/Arbeitsmarkt/Glossar/ erwerbstaetige.html;jsessionid=864D8BB58BB3956149861AC94 86F9971.live711 (30. Januar 2022).

Statistisches Bundesamt (2022b): Europa NUTS-Klassifikation. Die Einteilung der Europäischen Union in EU-Regionen. https://www.destatis.de/Europa/DE/Methoden-Metadaten/Klassifikationen/UebersichtKlassifikationen_NUTS.html (28. Mai 2022).

Stenpass, A. (2020): Pendlermobilität und partnerschaftliche Arbeitsteilung. Wiesbaden: Springer Fachmedien Wiesbaden.

Stouffer, S. A. (1940): Intervening opportunities: a theory relating mobility and distance. In: American Sociological Review 5, 845–867.

Teichert, C. / Niebuhr, A. / Otto, A. / Rossen, A. (2020): Work experience and graduate migration. An event history analysis of German data. In: Regional Studies 54 (10), 1413–1424. doi: 10.1080/00343404.2020.1716965.

Tharakan, J. / Tropeano, J.-P. (2009): On the impact of labor market matching on regional disparities. In: Journal of Regional Science 49 (1), 57–80. doi: 10.1111/j.1467-9787.2008.00600.x.

Thomas, T. / Tutert, S. (2013): An empirical model for trip distribution of commuters in the netherlands: transferability in time and space reconsidered. In: Journal of Transport Geography 26, 158–165. doi: 10.1016/j.jtrangeo.2012.09.005.

Tizzoni, M. / Bajardi, P. / Decuyper, A. / Kon Kam King, G. / Schneider, C. M. / Blondel, V. / Smoreda, Z. / González, M. / Colizza, V. (2014): On the use of human mobility proxies for modeling epidemics. In: PLoS Computational Biology 10, Article e1003716. doi: 10.1371/journal.pcbi.1003716.

Tuite, A. R. / Thomas-Bachli, A. / Acosta, H. / Bhatia, D. / Huber, C. / Petrasek, K. / Watts, A. / Yong, J. H. E. / Bogoch, I. I. / Khan, K. (2018): Infectious disease implications of large-scale migration of Venezuelan nationals. In: Journal of Travel Medicine 25 (1). doi: 10.1093/jtm/tay077.

Ullman, E. L. (1956): The role of transportation and the bases for interaction. In: Thomas, W. L. (Ed.) Man's Role in Changing the Face of Earth. Chicago, IL: The University of Chicago Press, 862–880.

van der Laan, L. (1998): Changing urban systems: an empirical analysis at two spatial levels. In: Regional Studies 32, 235–247. doi: 10.1080/00343409850119733.

Viboud, C. / Bjørnstad, O. N. / Smith, D. L. / Simonsen, L. / Miller, M. A. / Grenfell, B. T. (2006): Synchrony, waves, and spatial hierarchies in the spread of influenza. In: Science 312 (5772), 447–451. doi: 10.1126/science.1125237.

Volgmann, K. / Growe, A. / Münter, A. / Osterhage, F. (2022): Profitiert das Umland vom Boom der Großstädte? Reichweite und funktionale Differenzierung von Ausstrahlungseffekten in den deutschen Stadtregionen. In: Raumforschung und Raumordnung — Spatial Research and Planning 79 (6). doi: 10.14512/rur.144.

Vrontis, D. / Thrassou, A. / Melanthiou, Y. (2007): A contemporary higher education student-choice model for developed countries. In: Journal of Business Research 60 (9), 979–989. doi: 10.1016/j.jbusres.2007.01.023.

Walsh, S. / Flannery, D. / Cullinan, J. (2018): Analysing the preferences of prospective students for higher education institution attributes. In: Educ Econ 26 (2), 161–178. doi: 10.1080/09645292.2017.1335693.

Weisser, R. (2019): How Personality Shapes Study Location Choices. In: Research in Higher Education 61, 88–116. doi: 10.1007/s11162-019-09550-2.

Wilson, A. (1970): Entropy in urban and regional modelling. Pion, London.

Wiechmann, T. / Münter, A. / Danielzyk, R. (2016): Stand und Perspektiven der Erforschung polyzentraler Metropolregionen. In: Danielzyk, R./ Münter, A. / Wiechmann, T. (Hrsg.) (2016): Polyzentrale Metropolregionen. Lemgo: Rohn, 447–483.

Winkel, R. / Greiving, S. (2008): Zentrale-Orte-Konzept in praxisorientierter Weiterentwicklung. Monozentrale Orte, Städteverbünde und Zentralörtliche Funktionsräume. In: RaumPlanung 141, 263–268.

Winters, J. V. (2011): Human capital and population growth in nonmetropolitan US counties: The importance of college student migration. In: Economic Development Quarterly 25 (4), 353–365. doi: 10.1177/0891242411418221.

Wissenschaftliche Dienste des Deutschen Bundestages (2006): Der Studentenberg – Kollaps der Universitäten oder Illusion? Ein kritischer Beitrag zur aktuellen Diskussion. https://www.bundestag.de/resource/blob/418880/251afebe1c84c24d81ad39c8bbf34334/WD-8-212-06-pdf-data.pdf (02. April 2021).

Wolpert, J. (1965): Behavioral aspects of the decision to migrate. In: Papers in Regional Science 15 (1), 159–169.

Yan, X-Y. / Zhao, C. / Fan, Y. / Di, Z. / Wang, W-X. (2014): Universal predictability of mobility patterns in cities. In: Journal of Royal Society. Interface 11 (100). doi: 10.1098/rsif.2014.0834.

Yang, Y. / Herrera, C. / Eagle, N. / Gonzalez, M. C. (2014): Limits of predictability in commuting ows in the absence of data for calibration. In: Scientific Reports 4, Article 5662. doi: 10.1038/srep05662.

Zax, J. S. (1994): When is a move a migration? In: Regional Science and Urban Economics. 24 (3), 341–360. doi: 10.1016/0166-0462(93)02034-z.

Zipf, G. K. (1946): The P1 P2/D hypothesis: on the intercity movement of persons. In: American Sociological Review 11, 677–686.

Zipf, G. K. (1949): Human Behaviour and the Principles of Least Effort. Addison Wesley: Cambridge, MA.

Eidesstattliche Erklärung

Ich erkläre: Ich habe die vorgelegte Dissertation selbstständig und ohne unerlaubte fremde Hilfe und nur mit den Hilfen angefertigt, die ich in der Dissertation angegeben habe. Alle Textstellen, die wörtlich oder sinngemäß aus veröffentlichten Schriften entnommen sind, und alle Angaben, die auf mündlichen Auskünften beruhen, sind als solche kenntlich gemacht. Ich stimme einer evtl. Überprüfung meiner Dissertation durch eine Antiplagiat-Software zu. Bei den von mir durchgeführten und in der Dissertation erwähnten Untersuchungen habe ich die Grundsätze guter wissenschaftlicher Praxis, wie sie in der „Satzung der Justus-Liebig-Universität Gießen zur Sicherung guter wissenschaftlicher Praxis“ niedergelegt sind, eingehalten.

Datum: 22. November 2022 .....

(Unterschrift)