

Prädiktive Analyse: Vergleich von Methoden des maschinellen Lernens und statistischen Methoden für die Entwicklung eines interaktiven Tools zur Vorhersage der Bevölkerungsentwicklung in Neubaugebieten.

Bachelorarbeit

für die Prüfung zum
Bachelor of Science

des Studiengangs Wirtschaftsinformatik-Application Management
an der
Dualen Hochschule Baden-Württemberg Lörrach

Aikaterini Nakou

18. September 2019

Kurs	WWI16A-AM
Ausbildungsfirma	ITEOS, Freiburg
Betreuer der Ausbildungsfirma	Stefan Fuchs
Wissenschaftlicher Betreuer	Prof. Dr. Jan Michael Olaf

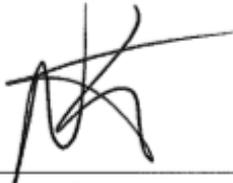
Ehrenwörtliche Erklärung

Ich versichere hiermit, dass ich meine Bachelorarbeit mit dem Thema:

Prädiktive Analyse: Vergleich von Methoden des maschinellen Lernens und statistischen Methoden für die Entwicklung eines interaktiven Tools zur Vorhersage der Bevölkerungsentwicklung in Neubaugebieten.

selbstständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe. Ich versichere zudem, dass die eingereichte elektronische Fassung mit der gedruckten Fassung übereinstimmt.

Freiburg, 18. September 2019



Aikaterini Nakou

Hinweis zum Umfang der Arbeit

Der Textteil der vorliegenden Arbeit - beginnend mit der Einleitung bis ausschließlich Quellenverzeichnis - umfasst 80 Seiten.

Freigabe der Arbeit

Die vorliegende Arbeit wurde durch das Ausbildungsunternehmen ITEOS, Freiburg inhaltlich geprüft und zur Vorlage an der DHBW Lörrach, Studiengang Wirtschaftsinformatik-Application Management, freigegeben.



Freiburg, 18. September 2019



Unterschrift Unternehmensvertreter

Kurzfassung

Für politische, gesellschaftliche und wirtschaftliche Entscheidungsprozesse, sowie die Planung von Infrastruktureinrichtungen und Versorgungsnetzen in Neubaugebieten einer Stadt stellen Vorhersagen der Bevölkerungsentwicklung und -struktur in die Zukunft gerichtete Basisinformationen dar. In dieser Arbeit wird die zu erwartende Entwicklung der Bevölkerung hinsichtlich der Einwohnerzahl und der demographischen Struktur (Altersgruppen und Geschlecht) in Neubaugebieten des Landes Baden-Württemberg prognostiziert. Auf Basis der Vorhersageergebnisse wird ein Bevölkerungsprognose-Tool als Prototyp entwickelt, um die kommunalen Entscheidungsträger bei der Infrastrukturplanung zu unterstützen.

Die Bevölkerungsvorausberechnungen werden mit Hilfe von Methoden der prädiktiven Analyse durchgeführt. Dabei wird eine Zeitreihenanalyse, sowohl durch Methoden des maschinellen Lernens (Neuronale Netze), sowie durch Methoden der klassischen Statistik (ARIMA-Modelle, Komponenten-Modelle) durchgeführt. Außerdem werden auch Kohorten-Komponenten-Modelle, als klassische Methode zur Bevölkerungsprojektionen berechnet.

Das Bevölkerungsprognose-Tool wird mit Hilfe des RShiny-Frameworks als interaktive Web-Applikation für die Erstellung von personalisierten Prognosen entwickelt. RShiny ermöglicht die schnelle und interaktive Entwicklung von statistischen Anwendungen und entsprechenden Visualisierungen, die online bereitgestellt werden können. Durch die Bereitstellung als Webanwendung ermöglicht diese Plattform Entscheidungsträgern in Echtzeit mit den Daten zu interagieren und nach Bedarf benutzerdefinierte Prognosen für einzelne Gemeinden zu erstellen.

Inhaltsverzeichnis

Ehrenwörtliche Erklärung	II
Hinweis zum Umfang der Arbeit	II
Freigabe der Arbeit	III
Kurzfassung	IIV
Inhaltsverzeichnis	V
Abkürzungsverzeichnis	VII
Abbildungsverzeichnis	VIII
Tabellen	XIV
1 Einleitung	1
1.1 Problemstellung	3
1.2 Ziel der Arbeit	4
1.3 Aufbau der Arbeit	5
2 Grundlagen	7
2.1 Bevölkerungsentwicklung und ihre Modellierung	7
2.2 Kohorten-Komponenten-Modelle	10
2.3 Zeitreihenanalyse – Klassische Statistik	12
2.3.1 Komponenten Modelle	15
2.3.2 ARIMA Modelle	17
2.4 Zeitreihenanalyse – Künstliche Neuronale Netze	18
3 Ist- und Problemanalyse	22
3.1 Istanalyse	22
3.2 Problemanalyse	24
3.3 Zusammenfassung	25
4 Lösungskonzept	27
4.1 Datengewinnung	27
4.2 Prädiktive Analyse der Bevölkerungsentwicklung	29

4.2.1	Trendextrapolationsmodelle	29
4.2.2	Kohorten-Komponenten-Modelle	37
4.2.3	Modellvergleich	37
4.3	Design der Webapplikation	38
5	Umsetzung	40
5.1	Implementierung der prädiktiven Analyse	40
5.2	Implementierung der Webapplikation	73
6	Fazit	78
6.1	Zusammenfassung	78
6.2	Ausblick	79
	Quellenverzeichnis	XI
	Anhang	XIV

Abkürzungsverzeichnis

ACF	Autocorrelation Function
ADF	Augmented-Dickey-Fuller-Test
AIC	Akaike Information Criterion
ARIMA	AutoRegressive Integrated Moving Average
GAM	Generalisierte Additive Modelle
KNN	Künstliche Neuronale Netze
MAPE	Mean Absolute Percentage Error
MLP	Multilayer Perzeptron

Abbildungsverzeichnis

Abbildung 1: Methoden der Bevölkerungsprognosen	8
Abbildung 2: Übersicht der Kohorten-Komponenten-Methode.....	11
Abbildung 3: Das MLP-Modell	19
Abbildung 4: Das Berechnungsprinzip von SIKURS	23
Abbildung 5: Gesamte Einwohnerzahl (Rieselfeld)	30
Abbildung 6: Gesamte Einwohnerzahl (Vauban)	31
Abbildung 7: Männliche Einwohnerzahl (Rieselfeld)	31
Abbildung 8: Männliche Einwohnerzahl (Vauban).....	32
Abbildung 9: Weibliche Einwohnerzahl (Rieselfeld)	32
Abbildung 10: Weibliche Einwohnerzahl (Vauban).....	33
Abbildung 11: Einwohnerzahl nach Altersgruppen (Rieselfeld).....	33
Abbildung 12: Einwohnerzahl nach Altersgruppen (Vauban)	34
Abbildung 13: Autokorrelationsfunktion der Gesamtbevölkerungsdaten (Rieselfeld) .	36
Abbildung 14: Autokorrelationsfunktion der Gesamtbevölkerungsdaten (Vauban).....	36
Abbildung 15: Rohdaten vs Vorhergesagte Werte für die Gesamtbevölkerung (Rieselfeld).....	43
Abbildung 16: Rohdaten vs Vorhergesagte Werte für die Gesamtbevölkerung (Vauban)	43
Abbildung 17: 30-jährige Prognose der Gesamtbevölkerung (Rieselfeld).....	44
Abbildung 18: 20-jährige Prognose der Gesamtbevölkerung (Vauban)	44
Abbildung 19: Rohdaten vs Vorhergesagte Werte für die weibliche Bevölkerung (Rieselfeld).....	45
Abbildung 20: Rohdaten vs Vorhergesagte Werte für die männliche Bevölkerung (Rieselfeld).....	45
Abbildung 21: Rohdaten vs Vorhergesagte Werte für die weibliche Bevölkerung (Vauban).....	46
Abbildung 22: Rohdaten vs Vorhergesagte Werte für die männliche Bevölkerung (Vauban)	46

Abbildung 23: 30-jährige Prognose der Bevölkerung getrennt nach Geschlecht (Rieselfeld).....	47
Abbildung 24: 20-jährige Prognose der Bevölkerung getrennt nach Geschlecht (Vauban)	47
Abbildung 25: Rohdaten vs Vorhergesagte Werte für die Bevölkerung der Altersgruppe 0 bis unter 3 Jahre (Rieselfeld)	48
Abbildung 26: Rohdaten vs Vorhergesagte Werte für die Bevölkerung der Altersgruppe 40 bis unter 70 Jahre (Rieselfeld)	49
Abbildung 27: Rohdaten vs Vorhergesagte Werte für die Bevölkerung der Altersgruppe 0 bis unter 3 Jahre (Vauban).....	49
Abbildung 28: Rohdaten vs Vorhergesagte Werte für die Bevölkerung der Altersgruppe 40 bis unter 70 Jahre (Vauban)	49
Abbildung 29: 30-jährige Prognose der Bevölkerung getrennt nach Altersgruppe (Rieselfeld).....	50
Abbildung 30: 20-jährige Prognose der Bevölkerung getrennt nach Altersgruppe (Vauban)	50
Abbildung 31: 30-jährige Prognose der männlichen Bevölkerung getrennt nach Altersgruppe (Rieselfeld)	52
Abbildung 32: 30-jährige Prognose der weiblichen Bevölkerung getrennt nach Altersgruppe (Rieselfeld).....	52
Abbildung 33: 20-jährige Prognose der männlichen Bevölkerung getrennt nach Altersgruppe (Vauban)	53
Abbildung 34: 20-jährige Prognose der weiblichen Bevölkerung getrennt nach Altersgruppe (Vauban)	53
Abbildung 35: Rohdaten vs Vorhergesagte Werte für die Gesamtbevölkerung (Rieselfeld).....	54
Abbildung 36: Rohdaten vs Vorhergesagte Werte für die Gesamtbevölkerung (Vauban)	55
Abbildung 37: 30-jährige Prognose der Gesamtbevölkerung (Rieselfeld)	55
Abbildung 38: 20-jährige Prognose der Gesamtbevölkerung (Vauban)	56

Abbildung 39: Rohdaten vs Vorhergesagte Werte für die männliche Bevölkerung (Rieselfeld).....	57
Abbildung 40: Rohdaten vs Vorhergesagte Werte für die weibliche Bevölkerung (Rieselfeld).....	57
Abbildung 41: Rohdaten vs Vorhergesagte Werte für die männliche Bevölkerung (Vauban).....	57
Abbildung 42: Rohdaten vs Vorhergesagte Werte für die weibliche Bevölkerung (Vauban).....	58
Abbildung 43: 30-jährige Prognose der Bevölkerung getrennt nach Geschlecht (Rieselfeld).....	58
Abbildung 44: 20-jährige Prognose der Bevölkerung getrennt nach Geschlecht (Vauban).....	59
Abbildung 45: 30-jährige Prognose der Bevölkerung getrennt nach Altersgruppe (Rieselfeld).....	60
Abbildung 46: 20-jährige Prognose der Bevölkerung getrennt nach Altersgruppe (Vauban).....	60
Abbildung 47: 30-jährige Prognose der männlichen Bevölkerung getrennt nach Altersgruppe (Rieselfeld).....	61
Abbildung 48: 30-jährige Prognose der weiblichen Bevölkerung getrennt nach Altersgruppe (Rieselfeld).....	62
Abbildung 49: 20-jährige Prognose der männlichen Bevölkerung getrennt nach Altersgruppe (Vauban).....	62
Abbildung 50: 20-jährige Prognose der weiblichen Bevölkerung getrennt nach Altersgruppe (Vauban).....	62
Abbildung 51: Rohdaten vs Vorhergesagte Werte für die Gesamtbevölkerung (Rieselfeld).....	63
Abbildung 52: Rohdaten vs Vorhergesagte Werte für die Gesamtbevölkerung (Vauban).....	64
Abbildung 53: 30-jährige Prognose der Gesamtbevölkerung (Rieselfeld).....	64
Abbildung 54: 20-jährige Prognose der Gesamtbevölkerung (Vauban).....	65

Abbildung 55: Rohdaten vs Vorhergesagte Werte für die weibliche Bevölkerung (Rieselfeld).....	66
Abbildung 56: Rohdaten vs Vorhergesagte Werte für die männliche Bevölkerung (Rieselfeld).....	66
Abbildung 57: Rohdaten vs Vorhergesagte Werte für die weibliche Bevölkerung (Vauban).....	66
Abbildung 58: Rohdaten vs Vorhergesagte Werte für die männliche Bevölkerung (Vauban).....	67
Abbildung 59: 30-jährige Prognose der Bevölkerung getrennt nach Geschlecht (Rieselfeld).....	67
Abbildung 60: 20-jährige Prognose der Bevölkerung getrennt nach Geschlecht (Vauban).....	68
Abbildung 61: 30-jährige Prognose der Bevölkerung getrennt nach Altersgruppe (Rieselfeld).....	69
Abbildung 62: 20-jährige Prognose der Bevölkerung getrennt nach Altersgruppe (Vauban).....	70
Abbildung 63: 30-jährige Prognose der männlichen Bevölkerung getrennt nach Altersgruppe (Rieselfeld).....	70
Abbildung 64: 30-jährige Prognose der weiblichen Bevölkerung getrennt nach Altersgruppe (Rieselfeld).....	70
Abbildung 65: 20-jährige Prognose der männlichen Bevölkerung getrennt nach Altersgruppe (Vauban).....	71
Abbildung 66: 20-jährige Prognose der weiblichen Bevölkerung getrennt nach Altersgruppe (Vauban).....	71
Abbildung 67: Vorhersage der Gesamtbevölkerung und getrennt nach Geschlecht in Fünfjahresperioden (Rieselfeld).....	73
Abbildung 68: Vorhersage der Gesamtbevölkerung und getrennt nach Geschlecht in Fünfjahresperioden (Vauban).....	73
Abbildung 69: Startseite der Webapplikation.....	75
Abbildung 70: Auswahl der gewünschten Neubaugebietsart.....	76
Abbildung 71: Darstellung der Bevölkerungsprognosen.....	76

Abbildung 72:Darstellung des Modellvergleichs.....	77
Abbildung 73: Möglichkeit eigene Daten zu laden und analysieren.....	77
Anhang-Abbildung 1: Autokorrelationsfunktion der männlichen Bevölkerungsdaten (Rieselfeld).....	XIV
Anhang-Abbildung 2: Autokorrelationsfunktion der weiblichen Bevölkerungsdaten (Rieselfeld).....	XIV
Anhang-Abbildung 3: Autokorrelationsfunktion der Bevölkerungsdaten der Altersgruppe 0-3 Jahre (Rieselfeld)	XV
Anhang-Abbildung 4: Autokorrelationsfunktion der Bevölkerungsdaten der Altersgruppe 3-6 Jahre (Rieselfeld)	XV
Anhang-Abbildung 5: Autokorrelationsfunktion der Bevölkerungsdaten der Altersgruppe 6-10 Jahre (Rieselfeld)	XVI
Anhang-Abbildung 6: Autokorrelationsfunktion der Bevölkerungsdaten der Altersgruppe 10-18 Jahre (Rieselfeld)	XVI
Anhang-Abbildung 7: Autokorrelationsfunktion der Bevölkerungsdaten der Altersgruppe 18-25 Jahre (Rieselfeld)	XVII
Anhang-Abbildung 8: Autokorrelationsfunktion der Bevölkerungsdaten der Altersgruppe 25-40 Jahre (Rieselfeld)	XVII
Anhang-Abbildung 9: Autokorrelationsfunktion der Bevölkerungsdaten der Altersgruppe 40-70 Jahre (Rieselfeld)	XVIII
Anhang-Abbildung 10: Autokorrelationsfunktion der Bevölkerungsdaten der Altersgruppe 70 Jahre oder älter (Rieselfeld)	XVIII
Anhang-Abbildung 11: Autokorrelationsfunktion der weiblichen Bevölkerungsdaten (Vauban)	XIX
Anhang-Abbildung 12: Autokorrelationsfunktion der männlichen Bevölkerungsdaten (Vauban)	XIX
Anhang-Abbildung 13: Autokorrelationsfunktion der Bevölkerungsdaten der Altersgruppe 0-3 Jahre (Vauban).....	XX
Anhang-Abbildung 14: Autokorrelationsfunktion der Bevölkerungsdaten der Altersgruppe 3-6 Jahre (Vauban).....	XX

Anhang-Abbildung 15: Autokorrelationsfunktion der Bevölkerungsdaten der Altersgruppe 6-10 Jahre (Vauban).....	XXI
Anhang-Abbildung 16: Autokorrelationsfunktion der Bevölkerungsdaten der Altersgruppe 10-18 Jahre (Vauban).....	XXI
Anhang-Abbildung 17: Autokorrelationsfunktion der Bevölkerungsdaten der Altersgruppe 18-25 Jahre (Vauban).....	XXII
Anhang-Abbildung 18: Autokorrelationsfunktion der Bevölkerungsdaten der Altersgruppe 25-40 Jahre (Vauban).....	XXII
Anhang-Abbildung 19: Autokorrelationsfunktion der Bevölkerungsdaten der Altersgruppe 40-70 Jahre (Vauban).....	XXIII
Anhang-Abbildung 20: Autokorrelationsfunktion der Bevölkerungsdaten der Altersgruppe 70 Jahre oder älter (Vauban).....	XXIII

Tabellen

Tabelle 1: Anpassungsgüte der GAM-Modelle nach Geschlecht	45
Tabelle 2: Anpassungsgüte der GAM-Modelle nach Altersgruppe	48
Tabelle 3: Anpassungsgüte der GAM-Modelle nach Altersgruppe und Geschlecht	51
Tabelle 4: Anpassungsgüte der KNN-Modelle nach Geschlecht	56
Tabelle 5: Anpassungsgüte der KNN-Modelle nach Altersgruppe	59
Tabelle 6: Anpassungsgüte der KNN-Modelle nach Altersgruppe und Geschlecht	61
Tabelle 7: Anpassungsgüte der ARIMA-Modelle nach Geschlecht	65
Tabelle 8: Anpassungsgüte der ARIMA-Modelle nach Altersgruppe	68
Tabelle 9: Anpassungsgüte der ARIMA-Modelle nach Altersgruppe und Geschlecht ...	69

1 Einleitung

Seit der Jahrtausendwende erleben viele Städte Nordamerikas und Europas eine bemerkenswerte Phase des Bevölkerungswachstums. Dieses neue demografische Wachstum der Städte wird in der Literatur als *Reurbanisierung* oder *Renaissance der Stadt* bezeichnet^{1, 2, 3, 4}. Von diesem Phänomen sind auch größere südwestdeutsche Städte betroffen. Trendanalysen des Statistischen Landesamts Baden-Württemberg zeigen, dass in Städten wie Stuttgart, Mannheim, Freiburg, Heidelberg und Ludwigshafen tatsächlich Reurbanisierungstendenzen nachgewiesen werden können^{5, 6}.

Diese Reurbanisierung der Metropolregionen hat mehrere Ursachen. Eine davon ist der soziale Wandel und seine vielfältigen Auswirkungen auf die Haushaltsbildung, die Arbeits- und Beschäftigungsverhältnisse sowie Geschlechternormen. Eine weitere ist der ökonomische Wandel und seine Folgen wie zum Beispiel die Entwicklung einer wachsenden wissensbasierten Wirtschaft. Auch die Bildungsexpansion wirkt sich positiv auf das städtische Wachstum aus. Ebenfalls ziehen internationale Migranten viel häufiger in Metropolen. Diese Konzentration ergibt sich aus der Größe des städtischen Arbeitsmarktes sowie differenzierten Beschäftigungsmöglichkeiten unabhängig vom Qualifikationsbedarf. Abschließend stärken die öffentliche Politik und die kommunalen Strategien zur Stadtentwicklung auch die Attraktivität städtischer Wohngebiete im Vergleich zum Leben in den umliegenden Vororten oder ländlichen Gebieten^{7, 8, 9}.

Wesentliche Folge dieses inter- und intraregionalen Wanderungsgeschehens ist der Wohnraumangel in solchen Städten. Um die Wohnraumknappheit zu bekämpfen,

¹ Vgl. Siedentop, 2008, S. 194-195

² Vgl. Frank, S. 2011, S. 64

³ Vgl. Sturn, G. 2012, S. 21

⁴ Vgl. Haase, A. et al. 2009, S. 443

⁵ Vgl. Brachat-Schwarz, W. 2008, S. 11

⁶ Vgl. Fricke, A. 2015, S. 172

⁷ Vgl. Geppert, K. und Gornig, M. 2010, S. 5-6

⁸ Vgl. Siedentop et al. 2017, S. 2-3

⁹ Vgl. Bundesinstitut für Bau- Stadt- und Raumforschung-Berichte KOMPAKT 2011, S. 1-4

entwickeln die Städte bereits geeignete Planungsstrategien und Instrumentarien, wie z.B. Programmplanung für die Entwicklung neuer Stadtquartiere^{10, 11, 12}.

Für die Planung solcher neuen Siedlungen, vor allem für den Neubau von Kindergärten, Schulen, Krankenhäusern, Straßen oder der Kapazitätserweiterung von Kanalisation und Kraftwerken sind detaillierte Informationen über die zukünftige Entwicklung wichtiger Variablen erforderlich. Eine dieser wichtigen Variablen, die in die lokale Planung einfließt, ist die Bevölkerungszahl. Dementsprechend werden Prognosen erstellt, auf deren Grundlage die Planung erfolgt^{13, 14, 15, 16}.

Für die Entwicklung von Bevölkerungsprognosen können viele Ansätze angewendet werden. Einige davon sind subjektiv und andere objektiv. Einige sind sehr einfach und andere sehr komplex. Einige sind streng extrapolierend und andere nutzen Strukturmodelle. Auch bei den Datenanforderungen unterscheiden sich die unterschiedlichen Ansätze stark. Zusammenfassend können die verschiedenen Ansätze zur Bevölkerungsentwicklung in drei verschiedene Gesamtkategorien klassifiziert werden¹⁷:

- Trendextrapolation
- Kohorten-Komponenten-Methoden
- Strukturmodelle

Die angewandten Extrapolationsmethoden können auf einer einfachen linearen oder exponentiellen Kurvenanpassung oder fortgeschrittenen Methoden basieren. In der Regel werden keine Faktoren analysiert, die zu Bevölkerungsveränderungen führen, z.B. Geburten, Todesfälle und Migration. Die Kohorten-Komponenten-Methoden dagegen berücksichtigen diese drei Komponenten des Bevölkerungswachstums und werden am häufigsten bei Bevölkerungsprognosen eingesetzt. Zuletzt betrachten die

¹⁰ Vgl. Breyer, F. et al. 2018, S. 5-6

¹¹ Vgl. Breuer, B. und Schmell, R. 2007, 4

¹² Vgl. Fricke, A. 2015, S. 178, 180, 181

¹³ Vgl. Smith et al. 2001, S. 7-11

¹⁴ Vgl. Hansen, H.S. 2010, S. 32-33

¹⁵ Vgl. Diamond, I. et al. 1990, S. 151

¹⁶ Vgl. Stycos, M. 1995, S. 205

¹⁷ Vgl. Smith et al. 2001, S. 5

Strukturmodelle die Bevölkerungsentwicklung als ein Zusammenspiel von demographischen und nicht-demographischen Faktoren¹⁸.

Jedes Jahr erstellen die nationalen statistischen Ämter Bevölkerungsprognosen für Nationen, Regionen und Gemeinden, aber für detaillierte Planungszwecke wie die Infrastrukturplanung von Neubaugebieten sind Prognosen auf niedrigerer räumlicher Ebene unbedingt erforderlich. Deshalb müssen viele Gemeinden Bevölkerungsprognosen auf Stadtteilebene erstellen^{19, 20}.

ITEOS als kommunaler IT-Dienstleister mit umfassenden Expertisen im Bereich der Digitalisierung und Automatisierung von Verwaltungsprozessen berät und begleitet Kommunen und Städte mit innovativen Produkten bei der Verwirklichung der digitalen Transformation. Im Verlauf dieser Bachelorarbeit wird die Möglichkeit untersucht, ein Bevölkerungsprognose-Tool für die Unterstützung der kommunalen Entscheidungsträger bei der Infrastrukturplanung für neue Baugebiete bereitzustellen. Das Tool wird auf Basis von Prognosen der Bevölkerungsentwicklung in ehemaligen Neubaugebieten entwickelt.

1.1 Problemstellung

Für politische, gesellschaftliche und wirtschaftliche Entscheidungsprozesse, sowie die Planung von Infrastruktureinrichtungen und Versorgungsnetzen in Neubaugebieten einer Stadt ist die Kenntnis über die Zahl und die Zusammensetzung der künftigen Bevölkerung eine unentbehrliche Voraussetzung. Um diese Voraussetzung zu erfüllen, werden Prognosen der Bevölkerungsentwicklung und -Struktur anhand von Bevölkerungsstrukturdaten ehemaliger Neubaugebiete benötigt.

Bevölkerungsprognosen werden oft von den statistischen Landesämtern durchgeführt. Solche Prognosen basieren allerdings auf der amtlichen Einwohnerzahl, die stark von

¹⁸ Vgl. Smith et al. 2001, S. 5-6

¹⁹ Vgl. Harting et al. 2011, S. 203-205

²⁰ Vgl. Hansen, H.S. 2010, S. 33

der jeweiligen Bevölkerungszahl im Melderegister abweicht. Beifolgend wird mit diesen Prognosen nur die Bevölkerungsanzahl nach Alter und Geschlecht für die gesamte Stadt in der Zukunft projiziert. Für die Zwecke der Neubaugebietsplanung sind aber kleinräumige Bevölkerungsprojektionen nach Stadtteil notwendig.

Eine ausschlaggebende Rolle bei Bevölkerungsprognosen spielen die historischen Datenreihen, auf denen sie basieren. Oft kann der Zugang zu kleinräumigen Daten ein ernsthaftes Hindernis darstellen. Außerdem werden je nach eingesetzter Prognostizierungsmethoden unterschiedliche Anforderungen an den Detaillierungsgrad der Daten vorausgesetzt. Besonders die Kohorten-Komponenten-Modelle, die am häufigsten von Städten und statistischen Landesämtern verwendet werden, haben sehr hohe Datenanforderungen.

Über die Jahre wurden verschiedene methodische Ansätze für Bevölkerungsprognosen, die sich in ihren Anforderungen sehr stark unterscheiden, entwickelt. Daher stellt die Auswahl der geeigneten Methode eine große Herausforderung an Zeit, Daten und Expertise dar. Somit sind für die Städte und Kommunen solche kleinräumigen Vorhersagen der Bevölkerungsentwicklung sehr zeit- und kostenintensiv.

1.2 Ziel der Arbeit

Um die Erstellung von Prognosen der Bevölkerungsentwicklung und -struktur in Neubaugebieten des Landes Baden-Württemberg zu vereinfachen und somit ihre Zeit- und Kostenintensivität zu verringern, wird ein Prototyp für ein Bevölkerungsprognose-Tool als Unterstützung der kommunalen Entscheidungsträger bei der Infrastrukturplanung von Neubaugebieten entwickelt.

Das Bevölkerungsprognose-Tool soll die zu erwartende Entwicklung der Bevölkerung nach planungsgerechten Altersgruppen und Geschlecht des beabsichtigten neuen Baugebietes hinsichtlich der Einwohnerzahl und der demographischen Struktur, wie auch die entsprechenden Visualisierungen online bereitstellen. Außerdem soll das Tool die Bevölkerungsentwicklung anhand von unterschiedlichen Prognostizierungsmetho-

den darstellen und ein Vergleich der Ergebnisse der Methoden anhand der Anpassungsgüte der Modelle anbieten. Zudem soll das Tool interaktiv mit dem Benutzer agieren. So soll es den Benutzer die Möglichkeit geben Angaben für das gewünschte Neubaugebiet, wie die Größe des Gebiets und das Erstbezugsjahr zu definieren und somit nach Bedarf benutzerdefinierte Prognosen zu erstellen.

Im Rahmen dieser Bachelorarbeit soll ein Prototyp für die Implementierung des Bevölkerungsprognose-Tools entwickelt und als Web-Applikation angeboten werden.

Für das Frontend ist als grobes Ziel ein Dashboard geplant, das den Benutzer zum Interagieren auffordert und die Ergebnisse seiner Anforderungen mittels interaktiver Visualisierungen darstellt.

Für das Backend werden Modelle mit verschiedenen Methoden der prädiktiven Analyse angepasst, die aus den Eingaben der Benutzer für einen bestimmten Zeithorizont Bevölkerungsprognosen erstellen.

1.3 Aufbau der Arbeit

Gemäß der Ziele und des Entwicklungsprozesses des Bevölkerungsprognose-Tools ist die Arbeit wie folgt gegliedert.

Im ersten Kapitel wird eine Einführung in den Themenbereich der Arbeit, wie auch ein Überblick der angegangenen Problempunkte und die daraus resultierten Ziele der Arbeit verfasst. Die Beschreibung der für die Arbeit relevanten theoretischen Bereiche erfolgt im zweiten Kapitel. In Kapitel drei wird der derzeitige Ablauf bei der Erstellung von Bevölkerungsprognosen beschrieben und die Ursache der identifizierten Probleme durchleuchtet.

Daraus wird im vierten Kapitel ein Lösungskonzept für die Entwicklung des Tools präsentiert. Insbesondere wird auf die Gewinnung der für die Populationsmodellierung notwendigen Daten, die prädiktive Analyse der Populationsentwicklung und das Design

der Webapplikation eingegangen. Im fünften Kapitel wird die Implementierung der entworfenen Lösung beschrieben.

Darüber hinaus werden im sechsten Kapitel die gewonnenen Erkenntnisse und der Stand der Implementierung zusammengefasst und abschließend ein Ausblick auf mögliche Erweiterungen und Entwicklungen gegeben.

2 Grundlagen

2.1 Bevölkerungsentwicklung und ihre Modellierung

Der Begriff der Bevölkerungsentwicklung kommt aus dem Bereich der Demographie, eine Wissenschaft, die Bevölkerungsstruktur und Bevölkerungswandel untersucht. Mit zunehmender Komplexität der modernen Gesellschaft wird es immer wichtiger, alle Aspekte der Veränderung von Größe und Zusammensetzung der Bevölkerung genau zu messen und somit abschätzen zu können, wie die zukünftige Größe und Zusammensetzung der Bevölkerung aussehen könnte²¹.

Die Analyse des demografischen Wandels basiert auf der Verfügbarkeit genauer Daten über relevante Populationsmerkmale und -prozesse. Da die Zukunft eng mit der Vergangenheit verbunden ist, werden Prognosen auf der Grundlage vergangener Trends und Entwicklungen erstellt. Beziehungen erhöhen das Verständnis für die Dynamik des Bevölkerungswachstums und liefern oft Prognosen für zukünftige Bevölkerungsveränderungen, die ausreichend genau sind, um eine gute Entscheidungsfindung zu unterstützen^{22, 23, 24}.

Wie schon in Kapitel 1 erwähnt wurde, können Bevölkerungsprognosen entweder mit subjektiven oder objektiven Methoden erstellt werden. Subjektive Methoden sind solche bei denen Daten, Techniken und Annahmen nicht eindeutig identifiziert werden können. Daher können andere Analysten sie nicht genau replizieren. Objektive Methoden sind solche für die Daten, Techniken und Annahmen eindeutig identifizierbar sind, so dass andere Analysten sie genau reproduzieren können (siehe Abb. 1)^{25, 26}.

²¹ Vgl. Hinde, A. 1998, S. 1

²² Vgl. George, M. et al. 2004, S. 561

²³ Vgl. Hinde, A. 1998, S. 2

²⁴ Vgl. Smith, S. et al. 2001, S. 3, 7

²⁵ Vgl. Smith, S. et al. 2001, S. 5

²⁶ Vgl. George, M. et al. 2004, S. 564-565

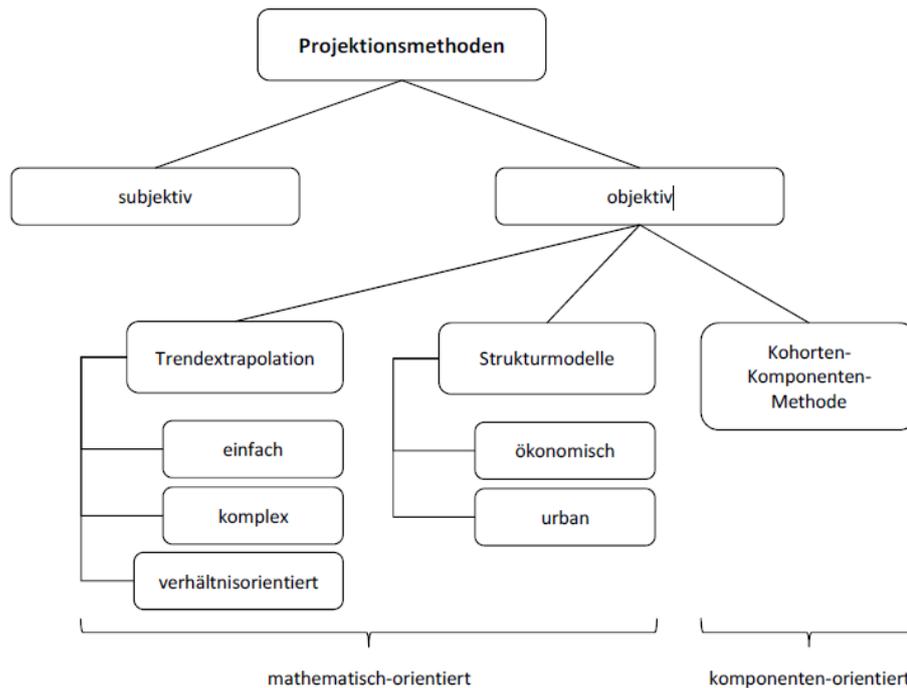


Abbildung 1: Methoden der Bevölkerungsprognosen

Quelle: Smith et al. (2001)

Trendextrapolationsmethoden beinhalten die Anpassung mathematischer Modelle an historische Daten und die Verwendung dieser Modelle zur Prognostizierung zukünftiger Bevölkerungswerte. Diese Methoden sind verbunden mit relativ niedrigen Kosten und geringen Datenanforderungen und somit in vielen Bereichen beliebt²⁷.

Trendextrapolationsmethoden können in drei Kategorien aufgeteilt werden^{28, 29}:

- Einfache Extrapolationsmethoden. Bei diesen Methoden wird einfach die Veränderung der Population in einem bestimmten Zeitraum berechnet und diese Veränderung wird über den Prognostizierungshorizont projiziert.
- Komplexe Extrapolationsmethoden. Bei diesen Methoden werden statistische Modelle der Zeitreihenanalyse, wie Komponenten-Modelle und ARIMA-Modelle angepasst.

²⁷ Vgl. George, M. et al. 2004, S. 572

²⁸ Vgl. George, M. et al. 2004, S. 573

²⁹ Vgl. Smith, S. et al. 2001, S. 5

- Verhältnis-Extrapolationsmethoden. Bei diesen Methoden werden die Bevölkerungsprognosen eines Gebietes anhand der Prognosen der Bevölkerung des „übergeordneten“ Gebietes abgeleitet.

Dementsprechend sind komplexe Extrapolationsmethoden für die Prognostizierung von Bevölkerungsentwicklungen besser geeignet, da sie auch nichtlinearen Bevölkerungsveränderungen nachbilden können und gleichzeitig eine quantitative Grundlage für die Konstruktion von Messungen der Prognoseunsicherheit bieten, da statistische Algorithmen zur Schätzung von Modellparametern verwendet werden³⁰.

Für die Erstellung einer Bevölkerungsprognose mit Anwendung komplexer Extrapolationsmethoden sind folgende Schritte erforderlich: Als Erstes erfolgt die Zusammenstellung der historischen Bevölkerungsdaten für verschiedene Zeitpunkte (Jahre) während des Basiszeitraums. Damit ein Modell gültig ist, müssen die Daten für jeden Zeitpunkt auf konsistent definierten geografischen Grenzen basieren. Danach wird ein geeigneter Modellierungsansatz (z.B. Komponenten-Modelle oder ARIMA-Modelle) ausgewählt und die Parameter des Modells werden geschätzt, um die Prognose zu erzeugen. Typischerweise werden Diagramme und statistische Messungen verwendet, um festzustellen, wie gut ein bestimmtes Modell zu den Daten des Basiszeitraumes passt. Abschließend werden die Bevölkerungsprognosen für den gegebenen Prognostizierungszeitraum anhand des angepassten Modells erstellt³¹.

Die Kohorten-Komponenten-Methode verwirklicht die Erstellung von Bevölkerungsprognosen, in dem sie die Bevölkerung in Alter-Geschlecht Kohorten aufteilt und die Fertilitäts-, Mortalitäts- und Migrationsrate jeder Kohorte berücksichtigt. Eine Vielzahl von Techniken kann verwendet werden, um jede der drei Komponenten des Bevölkerungswachstums zu projizieren^{32, 33}.

³⁰ Vgl. George, M. et al. 2004, S. 576

³¹ Vgl. George, M. et al. 2004, S. 576-578

³² Vgl. Smith, S. et al. 2001, S. 5

³³ Vgl. George, M. et al. 2004, S. 565

Eine ausführliche Beschreibung, der für diese Bachelorarbeit geeigneten Methoden, nämlich die Komplexe Extrapolationsmethode und die Kohorten-Komponenten Methode, folgt in den Unterkapiteln 2.2 bis 2.4.

2.2 Kohorten-Komponenten-Modelle

Die Kohorten-Komponenten-Methode ist eine flexible und leistungsfähige Methode zur Erstellung von Bevölkerungsprognosen. Da sie die Bevölkerung des Basisjahres getrennt nach Geschlecht-Altersgruppen-Kohorten bezieht und separat für jede Kohorte beim Durchlaufen des Projektionszeitraums das Fertilitäts-, Mortalitäts- und Migrationsrate berücksichtigt, liefert sie detailliertes Wissen über die Populationsdynamik. Außerdem kann sie eine Vielzahl von Annahmen berücksichtigen und auf jeder geographischen Ebene eingesetzt werden^{34, 35, 36, 37}.

Für die Erstellung der Bevölkerungsprognose mit Anwendung der Kohorten-Komponenten-Modelle sind folgende Schritte erforderlich^{38, 39, 40, 41}:

Als Erstes erfolgt die Aufteilung der Bevölkerung des Basisjahres in ein- oder fünfjährige Altersgruppen. Je nach dem was für Altersgruppen definiert werden, werden entsprechend auch die Projektionsintervalle bestimmt (z.B. Aufteilung nach fünfjährigen Altersgruppen verlangt auch fünfjährige Projektionsintervalle). Die älteste Altersgruppe ist praktisch immer "unbefristet", normalerweise 75+, 85+ oder 90+ Jahre. Diese Altersgruppen werden zunächst nach Geschlecht gegliedert.

Dann wird für jede von diesen Geschlecht-Altersgruppe-Kohorten die Bevölkerungszahl des Basisjahres bestimmt und die Anzahl der Personen berechnet, die bis zum Ende des Projektionsintervalls überleben werden. Dies geschieht durch die Anwendung altersspezifischer Überlebensraten auf jede Kohorte in der Population des Basisjahres.

³⁴ Vgl. George, M. et al. 2004, S. 591

³⁵ Vgl. Smith, S. et al. 2001, S. 43

³⁶ Vgl. Steinberg, J. und Doblhammer-Reiter, G. 2010, S. 393-394

³⁷ Vgl. Burch, T. 2018, S. 135-137

³⁸ Vgl. George, M. et al. 2004, S. 592

³⁹ Vgl. Smith, S. et al. 2001, S. 44-47

⁴⁰ Vgl. Steinberg, J. und Doblhammer-Reiter, G. 2010, S. 393-394

⁴¹ Vgl. Bretz, M. 2000, S. 647

Im Anschluss daran erfolgt die Berechnung der Migration mittels Migrationsraten während des Projektionsintervalls für jede Geschlecht-Altersgruppe-Kohorte. Die Migrationsraten stellen eine Projektion der Anzahl der Personen in jeder Kohorte dar, die während des Projektionsintervalls in ein Gebiet zuziehen oder wegziehen.

Anschließend wird die Anzahl der Geburten berechnet, die während des Projektionsintervalls auftreten. Dies wird durch die Anwendung altersspezifischer Geburtenraten auf die weibliche Bevölkerung in jeder Altersgruppe erreicht.

Schlussendlich erfolgt das Addieren der berechneten Anzahl der Geburten getrennt nach Geschlecht zu dem Rest der Bevölkerung. Diese Berechnungen liefern eine Projektion der Population nach Alter und Geschlecht am Ende des Projektionsintervalls. Diese Population dient dann als Ausgangspunkt für das folgende Intervall. Der Prozess wird wiederholt, bis das endgültige Prognostizierungszieljahr erreicht ist (Siehe Abb.2).

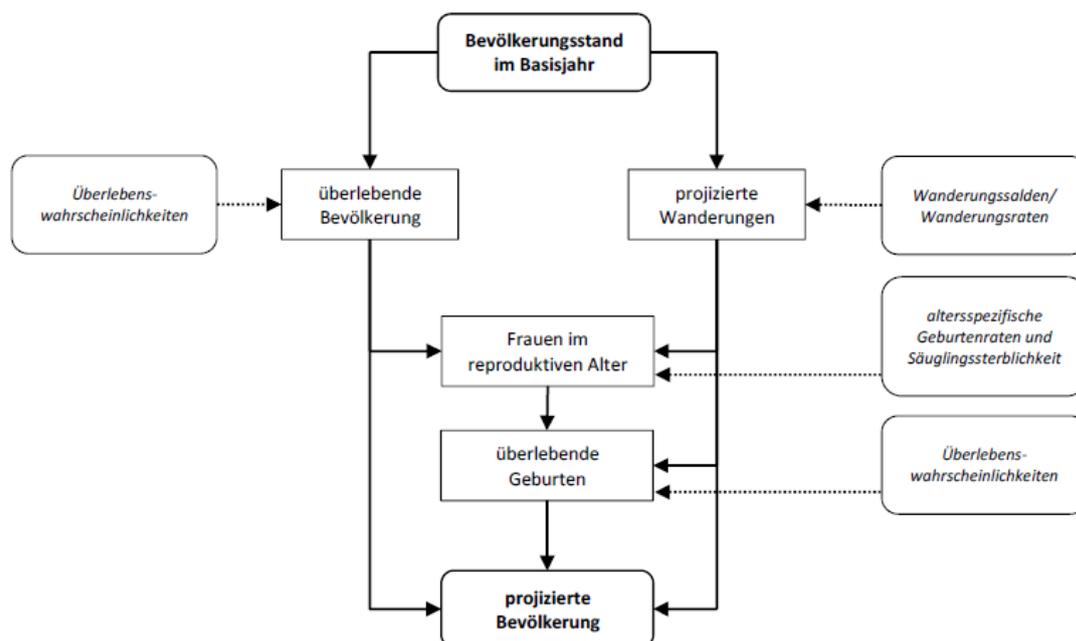


Abbildung 2: Übersicht der Kohorten-Komponenten-Methode

Quelle: Smith et al. (2001)

Abschließend lässt sich resultieren, dass Kohorten-Komponenten-Modelle^{42, 43, 44}:

- mathematisch einfache Modelle sind, die eine gewisse Komplexität, aber geringe inhärente Schwierigkeiten beinhalten.
- detaillierte Daten benötigen, um die Kohorten und die Komponenten Fertilität, Mortalität und Migration berechnen zu können.
- rechenintensiv sind, insbesondere wenn die Projektion über einen längeren Zeitraum erfolgt und kleine (z.B. einjährige) Altersintervalle verwendet werden. Dies liegt auch dran, dass es keine Abkürzungen gibt, durch die man nicht von Jahr t_0 auf Jahr t_n überspringen kann, ohne die Ergebnisse der Berechnung für alle dazwischen liegenden Jahre zu berechnen.
- detailliertes Wissen über die Populationsentwicklung liefern, da sie die Ausgangspopulationsstruktur und die Komponenten Fertilität, Mortalität und Migration berücksichtigen.
- prognostizierungstechnisch begrenzt sind, da die Komponenten Fertilität, Mortalität und Migration typischerweise durch die Annahme berechnet werden, dass sie eine Kontinuität mit der gegenwärtigen oder jüngeren Vergangenheit aufweisen werden. Für kurze Zeiträume ist dies oft eine vernünftige Annahme. Für längere Zeiträume kann diese Annahme jedoch nicht gültig sein.
- die Bevölkerungsentwicklung als linearen Trend berechnen.

2.3 Zeitreihenanalyse – Klassische Statistik

Eine Zeitreihe ist eine sequentielle Sammlung von Datenpunkten, die typischerweise über aufeinanderfolgende Zeiten gemessen werden. Sie ist mathematisch definiert als

⁴² Vgl. Burch, T.K. 2018, S. 137-140

⁴³ Vgl. Steinberg, J. und Doblhammer-Reiter, G. 2010, S. 394-397

⁴⁴ Vgl. George, M. et al. 2004, S. 602-604

eine Sammlung von Vektoren $y(t), t = 0, 1, 2, 3, \dots$ wobei t die Zeit darstellt. Die Variable $y(t)$ wird als Zufallsvariable behandelt^{45, 46, 47, 48, 49}.

Zeitreihen können univariat oder multivariat sein. Univariate Zeitreihen sind Zeitreihen, die Datensätze einer einzelnen Variablen enthalten, während multivariate Zeitreihen Datensätze von mehr als einer Variablen berücksichtigen. Außerdem kann eine Zeitreihe kontinuierlich oder diskret sein. In einer kontinuierlichen Zeitreihe werden Beobachtungen zu jedem Zeitpunkt gemessen, während eine diskrete Zeitreihe Beobachtungen enthält, die zu diskreten Zeitpunkten gemessen wurden. Beispielsweise können Temperaturmessungen, (Stoff-)Konzentration bei chemischen Prozessen usw. als kontinuierliche Zeitreihe aufgezeichnet werden. Andererseits stellt die Bevölkerung einer bestimmten Stadt oder die Produktion eines Unternehmens diskrete Zeitreihen dar. Darüber hinaus kann eine kontinuierliche Zeitreihe leicht in eine diskrete umgewandelt werden, indem Daten über ein bestimmtes Zeitintervall zusammengeführt werden^{50, 51, 52}.

Die Zeitreihenanalyse beschäftigt sich mit statistischen Methoden zur Analyse und Modellierung einer geordneten Folge von Beobachtungen (Zeitreihe). Diese Modellierung führt zu einem Prozessmodell für das System, das die Daten erzeugt hat. Somit können anhand dieses Modells zukünftige Ereignisse vorhergesagt werden. Die Zeitreihenprognose hat wichtige Anwendungen in verschiedenen Bereichen und auf Basis der prognostizierten Ergebnisse werden oft wertvolle strategische Entscheidungen und Vorsorgemaßnahmen getroffen. Daher ist es wichtig, eine gute Prognose zu erstellen, d.h. ein geeignetes Modell an eine Zeitreihe anzupassen^{53, 54, 55, 56}.

⁴⁵ Vgl. Backhaus, K. 2016, S. 136

⁴⁶ Vgl. Kirchgässner, G. et al. 2013, S. 1

⁴⁷ Vgl. Holland, H. 2010, S. 79

⁴⁸ Vgl. Bourier, G. 2012, S. 155

⁴⁹ Vgl. Shumway, R. und Stoffer, D. 2011, S. 13

⁵⁰ Vgl. Backhaus, K. 2016, S. 136

⁵¹ Vgl. Shumway, R. und Stoffer, D. 2011, S. 13

⁵² Vgl. Kirchgässner, G. et al. 2013, S. 1

⁵³ Vgl. Madsen, H. 2007, S. 1

⁵⁴ Vgl. Backhaus, K. 2016, S. 138

⁵⁵ Vgl. Bourier, G. 2012, S. 155

Die Prognoseverfahren der Zeitreihenanalyse lassen sich grob in zwei Gruppen einteilen, die Komponenten-Modelle und die ARIMA-Modelle. Bei den Komponenten-Modellen geht die Zeitreihenanalyse davon aus, dass die systematischen Komponenten einer Zeitreihe nicht von stochastischen Störungen beeinflusst werden und somit durch deterministische Funktionen der Zeit repräsentiert werden können. Im Gegensatz dazu gehen ARIMA-Modelle von stochastischen Prozessen aus. Diese stochastischen Prozesse sind mathematische Ausdrücke, die die Wahrscheinlichkeitsstruktur einer Zeitreihe beschreiben^{57, 58}.

Die allgemeine Vorgehensweise bei der Analyse und Prognose einer Zeitreihe lässt sich in fünf Schritte gliedern^{59, 60}:

- **Sammeln von Informationen.** Es sind immer mindestens zwei Arten von Informationen erforderlich: Zum einen die statistischen Daten und zum anderen das gesammelte Fachwissen der Personen, die die Daten erheben und die Prognosen verwenden. Oftmals ist es schwierig, genügend historische Daten zu erhalten, um ein gutes statistisches Modell zu erstellen.
- **Explorative Analyse.** Am Beginn der Zeitreihenanalyse sollte immer eine explorative Analyse mittels einer grafischen Darstellung der Zeitreihendaten stattfinden. Sie ist unerlässlich für die Formulierung eines geeigneten Modells, da sie visualisiert, ob einheitliche Muster oder ein signifikanter Trend aus den Daten erkennbar sind, ob es Saisonabhängigkeiten gibt oder ob Ausreißer in den Daten existieren, die von Experten erklärt werden müssen.
- **Auswahl und Anpassung der Modelle.** Das beste Modell hängt von der Verfügbarkeit historischer Daten, der Stärke der Beziehungen zwischen der Prognosevariablen und den erklärenden Variablen, sowie der Art und Weise ab, wie die Prognosen verwendet werden sollen. Es ist üblich, zwei oder drei potenzielle Modelle zu vergleichen. Die

⁵⁶ Vgl. Kirchgässner, G. et al. 2013, S. 2-3

⁵⁷ Vgl. Backhaus, K. 2016, S. 138

⁵⁸ Vgl. Kirchgässner, G. et al. 2013, S. 4

⁵⁹ Vgl. Backhaus, K. 2016, S. 140-159

⁶⁰ Vgl. Shumway, R. und Stoffer, D. 2011, S. 11

Modellformulierung enthält Parameter, die anhand der bekannten historischen Daten geschätzt werden müssen.

- **Einsatz des Prognosemodells.** Sobald ein Modell ausgewählt und seine Parameter geschätzt wurden, wird das Modell zur Erstellung von Prognosen verwendet.
- **Prüfung des Prognosemodells.** Die Prognosegüte des Modells kann durch die Gegenüberstellung der Prognosen mit der Realität bewertet werden (Modellvalidierung).

2.3.1 Komponenten Modelle

Die Komponenten-Modelle sind ein traditioneller Ansatz der Zeitreihenanalyse und ihre Anpassung erfolgt mit Hilfe deterministischer mathematischer Modelle, die die systematische Komponente der Zeitreihe umfassen. Diese Komponenten sind: Trend-, periodische oder saisonale Schwankungen und Restkomponenten. Die allgemeine Tendenz einer Zeitreihe, über einen längeren Zeitraum zu steigen, zu sinken oder zu stagnieren, wird als Trend bezeichnet. Saisonale Schwankungen in einer Zeitreihe sind regelmäßige, wiederkehrende Schwankungen um den Trend und sie sind am meisten durch natürliche Faktoren, wie z.B. Klima- und Wetterbedingungen oder unnatürliche Faktoren verursacht. In der Restkomponente werden alle einmaligen, nicht regelmäßigen Einflüsse zusammengefasst^{61, 62, 63}.

Hinsichtlich der Auswirkungen dieser drei Komponenten werden im Allgemeinen zwei verschiedene Arten von Komponenten-Modellen für eine Zeitreihe verwendet, nämlich Multiplikative und Additive Modelle^{64, 65}.

Additives Modell: $y(t) = T(t) + S(t) + R(t)$

Multiplikatives Modell: $y(t) = T(t) * S(t) * R(t)$

⁶¹ Vgl. Kirchgässner, G. et al. 2013, S. 3

⁶² Vgl. Bourier, G. 2012, S. 157

⁶³ Vgl. Holland, H. 2010, S. 79-80

⁶⁴ Vgl. Backhaus, K. 2016, S. 142

⁶⁵ Vgl. Bourier, G. 2012, S. 158-159

Dabei ist $y(t)$ ein Zeitreihenwert, $T(t)$ die Trend-Komponente, $S(t)$ die saisonale Schwankung und $R(t)$ die Restkomponente.

Um die Zeitreihe durch Komponenten-Modelle vorherzusagen, werden die saisonale Komponente und die Trendkomponente separat modelliert und angepasst. Es wird in der Regel davon ausgegangen, dass sich die saisonale Komponente unveränderlich oder extrem langsam ändert, so dass sie durch einfaches Auswerten des letzten Jahres der geschätzten Komponente prognostiziert wird. Mit anderen Worten für die saisonale Komponente wird eine saisonale Naive-Methode verwendet⁶⁶.

Da die Trend-Komponente linear oder nicht-linear sein kann, werden für ihre Modellierung Methoden der linearen Regression oder Glättungsmethoden eingesetzt^{67, 68, 69, 70}:

Lineare Regression

Das lineare Regressionsmodell erstellt die lineare Beziehung zwischen der Antwortvariablen y und der Prädiktorvariablen x mittels der Regressionsfunktion:

$$y = \alpha + \beta x + \varepsilon \text{ dar.}$$

Dabei sind α und β die Regressionskoeffizienten, ε der Fehler-Term und x die Prädiktorvariable, nämlich die Zeit als Zeitindex (1, 2, 3,...) oder als Jahreszahlen (2010, 2011, 2012,...).

Um die Regressionskoeffizienten zu schätzen wird die Methode der kleinsten-Quadrate angewendet und zur Beurteilung der Güte der geschätzten Regressionsfunktion das Bestimmtheitsmaß (R-Quadrat).

Generalisierte Additive Modelle (Glättungsmethode)

GAM ist eine Methode, die 1990 von Hastie und Tibshirani vorgeschlagen wurde. Es ist ein nichtlineares Modell, das durch die Summierung der Glättungsfunktionen der Kovarianz erklärt wird. Dieses Modell ist ein additives nichtlineares Modell, das der

⁶⁶ Vgl. Bourier, G. 2012, S. 180-185

⁶⁷ Vgl. Backhaus, K. 2016, S. 138, 142

⁶⁸ Vgl. Shumway, R. und Stoffer, D. 2011, S. 48, 72

⁶⁹ Vgl. Bourier, G. 2012, S. 169, 172

⁷⁰ Vgl. Wood, S. 2006, S. 121-139

Antwortvariablen y eine Verteilung der exponentiellen Familie, wie Normalverteilung, Binomialverteilung, Gammaverteilung, Poisson-Verteilung, inverse Gauß'sche Verteilungen, etc. folgt. Die folgende Gleichung stellt das GAM-Modell dar:

$$g(\mu_i) = X_i^* \theta + f_1(x_{1i}) + f_2(x_{2i}) + f_3(x_{3i}) + \dots$$

wo $\mu_i \equiv E(Y_i)$ und $Y_i \sim$ eine Verteilung der exponentiellen Familie

Dabei ist Y die Antwortvariable, X_i eine Reihe der Modellmatrix für alle streng parametrischen Komponenten, ϑ der entsprechende Parametervektor und f die Glättungsfunktion der Prädiktorvariable x , nämlich der Zeit als Zeitindex (1, 2, 3,...) oder als Jahreszahlen (2010, 2011, 2012, ...).

Das GAM-Modell ermöglicht eine flexible Spezifikation der Abhängigkeit der Antwortvariablen von der Prädiktorvariablen, indem es das Modell in Form der Glättungsfunktion spezifiziert.

Für die Anpassung des GAM-Modells wird die Methode der kleinsten-Quadrate oder die Maximum-Likelihood-Methode angewendet und zur Beurteilung der Güte der geschätzten Regressionsfunktion das Bestimmtheitsmaß (R-Quadrat).

2.3.2 ARIMA Modelle

Eines der wichtigsten und am weitesten verbreiteten Zeitreihenmodelle ist das *AutoRegressive Integrated Moving Average* (ARIMA)-Modell. Die Beliebtheit des ARIMA-Modells beruht auf seinen statistischen Eigenschaften sowie der bekannten Box-Jenkins-Methodik im Bildungsprozess des Modells. Somit wird im Gegensatz zum klassischen Ansatz nun davon ausgegangen, dass alle Komponenten einer Zeitreihe durch einfache stochastische Modelle dargestellt werden können. Da Zeitreihen außerdem nicht-stationär (wenn Mittelwert und Varianz der Zeitreihe über die Zeit nicht konstant sind) sein können, bieten ARIMA-Modelle im Vergleich mit anderen autoregressiven Methoden den Vorteil, dass Differenzierungsmethoden angewendet werden können, um eine nicht stationäre Zeitreihe in eine stationäre umzuwandeln^{71, 72, 73}.

⁷¹ Vgl. Zhang, P. 2003, S. 160-161

⁷² Vgl. Shumway, R. und Stoffer, D. 2011, S. 83

Ein ARIMA-Modell stellt mittels einer linearen Funktion die aktuellen Werte einer Zeitreihe in Bezug zu seinen eigenen Vergangenheitswerten (die autoregressive Komponente -AR-) und die Vergangenheitswerte des Fehler-Terms (Komponente des gleitenden Mittelwerts -MA-) dar. Die integrierte Komponente (I) bezieht sich auf die Anzahl der Wiederholungen, die eine Serie differenziert werden muss, um Stationarität zu induzieren. Die Struktur des Modells kann wie folgend dargestellt werden:^{74, 75}

$$y_t = \theta_0 + \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \dots + \varphi_p y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q}$$

Dabei ist y_t das aktuelle Wert der Zeitreihe, ε_t der Fehler-Term in Zeitperiode t , φ_i und θ_i Modellparameter, p und q ganze Zahlen, die als Ordnungen des Modells bezeichnet werden.

Für die Anpassung des ARIMA-Modells wird die Maximum-Likelihood-Methode angewendet.

2.4 Zeitreihenanalyse – Künstliche Neuronale Netze

Künstliche neuronale Netze (KNN) sind eine Methode des maschinellen Lernens und nutzen eine Reihe verschiedener mathematischer Modelle, die von entsprechenden biologischen Modellen, die das Verhalten von Neuronen im menschlichen Gehirn imitieren, inspiriert sind^{76, 77, 78}.

In den letzten Jahren wurden umfangreiche Forschungsarbeiten zur Anwendung neuronaler Netze für die Zeitreihenmodellierung und -prognose durchgeführt⁷⁹. Ein wesentlicher Vorteil der KNN-Modelle gegenüber anderen Klassen nichtlinearer Modelle ist, dass KNNs universelle Approximatoren sind, die eine große Klasse von Funktionen

⁷³ Vgl. Kirchgässner, G. et al. 2013, S. 4

⁷⁴ Vgl. Shumway, R. und Stoffer, D. 2011, S. 83-108

⁷⁵ Vgl. Zhang, P. 2003, S. 162

⁷⁶ Vgl. Mitchell 1997, S. 81,82

⁷⁷ Vgl. Backhaus et al. 2016, S. 604

⁷⁸ Vgl. Stecking 1999, S. 33

⁷⁹ Vgl. Zhang, P. et al. 1998, S. 35-62

mit hoher Genauigkeit approximieren können. Ihre Stärke liegt in der parallelen Verarbeitung der in den Daten erhaltenen Informationen. Im Modellbildungsprozess ist keine vorherige Annahme der Modellform erforderlich. Stattdessen wird das Netzwerkmodell weitgehend durch die Eigenschaften der Daten bestimmt und adaptiv gebildet. Aus diesem Grund sind KNNs von Natur aus datengesteuert und selbstadaptiv⁸⁰.

Die am weitesten verbreitete Modellform für die Modellierung und Prognose von Zeitreihen mittels Neuronaler Netze ist das mehrschichtige Feedforward Neuronale Netz⁸¹.

Mehrschichtige Feedforward Neuronale Netze (MLP)

Mehrschichtige Feedforward neuronale Netze bestehen aus einer Eingabeschicht, einer oder mehreren verdeckten Schichten und einer Ausgabeschicht⁸².

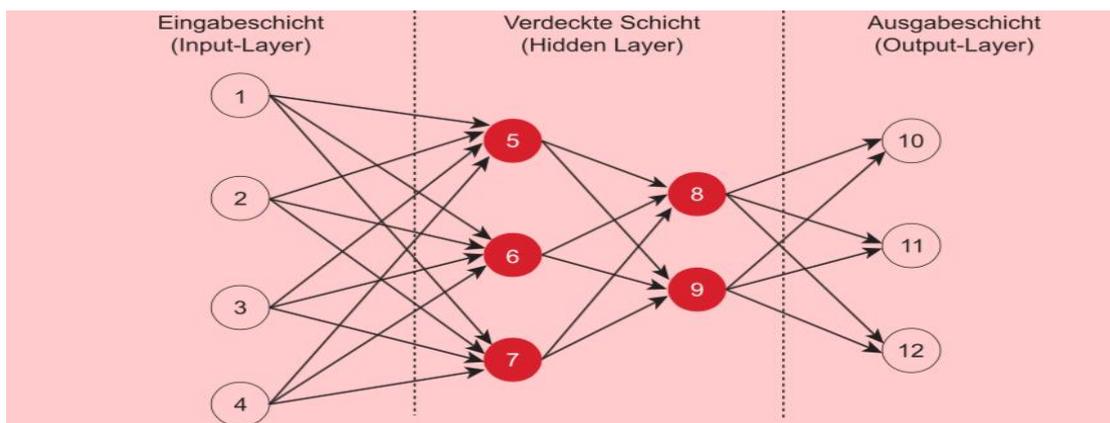


Abbildung 3: Das MLP-Modell

Quelle: Backhaus et al. (2015)

Der Informationsfluss beginnt bei dem MLP in der Eingabeschicht und endet bei der Ausgabeschicht. Die Anzahl der Variablen, die für die Lösung des Klassifizierungsproblems eine Rolle spielen, bestimmt die Anzahl der Neuronen in der Eingabeschicht. Die Neuronen der Eingabeschicht nehmen Informationen auf und leiten diese an alle Neuronen der ersten Zwischenschicht (oder verdeckten Schicht) weiter. Die Anzahl der

⁸⁰ Vgl. Zhang, P. 2003, S. 163

⁸¹ Vgl. Zhang, P. et al. 1998, S. 35-62

⁸² Vgl. Backhaus et al. 2015, S. 300

Zwischenschichten wie auch die Anzahl ihre Neuronen ist theoretisch unbegrenzt. Hinter der letzten Zwischenschicht befindet sich die Ausgabeschicht. Die Anzahl der Neuronen der Ausgabeschicht entspricht den Klassen der Zielvariablen. Diese sind mit allen Neuronen der letzten Zwischenschicht verbunden. In der Ausgabeschicht wird das Ergebnis der Informationsverarbeitung erzeugt. Zusammengefasst ist jedes Neuron einer Schicht mit jedem Neuron der nächsten Schicht verbunden und es bestehen keine Verbindungen zwischen den Neuronen der gleichen Schicht. Daher werden Informationen ständig von einer Schicht zur Nächsten weitergeleitet⁸³.

Die Informationsverarbeitung von einem Neuron einer Schicht zu einem Neuron der nächsten Schicht erfolgt durch Anwendung von Aktivierungsfunktionen. Dies können lineare, logistische, Tangenshyperbolicus- oder einfache Schwellenwert-Funktionen sein⁸⁴. Das Training des MLPs, um die Gewichtung für die Erreichung der bestmöglichen Klassifikation entsprechend anzupassen, geschieht mit der Anwendung des Backpropagation-Algorithmus. Bei diesem Algorithmus erfolgt die Anpassung der Gewichte rückwertig, nämlich von der Ausgabeschicht hin zur Eingabeschicht^{85, 86}. Die Basis für die Entwicklung des Backpropagation-Algorithmus bildet die Delta-Lernregel mit der Anwendung des Gradientenverfahrens⁸⁷.

Der Lernprozess eines MLP-Netzes, das aus n_{in} Neuronen in der Eingabeschicht, n_{hidden} Neuronen in der Zwischenschicht und n_{out} Neuronen in der Ausgabeschicht besteht, beginnt durch die Anwendung des Backpropagation-Algorithmus mit der Initialisierung aller Gewichte des Netzwerks mit niedrigen Zufallszahlen. Danach werden die Informationen vorwärts durch das Netzwerk weitergeleitet und die Ausgabewerte für jede Einheit des Netzwerks berechnet. Folglich werden die Fehler durch die Fehlerfunktion $E = \sum_{d \in D} (t_d - o_d)^2$ berechnet und rückwirkend über das Netzwerk weitergeleitet. Der Backpropagation-Algorithmus versucht die durchschnittlichen Gesamtfehler durch

⁸³ Vgl. Backhaus et al. 2015, S. 298-302

⁸⁴ Vgl. Backhaus et al. 2015, S. 317

⁸⁵ Vgl. Backhaus et al. 2015, S. 317

⁸⁶ Vgl. Mitchell 1997, S. 97

⁸⁷ Vgl. Mitchell 1997, S. 89

die Anpassung der einzelnen Gewichte zu minimieren und wird beendet, wenn die Abbruchkriterien erfüllt sind^{88, 89}.

Für die Optimierung eines KNNs, um das jeweilige Klassifikationsproblem bestmöglich zu lösen, sind folgende Parameter von großer Bedeutung: Anzahl der Zwischenschichten, Anzahl der Neuronen jeder Zwischenschicht, Höhe der Lernrate, Anzahl der Epochen und die Größe des Batches⁹⁰. Besonders bei der Zeitreihenmodellierung hat sich gezeigt, dass MLPs mit nur eine Zwischenschicht am besten geeignet sind, da mit einer geringen Anzahl von Zwischenschichten Overfitting-Effekte eliminiert werden können⁹¹.

⁸⁸ Vgl. Backhaus et al. 2015, S. 316-323

⁸⁹ Vgl. Mitchell 1997, S. 98

⁹⁰ Vgl. Backhaus et al. 2015, S. 309-341

⁹¹ Vgl. Zhang, P. 2003, S. 164

3 Ist- und Problemanalyse

3.1 Istanalyse

Kleinräumige Bevölkerungsprognosen nach Stadtteilen bilden die wichtigste Grundlage für eine Vielzahl von Planungsaktivitäten und unterstützen umfassend kommunale Träger, indem sie die Prognose von ehemaligen Neubaugebieten berücksichtigen, um entsprechende Entscheidungen für die Entwicklung von neuen Baugebieten zu treffen^{92, 93, 94, 95}.

Im Land Baden-Württemberg werden Bevölkerungsvorausrechnungen vom statistischen Landesamt erstellt. Da diese Prognosen nur Daten für die Stadt insgesamt nach Alter und Geschlecht ausweisen, müssen die einzelnen Städte selber kleinräumige Bevölkerungsprognosen berechnen. Die meisten Städte wie auch das statistische Landesamt verwenden für die Vorausberechnung der Bevölkerungszahl und -struktur das Prognosemodell SIKURS (Statistisches Informationssystem zur kleinräumigen Umliegung und Projektion einer regionalen Bevölkerungsstruktur)⁹⁶.

SIKURS ist ein kostenpflichtiges EDV-gestütztes Prognosemodell des KOSIS-Verbundes. Dieses Prognosemodell erstellt die Fortschreibung des aktuellen Bevölkerungsbestandes, indem es zu erwartenden Wanderungsbewegungen und natürliche Bevölkerungsbewegungen mit den aktuellen Bestandszahlen verrechnet. Für die Vorausberechnung benötigt SIKURS Angaben zu Volumen und Struktur des derzeitigen Bevölkerungsbestandes, zum generativen Verhalten, zum Wanderungsgeschehen und zur Verteilung der Sterbefälle der letzten Jahre. Diese Informationen werden in Dateien mit den Geburtenraten, Sterberaten, Bevölkerungsgruppenwechselraten, Binnenwegzugsraten (basierend auf innerstädtischen Umzügen), Außenwegzugsraten, Zuzugsquoten

⁹² Vgl. Wilson, T. 2015, S. 242

⁹³ Vgl. Dehler, K. 1976, S. 21

⁹⁴ Vgl. Hansen, H.S. 2010, S. 32-33

⁹⁵ Vgl. Stycos, M. 1995, S. 205

⁹⁶ Vgl. www.staedtestatistik.de

und Zuzugsvolumen sowie mit dem Bevölkerungsbestand zu Beginn des Prognosezeitraumes in absoluten Zahlen bereitgestellt (siehe Abb. 4)⁹⁷.

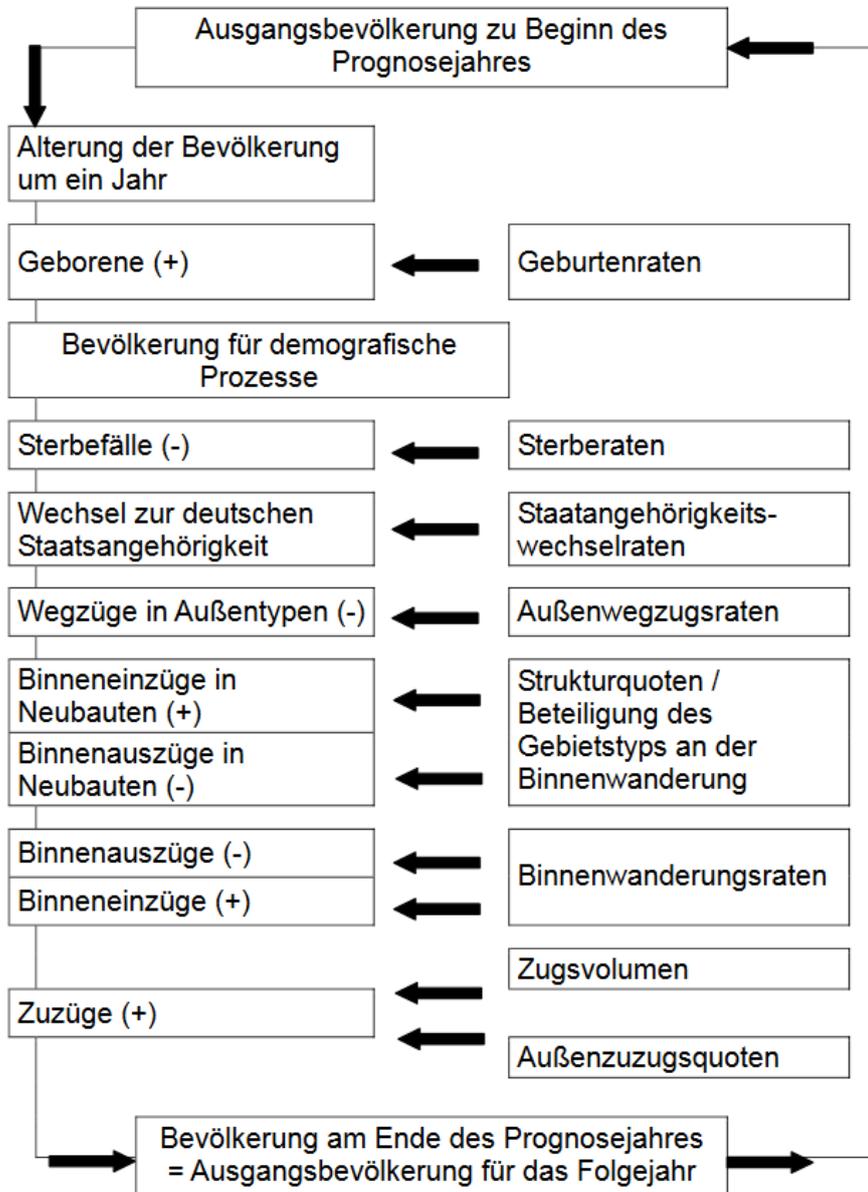


Abbildung 4: Das Berechnungsprinzip von SIKURS

Quelle: Willmann, T. (2014)

⁹⁷ Vgl. Willmann, T. 2014, S. 6

3.2 Problemanalyse

Die Prognose der Bevölkerungsentwicklung von ehemaligen neuen Baugebieten ist, wie gerade beschrieben, ein sehr zeitaufwändiger, kostenintensiver und komplizierter Prozess.

Das Prognosemodell, das von SIKURS angewendet wird, basiert auf der Kohorten-Komponenten-Methode. Diese Methode benötigt, wie schon in Kapitel 2 erwähnt wurde, aber auch wie im vorherigen Unterkapitel 3.1 zu sehen ist, eine sehr detaillierte Datengrundlage. Aber sehr oft und besonders auf kleineren räumlichen Ebenen ist der Zugang zu solchen Daten sehr schwer⁹⁸.

Für die Erstellung der Bevölkerungsprognose mittels der Kohorten-Komponenten-Methode ist es notwendig, die Bevölkerung des Basisjahres in ein- oder fünfjährige Altersgruppen aufzuteilen und je nachdem welche Altersgruppen definiert wurde, werden entsprechend auch die Projektionsintervalle bestimmt. Das führt dazu, dass diese Methode sehr rechenaufwändig ist. Da zudem die Altersgruppenintervalle ein wichtiges Merkmal für Planungsaufgaben sind (z.B. 3-6 Jahre Altersgruppe für Kindertageseinrichtungsplanung oder 6-10 Jahre Altersgruppe für Grundschulplanung), führt die Anwendung dieser Methode zu Beschränkungen^{99, 100, 101, 102}.

Außerdem gehen Kohorten-Komponenten-Methoden davon aus, dass die Überlebens- und Geburtenraten wie auch die Schätzungen des Wanderungssaldos während des gesamten Projektionszeitraums gleich bleiben. Somit sind sie sehr anfällig für Überschätzungen oder Unterschätzungen der Bevölkerungsentwicklung¹⁰³.

⁹⁸ Vgl. Hansen, H.S. 2010, S. 33

⁹⁹ Vgl. George, M. et al. 2004, S. 592

¹⁰⁰ Vgl. Smith, S. et al. 2001, S. 44-47

¹⁰¹ Vgl. Steinberg, J. und Doblhammer-Reiter, G. 2010, S. 393-394

¹⁰² Vgl. Bretz, M. 2000, S. 647

¹⁰³ Vgl. Burch, T.K. 2018, S. 137-140

Eine weitere Limitierung dieser Kohorten-Komponenten-Methode ist, dass sich ihre Modellanpassung nur durch lineare Funktionen verwirklichen lässt. Darüber hinaus können solche Modelle nicht die nicht-demographischen Faktoren, die das Bevölkerungswachstum oder den Bevölkerungsrückgang beeinflussen, berücksichtigen¹⁰⁴.

Abschließend werden Szenarien für die Prognostizierung der Population mit Anwendung der Kohorten-Komponenten-Methode und davon abgeleitete Prognosevarianten erstellt. Diese Prognosevarianten sind subjektiv, da die Szenarien auf Expertenwissen basieren. Außerdem ist es auch von Seite der Experten sehr schwer, solche Szenarien zu entwickeln. Zugleich sind diese Prognosevarianten für die Benutzer der Prognosen sehr verwirrend, da sie die Intervalle als Prognostizierungsgröße interpretieren können, obwohl diese „Intervalle“ keine tatsächliche probabilistische (statistische) Bedeutung haben¹⁰⁵.

3.3 Zusammenfassung

Es wird angestrebt die Erstellung von Prognosen der Bevölkerungsentwicklung und -struktur in Neubaugebieten des Landes Baden-Württemberg zu vereinfachen und somit ihre Zeit- und Kostenintensivität zu verringern.

Um das zu erreichen, wird in dieser Arbeit ein Prototyp für ein Bevölkerungsprognose-Tool als Unterstützung der kommunalen Entscheidungsträger bei der Infrastrukturplanung von Neubaugebieten entwickelt. Dieses Tool soll interaktiv mit dem Benutzer agieren. So soll es den Benutzer mit einfachen Angaben für das gewünschte Neubaugebiet, wie Größe des Gebiets und Erstbezugsjahr die zu erwartende Entwicklung der Bevölkerung des beabsichtigen neuen Baugebietes hinsichtlich der Einwohnerzahl und der demographischen Struktur, nämlich planungsgeeignete Altersgruppen und Geschlecht, wie auch die entsprechenden Visualisierungen online bereitstellen.

¹⁰⁴ Vgl. Burch, T.K. 2018, S. 138

¹⁰⁵ Vgl. Steinberg, J. und Doblhammer-Reiter, G. 2010, S. 394-397

Außerdem soll das Tool die Bevölkerungsentwicklung anhand von unterschiedlichen Prognostizierungsmethoden darstellen und einen Vergleich der Ergebnisse der Methoden anhand der Anpassungsgüte der Modelle anbieten, damit der Benutzer die freie Wahl hat, für seine Fragestellung die geeignete Prognose auszuwählen.

4 Lösungskonzept

Dieser Abschnitt umfasst die Beschreibung einer Lösungskonzeption für die Entwicklung eines prototypischen Bevölkerungsprognosen-Tools für Neubaugebiete des Landes Baden-Württemberg in Form einer interaktiven Webapplikation. Die Lösungsmöglichkeiten werden anhand der Anforderungen, die in Kapitel 1.2 „Ziel der Arbeit“ aufgezählt wurden, ausgewertet und überprüft. Beginnend im folgenden Abschnitt mit der Datengewinnung für die Bevölkerungsmodellierung folgt die prädiktive Analyse der Bevölkerungsentwicklung und abschließend die Vorstellung der Lösung für das Design der Webapplikation.

4.1 Datengewinnung

Wie bereits im zweiten Kapitel (Grundlagen) erwähnt wurde, basiert die Analyse der Bevölkerungsentwicklung auf der Verfügbarkeit genauer Daten über relevante Populationsmerkmale und -prozesse.

Um in dieser Arbeit die verschiedenen Arten von Bevölkerungsmodellen berechnen zu können, werden folgende Daten benötigt:

- **Bevölkerungsstrukturdaten** der jüngeren Vergangenheit in ehemaligen Neubaugebieten getrennt nach großen, mittleren/kleinen Baugebieten
- Mit Bevölkerungsstrukturdaten ist gemeint:
 - Bevölkerungsdaten nach Alter
 - Bevölkerungsdaten nach Geschlecht
- **Geburten** der jüngeren Vergangenheit in ehemaligen Neubaugebieten getrennt nach großen, mittleren/kleinen Baugebieten
 - Lebendgeborene nach Lebensalter der Mutter und Geschlecht des Kindes
- **Sterbefälle** der jüngeren Vergangenheit in ehemaligen Neubaugebieten getrennt nach großen, mittleren/kleinen Baugebieten
 - Sterbefälle nach Geschlecht und Alter

- **Wanderungen** der jüngeren Vergangenheit in ehemaligen Neubaugebieten getrennt nach großen, mittleren/kleinen Baugebieten
 - Zu- und Wegzüge nach Geschlecht und Alter

Da der Zugang zu solchen detaillierten Daten schwierig und zeitintensiv ist und die Verwirklichung dieser Arbeit einen festen Zeitraum voraussetzt, wurden Bevölkerungsdaten von ehemaligen Neubaugebieten der Stadt Freiburg verwendet. Die Stadt Freiburg wurde aus folgenden Gründen für die Zwecke dieser Arbeit als Datenlieferer ausgewählt, denn:

- Die Stadt Freiburg mit ihrem kommunalen Informationssystem FR.ITZ hält eine Vielzahl von Bevölkerungsdaten als OpenData bereit¹⁰⁶.
- Die Stadtentwicklung der Stadt Freiburg ähnelt vielen anderen Städten des Landes Baden-Württemberg, wie z.B. Heidelberg, Heilbronn, Ulm, Mannheim, etc.
- In der Stadt Freiburg gibt es große ehemalige Neubaugebiete, die auf die „Grüne Wiese“ gebaut wurden wie auch mittlere/kleine Neubaugebiete, die in schon bestehende Stadtteile integriert wurden.
- Durch eine intensive Zusammenarbeit zwischen ITEOS und der Stadt Freiburg war es möglich, durch das Amt für Statistik professionelle Informationen zum Umgang mit FR.ITZ zu erfahren.

Die ehemaligen Neubaugebiete, die für die Zwecke dieser Arbeit ausgewählt wurden, sind Rieselfeld und Vauban. Die Bebauung des Stadtteils Rieselfeld wurde auf einer Fläche von 70 Hektar mit 4.200 Wohnungen für 10.000 bis 12.000 Menschen im Jahr 1996 begonnen, in vier Bauabschnitte aufgeteilt und im Jahr 2012 abgeschlossen. Das Quartier Vauban besteht aus dem Neubaugebiet und dem bereits bewohnten Altbestand. Das Neubaugebiet mit Wohnungen für 5.000 bis 6.000 Menschen ist in drei Bauabschnitte eingeteilt. Die Bebauung des ersten Bauabschnitts begann im Jahr 1998

¹⁰⁶ <https://fritz.freiburg.de/Informationsportal>

und seit dem Jahr 2007 ist Vauban als statistischer Stadtteil in der Stadt Freiburg ausgewiesen¹⁰⁷.

4.2 Prädiktive Analyse der Bevölkerungsentwicklung

Bevölkerungsprognosen können mit zwei unterschiedlichen Arten von Methoden erstellt werden (siehe Kapitel 2.1). Da subjektive Projektionen einfach als "*wilde Vermutungen*" bezeichnet werden können und im besten Fall nur Eindrücke über zukünftige Bevölkerungstendenzen widerspiegeln¹⁰⁸, wird in dieser Arbeit die prädiktive Analyse der Bevölkerungsentwicklung in Neubaugebieten mit Anwendung von objektiven Methoden (Trendextrapolationsmodelle und Kohorten-Komponenten-Modelle) realisiert.

4.2.1 Trendextrapolationsmodelle

Um den Nachteil dieser Modelle im Vergleich zu den Kohorten-Komponenten-Modellen zu beseitigen und damit den gewünschten Detaillierungsgrad bei der Analyse der Bevölkerungsentwicklung zu erreichen, so dass die Planer und Entscheidungsträger von Neubaugebieten ihre Entscheidungen zuverlässig treffen können, werden in dieser Arbeit alle Bevölkerungsprognostizierungsmodelle nach planungsgerechten Altersgruppen und Geschlecht berechnet. Somit werden folgende acht Altersgruppen gebildet:

- 0 bis unter 3 Jahre (Krippenkinder)
- 3 bis unter 6 Jahre (Kindergartenkinder)
- 6 bis unter 10 Jahre (Grundschulkind)
- 10 bis unter 18 Jahre (Kinder und Jugendliche der mittleren Schulbildung)
- 18 bis unter 25 Jahre (Junge Erwachsene / Studenten)
- 25 bis unter 40 Jahre (Erwachsene)
- 40 bis unter 70 Jahre (ältere Erwachsene vor der Rente)
- 70 Jahre oder älter (ältere Erwachsene in Rente)

¹⁰⁷ <https://www.freiburg.de/pb/,Lde/618445.html>

¹⁰⁸ Wang, X. and vom Hofe, R. 2007, S. 55

Komponenten-Modelle

Da die Komponenten-Modelle die systematische Komponente einer Zeitreihe umfassen, wird als erstes in dieser Arbeit die Visualisierung der Bevölkerungsdaten wie auch die Berechnung der geeigneten statistischen Tests durchgeführt, um zu erkennen, ob bei den Daten deutliche Trends oder saisonale Muster vorliegen. Dieser Schritt ist von großer Bedeutung, um die geeigneten Modelltypen auszuwählen, die sich an den Verlauf der Zeitreihe gut anpassen lassen.

Wie bei den visuellen Beobachtungen der gesamten Einwohnerzahl in beiden ehemaligen Neubaugebieten (Rieselfeld und Vauban) zu erkennen ist, liegt ein deutlicher nicht-linearer Trend vor und es ist keine Saisonalität bei den Daten vorhanden (siehe Abb. 5 und Abb. 6). Die fehlende Saisonalität bestätigt auch der Weibel-Öllech-Saisonalitätstest. Die Null-Hypothese für den Weibel-Öllech-Saisonalitätstest ist, dass die Zeitreihe saisonal ist. Da der berechnete P-Wert größer als 0.05 ist, wird die Null-Hypothese abgelehnt und somit existiert keine saisonale Komponente.

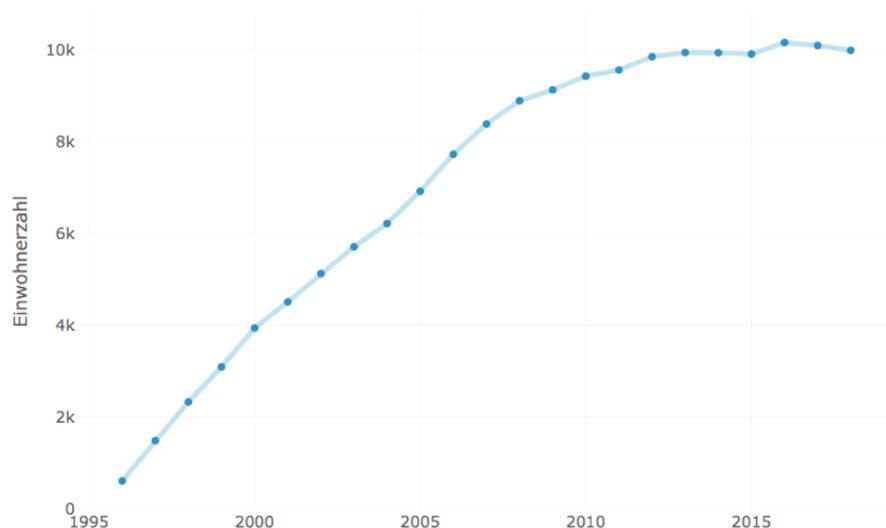


Abbildung 5: Gesamte Einwohnerzahl (Rieselfeld)

Quelle: Eigene Darstellung

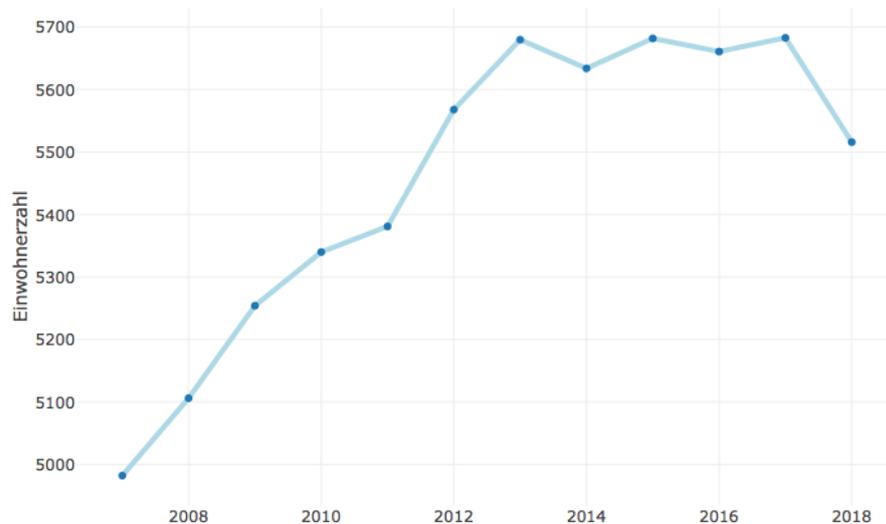


Abbildung 6: Gesamte Einwohnerzahl (Vauban)

Quelle: Eigene Darstellung

Die Abbildungen 7 bis 10 zeigen den Verlauf der Einwohnerzahl getrennt nach Geschlecht in Rieselfeld und Vauban. Auch in diesem Fall ist das Vorhandensein von nicht-linearen Trendkomponenten und das Fehlen von saisonalen Komponenten zu beobachten. Ebenso hat der Weibel-Ollech-Saisonalitätstest diese Annahme erwiesen.

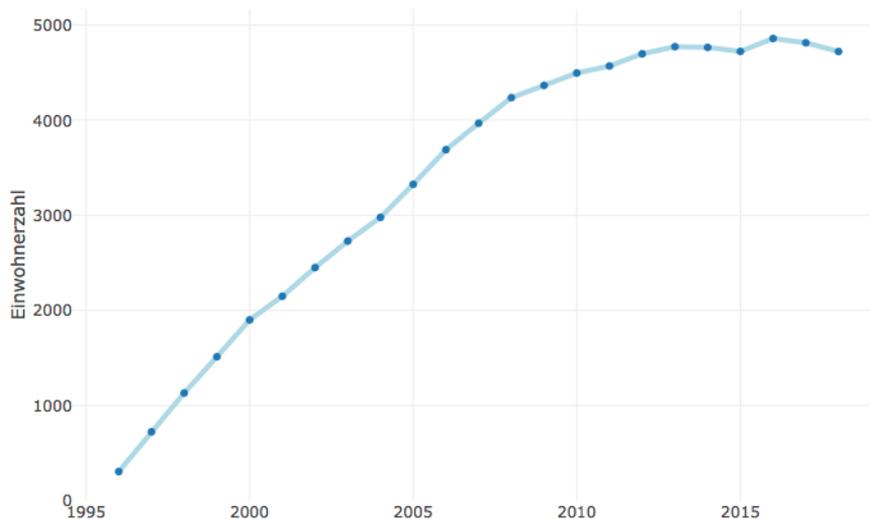


Abbildung 7: Männliche Einwohnerzahl (Rieselfeld)

Quelle: Eigene Darstellung

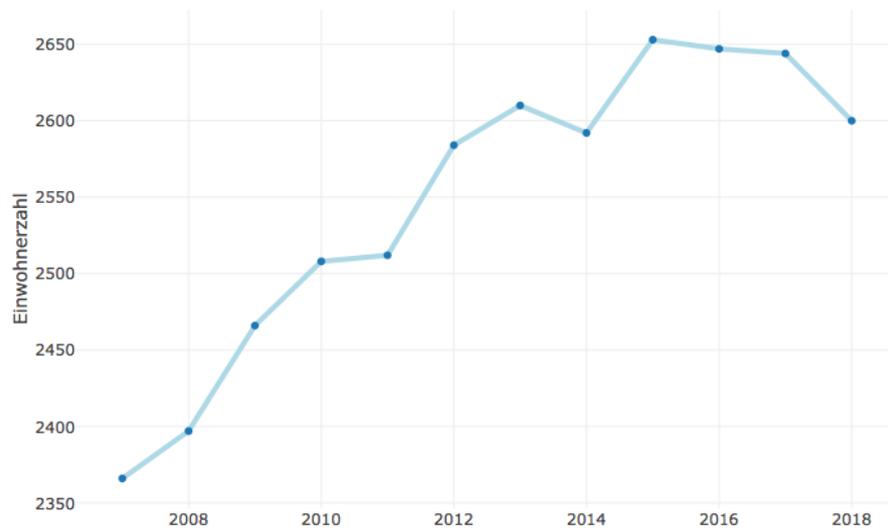


Abbildung 8: Männliche Einwohnerzahl (Vauban)

Quelle: Eigene Darstellung

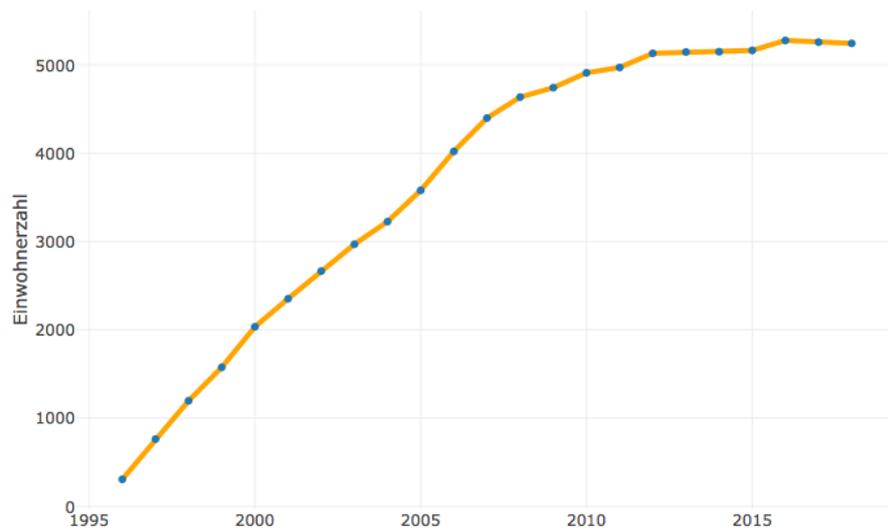


Abbildung 9: Weibliche Einwohnerzahl (Rieselfeld)

Quelle: Eigene Darstellung

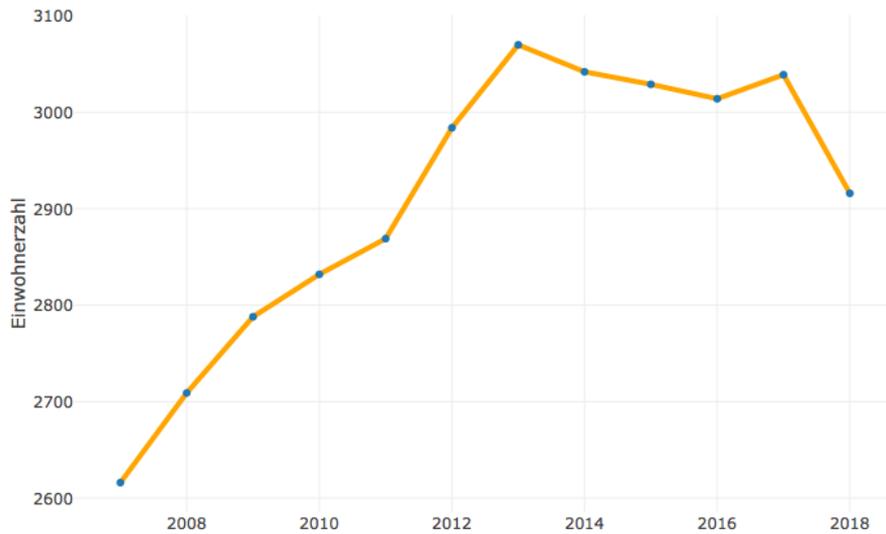


Abbildung 10: Weibliche Einwohnerzahl (Vauban)

Quelle: Eigene Darstellung

Ferner offenbaren die Visualisierungen und der Saisonalitätstest der Einwohnerzahl getrennt nach Altersgruppen, dass nur nicht-lineare Trendeffekte bei den Daten vorliegen (siehe Abb. 11 und Abb. 12)

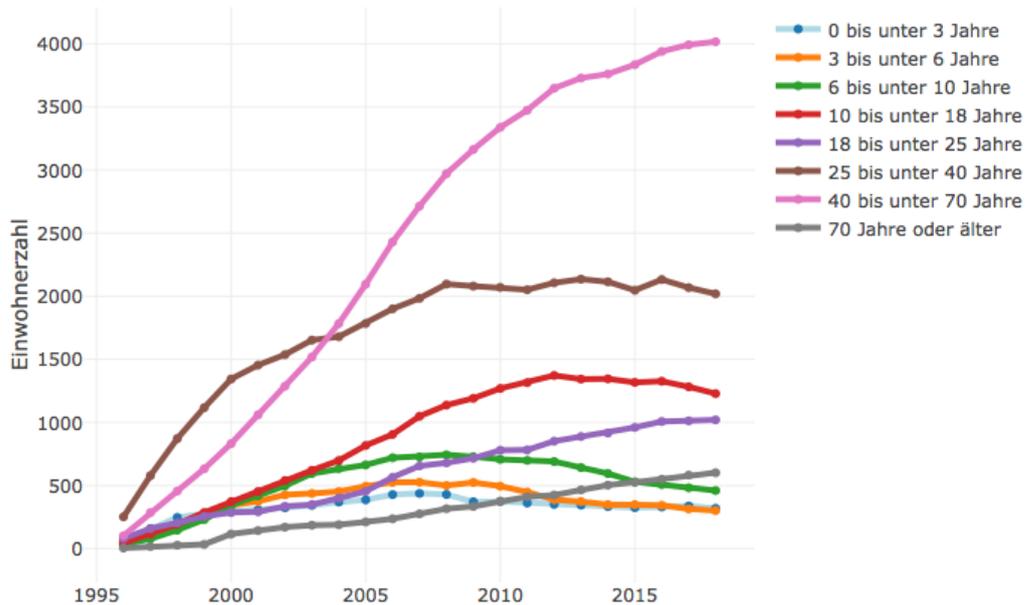


Abbildung 11: Einwohnerzahl nach Altersgruppen (Rieselfeld)

Quelle: Eigene Darstellung

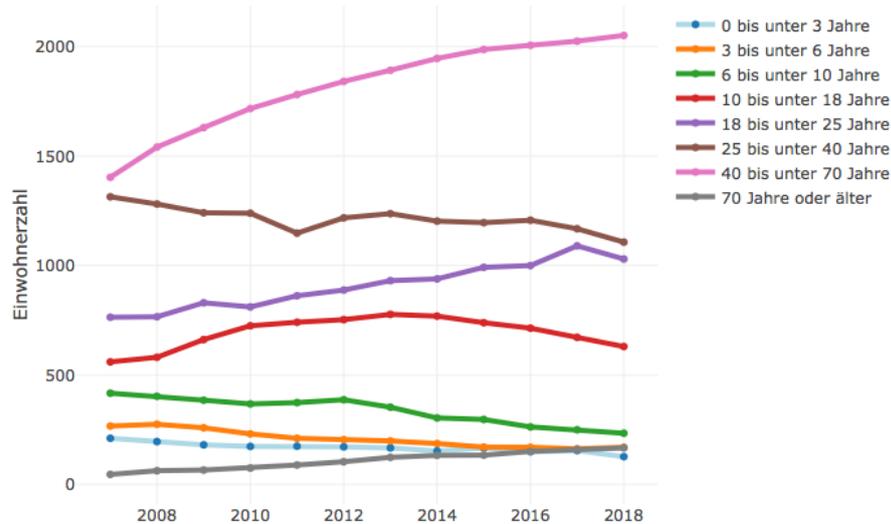


Abbildung 12: Einwohnerzahl nach Altersgruppen (Vauban)

Quelle: Eigene Darstellung

Anhand der Erkenntnisse, die durch die Visualisierung der Bevölkerungsentwicklungsdaten der beiden ehemaligen Neubaugebieten gewonnen wurden, werden in dieser Arbeit für die Bevölkerungsprognosen nach Altersgruppen und Geschlecht Komponenten-Modelle berechnet, die nichtlineare Trendverläufe darstellen können. Aus der klassischen Statistik werden GAM-Modelle angewendet. In der auf maschinellem Lernen basierenden Vergleichsprognose kommen KNNs zum Einsatz. Beide Modelltypen ermöglichen es, sehr komplexe nichtlineare Abhängigkeiten zu reproduzieren.

Autoregressive Modelle

Autoregressive Modelle stellen eine der Standard-Methoden der Zeitreihenanalyse dar, weshalb sie auch in dieser Arbeit implementiert werden. Für die Auswahl des geeigneten Modelltyps ist es wichtig, zu bestimmen, ob die Zeitreihe stationär ist oder nicht.

Die Analyse von Autokorrelationsfunktion (ACF)-Diagrammen ist eine aufschlussreiche Methode zur Identifizierung der Nichtstationarität von Zeitreihen. Solche Diagramme zeigen die Autokorrelationen, die die Beziehung zwischen Y_t und Y_{t-k} für verschiedene Werte von k beschreiben. Im Allgemeinen werden die Ergebnisse von ACF-

Diagrammen so interpretiert, dass für eine stationäre Zeitreihe der ACF relativ schnell auf null fällt, während der ACF einer nicht-stationären Zeitreihe langsam abnimmt¹⁰⁹.

Um die durch die ACF-Interpretation getroffene Annahme statistisch zu bestätigen, wird der Augmented-Dickey-Fuller-Test (ADF-Test) ausgeführt. Der ADF-Test ist einer der statistischen Einheitswurzeltests, der häufig zur Bestimmung der Zeitreihenstationarität verwendet wird. Die Null-Hypothese für den Test ist, dass die Daten nicht stationär sind. Normalerweise wird ein Schwellenwert von 5% verwendet, was bedeutet, dass die Null-Hypothese beibehalten wird, wenn der p-Wert kleiner als 0.05 ist¹¹⁰.

Die Darstellung von ACF der Gesamtbevölkerungsdaten zeigt für beide ehemaligen Neubaugebiete einen langsam abnehmenden Fortschritt mit klaren Anzeichen dafür, dass keine regelmäßig wiederkehrenden Muster existieren. Dies entspricht der Nichtstationarität der Zeitreihe, mit dem Vorhandensein von nur Trendkomponenten (siehe Abb. 13 und Abb. 14). Zusätzlich wurde der ADF-Test durchgeführt. Der p-Wert des Tests war für beide Gebiete 0.99. Das bedeutet, dass die Null-Hypothese des Tests verworfen werden kann, wodurch die Hypothese der Nichtstationarität der Zeitreihen erwiesen ist. Entsprechend sind die Aussagen der ACF-Diagramme der Bevölkerungszeitreihen getrennt nach Geschlecht und Altersgruppe (siehe Abbildungen 1 bis 20 im Anhang) und die Ergebnisse des ADF-Tests.

Infolgedessen werden in dieser Arbeit die ARIMA-Modelle angewendet, die unter Verwendung von Differenzierungsmethoden mit nichtstationären Zeitreihen umgehen können.

¹⁰⁹ Konasani, V.R. and Kadre S. 2015, S. 474-484

¹¹⁰ Konasani, V.R. and Kadre S. 2015, S. 456

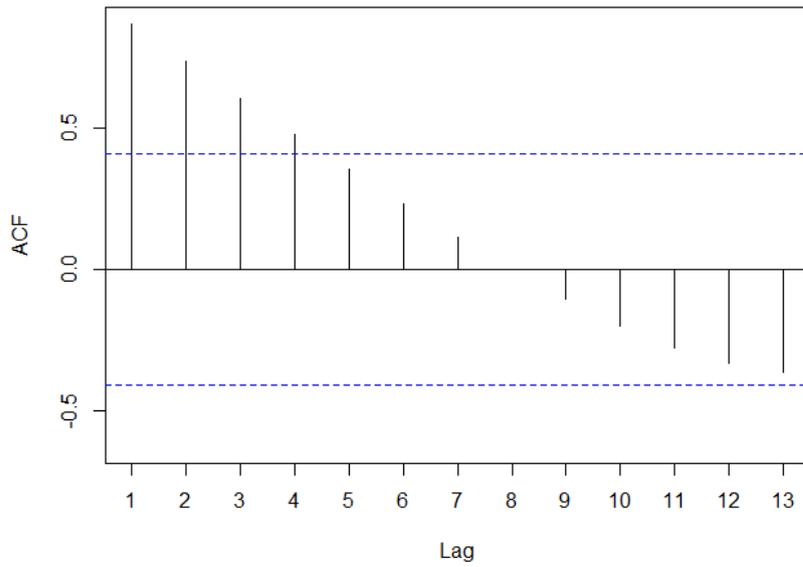


Abbildung 13: Autokorrelationsfunktion der Gesamtbevölkerungsdaten (Rieselfeld)

Quelle: Eigene Darstellung

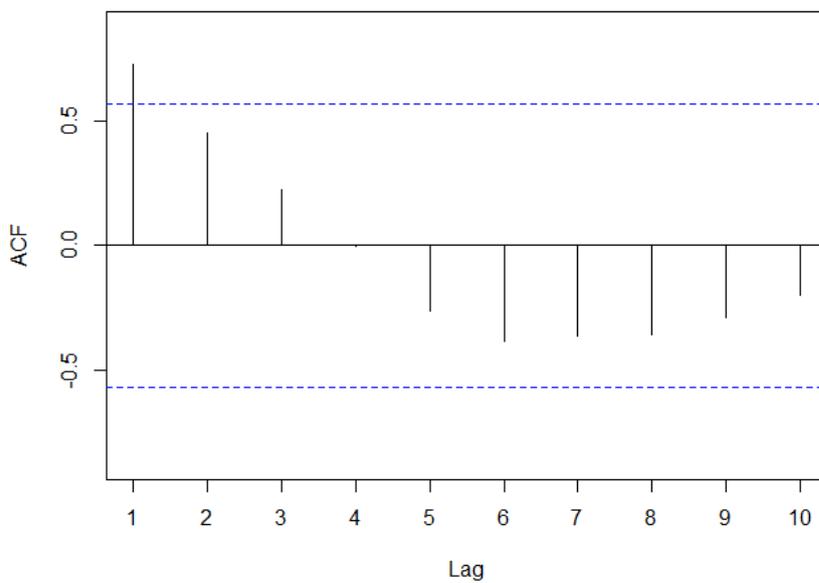


Abbildung 14: Autokorrelationsfunktion der Gesamtbevölkerungsdaten (Vauban)

Quelle: Eigene Darstellung

4.2.2 Kohorten-Komponenten-Modelle

Die Kohorten-Komponenten-Modelle sind das am meisten verwendete Werkzeug für die Prognostizierung der Bevölkerungsentwicklung. Daher wird in dieser Arbeit die erwartete Bevölkerungszahl und -struktur auch unter Verwendung dieser Methode berechnet.

Das hier berechnete Modell zielt darauf ab, Bevölkerungsprognosen für Neubaugebiete zu erstellen, die individuell verschiedenen demografischen Komponenten in Anspruch nehmen. Im Detail wird bei diesem Modell die Bevölkerung durch die Anwendung von zeitspezifischen Überlebensverhältnissen nach Alter und Geschlecht und die Ableitung von Geburten aus zeitspezifischen Fertilitätsraten von Frauen nach Alter vorhergesagt. Außerdem wird die Migration nach Alter und Geschlecht einbezogen.

Somit wird das Populationsprognostizierungsmodell durch die folgende allgemeine Bevölkerungsgleichung beschrieben, wobei sich das hochgestellte n auf das ehemalige Neubaugebiet bezieht:

$$B_{t+1,i,s}^n = B_{t,i,s}^n + G_{i,s}^n - T_{i,s}^n + M_{i,s}^n$$

Die Variablen $B_{t,i,s}$ und $B_{t+1,i,s}$ beziehen sich auf die Bevölkerung des Geschlechts s , Alter i zum Zeitpunkt t und $t+1$. G_s ist die Anzahl der neugeborenen Jungen ($s=1$) und Mädchen ($s=2$), während $T_{i,s}$ die Anzahl der Todesfälle für jede Altersgruppe und jedes Geschlecht bezeichnet. Die letzte Variable $M_{i,s}$ bezieht sich auf die Netto-Migration. Aufgrund der Rechenintensivität dieser Modelle wird die Bevölkerung in fünf-jährige Kohorten eingeteilt und damit beträgt das Vorhersageintervall fünf Jahre.

4.2.3 Modellvergleich

Eine der Anforderungen, die in dieser Arbeit erfüllt werden muss, ist dem Benutzer ein Vergleich der Ergebnisse der unterschiedlichen Methoden anzubieten, um ihm damit die „freie Wahl“ zu geben, die für ihn geeignete Methode auszuwählen. Dieser Modell-

vergleich wird in dieser Arbeit anhand zwei statistischer Kennzahlen der Anpassungsgüte der Modelle, nämlich den Prognosefehler und den Bestimmtheitsmaß (R-Quadrat) umgesetzt.

Es gibt viele Möglichkeiten, wie man das Maß des Prognosefehlers berechnet. Das in der Literatur am häufigsten verwendete Prognosefehlermaß ist der Prozentsatz des absoluten Fehlers des Erwartungswertes (Mean Absolute Percentage Error – MAPE –). MAPE ist ein leicht verständliches statistisches Maß, da es den prozentualen Fehler ausdrückt. Außerdem eignet es sich sehr gut für den Vergleich der Güte unterschiedlicher Prognosemethoden mit verschiedenen Testdaten.

4.3 Design der Webapplikation

Das Frontend des Bevölkerungsprognose-Tools wird in Form eines Dashboards realisiert, das den Benutzer zur Interaktion auffordert und die Ergebnisse seiner Anforderungen mittels interaktiver Visualisierungen darstellt.

Das Dashboard besteht aus zwei Komponenten, dem Sidebar-Menü und dem Hauptaktivitätsfenster. Die grundlegende Aufgabe des Sidebar-Menüs ist eine Übersicht aller angebotenen Funktionen des Bevölkerungsprognose-Tools zu zeigen und gleichzeitig den Benutzer durch die verschiedenen Prognostizierungsoptionen zu navigieren. Das Hauptaktivitätsfenster dient als „Leinwand“ zur Visualisierung der Prognosemodelle und zum Interagieren mit dem Benutzer.

Ins Detail gehend liefert die Startseite des Dashboards allgemeine Informationen über das Bevölkerungsprognose-Tool, wie z.B. wie und welche Modelle berechnet werden. Außerdem werden dort aktuelle Bevölkerungskennzahlen der Stadt, wie auch eine Bevölkerungspyramide der aktuellen Jahre bereitgestellt. Folglich wird der Benutzer aufgefordert, Informationen über das gewünschte Neubaugebiet, wie die Größe des Gebiets und das Erstbezugsjahr anzugeben und eine Prognostizierungsmethode auszuwählen. Anhand seiner Auswahl wird die Prognose der Bevölkerungsentwicklung für das definierte Neubaugebiet durch interaktive Grafiken bereitgestellt.

Da die Visualisierungen der Bevölkerungsprognose auf Daten von ehemaligen Neubaugebieten der Stadt Freiburg basieren (siehe Gründe in Kapitel 4.1), wird die Webapplikation dem Benutzer die Möglichkeit geben auch seine eigenen Daten von ehemaligen Neubaugebieten hochzuladen und zu analysieren. Dabei werden neue Modelle mit seinen Daten berechnet und verglichen, um fallbezogene Bevölkerungsprognosen zu erhalten.

Abschließend werden für das Design des Dashboard-Frontends Aspekte der Benutzerfreundlichkeit berücksichtigt. Somit wird das Dashboard so gestaltet sein, dass der Benutzer ohne vorherige Einweisung oder Schulung in der Bedienung mit dem Dashboard arbeiten kann. Außerdem wird bei der Realisierung des Bevölkerungsprognose-Dashboards ein erweiterbares Framework verwendet, so dass zukünftige Änderungen mit geringem Aufwand und ohne überflüssige technische Hindernisse umgesetzt werden können.

5 Umsetzung

Im Folgenden wird die Umsetzung des im vorherigen Kapitel vorgestellten Lösungskonzepts beschrieben.

5.1 Implementierung der prädiktiven Analyse

Wie bereits im vorherigen Kapitel dargestellt wurde, werden für die Realisierung der prädiktiven Analyse der Bevölkerungsentwicklung in Neubaugebieten verschiedene Arten von Modellen verwendet:

- Komponenten-Modelle (Trend-Modelle)
 - Klassische Statistik: Generalisierte Additive Modelle
 - Maschinelles Lernen: Mehrschichtige Feedforward Neuronale Netze
- ARIMA-Modelle
- Kohorten-Komponenten-Modelle

Anschließend werden diese miteinander verglichen.

Für die Implementierung der Zeitreihenanalyse gibt es mehrere Programmiersprachen, die angewendet werden können. In dieser Arbeit wurde die Programmiersprache R eingesetzt. R ist ein Open-Source-Projekt, das eine frei verfügbare und qualitativ hochwertige Computing-Umgebung mit einer großen Anzahl von Add-On-Bibliotheken (*R-Packages*) bietet. R vereint viele Jahre früherer Forschung im Bereich der statistischen und numerischen Datenverarbeitung und basiert somit auf einem soliden Fundament von statistischen und numerischen Kernalgorithmen. Die R-Programmiersprache ist eine funktionale und interaktive Skriptsprache, die zwei Ebenen der objektorientierten Programmierung bietet. Außerdem ist R die am häufigsten verwendete Statistik-Programmiersprache für die Visualisierung und Analyse von Zeitreihen¹¹¹. Auf diese Weise ist R das passende Werkzeug für die Verwirklichung der prädiktiven Analyse im Rahmen dieser Arbeit.

¹¹¹ McLeod et al. 2012, S. 662

Datenvorbereitung

Die notwendigen Bevölkerungsdaten von den zwei ehemaligen Neubaugebieten (Rieselfeld und Vauban) wurden von dem kommunalen Informationssystem FR.ITZ der Stadt Freiburg im csv-Format heruntergeladen. Für die Umwandlung dieser Rohdaten in für die Zeitreihenanalyse geeignete Daten wurden mehrere Schritte durchgeführt. Im ersten Schritt wurde die csv-Datei mithilfe der Funktion *read.csv2()* der R-Bibliothek *utils* in R eingelesen und daraus ein Dataframe¹¹² generiert.

Danach wurde das Dataframe nach Gebiet, Geschlecht und Altersgruppe gefiltert und es wurden unter Verwendung der Funktion *ts* der R-Bibliothek *stats* wurden die folgenden Einwohnerzahl-Zeitreihen gebildet:

- eine Zeitreihe für die Gesamtbevölkerung
- zwei Zeitreihen, je eine für jedes Geschlecht
- acht Zeitreihen, je eine für jede Altersgruppe
- sechszehn Zeitreihen, je eine für jede Altersgruppe getrennt nach Geschlecht

Dabei gelten die Zeitreihen für das Rieselfeld für den Zeitraum 1996 bis 2018 und für Vauban von 2007 bis 2018.

Generalisierte Additive Modelle

Die Berechnung der Trend-Modelle für beide Gebiete wurde mit Hilfe des statistischen Verfahrens der generalisierten additiven Modelle realisiert. Diese Modelle wurden durch Anwendung der Funktion *gam()* der R-Bibliothek *mgcv*¹¹³ implementiert. Die *gam()* Funktion passt ein generalisiertes additives Modell an die Zeitreihen an. Die Modelle werden durch einen quadratisch penalisierten Likelihood-Ansatz geschätzt. Der Grad der Glätte der Modellkomponenten wird im Rahmen der Anpassung geschätzt und Konfidenzintervalle werden berechnet.

Somit wurden folgende GAM-Modelle für die Zeitreihen des Gebietes Rieselfeld/Vauban gebildet:

¹¹² Dataframe ist ein besondere R-Begriff, der sich mit dem Begriff *Tabelle* in einer Datenbank ähnelt

¹¹³ Wood, S. N. 2017

- ein Modell mit der Antwortvariablen Gesamteinwohnerzahl und der Prädiktorvariablen Zeit in Jahreszahlen
- zwei Modelle, je eins für jedes Geschlecht. Dementsprechend mit der Antwortvariablen Einwohnerzahl pro Geschlecht und der Prädiktorvariablen Zeit in Jahreszahlen
- acht Modelle, je eins für jede Altersgruppe. Dementsprechend mit der Antwortvariablen Einwohnerzahl pro Altersgruppe und der Prädiktorvariablen Zeit in Jahreszahlen
- sechszehn Modelle, je eins für jede Altersgruppe getrennt nach Geschlecht. Dementsprechend mit der Antwortvariable Einwohnerzahl pro Altersgruppe und Geschlecht und der Prädiktorvariablen Zeit in Jahreszahlen

Die Bevölkerungsprognose jedes Modells wurde für einen Vorhersagehorizont von 30 Jahren für Rieselfeld und 20 Jahren für Vauban berechnet. Dies wurde mittels der Funktion *predict()* der R-Bibliothek *mgcv* implementiert. Für die Berechnung der Anpassungsgüte der Modelle, R-Quadrat, wurde die Funktion *summary()* der R-Bibliothek *mgcv* verwendet und für das MAPE die Funktion *accuracy()* der R-Bibliothek *forecast*¹¹⁴.

Die berechneten Modelle für die Entwicklung der Gesamtbevölkerung zeigten für beide Gebiete eine sehr hohe Vorhersagegenauigkeit mit einem R-Quadrat-Wert von 0.99 für Rieselfeld und 0.97 für Vauban und einem MAPE-Wert von 0.86% und 0.51%. Die hohe Anpassungsgüte der Modelle kann auch durch die folgenden Diagramme, auf denen die Rohzeitreihendaten und die vorhergesagten Werte der Modelle visualisiert sind, verdeutlicht werden (siehe Abb. 15 und Abb. 16).

¹¹⁴ Hyndman and Khandakar 2008

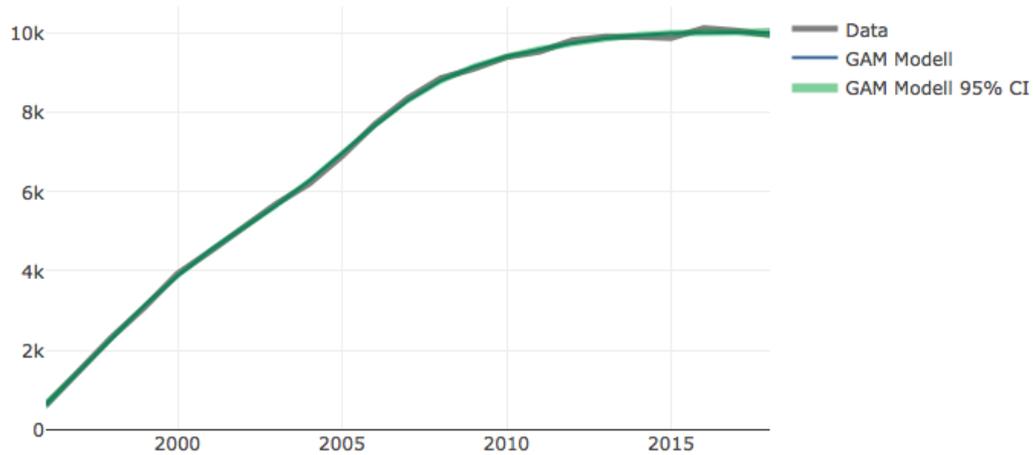


Abbildung 15: Rohdaten vs Vorhergesagte Werte für die Gesamtbevölkerung (Rieselfeld)

Quelle: Eigene Darstellung

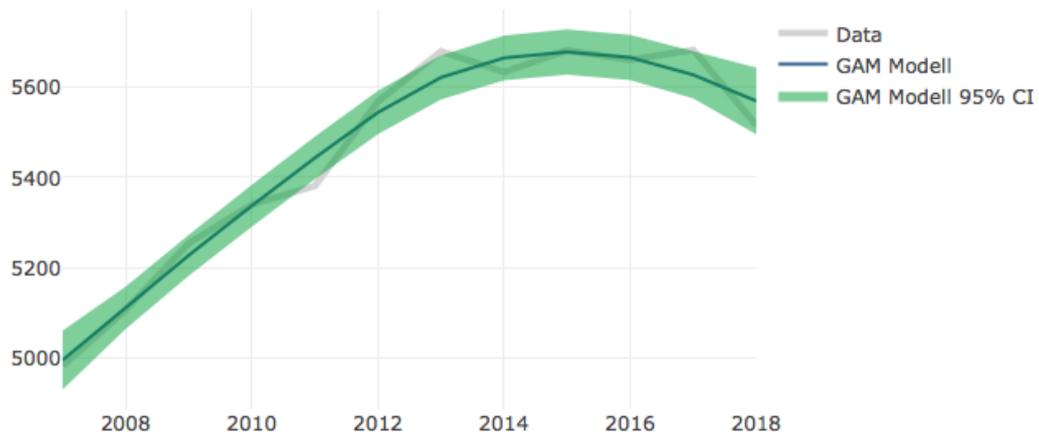


Abbildung 16: Rohdaten vs Vorhergesagte Werte für die Gesamtbevölkerung (Vauban)

Quelle: Eigene Darstellung

Die Abbildungen 17 und 18 zeigen die Bevölkerungsprognose der gesamten Einwohnerzahl für Rieselfeld (30 Jahre) und für Vauban (20 Jahre). Es ist deutlich zu sehen, dass in großen Neubaugebieten, die auf leere Flächen gebaut werden, die Einwohnerzahl in den ersten fünfzehn Jahren sehr stark zunimmt und danach, wenn die Wohnraumkapazität erschöpft ist, bis auf eine leichte Senkung stabil bleibt. Einen ähnlichen

Verlauf zeigt auch die Einwohnerzahl in kleinen/mittleren Neubaugebieten, die in schon bestehende Stadtteile gebaut werden, mit dem einzigen Unterschied, dass in solchen Gebieten die Einwohnerzahlentwicklung weniger dynamisch ist.

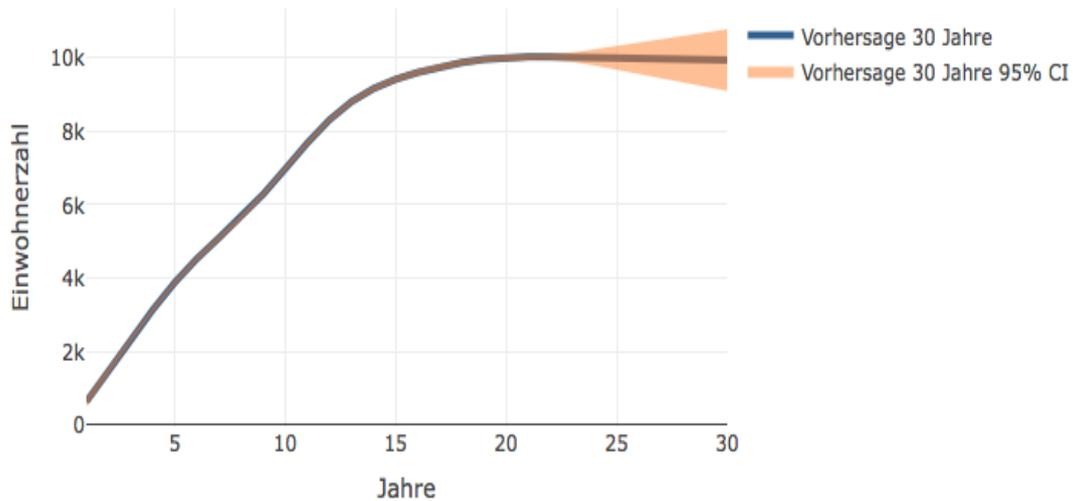


Abbildung 17: 30-jährige Prognose der Gesamtbevölkerung (Rieselfeld)

Quelle: Eigene Darstellung

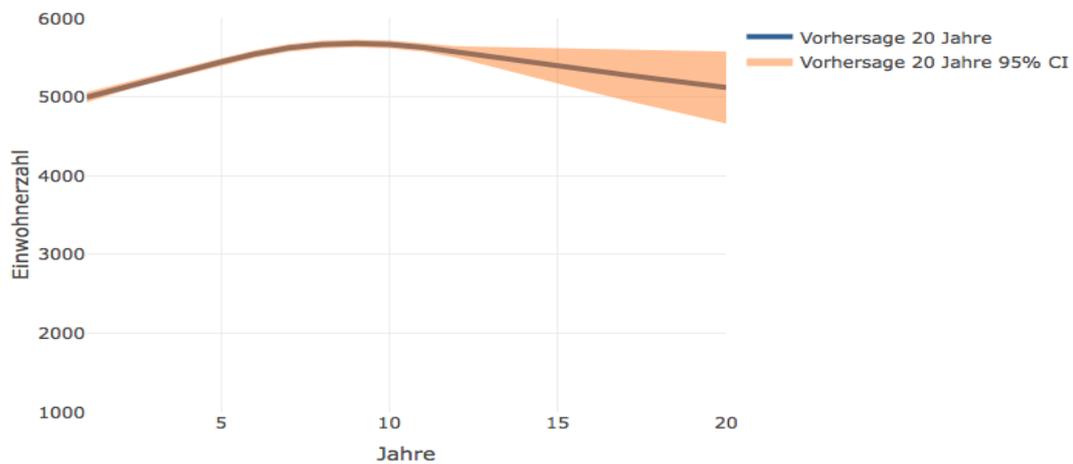


Abbildung 18: 20-jährige Prognose der Gesamtbevölkerung (Vauban)

Quelle: Eigene Darstellung

Auch bei der Prognostizierung der Einwohnerzahl getrennt nach Geschlecht haben sich die GAM-Modelle sehr gut an die Rohdaten angepasst. Das verdeutlichen die Werte der beiden Anpassungskennzahlen, die in Tabelle 1 zu sehen sind, wie auch die Visuali-

sierungen der Rohdaten im Vergleich zu den vom Modell vorhergesagten Werten (siehe Abb. 19 bis Abb. 22).

Tabelle 1: Anpassungsgüte der GAM-Modelle nach Geschlecht

	Rieselfeld		Vauban	
	R-Quadrat	MAPE	R-Quadrat	MAPE
Männliche Einwohnerzahl	0.99	0.72 %	0.96	0.57 %
Weibliche Einwohnerzahl	0.99	1.02 %	0.98	0.26 %

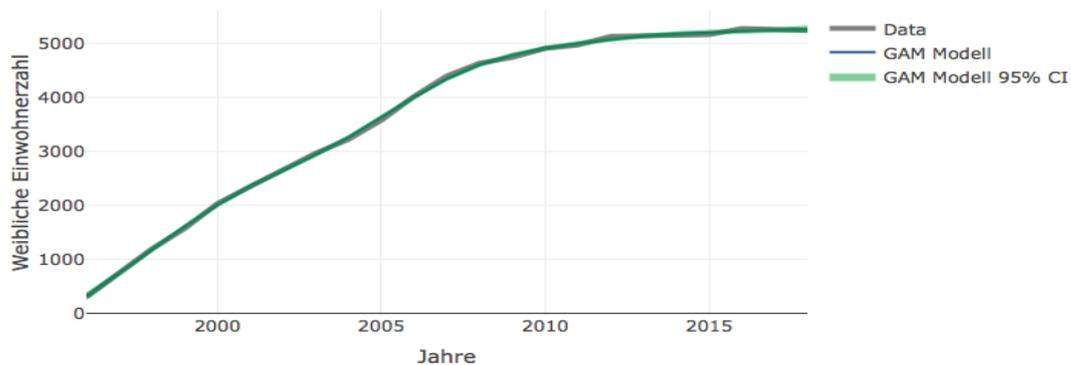


Abbildung 19: Rohdaten vs Vorhergesagte Werte für die weibliche Bevölkerung (Rieselfeld)

Quelle: Eigene Darstellung

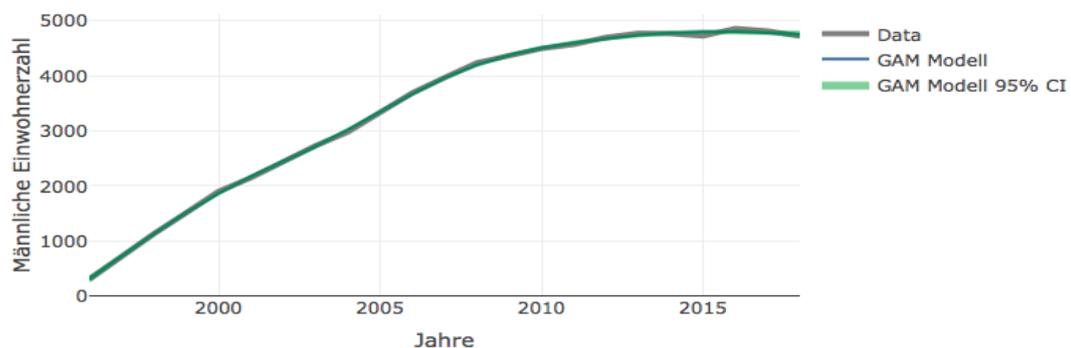


Abbildung 20: Rohdaten vs Vorhergesagte Werte für die männliche Bevölkerung (Rieselfeld)

Quelle: Eigene Darstellung



Abbildung 21: Rohdaten vs Vorhergesagte Werte für die weibliche Bevölkerung (Vauban)

Quelle: Eigene Darstellung

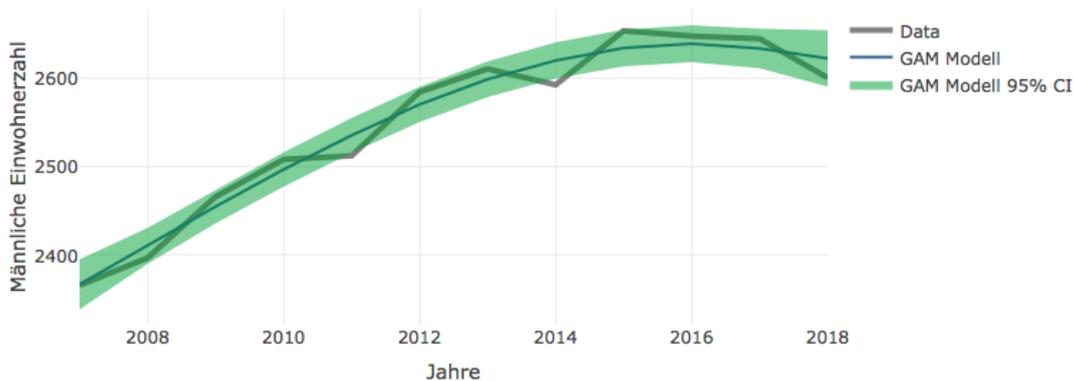


Abbildung 22: Rohdaten vs Vorhergesagte Werte für die männliche Bevölkerung (Vauban)

Quelle: Eigene Darstellung

Die nach Geschlecht getrennte Vorhersage der Bevölkerungsentwicklung zeigt bei großen Neubaugebieten auf, dass in den ersten Bewohnungsjahren der Anteil der beiden Geschlechter, wahrscheinlich wegen dem Einzug von jungen Familien, gleich ist und im Laufe der Zeit die Anzahl weiblicher Einwohnerinnen höher als die der männlichen steigt. In kleinen/mittleren Neubaugebieten ist wegen dem bereits bewohnten Altbestand von Anfang an der Anteil der weiblichen Einwohner höher allerdings mit einer im Laufe der Jahre sinkenden Tendenz (siehe Abb. 23 und Abb. 24).

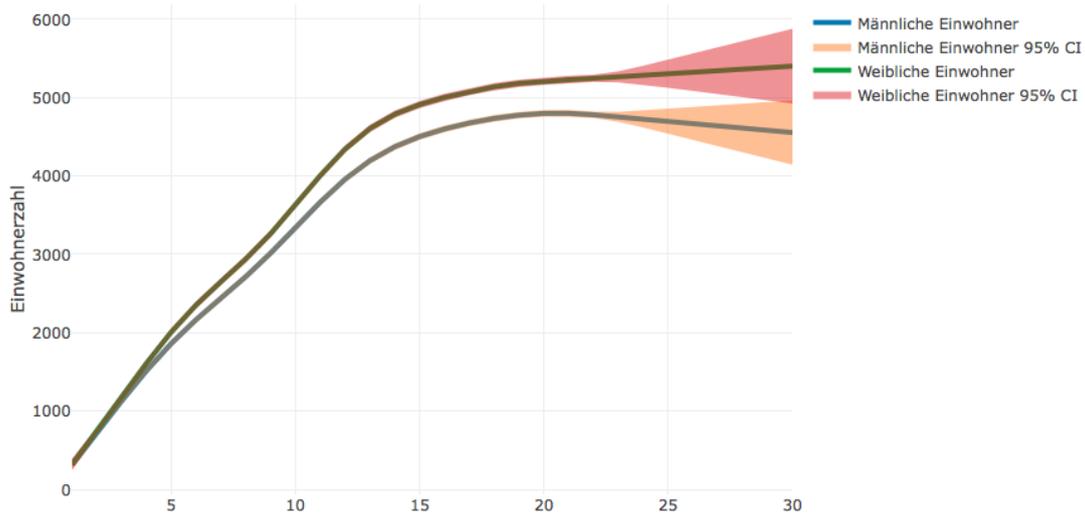


Abbildung 23: 30-jährige Prognose der Bevölkerung getrennt nach Geschlecht (Riesel-feld)

Quelle: Eigene Darstellung

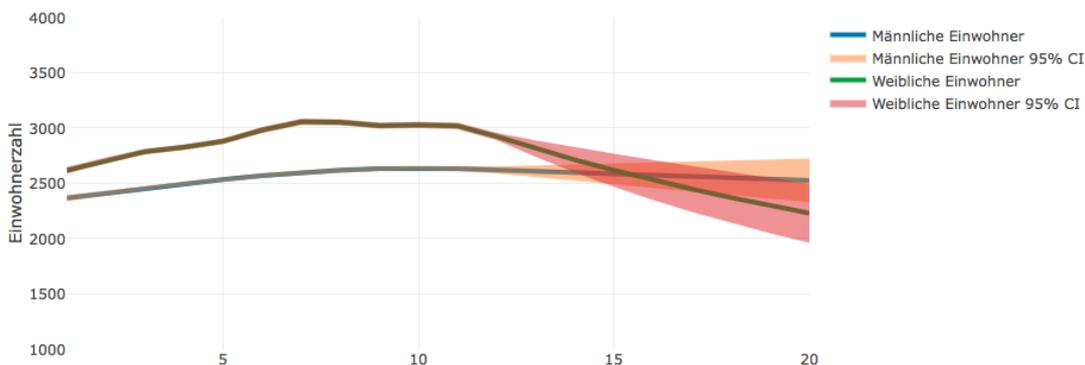


Abbildung 24: 20-jährige Prognose der Bevölkerung getrennt nach Geschlecht (Vau-ban)

Quelle: Eigene Darstellung

Die sehr hohen R-Quadrat-Werte von 0.98 bis 1 für Riesel-feld und von 0.77 bis 0.99 für Vau-ban, wie auch die sehr niedrigen MAPE-Werte für beide Gebiete weisen auch auf eine sehr hohe Vorhersagegenauigkeit der GAM-Modelle bei der Bevölkerungsprognose getrennt nach Altersgruppe hin (siehe Tabelle 2). Dies veranschaulichen auch die Grafiken der beobachteten versus vorhergesagten Werte (siehe exemplarisch Abb.25 bis Abb.28).

Tabelle 2: Anpassungsgüte der GAM-Modelle nach Altersgruppe

	Rieselfeld		Vauban	
	R-Quadrat	MAPE	R-Quadrat	MAPE
0 bis unter 3 Jahre	0.98	2.35 %	0.89	3.01 %
3 bis unter 6 Jahre	0.99	2.52 %	0.99	0.82 %
6 bis unter 10 Jahre	0.99	2.53 %	0.98	1.30 %
10 bis unter 18 Jahre	0.99	1.47 %	0.99	0.25 %
18 bis unter 25 Jahre	0.99	2.42 %	0.94	1.88 %
25 bis unter 40 Jahre	0.99	1.13 %	0.77	1.32 %
40 bis unter 70 Jahre	1	0.92 %	0.99	0.14 %
70 Jahre oder älter	0.99	9.19 %	0.99	0.47 %



Abbildung 25: Rohdaten vs Vorhergesagte Werte für die Bevölkerung der Altersgruppe 0 bis unter 3 Jahre (Rieselfeld)

Quelle: Eigene Darstellung

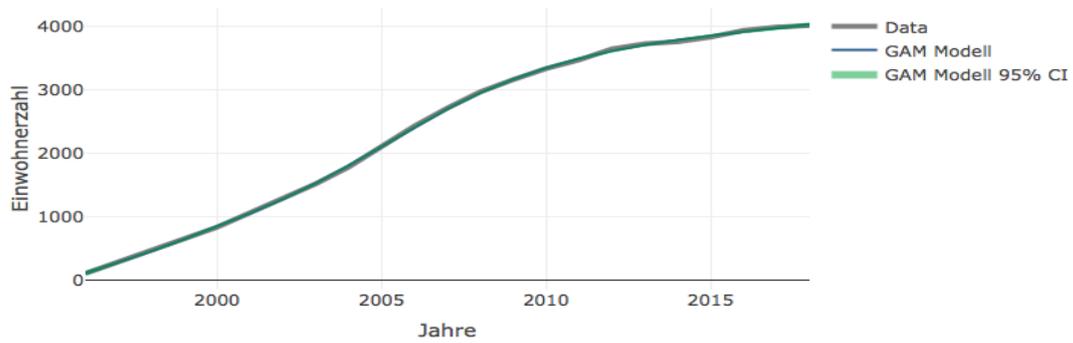


Abbildung 26: Rohdaten vs Vorhergesagte Werte für die Bevölkerung der Altersgruppe 40 bis unter 70 Jahre (Rieselfeld)

Quelle: Eigene Darstellung



Abbildung 27: Rohdaten vs Vorhergesagte Werte für die Bevölkerung der Altersgruppe 0 bis unter 3 Jahre (Vauban)

Quelle: Eigene Darstellung

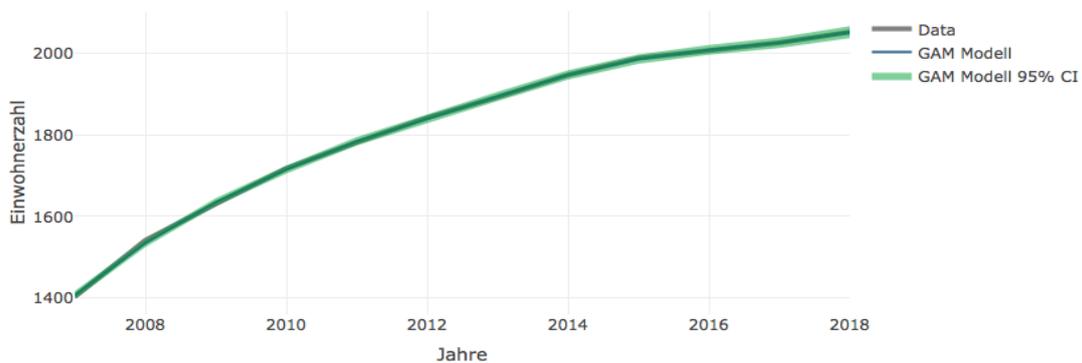


Abbildung 28: Rohdaten vs Vorhergesagte Werte für die Bevölkerung der Altersgruppe 40 bis unter 70 Jahre (Vauban)

Quelle: Eigene Darstellung

Bei den Zukunftsprognosen der Einwohnerzahlen für beide Arten von Neubaugebieten ist zu beobachten, dass der Anteil der älteren Erwachsenen (40 bis unter 70 Jahre und 70 Jahre oder älter) mit der Zeit sehr stark zunimmt und der Anteil der Kleinkinder (0 bis unter 3 Jahre und 3 bis 6 Jahre) ständig fällt. Dies zeigt eine Alterung der Bevölkerung in den Gebieten (siehe Abb. 29 und Abb. 30).

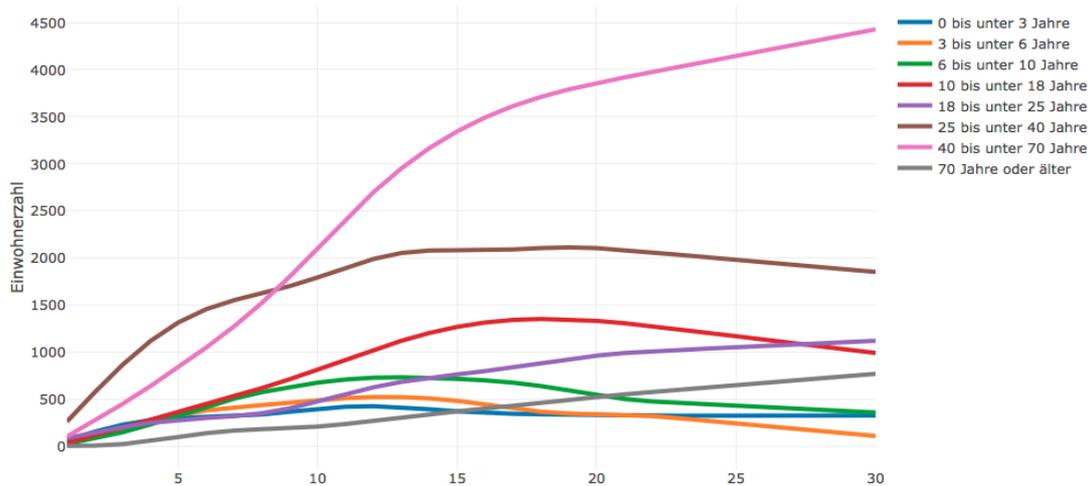


Abbildung 29: 30-jährige Prognose der Bevölkerung getrennt nach Altersgruppe (Rieselfeld)

Quelle: Eigene Darstellung

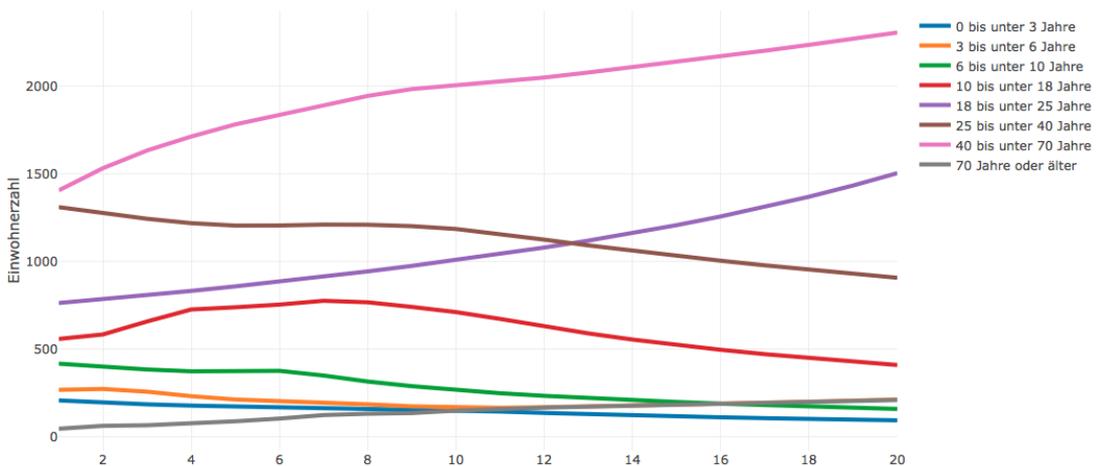


Abbildung 30: 20-jährige Prognose der Bevölkerung getrennt nach Altersgruppe (Vauban)

Quelle: Eigene Darstellung

Die Anpassung der Modelle zeigte auch bei der kleingliedrigen Einteilung der Einwohnerzahl-Zeitreihen nach Altersgruppe und Geschlecht sehr zufriedenstellende Ergebnisse mit R-Quadrat-Werten von 0.65 bis 1 und MAPE-Werten von 0.06% bis 15.13%, was bedeutet, dass die GAM-Modelle die Entwicklung der Bevölkerungsstruktur sehr gut prognostizieren können (siehe Tabelle 3).

Tabelle 3: Anpassungsgüte der GAM-Modelle nach Altersgruppe und Geschlecht

	Rieselfeld				Vauban			
	Weiblich		Männlich		Weiblich		Männlich	
	R ²	MAPE	R ²	MAPE	R ²	MAPE	R ²	MAPE
0-3 Jahre	0.95	5.14%	0.96	2.72%	0.85	3.88%	0.82	4.43%
3-6 Jahre	0.98	6.14%	0.97	5.08%	0.98	1.99%	0.97	1.45%
6-10 Jahre	0.99	1.77%	0.99	4.39%	0.99	1.03%	0.93	2.69%
10-18 Jahre	0.99	1.77%	0.99	2.23%	0.99	0.48%	0.95	1.22%
18-25 Jahre	0.99	3.01%	0.99	2.60%	0.83	2.79%	0.95	2.47%
25-40 Jahre	0.99	0.97%	0.99	2.07%	0.87	0.92%	0.65	2.09%
40-70 Jahre	1	1.52%	1	0.59%	1	0.06%	0.99	0.28%
70 Plus	0.99	15.13%	0.99	4.45%	0.96	6.25%	0.98	3.93%

Die Vorhersage für die in Zukunft zu erwartende Bevölkerungsentwicklung zeigten recht ähnliche Ergebnisse mit den getrennten Vorhersagen nach Geschlecht und nach Altersgruppe und erfüllten damit die geforderte Konsistenz (siehe Abb. 31 bis Abb.34).

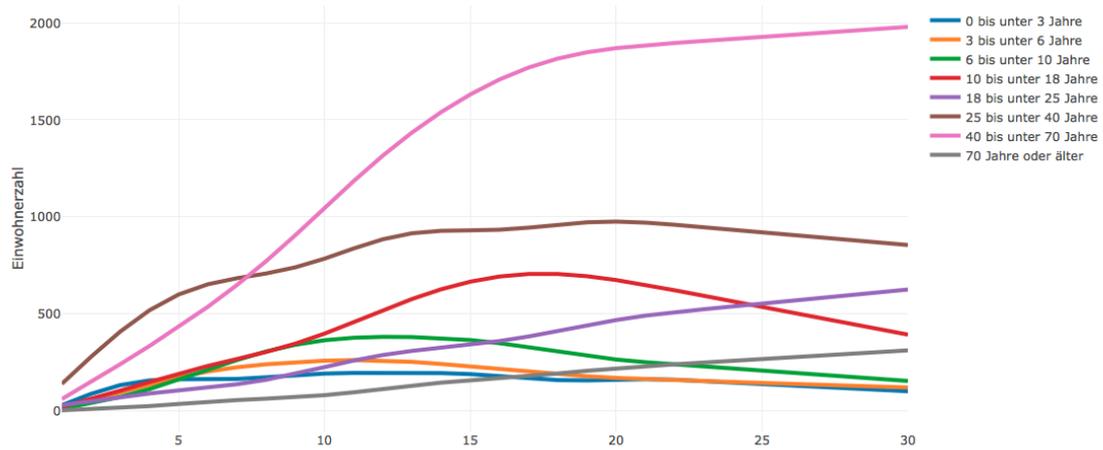


Abbildung 31: 30-jährige Prognose der männlichen Bevölkerung getrennt nach Altersgruppe (Rieselfeld)

Quelle: Eigene Darstellung

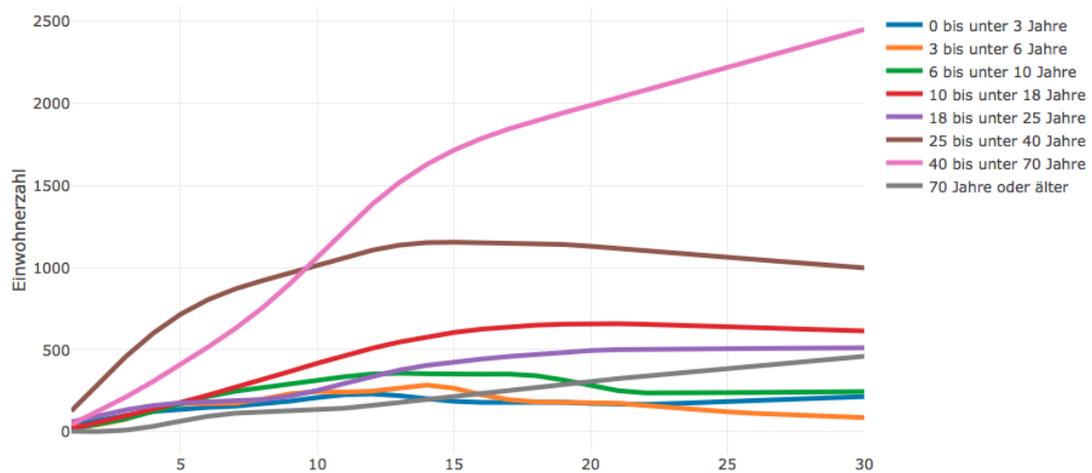


Abbildung 32: 30-jährige Prognose der weiblichen Bevölkerung getrennt nach Altersgruppe (Rieselfeld)

Quelle: Eigene Darstellung

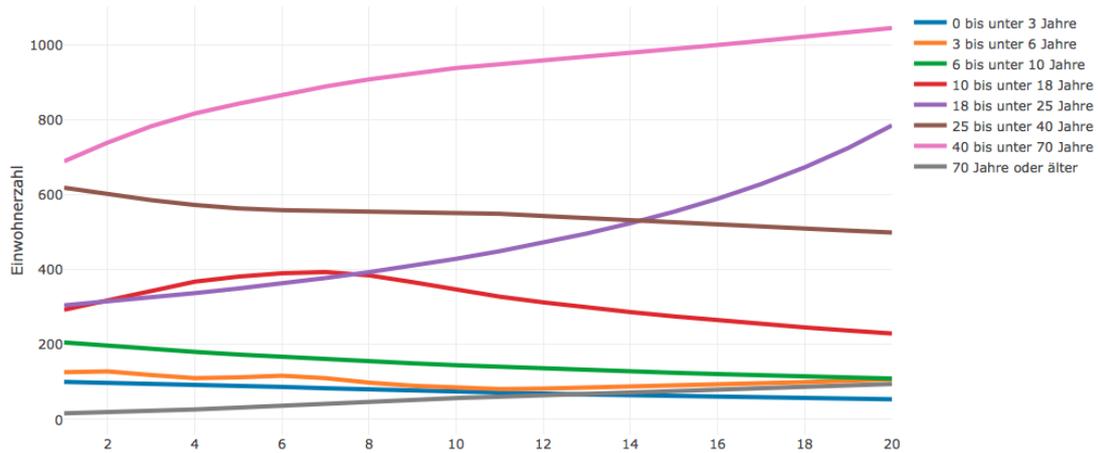


Abbildung 33: 20-jährige Prognose der männlichen Bevölkerung getrennt nach Altersgruppe (Vauban)

Quelle: Eigene Darstellung

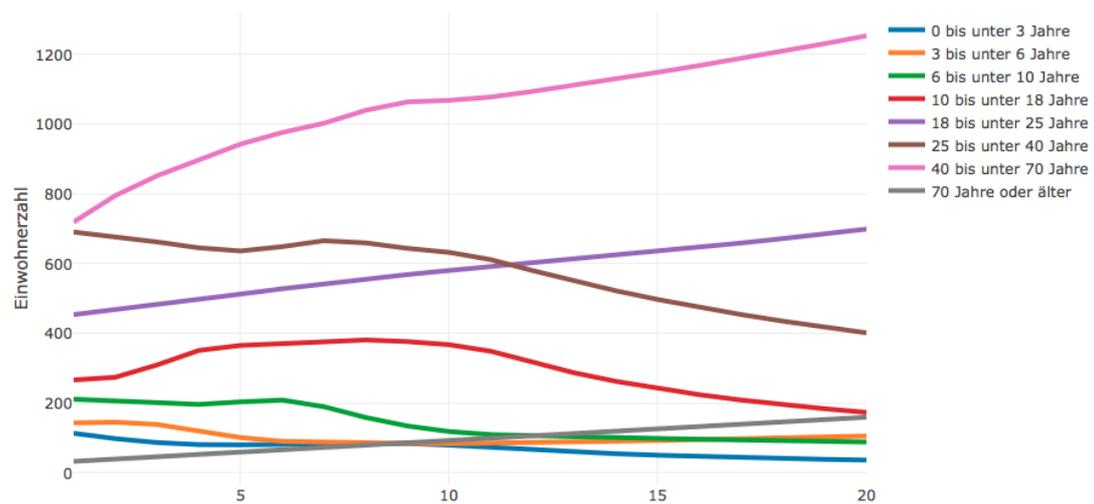


Abbildung 34: 20-jährige Prognose der weiblichen Bevölkerung getrennt nach Altersgruppe (Vauban)

Quelle: Eigene Darstellung

Mehrschichtige Feedforward Neuronale Netze

Trend-Modelle für beide Gebiete wurden auch mit Verfahren des maschinellen Lernens berechnet. Diese Modelle wurden unter Anwendung der Funktion *nnetar()* der R-Bibliothek *forecast* implementiert. Die *nnetar()* Funktion trainiert mehrschichtige Feedforward neuronale Netze mit einer einzigen versteckten Schicht und Lag-

Werten¹¹⁵ der Zeitreihe als Eingabeschicht. Für die Festlegung der geeigneten Netzarchitekturen und der Hyperparametern, wird das neuronale Netz zwanzigmal mit verschiedenen Startgewichten trainiert.

Für die Vorhersage der Entwicklung der Gesamtbevölkerung wurde als bestes Modell für beide Neubaugebiete ein Netz mit einem Neuron Lag1-Wertes in der Eingabeschicht und einem Neuron in Zwischenschicht ermittelt. Das bedeutet, dass jeder aktueller Wert einer Zeitreihe (X_t) in Bezug auf den letzten Vergangenheitswert von sich selbst (X_{t-1}) berechnet wird. Betrachtet man die Bestimmtheitsmaße und MAPE der angepassten Modelle, so ist zu sehen, dass die generelle Anpassung der Modelle sehr gut ist, da die Modelle sehr hohe Bestimmtheitsmaß-Werte (R-Quadrat-Riesel Feld = 0.99 und R-Quadrat-Vauban = 0.97) und sehr niedrige MAPE-Werte (Riesel Feld: 1.47%, Vauban: 0.97%) aufweisen. Diese Aussage bestätigen auch die Abbildungen 35 und 36.

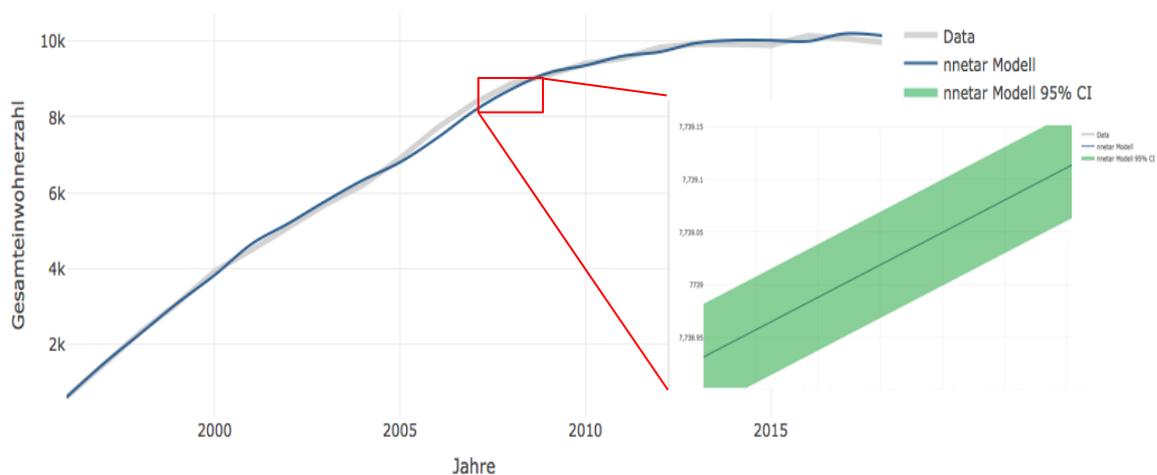


Abbildung 35: Rohdaten vs Vorhergesagte Werte für die Gesamtbevölkerung (Riesel Feld)

Quelle: Eigene Darstellung

¹¹⁵ Lag-Wert ist ein Begriff der Zeitreihenanalyse, der ursprünglich aus der englischsprachigen Literatur kommt und damit einen festen Wert gemeint, der die vergangenen Zeitpunkte der Zeitreihe beschreibt. Somit $t - 1$ entspricht "Lag 1", $t - 2$ "Lag 2", etc.

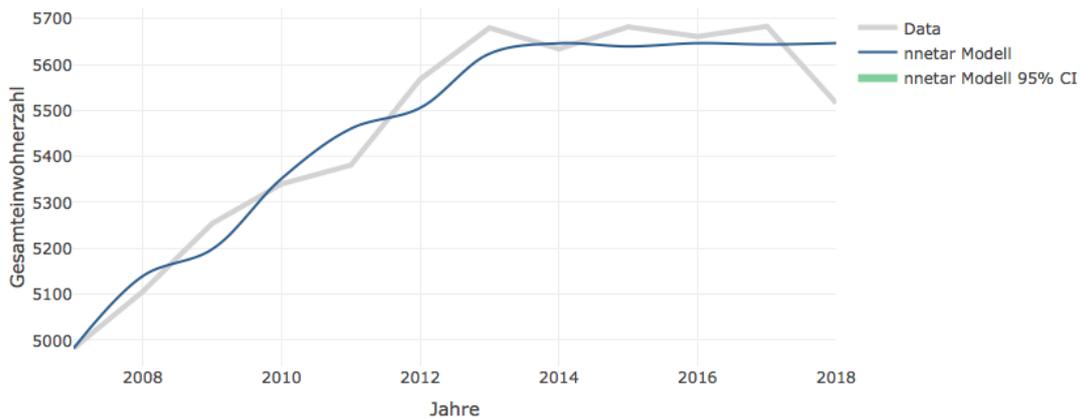


Abbildung 36: Rohdaten vs Vorhergesagte Werte für die Gesamtbevölkerung (Vauban)

Quelle: Eigene Darstellung

Die 30-jährige Prognose der KNN-Modelle für große Neubaugebiete ähnelt den Prognosen der GAM-Modelle. Ein Unterschied zeigt sich bei den kleinen/mittleren Neubaugebieten wo die KNN-Modelle nach den ersten 10 Jahren eine stabile Bevölkerungsentwicklung vorhersagen im Gegensatz zu den GAM-Modellen, die eine Senkung der Bevölkerung aufweisen. Somit ist die KNN-Prognose aus demografischer Sicht wahrscheinlicher (siehe Abb. 37 und 38).

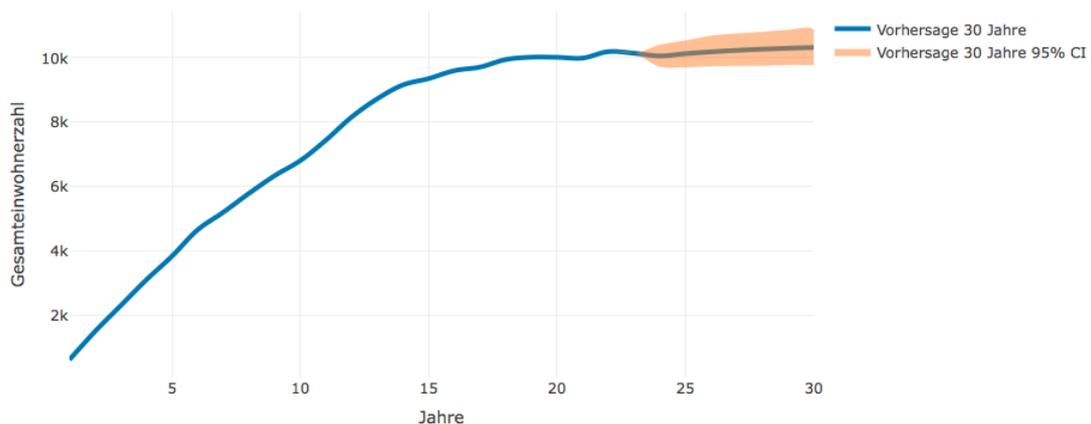


Abbildung 37: 30-jährige Prognose der Gesamtbevölkerung (Rieselfeld)

Quelle: Eigene Darstellung

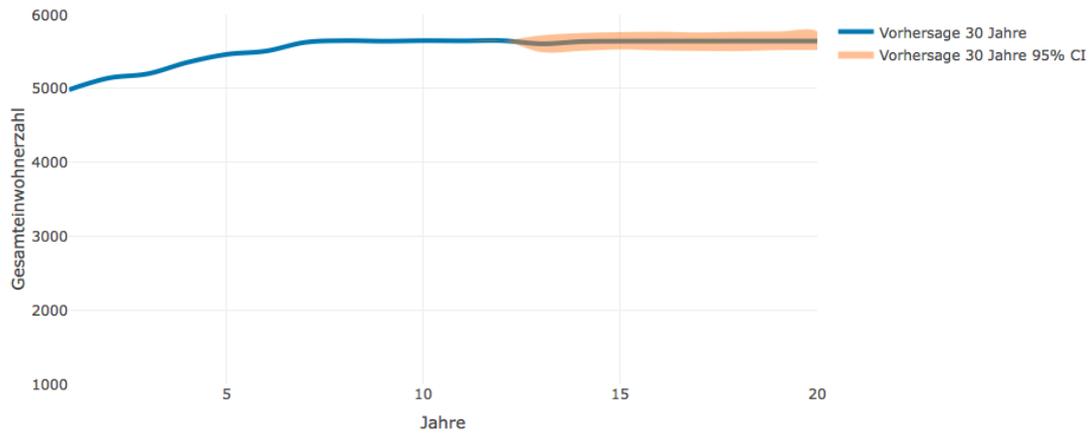


Abbildung 38: 20-jährige Prognose der Gesamtbevölkerung (Vauban)

Quelle: Eigene Darstellung

Die Tabelle 4 demonstriert, wie gut die KNN-Modelle an die Rohdaten der zwei Neubaugebiete angepasst wurden. Durch die R-Quadrat-Werte und der MAPE-Werte wird die Schlussfolgerung gezogen, dass KNN-Modelle sich auch sehr gut für die Vorhersage der Bevölkerungszahlen getrennt nach Geschlecht eignen. Das wird auch von den Abbildungen 39 bis 42 bestätigt.

Tabelle 4: Anpassungsgüte der KNN-Modelle nach Geschlecht

	Rieselfeld		Vauban	
	R-Quadrat	MAPE	R-Quadrat	MAPE
Männliche Einwohnerzahl	0.999	0.24 %	0.96	0.85 %
Weibliche Einwohnerzahl	0.999	1.52 %	0.96	0.51 %

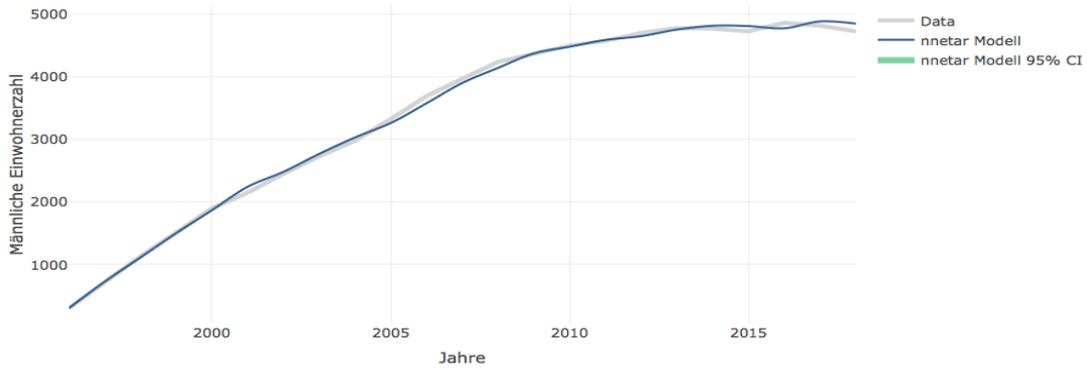


Abbildung 39: Rohdaten vs Vorhergesagte Werte für die männliche Bevölkerung (Rieselfeld)

Quelle: Eigene Darstellung



Abbildung 40: Rohdaten vs Vorhergesagte Werte für die weibliche Bevölkerung (Rieselfeld)

Quelle: Eigene Darstellung

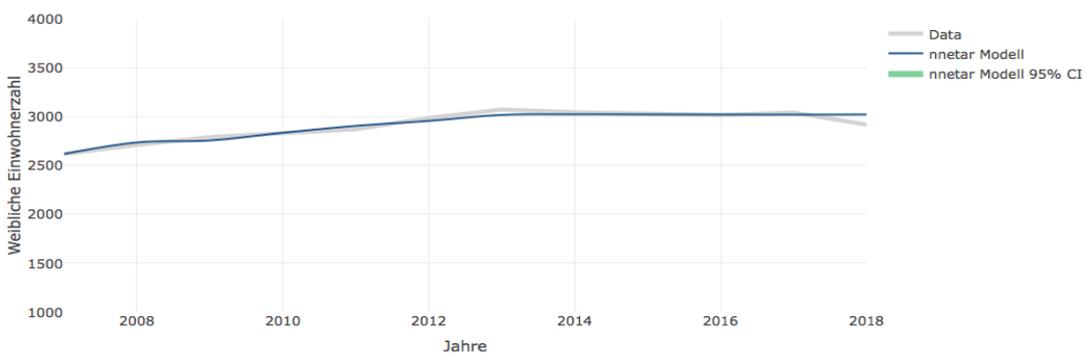


Abbildung 41: Rohdaten vs Vorhergesagte Werte für die männliche Bevölkerung (Vauban)

Quelle: Eigene Darstellung



Abbildung 42: Rohdaten vs Vorhergesagte Werte für die weibliche Bevölkerung (Vauban)

Quelle: Eigene Darstellung

Die erwartete Populationsentwicklung der männlichen und weiblichen Einwohner ist in den Abbildungen 43 und 44 zu sehen. In großen Neubaugebieten verlaufen männliche und weibliche Bevölkerungskurven ähnlich wie jene der GAM-Modelle. Dagegen prognostizieren die KNN-Modellen in kleinen/mittleren Neubaugebieten, dass die männliche Einwohnerzahl sich nach einer kurzen Zeit von sinkenden Werten wieder erholen wird.

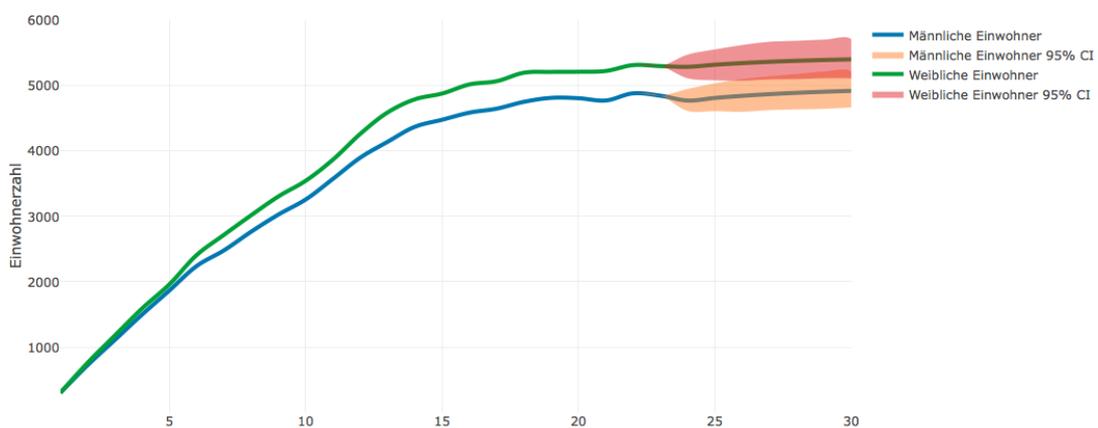


Abbildung 43: 30-jährige Prognose der Bevölkerung getrennt nach Geschlecht (Riesel-feld)

Quelle: Eigene Darstellung

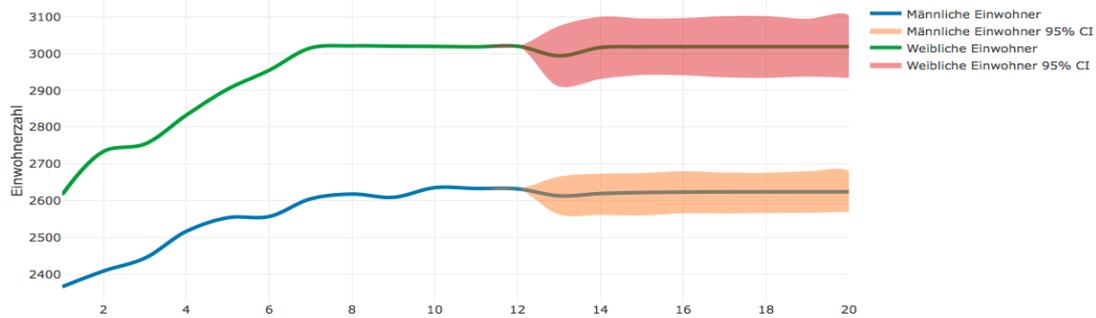


Abbildung 44: 20-jährige Prognose der Bevölkerung getrennt nach Geschlecht (Vauban)

Quelle: Eigene Darstellung

Auch die berechneten KNN-Modelle für die Entwicklung der Bevölkerung nach Altersgruppe zeigten für beide Gebiete ein sehr hohe Vorhersagegenauigkeit mit R-Quadrat-Werten von 1 bis 0.97 für Rieselfeld und 1 bis 0.77 für Vauban, sowie MAPE-Werten von 0.82% bis 8.58% und 0.27% bis 4.63%. Die Prognose anhand der angepassten Modelle für die zu erwartende Altersentwicklung der Bevölkerung ist in Abbildung 45 und 46 zu sehen.

Tabelle 5: Anpassungsgüte der KNN-Modelle nach Altersgruppe

	Rieselfeld		Vauban	
	R-Quadrat	MAPE	R-Quadrat	MAPE
0 bis unter 3 Jahre	0.97	4.32 %	0.89	4.63 %
3 bis unter 6 Jahre	0.97	7.78 %	0.998	0.65 %
6 bis unter 10 Jahre	0.98	6.10 %	0.97	3.88 %
10 bis unter 18 Jahre	0.998	2.30 %	0.89	4.31 %
18 bis unter 25 Jahre	0.997	3.44 %	0.96	2.99 %
25 bis unter 40 Jahre	0.997	1.99 %	0.77	2.35 %
40 bis unter 70 Jahre	1	0.82 %	1	0.27 %
70 Jahre oder älter	0.996	8.58 %	0.99	4.37 %

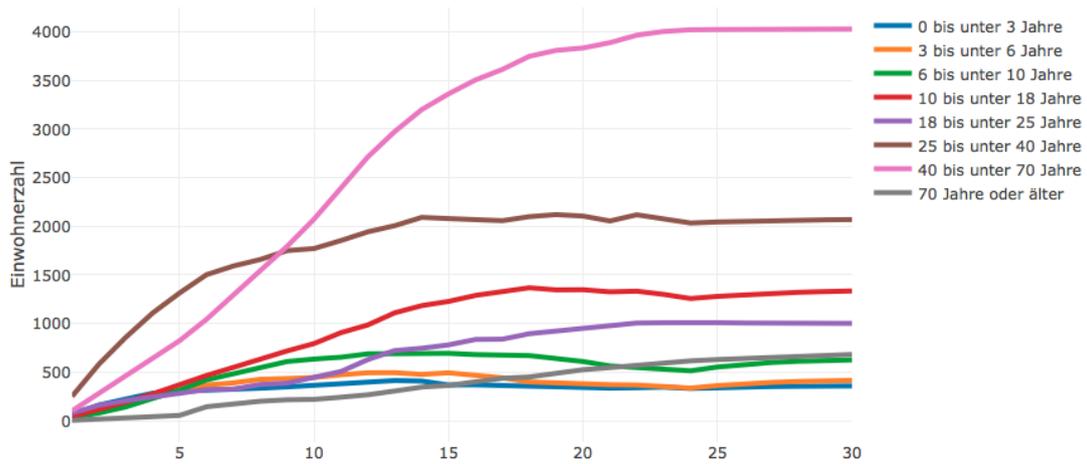


Abbildung 45: 30-jährige Prognose der Bevölkerung getrennt nach Altersgruppe (Rieselfeld)

Quelle: Eigene Darstellung

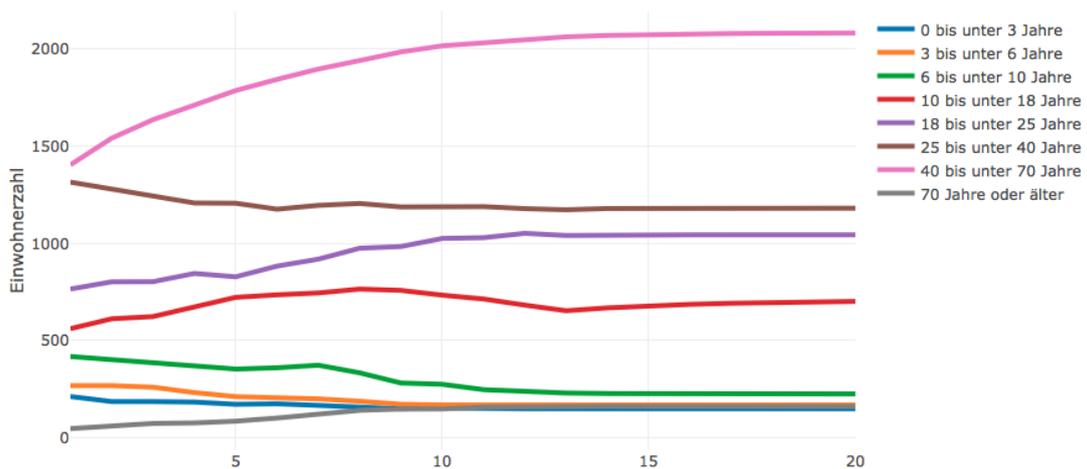


Abbildung 46: 20-jährige Prognose der Bevölkerung getrennt nach Altersgruppe (Vauban)

Quelle: Eigene Darstellung

Nach Betrachtung der Ergebnisse der R-Quadrat-Werte und der MAPE-Werte für Rieselfeld und Vauban wird festgestellt, dass auch im Fall der Modellierung der detaillierten demografischen Struktur die KNN-Modelle eine sehr hohe Anpassungsgüte haben (siehe Tabelle 6) und ihre Prognosen für die zukünftige Entwicklung den Erwartungen der grobgliebrigen Modelle entspricht (siehe Abb. 47 bis Abb. 50).

Tabelle 6: Anpassungsgüte der KNN-Modelle nach Altersgruppe und Geschlecht

	Rieselfeld				Vauban			
	Weiblich		Männlich		Weiblich		Männlich	
	R ²	MAPE	R ²	MAPE	R ²	MAPE	R ²	MAPE
0-3 Jahre	0.96	5.66%	0.95	5.70%	0.86	6.31%	0.93	3.93%
3-6 Jahre	0.97	7.53%	0.95	9.11%	0.99	3.11%	0.97	2.81%
6-10 Jahre	0.98	5.72%	0.97	7.29%	0.96	4.32%	0.94	4.02%
10-18 Jahre	0.999	1.80%	0.995	4.13%	0.92	3.40%	0.88	4.01%
18-25 Jahre	0.996	3.58%	0.995	4.37%	0.92	2.90%	0.96	3.38%
25-40 Jahre	0.997	1.82%	0.99	2.70%	0.73	2.87%	0.81	2.37%
40-70 Jahre	1	1.69%	1	0.89%	1	0.40%	1	0.36%
70 Plus	0.99	13.21%	0.998	5.27%	0.97	6.80%	0.98	6.32%

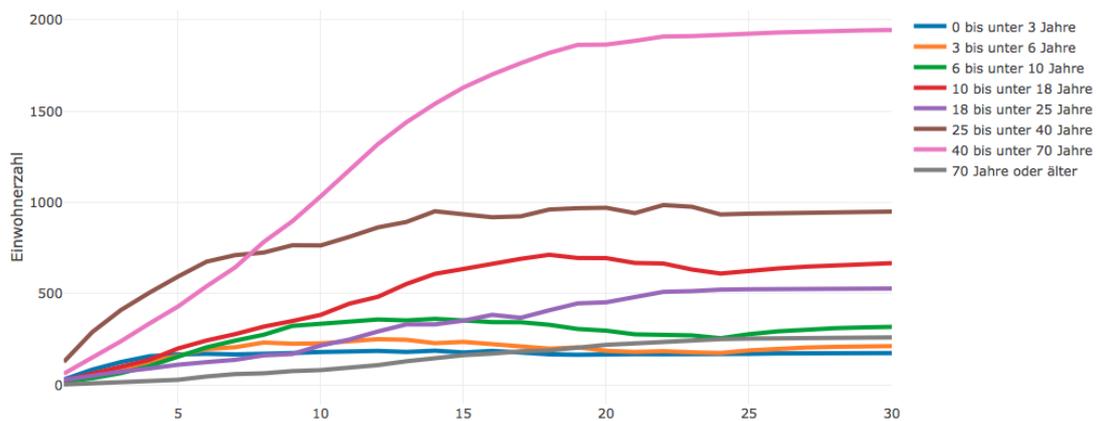


Abbildung 47: 30-jährige Prognose der männlichen Bevölkerung getrennt nach Altersgruppe (Rieselfeld)

Quelle: Eigene Darstellung

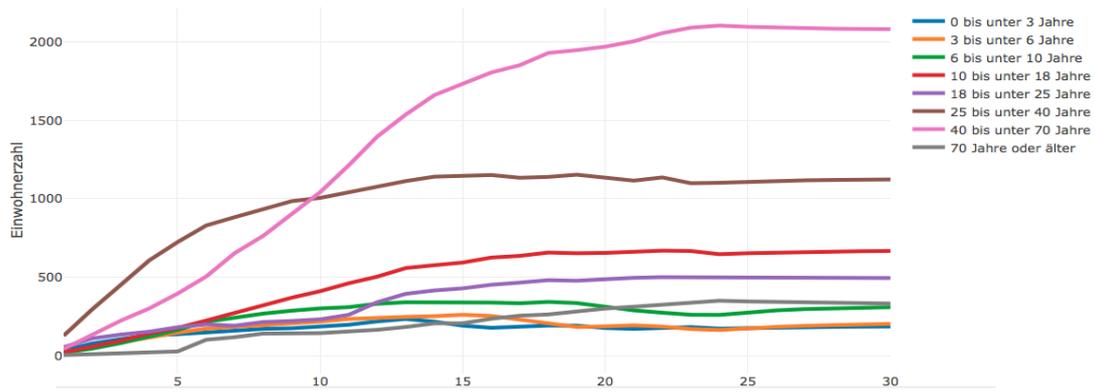


Abbildung 48: 30-jährige Prognose der weiblichen Bevölkerung getrennt nach Altersgruppe (Rieselfeld)

Quelle: Eigene Darstellung

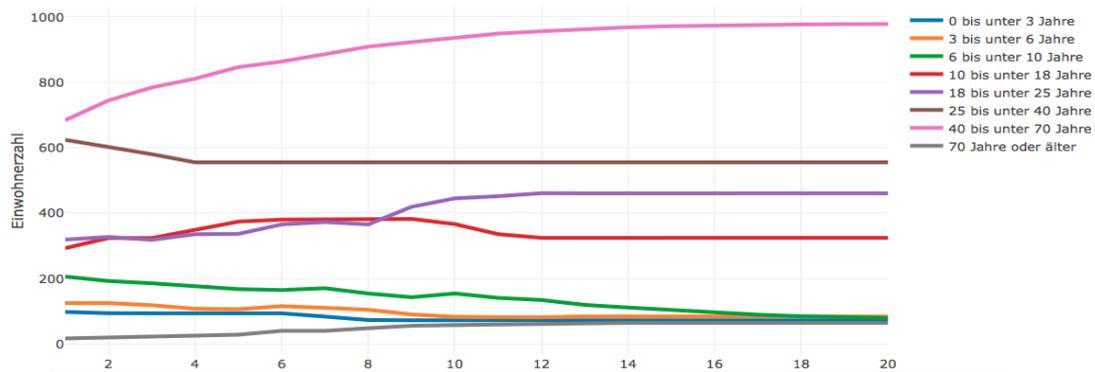


Abbildung 49: 20-jährige Prognose der männlichen Bevölkerung getrennt nach Altersgruppe (Vauban)

Quelle: Eigene Darstellung

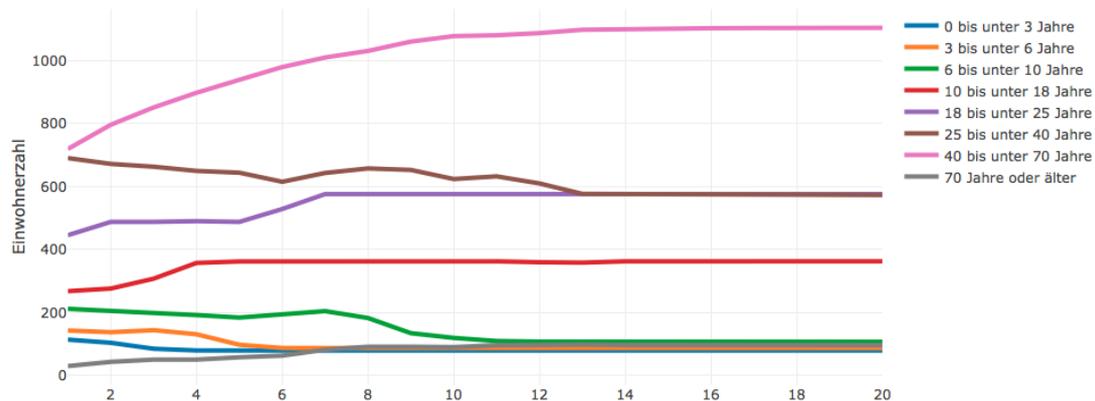


Abbildung 50: 20-jährige Prognose der weiblichen Bevölkerung getrennt nach Altersgruppe (Vauban)

Quelle: Eigene Darstellung

ARIMA-Modelle

Für die Implementierung der ARIMA-Modelle wurde die Funktion *auto.arima()* der R-Bibliothek *forecast* angewendet. Die *auto.arima()* Funktion verwendet eine Variation des Hyndman-Khandakar-Algorithmus, der Einheitswurzeltests und die Minimierung des Akaike Informationskriteriums (AIC) kombiniert, damit ein ARIMA-Modell mit der am besten angepassten (p, d, q) Ordnung ausgewählt wird.

Die ARIMA-Modelle weisen im Vergleich zu den anderen beiden Modellierungsansätzen bei der Entwicklung der Gesamtbevölkerung besonders für Vauban eine niedrigere Vorhersagegenauigkeit auf. Jedoch sind im Allgemeinen ihre Anpassungsergebnisse mit einem R-Quadrat-Wert von 0.99 für Rieselfeld und 0.93 für Vauban, wie auch mit MAPE-Werten von 1.70% und 1,37% zufriedenstellend. Dies verdeutlichen auch die zwei nächsten Grafiken (siehe Abb. 51 und Abb. 52).

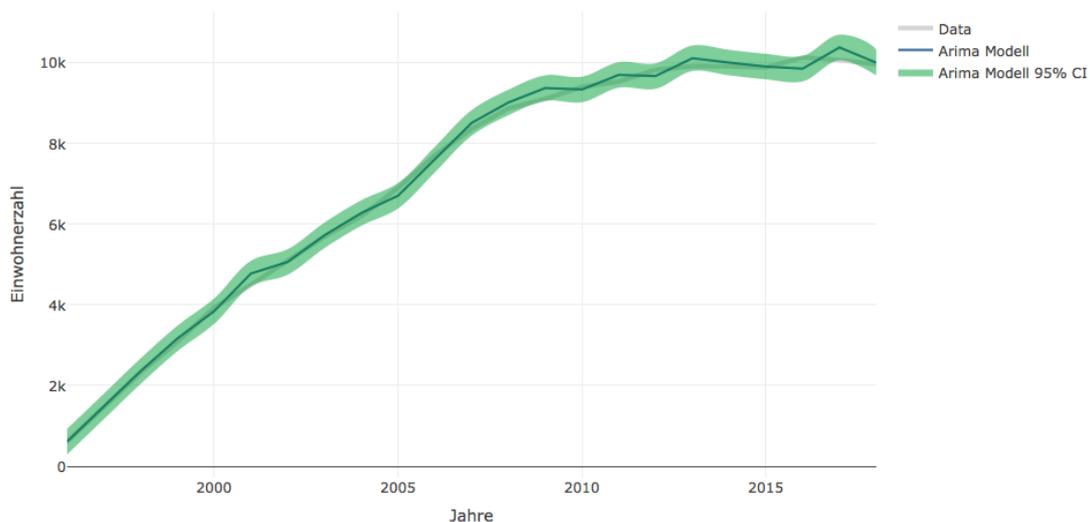


Abbildung 51: Rohdaten vs Vorhergesagte Werte für die Gesamtbevölkerung (Rieselfeld)

Quelle: Eigene Darstellung

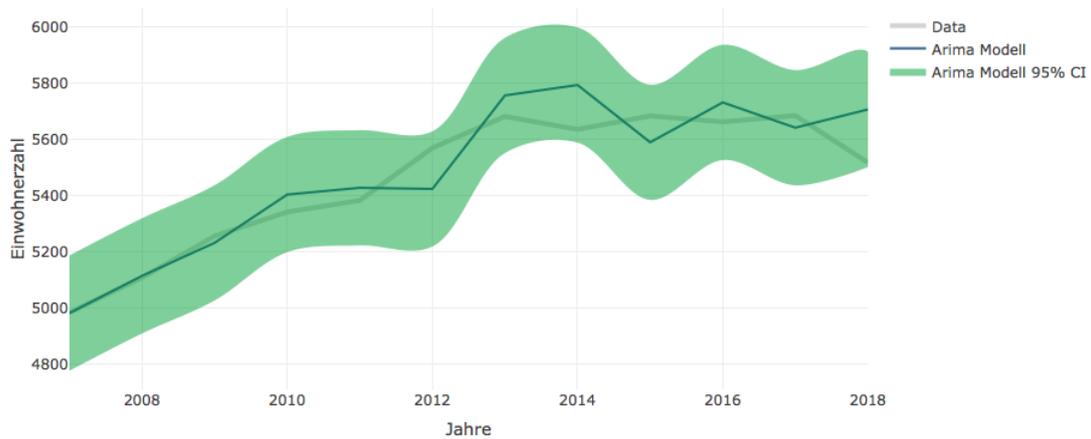


Abbildung 52: Rohdaten vs Vorhergesagte Werte für die Gesamtbevölkerung (Vauban)

Quelle: Eigene Darstellung

Die Abbildungen 53 und 54 veranschaulichen die Gesamtbevölkerungsprognose in den zwei Neubaugebietsarten. Hier ist deutlich die Problematik dieses Modellierungsansatzes zu erkennen, der die Trendverläufe linear anpasst. Die große Breite der Konfidenzintervalle besonders zu späteren Zeitpunkten drückt eine niedrige Präzision der langfristigen Schätzung aus.

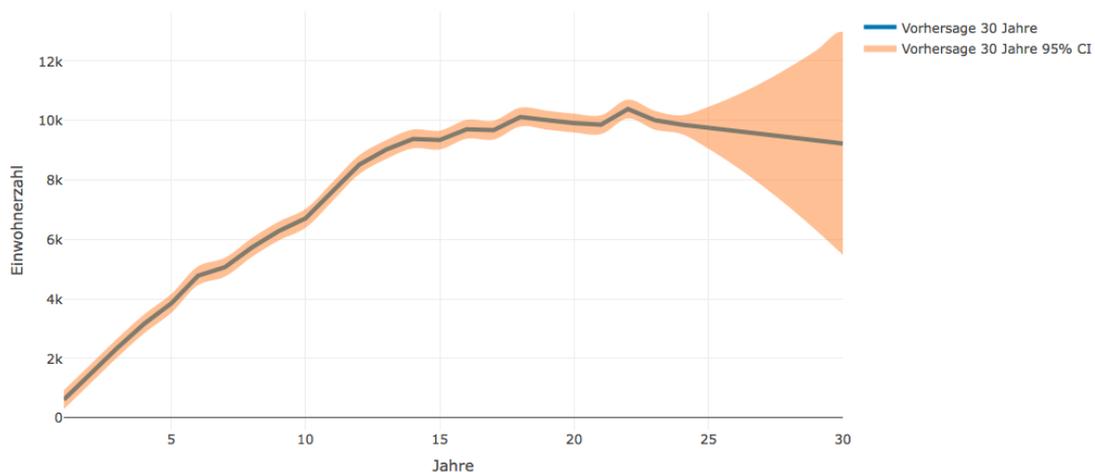


Abbildung 53: 30-jährige Prognose der Gesamtbevölkerung (Rieselfeld)

Quelle: Eigene Darstellung

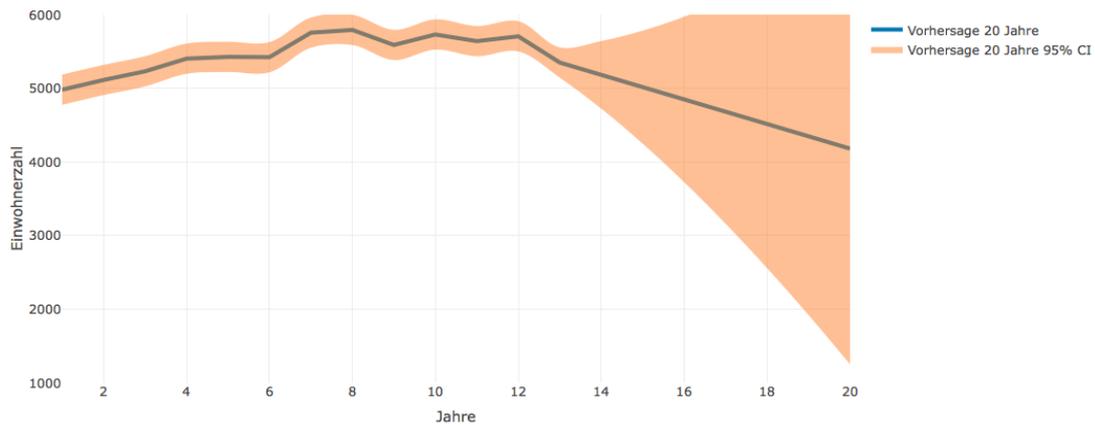


Abbildung 54: 20-jährige Prognose der Gesamtbevölkerung (Vauban)

Quelle: Eigene Darstellung

Bei der Prognostizierung der Einwohnerzahl getrennt nach Geschlecht zeigen die ARIMA-Modelle eine gute Anpassung an die Rohdaten (siehe Tabelle 7). Im Vergleich mit den GAM- und KNN-Modellen, die besser für die Modellierung von nichtlinearen Beziehungen geeignet sind, schneiden die ARIMA-Modelle schlechter ab. Das verdeutlichen die Visualisierungen der Rohdaten im Vergleich mit den vorhergesagten Werten des Modells (siehe Abb. 55 bis Abb. 58).

Tabelle 7: Anpassungsgüte der ARIMA-Modelle nach Geschlecht

	Rieselfeld		Vauban	
	R-Quadrat	MAPE	R-Quadrat	MAPE
Männliche Einwohnerzahl	0.99	1.60 %	0.95	1.13 %
Weibliche Einwohnerzahl	0.99	1.85 %	0.92	1.37 %



Abbildung 55: Rohdaten vs Vorhergesagte Werte für die weibliche Bevölkerung (Riesel-feld)

Quelle: Eigene Darstellung

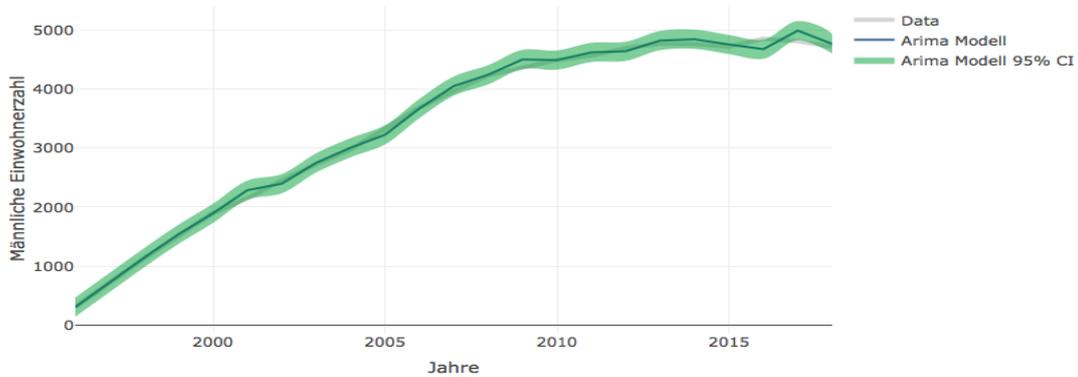


Abbildung 56: Rohdaten vs Vorhergesagte Werte für die männliche Bevölkerung (Riesel-feld)

Quelle: Eigene Darstellung

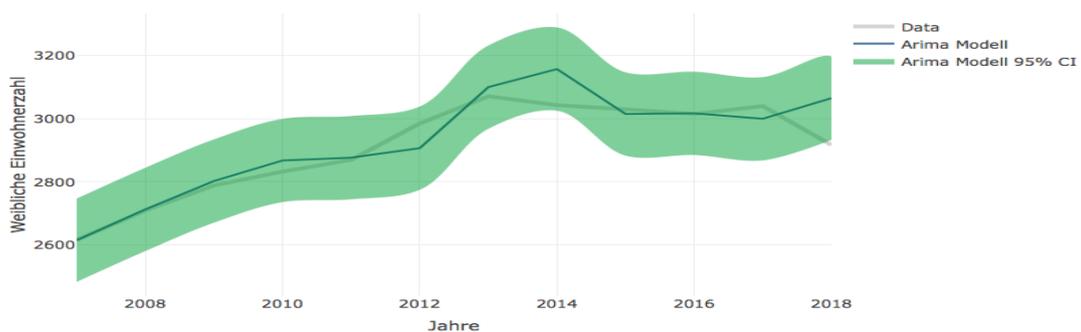


Abbildung 57: Rohdaten vs Vorhergesagte Werte für die weibliche Bevölkerung (Vau-ban)

Quelle: Eigene Darstellung

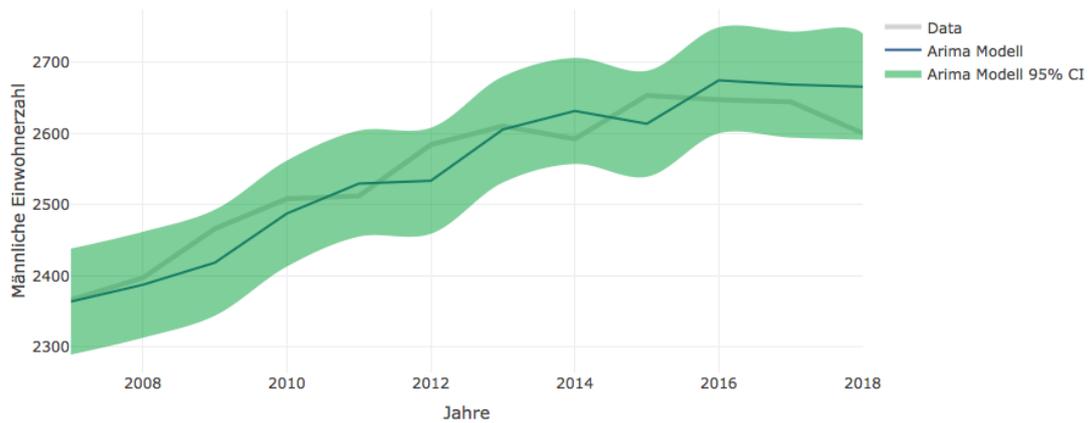


Abbildung 58: Rohdaten vs Vorhergesagte Werte für die männliche Bevölkerung (Vauban)

Quelle: Eigene Darstellung

Wie bei der Vorhersage der Bevölkerungsentwicklung angesichts der Gesamteinwohnerzahl ist auch bei den Bevölkerungsprognosen getrennt nach Altersgruppe eine größere Unsicherheit der Schätzungen als mit den anderen zwei Methoden zu beobachten (siehe Abb. 59 und Abb. 60).

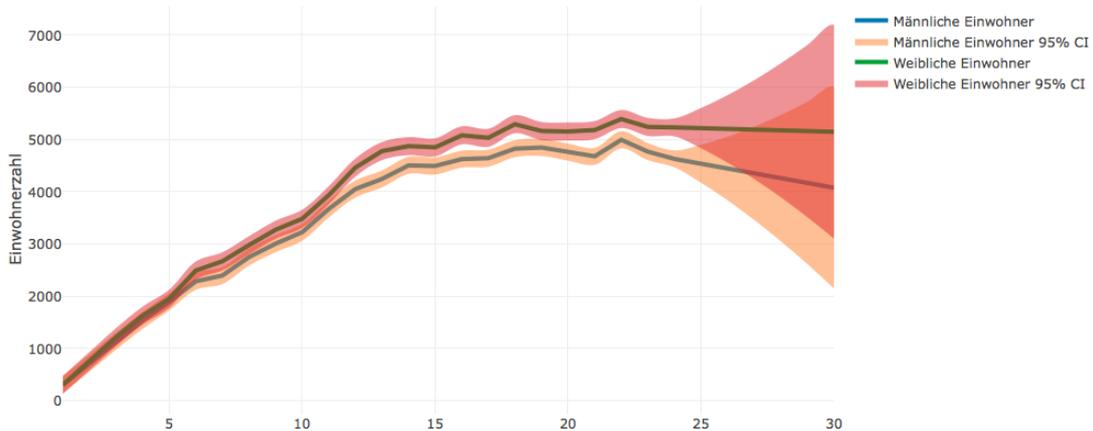


Abbildung 59: 30-jährige Prognose der Bevölkerung getrennt nach Geschlecht (Riesel)

Quelle: Eigene Darstellung

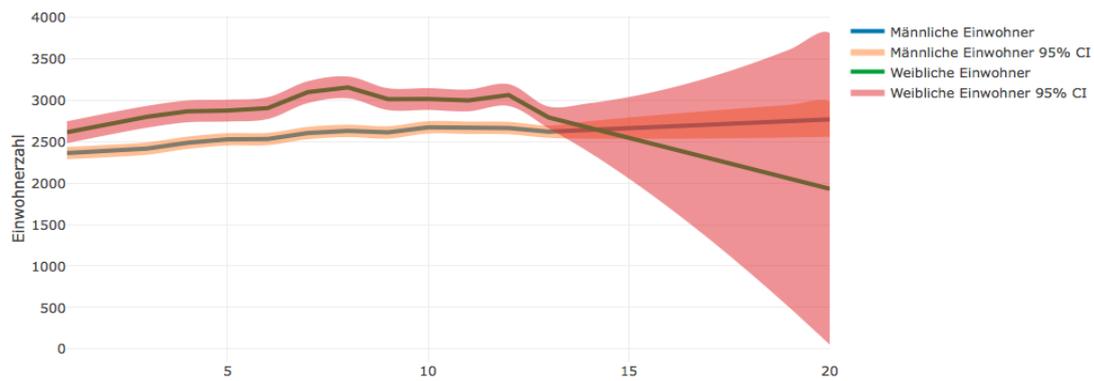


Abbildung 60: 20-jährige Prognose der Bevölkerung getrennt nach Geschlecht (Vauban)

Quelle: Eigene Darstellung

Ähnliche Schlussfolgerungen sind auch über die Vorhersagegenauigkeit der ARIMA-Modelle bei der Betrachtung der Einwohnerzahl getrennt nach Altersgruppe und nach Altersgruppe und Geschlecht zu ziehen. Die Anpassungsgüte der Modelle ist zwar schlechter als bei den nichtlinearen Modellen aber trotzdem ausreichend (siehe Tabelle 8 und 9).

Tabelle 8: Anpassungsgüte der ARIMA-Modelle nach Altersgruppe

	Rieselfeld		Vauban	
	R-Quadrat	MAPE	R-Quadrat	MAPE
0 bis unter 3 Jahre	0.96	4.61 %	0.87	5.32 %
3 bis unter 6 Jahre	0.98	7.87 %	0.96	3.86 %
6 bis unter 10 Jahre	0.99	4.18 %	0.96	3.50 %
10 bis unter 18 Jahre	0.99	2.11 %	0.95	2.40 %
18 bis unter 25 Jahre	0.99	4.45 %	0.93	3.24 %
25 bis unter 40 Jahre	0.99	0.70 %	0.70	2.87 %
40 bis unter 70 Jahre	0.99	1.29 %	0.99	0.62 %
70 Jahre oder älter	0.99	13.50 %	0.99	4.62 %

Tabelle 9: Anpassungsgüte der ARIMA-Modelle nach Altersgruppe und Geschlecht

	Rieselfeld				Vauban			
	Weiblich		Männlich		Weiblich		Männlich	
	R ²	MAPE	R ²	MAPE	R ²	MAPE	R ²	MAPE
0-3 Jahre	0.97	5.22%	0.92	9.78%	0.77	9.54%	0.84	6.62%
3-6 Jahre	0.98	5.11%	0.96	10.68%	0.93	7.11%	0.90	6.49%
6-10 Jahre	0.99	4.62%	0.99	5.88%	0.95	5.87%	0.93	3.83%
10-18 Jahre	0.99	2.52%	0.99	3.17%	0.93	3.37%	0.90	3.13%
18-25 Jahre	0.99	5.52%	0.99	4.93%	0.90	3.57%	0.95	3.67%
25-40 Jahre	0.99	2.52%	0.98	4.36%	0.72	3.12%	0.62	3.59%
40-70 Jahre	0.99	2.42%	0.99	1.45%	0.99	0.95%	0.99	0.84%
70 Plus	0.99	21.65%	0.99	7.89%	0.96	7.73%	0.98	7.54%

Abschließend demonstrieren die Abbildungen 61 und 62 die zu erwartende Einwohnerzahl getrennt nach Altersgruppe für große und kleine/mittlere Neubaugebiete. Hier ist wieder deutlich zu erkennen, dass bei den langfristigen Prognosen ein sehr starker linearer Trend geschätzt wurde. Gleiche Erkenntnisse wurden auch von der Prognostizierung der Entwicklung der demografischen Struktur gewonnen (siehe Abb. 63 bis Abb. 66).

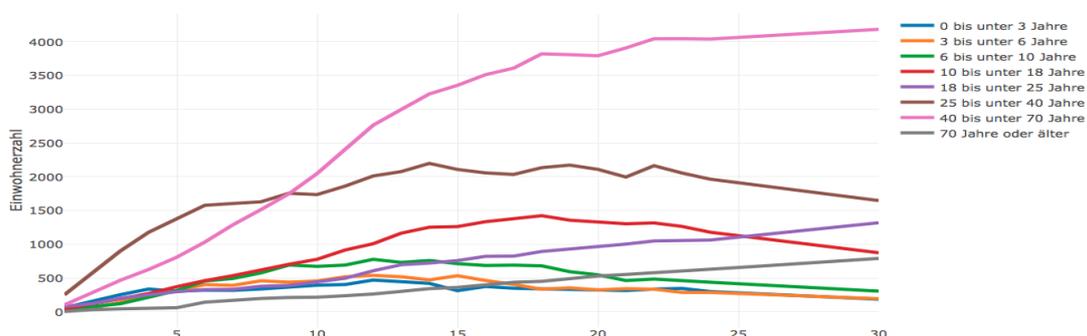


Abbildung 61: 30-jährige Prognose der Bevölkerung getrennt nach Altersgruppe (Rieselfeld)

Quelle: Eigene Darstellung

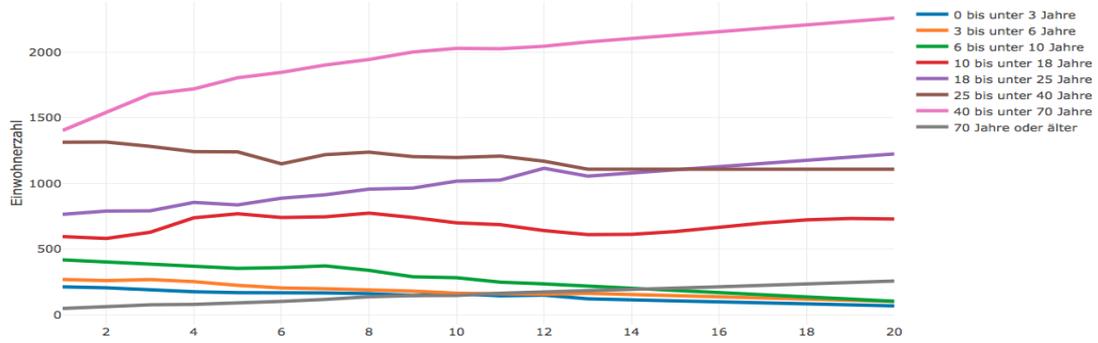


Abbildung 62: 20-jährige Prognose der Bevölkerung getrennt nach Altersgruppe (Vau-ban)

Quelle: Eigene Darstellung

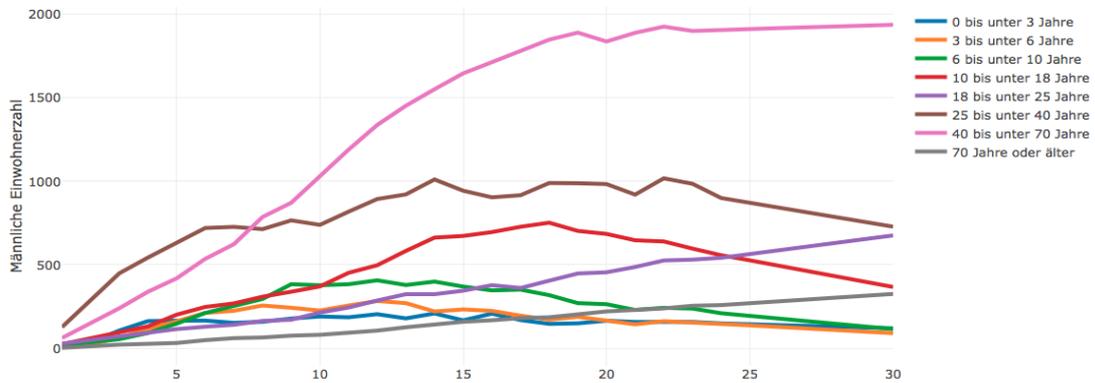


Abbildung 63: 30-jährige Prognose der männlichen Bevölkerung getrennt nach Altersgruppe (Rieselfeld)

Quelle: Eigene Darstellung

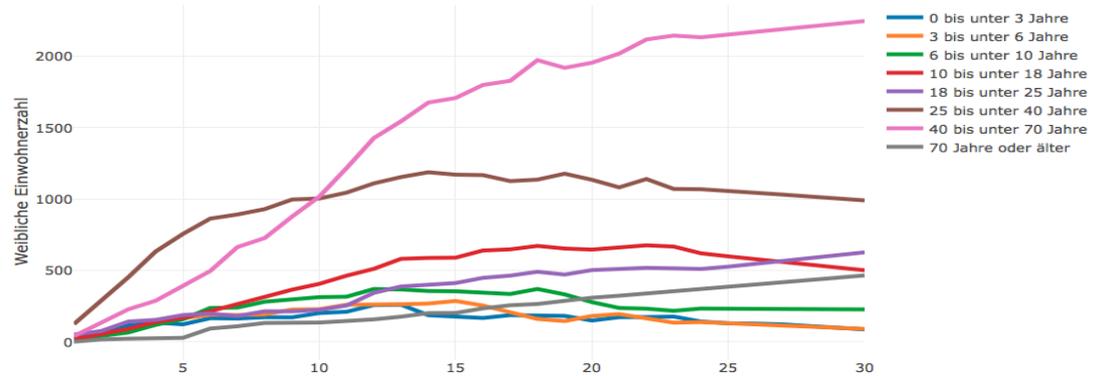


Abbildung 64: 30-jährige Prognose der weiblichen Bevölkerung getrennt nach Altersgruppe (Rieselfeld)

Quelle: Eigene Darstellung

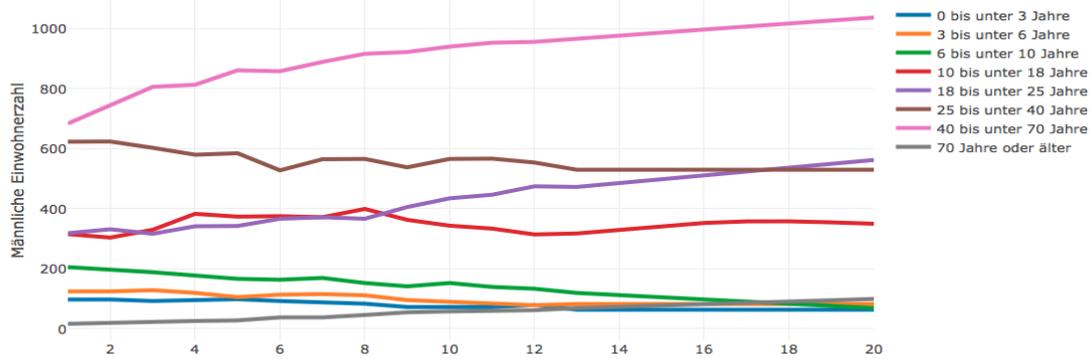


Abbildung 65: 20-jährige Prognose der männlichen Bevölkerung getrennt nach Altersgruppe (Vauban)

Quelle: Eigene Darstellung

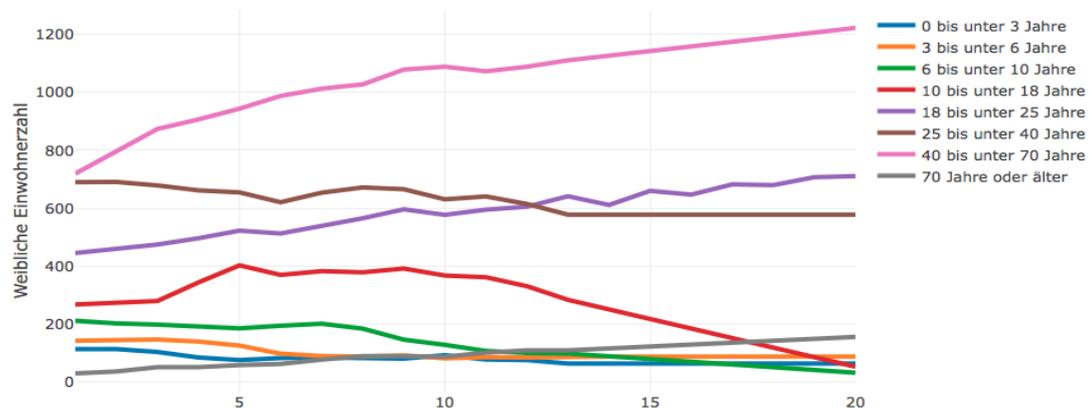


Abbildung 66: 20-jährige Prognose der weiblichen Bevölkerung getrennt nach Altersgruppe (Vauban)

Quelle: Eigene Darstellung

Kohorten-Komponenten-Modelle

Da sich die Kohorten-Komponenten-Modelle durch einfache mathematische Berechnungen verwirklichen lassen, wurden sie mit Anwendung eines Excel-Tools¹¹⁶ implementiert.

Dieses Tool enthält folgende Arbeitsblätter und Berechnungen:

- **Rohdaten:** Hier wurden die fünf-jährigen Alter-Geschlecht spezifischen Kohortendaten für die letzten drei fünf-jährigen Zeitperioden 2008, 2013 und 2018 eingetragen. Auf diese Daten werden die Datenblätter über Migration, Geburten und Projektionen zuzugreifen.

¹¹⁶ https://github.com/AppliedDemogToolbox/Chapin_CohortComponent

- **Projektionen:** Dieses Arbeitsblatt fasst das Modell zusammen, sobald alle anderen Arbeitsblätter fertig sind. Beobachtete Daten und Prognosen für jede der zwei fünfjährigen Projektionsperioden (2023 und 2028) werden sowohl nach Geschlecht als auch als Summe für die Neubaugebiete angezeigt.
- **2018-2028:** In diesem Arbeitsblatt findet durch die Übernahme der berechneten Migrationsraten-, Fertilitätsraten-, Überlebensraten- und Rohdaten, die tatsächliche Durchführung der Prognosen statt.
- **Überlebensraten:** Hier werden die spezifischen Überlebensraten der männlichen und weiblichen Kohorte zusammengefasst.
- **Fertilitätsraten:** Dieses Arbeitsblatt berechnet die Fertilitätsraten durch die eingebene Zahl der lebend geborenen Kinder je 1.000 Frauen im gebärfähigen Alter von 15 bis unter 45 Jahren innerhalb eines Jahres.
- **Neugeborene '08-'13, '13-'18:** Hier berechnet das Modell die Anzahl der Geburten für die Jahre 2008 bis 2018 anhand der berechneten Fertilitätsraten.
- **Migrationsraten:** Für die Verwendung in den Prognosen werden geschlechtsspezifische Migrationsraten berechnet, die über den gesamten Zeitraum als konstant angenommen werden. Da das Modell die beiden vorhergehenden Fünfjahresperioden verwendet, werden die Migrationsraten für jede der beiden vorhergehenden Fünfjahresperioden gemittelt, um altersspezifische durchschnittliche fünfjährige Migrationsraten zu erhalten.
- **W und M Überlebensraten:** Diese beiden Arbeitsblätter berechnen die Überlebensraten für die weibliche und männliche Kohorte durch die Eingabe der Stationären Population (Überlebende Population des Vorjahres minus Gestorbene des aktuellen Jahres) der einjährigen Kohorten.

Die Abbildungen 67 und 68 zeigen die zu erwartende Bevölkerung angesichts der Gesamteinwohnerzahl und der Einwohnerzahl nach Geschlecht für beide Neubaugebiete. Wie zu erkennen ist, wurde ein ständiges Wachstum der Bevölkerung vorhergesagt, dies ist zu erklären durch die in vorherigen Kapiteln beschriebene Problematik der Kohorten-Komponenten-Methoden: die einfache Fortschreibung der Bevölkerung von Basisjahr zum Prognostizierungsjahr und die Notwendigkeit subjektive Annahmen für die Entwicklung der Fertilitäts- und Migrationsraten zu treffen.

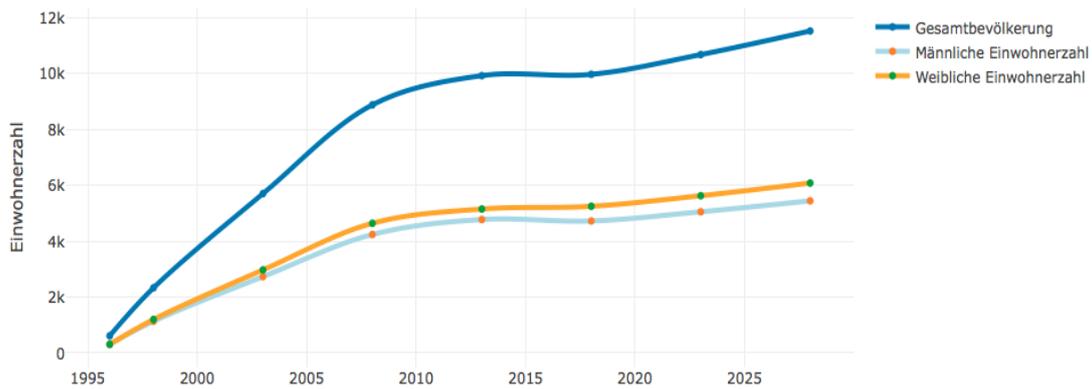


Abbildung 67: Vorhersage der Gesamtbevölkerung und getrennt nach Geschlecht in Fünfjahresperioden (Rieselfeld)

Quelle: Eigene Darstellung

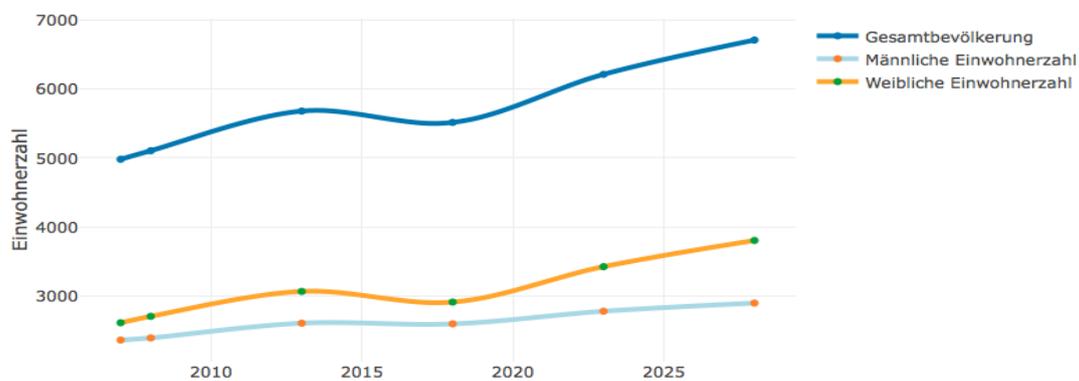


Abbildung 68: Vorhersage der Gesamtbevölkerung und getrennt nach Geschlecht in Fünfjahresperioden (Vauban)

Quelle: Eigene Darstellung

5.2 Implementierung der Webapplikation

Da für die Realisierung der Modellierungsansätze die Programmiersprache R verwendet wurde, wurde auch für die Entwicklung des Bevölkerungsprognose-Tools als Webapplikation ein in R-basiertes Framework verwendet. R-Shiny ist eine leistungsstarke Software-Entwicklungsplattform, die es Entwicklern ermöglicht, flexible, erweiterbare und interaktive Webanwendungen vollständig in R und JavaScript zu schreiben und für leistungsfähige moderne Browser zur Verfügung zu stellen¹¹⁷.

¹¹⁷ Chang et al. 2015

Somit wurde in dieser Arbeit mit R-Shiny die webbasierte grafische Benutzeroberfläche des Bevölkerungsprognose-Tools entwickelt, um die Bevölkerungsanalysen und ihre Visualisierungen mittels Schnittstellen veröffentlichen zu können.

Das Bevölkerungsprognose-Tool als Shiny-Anwendung besteht aus zwei Hauptkomponenten: einem Server, der festlegt, wie die Daten verarbeitet, analysiert und visualisiert werden, und einer Benutzeroberfläche, die das Gesamtbild der Shiny-Anwendung betrifft. Hierbei beinhaltet die Shiny-Anwendung Servercode in einem R-Skript mit der Name *server.R* und Benutzeroberflächencode in einem separaten Skript Namens *ui.R*.

Die Implementierung der Webapplikation in Form eines Dashboards wurde mithilfe der R-Bibliotheken *shinydashboard* und *shinythemes* realisiert. In *ui.R* wurde mithilfe der Funktionen wie *dashboardPage()*, *dashboardSidebar()*, *sidebarMenu()*, *dashboardBody()*, *tabsetPanel()*, *tabPanel()*, *textInput()*, und *radioButtons()* die Benutzeroberfläche gestaltet und modelliert. Das Skript *server.R* enthält die Anweisungen, die der Computer zum Erstellen einer Anwendung benötigt. In *server.R* werden die Funktionen wie *renderPlotly()*, *renderValueBox()*, *renderDataTable()* und *reactive()* für die Anwendungsentwicklung verwendet.

Auf der Startseite des Dashboards werden Informationen über das Tool, aktuelle demografische Daten und die aktuelle Bevölkerungspyramide der Stadt Freiburg angezeigt (siehe Abb. 69). Um die zu erwartende Entwicklung der Bevölkerung hinsichtlich der Einwohnerzahl und der demographischen Struktur (Altersgruppen und Geschlecht) anzeigen zu lassen, muss der Benutzer im Sidebar-Menü „*Baugebiet definieren*“ das Erstbezugsjahr eingeben und die Größe des geplanten Neubaugebiets auswählen (siehe Abb. 70). Im ausklappbaren Sidebar-Menü „*Vorhersage (Datenbasis Stadt Freiburg)*“ kann sich der Benutzer die Ergebnisse und einen Vergleich der verschiedenen Modellarten nach Gesamtbevölkerung, Geschlecht, Altersgruppe und Geschlecht/Altersgruppe (als Bevölkerungspyramide) in Form von interaktiven Grafiken anzeigen lassen (siehe Abb. 71 und Abb. 72). Abschließend gibt das Tool dem Benutzer die Möglichkeit, individuelle fallbezogene Prognosen zu bekommen, indem er durch

den Menüpunkt „Andere Stadt definieren“ seine eigenen Daten von ehemaligen Neubaugebieten hochladen (die Form und Aufbau der Datei ist mittels einer Beispieltabelle erklärt) und ähnlich wie bei den schon berechneten Arten von Neubaugebieten mit dem Menü „Vorhersage (Datenbasis eigene Daten)“ alle Modelle in Echtzeit trainieren und vergleichen lassen kann (siehe Abb. 73).



Abbildung 69: Startseite der Webapplikation

Quelle: Eigene Darstellung

BevölkerungsPrognose-Tool

Definition des neuen Baugebietes

Hier können Sie Angaben zu Ihrem geplanten Baugebiet machen, um eine korrekte Bevölkerungsvorhersage zu erhalten. Dafür müssen Sie folgendes tun:

- Zuerst wählen Sie das Jahr, ab dem das Baugebiet bewohnt werden soll (Erstbezugsjahr).
- Anschließend wählen Sie die geplante Größe des neuen Baugebietes.
- Die Auswertung erhalten Sie, indem Sie unter dem Menüpunkt 'Vorhersage (Datenbasis Stadt Freiburg)' eine beliebige Methode wählen.

Geben Sie bitte den Erstbezugsjahr

Wählen Sie die Größe des neuen Baugebietes

- Große Baugebiet (bis 12.000 Einwohner)
- Kleine/Mittlere Baugebiet (bis 6.000 Einwohner)

Abbildung 70: Auswahl der gewünschten Neubaugebietsart

Quelle: Eigene Darstellung

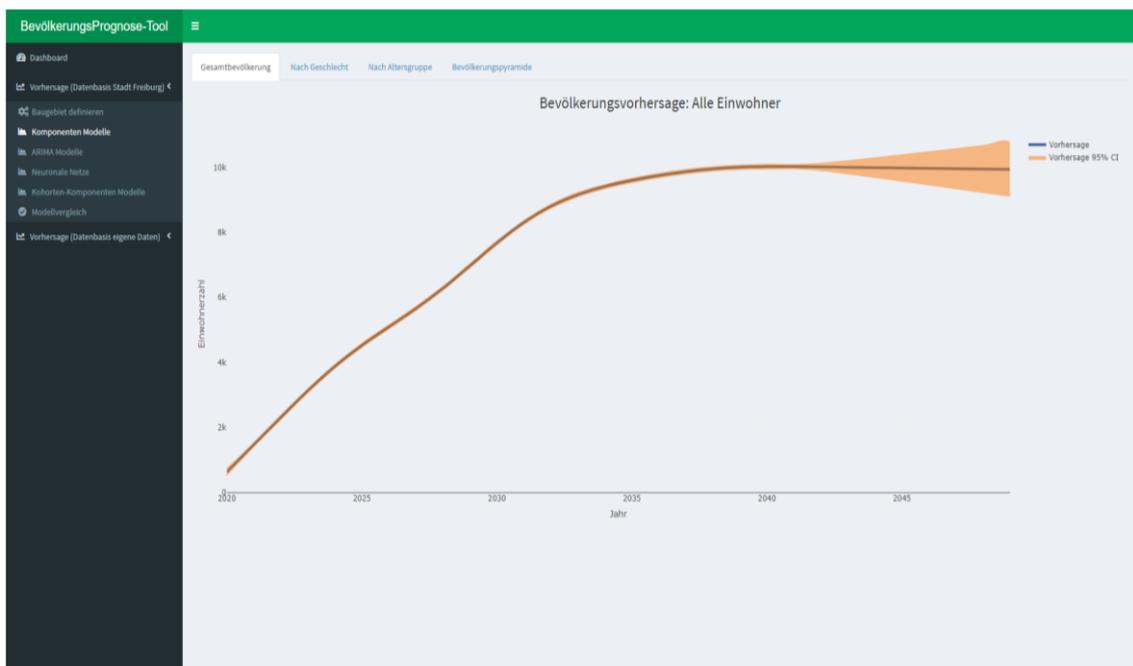


Abbildung 71: Darstellung der Bevölkerungsprognosen

Quelle: Eigene Darstellung



Abbildung 72: Darstellung des Modellvergleichs

Quelle: Eigene Darstellung

Year	Gender	Age Group	Population
1998	maennlich	0 bis unter 3 Jahre	11
1998	maennlich	0 bis unter 3 Jahre	1
1998	weiblich	0 bis unter 3 Jahre	7
1998	weiblich	0 bis unter 3 Jahre	1
1998	weiblich	0 bis unter 3 Jahre	1
1998	maennlich	0 bis unter 3 Jahre	7
1998	maennlich	0 bis unter 3 Jahre	2
1998	weiblich	0 bis unter 3 Jahre	11
1998	weiblich	0 bis unter 3 Jahre	2
1998	maennlich	0 bis unter 3 Jahre	14

Abbildung 73: Möglichkeit eigene Daten zu laden und analysieren

Quelle: Eigene Darstellung

Das Bevölkerungs-Prognose-Tool ist eine plattformübergreifende Anwendung und kann lokal von jedem Computer aus gestartet werden, auf dem R installiert ist. Außerdem kann die Webapplikation ohne R installiert zu haben gestartet werden, da sie auf der R-Cloud des shinyapps.io-Servers (https://bevoelkerungsprognose-tool.shinyapps.io/Bevoelkerungsentwicklung_app/) gehostet wurde (Erläuterungen für das Deployment siehe Anhang).

6 Fazit

6.1 Zusammenfassung

Ziel der vorliegenden Arbeit war die Entwicklung eines Bevölkerungsprognose-Tools als Prototyp, das kommunale Entscheidungsträger bei der Infrastrukturplanung von Neubaugebieten unterstützt. Das Bevölkerungsprognose-Tool sollte dazu dienen, die Erstellung von Prognosen der Bevölkerungsentwicklung und -struktur in Neubaugebieten des Landes Baden-Württemberg zu vereinfachen und somit ihre Zeit- und Kostenintensivität zu verringern. Dieses Bevölkerungsprognose-Tool sollte als frei nutzbare Lösung integrierbar sein und den Qualitätsstandards für Benutzerfreundlichkeit entsprechen.

Anhand der in Kapitel 2 vorgestellten Methoden und Möglichkeiten wurde eine passende Lösung für die Entwicklung des Bevölkerungsprognose-Tool konzipiert und implementiert, die den Zielen und den Anforderungen der Arbeit entspricht.

Für die Prognostizierung der zu erwartenden Entwicklung der Bevölkerung des neuen Baugebietes nach planungsgerechten Altersgruppen und Geschlecht hinsichtlich der Einwohnerzahl und der demographischen Struktur wurden vier unterschiedliche Modellierungsansätze (GAM-, ARIMA-, KNN-, Kohorten-Komponenten-Modelle) angewendet und miteinander verglichen. Somit wurden mittels der Open-Source-Computing-Umgebung R 27 GAM-Modelle (1 Modell für Gesamtbevölkerung, 8 Modelle für die Bevölkerungsentwicklung nach planungsgerechten Altersgruppen und 16 Modelle für die Bevölkerungsentwicklung nach Geschlecht und Altersgruppe), 27 ARIMA-Modelle und 27 KNN-Modelle pro Art des Neubaugebietes angepasst. Außerdem wurden 2 Kohorten-Komponenten-Modelle für die zwei Neubaugebietsarten mathematisch berechnet.

In dieser Bachelorarbeit wurden insgesamt 164 Modelle berechnet, bewertet und miteinander verglichen.

Der Modellvergleich der statistischen Modelle wurde anhand ihrer Anpassungsgüte (R-Quadrat und MAPE) realisiert. Da die Kohorten-Komponenten-Modelle keine statistischen Modelle sind, sondern durch die Anwendung einfacher mathematischer Formeln berechnet werden, konnten Sie nicht mit den anderen drei Arten von Modellen statistisch verglichen werden. Zusammengefasst kann man die Schlussfolgerung ziehen, dass die GAM-Modelle und besonders die KNN-Modelle am besten abschneiden. Dieses Ergebnis bestätigt die Vermutung, dass die Bevölkerungsentwicklung nicht linear verläuft, also Modelle die nichtlineare Beziehungen beschreiben am besten geeignet sind. Die Anpassung der ARIMA-Modelle war zwar schlechter als die von GAM und KNN aber trotzdem zufriedenstellend. Am schlechtesten waren die Kohorten-Komponenten-Modelle. Auch wurde in dieser Arbeit die in der Literatur beschriebene Problematik bezüglich der einfachen Fortschreibung der Bevölkerung von Basisjahr zum Prognostizierungsjahr und das Treffen von subjektiven Annahmen für die Entwicklung der Fertilitäts- und Migrationsraten deutlich erkennbar.

Für die Entwicklung der Webanwendung wurde die R-Shiny-Entwicklungsplattform eingesetzt und sie wurde anschließend unter der URL: https://bevoelkerungsprognose-tool.shinyapps.io/Bevoelkerungsentwicklung_app/ online zur Verfügung gestellt. Durch die Bereitstellung des Tools als interaktive Webanwendung wird Entscheidungsträgern die Möglichkeit gegeben, benutzerdefinierte Prognosen für die Bevölkerungsentwicklung in Neubaugebieten zu erhalten, ohne über die entsprechenden Daten zu verfügen. Außerdem wird ihnen angeboten ihre eigenen Daten von ehemaligen Neubaugebieten zu modellieren und zu vergleichen. Durch die Bereitstellung dieses Tools haben vor allem kleinere Gemeinden, denen keine eigenen statistischen Abteilungen zur Verfügung steht, eine kostengünstige Lösung, da sie für kleingliedrige Bevölkerungsentwicklungsprognosen nicht mehr auf externe Dienstleister zugreifen müssen.

6.2 Ausblick

Mit dem Abschluss der Bachelorarbeit wurde ein Bevölkerungsprognose-Tool als Werkzeug für die Planung von Neubaugebieten erstellt, das die Basis für die Entwicklung weiterer Bevölkerungsprognose-Tools bilden kann. Da diese Lösung bereits sehr

gute Ergebnisse liefert, bietet sie auch viel Potenzial für zukünftige Erweiterungen und Weiterentwicklungen.

Das Tool ermöglicht dem Benutzer in dieser ersten Umsetzung Unterstützung bei der Planung von zwei Arten von Neubaugebieten in Städten die Freiburg ähneln. Es wäre sinnvoll das Tool in Zukunft durch die Untersuchung und Modellierung von Daten anderer Städte und auch kleinerer oder mittlerer Kommunen zu erweitern.

Eine zusätzliche Erweiterung wäre auch Strukturmodelle für die Prognostizierung der Bevölkerungsentwicklung anzuwenden. Diese Modelle beziehen weitere Variablen (wie ökonomische und räumliche Aspekte) mit ein, die den Verlauf der Bevölkerungszitreihen möglicherweise erklären. Dadurch würde die Erklärungskraft der Modelle bestärkt.

In den letzten Jahren wurde in der Literatur ein Trend beobachtet, Zeitreihenanalyse mit Anwendung von rekurrenten neuronalen Netzen, wegen ihrer Eignung für die Modellierung sequentieller Daten durchzuführen. Da in dieser Arbeit die KNN-Modelle die besten Ergebnisse geliefert haben, wäre es eine weitere Ausbaustufe auch rekurrente Neuronale Netze auszuprobieren.

Abschließend wäre es sinnvoll die Webapplikation mit der Datenbank des Melderegisters zu koppeln und dadurch die Bevölkerungsprognosen zu validieren, indem die Modelle auf der aktuellen Datenbasis automatisiert aktualisiert werden und damit eine Überprüfung der Prognose möglich wird.

Quellenverzeichnis

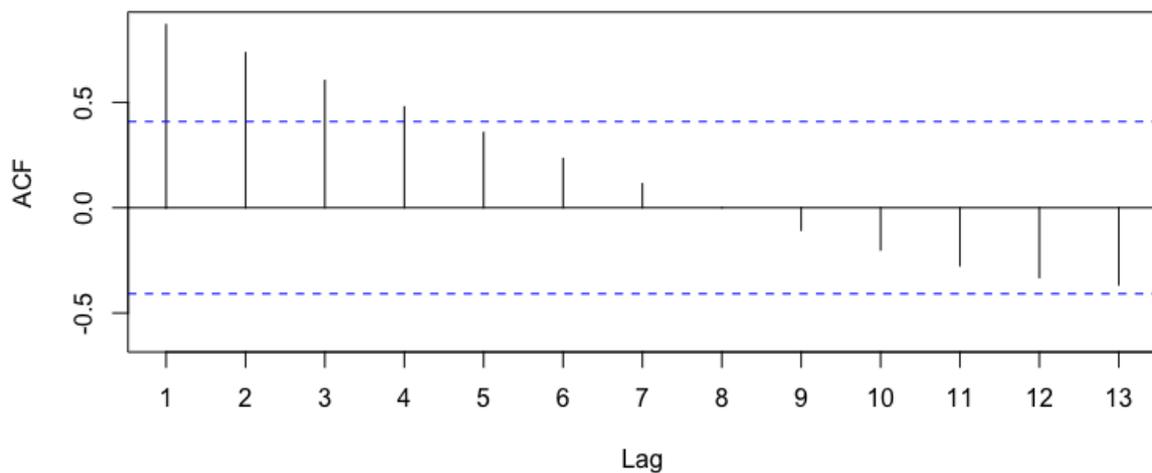
- Backhaus, K.; Erichson, B.; Weiber, R. (2015) *Neuronale Netze*. In: Fortgeschrittene Multivariate Analysemethoden. Springer Gabler, Berlin, Heidelberg.
- Backhaus, K.; Erichson, B.; Weiber, R.; Plinke, W. (2016). *Neuronale Netze*. In: Multivariate Analysemethoden. Springer Gabler, Berlin, Heidelberg.
- Backhaus, K.; Erichson, B.; Weiber, R.; Plinke, W. (2016). *Zeitreihenanalyse*. In: Multivariate Analysemethoden. Springer Gabler, Berlin, Heidelberg.
- Bourier G. (2012). *Zeitreihenanalyse*. In: Beschreibende Statistik. Gabler Verlag, pp 155-194.
- Brachat-Schwarz, W. (2008). *Reurbanisierung – Gibt es eine „Renaissance der Städte“ in Baden-Württemberg?* In: Statistisches Monatsheft Baden-Württemberg 11/2008, pp. 1-13.
- Bretz, M. (2000). *Methoden der Bevölkerungsvorausberechnung*. In: U. Mueller, B. Nauck und A. Diekmann, eds, Handbuch der Demographie. Modelle und Methoden (Band 1), Springer, Berlin/Heidelberg/New York, pp. 643-681.
- Breuer, B.; Schmall, R. (2007). *Neue Stadtquartiere. Bestand und städtebauliche Qualitäten. Vorgehen und Ergebnisse der laufenden Bestandserhebung des BBR zu neuen Stadtquartieren*. In: BBR-Online-Publikation, Nr. 01/2007, Herausgegeben vom Bundesamt für Bauwesen und Raumordnung, Bonn, pp. 1-47.
- Breyer, F. et al. (2018). *Scheitern der sozialen Wohnungspolitik: Wie bezahlbaren Wohnraum schaffen?* In: ifo Institut – Leibniz-Institut für Wirtschaftsforschung an der Universität München, München, Vol. 71, Iss. 21, pp. 3-30.
- Bundesinstitut für Bau- Stadt- und Raumforschung (2011). *Renaissance der Großstädte – eine Zwischenbilanz*. In: BBSR-Berichte KOMPAKT, pp 1-11.
- Burch T.K. (2018). *The Cohort-Component Population Projection: A Strange Attractor for Demographers*. In: Model-Based Demography. Demographic Research Monographs (A Series of the Max Planck Institute for Demographic Research). Springer, Cham.
- Chang, W.; Cheng, J.; Allaire, J.; Xie, Y.; McPherson, J. (2015). *Shiny: Web application framework for R*. <https://cran.r-project.org/web/packages/shiny/index.html>.
- Dehler, K. H. (1976). *Zielprognosen der Stadtentwicklung. Untersuchung am Beispiel kleinräumlicher Bevölkerungsprognosen*. Harald Boldt-Verlag.
- Diamond, I., Tesfaghiorghis, H. & Joshi, H. (1990). *The uses and users of population projections in Australia*. In: Journal of Population Research 7: 151, pp. 151-170

- Frank, S. (2011). *Die „Rückkehr“ der Familien in die Stadt*. In: Informationen zur Stadtgeschichte, 1/2011, pp 64-75.
- Fricke, A. (2015). *Reurbanisierung als stadtpolitische Herausforderung: Positionspapier zum Umgang mit dem Trend in der Stadt*. In: Fricke, Axel Siedentop, Stefan Zakrzewski, Philipp (Ed.): *Reurbanisierung in baden-württembergischen Stadtregionen*, Verlag der ARL - Akademie für Raumforschung und Landesplanung, Hannover, pp. 171-186.
- George, M.; Smith, S.; Swanson, D.; Tayman, J. (2004). *Population projections*. In: J. Siegel und D. Swanson, eds, "The Methods and Materials of Demography", Elsevier, San Diego/London, pp. 561–601.
- Geppert, K.; Gornig, M. (2010). *Mehr Jobs, mehr Menschen: Die Anziehungskraft der großen Städte wächst*. In: DIW-Wochenbericht 77 (19), pp 2-10.
- Haase, A. et al (2009). *Emergent Spaces of Reurbanisation: Exploring the Demographic Dimension of Inner-city Residential Change in a European Setting*. In: *Population, Space and Place*, Vol: 16, pp 443-463.
- Hansen, H.S. (2010). *Small-Area Population Projections - A Key Element in Knowledge Based e-Governance*. In: Andersen K.N., Francesconi E., Grönlund Å., van Engers T.M. (eds) *Electronic Government and the Information Systems Perspective. EGOVIS 2010. Lecture Notes in Computer Science*, vol 6267. Springer, Berlin, Heidelberg, pp. 32-46.
- Harding, A. et al (2011). *Demographic change and the needs-based planning of government services: projecting small area populations using spatial microsimulation*. In: *Journal of Population Research*, Vol. 28, pp. 203-224.
- Hinde, A. (1998). *Demographic Methods*. Arnold, London
- Holland H., Scharnbacher K. (2010). *Zeitreihenanalyse*. In: *Grundlagen der Statistik*. Gabler Verlag, pp 79-92.
- Hyndman, R.J.; Khandakar, Y. (2008). *Automatic time series forecasting: the forecast package for R*. *J. Stat. Softw.* 26(3), pp. 1-22.
- Kenny P. (2014). *Time Series*. In: *Better Business Decisions from Data*. Apress, Berkeley, CA, pp. 197-203.
- Kirchgässner, G.; Wolters, J.; Hassler, U. (2013). *Introduction to modern time series analysis*. Springer Verlag
- Konasani, V.R.; Kadre S. (2015). *Time-Series Analysis and Forecasting*. In: *Practical Business Analytics Using SAS*. Apress, Berkeley, CA.
- Madsen, H. (2007). *Time Series Analysis*. Chapman & Hall/CRC.

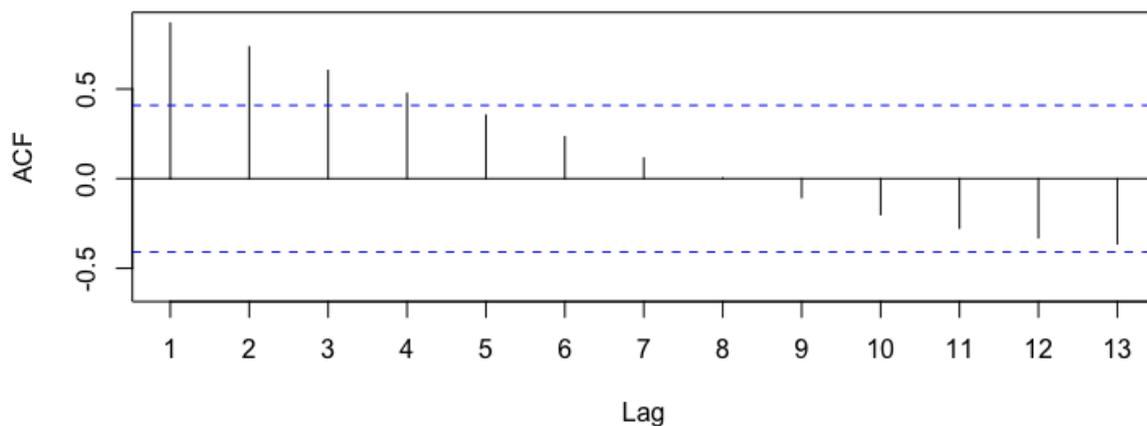
- McLeod, A.I., Yu; H.: Mahdi, E. (2012). *Time series analysis with R*. In: Subba Rao, T., Subba Rao, S., Rao, C.R. (eds.) *Handbook of Statistics*, vol. 30, pp. 661–712. Elsevier, Amsterdam.
- Shumway, R.; Stoffer, D. (2011). *Time Series Analysis and its Applications. With R Examples*. Springer Verlag.
- Siedentop, S. (2008). *Die Rückkehr der Städte? Zur Plausibilität der Reurbanisierungshypothese*. In: *Informationen zur Raumentwicklung*, 3/4.2008, pp 193-210.
- Siedentop, S. et al. (2017). *Wachsende Städte, junge Quartiere – Alterssegregation durch Reurbanisierung?* In: ILS (Institut für Landes- und Stadtentwicklungsforschung) – *Trends*, 2/17, pp. 1-8.
- Smith, S., Tayman, J., Swanson, D. (2001). *State and Local Population Projections – Methodology and Analysis*. Springer, Heidelberg.
- Steinberg, J. & Doblhammer-Reiter, G. (2010). *Demografische Bevölkerungsprognosen*. In: *Bundesgesundheitsblatt - Gesundheitsforschung – Gesundheitsschutz*, Vol. 53, pp. 393-403.
- Sturn, G. (2012). *Wachsende Großstädte in Deutschland – empirische Befunde aus der Stadtbeobachtung des BBSR*. In: *Mannheimer Schriften zu Wohnungswesen, Kreditwirtschaft und Raumplanung (Band 10)*, pp 19-31.
- Stycos, M. (1995). *Population, projections, and policy: A cautionary perspective*. In: *Population and Environment*, Vol. 16, pp. 205-219.
- Wang, X. and vom Hofe, R. (2007). *Research Methods in Urban and Regional Planning*. Tsinghua University Press, Beijing and Springer-Verlag GmbH Berlin Heidelberg.
- Willmann, T. (2014). *Kleinräumige Bevölkerungsvorausrechnung und Haushaltevorausrechnung für Freiburg 2014 bis 2030*. In: *Beiträge zur Statistik der Stadt Freiburg im Breisgau*.
URL:http://www.staedtestatistik.de/fileadmin/kosis/sikurs/Veroeffentlichungen/Prognose_Freiburg_2014Bis2030.pdf [Stand: 02.08.2019]
- Wilson, T. (2016). *Evaluation of Alternative Cohort-Component Models for Local Area Population Forecasts*. In: *Population Research and Policy Review*, Vol. 35, pp. 241-261.
- Wood, S.N. (2006). *Generalized additive models: an introduction with R*. Chapman and Hall/CRC.
- Zhang, P. (2003). *Time series forecasting using a hybrid ARIMA and neural network model*. In: *Neurocomputing*, Vol. 50, pp. 159-175.
- Zhang, P.; Patuwo, B.; Hu, Y. (1998). *Forecasting with artificial neural networks: the state of the art*. In: *Int. J. Forecasting*, Vol. 14, pp. 35-62.

Anhang

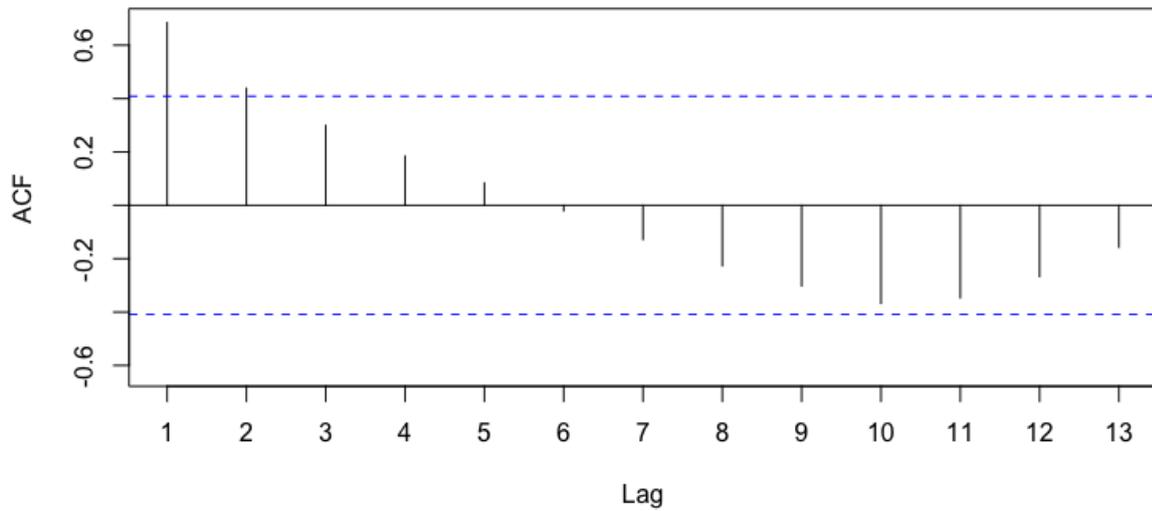
A. Autokorrelationsfunktionen der Zeitreihen



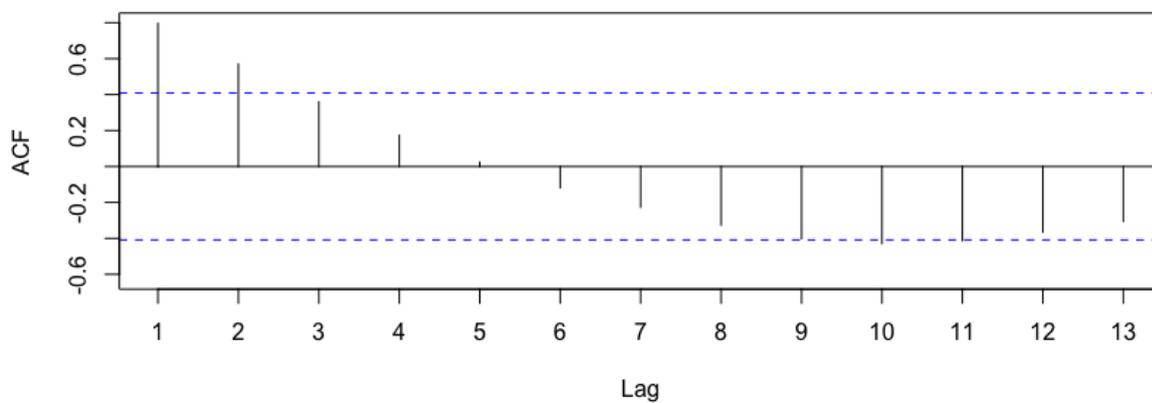
Anhang-Abbildung 1: Autokorrelationsfunktion der männlichen Bevölkerungsdaten (Rieselfeld) (eigene Darstellung)



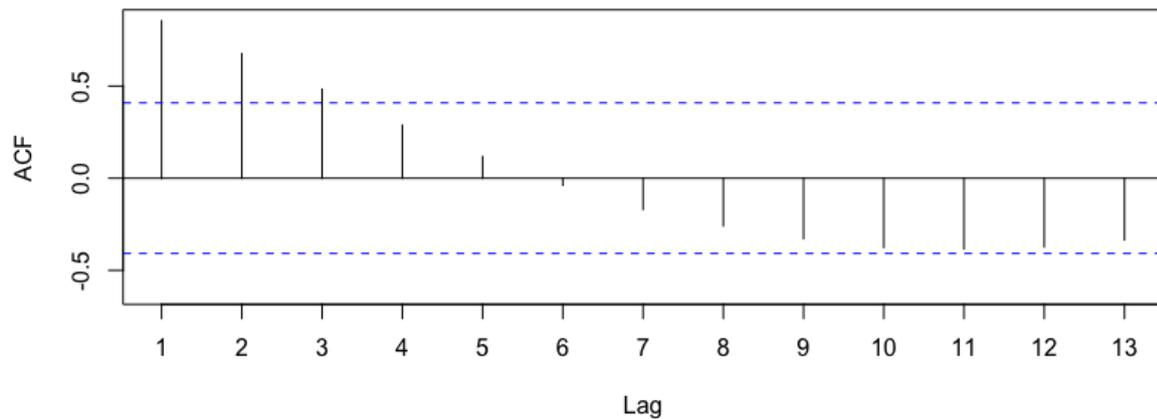
Anhang-Abbildung 2: Autokorrelationsfunktion der weiblichen Bevölkerungsdaten (Rieselfeld) (eigene Darstellung)



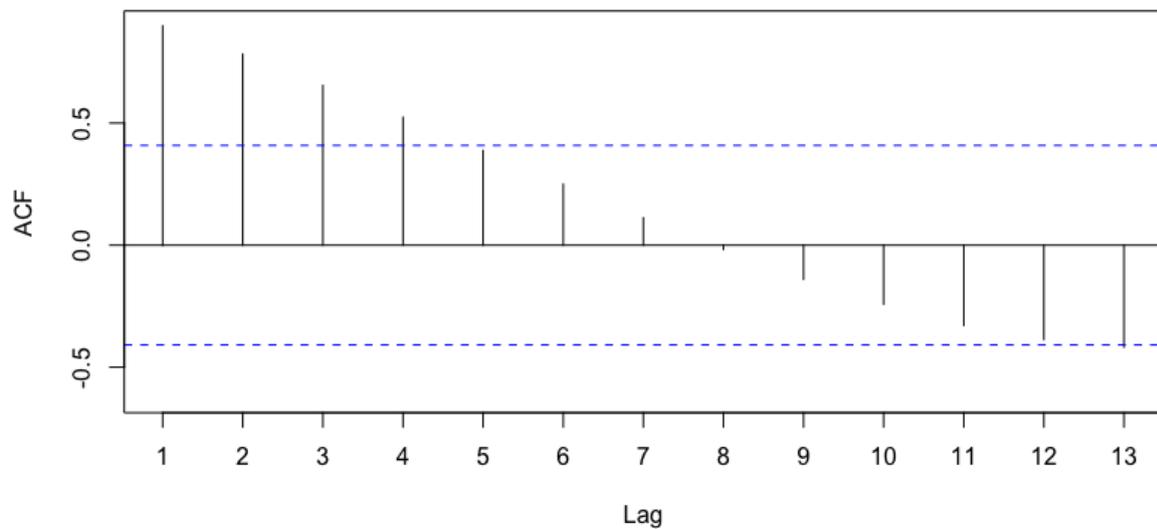
Anhang-Abbildung 3: Autokorrelationsfunktion der Bevölkerungsdaten der Altersgruppe 0-3 Jahre (Rieselfeld) (eigene Darstellung)



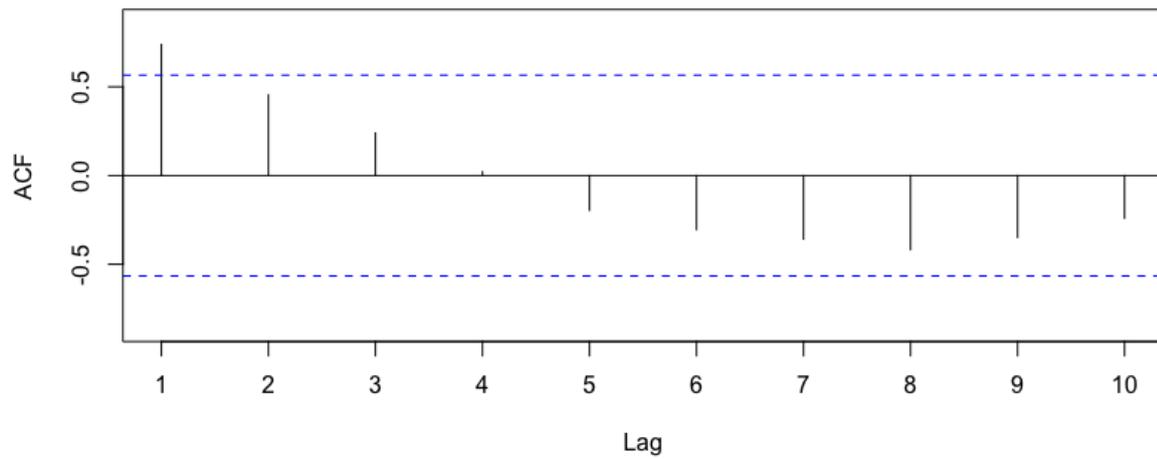
Anhang-Abbildung 4: Autokorrelationsfunktion der Bevölkerungsdaten der Altersgruppe 3-6 Jahre (Rieselfeld) (eigene Darstellung)



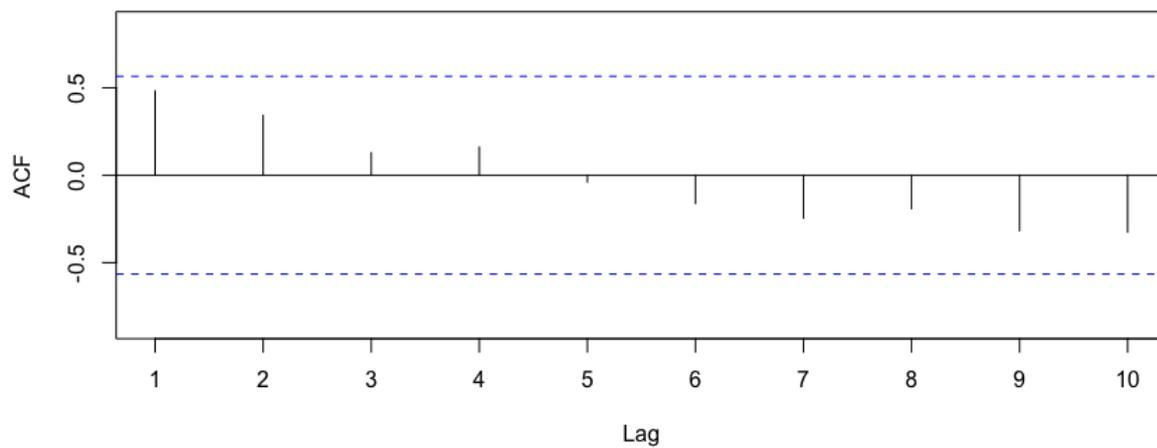
Anhang-Abbildung 5: Autokorrelationsfunktion der Bevölkerungsdaten der Altersgruppe 6-10 Jahre (Rieselfeld) (eigene Darstellung)



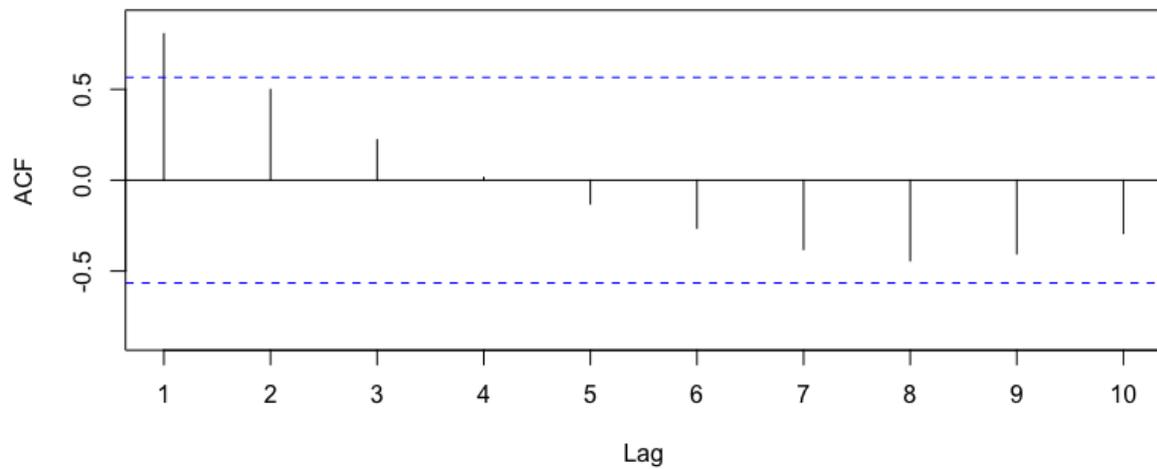
Anhang-Abbildung 6: Autokorrelationsfunktion der Bevölkerungsdaten der Altersgruppe 10-18 Jahre (Rieselfeld) (eigene Darstellung)



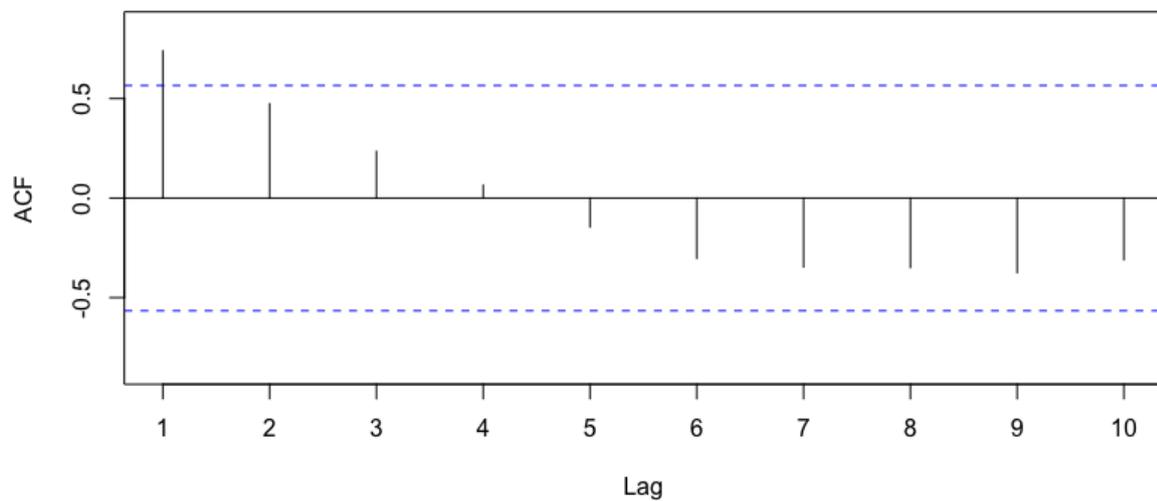
Anhang-Abbildung 7: Autokorrelationsfunktion der Bevölkerungsdaten der Altersgruppe 18-25 Jahre (Rieselfeld) (eigene Darstellung)



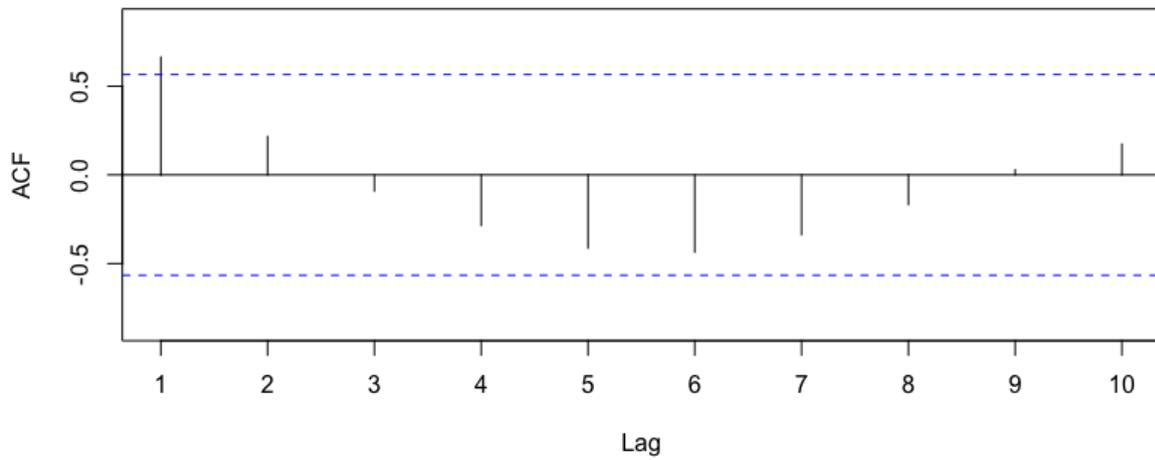
Anhang-Abbildung 8: Autokorrelationsfunktion der Bevölkerungsdaten der Altersgruppe 25-40 Jahre (Rieselfeld) (eigene Darstellung)



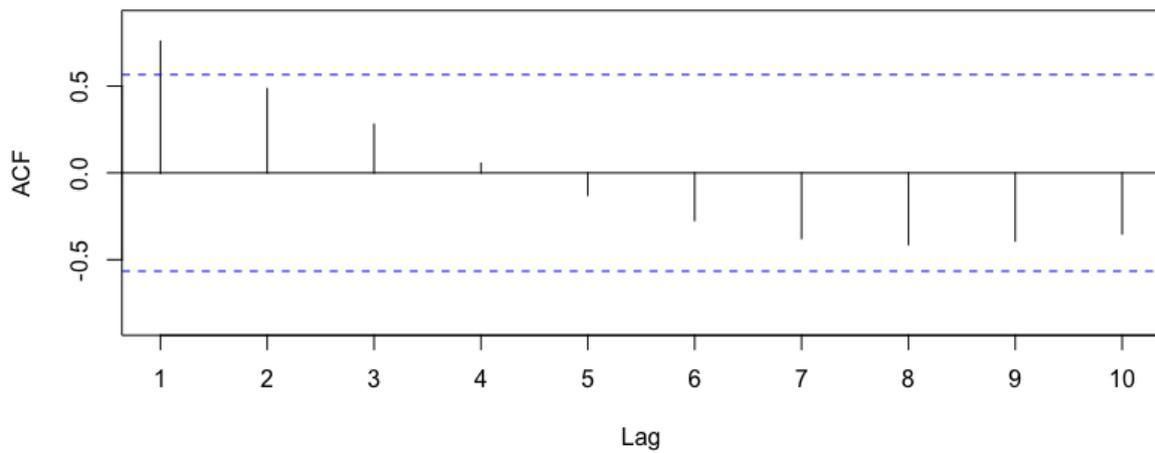
Anhang-Abbildung 9: Autokorrelationsfunktion der Bevölkerungsdaten der Altersgruppe 40-70 Jahre (Rieselfeld) (eigene Darstellung)



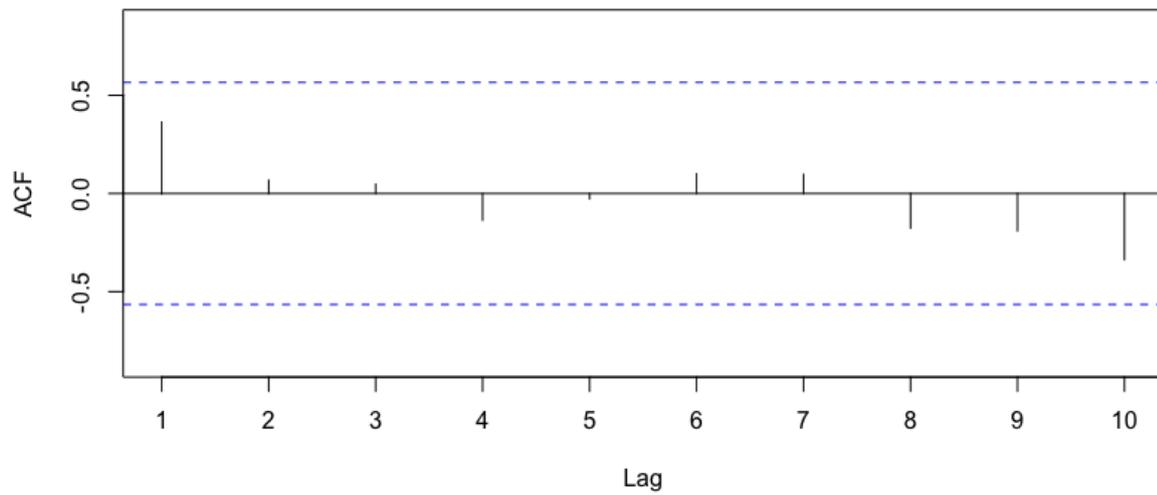
Anhang-Abbildung 10: Autokorrelationsfunktion der Bevölkerungsdaten der Altersgruppe 70 Jahre oder älter (Rieselfeld) (eigene Darstellung)



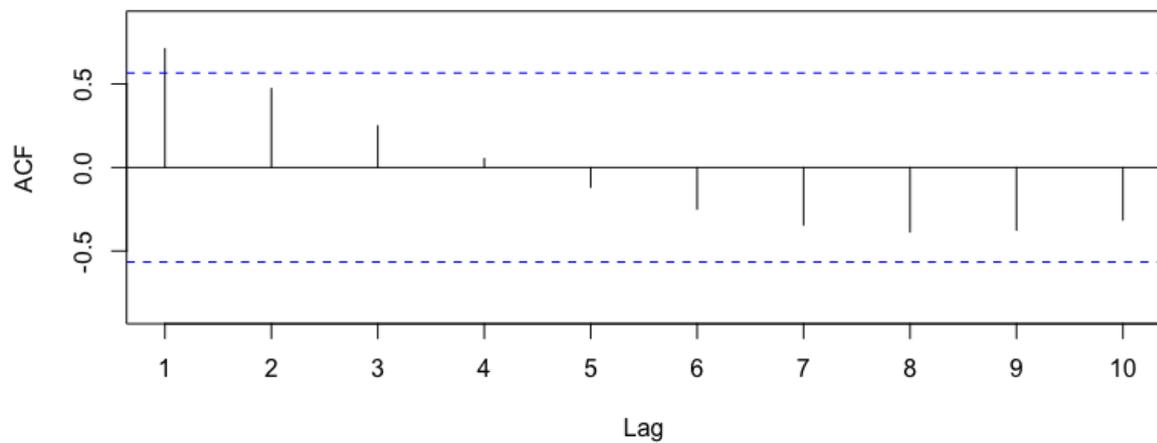
Anhang-Abbildung 11: Autokorrelationsfunktion der weiblichen Bevölkerungsdaten (Vauban) (eigene Darstellung)



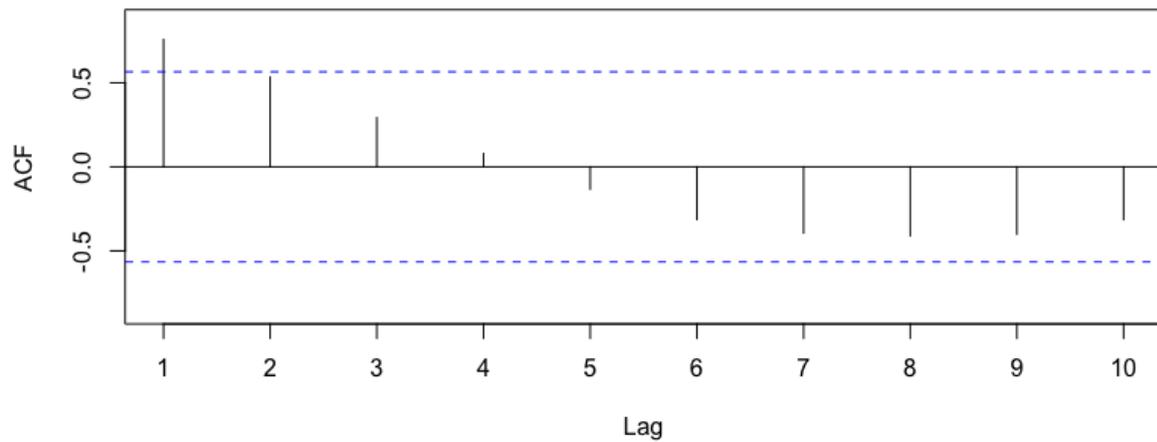
Anhang-Abbildung 12: Autokorrelationsfunktion der männlichen Bevölkerungsdaten (Vauban) (eigene Darstellung)



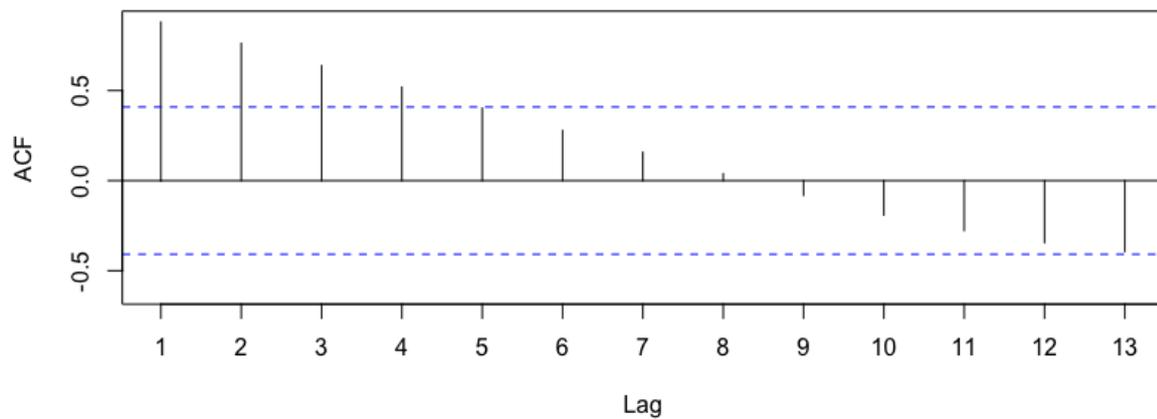
Anhang-Abbildung 13: Autokorrelationsfunktion der Bevölkerungsdaten der Altersgruppe 0-3 Jahre (Vauban) (eigene Darstellung)



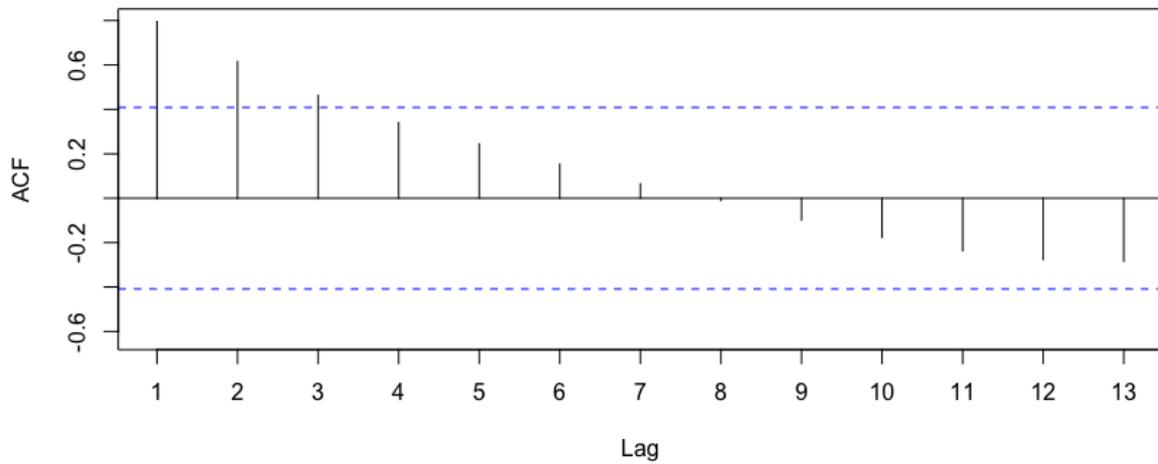
Anhang-Abbildung 14: Autokorrelationsfunktion der Bevölkerungsdaten der Altersgruppe 3-6 Jahre (Vauban) (eigene Darstellung)



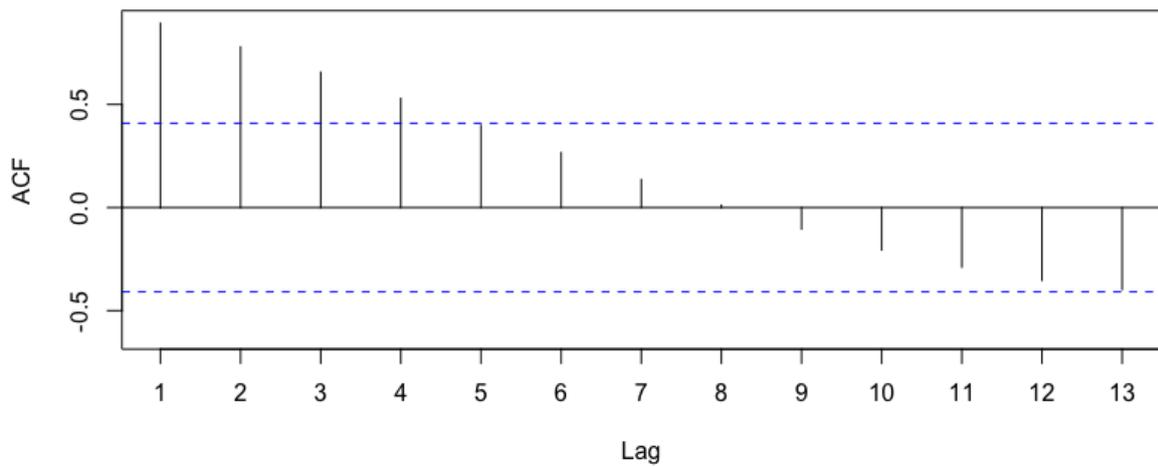
Anhang-Abbildung 15: Autokorrelationsfunktion der Bevölkerungsdaten der Altersgruppe 6-10 Jahre (Vauban) (eigene Darstellung)



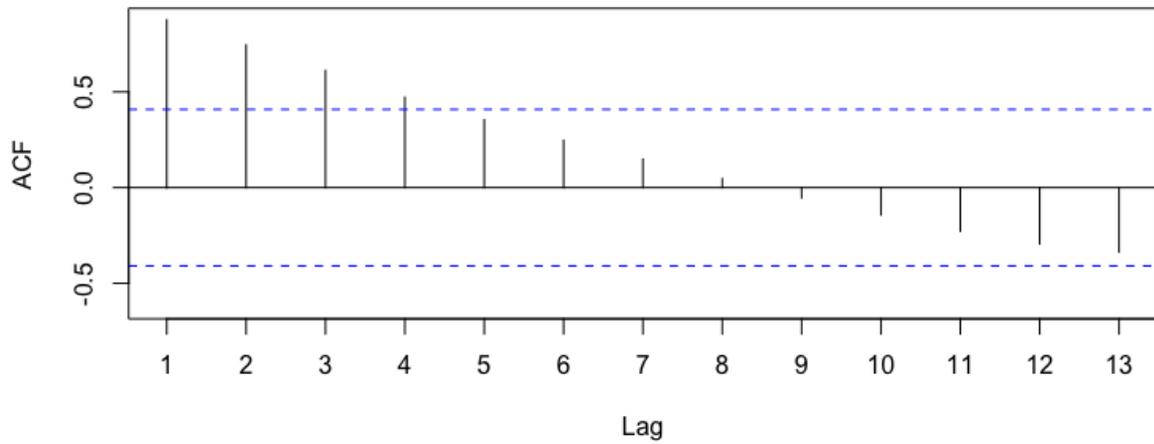
Anhang-Abbildung 16: Autokorrelationsfunktion der Bevölkerungsdaten der Altersgruppe 10-18 Jahre (Vauban) (eigene Darstellung)



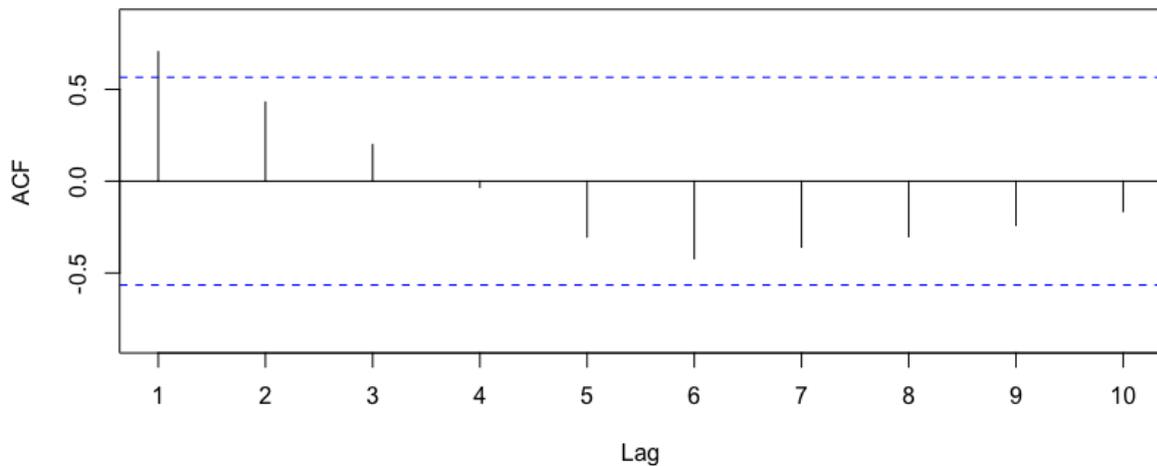
Anhang-Abbildung 17: Autokorrelationsfunktion der Bevölkerungsdaten der Altersgruppe 18-25 Jahre (Vauban) (eigene Darstellung)



Anhang-Abbildung 18: Autokorrelationsfunktion der Bevölkerungsdaten der Altersgruppe 25-40 Jahre (Vauban) (eigene Darstellung)



Anhang-Abbildung 19: Autokorrelationsfunktion der Bevölkerungsdaten der Altersgruppe 40-70 Jahre (Vauban) (eigene Darstellung)



Anhang-Abbildung 20: Autokorrelationsfunktion der Bevölkerungsdaten der Altersgruppe 70 Jahre oder älter (Vauban) (eigene Darstellung)