

Bachelorarbeit

Untersuchung von Verfahren für die Generierung von Kundenstandorten für ein Supply-Chain-Simulationsmodell

Daniel Stach
Matrikelnummer 171089
Studiengang Logistik

Ausgegeben am:
18.04.2018

Eingereicht am:
10.07.2018

Betreuer:
Prof. Dr.-Ing. Markus Rabe
M.Sc. Astrid Klüter

Inhaltsverzeichnis

1 Einleitung.....	1
2 Einordnung in wissenschaftlichen Kontext.....	4
2.1 Merkmale einer Supply Chain.....	4
2.2 Ausgewählte Transportmethoden einer Supply Chain	7
2.3 Aufbau eines Simulationsmodells	10
2.4 Distanzmessung in der Ebene.....	14
3 Verfahren bei unzureichender Datengrundlage	18
3.1 Die Standortplanung im Kontext fehlender Informationen	18
3.2 Gleichverteilung von Standorten	20
3.3 Statistische Verfahren bei fehlenden Daten	20
3.3.1 Statistische Schätzfunktionen.....	21
3.3.2 Imputation fehlender Werte.....	23
3.3.3 Mittelwertberechnungen	24
3.4 Data Farming und Data Mining.....	26
4 Entwicklung von Verfahren zur Standortannäherung.....	28
4.1 Anwendungsgrundlage und Bewertungsverfahren	28
4.2 Verfahrensentwicklung	32
4.2.1 Mittelpunkte als Standorte verwenden.....	32
4.2.2 Standortschätzung durch Geradenschnittpunkte	33
4.2.3 Standortannäherung durch Tourenabbildung	34
5 Anwendung von Approximationsverfahren	39
5.1 Gleichverteilung von Standorten.....	39
5.2 Mittelpunkte als Standorte verwenden	42
5.3 Standortschätzung durch Geradenschnittpunkte	45
5.4 Standortannäherung durch Tourenabbildung	47
6 Vergleich der Approximationsverfahren	50
7 Zusammenfassung und Ausblick.....	53
Literaturverzeichnis	X
Abbildungsverzeichnis	XIV
Tabellenverzeichnis	XV
Formelverzeichnis	XVI
Abkürzungsverzeichnis.....	XVII

Anhang 1: Erzeugung einer Datengrundlage.....	XVIII
Abfrage in Overpass Turbo.....	XVIII
Abgefragtes Gebiet	XVIII
Daten der ausgegebenen Postfilialen	XIX
Anhang 2: Ergebnisse der Approximationsverfahren	XXIII
Darstellung der Mittellinienapproximation	XXIII
Graphische Abbildung von Milk Runs	XXIV
Koordinaten der Ergebnisse	XXIV

1 Einleitung

Weltweit wächst die Bevölkerung stetig an. Nach einer Prognose der Vereinten Nationen (2015) wird dieser Trend in der Zukunft weiter fortgesetzt. Besonders in Städten und großen Agglomerationsräumen wächst die Population überproportional stark an. Im Jahr 2008 lebten deshalb erstmals mehr Menschen in Städten als auf dem Land (vgl. UN, 2008). Die steigende Bevölkerungsdichte in Städten und die damit resultierenden Verschärfungen des innerstädtischen Lebens, wie z.B. ein erhöhtes Verkehrsaufkommen und eine erhöhte Nachfrage nach Konsumgütern aller Art, führen zu veränderten Bedingungen für die urbane Versorgung. Versorgende Betriebe unterschiedlicher Ausrichtung und Größe stehen dadurch vor der Herausforderung andere Versorgungskonzepte zu entwickeln, die an die neuen Gegebenheiten besser angepasst sind. Zu diesen Einrichtungen können öffentliche Dienstleister, wie z.B. Krankenhäuser oder Feuerwachen, sowie auch gewinnorientierte Unternehmen zählen. Zu letzteren gehören auch Unternehmensnetzwerke, die eine organisiertes Liefernetzwerk zwischen ihren Mitgliedern, eine sogenannte Supply Chain (SC), bilden und in diesem Netzwerk nachfolgende Unternehmen beliefern müssen. Die Qualität dieser Liefernetzwerke bildet die Grundlage für wirtschaftlichen Erfolg (vgl. Werner, 2013).

Im Rahmen der Bevölkerungsvergrößerung soll das Versorgungskonzept einer bestehenden Supply Chain analysiert werden. Um diese Supply Chain und ihre Struktur untersuchen und bewerten zu können, soll diese simuliert werden. Für das Simulationsmodell werden aus allen Bereichen der Supply Chain qualitativ und quantitativ hochwertige Daten benötigt. Dazu gehören auch ausreichende Informationen zu den Distributionsstandorten der beliefernden SC-Mitglieder und zu den Standorten der zu beliefernden SC-Kunden, die am Ende eines solchen Liefernetzwerks stehen. In einigen Fällen kann es allerdings vorkommen, dass die Lage der Standorte nicht oder nur zu einem Teil bekannt ist bzw. die Informationen entsprechend begrenzt zur Verfügung stehen. Um die relevanten Daten zu Standorten und folglich zur Simulation zu erhalten, werden in der Regel umfassende Erhebungen zur Datenerfassung durchgeführt, die mit einem gewissen Aufwand verbunden sind. Anschließend erfolgt eine Datenaufbereitung der gesammelten Informationen, um damit im Folgenden effektiv arbeiten zu können. Erhebungen können Befragungen oder Messungen sein, bei denen in großem Umfang über einen längeren Zeitraum Daten gesammelt werden (Buchholz et al., 1998). In manchen Fällen ist es allerdings nicht möglich, diese Erhebungen überhaupt oder in vollem Umfang durchzuführen, da z.B. nicht genug Zeit bis zu einem gesetzten Stichtag vorhanden ist. In solchen Fällen existiert also eine nicht ausreichende Datengrundlage.

Im vorliegenden Fall sind die genauen Lokalisierungen der Kundenstandorte für eine zu simulierende Supply Chain unbekannt. Allerdings ist bekannt, wie viele Kundenstandorte die Supply Chain hat und in welchen Postleitzahlengebieten diese liegen. Um dennoch mittels einer Simulationsstudie die Supply Chain untersuchen und bewerten zu können, müssen die Daten aufbereitet oder ergänzt werden. Konkret heißt das also, die genauen Koordinaten der Kundenstandorte dieser Supply Chain müssen approximiert werden, damit sie in der Simulation

verwendet werden können. Dementsprechend müssen Verfahren gefunden und angewendet werden, die Standorte approximieren können. Hierbei ist zu untersuchen, ob auch Verfahren zur allgemeineren Approximation verwendet werden können, die nicht direkt Standorte annähern. Da es je nach angewendetem Verfahren Unterschiede in der Qualität des Outputs gibt, bietet es sich an, mehrere Methoden anzuwenden. Es ist also zu überprüfen, welche der betrachteten Verfahren die Realität am besten abbilden. Allgemein gibt es im Bereich der Datenergänzung und der Datengenerierung viele verschiedene Richtungen und Ansätze, die sich mit dem Problem von fehlenden Daten auseinandersetzen. Zu erwähnen sind in diesem Kontext beispielsweise das sogenannte Data Farming, das technische Hilfsmittel zur Datengenerierung einsetzt. Große Bedeutung haben auch Methoden der Statistik, die auf unterschiedlicher Weise fehlende Daten abschätzen sollen. Darüber hinaus wurde von dem Fachgebiet IT in Produktion und Logistik (ITPL) der TU Dortmund in diesem Zusammenhang ein weiteres, auf der Statistik aufbauendes Verfahren entwickelt, das speziell für ein derartiges Problem ausgerichtet ist und Kundenstandorte bei unzureichender Datenlage approximieren soll.

Das Ziel dieser Bachelorarbeit ist die Untersuchung von Verfahren bei einer unzureichenden Datenlage und die Überprüfung ihrer Tauglichkeit anhand der Anwendung auf ein konkretes Problem, bei dem Kundenstandorte einer Supply Chain zur Verwendung in einem Simulationsmodell geschätzt werden sollen. Dabei soll in erster Linie die Frage beantwortet werden, ob das vom ITPL neu entwickelte Verfahren zur Generierung von Kundenstandorten bei begrenzter Datenlage die Realität gut genug abbildet oder ob andere Verfahren zur Approximation von Standorten besser geeignet sind. Explizit sind dabei fünf Teilaufgaben zu bearbeiten. In der ersten Teilaufgabe werden verschiedene Grundlagen zur behandelten Problemstellung erörtert und wichtige Merkmale dazu herausgearbeitet. Danach folgt in der zweiten Teilaufgabe das Aufzeigen von Verfahren und Vorgehensweisen bei einer unzureichenden Datenlage. Diese Verfahren werden hinsichtlich der Problemstellung untersucht und bewertet, in wie weit sie oder Teile der Verfahren zur Standortapproximation geeignet sind. Die dritte Teilaufgabe umfasst die Erstellung einer Fallstudie, auf die unterschiedliche Verfahren bezogen werden können, und die Generierung eines Datensatzes, der als Referenz für die abschließende Bewertung der Güte der Ansätze dient. Da eventuell nicht alle Verfahren auf die Problemstellung zugeschnitten sind, werden in der vierten Teilaufgabe die Verfahren angepasst oder wenn nötig neu entwickelt. In der fünften Teilaufgabe werden diese Verfahren dann innerhalb der zuvor konstruierten Fallstudie angewendet und dabei untersucht, ob die Ergebnisse der Methoden für die Realität gute Näherungen darstellen.

Zur Problembehandlung soll im Folgenden die Vorgehensweise näher erläutert werden. Zunächst wird der Stand der Forschung verschiedener Themengebiete durch eine intensive Literaturrecherche genauer beleuchtet. Dabei werden wesentliche Grundlagen zu den Teilbereichen der Arbeit beschrieben und mathematische Definitionen aufgestellt, um später auf diesem Wissen aufbauen zu können. Da für das Supply-Chain-Simulationsmodell Daten zu der Verteilung der Standorte benötigt werden, diese aber nur in sehr begrenztem Umfang vorhanden sind, werden anschließend Verfahren beschrieben, die allgemein bei unzureichenden Daten angewendet werden, um diese sinnvoll zu ergänzen bzw. abzuschätzen. Hierbei sollen sie auf ihre Eignung für das vorliegende Problem untersucht und bewertet werden. Aus den aufgezeigten

Informationen und generellen Verfahren sollen Ansätze entwickelt werden, die speziell zur Standortapproximation geeignet sind. Dabei können vorher dargelegte Methoden auch entsprechend angepasst werden. Folglich werden die angepassten und entwickelten Verfahren auf das konkrete Problem einer unzureichenden Datenlage in einer Supply Chain angewendet und auf ihre Güte hin analysiert. Dazu wird die Problemstellung in eine konkrete Fallstudie überführt, welche als Basis für die Anwendung der Methoden dienen soll. Die Fallstudie zieht ihre Daten aus einem generierten Referenzdatensatz, der stellvertretend für eine Supply Chain stehen soll. Zur Bewertung der Verfahren soll ein eigens entwickeltes Bewertungsverfahren benutzt werden. Dieses Verfahren nutzt die reellen und die approximierten Daten, um einen Vergleich der Verfahren untereinander möglich zu machen. Abschließend wird ein Fazit zu der Tauglichkeit der entwickelten Verfahren gezogen und ein Ausblick zu der Problemstellung gegeben.

2 Einordnung in wissenschaftlichen Kontext

Das Problem umfasst die Approximation von Kundenstandorten einer Supply Chain, die zur Bewertung ihrer Versorgungsstrukturen simuliert werden soll. Aus der Problemformulierung sind drei Themengebiete abzuleiten, die das wissenschaftliche Fundament dieser Arbeit bilden sollen. Es sollen dementsprechend die Merkmale und die Organisation einer Supply Chain sowie deren zentralen Transportstrukturen aufgezeigt und erörtert werden. Dazu werden zunächst in Abschnitt 2.1 charakteristische Eigenschaften und die organisatorische Zusammensetzung von Supply Chains im Allgemeinen dargelegt. In diesem Zusammenhang folgt in Abschnitt 2.2 die Betrachtung gängiger Transportstrukturen und Transportmethoden, die zur Versorgung innerhalb von Supply Chains zum Einsatz kommen. Die Bewertung der Supply Chain soll durch eine Simulation erfolgen, daher ist der grundlegende Aufbau einer Simulationsstudie der dritte wichtige Themenbereich. In Abschnitt 2.3 werden deshalb die generellen Strukturen und die Arbeitsschritte in einer Simulationsstudie erläutert und veranschaulicht. Weil bei der Approximation von Werten die Abweichung zur Realität eine große Rolle spielt, sollen abschließend in diesem Kapitel Möglichkeiten zur Messung von Abständen untersucht werden. Die Annäherung beschränkt sich hier auf Standorte in einer Fläche, weshalb in Abschnitt 2.4 Methoden zur Distanzmessung und die dazugehörigen mathematischen Formeln vorgestellt werden sollen.

2.1 Merkmale einer Supply Chain

Eine große Bedeutung in der Logistik und den Wirtschaftswissenschaften kommt der Versorgungs- bzw. der Lieferkette zu, die im Allgemeinen besser unter ihrer englischen Bezeichnung Supply Chain bekannt ist. In der Literatur gibt es eine Vielzahl von Definitionen einer Supply Chain, die sich aber in einigen Kernelementen überschneiden. Nach Corsten und Gössinger (2008) ist eine Supply Chain ein Unternehmensnetzwerk, bei dem tendenziell stabile Zulieferer- und Abnehmerbeziehungen entlang der Wertschöpfungskette bestehen. Die Unternehmen in diesem Netzwerk sind autonom, arbeiten aber kooperativ an einer Leistungserstellung. Mentzer (2001) definiert die Supply Chain als ein Zusammenschluss von drei oder mehr Einheiten (Organisationen oder Individuen), die direkt an die vor- und nachgelagerten Produktions-, Service-, Finanz- und Informationsflüsse von einer Quelle zu einem Kunden beteiligt sind. Becker (2018) beschreibt die Supply Chain wiederum als eine Wertschöpfungskette, die aus Material-, Informations- und Werteflüssen vom Rohstofflieferanten bis zum Endkunden besteht. Eine Supply Chain zeichnet sich also dadurch aus, dass sie eine Lieferkette bzw. ein Liefernetzwerk von verschiedenen autonomen Organisationen ist, die kooperativ an einem Wertschöpfungsprozess arbeiten. Dabei erfolgen unter anderem Waren- und Informationsflüsse, die von einer Quelle bis zum Endkunden reichen.

Der von vielen Autoren verwendete Begriff einer Lieferkette impliziert eine sequentielle Folge, der Aufbau einer Supply Chain oftmals aber eher einen Netzwerkcharakter hat (vgl. Thaler,

2007). Im Folgenden soll daher von einer Supply Chain als ein Netzwerk ausgegangen werden. Die im folgenden Abschnitt aufgeführten Eigenschaften lassen sich aber auch auf eine sequentielle Versorgungskette übertragen oder an diese angleichen.

Der zentrale Vorgang in einer Supply Chain ist der zugrundeliegende Wertschöpfungsprozess, an dem gemeinschaftlich gearbeitet wird. Der Wertschöpfungsprozess einer Supply Chain deckt alle Stufen und Partner des Zusammenschlusses ab. Dazu erfolgt eine Dekomposition der komplexen Gesamtaufgabe in Teilaufgaben, die den einzelnen Supply-Chain-Partnern zugeordnet werden (vgl. Corsten und Gössinger, 2008). Jedes Unternehmen, das diesem Netzwerk angehört, trägt also einen Teil zu diesem Wertschöpfungsprozess bei. Dem Wertschöpfungsprozess soll hierbei eine ganzheitliche Betrachtung der Unternehmung durch die Teilnehmer und ein prozesskettenorientiertes Denken zugrunde liegen (vgl. Schulte, 2013). Am Ende des Wertschöpfungsprozesses steht der Verbraucher der Leistungserstellung, wodurch die eigentliche Aufgabe einer Supply Chain ersichtlich wird. Die Gesamtaufgabe bzw. der Zweck dieses Versorgungsnetzwerks besteht aus der optimalen Bedarfsabdeckung des Kunden (vgl. Corsten und Gössinger, 2008). Die wesentlichen Impulse zur Leistungserbringung gehen hierbei vom Kunden aus (vgl. Arndt, 2008). Für die Supply Chain bedeutet das, dass alle Aktivitäten und Prozesse innerhalb dieses Zusammenschlusses auf den Kunden hin ausgerichtet werden (vgl. Gudehus 2, 2007). Die Supply Chain zeichnet sich also durch eine hohe Kundenorientierung aus. Eng verbunden mit der Ausrichtung am Kunden in einer Supply Chain ist der sogenannte Peitschenschlageffekt (engl. Bullwhip-Effect), der die Vergrößerung von Nachfragemengen entlang der Kette beschreibt. Werden Informationen zum Konsum der Kunden innerhalb der Kette nicht im vollen Umfang oder nicht rechtzeitig an die Partner vorgestellter Stufen weitergegeben, vergrößert sich die nachgefragte Menge pro Stufe. Dieses Problem ist auf das Lagern von Sicherheitsbeständen der einzelnen Mitglieder zurückzuführen, die dadurch ein Abreißen der Kette verhindern wollen, allerdings diese Information nicht vollumfänglich weitergeben (vgl. Werner, 2013). Ein zu großer Lagerbestand schmälert aber den Gewinn, der durch die Zusammenarbeit der Unternehmen im Endeffekt entstehen soll. Der Zusammenschluss der Unternehmen soll letztendlich für jedes Mitglied bessere wirtschaftliche Konditionen bieten, als es die Organisationen ohne dieses Netzwerks vorfinden würden. Dabei ist zu beachten, dass ein Optimum der gesamten Kette nicht zwingend das Beste für die einzelnen Teilnehmer bedeutet. Durch ein Optimum der gesamten Supply Chain können unter Umständen einzelne Organisationen schlechter gestellt werden, als dies bei individueller Optimallösung der Fall wäre (vgl. Bacher, 2004). Deshalb sollten Kosteneinsparungen entlang der gesamten Kette auf die einzelnen Teilnehmer aufgeteilt werden, um sicherzustellen, dass jedes Mitglied davon profitieren kann (vgl. Poirier und Reiter, 1997). Verhandlungen zwischen den Akteuren regeln diese Kosten- und Gewinnverteilungen.

Innerhalb einer Supply Chain erfolgt ein abgestimmter Waren- und Informationsstrom zwischen allen Beteiligten. Der Warenstrom fließt von dem Rohstofflieferanten bis zum Endverbraucher über alle Stufen in eine Richtung (vgl. Heidtmann, 2008). Die Unternehmen einer Stufe in diesem Netzwerk beliefern dabei ihre Nachfolger der nächsten Stufe, die wiederum ihre nachfolgende Stufe beliefern. Für eine Supply Chain ist generell nicht festgelegt, wie viele Stufen oder Mitglieder sie besitzen muss. Sinnvollerweise hat sie aber mindestens drei Stufen oder

Teilnehmer, um als Lieferkette zu gelten. Die Kunden der Supply Chain sind in der stufenweisen Betrachtung inbegriffen. Jede Stufe einer Supply Chain kann wiederum aus mehreren Unternehmen bestehen. Die Unternehmen dieses Netzwerkes müssen dabei nicht zwangsläufig alle dem produzierenden Gewerbe angehören (vgl. Corsten und Gössinger, 2008). Eine Supply Chain kann beispielhaft auch aus einem fertigen Unternehmen und mehreren Händlern bestehen. Das Produkt wird dann von Produzenten über Großhändler an Einzelhändler geliefert. Am Ende jeder Supply Chain stehen dann aber Kunden, die die erstellten Produkte konsumieren. Neben dem Güterfluss spielt der Informationsfluss eine tragende Rolle. Die Informationsströme zwischen den Mitgliedern der Supply Chain erfolgen zu einem Teil parallel zum Warenstrom, zum anderen Teil dem Güterstrom entgegengesetzt. Der parallele Informationsstrom dient unter anderem zur Auftragsabwicklung, während der entgegengesetzte Fluss die Informationen zu planerischen Aufgaben, wie der Auslösung von Bestellungen, weiterleitet (vgl. Heidtmann, 2008). In der Praxis werden hierfür meistens moderne Informationssysteme und Modelle der elektronischen Datenverarbeitung benutzt (vgl. Thaler, 2007). Zusätzlich zu den Waren- und Informationsströmen fließen zwischen den Mitgliedern des Netzwerkes Finanzströme, die eine Bezahlung der Waren und mögliche Ausgleichszahlungen beinhalten. Die Ausgleichszahlungen sollen Kosteneinsparungen einzelner Supply-Chain-Mitglieder an andere Teilnehmer, die durch den Zusammenschluss schlechter gestellt werden als es bei ihrem Optimum bei individueller Marktpräsenz der Fall ist, weiterleiten. Die Finanzströme verlaufen wie die Informationsströme nicht nur in eine Richtung und integrieren alle Teilnehmer der Supply Chain.

In der folgenden Abbildung 1 wird ein möglicher Aufbau einer Supply Chain mit Waren-, Informations- und Finanzströmen schematisch dargestellt.

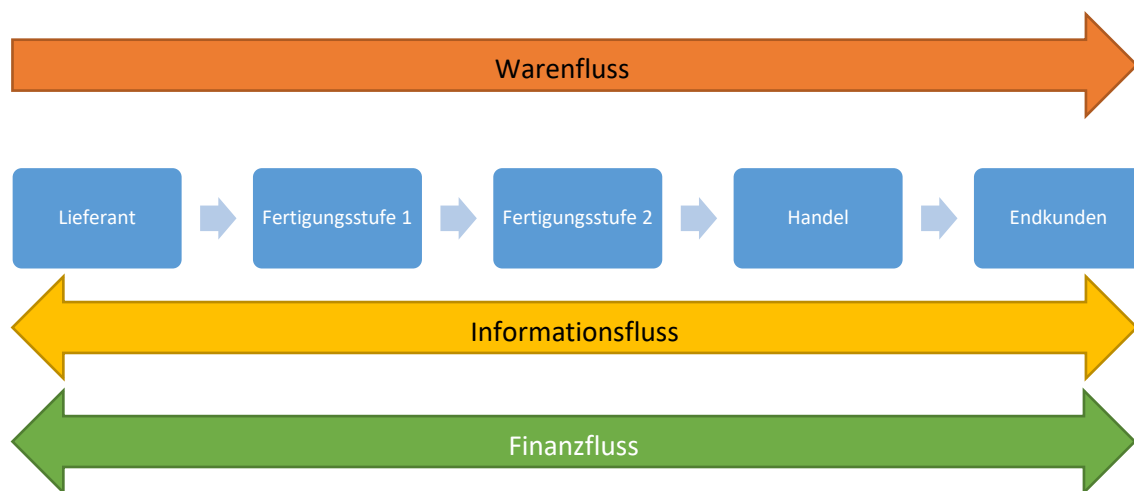


Abbildung 1: Möglicher Aufbau einer Supply Chain, eigene Darstellung nach Gadatsch (2012)

Um eine ganzheitliche Abstimmung der Mitglieder zu gewährleisten und einen reibungslosen Waren-, Informations- und Finanzstrom zu garantieren, ist eine dauerhafte Koordination der Supply Chain unerlässlich. In diesem Zusammenhang zeigt der Peitschenschlageffekt deutlich, dass ein unternehmensübergreifendes Management und eine umfassende Informationsweitergabe zentrale Punkte einer Supply Chain sind und diese Punkte einen großen Teil des Erfolgs ausmachen. Die Aufgabe der Steuerung kommt dem sogenannten Supply Chain Management

(SCM) zuteil. In der Literatur existiert keine allgemein anerkannte Definition des SCM. Für Corsten und Gössinger (2008) ist beispielsweise das Supply Chain Management ein Konzept, das unternehmensübergreifend auf eine optimale Gestaltung der Gesamtprozesse abzielt. Demgegenüber definiert Werner (2013) das Supply Chain Management als interne wie netzwerkgerichtete integrierte Unternehmensaktivität von Versorgung, Entsorgung und Recycling, inklusive begleitender Geld- und Informationsflüsse. Das zeigt, dass die Auffassungen des SCM in der Literatur weit auseinander gehen können. Allerdings sind sich viele Autoren in den Punkten einig, wie das Aufgabenspektrum dieses Managements aussieht und das dieses durch ein koordiniertes Zusammenarbeiten der Mitglieder erfüllt wird. Eine der Hauptaufgaben des SCM ist Abschwächung bzw. Vermeidung des Peitschenschlageffekts (vgl. Arndt, 2008). Dies wird unter anderem durch einen verbesserten Informationsaustausch erreicht. Eine weitere wichtige Aufgabe fällt der Produktions- und Losgrößenplanung zu, welche durch den Peitschenschlageffekt beeinflusst werden kann. Diese Planungen zielen darauf ab, die Produktions- und Losgrößen optimal zu gestalten, um die Gesamtkosten für die Supply Chain zu senken. Dazu ist es nötig, dass Informationen der verschiedenen Stufen verarbeitet werden. Oftmals steht aber eine Supply Chain vor dem Problem, dass einzelne Mitglieder opportunistisch handeln und wenige Informationen an andere Mitglieder preisgeben wollen (vgl. Müller, 2005). Dem muss das SCM zur Erfüllung seiner Aufgabe und zum Erfolg der gesamten Kette entgegenwirken. Die Mitglieder einer Supply Chain sind meistens über Verträge miteinander verbunden, die eine langfristige Zusammenarbeit sicherstellen sollen (vgl. Heidtmann, 2008). Durch die Verträge sollen auch Abhängigkeiten geschaffen werden, um opportunistisches Verhalten der beteiligten Akteure zu unterbinden (vgl. Corsten und Gössinger, 2008). Das Supply-Chain-Management hat außerdem die Aufgabe, die Transporte von Waren zwischen den Mitgliedern und zu den Kunden zu planen und zu koordinieren, also den Warenstrom übergeordnet zu verwalten. Dabei wird das Ziel verfolgt, die Transportkosten möglichst gering zu halten und Fehler zu unterbinden. Verschiedene Varianten zur Belieferung von Kunden werden in Abschnitt 2.2 diskutiert.

Eine Supply Chain ist also ein Unternehmensnetzwerk, das durch festgelegte Lieferbeziehungen einen Wertschöpfungsprozess vollzieht. Durch kooperative Zusammenarbeit sollen die Kunden der Supply-Chain optimal beliefert werden, weshalb ein unternehmensübergreifendes Management zur Planung und Organisation des Zusammenschlusses eingesetzt wird.

2.2 Ausgewählte Transportmethoden einer Supply Chain

Die Versorgung von Unternehmen einer nachfolgenden Stufe oder von Kunden innerhalb der Supply Chain kann auf verschiedene Weisen erfolgen. Dabei ist zunächst zwischen den grundlegenden Strukturen eines Netzwerkes zu unterscheiden. Die Struktur entscheidet, welche Form der Hauptlauf in einem Transportnetz annimmt. Als Hauptläufe werden die Fernverkehre bezeichnet, durch die ein überregionaler Transport von Waren erfolgt. Die Nahverkehre, die einem Hauptlauf vor- und nachgelagert sind, werden dementsprechend als Vor- und Nachläufe definiert. Gängige Varianten der Struktur sind Direktverkehrsnetze und Hub-and-Spoke-Netze (Speiche und Nabe Netze). Wie in Abbildung 2 dargestellt, sind Direktverkehre Transporte einer

Sendung, bei denen im Hauptlauf zwischen verschiedenen Depots kein Umschlag und damit keine Änderung des Verkehrsmittels erfolgt (vgl. Dabidian und Langkau, 2013). Solche Verkehre werden auch als Komplettladungstransport bezeichnet. Neben dem Komplettladungstransport gibt es den Teilladungstransport, welcher einen umschlagfreien Transport mehrerer Sendungen umfasst. Hub-and-Spoke-Netze sind dem Direktverkehrsnetz gegenüber durch unterbrochene Hauptläufe gekennzeichnet. Die Verbindung zwischen Depots wird durch mindestens einen Hub unterbrochen (vgl. Buchholz, 1998). Hubs sind Hauptumschlagbasen, in denen die Transportmittel gewechselt werden können, während Depots einfache Verteillager sein können. Der Einsatz von Hub-and-Spoke-Netzen lohnt sich im Vergleich zu Direktverkehrsnetzen bei schwach frequentierten Verbindungen zwischen einzelnen Depots, da hier eine Bündelung der Sendungen im Gegensatz zu einem Direkttransport wirtschaftlich sinnvoller ist. Direktverkehrsnetze werden dafür häufig angewendet, wenn viele Transporte zwischen den Depots gefahren werden müssen (vgl. Dabidian und Lankau, 2013).

Die nachfolgende Abbildung veranschaulicht den Unterschied zwischen einem Direktverkehrsnetz und einem Hub-and-Spoke-Netz.

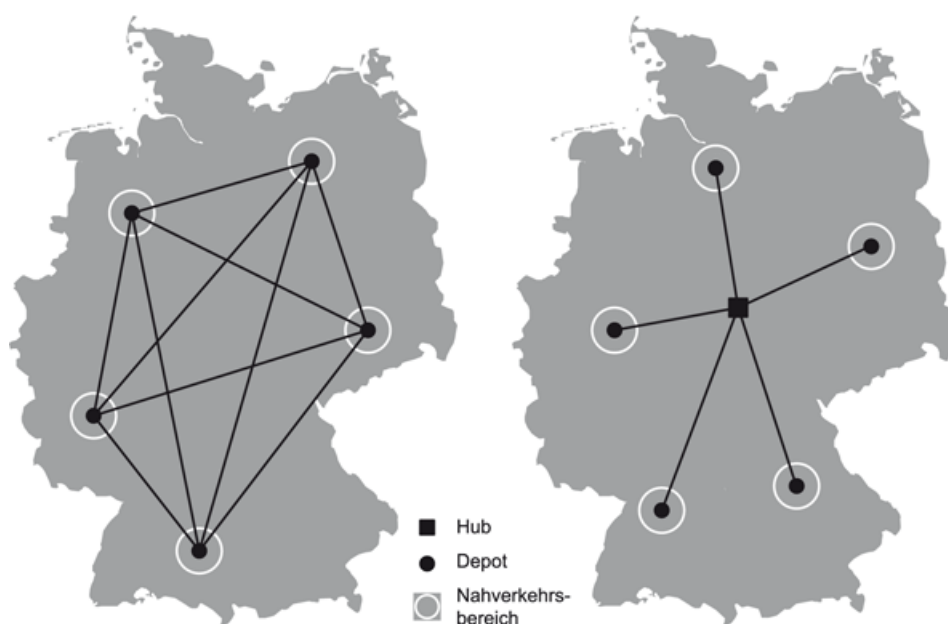


Abbildung 2: Direktverkehrsnetz und Hub-and Spoke-Netz nach Dabidian und Langkau (2013)

Während die Struktur eines Transportnetzes den Hauptlauf definiert, werden Vor- und Nachläufe durch die Touren- und Routenplanung bestimmt. Da sich in den Nahverkehrsbereichen der Depots meistens mehrere Organisationen befinden, von denen Waren einzusammeln sind oder an die Waren auszuliefern sind, erfolgt eine Planung, welche Standorte in einer Tour beliefert werden. Die Tourenplanung wird meistens täglich vorgenommen, um abhängig von dem Auftragsvolumen die optimale Tour fahren zu können (vgl. Metzler, 2013). Dazu können verschiedene Algorithmen und Verfahren verwendet werden, die man in Eröffnungs- und Verbesserungsverfahren unterteilen kann. Eröffnungsverfahren finden eine zulässige Lösung, welche von Verbesserungsverfahren verfeinert werden soll (vgl. Eufinger und Tesch, 2013). In

einem Eröffnungsverfahren können prinzipiell erst die abzufahrenden Standorte zu verschiedenen Touren zusammengefasst und dann die Reihenfolge der Belieferung bestimmt werden. Alternativ können Methoden zum Einsatz kommen, die beide genannten Schritte zusammenfassen. Ein Beispiel für ein solches Verfahren ist das Sweep-Verfahren. Das Sweep-Verfahren ist eine graphische Lösungsmöglichkeit zur Tourenplanung, welches ohne viele Informationen angewendet werden kann. Hierbei werden das Depot, von dem die Tour ausgehen soll, und die Kundenstandorte in ein Koordinatensystem eingeordnet. Als Nullpunkt des Systems wird das Depot verwendet, um das alle Kundenstandorte äquivalent zur Realität eingetragen werden. Beginnend mit der als Sweepline bezeichneten X-Achse werden die Kunden gegen den Uhrzeigersinn abgegangen. Die Kunden werden dabei gemäß der Größe ihres Winkels zum Ursprung sortiert. Eine Tour beginnt mit dem Kundenstandort, der den kleinsten Winkel besitzt und endet, wenn eine vorher festgelegte Restriktion erfüllt wird. Bei Domschke und Scholl (2010) wird diese Restriktion durch eine Zeitbeschränkung einer Tour dargestellt. Es eignen sich aber auch andere Möglichkeiten, wie z.B. eine maximale Anzahl an Standorten, die eine Tour umfassen darf. Falls zu dem Zeitpunkt noch nicht alle Standorte in einer Tour eingeschlossen sind, wird eine zweite Tour gebildet. Die zweite Tour beginnt mit dem Kundenstandort, welcher den kleinsten Winkel besitzt, der noch nicht Teil einer Tour ist. Das Verfahren endet, sobald alle Standorte in Touren eingebunden sind.

Das Ergebnis von solchen Algorithmen stellen sogenannte Milk Runs dar. Milk Runs sind Touren, die ausgehend von einem Umschlag- oder Verteillager mehrere Standorte in einer vorher geplanten Reihenfolge abfahren. Im Gegensatz zu Direkttransporten bieten Milk Runs Konsolidierungsmöglichkeiten, d.h. es ist durch einen Zusammenführen von kleineren Sendungen eine bessere Auslastung der Transportkapazität möglich. Deshalb werden besonders im Nahverkehrsbereich Milk Runs zum Sammeln oder Verteilen von Sendungen eingesetzt (vgl. Schulte, 2013). Je nach Organisation bestehen in der Tourenplanung unterschiedliche Zielsetzungen. Mögliche Ziele sind die Minimierung von Gesamtkosten, Fahrzeiten, Entfernungen oder Touren (vgl. Eufinger und Tesch, 2013). Die Tour selber kann dabei durch unterschiedliche Faktoren beschränkt werden. Dazu zählt die Beschränkung durch eine vorgegebene Maximalzeit oder durch eine Menge, die maximal transportiert werden darf. Letzteres wird durch die größtmögliche Kapazität des eingesetzten Fahrzeuges bestimmt. Besondere Bedeutung hat auch der maximale Radius einer Tour, welcher nicht über 100 km betragen sollte, da der Milk Run sonst unwirtschaftlich wird (vgl. Metzler, 2013). Weitere Faktoren sind Vorgaben zu der maximal zu fahrenden Strecke oder maximalen Anzahl an Standorten einer Tour. Nach Schulte (2013) sind besonders die Tourendichte, also der durchschnittliche Abstand zwischen zwei zu beliefernden Orten und das Auslieferungsvolumen pro Stopp relevant. Demnach sollten bei einem Milk Run die kürzesten Wege zwischen den Standorten gewählt werden, um effektive Versorgung zu gewährleisten.

Zur Belieferung können seitens des Kunden auch zeitliche Vorgaben wie z.B. Just In Time festgelegt werden. Just In Time ist ein Konzept, bei dem die benötigten Waren genau zur richtigen Zeit an einem vorher bestimmten Punkt angeliefert werden müssen. Der große Vorteil des Konzepts ist die Vermeidung von Puffern und Lagern. Allerdings wird aufgrund der Voraussetzungen für das Just In Time Prinzip, wie hohe Termintreue und hohe Ausfallsicherheit,

dieses vorwiegend in der Produktion höherwertiger Konsumgüter, z.B. bei der Produktion von Autos, verwendet (vgl. Gudehus 1, 2007).

Zusammenfassend kann man sagen, dass Supply Chains abhängig von dem Sendungsaufkommen unterschiedliche Transportstrukturen zur überregionalen Belieferung von Depots besitzen. Im Nahverkehrsbereich wird aber häufig auf Milk Runs zur Sammlung oder Auslieferung von Waren gesetzt, da diese bei kleineren Sendungen eine bessere Auslastung der Transportkapazität versprechen.

2.3 Aufbau eines Simulationsmodells

Um zu überprüfen, ob ein reales oder geplantes System gut funktioniert, muss es hinreichend getestet werden. Ein System ist in diesem Fall eine Menge miteinander in Beziehung stehender Elemente, die in einem bestimmten Zusammenhang als Ganzes gesehen und als von ihrer Umgebung abgegrenzt betrachtet werden (vgl. DIN Norm, 2014). Für das Testen des Systems ist es theoretisch möglich, am realen System Experimente vorzunehmen, die im Rahmen von aufwendigen Messungen und Erhebungen durchgeführt und dokumentiert werden. In der Praxis ist dies aber nicht unbedingt sinnvoll, da solche Untersuchungen viel Zeit und Geld kosten und den normalen Arbeitsablauf stark beeinträchtigen können. Außerdem können diese Untersuchungen zu umfangreich sein, um in einer vorgegeben Zeit überhaupt zu einem Ergebnis zu kommen. Als Alternative zu solchen Erhebungen bietet sich die Simulation an (vgl. Bungartz et al., 2013). Die Simulation ist eine Problemlösungsmethode, bei der durch Experimente mit Simulationsmodellen Aussagen über das Verhalten der durch die Modelle beschriebenen Systeme gewonnen werden (Rabe et al., 2017). Der Verein Deutscher Ingenieure (VDI, 2014) definiert die Simulation genauer als ein Nachbilden eines dynamischen Prozesses in einem System mit Hilfe eines experimentierfähigen Modells, um zu Erkenntnissen zu gelangen, die auf die Wirklichkeit übertragbar sind, mit der Ergänzung, dass die Prozesse über die Zeit entwickelt werden. Es gibt also keine eindeutige Definition der Simulation, aber in der Literatur ist man sich weitestgehend über entscheidende Merkmale einig, die die Simulation charakterisieren. Demnach ist die hier betrachtete Simulation durch drei wesentliche Punkte gekennzeichnet:

- Erstellen eines Modells für das betrachtete System
- Experimente an diesem Modell durchführen
- Rückschlüsse aus diesen Experimenten für die Anwendung auf das System in der Realität ziehen

Der oft genannte Zusatz der Entwicklung über die Zeit ist für die Simulation nicht allgemeingültig, da es auch Varianten gibt, die die Zeit nicht berücksichtigen, z.B. die sogenannte Monte-Carlo-Simulation (vgl. Rabe et al., 2017). Das Modell ist in diesem Fall eine abstrakte Abbildung des realen Systems mit seinen für die Simulation relevanten Prozessen und Eigenschaften. Es stellt dabei die Grundlage für die Experimente dar (vgl. Bungartz et al. 2013). Da die Simulation computerbasiert abläuft, wird für diese Methode ein ausführbares Computermodell benötigt, das aus einem formalen Modell entwickelt wird. In dem Kontext definiert der VDI (2014) das Experimentieren in Simulationen als gezielte empirische Untersuchung des Modellverhaltens auf Basis wiederholter Simulationsläufe, wobei das

Simulationsmodell systematisch hinsichtlich seiner Parameter oder seiner Struktur variiert werden kann. Eine Simulation hat demnach mehrere Durchläufe, bei denen der Aufbau oder die Parameter von Durchlauf zu Durchlauf verändert werden können.

Der VDI hat in der Richtlinie 3633 Blatt 1 (2014) einen prinzipiellen Ablauf einer Simulation dargestellt. Der generelle Ablauf der Simulation entspricht hierbei einem Kreislauf, der aus vier wesentlichen Punkten bzw. Schritten besteht. Ausgehend von dem realen System wird in Schritt 1 durch eine Abstraktion ein formales Simulationsmodell erstellt. Anhand dieses formalen Modells werden in Schritt 2 Experimente vorgenommen, die formale Ergebnisse liefern. Die Ergebnisse müssen nun in Schritt 3 interpretiert werden, damit man Folgerungen für das reale System erhält. Diese können dann in Schritt 4 auf das System in der Wirklichkeit angewendet werden. Alternativ zu dem Schritt kann man aber auch Parameter variieren und auf das formale Modell anwenden. Dann würde man ab Schritt 2 den Kreislauf erneut durchlaufen und andere Ergebnisse erzielen. Simulationsmodelle können aufgrund ihres Verhaltens in Hinblick auf Veränderungen ins statisch und dynamisch differenziert werden. Statische Systeme unterliegen im Gegensatz zu dynamischen Systemen keiner Betrachtung der Zeit (vgl. Waldmann und Helm, 2016). Dynamische Systeme und Modelle werden anhand ihrer Variablenfortschreibung über die Zeit weiter in kontinuierliche und diskrete Modelle unterschieden (vgl. Rabe et al., 2017). Kontinuierliche Modelle besitzen kontinuierliche, d.h. fortlaufende Funktionen der Simulationszeit, sie haben also keine klar voneinander abgegrenzten Objektzustände. Demgegenüber haben diskrete Modelle klar abgegrenzte Objektzustände, da die Variablen ihre Werte nur zu bestimmten, diskreten Zeitpunkten, sogenannten Ereignissen oder Events, ändern können (vgl. Liebl, 1995). Beide Modelle werden weiter in deterministische und stochastische Modelle unterschieden. Bei stochastischen Modellen unterliegen die Ereignisse einer stochastischen Verteilung und können daher nicht vorhergesagt werden. Andernfalls bedeutet dies ein deterministisches Modell. Diskrete Modelle haben im Gegensatz zu der kontinuierliche Simulation eine übergeordnete Rolle bei der Analyse in der Produktion und der Logistik (vgl. Rabe et al., 2017). Diese Simulationsmodelle können zusätzlich noch in zeitdiskret und ereignisdiskret aufgeteilt werden. Während bei der ersten Variante festgelegte Zeitabstände zur Zustandsänderung führen, treten Änderungen bei der zweiten Variante zu unregelmäßigen Zeitpunkten ein. Die folgende Abbildung 3 veranschaulicht eine derartige Unterteilung eines Systems in seine verschiedenen Modellformen.

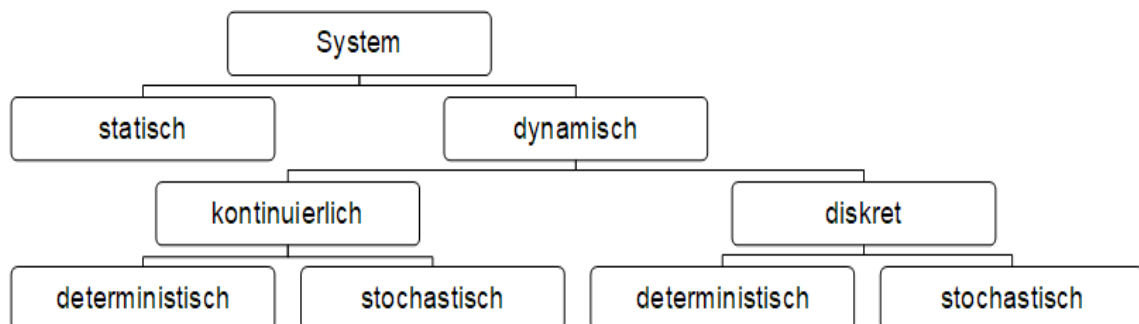


Abbildung 3: Aufteilung eines Systems nach Hedtstück (2013)

Die Simulation wird durch verschiedene wissenschaftliche Bereiche wie der Informatik und der Statistik bis zum Operations Research und der Logistik hin beeinflusst. Die Statistik und die Informatik sind in diesem Zusammenhang von besonderer Bedeutung, da die computerbasierte Simulation auf statistischen Verteilungen und der damit verbundenen Generierung von Zufallszahlen aufbaut (vgl. Waldmann und Helm, 2016). Die Simulation kann ihre Anwendung in vielen verschiedenen Fällen finden. Falls z.B. ein neu entwickeltes System untersucht werden soll oder die Komplexität des System die Grenzen analytischer Methoden übersteigen und die menschliche Vorstellungskraft überfordern würde, bietet die Simulation Lösungen für diese Probleme (vgl. Goedicke, 2013). Die eingangs erwähnte Möglichkeit, dass Untersuchungen am realen System nicht möglich sind, z.B. weil sie einen erheblichen Eingriff in den Tagesablauf darstellen, ist ein weiterer häufig vorkommender Anwendungsgrund für die Simulation. Dabei führt eine Simulation zu einer Reihe von Vorteilen, die für ihre Nutzung sprechen. So kann die Simulation zu einer höheren Planungssicherheit und zu einer höheren Qualität der Planungsergebnisse führen, da eventuell auftretende Fehlentwicklungen frühzeitig erkannt werden können (vgl. Kuhn et al., 1993). Für einfache Mitarbeiter aber auch für Manager kann durch eine Simulation und eine Visualisierung oder Animation der Ergebnisse ein besseres Verständnis für das betrachtete System und seine Abläufe erzielt werden (vgl. Rabe et al., 2017). Durch das Vereinfachen von Systemelementen als Konsequenz aus den Simulationsergebnissen kann diese Methode auch zur Reduktion der Gesamtkosten beitragen. Außerdem kann die Simulation bei der Optimierung von Arbeitsabläufen oder von Einflussgrößen, wie z.B. Lagerbeständen, helfen. Allerdings ist die Simulation selber kein Optimierungsverfahren. Sie ist ein Hilfsmittel für Planungs- und Entscheidungsaufgaben, das durch Experimente eine Bewertung eines Systems, aber keine Lösung für Probleme, bereitstellt (vgl. Gudehus 1, 2007). Der Planer oder der Anwender der Simulation muss selbst Konzepte erarbeiten, um eine Simulation aufbauen und damit arbeiten zu können. Auf Basis der Simulationsergebnisse muss der Planer dann auch eigenständig die Konzepte und das System weiterentwickeln und optimieren. Im Vorfeld des Einsatzes einer Simulation sollten immer die Zeit- und Kostenrahmen berücksichtigt werden, da eine Simulation zwar viele Vorteile bringt, aber nicht in jedem Fall nötig oder aus finanzieller Sicht sinnvoll ist.

Es kann vorkommen, dass eine Simulationsstudie nicht unternehmensintern durchgeführt wird, sondern an externe Experten dieses Fachgebiets vergeben wird. Vor allem in so einem Fall hat es große Priorität bereits zu Beginn der Studie ausführlich mit allen Beteiligten zu kommunizieren und viel Arbeit zu investieren, um Fehler in der Planung der Studie auszuschließen oder frühzeitig verbessern zu können. Die Kosten der Fehlerbehebung können sich in einer Simulationsstudie um den Faktor zehn pro Entwicklungsstufe erhöhen, in der sie unentdeckt bleiben (vgl. Rabe et al., 2008). Die genaue Abstimmung zwischen Auftraggeber und Planer der Studie ist daher von zentraler Bedeutung. Hier eignet sich der Einsatz eines Vorgehensmodells zur strukturierten Arbeit. Dabei obliegt es dem Anwender bzw. Planer der Simulationsstudie, welches Vorgehensmodell zur Durchführung genutzt werden soll. Entscheidend für ein nachvollziehbares Vorgehen ist, dass innerhalb der Simulationsstudie überhaupt ein Vorgehensmodell angewendet wird und dass nach diesem Modell strukturiert gearbeitet wird. Das Vorgehensmodell nach Rabe et al. (2008) besteht aus fünf Phasen, die iterativ durchlaufen werden sollen (vgl. Abbildung 3).

Entlang dieser Phasen sollten Dokumente angefertigt werden, die Zwischenergebnisse und wichtige Informationen festhalten. Bevor die eigentlichen Phasen durchlaufen werden, muss eine Zielbeschreibung ausgearbeitet werden, die die Aufgabe aus Sicht des Auftraggebers beschreibt. Danach wird in der ersten Phase eine Aufgabendefinition erarbeitet. Diese soll nach Meinung aller Beteiligten das zu lösende Problem beschreiben und die vorgesehenen Rahmenbedingungen zu der Studie fixieren. Aus der folgenden Systemanalyse entsteht ein Konzeptmodell, welches den Übergang von der Problemformulierung zu der Beschreibung, wie dieses Problem konkret zu lösen ist, darstellt. Mit der Modellformalisierung entsteht aus dem Konzeptmodell ein formales Modell, das einen tieferen Detaillierungsgrad besitzt. Durch die Implementierung erhält man dann ein ausführbares Modell, welches wiederum detaillierter als sein Vorgängermodell ist. Die fünfte Phase besteht aus Experimenten und Analysen. Diese führen zu Simulationsergebnissen aus denen Folgerungen für das reale System abgeleitet werden können. Parallel zu den fünf Phasen müssen Daten beschafft und aufbereitet werden, damit die Simulation eine Datengrundlage hat, auf der sie aufbauen kann. Die Datenbeschaffung aus existierenden Datenquellen (Sekundärerhebung) oder durch Messungen und Beobachtungen (Primärerhebung) führt zunächst zu Rohdaten, die nachfolgend durch verschiedene Methoden, wie z.B. der Ergänzung oder Eliminierung von Daten, aufbereitet werden müssen. Die Aufbereitung der Daten erfolgt unter dem Aspekt, dass nicht alle gesammelten Werte relevant sind oder der Richtigkeit entsprechen. Über den gesamten Zeitraum der Simulationsstudie besitzen die Aspekte der Verifikation und der Validierung große Wichtigkeit. Sie sollen sicherstellen, dass die aus der Simulation erhaltenen Ergebnisse für die Beantwortung der gegebenen Fragestellung verwendet werden dürfen (Rabe et al., 2017). Hier sind die Fragen „Ist das Modell richtig?“ (Verifikation) und „Ist es das richtige Modell (im Hinblick auf die Aufgabenstellung)?“ (Validierung) zu beantworten (vgl. Balci, 2003). Dazu werden Tests durchgeführt, die einerseits auf die Gültigkeit des Modells als Phasenergebnis und andererseits auf den Prozess zur Erstellung dieses Ergebnisses abzielen (vgl. Rabe et al., 2008).

In der nachfolgenden Abbildung werden die einzelnen Arbeitsschritte einer Simulationsstudie nach Rabe et al. (2008) übersichtlich dargestellt.

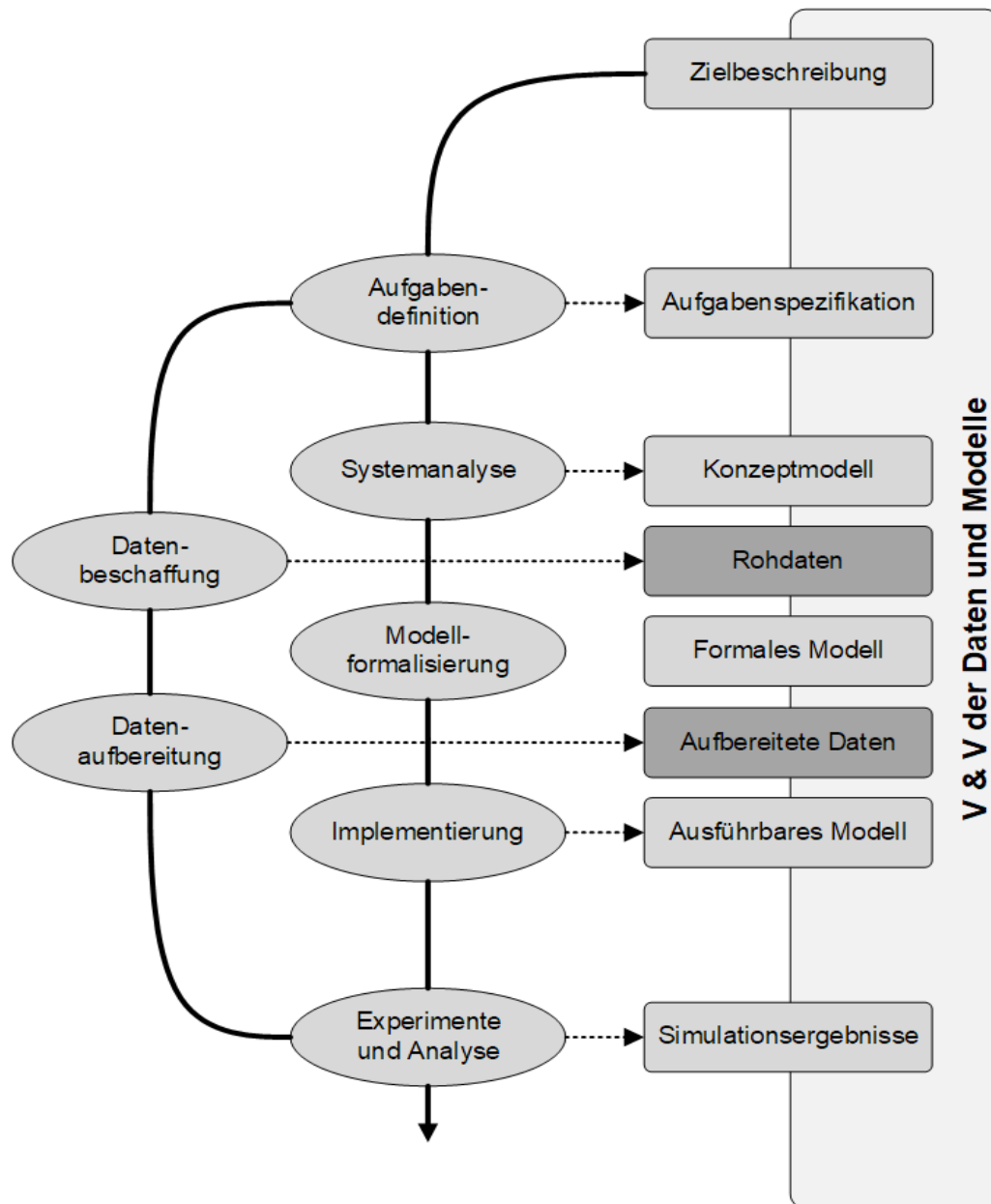


Abbildung 4: Vorgehensmodell bei Simulationsstudien nach Rabe et al. (2008)

Für die Simulation kann also festgehalten werden, dass sie bei verschiedenen Problemstellung, wie der Bewertung von Systemen, helfen kann, aber stets eine Interpretation der Ergebnisse voraussetzt, da sie kein Optimierungsverfahren im herkömmlichen Sinn ist. Während einer Simulationsstudie sollte immer ein Vorgehensmodell zur Strukturierung der Arbeit angewendet werden.

2.4 Distanzmessung in der Ebene

Die Entfernung zwischen zwei Punkten ist für viele Problembehandlungen von entscheidender Bedeutung. So ist z.B. in der Tourenplanung die Distanz zwischen verschiedenen Standorten maßgebend zur Festlegung der Route und folglich der Tour. Dieser Umstand ist später relevant für die Entwicklung von Verfahren zur Standortapproximation in dieser Arbeit. Um für diese Überlegung eine mathematische Grundlage zu haben, sollen verschiedene Arten der

Entfernungsmessung aufgezeigt werden. In der Literatur wird in den meisten Fällen mit der Entfernung zwischen zwei Punkten die kürzeste Distanz zwischen besagten Punkten bezeichnet. Dabei gilt, ist für ein vorliegendes Problem die Distanz nicht vorgegeben, wie es z.B. in Graphen häufig der Fall ist, so muss diese ermittelt werden. Dafür wurden in der Wissenschaft eine Reihe verschiedener Möglichkeiten entwickelt. Im Folgenden sollen die wichtigsten Vertreter zur Entfernungsmessung vorgestellt werden. Dabei soll sich auf die Distanzmessung in der Ebene beschränkt werden. Metriken für dreidimensionale Fälle oder die Distanzmessung in Graphen werden in dieser Arbeit nicht betrachtet. Da sich in der Literatur die Variablenbezeichnungen je nach Autor und Quelle unterscheiden können, wurden diese für eine bessere Vergleichbarkeit der Formeln entsprechend angepasst. Die Punkte i und j , ausgedrückt durch ihre jeweiligen Ausprägungen x_i, y_i und x_j, y_j , bezeichnen zwei beliebige Punkte innerhalb einer Ebene. Ein geographischer Punkt wird durch seine Koordinaten eindeutig definiert. Die Werte der Ausprägungen lassen sich deshalb durch die geographische Breite (engl. Latitude) und die geographische Länge (engl. Longitude) bestimmen. Eine Ebene bedeutet, dass in der betrachteten Fläche keine Erhebungen oder Tiefen Berücksichtigung finden (vgl. Duden, 2018). Daher lässt sich eine Ebene mit einem 2-dimensionalen Koordinatensystem vergleichen. Diese Eigenschaft macht sich die rechtwinklige Entfernungsmessung zu Nutze. Die rechtwinklige Entfernung entspricht den Abständen zweier Punkte, wenn man parallel zu den Koordinatenachsen vorgehen würde. In der Mathematik ist das die Summe des Abstandes zweier Punkte in x -Achsenrichtung und des Abstandes derselben Punkte in y -Achsenrichtung. Übertragen bedeutet das, die Größe

$$d_{i,j}^1 := |x_i - x_j| + |y_i - y_j| \quad (1)$$

bezeichnet die rechtwinklige Entfernung zwischen zwei Punkten i und j in der Ebene (vgl. Domschke und Drexl, 1996). In Manhattan entspricht durch die dortige schachbrettartige Straßenanordnung eine Fahrtstrecke zwischen zwei Punkten dieser Metrik. Daraus folgt der geläufige Beiname Manhattan-Distanz. Diese Entfernung wird vor allem bei der innerbetrieblichen Layoutplanung und bei der innerstädtischen Standortplanung verwendet (vgl. Thonemann, 2015). Bei Entfernungsberechnungen außerhalb von Städten ist diese Metrik weniger sinnvoll, da die Abweichung zu einer realen Fahrtstrecke zum Teil sehr groß ausfallen kann. Eine andere, sehr verbreitete Entfernungsmessung ist die euklidische Distanz. Sie entspricht der Luftlinienentfernung zwischen zwei Punkten. In einem Koordinatensystem wäre es die direkte Verbindungslinie zwischen zwei Punkten. In einem rechtwinkligen Dreieck stellt diese Metrik die Hypotenuse dar. Mathematisch ausgedrückt bezeichnet die Größe

$$d_{i,j}^2 := \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2} \quad (2)$$

die euklidische Distanz zwischen zwei Punkten i und j in einer Ebene. Die euklidische Distanz hat keine Beschränkungen in ihrer Anwendung, aber aufgrund der Tatsache, dass die

Luftlinienentfernung nicht für jeden Fall passend ist, ist eine unterschiedliche Güte hinsichtlich des Ergebnisses möglich. So kann die tatsächliche Fahrtstrecke in innerstädtischen Gebieten aufgrund der Bebauung deutlich von der Luftlinienentfernung abweichen. Nach Thonemann (2015) wird diese Metrik daher besonders in der außerstädtischen Standortplanung angewendet, da sich mit zunehmender Entfernung die Fahrtstrecke der Luftlinie annähert. Die euklidische Entfernung ist Grundlage für diverse Variationen. Wird z.B. in der ursprünglichen Formel unter der Wurzel ein beliebig kleiner positiver Wert ϵ addiert, erhält man die modifizierte euklidische Entfernung. Diese wird häufig im Rahmen von Lösungsverfahren der Standortplanung verwendet (vgl. Domschke und Drexl, 1996). Die Addition des Wertes ϵ soll Abweichungen miteinbeziehen, und so die Metrik genauer machen. In anderer Form wird dies durch die gewichtete euklidische Distanz ebenfalls versucht. Für eine genauere Entfernungsmessung als die Luftlinie zwischen zwei Punkten wird hier ein Faktor f_{umw} mit der einfachen euklidischen Distanz multipliziert. Die Größe

$$d_{i,j}^3 := f_{umw} \times \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2} \quad (3)$$

entspricht der gewichteten euklidischen Entfernung zwischen zwei Punkten i und j in der Ebene. Der sogenannte Umwegfaktor f_{umw} gibt dabei die mittlere Abweichung der Fahrtstrecke von der Luftlinie an (vgl. Gudehus 2, 2007), d.h. den Mehrweg bei einer Fahrt über die kürzeste Strecke von einem zum anderen Punkt im Verhältnis zur Luftlinie. Gudehus (2, 2007) verwendet für den Umwegfaktor einen Wert von 1,3. Statt f_{umw} kann auch ein beliebiger anderer Wert g für die Gewichtung der euklidischen Distanz verwendet werden. In der Literatur werden noch weitere mögliche Arten der Distanzmessung untersucht, z.B. die Minkowski-Distanz, welche aber in dieser Arbeit nicht vorgestellt werden sollen. Diesbezüglich sei unter anderem auf die Arbeit von Thonemann (2015) hingewiesen, der im Rahmen der Betrachtung von Standortverfahren verschiedene Metriken vorstellt. Im selben Kontext zeigen auch Domschke und Drexl (1996) Arten der Entfernungsmessung auf.

Bei der Berechnung von Distanzen müssen diverse Faktoren beachtet werden, um Fehler zu vermeiden und mit den Entfernungen problemlos weiterarbeiten zu können. Dazu gehört die Berücksichtigung der Einheiten, in denen die Werte zur Berechnung vorliegen. Die Einheiten des Inputs bestimmen dabei die Einheit des Outputs. Bei der Entfernungsberechnung mit Koordinaten ist eine Umrechnung der Werte des Inputs erforderlich, um ein Ergebnis in der gewünschten Einheit Kilometer zu erhalten. Ohne eine Umrechnung würde das Ergebnis eine Dezimalzahl eines Koordinatenunterschieds sein, die zum einen stark verzerrt wäre und zum anderen nicht als Maß für eine Entfernung geeignet ist. Die Verzerrung tritt auf, weil durch die Erdkrümmung die geographischen Längen und Breiten unterschiedliche Abstände zueinander haben (vgl. de Lange, 2013). Je nachdem, welche Quelle untersucht wird, erhält man verschiedene Angaben zu den Abständen zwischen den Längen- und Breitengraden. OpenStreetMap gibt in seiner Info an, dass der Abstand allgemein 111,2 km zwischen den Graden beträgt. Laut komp.f.de haben die Breitengrade einen Abstand von 111,3 km und die Längengrade in Deutschland einen

durchschnittlichen Abstand von 71,5 km zueinander. Ähnliche Werte gibt de Lange (2013) auch an, allgemeingültige Werte wurden aber nicht gefunden. Für die nachfolgenden Berechnungen wird deshalb ein Mittelwert der gefundenen Daten als Grundlage genommen. Es wird angenommen, dass der Abstand zwischen den Breitengraden 111,2 km und der Abstand zwischen den Längengraden 71,5 km beträgt. Um nun eine euklidische Distanz zwischen zwei durch Koordinaten definierten Punkten in Kilometer umzurechnen, werden die Differenzen unter der Wurzel mit dem jeweiligen Faktor multipliziert. Die Differenz der Breitengrade wird also mit 111,2 multipliziert, die der Längengrade dementsprechend mit 71,5 km. Dadurch wird eine Verzerrung der Distanz vermieden und das Ergebnis hat mit Kilometer eine Einheit, die ideal zur Entfernungsdarstellung ist.

Zusammenfassend ist festzuhalten, dass es verschiedene Möglichkeiten der Entfernungsmessung gibt, die je nach Situation Vor- und Nachteile haben. Die euklidische Distanz ist für den weiteren Verlauf der Arbeit besonders relevant. Bei der Distanzberechnung zweier durch Koordinaten ausgedrückter Punkte ist die Umrechnung der Einheit wichtig, um Verzerrungen zu vermeiden.

3 Verfahren bei unzureichender Datengrundlage

Nachdem in dem vorherigen Kapitel Grundlagen zu den verschiedenen Themenbereichen dieser Arbeit aufgezeigt und erklärt worden sind, soll nun der Stand der Forschung hinsichtlich der Vorgehensweise bei einer unzureichenden Datenlage dargestellt werden. In erster Linie geht es darum Verfahren zu finden, die angewendet werden, wenn keine oder zu wenig Daten vorhanden sind, um normal weiterarbeiten zu können. Diese Verfahren sollten im Idealfall auch dazu in der Lage sein, für das Problem der Standortapproximation angewendet werden zu können. Deshalb werden diese Verfahren dahingehend untersucht, in wie weit sie auf das vorliegende Problem anwendbar sind bzw. ob man Teile dieser Verfahren für eigene Ansätze weiterverwenden kann. Dabei soll zunächst das Forschungsgebiet der Standortplanung und Standortfindung betrachtet werden. Dabei erfolgt eine Abgrenzung des Begriffes von der hier vorliegenden Standortapproximation. Im darauffolgenden Abschnitt wird ein neues Verfahren vorgestellt, welches speziell für die Annäherung von Standorten bei unzureichenden Daten entwickelt worden ist. Dann wird die Beschränkung auf die Bestimmung von Standorten aufgelöst, sodass allgemeinere Verfahren zur Approximation von Werten begutachtet werden können. Hier werden vor allem verschiedene Möglichkeiten der Statistik zum Umgang mit mangelnden Daten betrachtet. Dabei wird zuerst die Imputation erläutert, bevor allgemeine statistische Schätzfunktionen beschrieben werden. Anschließend wird die Mittelwertberechnung zur Bestimmung von Lageparametern untersucht. Das Kapitel schließt mit der Vorstellung und Bewertung der aus der Informatik stammenden Themenbereiche des Data Farming und Data Mining ab.

3.1 Die Standortplanung im Kontext fehlender Informationen

In den Wirtschaftswissenschaften spielt die Optimierung von Parametern eine große Rolle. Dazu gehört auch den optimalen Standort zur Versorgung anderer unternehmensinterner oder externer Standorte zu finden. Die sogenannte Standortplanung oder auch Standortfindung umfasst hierbei sämtliche betriebswirtschaftliche und informatorische Forschungsansätze, die sich genau mit dieser Problematik auseinandersetzen und ist daher ein Teil des Operations Research. Die Standortplanung lässt sich in diskrete und in kontinuierliche Standortplanung unterteilen (vgl. Domschke und Drexl, 1996). In der diskreten Standortplanung werden aus einer vorausgewählten Menge an potenziellen Standortkandidaten derjenige Standort oder diejenigen Standorte ausgewählt, welche unter betriebswirtschaftlichen Gesichtspunkten ein Optimum für die betrachtete Situation darstellen (vgl. Corsten und Gössinger, 2008). Dafür wird in den meisten Fällen zunächst die Kandidatenanzahl durch harte Standortfaktoren (quantitative Faktoren, wie z.B. gegebene Infrastruktur) und weiche Standortfaktoren (qualitative Faktoren, wie z.B. Lebensqualität für Arbeiter) begrenzt, um danach ein für den Sachverhalt passendes mathematisches Verfahren anzuwenden, welches meistens das Ziel hat, den transportkostenminimalen Standort zu finden (vgl. Thonemann, 2015). Zur Lösung der Verfahren werden eine zu minimierende Zielfunktion und mehrere Nebenbedingungen aufgestellt. In den

Ziel- und Nebenfunktionen erfolgt dann die Betrachtung von verschiedenen messbaren Parametern wie beispielsweise den Transportkosten und den Bedarfen der einzelnen Kunden. Bei diesen Optimierungsmethoden ist es unerheblich, um welche Art von Unternehmen oder Organisationen es sich handelt. Voraussetzung zur Anwendung der diskreten Verfahren sind aber gegebene Standorte und dazugehörige Informationen oder Werte zu bestimmten Parametern. Diskrete Standortplanungsverfahren sind unter anderem die sogenannten Zentren- und Medianprobleme, die ursprünglich aus der Statistik abgeleitet worden sind. Ein weiteres wichtiges Verfahren ist das Warehouse-Location-Problem, welches anhand von fixen und variablen Kosten den optimalen Standort finden soll (vgl. Domschke und Drexl, 1996). Je nach Aufgabenstellung oder Kundenbedarf ist es möglich, einen Standort oder mehrere Standorte zur Versorgung zu bestimmen. Für beide Fälle wurden Varianten der Verfahren entwickelt, die sich in der Formulierung der Funktionen und ihrer Anzahl nur gering vom grundlegenden Modell unterscheiden. Da das Berechnen einer exakten Lösung der gängigen Verfahren zum Teil sehr rechenaufwendig ist, kommen oft heuristische Lösungsmöglichkeiten zum Einsatz (Baudach et al., 2013). Der Vorteil dieser Verfahren ist, dass sie relative simpel anzuwenden sind und für das Problem hinreichend gute Lösungen erzielen.

Demgegenüber steht die kontinuierliche Standortplanung. Hier kommen alle Punkte in einer homogenen Ebene als Standortkandidaten in Frage (vgl. Domschke und Drexl, 1996). Dieses Problem benötigt im Gegensatz zu den diskreten Verfahren der Standortplanung daher keine vorausgewählten Standortkandidaten, sondern nur Werte zu bestimmten Parametern, wie den Transportkosten, anhand derer die Standorte an beliebigen Punkten in der Ebene platziert werden können. Die relevanten Distanzen zwischen den Standorten werden hier mittels verschiedener Metriken berechnet, wie sie auch in Abschnitt 2.4 vorgestellt worden sind. Der Ursprung geht dabei auf das sogenannte Steiner-Weber-Problem zurück, welches die Standortfindung in der Ebene mittels euklidischer Distanzen behandelt.

Verfahren der Standortplanung sind im allgemeinen Optimierungsverfahren, die von einer bestehenden Datengrundlage ausgehen. Das bedeutet diese Verfahren benötigen Informationen mit denen sie arbeiten können. Eine Approximation von Standorten bei fehlenden Daten, wie es Ziel dieser Arbeit ist, unterscheidet sich daher im Datenbedarf grundlegend von den gängigen Verfahren der Standortplanung. Während in dieser Arbeit Standorte in einer Fläche approximativ verteilt werden sollen, um die Realität möglichst genau abzubilden, zielen diese Verfahren darauf ab, durch gegebene Informationen zu Standorten in der Realität ein optimales Ergebnis zu erzielen, also Kosten zu minimieren. Die Standortplanung ist also von Informationen zu Standorten abhängig, während die Standortapproximation ohne Informationen auskommen muss. Trotz des Unterschieds wurde im Rahmen der Literaturrecherche nach Verfahren dieser Disziplin gesucht, die auch bei fehlenden Daten angewendet werden können. Hier ist festzuhalten, dass in der Literatur zur Standortplanung keine Verfahren oder Modelle gefunden wurden, die bei unzureichender Datenlage arbeiten können. Die oben beschriebene Standortplanung soll in dieser Arbeit deshalb nicht weiter betrachtet werden. Stattdessen wird an dieser Stelle auf andere Arbeiten verwiesen. Domschke und Drexl (1996) geben in ihrer Arbeit eine sehr gute Übersicht mit mathematischen Erklärungen über die wichtigsten Verfahren der Standortplanung. Die betriebliche Standortwahl in der Ebene wurde Weber (1909) begründet. Ebenfalls aus

betriebswirtschaftlicher Sicht zeigt Hummeltenberg (1981) Verfahren auf, während Thonemann (2015) eine allgemeinere Herangehensweise an Standortprobleme in der Ebene hat. Ansätze zu den Verfahren werden auch in der Arbeit von Corsten und Gössinger (2001) diskutiert.

3.2 Gleichverteilung von Standorten

Die Problematik unzureichender Daten bei der Bestimmung von Standorten wurde bereits im Vorfeld dieser Arbeit von einem Logistiklehrstuhl der Universität Dortmund untersucht. Das Fachgebiet IT in Produktion und Logistik (ITPL) der TU Dortmund hat sich im Rahmen zu Vorarbeiten einer Simulationsstudie mit der Lokalisierung von Standorten beschäftigt, wenn wie im vorliegenden Fall eine unzureichende Datengrundlage existiert und die Lage der Standorte unbekannt ist. Da keine Erhebungen möglich sein sollten, war das Ziel die Standorte zu approximieren. Aufgrund der begrenzten Datenlage gab es keinen Anhaltspunkt, wo sich die Standorte im jeweiligen Postleitzahlengebiet befinden konnten. Deshalb kam theoretisch jeder Punkt der Flächen als möglicher Standort in Frage. Das ITPL verfolgte daher den Ansatz, die gesuchten Standorte zufällig in die Postleitzahlengebiete zu legen. Dies hatte den Vorteil, ohne großen Aufwand Approximationen für die Standorte zu erlangen. Die Punkte der betrachteten Flächen hatten zueinander keine Unterschiede. Es war also für alle Punkte gleichwahrscheinlich als Standort ausgewählt zu werden. Deshalb sollte die Verteilung der Standorte in den Gebieten mittels einer Gleichverteilung erfolgen. Die Gleichverteilung ist statistische Verteilung, bei der alle möglichen Ausprägungen von Zufallszahlen die gleiche Wahrscheinlichkeit besitzen. Im diskreten Fall wird sie durch die gleiche Wahrscheinlichkeit jedes potentiellen Ergebnisses charakterisiert, während im stetigen Fall die Wahrscheinlichkeitsdichte einen konstanten Wert annimmt. Eine Gleichverteilung bevorzugt also keinen Wert gegenüber einem anderen. Daher bietet sich diese Form einer statistischen Verteilung für eine Approximation in einer Fläche an, deren Punkte sich nur in den Koordinatenwerten unterscheiden. In der Praxis lässt sich die Standortschätzung durch eine Gleichverteilung leicht umsetzen. Eine einfache Variante für die Anwendung ist die Nutzung eines Zufallsgenerators, der aufbauend auf einer Gleichverteilung Zufallszahlen erzeugt. Die generierten Zufallszahlen können daraufhin in Koordinaten übersetzt werden, um diese in der vorgegebenen Fläche zu verteilen. Zwei Zufallszahlen entsprechen hierbei den Koordinaten eines Standortes, eine dem Längen- und eine dem Breitengrad. Pro Postleitzahlengebiet sollten deshalb doppelt so viele Zufallszahlen durch den Generator erzeugt werden, wie Standorte in dem Gebiet zu approximieren sind. Die Standortbestimmung durch eine Gleichverteilung ist passend für das vorliegende Problem zugeschnitten. Deshalb soll die Methode im Folgenden zur Anwendung kommen und hinsichtlich der Problemstellung getestet werden.

3.3 Statistische Verfahren bei fehlenden Daten

Der richtige Umgang mit Daten ist in allen Bereichen der Wissenschaft die Grundlage, um aus den Daten Informationen zu gewinnen. Die Analyse der Daten ist die Hauptaufgabe der Statistik (vgl. Sibbertsen und Lehne, 2015). Mit statistischen Methoden können aufgrund der Daten wichtige Erkenntnisse erlangt werden, die für strategische und operative Entscheidungen relevant

sind. In der Statistik kann aber das Problem unzureichender Daten zu großen Problemen führen. Als Lösung wurden deshalb Verfahren entwickelt, die fehlende oder unbekannte Daten ersetzen oder schätzen sollen. Drei Möglichkeiten für die Berechnung solcher Schätzwerte sollen im Folgenden vorgestellt und im Kontext der Standortapproximation bei unzureichender Datenlage untersucht werden.

3.3.1 Statistische Schätzfunktionen

In der Statistik steht man häufig vor dem Problem, dass man Kenngrößen, wie z.B. Erwartungswerte oder Varianzen, für eine zu untersuchende Verteilung von Beobachtungen nicht kennt. Oft kann man die gesuchten Werte nicht einfach erheben oder errechnen, da die statistische Grundgesamtheit, also die gesamte interessierende Menge, auf die sich die Parameter beziehen sollen, zu groß ist um alle benötigten Werte zur Berechnung messen zu können (vgl. Sibbertsen und Lehne, 2015). Diese Kennwerte werden aber benötigt, um sinnvolle Aussagen über die zugrunde liegende Verteilung der Grundgesamtheit treffen zu können. Da einfache Annahmen für die Parameter willkürlich sind und selten zu guten Ergebnissen führen, werden diese Größen in der induktiven oder auch schließenden Statistik approximiert. Dafür werden sogenannte Schätzfunktionen und Schätzwerte benutzt. Kamps (2018) definiert eine Schätzfunktion als eine spezielle Funktion, die aufgrund ihrer Eigenschaften (wie Erwartungstreue, Wirksamkeit oder Konsistenz) zur Schätzung eines Parameters der Grundgesamtheit geeignet ist. Die Schätzfunktion soll also aus einer erhobenen Stichprobe einer statistischen Grundgesamtheit einen passenden Schätzwert für den gesuchten Parameter ermitteln. Dies geschieht im Allgemeinen mit der Absicht, Rückschlüsse auf die interessierende Grundgesamtheit ziehen zu können (vgl. Bamberg et al., 2012). Die Stichprobe x_1, \dots, x_n der Grundgesamtheit entspricht hierbei einer Ziehung von n unabhängigen Zufallsvariablen X_1, \dots, X_n , die die gleiche Wahrscheinlichkeit und Verteilung haben sollen. Eine Stichprobe ist also eine Teilmenge der statistischen Grundgesamtheit, welche Zufallsvariablen enthält. Zufallsvariablen ordnen einem Ergebnis eines Zufallsexperiments eine reelle Zahl als Wert zu (vgl. Rabe et al., 2017). Die Werte x_1, \dots, x_n gelten in dem Fall als die Realisationen der Zufallsvariablen. Die Eigenschaften Erwartungstreue, Wirksamkeit und Konsistenz sollen sicherstellen, dass die Schätzfunktion möglichst gute Schätzwerte für die gesuchten Parameter liefert. Erwartungstreue bedeutet, dass der Schätzer im Mittel dem tatsächlichen Wert entsprechen soll. Wird der Abstand zwischen dem Schätzer und dem tatsächlichen Wert mit steigendem Stichprobenumfang geringer, so ist der Schätzer zusätzlich konsistent (vgl. Fahrmeir et al., 2016). Ein Schätzer ist wirksamer als andere, wenn er eine geringere Varianz hat als andere Schätzer. Werden alle diese Eigenschaften erfüllt, so stellt der betrachtete Schätzer eine gute Annäherung dar.

Schätzfunktionen werden in zwei Kategorien unterschieden, in Punktschätzer und Intervallschätzer. Allgemein bilden Punktschätzer die Stichprobenvariablen durch eine Funktion in die Menge der reellen Zahlen ab, um den gesuchten Parameter zu approximieren (vgl. Sibbertsen und Lehne, 2015). Die allgemeine mathematische Definition eines Punktschätzers lautet daher $T = g(X_1, \dots, X_n)$. Der Schätzer T kann auch als Schätzverfahren oder Schätzfunktion bezeichnet werden. Der sich aus dem Einsetzen der Realisationen x_1, \dots, x_n ergebende Wert $t = g(x_1, \dots, x_n)$ entspricht dann dem eigentlichen Schätzwert für den gesuchten

Parameter. Aufgrund von verschiedenen Eigenschaften hinsichtlich der Anforderungen, wie z.B. Erwartungstreue oder Konsistenz, eignen sich unterschiedliche Schätzer für unterschiedliche statistische Werte. So ist beispielsweise das arithmetische Mittel ein Schätzer für den Erwartungswert und die empirische Varianz ein Schätzer für die Varianz. Zur Konstruktion von Punktschätzern können verschiedene Prinzipien eingesetzt werden. Laut Sibbertsen und Lehne (2015) ist die Maximum-Likelihood-Schätzung die populärste Methode. Diese Methode folgt dem Prinzip, zu den Realisationen x_1, \dots, x_n denjenigen Parameter auszuwählen, für den die Wahrscheinlichkeit, dass gerade diese Werte x_1, \dots, x_n auftreten, maximal wird. (vgl. Fahrmeir et al., 2016) Eine große Bedeutung kommt auch der Kleinste-Quadrate-Methode zuteil. Dieses Verfahren minimiert die quadratischen Abweichungen zwischen den Beobachtungswerten und dem geschätzten Wert.

Die andere Möglichkeit Parameter zu schätzen ist durch Intervallschätzer gegeben. Intervallschätzer schätzen unbekannte Parameter im Gegensatz zu den Punktschätzern nicht durch eine einzige Zahl ab, sondern durch ein Intervall, in dem der Parameter mit einer bestimmten Wahrscheinlichkeit liegen soll (vgl. Bamberg et al., 2012). Solche Intervalle $[U;O]$ werden auch Konfidenzintervalle genannt. Intervallschätzer bzw. Konfidenzintervalle werden angewendet, da bei Punktschätzern keine Aussage darüber getroffen werden kann, wie weit der Schätzer von der gesuchten Größe entfernt ist (vgl. Rabe et al., 2017). Ist es von Interesse, ob der gesuchte Wert einen bestimmten Grenzwert unter- oder überschreitet, so kommen Intervallschätzer zum Einsatz. Die Intervallgrenzen U und O entsprechen wiederum Punktschätzer der Zufallsvariablen X_1, \dots, X_n .

Die Schätzung unbekannter Parameter durch Schätzfunktionen wird in der Statistik häufig angewendet, da die daraus resultierenden Schätzer die gesuchten Kenngrößen zum Teil gut approximieren können und man dadurch Aussagen über eine Grundgesamtheit treffen kann. Allerdings lassen sich die statistischen Schätzfunktionen schlecht auf die betrachtete Standortapproximation übertragen. Eine Stichprobe von Zufallsvariablen mit ihren zugehörigen Realisationen, wie sie für die Konstruktion von Schätzfunktionen benötigt wird, ist in dem vorliegenden Fall nicht bzw. nur begrenzt vorhanden. Zwar kann man die Anzahl der Kundenstandorte der Supply Chain mit der Anzahl der benötigten Zufallsvariablen gleichsetzen, aber es würden in dem Fall immer noch die Realisationen zu den Variablen fehlen, auf denen man die Schätzfunktionen konstruieren und anwenden kann. Die statistischen Schätzfunktionen sind daher hier nicht zielführend. Aufgrund ihrer Bedeutung in der Statistik sei auf weiterführende Arbeiten zu dem Thema verwiesen. Sibbertsen und Lehne führen in ihrer Arbeit (2015) allgemein in die Statistik ein und betrachten dabei ausführlich auch die schließende Statistik mitsamt Punkt- und Intervallschätzern. Fahrmeir et al. diskutieren in ihrer Arbeit (2016) die induktive Statistik als ein Mittel zur Datenanalyse. Ähnlich geht auch Caspary in seiner Arbeit (2013) vor, der besonders auf mögliche Fehler in Messreihen eingeht und in dem Zusammenhang Schätzverfahren anwendet. Im Rahmen der Wahrscheinlichkeitsrechnung für die Simulation zeigen auch Rabe et al (2017) wichtige Grundlagen zu Schätzfunktionen auf.

3.3.2 Imputation fehlender Werte

Datenerhebungen wie z.B. Umfragen werden in verschiedenen Bereichen der Statistik eingesetzt, um Informationen zu sammeln. Dabei kann es immer wieder vorkommen, dass die erhobene Datenmenge aus unterschiedlichen Gründen unvollständig ist. Fehlende Werte (engl. Missing Data) können die Analyse der Erhebung verkomplizieren oder gar unmöglich machen (vgl. Carpenter und Kenward, 2013). Um solche Probleme zu vermeiden, wurden in der Statistik verschiedene Möglichkeiten zum Umgang mit fehlenden Werten entwickelt. Eine dieser Methoden ist die sogenannte Imputation (auch Ersetzungsverfahren genannt), die fehlende Werte ergänzen sollen. Aber auch Eliminierungsverfahren werden in dem Zusammenhang häufig angewendet, um Objekte oder Merkmale mit fehlenden Daten aus der Untersuchung auszuschließen (vgl. Bankhofer, 1995). Im Kontext dieser Arbeit soll sich auf die Betrachtung von Imputationsvarianten und der Vorarbeit zu ihrer Anwendung beschränkt werden, da Eliminierungen konträr zu Schätzungen von Parametern, wie z.B. Standorten, sind. Fehlende Werte können verschiedene Gründe haben. In der Literatur wird die Ursache des Fehlens von Daten nach ihren Abhängigkeiten von anderen Werten in die drei Kategorien, MCAR, MAR und NMAR unterteilt. Missing Completely At Random (MCAR) bedeutet, dass der fehlende Wert in einer Variablen weder von den übrigen Ausprägungen dieser Variablen noch von den anderen Variablen abhängig ist (vgl. Graham, 2009). Der fehlende Wert ist also völlig zufällig. Missing At Random (MAR) ist ein fehlender Wert, wenn dieser von den restlichen Variablen des Datensatzes abhängig sein kann, aber nicht von anderen fehlenden Ausprägungen (vgl. Molenberghs und Kenward, 2007). Not Missing At Random (NMAR) sind fehlende Werte die nicht MCAR oder MAR sind. Um die fehlenden Daten besser behandeln zu können bzw. das richtige Verfahren für den vorliegenden Fall anwenden zu können, müssen die fehlenden Daten anhand ihrer Struktur analysiert werden. Mittels verschiedener Verfahren werden ausgehend von einer Indikatormatrix beispielsweise Muster und Korrelationen zwischen Werten und Variablen gesucht, um zu überprüfen, welchen Typ die fehlenden Werte besitzen. Aufbauend auf die Strukturanalyse und ihre Ergebnisse können dann die entsprechenden Verfahren zur Ergänzung fehlender Werte, wie die Imputation, eingesetzt werden (vgl. Carpenter und Kenward, 2013).

Die Imputation ist eine Methode, bei der fehlende Werte durch bestimmte Schätzwerte ersetzt werden. Die Methode selber wird grundlegend in singuläre und multiple Imputation unterteilt. In der singulären Imputation existieren verschiedene Ansätze, um einen geeigneten Schätzwert für den fehlenden Wert zu finden. Eine Möglichkeit besteht darin, den fehlenden Wert durch den Mittelwert der vorhandenen Werte des Datensatzes zu ersetzen. Diese Variante lässt sich einfach anwenden, verursacht aber leicht Verzerrungen des Ergebnisses (vgl. Molenberghs und Kenward, 2007). Ähnlich simple Möglichkeiten sind Imputationen durch Verhältnisschätzer der vorhandenen Ausprägungen oder Hot- und Cold-Deck-Verfahren. Bei letzteren wird ein fehlender Wert eines Merkmals durch eine andere, vorhandene Ausprägung desselben Merkmals ausgetauscht. In dem Kontext wird häufig auch ein Regressionsverfahren verwendet (vgl. Carpenter und Kenward, 2013). Dieses Verfahren zielt darauf ab, Zusammenhänge zwischen verschiedenen Merkmalen zu nutzen, um einen Schätzwert zu generieren. Dafür wird z.B. die Methode der kleinsten Quadrate aus der linearen Regression verwendet. Der sogenannte EM-Algorithmus wird im Bereich der statistischen Datenergänzung ebenfalls häufig angewendet (vgl.

Bankhofer, 1995). Dieser Algorithmus ist ein iteratives Vorgehen zur Parameterschätzung, das aus zwei wesentlichen Schritten besteht. Zuerst wird mittels einer Maximum-Likelihood-Schätzung aus den unvollständigen Daten ein Schätzer erzeugt. Dann wird die Schätzfunktion der unvollständigen Daten maximiert, indem iterativ die Schätzfunktion der vollständigen Daten, also mit dem Schätzwert inbegriffen, maximiert wird (vgl. Ueda und Nakano, 1998)

Den singulären Imputationsmethoden gegenüber steht die multiple Imputation. Die multiple Imputation generiert nicht nur einen, sondern mehrere Schätzwerte für einen fehlenden Wert. Dazu werden zunächst verschiedene singuläre Imputationsmethoden auf denselben unvollständigen Datensatz angewendet, um fehlende Werte zu ersetzen (vgl. van Buuren, 2012). Pro fehlendem Wert werden mehrere Werte generiert. Danach werden die so erzeugten Datensätze mit den unterschiedlichen generierten Werten analysiert, um daraus geeignete Schätzwerte zu entwickeln. Zum Schluss werden dann die Schätzwerte zu einem einzigen Schätzwert zusammengeführt. Dies kann beispielsweise durch die Berechnung eines Mittelwertes aus den verschiedenen Schätzwerten erfolgen.

Je nach Verfahren gibt es unterschiedliche Vor- und Nachteile, die für oder gegen eine Anwendung sprechen. Der EM-Algorithmus und die multiple Imputation liefern im Gegensatz zu den anderen Verfahren der singulären Imputation bessere Ergebnisse, haben aber auch einen höheren Aufwand, diese zu berechnen. Die anderen Verfahren sind dagegen ohne großen Aufwand anzuwenden, erzeugen aber auch für die meisten Anwendungsfälle schlechtere Ergebnisse. Unabhängig von der Art des Verfahrens ist diese Form der Datengenerierung bzw. Datenergänzung für die Problemstellung dieser Arbeit allerdings nicht zielführend. Die verschiedenen Verfahren der Imputation haben gemein, dass sie von vorhandenen Ausprägungen eines Merkmals ausgehen, um fehlende Werte eines Datensatzes ersetzen zu können. Im vorliegenden Fall sind aber nur Ausprägungen des Merkmals Postleitzahl des Kundenstandorts vorhanden. Die Daten dieses Merkmals sind aber vollständig, hier werden keine Ergänzungen oder Imputationen benötigt. Die Generierung von Schätzwerten wird für das Merkmal Adresse oder Koordinaten des Kundenstandorts gebraucht. Die Daten dieses Merkmals fehlen aber vollständig, sodass hier keine Anwendung der beschriebenen Methoden erfolgen kann. Daher können die Verfahren der Imputation nicht für die interessierende Standortapproximation angewendet werden. Es sei dennoch auf weitere Literatur hingewiesen, die sich mit diesen Methoden zum Umgang mit fehlenden Daten tiefergehend befassen. Bankhofer (1995) hat sich in seiner Arbeit mit Missing Data Techniken in Bezug auf Daten- und Distanzmatrizen auseinandergesetzt. Eine Übersicht über verschiedene Verfahren im Bereich der fehlenden Daten gibt auch van Buuren (2012). Carpenter und Kenward (2013) beschränken sich in ihrer Arbeit auf die Multiple Imputation, während sich Molenberghs und Kenward (2007) auf Verfahren zum Umgang mit fehlenden Daten in medizinischen Untersuchungen fokussieren.

3.3.3 Mittelwertberechnungen

In der Logistik und dem Operations Research werden vielfach mathematische Formeln oder Algorithmen angewendet, um gesuchte Größen zu ermitteln. Die gesuchten Werte spielen für Planungs- und Optimierungsvorgehen eine große Rolle. Viele der Formeln haben ihren Ursprung in der Statistik, sind aber auf viele verschiedene Themenbereiche übertragbar. Im Rahmen der

Disposition und der Dimensionierungsrechnung kommt so z.B. die Berechnung von Mittelwerten zum Einsatz. Der Mittelwert soll dabei einen Repräsentanten von stochastischen Größen wie beispielsweise Auftragseingängen, Auftragsstrukturen oder Leistungsdurchsätzen darstellen (vgl. Gudehus 1, 2007). Zur Berechnung eines Mittelwertes lässt sich das arithmetische Mittel verwenden. Das arithmetische Mittel ist auch als Durchschnitt bekannt und kann auf verschiedene Zahlenmengen angewendet werden. Das Vorgehen zur Berechnung ist bei allen Mengen an Werten dasselbe. Alle Zahlenwerte einer Menge werden aufsummiert und anschließend durch die Anzahl der Werte dividiert. Die mathematische Formel des arithmetischen Mittels einer Zahlenmenge x_1, x_2, \dots, x_n ist

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad (4)$$

wobei n der Anzahl der Zahlenwerte entspricht. Der Vorteil des arithmetischen Mittels ist die Angabe eines charakteristischen Mittelwerts bei einer einfachen Anwendung, ohne dass hierzu bestimmte Voraussetzungen zu erfüllen sind.

Eine andere Möglichkeit einen Mittelwert zu erhalten ist die Berechnung des Medians. Wie das arithmetische Mittel ist der Median eine statistische Größe, die einen Lageparameter für eine Zahlenmenge darstellt. In der Logistik wird der Median unter anderem in der Standortplanung verwendet (vgl. Domschke und Drexl, 1996). Die Bestimmung des Medians ist allgemein sehr einfach. In einer Zahlenmenge, die aufsteigend nach ihren Werten sortiert wird, entspricht der Median genau dem Wert, der in der Mitte der geordneten Menge steht. Da aber nur bei einer ungeraden Anzahl an Werten in einer Menge eine exakte Mitte existiert, muss bei einer geraden Anzahl an Zahlen ein anderer Wert als Median bestimmt werden. Dieser Wert wird durch das arithmetische Mittel der sich an den beiden mittleren Stellen der sortierten Zahlenmenge befindlichen Werte berechnet. Mathematisch bedeutet das, der Median einer geordneten Zahlenmenge x_1, x_2, \dots, x_n ist

$$\tilde{x} = \begin{cases} x_{\frac{n+1}{2}} & \text{falls } n \text{ gerade} \\ \frac{1}{2} \left(x_{\frac{n}{2}} + x_{\frac{n}{2}+1} \right) & \text{falls } n \text{ ungerade} \end{cases} \quad (5)$$

wobei n der Anzahl der Zahlen entspricht. Wie beim arithmetische Mittel besteht der Vorteil in der Anwendung dieses Mittelwerts in seiner Einfachheit. Man kann also mit dem Median schnell einen Lageparameter bestimmen, der die zugrundeliegende Zahlenmenge charakterisiert. Allerdings kann der Median nur angewendet werden, wenn die betrachteten Werte sortierbar sind. Zur Mittelwertberechnung können auch andere mathematische und statistische Größen zu Rate gezogen werden, wie z.B. das getrimmte Mittel (vgl. Caspary, 2013). Als Beispiel für eine

Verteilung beschreibende Lageparameter reichen aber die aufgezeigten Mittelwerte aus, weshalb auf andere Berechnungsmethoden in dieser Arbeit nicht explizit eingegangen werden soll.

Das arithmetische Mittel und der Median bilden einfache Mittelwerte, die Lageparameter für Verteilungen darstellen. Die Anwendung kann auf verschiedene Zahlenmengen angewendet werden, auch wenn die zugrundeliegende Verteilung unbekannt ist. Die Lageparameter können wiederum Approximationen für Werte innerhalb der Zahlenmenge verkörpern (vgl. Sibbertsen und Lehne, 2015). Daher kann die Berechnung von Mittelwerten auf das vorliegende Problem der Standortapproximation abstrahiert werden. Die Methode soll deshalb in den folgenden Kapiteln erst an die vorliegende Problemstellung angepasst werden und dann zur Anwendung kommen.

3.4 Data Farming und Data Mining

Außerhalb der Statistik gibt es noch andere Möglichkeiten fehlende oder unbekannte Daten zu ergänzen. Im Bereich der Informatik wird für das Problem der Datenergänzung oder Datengenerierung häufig auf die Themengebiete Data Mining und Data Farming hingewiesen. Die gemeinschaftliche Anwendung dieser beiden aus der Informatik und Statistik stammenden Methoden soll im Allgemeinen dabei helfen, Daten, besonders nützliche Wirkungszusammenhänge, aufzuzeigen und zu verarbeiten (vgl. Feldkamp et al., 2017). Dazu sollen mit Hilfe eines Simulationsmodells Daten erzeugt werden, die anschließend mit verschiedenen Verfahren untersucht und verarbeitet werden. Die eigentliche Datengenerierung ist hierbei Aufgabe des Data Farming. Die Verarbeitung der produzierten Daten erfolgt durch das Data Mining. Data Farming ist eine Methode, die basierend auf einem Simulationsmodell einen ganzheitlichen und iterativen Ansatz zur Analyse von komplexen Systemen darstellt (vgl. Kallfass und Schlaak, 2012). Feldkamp et al. (2017) gehen weiter und beschreiben Data Farming als eine Methodik für die Verwendung eines Simulationsmodells als Datengenerator mit Hilfe von effizientem Experimentdesign und High Performance Computing mit dem Ziel, das möglichst vollständige Spektrum von Ergebnisdaten zu erhalten und somit den Informationsgewinn zu verbessern. Bildlich gesprochen heißt das, durch diese Methode sollen Daten anwachsen und ihre Erträge maximiert werden, ähnlich, wie es bei einem Farmer und der Bestellung seines Landes aussieht. Ursprünglich wurde diese Methode für militärische Zwecke entwickelt, sie wird aber auch zunehmend im zivilen angewendet, um komplexe Systeme und deren Verhalten zu untersuchen (vgl. Horne und Meyer, 2016). Dabei werden nicht zwingend nur die gestellten Fragen beantwortet, sondern auch nicht betrachtete Probleme und Lösungen zur Entscheidungsfindung aufgedeckt. Eine steigende Rechenleistung von Computern, zunehmend vorhandene Big-Data-Infrastrukturen und neue Ansätze für die Gestaltung von Simulationsexperimenten ermöglichen die Abdeckung eines großen Wirkungsraums von einem System und seinem Verhalten (vgl. Feldkamp et al., 2017).

Die durch Data Farming erzeugten Simulationsdaten für ein System werden dann in vielen Fällen durch Data Mining Methoden verarbeitet. Lackes (2018) definiert Data Mining als die Anwendung von Methoden und Algorithmen zur möglichst automatischen Extraktion empirischer Zusammenhänge zwischen Planungsobjekten, deren Daten in einer hierfür aufgebauten Datenbasis bereitgestellt werden. Dieses Verfahren zielt also darauf ab,

Zusammenhänge und Abhängigkeiten zwischen Daten zu ermitteln, also Informationen aus Daten „heraus zu schürfen“. Dazu kommen verschiedene Verfahren der künstlichen Intelligenz und der Statistik zum Einsatz, wie z.B. Clustering-Verfahren (vgl. Feldkamp et al, 2016). Data Mining ist also eine Form von Informationsgenerierung und -verarbeitung, die z.B. für Prognosemodelle relevant ist. Anschließend an die Verarbeitung der erzeugten Daten durch Data Mining Verfahren folgt eine Analyse durch eine Visualisierung der Ergebnisse (vgl. Feldkamp et al., 2017). Prinzipiell sind die Verfahren unabhängig voneinander, daher kann in der Praxis sowohl eine Kombination der Verfahren, als auch eine einzelne Anwendung bestimmter Methoden erfolgen.

Das Zusammenspiel dieser Methoden hat den großen Vorteil auch bei komplexen Systemen Daten generieren und verarbeiten zu können. Nachteilig in Bezug auf die vorliegende Problemstellung dieser Arbeit ist aber, dass zur Anwendung von Data Farming ein valides Simulationsmodell zur Datengenerierung benötigt wird. Dieses ist im vorliegenden Fall aber nicht vorhanden, daher kann dieses Konzept in diesem Fall nicht angewendet werden. Wäre ein Simulationsmodell vorhanden, könnte dieses Verfahren unter Umständen auch für die Standortapproximation verwendet werden. Da Data Mining aus vorhandenen Daten relevante Informationen herauszieht, ist diese Methode ebenfalls nicht zielführend. Die Daten im betrachteten Fall sind zu wenig, dass sich Data Mining hier lohnen würde bzw. überhaupt funktionieren würde. Beide Themengebiete stellen für sich und in ihrer gemeinschaftlichen Anwendung gute Methoden zur Informationsgenerierung dar, eignen sich aber nicht für die betrachtete Problemstellung. Dennoch sei wegen der zunehmenden Bedeutung von Data Farming und Data Mining aber auf andere Arbeiten hingewiesen, die sich nähergehend mit diesen Methoden befassen. Grundlagen zum Data Farming haben unter anderem Kallfass und Schlaak (2012) sowie Horne und Meyer (2016) betrachtet. Feldkamp et al. (2016) haben in ihrer Arbeit sowohl die Anwendung von Data Farming als auch Data Mining Methoden betrachtet. In einer anderen Arbeit (2017) haben Feldkamp et al. Data Farming in Bezug zu Produktion und Logistik gesetzt und mit einer Fallstudie veranschaulicht. Ausgehend von dem militärischen Ursprung hat die NATO (2010) ebenfalls eine Zusammenfassung zum Data Farming erarbeitet.

4 Entwicklung von Verfahren zur Standortannäherung

In Kapitel 3 wurden verschiedene Methoden zum Umgang mit einer nicht ausreichenden Datenlage und zur Schätzung fehlender Werte betrachtet und hinsichtlich ihrer Praktikabilität für die Standortapproximation bewertet. Es wurde festgestellt, dass nur ein Verfahren ohne Anpassung auf die vorliegende Problemstellung anwendbar ist. Die Mehrheit der Verfahren stellt zwar gute Schätzmethode dar, eignet sich aber aufgrund diverser Eigenschaften nicht zur Standortapproximation. Um die Güte der Ergebnisse einordnen zu können, werden weitere Verfahren zum Vergleich benötigt. Deshalb sollen im Folgenden neue Methoden zur Annäherung von Standorten entwickelt werden. Diese Verfahren sollen ebenfalls von einer begrenzten Datengrundlage ausgehen und haben daher die Aufgabe die unbekannt Lage von Standorten mit wenigen Daten zu approximieren. Hinsichtlich dieser Zielsetzung wird zunächst in Abschnitt 4.1 wiederholt, wie die allgemeine Problemstellung dieser Arbeit aussieht und daraufhin konkretisiert, wie die daraus resultierende Fallstudie aufgebaut ist, auf die sämtliche Verfahren angewendet werden können. In dem Zuge wird erläutert, wie der Referenzdatensatz erzeugt worden ist, der den Bezugspunkt für die Fallstudie und die Auswertung der Approximationsverfahren darstellt. Danach wird das Bewertungsmuster zur Beurteilung der einzelnen Verfahren entwickelt und erläutert, warum sich dieses Muster für die Bewertung der Methoden eignet. Abschließend folgt die Entwicklung von drei Ansätzen zur Standortverteilung bei unzureichender Datenlage. Für jeden dieser Ansätze werden eine grundlegende Herleitung und ein theoretisches Vorgehen dargelegt, welche dann folgend in Kapitel 5 zur praktischen Anwendung kommen.

4.1 Anwendungsgrundlage und Bewertungsverfahren

Für die vollständige Verwendung und Überprüfung von Verfahren, bei der keine Variablen in den Ergebnissen vorhanden sein sollen, wird eine konkrete Situation benötigt, auf die sich die unterschiedlichen Ansätze beziehen können. Die Situation soll durch eine konkretisierte Fallstudie mit einer fiktionalen Supply Chain wiedergegeben werden, die aus der allgemeinen Problematik der Arbeit hergeleitet werden soll. Dazu wird die vorliegende Problemstellung noch einmal ausführlich wiederholt. Im betrachteten Fall soll eine nicht weiter definierte Supply Chain hinsichtlich ihrer Versorgungsstruktur bewertet werden. Dafür ist eine Simulation der Lieferkette geplant, die Erkenntnisse zur Güte der bestehenden Strukturen und Prozesse liefern soll. Für die Simulation werden qualitativ und quantitativ hochwertige Daten benötigt. Dazu zählen auch Daten zu allen Kundenstandorten der Supply Chain. Von diesen Standorten sind aber weder die genauen Adressen noch ihre Geokoordinaten hinterlegt. Einzig die Postleitzahlen der Gebiete, in denen die Standorte jeweils lokalisiert sind, sind vorhanden. Zu weiteren Daten, wie die Bedarfe der Kundenstandorte oder Transportkosten zur ihrer Belieferung, wurden keine Aussagen getroffen. Für den zu untersuchenden Sachverhalt sollen diese Informationen als nicht relevant erachtet werden, da sie keinen Einfluss auf die Schätzung der Lage der Standorte haben. Deshalb werden diese Parameter im Folgenden nicht berücksichtigt. Um fehlende Daten für eine

Simulation zu ergänzen werden in der Regel Erhebungen eingesetzt. Erhebungen sollen aber aufgrund von außenstehenden Faktoren in diesem Fall nicht möglich sein. Es besteht hier also eine unzureichende Datenlage. Um dennoch eine Simulation durchführen zu können, müssen die fehlenden Daten hierfür geschätzt werden. Die Schätzung soll durch Verfahren, die Standorte in einer Fläche approximativ verteilen können, erfolgen. Weil die Anwendung eines Verfahrens nicht notwendigerweise ein gutes Ergebnis zur Folge hat, sollen mehrere Verfahren getestet werden. Durch einen Vergleich der Ergebnisse ist auch eine Einordnung der Qualität der Methoden möglich. Diese Arbeit hat daher das Ziel verschiedene Verfahren zu untersuchen, die Kundenstandorte einer Supply Chain in einer gegebenen Fläche approximieren sollen, wenn die genauen Standortinformationen nicht vorhanden sind. Dies soll durch die Betrachtung einer konkreten, aber fiktionalen Supply Chain innerhalb einer Fallstudie geschehen. Die Fallstudie steht dabei stellvertretend für ähnliche Problemstellungen mit unbekanntem Positionierungen von Standorten. Durch die zu den Standorten gehörenden Postleitzahlen ist bekannt, wie die Kundenstandorte innerhalb eines Postleitzahlengebietes vorhanden sind. Der Einfachheit halber soll sich für die Fallstudie auf eine Supply Chain in Nordrhein Westfalen, einem Bundesland Deutschlands, konzentriert werden. Die betrachteten Kundenstandorte sollen also alle in Postleitzahlengebieten Nordrhein Westfalens lokalisiert sein. Dies hat den großen Vorteil, dass sich der Aufwand zur Generierung der Referenzdaten, aber besonders zur späteren Darstellung der Ergebnisse reduziert. Das betrachtete Gebiet umfasst die Stadt Dortmund und die räumliche Umgebung mit angrenzenden Städten, wie Witten, Kamen und Bochum. Die untersuchten Verfahren können aber leicht auf ähnliche Problemstellungen in anderen oder größeren Gebieten und Ländern übertragen werden.

Zur Ausübung der unterschiedlichen Ansätze werden Referenzdaten benötigt, die als Ausgangspunkte für die Anwendung und die Bewertung der Verfahren dienen sollen. Die fiktionale Supply Chain soll eine große Anzahl an Kunden in mehreren Postleitzahlengebieten besitzen, um die Verfahren in ausreichender Weise testen zu können. Mehrere Postleitzahlengebiete mit unterschiedlich vielen zu approximierenden Standorten verdeutlichen Stärken und Schwächen der Ansätze. Dabei soll die Anzahl der Kundenstandorte aber auch nicht zu groß ausfallen, um nicht mehr Aufwand zu verursachen, als zur Validierung der Verfahren benötigt werden. Die Supply Chain soll deshalb etwa 100 Standorte besitzen. Da sich hierbei auf den Großraum Dortmund beschränkt wurde, werden Daten für Standorte innerhalb der Region gesucht. Zur Generierung der Referenzdaten wird auf die Onlineplattform Overpass Turbo zurückgegriffen. Das kostenlose Online Tool basiert auf dem frei zugänglichen Kartendienst OpenStreetMap. Es ermöglicht einfache Abfragen und Ausgaben zu Standorten und Geoinformationen aus verschiedenen Bereichen. Abfragen geben die gesuchten Informationen oder Standorte in einem selbst gewählten Gebiet aus. Das selbst gewählte Gebiet soll die Stadt Dortmund und deren Umgebung sein. Hierfür wird der betrachtete Bereich manuell eingegrenzt. Um eine Supply Chain als Referenz zu erstellen, werden in etwa 100 Standorte benötigt, die in der Fläche beliebig verteilt sind. Mehrere Abfragen zu unterschiedlichen Einrichtungstypen wurden ausgeführt, um einen Typus mit passender Anzahl an Standorten zu finden. Die Abfrage nach Postfilialen (in Overpass Turbo wird durch die Programmiersprache das englische Synonym `post_office` benutzt) stellt hier eine gute Referenzgröße aufgrund der Anzahl der Standorte und

deren Verteilung in unterschiedlichen Postleitzahlengebieten dar (vgl. Abbildung in Anhang 1). Die Standorte der Postfilialen in dem eingegrenzten Gebiet sollen stellvertretend für die Kundenstandorte der Supply Chain stehen. Overpass Turbo ermöglicht es dem Anwender, abgefragte Standorte und dazugehörige Daten im CSV-Dateiformat auszugeben. Für den Referenzdatensatz werden die Postleitzahlen und die genauen Koordinaten der Standorte, die unter dem Tag `post_office` abgespeichert sind, abgefragt. Die Koordinaten von Standorten werden durch die geographische Breite (engl. Latitude) und die geographische Länge (Longitude) definiert. Zusammen legen die Koordinaten einen Standort eindeutig fest. Die durch die Abfrage in Overpass Turbo ausgegebenen Daten befinden sich in Tabellenform im Anhang 1. Weil die Daten in OpenStreetMap bei einigen Standorten zum Zeitpunkt der Abfrage nicht vollständig gepflegt waren, mussten einige Zuordnungen der Standorte zu Postleitzahlengebieten manuell ergänzt werden. Insgesamt enthalten der Referenzdatensatz und damit die fiktive Supply Chain 110 Kundenstandorte, welche in 35 Postleitzahlengebieten Dortmunds und der angrenzenden Städte verteilt sind. Die Anzahl der Standorte pro Postleitzahlengebiet variiert zwischen einem und 19 Standorten, wobei 17 der Gebiete nur einen einzigen Standort beherbergen. Bei der Anwendung der nachfolgend entwickelten Approximationsverfahren in Kapitel 5 werden nur die Postleitzahlen des Datensatzes verwendet, um eine unzureichende Datenlage zu erreichen. Die genauen Koordinaten der Referenzstandorte sollen dann im Rahmen der Auswertung der Ansätze verwendet werden.

Die aus der beschriebenen Problemstellung und den Referenzdaten hergeleitete Fallstudie sieht somit wie folgt aus: Eine Supply Chain soll durch eine Simulation untersucht werden. Die Unternehmung hat 110 Kunden, die beliefert werden müssen. Sämtliche Kunden sind in dem Großraum Dortmund angesiedelt. Die genauen Koordinaten der Kunden sind nicht bekannt, werden aber für die geplante Simulation der Supply Chain benötigt. Daher sollen diese Standorte approximiert werden. Ausgehend von den bekannten Postleitzahlen der Kundenstandorte sollen diese in den einzelnen Gebieten verteilt werden. Die Verteilung innerhalb der PLZ soll durch bereits vorhandene und durch neu entwickelte Ansätze erfolgen. Es sollen insgesamt vier Ansätze anhand der Supply Chain getestet werden, um zu überprüfen, welche von den Ansätzen die Realität gut abbilden.

Unterschiedliche Verfahren können auch bei einer Anwendung auf denselben Datensatz unterschiedliche Ergebnisse liefern. Um zu ermitteln, welche Verfahren die Realität gut abbilden, wird eine allgemeingültige Bewertungsmethode gebraucht, die einen nachvollziehbaren Vergleich der verschiedenen Approximationsverfahren erlaubt. Eine logische Schlussfolgerung aus der Problemstellung der Standortapproximation ist die Bewertung der Verfahren anhand der Abweichung der Koordinaten der approximierten Standorte zu den Koordinaten der realen Standorte des Referenzdatensatzes und damit der Supply Chain. Je geringer also die Abweichung zwischen den realen und angenäherten Standorten eines Verfahrens ist, desto besser bildet das verwendete Verfahren demnach die Realität ab. Eine Bewertung zum Aufwand eines Verfahrens wird bei dieser Bewertungsmethode vernachlässigt, d.h. es wird hier nur das Ergebnis bewertet. Über andere eventuell negative Eigenschaften eines Verfahrens, wie z.B. eine Beschränkung bei der Anwendung eines Verfahrens, wird ebenfalls keine Aussage getroffen.

Die Abweichungen der Standorte können mithilfe der einfachen euklidischen Distanz berechnet werden. Die euklidische Entfernung bietet sich hier an, da die Abweichung zwischen realem Standort und Approximation unabhängig von geographischen Faktoren ist. Es kann also die direkte Luftlinie zwischen den Standorten als Wert für die Abweichung verwendet werden. Da in vielen Postleitzahlengebieten mehrere Standorte zu approximieren sind und daher auch entsprechend viele Standorteschätzungen vorgenommen werden müssen, existiert in diesen Gebieten ein Zuordnungsproblem. Theoretisch könnte jeder reale Standort mit jedem approximierten Standort verglichen werden. Allerdings sind der mehrfache Vergleich und die damit verbundene mehrfache Berechnung der Abweichung wenig sinnvoll, weil ein realer Standort nur durch genau eine Standortsschätzung dargestellt werden kann. Die Distanzen zwischen den anderen Standorten sind deshalb für den vorliegenden Fall unerheblich. Daraus ergibt sich, dass festgelegt werden muss, zwischen welchen Standorten eine Abweichungen gemessen werden soll und zwischen welchen nicht.

Eine Zuordnungsmöglichkeit besteht darin, die approximierten mit den realen Standorten in Beziehung zu setzen, die die geringste euklidische Distanz zueinander haben. So werden nur die Kombinationen gewichtet, bei denen die Approximationen der Realität am ehesten entsprechen. Es werden also die minimalen euklidischen Distanzen zwischen zwei Standorten innerhalb eines Postleitzahlengebiets gesucht. Mathematisch bedeutet das, es wird das $\min d_{ij}(x, y)$ für alle i gesucht, die innerhalb eines Postleitzahlengebietes lokalisiert sind. Das i steht für die realen Standorte, während das j die approximierten Standorte bezeichnet. Die beiden Indizes sind insgesamt durch die Anzahl der Standorte n limitiert. Die Variablen x und y definieren die Breiten- und Längengrade der Standortkoordinaten. Die Distanz $d_{ij}(x, y)$ wird wie in Abschnitt 2.4 vorgestellt mittels der einfachen euklidischen Distanz (2) berechnet. Die Koordinaten der realen Standorte entstammen dem Referenzdatensatz, die Koordinaten der angenäherten Standorte werden durch die unterschiedlichen Approximationsverfahren ermittelt. Ein Wert von Null für eine Abweichung würde ein Übereinstimmen des geschätzten Standortes mit dem tatsächlichen Standort bedeuten, ein hoher Wert zeigt wiederum eine starke Differenz von der Realität.

Damit die Verfahren untereinander verglichen werden können, wird ein Wert benötigt, der eindeutig Aufschluss darüber gibt, wie gut das Verfahren die Realität abbildet. Einen solchen Wert findet man in der Summe aller Abweichungen zwischen den Standorten, d.h. die Summe aller minimaler Distanzen zwischen den angenäherten und realen Standorten ergibt die absolute Abweichung zwischen den Standorten durch die Anwendung eines Verfahrens. Für die Summe ist es unerheblich, welchem Postleitzahlengebiet die Standorte angehören, da die relevanten Abweichungen aller Standorte einbezogen werden. Mathematisch wird also

$$\sum_{j=1}^n \min d_{ij}(x, y) \tag{6}$$

gesucht. Je geringer die absolute Abweichung, also die Summe aller Abweichungen, bei einem Approximationsverfahren ausfällt, desto besser bilden die Ergebnisse des Verfahrens die Realität ab. Ein Wert von Null würde ein Übereinstimmen aller approximierten Standorte mit den realen bedeuten. Demgegenüber würde durch viele große Abweichungen oder einiger sehr großer die Summe entsprechend hoch ausfallen, was insgesamt eine schlechte Abbildung der Realität durch das Annäherungsverfahren bedeuten würde. In dem Zuge können auch andere Lageparameter wie beispielsweise das arithmetische Mittel und der Median eingesetzt werden. Werden die Parameter auf die gesamte Anzahl der Abweichungen angewendet, so können sie Aufschluss über die Verteilung der Abweichungen geben. Trotz der Vernachlässigung anderer Faktoren zur Beurteilung, wie dem Aufwand bei einer Anwendung, eignet sich die beschriebene Bewertungsmethode für den vorliegenden Fall, da die Realitätstreue das maßgebliche Kriterium für ein Approximationsverfahren von Standorten darstellt. Hinzu kommt, dass die Bewertungsmethode einen einfachen Vergleich der unterschiedlichen Ansätze erlaubt. Die absoluten Abweichungen geben hierbei nicht nur Aufschluss darüber, ob ein Verfahren relativ besser ist als ein anderes, es erlaubt auch eine Beurteilung, ob ein Verfahren an sich die Realität gut genug abbildet oder nicht.

4.2 Verfahrensentwicklung

Für die Simulation sollen die Kundenstandorte approximiert werden. Wie bereits erwähnt, wird dieses Problem der Schätzung von Standorten bei unzureichenden Daten in der Literatur fast vollständig vernachlässigt. Einzig die Methode, Standorte mittels einer Gleichverteilung in einem Postleitzahlengebiet zu verteilen, ist für das vorliegende Problem ohne Abwandlung geeignet. Weil das Ergebnis des Verfahrens nur eine mögliche, eventuell verbesserungswürdige Annäherung darstellt, ist es sinnvoll, mehrere Verfahren für die betrachtete Problematik zu untersuchen. Aus diesem Grund müssen weitere Ansätze speziell zur Standortapproximation entwickelt werden. In den folgenden Abschnitten sollen drei verschiedene Möglichkeiten, Standorte zu schätzen, entwickelt und beschrieben werden. Dazu soll für jede Variante zunächst die Herleitung des Verfahrens erörtert werden, um anschließend ein theoretisches Vorgehen zur Standortermittlung aufzuzeigen, welches später in die Tat umgesetzt werden soll.

4.2.1 Mittelpunkte als Standorte verwenden

Eine Möglichkeit eine Tendenz einer Verteilung zum Ausdruck zu bringen ist die Berechnung von Lageparametern. In Abschnitt 3.5 wurden mit dem arithmetischen Mittel und dem Median zwei verschiedene Mittelwerte aufgezeigt, die Lageparameter für eine Verteilung darstellen. Da die Mittelwerte selber auch Approximationen für Werte der zugrundeliegenden Zahlenmenge darstellen können, soll die Mittelwertberechnung auf die betrachtete Standortapproximation übertragen werden. Der zweite Verfahrensansatz besteht also aus dem Gedanken, einen Mittelwert bzw. einen Mittelpunkt eines Postleitzahlengebiets als Approximation für die sich in dem Gebiet befindlichen Kundenstandorte zu verwenden. Die Bestimmung der Mittelpunkte soll durch die Berechnung von Mittelwerten anhand der vorgestellten Formeln zum arithmetischen Mittel (4) und Median (5) erfolgen. Für die Positionierung der Mittelpunkte sollen beide Berechnungsmethoden verwendet und getestet werden. So können pro Postleitzahlengebiet bis zu

zwei unterschiedliche Mittelpunkte berechnet werden. Im Rahmen der Annäherung der Standorte sollen die Methoden getrennt voneinander angewendet und ausgewertet werden. Die betrachteten Zahlenmengen, aus denen die Mittelwerte bestimmt werden, sollen durch die geographischen Koordinaten, die die Grenzen eines Gebiets definieren, gegeben sein. Für den Fall, dass es mehrere Standorte in einem Postleitzahlgebiet gibt, so werden diese Standorte alle durch den Mittelpunkt approximiert und gebündelt dargestellt. Es gibt also nur eine Annäherung für alle Standorte eines Postleitzahlgebiets. In der Realität würde eine derartige Bündelung mehrerer Kunden einem Industrie- oder Technologiezentrum entsprechen, in welches sich mehrere verschiedene Einrichtungen niedergelassen haben. Die unterschiedlichen Organisationen in diesem Zentrum haben für beliefernde Betriebe alle quasi dieselbe Adresse. Die Anwendung des Ansatzes zur Approximation der Kundenstandorte durch Mittelpunkte eines Postleitzahlgebiets soll in Abschnitt 5.2 erfolgen.

4.2.2 Standortschätzung durch Geradenschnittpunkte

Eine weitere Möglichkeit zur Approximation von Standorten bei unzureichender Datenlage ist ein graphisches Vorgehen. Der zugrundeliegende Ansatz geht auf den Gedanken zurück, dass Flächen zur besseren Einteilung unterteilt werden können. Da Postleitzahlgebiete annähernd eine ebene Fläche darstellen, ist es daher möglich diese Gebiete einfach durch graphische Hilfsmittel aufzuteilen. Das graphische Hilfsmittel soll in diesem Fall ein Einzeichnen von Rechtecken und Geraden sein. Die Geraden entsprechen hierbei den Mittellinien von Rechtecken in horizontaler und vertikaler Richtung. Eine Mittellinie besteht aus der Verbindung der Mittelpunkte zweier gegenüberliegender Kanten eines Rechtecks. Durch eine Konstruktion der Mittellinien beider Kanten eines Rechtecks ist es möglich, die Fläche des Rechtecks in vier gleich große Bereiche aufzuteilen, die wiederum Rechtecke darstellen. Auf ein Postleitzahlgebiet übertragen bedeutet das, durch ein Rechteck, welches das betrachtete Gebiet umschließt, und die Mittellinien der Kanten kann jenes Gebiet in vier Teilflächen unterteilt werden. Die neuen Rechtecke, die die Teilflächen abgrenzen, können anhand von Mittellinien erneut aufgeteilt werden. Der Vorteil eines solchen graphischen Unterteilens ist eine gleichmäßige Aufteilung eines Gebietes, das quasi grenzenlos fortgeführt werden kann.

Die Approximation der Standorte erfolgt bei dieser Aufteilung durch die Schnittpunkte der konstruierten Mittellinien. Ein Schnittpunkt, welcher durch die von den Kanten ausgehenden Mittellinien eines Rechtecks entsteht, soll die Annäherung für einen gesuchten Standort darstellen. Dabei umschließt das äußerste Rechteck, welches die Grundlage zur weiteren Aufteilung bildet, das betrachtete Postleitzahlgebiet dermaßen, dass die maximalen Ausprägungen der Längen- und der Breitengrade an das Rechteck grenzen. Das bedeutet, das äußerste Rechteck wird durch die nördlichsten, südlichsten, westlichsten und östlichsten Punkte eines Postleitzahlgebiets definiert. Eine einfache Aufteilung des Rechtecks durch die Mittellinien generiert genau einen Geradenschnittpunkt, welcher eine Approximation eines Standortes verkörpert. Weitere Schnittpunkte und damit weitere Approximationen werden durch eine fortführende Aufteilung der Rechtecke erreicht. In der folgenden Abbildung 5 wird beispielhaft die Aufstellung von Mittellinien, die von einem umschließenden Rechteck des

betrachteten Gebiets aus konstruiert werden, veranschaulicht. Dabei wird die Anfertigung eines einzelnen und dreier Schnittpunkte durch sich schneidende Mittellinien gezeigt.

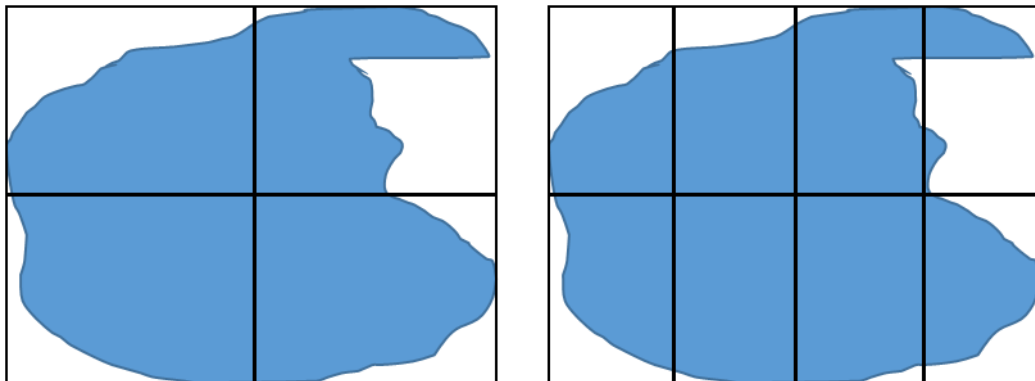


Abbildung 5: Konstruktion von Mittellinien zur Standorterschätzung

Die weitere Aufteilung muss nicht zwingend ein vollständiges Unterteilen aller Rechtecke durch alle Mittellinien beinhalten. Durch eine Beschränkung auf entweder die horizontalen oder die vertikalen Mittellinien kann eine abgeschwächte Aufteilung erzielt werden, die dennoch neue Schnittpunkte erzeugt (vgl. Abbildung 5). Eine solche Aufteilung hat den Vorteil, nur zwei und nicht direkt vier neue Schnittpunkte zu schaffen. Dadurch kann die Anzahl der anzunähernden Standorte genauer getroffen werden. Werden durch eine Aufteilung der Rechtecke mehr Schnittpunkte erzeugt als zur Approximation der Standorte benötigt werden, so werden die Punkte als Annäherung bevorzugt, die näher am absoluten Mittelpunkt, d.h. am Schnittpunkt der Mittellinien des umschließenden Rechtecks, lokalisiert sind. Sollten für den Fall mehrere Schnittpunkte in Frage kommen, können zuerst die Punkte ausgewählt werden, die auf den Mittellinien des äußersten Rechtecks liegen. Übriggebliebene Schnittpunkte werden dann ignoriert. Die entstehenden Rechtecke werden so lange weiter aufgeteilt, bis genügend Annäherungen für die gesuchten Standorte vorhanden sind. Bei dem Verfahren kann es vorkommen, dass Schnittpunkte von Mittellinien nicht im untersuchten Postleitzahlengebiet liegen. Der Grund dafür ist die Form des jeweiligen Gebiets, welches das umschließende Rechteck nicht voll ausfüllt. Solche außenliegenden Schnittpunkte können keine Approximation eines Standorts darstellen und sollen daher nicht weiter betrachtet werden. Die Annäherung der Kundenstandorte durch Mittellinienschnittpunkte eines Rechtecks soll praktisch in Abschnitt 5.3 vorgenommen werden.

4.2.3 Standortannäherung durch Tourenabbildung

Der dritte Ansatz ist wie die zuvor beschriebene Methode ein graphisches Lösungsverfahren, das auf einen Spezialfall einer unzureichenden Datenlage zugeschnitten ist. Diese Methode ist bei seiner Anwendung daher auf bestimmte Voraussetzungen angewiesen. Die Entwicklung des Ansatzes zur Standortapproximation geht auf mögliche Transportmethoden einer Supply Chain zurück, wie sie in den Abschnitten 2.1 und 2.2. aufgezeigt und erläutert wurden. Demnach werden zur Versorgung von Kunden einer Supply Chain in Nahverkehrsbereichen Touren und Routen geplant. Die Touren, d.h. welche Standorte bei einer Fahrt beliefert werden, und die Routen, also

die Reihenfolge der Standorte innerhalb einer Tour, sollen für diesen speziellen Fall der Standortapproximation als bekannt vorausgesetzt werden. Das Gleiche gilt für die zurückzulegenden Distanzen zwischen den Standorten einer Tour. Die Touren zu den Kundenstandorten können auf unterschiedlicher Art und Weise erfolgen. In Abschnitt 2.2. wurden verschiedene Möglichkeiten, wie das allgemeine Transportkonzept einer Supply Chain aussehen kann, beschrieben. Um das im Folgenden aufgezeigte Approximationsverfahren anwenden zu können, soll das Transportkonzept der zugrundeliegenden Supply Chain eine bestimmte Form aufweisen. Das hat den Hintergrund, dass diese graphische Vorgehensweise durch Eigenschaften bestimmter Transportmethoden determiniert wird und daher nicht bei jedem möglichen Transportkonzept, das in einer Supply Chain verwendet wird, anwendbar ist. Die für das Verfahren vorausgesetzte Transportmethode ist der Einsatz von Milk Runs zur Versorgung der Nahverkehrsbereiche. Die Milk Runs erfolgen von Depots aus, deren genaue Lokalisierung analog zu den Kundenstandorten nicht bekannt ist. Die Form der überregionalen Versorgung hat keinen Einfluss auf das hier behandelte Verfahren. Bei allen grundlegenden Strukturen, die den überregionalen Transport definieren, werden Depots verwendet. Der Einsatz von Milk Runs ist ein weit verbreitetes Konzept, welches bei vielen Supply Chains verwendet wird (vgl. Abschnitt 2.2). Daher hat dieses auf einen bestimmten Fall zugeschnittene Approximationsverfahren dennoch seine Relevanz bei der Annäherung von Kundenstandorten einer Supply Chain.

Die Zusammenstellung der Touren, die abzufahrende Reihenfolge und die zurückzulegenden Distanzen zwischen den Standorten sollen im Rahmen der Untersuchung einer fiktionalen Supply Chain bekannt sein. Für die vorliegende Problemstellung wurde davon ausgegangen, dass die Lage der Standorte unbekannt ist und nicht durch Erhebungen nachträglich ausgemacht werden kann. Die Postleitzahlen der Standorte sind gegeben, aber zu den anderen Informationen zu der Supply Chain wurden keine grundlegenden Angaben gemacht. Es kann deshalb angenommen werden, dass Daten, die nicht direkt die Position der Standorte betreffen bekannt sind. Dazu gehören die Distanzen zwischen den Standorten, die zum Beispiel im Rahmen einer Fahrtkostenkalkulation ermittelt wurden. Die Touren und Routen werden wie in Abschnitt 2.2. beschreiben normalerweise täglich neu geplant, um Auftrags- und Bedarfsschwankungen optimal begegnen zu können. Im vorliegenden Fall sind aber keine Daten zu den Bedarfen der Kunden gegeben, daher wird von einem Ausschnitt der täglichen Planung ausgegangen, bei dem ohne eine konkrete Nachfrage die Standorte abgefahren werden. Die Touren und Routen können ebenfalls der Fahrtkostenkalkulation entstammen, aus der die Distanzen genommen werden. Die expliziten Werte der Daten zu den Touren, Routen und Distanzen sollen für den vorliegenden Fall durch den Referenzdatensatz ermittelt werden. Die Ermittlung der Werte erfolgt später im Vorfeld der Anwendung des Verfahrens.

Ist eine Nahverkehrsbeförderung durch Milk Runs gegeben und die Distanzen zwischen den Standorten sind bekannt, so ist es möglich, die Standorte mit einem graphischen Vorgehen in einem Postleitzahlengebiet zu verteilen. In der Abbildung 5 wird das Vorgehen des graphischen Verfahrens schrittweise dargestellt. Dazu wird zuerst ein Standort, der mit einem Depot verbunden ist, beliebig in das betrachtete Gebiet gelegt. Um diesen Standort wird dann ein virtueller Kreis mit einem Radius gezogen, welcher der Entfernung zum zweiten Standort entsprechen soll. Der zweite Standort wird daraufhin beliebig auf den zuvor angelegten Kreis des

ersten Standortes gesetzt. Zweiter Standort bedeutet hierbei, dass dieser Kundenstandort einem mit dem Depot verbundenen Standort nachfolgt, also der zweite zu beliefernde Standort innerhalb eines Milk Runs ist. Sind zwei Standorte in die Fläche eingetragen, werden diese durch eine virtuelle Gerade miteinander verbunden. Um den neuen Standort wird wiederum ein virtueller Kreis mit einem Radius gleich der Entfernung zum nächsten Standort der Tour gezogen. Die Lokalisierung der weiteren Standorte erfolgt nach demselben Prinzip. Bei dem Einsetzen der Standorte in die Fläche ist zu beachten, dass sich die Verbindungslinien zwischen den einzelnen Standorten nicht überschneiden. Dies ist eine Einschränkung, die auf die generelle Route eines Milk Runs zurückzuführen ist. Wie in Abschnitt 2.2 beschrieben, werden in der Regel bei Milk Runs die kürzesten Wege zwischen benachbarten Standorten abgefahren, um insgesamt die kürzeste mögliche Route zu erhalten. Ein Überschneiden der Geraden würde aber bedeuten, dass zwischen zwei nicht direkt verbundenen Standorten eine kürzere Distanz vorhanden ist, als zwischen den verbundenen Standorten. Die Festlegung auf einen Milk Run als Distributionsmethode der Supply Chain hat eine weitere Folge für die Anordnung der Kundenstandorte. In einem Milk Run sind der Ausgangs- und der Endpunkt einer Tour identisch. Beide Punkte entsprechen dem beliefernden Depot, d.h. der erste und der letzte Kundenstandort der Tour müssen folglich mit dem Depot verbunden sein. Zusätzlich zu den bereits vorhandenen Kreisen müssen deshalb bei den ersten und letzten Kundenstandorten noch jeweils ein weiterer Kreis konstruiert werden. Die Kreise haben als Radius die jeweilige Distanz des Standorts zum Depot. Auf diesen Kreisen muss sich demnach das Depot befinden. Da es nur ein Depot für jeden Milk Run gibt, müssen sich die zweiten Kreise des ersten und des letzten Kundenstandorts in einem Punkt schneiden. An diesem Schnittpunkt muss sich das Depot befinden. Die Anordnung der Standorte wird also bildlich selbst einem Kreis oder einer Ellipse ähnlich sein.

Kommt unter den genannten Bedingungen kein Schnittpunkt der Kreise mit Depot-Distanz-Radius des ersten und des letzten Kundenstandorts zustande, so müssen die Lokalisierungen der Standorte verändert werden. Hierbei soll von dem letzten Standort der Tour ausgegangen werden. Dieser soll solange auf dem Kreis seines Vorgängers bewegt werden, bis die Bedingungen erfüllt sind. Ist es trotz des Verschiebens des Punktes nicht möglich unter den Gegebenheiten einen Schnittpunkt zu erreichen, so wird der vorletzte Standort umgelegt. Der letzte Standort verändert dabei analog seine Position, wobei der ursprüngliche Winkel zwischen den beiden Geraden, die mit dem vorletzten Standort verbundenen sind, erhalten bleiben soll. Der vorletzte Standort wird solange auf dem Kreis seines Vorgängers bewegt, bis unter den gegebenen Bedingungen ein Schnittpunkt der Kreise mit der Distanz zum Depot als Radius des ersten und des letzten Standortes möglich ist. Ist dies erneut nicht möglich, wird die Erhaltung des Winkels aufgegeben und der letzte Standort auf dem Kreis seines Vorgängers ebenfalls bewegt. Eine Änderung der vorherigen Standorte erfolgt nach demselben Prinzip, falls ein Schnittpunkt durch eine Änderung der letzten beiden Standorte nicht zu realisieren ist. Ist eine Anordnung der Standorte unter Einhaltung der Bedingungen, d.h. ohne sich überschneidende Geraden zwischen den Punkten, aber mit sich überschneidenden Depot-Distanz-Kreisen des ersten und des letzten Kundenstandortes gefunden, so gilt diese Konstellation für das vorliegende Verfahren als eine Approximation der Kundenstandorte.

In der folgenden Abbildung wird das Vorgehen der Standortapproximation durch eine graphische Tourenaufstellung veranschaulicht. Dabei werden die einzelnen Schritte des Verfahrens bis zu einer Konstellation mit drei Standortapproximationen bebildert. In dem letzten Kästchen der Abbildung wird eine Überschneidung zweier Geraden gezeigt, was es bei der Methode zu verhindern gilt.

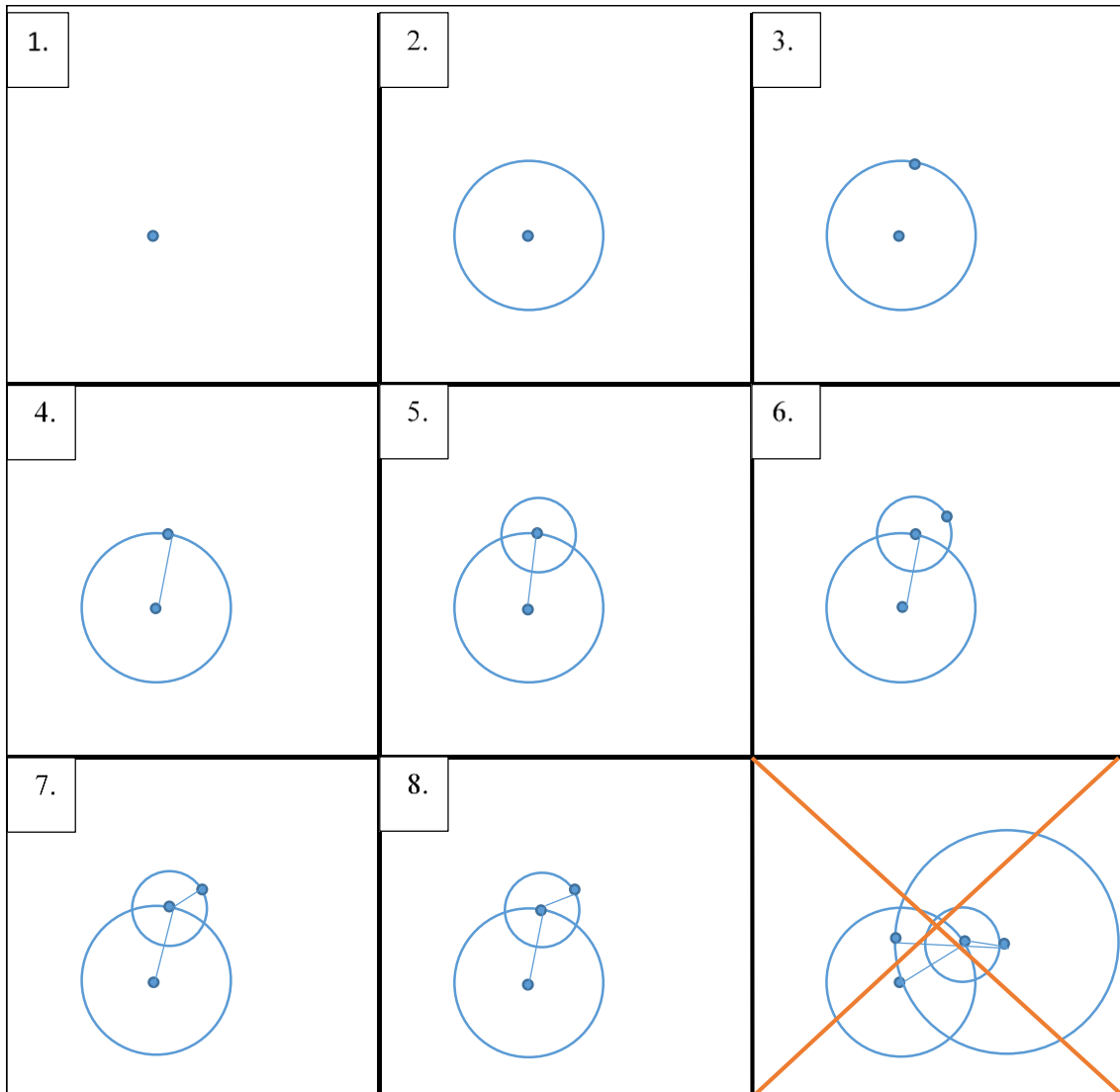


Abbildung 6: Vorgehen bei der graphischen Tourenabbildung

In diesem Zusammenhang muss festgelegt werden, wie viele Standorte in einem Milk Run maximal abgefahren werden dürfen, damit das Verfahren anwendbar bleibt und keinen zu großen Aufwand erzeugt, der grundlegend gegen eine Nutzung sprechen würde. In Abschnitt 2.2 wurde erläutert, dass Milk Runs im Allgemeinen Beschränkungen unterliegen, die den Umfang einer Tour bestimmen. Zu den genannten Faktoren wurde keine Aussage getroffen. Es lässt sich aber aus der Menge der Kundenstandorte, die in einem Postleitzahlengebiet zu approximieren sind, die Kundendichte ableiten bzw. festlegen, wie viele Kunden maximal in einer Tour beliefert werden sollen. Aus dem Referenzdatensatz ist ersichtlich, dass zwischen einem und 19 Standorte in einem Postleitzahlengebiet lokalisiert sind (vgl. Anhang 1). Der Großteil der Gebiete beinhaltet einen bis vier Standorte. Aufgrund der Anzahl und der Aufteilung der Standorte bietet es sich an, die

Grenze auf vier Standorte pro Tour zu setzen. Das Verfahren selber ist grundlegend nicht durch eine gewisse Anzahl an Standorten pro Tour beschränkt. Allerdings würde eine zu große Anzahl an Standorten in einer Tour die graphische Abbildung deutlich erschweren. Deshalb wird die zugrundeliegende Beschränkung der Standorte pro Milk Run auf vier gesetzt. Bei der Anzahl kann das Verfahren ohne Probleme angewendet werden. Es können theoretisch aber auch weniger Standorte angefahren werden. Die minimale Anzahl an Standorten ist mit zwei beziffert, um noch als Milk Run zu gelten. In Gebieten mit mehr als vier Standorten müssen durch die vorgenommene Beschränkung mehrere Touren gefahren werden.

Durch das prinzipielle Vorgehen ist bereits ersichtlich, dass dieses Verfahren nicht für jede beliebige Anzahl an zu approximierenden Standorten in einem Postleitzahlengebiet geeignet ist. Im Fall nur eines Standortes in einem Postleitzahlengebiet würde die Annäherung zufällig durch den Anwender in der Fläche platziert werden. Das eigentliche Verfahren bleibt dabei ungerührt, da keine weiteren Standorte, und damit keine Tour im herkömmlichen Sinn vorhanden sind. Bei zwei zu approximierenden Standorten in einem Postleitzahlengebiet ist wiederum ein Milk Run möglich. Dadurch kann das Verfahren grundlegend angewendet werden. Eine Beachtung der Geraden zwischen den Punkten kann hier aber vernachlässigt werden, da bei nur einer potentiellen Geraden keine Überschneidung möglich ist. Eine vollständige Anwendung dieses Schnittpunktmodells ist deshalb erst ab drei Standorten in einem Postleitzahlengebiet möglich. Eine Lösung dieses Problems würde die Ausweitung des Verfahrens auf mehrere Postleitzahlengebiete bieten. Werden mehrere Postleitzahlengebiete mit jeweils einem oder zwei zu approximierenden Standorten zusammengefasst, kann das Modellverfahren auf dieses vergrößerte Gebiet angewendet werden. Dabei muss beachtet werden, dass die Platzierung eines Standortes in dem zugehörigen Postleitzahlengebiet erfolgt, in dem der jeweilige Standort zu approximieren ist. Diese Form der Zusammenlegung bietet sich für Postleitzahlengebiete mit mehr als zwei Standorten nicht an, da hier das eigentliche Standortschätzungsverfahren für die einzelnen Gebiete angewendet werden kann. Allerdings ist es auch möglich, dass Touren in Postleitzahlengebieten mit genügend Standorten, sprich drei Standorten, durch einen Standort eines anderen Gebietes ergänzt werden, falls für diesen einzelnen Standort keine andere sinnvolle Unterbringung in einer Tour möglich ist. Die Standortapproximation durch eine graphische Abbildung eines Milk Runs soll in Abschnitt 5.4 vorgenommen werden.

Zusammengefasst wurden in diesem Kapitel drei Ansätze zur Standortapproximation entwickelt und beschrieben. Der erste entwickelte Ansatz soll die Standorte durch einen Mittelpunkt pro Postleitzahlengebiet annähern. Hierbei sollen zwei Arten von Mittelwerten zur Berechnung der Mittelpunkte benutzt werden. Der zweite Ansatz besteht darin, mittels der Schnittpunkte von Mittellinien die Standorte abzuschätzen. Der dritte Ansatz soll die Kundenstandorte durch Eigenschaften eines Milk Runs approximieren, indem die Standorte graphisch der gegebenen Tour angenähert werden. Diese drei Ansätze sollen nun im nachfolgenden Kapitel auf ein konkretes Problem einer Supply Chain mit unbekanntem Kundenstandorten angewendet und dadurch getestet werden.

5 Anwendung von Approximationsverfahren

In den Kapiteln 2 und 3 wurden die Grundlagen zu Supply Chains und ihren Transportmethoden, Simulationsmodellen und der Distanzmessung zwischen verschiedenen Punkten geschaffen, sowie Verfahren zum Umgang mit fehlenden Daten analysiert. Hier wurde unter anderem die Möglichkeit, Standorte durch eine Gleichverteilung in einer Fläche zu platzieren, aufgezeigt. Aufbauend auf den Grundlagen und den diskutierten Methoden zur Datenergänzung wurden in Kapitel 4 Verfahren zur Standortapproximation entwickelt. Im Vorfeld wurde das vorliegende Problem detailliert aufgezeigt und ein Referenzdatensatz geschaffen, auf den sich die weiteren Arbeiten beziehen sollen. Aus dem Problem und dem Datensatz wurde dann eine Fallstudie entwickelt, die die Grundlage zur Anwendung der Verfahren darstellt. In diesem Kapitel sollen nun neben der Methode der gleichverteilten Platzierung von Standorten in der Fläche die selbst entwickelten Ansätze für das konkretisierte Problem der Standortapproximation zum Einsatz kommen. Anschließend werden die Ergebnisse jeder Methode ausgewertet und mit dem in Abschnitt 4.1 vorgestellten Bewertungsverfahren beurteilt, um diese nachfolgend vergleichen zu können. Insgesamt vier Verfahren zur Approximation kommen hierbei zum Einsatz. Zuerst wird in Abschnitt 5.1 die Methode getestet, Standorte mittels einer Gleichverteilung in einem Postleitzahlengebiet zu verteilen. Danach folgt in Abschnitt 5.2 die Approximation von Standorten durch einen Mittelpunkt eines Gebietes. Nachdem in Abschnitt 5.3 eine Annäherung der Standorte durch Schnittpunkte von Mittellinien vorgenommen wird, werden abschließend in Abschnitt 5.4 Standorte durch die graphische Darstellung eines Milk Runs approximiert.

5.1 Gleichverteilung von Standorten

Der erste Ansatz, der auf das konkretisierte Problem einer unzureichenden Datengrundlage hinsichtlich der Kundenstandorte in einer Supply Chain angewendet werden soll, ist die Standortapproximation durch gleichverteilte Zufallswerte. Das theoretische Vorgehen dazu wurde in Abschnitt 3.2 vorgestellt. Um die Standorte in der Fläche mit einer Gleichverteilung platzieren zu können, müssen vor der eigentlichen Anwendung Daten zu den relevanten Gebieten gesammelt werden. Die Daten werden benötigt damit der Zufallsgenerator für die Postleitzahlengebiete adäquate Werte zur Approximation erzeugen kann. Deshalb wurden zunächst für jedes Postleitzahlengebiet, in dem Standorte angenähert werden sollen, Koordinaten ermittelt, die das Gebiet eindeutig definieren. Diese Koordinaten sollen die Grenzen des betrachteten Gebiets darstellen. Die das Postleitzahlengebiet kennzeichnenden Koordinaten sind also durch alle Punkte und Wege definiert, die in der Onlineplattform OpenStreetMap das jeweilige Gebiet abgrenzen. Die Ermittlung dieser Koordinaten erfolgte ähnlich zu der Datensatzgenerierung in Abschnitt 4.1 durch eine entsprechende Abfrage bei dem auf OpenStreetMap basierenden Internet Tool Overpass Turbo. Die Abfragen zu den aus den Punkten und Wegen bestehenden Relationen wurden zur besseren Abgrenzung für jedes Postleitzahlengebiet einzeln vorgenommen. Die relevanten Postleitzahlengebiete sind aus dem in Abschnitt 4.1 erstellten Datensatz, der als Referenzgröße dient, ersichtlich. Die Genauigkeit der

hier erhaltenen Koordinaten ist von der Qualität der Datenpflege der Onlineplattform abhängig. Es ist deshalb möglich, dass kleine Abweichungen zum tatsächlichen Fall auftreten können.

Zur eigentlichen Generierung der Zufallszahlen, die die gesuchten Standorte approximieren sollen, wird ein Zufallsgenerator genutzt. Für den vorliegenden Fall wird der im Internet frei zugängliche Zufallsgenerator von Zufallsgenerator.net verwendet. Dieser Zufallsgenerator bietet sich für die Anwendung in diesem Zusammenhang an, da hier die Zahlen durch eine Gleichverteilung generiert werden. Darüber hinaus ist es möglich, den Wertebereich zur Zufallszahlenerstellung für den jeweiligen Bedarf anzugleichen. Indem der zugrunde liegende Wertebereich des Zufallsgenerators dem Wertebereich, der die Grenzen darstellenden Koordinaten des jeweiligen Postleitzahlengebietes umfasst, angepasst wird, kann dieser Generator passende Zufallszahlen für das Gebiet erzeugen. Weil es nicht möglich ist, die gesamte Anzahl an kennzeichnenden Werten, und damit eine genaue Abgrenzung des Postleitzahlengebiets, als Wertebereich des Zufallsgenerators anzugeben, wird sich hierfür auf die wichtigsten Werte beschränkt. Die wichtigsten Werte sind in diesem Fall die Extrempunkte eines Gebietes, d.h. die maximalen und minimalen Ausprägungen des Längen- und des Breitengrades eines Gebiets. Diese Ausprägungen sollen die Grenzen des Wertebereichs für den Zufallsgenerators darstellen. Die ursprüngliche Punktierung der Koordinaten in Dezimalzahlen wird für die Wertebereiche der Zufallszahlen aufgehoben, um Fehler bei der Dezimaltrennung der erzeugten Zahlen zu vermeiden. Die erzeugten Zufallszahlen können anschließend durch eine entsprechende Punktierung leicht in Koordinaten in Dezimalzahlen umgewandelt werden. Der Zufallsgenerator kann nur eine Zahl in einem Durchgang generieren. Koordinaten bestehen aber mit dem Längen- und dem Breitengrad aus zwei Werten, die die Lage eindeutig charakterisieren. Das bedeutet, dass pro Standort dementsprechend zwei Zufallszahlen erzeugt werden müssen. Die erste Zufallszahl soll hierbei den Längengrad annähern, während die zweite dem Breitengrad entsprechen soll. Daher müssen die Wertebereiche des Zufallsgenerators jeweils für den Längen- und den Breitengrad durch die Extremwerte des Gebiets angepasst werden. Zusammen ergeben dann die generierten Zufallszahlen die Koordinaten eines Standorts. Es ist bekannt, wie viele Standorte in einem Postleitzahlengebiet vorhanden und damit zu approximieren sind. Entsprechend viele Koordinaten bzw. doppelt so viele Zufallszahlen müssen hier nach dem oben beschriebenen Vorgehen in den Wertebereichen eines Gebiets erzeugt werden. Für jedes Postleitzahlengebiet der Supply Chain wird dieses Vorgehen wiederholt.

Wegen der Beschränkung des Wertebereichs durch die Extrempunkte eines Postleitzahlengebiets kann es vorkommen, dass die erzeugten Koordinaten, oder einer der beiden Werte, außerhalb des betrachteten Gebietes lokalisiert sind. Der Grund dafür liegt in der Form des Gebietes. Die Extrempunkte bilden zwar grob die Postleitzahlengebiete in alle Richtungen ab, aber die Gebiete können durch geographische oder verwaltungstechnische Besonderheiten unterschiedliche Formen annehmen, und daher auch Buchten, Spitzen oder ähnliches enthalten. Diese Besonderheiten können dazu führen, dass generierte Koordinaten trotz Einhaltung des Wertebereichs außerhalb des eigentlichen Gebiets liegen. Daher werden alle erzeugten Punkte auf ihre Lage im Postleitzahlengebiet hin überprüft. Befinden sich beide koordinatenbildende Werte im Postleitzahlengebiet so stellt dieser Punkt eine Approximation eines Standorts dar. Befindet sich der Punkt aufgrund eines Koordinatenwertes außerhalb des Gebiets, so wird für diesen Wert

eine neue Zufallszahl im entsprechenden Wertebereich erzeugt. Diese Vorgehensweise wird für alle generierten Koordinaten wiederholt, bis genug Punkte als Standortapproximationen vorhanden sind. Die erzeugten Zufallszahlen bzw. Koordinaten befinden sich in Anhang 2.

Damit die untersuchten Ansätze ausgewertet und anschließend untereinander verglichen werden können, muss für jedes Verfahren die Abweichung zur Realität bzw. zum erzeugten Referenzdatensatz geprüft werden. In Abschnitt 4.1 wurde dazu eine Formel aufgestellt, mit der die absolute Abweichung der Ergebnisse eines Verfahrens von den Referenzdaten berechnet wird. Für eine einfache Umsetzung dieser Berechnung wird schrittweise vorgegangen. Zuerst wurden die geschätzten Standorte eines Postleitzahlengebiets den realen gegenübergestellt, um anhand ihrer Distanzen zueinander eine Zuordnung zu erreichen. Die Distanzen wurden wie in Abschnitt 2.4 beschrieben mit der einfachen euklidischen Distanz (2) ermittelt. Die generierten Standorte wurden den realen Standorten zugeordnet, zwischen denen die geringste Entfernung vorhanden ist. Dieser Schritt entspricht der Berechnung des $\min d_{ij}(x, y)$. Bei Postleitzahlengebieten mit nur einem zu approximierenden Standort entfällt der Schritt der Zuordnung. Nachdem allen geschätzten Standorten ein reales Gegenstück zugewiesen wurde, erfolgt die Berechnung der Abweichung. Dies geschah mit Hilfe von Excel. Durch eine tabellenartige Auflistung der Standorte kann die Berechnung hierbei vereinfacht werden. Alle Werte, die Abweichungen und damit die Distanzen zwischen realen und geschätzten Standorten verkörpern, wurden zur besseren Vergleichbarkeit bis auf eine Ziffer nach der Dezimaltrennung gerundet. Die Ergebnisse der Auswertung des Ansatzes sind für eine übersichtliche Darstellung in einem Box-Whisker-Plot in Abbildung 7 festgehalten.

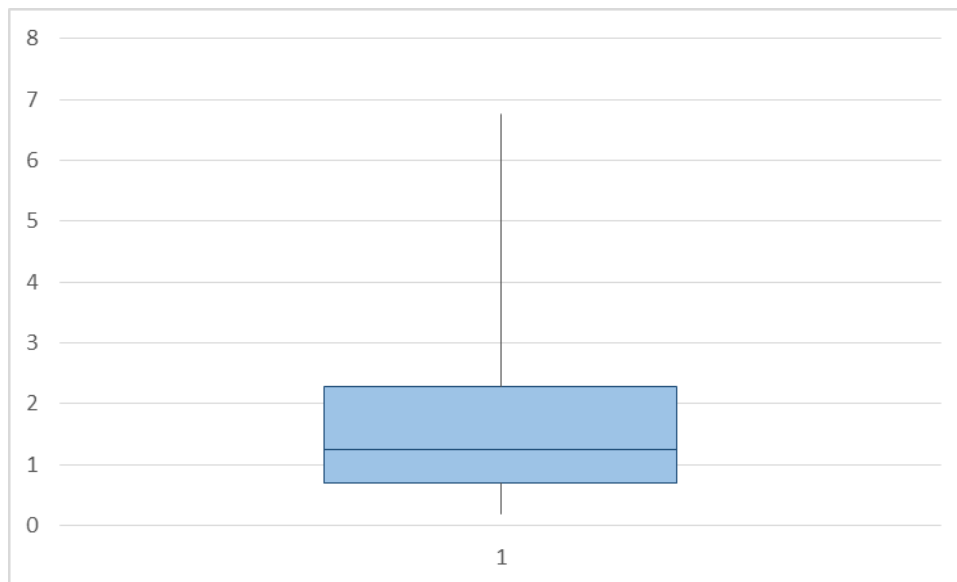


Abbildung 7: Boxplot der Abweichungen bei einer Gleichverteilung der Standorte

Der Ansatz, Standorte mit einer Gleichverteilung in einem Postleitzahlengebiet zu verteilen, liefert bei der Anwendung auf die konkrete Fallstudie starke Unterschiede in der Genauigkeit der Approximationen. Die minimale Abweichung eines geschätzten Standorts zu seinem entsprechenden realen Gegenstück ist mit 198 Metern vermeintlich gering. Allerdings steht dieser Abweichung eine maximale Abweichung von mehr als 6,7 km gegenüber. Bezogen auf die Größe einiger Postleitzahlengebiete, welche zum Teil keine fünf Quadratkilometer Fläche haben, ist

diese Abweichung immens. Die aus den Extremwerten resultierende Spannweite verdeutlicht mit einer Größe von ca. 6,5 km wie stark die Streuung der Ergebnisse in ihrer Genauigkeit ist. Werden die Abweichungen pro Postleitzahlengebiet betrachtet, so ist keine Tendenz zu erkennen, dass die Streuung durch die Anzahl der zu approximierenden Standorte beeinflusst wird. Durch die größere Fläche einiger Gebiete kann die Abweichung der Standorte in absoluten Zahlen entsprechend größer ausfallen, allerdings ist auch hier keine Tendenz festzustellen, dass in größeren Gebieten überwiegend schlechtere Ergebnisse erzielt werden. Bei der Untersuchung der Abweichung im Verhältnis zur Größe des Gebiets sind also keine Auffälligkeiten zu erkennen. Die Güte der Approximation ist relativ zum Flächeninhalt eines Postleitzahlengebiets in etwa gleich. Im Durchschnitt haben die geschätzten Standorte eine Abweichung von knapp 1,7 km zu Realität. Der Median der Werte liegt hier nur bei ca. 1,3 km. Die Hälfte der Ergebnisse hat also eine Abweichung von unter 1,3 km. Ist der Median unter dem arithmetischen Mittel angesiedelt, so zeigt dies, dass einige Ergebnisse in ihren Abweichungen so große Werte annehmen, dass der Durchschnitt deutlich negativ beeinflusst wird. Dass das dritte Quartil der Abweichungen bei gut 2,3 km liegt verdeutlicht, dass sehr starke Abweichungen von mehr als drei Kilometern Differenz zur Realität bei diesem Verfahren selten vorkommen. Dennoch sind solche Abweichungen möglich und diese können mitunter beträchtliche Werte annehmen. Die Möglichkeit der Standortapproximation durch gleichverteilte Zufallszahlen ist also ein einfaches Verfahren, welches überwiegend gute Ergebnisse generiert, aber auch durch eine starke Streuung gravierende Fehlschätzungen erzeugen kann. Insgesamt haben die 110 angenäherten Standorte dieses Ansatzes eine absolute Abweichung von 185,6 km zu den tatsächlichen Lokalisierungen.

5.2 Mittelpunkte als Standorte verwenden

Der zweite zu testende Ansatz ist die Standortapproximation durch einen zentralen Punkt einer Fläche. In anderen Worten bedeutet dies eine Annäherung der gesuchten Standorte innerhalb einer Fläche durch einen mathematischen Mittelpunkt dieses Gebiets. Für das vorliegende Problem sollen zwei Arten von Mittelpunkten getestet werden. Beide werden durch die Berechnung eines Mittelwerts bestimmt. Es soll zum einen das arithmetische Mittel und zum anderen der Median für jedes Gebiet ermittelt werden. Die grundlegenden Formeln zur Berechnung dieser Mittelpunkte wurden in Abschnitt 3.3.3 vorgestellt. Um diese Mittelpunkte für den vorliegenden Fall zu bestimmen, werden Daten für jedes der zu betrachtenden Postleitzahlengebiete benötigt. Zur Berechnung der Mittelwerte können unterschiedliche Datenmengen eines Datensatzes verwendet werden. Der zugrundeliegende Datensatz wird, wie im vorherigen Abschnitt 5.1 beschrieben, durch eine Abfrage bei Overpass Turbo generiert. Es gibt hier die Möglichkeiten, sämtliche in OpenStreetMap hinterlegten Daten zu einem Postleitzahlengebiet abzufragen oder sich auf eine Teilmenge der Daten zu beschränken. Die Teilmenge kann beispielsweise nur die Relationen, die die vollständigen Grenzen eines Postleitzahlengebiets darstellen, beinhalten oder nur durch besondere Werte ausgedrückt werden. Dies können z.B. die Extremwerte der Längen- und Breitengrade sein. Für die Untersuchung des Ansatzes wurden zwei Varianten von Teilmengen der Daten verwendet. Zum einen wurden die als Grenzen fungierenden Relationen ausgewählt, zum anderen wurde mit der Beschränkung auf die Extremwerte in horizontaler und vertikaler Richtung gearbeitet. Dies hat den Hintergrund, die

Spannbreite des Verfahrens durch möglichst viele und möglichst wenige Daten als Berechnungsgrundlage aufzuzeigen. Durch eine tabellarische Auflistung der jeweiligen Daten in Längen- und Breitengrad pro Postleitzahlengebiet lassen sich die Berechnungen der Mittelwerte in Excel anschließend einfach durchführen. Für jede Spalte, also für jede gesammelte Menge an entweder Längen- oder Breitengraden der durch die Abfrage gefundenen Punkte eines Postleitzahlengebiets, können durch die Eingabe der entsprechenden Formeln und die Begrenzung auf die betrachtete Spalte die Mittelwerte berechnet werden. Es werden sowohl für den vollständigen, als auch für die Teilmenge des Datensatzes beide Möglichkeiten zur Mittelwertberechnung angewendet.

Die berechneten Mittelpunkte stellen die Approximation für alle Standorte in dem jeweiligen Gebiet dar, d.h. alle Standorte in einem Gebiet werden durch den Mittelpunkt ausgedrückt. Deshalb kann bei der Auswertung des Verfahrens auf eine Zuordnung der geschätzten Standorte zu den tatsächlichen Standorten verzichtet werden. Für die 35 Postleitzahlengebiete wurden insgesamt 140 Mittelpunkte berechnet, welche sich aus jeweils 35 Werten, die aus dem arithmetischen Mittel und dem Median sowohl bei vielen und als auch bei wenigen Daten als Berechnungsgrundlage ermittelt wurden, zusammensetzen. Die genauen Koordinaten der Mittelpunkte befinden sich in Anhang 2. Das arithmetische Mittel und der Median geben für den vollständigen Datensatz bei allen untersuchten Gebieten unterschiedliche Werte aus. Bei der Nutzung der Extremwerte als Berechnungsgrundlage sind die beiden Mittelwerte allerdings identisch. Dies ist der mathematischen Formulierung des Medians geschuldet, die für größere Zahlenmengen ausgelegt ist und bei zwei zugrundeliegenden Werten auf die Formel des arithmetischen Mittels zurückgreift. Daher kann die Auswertung der Ergebnisse bei der Nutzung der Extremwerte als Berechnungsgrundlage auf das arithmetische Mittel beschränkt werden.

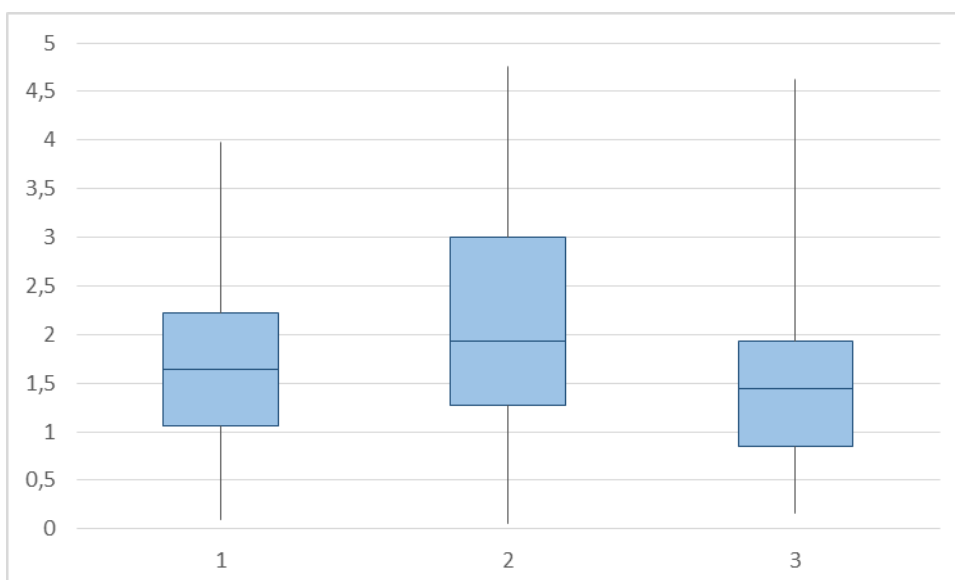


Abbildung 8: Boxplots der Abweichungen bei der Standortapproximation durch Mittelpunkte, 1 arithmetisches Mittel, 2 Median, 3 arithmetisches Mittel und Median bei minimaler Datengrundlage

Die Abweichungen der verschiedenen Methoden sind zum besseren Vergleich übersichtlich in Boxplot-Diagrammen in Abbildung 8 zusammengefasst. Damit die Berechnungsmethoden der

Standortapproximation sprachlich nicht mit den Berechnungsmethoden, die in der Auswertung angewendet werden, kollidieren, sollen im Folgenden die Annäherungsverfahren gemäß ihrer Anwendungsreihenfolge durchnummeriert werden. Die Verwendung des arithmetischen Mittels zur Mittelpunktberechnung, welche alle Relationen als Datengrundlage nutzt, soll als Verfahren 1 bezeichnet werden, die analoge Verwendung des Medians als Verfahren 2 und die Anwendung des arithmetischen Mittels zur Mittelpunktbestimmung auf die Extremwerte als Berechnungsgrundlage soll als Verfahren 3 gelten.

Die Bestimmung von Mittelpunkten gibt je nach angewandter Berechnungsmethode und zugrundeliegendem Datenumfang unterschiedlich gute Ergebnisse. Die geringste minimale Abweichung eines Mittelpunkts von einem realen Standort liefert das Verfahren 2 mit 51 m deutlich vor den anderen Verfahren. Aber auch deren geringste Abweichungen stellen mit 91 m (Verfahren 2) und 162 m (Verfahren 3) passable Approximationen dar. Demgegenüber finden sich aber bei jedem Verfahren auch schlechte Annäherungen. Am niedrigsten ist die maximale Abweichung bei Verfahren 1 mit knapp vier Kilometern. Das zeigt, dass die Spannweite und damit die Streuung der Abweichungen im Verhältnis gering ausfallen. Etwas anders sieht es bei Verfahren 2 und 3 aus. Hier sind die maximalen Ausprägungen gut 750 m bzw. 630 m größer. Die Spannweite der Verfahren ist dementsprechend ebenfalls größer. In den Verfahren 2 und 3 streuen die Ergebnisse also mehr als in Verfahren 1. Bei keinem der drei Verfahren ist eine Tendenz zu erkennen, dass die Anzahl der zu approximierenden Standorte oder die Größe des zugrundeliegenden Postleitzahlgebietes die Genauigkeit der Verfahren beeinflusst. Die Größe der Gebiete determiniert zwar die maximal mögliche Abweichung in absoluten Zahlen, aber da es sich hier um Mittelpunkte als Approximationen handelt sind relativ gesehen keine Unterschiede in der Güte der Annäherungen bei größeren und kleineren Postleitzahlgebieten erkennbar. Durch die Verwendung von Mittelpunkten werden aber an den Grenzen der Gebiete gelegene Standorte schlechter angenähert, als dem Zentrum nähergelegene Standorte. Im Schnitt haben die Mittelpunkte bei dem Verfahren 3 eine Abweichung von 1,4 km zu den realen Standorten. Bei den Verfahren 1 und 2 ist der Durchschnitt mit 1,7 km und 2,1 km klar größer. Die Mediane der Verfahren weichen nur geringfügig von den arithmetischen Mitteln ab, d.h. es gibt bei keinem Verfahren so schwerwiegende Ausreißer, dass der Durchschnitt stark verzerrt wird. Allerdings sind bei Verfahren 2 über die Hälfte der Abweichungen über zwei Kilometer groß, während bei Verfahren 3 drei Viertel der Abweichungen unter dieser Grenze liegen. Ein Viertel der Abweichungen bei Verfahren 2 ist sogar größer als drei Kilometer. Das Verfahren 2 ist also in der Breite deutlich schlechter als die beiden anderen Möglichkeiten, obwohl es die insgesamt beste Approximation der drei Ansätze generiert hat. Verfahren 3 ist besser als Verfahren 1, da die Differenzen zur Realität insgesamt geringer ausfallen. Die Abweichungen der Mittelpunkte zu den realen Standorten betragen bei Verfahren 3 zusammengerechnet 156,3 km. Verfahren 1 hat mit einer absoluten Abweichung von 182,7 km eine klar größere Diskrepanz, während Verfahren 2 mit 227 km insgesamt ein sehr schlechtes Ergebnis hat. Bei der Bestimmung von Mittelpunkten anhand der mathematischen Berechnung von Mittelwerten sollte daher das arithmetische Mittel auf einen reduzierten Datensatz als Berechnungsgrundlage verwendet werden. Diese Variante bietet eine einfach anzuwendende Approximationsmöglichkeit, die relativ gute Ergebnisse erzielen kann.

5.3 Standortschätzung durch Geradenschnittpunkte

Als drittes soll der Ansatz der Standortapproximation durch Schnittpunkte der Mittellinien von Rechtecken untersucht werden. Nach der Herleitung und der Darlegung des theoretischen Vorgehens in Abschnitt 4.2.2 wird dieses Verfahren nun praktisch angewendet. Da dies ein graphisches Verfahren ist, werden für die Konstruktion von Rechtecken und Mittellinien zunächst Abbildungen der interessierenden Postleitzahlengebiete benötigt. Aus dem Referenzdatensatz ist ersichtlich, welche Gebiete betrachtet werden müssen. Die 35 relevanten Postleitzahlengebiete werden durch entsprechende Abfragen bei Overpass Turbo in jeweils eine Karte projiziert. Die Abfragen zielen nur auf die Relationen, die die Grenzen des Gebiets definieren, ab. Die Darstellungen der Postleitzahlengebiete können dann in ein Programm übertragen werden, mit dessen Hilfe sich das graphische Vorgehen umsetzen lässt. Eine einfache Lösung stellt das Programm Paint dar. Mit dem Programm können durch seine Zeichentools simpel die Rechtecke und die Mittellinien konstruiert werden. Durch ein voreingestelltes Raster, welches automatisch in den Hintergrund der Darstellung gelegt wird, kann die Erzeugung der Rechtecke und Linien vereinfacht werden.

Ausgangspunkt für jedes Postleitzahlengebiet ist die Anfertigung eines das Gebiet umschließenden Rechtecks. Dieses Rechteck soll an die horizontalen und vertikalen Spitzen des Gebiets, d.h. an die maximalen Ausprägungen in den vier Himmelsrichtungen, grenzen. Alternativ zu der Festlegung des umschließenden Rechtecks kann die Darstellung des Gebietes auch auf die maximalen Ausprägungen der Grenzen zugeschnitten werden. Anschließend können dann die Mittellinien konstruiert werden. Dazu wird für jede der vier Kanten der jeweilige Mittelpunkt gesucht und markiert. Zwischen den gegenüber liegenden Mittelpunkten zweier Kanten kann dann eine Gerade gezogen werden. Die hierdurch entstehenden Mittellinien schneiden sich in einem Punkt in der Mitte des Rechtecks. Dieser Schnittpunkt soll als erste Approximation eines Standorts gelten. Für Postleitzahlen mit nur einem zu approximierenden Standort endet das Verfahren mit der Lokalisierung des Schnittpunktes. Für die anderen Gebiete werden so lange die Rechtecke durch Mittellinien aufgeteilt, bis die Anzahl der Schnittpunkte der Anzahl der anzunähernden Standorte entspricht. In Abschnitt 4.2.2 wurde dieses Vorgehen der Aufteilung detailliert beschreiben. Eine Beschränkung bei der weiteren Aufteilung auf die Konstruktion nur von horizontalen oder nur von vertikalen Mittellinien ist dabei zulässig, um der Anzahl der zu approximierenden Standorten eher zu entsprechen. Schnittpunkte, die außerhalb des betrachteten Postleitzahlengebiets liegen, werden für die Approximation ignoriert. Durch die Approximation der Standorte in dem Zeichenprogramm ist keine direkte Lokalisation der Standorte möglich, d.h. die Koordinaten lassen sich nicht direkt ablesen. Deshalb ist es nötig, die erzeugten Schnittpunkte wieder in ein System zu transferieren, welches Koordinaten anzeigen kann. Hierbei wurde das Online Tool Tim-Online verwendet, welches Koordinaten bis auf die fünfte Stelle hinter der Dezimaltrennung genau anzeigt. Durch einen Abgleich mit den modifizierten Abbildungen der Postleitzahlengebiete konnten die Koordinaten der Schnittpunkte bestimmt werden. Weil durch das hier verwendete Programm zwei Nach Komma Stellen im Vergleich zu den Koordinatenwerten des Referenzdatensatzes fehlen, wurden diese durch zwei Nullen ergänzt. Damit waren die Koordinaten vergleichbar, ohne dass sich die Werte verändert haben.

Die Daten der geschätzten Standorte sind ausführlich in Anhang 2 hinterlegt. Die Anzahl der Schnittpunkte entspricht der Anzahl der zu approximierenden Standorte, deshalb muss zur Auswertung des Verfahrens eine Zuordnung der Standorte vorgenommen werden. Dies geschieht analog zu dem Vorgehen der Zuordnung, wie es in Abschnitt 5.1 beschrieben wurde. Die realen und geschätzten Standorte, die die geringste Entfernung zueinander haben, werden einander zugewiesen, um zu vermeiden, dass überflüssige Kombinationen die Auswertung verzerren. Anschließend erfolgt die Berechnung der Abweichungen der Annäherungen zur Realität. Das Vorgehen entspricht hierbei den im gleichen Abschnitt erklärten Schritten. Die daraus resultierenden Ergebnisse wurden zur Veranschaulichung in einem Boxplot-Diagramm in der folgenden Abbildung 9 festgehalten.

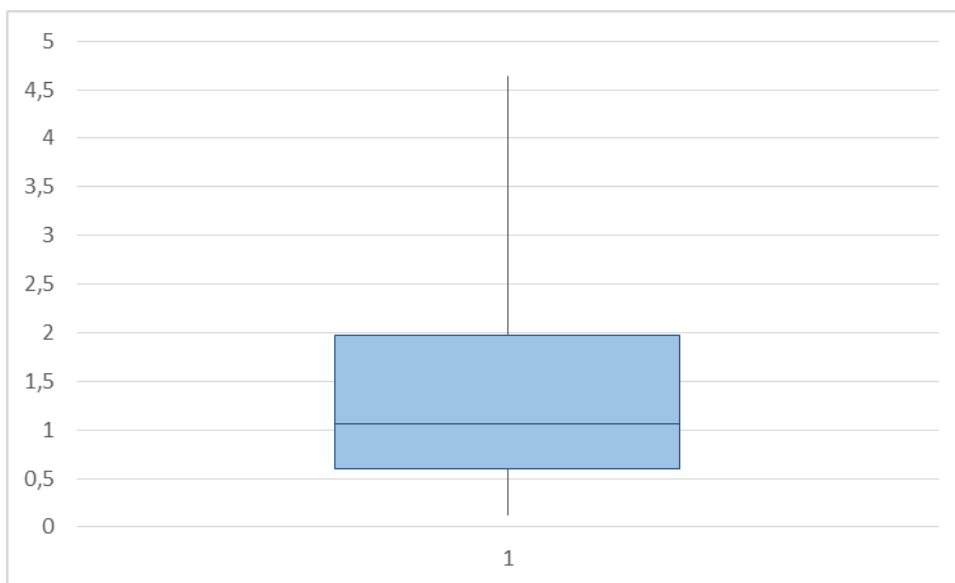


Abbildung 9: Boxplot der Abweichungen bei einer Standortschätzung mittels Mittellinienschnittpunkten

Der Ansatz, Standorte graphisch durch Schnittpunkte von Mittellinien zu approximieren, liefert bei der Anwendung auf die Fallstudie überwiegend gute Ergebnisse. Das beste Ergebnis ist eine Abweichung von nur 127 m zwischen einem Schnittpunkt und einem realen Standort. Das erste Quartil und damit ein Viertel der Abweichungen liegen bei unter 600 m. Bezogen auf die Größe der Postleitzahlengebiete stellt dies einen guten Wert dar. Der Durchschnitt der Abweichungen ist mit knapp 1,3 km beziffert, während der Median mit 1,1 km darunter liegt. Der niedrigere Wert des Medians im Vergleich zum Durchschnitt zeigt, dass bei diesem Verfahren einige wenige Abweichungen so große Werte annehmen, dass das arithmetische Mittel dadurch merklich verschlechtert wird. Die maximale Abweichung zwischen Approximation und Realität beträgt bei diesem Verfahren 4,6 km. Die daraus resultierende Spannweite von knapp 4,5 km verdeutlicht, dass generell eine große Streuung möglich ist. Allerdings zeigt das dritte Quartil, dass gut drei Viertel der Ergebnisse eine Abweichung von unter zwei Kilometern hat und damit verhältnismäßig gute Approximationen darstellen. Der Großteil der Abweichungen hat also eine deutlich geringe Streuung bei einer Spannweite von ca. 1,8 km. Abweichungen von mehr als drei Kilometern kommen also bei den Approximationen des Verfahrens verhältnismäßig selten vor. Betrachtet man die Abweichungen pro Postleitzahlengebiet, so ist keine Tendenz der Güte des

Ansatzes bezüglich der Anzahl der zu approximierenden Standorte oder der Größe eines Gebiets festzustellen. Allerdings bleibt festzuhalten, dass in größeren Postleitzahlengebieten theoretisch auch größere Abweichungen möglich sein können. Das Verfahren selber hat eine absolute Abweichung aller 110 Approximationen von den entsprechenden Standorten von 153,5 km. Der Ansatz der Standortschätzung durch Schnittpunkte von Mittellinien erzeugt also überwiegend gute Annäherungen bei einer leichten Anwendung des Verfahrens.

5.4 Standortannäherung durch Tourenabbildung

Den vierten zu testenden Ansatz stellt die Annäherung der Standorte durch graphische Abbildungen von Milk Runs dar. Die Annahmen und Erklärungen zur grundlegenden Vorgehensweise wurden in Abschnitt 4.2.3 besprochen. Für die Anwendung dieses Standortapproximationsverfahrens werden die Touren und Routen der Supply Chain sowie die Distanzen zwischen den Kundenstandorten benötigt. Es wird angenommen, dass diese Daten bekannt sind. Die genauen Werte der Distanzen sowie die Touren und Routen lassen sich aus dem Referenzdatensatz für den vorliegenden Fall ableiten. In dem Datensatz sind die Standorte samt Koordinaten und ihren Zuordnungen zu den Postleitzahlengebieten eingetragen. Diese Daten sollen die Grundlage zur Erstellung der Touren bilden. Die Planung eines Milk Runs kann durch ein gängiges Verfahren der Tourenplanung erfolgen. Da allerdings die genauen Lokalisierungen der Standorte nicht bekannt sind, kann nicht jedes Verfahren angewendet werden. Ein Verfahren, das mit der begrenzten Anzahl an Daten arbeiten kann, ist das in Abschnitt 2.2 vorgestellte Sweep-Verfahren. Deshalb sollen mit dieser Methode die Touren und Routen festgelegt werden. Zu jedem Postleitzahlengebiet mit mindestens drei Standorten wird das Verfahren eingesetzt. Daraus ergeben sich die Zusammenstellung der Standorte zu einer Tour und die Route, wie die Standorte abzufahren sind. Bei Postleitzahlengebieten mit weniger als drei Standorten werden benachbarte Gebiete verbunden und als ein großes Gebiet aufgefasst. In dem zusammengelegten Gebiet wird dann wieder das Sweep-Verfahren angewendet, um einen Milk Run zu planen. So konnten für die meisten Standorte Touren gebildet werden. In zwei Fällen konnten nur Touren aus Standorten von Postleitzahlengebieten gebildet werden, die nicht benachbart sind. In einem der beiden Fälle mussten drei Postleitzahlengebiete mit jeweils einem Standort zusammengefasst werden, die einen sehr großen Abstand zueinander haben. Die Gebiete befinden sich an dem Rand der gesamten betrachteten Fläche. Dies war dennoch die sinnvollste Möglichkeit die Standorte in Touren einzugliedern, ohne andere Touren über viele Postleitzahlengebiete zu verteilen. Insgesamt wurden die 110 Standorte in 31 Touren aufgeteilt, von denen 13 drei Standorte und 18 vier Standorte umfassen.

Die Distanzen der Standorte zueinander wurden mit der einfachen euklidischen Distanz berechnet, weil durch die Kreise der Standorte die Luftlinienentfernung zu dem nächsten Standort angezeigt wird. Die Formel dazu wurde in Abschnitt 2.4 vorgestellt. Dabei ist es nötig, die Distanzen in Kilometer umzurechnen. In dem Abschnitt wurde ebenfalls das generelle Vorgehen dazu beschrieben.

Durch die Anwendung des Sweep-Verfahrens wurde für den vorliegenden Fall festgelegt, dass es für jede Tour ein Depot gibt, welches in der Mitte des Postleitzahlengebiets lokalisiert sein soll.

Erfolgen die Touren über mehrere Gebiete ist der Mittelpunkt des zusammengelegten Gebiets der Standort des Depots. Zur Berechnung für die Distanzen der Depots zu den Kundenstandorten können die Mittelpunkte verwendet werden, die in Abschnitt 5.2 durch das arithmetische Mittel berechnet wurden. Bei zusammengelegten Gebieten können die Mittelpunkte analog berechnet werden. In diesem Fall wird auch das arithmetische Mittel verwendet.

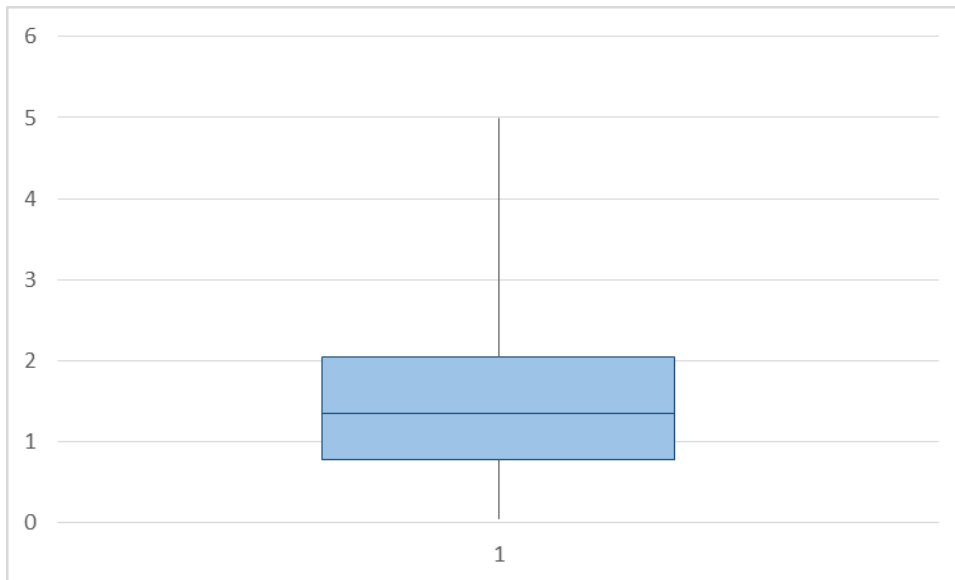


Abbildung 10: Boxplot der Abweichungen bei einer Approximation durch Tourenabbildungen

Die Approximation der Standorte erfolgt durch die graphische Abbildung der Milk Runs. Dazu wurden wie in Abschnitt 5.3 Abbildungen der einzelnen und der zusammengelegten Postleitzahlengebiete benötigt. Wie in dem Abschnitt beschrieben, wurden auch hier diese Abbildungen durch Abfragen bei Overpass Turbo generiert. In den daraus resultierenden Darstellungen konnten dann mittels des Zeichenprogramms Paint die Milk Runs graphisch veranschaulicht werden. Das genaue Vorgehen wurde in Abschnitt 4.2.3 erläutert. Zur Bestimmung der Koordinaten wurde anschließend wieder das Online Tool Tim-Online verwendet. Wie in Abschnitt 5.3 konnten mittels eines Abgleichs die Koordinaten bis auf fünf Stellen hinter der Dezimaltrennung genau identifiziert werden. Um die gleiche Anzahl an Nach Komma Stellen wie die Referenzdaten zu besitzen wurden die Koordinaten um zwei Nullen erweitert. Für die Bewertung des Verfahrens wird keine Zuordnung gemäß Abschnitt 4.1 benötigt, da die Standorte durch die Zusammenstellung und die festgelegte Reihenfolge innerhalb einer Tour bereits einander zugeteilt sind. Der i -te angenäherte Standort einer Tour entspricht also dem i -ten Standort des Referenzdatensatzes in derselben Tour. Zur Veranschaulichung wurden die Abweichungen der realen von den approximierten Standorten in einem Boxplot-Diagramm in Abbildung 10 festgehalten. Eine ausführliche Auflistung der Ergebnisse ist in Anhang 2 enthalten.

Betrachtet man die Ergebnisse des Ansatzes, so lassen sich mehr positive als negative Aspekte festhalten. So beträgt die minimale Abweichung eines geschätzten Standorts zur Realität bei diesem Verfahren nur 46 m. Auch das erste Quartil hat mit 772 m einen niedrigen Wert. Das bedeutet, ein Viertel der angenäherten Standorte stellt bezogen auf den Flächeninhalt der Postleitzahlengebiete gute Approximationen dar. Der Durchschnitt der Abweichungen liegt bei

ca. 1,5 km, während der Median mit ca. 1,3 km etwas geringer ausfällt. Der im Verhältnis zum arithmetischen Mittel niedrigere Median deutet an, dass der Durchschnitt durch einige wenige große Ausreißer entscheidend verschlechtert wird. Das Verfahren erzeugt also tendenziell mehr gute Ergebnisse mit niedrigen Abweichungen als schlechte Ergebnisse mit hohen Abweichungen. Das dritte Quartil bestätigt mit einem Wert von zwei Kilometern diesen Trend. Es haben also drei Viertel der Ergebnisse eine Abweichung von unter zwei Kilometern. Die maximale Abweichung einer Approximation liegt bei knapp fünf Kilometern. Eine Tendenz bezüglich des Einflusses der Anzahl der anzunähernden Standorte oder der Flächengröße eines Postleitzahlgebiets auf die Genauigkeit der Ergebnisse ist nicht festzustellen. Auffällig ist aber, dass falls der erste Standort einer Tour ungünstig in dem relevanten Gebiet platziert worden ist, d.h. eine große Abweichung zur Realität hat, alle weiteren Standorte ebenfalls größere Abweichungen zu den tatsächlichen Standorten besitzen. Die Güte dieses Verfahrens hängt also offensichtlich stark von der Platzierung des ersten Standortes einer Tour ab. Insgesamt haben die Abweichungen des Ansatzes, Standorte durch eine graphische Abbildung von Milk Runs zu approximieren, einen Wert von 170,5 km.

6 Vergleich der Approximationsverfahren

In dem vorangegangenen Kapitel 5 wurden vier Ansätze zur Standortapproximation auf einen konkreten Fall bezogen angewendet und somit hinsichtlich ihrer Realitätstreue getestet. Die verschiedenen Möglichkeiten und ihre Ergebnisse wurden anschließend einzeln ausgewertet. Die Auswertung erfolgte anhand der in Abschnitt 4.1 vorgestellten Bewertungsmethode und anhand verschiedener Lageparameter zur Einschätzung der Güte der Ergebnisse. Die Bestimmung der Parameter stellt nun die Grundlage zum Vergleich der Verfahren dar. Das Ziel ist dabei herauszufinden, welches der Verfahren die Realität am besten abbildet.

Führt man die Ergebnisse der einzelnen Auswertungen zusammen, so ist festzustellen, dass bei allen Verfahren ein Großteil der Abweichungen zwischen Approximationen und realen Standorten in derselben Größenordnung liegt. Bis auf den Ansatz der Mittelpunktberechnung durch den Median bei einer umfangreichen Datengrundlage erzielen die Verfahren überwiegend Annäherungen, die Abweichungen zwischen 800 Metern und zwei Kilometern besitzen. Das bedeutet gut die Hälfte der Differenzen zur Realität der Verfahren befindet sich in diesem Zahlenbereich. Die Mittelwertberechnung durch den Median schneidet hier deutlich schlechter ab, hier entstehen vorwiegend Abweichungen zwischen 1,3 und 3 km. Dazu kommt, dass ein Viertel der Ergebnisse sogar eine Differenz von mehr als drei Kilometern zu den realen Standorten hat. Zusammengerechnet ergibt das eine absolute Abweichung von über 227 km bei 110 Standorten, was einen Durchschnitt von etwas mehr als zwei Kilometern ergibt. In diesen Kategorien sind das die schlechtesten Resultate. Dieser Ansatz eignet sich daher weniger gut zur Standortapproximation. Zwar hat das Verfahren auch einige gute Ergebnisse erzeugt, allerdings ist ein beträchtlicher Teil der Abweichungen zu groß um im Hinblick auf das vorliegende Problem eine gute Methode darzustellen. Die Bestimmung von Mittelpunkten mit Hilfe des arithmetischen Mittels ist demgegenüber sowohl bei einem umfangreicheren als auch bei einem auf die Extrempunkte beschränkten Datensatz als Berechnungsgrundlage bei allen Lageparametern besser. Die anderen Verfahren haben ebenfalls in fast allen Bereichen bessere Werte.

Wird die Betrachtung auf die komplette Spannweite der Abweichungen ausgeweitet, so fällt der Ansatz durch gleichverteilte Zufallszahlen die Standorte anzunähern besonders schlecht auf. Die Spannweite des Ansatzes ist über 1,5 km größer als die zweitgrößte Spannweite aller Verfahren. Der Ansatz hat also die eindeutig größte Streuung, welche von niedrigen Abweichungen mit knapp unter 200 Metern bis zu den größten gemessenen Werten aller Verfahren mit fast 6,8 km reicht. Da aber gut die Hälfte aller Abweichungen der Approximationen unter 1,2 km Länge haben, stellt der überwiegende Teil der Ergebnisse dennoch akzeptable Annäherungen dar. Das Verfahren hat aber mit einer absoluten Abweichung von 185,5 km das zweitschlechteste Gesamtergebnis. Die kürzeste Spannweite und damit die niedrigste Streuung hat demgegenüber der Ansatz der Mittelpunktberechnung durch das arithmetische Mittel bei dem umfangreichen Datensatz als Berechnungsgrundlage. Hier streuen die Abweichungen nur zwischen 97 m und knapp 3,9 km. Allerdings besitzen nur knapp ein Viertel der Abweichungen innerhalb dieser Menge weniger als einen Kilometer Länge. Die Hälfte liegt sogar über 1,6 km. Anders sieht es

aus, wenn die Bestimmung der Mittelpunkte anhand einer auf die Extrempunkte beschränkten Berechnungsgrundlage erfolgt. Wie in Abschnitt 5.2 erklärt macht es hierbei keinen Unterschied, ob das arithmetische Mittel oder der Median verwendet wird. Bei einer derartigen Standortapproximation ist die Spannweite der Abweichungen zwar insgesamt über einen halben Kilometer größer, aber im Durchschnitt liegen die Schätzungen gut 200 Meter näher an den entsprechenden realen Standorten. Andere Lagerparameter fallen im direkten Vergleich der beiden Möglichkeiten zur Mittelpunktbestimmung bei der begrenzten Datengrundlage ebenfalls um die 200 Meter geringer aus als bei der umfangreichen Berechnungsgrundlage. Dadurch ist auch die absolute Abweichung bei dieser Variante mit 156,3 km zu 182,7 km merklich besser als bei der Mittelpunktberechnung mit einer umfangreichen Berechnungsgrundlage. Der Ansatz, Mittelpunkte als Approximation zu verwenden, ist in der Variante wenige Daten zur Berechnung zu nutzen am besten. Bei der Betrachtung aller Verfahren stellt diese Möglichkeit das in absoluten Zahlen zweitbeste Verfahren dar.

Etwas schlechter als bei der Mittelpunktbestimmung bei begrenzter Berechnungsgrundlage fallen die Ergebnisse bei dem Ansatz die Standorte mittels einer graphischen Abbildung von Milk Runs zu schätzen aus. Die Hälfte der Abweichungen ist hier niedriger als bei den Mittelpunkten, allerdings ist die andere Hälfte dafür eindeutig schlechter. Das führt dazu, dass im direkten Vergleich der beiden Möglichkeiten die Spannweite und damit die Streuung fast einen halben Kilometer größer ist. Die schlechtere Hälfte der Approximationen beeinflusst das Gesamtergebnis derartig, dass insgesamt die absolute Abweichung aller Annäherungen bei dem graphischen Verfahren mit 170,5 km knapp 14 km höher beziffert ist als bei den Mittelpunkten mit begrenzter Berechnungsgrundlage.

Die besten Ergebnisse bei der Anwendung auf die Fallstudie hat das Approximationsverfahren, das Schnittpunkte von Mittellinien als Annäherungen für die gesuchten Standorte verwendet. Hier sind die Spannweite und die Streuung ähnlich groß, wie bei der Mittelpunktbestimmung, aber ein Großteil der Ergebnisse hat insgesamt geringere Abweichungen als die Ergebnisse der anderen Verfahren. Der Median hat hier einem Wert von ca. einem Kilometer, was mit Abstand der niedrigste Wert dieses Lageparameters von allen Verfahren ist. Ein Viertel der Abweichungen hat sogar nur eine maximale Länge von 600 Metern. Diese niedrigen Abweichungen sorgen dafür, dass trotz einiger hohen Werte die Summe aller Abweichungen des Verfahrens mit 153,5 km das beste Resultat der Untersuchungen darstellt. Bei diesem Verfahren sind, wie bei allen anderen auch, keine Tendenzen zu erkennen, dass Faktoren wie die Anzahl der zu approximierenden Standorte oder die Größe eines Postleitzahlgebietes, die Güte des Ergebnisses beeinflussen. Einen Vorzug aufgrund bestimmter Rahmenbedingungen für ein Verfahren ist also hier nicht zu rechtfertigen. Ein Ausschlusskriterium ist nur bei der graphischen Abbildung der Milk Runs gegeben, indem zur Anwendung des Verfahrens die Distanzen, Touren und Routen bekannt sein müssen. Sind die Voraussetzungen nicht erfüllt so kann dieses Verfahren nicht für eine Standortapproximation verwendet werden. Abgesehen von dieser Einschränkung wird der Einsatz eines der getesteten Approximationsverfahren aber nur durch den Aufwand und die Genauigkeit der Ergebnisse bestimmt. Der Aufwand unterscheidet sich bei den Verfahren nur geringfügig. Jedes der Verfahren lässt sich einfach und ohne großen Aufwand praktisch umsetzen.

Zusammenfassend lässt sich sagen, dass das Verfahren, Schnittpunkte von Mittellinien als Approximation zu verwenden, aufgrund der vorliegenden Ergebnisse und des niedrigen Aufwands zur Nutzung eine gute Möglichkeit Standorte bei unzureichenden Daten zu schätzen darstellt. Dieses Verfahren bildet die Realität also am besten ab. Eine weitere gute Variante ist die Annäherung durch Mittelpunkte, welche bei einer begrenzten Berechnungsgrundlage bestimmt werden. Die Ergebnisse der Methode sind nur geringfügig schlechter als die Approximation durch die Schnittpunkte, bei einem ähnlich niedrigen Anwendungsaufwand. Die anderen Verfahren eignen sich auch generell zur Standortapproximation, sind aber in ihrer Genauigkeit weniger gut. Eine Ausnahme bildet hier die Annäherung durch Mittelpunkte, die mit dem Median und einer umfangreichen Berechnungsgrundlage bestimmt worden sind. Das Mittelpunktverfahren ist klar schlechter als die anderen Verfahren und für die Annäherung auch aufgrund der Abweichungen in absoluten Zahlen ungeeignet.

7 Zusammenfassung und Ausblick

Durch Faktoren wie der zunehmenden Verstadterung der Bevolkung wird die Bedeutung von Supply Chains zur Versorgung verschiedenster Kunden in urbanen Raumen voraussichtlich weiter ansteigen. Die Supply Chain ist ein Liefernetzwerk, welches sich durch eine starke Zusammenarbeit in einem Wertschopfungsprozess, welcher die Versorgung der Kunden zum Ziel hat, auszeichnet. Fur eine effektive Versorgung spielen die Transportstrukturen der Netzwerke eine entscheidende Rolle. Wie in den Abschnitten 2.1 und 2.2 aufgezeigt konnen sich diese grundlegenden Strukturen bei den einzelnen Supply Chains unterscheiden. Dazu kommen unterschiedliche Konzepte in der Last Mile Distribution. Aufgrund der verschiedenen Strukturen und Konzepte ist es notig, die Qualitat der Supply Chains zu untersuchen und zu bewerten. Zur Analyse und Bewertung der Netzwerke und ihrer Strukturen eignen sich insbesondere Simulationen. Simulationen sind Vorgange, bei denen an modellartigen Abbildungen eines realen Systems Experimente durchgefuhrt werden, um Ruckschlusse auf die Realitat ziehen zu konnen. Durch die gewonnenen Erkenntnisse konnen dann die realen Systeme verbessert werden. Eine Simulation benotigt eine umfassende Datengrundlage, durch die das Modell erstellt werden kann. Wenn wie im vorliegenden Fall zu einem bestimmten Bereich wie den Standorten der Kunden einer Supply Chain keine Daten gegeben und Erhebungen nicht moglich sind, so ist es unerlasslich, einen sinnvollen Umgang mit den fehlenden Werten zu finden. Es wurde gezeigt, dass es verschiedene Verfahren gibt, Approximationen zu generieren. Vor allen aus der Statistik stammen einfache Moglichkeiten fehlende Werte zu schatzen. Die verbreiteten Methoden sind aber nicht unbedingt zur Annaherung von Standorten geeignet. Deshalb wurden in dieser Arbeit drei Verfahren zur Standortapproximation neu entwickelt, um bei einer unzureichenden Datenlage Standorte schatzen zu konnen. Zusammen mit einer bereits entwickelten Methode wurden die Verfahren auf eine aus dem allgemeinen Problem hergeleitete Fallstudie angewendet und getestet, um zu uberprufen, welches der Verfahren die Realitat am besten abbildet und sich damit fur einen Einsatz in einer ahnlichen Situation am ehesten empfiehlt.

Im vorherigen Kapitel 6 wurden die Ergebnisse der Verfahrensanwendungen verglichen. Dabei wurde festgestellt, dass zwei Methoden eindeutig bessere Resultate erzielt haben, als die anderen Verfahren. Diese Verfahren sind die Approximation durch Schnittpunkte von Mittellinien und die Annaherung durch Mittelpunkte, die mit Daten, die auf Extrempunkte eines Postleitzahlengebiets beschrankt sind, berechnet wurden. Beide Methoden bieten uberwiegend gute Annaherungen bei verhaltnismaig geringem Aufwand. Fur die konkrete Situation, bei der fur die Simulation einer Supply Chain Standorte zu approximieren sind, eignen sich daher besonders diese beiden Verfahren. Die anderen Verfahren konnen generell auch verwendet werden, sind aber weniger zu empfehlen, da die Annaherungen in der Breite weniger genau sind. Allgemein lassen sich alle betrachteten Verfahren, mit Ausnahme der graphischen Abbildung von Touren, auf ahnliche Problemstellungen, bei denen aufgrund von fehlenden Daten Standorte oder auch einfach Punkte in einer Flache zu schatzen sind, ubertragen. Da aber auch eine gewisse Streuung bei den Ergebnissen vorhanden ist, sind die untersuchten Verfahren nicht optimal. Deshalb sollte, falls

dies möglich ist, immer eine umfassende Datenerhebung zu fehlenden Informationen getätigt werden. Für den Fall, dass keine Daten zu Lokalisierungen erhoben werden können, bieten die genannten Verfahren eine gute Lösungsmöglichkeit.

Die Simulation wird aufgrund ihrer vielfältigen Möglichkeiten in Zukunft voraussichtlich weiter an Bedeutung gewinnen. Die gesellschaftliche und wirtschaftliche Entwicklung führt dazu, dass viele bestehende Strukturen unterschiedlicher Einrichtungen analysiert, bewertet und darauf aufbauend verbessert werden müssen, um einen Effizienzverlust zu vermeiden. Die Simulation stellt zur Bewertung ein umfassendes Hilfsmittel dar. Die zunehmende Digitalisierung beschleunigt dabei die Entwicklung, computergestützte Analysemethoden in den Planungs- und Arbeitsabläufen zu integrieren. Gerade im Bereich von Liefernetzwerken wie Supply Chains kann daher eine höhere Verknüpfung der Wissenschaftsbereiche erwartet werden.

Literaturverzeichnis

- Arndt, Holger: Supply Chain Management: Optimierung logistischer Prozesse. Gabler Verlag, Wiesbaden, 2008
- Bacher, Andreas: Instrumente des Supply Chain Controlling: Theoretische Herleitung und Überprüfung der Anwendbarkeit in der Unternehmenspraxis. Deutscher Universitäts-Verlag, 2004
- Balci, Osman: Verification, Validation, and Certification of Modeling and Simulation Applications. In: Chick, S.; Sanchez, P.J.; Ferrin, D.; Morrice, D.J. (Hrsg.): Proceedings of the 2003 Winter Simulation Conference
- Bamberg, Günter; Baur, Franz; Krapp, Michael: Statistik. Oldenbourg Verlag, München, 2012
- Bankhofer, Udo: Unvollständige Daten- und Distanzmatrizen in der Multivariaten Datenanalyse. Josef Eul Verlag, Bergisch Gladbach Köln, 1995
- Baudach, Jens; Voll, Robert; Eufinger, Lars; Meier, Fabian; Sender, Julia; Goedicke, Ina; Thaller, Carina: Modellentwicklung. In: Clausen, Uwe; Geiger, Christiane (Hrsg.): Verkehrs- und Transportlogistik. Springer Verlag, Berlin Heidelberg, 2013
- Becker, Torsten: Prozesse in Produktion und Supply Chain optimieren. Springer Verlag, 2018
- Bol, Georg: Induktive Statistik. Oldenbourg Verlag, München Wien, 2003
- Buchholz, Jonas; Clausen, Uwe; Vastag, Alex: Handbuch der Verkehrslogistik. Springer Verlag, Berlin Heidelberg, 1998
- Bungartz, Hans-Joachim; Zimmer, Stefan; Buchholz, Martin; Pflüger: Modellbildung und Simulation: Eine anwendungsorientierte Einführung. Springer Verlag, Berlin Heidelberg, 2013
- Carpenter, James R.; Kenward, Michael G.: Multiple Imputation and its Application. John Wiley & Sons, Ltd, 2013
- Caspary, Wilhelm: Fehlertolerante Auswertung von Messdaten: Daten- und Modellanalyse, robuste Schätzung. Oldenbourg Verlag, München, 2013
- Corsten, Hans; Gössinger, Ralf: Einführung in das Supply Chain Management. Oldenbourg Verlag, München, 2008
- Dabidian, Peiman; Langkau, Sven: Straßengüterverkehr. In: Clausen, Uwe; Geiger, Christiane (Hrsg.): Verkehrs- und Transportlogistik. Springer Verlag, Berlin Heidelberg, 2013
- De Lange, Norbert, Geoinformatik. Springer Verlag, 2013
- DIN IEC 60050-351: Internationales Elektrotechnisches Wörterbuch: Teil 351: Leittechnik. Beuth, Berlin, 2014
- Domschke, Wolfgang; Drexl, Andreas: Logistik: Standorte. Oldenbourg Verlag, München 1996

- Domschke, Wolfgang; Scholl, Armin: Logistik: Rundreisen und Touren. Oldenbourg Verlag, München, 2010
- Elmegreen, Bruce G.; Sanchez, Susan M.; Szalay, Alexander S.: The Future of Computerized Decision Making. In: Tolk, A.; Diallo, I. O.; Yilmaz, I.; Buckley, S.; Miller, J. A.: Proceedings of the 2014 Winter Simulation Conference. 2014
- Eufinger, Lars; Tesch, Christian: Tourenplanung. In: Clausen, Uwe; Geiger, Christiane (Hrsg.): Verkehrs- und Transportlogistik. Springer Verlag, Berlin Heidelberg, 2013
- Fahrmeir, Ludwig; Heumann, Christian; Künstler, Rita; Pigeot, Iris; Tutz, Gerhard: Statistik: Der Weg zur Datenanalyse. Springer Verlag, Berlin Heidelberg, 2016
- Feldkamp, Niclas; Bergmann, Sören; Strassburger, Steffen: Data Farming im Kontext von Produktion und Logistik. In: Wenzel, Sigrid; Peter, Tim (Hrsg.): Simulation in Produktion und Logistik 2017. Kassel university press, Kassel, 2017
- Feldkamp, Niclas; Bergmann, Sören; Strassburger, Steffen: Knowledge Discovery in Manufacturing Simulations. 2016
- Gagatsch, Andreas: Grundkurs Geschäftsprozess – Management: Methoden und Werkzeuge für die IT-Praxis: Eine Einführung für Studenten und Praktiker. Springer Verlag, 2012
- Goedicke, Ina: Simulation logistischer Prozesse. In: Clausen, Uwe; Geiger, Christiane (Hrsg.): Verkehrs- und Transportlogistik. Springer Verlag, Berlin Heidelberg, 2013
- Graham, John W.: Missing Data Analysis: Making It Work in the Real World. In: Annual Review of Psychology, 2009
- Gudehus, Tim: Logistik 1: Grundlagen, Verfahren und Strategien. Springer Verlag Berlin Heidelberg, 2007
- Gudehus, Tim: Logistik 2: Netzwerke, Systeme und Lieferketten. Springer Verlag, Berlin Heidelberg, 2007
- Hedstück, Ulrich: Simulation diskreter Prozesse. Springer Verlag, 2013
- Heidtmann, Volker: Organisation von Supply Chain Management: Theoretische Konzeption und empirische Untersuchung in der deutschen Automobilindustrie. Gabler Verlag, Wiesbaden, 2008
- Horne, Gary; Meyer, Theodore: Data Farming Process and Initial Network Analysis Capabilities. MDPI, Basel, 2016
- Hummeltenberg, Wilhelm: Optimierungsmethoden zur betrieblichen Standortwahl. Physica-Verlag, Würzburg Wien, 1981
- Kallfass, Daniel; Schlaak, Tobias: NATO MSG-088 Case Study Results to Demonstrate the Benefit of Using Data Farming for Military Decision Support. In: Proceedings of the 2012 Winter Simulation Conference. 2012

- Kamps, Udo: Definition Schätzfunktion. In: Gabler Wirtschaftslexikon, <https://wirtschaftslexikon.gabler.de/definition/schaetzfunktion-44631> (zuletzt geprüft am 09.07.2018)
- Kompf.de: <https://www.kompf.de/gps/distcalc.html> (zuletzt geprüft am 09.07.2018)
- Kuhn, Axel; Reinhardt, Adolf; Wiendahl, Hans-Peter: Handbuch Simulationsanwendungen in Produktion und Logistik. Vieweg Verlag, 2013
- Lackes, Richard: Definition Data Mining. In: Gabler Wirtschaftslexikon, <https://wirtschaftslexikon.gabler.de/definition/data-mining-28709> (zuletzt geprüft am 09.07.2018)
- Liebl, Franz: Simulation: Problemorientierte Einführung. Oldenbourg Verlag, München, 1995
- Mentzer, John T.; DeWitt, William; Keebler, James S.; Min, Soonhong; Nix, Nancy W.; Smith, Carlo D.; Zacharia, Zach G.: Defining Supply Chain Management. In: Journal of Business Logistics, Vol. 22, No. 2, 2001
- Metzler, Ute: Anwendungsbereiche der Transportplanung. In: Clausen, Uwe; Geiger, Christiane (Hrsg.): Verkehrs- und Transportlogistik. Springer Verlag, Berlin Heidelberg, 2013
- Mohlenberghs, Geert; Kenward, Michael G.: Missing Data in Clinical Studies. John Wiley & Sons, Ltd, 2007
- Müller, Martin: Informationstransfer im Supply Chain Management: Analyse aus Sicht der Neuen Institutionenökonomie. Deutscher Universitäts-Verlag, 2005
- Poirier, Charles C.; Reiter, Stephen E.: Die optimale Wertschöpfungskette: Wie Lieferanten, Produzenten und Handel bestens zusammenarbeiten. Campus Verlag, 1997
- Rabe, Markus; Gutenschwager, Kai; Spieckermann, Sven; Wenzel, Sigrid: Simulation in Produktion und Logistik: Grundlagen und Anwendungen. Springer Verlag, 2017
- Rabe, Markus; Spieckermann, Sven; Wenzel, Sigrid: Verifikation und Validierung für die Simulation in Produktion und Logistik: Vorgehensmodelle und Techniken. Springer Verlag, Berlin Heidelberg, 2008
- Raps, Johannes: Bewertung von Distanzmetriken für ein Supply-Chain-Simulationsmodell. 2017
- Richtlinie des Vereins Deutscher Ingenieure 3633 Blatt 1: Simulation von Logistik-, Materialfluss- und Produktionssystemen (Grundlagen). Beuth, Berlin, 2014
- Richtlinie des Vereins Deutscher Ingenieure 4465 Blatt 1: Modellierung und Simulation (Modellbildungsprozess). Beuth, Berlin, 2016
- Schulte, Christof: Logistik: Wege zur Optimierung einer Supply Chain. Franz Vahlen Verlag, München 2013
- Sibbertsen, Philipp; Lehne, Hartmut: Statistik: Einführung für Wirtschafts- und Sozialwissenschaftler. Springer Verlag, Berlin Heidelberg, 2015

- Thaler, Klaus: Supply Chain Management: Prozessoptimierung in der logistischen Kette. Bildungsverlag EINS, Troisdorf, 2007
- Thonemann, Ulrich: Operations Management: Konzepte, Methoden und Anwendungen. Pearson Verlag, 2015
- Ueda, Naonori; Nakano, Ryohei: Deterministic annealing EM algorithm. In: Neural Networks 11, 1998
- Van Buuren, Stef: Flexible Imputation of Missing Data. Taylor & Francis Group, 2012
- Vereinigte Nationen: World Population Prospects: The 2008 Revision, Highlights, New York, 2009
- Vereinigte Nationen: World Population Prospects: The 2015 Revision, Key Findings and Advance Tables, New York, 2015
- Waldmann, Karl-Heinz; Helm, Werner E.: Simulation stochastischer Systeme: Eine anwendungsorientierte Einführung. Springer Verlag, Berlin Heidelberg, 2016
- Werner, Hartmut: Supply Chain Management: Grundlagen, Strategien, Instrumente und Controlling. Springer Verlag, Wiesbaden, 2013
- Wiki.openstreetmap.org:
https://wiki.openstreetmap.org/wiki/DE:Genauigkeit_von_Koordinaten (01.07.2018)

Abbildungsverzeichnis

Abbildung 1: Möglicher Aufbau einer Supply Chain, eigene Darstellung nach Gadatsch (2012) .	6
Abbildung 2: Direktverkehrsnetz und Hub-and Spoke-Netz nach Dabidian und Langkau (2013)	8
Abbildung 3: Aufteilung eines Systems nach Hedtstück (2013)	11
Abbildung 4: Vorgehensmodell bei Simulationsstudien nach Rabe et al. (2008).....	14
Abbildung 5: Konstruktion von Mittellinien zur Standortschätzung.....	34
Abbildung 6: Vorgehen bei der graphischen Tourenabbildung	37
Abbildung 7: Boxplot der Abweichungen bei einer Gleichverteilung der Standorte.....	41
Abbildung 8: Boxplots der Abweichungen bei der Standortapproximation durch Mittelpunkte, 1 arithmetisches Mittel, 2 Median, 3 arithmetisches Mittel und Median bei minimaler Datengrundlage.....	43
Abbildung 9: Boxplot der Abweichungen bei einer Standortschätzung mittels Mittellinienschnittpunkten	46
Abbildung 10: Boxplot der Abweichungen bei einer Approximation durch Tourenabbildungen	48
Abbildung 11: Datenabfrage in Overpass Turbo.....	XVIII
Abbildung 12: Betrachtetes Gebiet des Referenzdatensatzes mit abgefragten Standorten.....	XIX
Abbildung 13: Beispielhafte Darstellung der Standortannäherung durch Mittellinien.....	XXIII
Abbildung 14: Beispielhafte Darstellung der Standortannäherung durch eine graphische Abbildung von Touren.....	XXIV

Tabellenverzeichnis

Tabelle 1: Ausgegebene Daten zu den abgefragten Standorten	XIX
Tabelle 2: Koordinaten der Approximationen Teil 1	XXIV
Tabelle 3: Koordinaten der Approximationen Teil 2	XXVII

Formelverzeichnis

Formel (1): rechtwinklige Distanz	15
Formel (2): einfache euklidische Distanz	15
Formel (3): gewichtete euklidische Distanz.....	16
Formel (4): arithmetisches Mittel.....	25
Formel (5): Median	25
Formel (6): absolute Abweichung.....	31

Abkürzungsverzeichnis

CSV	Comma-Separated Values
DIN	Deutsches Institut für Normung
IT	Informationstechnik
MAR	Missing At Random
MCAR	Missing Completely At Random
NMAR	Not Missing At Random
SC	Supply Chain
SCM	Supply Chain Management
UN	United Nations (Vereinigte Nationen)
VDI	Verein Deutscher Ingenieure

Anhang 1: Erzeugung einer Datengrundlage

Abfrage in Overpass Turbo

Für die Generierung des Referenzdatensatzes wurde wie in Abschnitt 4.1 beschrieben das Online Tool Overpass Turbo (www.overpass-turbo.eu) verwendet. Eine mögliche Datenabfrage nach Postfilialen in einem manuell begrenzten Bereich wird in Abbildung 11 gezeigt.

```
1 //den gewünschten Bereich manuell begrenzen
2
3 [out:csv(;;id,;;lat,;;lon)] //Ausgabe der Daten in CSV-Format definieren
4
5 node
6   [amenity=post_office] //Abfrage nach Einrichtungstyp Postfilialen
7   ({{bbox}}); //Abfrage innerhalb des manuell begrenzten Bereichs
8 out; //Ausgabe der Informationen
```

Abbildung 11: Datenabfrage in Overpass Turbo

Abgefragtes Gebiet

Die abgefragte Region zur Referenzdatengenerierung wurde wie beschrieben manuell festgelegt. In der Abbildung 12 ist das eingegrenzte Gebiet um den Großraum Dortmund mitsamt den abgefragten Postfilialen dargestellt.

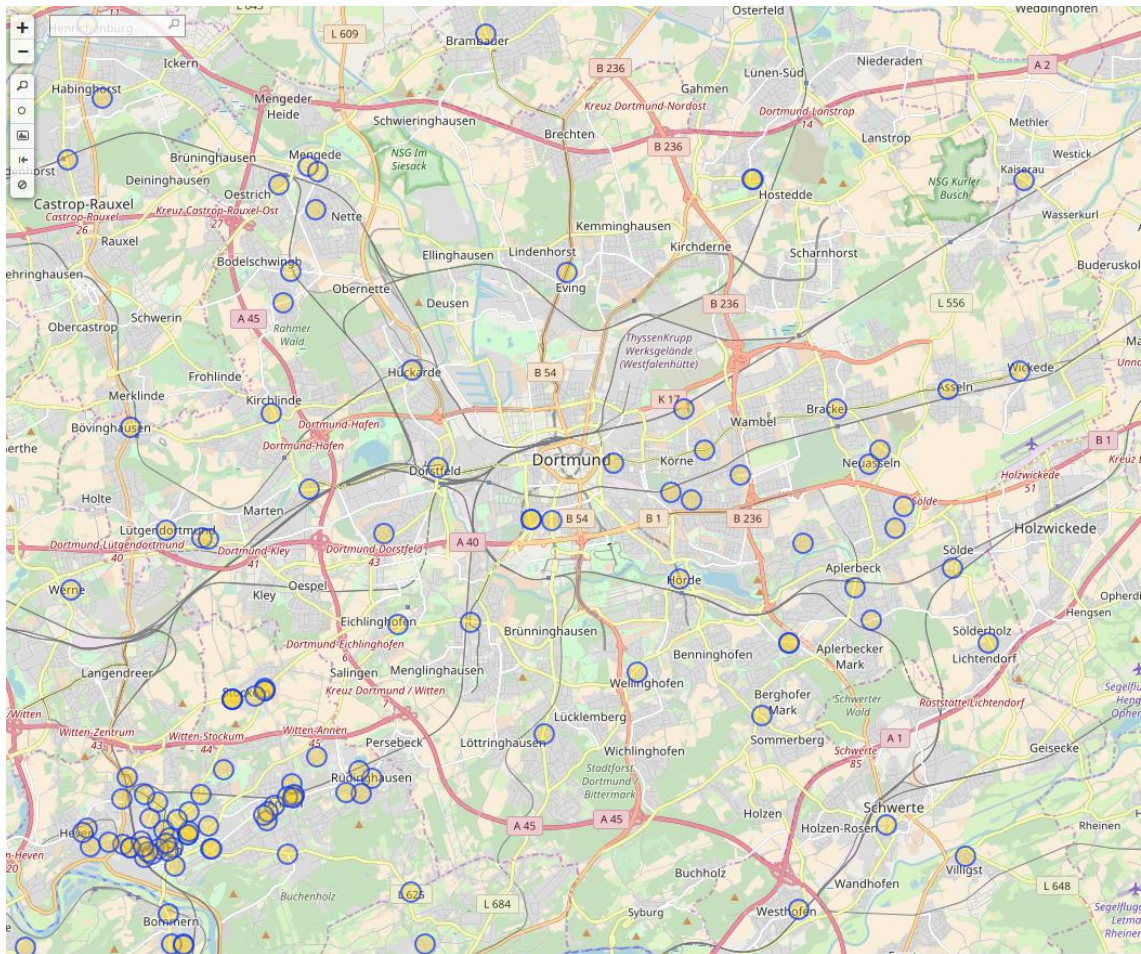


Abbildung 12: Betrachtetes Gebiet des Referenzdatensatzes mit abgefragten Standorten

Daten der ausgegebenen Postfilialen

Bei der Abfrage der Postfilialen wurden nur die direkt relevanten Informationen zu den Standorten ausgegeben. Die Ausgabe erfolgte im CSV-Dateiformat. Die nachfolgende Tabelle enthält die ausgegebenen Daten zu den 110 Postfilialen, welche den in Abschnitt 4.1 beschriebenen Referenzdatensatz darstellen.

Tabelle 1: Ausgegebene Daten zu den abgefragten Standorten

Nummer	Breitengrad	Längengrad	PLZ
1	515.134.156	74.784.997	44135
2	515.023.583	74.525.007	44137
3	515.021.018	74.590.693	44139
4	515.062.195	75.033.263	44141
5	515.076.276	74.969.050	44141
6	515.110.257	75.188.951	44141
7	515.160.793	75.076.074	44143
8	515.240.756	75.010.779	44143
9	515.126.640	74.231.335	44149

10	514.820.219	74.333.475	44225
11	514.600.434	74.567.919	44229
12	514.904.776	74.996.734	44263
13	514.721.853	74.862.409	44265
14	514.636.104	75.255.858	44267
15	514.779.652	75.343.765	44269
16	514.976.200	75.385.462	44269
17	514.823.549	75.602.756	44287
18	514.888.108	75.552.173	44287
19	515.007.135	75.677.471	44287
20	515.050.014	75.705.002	44287
21	514.778.602	75.974.933	44289
22	514.928.001	75.861.268	44289
23	515.160.078	75.632.020	44309
24	515.241.345	75.494.010	44309
25	515.279.006	75.846.438	44319
26	515.315.111	76.074.058	44319
27	515.693.524	75.227.092	44329
28	515.693.588	75.229.179	44329
29	515.510.938	74.637.637	44339
30	515.451.516	73.738.678	44357
31	515.513.351	73.762.180	44357
32	515.683.777	73.726.071	44357
33	515.634.026	73.842.358	44359
34	515.708.731	73.849.907	44359
35	515.720.272	73.819.484	44359
36	515.317.647	74.146.658	44369
37	515.083.289	73.819.607	44379
38	515.232.168	73.700.163	44379
39	514.984.658	73.503.465	44388
40	514.987.566	73.479.810	44388
41	515.002.644	73.366.992	44388
42	515.204.884	73.256.284	44388
43	515.979.624	74.379.443	44536
44	515.732.728	73.054.438	44579
45	515.852.856	73.164.857	44579
46	515.998.299	73.118.976	44581
47	514.884.540	73.065.654	44894
48	514.252.430	75.372.025	58239
49	514.357.176	75.901.432	58239
50	514.420.896	75.651.005	58239
51	514.184.981	74.189.765	58313
52	514.183.061	73.422.557	58452
53	514.185.525	73.421.618	58452
54	514.245.118	73.374.508	58452

55	514.339.376	73.394.299	58452
56	514.355.289	73.300.140	58452
57	514.363.122	73.323.441	58452
58	514.366.710	73.310.948	58452
59	514.367.837	73.375.342	58452
60	514.373.020	73.345.115	58452
61	514.375.152	73.390.254	58452
62	514.376.965	73.295.608	58452
63	514.384.709	73.369.513	58452
64	514.387.259	73.290.011	58452
65	514.395.185	73.378.737	58452
66	514.401.568	73.437.068	58452
67	514.401.786	73.437.064	58452
68	514.407.305	73.439.639	58452
69	514.409.074	73.361.940	58452
70	514.433.285	73.315.581	58452
71	514.361.518	73.752.265	58453
72	514.372.400	73.510.234	58453
73	514.372.832	73.511.708	58453
74	514.438.848	73.677.941	58453
75	514.446.616	73.689.109	58453
76	514.458.332	73.711.164	58453
77	514.473.077	73.770.745	58453
78	514.476.470	73.751.416	58453
79	514.477.158	73.772.852	58453
80	514.479.407	73.771.919	58453
81	514.429.746	73.687.700	58453
82	514.287.515	74.143.483	58454
83	514.483.033	73.985.602	58454
84	514.485.048	73.936.912	58454
85	514.512.317	74.018.671	58454
86	514.526.894	73.980.276	58454
87	514.528.763	73.548.403	58454
88	514.555.845	73.846.011	58454
89	514.667.534	73.578.357	58454
90	514.667.668	73.577.771	58454
91	514.668.026	73.578.062	58454
92	514.675.012	73.648.856	58454
93	514.685.013	73.680.140	58454
94	514.685.406	73.680.764	58454
95	514.689.581	73.680.560	58454
96	514.374.853	73.254.960	58455
97	514.375.313	73.254.970	58455
98	514.376.904	73.127.205	58455
99	514.382.325	73.231.623	58455

100	514.386.966	73.184.901	58455
101	514.414.246	73.118.314	58455
102	514.417.221	73.501.665	58455
103	514.445.801	73.440.251	58455
104	514.465.689	73.342.245	58455
105	514.471.990	73.228.586	58455
106	514.480.823	73.478.510	58455
107	514.480.968	73.300.578	58455
108	514.514.933	73.245.648	58455
109	514.178.839	72.923.338	58456
110	515.691.013	76.086.540	59174

Anhang 2: Ergebnisse der Approximationsverfahren

Darstellung der Mittellinienapproximation

Wie im Text erläutert wurden die Postleitzahlgebiete für die Annäherung der Standorte durch Mittellinien durch ein umschließendes Rechteck eingegrenzt, um anschließend Mittellinien konstruieren zu können. Die Abbildung 13 zeigt beispielhaft die Approximation von drei Standorten innerhalb eines Postleitzahlgebiets nach dem beschriebenen Prinzip.

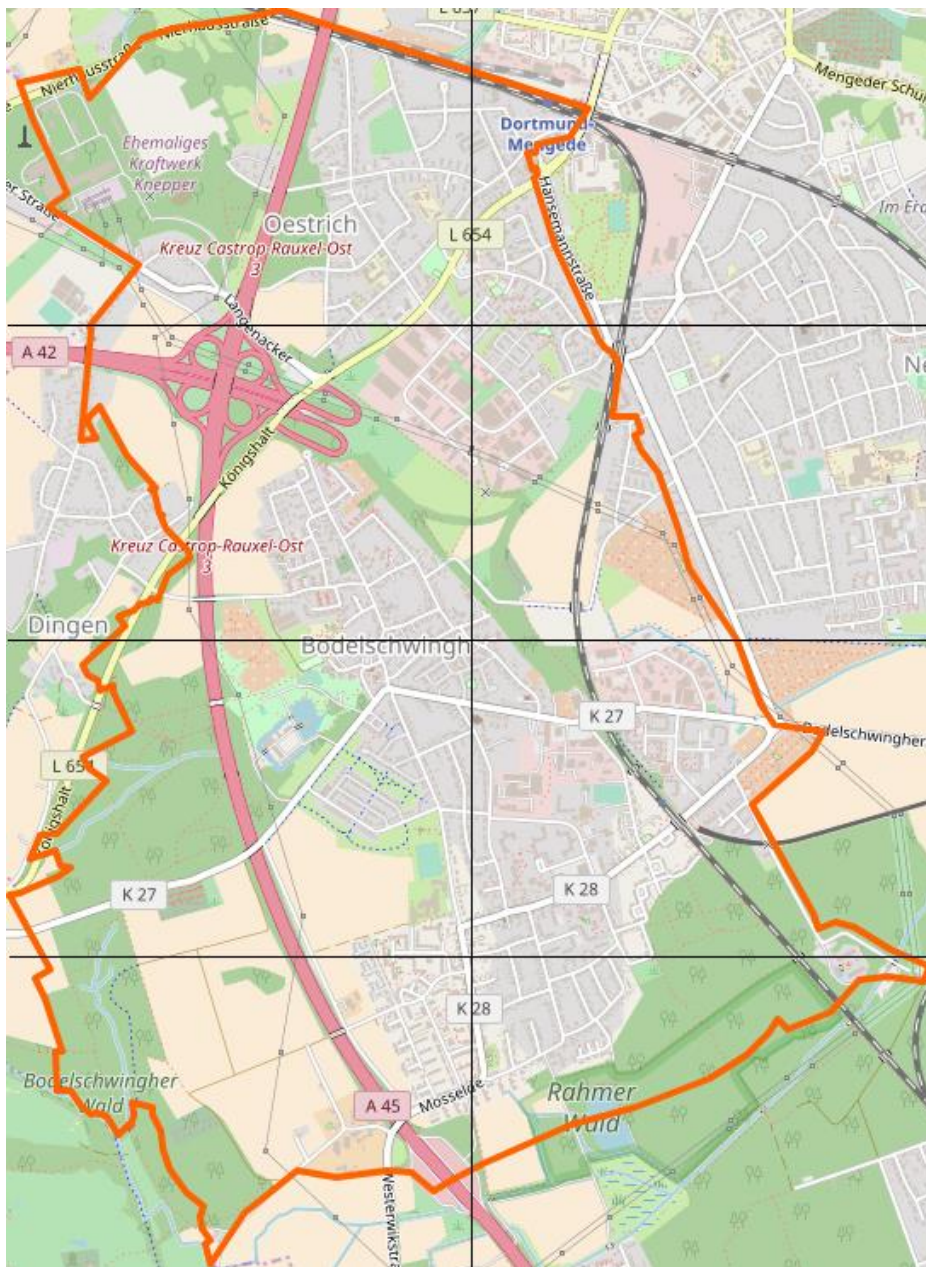


Abbildung 13: Beispielhafte Darstellung der Standortannäherung durch Mittellinien

Graphische Abbildung von Milk Runs

Für die Approximation von Standorten mittels der Abbildung von Milk Runs wurden die Touren graphisch in der Fläche angenähert. Ein Beispiel, wie die graphische Abbildung von Touren und damit die Standortannäherung aussehen können, zeigt die Abbildung 14.

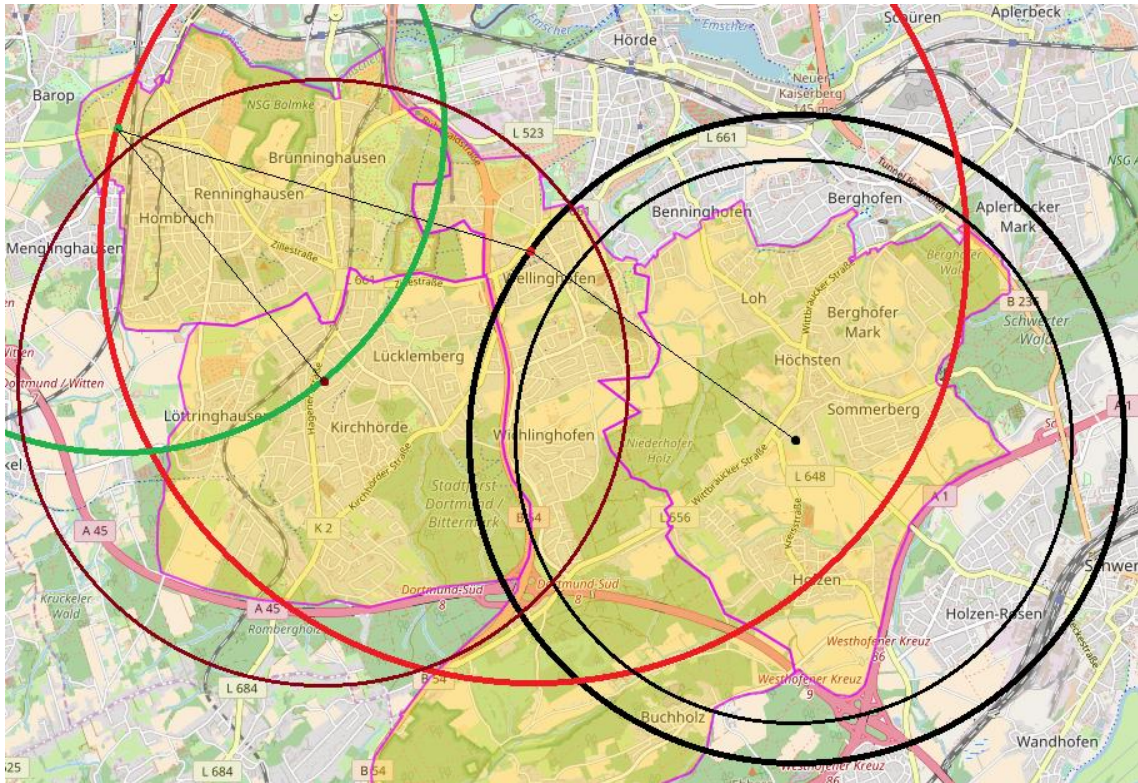


Abbildung 14: Beispielhafte Darstellung der Standortannäherung durch eine graphische Abbildung von Touren

Koordinaten der Ergebnisse

Durch die Anwendung der Verfahren sind entsprechend viele Approximationen zu den Standorten erzeugt worden. In der folgenden Tabelle sind die Koordinaten der einzelnen Annäherungen aufgelistet.

Tabelle 2: Koordinaten der Approximationen Teil 1

Gleichverteilung		Mittellinien		Tourenabbildung	
515.125.029	74.658.543	515.139.200	74.763.000	515.136.600	74.862.400
515.079.834	74.598.489	515.079.800	74.508.700	515.013.000	74.499.300
514.984.441	74.561.558	514.972.400	74.607.800	514.993.500	74.562.600
514.984.945	74.946.952	515.039.100	74.960.900	515.057.700	74.987.300
515.065.745	74.906.370	515.039.100	74.824.100	514.983.000	74.857.300
515.020.804	74.796.505	515.039.100	75.099.000	515.029.300	75.076.000
515.283.503	75.189.157	515.194.900	75.143.700	515.262.400	75.100.300
515.274.745	75.146.355	515.194.900	74.941.700	515.173.300	75.156.400

514.835.621	73.748.820	515.000.100	74.137.300	515.014.600	74.099.400
514.668.115	74.321.348	514.785.000	74.520.000	514.826.600	74.300.500
514.642.527	74.421.482	514.565.400	74.611.400	514.610.900	74.584.400
514.875.234	74.951.014	514.880.600	74.983.000	514.909.400	75.025.900
514.181.290	74.642.676	514.499.300	74.933.600	514.726.000	74.863.200
514.489.327	75.068.592	514.557.000	75.244.600	514.562.600	75.225.900
514.761.076	75.435.100	514.889.800	75.359.300	515.227.100	75.337.600
515.041.153	75.247.637	514.982.000	75.359.300	515.180.100	75.661.600
514.770.173	75.621.191	514.897.900	75.693.600	514.826.100	75.507.400
514.793.318	75.578.532	514.897.900	75.579.000	514.873.900	75.516.200
514.953.083	75.629.273	515.002.200	75.693.600	515.071.200	75.565.800
514.755.650	75.498.404	515.002.200	75.579.000	515.031.100	75.666.800
514.797.713	75.994.301	514.861.900	75.917.400	515.033.800	75.953.700
515.010.789	75.769.623	514.973.400	75.917.400	514.855.800	75.938.200
515.351.438	75.290.722	515.172.000	75.546.600	514.817.300	75.328.100
515.268.386	75.444.558	515.282.200	75.546.600	514.938.400	75.328.700
515.383.441	75.774.719	515.220.200	76.023.400	515.236.200	76.140.000
515.536.809	76.029.299	515.378.400	76.023.400	515.383.400	76.034.200
515.748.333	75.880.728	515.656.800	75.166.600	515.606.700	75.237.200
515.650.705	75.256.625	515.656.800	75.432.400	515.609.900	75.233.000
515.508.618	74.492.740	515.662.900	74.635.700	515.661.200	74.537.500
515.479.642	73.763.837	515.433.400	73.706.500	515.575.500	73.782.000
515.523.981	73.669.174	515.534.800	73.706.500	515.444.300	73.574.100
515.508.917	73.564.217	515.633.200	73.706.500	515.416.900	73.674.300
515.520.439	73.929.255	515.571.700	73.877.600	515.801.700	73.801.500
515.655.071	73.941.455	515.717.400	73.877.600	515.718.200	73.839.600
515.696.099	73.950.009	515.860.600	73.877.600	515.744.500	73.955.600
515.351.369	73.924.568	515.336.000	74.032.600	515.268.800	74.086.600
515.079.362	73.865.413	515.056.100	73.785.000	515.301.200	73.688.900
515.151.065	73.648.274	515.159.700	73.785.000	515.121.500	73.681.400
515.147.416	73.494.893	515.141.200	73.343.200	515.194.900	73.512.200
514.964.373	73.234.294	515.033.000	73.184.600	515.192.400	73.151.400
514.983.486	73.346.935	515.033.000	73.343.200	515.122.900	73.101.700
515.188.820	73.268.809	515.141.200	73.184.600	515.110.100	73.118.300
516.006.257	74.315.760	516.028.400	74.693.000	516.075.300	74.611.800
515.809.870	72.582.197	515.741.500	72.885.000	515.817.000	73.195.900
515.672.620	72.846.064	515.741.500	73.114.600	515.902.100	73.002.400
516.056.287	72.814.815	516.016.800	73.262.700	516.072.300	73.037.100
514.955.037	73.153.009	514.869.900	73.108.600	514.896.600	73.344.700
513.902.224	76.145.616	514.076.000	75.692.300	514.169.600	75.313.100
514.419.060	75.929.047	514.302.000	75.692.300	514.105.800	75.591.300
514.248.353	75.980.442	514.534.200	75.692.300	514.309.700	75.841.900
514.292.927	74.591.072	514.129.500	74.106.000	514.183.500	74.603.200
514.143.170	73.411.999	514.124.700	73.404.000	514.398.500	73.305.100
514.109.790	73.466.357	514.124.700	73.334.600	514.322.700	73.202.000

514.231.703	73.163.410	514.124.700	73.265.200	514.256.000	73.166.800
514.376.847	73.488.903	514.183.400	73.404.000	514.254.900	73.172.900
514.148.484	73.060.459	514.124.700	73.195.800	514.100.500	73.238.300
514.196.204	73.378.494	514.183.400	73.334.600	514.112.800	73.309.500
514.195.730	73.286.061	514.124.700	73.126.400	514.115.900	73.330.800
514.140.190	73.505.867	514.242.100	73.404.000	514.117.200	73.377.800
514.206.597	73.325.365	514.183.400	73.195.800	514.219.900	73.437.800
514.183.749	73.394.287	514.242.100	73.334.600	514.237.000	73.416.100
514.082.072	73.247.603	514.183.400	73.265.200	514.250.800	73.359.900
514.278.003	73.140.739	514.242.100	73.265.200	514.196.800	73.350.700
514.168.545	73.187.918	514.183.400	73.126.400	514.401.900	73.386.200
514.259.828	73.314.093	514.242.100	73.195.800	514.363.400	73.290.200
514.386.053	73.450.645	514.300.800	73.334.600	514.289.700	73.352.700
514.044.489	73.221.244	514.300.800	73.195.800	514.272.600	73.325.800
514.072.562	73.228.387	514.300.800	73.126.400	514.342.000	73.278.400
514.330.514	73.220.014	514.242.100	73.126.400	514.111.900	73.256.900
514.291.545	73.085.917	514.300.800	73.265.200	514.111.500	73.268.800
514.269.052	73.886.113	514.350.400	73.756.500	514.449.400	73.586.500
514.286.070	73.557.108	514.350.400	73.567.300	514.408.900	73.763.900
514.152.113	73.632.487	514.350.400	73.567.300	514.399.400	73.756.000
514.150.690	73.676.804	514.276.100	73.756.500	514.395.300	73.755.800
514.495.743	73.625.655	514.276.100	73.567.300	514.407.700	73.736.100
514.366.323	73.574.568	514.424.700	73.756.500	514.407.600	73.756.700
514.318.542	73.616.464	514.201.800	73.756.500	514.402.700	73.795.400
514.503.918	73.696.882	514.201.800	73.567.300	514.373.900	73.811.200
514.313.537	73.663.277	514.127.500	73.567.300	514.494.000	73.720.300
514.492.719	73.806.769	514.127.500	73.756.500	514.364.800	73.722.800
514.426.925	73.734.432	514.276.100	73.567.300	514.363.900	73.734.200
514.377.586	74.069.414	514.433.500	73.825.600	514.726.400	73.693.500
514.449.450	74.054.121	514.501.700	74.022.400	514.719.200	73.689.900
514.463.234	73.946.496	514.501.700	73.924.000	514.718.900	73.703.200
514.501.510	74.133.205	514.569.900	73.924.000	514.717.300	73.733.300
514.445.562	74.132.218	514.569.900	73.825.600	514.644.500	73.673.900
514.530.237	73.519.161	514.569.900	73.628.800	514.641.700	73.684.600
514.538.697	73.880.936	514.638.100	73.825.600	514.644.800	73.691.200
514.657.249	73.687.505	514.638.100	73.628.800	514.556.800	73.880.500
514.351.635	74.150.639	514.433.500	73.924.000	514.694.400	73.657.300
514.489.091	74.017.245	514.433.500	74.022.400	514.620.000	73.757.600
514.465.916	74.081.541	514.706.300	73.628.800	514.438.800	74.028.400
514.623.564	73.824.408	514.638.100	73.727.200	514.459.600	73.981.100
514.646.361	73.798.689	514.706.300	73.825.600	514.497.700	73.980.600
514.675.892	73.761.667	514.706.300	73.727.200	514.521.800	73.944.800
514.349.201	73.082.843	514.406.100	73.184.100	514.330.300	72.953.800
514.400.388	73.250.665	514.406.100	73.076.000	514.424.800	72.954.500
514.337.025	73.045.238	514.450.700	72.967.900	514.451.600	72.988.200

514.418.398	73.226.600	514.406.100	72.967.900	514.491.600	73.031.900
514.549.928	72.942.940	514.450.700	73.076.000	514.540.900	73.262.600
514.534.662	72.894.472	514.495.300	73.076.000	514.481.300	73.066.700
514.546.133	73.393.882	514.361.500	73.184.100	514.443.100	73.063.500
514.524.242	73.391.799	514.361.500	73.076.000	514.486.300	73.295.900
514.331.049	72.840.909	514.361.500	72.967.900	514.502.400	73.339.300
514.508.048	73.056.511	514.495.300	73.184.100	514.520.800	73.348.900
514.560.246	73.416.396	514.316.900	72.967.900	514.374.600	73.209.400
514.480.880	73.192.814	514.450.700	73.184.100	514.545.500	73.254.200
514.514.546	73.170.984	514.495.300	73.292.200	514.523.600	73.329.200
513.885.620	73.085.462	514.025.600	72.833.700	514.247.100	72.770.800
515.909.809	76.394.041	515.816.800	76.705.600	515.848.000	76.589.600

Tabelle 3: Koordinaten der Approximationen Teil 2

arithmetisches Mittel		Median		begrenzte Berechnungsgrundlage	
515.143.396	74.753.658	515.147.400	74.789.051	515.138.937	74.763.127
515.084.355	74.577.659	515.087.728	74.604.186	515.079.898	74.508.776
515.009.046	74.602.676	515.019.327	74.598.859	514.972.028	74.608.201
515.032.708	74.913.768	515.020.143	74.896.359	515.038.737	74.960.797
515.150.512	75.085.864	515.127.729	75.014.108	515.196.886	75.143.872
514.930.604	73.849.060	514.875.980	73.806.076	514.999.287	73.935.870
514.773.681	74.487.683	514.771.795	74.430.508	514.786.362	74.523.032
514.586.090	74.593.244	514.619.038	74.580.746	514.564.554	74.609.812
514.878.695	75.006.317	514.861.157	74.977.592	514.881.402	74.980.137
514.388.997	74.906.620	514.296.540	74.924.741	514.500.591	74.933.180
514.569.743	75.313.732	514.582.482	75.350.765	514.554.967	75.242.072
514.827.797	75.250.251	514.810.149	75.235.949	514.890.006	75.220.975
514.827.838	75.590.910	514.749.374	75.607.713	514.896.263	75.576.943
514.806.514	75.932.910	514.782.566	75.971.393	514.863.022	75.916.051
515.242.705	75.557.857	515.207.656	75.550.376	515.284.333	75.547.128
515.440.444	76.120.561	515.491.011	76.078.425	515.378.667	76.022.328
515.737.954	75.557.944	515.780.128	75.637.242	515.653.510	75.432.199
515.745.196	74.687.832	515.817.621	74.658.153	515.661.783	74.631.260
515.506.053	73.605.663	515.497.355	73.545.644	515.535.083	73.705.585
515.849.191	73.894.954	515.880.984	73.921.086	515.716.221	73.874.914
515.312.108	74.023.991	515.314.493	73.975.167	515.335.741	74.031.079
515.161.763	73.736.174	515.169.301	73.664.368	515.160.873	73.785.701
515.034.335	73.294.371	515.035.274	73.279.025	515.034.961	73.341.330
516.067.077	74.634.567	516.090.193	74.587.606	516.029.557	74.691.066
515.756.871	72.796.400	515.782.361	72.730.042	515.741.806	72.881.977
516.002.408	73.249.466	516.002.092	73.214.884	516.017.324	73.259.086
514.903.349	73.152.975	514.913.087	73.176.101	514.868.678	73.107.897

514.222.922	75.743.361	514.126.975	75.764.373	514.306.184	75.691.379
514.053.839	74.018.794	513.992.144	73.927.872	514.130.072	74.104.226
514.207.771	73.362.341	514.095.504	73.370.974	514.240.334	73.266.276
514.225.273	73.711.096	514.228.542	73.612.258	514.276.228	73.758.549
514.585.587	73.835.875	514.584.221	73.840.912	514.564.311	73.827.572
514.473.613	73.135.710	514.478.852	73.047.220	514.405.610	73.183.784
513.879.721	72.777.626	513.832.497	72.697.437	514.025.544	72.828.829
515.729.496	76.287.888	515.671.906	76.064.576	515.817.415	76.702.946