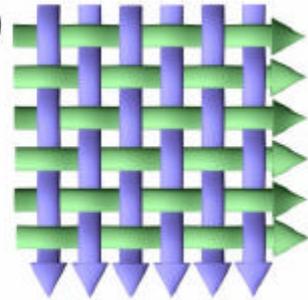


12. Dezember 2003

Sonderforschungsbereich 559  
**Modellierung großer  
Netze in der Logistik**



Technical Report 03014

ISSN 1612-1376

**Experimentelle Anwendung statistischer Verfahren  
und Visualisierungsmethoden zur Gewinnung  
ausgesuchter Eingangsdaten im Kontext von GNL**

Teilprojekt M9:

Jochen Bernhard

Fraunhofer IML

Joseph-von-Fraunhofer-Strasse 2-4

44227 Dortmund

Thomas Fender

Lehrstuhl Mathematische Statistik und  
industrielle Anwendungen

Vogelpothsweg 87

44221 Dortmund

Dortmund, den 12. Dezember 2003

## Inhalt

<b>1</b>	<b>Einleitung</b> .....	<b>3</b>
<b>2</b>	<b>Experimente am Anwendungsfall des Teilprojektes A5</b> .....	<b>4</b>
2.1	Kurzbeschreibung der Aufgabenstellung A5 .....	4
2.2	Zieldefinition .....	4
2.3	Informationsidentifikation .....	5
2.4	Erhebungsvorbereitung, Erhebung und Datenerfassung.....	6
2.5	Datenanalyse.....	6
2.6	Validierung und Visualisierung der Ergebnisse.....	7
<b>3</b>	<b>Experimente am Anwendungsfall „Parkhaus“</b> .....	<b>14</b>
3.1	Kurzbeschreibung der Aufgabenstellung .....	14
3.2	Zieldefinition .....	14
3.3	Informationsidentifikation .....	14
3.4	Erhebungsvorbereitung, Erhebung und Datenerfassung.....	15
3.5	Datenanalyse.....	15
3.6	Validierung und Visualisierung der Ergebnisse.....	15
<b>4</b>	<b>Ausblick</b> .....	<b>17</b>
<b>5</b>	<b>Literatur</b> .....	<b>18</b>

## 1 Einleitung

Zur Überprüfung des innerhalb des Methodenteilprojektes M9 entwickelten Vorgehensmodells zur Informationsgewinnung für die Anwendung in GNL wurde sowohl auf der Basis einer vorhandenen, außerhalb des SFB 559 angesiedelten Aufgabenstellung (Parkhaus) als auch in Zusammenarbeit mit dem Anwendungsteilprojekt A5 im SFB 559 ein Experimentierraum aufgespannt. Dieser Raum umfasst Untersuchungen mit unterschiedlichem Informationsbedarf und Informationsaufkommen in Abhängigkeit von den Modellierungsaufgaben und -methoden in den jeweiligen Projekten. Hierbei wurde speziell die Eignung der Methoden aus den Teildisziplinen Visualisierung und Statistik experimentell untersucht. Dieses Vorgehen diente der Methodenauswahl, dem Methodeneinsatz und der optionalen Adaption von Methoden und führte letztlich für die jeweiligen Methoden der Visualisierung und der statistischen Analyse zu einer Klassifikation und deren Überprüfung.

Primäres Entscheidungskriterium für die Auswahl der Experimentierszenarien war die Intention für die Informationsgewinnung. Einerseits dienten die gewonnenen Daten zur Entscheidungsgrundlage für die Definition der Systemgrenzen und der Einschränkung des Betrachtungs-/Untersuchungsraums auf relevante Systemkomponenten andererseits zur Eingangsdatenermittlung für die modellgestützte Analyse auf der Basis der Simulation.

Der Anwendungsfall des Teilprojektes A5 „Luffrachtknoten“ wurde wegen der Aufgabe einer Zustandsmodellierung (Gewinnung von Eingangsdaten für ein Simulationsmodell) und Systemverbesserung (Organisation und Lenkung des Frachtumschlags an einem Hub unter Berücksichtigung weiterer Knoten im Netz) ausgewählt. Im Rahmen dieser Anwendung waren für die Modellbildung zahlreiche Informationen in Form von Daten aber auch Entscheidungshilfen notwendig. Vor diesem Hintergrund wird in diesem Abschnitt das Vorgehensmodell zur Informationsgewinnung experimentell angewandt, um sowohl das Modell aber auch die hier genutzten Methoden der einzelnen Disziplinen zu evaluieren.

Der SFB-externe Anwendungsfall „Parkhaus“ wurde wegen einer benötigten systembezogenen Prognose gewählt. In diesem Fall waren für eine Modellbildung nicht nur Informationen in Form von Daten notwendig, sondern vor allem auch Informationen zu den Annahmen, die dem jeweiligen Modell zu Grunde liegen müssen, damit die Aussagekraft der Modellierungsergebnisse erhalten bleibt. Im vorliegenden Fall einer Simulation eines automatischen Parksystems wird das Vorgehensmodell experimentell zur Informationsgewinnung angewandt, um vor allem die Notwendigkeit der Nutzung statistischer Analysemethoden und Visualisierungsverfahren zu validieren.

Die in diesem Bericht vorgestellten Beispiele sind auf Teilaspekte einer Aufgabenstellung eingeschränkt, da die Darstellung einer ganzheitlichen Durchführung des Informationsprozesses unter Anwendung des Vorgehensmodells im Vordergrund stand. Von besonderer Bedeutung ist allerdings, dass hier das Modell vollständig auf Beispiele mit verschiedenen Zielvorgaben angewendet wurde, so dass die Generalisierungsfähigkeit und Übertragbarkeit des Vorgehensmodells experimentell bestätigt werden konnte.

## 2 Experimente am Anwendungsfall des Teilprojektes A5

Die Untersuchung isolierter Luftfrachtknoten hat gezeigt, dass ein wesentlicher Einflussfaktor, welcher auf den Frachtumschlag-Prozess an einem Knoten einwirkt, die erbrachten Vorleistungen vorhergehender Knoten sind. Aufbauend auf diesen Ergebnissen der ersten Phase des SFB hat das Teilprojekt A5 ein erweitertes Modell entwickelt, welches das für die Beziehungen und Wechselwirkungen zwischen Flughäfen relevante Teilsystem im Sinne eines Ausschnittes des Gesamtnetzes abbildet. Für die konkrete Modellbildung sind zahlreiche Informationen in Form von Daten aber auch Entscheidungshilfen notwendig. Vor diesem Hintergrund wird im Folgenden die Anwendung des Vorgehensmodells zur Informationsgewinnung experimentiert, um sowohl das Modell aber auch die hier genutzten Methoden der einzelnen Disziplinen zu evaluieren.

### 2.1 Kurzbeschreibung der Aufgabenstellung A5

Im Rahmen der zweiten Phase des SFB-Teilprojekts A5 sollen die Wechselwirkungen zwischen Knoten im Netz untersucht werden. Dazu sind untereinander in Beziehung stehende Knoten in einem Modell abzubilden. Den elementaren Prototyp für diese Wechselwirkungen liefern die Beziehungen zwischen zwei zeitpunktgeführten Knoten. Die Verkettung von Knoten erfolgt grundsätzlich entlang einer Transportkette, die über einen Teil des Netzes abgewickelt wird.

Eine typische Transportkette für Luftfrachttransporte ist gekennzeichnet durch den Vorlauf, den Hauptlauf und den Nachlauf. Diese Kette benutzt insgesamt vier Umschlagknoten, von denen der erste die Quelle und der letzte die Senke der zeitpunktgeführten Prozesse darstellt. Die beiden Umschlagknoten innerhalb der Kette besitzen jeweils eine Hubfunktion, so dass damit eine direkte Beziehung zwischen zwei zeitpunktgeführten Hubknoten Hub B und Hub C vorliegt. Die Erweiterung im Vergleich zum Modell aus der ersten Phase des Teilprojektes ist darin zu sehen, dass die Ausprägung und die räumliche Zuordnung von Umschlagaktivitäten nicht für einen Knoten determiniert ist, sondern zwischen zwei Knoten variiert werden kann.

Jeder Feeder-Flughafen hat ein spezifisches Frachtaufkommen mit unterschiedlichen Zielflughafen-Anteilen. Diese kommen in der Summe an einem Hub B an, wo entschieden werden muss, ob die Luftfrachtsendungen mit noch zu erwartenden weiteren Sendungen gesammelt und gebündelt oder gleich zum Hub C weiter transportiert werden sollen. Entsprechendes gilt für das Auflösen und Verteilen am Hub C.

### 2.2 Zieldefinition

Ziel der Untersuchungen ist es, bezogen auf ein abgegrenztes Untersuchungsfeld, mit Hilfe des erweiterten Modells die Einflüsse und Auswirkungen von einzelnen Maßnahmen (Parameteränderungen) an einem der Hubknoten auf den jeweils anderen Knoten zu untersuchen. Im Gegensatz zu SFB-Anwendungsprojekten, die ein logistisches Netz untersuchen, erfolgt hier keine Netz- oder Routenoptimierung, sondern eine flughafenübergreifende Optimierung der standortinternen Umschlagaktivitäten hinsichtlich Kosten, Durchlaufzeiten und Ressourcenverbrauch.

## 2.3 Informationsidentifikation

### Informationsbedarfsanalyse

Da sich die zu treffenden Aussagen auf den zeitpunktgeführten Frachtumschlag von einzelnen Sendungen in dem oben beschriebenen Netz auf zwei Hubs beziehen sollen, ergeben sich als notwendiger Informationsbedarf für die geplanten Untersuchungen Aussagen pro Hub zu den Sendungen, zum Flugplan und zu den eingesetzten Flugzeugtypen:

#### Sendungen

- Start-, Zielflughafen, Routing:  
Festlegung der Route über die Umschlagflughäfen der Fracht
- Startzeitpunkt, spätester Zielzeitpunkt zur Modellierung von Termintreue
- Gewicht, Anzahl pro Sendung, evtl. Volumen

#### Flugplan

- Start- und Landzeitpunkt der Flugzeuge an den Umschlagflughäfen
- Routing der Flugzeuge
- Flugzeugtyp (Kapazität bzw. Zuladungsvolumen und –gewicht, Geschwindigkeit, verwendbare Frachtladungsträger), im Speziellen das reine Frachtzuladevolumen
- Anteil der Frachtsendungen an der Gesamtfracht pro Flugzeug

#### Flugzeugtyp

- Zuladungsvolumen (Fracht)
- Zuladungsgewicht (Fracht)
- Geschwindigkeit
- verwendbare Frachtladungsträger

### Informationsangebotsanalyse

Informationen bzgl. der Sendungsdaten für den ausgesuchten Hub B sind nur für einen Carrier vorhanden. Für diesen sind allerdings die meisten Informationen aus verschiedenen Datenbanken abrufbar: Angaben zu den Sendungen, den Start- und Zielflughäfen, sowie für Zwischenlandungen auf bis zu zwei Hubs. Es fehlen aber Informationen zu weiteren Zwischenlandungen, so dass das vollständige Routing nicht immer ermittelt werden kann. Ein weiterer wichtiger Punkt ist das Fehlen eines exakten Startzeitpunktes einer Sendung, so dass hierfür die Abflugzeit am Startflughafen genutzt werden muss. Zusammengefasst stehen dem Anwendungsprojekt A5 folgende Informationen zur Verfügung:

#### Sendungen (Daten für Hub B, eines Transportanbieters über einen Monat)

- Start-, Zielflughafen
- Flugnummer und Datum der Flüge, mit denen die Fracht an Hub B ankommt bzw. abtransportiert wird
- Gewicht, Anzahl der Frachtstücke pro Sendung

#### Flugplan (Daten des Hub B, alle Ankunfts- und Abflugzeiten über einen Monat)

- Flugnummer
- Datum und Uhrzeit (tatsächlich / geplant)

- Routing des Flugs
- Flugzeugtyp, mit Abfluggewicht

Aktuelle Flugpläne beliebiger Flughäfen (Plandaten pro Flughafen oder Fluggesellschaft)

- Start- und Landezeit
- Flugnummer

Flugzeugtyp (Hersteller, Flughafenumfeld)

- Zuladungsvolumen und -gewicht (gesamt)
- Geschwindigkeit
- verwendbare Frachtladungsträger

Flughafen-Code-Zuordnung

- IATA, ICAO, Klartextbeschreibung

Road-Feeding-Systematik

- LKW-Anlieferung und Abtransport per Flugnummer identifizierbar

### **Nutzbarkeitsprüfung – Informationsqualität**

Der Flugplan an Hub B ist für die formulierte Aufgabenstellung und damit für den ermittelten Informationsbedarf vollständig, die Flugpläne für den weiteren Hub C oder für Start- und Zielflughäfen sind nur in eingeschränkter Form über sekundäre Informationsquellen, wie beispielsweise das Internet, verfügbar. Da in diesen Quellen auch nur aktuelle Daten zur Verfügung stehen, passen diese zeitlich nicht zu den Flugplandaten an Hub B. Außerdem sind in dieser Art Informationsquellen auch nur die Daten für Passagiermaschinen abgelegt.

Die Flugzeugtypen sind nur im Flugplan für Hub B aufgeführt, ebenso das zugeladene, gesamte Frachtgewicht. Spezielle Daten für Fracht existieren nicht.

## **2.4 Erhebungsvorbereitung, Erhebung und Datenerfassung**

Die Daten wurden von den verschiedenen Ansprechpartnern und Datenquellen vollständig in digitaler Form erhoben. Anschließend erfolgte eine einheitliche Ablage der Daten mittels quellenspezifischer Importfunktionen, eine parameterspezifische Benennung und Datentyp-anpassung, Indizierung und Definition von Relationen in einem relationalen Datenbanksystem. Als vorbereitende Hilfsmittel für die Aufbereitung und Analyse der Daten wurden des Weiteren verschiedene SQL-Abfragen und Filter innerhalb der Datenbank definiert und zur Verfügung gestellt.

## **2.5 Datenanalyse**

Für eine Modellierung gemäß der Aufgabenstellung dieser Anwendung ist die Ermittlung der potentiell nutzbaren Eingangsdaten aus den bereits erfassten Daten unabdingbar. Dazu ist eine zielorientierte, der Aufgabenstellung angepasste Anwendung und Verknüpfung von statistischen Datenanalyseverfahren und Visualisierungsverfahren notwendig.

Mit Hilfe von Datenbankabfragen und multivariater deskriptiver Statistik kann eine Häufigkeitsverteilung der Sendungsströme mit Fokus darauf, dass alle Sendungen über den Hub B abgefertigt werden, erstellt werden. In den daraus entwickelten Sankey-Diagrammen sind neben Hub B sowohl die Start- und Zielflughäfen der Sendungen als auch weiterer Flughäfen, die als Hub C in Frage kommen, mit aufgenommen. Die Visualisierung erfolgt über zweidimensionale Häufigkeitstabellen und zweidimensionale Block-Charts.

Das Ankunftsverhalten der Sendungen wird mit deskriptiver Zeitreihenanalyse auf saisonale Effekte, wie z.B. Tages- bzw. Wochenrhythmus, und Trends untersucht. Dabei werden sowohl Ähnlichkeiten bei den Mengenniveaus als auch im strukturellen Verteilungsverhalten im zeitlichen Verlauf aufgedeckt.

Es muss eine Klassifizierung der Flugzeugtypen nach der realistischen Frachtzuladung im Gegensatz zu einer Klassifizierung nach dem Abfluggewicht vorgenommen werden, um eine Verfälschung bei der möglichen Zuladung der Fracht pro Flugzeugtyp zu vermeiden. Dies geschieht mit statistischen Verfahren zur Clusteranalyse aus den Ankunfts- und Abflugdaten in Hub B. Dabei sind die Größe des Datensatzes, sowie fehlerhaft eingegebene oder fehlende Daten eine besondere Herausforderung. Dazu sind Clustermethoden nötig mit schnellen Algorithmen und guten robusten Eigenschaften.

## 2.6 Validierung und Visualisierung der Ergebnisse

### Sendungen Hub B

Zur Bestimmung des Ankunftszeitverhaltens von Sendungen wurden zunächst alle Ankünfte von Sendungen innerhalb des betrachteten Zeitraums von 4 Wochen und deren Gewicht in ein Halbstundenraster eingeteilt und visualisiert (siehe Abbildung 1). Die Darstellung der Ankünfte mit halbstündiger Genauigkeit ist zur Auswertung ausreichend, da für eine kleinere Aufteilung nicht genügend Messwerte und damit Lücken in der Darstellung und eine größere Aufteilung ein zu grobes Abbild der Ankunftsverteilung ergeben. Des Weiteren erfolgt die anschließende Bearbeitung der ankommenden Sendungen im Umschlagterminal nicht minutengenau.

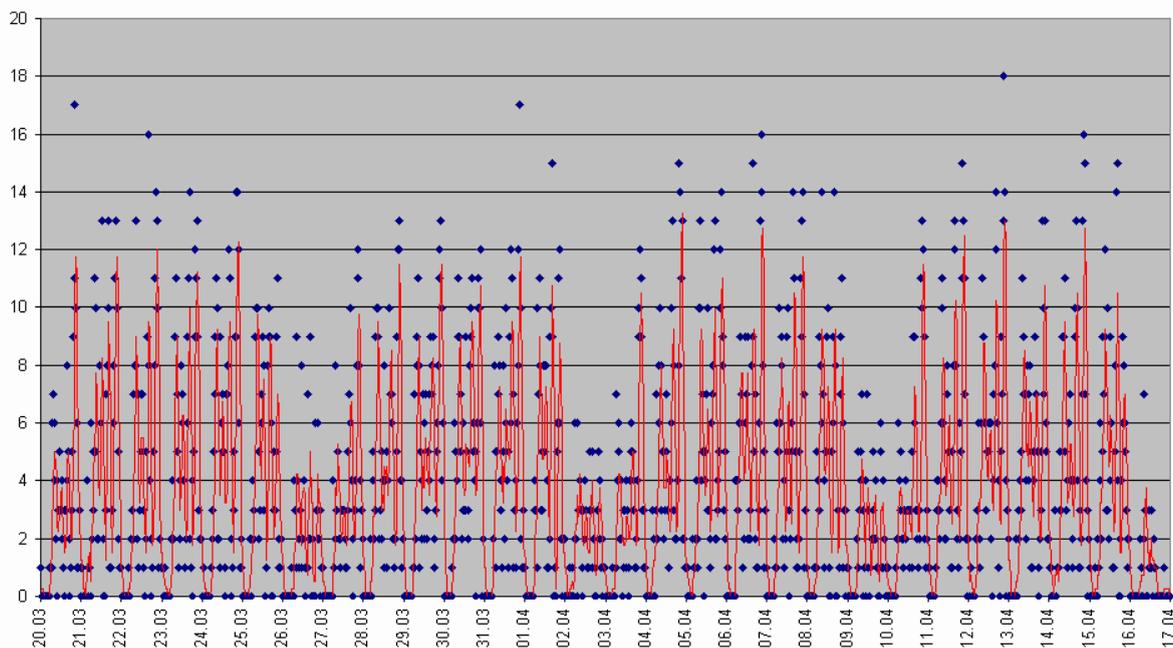


Abbildung 1: Ankommende Sendungen im Halbstundenraster an Hub B.

Anhand dieser deskriptiven Darstellung konnte sowohl ein typisches Tagesprofil als auch ein typisches Wochenprofil bezüglich der Anzahl ankommender Sendungen auf dem Hub B identifiziert werden. Die jeweiligen Tagesprofile unterscheiden sich demnach nicht strukturell, aber sehr wohl in der absoluten Menge ankommender Sendungen. Deshalb wurden mit Hilfe des Wochenprofils die Unterschiede des mittleren Sendungsaufkommens und der Streuung an den jeweiligen Tagen ermittelt (siehe Abbildung 2). Dies diente als Basis zur Normierung der Wochentage, um ein einheitliches Niveau für ein typisches relatives Tagesprofil zu erhalten (siehe Abbildung 3). Zur Ermittlung eines spezifischen Sendungsaufkommens zu

einer bestimmten Tageszeit an einem bestimmten Wochentag muss der Tagesmittelwert mit dem relativen Tagesprofil verrechnet werden.

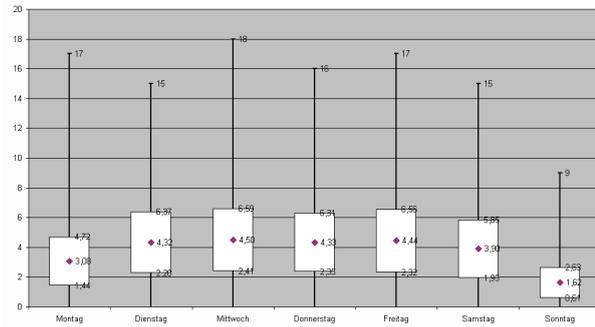


Abbildung 2: Tagesverteilung der Anzahl der Sendungen im Wochenprofil (Ankunft)

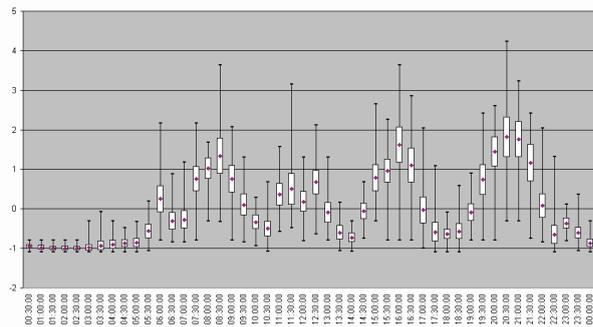


Abbildung 3: relatives Tagesprofil der Anzahl der Sendungen im Halbstundenraster (Ankunft)

Durch diese Aufbereitung der Daten wird eine deutliche Vereinfachung der Eingangsdaten für die Simulation gewährleistet. Hierbei reicht es aus, das relative Tagesprofil und die jeweiligen Niveaus der Wochentage zu verwenden. Gleichzeitig wurde verifiziert, dass die Ankünfte der Sendungen über vier Wochen ein ähnliches Verhalten aufweisen, so dass die Datenbasis dahingehend verbreitert werden kann, dass pro Messzeitpunkt (d. h. pro halbe Stunde) 28 Messwerte (4 Wochen x 7 Tage) zur Verfügung stehen. Damit können sowohl die Mittelwerte als auch die Standardabweichungen im relativen Tagesprofil als aussagekräftig bezeichnet werden.

### Flugplan Hub B

Zur Abbildung des Flugplans auf Hub B, d.h. des Ankunftszeitverhaltens bezüglich landender und startender Flugzeuge, wurden wiederum alle Landungen und Starts innerhalb des betrachteten Zeitraums von vier Wochen in ein Halbstundenraster eingeteilt und visualisiert (siehe Abbildung 4 und Abbildung 5).

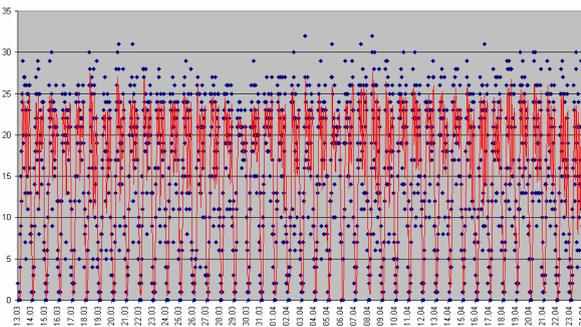


Abbildung 4: Landungen auf Hub B im Halbstundenraster

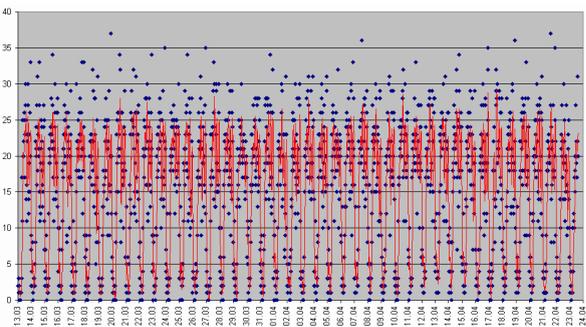


Abbildung 5: Starts auf Hub B im Halbstundenraster

Die deskriptive statistische Analyse ergab, dass im Gegensatz zum Sendungsaufkommen kein oder ein zu vernachlässigendes Wochenprofil zu erkennen ist, welches in der Varianz des anschließend bestimmten Tagesprofils mit eingeschlossen ist (siehe Abbildung 6 und Abbildung 7). Die Normierung der Werte kann damit entfallen, trotzdem ist die Verbreiterung der Datenbasis auf 28 Messwerte pro halbe Stunde auch hier möglich.

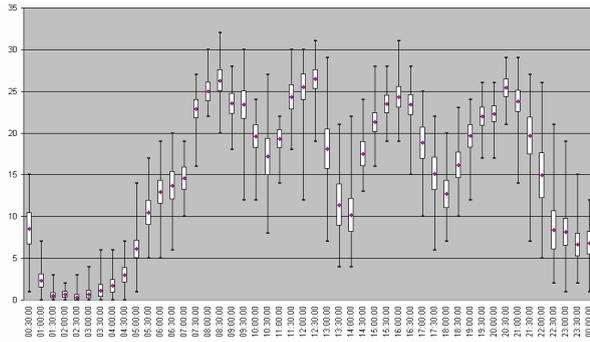


Abbildung 6: Tagesprofil Landungen pro halbe Stunde pro Tag

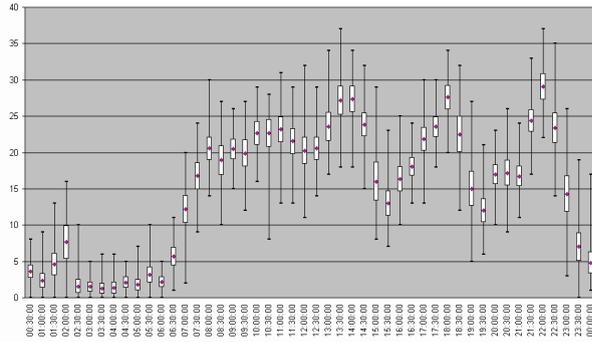


Abbildung 7: Tagesprofil Starts pro halbe Stunde pro Tag

### Auswahl Hub C

Die Auswahl eines geeigneten Flughafens als Hub C, um mittels Daten die Sendungsströme (Origin-Destination-Relation der Sendungen) über vier Umschlagknoten zu betrachten, hängt im Wesentlichen davon ab, dass eine ausreichende Menge an Fracht in den Sendungsdatenbanken des Hub B identifizierbar sind, die nicht nur in Hub B, sondern auch in Hub C umgeschlagen werden, so genannte Transfersendungen.

Nach Bereinigung der Sendungsdatenbanken (Sendungen Inbound und Sendungen Outbound) von den Sendungen, die nicht eindeutig einem Transport vom bzw. zum Hub B zuzuordnen waren, blieben 63817 Transfersendungen, die innerhalb von vier Wochen über Hub B umgeschlagen wurden (Abbildung 8). Das Sendungsaufkommen aus den einzelnen Start- und Zielflughäfen, das an Hub B umgeschlagen wird, ist ausreichend hoch (Abbildung 8(a)), aus Abbildung 8(b) wird allerdings deutlich, dass nur bei wenigen Relationen ein hohes Frachtaufkommen vorhanden ist.

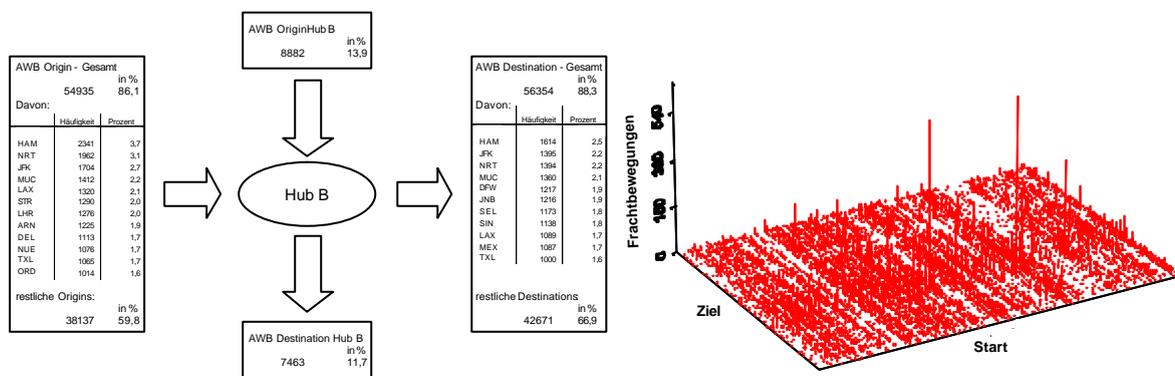


Abbildung 8: (a) Sendungsströme am Hub B (gesamt 63817 Sendungen (AWB)) mit den häufigsten Start- bzw. Zielflughäfen. (b) Visualisierung der 63817 Origin-Destination-Relationen über Hub B.

Abbildung 9 ist ein Ausschnitt aus Abbildung 8(b) und zeigt die 30 Relationen mit mehr als 100 Transfersendungen. Diese Grafik macht bereits deutlich, dass trotz der hohen Anzahl an Fracht, die an Hub B umgeschlagen wird, viele Sendungsströme nur schwach frequentiert sind. Ein Routing über mindestens vier Umschlagknoten mit zu Hub B fest gewähltem Hubknoten Hub C, wie es hier in der Zieldefinition gefordert wird, wird deshalb nur von wenigen Sendungsrelationen eingehalten.

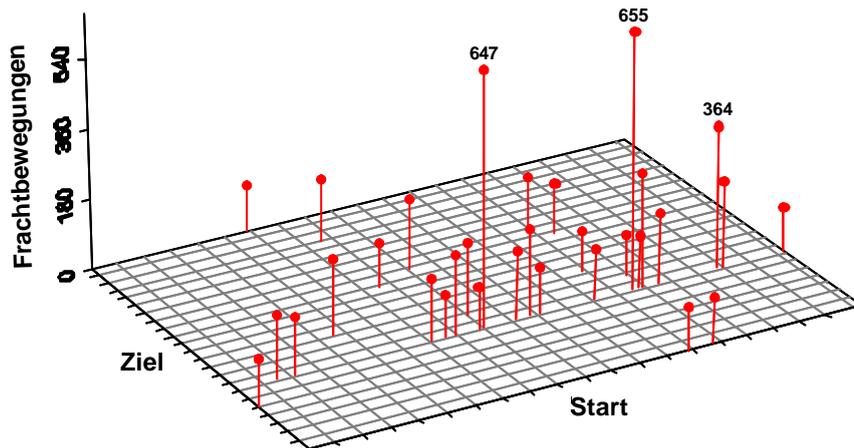


Abbildung 9: Origin-Destination-Relationen mit mehr als 100 Transfersendungen im Beobachtungszeitraum.

Die beispielhafte Visualisierung der Sendungsströme von einem Startflughafen A über einen Hubknoten B und einen Hubknoten C zum Zielflughafen D in Abbildung 10 macht dementsprechend deutlich, dass nur wenige Relationen stark genug besetzt sind, um diese als Eingangsdaten für eine Simulation nutzen zu können. Außerdem muss die Tatsache Beachtung finden, dass bei einer Vielzahl von Relationen entweder Hub B der Start- bzw. Hub C der Zielflughafen der Fracht ist. Die grün gekennzeichneten Sendungsströme in Abbildung 10 dagegen stellen im Sinne der Zieldefinition die interessanten Relationen in ausreichender Anzahl dar.

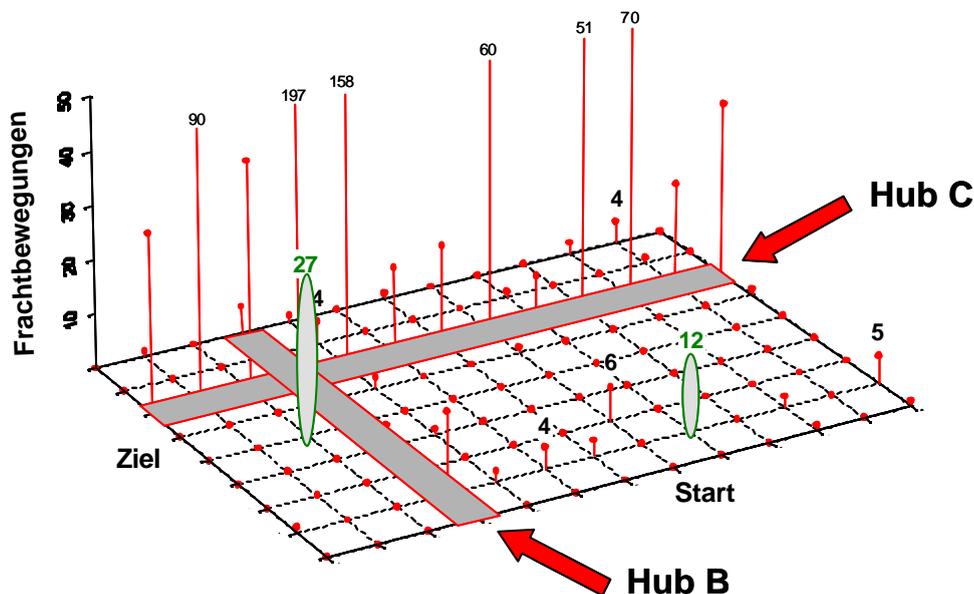


Abbildung 10: Origin-Destination-Relationen über zwei Hubknoten Hub B und Hub C.

Wie in Abbildung 10 beispielhaft aufgezeigt, ist ein wichtiges Ergebnis dieser visuellen Analyse für verschiedene Hubknoten Hub C, dass die Origin-Destination-Matrizen grundsätzlich immer eine gleiche Struktur haben, so dass für eine simulationsbasierte Analyse es ausreicht, die Sendungsströme über zwei Flughäfen als Hubknoten Hub B bzw. Hub C zu betrachten.

Für die Auswahl eines geeigneten Hub C werden hier beispielhaft die Sendungsströme unter Berücksichtigung des Hub B für vier Flughäfen die als Hubknoten in Frage kommen detailliert aufgeführt (Abbildung 11). Grundsätzlich muss bei der Darstellung der Sendungsströme als

potentielle Systemlast des Simulationsmodells sowohl zwischen der auf die Zielstellung bezogene Last als auch der an den Hubknoten zusätzlich auftretende Grundlast differenziert werden, um die Umschlagskapazitäten an den Hubs korrekt darstellen zu können. Relevante Informationen, die nicht aus den zur Verfügung stehenden Datenbanken extrahierbar sind, sind in Abbildung 11 jeweils mit einem Fragezeichen gekennzeichnet und müssen über Sekundärinformationen oder durch geeignete Generierung erstellt werden. Die rot gekennzeichneten Start und Zielflughäfen (AWB Origin bzw. AWB Destination) geben die Flughäfen an, bei denen in der Origin-Destination-Matrix ausreichend hohe Frachtbewegungen stattfinden.

Grundsätzlich machen die Sendungen, die direkt im Hub B ihr Origin haben bzw. direkt im Hub C ihre Destination einen großen Teil der Sendungen in der systembezogenen Last aus, so dass sich die in der Zielstellung gewünschten Relationen über vier Umschlagknoten hier auf drei oder sogar zwei Knoten reduzieren. Die Entscheidung, welcher Hub C der für die Zieldefinition am Besten geeignet ist, kann deshalb nicht eindimensional entschieden werden. Nicht nur die Anzahl der Sendungen, die überhaupt von Hub B zu einem Hub C transportiert werden, ist für eine sinnvolle Simulation bedeutend, sondern auch, die Anzahl der Sendungen von einem Startflughafen zu mindestens dem Hub C und die Anzahl der darüber hinaus gehenden Sendungen zu einem vom Hub C unterschiedlichen Zielflughafen. Ebenso müssen die Gesamtanzahlen der Sendungen von den AWB-Origins zum Hub B berücksichtigt werden, die einen Großteil der Belastung beim Umschlag der Sendungen auf diesen Routen ausmachen.

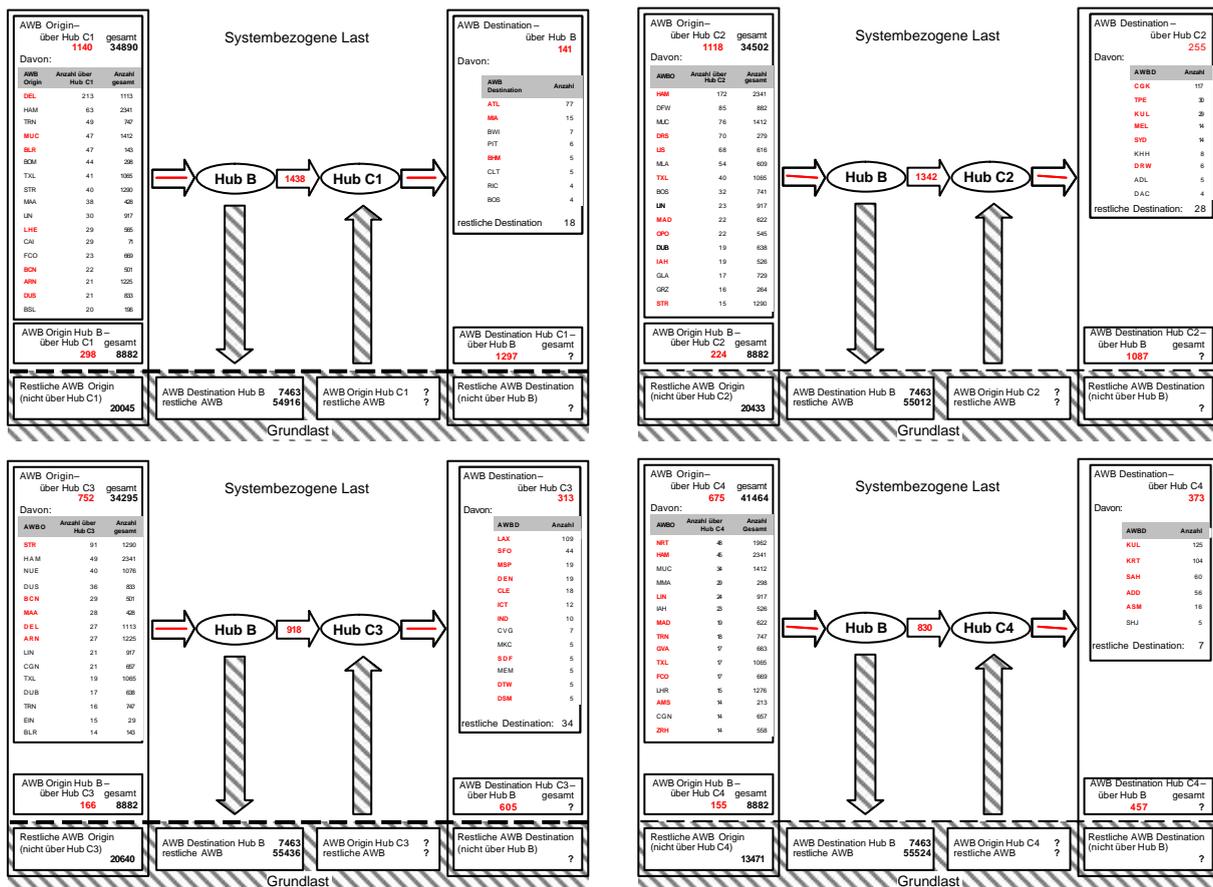


Abbildung 11: Origin-Destination-Relationen von Frachtsendungen (AWB) über einen Hubknoten Hub B und einen weiteren Hubknoten (a) Hub C1, (b) Hub C2, (c) Hub C3 bzw. (d) Hub C4.

## Flugplan und Sendungen Hub C

An dem zu wählenden Hubknoten Hub C ist zur Zeitpunkt-Steuerung ein Flugplan notwendig. Da keine expliziten Daten für den von Hub B bekannten Beobachtungszeitraum vom Anwender genutzt werden können, muss in diesem Fall entweder auf geeignete Sekundärinformationen zurückgegriffen werden, wie z.B. aktuelle Flugpläne aus dem Internet. Oder es kann auf der Basis des Flugplan-Tagesprofils von Hub B bzw. eines anderen Tagesprofils mit geeigneten Daten zur Frequentierung am Hub C ein Flugplan durch Hochrechnung und Prognose generiert werden.

Ebenfalls liegen die Lastdaten, die an Hub C anfallen, nicht vor. Nur die aus Hub B anfallenden Sendungen können als Basis genutzt werden. Damit fehlen speziell auch Informationen bezüglich der zu betrachtenden Grundlast in Hub C (siehe Abbildung 11). Neben der Beschaffung von Sekundärinformationen, gibt es in diesem Fall nur die Möglichkeit einer sehr groben Verteilungsschätzung der Sendungsströme.

Die Problematik der fehlenden realen Datenbasis sollte bei der Entscheidung, welcher Flughafen als Hubknoten Hub C ausgewählt wird, mit berücksichtigt werden. Gerade Sekundärinformationen, mit denen auf das Frachtaufkommen an Hub C geschlossen werden kann, würden die Abbildungsgültigkeit enorm erhöhen.

## Flugzeugtyp

Die Beschreibung der Transportkapazitäten beim An- und Abflug an den Hubknoten hängt entscheidend von der Frachtkapazität der jeweiligen Flugzeugtypen ab. In der Flugbewegungsdatenbank sind 149 verschiedene Flugzeugtypen registriert, inklusive der Spezialtypencodes, die häufig aus konstruktionstechnischen Gründen maßgeblich unterschiedliche Frachtkapazitäten beschreiben, erhöht sich die Anzahl auf 349 Typen.

Da andererseits das maximale Abfluggewicht und auch die tatsächlich transportierte Fracht bei vielen unterschiedlichen Typen sehr ähnlich sind, ist es sinnvoll die Flugzeugtypen bzgl. ihrer Frachtkapazität in Gruppen zu klassifizieren. Dabei muss notwendigerweise die real transportierte Fracht im Vordergrund stehen. Das maximale Abfluggewicht der Flugzeugtypen stellt nur einen unzureichenden Klassifizierungsparameter dar, denn sowohl das Flugzeug- als auch das Gepäckgewicht und die Anzahl der Passagiere verringern die tatsächliche Frachtkapazität.

Mit Hilfe der Daten zum Frachtgewicht, sowie als Kovariablen die Gepäckgewichts-, Postgewichtsdaten und die Anzahl der Passagiere aus der Flugbewegungsdatenbank von Hub B kann mit einer statistischen Clusteranalyse eine Klassifizierung bzgl. der Transportkapazitäten durchgeführt werden. Da die Datenaufnahme zu den oben beschriebenen Variablen oft manuell durchgeführt wird, gibt es zahlreiche fehlende oder unplausible Datenwerte. Deshalb müssen robuste Klassifizierungsverfahren wie CLARA [KR09] oder robustifizierte Verfahren – in diesem Anwendungsfall eine klassische Clusteranalyse mit k-means-Algorithmus und robuste Vorverarbeitung durch geeignete Parameterschätzung [RDr99] – zum Einsatz kommen. Die Durchführung beider Verfahren führt zu ähnlichen Segmentierungen der Flugzeugtypen bezüglich ihrer tatsächlichen Frachtkapazität. Sowohl mit dem Algorithmus CLARA als auch bei Durchführung des k-means-Algorithmus ergibt sich als bestes Clusterergebnis eine Aufteilung in sechs Flugzeugklassen (vgl. [Kuh03]). Dabei werden mit dem Verfahren CLARA die einzelnen Typen mit unterschiedlicher Häufigkeit den Clustern (vgl. Abbildung 12) zugeordnet, so dass sich eine Verteilungsvorschrift ergibt, mit welcher Häufigkeit ein Flugzeugtyp einem Cluster zugeordnet werden kann.

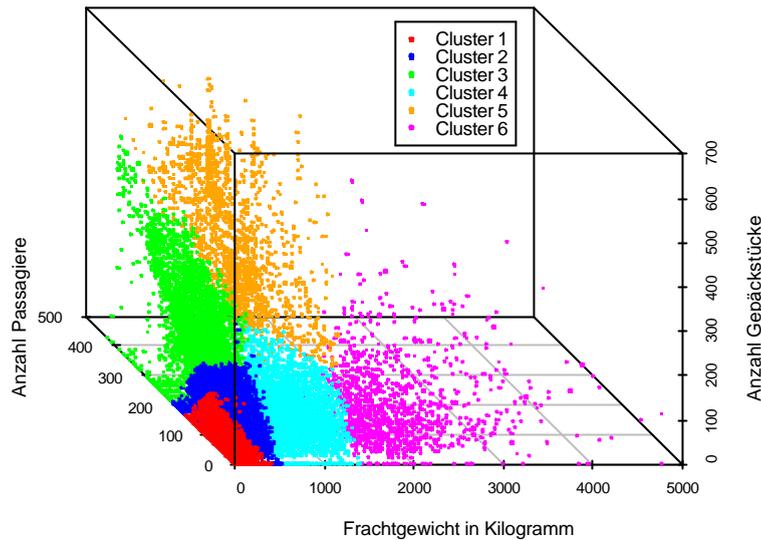


Abbildung 12: Klassifizierung der Flüge an Hub B in 6 Cluster bezüglich Frachtgewicht (in kg), Anzahl der Gepäckstücke und Anzahl der Passagiere mit dem Algorithmus CLARA.

Im Gegensatz zu dieser sensitiven Klassifizierung liefert das robustifizierte k-means-Verfahren eine eindeutige Aufteilung auf die sechs Cluster. Dabei gehen für alle oben beschriebenen Einflussvariablen sowohl Lage- als auch Streuungsparameter der Flugzeugtypen ein (vgl. Abbildung 13). Die Eindeutigkeit der Klassifizierung vereinfacht somit auch die Anwendbarkeit dieser Ergebnisse als Eingangsdaten für eine Modellierung, da jeder Flugzeugtyp exakt durch eine Frachtkapazitätsklasse modelliert werden kann. Andererseits liefert die sensitive Klassifizierung exaktere Eingangsdaten in Form einer Verteilung über die sechs Cluster für jeden Typ.

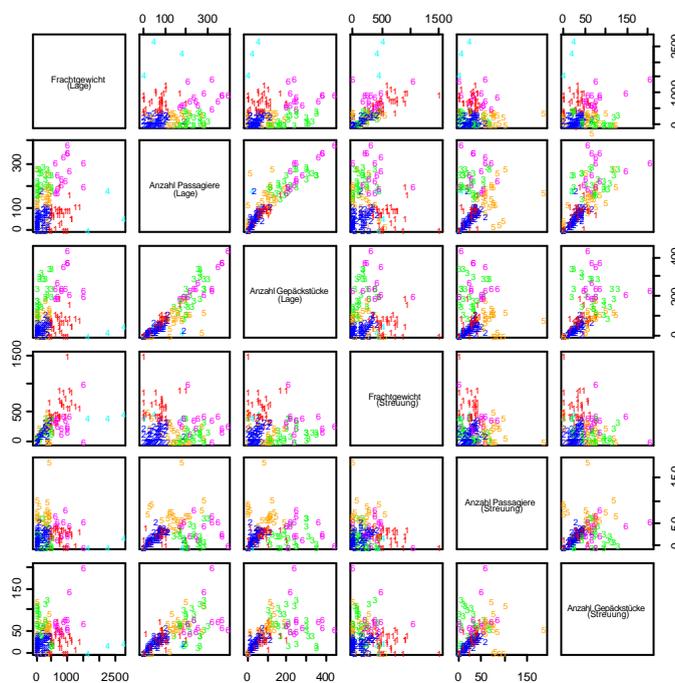


Abbildung 13: Klassifizierung der auftretenden Flugzeugtypen an Hub B mit k-means-Algorithmus in 6 Cluster bezüglich Frachtgewicht (in kg), Anzahl der Gepäckstücke und Anzahl der Passagiere unter Verwendung robuster Lage- und Streuungsmaße für jeden Flugzeugtyp.

### **3 Experimente am Anwendungsfall „Parkhaus“**

Zur weiteren Verdeutlichung der Notwendigkeit einer integrativen Informationsgewinnung wird in diesem Abschnitt anhand der Problematik systembezogener Prognosen aufgezeigt, dass für eine konkrete Modellbildung nicht nur Informationen in Form von eigenschaftsbestimmenden Daten notwendig sind, sondern vor allem auch Informationen zu den Annahmen, Restriktionen und Lenkungsalgorithmen, die dem zu modellierenden System zu Grunde liegen müssen. Im vorliegenden Fall einer Simulation eines automatischen Parksystems wird das Vorgehensmodell experimentell zur Informationsgewinnung angewandt, um vor allem die Notwendigkeit der Nutzung statistischer Analysemethoden und Visualisierungsverfahren zu validieren.

#### **3.1 Kurzbeschreibung der Aufgabenstellung**

Im Gegensatz zu herkömmlichen Parkhäusern müssen bei der Planung von automatisierten Parkhäusern zusätzlich die Beförderungskapazitäten zur Ein- und Auslagerung der Fahrzeuge berücksichtigt werden, da gerade in Innenstädten Rückstauungen auf die öffentlichen Verkehrswege vermieden werden müssen. Die Vorgaben an die Kapazität des automatischen Förderers der Fahrzeuge tragen außerdem einen großen Anteil an den Kosten des Parkhauses, so dass eine möglichst exakte Planung mit Hilfe der Simulation unter besonderer Berücksichtigung der Ein- und Ausfahrten der Fahrzeuge wünschenswert ist.

#### **3.2 Zieldefinition**

Für die Simulation eines geplanten automatischen Parksystems in der Innenstadt soll das Ein- und Ausfahrtverhalten der Kunden modelliert werden. Dabei ist insbesondere die Anzahl der gleichzeitigen Ein- und Ausfahrten von hohem Interesse, denn diese tragen wesentlich zur Auslastung des automatischen Förderers bei. Bei diesen so genannten Doppelspielen treten fast keine Leerfahrten des Förderers auf, was einer optimalen Auslastung entspricht.

#### **3.3 Informationsidentifikation**

##### **Informationsbedarfsanalyse**

Es werden Informationen über das Parkaufkommen an dem Standort und den potenziellen Bedarf an Parkplätzen benötigt. Um eine möglichst exakte Modellierung unter Berücksichtigung des unterschiedlichen Parkverhaltens im Tages- und Wochenverlauf gewährleisten zu können, werden diese Informationen im zeitlichen Kontext benötigt. Zusätzlich müssen Informationen über die gleichzeitigen Ein- und Ausfahrten im zeitlichen Verlauf bereitgestellt werden.

##### **Informationsangebotsanalyse**

Es liegen jeweils unabhängige Informationen über die Ein- und Ausfahrten im zeitlichen Verlauf aus einem konventionellen Parksystems in ähnlicher Standortlage vor. Aufgrund der fehlenden Identifizierung einzelner Fahrzeuge beinhalten diese Informationen keinerlei Angaben über die Standzeiten einzelner Fahrzeuge sondern ausschließlich quantitative Angaben bezüglich der Parkhausbelegung.

Informationen zum Stellflächenbedarf und dem Verkehrsaufkommen sowohl an dem gewählten Standort des automatischen Parksystems als auch an dem des konventionellen Systems müssen durch Primärerhebungen erzeugt oder durch Sekundärerhebungen an anderen vorhandenen Informationsquellen bereitgestellt werden.

## **Nutzbarkeitsprüfung – Informationsqualität**

Die Ein- und Ausfahrten des konventionellen Parksystems können als Einzeldaten eines jeden Fahrzeugs mit genauer Zeitangabe exakt erhoben werden.

Informationen zum Verkehrsaufkommen in Innenstädten sind, obwohl sekundär zu erhebende Information, exakt und einfach abfragbar, da diese im Zuge stadtplanerischer Tätigkeiten in der Regel in guter Qualität vorliegen. Der Stellplatzbedarf am neuen Standort muss den Gegebenheiten angepasst und prognostiziert werden, die Anzahl der Stellplätze des konventionellen Parksystems liegen exakt vor.

### **3.4 Erhebungsvorbereitung, Erhebung und Datenerfassung**

In dem dem geplanten Standort benachbarten, konventionellen Parksystem werden die von dem Schrankensystem gespeicherten Einfahrt- und Ausfahrtzeitpunkte digital ausgelesen und in einer Datenbank direkt weiterverarbeitet. Für die Verwendung der Daten zur Dimensionierung eines konventionellen Parksystems können diese beispielsweise in Stundenintervallen zusammengefasst werden. Die Sekundärdaten zum Verkehrsaufkommen und den Stellplatzbedarfen müssen erfasst und in die Datenbank eingespeist werden.

### **3.5 Datenanalyse**

Mit Hilfe der gespeicherten Einfahrt- und Ausfahrtzeitpunkte der Fahrzeuge über die Zeit hinweg ist die Schätzung von Ankunfts- und Abfahrtsverteilungen für die Simulation direkt möglich. Aus diesen Daten kann zusätzlich auch das Parkverhalten bestimmt werden. Insbesondere ist die Bestimmung einer Verteilung der Parkdauer oder einer Verteilung über die Anzahl der Doppelspiele des Regalbediengerätes möglich. Bei der Übertragung der Daten aus dem konventionellen Parksystems auf die Planung des automatischen Systems sind allerdings die zusätzlichen Annahmen zum Verkehrsaufkommen und den Stellplatzbedarfen zu treffen. Mit diesen Annahmen ist dann eine systembezogene Prognose machbar. Potenzielle Fehlerquellen hierbei liegen in der Formulierung falscher Annahmen, in der Vernachlässigung bzw. dem Nichterkennen relevanter Information (Auswirkungen des Parkdauerhaltens auf die Umschlagleistung des automatischen Förderers) bzw. dem Ableiten falscher Schlussfolgerungen aus den getroffenen Annahmen (Korrekte Berechnung einer Verteilung der Doppelspiele aus den vorhandenen Daten).

### **3.6 Validierung und Visualisierung der Ergebnisse**

Die Bestimmung der Ein- und Ausfahrtverteilung im zeitlichen Kontext und damit die Lastverteilung für das Regalbediengerät ergibt sich direkt aus den vorliegenden Daten. Darauf aufbauend kann eine systembezogene Prognose für den vorgegebenen neuen Standort abgeleitet werden. Problematischer ist dagegen die Schätzung einer Verteilung für die Anzahl der Doppelspiele allein auf der Basis des Verhältnisses der Ein- und Auslagerungen der Fahrzeuge im automatischen Parksystem (vgl. [WBJ01]). Die Annahme, dass die Umschlagleistung linear von dem Verhältnis der Ein- und Auslagerungen abhängt (Abbildung 14a) ist keinesfalls haltbar, denn die über den zeitlichen Verlauf hinweg unterschiedlichen Wahrscheinlichkeiten für die Gleichzeitigkeit der Einfahr- und Ausfahrereignisse muss bei der Ermittlung der Doppelspiele berücksichtigt werden. Bereits die Annahme einer Gleichverteilung potenziell gleichzeitiger Ereignisse in einem Zeitintervall ergibt dahingehend ein substantiell verschiedenes Resultat, dass die Anzahl der Doppelspiele in der Regel sehr viel geringer ausfällt (Abbildung 14b). Noch deutlicher fällt die Veränderung aus, wenn eine an das Verhältnis der Einfahr- und Ausfahrereignisse angepasste Verteilung über die potenziell gleichzeitigen Ereignisse aus den vorliegenden Daten geschätzt wird (Abbildung 14c). Das zu wählende Zeitintervall ist dabei abhängig von der Dauer des Umsetzvorgangs im Parksystem und der Verteilung der Ankunfts- und Abfahrtszeiten der Fahrzeuge in dem zu betrachtenden Gesamtintervall.

Für jede systembezogene Prognose zur Kapazitätsauslastung des automatischen Regalbediengeräts in diesem Kontext müssen demnach neben den Einfahr- und Ausfahrverteilungen insbesondere die Annahmen über die Anzahl der Doppelspiele einfließen.

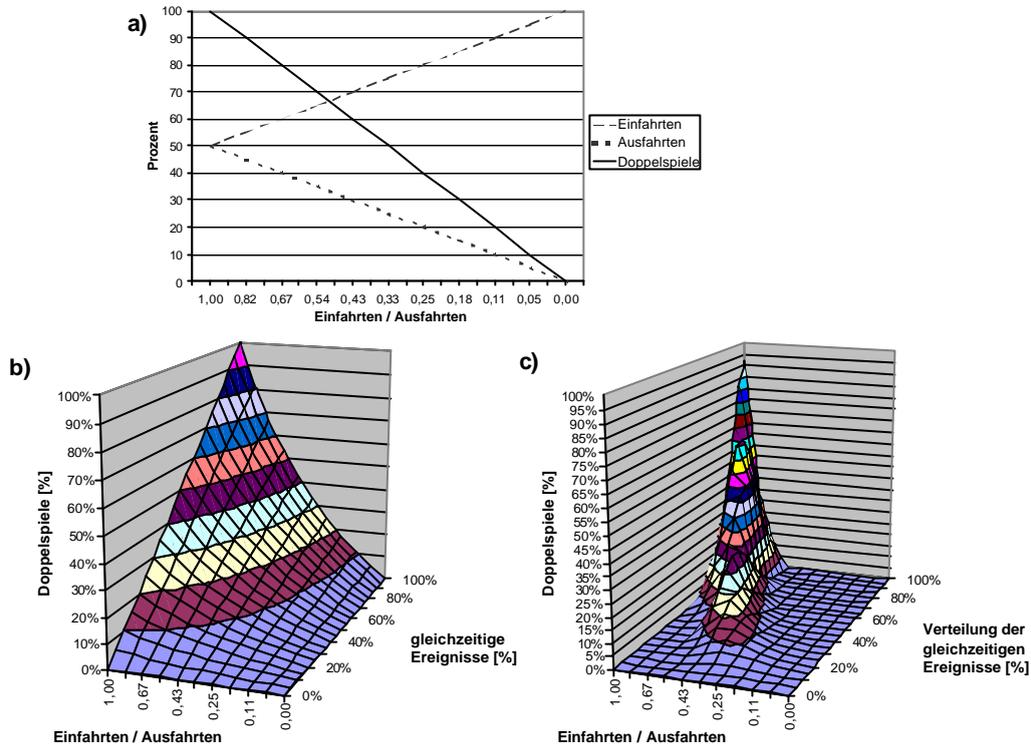


Abbildung 14: Prozentualer Anteil Doppelspiele in Abhängigkeit von dem Verhältnis der Ein- und Ausfahrten, mit zusätzlicher Berücksichtigung der potenziell gleichzeitigen Ereignisse, mit Verteilungsannahmen für die potenzielle Gleichzeitigkeit der Ereignisse

## 4 Ausblick

Insbesondere der Einsatz des entwickelten Vorgehensmodells zur Informationsgewinnung innerhalb beider Anwendungsfälle hat gezeigt, dass dieses Modell durch die strukturierten Vorgaben und die konsequente Aufgabenausrichtung hervorragend anwendbar und zielführend ist. Die Auswahl und Nutzung der unterschiedlichen, aufgabenbezogenen Methoden führte zu experimenteller Bestätigung erster Regeln bzgl. der Anwendbarkeit der Methoden aus den Bereichen Statistik und Visualisierung, so dass die bis zu diesem Zeitpunkt entwickelten Klassifikationskriterien und damit die Methodenklassifikationen anwendungsorientiert erweitert und validiert wurden (vgl. [WBJ03], [BDW03], [Kuh03]).

Weiterhin hat die konsequente Verknüpfung der Methoden aus den Bereichen Statistik und Visualisierung in beiden Beispielen die Validierung und Bewertung der Resultate aus der Datenanalyse vereinfacht oder sogar erst ermöglicht. Damit ist experimentell nachgewiesen, dass nicht nur die zielführende Anwendung aufgabenbezogener Methoden innerhalb der Prozessschritte der Informationsgewinnung von herausragender Bedeutung ist, sondern auch die integrative Nutzung dieser Methoden.

Ein weiterer wichtiger Aspekt dieser Arbeit ist die Erkenntnis, dass durch das Treffen falscher Annahmen bei der statistischen Datenanalyse oder das Nichtberücksichtigen von statistisch relevanten Faktoren bei der endgültigen Ermittlung der zu verwendenden Eingangsdaten für die Modellierung von GNL gravierende Fehler auftreten können. Aufgrund des maßgeblichen Einflusses der Eingangsdaten auf die Ergebnisse einer modellgestützten Analyse beinhalten letztlich auch diese Ergebnisse keinerlei Aussagekraft. Dies führt zu der Schlussfolgerung, dass sowohl die Qualität der Informationsquellen als auch die Informationsgüte der nutzbaren Eingangsdaten bestimmt werden muss, um damit zu Aussagen bzgl. Verlässlichkeit und Güte der Analyseergebnisse zu gelangen.

Insgesamt zeigt die Nutzung der Methoden der Statistik und Visualisierung anhand des Vorgehensmodells für die Informationsgewinnung in der konkreten Anwendung die Notwendigkeit eines integrativen Ansatzes zur zielgerichteten Informationsgewinnung für die Modellierung und Analyse großer Netze der Logistik auf.

## 5 Literatur

- [BDW03] Bernhard, J.; Dragan, M.; Wenzel, S.: Klassifikation von Visualisierungsverfahren für GNL. Technical Report – Sonderforschungsbereich 559 „Modellierung großer Netze in der Logistik“ 03005, 2003, ISSN 1612-1376.
- [KR090] Kaufman, L.; Rousseeuw, P.J.: Finding Groups in Data: An Introduction to Cluster Analysis, Wiley, New York, 1990.
- [Kuh03] Kuhls, S.: Robuste Clusteranalyseverfahren zur Klassifikation von Flugzeugtypen gemäß transportiertem Frachtgewicht. Diplomarbeit, Universität Dortmund, Fachbereich Statistik, Dortmund, 2003.
- [RDr99] Rousseeuw, P.J.; Driessen, K. van: A Fast Algorithm for the Minimum Covariance Determinant Estimator, *Technometrics* 41, 1999, S. 212-223.
- [WBJ01] Wenzel, S.; Becker, C.; Jodin, D.: Methoden der Informationsgewinnung zur Bestimmung des Eingangsdatenraums für Simulationsmodelle in der Logistik. In: K. Panreck, F. Dörrscheidt (Hrsg.): *Simulationstechnik. Tagungsband zum 15. Symposium ASIM 2001*, Paderborn, 11.-14. September, 2001. Reihe: Fortschrittsberichte Simulation Frontiers in Simulation, SCS-Europe BVBA, Ghent, 2001, S. 417-422.
- [WBJ03] Wenzel, S.; Bernhard, J.; Jessen, U.: A Taxonomy of Visualization Techniques for Simulation in Production and Logistics. In: Chick, S.; Sanchez, P. J.; Ferrin, D.; Morrice, D. J. (Hrsg.): *Proceedings of the 2003 Winter Simulation Conference*. New Orleans, Louisiana, USA, 2003, S. 729-736.