

# Analyse der Robustheit von Flugtrajektorien gegenüber Unsicherheiten in der klimaoptimierten Flugplanung

IB-M28-2022-09 (TUHH)  
DLR-IB-LY-HA-2022-116 (DLR)

**Masterarbeit**  
im Fachgebiet *Lufttransportsysteme*  
Studiengang *Theoretischer Maschinenbau*  
der  
Technischen Universität Hamburg  
Institut für Lufttransportsysteme

**Sabrina Thiessen**  
Matrikel- Nr.: 21599110

Erstprüfer:	Prof. Dr.-Ing. Volker Gollnick
Zweitprüfer:	Dr.-Ing. Florian Linke
Betreuer:	Dr.-Ing. Florian Linke

Bearbeitungszeitraum: 25.03.2022 - 26.09.2022

Hamburg, 23. September 2022

---



Die selbständige und eigenhändige Anfertigung  
versichere ich an Eides statt.  
Hamburg, den 23. September 2022

---

Unterschrift



**Aufgabenstellung für  
Sabrina Thiessen**

Matrikel Nr. 21599110

**Analyse der Robustheit von Flugzeugtrajektorien gegenüber  
Unsicherheiten in der klimaoptimierten Flugplanung**

**Robustness analysis of aircraft trajectories with respect to uncertainties  
in climate-optimal flight planning**

**Masterarbeit**

**Einleitung**

Vor dem Hintergrund des Klimawandels und des zunehmenden Bewusstseins der Nachhaltigkeit in der Gesellschaft wächst auch der Druck auf den Luftverkehr, deutlich umweltverträglicher zu werden. Interessensverbände haben hierzu entsprechende Ziele formuliert, die mit Hilfe von technologischen, operationellen und umweltpolitischen Maßnahmen erreicht werden sollen. In Zusammenarbeit mit dem DLR forscht das Institut für Lufttransportsysteme an derartigen Mitigationsmaßnahmen.

Ein vielversprechender operationeller Ansatz, der auf eine Verringerung der Klimawirkung des Luftverkehrs abzielt, ist die Planung klimaoptimierter Trajektorien im Zuge der Flugplanung. Eine solche Planung soll gleichermaßen die CO<sub>2</sub>-Emissionen eines Fluges wie auch die so genannten, z.B. von NO<sub>x</sub>-Emissionen oder Kondensstreifen verursachten Nicht-CO<sub>2</sub>-Effekte adressieren und die zu erwartenden Wetterbedingungen entlang des Fluges berücksichtigen. Dabei bestehen sowohl im Hinblick auf die Wettervorhersage als auch bezüglich der Genauigkeit, mit der bestimmte Klimaeffekte bestimmt werden können, Unsicherheiten, die bei der Flugplanung derart mit einzubeziehen sind, dass möglichst robuste Flugpläne erzeugt werden können. Mithilfe geeigneter Metriken können Trajektorien miteinander verglichen werden und so die Änderungen der resultierenden Flugpläne bei sich ändernden Randbedingungen in der Optimierung in Form von Sensitivitäten sichtbar gemacht werden.

**Aufgabe**

Die vorliegende Masterarbeit soll ein Verfahren zur Analyse der Robustheit von Flugzeugtrajektorien gegenüber Unsicherheiten in der klimaoptimierten Flugplanung entwickeln und dieses auf Datensätze optimierter Trajektorien anwenden. Diese Datensätze stammen z.B. aus den europäischen Forschungsprojekten ATM4E (ATM for environment) und FlyATM4E (Flying ATM for environment), die die prinzipielle Machbarkeit der klimaoptimierten Flugplanung und deren Potenziale untersuchen. In dem vom DLR geleiteten und nunmehr abgeschlossenen Projekt ATM4E wurde die Entwicklung so genannter algorithmischer Klimawirkungsfunktionen vorangetrieben.

Diese können von Trajektorienoptimierern, wie sie in der Flugplanung Verwendung finden, in der Zielfunktion derart genutzt werden, dass ein Kostenfunktional, welches sich aus Klimawirkung

und Betriebskosten (u.a. beeinflusst durch Kraftstoffbedarf und Flugzeit) zusammensetzt, minimiert werden kann. Durch eine Änderung der Gewichtung beider Kostenanteile können verschiedene Optimierungsstrategien (von kostenminimal bis klimaoptimal) getestet werden. Die Variabilität der sich dabei ergebenden Trajektorien kann als Maß für die Robustheit der Planung gegenüber Änderungen in der Strategie verstanden werden und mittels geeigneter Ähnlichkeitsmetriken analysiert werden. In ähnlicher Weise lässt sich die Variabilität in der Flugplanung auch bei sich ändernden Wetterbedingungen untersuchen, wodurch sich Aussagen zur Robustheit gegenüber Unsicherheiten in der Wettervorhersage generieren lassen.

Auf Basis vorangegangener Arbeiten zur Entwicklung von Metriken für den Vergleich der Ähnlichkeit von Trajektorien, soll nun ein Verfahren zur Anwendung entwickelt werden. Hierzu sind im Einzelnen die folgenden Arbeitsschritte durchzuführen:


### Arbeitsschritte

- Einarbeitung in die klimaoptimierte Flugplanung, z.B. durch Hintergrundrecherche und Lektüre einschlägiger Publikationen und Berichte (aus o.g. Projekten)
- Sichtung und Gegenüberstellung verschiedener Ähnlichkeitsmetriken und Bewertung deren Eignung für den vorliegenden Anwendungsfall
- Auswahl und Weiterentwicklung einer (oder mehrerer) geeigneter Metriken, ggf. durch Verknüpfung verschiedener Teilmetriken, und Implementierung von effizienten Berechnungsalgorithmen
- Aufbereitung der bereitgestellten Datensätze
- Erarbeitung eines intuitiven Visualisierungsansatzes zur Darstellung der Ähnlichkeit von Metriken und Demonstration anhand von realen oder konstruierten Beispieltrajektorien
- Entwicklung des Verfahrens zur Analyse der Robustheit von Trajektorien durch Auswertung der Ähnlichkeitsmetriken für die bereitgestellten Daten
- Implementierung des Verfahrens in MATLAB
- Anwendung des Verfahrens auf den gesamten Flugplan sowie statistische Auswertung und Aufbereitung der Ergebnisse, inkl. Visualisierung
- Umfassende und strukturierte Dokumentation

Die Arbeit soll, sofern möglich, von vertraulichen Inhalten frei sein, so dass eine Veröffentlichung nach Abgabe nicht behindert wird.

Ausgabedatum: 25.3.22

Abgabedatum: 26.9.22



Prof. Dr.-Ing. V. Gollnick



Externer Betreuer (DLR)



Studentin

# Inhaltsverzeichnis

<b>Aufgabenstellung</b>	<b>V</b>
<b>1 Einleitung</b>	<b>5</b>
<b>2 Grundlagen</b>	<b>7</b>
2.1 Hintergründe zur Klimaoptimierten Planung . . . . .	7
2.1.1 Grundlegende Modelle und Annahmen . . . . .	10
2.1.2 Air Traffic Management for Environment (ATM4E) . . . . .	13
2.1.3 Fly Air Traffic Management for Environment (FlyATM4E) . . . . .	15
2.2 Ähnlichkeitsvergleiche . . . . .	16
2.2.1 Metriken allgemein . . . . .	17
2.2.2 Euklidische Metrik . . . . .	18
2.2.3 Hausdorff Metrik . . . . .	19
2.2.4 Fréchet Metrik/ „Dog-man“ Metrik . . . . .	19
2.2.5 Editierdistanz / Levenshtein Metrik . . . . .	20
2.3 Energieminimierungsverfahren . . . . .	21
2.4 Vergleichsgröße . . . . .	23
2.5 Clusteranalyse . . . . .	23
2.5.1 Harte und weiche Cluster . . . . .	24
2.5.2 Löschoptionen . . . . .	26
2.5.3 Anzahl von Clustern . . . . .	26
2.5.4 Hierarchische Cluster . . . . .	29
2.5.5 Variierendes Mittel . . . . .	30
2.5.6 Form der Cluster . . . . .	30
2.6 Definition von Robustheit . . . . .	30

<b>3</b>	<b>Robustheitsanalyse</b>	<b>35</b>
3.1	Aufbereitung der Datensätze . . . . .	36
3.2	Gewichtung der Kosten- und Umweltparameter . . . . .	37
3.3	Funktion der MATLAB-Skripte für die Berechnung von Ähnlichkeit . .	38
3.4	Verbesserung der ähnlichkeitsberechnenden MATLAB-Skripte . . . . .	44
3.4.1	Vektorisierung . . . . .	44
3.4.2	Auswertung mehrerer Trajektorien . . . . .	45
3.4.3	Effizientes Laden der Daten . . . . .	45
3.4.4	Berechnung durch Energieminimierung . . . . .	46
3.4.5	Berechnung durch Metriken . . . . .	47
3.4.6	Darstellung der Trajektorien . . . . .	47
3.4.7	Interpolation der Trajektorien . . . . .	48
3.4.8	Nicht verwendete MATLAB-Funktionen . . . . .	52
3.5	Konstruierte Beispiele für Ähnlichkeit . . . . .	53
3.5.1	Diskrete und kontinuierliche Verfahren . . . . .	53
3.5.2	Vergleichbarkeit . . . . .	56
3.5.3	Einfluss der Datensatzlängen - gleiche Länge . . . . .	56
3.5.4	Auflösung . . . . .	58
3.5.5	Trajektorien mit Zeitversatz . . . . .	59
3.5.6	Gegenläufige Trajektorien . . . . .	60
3.5.7	Versetzte Datenpunkte . . . . .	62
3.5.8	Filternde Eigenschaften . . . . .	62
3.5.9	Verhalten bei Instabilitäten . . . . .	63
3.5.10	Vorteile gegenüber anderen Methoden . . . . .	64
3.5.11	Bezugsgröße . . . . .	66
3.6	Zusammenfassung der Ähnlichkeitsberechnung . . . . .	67
3.7	Anpassen an einen Zeichensatz . . . . .	68
3.7.1	Funktionen der MATLAB-Skripte . . . . .	68
3.7.2	Auftretende Probleme während der Berechnung . . . . .	69
3.8	Konstruierte Beispiele für Zeichensätze . . . . .	70
3.8.1	Abhängigkeit von Zeichensatzwahl . . . . .	70
3.8.2	Zeichensatziteration . . . . .	71



3.8.3	Zeichengröße . . . . .	73
3.8.4	Toleranz beim Anpassen . . . . .	74
3.8.5	Anpassungsverfahren . . . . .	75
3.8.6	Verschiedene Zeichenkettenlängen . . . . .	77
3.8.7	Gestalt des Zeichens . . . . .	77
3.8.8	Konvergenzproblem . . . . .	78
3.9	Clusteranalyse . . . . .	80
3.9.1	Harte oder weiche Cluster . . . . .	80
3.9.2	Anzahl der Cluster . . . . .	81
3.9.3	Löschen von Daten . . . . .	83
3.9.4	Variierendes Mittel . . . . .	84
3.9.5	Form . . . . .	84
3.9.6	Hierarchische Aufteilung . . . . .	85
3.9.7	Zusammenfassung . . . . .	85
3.10	Konstruierte Beispiele für die Clusteranalyse . . . . .	85
3.10.1	Vergleichbarkeit . . . . .	85
3.10.2	Relative Distanzen . . . . .	86
3.10.3	Geordnete Daten . . . . .	87
3.11	Umsetzung der Clusteranalyse in MATLAB . . . . .	88
<b>4</b>	<b>Ergebnisse</b>	<b>91</b>
4.1	Ergebnisse mit ATM4E . . . . .	91
4.1.1	Plots der Pareto-Fronten . . . . .	91
4.1.2	Ergebnisse der Clusteranalyse . . . . .	94
4.1.3	Ergebnisse der Ähnlichkeitsberechnung . . . . .	104
4.1.4	Zusammenfassung für ATM4E . . . . .	109
4.2	Ergebnisse mit FlyATM4E . . . . .	110
4.2.1	Ergebnisse der Clusteranalyse . . . . .	110
4.2.2	Ergebnisse der Ähnlichkeitsberechnung . . . . .	112
4.2.3	Plots der Pareto-Fronten . . . . .	113
4.2.4	Zusammenfassung für FlyATM4E . . . . .	114
<b>5</b>	<b>Zusammenfassung und Ausblick</b>	<b>115</b>

<b>Literaturverzeichnis</b>	<b>117</b>
<b>Abbildungsverzeichnis</b>	<b>123</b>
<b>A Anhang</b>	<b>127</b>

# 1. Einleitung

Der Druck auf den Luftverkehr, umweltfreundlicher zu werden, nimmt zu, da sich durch den Klimawandel das Nachhaltigkeitsbewusstsein in der Bevölkerung vergrößert. Auch Interessenverbände wie zum Beispiel die International Air Transport Association (IATA) haben Maßnahmen zusammengefasst, wie der zivile Flugverkehr trotz steigender Auslastung umweltverträglicher und schließlich klimaneutral werden soll. Neben Treibstoffalternativen zu Kerosin, elektrischen Antrieben und der Filterung von  $\text{CO}_2$  aus den Abgasen direkt nach dem Verbrennungsprozess ist die Planung emissionsarmer Flugrouten eine zentrale Maßnahme. Das Deutsche Zentrum für Luft- und Raumfahrt (DLR) forscht zusammen mit dem Institut für Lufttransportsysteme (ILT) der Technischen Universität Hamburg (TUHH) an Methoden, die Planung von Flugrouten zu optimieren. Eine optimale Planung berücksichtigt sowohl  $\text{CO}_2$ -Emissionen als auch Nicht- $\text{CO}_2$ -Emissionen wie  $\text{NO}_x$ - und  $\text{H}_2\text{O}$ -Emissionen, Kondensstreifen- und Wolkenbildung, den Ausstoß von Aerosolen und viele weitere Faktoren. Dabei beeinflusst das Wetter die Stärke der Auswirkungen dieser Faktoren. In bestimmten Regionen und Luftschichten werden zum Beispiel Bildung von Ozon durch  $\text{NO}_x$  begünstigt. Da dies neben weitere Effekten in großer Menge und langfristig zur Klimaerwärmung führt, gilt es, diese Regionen zu vermeiden. Andererseits gibt es Regionen, in denen sich Kondensstreifen bilden, die das Sonnenlicht tagsüber reflektieren. Das Durchfliegen solcher Regionen kann trotz zusätzlichen Treibstoffverbrauchs eine in Summe kühlende Wirkung auf das Klima haben. Allgemein weisen Kondensstreifen aber eine wärmende Wirkung auf und sind in der Regel zu vermeiden. Bei der Wettervorhersage bestehen aber Unsicherheiten bezüglich der Genauigkeit, mit der bestimmte Effekte und deren Auswirkungen vorausgesagt werden können. Auch die Bereiche, in denen bestimmte Effekte auftreten, unterliegen einer Unsicherheit. Daher ist eine robuste Flugplanung ein wichtiges Ziel, in der solche Unsicherheiten kaum eine Rolle spielen.

Im Rahmen der von der Europäischen Union geförderten Projekte Air Traffic Management for Environment (ATM4E) und Fly Air Traffic Management for Environment (FlyATM4E) wurden optimale Routen berechnet. Bei ATM4E wird eine einzige, real aufgetretene und in der Vergangenheit vorgekommene Wetterlage betrachtet. Zu verschiedenen Start- und Zielflughäfen wurden jeweils 100 optimale Trajektorien berechnet. Die 100 verschiedenen und gleichzeitig optimalen Trajektorienverläufe ergeben sich aus unterschiedlich starker Gewichtung der Umweltverträglichkeit gegenüber den Flugkosten wie zum Beispiel Treibstoffverbrauch und Flugdauer. Bei FlyATM4E wurden verschiedene in der Vergangenheit aufgetretene Wetterlagen für dieselben Flughäfen betrachtet. Auch hier wurden Umweltverträglichkeit und Kosten gewichtet, sodass 100 Routen je Verbindung bei unterschiedlichen Wetterlagen berechnet wurden.

Im Rahmen dieser Arbeit wird untersucht, wie robust die in den jeweiligen Projekten berechneten Routen sind. Robustheit meint hier Nähe und Ähnlichkeit zu weiteren Trajektorien, die durch eine abweichende Gewichtung oder verschiedene Wetterensembles entstanden sind. Zur Feststellung der Trajektorienrobustheit werden verschiedene Konzepte wie die in der Projektarbeit [62] erläuterten Ähnlichkeitsberechnungen für Flugtrajektorien verwendet. Da es sich in der Projektarbeit [62] um eine prototypische Ausarbeitung handelt, wird das Konzept und die Umsetzung in dieser Arbeit erweitert, verbessert und schließlich zur Robustheitsberechnung angewandt. Zudem stehen mehrere Verfahren zur Berechnung von Ähnlichkeit zur Auswahl. Es wird analysiert, welche Verfahren sich bezüglich Robustheit am besten eignen und eine entsprechende Auswahl getroffen. Auch Clusteranalysen zur Bestimmung geometrischer Nähe der Trajektorien werden zur Feststellung der Robustheit herangezogen. Die verschiedenen Konzepte werden zur Quantifizierung in einer Zahl zusammengefasst, die die Robustheit der jeweiligen Trajektorie bemisst. Langfristiges Ziel ist es, die optimierten Trajektorien in die real vorgenommene Flugplanung zu integrieren. Leider kann nicht einfach eine Trajektorie aus ATM4E ausgewählt werden, da das Wetter für dieses Projekt als unveränderlich angenommen wurde. Das Wetter ist aber nur eine Vorhersage und weist gewisse Unsicherheiten auf. Daher wird die in dieser Arbeit berechnete Robustheit der Trajektorien benötigt um zu belegen, welche Trajektorien trotz Unsicherheiten und Schwankungen weiterhin einen kaum abweichenden Verlauf und damit verbunden identische Auswirkungen aufweisen werden.

## 2. Grundlagen

Ziel dieser Arbeit ist es, die Robustheit von Flugtrajektorien zu bestimmen. Die Trajektoriendaten, die für diese Arbeit verwendet wurden, sind Ergebnisse der Projekte Air Traffic Management for Environment (ATM4E) und Fly Air Traffic Management for Environment (FlyATM4E). In Abschnitt 2.1 werden daher zunächst allgemeine Hintergründe zu Projekten mit dem Ziel der Flugtrajektorienoptimierung erläutert und in 2.1.2 und 2.1.3 speziell auf die beiden Projekte ATM4E und FlyATM4E eingegangen.

Um Ähnlichkeit und schließlich Robustheit von Trajektorien zu berechnen, werden drei Faktoren benötigt: Metrik, Methode und Vergleichsgröße. Eine detaillierte Ausführung zu Ähnlichkeitsvergleichen und den drei Faktoren ist in [62] zu finden. Für die Verarbeitung der gegebenen Trajektorien der Projekte ATM4E und FlyATM4E werden verschiedene Metriken und Methoden benötigt sowie Vergleichsgrößen festgelegt. Die Begründung der Auswahl ist ebenfalls in [62] zu finden. In dieser Arbeit wird lediglich auf die verwendbaren, erfolgversprechenden Größen eingegangen. Mathematische Definitionen und grundlegende Eigenschaften der betrachteten Metriken sind in Abschnitt 2.2 aufgeführt.

In einer dieser Arbeit vorangegangenen Projektarbeit wurde die Methode der Energieminimierung für den Vergleich von Trajektorien als passend eingestuft [62]. Auf das ausgewählte Verfahren wird in Abschnitt 2.3 eingegangen. In Abschnitt 2.4 wird die Auswahl einer Vergleichsgröße erläutert. Abschnitt 2.5 beschäftigt sich mit den Grundlagen und Eigenschaften von Clustern. Hierbei handelt es sich um eine andere Teilberechnung der Robustheit von Trajektorien. Ähnlichkeit und die durch Clusteranalysen bestimmte räumliche Nähe von Trajektorien ergeben zusammengefasst die Robustheit einer Trajektorie. Geometrische Nähe ist für die robuste Flugplanung eine entscheidende Größe. Mit Hilfe der Clusteranalyse werden die gegebenen Trajektorien in Gruppen geteilt, die nach festgelegten Parametern als nahe beieinander gelten. Da das Ziel dieser Arbeit eine Robustheitsanalyse ist, es aber unterschiedliche Definitionen für Robustheit gibt, wird in Abschnitt 2.6 geklärt, was in dieser Arbeit unter dem Begriff der Robustheit verstanden wird.

### 2.1 Hintergründe zur Klimaaptimierten Planung

$\text{NO}_x$ -Emissionen haben für sich genommen keine Auswirkung auf das Klima. Sie reagieren erst unter Einfluss von Sonnenlicht zu Radikalen, die die Bildung von Ozon ( $\text{O}_3$ ) fördern, was einen wärmenden Effekt auf das Klima hat. Ein positiver Nebeneffekt von  $\text{NO}_x$ -Emissionen ist die Oxidation von Methan  $\text{CH}_4$ . Methan wirkt als Treibhausgas. Da die  $\text{NO}_x$ -Emissionen mit dem Methan reagieren, wird der Anteil

von Methan in der Atmosphäre reduziert und trägt somit weniger zur Klimaerwärmung bei. Die negativen Aspekte der  $\text{NO}_x$ -Emissionen durch die begünstigte Bildung von Ozon überwiegen jedoch deutlich [40].

Durch Abgase entstandene Kondensstreifen haben in sehr kalter Umgebung eine reflektierende Wirkung. Dabei wird sowohl kurzweilige Sonnenstrahlung reflektiert als auch terrestrische Strahlung. Die Reflexion des Sonnenlichtes hat eine kühlende Wirkung, die Reflexion der terrestrischen Strahlung eine wärmende. Dabei gibt es je nach Tag, Wetterlage und Höhe Kondensstreifen mit in Summe erdkühlender oder erderwärmender Wirkung [43]. Erdkühlende Kondensstreifen sind wünschenswert, während Kondensstreifen mit wärmender Wirkung zu vermeiden sind. Die Voraussage solcher Kondensstreifen ist sehr schwierig und ändert sich abhängig von der Tageszeit stark [40]. Da die Kondensstreifen im Mittel eine wärmende Wirkung haben [4], gilt es, Kondensstreifen in der Regel zu verhindern. Besonders nachts besitzen die Kondensstreifen eine wärmende Wirkung, die es zu vermeiden gilt.

Aerosole dienen als Keime für Wolken und Kondensstreifen. Indem sich Anlagerungen bilden, begünstigen Aerosole die Wolkenbildung [40]. Die Wolken wirken ähnlich wie die Kondensstreifen reflektierend. Einige Reflexionen sind wünschenswert, im Durchschnitt wirken Wolken aber ebenfalls erderwärmend und sind daher eher zu vermeiden.

$\text{CO}_2$ -Emissionen korrelieren größtenteils mit dem Treibstoffverbrauch. Durch einen geringeren  $\text{CO}_2$ -Ausstoß wird die klimaerwärmende Wirkung eines Fluges verringert [32].  $\text{CO}_2$  wirkt als Treibhausgas. Es reagiert in der Atmosphäre nur sehr wenig mit anderen Komponenten und bleibt daher sehr lange in der Atmosphäre bestehen. All diese Faktoren berücksichtigend, beträgt der Anteil der zivilen Luftfahrt an klimaverändernden anthropogen verursachten Einflüssen etwa 3,5% [38]. Um den Einfluss auf das Klima zu verringern, wurden weltweit viele Projekte ins Leben gerufen [28].

Zur IATA (International Air Transport Association) gehören 290 Fluggesellschaften, die zusammen 83% des globalen zivilen Luftverkehrs kontrollieren. Damit ist die IATA die größte Vereinigung von Fluggesellschaften. Im Oktober 2021 wurde sich beim 77. *IATA Annual General Meeting* in Boston darauf geeinigt, zu versuchen, den Flugverkehr bis 2050 emissionsneutral zu gestalten [29]. Mitglieder der Assoziation haben sich entsprechend ihrer Größe an zielgerichteten Projekten zu beteiligen. Um die Emissionsneutralität bis 2050 zu erreichen, wird bezüglich verschiedener Teilaspekte geforscht [28]. Die jeweiligen Themengebiete sind:

- **Alternative Quellen für Treibstoff, auf Englisch „Sustainable Aviation Fuel“ (SAF):**

Bei dem bisher verwendeten Kerosin handelt es sich um keinen nachwachsenden Rohstoff. Zudem ist ein Produkt bei der Verbrennung von Kerosin  $\text{CO}_2$ , das den Klimawandel begünstigt und das es zu vermeiden gilt. Bis 2050 sollen Flugzeuge ausschließlich durch nachwachsende Rohstoffe betrieben werden [28]. Da der ausschließliche Einsatz von wasserstoffbasierten und elektrischen Antrieben bis dahin als unrealistisch angesehen wird [25], könnte Kerosin zum Beispiel durch Speiseöl als SAF substituiert werden [12]. Derzeit ist noch unklar, welche Alternativen zu Kerosin kurzfristig am vielversprechendsten sind und wie ein sicherer Einsatz gewährleistet wird [55].

- **Entwicklung neuer Technologie mit Fokus auf elektrischen und wasserstoffbetriebenen Entwicklungen:**

Auch wenn dieses Vorhaben bis 2050 als unrealistisch angesehen wird [28], wird bereits jetzt mit der Forschung begonnen, um entsprechende Grundlagen zu erforschen. Neben der Grundlagenforschung besteht die Hoffnung, die neuen Entwicklungen zeitnah in die Flugzeugsysteme integrieren zu können und so Energie und damit Treibstoff einzusparen [7].

- **Effizientere Infrastrukturen und Routenplanung:**

In Abhängigkeit der Wetterlage ändert sich die Umweltauswirkung eines Fluges stark [14]. Um die Auswirkungen so gering wie möglich zu halten, bedarf es präziser Wettervorhersagen. Außerdem werden Programme benötigt, die aus den Wetterdaten eine optimale Route berechnen. Das Sammeln vieler Daten, genaue Wettervorhersagen über lange Zeiträume zu treffen und Software, die aus den erhobenen Daten schnell und zuverlässig robuste Trajektorien berechnet, sind Teil dieses Forschungsaspektes. Die Projekte ATM4E und FlyAMT4E, auf die in 2.1.2 und 2.1.3 eingegangen wird, befassen sich mit der optimierten Trajektorienplanung. Auch wenn beim 77. *IATA Annual General Meeting* in Boston dieser Punkt als wichtiger Forschungsaspekt beschlossen wurde, ist zu erwähnen, dass die Projekte ATM4E und FlyAMT4E bereits vorher gestartet wurden und unabhängig der Beschlüsse der IATA durchgeführt wurden.

- **Filterung und Einlagerung von CO<sub>2</sub>:**

Bei der Verbrennung von Treibstoff entsteht unter Anderem CO<sub>2</sub> als Abfallprodukt. Deshalb gibt es verschiedene Ansätze zur Vermeidung und Verminderung des Ausstoßes von CO<sub>2</sub>. Nach der Verbrennung könnte das CO<sub>2</sub> getrennt und für die restliche Flugdauer in einem Tank gespeichert werden. Nach der Landung wird sich um die Einlagerung des CO<sub>2</sub> gekümmert [32]. Ein anderer Ansatz beschäftigt sich mit der Filterung von CO<sub>2</sub> aus der Luft [4]. Zusätzliche Vorrichtungen filtern während des Fluges die Luft, die das Flugzeug umströmt. Das herausgefilterte CO<sub>2</sub> wird ebenfalls kurzfristig in einem Tank gelagert und am Boden für eine langfristige Lagerung vorbereitet.

Schon vor den erklärten Ziele der IATA gab es europäische Projekte, an denen das Deutsche Zentrum für Luft- und Raumfahrt (DLR), The University of Reading (UREAD), Envisa SAS (ENVISA), die Manchester Metropolitan University (MMU), die Technische Universität Delft (TUD) und die Technische Universität Hamburg (TUHH) mitgewirkt haben [44]. Seit 1999 existiert ein durch die Europäische Kommission gestartetes Projekt namens Single European Sky Air Traffic Management Research, kurz SESAR 2020 [15] [36]. Primäres Ziel von SESAR 2020 ist die Entwicklung einheitlicher Systeme für den europäischen Luftverkehr. Da das Projekt aber auch den Anspruch hat, den Herausforderungen eines wachsenden Luftverkehrs gerecht zu werden, wurde 2016 das Projekt Air Traffic Management for Environment, kurz ATM4E, gestartet [45]. Das Projekt wurde 2018 abgeschlossen und zwei Jahre später als Nachfolgeprojekt FlyATM4E gestartet, wenn auch mit veränderten Partnern [46]. Neben diesen beiden Projekten gibt es weitere von der EU geförderte Projekte wie zum Beispiel Advancing the Science for Aviation and Climate (ACACIA) [65]. Dieses Projekt beschäftigt sich mit Unsicherheiten bei der Wettervorhersage und der Schwierigkeit, die Umweltauswirkungen eines Fluges genau zu quantifizieren [65] [43]. Eine Vielzahl von Modellen, von denen einige derzeit

weiterentwickelt werden, ist durch viele Projekte entstanden. Diverse dieser Modelle werden von den Projekten ATM4E und FlyATM4E verwendet, weshalb in 2.1.1 aufgelistet und erklärt wird, welche Modelle Anwendung finden. Auch wenn grundlegende Modelle noch verändert werden, sind die Ergebnisse der beiden Projekte trotzdem verwendbar. Auf die Gründe dafür wird ebenfalls in 2.1.1 eingegangen. Die Ergebnisse von ATM4E und FlyATM4E sind Grundlage für diese Arbeit, weshalb in 2.1.2 und in 2.1.3 speziell auf diese beiden Projekte eingegangen wird.

### 2.1.1 Grundlegende Modelle und Annahmen

In den 1960er Jahren wurde mit der Optimierung von Flugrouten begonnen und wird seitdem durch immer leistungstärkere Rechner stetig weiter verbessert [58]. Umweltaspekte werden außer Wetter in der aktuell vorgenommenen Flugplanung jedoch kaum berücksichtigt [47]. Wirtschaftliche Interessen stehen im Vordergrund. Die Optimierung von Flugrouten bezieht sich hauptsächlich auf geringe Kosten, so genannten „Cash Operating Costs“ (COC). Da ein möglichst geringer Treibstoffverbrauch und eine kurze Flugdauer mit geringen COC korrelieren, ist dies ein Umweltaspekt, der indirekt in die Optimierung von Flugrouten einfließt. Ein geringer Treibstoffverbrauch bedeutet entsprechend weniger CO<sub>2</sub>-Emissionen und damit eine geringere Umweltbelastung. Andere Faktoren der so genannten Nicht-CO<sub>2</sub>-Emissionen wie Kondensstreifen- und Wolkenbildung, der Ausstoß von Partikeln oder NO<sub>x</sub>-Emissionen sind ebenfalls von wirtschaftlichem Interesse, da bei zu hoher Emission Strafen bezahlt werden müssen [8] und so die COC erhöhen. Unterhalb dieser Grenzwerte sind die Nicht-CO<sub>2</sub>-Emissionen von geringem Interesse. Alle genannten wirtschaftlichen Faktoren und Umweltaspekte fließen bei der Trajektorienoptimierung im Rahmen der Projekte ATM4E und FlyATM4E ein.

Für die Berechnung einer Trajektorie wird zunächst ein Modell der Realität benötigt. Dazu gehört ein Modell des Flugzeugs und die Lage von Start und Ziel der Flugroute. Für die Abschätzung der Umweltauswirkungen ist außerdem relevant, welche Luftschichten durchflogen werden, wie der Treibstoff verbrannt wird, die anschließende Vermischung mit der Luft und die daraus resultierenden Reaktionen. Für all diese Teilaspekte wurden durch verschiedene Projekte Modelle erstellt beziehungsweise Daten gesammelt. In so genannten Climate Change Functions (CCFs) werden Wetterdaten zusammengefasst [47]. Die CCFs bilden die Grundlage für die Trajektorienberechnung, die bei den Projekten ATM4E, in 2.1.2 aufgegriffen, und FlyATM4E, in 2.1.3 erklärt, vorgenommen wird. Die Modelle im Einzelnen sind:

- *The Modular Earth Submodel System* (MESSy) [33], genauer wurde das Submodell *AirTraf* verwendet [67], das für den Flug relevante Teile von MESSy aufgreift und das Flugzeugmodell berücksichtigt. MESSy stellt das Grundgerüst dar, in das alle anderen Modelle eingearbeitet werden können [34]. Neben der Kombination von Modellen, für die geeignete Schnittstellen benötigt werden, bietet MESSy auch schnelle und effiziente Berechnungsmöglichkeiten [10].
- *EUROCONTROL's Base of Aircraft Data (BADA) 4.2 aircraft performance models*, welche Modelle der Flugzeuge und deren Flugverhalten beinhaltet [47]. Entsprechend des zu betrachtenden Fluges wird ein passendes Flugzeugmodell ausgewählt und in MESSy verwendet.



- *European Center Hamburg general circulation model* ist ein Modell für die Modellierung der verschiedenen Atmosphärenschichten [10]. Verschiedene Atmosphärenschichten weisen unterschiedliche chemische Zusammensetzungen auf. Die Produkte der Treibstoffverbrennung reagieren je nach Zusammensetzung unterschiedlich und haben anderen Auswirkungen auf die Umwelt. Zudem verschieben sich die Schichten zueinander in Abhängigkeit der Wetterlage, was ebenfalls in diesem Modell enthalten ist [56].
- Wetterdaten des *European Centre for Medium-Range Weather Forecasts* [50], die ebenfalls in MESSy eingearbeitet werden. Aus den Wetterdaten können Durchschnittswerte des Wetters berechnet werden, typische Wetterlagen für die Region ausgewählt werden oder das aktuelle Wetter betrachtet werden. Entsprechen ausgewählte Wetterdaten beeinflussen und verändern die anderen Modelle.
- *AUSTAL2000* ist ein Modell zur Simulation der Dispersion von Luftschadstoffen in der Atmosphäre [47]. Neben den unmittelbaren Reaktionen der Verbrennungsprodukte mit der Umgebungsluft können die Verbrennungsprodukte sich auch zunächst nur mit der Luft vermischen und zu einem späteren Zeitpunkt reagieren. Ob eine und welche Reaktion stattfindet, wird durch die Durchmischung und die Umgebungsbedingungen bestimmt und durch *AUSTAL2000* in das übergeordnete MESSy-Modell überführt.
- *ICAO Exhaust Emission Databank* ist eine Datenbank für die Verbrennungsprozesse der Triebwerke [47]. Ergänzend zum *European Center Hamburg general circulation model* kann so modelliert werden, welche Verbrennungsprodukte entstehen und wie diese in den durchflogenen Atmosphärenschichten weiter reagieren werden. Die *ICAO Exhaust Emission Databank* liefert Messdaten real abgeflogener Trajektorien. Mit dieser Datengrundlagen kann sinnvoll der Verbrennungsprozess und dessen Produkte modelliert werden.
- *FOCA Databank for Pistons*: Datensammlung von real gemessenen Abgasausstößen [47]. Genau wie die *ICAO Exhaust Emission Databank* ist die *FOCA Databank for Pistons* eine Sammlung von Abgasdaten. Beide Datenbanken werden zur gegenseitigen Verifikation benutzt.
- *DLR Fuel Flow Method* [14] und *Boeing Fuel Flow Method 2* [47] für die Vermischung der Abgase nach der Verbrennung mit der das Flugzeug umströmenden Luft. Auch hier werden zwei Modelle genutzt, um die gegenseitigen Ergebnisse zu verifizieren.
- *AIRTRAC* und *CONTRAIL*, die die Kondensstreifenbildung modellieren [16]. *CONTRAIL* ist Teil von *AIRTRAC*. Die Modellnutzung beschränkt sich aber nicht nur auf *CONTRAIL*, sondern benutzt auch Aspekte von *AIRTRAC*. In *CONTRAIL* wird modelliert, ob und an welchen Orten Kondensstreifen entstehen sowie deren Auswirkungen auf das Klima bestimmt. Um die Kondensstreifenbildung zuverlässiger berechnen zu können, werden weitere Daten aus *AIRTRAC* herangezogen.
- *EUROCONTROL Open - Airport Local Air Quality Studies tool*: Daten der Luftqualität [47]. Neben der variierenden Zusammensetzung verschiedener Luftschichten gibt es lokale Unterschiede der Luftzusammensetzung. So beinhaltet

die Luft über Ballungsräumen zum Beispiel mehr Feinstaub als die Luft über ländlichen Regionen [64]. Chemische Reaktionen, die durch verbrannte und ausgestoßene Treibstoffe entstehen, können so beeinflusst werden.

- *Aircraft Noise and Performance Database (ANP2.1)*, um den Einflussfaktor des Lärms durch Flugzeuge zu berücksichtigen [47]. Diese Daten sind hauptsächlich aber nicht nur für die nähere Umgebung von Flughäfen verfügbar. Da eine Gesamtoptimierung der Trajektorien vorgenommen werden soll, werden lokale Einflüsse wie die Lärmbelastung an den Flughäfen kaum berücksichtigt. Für die Lärmbelastung während des Fluges sind einige Daten verfügbar, es gilt aber noch einen Großteil solcher Daten zu sammeln [21].

In der Modellannahme wird von fehlerfreien Wetterdaten ausgegangen. Da die verwendeten Daten in der Vergangenheit liegen, ist dies gewährleistet. Ziel ist es aber, optimale Trajektorien, die in der Zukunft liegen, zu berechnen. Es muss daher geprüft werden, wie die optimierten Routen auf Unsicherheiten in der Wettervorhersage reagieren. Diesen Aspekt wird diese Materarbeit untersuchen. Zudem wird angenommen, dass die Route ohne Abweichungen geflogen wird, was in den meisten Fällen ebenfalls gewährleistet ist.

Um eine umweltoptimale Trajektorie zu erhalten, werden verlässliche Wetterdaten benötigt. Hier gibt es drei verschiedene Ansätze, wie Wetterdaten ausgewertet werden und so den Einfluss der Unsicherheiten der aktuellen Vorhersage zu verringern [52] [21]: Über einen längeren Zeitraum werden Durchschnittswerte von Wetterdaten verwendet. Genauso ist es möglich, Wetterlagen bestimmter Tage als typisch für diese Region zu definieren und umweltoptimale Trajektorien für diese Wetterlage auszurechnen. Wünschenswert wäre eine Online-Berechnung, die abhängig von der aktuellen Wetterlage die optimale Flugroute berechnet. ATM4E und FlyATM4E versuchen, ein Umweltoptimal entsprechend der vorhandenen Wetterlage zu berechnen [47]. Zu Testzwecken und Verifikation der Rechnungen wurden aber zunächst gemittelte Daten verwendet, die Ergebnis des Projektes „Reducing emissions from aviation by changing trajectories for the benefit of climate“ (REACT4C) sind [17]. Die meisten der Modelle werden stetig überarbeitet und weiterentwickelt [10]. Zusätzlich werden die Datenbanken durch neue Daten erweitert [68] und alte Daten gegebenenfalls korrigiert und angepasst. Neben neuen wissenschaftlichen Erkenntnissen, die in die Modelle einfließen und Unsicherheiten verändern, ist das aktuelle Wetter eine sich ständig verändernde Größe. Die Wettervorhersage ist nur für einen geringen Zeitraum mit wenigen Unsicherheiten zu treffen. Über einen größeren Zeitraum wächst die Unsicherheit der Vorhersage.

Die oben genannten Modelle werden durch Climate Change Functions (CCFs) zusammengefasst [59]. Der Begriff CCF ersetzt den Begriff „Kostenfunktion“, der sonst für die Gewichtung verschiedener zu erreichender Ziele verwendet wird. Kostenfunktion und Kosten des Fluges können so nicht mehr verwechselt werden. Neue wissenschaftliche Erkenntnisse und Modelle verändern ausschließlich die CCFs, nicht aber den Berechnungsprozess [69]. So werden mit jetzigem Kenntnisstand in den Projekten ATM4E und FlyATM4E aus vorgegebenen CCFs Trajektorien berechnet. Mit anderen CCFs sind auch die Trajektorien verändert. Der Prozess, aus CCFs optimale Trajektorien zu generieren, bleibt aber gleich [42] [22] [54]. Bekannte Unsicherheiten können verschieden interpretiert werden, sodass auch mit dem jetzigen Kenntnisstand Möglichkeiten bestehen, andere CCFs zu generieren und gleichwertig

zu verwenden. Werden andere Metriken für das Klima (zum Beispiel *Global Warming Potential* oder *Average Temperature Response*) verwendet, verändert sich die CCF entsprechend. Beide Metriken unterliegen Unsicherheiten. Laut dem aktuellen Forschungsstand ist aber nicht eindeutig zu sagen, welche der Metriken genauer ist [66]. Abweichungen entstehen außerdem, da Luftqualität, Klimawirkung und Lärm die CCFs beeinflussen. Die Gewichtung der einzelnen Faktoren kann aber verschieden stark erfolgen [69].

### 2.1.2 Air Traffic Management for Environment (ATM4E)

ATM4E ist Teil des Projektes SESAR 2020. 2016 wurde mit dem Projekt begonnen und ATM4E wurde 2018 abgeschlossen [45]. Da verschiedene Untersuchungen [18] [21] [23] [31] [48] [52] zeigten, dass eine erhebliche Verbesserung der Umweltauswirkungen eines Fluges und der Kosten durch optimierte Flugrouten erzielt werden kann, wurden für eine Vielzahl von Paarungen von Start- und Zielflughäfen Sätze von optimalen Trajektorien berechnet. In der bisherigen Flugroutenplanung wird beispielsweise nicht berücksichtigt, dass längere Routen und ein damit höherer Treibstoffverbrauch umweltfreundlicher sein kann, wenn dabei Regionen umflogen werden, in denen sich Kondensstreifen mit erderwärmenden Effekten bilden [50]. Die summierte Klimabilanz des Fluges kann trotz längerer Flugdauer und -strecke verbessert sein. Berücksichtigt wird außerdem die Länge der Auswirkungen. So ist das CO<sub>2</sub> aus den Abgasen Jahrzehnte lang in der Atmosphäre vorhanden und verändert so das Klima. Kondensstreifen hingegen sind für Stunden vorhanden, Ozonveränderungen sind monatelang messbar und chemische Reaktionen mit Methan können bis zu einem Jahr nach einem Flug stattfinden. Die verschiedenen Zeitspannen werden durch unterschiedlich starke Gewichtung der Faktoren in den CCFs berücksichtigt. Außerdem sind die Wirkungen des CO<sub>2</sub> ortsunabhängig, während Nicht-CO<sub>2</sub>-Effekte sich auf bestimmte Regionen beschränken [50] [51]. Auch dieser Einfluss wird berücksichtigt. Zusätzlich ist zu erwähnen, dass die globalen Auswirkungen verbessert werden sollen. Lokal können sich die Umstände trotz optimierter Trajektorie verschlechtern, zum Beispiel durch eine schlechtere Luftqualität und höhere Lärmbelastung in der Nähe von Flughäfen [50].

Die Berechnung der Trajektorien geschieht auf Basis einer CCF. Für das Projekt ATM4E wurde eine unveränderliche Wetterlage eines ausgewählten Tages betrachtet, die typisches Wetter für die durchflogene Region darstellt. Bei den Projekte ATM4E und FlyATM4E werden je eine kosten- und eine umweltoptimale Flugroute zu gegebenem Start und Ziel berechnet. Zudem werden COC und Umwelteinflüsse verschieden stark gewichtet, sodass sich neben den beiden Optimalrouten weitere Routen ergeben, die einen Kompromiss zwischen beiden Optimalen darstellen [47]. Da die Gewichtung von Null bis Eins geht und in 0,01-Schritten gewichtet wird, entstehen in der Regel 101 verschiedene Trajektorien. In einigen Fällen kann es zu Fehlern kommen, da Routen nicht berechenbar sind. Dann besteht ein Datensatz entsprechend aus weniger Trajektorien.

Dabei heißt ein Kompromiss zwischen kosten- und umweltoptimaler Route nicht, dass sich die Trajektorien im Raum zwischen der umweltoptimalen Route und der kostenoptimalen Trajektorie befinden. In Abbildung 2.1 ist ein Trajektoriensatz für den Flug von Gran Canaria, Spanien, nach Trondheim, Norwegen, zu sehen. Dies entspricht Datensatz „00005\_GCLP\_ENVA“ aus dem Projekt ATM4E.

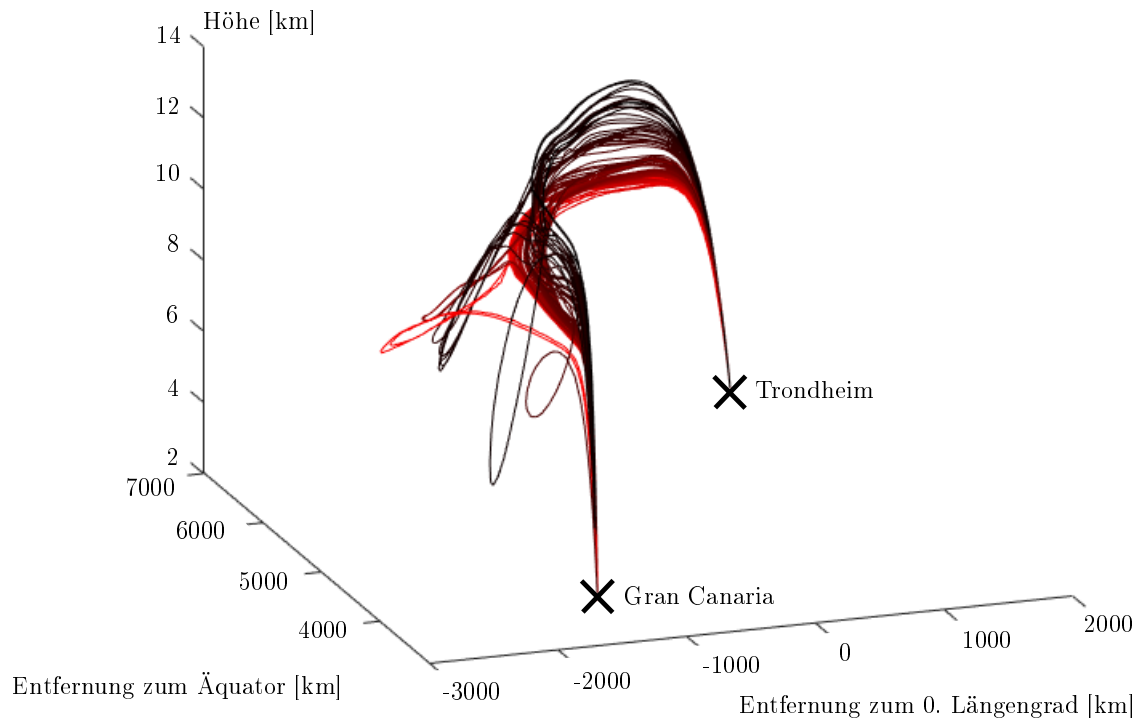


Abbildung 2.1: 101 optimierte Trajektorien für einen Flug von Gran Canaria nach Trondheim.

Die Trajektorien sind in verschiedenen Rottönen eingefärbt. Rot steht für eine umweltoptimale Trajektorie. Je dunkler die Trajektorie wird, desto stärker wird der COC-Faktor gewichtet. Entsprechend steht eine schwarze Trajektorie für eine kostenoptimale Route. Zu sehen ist, dass die meisten Trajektorien eine direkte Verbindung darstellen. Einzelne Trajektorien weichen aber stark von der direkten Verbindung ab. Bestimmte Wetterlagen können dazu führen, dass ein Flug durch gewisse Regionen eine kühlende Wirkung auf das Klima hat [49]. Die Kühlwirkung ist so groß, dass der zusätzliche Treibstoffverbrauch dadurch kompensiert wird [50] [39]. Trotz gleicher CCF auf der die Trajektorienberechnungen basieren, sorgt eine leicht verschiedene Gewichtung der Faktoren Umwelt und Kosten für sehr stark voneinander abweichende Trajektorien. Die umweltoptimale Route befindet sich in der geringsten Höhe, COC-optimale Routen in größerer Höhe. Diese Tendenz ist zu erwarten, denn das Fliegen in größerer Höhe verringert die Luftreibung, damit verbunden die Flugdauer und den Treibstoffverbrauch [51]. Im Gegensatz dazu haben umweltbelastende Reaktionen der Luft mit den Treibstoffabfallprodukten in geringerer Höhe einen kleineren Einfluss auf das Klima [20]. Dass sich die umweltoptimalen Trajektorien in geringerer Höhe befinden, ist daher ebenfalls zu erwarten.

Ziel des Projektes ATM4E waren bewusst visualisierbare Trajektorien und kein Code, der eine optimierte Route ausgibt [45]. Durch die Trajektoriensätze wird anschaulich dargestellt, innerhalb welches Korridors sich die Trajektorien befinden, ob und welche Ausreißer vorhanden sind und welche Auswirkungen kleine Änderungen in der Gewichtung von Kosten- und Umweltfaktoren haben. Ausreißer wie in Abbildung 2.1 fallen durch die Darstellung durch Trajektorien erst auf.

Als Datenbasis wurden CCFs verwendet, die Ergebnis des Projektes REACT4C sind [10] [17]. Da die Verwendung von CCFs sehr rechen- und damit kostenintensiv

ist, wurden während des Projektes ATM4E Algorithmic Climate Change Functions entwickelt (aCCFs) [42]. Die Verwendung von aCCFs erlaubt die Online-Berechnung der Trajektorien aus aktuellen Wetterdaten. Bei ATM4E wurden nur bestimmte konstante Wetterlagen betrachtet. Da in Zukunft die optimierten Trajektorien zur Planung von Flugrouten verwendet werden sollen, ist eine schnelle Online-Berechnung von Trajektorien unerlässlich.

Bisher wurden die Trajektoriensätze nicht ausgewertet. Daher beschäftigt sich diese Arbeit mit der Robustheit der verschiedenen Trajektorien. Sie beantwortet die Fragen, wie sehr die gegebenen Trajektorien voneinander abweichen und wie ähnlich sich die Trajektorien sind.

### 2.1.3 Fly Air Traffic Management for Environment (FlyATM4E)

Genau wie ATM4E ist FlyATM4E Teil des Projektes SESAR2020 [46]. Im Jahr 2020 wurde FlyATM4E als Nachfolger des 2018 abgeschlossenen Projektes ATM4E gestartet. Das Projekt läuft derzeit noch, soll aber Ende 2022 abgeschlossen werden [46]. Im Rahmen des Projektes wird, anders als bei ATM4E, kein konstantes Wetter angenommen. Bei FlyATM4E werden zwischen demselben Start und Ziel optimale Trajektorien zu verschiedenen Wetterlagen, die in der Realität so aufgetreten sind und als repräsentative Tage für das Wetter in der Region angesehen werden können, optimale Routen berechnet [10]. Abbildung 2.2 zeigt die Trajektorien zu verschiedenen Wetterlagen.

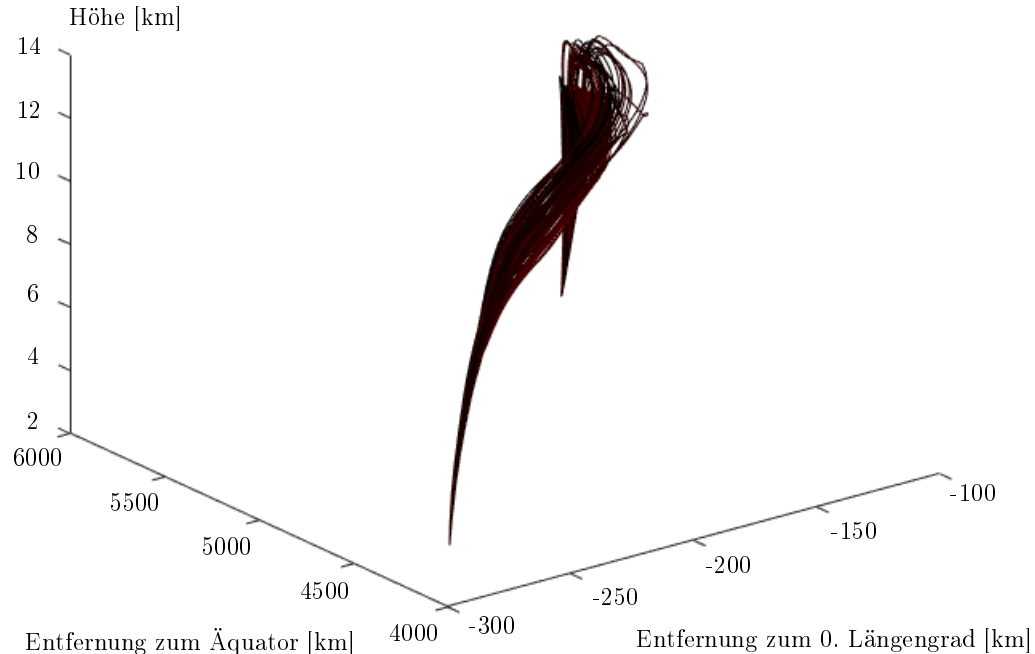


Abbildung 2.2: 510 optimierte Trajektorien für einen Flug von Granada nach London bei zehn verschiedenen Wetterlagen.

Zu zehn verschiedenen, als repräsentativ für diese Region anzusehenden Wetterlagen wurden jeweils 51 Trajektorien berechnet, die eine kosten- oder umweltoptimale

Route darstellen beziehungsweise eine Abstufung zwischen beiden Optimalen sind. Genau wie bei ATM4E ist ein Farbverlauf von rot nach schwarz im Diagramm zu erkennen. Rot repräsentiert die umweltoptimale Trajektorie, schwarz die kostenoptimale. Die Farbtöne zwischen rot und schwarz sind für die entsprechenden Abstufungen der Gewichtung.

Die insgesamt 510 Trajektorien liegen dicht beieinander. Da in MATLAB mit dem Plotten der umweltoptimalen Trajektorien in rot begonnen wird und danach die dunkler werdenden Trajektorien, das jeweilige Kostenoptimal, geplottet werden, überlagern die dunkel geplotteten Trajektorien hier die roten Trajektorien. Was dies genau bedeutet, wird in Abschnitt 4 aufgegriffen. Genau wie die Trajektorien des Projektes ATM4E wurden die Daten des Projektes FlyATM4E bisher nicht ausgewertet. Grundsätzlich sind die Daten beider Projekte vom gleichen Typ. Bei FlyATM4E muss zusätzlich unterschieden werden, bei welcher Wetterlage die Trajektorie berechnet wurde.

Vorangegangene Studien wie in [59] zeigten bereits, dass eine Optimierung von Flugrouten in Abhängigkeit des Wetters ein großes Einsparpotenzial bietet, sowohl bezüglich COC als auch bezüglich Umweltbelastung. Ziel von FlyATM4E ist es, Routen zu finden, die bei vielen möglichen Wetterlagen einen ähnlichen Verlauf aufweisen und somit robust bezüglich Unsicherheiten in der Wettervorhersage sind. Nach einem erfolgreich abgeschlossenen Projekt soll eine Auswahl an Flugrouten zur Verfügung stehen, die bezüglich verschiedener Aspekte optimal sind [19]. Dabei besonders von Interesse sind Routen, die die selben COC verglichen mit der jetzigen Situation aufweisen, jedoch eine geringere Auswirkung auf die Umwelt haben. Auch umweltoptimale Routen und die kostengünstigste Route ist von großem Interesse [9].

## 2.2 Ähnlichkeitsvergleiche

Um zwei Trajektorien miteinander vergleichen zu können, muss eine Kombination von Methode, Metrik und Vergleichsgröße gewählt werden. In der dieser Masterarbeit vorangegangenen Projektarbeit [62] wurde ausführlich dargelegt, warum diese drei Größen benötigt werden und welche Methoden, Metriken und Vergleichsgrößen in welcher Kombination für den Vergleich der gegebenen Flugtrajektorien in Frage kommen. Ergebnis der Projektarbeit ist, dass sich Euklidische Metrik, Hausdorff Metrik und Fréchet Metrik am besten für den Vergleich eignen, weshalb im Folgenden die geeigneten Metriken erneut aufgegriffen und erklärt werden. In 2.2.1 wird auf Metriken allgemein eingegangen. Für das weitere Verständnis werden die mathematischen Definitionen dieser Metriken benötigt, weshalb in 2.2.2 auf die Euklidische Metrik, in 2.2.3 auf die Hausdorff Metrik und in 2.2.4 auf die Fréchet Metrik mit ihren jeweiligen Definitionen eingegangen wird. Im Rahmen der Projektarbeit [62] wurde außerdem die Editierdistanz als mögliche Metrik vorgeschlagen, aufgrund des zu hohen Arbeitsaufwands und vielerlei Herausforderungen für die Projektarbeit aber verworfen. Die Editierdistanz wird in 2.2.5 erneut aufgegriffen, da mehr Zeit zur Verfügung steht und Skripte, die im Rahmen der Projektarbeit [62] prototypisch implementiert wurden, nun als Grundlage dienen können. Außerdem wurde das Verfahren der Energieminimierung in der Projektarbeit [62] als vielversprechend angesehen, weshalb es in Abschnitt 2.3 erläutert wird. In Abschnitt 2.4 wird auf die Vergleichsgröße eingegangen. Zusammen ergibt die Kombination von Metrik, Metho-

de und Vergleichsgröße eine Zahl, die für den Vergleich von Ähnlichkeit verwendet werden kann. Diese Zahl wird in späteren Untersuchungen einen zu berücksichtigenden Faktor für die Berechnung der Robustheit darstellen.

### 2.2.1 Metriken allgemein

Es gibt für Mengen verschiedene Strukturen, die unterschiedlich vieler Voraussetzungen bedürfen und dadurch bestimmte Rechenoperationen zulassen. Für eine Menge ist eine Topologie die allgemeinste Ordnungsstruktur. Eine engere Definition einer Ordnung wird durch eine Metrik definiert. Wie Abbildung 2.3 zeigt, inkludiert eine Metrik alle Eigenschaften einer Topologie.

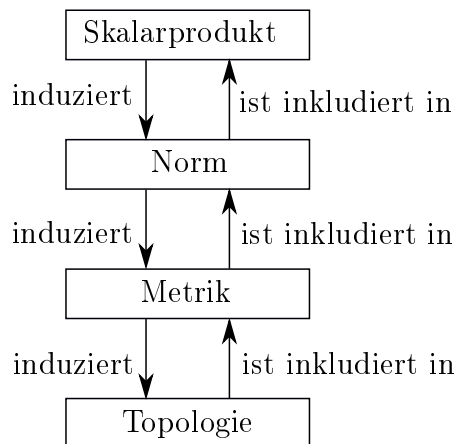


Abbildung 2.3: Hierarchie der Ordnungsstrukturen für Mengen.

Ist eine Metrik gegeben, wird eine Topologie automatisch induziert. Normen setzen wiederum mehr erfüllte Axiome voraus und beinhalten alle Eigenschaften einer Metrik. Eine gegebene Norm induziert eine Metrik, die wiederum eine Topologie induziert. Skalarprodukte weisen weitere Eigenschaften auf und beinhalten alle Axiome einer Norm, damit auch alle Eigenschaften von einer Metrik und einer Topologie. Skalarprodukte sind die schärfste, allgemein verwendete Definition von Ordnung für eine Menge. Weitere Eigenschaften könnten zusätzlich beliebig definiert werden.

Um eine Topologie auf einer Menge  $\mathcal{X}$  zu definieren, müssen folgende Axiome erfüllt sein [37]:

- i) Die Teilmengen leere Menge  $\emptyset$  und  $\mathcal{X}$  sind offen.
- ii) Der Durchschnitt endlich vieler offener Mengen ist offen.
- iii) Die Vereinigung beliebig vieler offener Mengen ist offen.

In der grundlegenden Projektarbeit für diese Masterarbeit [62] wurde ausführlich diskutiert, warum und wie die gegebenen Flugtrajektorien die Voraussetzungen einer Topologie erfüllen. Die Voraussetzungen für eine Topologie werden von den Trajektorien immer erfüllt. Somit stellt eine Topologie keine Einschränkung bezüglich der Anwendung auf Flugtrajektorien dar.

In Topologien kann aber noch kein Abstand als Zahl benannt werden. Das Vergleichen von Objekten ist innerhalb einer Topologie hingegen möglich. Jedoch kann die Relation von Objekten nur als „näher“, „gleich“ oder „weiter entfernt“ beschrieben werden. Die Menge  $\mathcal{X}$  enthalte die Objekte  $x, y$  und  $z$ . Um einen Abstand, in den folgenden Axiomen  $d(x, y)$  genannt, als Zahl ausdrücken zu können, bedarf es einer Metrik. Zusätzlich zu den Axiomen einer Topologie sind folgende Axiome in einer Metrik zu erfüllen [37]:

- i) Symmetrie:  $d(y, x) = d(x, y) \forall x, y$ .
- ii) Positive Definitheit:  $\forall x, y : d(x, y) \geq 0$  und  $d(x, y) = 0 \Leftrightarrow x = y$ .
- iii) Dreiecksungleichung: Für je drei Objekte  $x, y, z$  ist  $d(x, z) \leq d(x, y) + d(y, z)$  erfüllt.

In [62] wurde gezeigt, dass die gegebenen Trajektorien nicht immer alle Eigenschaften einer Metrik erfüllen. So dürfen die Trajektorien nicht in ihrer Gesamtheit betrachtet werden, sondern müssen in Teilabschnitte zerlegt werden. Genauere Ausführungen zu den Anforderungen, die eine Metrik stellt, sind in Abschnitt 3.5 aufgelistet. Wie gefordert, ist bei der Definition einer Metrik auf einer Menge die Abstandsbemessung als Zahl möglich. Bei Normen und Skalarprodukten handelt es sich um Metriken mit weiteren Axiomen, die erfüllt werden müssen. Da eine Metrik bereits ausreichend und notwendig ist, um Abständen eine Zahl zuzuordnen, werden Normen und Skalarprodukte hier nicht weiter betrachtet, da sie lediglich weitere Einschränkungen, aber keinen weiteren Nutzen liefern.

In [62] wurden verschiedene Metriken mit ihren Vor- und Nachteilen betrachtet. Euklidische Metrik, Hausdorff Metrik, Fréchet Metrik und die Editierdistanz haben sich im Rahmen der Projektarbeit [62] als erfolgversprechend erwiesen und werden daher hier aufgegriffen.

## 2.2.2 Euklidische Metrik

Die Euklidische Metrik gehört zu den durch eine Norm induzierten Metriken. Das heißt, dass mit der Definition des Abstandes nicht nur alle Voraussetzungen für eine Metrik, sondern auch für eine Norm erfüllt werden [3]. Die zusätzlichen Eigenschaften einer Norm sind für die Anwendung hier jedoch irrelevant. Die Euklidische Metrik bemisst Abstände  $d(x, y)$  durch

$$d(x, y) = \sqrt{\sum_i (x_i - y_i)^2} \quad (2.1)$$

wobei  $i \in \mathbb{N}$  die Anzahl der Richtungskomponenten, also die Dimension, innerhalb der Metrik indiziert. Die Metrik könnte durch jede andere durch eine Norm induzierte Metrik gleichwertig ersetzt werden, was durch die Hölder-Ungleichung bewiesen wird und zum Beispiel in [37] nachzulesen ist. Die Euklidische Metrik wurde gewählt, da sie dem alltäglich verwendeten Abstandsbegriff entspricht und damit anschaulicher ist.



### 2.2.3 Hausdorff Metrik

Für die Definition der Hausdorff Metrik wird ein Hilfsabstand benutzt. Als Hilfsabstand wird  $D$  zwischen einem Objekt  $x$  und einer nichtleeren kompakten Teilmenge  $K$  definiert:

$$D(x, K) := \min\{d(x, k) \mid k \in K\}. \quad (2.2)$$

mit  $d$  als Abstand einer durch eine Norm induzierten Metrik. Der Hausdorff-Abstand  $\delta$  zwischen zwei nichtleeren kompakten Teilmengen  $A$  und  $B$  ist definiert als

$$\delta(A, B) := \max\{\max\{D(a, B) \mid a \in A\}, \max\{D(b, A) \mid b \in B\}\}. \quad (2.3)$$

Die Definition bedeutet, dass allen Punkten aus  $A$  jeder Punkt in  $B$  zugeordnet werden darf und umgekehrt. Es gibt keine unzulässige Wahl. Außerdem gibt es einen Punkt  $a$  in  $A$  und ein dazugehöriges  $b$  in  $B$ , sodass der Abstand genau (dem frei wählbaren)  $D$  entspricht [2].

Die Hausdorff Metrik vergleicht zwei Trajektorien, indem zu jedem Trajektorienpunkt oder -element der Punkt mit dem geringsten Abstand ausgewählt wird. Der Abstand bezeichnet hier den Durchmesser des kleinstmöglichen Kreises, der beide zu vergleichenden Elemente beinhaltet. Da die Auswahl beliebig ist, wird eine mögliche Zeitabhängigkeit in der Hausdorff Metrik nicht berücksichtigt. Mit Bezug auf Flugtrajektorien wird für die Verwendung dieser Metrik angenommen, dass Zeit und Ort kausal voneinander anhängig sind und eine geringe geometrische Distanz einer kleinen zeitlichen Differenz entsprechen.

In [62] wurde erklärt, dass diese Metrik auf die gegebenen Flugtrajektorien immer anwendbar ist, sofern diese nicht in ihrer Gesamtheit, sondern detaillierter, zum Beispiel als Menge von Punkten oder geteilt in verschiedene Unterelemente, betrachtet werden. Aus der Hausdorff Metrik ergeben sich daher keine Einschränkungen für die Darstellung. Um den vergleichsweise großen Rechenaufwand, der durch die Hausdorff Metrik entsteht, gering zu halten, bedarf es spezieller Algorithmen [27] [30]. Auf diese wird in Abschnitt 3.4 eingegangen.

### 2.2.4 Fréchet Metrik/ „Dog-man“ Metrik

Für die Definition der Fréchet Metrik wird in diskrete und kontinuierliche Fälle unterschieden. Für diskrete Objekte lautet sie [37]:

$$\delta = \{((a_i, b_j), (a_{i+1}, b_j)) \mid \|a_i - b_j\|, \|a_{i+1} - b_j\| \leq D\} \cup \{((a_i, b_j), (a_i, b_{j+1})) \mid \|a_i - b_j\|, \|a_i - b_{j+1}\| \leq D\} \quad (2.4)$$

wobei die Bezeichnungen aus der Definition der Hausdorff Metrik übernommen wurden.  $S$  besteht aus Elementen  $a_1$  bis  $a_m$  mit  $m \in \mathbb{N}$  und  $i \in \mathbb{N}$  ist die Laufvariable. Analog sind  $b_1$  bis  $b_n$  mit  $n \in \mathbb{N}$  Elemente von  $B$  und werden von  $j \in \mathbb{N}$  und o.B.d.A.  $i \leq j$  gezählt.

Für kontinuierliche Objekte ist die Fréchet Metrik durch

$$\delta(f, g) = \inf \max\{\|f(a(t)) - g(b(t))\|\} \quad (2.5)$$

definiert. Die Zeit wird durch  $t$  angegeben,  $f$  und  $g$  sind kontinuierliche Funktionen [2].

Die Fréchet Metrik ist der Hausdorff Metrik sehr ähnlich. Die Fréchet Metrik berücksichtigt verglichen mit der Hausdorff Metrik eine Zeitabhängigkeit der Daten. Deshalb wird die Fréchet Metrik auch „Dog-Man“ Metrik genannt. Der Vergleich der Trajektorien mit Hilfe der Fréchet Metrik kann sich bildlich gesprochen als Mann mit Hund vorgestellt werden und wird hier anhand dieser Analogie erläutert. Der Mann läuft auf seinem Spaziergang eine bestimmte zeitabhängige Route entlang. Der Weg des Mannes entspricht der einen zu vergleichenden Trajektorie. Der Hund wird vom Mann an der Leine mit einer bestimmten konstanten Länge geführt. In Abhängigkeit der Position des Mannes kann der Hund sich an alle Punkte innerhalb eines Kreises bewegen und dort den interessantesten Punkt auswählen. Beim Vergleich der Trajektorien wird genauso zur Position auf der einen Trajektorie der beste Punkt der anderen Trajektorie innerhalb eines bestimmten Radius ausgewählt. Dabei muss sichergestellt werden, dass der Radius groß genug ist, sodass sich immer mindestens ein Punkt der anderen Trajektorie innerhalb der Kreises beziehungsweise im Dreidimensionalen innerhalb der Kugel befindet. Ist ein solcher Punkt gewählt, entspricht die tatsächlich benötigte Leinenlänge dem Abstand. Bewegt sich der Mann nun zum nächsten Punkt auf der Trajektorie, muss der Hund aus einer anderen Menge erreichbarer Orte wählen. Dabei ist es zulässig, dass der Hund stehen bleibt. Sofern ausgewählte Punkte weiterhin den geringsten Abstand ausweisen, dürfen sie mehrfach benutzt werden. Dem kleineren Datensatz wird je ein Punkt des längeren Datensatzes zugeordnet. Diese Zuordnung geschieht per Definition der Hausdorff Metrik. Die Symmetrie der Metrik wird durch diese Vorschrift nicht verletzt, da sich die Symmetrie auf den Abstand und nicht auf eine gesetzte Zuordnungsvorschrift bezieht [2].

### 2.2.5 Editierdistanz / Levenshtein Metrik

Levenshtein Metrik und Editierdistanz bezeichnen die gleiche Metrik. Der Ausdruck Levenshtein Metrik wird üblicherweise in der Mathematik benutzt, während die Metrik in der Informatik als Editierdistanz bezeichnet wird.

Die Editierdistanz wird zum Vergleichen von Zeichenketten verwendet. Die gegebenen Flugtrajektorien müssen entsprechend als eine diskrete Folge von Zeichen dargestellt werden. Für die Zeichenketten stehen die Operatoren „Einfügen“, „Löschen“ und „Ersetzen“ zur Verfügung. Zeichen können an beliebiger Position der Zeichenkette eingefügt oder gelöscht werden. Bei „Ersetzen“ wird die Position des zu ersetzenden Zeichens beibehalten und das Zeichen durch ein anderes überschrieben. Die drei Operatoren sind gleich gewichtet, das heißt, jede Operation verursacht denselben Aufwand. Der Abstand bezeichnet die kleinstmögliche Anzahl von Operationen, um identische Zeichenketten zu erhalten.

Bei den gegebenen Zeichenketten „1011“ und „10011“ beträgt der Abstand  $d = 1$ . An zweiter oder dritter Position muss in die erste Zeichenkette eine Null eingefügt werden, um identische Ketten zu erhalten. Am Beispiel ist zu sehen, dass zwar der Abstand mit  $d = 1$  eindeutig ist, nicht aber auf welche Position Operatoren angewendet werden. Soll aus „Metrki“ das Wort „Metrik“ gebildet werden, ist eindeutig  $d = 2$ . Der Abstand kann gefunden werden, indem das k durch ein i ersetzt wird und das i durch ein k. Es könnte aber genauso das k gelöscht werden und ans Ende eingefügt werden.

Bei kurzen Zeichenketten ist es noch möglich, den Abstand intuitiv zu berechnen. So ist für das erste Beispiel klar, dass  $d \neq 0$ , da die Zeichenketten nicht identisch sind. Für den nächstgrößeren Abstand  $d = 1$  ist ein zugehöriges Verfahren schnell gefunden. Für das zweite Beispiel muss  $d \leq 3$  sein, da die Zeichenketten dieselbe Länge besitzen und sich nur drei Zeichen unterscheiden. Eine weitere Eingrenzung für  $d$  kann nur per Exhaustive Search Verfahren gefunden werden, was für das Beispiel vergleichsweise schnell erfolgen kann. Bei computergestützten Berechnungen müssen alle Kombinationen per Exhaustive Search Verfahren durchgegangen werden. Zwar existieren Ansätze, um bereits verglichene Blöcke von Zeichen abzuspeichern und auf einen erneuten Abgleich zu verzichten, der Aufwand wird dadurch aber kaum reduziert [5].

## 2.3 Energieminimierungsverfahren

Methoden, die auf Energieminimierung beruhen, werden in diskrete und kontinuierliche unterschieden [13]. Voraussetzung für beide Arten ist, dass die Trajektorie knick- und sprunfrei verläuft, da Änderungen an Sprüngen und Knicken nicht eindeutig zu bestimmen sind und so Fehler bei der computergestützten Berechnung auftreten. Es gibt außerdem Varianten der Energieminimierung, die zeitliche Abhängigkeit berücksichtigen und welche, die dies nicht tun. Da Flugtrajektorien verglichen werden sollen, wurden nur Energieminimierungsverfahren betrachtet, die Zeitabhängigkeit berücksichtigen. Innerhalb einer Trajektorie ist es physikalisch nicht möglich, Punkte beliebig zu vertauschen. Die Punkte müssen in ihrer Reihenfolge beibehalten werden, da das Flugzeug sie nacheinander abfliegen muss. Sprünge bezüglich der Position sind nicht möglich.

**Diskrete Variante:** Die zu untersuchenden Trajektorien werden, wenn nicht bereits in passender Form vorliegend, in Punkte unterteilt. Ist das diskrete Modell von einem kontinuierlichen abgeleitet, erhält jeder Punkt entsprechend der im kontinuierlichen gebildeten Ableitungen zusätzlich einen Wert für seine Änderung an der jeweiligen Stelle. Liegen nur Punkte vor, wird die Änderung mit Hilfe der Nachbarpunkte berechnet: Die lineare Verbindung zweier benachbarter Punkte entspricht der Änderung. In einer passenden Metrik wird jeweils der Abstand der Punktpaare berechnet und der Abstand gespeichert. Die Abstände sind ein Maß für Ähnlichkeit [70].

Abbildung 2.4 zeigt zwei kontinuierliche Trajektorien, je eine rote und eine blaue, und ihre Diskretisierung durch die Punkte  $A$  bis  $D$  beziehungsweise  $A'$  bis  $D'$ . Angenommen wird, dass die Trajektorien senkrecht verliefen, bevor sie Punkt  $A$  beziehungsweise  $A'$  durchqueren. Ebenso wird angenommen, dass die eine Trajektorie waagrecht verläuft nachdem sie  $D$  erreicht hat und die andere Trajektorie nach  $D'$  weiter senkrecht verläuft.

Bei einer Ähnlichkeitsbestimmung über den Vergleich der Abstände von Punkten sind  $A$  und  $A'$  sowie  $B$  und  $B'$  gleich. Die Punkte  $C$  und  $C'$  weisen eine kleine Differenz auf. Sie sind nicht gleich, unterscheiden sich aber nur wenig voneinander. Die Punkte  $D$  und  $D'$  besitzen einen großen Abstand zueinander und sind damit deutlich verschieden. Anders wird bei der Methode der minimale Energie gemessen.

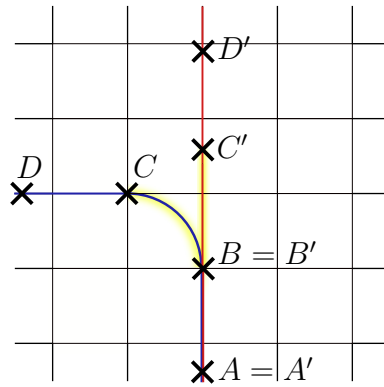


Abbildung 2.4: Methode der minimalen Energie mit diskretisierten und kontinuierlichen Trajektorien.

Die Richtung ist vor und beim Durchqueren von  $A = A'$  dieselbe. Die Punkte  $B$  und  $B'$  sind verschieden, da die Richtung der Verbindungslinien von  $A$  nach  $C$  und von  $A'$  nach  $C'$  eine andere ist. Aus demselben Grund sind  $C$  und  $C'$  verschieden. Während in  $C$  eine Richtungsänderung stattfindet, bleibt der Verlauf in  $C'$  unverändert. Hingegen sind  $D$  und  $D'$  nach der Energieminimierungsmethode gleich, da keine Richtungsänderung vorhanden ist.

Veranschaulicht dargestellt entspricht die Methode dem Verbiegen eines Stücks Draht. Würde hier der Abschnitt  $\overline{BC}$  gerade gebogen werden, wären die Verläufe identisch. Da nur Punkte betrachtet werden, müssen für einen identischen Verlauf die Form in  $B$  und  $C$  verändert werden, um die Gestalt von  $B'$  und  $C'$  zu erreichen. Beim Verbiegen werden folgende Abschnitte anders positioniert, sodass Teilstücke als ähnlich gelten, auch wenn sie vorher in eine andere Richtung gezeigt haben. Sowohl in  $D$  als auch in  $D'$  erfährt der Draht keine Gestaltveränderung, sodass dieser Abschnitt als identisch eingestuft wird.

**Kontinuierliche Variante**, auch „Curve Alignment“ genannt [13]: Die diskrete Variante der Energieminimierung bildet, da nur Punkte betrachtet werden, die Vorstellung vom Verbiegen eines Drahtes nur etwa ab. Für die kontinuierliche Variante ist diese Vorstellung die Grundlage und kann direkt übernommen werden. Um das Verfahren auf Flugtrajektorien anzuwenden, müssen die Trajektorien zunächst auf dieselbe Länge skaliert werden. Allgemein betrachtet ist ein Skalierungsfaktor verschieden von Eins bereits eine Charakteristik für die Ähnlichkeit der Trajektorien [57]. Eine solche Skalierung ist unter dem Namen „Dynamic Time Warping“ bekannt [35], weil physikalische Einheiten wie Sekunden durch die Skalierung nicht mehr einer Sekunde der anderen Trajektorie entsprechen, sondern durch die Skalierung verzerrt werden [63]. Für die Flugtrajektorien wird angenommen, dass die Skalierung nur wenig vom Faktor Eins abweicht. Im Verhältnis zur gesamten Fluglänge ist die Skalierung sehr klein und wird somit als vernachlässigbar angenommen. In Abbildung 2.4 handelt es sich bei dem gelb hinterlegten Stück um den Abschnitt, der verändert werden muss. Um die benötigte Energiemenge zu berechnen, die benötigt wird, um aus den beiden Trajektorien deckungsgleiche zu formen, wird für die Trajektorien ein beliebiges Material definiert. Per Finiten Elemente Methode (FEM) wird berechnet, welche Verformungsenergie benötigt wird, um die eine Trajektorie zur anderen zu formen. Wird diese Energie pro Steifigkeitswert des Materials

genommen, ist die Energiemenge konstant und somit unabhängig vom zuvor willkürlich verwendeten Material [35]. Für ein steifes Material wird viel Energie zur Verformung benötigt. Jedoch ist die Steifigkeit ebenfalls groß, sodass das Verhältnis von benötigter Energie zur Steifigkeit immer dasselbe ist.

Das Verfahren per FEM für diese Methode ist streng genommen auch diskret. Für die Berechnung per FEM muss für eine kontinuierliche Trajektorie eine Diskretisierung vorgenommen werden. Diese ist jedoch sehr fein, weshalb trotz dieser notwendigen Diskretisierung für die FEM von einer kontinuierlichen Methode gesprochen wird [6] [13].

## 2.4 Vergleichsgröße

Die Flugtrajektorien sind als vierdimensionale Daten gegeben: Drei Dimensionen beschreiben die räumliche Position, eine weitere Dimension gibt die zugehörige Zeit an. Durch Interpolation zwischen den Punkten kann die zurückgelegte Strecke bestimmt werden. Anhand der zugehörigen Zeiten und deren Differenz kann Geschwindigkeit und Beschleunigung des Flugzeugs berechnet werden. Für einen knick- und sprungfreien Verlauf von Geschwindigkeit und Beschleunigung muss die Strecke mindestens durch ein Polynom zweiten Grades angenähert werden. Bei einer linearen Interpolation entstehen Knicke beim Streckenverlauf, damit Sprünge in der Geschwindigkeit. Dies führt dazu, dass die Beschleunigung an einigen Stellen nicht definiert ist. Entsprechend entstehen bei der computergestützten Berechnung Fehler bei der Ausgabe der Beschleunigung.

Bei der Geschwindigkeit und der Beschleunigung handelt es sich, abgesehen von den Anfangsbedingungen, um die gleichen Informationen wie bei der Strecke. Welche dieser Größen verglichen wird, ist daher nicht relevant. Der Streckenverlauf beinhaltet alle Informationen direkt. Im Gegensatz zu Geschwindigkeit und Beschleunigung müssen keine zusätzlichen Anfangsbedingungen für denselben Informationsgehalt verwendet werden. Um Trajektorien bezüglich Robustheit zu klassifizieren, ist ebenfalls der geometrische Abstand von den genannten Größen am interessantesten. Ob Geschwindigkeit und Beschleunigung geringe Abweichungen aufweisen, ist weniger wichtig als eine geringe räumliche Distanz. Daher werden die Ähnlichkeitsberechnenden Verfahren bezüglich der Strecke angewendet.

## 2.5 Clusteranalyse

Clusteranalysen werden verwendet, wenn eine Vielzahl von Daten vorliegt und diese geordnet werden sollen. Dabei ist die Datenmenge oft so groß, dass nicht alle Daten detailliert miteinander verglichen werden, sondern die Daten anhand vorher festgelegter Merkmale in Gruppen sortiert werden. Diese Gruppen werden Cluster genannt. In MATLAB gibt es viele bereits implementierte Funktionen und eine Vielzahl wählbarer Parameter, die für eine Clusteranalyse eingesetzt werden können [60]. Was diese Funktionen bewirken und welche Parameter wählbar sind, wird im Folgenden erläutert. So wird in 2.5.1 die Unterscheidung von harter und weicher

Clusteranalyse erklärt. In 2.5.2 wird auf das mögliche Löschen von Daten eingegangen. Außerdem ist die Anzahl der Cluster wählbar, was in 2.5.3 beschrieben wird. Hierbei kann es sich um eine Zahl handeln oder einen Zahlenraum. In 2.5.4 wird eine Datenstruktur festgelegt. Mit verschiedenen Optionen kann entschieden werden, ob und wie fein aufgegliederte hierarchische Strukturen gebildet werden sollen. Mögliche Verfahren, wie Cluster gebildet werden, werden in 2.5.5 erklärt. Mittels einer Verteilungshypothese kann festgelegt werden, dass das Cluster eine bestimmte Verteilung berücksichtigen soll, was in 2.5.6 erklärt wird.

### 2.5.1 Harte und weiche Cluster

Bei einer Clusteranalyse muss festgelegt werden, ob eine harte oder weiche Unterteilung von Clustern durchgeführt werden soll. Es muss eine der beiden Optionen gewählt werden, keine andere ist möglich.

#### Harte Cluster

Jedes Objekt, hier eine Trajektorie, wird bei einer harten Clusteranalyse genau einem Cluster zugeordnet. Eine Trajektorie kann nicht Teil von mehreren Clustern sein. Es ist dabei denkbar, dass ein Cluster nur eine einzige Trajektorie beinhaltet. Abbildung 2.5 zeigt den gefilterten Trajektoriensatz “00005\_GCLP\_ENVA” des Projektes ATM4E mit einer harten Clusterbildung in zehn Cluster. Speziell für

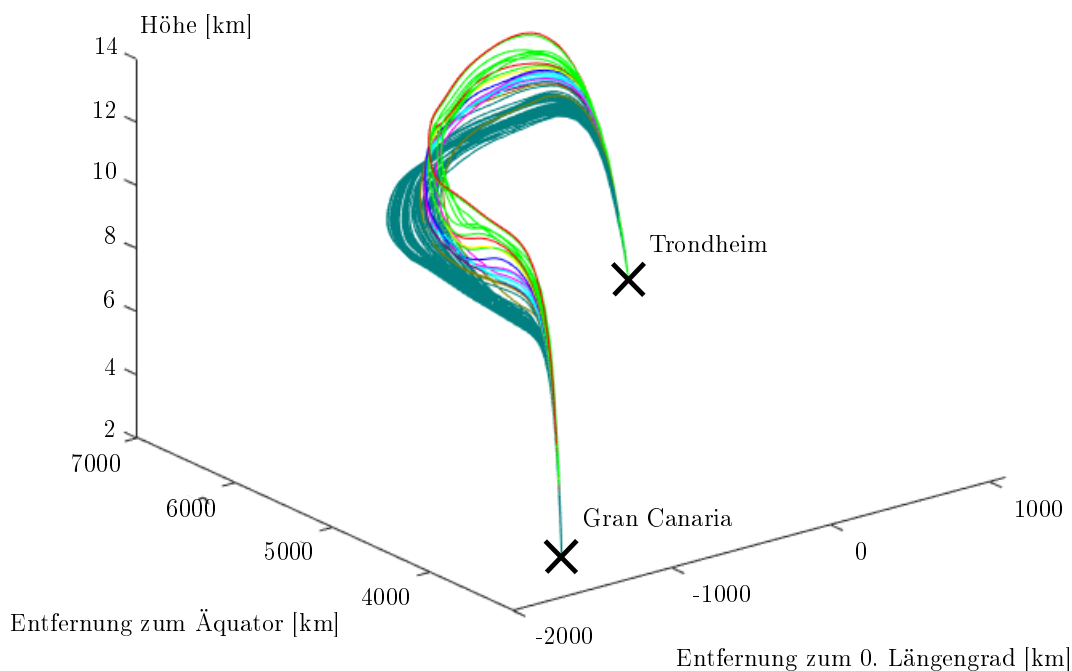


Abbildung 2.5: Trajektoriensatz “00005\_GCLP\_ENVA” mit der Unterscheidung in zehn harte Cluster.

diesen Trajektoriensatz ist auffällig, dass sich durch die festgelegte Anzahl von zehn Clustern ein großes Cluster in grün-blau und neun weitere Cluster jeweils aus sehr wenigen Trajektorien bilden.

### Weiche Cluster

Bei einer weichen Clusterbildung dürfen Objekte mehreren Clustern angehören. Sie müssen mindestens zu einem Cluster zugehörig sein und dürfen, je nach festgelegten Parametern, in beliebig vielen Clustern enthalten sein. Auch hier dürfen Cluster aus einer einzigen Trajektorie bestehen. Dies geschieht aber nur, wenn die ausgewählte Trajektorie keinem anderen Cluster zugehörig ist. Sollte die Trajektorie anderweitig zuzuordnen sein, wird kein Cluster mit einer einzelnen Trajektorie gebildet.

Das Zuordnen zu mehreren Gruppen stellt in MATLAB kein Problem dar, jedoch ist die Visualisierung mehrfach zugeordneter Daten eine Herausforderung. Da es sich um kontinuierliche Trajektorien handelt, ist die Darstellung durch verschiedene Symbole wie zum Beispiel „X“ und „O“ nicht möglich. Eine weitere Möglichkeit wäre, mehrfach zugeordnete Trajektorien in der Mischfarbe der zugehörigen Cluster darzustellen. Hier kann aber nicht eindeutig unterschieden werden, aus welchen Clustern die Trajektorie sich zusammensetzt oder ob sie gegebenenfalls ein neues, nicht zugehöriges Cluster darstellt. Zur Unterscheidung von mehrfach zugeordneten Daten wurden für den Trajektoriensatz in Abbildung 2.6 neben verschiedener Farben, die sich zum Teil überlappen, zusätzliche Meldungen in MATLAB ausgegeben.

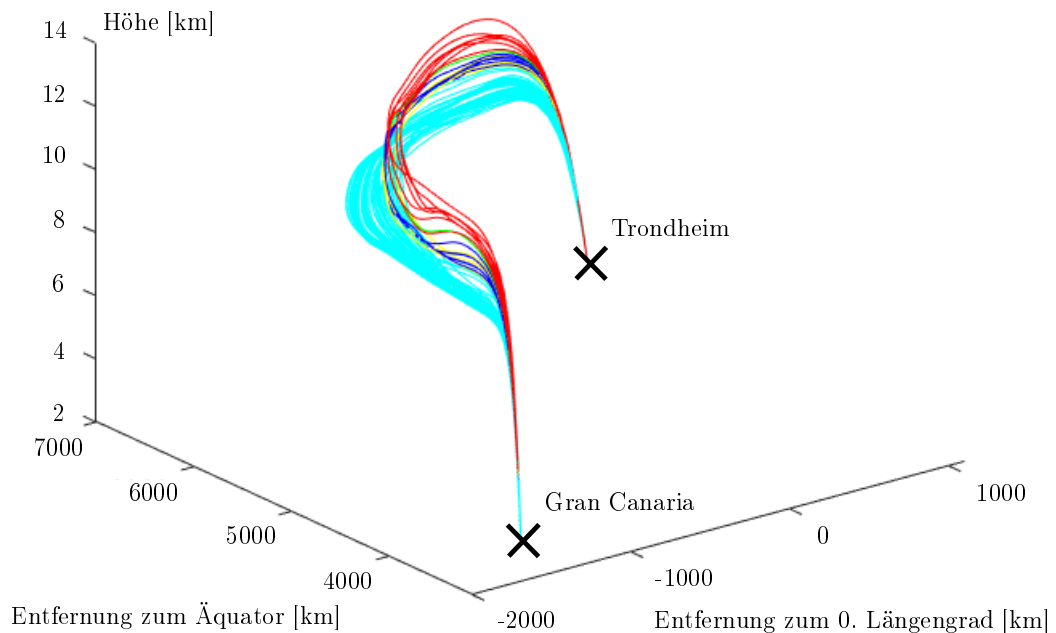


Abbildung 2.6: Trajektoriensatz “00005\_GCLP\_ENVA“ mit der Unterscheidung in fünf weiche Cluster.

Meldungen in MATLAB:

```
ans = Cluster 1 und Cluster 2 überschneiden sich zu 67 Prozent.
ans = Cluster 1 und Cluster 3 überschneiden sich zu 68 Prozent.
ans = Cluster 2 und Cluster 3 überschneiden sich zu 30 Prozent.
```

Cluster 1 ist in Abbildung 2.6 grün dargestellt, Cluster 2 gelb eingefärbt und Cluster 3 mit blauer Farbe markiert. Als Mischfarbe der Cluster ergibt sich das große hellblaue Cluster.

### 2.5.2 Löschoptionen

Unterhalb einer festzulegenden Anzahl von Elementen innerhalb eines Clusters können Elemente gelöscht werden. Bestehen Cluster zum Beispiel aus einzelnen Trajektorien, kann entschieden werden, dieses Cluster zu behalten oder die enthaltenen Daten zu löschen, da es sich wahrscheinlich um einen Ausreißer handelt. Bezogen auf die Flugtrajektorien handelt es sich um Routen, in deren Nähe sich keine andere optimale Route befindet. Das Verwerfen von Daten ist nicht nur auf Trajektorien beschränkt, die einzeln vorkommen. Es kann eine beliebige Zahl festgelegt werden, aus wie vielen Trajektorien ein Cluster bestehen muss, um nicht gelöscht zu werden. Abbildung 2.7 zeigt den vollständigen Datensatz des Fluges „00005\_GCLP\_ENVA“. Ein Farbverlauf von rot nach schwarz zeigt dabei an, um welche Trajektorie es sich

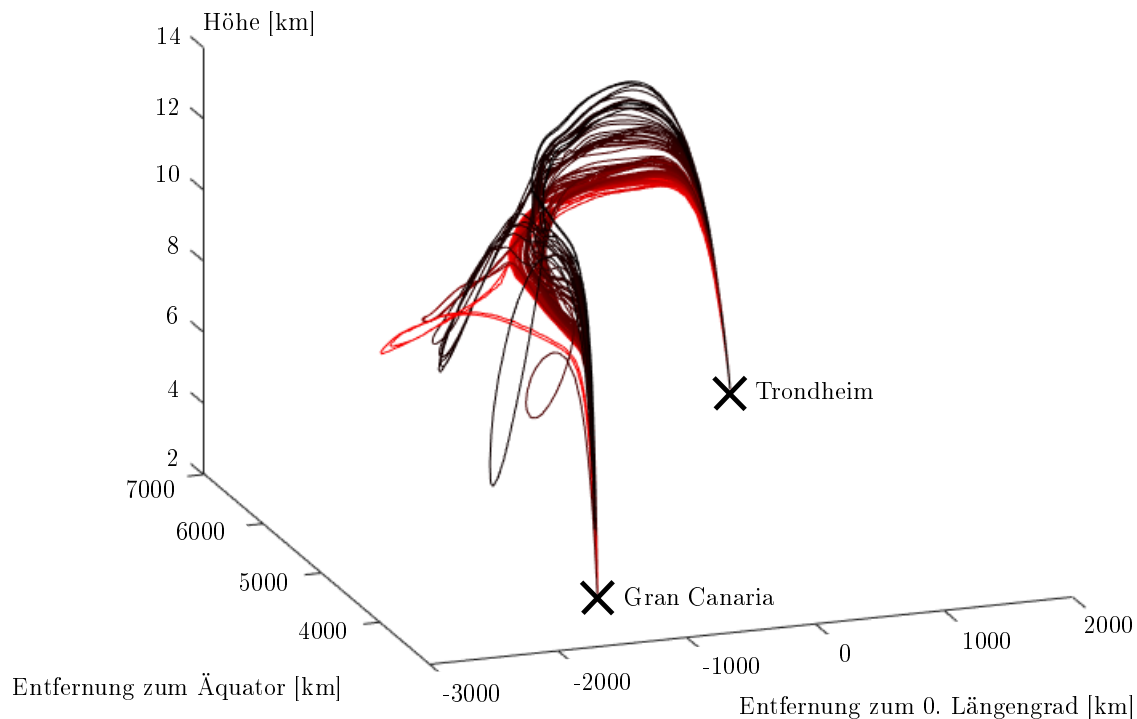


Abbildung 2.7: Trajektoriensatz „00005\_GCLP\_ENVA“ ohne Cluster.

handelt. Die für die Umwelt optimale Route ist rot markiert, die kostengünstigste Trajektorie ist schwarz dargestellt. Entsprechend ihrer Einfärbung entstehen die anderen Trajektorien durch verschiedene Abwägungen zwischen Kosten- und Umweltoptimal. Abbildung 2.8 zeigt einen Fall, bei dem Cluster, die aus weniger als fünf Trajektorien bestehen, verworfen werden.

### 2.5.3 Anzahl von Clustern

Ein Datensatz wird in eine bestimmte Clusteranzahl aufgeteilt oder ein maximaler Abstand zwischen allen Daten innerhalb eines Clusters festgelegt. Der maximale Abstand ist eine konkurrierende Größe zur Anzahl der Cluster. Wird eine der Größen festgelegt, ergibt sich daraus die andere. Wird die Anzahl der Cluster vorgegeben,



ist dies eine statische Berechnung genannt, während die Vorgabe der maximalen Abstände eine dynamischer Berechnung ist.

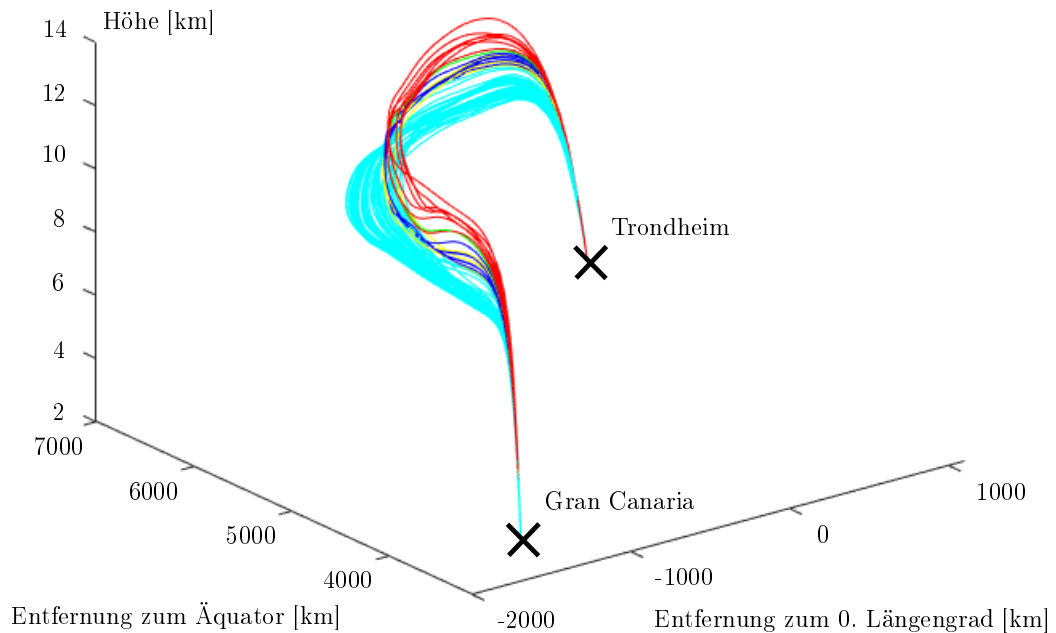


Abbildung 2.8: Trajektoriensatz „00005\_GCLP\_ENVA“ mit gelöschten Elementen und der Einteilung in fünf Cluster.

**Statische Berechnung:** Bei einer statischen Aufteilung wird vor Programmstart festgelegt, in wie viele Cluster die gegebenen Daten aufzuteilen sind. Eine festgelegte Anzahl von Clustern bedeutet, dass der maximale Abstand zwischen den Trajektorien dadurch festgelegt wird. Abbildung 2.9 zeigt drei Fälle, in denen gegebene Daten, durch „X“ dargestellt, in  $c = 4$ ,  $c = 3$  beziehungsweise  $c = 2$  Cluster aufgeteilt werden sollen. Für  $c = 4$  ergibt sich der maximale Abstand von einem Kästchen zwischen den Datenpunkten. Hier wurde eine weiche Clusteranalyse durchgeführt, da ein Datenpunkt zu mehreren Clustern gehört. Zwei Cluster bestehen aus einer einzigen Trajektorie. Für  $c = 3$  und  $c = 2$  ergeben sich zwei Cluster. Hier beträgt der maximale Abstand zwischen den Daten zwei Kästchen innerhalb beider Cluster. Die Verteilungen sind identisch, da bei  $c = 2$  beide Cluster denselben maximalen Abstand aufweisen. Durch ein weiteres Cluster wird der maximale Abstand insgesamt nicht verkleinert. Erst bei  $c = 4$  kann eine weitere Verfeinerung der Cluster vorgenommen werden.

**Dynamische Berechnung:** Bei einer dynamischen Berechnung wird der maximale Abstand zwischen den Elementen vorgegeben. Die Anzahl der Cluster ergibt sich aus den durch die Elemente gegebenen Abständen. Abbildung 2.10 zeigt die Clusterbildung, wenn der maximal tolerierte Abstand  $a = 0,5$ ,  $a = 1$  und  $a = 2$  beträgt. Beim maximalen Abstand  $a = 0,5$  können keine Cluster aus mehr als einer Trajektorie gebildet werden. Jede Trajektorie für sich ist eine eigene Gruppe. Im Beispiel ergeben sich fünf Cluster. Für den maximal zulässigen Abstand  $a = 1$  ergeben sich

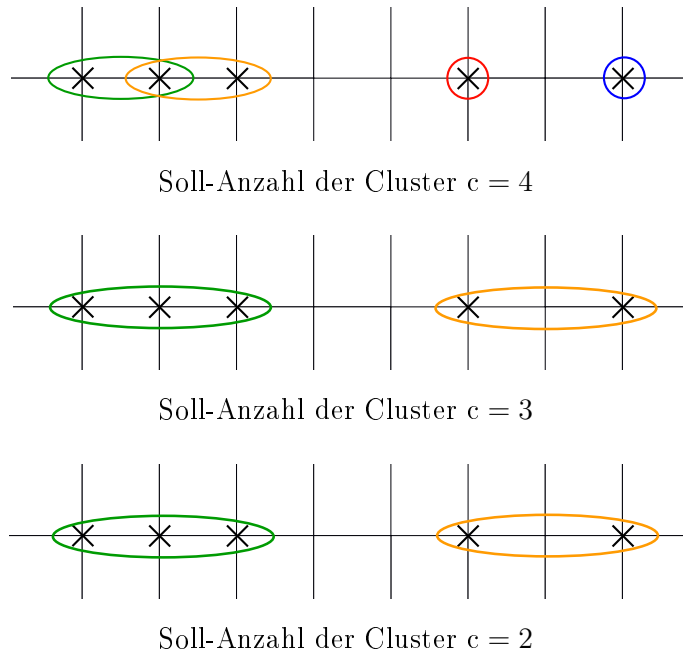


Abbildung 2.9: Beispiele für Clusterbildung für eine festgelegte Anzahl von Clustern.

vier Cluster, davon zwei einzelne Trajektorien und zwei Cluster mit zwei Elementen. Auch in Abbildung 2.10 wurde wie oben eine weiche Clusteranalyse durchgeführt, da ein Datenpunkt zu zwei Clustern gehört. Auch der maximale Abstand von  $a = 2$  entspricht einem Fall aus Abbildung 2.9. Es ergeben sich zwei Cluster mit zwei und drei Elementen, was dem Fall  $c = 2$  von oben entspricht.

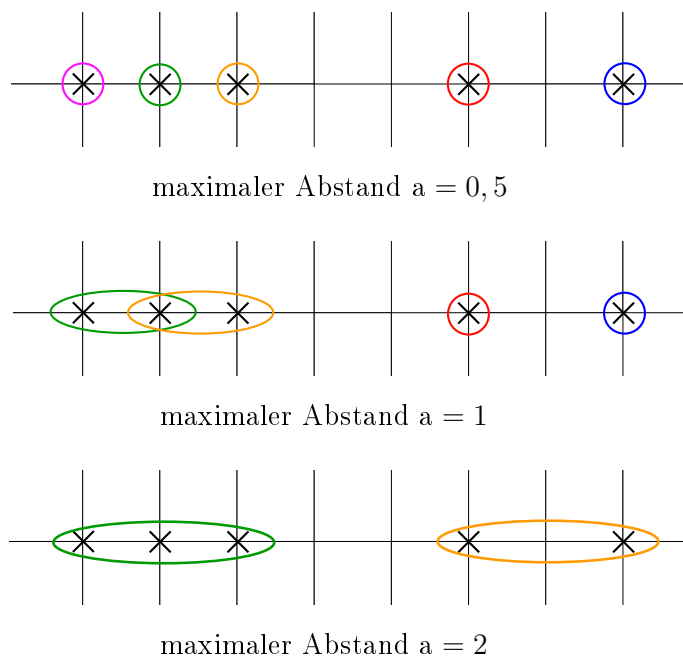


Abbildung 2.10: Beispiele für Clusterbildung für festgelegte maximale Abstände.

### 2.5.4 Hierarchische Cluster

Neben dem Zuteilen von Daten zu einem Cluster gibt es die Möglichkeit, Daten innerhalb eines Clusters weiter zu unterteilen. Alternativ werden alle Daten einem oder keinem Cluster zugeordnet. Innerhalb eines Clusters wird keine weitere Unterscheidung der Daten vorgenommen. Abbildung 2.11 zeigt die Clusterbildung ohne weitere Unterteilung.

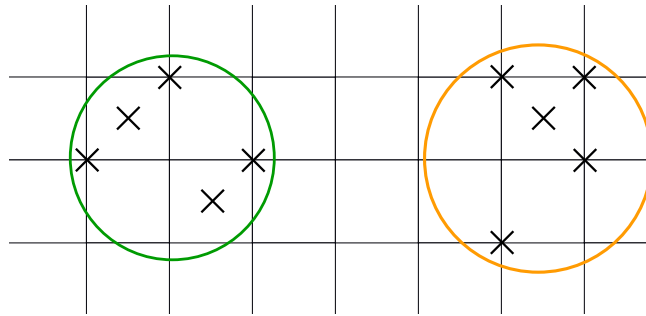


Abbildung 2.11: Beispiel für Clusterbildung ohne hierarchische Struktur.

Abhängig davon, ob eine harte oder weiche Clusteranalyse vorgenommen wird, können Daten zu einem oder mehreren Clustern gehören. Innerhalb eines Clusters wird aber nicht weiter unterschieden. Eine feinere Unterteilung der Cluster wird nur vorgenommen, wenn eine hierarchische Clusterbildung gewählt wurde. Abbildung 2.12 zeigt, wie die weitere Unterteilung beispielsweise aussehen könnte.

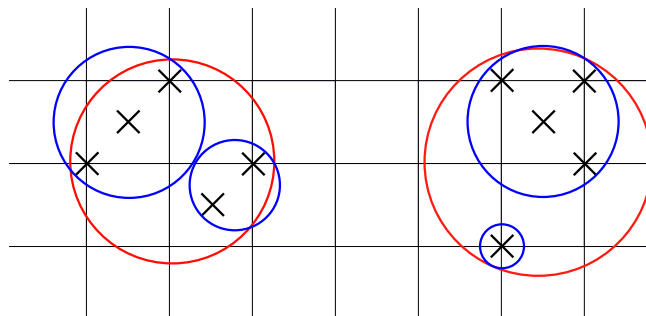


Abbildung 2.12: Beispiel für Clusterbildung mit hierarchischer Struktur.

Hier wurde festgelegt, dass die Unterteilung in zwei Hierarchieebenen zu je zwei Clustern stattfinden soll. Die Aufteilung innerhalb der Cluster wird über verschiedene Parameter bestimmt: So kann festgelegt werden, wie viele Cluster sich innerhalb eines Clusters befinden sollen oder wie groß der maximale Abstand zwischen Daten der Untergruppen sein soll. Auch die Anzahl der Ebenen, in die geteilt wird, kann festgelegt werden. In Abbildung 2.12 wurde in zwei Ebenen aufgeteilt, die jeweils rot und blau markiert wurden.

### 2.5.5 Variierendes Mittel

In 2.5.3 wurde erklärt, dass Cluster entweder entsprechend einer vorgegebenen Anzahl oder basierend auf maximale zulässigen Abständen festgelegt werden können. Maximal zulässiger Abstand und Anzahl der Cluster sind konkurrierende Größen. Gibt es für keine der beiden Optionen einen Anhaltspunkt, welche Zahl zu wählen ist, kann die Clusterbildung über variierende Mittel erfolgen. Hierbei werden alle Möglichkeiten für die Anzahl der Cluster, also ein Cluster bis hin zur Anzahl der Elemente, berechnet und der dazugehörige maximale Abstand gespeichert. Über eine Kostenfunktion wird gewichtet, welche Wichtigkeit Anzahl der Cluster und der Abstand gegeneinander abgewogen besitzen. Je nach Gewichtung heißen die Methoden „k-means“-Clusterbildung, „k-medoids“-Clusterbildung, Clusterbildung durch Mahalanobis Abstand, „Kd-tree“-Clusterbildung, oder „Hidden Markov“-Clusterbildung. Dies sind die Methoden, die MATLAB als Standard für die Clusteranalyse anbietet. Details zu den Verfahren können in [60] nachgelesen werden. Im Rahmen dieser Arbeit werden die Verfahren nur implizit eingesetzt. Da es zunächst keine Anhaltspunkte gibt, welcher Abstand oder welche Anzahl von Clustern zu wählen ist, werden verschiedene Anzahlen von Clustern gebildet. Die Auswertung geschieht aber nicht über einen in MATLAB implementierten und damit wenig nachvollziehbaren Algorithmus. Die Auswertung geschieht per eigenem Skript, sodass die Clusterbildung ausführlicher und nachvollziehbarer stattfinden kann.

### 2.5.6 Form der Cluster

Werden minimale Abstände zwischen Datengruppen gewünscht, ergibt sich je nach Dimension des Raumes, natürlicherweise ein Kreis oder eine Kugel. Nicht alle Cluster müssen die Form eines Kreises (zweidimensional) beziehungsweise im dreidimensionalen Raum einer Kugel haben. Mit Hilfe des Gaussian Mixture Models können abweichende Formen von Clustern erstellt werden. Dafür muss eine erwartete Verteilung durch Mittelwert und Standardabweichung angegeben werden. Um den Mittelwert müssen sich viele Daten befinden. Wird sich vom Mittelwert entfernt, darf der Abstand der Daten zunehmen. Der Abstand der Daten unter sich ist irrelevant. Es muss lediglich eine ausreichende Anzahl entsprechend der Erwartungsverteilung existieren. Die so zu einem Cluster zusammengefassten Daten können abweichend von einem Kreis beziehungsweise einer Kugel sein. Da für die gegebenen Flugtrajektorien kein Erwartungswert und keine Standardabweichung angegeben werden kann, ist auch diese Eigenschaft für die Clusterbildung hier irrelevant.

In diesem Grundlagenkapitel wurden die wählbaren Eigenschaften von Clustern in MATLAB aufgezählt. In Kapitel 3.9 werden diese Eigenschaften aufgegriffen und mit Bezug auf die gegebenen Flugdaten diskutiert, welche Eigenschaften für die Analyse von Robustheit zu wählen sind.

## 2.6 Definition von Robustheit

Robustheit allgemein wird in der Regelungstechnik als Eigenschaft eines Systems verstanden, Störungen gegenüber unsensibel zu sein [24]. Dabei muss für jeden Anwen-

dungsfall genauer definiert werden, was „Störung“ und „unsensibel“ genau bedeutet. Andere Definitionen von Robustheit setzen voraus, dass das zu betrachtende System anfänglich stabil ist. Robust ist ein System, wenn Veränderungen die Stabilität des Systems nicht gefährden können [11]. Für statistische Verfahren wird Robustheit als Eigenschaft definiert, bei Abweichungen von den für die Verfahren notwendigen Voraussetzungen noch hinreichend zuverlässige Ergebnisse zu liefern [26]. Dies wird auch auf die Verarbeitung idealer Daten bezogen, die in der Realität Abweichungen aufweisen. Trotz der idealen Annahmen soll ein robustes System zutreffende Vorhersagen für ein reales System liefern, auch wenn die Annahmen nicht genau erfüllt sind [1]. In der Systemtheorie existiert eine weitere Definition für Robustheit: Ein robustes System fängt fehlerhafte Benutzereingaben ab, verhindert undefinierte Systemzustände und reagiert tolerant [41]. Auch hier muss wie in allen genannten Fällen speziell für jeden Anwendungsfall definiert werden, was die einzelnen Begriffe „fehlerhafte Eingabe“, „undefinierter Zustand“ und „sinnvolle Systemreaktion“ genau meinen.

Die aufgezählten Eigenschaften sind nicht alle auf die vorliegenden Flugtrajektorien anzuwenden. Folgende Eigenschaften können als Indikatoren für Robustheit verstanden werden:

- **Stabilität:** Die verschiedenen Trajektorien eines Datensatzes entstehen durch unterschiedliche Gewichtungen von Kostenoptimal und Umweltoptimal. Diese Gewichtung ist diskret und erzeugt diskrete Trajektorien als Ergebnis. Erwartbar wäre für eine kontinuierliche Gewichtung ein kontinuierlicher Raum, der durch die Ergebnistrajektorien aufgespannt wird. Jedoch sind die Ausgangsdaten für die Berechnung der Trajektorien nicht kontinuierlich, weshalb nur nahezu kontinuierliche Bereiche abgedeckt werden können. Abbildung 2.13 zeigt einen solchen Trajektoriensatz.

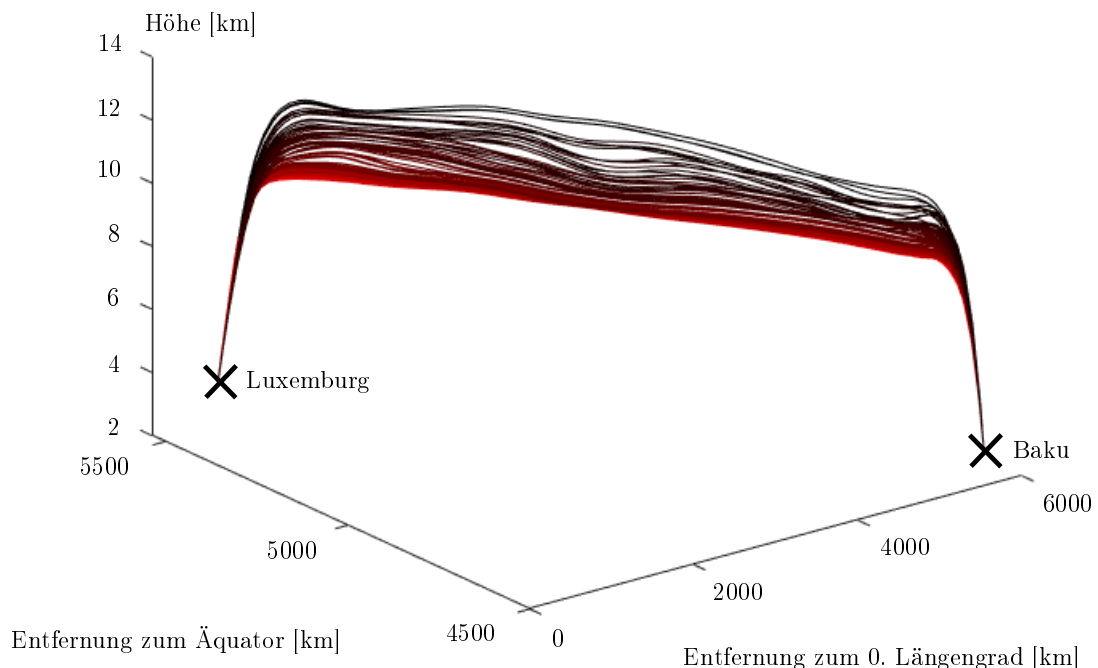


Abbildung 2.13: Trajektoriensatz „00001\_UBBB\_ELLX“ mit ausschließlich stabilen Trajektorien.

Zwar gibt es einige Lücken zwischen den Trajektorien, insgesamt sind aber keine großen Sprünge vorhanden und zu jeder Trajektorie gibt es weitere Trajektorien, die intuitiv als nahe beieinander eingestuft werden können.

Anders verhält sich der Trajektoriensatz in Abbildung 2.14. Hier gibt es einige Trajektorien, die einen anderen Verlauf besitzen als der Rest. Auch wenn die meisten Trajektorien einen etwa zusammenhängenden Bereich abdecken, gibt es sprunghaften Wechsel zwischen einigen Trajektorien. Die vereinzelt Trajektorien sind offensichtlich in ihrem Verhalten nicht stabil und damit nicht robust, denn Stabilität ist nicht gegeben. Entsprechend müssen instabile Trajektorien aussortiert werden. So entstehen Lücken zwischen den Trajektorien. Die übrigen Trajektorien sind stabil oder semi-stabil, das heißt Stabilität ist durch bestimmte Abweichungen gegeben, andere Abweichungen führen zu instabilem Verhalten.

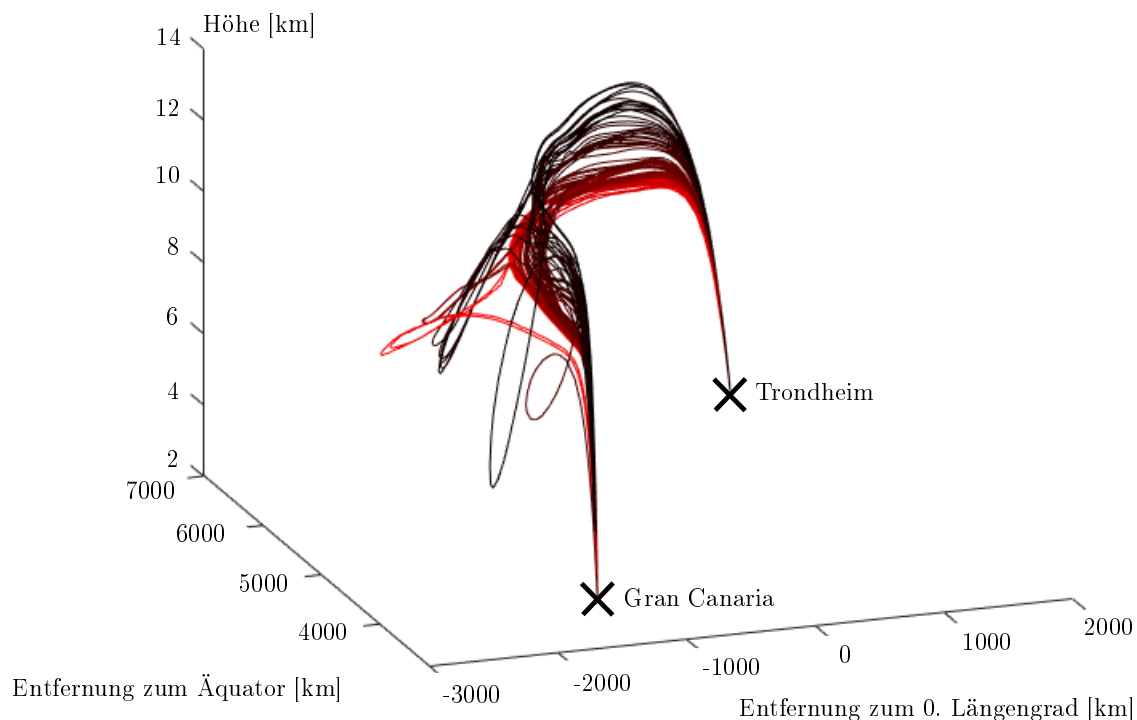


Abbildung 2.14: Trajektoriensatz „00005\_GCLP\_ENVA“ mit einigen instabilen Trajektorien.

Stabilität als Kriterium für Robustheit ist für die vorliegenden Trajektorien anwendbar und ein wichtiges Kriterium. Ist keine Stabilität für den gesamten Trajektoriensatz gegeben, müssen die instabilen Trajektorien gelöscht werden und die so entstehenden semi-stabilen Trajektorien gesondert behandelt werden wie in Abschnitt 3.5 beschrieben. Stabilität ist hier ein notwendiges, aber nicht hinreichendes Kriterium für Robustheit.

- **Störungen:** Bei den gegebenen Flugtrajektorien handelt es sich um ideale, störungsfreie Daten. Störungen sind auf die Abweichung von Idealdaten und dem in der Realität auftretenden Wetter zu beziehen. Das Wetter stellt die konstanten Eingangsdaten dar, auf deren Basis die verschiedenen Trajektorien berechnet werden. Ein robustes System verändert das berechnete Ergebnis nur

wenig, auch wenn sich die Eingangsdaten, also die gegebene Wetterlage, verändern. Damit ist eine geringe Reaktion auf Störungen ein Kernaspekt der hier geforderten Robustheit. Würden sich die Trajektorien durch andere Wetterdaten stark verändern, wären die theoretischen Berechnungen der Trajektorien durch die Projekte ATM4E und FlyATM4E für die Analyse von robusten Trajektorien nicht verwendbar.

- **Idealannahmen:** Die Definition von Robustheit über die Reaktion auf die Störungen entspricht beim betrachteten Sachverhalt dem Punkt der Idealannahmen. Die Störungen bezeichnen die Diskrepanz zwischen Modellannahmen für konstantes Wetter und dem in der Realität eintretenden, sich verändernden Wetter. Entsprechend ist die Annahme von Idealen und eine geringe Abweichung im Ergebnis durch verschiedene Ausgangsdaten genau wie der Punkt der Störungen eine wichtige Voraussetzung, um robuste Trajektorien zu identifizieren.
- **Sensibilität:** Ein unsensibles System steht für ein robustes System. Das heißt, geringe Änderungen der Eingabe beeinflussen das Ergebnis nur wenig. Das Ergebnis sind die Trajektorien. Bei der Stabilität wurde bereits ausgeführt, dass ein (nahezu) kontinuierlicher Raum für die berechneten Trajektorien wünschenswert ist. Dies spiegelt sich auch in der Sensibilität wider. Sprünge in den Trajektorienverläufen stellen große Veränderungen dar. Auch wenn sich die Eingangsdaten nur wenig ändern, hat dies ein völlig anderes Ergebnis zur Folge. Das System reagiert sehr sensibel auf Veränderungen und ist damit nicht robust.  
Die geringe Sensibilität des Systems ist damit ein gutes Kriterium, um Robustheit der Flugtrajektorien zu bemessen. Mit der Sensibilität kann sogar nicht nur klassifiziert werden, ob es sich um ein robustes System handelt, sondern verschiedene Trajektoriensätze oder Ausschnitte der Trajektoriensätze können bezüglich ihrer Robustheit verglichen werden: Wird von der gleichen Anzahl Trajektorien ein kleinerer Ergebnisraum aufgespannt, sind diese Trajektorien robuster als diejenigen, mit denen sie verglichen wurden.
- **Fehlerhafte Benutzereingabe:** Fehlerhafte Eingaben sind im betrachteten Fall kaum möglich. Sind Daten in irgendeiner Weise fehlerhaft, nicht berechenbar oder führen zu keinem Ergebnis, wird dies bereits beim Berechnen der Trajektorien festgestellt. Bei den gegebenen Trajektorien ist immer davon auszugehen, dass sie korrekt berechnet wurden. Entsprechend ist das Abfangen fehlerhafter Eingaben hier kein Kriterium für Robustheit.
- **Undefinierte Zustände:** Wie fehlerhafte Benutzereingaben können auch keine undefinierten Zustände innerhalb des gegebenen Systems entstehen. Die Trajektorien wurden berechnet und haben einen bestimmten Verlauf. In einigen Fällen kann es zu Fehlern während der Trajektorienberechnung gekommen sein und es wurde keine Trajektorie berechnet. Entsprechend gibt es für diese Berechnungen keine Trajektorien als Ergebnis. Nur der Zustand von korrekt berechneten oder nicht vorhandenen Trajektorien ist möglich. Das Abfangen undefinierter Systemzustände ist hier keine Notwendigkeit und damit kein Kriterium für Robustheit.

- **Vorhersehbare Systemreaktion:** Eine sinnvolle Reaktion beziehungsweise Systemausgabe sind Trajektorien, die einen (nahezu) kontinuierlichen Korridor aufspannen. Sprunghaftes Verhalten kann auftreten, ist bezüglich Robustheit aber schwierig verwertbar und wird damit nicht als sinnvolle Systemreaktion eingestuft. Damit ist eine sinnvolle Systemreaktion ein Kriterium für Robustheit. Es ähnelt sehr den Kriterien Stabilität und Sensibilität. Als Charakteristik für sich kann die sinnvolle Systemreaktion aber hier auch als Indiz für Robustheit verstanden werden.

**Zusammenfassung:** Stabilität und die Reaktionen auf Störungen sind hier die Hauptcharakteristiken für Robustheit. Ein nahezu kontinuierlicher Raum von Flugtrajektorien erfüllt diese Kriterien. Zusätzlich kann anhand der Größe des aufgespannten Ergebnisraumes verglichen werden, welcher Trajektoriensatz oder welche Gruppe von Trajektorien robuster sind als andere. Entsprechend sind die Abstände zwischen den Trajektorien eng mit der Robustheit verbunden: Ist der Abstand auf zu vergleichenden Trajektorien durchgängig gering, gibt es Trajektorien in unmittelbarer Umgebung. Die Trajektorien scheinen die benötigten Charakteristiken für Robustheit aufzuweisen.



### 3. Robustheitsanalyse

Im Grundlagenkapitel 2.6 wurde der Zusammenhang zwischen dem geometrischen Abstand der Trajektorien und der Robustheit erklärt. Je mehr Trajektorien sich nahe beieinander befinden, desto robuster sind die jeweiligen Trajektorien. Ob sich Trajektorien nahe beieinander befinden, kann per Ähnlichkeitsberechnung erfolgen. Auch eine Clusteranalyse oder die Definition eines Alphabets und das Vergleichen entsprechender Zeichen charakterisiert, ob Trajektorien sich nahe beieinander befinden.

Zur Berechnung von Robustheit sollen im Rahmen dieser Arbeit insgesamt drei Ansätze die Nähe der Trajektorien berechnen. Die dieser Masterarbeit vorangegangene Projektarbeit [62] berechnet Ähnlichkeit und damit eine Art von Nähe mit durch eine Kombination aus Methode, Metrik und Vergleichsgröße. Die Projektarbeit prüft allgemein, welche Verfahren sich für die Berechnung von Ähnlichkeit eignen und bestimmt 15 potentielle Verfahren. Für diese Verfahren ist nun zu untersuchen, ob und welche von ihnen sich bezüglich der Analyse von Robustheit eignen. Dafür müssen die gegebenen Daten entsprechend aufbereitet werden, was in Abschnitt 3.1 erklärt wird. Die Trajektorien sind verschieden zwischen Kosten- und Umweltoptimal gewichtet. Für die Auswertung ist es hilfreich, die Verteilung der Gewichtung zu kennen, weshalb zur Gewichtung Plots erstellt werden. Das Verfahren wird in Abschnitt 3.2 erklärt. Die Funktionen der ähnlichkeitsberechnenden MATLAB-Skripte werden in Abschnitt 3.3 erklärt. In der Projektarbeit [62] wurden die Skripte in MATLAB nur prototypisch implementiert. Jetzt vorgenommene Verbesserungen werden in Abschnitt 3.4 aufgeführt. Um geeignete Kombinationen aus Methode, Metrik und Vergleichsgröße zu finden, werden Grenzfälle betrachtet und deren Eignung diskutiert. Dies geschieht ausführlich in Abschnitt 3.5. Die Ergebnisse der Ähnlichkeitsberechnung werden in 3.6 erklärt.

In Abschnitt 3.7 wird die Idee eines Alphabets zur Robustheitsanalyse verfolgt. Nach der Erklärung des Ansatzes und dem allgemeinen Vorgehen werden MATLAB-Skripte erstellt und erklärt und die Ergebnisse der Skripte erläutert. In Abschnitt 3.8 folgt eine ausführliche Diskussion von allgemeinen Beispielen und Grenzfällen, die die Stärken und Schwächen der Robustheitsanalyse durch ein Alphabet illustrieren.

Abschnitt 3.9 beschäftigt sich mit der dritten Möglichkeit, die Robustheit und damit verbunden geometrische Nähe zu charakterisieren. Bezüglich des langfristigen Ziels, Empfehlungen für Flugtrajektorien zu erstellen, ist ein geringer geometrischer Abstand für eine robuste Planung relevant. So kann bei sich verändernder Datenlage eine andere Trajektorie gewählt werden, die nahe bei der bereits abgefliegenen Trajektorie liegt. Große Änderungen an der geflogenen Route sind nicht notwendig. Eine solche Berechnung wird mit Hilfe einer Clusteranalyse durchgeführt, auf die ebenfalls in Abschnitt 3.9 eingegangen wird. Auch hier folgen in Abschnitt 3.10 Beispiele, um die Vor- und Nachteile der so berechneten geometrischen Nähe zu cha-

rakterisieren. Auf die Umsetzung der Clusteranalyse in MATLAB wird in Abschnitt 3.11 eingegangen. Ergebnisse und deren Bedeutung folgen in Kapitel 4.

### 3.1 Aufbereitung der Datensätze

Ob es sich um Daten aus dem Projekt ATM4E oder FlyATM4E handelt, ist zunächst nicht relevant. Die Trajektoriendaten müssen in beiden Fällen ähnlich bearbeitet werden. Die Benennung der FlyATM4E Daten weicht von der folgenden Erklärung ab. Das erklärte Verfahren zur Benennung der Daten ist aber auch für FlyATM4E gültig: Alle Dateien müssen die gleiche Benennung mit fortlaufender Nummerierung erhalten.

Für die Verarbeitung in MATLAB wird vorausgesetzt, dass die Ordner in unkomprimierter Form vorliegen. Die Dateien müssen also extrahiert werden, da die Daten als zip-Ordner gespeichert vorliegen. Die Benennung der einzelnen Trajektorien in ATM4E setzt sich wie folgt zusammen:

$$\underbrace{00001}_I - \underbrace{UBBB}_II - \underbrace{ELLX}_III - \underbrace{001}_IV - \underbrace{20171025T235513}_V \quad (3.1)$$

- I: Nummer des Datensatzes; fortlaufende Nummerierung
- II: Startflughafen in ICAO-Code
- III: Zielflughafen in ICAO-Code
- IV: Nummer der Trajektorie; fortlaufende Nummerierung
- V: Erstellungsdatum und Codierung wichtiger Trajektoriendaten

Teil I bis IV folgen einer systematischen Benennung. Während es sich bei I und IV um eine aufsteigende fortlaufende Nummerierung handelt, ist die Benennung von II und III schon vor dem Einlesen der Trajektoriendaten bekannt. Es handelt sich hierbei um die ICAO-Codierung des Start- und Zielflughafens. Da Teil V nicht fortlaufend ist, sondern entsprechend der Ergebnisse benannt ist beziehungsweise ein in diesem Zusammenhang zufälliges Datum erhalten hat, ist es nicht möglich, diesen Teil systematisch in MATLAB einzulesen. Um die Benennung nachzuvollziehen, müssten die Trajektoriendaten bereits geladen sein. Dies ist aber ohne die Angabe des passenden Dateipfades nicht möglich. Es besteht die Herausforderung, dass die Dateien nach ihrem noch nicht bekannten Inhalt benannt sind. Deshalb wird Teil V der Benennung gelöscht. Um dies nicht für jede einzelne Datei durchzuführen, können alle Trajektorien innerhalb eines Ordners ausgewählt werden. Bei der Umbenennung werden IV und V gelöscht. Da die Dateien nun dieselbe Benennung hätten, fügt Windows automatisch eine fortlaufende Nummerierung hinzu, sodass die Benennung nun dieser Form sind:

$$\underbrace{00001}_I - \underbrace{UBBB}_II - \underbrace{ELLX}_III \quad (1) \quad (3.2)$$

Dabei handelt es sich bei der eingeklammerten Zahl am Ende der Benennung um eine fortlaufend aufwärts zählende Nummerierung für alle zugehörigen Daten. Für das MATLAB-Skript `Flugtrajektorien_durchgehen_ATM4E.m` für die Daten aus ATM4E beziehungsweise das Skript `Flugtrajektorien_durchgehen_FlyATM4E.m`

für die Daten aus FlyATM4E sind die umbenannten Dateien so einlesbar. Nach gegebenenfalls einmaligem Anpassen des Speicherverzeichnisses muss für das Laden eines Ordner nur der Wert „Datei“ geändert werden, der der Benennung von Start und Ziel entspricht.

## 3.2 Gewichtung der Kosten- und Umweltparameter

Die Trajektorien entstehen durch verschiedene Gewichtungen von Umwelt- und Kostenfaktoren. Beide Parameter gleichzeitig zu optimieren, ist in der Regel nicht möglich. Es können wahlweise die Kosten minimiert werden oder der Einfluss auf die Umwelt so gering wie möglich gehalten werden. Zwischen beiden Optimalen sind Kompromisse möglich. So können die Kosten für einen Flug gesenkt werden und die negativen Einflüsse auf die Umwelt nur wenig ansteigen. Die verwendete Gewichtung für den Trajektoriensatz „00001\_UBBB\_ELLX“ zeigt Abbildung 3.1.

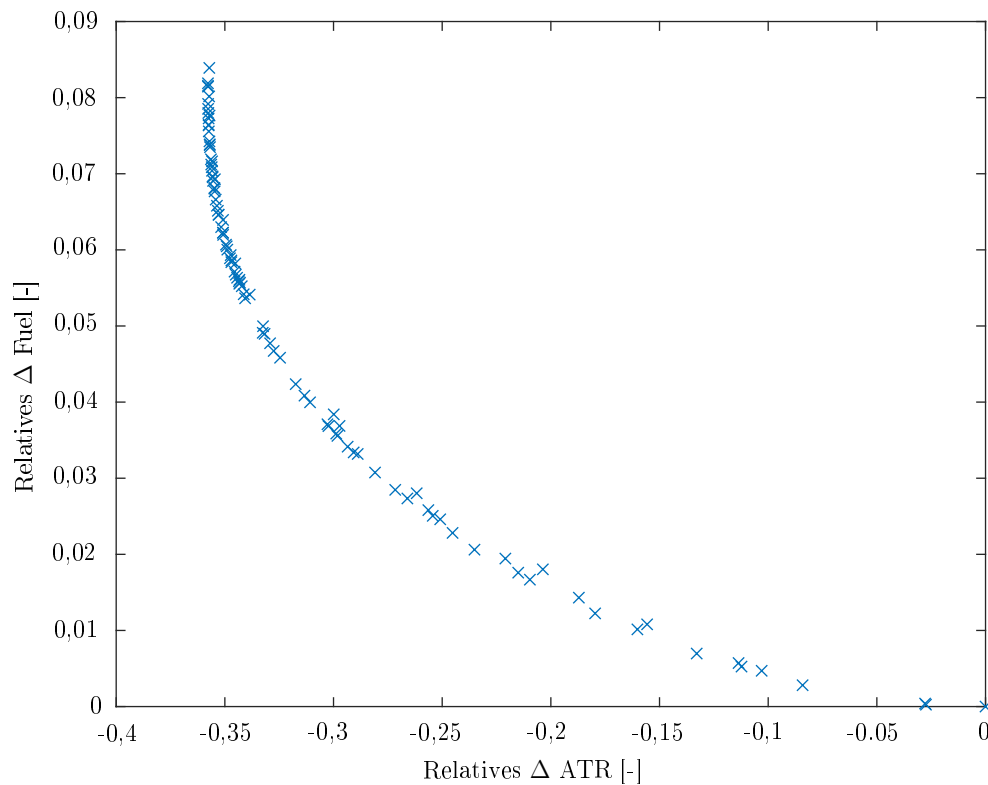


Abbildung 3.1: Gewichtung von Kosten- und Umweltparametern für Datensatz „00008\_EFHK\_GCTS“.

Jedes blaue Kreuz repräsentiert einen Fall, der für die Trajektorienberechnung verwendet wird. Umweltoptimale Trajektorien werden durch die Kreuze links oben markiert. Der Unterschied bezüglich der Kosten ist für diese Trajektorien groß und bezüglich der kostenoptimalen Route um so viel größer, wie die y-Achse angibt. Bezüglich ihrer Umweltauswirkung ist der Einfluss der Trajektorien aber so gering

wie möglich. Entsprechend sind die kostenoptimalen Trajektorien rechts unten zu sehen. Hier sind die Kosten so gering wie möglich, jedoch ist die Umweltauswirkung deutlich größer. Das Einsparpotenzial bezüglich der Umweltauswirkung ist auf der x-Achse angegeben. Hier wird der kostenoptimale Fall als Referenz genommen und alle anderen Flüge mit ihrem Einsparpotenzial ihm gegenüber angegeben.

Dazwischen ergibt sich eine Kurve, die bestmögliche Kompromisse aus beiden Parametern repräsentiert. Punkte links und unterhalb der Kreuze sind in der gegebenen Situation unmöglich zu erreichen. Rechts oberhalb der Kreuze existieren weitere Möglichkeiten. Sie sind aber nicht interessant, da sie optimaler zu gestalten sind. Von Interesse ist lediglich die Grenze der Bereiche und damit alle Fälle, in denen der Kostenparameter nur durch Verschlechterung der Umweltauswirkung beziehungsweise der Umweltparameter nur durch Verschlechterung des Kostenparameters verbessert werden kann. Diese Grenzfälle bilden eine so genannte Pareto-Front. In Abbildung 3.1 ergeben die Kreuze in etwa einen hyperbelähnlichen Verlauf, was typisch für Pareto-Kurven ist.

Die Kreuze weisen in Abbildung 3.1 eine Anhäufung bei den umweltoptimalen Trajektorien auf. Die Abstände zwischen den Kreuzen der kostenoptimalen Trajektorien sind deutlich größer. Dies beeinflusst die Häufung der Trajektorien. Werden zu vielen Gewichtungungen mit geringen Änderungen Trajektorien erstellt, verändern sich diese Trajektorien erwartungsgemäß wenig. Ist die Variation der Gewichtungsfaktoren größer, sind auch die dazugehörigen Trajektorien verschiedener. Inwieweit die Verteilung der Kreuze Auswirkungen auf das Ergebnis hat, wird in 4.1.1 diskutiert.

### 3.3 Funktion der MATLAB-Skripte für die Berechnung von Ähnlichkeit

Insgesamt soll die Ähnlichkeit der Trajektorien über eine Clusteranalyse und Ähnlichkeitsberechnungen wie in [62] bestimmt werden. Prototypische Skripte zur Berechnung von Ähnlichkeit wurden bereits im Rahmen der Projektarbeit [62] erstellt und bilden die Grundlage für die weiteren Skripte. Hier wird auf die Ähnlichkeitsberechnung und ihre Verbesserungen eingegangen. Die Erklärungen zur Clusteranalyse folgen in 3.9.

Gegenüber der Projektarbeit [62] hat sich die Struktur der Programme und Funktionen in MATLAB nicht verändert. Die Funktionen wurden überarbeitet, verbessert und in einem Hauptskript zusammengefasst. Abbildung 3.2 zeigt das Ablaufdiagramm, das dem Ablaufdiagramm der Projektarbeit [62] sehr ähnlich ist, aber kleine Änderungen erfahren hat.

Auf die Verbesserungen wird im Detail in Abschnitt 3.4 eingegangen. Die Skripte `Flugtrajektorien_durchgehen_ATM4E.m` für die Daten aus ATM4E beziehungsweise `Flugtrajektorien_durchgehen_FlyATM4E.m` für die Daten aus FlyATM4E startet die Ähnlichkeitsberechnung. Es sind zwei verschiedene Skripte erforderlich, da für ATM4E nur ein Datensatz bei unveränderlichem Wetter geladen wird. Das Skript für FlyATM4E lädt weitere Parameter der gegebenen Wetterdaten und unterscheidet zwischen bis zu zehn verschiedenen Datensätzen. Die notwendige Unterscheidung von Daten ist für FlyATM4E zwingend erforderlich, für ATM4E wird durch das komplexere Abspeichern von mehr Daten die Laufzeit unnötig verlängert. Die weiteren Teilfunktionen, die beide Skripte besitzen, sind:

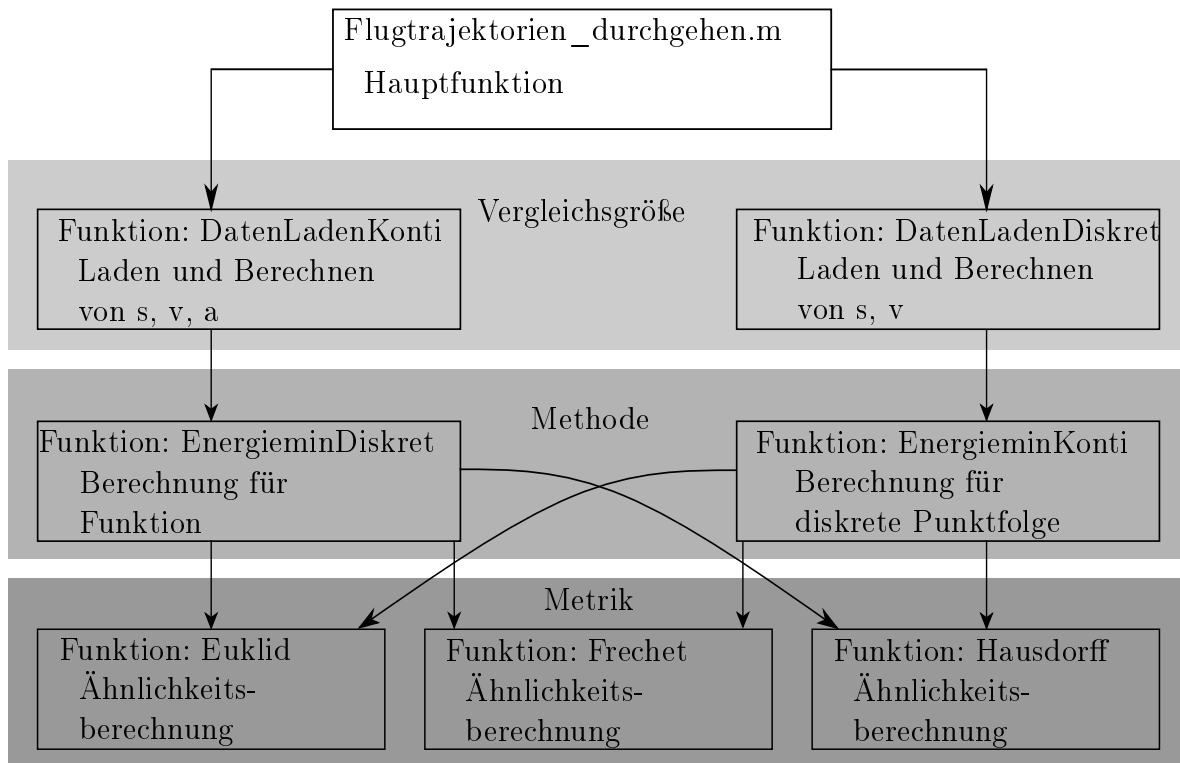


Abbildung 3.2: Verknüpfung der Funktionen für Ähnlichkeitsberechnung innerhalb von MATLAB.

- Vergleichsgröße

- Funktion: `DatenLadenDiskret`

**Eingabe:** Der Funktion wird das Verzeichnis des Datensatzes, der geladen werden soll, übergeben. Dem Gesamtskript wurde bereits das Verzeichnis übergeben und dieses gibt die Information an die Funktion weiter, sodass die Funktion nur indirekt einer Eingabe bedarf.

**Verarbeitung:** In dem entsprechenden Verzeichnis, das der Funktion übergeben wurde, existiert pro Ordner darin ein MATLAB-File namens `Output.m`. Sollte dieser Ordner nicht existieren, werden alle benötigten Daten anstelle der auszulesenden Werte mit Nullen aufgefüllt. Innerhalb des MATLAB-Files `Output.m` gibt es viele Datensätze. Relevant sind hier `Output.states.H`, `Output.states.phi` und `Output.states.lambda` sowie `Output.time`. Für Daten von FlyATM4E wird zusätzlich das Verzeichnis abgespeichert, in dem sich die ausgelesene Datei befindet. So ist später zu unterscheiden, zu welcher Wetterlage die Daten gehören. In genannter Reihenfolge wird durch die Daten die Flughöhe, der Längengrad, der Breitengrad und die auf dem gesamten Flug verstrichene Zeit angegeben. Das Auslesen dieser Daten muss für jede Trajektorie einzeln erfolgen. Die ausgelesenen Daten werden in ein gemeinsames Array gespeichert. Zur Visualisierung der Daten wird innerhalb dieser Funktion jede Trajektorie als Punkte geplottet. Trajektorien, bei denen die Berechnung abgebrochen wurde oder Trajektorien mit sehr abweichendem Streckenverlauf verglichen mit den anderen Trajektorien können so be-

reits erkannt werden. Zur besseren Unterscheidung der Trajektorien gibt es innerhalb des Plots einen Farbverlauf: Die kostenoptimale Route ist schwarz, die umweltoptimale Trajektorie rot. Gewichtungen dazwischen sind eine entsprechende Mischung der Farben rot und schwarz. Da die Trajektorien unterschiedlich viele Datenpunkte besitzen, müssen die Trajektorien mit weniger Punkten auf die Länge des längsten Datensatzes gebracht werden. Dies geschieht durch das Auffüllen mit weiteren Nullen am Ende der gegebenen Daten.

**Ausgabe:** Alle Daten, die aus verschiedenen Unterordnern herauskopiert wurden, wurden in ein einziges Array gespeichert. Dieses beinhaltet für jede Trajektorie einen vollständigen Satz aus drei Raumkoordinaten und einer jeweils zugehörigen Zeit. Für FlyATM4E werden Verzeichnisnamen ebenfalls abgespeichert.

– Funktion: `DatenLadenKonti`

**Eingabe:** Der Funktion wird das Verzeichnis des Datensatzes, der geladen werden soll, übergeben. Dem Gesamtskript wurde bereits das Verzeichnis übergeben und dieses gibt die Information an die Funktion weiter, sodass die Funktion nur indirekt einer Eingabe bedarf.

**Verarbeitung:** Wie bei der Funktion `DatenLadenDiskret` werden alle benötigten Daten aus `Output.m` ausgelesen, als Punkte geplottet und in ein Gesamtarray gespeichert. Der Unterschied der Skripte besteht auch hier darin, dass für FlyATM4E zusätzlich Verzeichnisbenennungen gespeichert werden. Der Aufbau des Skriptes im aktuellen Zustand würde zulassen, die Funktion `DatenLadenDiskret` zuerst auszuführen und die Ausgabe an diese Funktion zu übergeben. Für weitere Anwendungen sollen die Funktionen gegebenenfalls separat verwendet werden. Daher berechnen beide Funktionen dasselbe Ergebnis, um auch getrennt voneinander verwendbar zu sein.

Um im Raum dreidimensionale kontinuierliche Trajektorien aus den gegebenen Punkten zu erzeugen, wird die MATLAB-Funktion `fitobject` verwendet. Für den dreidimensionalen Raum ist eine Näherung durch Polynom vom Grad eins bis neun wählbar. Welcher Grad zu verwenden ist, wird durch die in MATLAB integrierte Funktion frei entschieden. Es wird je die Interpolation durch ein Polynom nächsthöheren Grades probiert, bis die Abweichung des nächsthöheren Polynoms gering ist. Das Polynom niedrigeren Grades wird verwendet. Die Abweichung ist gering, wenn das Flächenintegral der beiden Polynome sich um weniger als 10 Prozent unterscheidet [61]. Durch die Taylor-Reihenentwicklung ist gegeben, dass alle Polynome höheren Grades dieselbe oder eine geringere Abweichung aufweisen, sodass Polynome höheren Grades nicht zu testen sind.

Zur Überprüfbarkeit der Interpolation werden die Funktionen ebenfalls geplottet. Auch sie erhalten wie die einzelnen Punkte einen Farbverlauf zur besseren Unterscheidung. Rot stellt die umweltoptimale Trajektorie dar, schwarz die kostenoptimale. Die Mischung der Farben steht für eine entsprechende Gewichtung der beiden Faktoren Kosten und Umwelt.

**Ausgabe:** Aus den Datenpunkten wurden Funktionen interpoliert. Die Funktionen werden übergeben, indem die Koeffizienten der Polynome

in ein Array gespeichert werden und entsprechend eine Aufschlüsselung übergeben wird, dass das gespeicherte Array Koeffizienten einer Funktion erhält. Für die weitere Verwendung müssen die Koeffizienten erneut in die Funktion eingesetzt werden und der gewünschte Abschnitt neu berechnet werden. Zwar wird so die Berechnungszeit durch das mehrfache Berechnen der Funktion verlängert, durch das Übergeben der Koeffizienten anstelle einer Funktion wird aber Speicherplatz eingespart. Da es ebenfalls viel Zeit kostet, auf einen großen Speicher zuzugreifen, gleicht dies etwa die mehrfache Berechnung der Funktion aus.

- Methode

- Funktion: **EnergieminDiskret**

**Eingabe:** Der Funktion werden alle Trajektorien mit ihren Datenpunkten durch ein Array übergeben. Für einen Trajektorienpunkt liegen drei Koordinaten für die Position im Raum und eine zugehörige Zeitangabe vor.

**Verarbeitung:** Die Differenz der ersten beiden Komponenten wird berechnet und abgespeichert. Durch die Differenz werden entsprechend der Energieminimierungsmethode alle weiteren Punkte verändert. Danach wird die Differenz des zweiten und dritten Punktes berechnet, abgespeichert und das übrige Array entsprechend verändert. Die Berechnung setzt sich fort, bis der letzte Datenpunkt der Trajektorie erreicht ist. Die Berechnung der Differenzen erfolgt pro Raumrichtung.

**Ausgabe:** Pro Raumrichtung wird ein Array mit Differenzen der jeweiligen Punkt übergeben. Die Differenzen entsprechen nicht dem geometrischen Abstand, sondern sind durch diskrete Energieminimierung berechnet worden.

- Funktion: **EnergieminKonti**

**Eingabe:** Hier werden sowohl die einzelnen Trajektorienpunkte mit allen vier Daten als auch die Koeffizienten einer Funktion, die die im Raum dreidimensionale Trajektorie repräsentiert, übergeben. Zu den Koeffizienten gehört ebenfalls eine Codierung, wie aus den Koeffizienten eine Funktion zu bilden ist.

**Verarbeitung:** Durch die Ableitung der Funktion und bekannte Zeitpunkte auf der Trajektorie lassen sich die gemittelte Geschwindigkeit zwischen den jeweiligen Datenpunkten bestimmen. Die Bestimmung der gemittelten Beschleunigung erfolgt über die Ableitung der Geschwindigkeit und dieselben bekannten Zeitpunkte. Für den Vergleich von zwei Trajektorien wird die Differenzfunktion der beiden die Trajektorie repräsentierenden Funktionen verwendet. Dies geschieht separat für die Funktionen von Strecke, Geschwindigkeit und Beschleunigung. Mit Hilfe dieser Differenzfunktion wird zwischen zwei Punkten mit Zeitangabe jeweils das Flächenintegral für jede Raumrichtung berechnet. Die Fläche ist ein Maß für aufzubringende Energie. Den Trajektorien wird ein willkürlicher Materialwert zugewiesen. Per Methode der Finiten Elemente werden die Trajektorien mit Materialwert, die man sich wie zwei Drähte vorstellen kann, so verbogen, dass sie deckungsgleich werden. Die benötigte Energie für diesen Vorgang wird berechnet und durch den willkürlich gewählten

Materialwert geteilt. So ist das Ergebnis unabhängig von den Materialwerten. Eine hohe Steifigkeit führt zu einer großen Verformungsenergie. Der Quotient bleibt (zumindest für die angenommene kleine Verformung) konstant. Zum Berechnungsvorgang gehört wie bei der diskreten Methode, dass die Veränderung eines Abschnittes Auswirkungen auf den Rest hat. Auch hier können sich die Trajektorien wie ein Draht vorgestellt werden: Wird ein Abschnitt verbogen, so verändert sich Position und Ausrichtung der weiteren Elemente.

**Ausgabe:** Pro Raumrichtung wird eine benötigte Menge Energie, die für deckungsgleiche Trajektorien aufzubringen ist, berechnet. Um einen konstanten, unabhängig vom willkürlich gewählten Materialwert Wert zu erhalten, wird die Energie durch die Steifigkeit des Materials geteilt. Die Energien werden pro Abschnitt in einem Array gespeichert. Da die Berechnung für Strecke, Geschwindigkeit und Beschleunigung durchgeführt werden, berechnet die Funktion drei Arrays, deren Inhalt abschnittsweise die Energien sind.

- Metrik

- Funktion: **Euklid**

**Eingabe:** Der Metrik müssen Differenzen pro Raumrichtung übergeben werden. Die Einheit der Differenzen ist dabei egal, sodass sowohl die diskret berechneten Abstände als auch die durch die kontinuierliche Energieminimierung berechneten Energien übergeben werden können. Ob die Differenzen bezüglich Strecke, Geschwindigkeit oder Beschleunigung berechnet wurden, spielt ebenfalls keine Rolle. Eine Zeitabhängigkeit wird durch die Euklidische Metrik nicht direkt berücksichtigt. Indirekt kann die Zeit durch die zuvor durch die Energieminimierung berechneten Differenzen berücksichtigt werden. Um Funktionsaufrufe zu sparen, werden der Funktion alle Trajektorien übergeben und nicht nur Trajektorienpaare.

**Verarbeitung:** Für das Paar zu vergleichender Trajektorien wird von beiden Trajektorien der Wert für die erste Differenz verwendet. Entsprechend der Euklidischen Metrik werden beide Werte quadriert, zueinander addiert und die positive Wurzel von ihnen genommen. So werden die Daten Wert für Wert durchgegangen. Sollten die Datensätze unterschiedlich lang sein, wird das Verfahren beim Erreichen des Endes des kürzeren Datensatzes gestoppt. Weitere Werte des längeren Datensatzes werden nicht berücksichtigt. Alle durch die Euklidische Metrik berechneten Werte werden am Ende addiert. Die Summe ist die Maßzahl für Ähnlichkeit. Die Trajektorien werden paarweise verglichen. Das Ergebnis wird in einem Array festgehalten. Die Position im Array gibt an, zu welchen Trajektorien der Ähnlichkeitswert gehört. Vergleiche derselben Trajektorie werden nicht berechnet. Das Array wird an der entsprechenden Stelle mit Nullen aufgefüllt. (Alternativ kann die Ähnlichkeit zweier identischer Trajektorien auch berechnet werden. Das Ergebnis lautet ebenfalls Null.) Genauso werden vertauschte Trajektorienpaare nicht berechnet. Das Ergebnis der Vergleichs von Trajektorie 1 und 2 ist dasselbe wie von Trajektorie 2 und 1, da Metriken immer symmetrisch sind. Das Ergebnisarray wird



daher nur oberhalb der Diagonalen mit berechneten Ähnlichkeitswerten gefüllt. Der Rest ist mit Nullen besetzt.

**Ausgabe:** Ein Array enthält eine Maßzahl für Ähnlichkeit für jeden Vergleich von einem Trajektorienpaar. Die Gestalt des Arrays ist eine linke obere Dreiecksmatrix, da Vergleiche von denselben Trajektorien in anderer Reihenfolge nicht berechnet werden.

– Funktion: **Hausdorff**

**Eingabe:** Die durch die Energieminimierung berechneten Differenzen sowie die Trajektorienverläufe werden der Funktion übergeben. Auch hier ist, wie bei der Euklidischen Metrik, nicht relevant, ob die Differenzen durch die diskrete oder kontinuierliche Energieminimierung entstanden sind und bezüglich welcher Größe (Strecke, Geschwindigkeit, Beschleunigung) sie berechnet wurden. Außerdem wird die Zeitabhängigkeit nicht direkt berücksichtigt.

**Verarbeitung:** Innerhalb des kürzeren Datensatzes wird für jeden Trajektorienpunkt bestimmt, welcher Punkt des anderen Datensatzes am nächstgelegenen ist. Dies geschieht durch eine Kugel, deren Mittelpunkt der zu untersuchende Trajektorienpunkt ist. Die Kugel wird vergrößert, bis sich innerhalb von ihr ein Punkt des anderen Datensatzes befindet. Der Radius der benötigten Kugel wird gespeichert. Punkte können mehrfach als nächstgelegener Punkt ausgewählt werden. Die gespeicherten Radien dienen als Gewichtung für die übergebenen Differenzen. Wird eine große Kugel benötigt, um den nächstgelegenen Punkt zu erreichen, wird dieser Abschnitt stark gewichtet. Jede Differenz wird mit ihrer Gewichtung multipliziert und anschließend alle Produkte addiert. Die Summe ist ein Maß für Ähnlichkeit.

**Ausgabe:** Wie bei der Funktion **Euklid** werden in einem großen Array alle Werte für Ähnlichkeit abgespeichert. Auch hier entsteht eine linke obere Dreiecksmatrix, da Ähnlichkeiten nicht mehrfach berechnet werden.

– Funktion: **Frechet**

**Eingabe:** Die durch die Energieminimierung berechneten Differenzen und die räumlichen Trajektorienpunkte mit ihren jeweiligen Zeiten werden der Funktion übergeben.

**Verarbeitung:** Die Fréchet Metrik verfährt ähnlich zur Hausdorff Metrik. Auch hier werden um die Punkte des kürzeren Datensatzes Kugeln gebildet, bis sich ein Punkt innerhalb einer Kugel befindet. Anders als bei der Hausdorff Metrik stehen hier aber nicht alle Punkte zur Verfügung. Die möglichen Punkte werden durch eine maximale Zeitdifferenz, die zum betrachteten Punkt zulässig ist, eingeschränkt. Nur in der Fréchet Metrik wird die Zeitabhängigkeit dadurch direkt berücksichtigt.

**Ausgabe:** Wie bei den anderen Metriken werden die Zahlen für Ähnlichkeit in einem großen Array gespeichert. Die Gestalt ist dieselbe.

Anmerkung zu den Metrik-Funktionen: Die Funktionen für Metriken **Euklid**, **Frechet** und **Hausdorff** haben ein Array als Ergebnis. Um die Vielzahl von Ähnlichkeitsmaßen auszuwerten, wird eine Clusteranalyse durchgeführt. Dafür werden die Ähnlichkeiten innerhalb der Arrays als dreidimensional aufgefasst. Zeile und Spalte des

Arrays, an dem sich die Zahl befindet, werden als zwei Raumkoordinaten aufgefasst. Eine dritte Koordinate ergibt sich durch den Wert für Ähnlichkeit. Innerhalb des so entstandenen dreidimensionalen Raums werden die Ähnlichkeiten durch eine Clusteranalyse in Gruppen aufgeteilt. Zu allen Ähnlichkeiten innerhalb einer Gruppe werden die jeweils beiden zugehörigen Trajektorien geplottet und so veranschaulicht, welche Trajektorien laut der Ähnlichkeitsberechnung als nahe beieinander und somit robust einzustufen sind.

## 3.4 Verbesserung der ähnlichkeitsberechnenden MATLAB-Skripte

Im Rahmen einer vorausgegangenen Projektarbeit [62] wurden in MATLAB prototypische Verfahren implementiert, um geeignete Metriken zu testen. Als Ausgangspunkt wurden die MATLAB-Skripte im Anhang der Projektarbeit [62] verwendet. Änderungen und Verbesserungen, die im Rahmen dieser Arbeit vorgenommen wurden, beinhalten die in 3.4.1 beschriebene Verbesserung der Laufzeit. Dazu gehören ein effizientes Laden der Daten, auf das in 3.4.3 eingegangen wird und eine schnellere Berechnung der Energieminimierung und der Metriken, die in 3.4.4 und 3.4.5 erklärt werden. Trajektorien können nun allgemeingültig dargestellt werden, was in 3.4.6 ausgeführt wird. Das Interpolationsverfahren wurde aufgrund der Trajektorienverläufe geändert, was in 3.4.7 beschrieben wird. Viele neue Funktionen der Skripte, die in MATLAB verfügbar sind, wurden nicht benutzt, sondern stattdessen selbst programmiert. Gründe dafür und welche Funktionen bewusst nicht verwendet wurden, werden in 3.4.8 genannt.

### 3.4.1 Vektorisierung

Dem Skript der Projektarbeit [62] `Datenauswertung.m` wurde eine Zeitmessung per MATLAB-Befehl `tic toc` hinzugefügt. Das veränderte Skript `DatenauswertungPA.m` ist im Anhang beigelegt. So ergibt sich beim Laden und Auswerten einer Trajektorie eine Berechnungszeit von knapp 0,6 s. Je nach Rechner und parallel ausgeführten Programmen kann diese Zeit variieren. Das Skript der Projektarbeit lädt und verarbeitet die Daten für eine Trajektorie. Da ein Datensatz meistens aus 101 Trajektorien besteht, wurde für die Berechnungszeit eines gesamten Trajektoriensatzes  $0,6\text{ s} \cdot 101 = 60,6\text{ s}$  angenommen.

Die Überarbeitung und Erweiterung des Skriptes `Datenauswertung.m` ist das Skript `Flugtrajektorien_durchgehen_ATM4E.m` für die Daten des Projektes ATM4E beziehungsweise `Flugtrajektorien_durchgehen_FlyATM4E.m` aus FlyATM4E. Diese Skripte laden und verarbeitet die Daten eines gesamten Trajektoriensatzes. Das heißt, mit einem Aufruf werden alle 101 Trajektorien miteinander verglichen. Eine Zeitmessung per MATLAB-Befehl `tic toc` ergibt hier knapp 23 s für einen Durchlauf.

Damit wurde die Berechnungszeit auf ein Drittel reduziert. Grund dafür ist, dass viele Vergleiche nicht berechnet werden müssen. Während die erste Trajektorie mit allen anderen verglichen werden muss, kann bei der zweiten Trajektorie auf den Vergleich mit der ersten verzichtet werden. Das Ergebnis ist dasselbe, wie bereits im

Vergleich von der ersten Trajektorie mit der zweiten. Entsprechend kann bei der dritten Trajektorie auf den Vergleich mit den ersten beiden verzichtet werden. Für die letzte Trajektorie muss keine Berechnung mehr durchgeführt werden, da sie bereits mit allen anderen Trajektorien verglichen wurde. So wird etwa die Hälfte aller notwendigen Berechnungen und damit die Hälfte der Rechenzeit eingespart. Zudem werden die Trajektorien nicht mehr mit sich selbst verglichen. Diese Option kann im MATLAB-Skript gewählt werden, um die korrekte Funktion der Skripte zu überprüfen. Identische Trajektorien müssen das Ergebnis Null liefern. Bei der eigentlichen Berechnung werden diese Vergleiche aber weggelassen, um Rechenzeit zu sparen. Das Weglassen von gut der Hälfte der sonst gemachten Berechnungen ergibt eine Berechnungszeit von etwa 23 s.

### 3.4.2 Auswertung mehrerer Trajektorien

In der Projektarbeit [62] wurde bei jedem Start des Matlab-Skriptes eine Trajektorie geladen oder gespeichert. Je nach Wunsch des Anwenders konnte eine weitere Trajektorie geladen werden und die beiden Trajektorien miteinander verglichen. In der überarbeiteten Variante wird ein gesamter Datensatz von Trajektorien verwendet und die Ähnlichkeit von allen Paarungen berechnet. Während es in der Projektarbeit noch um das Verfahren und dessen Überprüfung selbst ging, wird jetzt die Auswertung eines ganzen Ensembles möglich.

### 3.4.3 Effizientes Laden der Daten

In der unbearbeiteten Form gibt es zu jeder Trajektorie eine Datei `Output.m`, die alle Daten zur Trajektorie enthält. Davon wird nur ein Bruchteil benötigt. Da MATLAB für die Berechnung mit großen Datenmengen optimiert wurde, werden die jeweils benötigten Variablen aus den zu den Trajektorien gehörigen `Output.m`-Files ausgelesen und in ein Array gespeichert. Die Zugriffszeit für die Variablen kann so reduziert werden und auch Berechnungen laufen effizienter ab.

Eine Herausforderung stellt dabei die unterschiedliche Länge der Trajektorien dar. Die Anzahl von Datenpunkten variiert. Das Array kann während des Beschreibens mit Daten eine beliebige Form annehmen, muss zur Übergabe an weitere Funktionen aber eine rechteckige Form annehmen. Daher ist es erforderlich, dass die kürzeren Trajektorien einen Ausgleich erhalten, um das Array korrekt zu formatieren. Dies geschieht hier, indem nach dem Ende der Datenpunkte weiter mit Nullen aufgefüllt wird und so alle Trajektorien die Dimension der längsten Trajektorie annehmen. Das Auffüllen geschieht am Ende der Funktion, da erst dann die Größe des längsten Datensatzes bekannt ist.

Für die weitere Verarbeitung dürfen die hinzugefügten Nullen nicht mit den gegebenen Datenpunkten verwechselt werden. Die Funktionen zur Energieminimierung und die Funktionen der Metriken besitzen daher eine Teilfunktion, die Daten mit der Position und gleichzeitig dem Zeitstempel Null herausfiltert. Obwohl die Filterung der zusätzlich hinzugefügten Daten einen Mehraufwand bedeutet, wird durch die Verwendung eines einzigen Arrays für alle Trajektorien dennoch Rechenzeit eingespart.

### 3.4.4 Berechnung durch Energieminimierung

Die beiden Funktionen `EnergieminDiskret` und `EnergieminKonti` nutzen jeweils dynamische Programmierung, das heißt das Problem wird in Teilprobleme zerlegt und Zwischenergebnisse systematisch gespeichert [5]. Konkret wird dynamische Programmierung benutzt, um die minimale Energie für die jeweiligen Trajektorien zu finden.

Dazu müssen die Trajektorien für die diskrete Energieminimierung skaliert werden. Sie besitzen in der Regel nicht dieselbe Länge. Bei der diskreten Energieminimierung wird die Trajektorie in Abschnitte aufgeteilt. Eine längere Trajektorie enthält entsprechend mehr Abschnitte. Damit sind die beiden zu vergleichenden Trajektorien nicht in Deckungsgleichheit zu bringen. Daher wird die längere Trajektorie in allen drei Raumdimensionen gleichmäßig gestaucht, um die Trajektorien auf die gleiche Länge zu bringen und sie erst danach in gleich viele Abschnitte aufzuteilen. Alternativ könnte die kürzere Trajektorie gestreckt werden. Willkürlich wurde sich für die Stauchung entschieden. Angenommen wird dabei, dass die Trajektorien selbst sehr lang sind und die Stauchung nur einen geringen Unterschied macht. Damit ist die Stauchung in weiteren Berechnungen vernachlässigbar klein und muss nicht weiter betrachtet werden.

Anschließend ist die minimale Anzahl und Verschiebungsgröße von zu verändernden Elementen zu bestimmen, sodass die beiden zu vergleichenden Trajektorien deckungsgleich werden. Die Veränderung eines Elementes beeinflusst alle folgenden Elemente, weshalb kein allgemeingültiges System gefunden werden kann, nach dem die beste Kombination von Änderungen zu bestimmen ist [5]. Zunächst werden die benötigten Verschiebungen berechnet, wenn die Elemente in ihrer Reihenfolge unverändert von Start bis Ziel durchgegangen werden. Dieser Wert dient als Referenz. Anschließend werden alle anderen Kombinationen berechnet. Hierbei wird dynamische Programmierung verwendet. Teile von Kombinationen werden gespeichert und für andere Kombinationen erneut verwendet. Ist eine Kombination besser als der Referenzwert, wird der Referenzwert durch sie ersetzt. Ist eine gewählte Reihenfolge bereits während der Berechnung schlechter als der Referenzwert, so wird sie verworfen. Nach diesen Berechnungen ist ein Minimum gefunden.

Bei der kontinuierlichen Variante des Energieminimierungsverfahrens ist keine gesonderte Skalierung notwendig. Der Trajektorie wird ein fiktiver Materialwert zugewiesen, sodass sich das Anpassen der Trajektorien wie das Verbiegen eines Drahtes vorgestellt werden kann. Ohne diesen Wert können keine Berechnungen durchgeführt werden. In einem gewissen Maße ist das Material stauch- und dehnbar, sodass die Länge der Trajektorien so angepasst werden können. Auch hier gilt die Annahme, dass die Verzerrung im Verhältnis zur Gesamtlänge sehr klein bleibt und die zusätzlich benötigte Verformungsenergie das Ergebnis nicht wesentlich verändert.

Per Methode der Finiten Elemente wird berechnet, welche Energiemenge es benötigen würde, um die Trajektorien in Deckungsgleichheit zu bringen. Da auch hier die Veränderung eines Segments andere Segmente beeinflusst, wird erneut dynamische Programmierung verwendet. Wie bei der diskreten Energieminimierung wird ein Referenzwert berechnet, indem alle Elemente in ihrer Reihenfolge vom Start bis zum Ziel durchgegangen werden. Abschnitte werden abgespeichert und für weitere Berechnungen erneut genutzt. Als Abschnitt gelten alle Elemente, die nach Veränderung eines Elementes keiner weiteren Veränderung bedürfen. Ein geringerer Wert der benötigten Energie ersetzt den Referenzwert. Berechnungen, die währenddes-

sen mehr Energie benötigen als der Referenzwert, werden verworfen. Der gefundene minimale Wert für die benötigte Energie wird durch den willkürlich festgelegten Materialwert für Steifigkeit geteilt. So ist das Ergebnis nicht mehr abhängig von der Wahl. Ein steifes Material benötigt eine große Menge Energie zum Verbiegen, besitzt aber gleichzeitig einen großen Wert für Steifigkeit. Weniger Energie wird für ein weniger steifes Material benötigt, der Materialwert ist entsprechend kleiner und damit der Quotient aus Energie und Materialwert konstant.

### 3.4.5 Berechnung durch Metriken

Die Berechnung, die in den Funktionen `Euklid`, `Frechet` und `Hausdorff` durchgeführt wird, ist durch die Verwendung eines Arrays, in dem alle Trajektoriendaten abgespeichert sind, deutlich effizienter geworden. In der Projektarbeit [62] wird für einen Vergleich von zwei Trajektorien durch die Euklidische Metrik etwa eine Zeit von  $5 \cdot 10^{-5}$  s gebraucht. Die überarbeiteten MATLAB-Skripte berechnen mit einem Aufruf alle Trajektorienvergleiche. Wie in 3.4.1 beschrieben, werden jedoch nicht alle Kombinationen miteinander verglichen. Für einen Satz aus  $n$  Trajektorien ergeben sich

$$m = \underbrace{n^2}_{\text{I}} - \underbrace{\frac{n^2}{2}}_{\text{II}} - \underbrace{\frac{n}{2}}_{\text{III}} \quad (3.3)$$

I: Gesamtzahl an möglichen Vergleichen.

II: Gleiche Kombinationen in anderer Reihenfolge werden nicht berechnet.

III: Trajektorien werden nicht mit sich selbst verglichen. Die Hälfte der Diagonalen ist bereits durch II abgezogen worden.

Paare, die bezüglich ihrer Ähnlichkeit berechnet werden. Für ein Beispiel von  $n = 101$  Trajektorien (was den meisten Datensatzlängen in ATM4E entspricht) werden

$$m_{101} = 101^2 - \frac{101^2}{2} - \frac{101}{2} = 5050 \quad (3.4)$$

Vergleiche durchgeführt. Entsprechend würde das Skript aus der Projektarbeit [62]  $5050 \cdot 10^{-5} = 0,2525$  s benötigen. Die neue Funktion `Euklid` wird in einem Skriptdurchlauf fünffach aufgerufen und braucht im Durchschnitt etwa 0,0285 s, abhängig vom verwendeten Rechner und parallel laufender Prozesse. Grund für die Verkürzung der Berechnungszeit auf etwa ein Zehntel ist, dass im neuen Skript die Trajektorienpunkte nicht einzeln aufgerufen werden, sondern bereits in einem einzigen Array abgespeichert sind. Da MATLAB auf die Berechnung mit Matrizen und mehrdimensionalen Arrays optimiert ist, erfolgt die Berechnung so effizienter als der Aufruf einzelner Daten über `for`-Schleifen.

### 3.4.6 Darstellung der Trajektorien

Das Koordinatensystem, in dem sich die Trajektorien befinden durften, war in der Projektarbeit [62] eingeschränkt. Nur Trajektorien, die nördlich des Äquators sowie östlich des 0. Längengrades verlaufen, konnten richtig ausgewertet werden. Das

Koordinatensystem ist nun für beliebige Flüge gültig. Abbildung 3.3 zeigt den Datensatz „00115\_EKYT\_GCLP“, also einen Flug von Aalborg nach Gran Canaria. Dabei wird der 0. Längengrad überschritten. Entfernungen westlich des 0. Längengrades werden negativ angegeben. Genauso werden Positionen südlich des Äquators negativ gerechnet.

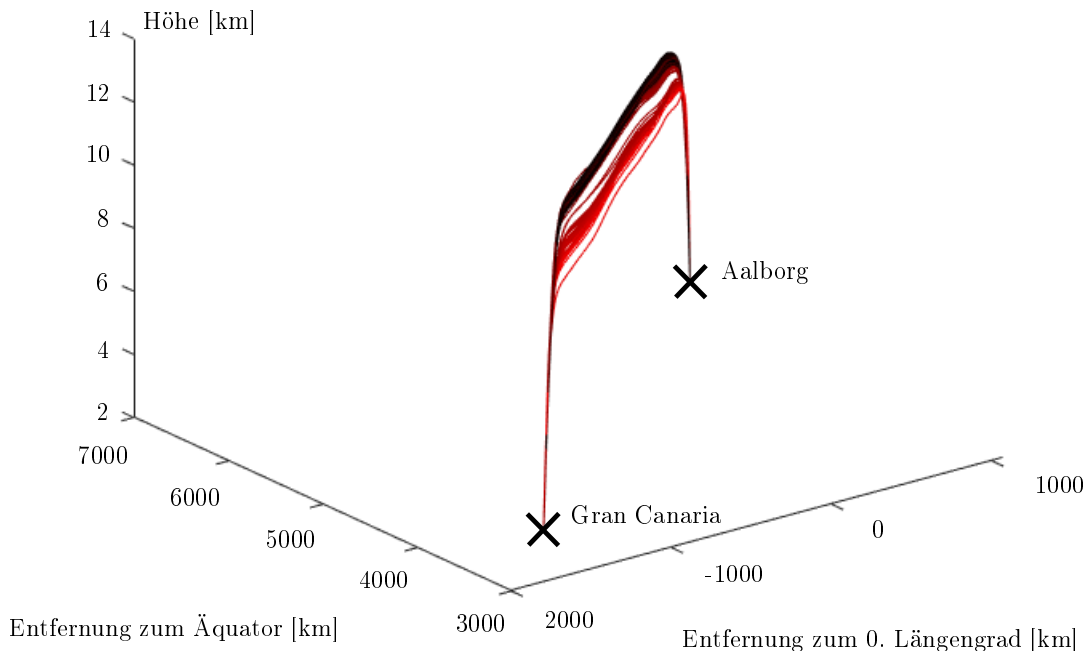


Abbildung 3.3: Trajektoriensatz „00115\_EKYT\_GCLP“ mit Überflug des 0. Längengrades.

Werden 0. Längen- oder Breitengrad überschritten, wird mit einer Fallunterscheidung gearbeitet. Fast alle Funktionen wie die der Energieminimierung und die Metriken können nicht mit negativen Zahlen umgehen. Die Trajektorien werden je bis zum Überschreiten der Null ausgewertet. Für die folgenden Positionen wird das Vorzeichen umgekehrt und dann die Auswertung fortgesetzt. Bei Bedarf können 0. Längen- und Breitengrad auch mehrfach überschritten und die Vorzeichen passend umgerechnet werden.

### 3.4.7 Interpolation der Trajektorien

Für die kontinuierlichen Verfahren ist es erforderlich, aus den gegebenen Trajektorienpunkten einen kontinuierlichen Verlauf zu interpolieren. Dies wurde in der vorangegangenen Projektarbeit [62] abschnittsweise, also je zwischen zwei aufeinanderfolgenden Punkten mit Berücksichtigung der benachbarten Punkte, mit einem quadratischen Polynom durchgeführt. Da die Differenzen für den gegebenen Trajektoriensatz zwischen quadratischer und kubischer Näherung gering waren, wurde davon ausgegangen, dass ein Polynom zweiten Grades ausreichend sei. Für das in der Projektarbeit untersuchte Beispiel für den Trajektoriensatz von Baku nach Luxemburg trifft dies zu. Abbildung 3.4 zeigt die gegebenen Datenpunkte und Abbildung 3.5 die interpolierten Trajektorien des Trajektoriensatzes „00001\_UBBB\_ELLX“ von Baku nach Luxemburg. Die verschiedenen Trajektorien, dargestellt durch einen

Farbverlauf von rot nach schwarz, unterscheiden sich durch ihre Höhe. Abweichungen zur Seite sind nur in sehr geringem Maße vorhanden. Neben den korrekten Trajektorienpunkten ist ein Punkt im Ursprung des Koordinatensystems vorhanden. Abbildung 3.4 wirkt daher gestaucht, um den Punkt im Ursprung ebenfalls darstellen zu können.

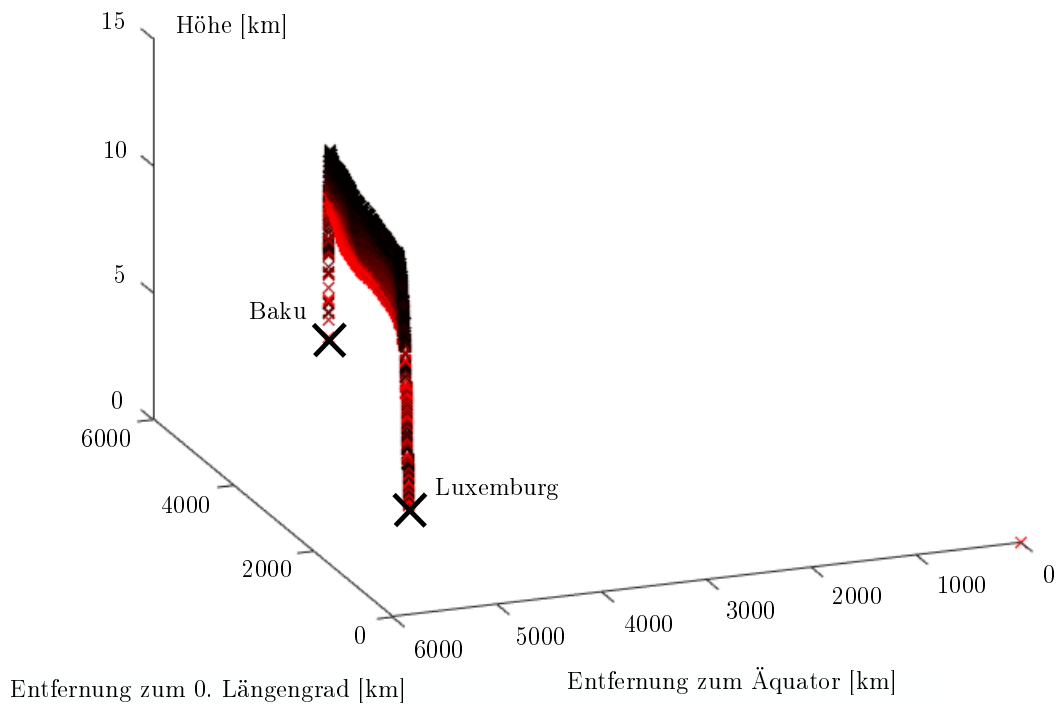


Abbildung 3.4: Trajektorienpunkte zu Datensatz „00001\_UBBB\_ELLX“.

Passend zu den Datenpunkten sind auch die interpolierten Trajektorien in Abbildung 3.5 nur in ihrer Höhe deutlich zu unterscheiden. Da die Auflösung feiner gewählt wurde und in MATLAB die kontinuierlichen Linien weniger auftragen als die Kreuze zur Markierung der Datenpunkte, sind Abstände zwischen den Trajektorien sichtbar sowie eine geringe Vertikalbewegung bei einigen Trajektorien. Der Punkt im Ursprung wurde entfernt, da er sich nicht ohne große Veränderungen in den Trajektorienverlauf einfügt.

Da sich die Trajektorien voneinander so wenig unterscheiden, war die Interpolation mit einem Polynom zweiten Grades für die Projektarbeit ausreichend. Da aber alle Daten der Projekte ATM4E und FlyATM4E untersucht werden sollen, ist die Interpolation so nicht mehr genügend. Abbildung 3.6 zeigt den Datensatz „00113\_EGCC\_LTBA“ aus dem Projekt ATM4E.

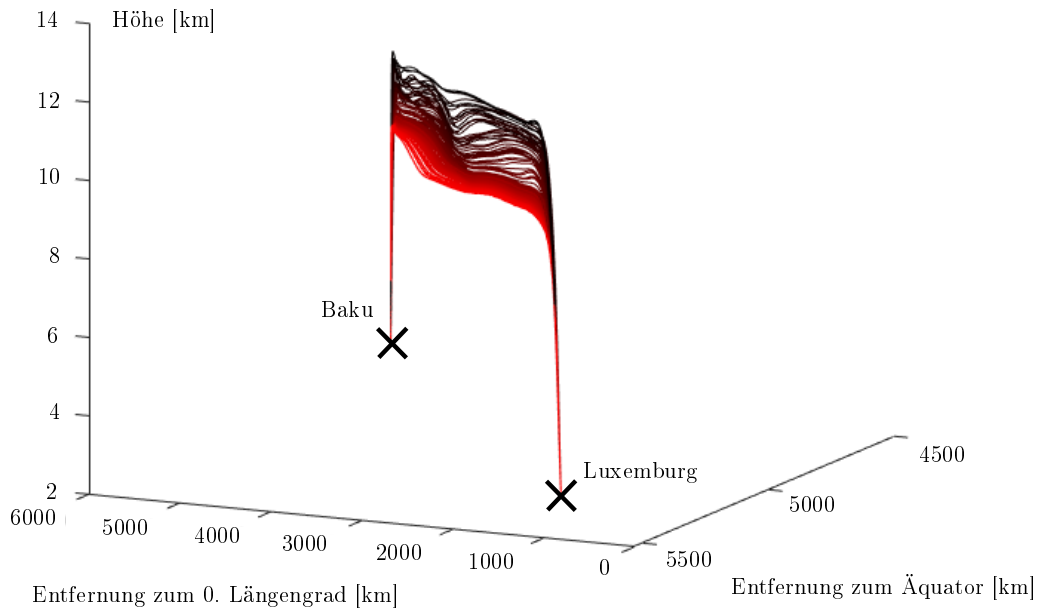


Abbildung 3.5: Interpolierte Trajektorien zu Datensatz „00001\_UBBB\_ELLX“.

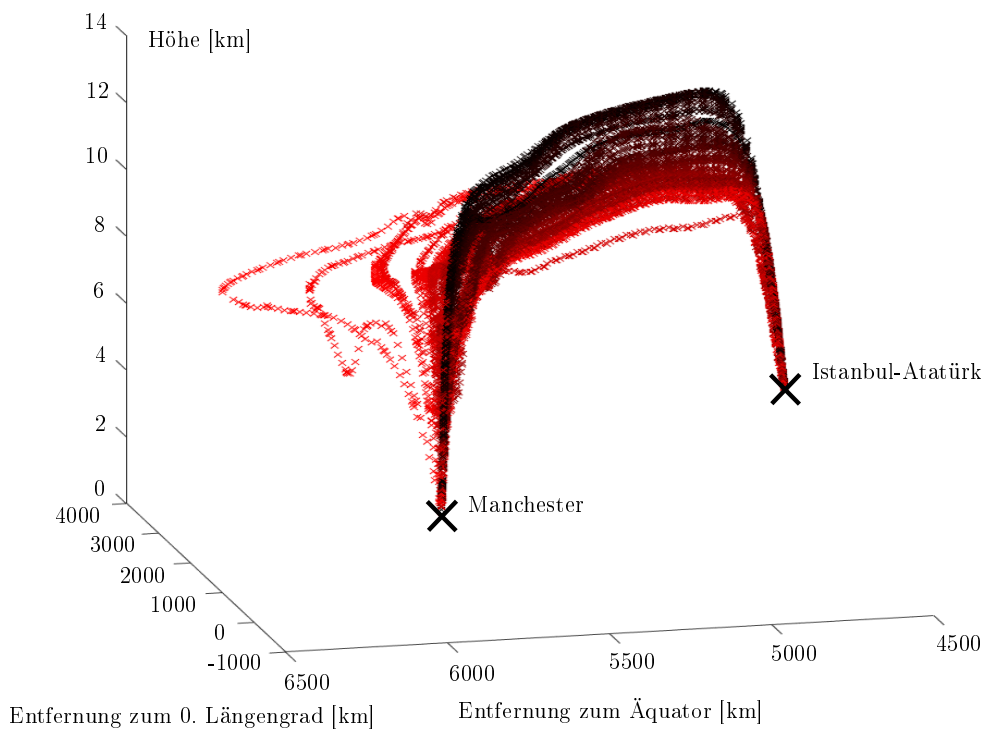


Abbildung 3.6: Trajektorienpunkte zu Datensatz „00113\_EGCC\_LTBA“.

Hierbei handelt es sich um einen Trajektoriensatz für den Flug von Manchester nach Istanbul-Atatürk. Neben unterschiedlichen Höhen der Trajektorien sind hier deutliche horizontale Abweichungen zu erkennen. Einzelne Trajektorien weisen völlig andere Verläufe als der Rest auf. Die Trajektorien liegen verglichen mit dem Datensatz aus der Projektarbeit „00001\_UBBB\_ELLX“ weiter auseinander.



Aufgrund der abweichenden Gestalt benötigt der Trajektoriensatz von Manchester nach Istanbul-Atatürk „00113\_EGCC\_LTBA“ einen höheren Grad für die Polynome, mit denen interpoliert wird. Der Grad des Polynoms wird von MATLAB flexibel ausgewählt, ist aber nie höher als neunten Grades. Abbildung 3.7 zeigt die entsprechend interpolierten Trajektorien. Hier wurden von MATLAB Teilstrecken siebten Grades als passend erachtet und entsprechend ausgewählt und knickfrei aneinandergereiht. Die Auswahl geeigneter Polynome beginnt bei einer quadratischen Interpolation. Eine lineare Interpolation verläuft an den Übergängen nicht knickfrei und ist für weitere Berechnungen nicht verwendbar. Es wird stets der nächsthöhere Grad verwendet, bis entweder der neunte Grad für ein Polynom erreicht ist oder die interpolierten Trajektorien in einem Plot identisch gezeichnet werden würden. Der neunte Grad ist willkürlich gewählt. Bei allen getesteten Trajektorien kam es nie vor, dass eine Trajektorie mit einem Polynom neunten Grades interpoliert wurde. In allen Fällen war siebten Grades oder weniger ausreichend.

Die Abbruchbedingung, siehe 3.3, bezüglich der geplotteten Linie wurde gewählt, um keine absolute Differenz der Polynome festlegen zu müssen, sondern um die Grenze relativ zu setzen. Liegen Start und Ziel nahe beieinander, ist die Trajektorie im Plot sehr fein aufgelöst. Differenzen fallen hier stärker ins Gewicht, insgesamt werden die Trajektorien genauer angepasst. Für größere Distanzen werden die Trajektorien weniger genau interpoliert. Es wird angenommen, dass die Genauigkeit aber relativ dieselbe bleibt.

Wird ein Polynom zu hohen Grades verwendet, ergeben sich zudem Trajektorienverläufe, die bei Betrachtung der gegebenen Punkte nicht realistisch zu sein scheinen. In Abbildung 3.7 sind einige Stellen gelb markiert, deren Verlauf sich so nicht in den Trajektorienpunkten widerspiegelt. Die interpolierten Verläufe sind nicht unmöglich. Realistischer wäre jedoch ein glatter Verlauf. Gerade bei der einzelnen Trajektorie mit der geringsten Höhe ist nicht anzunehmen, dass dieser Verlauf keine reale Trajektorie beschreibt.

Stellen wie die in Abbildung 3.7 gelb markierten treten auf, wenn ein Polynom zu hohen Grades zur Interpolation verwendet wurde. Der mit jedem weiteren Grad hinzukommende Freiheitsgrad ermöglicht die genauere Anpassung an gegebene Punkte, gibt jedoch auch die Möglichkeit für Überschwinger, die in den gegebenen Daten nicht vorhanden sind. Gleichzeitig wird der hohe Grad benötigt, um die Vielfalt der Trajektorienverläufe zu erfassen und den kontinuierlichen Verlauf gut an die gegebenen Punkte anzupassen.

Die gewählte Interpolationsmethode stellt damit einen Kompromiss zwischen Allgemeingültigkeit und Genauigkeit dar. Einerseits sollen die MATLAB-Skripte für alle Daten aus ATM4E und FlyATM4E verwendbar sein. Andererseits sollte das Polynom für die Interpolation einen nicht allzu hohen Grad haben, um einen glatten Trajektorienverlauf ohne Überschwinger sicher zu stellen. Für die meisten Sätze ist die Interpolation ausreichend genau. Nur in wenigen Trajektoriensätzen wie zum Beispiel „00113\_EGCC\_LTBA“ ist eine Abwägung zwischen der Genauigkeit der Anpassung an die Datenpunkte und einem Verlauf, der durch die Trajektorienpunkte gegeben ist, notwendig.

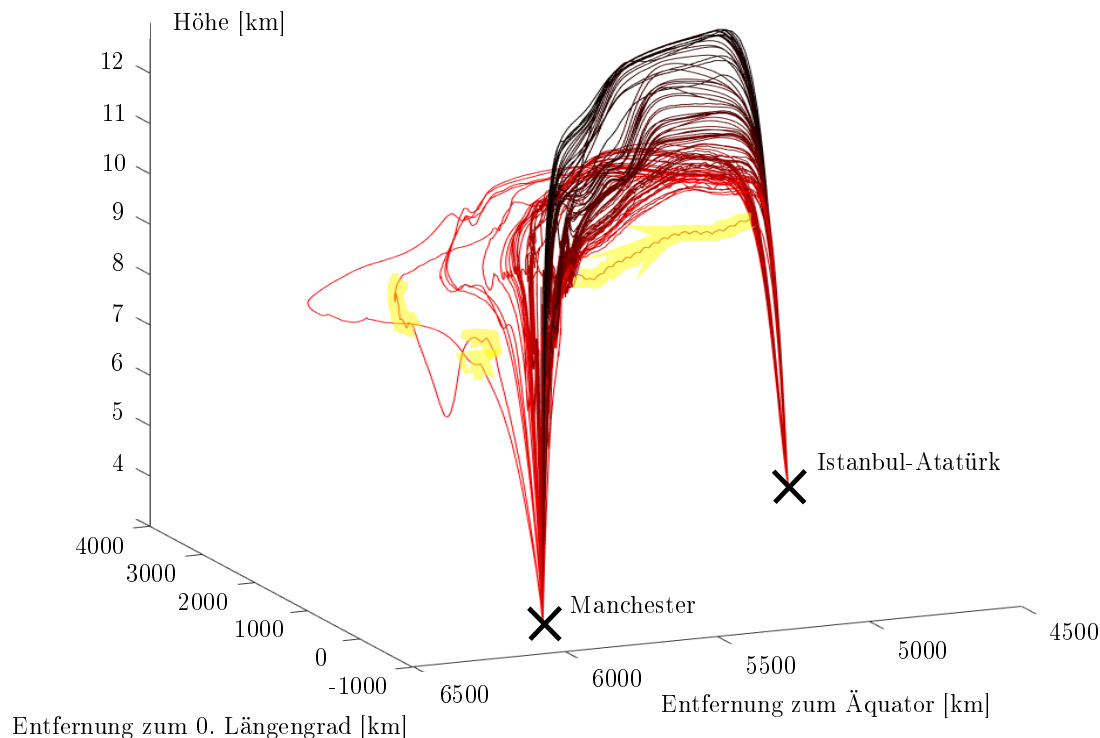


Abbildung 3.7: Interpolierte Trajektorien zu Datensatz „00113\_EGCC\_LTBA“.

### 3.4.8 Nicht verwendete MATLAB-Funktionen

MATLAB verfügt über eine Vielzahl von Funktionen, die in ihrer Berechnungsweise optimiert sind. Funktionen wie `polyfit` zur Anpassung eines Polynoms an gegebene Punkt im Raum wurde benutzt. Diese Funktion stellt eine recht einfache Form von Funktion dar. Mit vorgegebenem Grad wird ein Polynom interpoliert, dass möglichst gut zu den übergebenen Daten passt. Die Auswahl, welcher Grad zu verwenden ist, wird durch eigene Programmierung vorgenommen. Die verschiedenen Funktionen werden geplottet und verglichen, ob sie im Plot voneinander abweichen. Falls nein, ist der verwendete Grad für das Interpolationspolynom ausreichend.

Auf bestehende Funktionen zur Clusteranalyse wurde hingegen komplett verzichtet. MATLAB bietet für Clusteranalysen eine Vielzahl an Funktionen und auszuwählenden Parametern an [60]. Hier liegt aber die Schwierigkeit in der Verwendung dieser Funktionen. Allein zur Bestimmung von in der Nähe liegenden Punkten werden die vier verschiedenen Funktionen `ExhaustiveSearcher`, `KDTreeSearcher`, `knnsearch` und `rangesearch` angeboten. Allen müssen unterschiedliche Parameter in anderer Reihenfolge übergeben werden. Für weitere Gruppierungen gibt es weitere sieben verschiedene Funktionen (`cluster`, `clusterdata`, `cophenet`, `inconsistent`, `linkage`, `pdist`, `squareform`), die alle auf andere Weise mit den vorherigen Funktionen kompatibel sind. Mit den Funktionen einher gehen unterschiedliche Parameter, die durch die Verwendung implizit gesetzt werden.

Da zunächst unklar ist, wie die Cluster genau aufgeteilt werden sollen, ist ein modularer Aufbau für die Parameter und durchgeführten Berechnungen wünschenswert. Damit können die Änderung einzelner Parameter und dessen Auswirkungen auf die Gestalt der Cluster analysiert werden. Die von MATLAB gegebenen Funktionen las-

sen dies jedoch nicht zu, da die jeweiligen Funktionen nicht äquivalent zu ersetzen sind. Da die Funktionen für spezielle Anwendungsfälle kreiert wurden, ist nicht vorgesehen, sie beliebig miteinander zu kombinieren. Eine andere Funktion bedeutet, dass das MATLAB-Skript in seinem gesamten Aufbau verändert werden muss. Daher wurde auf die MATLAB-Funktionen verzichtet und die Clusteranalyse selbst programmiert, sodass jeder Parameter der Clusteranalyse, siehe Abschnitt 2.5, einzeln verändert werden kann. Parameter, die nicht explizit gesetzt werden, können damit nicht übersehen werden. Insgesamt wird auf diesem Wege mehr Übersichtlichkeit und Verständnis erzielt. Zwar sind die Berechnungen nicht bezüglich der Zeit optimiert, dafür wird durch die eigene Programmierung der modularisierte Aufbau möglich und die Kontrolle über die einzelnen Funktionen ist deutlich höher als bei der Verwendung der bereits implementierten MATLAB-Funktionen.

## 3.5 Konstruierte Beispiele für Ähnlichkeit

Die durchgeführte Berechnung von Ähnlichkeit weist wie jedes Verfahren bestimmte Stärken und Schwächen auf. In seiner Gesamtheit wird das Verfahren als für den Vergleich von Flugtrajektorien geeignet befunden. Jedoch gibt es einige Spezialfälle, in denen nicht exakt das gewünschte Ergebnis erzielt wird. Diese sind zwar selten, im Folgenden soll aber auf genau diese Spezialfälle eingegangen werden, um Stärken und Schwächen dieses Verfahren, auch im Vergleich mit dem Anpassen an einen Zeichensatz (Abschnitt 3.7) und mit der Clusteranalyse (Abschnitt 3.9), aufzuzeigen. Warum es überhaupt Sinn ergibt, sowohl diskrete als auch kontinuierliche Verfahren zu verwenden und zu vergleichen, wird in 3.5.1 diskutiert. Charakteristisch für die Berechnung von Ähnlichkeit ist zudem, dass sie nur für Trajektorien mit gleichem Start- und Zielpunkt angewendet werden kann, worauf in 3.5.2 eingegangen wird. Es folgen bestimmte Spezialfälle, in denen gewisse Metriken deutlich sinnvollere Ergebnisse liefern als andere oder die Anwendbarkeit des hier verwendeten Verfahrens nicht mehr gegeben ist. Auf diese Fälle wird in 3.5.3 anhand von Datensätzen gleicher Länge, in 3.5.4 auf die unterschiedliche Auflösung der Daten, in 3.5.5 auf Daten mit einem Zeitversatz, in 3.5.6 auf den Fall des entgegengesetzten Fluges, in 3.5.7 auf spezielle ungünstige Verteilungen und in 3.5.8 auf die filternde und datenglättende Wirkung einiger Metriken eingegangen. Die Verarbeitung stark von der Mehrheit abweichender Trajektorien wird in 3.5.9 diskutiert und in 3.5.10 Vor- und Nachteile der Ähnlichkeitsberechnung gegenüber anderen Verfahren aufgezeigt. In 3.5.11 wird diskutiert, bezüglich welcher Dimension (Strecke, Geschwindigkeit, Beschleunigung) verglichen wird, ob die Wahl Auswirkungen auf die Ähnlichkeitsbestimmung hat und welche Dimension für die Bestimmung der Robustheit von Flugtrajektorien gewählt werden sollte.

### 3.5.1 Diskrete und kontinuierliche Verfahren

Werden Trajektorien in ihrer Gesamtheit betrachtet, sind Start und Ziel innerhalb eines Satzes identisch. Diese Aussage ist richtig, bezüglich der Flugroutenplanung jedoch nicht hilfreich. Die Gestalt bestimmter Trajektorienabschnitte ist von Interesse, weshalb die Trajektorien in kleinere Elemente zerlegt werden und diese miteinander verglichen. Die gegebenen Trajektorien sind als vierdimensionale Punktfolge

(drei Raumdimensionen und Zeit) gegeben, weshalb es sich anbietet, diese Punkte miteinander zu vergleichen. Der Verlauf zwischen den Punkten muss bei Bedarf interpoliert werden. Es liegen keine Informationen vor, wie diese Interpolation im Detail auszusehen hat. Es gibt lediglich Verläufe, die wahrscheinlicher sind als andere. Die diskreten Punkte miteinander zu vergleichen ist damit die Methode, die am wenigsten ungesicherte Informationen in das Berechnungsverfahren einbringt, gleichzeitig die Realität weniger gut abbildet als kontinuierliche Verfahren.

Kontinuierliche Trajektorien entsprechen hingegen der in der Realität abgeflogenen Trajektorie. Das Flugzeug hat zu jedem Zeitpunkt eine bestimmte Position im Raum. Die Auswertungspunkte können beliebig nah aneinander grenzen. Damit gibt es sowohl für die diskreten ähnlichkeitsberechnenden Verfahren als auch für die kontinuierlichen Gründe, sie zu verwenden.

Ein Zwischenfazit der Projektarbeit [62] ist, dass diskrete Verfahren eher die Gestalt der Trajektorien in den Vordergrund rücken, kontinuierliche Verfahren den Abstand zwischen den Trajektorien stärker gewichten. Der Grund dafür liegt in der Anzahl der betrachteten Punkte. Abbildung 3.8 zeigt ein Beispiel zur Veranschaulichung diskreter Berechnung, Abbildung 3.9 dasselbe Beispiel für ein kontinuierliches Verfahren. Eine kontinuierliche Trajektorie wird hier als aus unendlich vielen Punkten mit infinitesimal kleinem Abstand aufgefasst. Es liegen jeweils drei vorgegebene Datenpunkte, je rot und blau, vor. Werden die Trajektorienverläufe interpoliert, werden die jeweiligen linearen Verbindungen zwischen den Punkten als Verlauf angenommen.

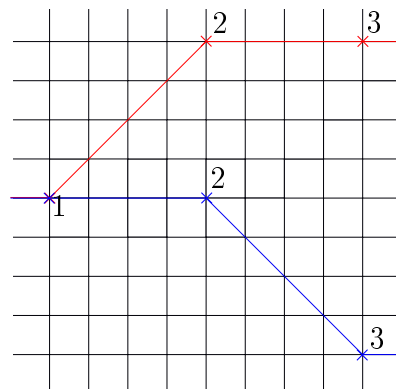


Abbildung 3.8: Schematische Darstellung der Berechnung durch ein diskretes Verfahren.

Wird der Abstand verglichen, ist für den ersten Vergleich das Ergebnis Null. Die Kreuze haben identische Positionen. Die zweiten Kreuze haben den Abstand von vier Kästchenlängen voneinander, die dritten Kreuze haben den Abstand von acht. In Summe beträgt der Abstand zwölf Kästchenlängen. Betrachtet man die Änderungen, so ändert sich im ersten Vergleichspunkt der Verlauf der Linie um  $45^\circ$ . Beide verliefen vorher parallel, hier sogar identisch. Im zweiten Vergleichspunkt weisen beide Trajektorien einen Knick von  $-45^\circ$  auf, die Differenz ist damit Null. Im dritten Vergleichspunkt erfährt die blaue Trajektorie eine Änderung von  $45^\circ$ , die rote bleibt unverändert. In Summe beträgt die Veränderung in den betrachteten Punkten  $90^\circ$ . In Abbildung 3.9 wurden zwischen den drei gegebenen Datenpunkten 1 bis 3 weitere Punkte a bis f interpoliert. Da die neuen Punkte an Stellen hinzugefügt wurden, an

denen der Verlauf keine Änderung erfährt, unterscheiden sich die rote und die blaue Trajektorie weiterhin um  $90^\circ$ . Neben dem bereits berechneten Abstand der Punkte 1 bis 3 von insgesamt zwölf Kästchenlängen kommen nun die Abstände zwischen den neu eingefügten Punkten hinzu. Die Punkte a bis c ergeben summiert einen Abstand von sechs Kästchenlängen. Für die Punkte d bis f kommen weitere 18 hinzu. In Summe beträgt der Abstand des kontinuierlichen (hier zumindest als solches angenommen) Verfahrens 36 Längen, die Änderung beträgt  $90^\circ$ .

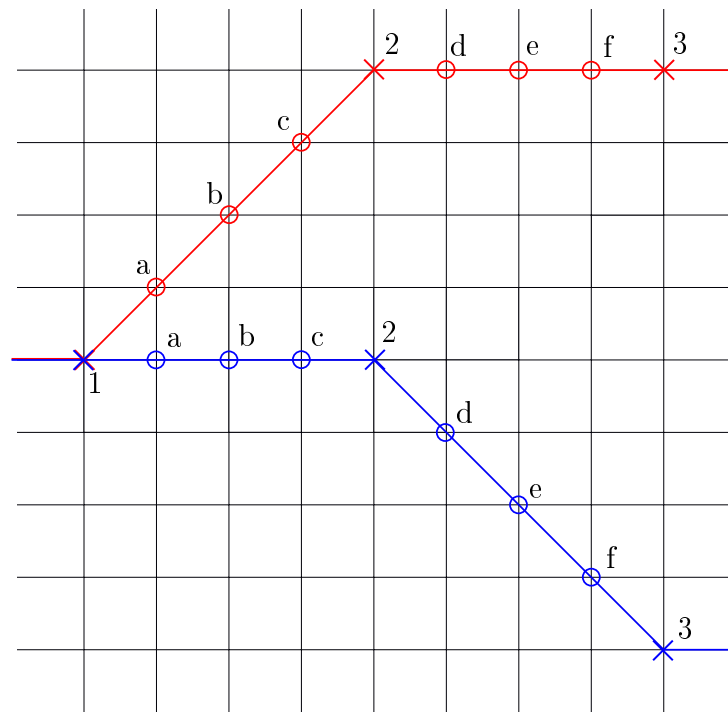


Abbildung 3.9: Schematische Darstellung der Berechnung durch ein (nahezu) kontinuierliches Verfahren.

Am Beispiel wird erkenntlich, worin der Unterschied zwischen einem diskreten und einem kontinuierlichen Vergleich besteht. Während die Änderungsrate über die gesamte Trajektorie gleich bleibt, erhöht sich der aufsummierte Abstand, wenn mehr Punkte verglichen werden. Da bei der diskreten Betrachtung verhältnismäßig wenig Punkte betrachtet werden, besitzt die Veränderung der Gestalt gegenüber dem Abstand einen großen Einfluss. Bei einem kontinuierlichen Vergleich von unendlich vielen Punkten, nimmt der Einfluss des Abstandes gegenüber der Veränderung der Gestalt deutlich zu. Entsprechend gewichten kontinuierliche Verfahren den Abstand stärker als diskrete Verfahren. Bei diskreten Verfahren ist eher die Gestalt der Trajektorie von größerer Relevanz, weniger ihre Position im Raum.

Insbesondere die Charakteristik des kontinuierlichen Verfahrens ist von Vorteil. Wie in 3.4.7 beschrieben, stellt die verwendete Interpolation einen Kompromiss aus Genauigkeit und überschwingendem Verhalten dar. Da aber bei der kontinuierlichen Variante die Gestalt als solche weniger wichtig ist, fallen die Überschwinger in einzelnen Funktionsabschnitten wenig ins Gewicht.

Bezüglich des langfristigen Ziels, Flugrouten optimal zu planen, ist es sinnvoll, den Fokus auf die Distanz zwischen den verschiedenen Trajektorien zu legen. Kurven in

ihrer Gestalt anzupassen, ist einfach möglich. Befindet sich in der Realität ein Flugzeug auf der Trajektorie, kann ihm jederzeit der Befehl zur Kursänderung gegeben werden. Hingegen ist es nicht möglich, ohne zeitliche Verzögerung die Position zu verändern. Eine Trajektorie ähnlicher Gestalt aber in weiter Entfernung kann nicht kurzfristig die bisherige Flugroute ersetzen. Damit sollte den kontinuierliche Verfahren in Bezug auf die Robustheit von Trajektorien gegenüber den diskreten Verfahren eine höhere Gewichtung gegeben werden.

### 3.5.2 Vergleichbarkeit

Die gesamte Ähnlichkeitsberechnung ist nur anwendbar, wenn Start und Ziel der Trajektorie dieselben sind. Bedingt ist dies durch die vorgenommene Skalierung. Für Trajektorien mit gleichem Start und Ziel wird angenommen, dass sich die Gesamtlänge nicht wesentlich unterscheidet. Die Trajektorien müssen für den Vergleich auf dieselbe Länge skaliert werden. Bei etwa gleich langen Trajektorien ist die Skalierung vernachlässigbar klein. Für deutlich verschiedene Längen, wie es bei beliebigen Trajektorien der Fall ist, kann diese Annahme nicht getroffen werden und die gesamte Ähnlichkeitsberechnung ist in der hier vorgestellten Variante nicht zielführend. Dies ist einer der Hauptnachteile dieses Verfahrens. Die beiden anderen betrachteten Methoden wie das Anpassen an einen Zeichensatz (Abschnitt 3.7) oder eine Clusteranalyse (Abschnitt 3.9) können für Trajektorien mit beliebigem Start und Ziel durchgeführt werden.

Ein einfaches Beispiel, welches die nicht zielführende Vergleichbarkeit illustriert, zeigt Abbildung 3.10. Identische Trajektorien wie die rote und blaue besitzen die in MATLAB berechnete Verschiedenheit Null. Hier stimmen Lage und Gestalt der Trajektorien überein. Die grüne Trajektorie besitzt die gleiche Form, jedoch eine andere Lage. Durch die identische Form wird jedoch ebenfalls eine Verschiedenheit von Null und damit Gleichheit berechnet. Auch die orange Trajektorie ist identisch, da ihr Verlauf derselbe ist. Diese Ergebnisse können nicht sinnvoll sein. Entsprechend ist es notwendig, zwei Punkte festzulegen, an denen die Trajektorien beginnen und enden sollen. Somit wird die Rotation und Verschiebung von Trajektorien verhindert.

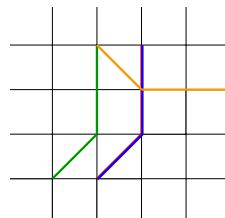


Abbildung 3.10: Beispielhafter Vergleich von Trajektorien mit verschiedener Lage.

### 3.5.3 Einfluss der Datensatzlängen - gleiche Länge

Die Berechnungsvorschrift für Metriken gibt vor, dass jeweils der kürzere Datensatz verwendet wird und dieser mit dem längeren Datensatz verglichen wird. Die Euklidische Metrik berücksichtigt die überschüssigen Datenpunkte nicht, wie in 3.3 erklärt

wird. Die Hausdorff und Fréchet Metriken wählen Datenpunkte aus, die sie nicht berücksichtigen. So entsteht ein Verhalten, dass Fehler glättet, aber in ungünstigen Fällen nicht sinnvolle Ergebnisse für den Ähnlichkeitsvergleich liefert. Details dazu sind in Abschnitt 3.5.8 aufgeführt. Bestehen die zu vergleichenden Trajektorien aus gleich vielen Punkten, ist dies für den Vergleich mit der Euler Metrik der bestmögliche Fall. Keine Datenpunkte entfallen. Welche Trajektorie als Referenz verwendet wird und welche mit ihr verglichen wird, ist irrelevant. In beiden Fällen ist das Ergebnis gleich.

Für den Vergleich mit der Hausdorff und Fréchet Metrik ist die exakt gleichlange Punktfolge ein ungünstiger Fall. Da nicht eindeutig eine kürzere Trajektorie bestimmt werden kann, wird die Wahl willkürlich getroffen. Um dem Problem in MATLAB zu begegnen, wird in solchen Fällen die Trajektorie mit der niedrigeren Nummer im Datensatz zur kürzeren und damit Referenztrajektorie bestimmt. Dieser Fall tritt jedoch selten auf.

In Abhängigkeit von der Wahl der Referenztrajektorie ändert sich das Ergebnis für die Ähnlichkeit. Da die Fréchet Metrik eine spezielle Form der Hausdorff Metrik ist (und zwar wenn das Zeitfenster so gewählt wird, dass es den gesamten Datensatz umfasst), wird folgendes Beispiel nur für die Hausdorff Metrik ausgeführt. Abbildung 3.11 zeigt den Vergleich von zwei Trajektorien, hier auf eine Folge von vier Punkten vereinfacht. Die rote und die blaue Trajektorie sind miteinander zu vergleichen. Beide Trajektorien haben gleich viele Datenpunkte, daher wird (ohne zusätzliche Vorgabe) willkürlich eine Trajektorie gewählt. Links ist die blaue Trajektorie als Referenztrajektorie gewählt, rechts die rote. Links im Bild befindet sich Punkt a von den roten Datenpunkten am nächsten an Punkt 1 der blauen Trajektorie. Am nächsten an Punkt 2 befindet sich Punkt b, an Punkt 3 c und an Punkt 4 d. Alle Punkte werden genau einmal für einen Vergleich durch die Hausdorff Metrik verwendet. Wird derselbe Vergleich mit der roten Trajektorie als Referenz ausgeführt, ergibt sich ein anderes Bild. Wie rechts in Abbildung 3.11 aufgeführt, bilden wieder Punkt 1 und Punkt a ein Paar, da sie den geringsten Abstand aufweisen. Wird der Datenpunkt gesucht, der zu Punkt 2 den geringsten Abstand hat, ist es ebenfalls Punkt a. Auch zu den nächsten beiden Punkten gehört derselbe Punkt. Für den Vergleich wurde hier nur ein einziger Punkt aus vier möglichen gewählt. Obwohl es sich um die gleichen Trajektorien handelt, kommt der Vergleich zu unterschiedlichen Ergebnissen.

Eine grundlegende Eigenschaft einer Metrik ist Symmetrie. Wird ein Objekt mit einem anderen verglichen, so ist das Ergebnis unabhängig der Reihenfolge der verglichenen Objekte identisch. Symmetrie scheint hier aber verletzt zu werden, denn das Ergebnis hängt von der Wahl der Referenztrajektorie ab. In den Definitionen der Hausdorff und Fréchet Metrik ist implizit enthalten, dass kein Fall auftreten darf, in dem die Referenz nicht eindeutig zu bestimmen ist. Dies beinhaltet der in 2.2.3 und 2.2.4 definierte Hilfsabstand  $D$ . Dieser ist frei wählbar, jedoch muss es genau ein eindeutig zuordnbares  $D$  geben. Bei gleich langen Trajektorien ist der Wert von  $D$  zwar eindeutig zu bestimmen, jedoch nicht die Zuordnung. Um immer einen eindeutigen Hilfsabstand und eine eindeutige Position zu bestimmen, wird dazu im MATLAB der Weg über die Datensatznummer gegangen. Sind Datensätze gleich lang, wird immer die niedrigere Datensatznummer als Referenz verwendet. Diese Wahl ist willkürlich, garantiert jedoch die Eindeutigkeit und damit, dass Hausdorff und Fréchet Metrik auf diese Fälle überhaupt anwendbar sind.

Auch bei ungleich langen Trajektorien bedingt die Wahl der Referenztrajektorie das

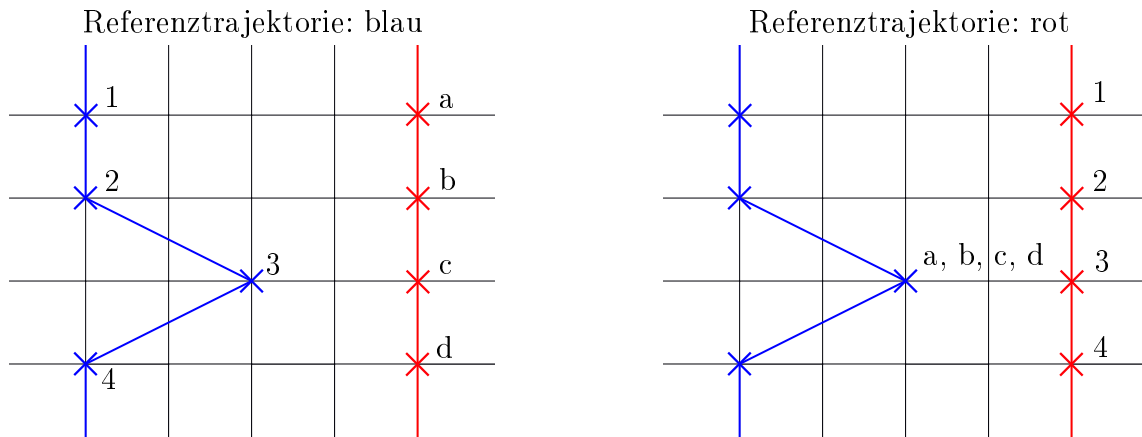


Abbildung 3.11: Vergleich von Trajektorien mit unterschiedlicher Referenz.

Ergebnis. Hier ist die Wahl, ob es der in Abbildung 3.11 linke oder rechte Fall wird, jedoch eindeutig. Welche Variante geeigneter ist, ist allgemein nicht zu sagen. Beide Varianten führen lediglich zu verschiedenen Ergebnissen. Dennoch wird im Allgemeinen angenommen, dass die Auswahl der Referenz einen kleinen Einfluss auf das Ergebnis nimmt. Insgesamt sind in den gegebenen Datensätzen deutlich mehr als vier Punkte vorhanden. Der unterschiedliche Vergleich von wenigen Punkten fällt damit verhältnismäßig kaum ins Gewicht. Zudem ist das Verhältnis der Ausreißergröße im Beispiel nicht so gewählt, wie es in den untersuchten Trajektorien vorkommt. Die tatsächlich untersuchten Trajektorien besitzen einen glatten Verlauf. Die Länge der Trajektorie ist deutlich größer als die Dimension der Ausreißer. Hier im Beispiel ist die seitliche Abweichung etwa halb so groß wie die gesamte Trajektorienlänge. Dies ist in den Flugtrajektorien nicht der Fall, weshalb durch diesen Effekt Abweichungen entstehen, diese aber insgesamt sehr klein sind.

### 3.5.4 Auflösung

Es wird davon ausgegangen, dass die zu vergleichenden Trajektorien ähnliche Auflösungen besitzen, das heißt Punkte im Raum, die zueinander etwa den gleichen Zeitabstand ausweisen. Kleinere Abweichungen sind kein Problem, da davon auszugehen ist, dass der gesamte betrachtete Zeitraum groß im Verhältnis zu den Schwankungen ist. Abbildung 3.12 zeigt ein Beispiel, in dem die Auflösung nicht dieselbe ist. Zur besseren Unterscheidung wurden die Trajektorien parallel versetzt gezeichnet. Gemeint ist jedoch, dass die Trajektorien in ihrer räumlichen Position identisch sind. Dargestellte Abweichung in der Auflösung dürfen bei den zu vergleichenden Trajektorien nicht vorkommen. Jeder Datenpunkt wird durch ein X markiert. Die Auflösung der roten Trajektorie ist doppelt so fein wie die der blauen bei demselben Verlauf der Trajektorien.

Wird die Ähnlichkeitsberechnung mit der Fréchet oder Hausdorff Metrik durchgeführt, stellt die Auflösung in diesem Beispiel kein Problem dar. Wie in 3.5.3 beschrieben, handelt es sich bei der blauen Trajektorie wahrscheinlich um die kürzere. Damit entstehen die Paarungen 1-2, 2-4, 3-6 und 4-8 und die Trajektorien werden (richtigerweise) als identisch berechnet, auch wenn es mehr Punkte gibt. Es könnte



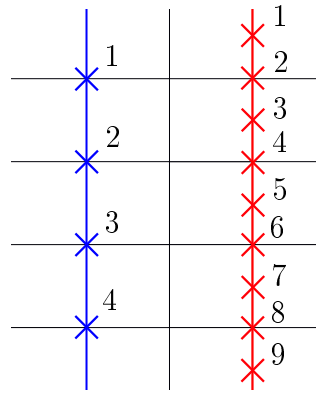


Abbildung 3.12: Vergleich identischer Trajektorien mit unterschiedlicher Auflösung.

aber auch sein, dass die blaue Trajektorie in einem anderen Abschnitt eine höhere Punktdichte aufweist und somit die rote die kürzere Trajektorie ist. In diesem Fall kann zu den ungeraden Punkten der roten Trajektorie kein identischer Punkt gefunden werden und die eigentlich identischen Trajektorien werden als verschieden berechnet.

Die Euklidische Metrik hingegen reagiert anders auf die verschiedenen Auflösungen. Hier wird je Punkt 1 zu Punkt 1, 2-2, 3-3 und so weiter so zugeordnet. Durch die abweichende Auflösung wird der Abstand der Punkte bei jeder weiteren Zuordnung größer. Je länger die Trajektorien sind, desto verschiedener werden sie berechnet, obwohl sie denselben Verlauf besitzen und damit identisch sein müssten. Die Fréchet und Hausdorff Metrik sind dadurch aber nicht als in jedem Fall besser als die Euklidische Metrik zu kategorisieren. Welche Probleme ihre Vergleichsweise mit sich bringt, wird in 3.5.8 aufgegriffen.

Dieses Problem tritt nur bei der diskreten Betrachtung auf. Bei der kontinuierlichen Betrachtung gibt es unendlich viele Punkte, die miteinander verglichen werden können. Bei identischen Gesamtverläufen gibt es auch immer einen identischen Punkt und die Berechnung wird richtig durchgeführt.

### 3.5.5 Trajektorien mit Zeitversatz

Die gegebenen Trajektorien beginnen alle mit der Zeit Null und zählen dann aufwärts. Je nach verwendeter Metrik hat eine Zeitmessung, die nicht bei Null beginnt, verschiedene Auswirkungen auf die Ähnlichkeit. Die Euklidische Metrik ist zeitunabhängig. Für sie bedeutet ein Zeitversatz keinen Unterschied. Ebenso ist die Hausdorff Metrik zeitunabhängig. Wird sie verwendet, ergeben sich ebenfalls keine Änderungen. Für die Fréchet Metrik bedeutet ein Zeitversatz hingegen erhebliche Veränderungen in der Berechnung. Abbildung 3.13 zeigt das Vorgehen bei der Berechnung mit Hilfe der Fréchet Metrik. Die beiden Trajektorien haben im Raum einen identischen Verlauf. Sie wurden in Abbildung 3.13 nur der Übersichtlichkeit halber nebeneinander gezeichnet. Zu erwarten ist in MATLAB das Ergebnis Null, wenn die Euklidische Metrik oder die Hausdorff Metrik eingesetzt werden. Die Trajektorien sind bezüglich ihrer Lage im Raum identisch, was entsprechend berechnet wird. Je nach Anwendungsfall ist das Ergebnis sinnvoll. Die Verläufe liegen unend-

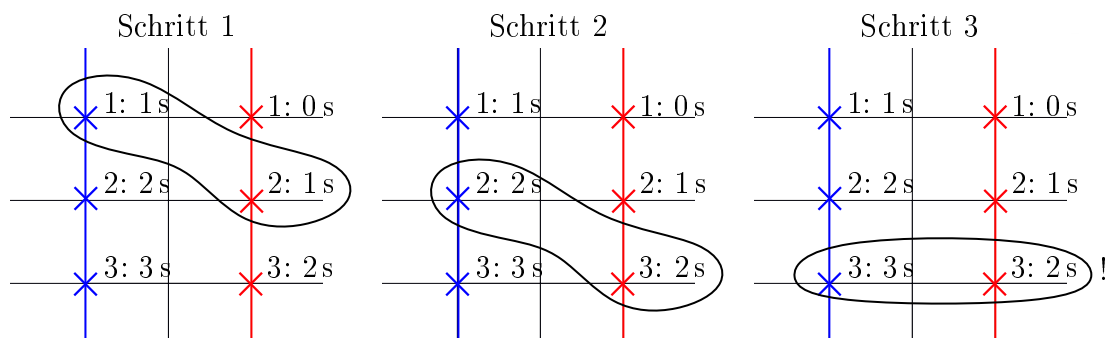


Abbildung 3.13: Vergleichsschritte mit der Fréchet Metrik.

lich nah beieinander. Jedoch könnten sich die Trajektorien bezüglich der zeitlichen Dimension ähnlicher sein. Der Startzeitpunkt der blauen Trajektorie könnte bei weniger als 10 s liegen. Damit wäre die blaue Trajektorie der roten noch ähnlicher. Dies spiegelt sich aber nicht in den Berechnungen wieder.

Die Fréchet Metrik berücksichtigt eine zeitliche Verschiebung. Für die Anwendung muss ein Zeitfenster festgelegt werden, in dem sich mögliche Punkte für einen Vergleich befinden. Für das Beispiel wird ein Zeitfenster von 0,5 s angenommen. Im ersten Schritt wird der zweite Punkt der roten Trajektorie als am besten geeignet zugeordnet. Er ist der einzige Punkt, der innerhalb des Zeitfensters liegt. Analog wird beim zweiten Schritt vorgegangen. Zu Punkt 2 der blauen Trajektorie befindet sich nur Punkt 3 der roten Trajektorie im selben Zeitfenster. Im dritten Schritt kann kein Punkt innerhalb des vorgegebenen Zeitfensters gefunden werden. Der Punkt, der zeitlich den geringsten Abstand hat, ist Punkt 3 der roten Trajektorie. Räumlich gesehen wird durch diese Zuordnung der Abstand von zwei Kästchenlängen erreicht, je in Schritt 1 und 2 eine Kästchenlänge. Wie mit Punkten umgegangen wird, die sich nicht innerhalb des vorgegebenen Zeitfensters befinden, ist im Einzelfall zu entscheiden. In jedem Fall kommt die Fréchet Metrik zu dem Ergebnis, dass die Trajektorien sich deutlich voneinander unterscheiden.

Zusammenfassend lässt sich sagen, dass die Euklidische und Hausdorff Metrik die Zeit vernachlässigen, dafür aber die räumliche Lage priorisieren. Die Ähnlichkeit wird nur innerhalb von drei Dimensionen, den Raumrichtungen, berechnet. Die Fréchet Metrik berücksichtigt eine vierte Dimension, die Zeit. In bestimmten Fällen kann die Ähnlichkeitsberechnung mit der Fréchet Metrik aber zu sehr überraschenden, nicht intuitiven Ergebnissen führen.

Auch wenn hier diskrete Punkte betrachtet werden, tritt das Problem gleichermaßen bei kontinuierlichen Trajektorien auf. Ort und Zeit müssen nicht zwangsläufig voneinander abhängen. Die kleinste räumliche Distanz zwischen zwei Trajektorien ist nicht unbedingt die mit der zeitlich geringsten Differenz und umgekehrt.

### 3.5.6 Gegenläufige Trajektorien

Viele der betrachteten Trajektorien verlaufen gradlinig zwischen Start und Ziel und in 3.5.2 wird festgelegt, dass die Trajektorien alle denselben Start- und Zielpunkt

haben müssen. Dabei dürften Start und Ziel vertauscht werden. Die Frage lautet daher, wie ähnlich sich Trajektorien sind, die in entgegengesetzter Richtung abgeflogen werden. Abbildung 3.14 zeigt zur Illustration eines solchen Falles zwei Trajektorien. Die beiden Trajektorien sollen deckungsgleich sein. Zur Übersicht wurden die Trajektorien in Abbildung 3.14 aber parallel verlaufend dargestellt. Die Trajektorien verlaufen jeweils von Punkt 1 nach Punkt 5 beziehungsweise von Punkt a nach Punkt e.

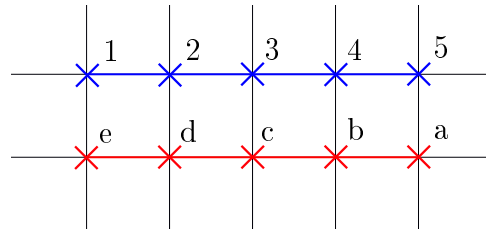


Abbildung 3.14: Beispiel für Trajektorien in umgekehrter zeitlicher Reihenfolge.

Die Euklidische Metrik ordnet die Punkte in ihrer auftretenden Reihenfolge zu. Hier ergeben sich die Punktpaare 1 a, 2 b, 3 c, 4 d und 5 e. In Summe ergeben sich damit zwölf Kästchenlängen Abstand. Je vier von den Paarungen mit Punkt 1 und Punkt 5, je zwei Kästchen bei Punkt 2 und Punkt 4 und Punkt 3 und Punkt c sind identisch. Damit werden die beiden Trajektorien als deutlich voneinander verschieden bewertet. Würden die Trajektorien zum Beispiel in dieselbe Richtung verlaufen, aber wie für die Zeichnung verwendet tatsächlich eine Parallelverschiebung aufweisen, wären die nur fünf Kästchenlängen Abstand und die parallel versetzte Trajektorie der Referenz nach dieser Metrik sehr viel ähnlicher als dieselbe Trajektorie in anderer Richtung.

Anders wird der Abstand durch die Hausdorff Metrik berechnet. Hier wird unabhängig von der Zeit der nächste Punkt zugeordnet. Damit wird zu Punkt 1 e zugeordnet, 2 zu d, 3 c, 4 b und 5 a. Zu jedem Punkt ist jeweils ein identischer Punkt der anderen Trajektorie zu finden. Die Hausdorff Metrik bewertet daher die beiden Trajektorien als identisch. Ob das Ergebnis der Fréchet Metrik eher mit der Hausdorff Metrik oder der Euklidischen Metrik überein stimmt, hängt vom gewählten Zeitfenster ab, was in 3.5.5 erläutert wurde. Allgemein ist dazu keine Aussage zu treffen.

In diesem speziellen Fall sind die berechneten Ähnlichkeiten durch verschiedene Metriken deutlich verschieden. Die Vergleichbarkeit ist nur gegeben, wenn Start und Ziel gleich sind und nicht bei einigen Trajektorien miteinander vertauscht werden. Das Beispiel veranschaulicht das Phänomen im Falle diskreter Berechnungen. Gleichmaßen ist das Beispiel aber auch auf kontinuierliche Trajektorien zu übertragen. Eine Euklidische Metrik bemisst die beiden Trajektorien auch im Kontinuierlichen als deutlich verschieden, während die Hausdorff Metrik zu dem Ergebnis der identischen Trajektorien kommt.

### 3.5.7 Versetzte Datenpunkte

Dieses Problem beschränkt sich auf diskrete Vergleiche. Kontinuierliche Trajektorien zu vergleichen bedeutet zwar mehr Aufwand, das Problem von verschiedenen versetzten Datenpunkten kann dort aber nicht auftreten.

Abbildung 3.15 zeigt dieselben Trajektorienverläufe wie die aus 3.5.6. Die Trajektorien sind hier aber nicht mehr gegenläufig. Beide Trajektorien starten im selben Punkt 1 beziehungsweise a und enden bei Punkt 5 und e. Sie sind nur der Übersichtlichkeit halber nebeneinander gezeichnet und sollen deckungsgleich sein.

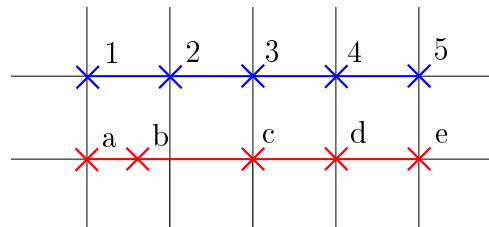


Abbildung 3.15: Beispiel für Trajektorie mit einem versetzten Datenpunkt.

Egal ob die Euklidische, Hausdorff oder Fréchet Metrik verwendet wird, alle Metriken finden zu den Punkten 1, 3, 4 und 5 die zugehörigen identischen Punkte der anderen Trajektorie a, c, d und e. Nur bei Punkt 2 tritt eine Abweichung zu b auf. Dies geschieht in der Euklidischen Metrik, da beides die zweiten Punkte eines Datensatzes sind. Hausdorff und Fréchet Metrik suchen den nächsten Punkt, was hier b ist. Alle Metriken berechnen eine gewisse Abweichung der Trajektorien. Eigentlich sind diese aber identisch. Nur weil ein einzelner Punkt einen räumlichen Versatz aufweist, werden die Trajektorien als unterschiedlich berechnet. Das Problem kann umgangen werden, indem die Ähnlichkeitsberechnungen für kontinuierliche Verläufe durchgeführt werden. Für diskrete Fälle kann angenommen werden, dass die Abweichungen klein sind und das Ergebnis nicht maßgeblich verändern. Ein Datensatz aus ATM4E beziehungsweise FlyATM4E besteht aus weit mehr als fünf Datenpunkten. Die Distanz der gesamten Trajektorie ist deutlich länger als der Abstand zwischen zwei zu vergleichenden Trajektorien. Eine Abweichung wie in Abbildung 3.15 wird das Ergebnis daher nur minimal verändern. Hier fällt der Versatz nur besonders ins Gewicht, da er den Unterschied zwischen dem Erkennen identischer Trajektorien ausmacht.

### 3.5.8 Filternde Eigenschaften

In 3.5.3 wird gezeigt, dass Ausreißer oder scharfe Krümmungen den Vergleich von Trajektorien negativ beeinflussen können. Bei fehlerhaften Daten kann genau diese Eigenschaft von Hausdorff und Fréchet Metrik das Ergebnis aber positiv beeinflussen. Der im folgenden Beispiel dargestellte Effekt gilt für diskrete Trajektorien. Mit Bezug auf die in 3.4.7 ausgeführte Interpolation und deren Schwierigkeiten können Filtereffekte aber auch für kontinuierliche Trajektorien sinnvoll eingesetzt werden. Hier gilt die Annahme, dass ein Abschnitt aus unendlich vielen Datenpunkten besteht und das gezeigte Beispiel in Abbildung 3.16 in seiner Auflösung beliebig klein

dafür einzusetzen ist, wodurch der Effekt auf den gesamten Vergleich klein ist. Bei der Euklidischen Metrik tritt der im Folgenden beschriebene Effekt nicht auf. Erneut sollen die Trajektorien in Abbildung 3.16 einen identischen Verlauf aufweisen und sind nur zur besseren Unterscheidung nebeneinander gezeichnet worden. Punkt b ist ein fehlerhafter Punkt innerhalb der blauen Trajektorie. Der eigentlich korrekte Verlauf ist der Verlauf der roten Trajektorie. Mit Ähnlichkeitsberechnungen durch die Euklidische Metrik wird zu 1 a, 2 b, 3 d und 4 e zugeordnet. Nur das erste Punktpaar ist identisch, die anderen drei Punkte besitzen einen Abstand von mindestens einer Kästchenlänge. Die Trajektorien werden als voneinander verschieden eingestuft. Bei der Hausdorff und Fréchet Metrik kann eine Eigenschaft dieser Me-

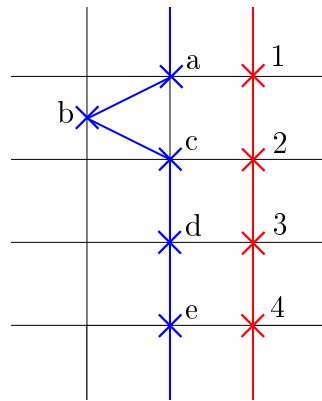


Abbildung 3.16: Beispiel für Trajektorie mit fehlerhaftem Datenpunkt.

triken in einem solchen Fall positive Auswirkungen haben. In beiden Metriken wird (bei der Fréchet Metrik natürlich mit entsprechend großem Zeitfenster) zu Punkt 1 a, 2 c, 3 d und 4 e zugeordnet. Es handelt sich hierbei jeweils um die identisch gelegenen Punkte. Die Trajektorien werden trotz fehlerhafter Daten als gleich erkannt. Was in diesem Fall von Vorteil ist, kann aber auch ein Nachteil sein. Handelt es sich bei den Trajektorien aus Abbildung 3.16 nicht um fehlerhafte Daten, sondern den tatsächlichen Verlauf der Trajektorien, ist die filternde Eigenschaft der Hausdorff und Fréchet Metrik eher ungünstig und führt zu einem falschen Ergebnis.

### 3.5.9 Verhalten bei Instabilitäten

Wie in Abschnitt 2.6 beschrieben, können Instabilitäten und Sprünge als Resultat von nicht kontinuierlichen Wetterlagen und Sprüngen in der Gewichtung zwischen Kosten- und Umweltfaktoren zwischen den Trajektorien auftreten. Situationen, in denen solche Sprünge auftreten, müssen erkannt und vermieden werden. Sie sind im Sinne der Flugplanung nicht robust, denn kleine Veränderungen in den Annahmen verändern das Ergebnis erheblich. Bezüglich der räumlichen Lage kann die durchgeführte Ähnlichkeitsberechnung solche Trajektorien erkennen, da Lage und Gestalt erheblich abweichen. Diese Eigenschaft weisen aber auch das Anpassen an einen Zeichensatz (Abschnitt 3.7) und die Clusteranalyse (Abschnitt 3.9) auf. Was nicht berücksichtigt wird, ist die Lage der zugehörigen Punkte im Diagramm von Kosten- und Umweltfaktoren. Wünschenswert ist eine große Bandbreite möglicher

Gewichtungen, jedoch nur wenig Abweichung in den daraus resultierenden Trajektorien. Dies stellt einen Faktor dar, den bisher keine Methode berücksichtigt. Diese Problemstellung wird ausführlich in Abschnitt 4 behandelt. An dieser Stelle lässt sich nur die Aussage treffen, dass die Berechnung von Ähnlichkeit und alle anderen betrachteten Methoden mit Instabilitäten umgehen können.

### 3.5.10 Vorteile gegenüber anderen Methoden

In den vorherigen Kapiteln werden viele Beispiele konstruiert, bei denen Ähnlichkeitsberechnungen Fehler ausweisen oder kein sinnvolles Ergebnis liefern. Alternative Verfahren werden in [62] ausführlich diskutiert. Viele Verfahren zur Feststellung von Ähnlichkeit beschäftigen sich mit nicht numerischen Eigenschaften und sind daher ungeeignet für die Ähnlichkeitsberechnung von Trajektorien. Nur wenige Verfahren sind zur Feststellung der Ähnlichkeit von Positionen und Verläufen vorgesehen.

Mathematische Ähnlichkeit wäre ein geeignetes Konzept, da es für geometrische Objekte beliebiger Gestalt gedacht ist. Jedoch wird hier mit Verschiebung, Skalierung und Drehung gearbeitet, um die Ähnlichkeit zweier Objekte festzustellen. Sind Objekte durch die drei genannten Operationen in Deckungsgleichheit zu bringen, sind sie sich ähnlich. Alle zu vergleichenden Trajektorien besitzen bereits denselben Start- und Zielpunkt. Verschiebungen der Trajektorien sind deshalb nicht mehr möglich. Mindestens Start und Ziel befinden sich in Deckungsgleichheit, während der Trajektorienverlauf dazwischen dies in der Regel nicht erfüllt. Auch die Skalierung der Trajektorien ist stark eingeschränkt. Abbildung 3.17 zeigt zwei Trajektorien, die bezüglich mathematischer Ähnlichkeit überprüft werden sollen. Ob die Trajektorien in Deckungsgleichheit gebracht werden können, ist von der Wahl des Koordinatensystems abhängig. Im Koordinatensystem  $xyz$  ist keine Skalierung in Richtung der  $x$ -Achse möglich, hier links in Abbildung 3.17 dargestellt. Eine Skalierung würden das Ziel „Z“ verschieben. Gleiches gilt für die Skalierung in Richtung der  $y$ -Achse. Nur die Skalierung der  $z$ -Achse ist möglich, in Abbildung 3.17 durch den grünen Pfeil markiert. Der Grund dafür ist, dass sich der Start „S“ und das Ziel „Z“ beide auf gleicher Höhe der  $z$ -Achse befinden. Wird statt des  $xyz$ -Koordinatensystems das  $x'y'z$ -Koordinatensystem gewählt (rechts), also ein verdrehtes Koordinatensystem um die  $z$ -Achse, wird die Skalierung in  $x'$ -Richtung möglich. Grund dafür ist die neue Lage der Punkte, die sich beide an der  $x'$ -Position Null befinden. Abbildung 3.17 verdeutlicht den neuen Freiheitsgrad durch einen weiteren grünen Pfeil. Analog verhält sich die Drehung der Trajektorien. Abhängig von der Wahl des Koordinatensystems sind die Trajektorien drehbar.

Dass die Ähnlichkeit von einer willkürlichen Wahl eines Koordinatensystems abhängt, ist keine positive Eigenschaft. In den meisten Fällen sind weder Skalierung noch Drehung möglich. Das Ergebnis mathematischer Ähnlichkeit ist zudem, dass sich die Trajektorien entweder ähnlich sind oder unähnlich sind. Allgemein wird mangels Anpassbarkeit das Ergebnis nicht ähnlicher Trajektorien berechnet. Um die willkürliche Wahl zu umgehen, könnte ein Koordinatensystem festgelegt werden, das alle Freiheitsgrade der Anpassbarkeit erlaubt. Abbildung 3.18 zeigt die Wahl eines solchen Koordinatensystems. Durch die Freiheitsgrade bei der Skalierung sind beliebige Trajektorien auf die direkte Verbindungslinie zwischen den Punkten zu stauchen. Die beiden Schritte in Abbildung 3.18 illustrieren den Vorgang. Trajektorien beliebiger Gestalt sind damit in Deckungsgleichheit zu bringen und sich damit

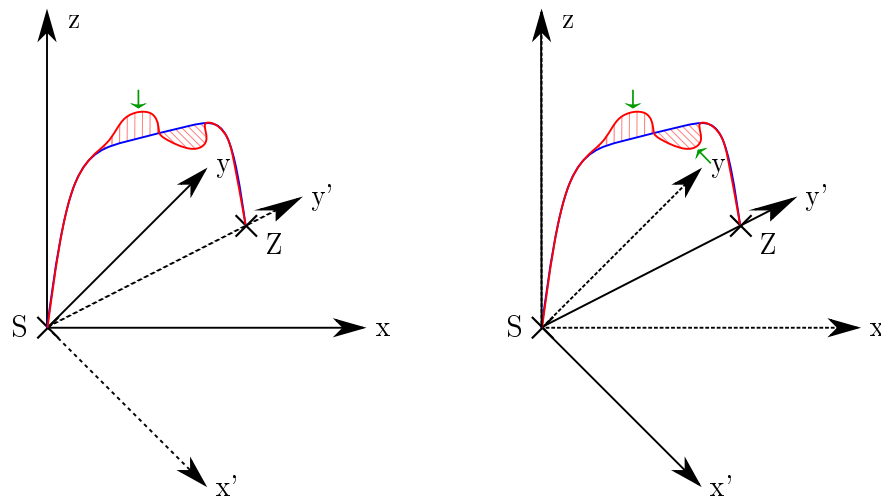


Abbildung 3.17: Skalierbarkeit bei unterschiedlicher Wahl des Koordinatensystems.

ähnlich. Daher ist das Stauchen auf eine gerade Linie in einem solchen Koordinatensystems nicht sinnvoll.

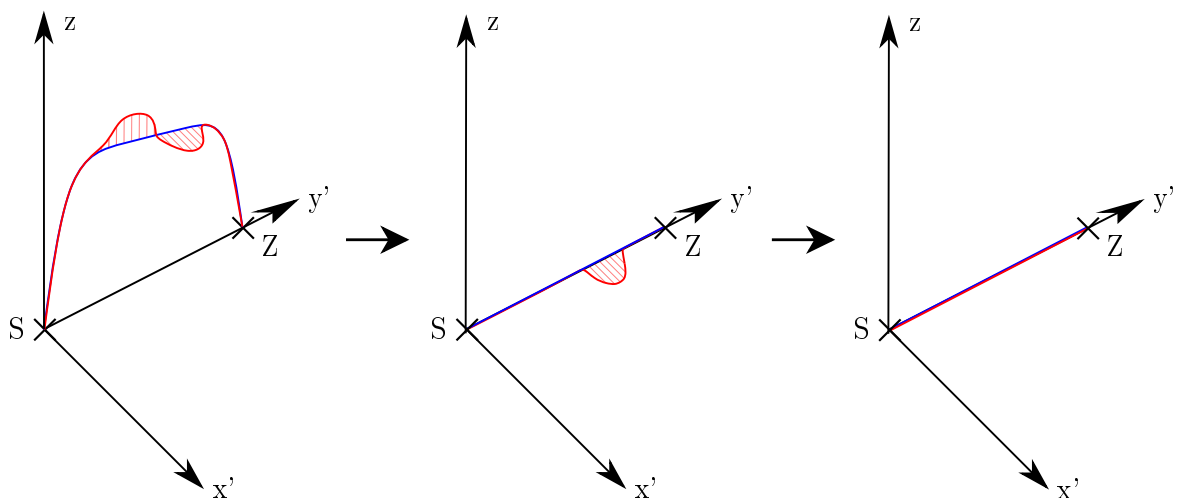


Abbildung 3.18: Stauchen von zwei Trajektorien.

Auch Flächenintegrale scheinen auf den ersten Blick zielführend. Abbildung 3.19 zeigt jedoch, warum Integrale kein gutes Maß für die Ähnlichkeit von Flugtrajektorien darstellen. Es gibt die Möglichkeit, das Integral zu verwenden, dass durch die Integration der interpolierten Funktion entsteht. Sowohl die rote als auch die blaue Trajektorie verlaufen von Punkt 1 zu Punkt 2. Die rote Trajektorie verläuft gerade auf direktem Weg, die blaue Trajektorie in einer sinusförmigen Welle. Wird nun die blaue Trajektorie mit Bezug auf die rote integriert, ist der Wert des Integrals Null. Ein Teil der Fläche befindet sich oberhalb der roten Kurve. Ein anderer gleich großer Teil unterhalb der roten Trajektorie und wird somit abgezogen. Die Skalierung der Abweichung nach oben und unten kann beliebig groß werden, solange die Streckung für beide Kurvenabschnitte, oberhalb und unterhalb, vorgenommen wird. Dass solche Trajektorien als ähnlich beziehungsweise hier sogar identisch eingestuft werden, ist kein hilfreiches Ergebnis.

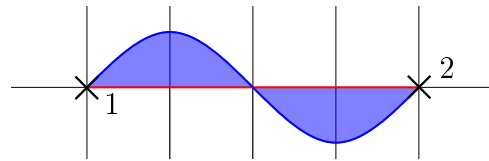


Abbildung 3.19: Zwei Trajektorien, die durch Integration verglichen werden.

Jedoch gibt es die Möglichkeit, Betragsintegrale zu verwenden. Unabhängig ihrer Ausrichtung wird die Flächendifferenz in jedem Abschnitt addiert. Im Beispiel aus Abbildung 3.19 würde sich damit die blaue Fläche nicht mehr auslöschen, sondern zwei Mal die Größe der Fläche den Wert für Ähnlichkeit beschreiben. Werden die Flächen gestaucht, werden die Trajektorien als ähnlicher zueinander bestimmt, mit einer Streckung als weniger ähnlich berechnet.

Mathematische Ähnlichkeit und einfache Integrale eignen sich daher nicht, um die gegebenen Trajektorien miteinander zu vergleichen. Betragsintegrale sind in ihrer Funktionsweise zu den Ähnlichkeitsberechnungen identisch. In beiden Fällen wird je Abschnittsweise verglichen und der Abstand je zum Ergebnis hinzugefügt, niemals abgezogen. Die Betragsintegrale werden nicht weiter explizit genannt. Sie werden lediglich implizit in der Ähnlichkeitsberechnung verwendet.

### 3.5.11 Bezugsgröße

Um eine Bewegung zu beschreiben, können Positionen, Geschwindigkeiten und Beschleunigungen zusammen mit Start- oder Randbedingungen gleichwertig eingesetzt werden. Jedoch ergeben sich andere Ähnlichkeiten in Abhängigkeit der gewählten Bezugsgröße. In Abbildung 3.20 sind drei verschiedene Körper zu sehen, je in blau, rot und grün. Sie bewegen sich in Richtung der Pfeile. Für den blauen und den roten Körper handelt es sich um zwei Momentaufnahmen. Zunächst befinden sich beide Körper an den äußeren Positionen und bewegen sich in Pfeilrichtung. Im nächsten Moment haben sich die Körper aufeinander zu bewegt und sind schließlich zusammengestoßen. Der grüne Körper hat drei Momentaufnahmen. Er startet links im Bild und bewegt sich in Richtung des Pfeils nach rechts. Die Geschwindigkeiten aller Körper sind dieselben. Zwischen dem roten und blauen Körper und dem grünen Körper wurden die Linien unterbrochen. Der grüne Körper befindet sich nicht genau ein Kästchen vom roten und blauen entfernt, sondern in einer etwas größeren Entfernung.

Werden der blaue und der grüne Körper bezüglich ihrer Ähnlichkeit verglichen, ist die Bezugsgröße entscheidend für das Ergebnis. Bezüglich ihrer Position sind sie zu jedem Zeitpunkt weit voneinander entfernt und damit kaum ähnlich. Bezüglich ihrer Geschwindigkeit sind die Verhalten von blauem und grünem Körper bis zum Zusammenstoß mit dem roten Körper identisch. Würde die Ähnlichkeit bezüglich der Geschwindigkeit berechnet werden, würden die Bewegungsgeschwindigkeiten von blauem und grünem Körper als sehr ähnlich eingestuft werden. Wird wiederum die Beschleunigung als Vergleichsgröße gewählt, sind sogar die Bewegungen von rotem, blauem und grünem Körper zeitweise identisch. Bis zu dem Zusammenstoß von rotem und blauem Körper beträgt die Beschleunigung für alle Körper konstant Null.



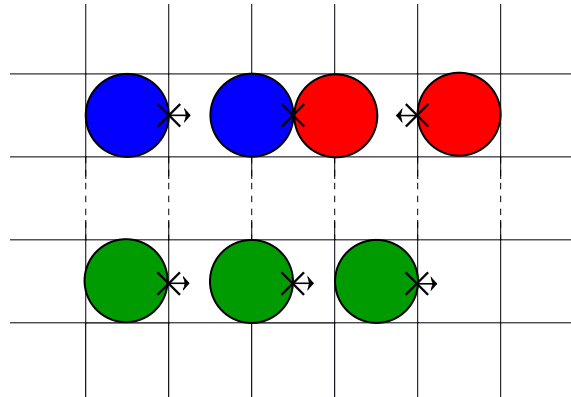


Abbildung 3.20: Körper mit verschiedenen Positionen und Geschwindigkeiten.

Es gibt aber auch den gegenteiligen Fall, dass sich Körper bezüglich ihrer Geschwindigkeit oder Beschleunigung unterscheiden, sich in ihrer Position aber sehr ähnlich sind. Einen solchen Fall zeigen der blaue und der rote Körper. Ihre Geschwindigkeiten sind zunächst betragsgleich, aber in entgegengesetzter Richtung. Die Körper stoßen zusammen und ändern sprunghaft ihre Geschwindigkeit in die entgegengesetzte Richtung. Bezüglich ihrer Geschwindigkeit sind sich die Körper nicht ähnlich. Bezüglich ihrer Position kann die Entfernung zu Beginn der Betrachtung bereits klein sein. Sie verringert sich auf Null, wenn die mit dem X markierten Punkte als Bezugspunkte herangezogen werden. Nach dem Zusammenstoß steigt die Distanz wieder an und die Ähnlichkeit der Position nimmt ab. Bei geringer Entfernung und hoher Geschwindigkeit sind sich die Körper bezüglich ihrer Position sehr ähnlich, bezüglich ihrer Geschwindigkeit kaum.

Ob Position, Geschwindigkeit oder Beschleunigung für die Ähnlichkeitsberechnung verwendet wird, hat damit einen maßgeblichen Einfluss auf das Ergebnis. Gegeben sind die Daten als Position. Mit der Verwendung von den Positionsdaten wird Berechnungsaufwand eingespart. Zudem entstehen durch die Interpolation wie in 3.4.6 beschrieben teilweise Verläufe, die so nicht direkt gegeben und nicht plausibel sind. Durch Differentiation wird der Einfluss dieser Stellen größer und verfälscht das Ergebnis mehr. Auch dies ist ein Grund für die Verwendung der Positionsdaten. Da für eine Aussage über Robustheit der Abstand der Trajektorien entscheidend ist und weniger die Gestalt der Trajektorie, ist dies ein weiteres Argument für die Verwendung der Positionsdaten. Für die weiteren Berechnungen wird daher immer die Position verwendet.

## 3.6 Zusammenfassung der Ähnlichkeitsberechnung

Die Beispiele zeigen viele Fälle, für die die ausgewählte Ähnlichkeitsberechnung kein sinnvolles Ergebnis liefert. Alternative Ansätze wie in der vorangegangenen Projektarbeit [62] aufgelistet, sind jedoch oft gar nicht oder stark eingeschränkt anwendbar. Andere anwendbare Ansätze wie die Verwendung eines Integrals liefern keine sinnvollen Ergebnisse. Da die jeweiligen Beispiele Extremfälle sind, die in dieser Ausprägung in den gegebenen Daten nicht auftreten, wird angenommen, dass die

Ergebnisse der Ähnlichkeitsberechnung durchgehend als brauchbar einzustufen sind. Für die Gestalt, die die meisten Trajektorien aufweisen, ist eine dreischrittige Ähnlichkeitsberechnung wie in der Projektarbeit [62] ausführlich beschrieben am besten geeignet. Die später erläuterten Ansätze über das Anpassen an einen Zeichensatz (Abschnitt 3.7) und die in Abschnitt 3.9 erklärte Clusteranalyse finden ebenfalls Anwendung, da es hier nicht mehr nur um ein geeignetes Verfahren zur Bestimmung von Ähnlichkeit geht, sondern Robustheit Bezug auf räumliche Distanzen nimmt. Ist explizit eine Abstandsbestimmung gewünscht, sind beide Verfahren neben der Ähnlichkeitsberechnung geeignet. Hingegen bezieht sich das Verfahren aus der Projektarbeit [62] allgemein auf Ähnlichkeit von Flugtrajektorien.

Auf die Bedeutung der Ergebnisse wird in Abschnitt 4 ausführlich eingegangen. Allgemein ist das Ergebnis der Ähnlichkeitsberechnung ein Array mit Zahlen, die die Ähnlichkeit jeder gegebenen Trajektorie mit einer anderen beziffern. Diese Zahlen für sich genommen auszuwerten und Aussagen zur Robustheit zu treffen, ist schwierig. Daher wird das Verfahren nicht alleine angewendet, sondern mit der Clusteranalyse kombiniert. Warum und wie dies geschieht, wird in 3.9 erklärt.

## 3.7 Anpassen an einen Zeichensatz

In Anlehnung an die in 2.2.5 beschriebene Editierdistanz soll untersucht werden, ob die Trajektorien sinnvoll durch einen festzulegenden Zeichensatz zu approximieren sind und damit Ähnlichkeit der Trajektorien bestimmt werden kann. Dafür wird ein MATLAB-Skript geschrieben, dessen Funktionen in 3.7.1 erläutert werden. Die Implementierung wird nur prototypisch durchgeführt, da während des Prozesses viele Probleme auftauchten, die das Verfahren noch vor vollständiger Implementierung als ungeeignet herausgestellt haben. Trotzdem wird der MATLAB-Code fertiggestellt, sodass durch den Code die auftretenden Probleme, die in 3.7.2 beschrieben werden, deutlich werden. Ob und wie die systematisch auftretenden Probleme zu lösen sind, wird in Abschnitt 3.8 ausführlich erklärt. Vorweg genommen werden kann, dass das Anpassen an einen Zeichensatz in jedem Fall nicht zielführend eingestuft wird, da mögliche Lösungen nicht konvergieren. Das Verfahren wird daher in der Auswertung nicht berücksichtigt.

### 3.7.1 Funktionen der MATLAB-Skripte

Das vollständige MATLAB-Skript `Zeichensatz.m` ist im Anhang beigelegt. Den Ablauf des Verfahrens in MATLAB stellt Abbildung 3.21 dar. Das Skript `Zeichensatz.m` bündelt alle Einzelfunktionen und wird ausgeführt. In einem ersten Schritt werden die Daten von einem gesamten Trajektoriensatz geladen. Diese Funktion entspricht der Ladefunktion der diskreten Daten aus den anderen Skripten zur Ähnlichkeitsberechnung. In diesen gibt es wie in Abschnitt 3.3 beschrieben jeweils eine Funktion zum Laden kontinuierlicher und diskreter Trajektoriendaten. Da für die Bestimmung der zu verwendenden Zeichen nur die Einzelpunkte und kein kontinuierlicher Verlauf erforderlich sind, wurde die Funktion `DatenLadenDiskret` kopiert. Als nächstes werden die zu verwendenden Zeichen festgelegt. Hier kann eine beliebige Auswahl an Form, Größe und Anzahl der Zeichen getroffen werden. Willkürlich

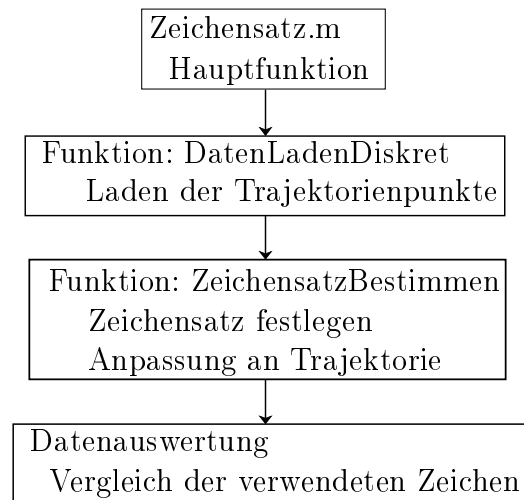


Abbildung 3.21: Funktionsverknüpfung für das Anpassen an einen Zeichensatz.

gewählt wurden für die Durchläufe Zeichen, die von einem gegebenen Datenpunkt zum nächsten reichen. Es dürfen drei Arten (ebenfalls willkürliche Wahl) von Zeichen verwendet werden. Dies sind jeweils gerade Strecken entlang der Achsen des Koordinatensystems. Mit den festgelegten Zeichen wird die Trajektorie approximiert. Es wird getestet, welches Zeichen dem nächsten Abschnitt jeweils am besten entspricht und eine der drei Optionen ausgewählt. Ohne den vorherigen oder nächsten Abschnitt zu berücksichtigen, wird so für jeden Abstand verfahren.

Die Auswertung besteht im Vergleich der Anzahl und Reihenfolge der verwendeten Zeichen. Die Reihenfolge ist ein theoretisch verfügbarer Wert. In der prototypischen Version des MATLAB-Skriptes wurde sich auf die jeweilige Anzahl der verwendeten Zeichen beschränkt, da hier bereits Probleme bezüglich der Vergleichbarkeit auftreten. Details dazu sind in 3.7.2 nachzulesen.

### 3.7.2 Auftretende Probleme während der Berechnung

Die größte Herausforderung stellt das Finden eines geeigneten Zeichensatzes zur Approximation dar. Die im Skript `Zeichensatz.m` festgelegte Wahl von geraden Linien entlang der drei Raumachsen liefert für den Datensatz „00001\_UBBB\_ELLX“ aus dem Projekt ATM4E Ergebnisse, die nie das dritte Zeichen (`ZeichenL`) benutzen. Obwohl die Trajektorien dreidimensional sind, stehen hier die Höhe und der Breitengrad im Vordergrund. Jeder Abschnitt ist eher mit einer Linie in einer dieser beiden Dimensionen anzupassen. Werden alle drei Zeichen berücksichtigt, sind sich die Trajektorien sehr ähnlich, da sie alle mit nur zwei von drei möglichen Zeichen gut approximiert werden können. Wird das nicht verwendete Zeichen entfernt, sind die Trajektorien weniger ähnlich zueinander, da es nur noch zwei Zeichen gibt, bezüglich derer sie verglichen werden können und alle möglichen Zeichen benötigt wurden. Es könnten aber auch weitere Zeichen eingeführt werden, die fast `ZeichenL` entsprechen. Diese Zeichen werden weiterhin nicht verwendet und so werden sich die Trajektorien in Abhängigkeit der Zeichensatzwahl immer ähnlicher. Es werden für den Datensatz „00001\_UBBB\_ELLX“ lediglich zwei aus beliebig vielen Zeichen verwendet. Für andere Datensätze sind die deutlich mehr benötigten Zeichen.

Alle drei festgelegten Zeichen werden zum Beispiel bei Datensatz „00002\_ENVA\_-GCLP“ aus ATM4E benötigt. Wird hier ein weiteres Zeichen beliebiger Gestalt hinzugefügt, verändert sich die verwendeten Zeichenhäufigkeit. Das Hinzufügen oder Weglassen von einem oder mehreren Zeichen verändert hier in jedem Fall die Ähnlichkeit. Dieser Fall wird in 3.8.1 noch einmal anhand eines einfacheren Beispiels veranschaulicht. Insgesamt sind zu wählender Zeichensatz und Trajektoriensatz voneinander abhängig. Je nach Auswahl des Zeichensatzes ergeben sich verschiedene Ähnlichkeiten für die zu überprüfenden Trajektorien. Ein Zeichensatz kann aber nur mit Kenntnis der Trajektoriengestalt sinnvoll gewählt werden und muss zudem für jeden Vergleich neu angepasst werden. Ein Zeichensatz ist somit weder allgemein festzulegen noch kann die Festlegung ohne Iterationsschritte erfolgen. Eine Iteration der Zeichenauswahl birgt aber das Problem, dass diese nur in wenigen Fällen zufällig konvergiert. In dieser Anwendung ist keine Konvergenz des Verfahrens gegeben. Warum das Verfahren allgemein nicht konvergieren kann, wird in 3.8.8 erklärt. Es gibt damit keine optimale oder empfehlenswerte Wahl der Zeichen. In Abhängigkeit der gewählten Zeichen variiert auch die Ähnlichkeit der Trajektorien. Das Verfahren ist somit nicht objektiv und wird in der weiteren Auswertung verworfen.

## 3.8 Konstruierte Beispiele für Zeichensätze

Bereits in 3.7.1 wird beschrieben, dass die Implementierung nur prototypisch durchgeführt wird, da das Anpassen an einen Zeichensatz auf zu viele Probleme gestoßen ist und damit nicht als sinnvoll für die Ähnlichkeitsberechnung anzunehmen ist. In 3.8.1 wird auf die Abhängigkeit der Auswahl des Zeichensatzes und den Trajektorien voneinander eingegangen. Wird der zu vergleichende Trajektoriensatz um weitere Trajektorien ergänzt, hat dies ebenfalls Auswirkungen auf die Ähnlichkeit, was in 3.8.2 thematisiert wird. Auch die Größe der verwendeten Zeichen spielt eine Rolle, was in 3.8.3 erklärt wird. Da die Trajektorien nie perfekt mit einem Zeichen zu approximieren sind, müssen Toleranzen festgelegt werden. Auch diese beeinflussen das Ergebnis, was in 3.8.4 beschrieben wird. Es gibt zudem verschiedene Methoden, anhand welcher das nächste Zeichen auszuwählen ist und die Auswahl gegebenenfalls voneinander abweicht. Dies wird anhand eines Beispiels in 3.8.5 erklärt. Außerdem treten Probleme bezüglich der Vergleichbarkeit auf, wenn Trajektorien mit verschiedenen vielen Zeichen approximiert werden, dies wird in 3.8.6 erläutert. Die Gestalt der Zeichen spielt ebenfalls eine wichtige Rolle, was in 3.8.7 erläutert wird. Das Hauptproblem des Verfahrens ist die fehlende Konvergenz, die in 3.8.8 erklärt wird.

### 3.8.1 Abhängigkeit von Zeichensatzwahl

Je nach Wahl des Zeichensatzes, verändert sich die Ähnlichkeit von zu vergleichenden Trajektorien. Für das Beispiel werden zwei Zeichen festgelegt, ohne die Gestalt der Trajektorien zu kennen. In Abbildung 3.22 sind dieselben Trajektorien links und rechts durch unterschiedliche Zeichen zu approximieren. Im linken Fall stehen eine senkrechte und eine waagerechte Linie zur Auswahl. Im rechten Fall dürfen eine waagerechte Linie und eine leicht geneigte Linie verwendet werden. Im linken Fall entspricht die blaue Trajektorie genau der Aneinanderreihung waagerechter Linien.

Für die rote Trajektorie entspricht jedes einzelne Element eher einer waagerechten als einer senkrechten Linie. Deshalb werden auch für die rote Trajektorie nur waagerechte Elemente verwendet. Bei Verwendung dieses Zeichensatzes sind die beiden Trajektorien identisch. Anders verhält sich der rechte Fall. Die blaue Trajektorie kann erneut mit dem waagerechten Zeichen angenähert werden. Für die rote Trajektorie sind in zwei Abschnitten die geneigte Linie eine bessere Näherung als die waagerechte Linie, da sie mit dem Verlauf der Trajektorie übereinstimmen. Für die anderen beiden Abschnitte ist das waagerechte Zeichen weiterhin die beste Näherung. Die Trajektorien sind, da sie in zwei Abschnitten verschiedene Zeichen verwenden, hier nicht mehr identisch.

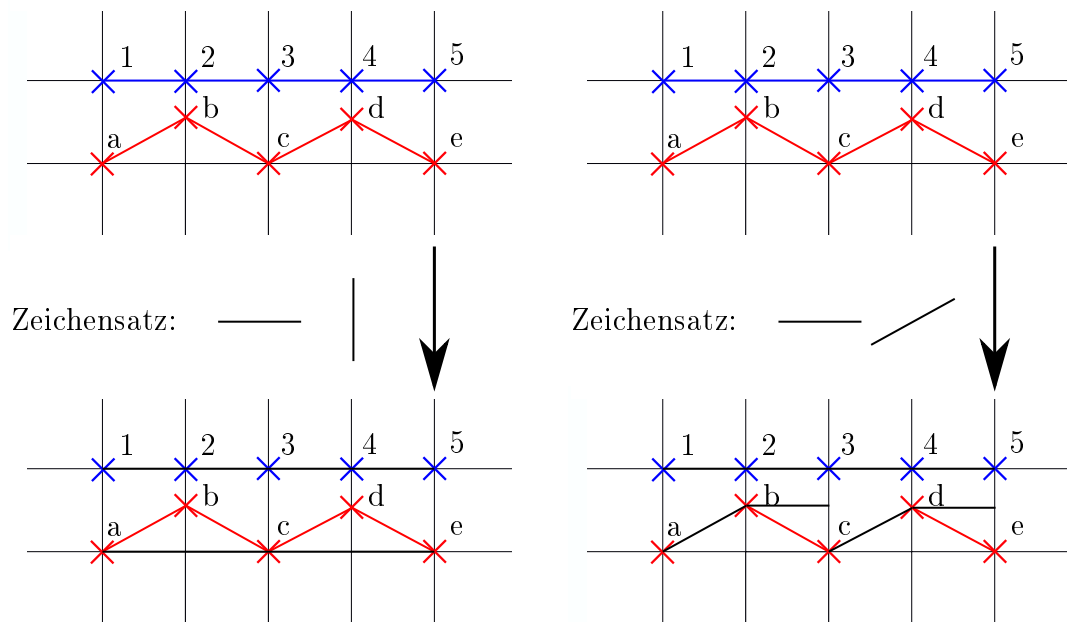


Abbildung 3.22: Beispiel für verschiedene Auswahl von Zeichensätzen.

In Abhängigkeit von der Wahl des Zeichensatzes verändert sich die berechnete Ähnlichkeit. Hier wurden die Wahl jeweils willkürlich ohne Kenntnis über die Trajektoriengestalt getroffen. Alternativ kann mit der bekannten Gestalt der Trajektorie ein Zeichensatz definiert werden beziehungsweise der Prozess von Zeichenauswahl und Trajektoriengestalt und -anpassung beliebig häufig iteriert werden. Dieser Fall wird im folgenden Kapitel 3.8.2 betrachtet.

### 3.8.2 Zeichensatziteration

Die Gestalt der Trajektorien kann als Anhaltspunkt für den zu wählenden Zeichensatz dienen. Abbildung 3.23 zeigt einen solchen Vorgang. Hierbei ist zunächst ausschlaggebend, welche Trajektorie als Referenz gewählt wird. Im Beispiel soll dies die blaue Trajektorie sein. Die blaue Trajektorie wird in ihrer gesamten Gestalt durch die vierfache Wiederholung einer waagerechten Linie beschrieben. Damit ist das erste Zeichen festgelegt. Wird nun versucht, die rote Trajektorie mit diesem Zeichen zu beschreiben, steht nur ein Zeichen zur Auswahl. Mit nur einem einzigen Zeichen

als mögliche Option sind beliebige Trajektorien zwangsläufig identisch. Daher wird mindestens ein weiteres Zeichen benötigt. In einem zweiten Schritt könnte ein weiteres Zeichen so gewählt werden, dass es dem ersten Element der roten Trajektorie entspricht. Die blaue Trajektorie wird weiterhin durch das erste Zeichen beschrieben. Die rote Trajektorie wird in ihrem ersten und dritten Element besser durch das neue Zeichen genähert. In Abschnitt zwei und vier bleibt jedoch die Waagerechte die bessere Näherung. Die Trajektorien sind nun durch zwei identische Zeichen an gleicher Position und zwei verschiedene Zeichen beschrieben. Wird ein weiteres Zeichen eingeführt, das zum Beispiel wie die Verbindung zwischen Punkt b und Punkt c aussieht, werden beide Trajektorien durch komplett verschiedene Zeichen beschrieben.

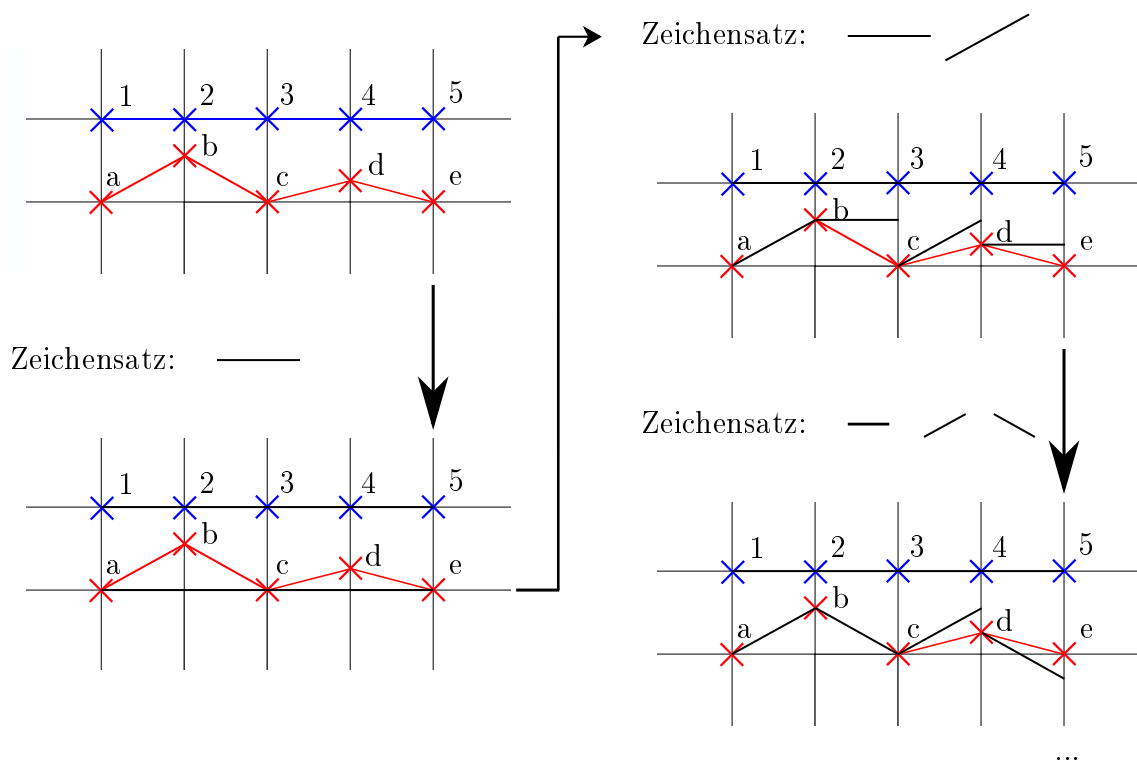


Abbildung 3.23: Iterationsschritt für die Festlegung eines Zeichensatzes.

Die Iteration kann beliebig fortgeführt werden. Die Ähnlichkeit hängt vom Zeitpunkt des Abbruchs ab. Während die Trajektorien im ersten Schritt, wenn auch hier mit nur einem Zeichen wenig sinnvoll, als identisch eingestuft werden, nimmt die Ähnlichkeit mit einer wachsenden Anzahl von Zeichen in den meisten Fällen ab. Insgesamt existiert die Tendenz, dass ein großer Zeichensatz eine geringe Ähnlichkeit bedingt. Dies ist aber nicht in allen Fällen gültig, was in 3.8.8 gezeigt wird. Daher kann für die Iteration kein mathematisches Abbruchkriterium gefunden werden. Hier wird zwar nicht der Zeichensatz direkt willkürlich festgelegt, die Anzahl der Iteration und die Wahl der Referenztrajektorie sind jedoch Variablen, für die es keine allgemeingültige richtige oder gute Wahl gibt.

### 3.8.3 Zeichengröße

Für die Implementierung in MATLAB wurde eine dynamische Größe der Zeichen, je von gegebenem Datenpunkt zum nächsten, gewählt. Da die Punkte nicht äquidistant verteilt sind, verändert sich die Größe der Zeichen in jedem Abschnitt. Ob eine dynamische oder statische Wahl der Größe getroffen wird, verändert bereits das Ergebnis. Abbildung 3.24 zeigt links die Wahl einer dynamischen Größe. Die gestrichelten Linien markieren dabei die Trennung zwischen einzelnen Zeichen. Die Punkte a bis g sind verschieden weit voneinander entfernt. Per Definition soll zwischen jedem gegebenen Punkt ein Zeichen verwendet werden. Bei der dynamischen Größenwahl sind dies zunächst vier waagerechte Linien, dann eine leicht aufwärts gerichtete und eine leicht abwärts gerichtete. Die waagerechten Linien sind durch den geringeren Abstand der Punkte kleiner. Für die blaue Trajektorie werden sechs waagerechte Linien verwendet, um jeweils den Abschnitt zwischen den Punkten 1 bis 7 durch ein Zeichen anzunähern. Mit der dynamischen Größenwahl sind die ersten vier Zeichen für beide Trajektorien gleich, die letzten beiden Zeichen unterscheiden sich.

Für die statische Zeichenwahl, rechts in Abbildung 3.24, sind feste Längen für ein Zeichen vorgegeben. Hier besteht eine Zeichengröße aus einer Kästchenlänge. Die blaue Trajektorie wird somit in vier waagerechte Zeichen aufgeteilt. Die rote Linie wird durch zwei waagerechte, eine aufwärts gerichtete und eine abwärts gerichtete Linie angenähert. Hier sind zwei von insgesamt vier verwendeten Zeichen identisch.

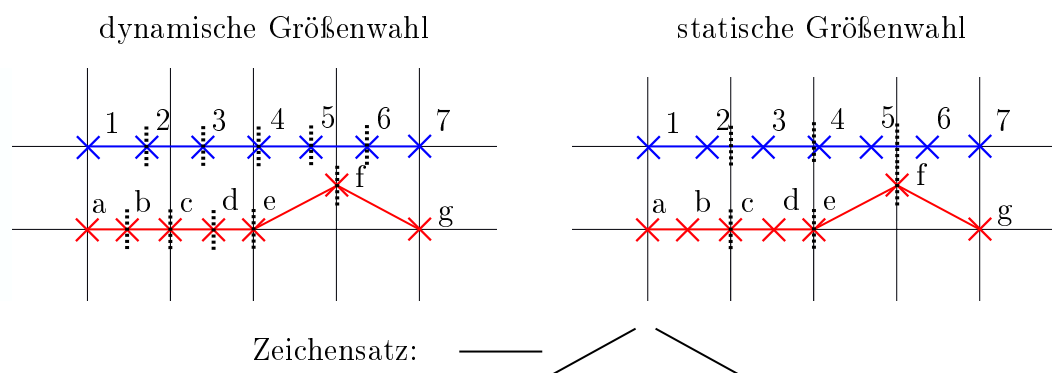


Abbildung 3.24: Vergleich von dynamischer und statischer Zeichengröße.

Im Beispiel sind im statischen Fall zwei von vier Zeichen identisch, im dynamischen Fall vier von insgesamt sechs Zeichen gleich. Für den dynamischen Fall könnte die Punktdichte zwischen Punkt 1 und 4 beziehungsweise zwischen Punkt a und e weiter erhöht werden, sodass mehr Zeichen gleich sind und nur die letzten zwei Zeichen der gesamten Zeichenkette abweichen. Welche Wahl, statisch oder dynamisch, die bessere ist, ist allgemein nicht zu sagen. Beide Verfahren berechnen für dieselben Trajektorien verschiedene Ähnlichkeiten.

Wird eine feste Größe der Zeichen, also ein statisches Verfahren gewählt, ändert sich die Ähnlichkeit in Abhängigkeit der Größenwahl. In Abbildung 3.25 wurden zwei verschiedene Zeichengrößen gewählt. Für die blaue Trajektorie spielt es keine Rolle, welche Zeichengröße gewählt wird. Im linken Bild, bei einem Zeichen pro vier Kästchen, benötigt die blaue Trajektorie die waagerechte Linie einmalig. Im rechten Fall werden vier waagerechte Linien aneinandergereiht, um die blaue Trajektorie zu

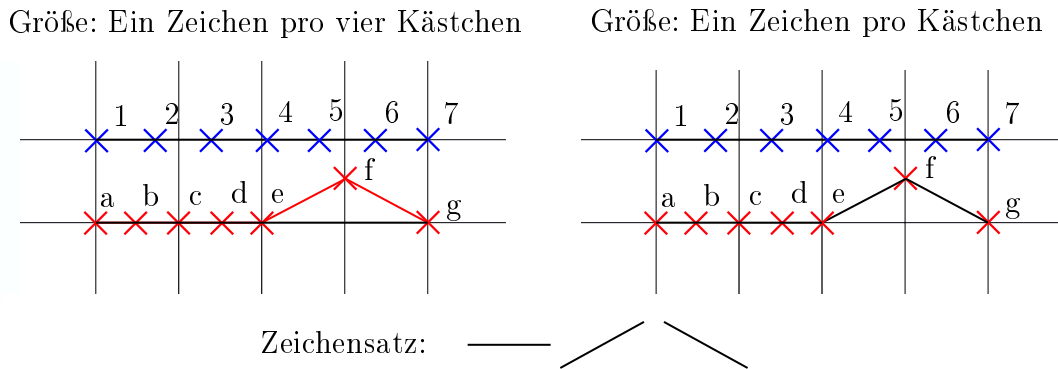


Abbildung 3.25: Wahl von verschiedenen Zeichengrößen.

approximieren. Anders verhält es sich bei der roten Trajektorie. Im rechten Fall, bei einem Zeichen pro Kästchen, wird die rote Trajektorie durch zwei waagerechte Linien, dann eine aufwärts gerichtete gefolgt von einer abwärts gerichteten Linie approximiert. Soll nur ein Zeichen pro vier Kästchen verwendet werden, kann die rote Trajektorie nicht ohne Abweichung angenähert werden. Am besten entspricht eine waagerechte Linie dem Verlauf der Trajektorie. Die erste Hälfte der Trajektorie entspricht genau dem Zeichen. Die zweite Hälfte kann mit dem aufwärts- und abwärts gerichteten Teil zusammen genommen auch als waagerechter Verlauf approximiert werden. Entsprechend ist das Ergebnis im linken Fall, dass die Trajektorien identisch sind. Beide Trajektorien werden mit demselben Zeichen angenähert. Im rechten Fall sind die Trajektorien nur in der ersten Hälfte gleich und im zweiten Teil verschieden voneinander. Die Wahl der Zeichengröße verändert somit die Ähnlichkeit.

Wird die Auswahl der Zeichen ohne Kenntnisse der zu vergleichenden Trajektorien getroffen, ist diese Wahl willkürlich und liefert in Abhängigkeit der Trajektoriengestalt verschiedene Ergebnisse. Erst mit Kenntnis der Trajektorien kann die Größe der zu verwendenden Zeichen festgelegt werden. So können als relevant erachtete Abschnitte entsprechend stärker gewichtet werden als andere Abschnitte. Ein solches Verfahren ist aber nicht allgemein gültig. Es muss immer Kenntnis über die Gestalt der Trajektorien vorhanden sein. Erst dann kann die Größe festgelegt werden. Mit weiteren Trajektorien verändert sich die zu treffende Wahl. Dadurch ist keine Vergleichbarkeit von verschiedenen Trajektoriensätzen gegeben. Insgesamt stellt die passende Größenauswahl eine Herausforderung dar, die für jeden neuen Datensatz angepasst werden muss und nie allgemein gültig, abgesehen von einer willkürlichen Wahl die beliebige Ergebnisse liefert, angewendet werden kann.

### 3.8.4 Toleranz beim Anpassen

Der Trajektorienverlauf kann nicht immer perfekt durch ein vorgegebenes Zeichen angepasst werden. Würde man dies tun, bräuhete es unendlich viele Zeichen oder unendlich kleine Zeichen, um die Trajektorienverläufe ohne Abweichung zu approximieren. Auch bei einem fest vorgegebenen Zeichensatz müsste nur die Frage nach dem Zeichen, das den betrachteten Abschnitt am besten approximiert, gestellt werden, nicht aber danach, wie groß die Abweichung von Trajektorie und Zeichen ist. Um überhaupt einen geeigneten Zeichensatz festlegen zu können, können die Trajektorien herangezogen werden. Dass es nicht zielführend ist, einen (beliebigen) Zei-



chensatz für den Vergleich festzulegen, wurde bereits in 3.8.2 ausführlich erklärt. Alternativ kann ein Zeichen festgelegt werden, das die Trajektorie in einem ersten Abschnitt sehr gut approximiert. Eine gute Approximation befindet sich innerhalb der vorgegebenen (willkürlich gewählten) Toleranz. Für den zweiten Abschnitt wird verglichen, ob das Zeichen des ersten Abschnitts diesem entsprechen kann oder ob ein neues Zeichen definiert werden muss. So wird die gesamte Trajektorie durchgegangen und für jeden weiteren Abschnitt ein bereits vorhandenes Zeichen benutzt oder ein neues festgelegt.

Der Wert für die Toleranz ist subjektiv und stark von den Trajektorien abhängig. Entsprechend der vorgegebenen Toleranz und Zeichengröße wird die Ähnlichkeit beeinflusst und liefert andere Ergebnisse. Daher ist es schwierig, für die Toleranzen einen allgemeingültigen Maßstab zu finden.

### 3.8.5 Anpassungsverfahren

Es gibt zwei wesentliche Verfahren, nach denen die Zeichen für die Approximation einer Trajektorie ausgewählt werden können. Entweder wird die Trajektorie für sich in Abschnitte geteilt und für jeden Abschnitt wird ein Zeichen festgelegt, das die Gestalt der Trajektorie in diesem Abschnitt bestmöglich annähert oder für jeden Schritt wird das Zeichen verwendet, das nach der Durchführung dieses Schrittes die geringste Differenz zwischen gegebener Trajektorie und Zeichenkette zur Folge hat. In den folgenden zwei Absätzen wird auf die beiden Verfahren und deren Unterschiede detailliert eingegangen. Die Entscheidung für ein Verfahren verändert die berechnete Ähnlichkeit.

**Gestaltnäherung pro Abschnitt:** Abbildung 3.26 zeigt in rot eine Trajektorie, die mit dem gegebenen Zeichensatz aus drei Zeichen approximiert werden soll. Die Trajektorie wurde dafür in Abschnitte mit je einer Kästchenlänge und -höhe eingeteilt. Abgesehen von zwei Abschnitten wird die Trajektorie mit den gegebenen Zeichen genau getroffen. Die nicht genau passenden Abschnitte werden am besten durch die waagerechte Linie genähert. Damit kommt es bei der Anpassung durch

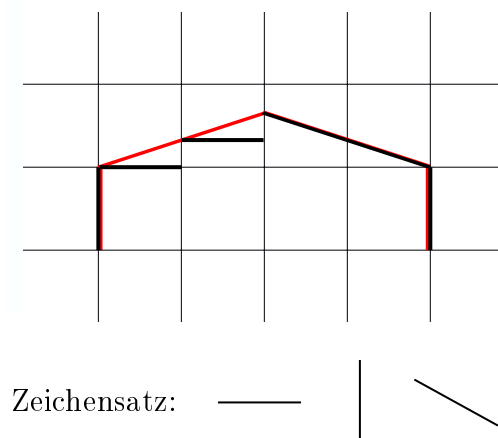


Abbildung 3.26: Abschnittsweise Anpassung an Trajektorie.

einen Zeichensatz zu zwei Sprüngen im approximierten Trajektorienverlauf. Der Anknüpfungspunkt für jedes neue Zeichen ist hier immer der Trajektorienbeginn im jeweiligen Abschnitt. Da das Zeichen zuvor nicht deckungsgleich war, kommt es zu Sprüngen, da Trajektorie und Zeichen an unterschiedliche Punkten enden. Dies ist kein Problem, da für das Verfahren lediglich die Aneinanderreihung von Zeichen relevant ist. Insgesamt wird die Trajektorie durch die Verwendung des zweiten Zeichens, einer senkrechten Linie, der zweifachen Verwendung der waagerechten Linie, doppelter Verwendung der abwärts gerichteten Linie und einer weiteren senkrechten Linie genähert.

**Kleinste Differenz:** Dieselbe Trajektorie wird nun mit der jeweils kleinsten Differenz angenähert. Abbildung 3.27 zeigt die entsprechende Trajektorie, auch hier wieder rot gezeichnet. Angenommen wird, dass die Trajektorie links unten beginnt und rechts unten endet. Die ersten drei Elemente sind verglichen mit der Gestaltanpassung unverändert geblieben, wenn auch die Position des dritten Zeichens eine andere ist. Der Anknüpfungspunkt ist bei diesem Verfahren immer das Ende des zuletzt verwendeten Zeichens. Das erste Element trifft die Trajektorie exakt, während die nächsten vier Elemente abweichen. Zufällig sind die ersten beiden abweichenden Elemente auch die, die bei der Gestaltanpassung verwendet wurden. In der zweiten Hälfte weichen die Zeichen für das Verfahren der kleinsten Differenz vom Verfahren der Gestaltanpassung ab, während das letzte Zeichen in beiden Verfahren wieder identisch ist.

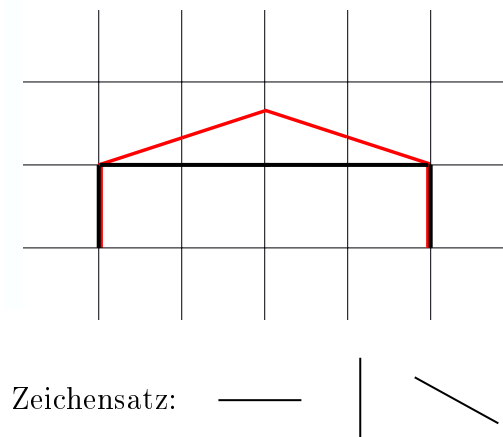


Abbildung 3.27: Anpassung per Verfahren der kleinsten Differenz.

Während die Näherung pro Abschnitt die Gestalt der Trajektorien in ihrer Wichtigkeit in den Vordergrund stellt, ist das Verfahren mit der kleinsten Differenz auf den geometrisch kleinsten Abstand zwischen den Trajektorien und dem Zeichensatz bedacht. Beide Verfahren können zu völlig unterschiedlichen Ergebnissen führen. Hier wird dieselbe Trajektorie approximiert, zwei von sechs verwendeten Zeichen sind in Abhängigkeit vom gewählten Verfahren anders. Das geeignetere Verfahren bezüglich Robustheit ist das Anstreben einer kleinen Differenz zwischen Zeichen und Trajektorie. Jedoch kann wie in Abbildung 3.27 eine ungünstige Kombination von Zeichen

dazu führen, dass das Verfahren der Gestaltanpassung die bessere Näherung darstellt. Während in Abbildung 3.26 vier von sechs Zeichen den Trajektorienverlauf exakt treffen, sind es bei der kleinsten Differenz nur zwei von sechs Zeichen. Die abweichenden vier Zeichen weisen zudem eine größere Differenz zum Trajektorienverlauf auf als die Approximation in Abbildung 3.26.

### 3.8.6 Verschiedene Zeichenkettenlängen

Bestehen Zeichenketten aus unterschiedlich vielen Zeichen, gibt es in der Levenshtein Metrik beziehungsweise Editierdistanz die Möglichkeit, Zeichen hinzuzufügen, zu löschen oder zu ersetzen. Üblicherweise sind diese drei Operationen gleich gewichtet. Bezüglich Flugtrajektorien stellt sich die Frage, ob die gleiche Gewichtung der Operationen sinnvoll ist beziehungsweise die Operationen überhaupt anwendbar sind. Einen Bereich zu ersetzen, ist bezüglich einer Trajektorie entlang derer ein Flugzeug fliegt, gut vorstellbar. An Start und Ziel wird keine Veränderung vorgenommen. Lediglich der Verlauf zwischen ausgewählten Punkten wird verändert. Auch das Hinzufügen von Zeichen ist kein Problem. Dem bisherigen Trajektorienverlauf wird weitere Strecke hinzugefügt. Auch dies ist in der Realität umsetzbar. Problematisch hingegen ist das Löschen von Zeichen. Innerhalb der Zeichenkette selbst kann diese Operation durchgeführt werden. Ein Flugzeug kann im Allgemeinen eine solche Trajektorie nicht mehr abfliegen.

Entsprechend müssen bei unterschiedlicher Länge der Zeichenketten dem kürzeren Datensatz Zeichen hinzugefügt und auf die Lösch-Operation verzichtet werden, da diese im Sachzusammenhang nicht sinnvoll ist. Die Operation „Ersetzen“ darf ebenfalls angewendet werden. Das Hinzufügen beziehungsweise Ersetzen von Zeichen ist aber auch kritisch zu sehen, denn die Gestalt des Zeichens selbst, wie im folgenden Abschnitt 3.8.7 erklärt, spielt dabei eine wichtige Rolle.

### 3.8.7 Gestalt des Zeichens

Im lateinischen Alphabet gibt es 26 Buchstaben, vier weitere wenn die Zeichen ä, ö, ü und ß hinzugezählt werden. Jedes dieser Zeichen hat eine feste Bedeutung: Wie es alleine ausgesprochen wird, wo es in Wörtern steht, welche Bedeutung die Wörter durch genau diese Aneinanderreihung von Zeichen erhalten. Die Gestalt der Buchstaben spielt dabei keine Rolle. Aus Gewohnheit und Erziehung weisen wir dem Buchstaben „a“ bestimmte Eigenschaften zu. Dass „A“ und „a“ eng miteinander verknüpft sind, ist ebenfalls erlerntes Wissen und an der Gestalt nicht zu erkennen. Ein Zeichen besteht aus drei Linien, das andere aus einem Kreis mit einem Strich daran, je nach Schriftart noch weiteren Elementen. Alleine durch die Form existiert kein Zusammenhang. Es wäre problemlos möglich, in einem Text alle „a“ und „b“ miteinander zu vertauschen. Der Leser muss lediglich wissen, dass diese Änderung vorgenommen wurde und braucht zum Lesen entsprechend länger, da die Veränderung nicht seiner Gewohnheit entspricht. Die Informationen des Textes bleiben durch den Tausch aber unverändert.

Die Annäherung der Trajektorien durch einen Zeichensatz sieht ebenfalls keinen Zusammenhang zwischen den Zeichen vor. Die Zeichen sind frei gewählt und bestimmte Abschnitte werden dem bestmöglich passenden Zeichen zugeordnet. Das

Ersetzen beliebiger Zeichen ist möglich. So ist das Ersetzen von „a“ mit „b“ gleichwertig gegenüber dem Tausch von „c“ mit „d“. Genau diese Eigenschaft besitzen Flugtrajektorien nicht, wodurch das Anpassen an einen Zeichensatz als problematisch zu sehen ist. Abbildung 3.28 zeigt drei Trajektorienverläufe, die durch vier vorgegebene Zeichen approximiert werden sollen. Jeder Trajektorienausschnitt wird durch drei Zeichen genau beschrieben. Dieselben Zeichen sind jeweils in derselben Farbe eingezeichnet. Werden der mittlere und der rechte Trajektorienverlauf verglichen, sind zwei Zeichen identisch und das mittlere Zeichen verschieden. Verglichen mit dem Verlauf links sind die beiden Trajektorien sich nach dieser Vergleichsmethode ähnlicher. Mit der linken Zeichenkette stimmt nur das erste Zeichen überein und die anderen beiden sind verschieden. Die Gestalt der beiden Kurvenelemente ist jedoch entscheidend. Im Beispiel verursacht das Ersetzen der einen Kurve durch die andere eine Richtungsänderung von  $180^\circ$ . Alle folgenden Elemente erfahren durch das Austauschen eines Zeichens eine erhebliche Veränderung, hier eine Richtungsumkehrung. Die links abgebildete, gerade verlaufende Trajektorie ist den beiden anderen Trajektorien ähnlicher, da sie nur eine Richtungsänderung von  $90^\circ$  bedeutet.

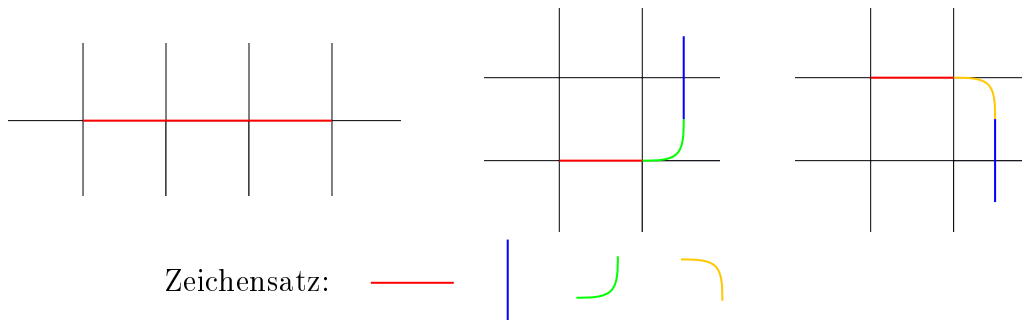


Abbildung 3.28: Trajektorien approximiert durch Zeichensatz mit vier Zeichen.

Für den Vergleich von Trajektorien ist damit die Gestalt der Zeichen und durch welches Zeichen sie jeweils ersetzt werden, entscheidend. Anders als bei einem Alphabet hängt hier die Gestalt jedes einzelnen Zeichens mit der erzeugten Gesamtform direkt zusammen. Es müssen Abstufungen vorgesehen werden, welche Zeichen sich ähnlicher sind als andere. Im konstruierten Extremfall sind sonst eine  $180^\circ$ -Kehre und eine Kursveränderung von  $1^\circ$  als gleichwertig einzustufen, was sie im Flugbetrieb nicht sind. Für die Feststellung der Ähnlichkeit der Zeichen kann zum Beispiel das Verfahren der Energieminimierung eingesetzt werden, was es in 3.4.4 bereits wird. Damit geht aber der Mehrwert der Zeichen verloren.

### 3.8.8 Konvergenzproblem

Werden die Trajektorien als Ganzes verglichen, wird die Übereinstimmung im Allgemeinen gering sein. Je feiner die Trajektorien zerteilt und per festgelegten Zeichen angenähert werden, desto ähnlicher sollten sich die Trajektorien werden, denn bei einer unendlich feinen Unterteilung ist der überwiegende Teil der Trajektorien mit demselben Zeichen zu approximieren. Dies ist eine notwendige Eigenschaft für ein sinnvoll einsetzbares Verfahren. Wie zum Beispiel bei der Methode der Finiten Elemente gibt es die theoretische Überlegung, dass eine unendlich feine Zerteilung zu

einer perfekten Berechnung führt. Da in der Praxis eine unendlich kleine Unterteilung nicht möglich ist, werden hinreichend kleine Elemente verwendet, die ein ausreichend genaues Ergebnis liefern. Wird diese Konvergenz auf die Anpassung durch einen Zeichensatz übertragen, bedeutet dies, dass eine feinere Aufteilung zu einer höheren Ähnlichkeit führen sollte. Ein Gegenbeispiel ist in Abbildung 3.29 zu sehen. Der Trajektorienausschnitt ist beliebig klein gewählt. Wird dieser Abschnitt in nur ein Zeichen unterteilt, sind die beiden Zeichen verschieden voneinander. Es folgt die jeweils feinere Einteilung in zwei, drei, vier und sechs gleich große Zeichen. Identische Zeichen an gleicher Position sind jeweils rot markiert, unterschiedliche Zeichen in blau. Bei zwei Zeichen entspricht ein Zeichen dem verwendeten in der anderen Trajektorie. Bei drei Zeichen ist es ebenfalls eins, das übereinstimmt. Bei vier Zeichen insgesamt sind es drei gleiche Zeichen und für die Unterteilung in sechs Zeichen sind vier identisch. Werden alle Anteile auf das kleinste gemeinsame Vielfache gebracht, ist zu sehen, dass keine Konvergenz vorliegt. Mit einer feineren Unterteilung springen die Werte zwischen mehr und weniger Ähnlichkeit.

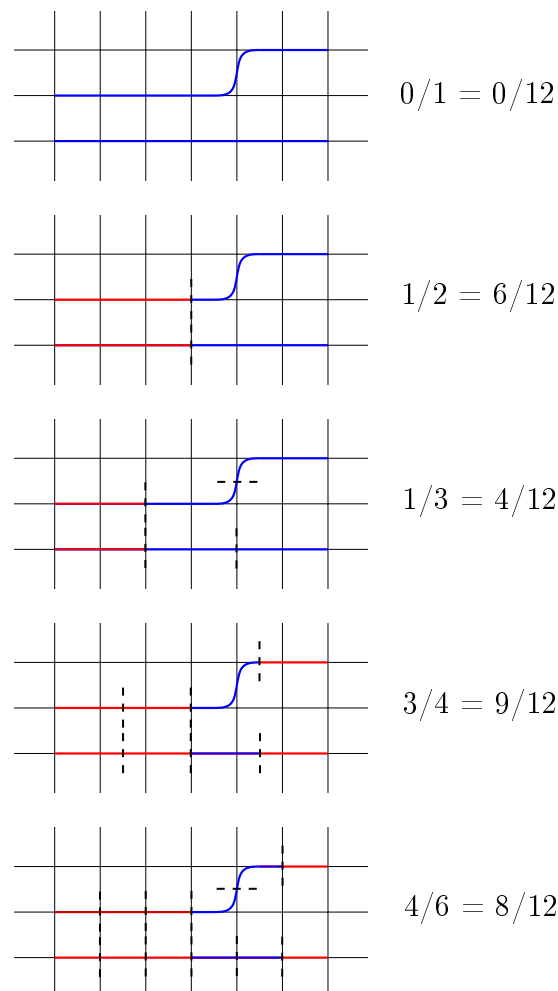


Abbildung 3.29: Schrittweise Verwendung von kleineren Zeichen.

In Abschnitt 3.8.1 bis 3.8.7 wurden in vielen Beispielen gezeigt, dass die Ähnlichkeit durch das Anpassen an einen Zeichensatz von einer Wahl vieler Parameter abhängt. Für die Bestimmung dieser Parameter gibt es keine allgemeingültige Wahl, sondern

die Parameter müssen immer auf die gegebenen Trajektorien angepasst werden und sind auch dann nicht eindeutig. Es verbleiben viele willkürlich anzunehmende Parameter. Da die Ähnlichkeit von den gewählten Parametern direkt abhängt, ist auch die Ähnlichkeit ein subjektiver Wert. Da das Verfahren zudem keine Konvergenz aufweist, wird der gesamte Ansatz als nicht passend für den Vergleich der gegebenen Flugtrajektorien eingestuft und das Verfahren nicht weiter betrachtet.

## 3.9 Clusteranalyse

Es besteht ein großes Interesse, Flugtrajektorien bezüglich ihrer geometrischen Nähe zu untersuchen. Befinden sich weitere Trajektorien nahe an einer untersuchten Trajektorie, gilt diese als robust. Dafür muss aber genau definiert werden, welche Trajektorien als nahe beieinander oder voneinander entfernt liegen. Dieser Wert ist nicht absolut festzulegen, sondern immer individuell auf den gegebenen Datensatz anzupassen. Dafür geeignet ist neben der Ähnlichkeitsberechnung eine Clusteranalyse. Aufgrund verschiedener einstellbarer Parameter werden die Trajektorien gruppiert. So kann definiert werden, welche Trajektorien nahe beisammen sind und welche nicht. Wie in Abschnitt 2.5 gibt es sehr viele Optionen und Möglichkeiten bei der Clusterbildung. In 3.9.1 wird diskutiert, ob eine harte oder eine weiche Clusterbildung für die Anwendung besser geeignet ist. Eine geeignete Anzahl von Clustern wird in 3.9.2 bestimmt und in 3.9.3 wird sich mit dem Löschen von Daten auseinander gesetzt. Das variierende Mittel ist eine weitere Variante der Clusterbildung. Ob und wie sie eingesetzt werden kann, wird in 3.9.4 diskutiert bevor in 3.9.5 auf die Clusterform eingegangen wird. Hierarchische Cluster werden in 3.9.6 erläutert. Schlussendlich werden die gewählten Optionen in 3.9.7 zusammengefasst.

### 3.9.1 Harte oder weiche Cluster

Bei harten Clustern gehören die zu unterscheidenden Daten in genau ein Cluster. Bei weichen Clustern dürfen Daten beliebig vielen Clustern angehören. Anhand eines Beispiels wird gezeigt, dass weiche Cluster für den vorgesehenen Anwendungsfall in dieser Arbeit die bessere Wahl sind als harte Cluster. Dafür wurden die Trajektorien zu Punkten vereinfacht. Wie die zu vergleichenden Objekte aussehen, ist für die Erklärung des Verfahrens unerheblich. Abbildung 3.30 zeigt verschiedene Punkte, durch X markiert, die es bezüglich ihrer Robustheit zu untersuchen gilt. Robust sind alle Punkte, die weitere Punkte in ihrer Nähe haben. Für das Beispiel sollen drei Cluster gebildet werden. Bei einer harten Unterscheidung werden vier Punkte zu einem Cluster zusammengefasst. Alle beinhalteten Punkte sind durch einen Kreis mit dem Radius von einer Kästchenlänge zu erfassen. Es verbleiben zwei Cluster für zwei weitere Punkte. Diese Punkte bilden daher alleine ein Cluster. Bei der Unterteilung in weiche Cluster wird das rote Cluster auf zwei Punkte erweitert. Ein Punkt erhält hier eine doppelte Zugehörigkeit. Da das rot eingezeichnete Cluster denselben Kreisradius besitzt wie das blau eingezeichnete Cluster und nicht größer ist, darf diese Erweiterung vorgenommen werden. Es müssen die kleinstmöglichen Kreise gewählt werden. Ein größerer Radius als eine Kästchenlänge wäre unzulässig. Gleich große Kreise oder geringfügig kleinere Kreise sind als Erweiterung von einzelnen Clustern

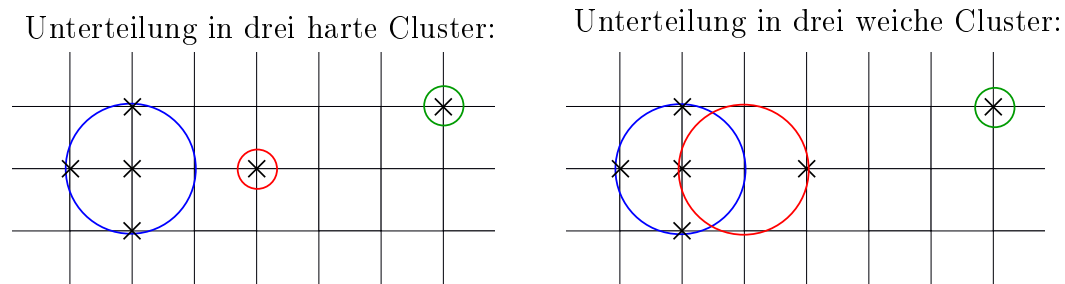


Abbildung 3.30: Beispiel für Clusterbildung für gegebene Punkte.

immer zulässig. Das grün eingezeichnete Cluster bleibt unverändert. Wird die harte Clusteranalyse angewandt, gibt es vier Punkte, die nah beieinander liegen und zwei Punkte, die als einzeln und damit nicht robust erkannt werden. Trotzdem ist der rot eingekreiste Punkt näher an anderen Punkten als der grüne und damit robuster als der grün markierte Punkt, wenn er auch nicht so robust ist wie die Punkte innerhalb des blauen Clusters. Diese Abstufung erkennt die weiche Clusterbildung. Die blau markierten Punkte sind am robustesten, da sich innerhalb des Clusters vier Punkte befinden. Im roten Cluster befinden sich zwei Punkte, was weniger dicht ist als im blauen Cluster, jedoch einen großen Unterschied gegenüber dem Einzelpunkt ausmacht.

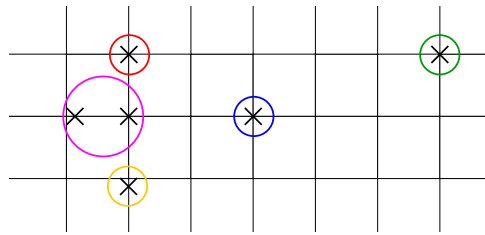
Es werden daher weiche Cluster verwendet, um diese Abstufung der Robustheit zu sehen. Da verschiedene Fälle untersucht werden, geht es nicht nur um einen optimalen, besonders robusten Punkt, den eine harte Clusteranalyse findet, sondern insbesondere um die Abstufung der Robustheit und Nähe, die durch eine weiche Clustereinteilung bestimmt wird.

### 3.9.2 Anzahl der Cluster

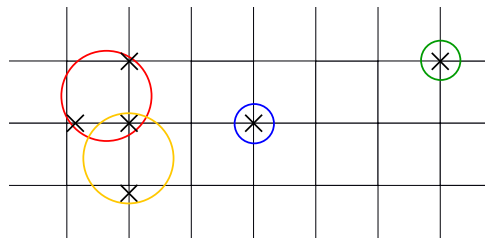
Einfach wäre es, eine Zahl vorzugeben und für beliebige Datensätze die Daten in entsprechend viele Cluster zu unterteilen. Da die Daten aber in ihrer Anzahl und ihrer Art sehr verschieden sein können, ist dies im Allgemeinen keine sinnvolle Herangehensweise. Die Anzahl der Cluster wird in Abhängigkeit der Daten festgelegt. In den beiden MATLAB-Skripten `Flugtrajektorien_durchgehen_ATM4E.m` beziehungsweise `Flugtrajektorien_durchgehen_FlyATM4E.m` wird immer mit der Einteilung in fünf Cluster begonnen. Dieser Startwert hat sich in vielen Tests mit den Daten aus ATM4E und FlyATM4E als ein guter Ausgangspunkt erwiesen. Sollte dieser Wert unpassend sein, wird dies erkannt und der Wert entsprechend verändert. Denn in Abhängigkeit der Überlappung der Cluster und der Anzahl der Daten in den Clustern wird anschließend in mehr oder weniger Cluster aufgeteilt und die Berechnung erneut durchgeführt. Abbildung 3.31 zeigt ein Beispiel, wie durch Bestimmung der Überlappung und die Anzahl der Cluster mit Einzelpunkten festgestellt werden kann, ob mehr oder weniger Cluster erforderlich sind.

Begonnen wird mit der Einteilung in fünf Cluster. Dies zeigt der oberste Abschnitt in Abbildung 3.31. Dabei entstehen vier Cluster mit nur einem Datenpunkt. Viele Cluster mit wenigen Datenpunkten sind ein Indiz dafür, dass in zu viele Cluster aufgeteilt wurde. Zudem überlappen die Cluster trotz zulässiger weicher Einteilung

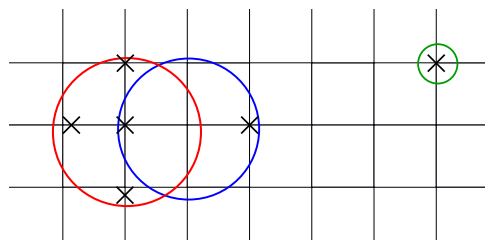
Einteilung in fünf Cluster:



Unterteilung in vier Cluster:



Unterteilung in drei Cluster:



Unterteilung in zwei Cluster:

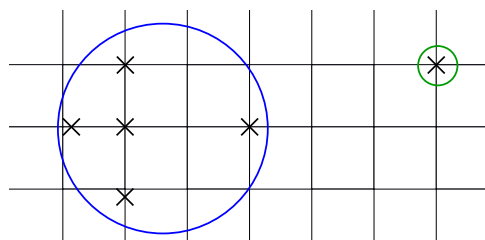


Abbildung 3.31: Bestimmung einer geeigneten Clusteranzahl.

nicht. Dies kann ebenfalls ein Hinweis darauf sein, dass mehr Cluster als benötigt verwendet wurden. Daher wird im nächsten Schritt in vier Cluster aufgeteilt. Dadurch bleiben zwei Datenpunkte einzeln. Zwei andere Punkte bilden aber nun ein gemeinsames Cluster und eine Überlappung ist neu hinzugekommen. Die Einteilung in vier Cluster liefert hier ein recht gutes Ergebnis, trotzdem wird auch die Einteilung in weniger Cluster betrachtet. Bei drei Clustern existiert nur noch ein Cluster mit einem einzigen Punkt, bei zwei Clustern wachsen alle bis auf einen Punkt zu einem Cluster zusammen. Hier wird deutlich, dass der Punkt rechts sehr weit von allen anderen Punkten entfernt ist und die Option des Löschens, im folgenden Abschnitt 3.9.3 erklärt, für ihn in Frage kommt. Da nun keine Cluster mehr überlappen, ist die Einteilung in zwei Cluster möglicherweise zu wenig. Eine gute Einteilung besteht daher für diese Daten in drei bis vier Clustern. Aber auch die anderen Aufteilungen liefern wichtige Informationen. So wird nicht nur eine Anzahl für Cluster festgelegt, sondern ein passender Bereich betrachtet. Denn nur in Kombination der Aufteilung



wird deutlich, dass der rechte Punkt sich sehr weit entfernt von allen anderen befindet. Die linken Punkte werden schnell zu einem Cluster gruppiert, befinden sich also nahe beieinander und sind daher robust. Der mittlere Punkt wird erst bei geringer Clusteranzahl hinzugezählt. Damit ist er weniger robust als die linken Punkte, trotzdem ist er robuster als der sich ganz rechts befindende Punkt.

### 3.9.3 Löschen von Daten

Bezüglich der Robustheitsuntersuchung ist das Löschen einzelner Trajektorien eine zielführende Option. Per Definition besitzen robuste Trajektorien in ihrer Nähe weitere Trajektorien, die durch eine ähnliche Gewichtung von Kosten- und Umweltfaktoren entstanden sind. Sind Trajektorien als weit entfernt von allen anderen Trajektorien einzustufen (wobei „weit entfernt“ einer weiteren Definition bedarf), können diese nicht robust sein und sind nicht von Interesse. Einzelne Trajektorien behindern zudem die Clusterbildung, denn zum Beispiel bei einer festgelegten Anzahl von Clustern (3.9.2) benötigen vereinzelte Trajektorien ihr eigenes Cluster. Wie in Abbildung 3.32 zu sehen, wurden für den Trajektoriensatz „00005\_GCLP\_ENVA“ die vereinzelten Trajektorien gelöscht. Die übrigen Trajektorien wurden in fünf Cluster aufgeteilt. Hier gibt es aus den eher gehäuft auftretenden Trajektorien verschiedene Cluster.

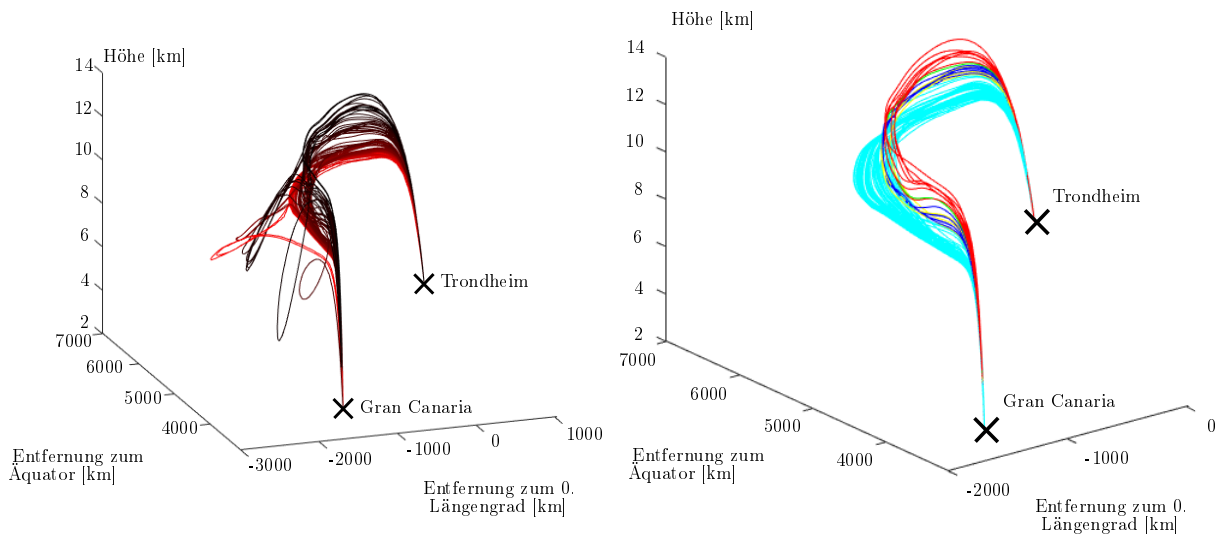


Abbildung 3.32: Clusterbildung am Beispiel von Datensatz „00005\_GCLP\_ENVA“ mit Löschen von Daten.

Anders verhalten sich die Trajektorien in Abbildung 3.33, bei denen keine Trajektorien gelöscht werden. Hier werden die Cluster für die Einzeltrajektorien verwendet und die Häufung der Trajektorien nicht feiner aufgeteilt. Eben dieser Bereich ist aber von Interesse. Daher wird sich für eine Löschoption entschieden. Es wird zunächst in MATLAB die festgelegte Anzahl von Clustern gebildet. Sollte nach der Clusterbildung ein Cluster aus einer einzelnen Trajektorie bestehen, so wird diese gelöscht und die Clusterbildung erneut durchgeführt. Dieser Vorgang wiederholt sich, bis in jedem Cluster mindestens zwei Trajektorien vorhanden sind.

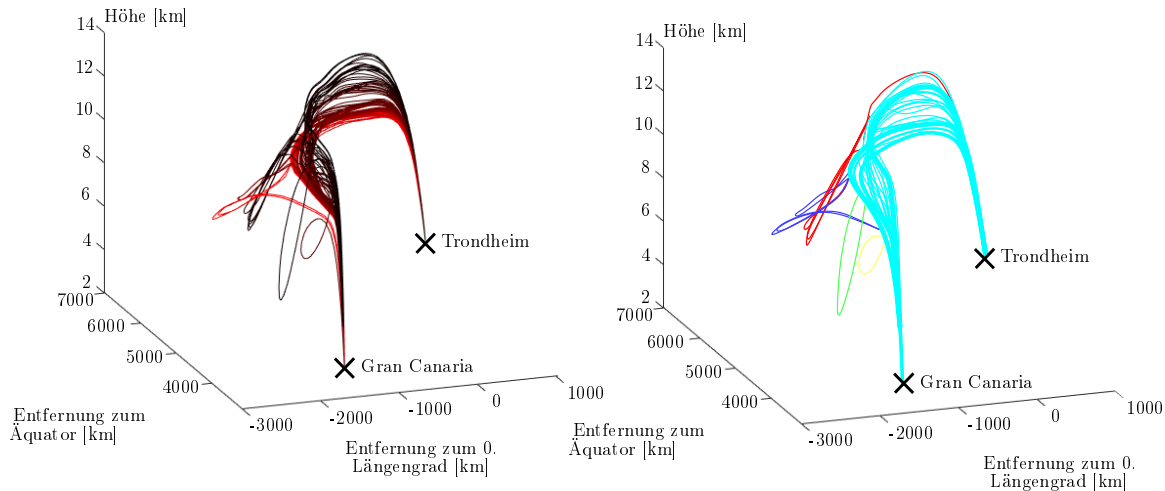


Abbildung 3.33: Clusterbildung am Beispiel von Datensatz „00005\_GCLP\_ENVA“ ohne Löschen von Daten.

### 3.9.4 Variierendes Mittel

Wie im Grundlagenkapitel 2.5.5 beschrieben, sind die Anzahl der Cluster und der maximale Abstand von Daten innerhalb eines Clusters konkurrierende Größen. Wird eine der beiden Größen festgelegt, ist die andere dadurch bestimmt. In 3.9.2 wird erklärt, dass in dieser Arbeit der Weg über die Anzahl der Cluster gewählt wird. In den Tests haben sich fünf Cluster als Ausgangswert als häufig passend erwiesen. Allgemein einen Startabstand festzulegen, erwies sich in den Tests als weniger geeignet, da dieser in Abhängigkeit des betrachteten Datensatzes sehr verschieden ist. Entsprechend werden keine MATLAB-Funktionen für das variierende Mittel verwendet, sondern die eigene Programmierung für die Anzahl der Cluster benutzt.

### 3.9.5 Form

Cluster können an verschiedene Formen angenähert werden. Die Grundform besteht im Zweidimensionalen aus einem Kreis und im Dreidimensionalen aus einer Kugel. Dadurch werden alle Daten erfasst, die sich in einem bestimmten Abstand zueinander befinden. Voraussetzung dafür ist, dass alle drei Dimensionen gleichwertig sind, was hier bei den betrachteten Daten der Fall ist. Eine Alternative wäre zum Beispiel der Höhe eine größere Gewichtung als der seitlichen Abweichungen zu geben. Entsprechend dieser Gewichtung würde die Kugel nun zu einem Ellipsoid gestaucht werden, da Daten mit größerer Höhenabweichung auszuschließen sind als seitlich abweichende Trajektorien. Da aber keine unterschiedliche Gewichtung vorgenommen wird, bleiben die Cluster in ihrer Grundform.

Zusätzlich bietet MATLAB Funktionen, um innerhalb der Cluster Datenverteilungen entsprechend vorgegebener Wahrscheinlichkeiten zu ordnen und so auch zu von Kreisen und Kugeln abweichende Formen zu erlangen. Dass die Flugtrajektorien in ATM4E und FlyATM4E bestimmten Verteilungen entsprechen und daran angepasst werden sollten, ist nicht bekannt. Daher werden die Cluster ohne Verteilungsfunktion entsprechend ihrer Grundform gebildet.

### 3.9.6 Hierarchische Aufteilung

Eine in MATLAB implementierte Funktion zur hierarchischen Aufteilung innerhalb der Cluster wird nicht verwendet. Indirekt findet die hierarchische Betrachtung aber Anwendung, denn durch das in 3.9.2 beschriebene Vergleichen von verschieden vielen Clustern, der Anzahl an Daten innerhalb eines Clusters sowie dem Abgleich der Überlappung werden die zunächst groben Cluster in feinere Strukturen aufgespalten. So wird in Abbildung 3.31 in zwei Cluster unterschieden. Mit der Verwendung von mehr Clustern wird das große Cluster aufgespalten und in weitere Teile zerlegt, deren Eigenschaften das weitere Vorgehen bestimmen und somit eine gegliederte Verteilung innerhalb der Cluster darstellen.

### 3.9.7 Zusammenfassung

Die Verwendung von weichen Clustern anstelle von einer harten Aufteilung ist für diesen Anwendungsfall sinnvoll. Durch verschiedene Tests wird herausgefunden, dass ein erster Ansatz mit dem Aufteilen der Daten in fünf Cluster oft sinnvoll ist. Ausgehend von diesem Startwert wird ein Bereich für die Anzahl der Cluster, in die die Daten sinnvoll aufgeteilt werden sollen, festgelegt. Veränderung von Clustern, Clusterüberlappungen und Einzeldaten innerhalb eines Clusters geben dabei Auskunft über das weitere Vorgehen. Wird so bestimmt, dass Daten gelöscht werden sollen, so wird dies durchgeführt und die gelöschten Daten werden als in keinem Fall robust klassifiziert. Eine detaillierte Erklärung zu dem programmierten MATLAB-Skript sowie ein Ablaufdiagramm folgen in Abschnitt 3.11.

## 3.10 Konstruierte Beispiele für die Clusteranalyse

Anders als die in Abschnitt 3.5 behandelte Ähnlichkeitsberechnung muss die Ausgangssituation für die Clusteranalyse deutlich weniger speziell sein. Damit gibt es auch viel weniger Fälle, in denen die Clusteranalyse scheitert oder kein sinnvolles Ergebnis liefert. Zudem kann die Analyse durch viele wählbare Optionen sehr gut auf die gegebenen Daten angepasst werden. In 3.10.1 wird auf die Vergleichbarkeit verschiedener Flüge eingegangen. In 3.10.2 wird erklärt, welche Vor- und Nachteile die relative Messung von Distanzen birgt. Problematisch in der Auswertung sind Daten mit einer Ordnungsstruktur, auf die in 3.10.3 eingegangen wird.

### 3.10.1 Vergleichbarkeit

Eine Clusteranalyse kann für sehr allgemeine Daten durchgeführt werden. Die Ähnlichkeitsberechnung aus 3.3 erfordert zum Beispiel, dass Trajektorien etwa die gleiche Länge haben, Start und Ziel identisch sind und keine Abweichungen durch zum Beispiel Schleifen vorhanden sind. Diese Voraussetzungen sind für eine Clusteranalyse nicht erforderlich. Hier können beliebige Trajektorien miteinander verglichen werden. Abbildung 3.34 zeigt den ersten Datensatz aus zehn Trajektorienensembles „Atmosphäre-2018-06-13-00\_Ensemble\_01\_00001\_CG\_FG“ aus FlyATM4E.

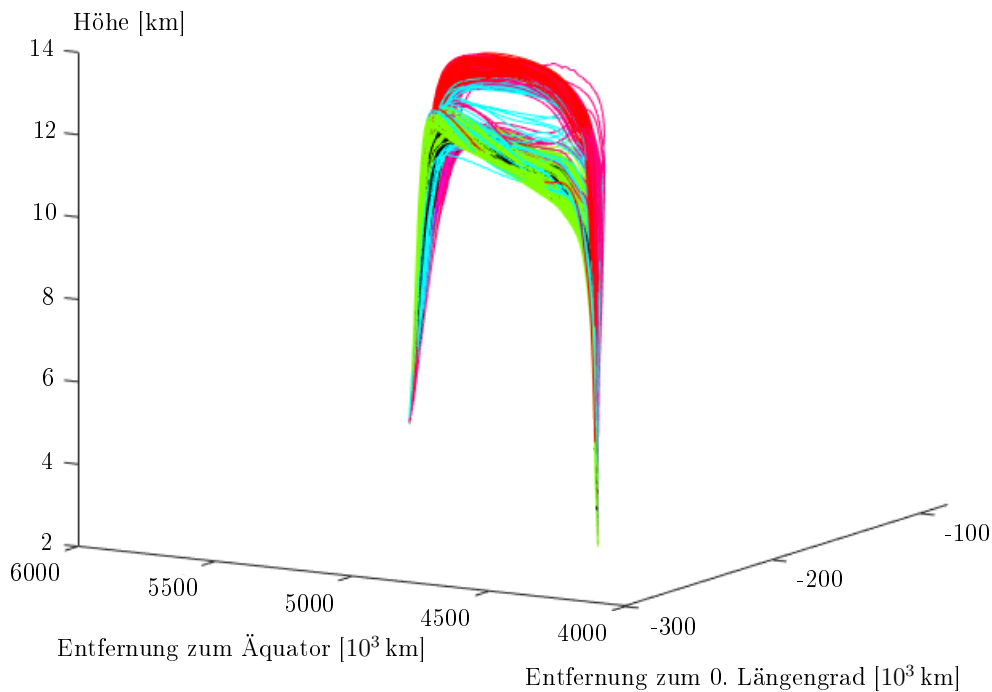


Abbildung 3.34: Gleichzeitige Clusteranalyse von 10 Datensätzen aus FlyATM4E.

Zwar sind Start und Ziel in diesem Datensatz identisch, aufgrund verschiedener Wetterlagen sind die Trajektorien aber sehr verschieden voneinander. Für die Clusteranalyse ist dies nicht relevant. Ähnliche Verläufe gehören zum selben Cluster, unabhängig davon, zu welchen Datensatz sie gehören. In Abbildung 3.34 sind die Trajektorien so in sieben Cluster aufgeteilt. Die in 3.9.3 erwähnte Option zum Löschen einzelner Trajektorien ist ebenfalls angewandt.

Einerseits ist die Vergleichbarkeit beliebiger Trajektorien ein Vorteil, denn die Daten müssen nicht genau spezifiziert werden oder bestimmte Voraussetzungen erfüllen. Andererseits kann von Nachteil sein, dass in jeder Gruppe von Daten eine relative Einteilung durchgeführt wird. So verändert sich das Ergebnis aller Daten, wenn weitere Trajektorien hinzugefügt oder Trajektorien gelöscht werden. Auf dieses Problem wird in 3.10.2 genauer eingegangen.

### 3.10.2 Relative Distanzen

Das Ergebnis der Clusteranalyse ist immer ein relatives. Mit dem Entfernen oder Hinzufügen von Daten verändert sich das Ergebnis. Abbildung 3.35 zeigt die separate Analyse von zwei Datensätzen. In beiden Fällen bilden die vier Datenpunkte ein Cluster aus drei Punkten und ein Cluster mit einem einzelnen Punkt. Zu beachten ist hierbei der Maßstab, der an einer Kästchenlänge gemessen wird. Das Ergebnis ist in beiden Fällen, dass drei Punkte robust sind, da sie alle etwa dieselbe Distanz zueinander aufweisen und ein Punkt weit entfernt vom Rest liegt. Im Anschluss sollen Datensatz 1 und Datensatz 2 miteinander verglichen werden. Die Daten könnten so wie im unteren Bild von Abbildung 3.35 liegen. Das Ergebnis ist nun ein völlig anderes. Die Punkte, die in der Einzelanalyse von Datensatz 1 noch auf zwei Cluster

aufgeteilt wurden, befinden sich jetzt in einem. Dafür wurde das Cluster in Datensatz 2 durch das Hinzufügen aufgespalten. Andere Punkte haben nun viele weitere Punkte in ihrer Nähe und werden als robust klassifiziert.

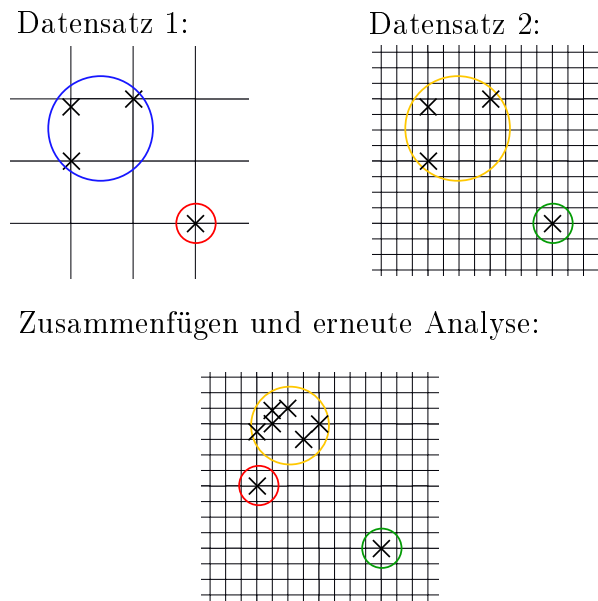


Abbildung 3.35: Beispiel für Einzelvergleich und anschließendes Zusammenfügen.

Das sich relativ verändernde Ergebnis ist Vor- und Nachteil zugleich. Beliebige Daten können analysiert werden. Vergleichsweise benötigt die Ähnlichkeitsberechnung Daten, die viel mehr Voraussetzungen erfüllen müssen. Andererseits ist es nicht möglich, gegebene Daten als nicht robust zu klassifizieren. Da die Distanzbestimmung relativ und nicht absolut erfolgt, werden wie in Abbildung 3.35 gegebenenfalls völlig verschiedene Größenordnungen für den Vergleich benutzt. Die verschiedenen Größen sind im Ergebnis nicht unbedingt nachvollziehbar.

### 3.10.3 Geordnete Daten

In sehr wenigen Fällen kann die Clusteranalyse scheitern. Dies geschieht vorwiegend, wenn die gegebenen Daten bereits eine Struktur aufweisen beziehungsweise gleich weit voneinander entfernte Punkte aufweisen. Dabei ist entscheidend, dass Entfernungen zwischen Punkten nicht nur wenige Male mehrfach auftreten, sondern derselbe Abstand oft auftritt. Abbildung 3.36 zeigt neun Datenpunkte, die per Clusteranalyse beurteilt werden sollen. Die Berechnungen starten mit fünf Clustern, da dies der in 3.11 benutzte Startwert ist.

Es folgt eine Einteilung in vier Cluster mit je zwei Punkten und ein Cluster mit nur einem Punkt. Die Anordnung der Cluster ist dabei willkürlich. Da in jedem Fall ein Einzelpunkt gefunden wird, wird die nächste Einteilung mit vier Clustern vorgenommen. Die gefundene Lösung in Abbildung 3.36 ist bezüglich ihrer Anordnung der Cluster erneut nicht eindeutig. Diesmal können sogar die Clustergrößen variieren. Das MATLAB-Skript soll die Cluster immer so klein wie möglich gestalten. Damit wird der linke Fall, ggf. in einer anderen Anordnung der Cluster, gefunden.

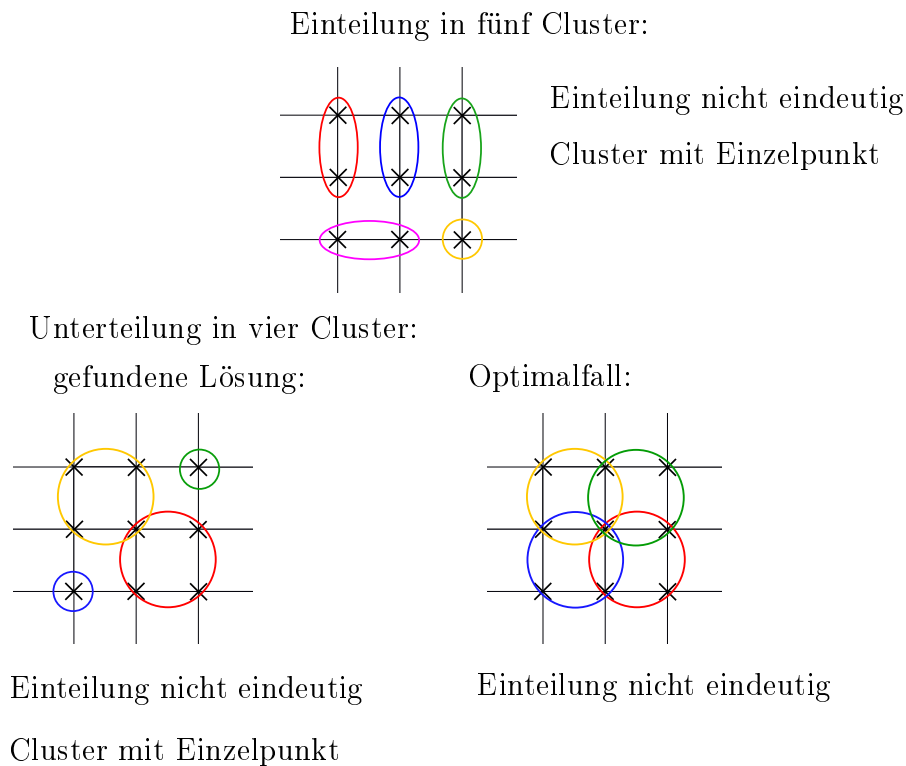


Abbildung 3.36: Clusteranalyse für Daten mit gleichen Abständen.

Es folgt die Einteilung in ein Cluster weniger. Bei diesem Vorgang gehen jedoch wichtige Informationen verloren. Optimalerweise würde die rechte Unterteilung bei vier Clustern gefunden werden. Damit ist kein Einzelpunkt vorhanden und da der mittlere Punkt in allen vier Clustern vorhanden ist, ist er am robustesten. Diese Information geht im linken Fall verloren und kann auch nicht bei der Aufteilung in eine andere Anzahl zurückgeholt werden.

Im Allgemeinen kann aber davon ausgegangen werden, dass die Datenpunkte der Projekte ATM4E und FlyATM4E keine Ordnung oder gleiche Distanzen aufweisen, sondern die Trajektorienpunkte verschiedene Abstände zueinander besitzen. In sehr wenigen Fällen, wenn überhaupt, sollte ein Informationsverlust aufgrund von Ordnungsstrukturen auftreten.

### 3.11 Umsetzung der Clusteranalyse in MATLAB

Der Ausgangspunkt der Clusteranalyse ist das Einteilen der Daten in fünf Cluster. Diese Anzahl wurde gewählt, da sich bei den gegebenen Daten aus ATM4E und FlyATM4E der Start bei fünf Clustern oft als sinnvoll erwiesen hat. Abbildung 3.37 zeigt das Ablaufdiagramm des MATLAB-Programms.

Nachdem eine Clusteranalyse durchgeführt wurde, werden bestimmte Eigenschaften der Cluster untersucht. Zu den Kriterien gehören Cluster mit Einzelpunkten, Überlappung von Clustern oder im Vergleich mit zuvor durchgeführten Analysen die Aufspaltung beziehungsweise Zusammenführung von Clustern und allgemeine Veränderungen der Cluster. Nach Auswertung dieser Kriterien wird eine neue Anzahl von Clustern bestimmt, in die die Daten neu aufzuteilen sind. Wurde die so be-

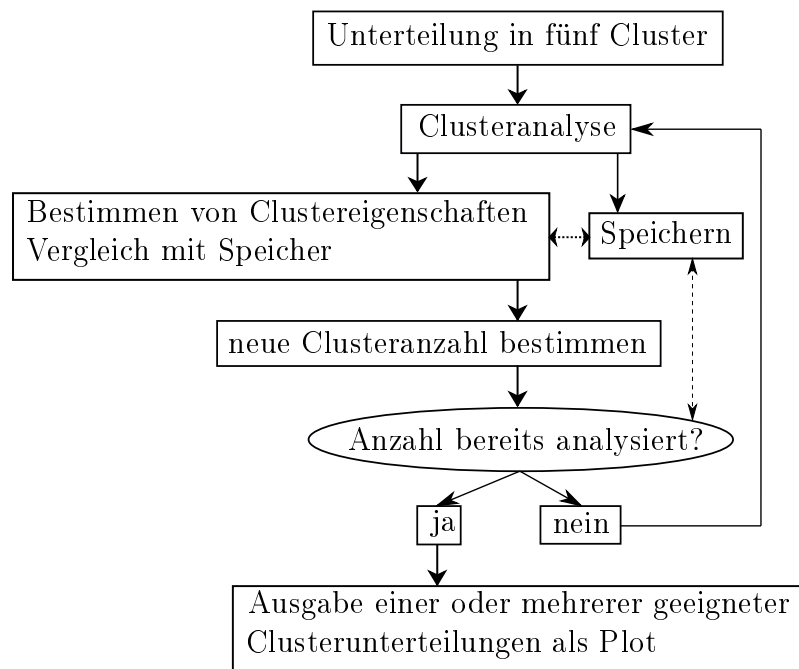


Abbildung 3.37: Ablaufdiagramm der Clusteranalyse in MATLAB.

stimmte Anzahl noch nicht analysiert, wird eine neue Clusteranalyse durchgeführt und der Prozess beginnt erneut. Sollte die neue Anzahl bereits analysiert worden sein, folgt die Zusammenfassung aller durchgeführten Analysen. Neben einem oder mehreren Plots mit einer aufgrund der Kriterien als sinnvoll erachteten Clusteranzahl werden Eigenschaften wie Überlappung von Clustern ausgegeben oder das Löschen von Daten veranlasst. Wird die Löschoption gewählt, folgt meist eine erneute Clusteranalyse mit einem Ausgangspunkt von fünf Clustern.

MATLAB-Funktionen zur Clusteranalyse werden nicht verwendet. Trotz der Dokumentation der Clusterfunktionen [60] wird nicht die vollständige Funktionsweise klar. So gibt MATLAB eine bestmögliche Anzahl für Cluster aus, der Weg dorthin bleibt jedoch unbekannt. Wie in 3.9.2 beschrieben wird mit den selbst programmierten Funktionen eine einzige Zahl oder ein Bereich für die Anzahl der Cluster festgelegt. Die verschiedenen Kriterien, die diesen Bereich festlegen, sind klar zu ermitteln. Ähnlich verhält es sich bei der in MATLAB implementierten Löschoption. Trajektorien können gelöscht werden, jedoch ist unklar, nach welchen Kriterien dies erfolgt. Daher wird auch hier eine eigene Funktion verwendet. Andere in MATLAB zur Verfügung stehende Optionen wie Filter sind für die Anwendung nicht relevant und wurden daher auch nicht verwendet.





## 4. Ergebnisse

Insgesamt werden in dieser Arbeit Trajektoriendaten aus zwei verschiedenen Projekten, ATM4E und FlyATM4E analysiert. Bei ATM4E sind Flüge von verschiedenen Start- und Zielorten gegeben, das Wetter bleibt innerhalb eines Trajektoriensatzes jeweils konstant. Die Ergebnisse zu diesem Projekt sind in Abschnitt 4.1 aufgeführt. Bei dem Projekt FlyATM4E blieben innerhalb eines Datensatzes Start und Ziel unverändert, jedoch werden verschiedene Wetterlagen betrachtet. Die Ergebnisse dieser Analyse folgen in Abschnitt 4.2. In beiden Fällen wird dieselbe Methode zur Analyse der Trajektoriendaten verwendet, jedoch unterscheiden sich die Ergebnisse der beiden Projekte. Aussagen über Robustheit variieren je nach Projekt und sind in den folgenden Abschnitten zusammengefasst.

### 4.1 Ergebnisse mit ATM4E

Zur Analyse der Daten stehen verschiedene Plots und die zugehörigen Daten zur Verfügung. Zunächst werden in 4.1.1 die Gewichtungen analysiert, die einen Satz von Trajektorien erzeugen. Entsprechend dieser Plots können die Trajektorien in drei Gruppen aufgeteilt werden. Jede Gruppe hat bezüglich der groben Einteilung von Clustern andere Charakteristiken, weshalb in 4.1.2 auf die Clusteranalyse der jeweiligen Gruppen eingegangen wird. In 4.1.3 folgt diese Betrachtung bezüglich der berechneten Ähnlichkeit. Alle Ergebnisse für die Trajektorien aus ATM4E werden anschließend in 4.1.4 zusammengefasst.

#### 4.1.1 Plots der Pareto-Fronten

Zu jedem Trajektoriensatz gehört ein Plot zur Gewichtung zwischen Kosten- und Umweltfaktoren. Die Kosten werden in den Diagrammen als  $\Delta$  Fuel gemessen, also dem Unterschied des Kraftstoffverbrauches zu einer bestimmten Referenztrajektorie. Die Umweltauswirkungen werden in  $\Delta$  ATR, der Average Temperature Response angegeben, auch hier als Differenz zwischen einer ausgewählten Referenztrajektorie. Jede Trajektorie ist für sich genommen optimal. Es entstehen aber verschiedene Trajektorienverläufe, weil die beiden Faktoren  $\Delta$  Fuel und  $\Delta$  ATR verschieden stark berücksichtigt werden. Da die gegebenen Daten bereits mit einem Pareto-Filter bereinigt wurden, wurde der Filter wieder entfernt. Mit dem Filter weisen alle Trajektoriensätze eine nahezu perfekte Form auf, wenn auch in einigen Fällen nur sehr wenige Punkte nach dem Anwenden eines Filters verbleiben. Ohne Filter sind die Daten gut in drei verschiedene Gestalten der Pareto-Fronten zu unterscheiden. Die

drei verschiedenen Typen werden als Gruppe 1, Gruppe 2 und Gruppe 3 bezeichnet. Je nach Gruppe ergeben sich verschiedene Eigenschaften für die Robustheit des Trajektoriensatzes. Daher wird hier in drei Gruppen unterschieden, die bei weiteren Betrachtungen mit Hilfe der Clusteranalyse und der Ähnlichkeitsberechnung erneut aufgegriffen werden.

**Gruppe 1:** Idealerweise zeigt die Pareto-Front einen hyperbelartigen Verlauf. Dieses Verhalten weist zum Beispiel der Datensatz „00001\_UBBB\_ELLX“ für den Flug von Baku nach Luxemburg auf. Auch der Pareto-Filter in MATLAB versucht, die Daten einer solchen Form anzugleichen. Ob ein solcher Filter angewandt wird, ist für Daten dieser Gruppe nicht relevant, da keine bis sehr wenige Punkte herausgefiltert werden. Abbildung 4.1 zeigt das entsprechende Plot für den Flug von Baku nach Luxemburg. In dieser Gruppe dürfen die Dichte der Kreuze sowie die Skalen variieren und die Kurve unterschiedlich scharf gestaltet sein. Jedoch muss sich ein im Plot weiter rechts befindlicher Punkt immer unterhalb aller Datenpunkte links von ihm befinden. Ein solcher Verlauf lässt auf eine kontinuierliche Pareto-Front schließen. Zu dieser Gruppe gehört die Mehrheit der Trajektorien. Von den 116 gegebenen Trajektorien zählen 52 zu dieser Gruppe.

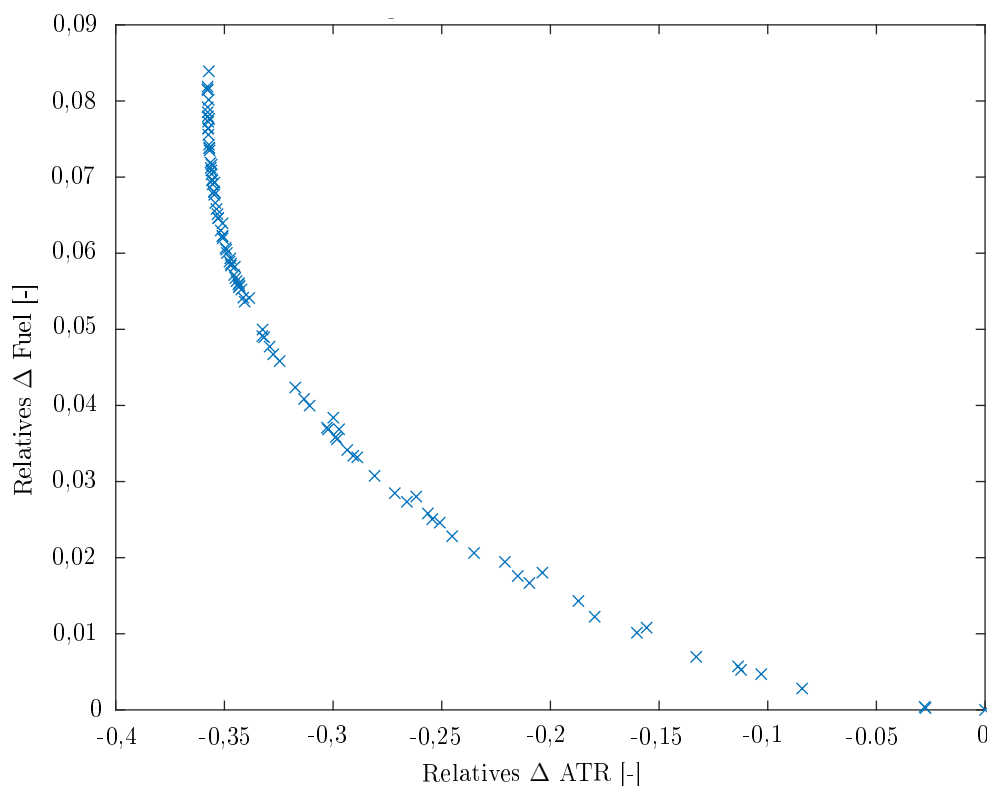


Abbildung 4.1: Pareto-Front des Trajektoriensatzes „00001\_UBBB\_ELLX“ aus ATM4E.

**Gruppe 2:** Trajektorien, die zu dieser Gruppe gehören, besitzen Sprünge innerhalb der Pareto-Front. Abbildung 4.2 zeigt ein solches Beispiel durch Trajektoriensatz „00025\_ESSA\_GCRR“. Das relative  $\Delta$  Fuel verändert sich, während das relative  $\Delta$  ATR nahezu konstant bleibt. Im Plot entsteht eine Gruppierung von senkrecht untereinander liegenden Kreuzen. Das relative  $\Delta$  ATR verändert sich nach dieser Anordnung sprunghaft. Es entsteht eine neue Gruppierung ähnlicher Form, jedoch

mit deutlich verändertem  $\Delta$  ATR. Von diesen Sprüngen treten mehrere auf, jedoch ist die genaue Anzahl nicht festgelegt. Oft sind diese Plots von einigen Ausreißern begleitet. Neben den beschriebenen Gruppierungen treten weitere einzelne Datenpunkte auf. Es bleibt zunächst unklar, ob diese Punkte durch ein sehr sprunghaftes Verhalten der Pareto-Front in ein oder beiden Dimensionen verursacht wird oder ob es sich bei diesen Punkten um Fehler handelt. Wird der Pareto-Filter in MATLAB angewandt, werden hier einige Punkte gelöscht. Jedoch verbleiben die meisten Punkte auch mit einem Filter im Plot. Von den untersuchten 116 Datensätzen gehören 26 zu Gruppe 2.

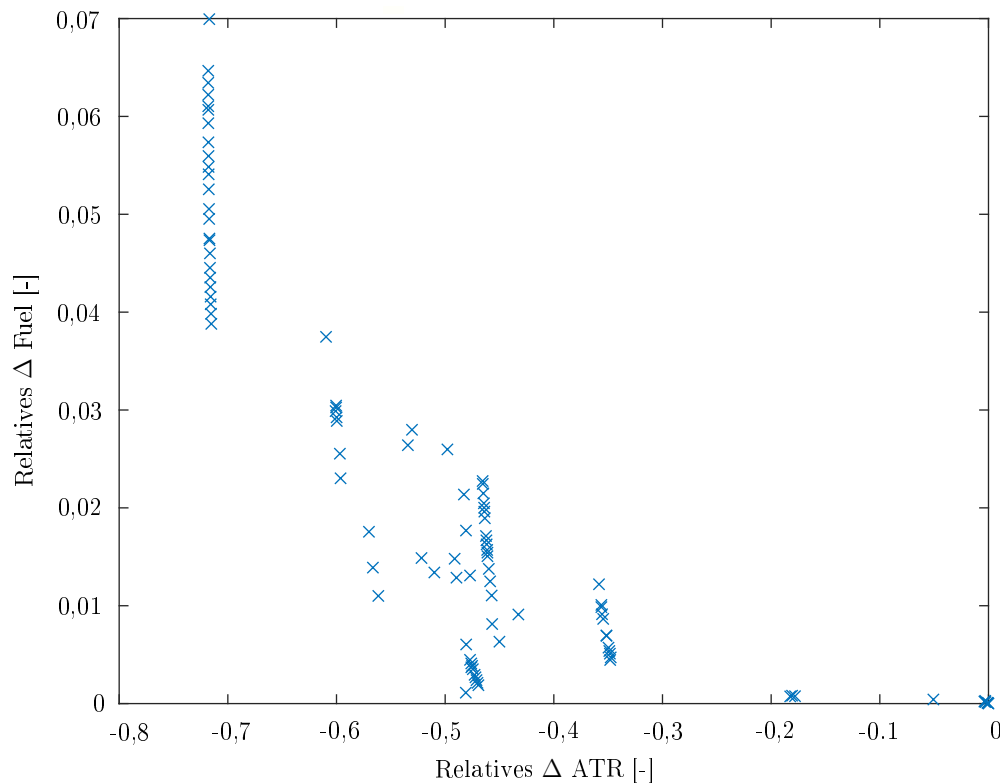


Abbildung 4.2: Pareto-Front des Trajektoriensatzes „00025\_ESSA\_GCRR“ aus ATM4E.

**Gruppe 3:** In dieser Gruppe folgen die Datenpunkte ohne Pareto-Filter keinem offensichtlichen Kurvenverlauf. Viel mehr handelt es sich um eine Punktwolke. In Abbildung 4.3 wird die Gewichtung für Datensatz „00045\_GCLP\_ENBO“ dargestellt, der ein Beispiel für Gruppe 3 ist. Wird ein Pareto-Filter angewandt, wird ein Großteil der Punkte herausgefiltert. Auch hier ist wie in Gruppe 2 unklar, ob es sich um sehr fein gestreute Sprünge oder fehlerhafte Daten handelt. 38 der 116 untersuchten Trajektoriensätze gehören zu dieser Gruppe, was sie nach Gruppe 1 zur zweitgrößten macht.

Aussagen über Robustheit sind innerhalb einer Gruppe sehr ähnlich, zwischen den Gruppen variieren sie. In 4.1.2 und 4.1.3 findet die weitere Betrachtung daher nach den drei Gruppen getrennt statt.

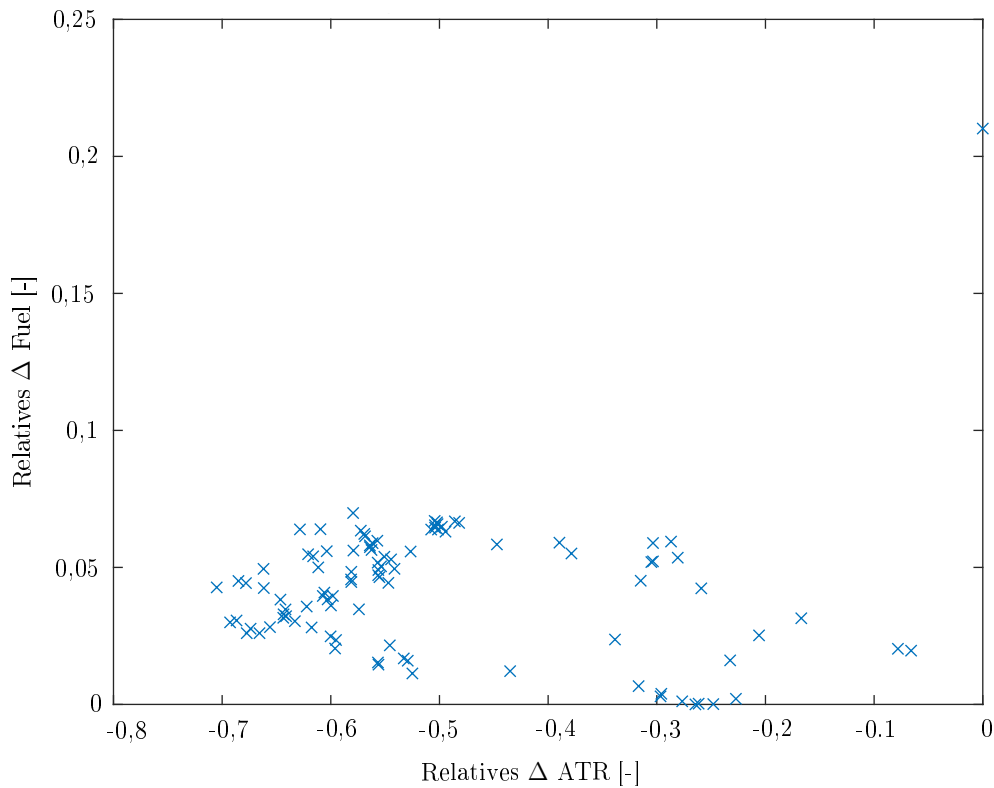


Abbildung 4.3: Pareto-Front des Trajektoriensatzes „00045\_GCLP\_ENBO“ aus ATM4E.

### 4.1.2 Ergebnisse der Clusteranalyse

Entsprechend der in 4.1.1 eingeteilten Gruppen werden auch hier die Ergebnisse für Gruppe 1 bis 3 separat diskutiert.

**Gruppe 1:** Trajektorien, die aus einer Pareto-Front ohne Sprünge resultieren, besitzen in der Regel keine großen Abweichungen zwischen den Trajektorien. Abbildung 4.4 zeigt die Clusteranalyse für den Datensatz „00001\_UBBB\_ELLX“ von Baku nach Luxemburg, der in 4.1.1 bereits als Beispiel für eine kontinuierliche Pareto-Front verwendet wurde. Die Trajektorien liegen alle innerhalb eines engen Korridors. Hauptsächlich unterscheiden sich die Trajektorien in der Höhe, die von gut 10 000 m bis knapp 12 000 m variiert. Die Trajektorien wurden in insgesamt drei Cluster, rot, gelb und türkis, aufgeteilt. Dabei wurde eine der größten Lücken in Abbildung 4.4 markiert. Die schwarze Linie innerhalb der Markierung entspricht einer Entfernung von 400 m. Bezogen auf die Dimensionen eines Fluges ist eine Strecke von 400 m sehr gering. Insgesamt liegen die Trajektorien sehr nah beieinander. Da das MATLAB-Skript aber zu einer Unterscheidung in mehr als ein Cluster gezwungen wird, sind hier drei Cluster entstanden, obwohl die Abstände sehr gering sind. Da die Berechnung zudem relativ ist, wird ein türkises Cluster mit teilweise Abständen von lediglich 10 m gebildet. Zudem wird in Abbildung 4.5 ersichtlich, dass für die türkisen Trajektorien eine Abweichung zur Seite existiert und daher eine Trennung der Cluster vorgenommen wurde. Das gelbe Cluster liegt im Übergang zwischen den beiden Clustern.

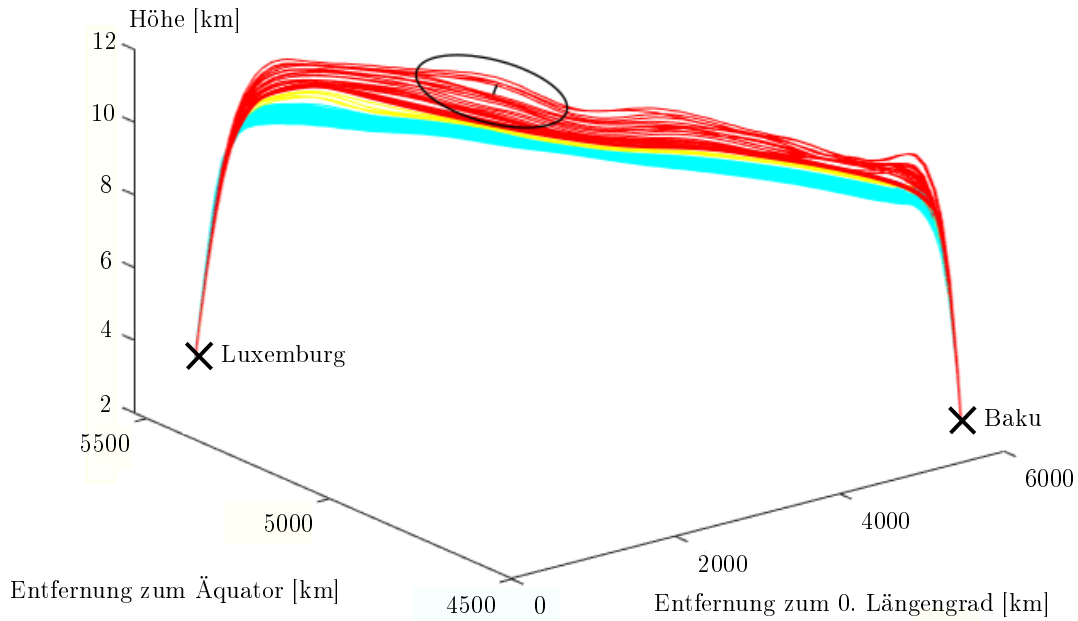


Abbildung 4.4: Endgültige Clustereinteilung für Datensatz „00001\_UBBB\_ELLX“.

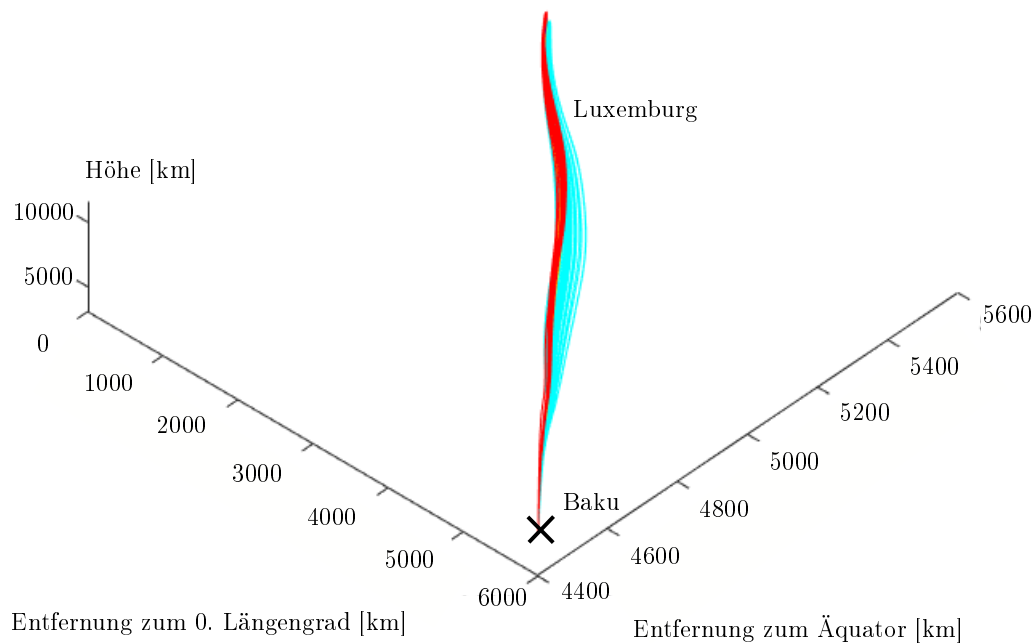


Abbildung 4.5: Endgültige Clustereinteilung für Datensatz „00001\_UBBB\_ELLX“ gedreht.

Ein weiteres Phänomen, das verstärkt bei Gruppe 1 auftritt, ist die Häufung von Trajektorien in geringer Höhe. Der für das Beispiel verwendete Trajektoriensatz „00001\_UBBB\_ELLX“ weist diese Charakteristik auf, jedoch ist in anderen Datensätzen die Häufung noch stärker ausgeprägt. Bereits in Abbildung 4.4 ist zu sehen, dass die türkisen Trajektorien im Plot nicht mehr auseinander zu halten sind, während zwischen den roten Trajektorien noch Lücken erkennbar sind. Bei Daten-

satz „00016\_EDDH\_GCLP“ von Hamburg nach Gran Canaria ist diese Eigenschaft besonders ausgeprägt, wie in Abbildung 4.6 zu sehen. Hier wurde in vier Cluster unterschieden, wobei das türkise Cluster besonders groß ist.

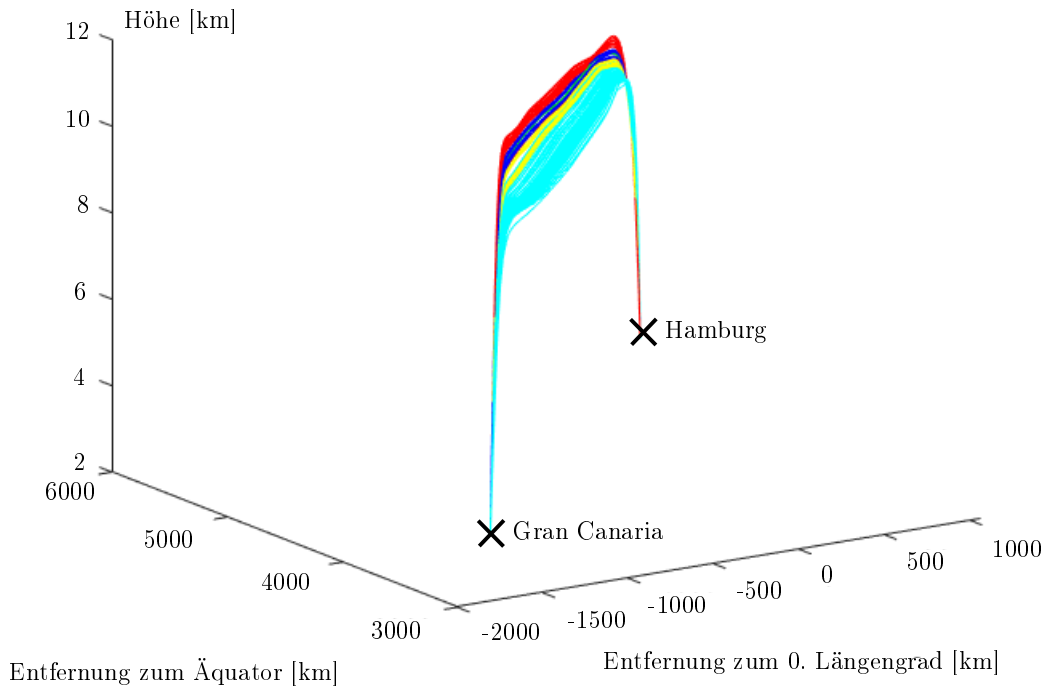


Abbildung 4.6: Endgültige Clustereinteilung für Datensatz „00016\_EDDH\_GCLP“.

Die Erklärung dafür liegt in der Pareto-Front begründet. Abbildung 4.7 zeigt die zugehörige Pareto-Front für Datensatz „00016\_EDDH\_GCLP“. Die Dichte der Kreuze nimmt zur linken Seite des Plots deutlich zu. Hier befinden sich die Trajektorien, die geringst möglichen Umweltauswirkungen aufweisen, jedoch die Kosten des Fluges erhöhen. Die Abstufung der Umweltoptimalen Trajektorien ist damit viel feiner als die der Kostenoptimalen, was in Abbildung 4.6 zu sehen ist. Die umweltoptimalen Trajektorien befinden sich in geringerer Höhe. Für die Trajektorienoptimierung wurde eine Mindesthöhe von 10 500 m festgelegt, die außer beim Start und bei der Landung nicht unterschritten werden darf. Aufgrund der verschiedenen Luftschichten und den entsprechenden Reaktionen mit Flugzeugabgasen, ergibt eine geringere Höhe in der Regel weniger negative Umwelteinflüsse. Die festgelegte Mindesthöhe und die feine Abstufung der Trajektorien ergibt so sehr nah beieinander liegende Trajektorien, die zu einem Cluster zusammengefasst werden.

Zu erwarten wäre eine Häufung bei den kostenoptimalen Trajektorien. Während die Umwelteinflüsse schwierig zu berechnen sind, orientieren sich die Kosten eines Fluges hauptsächlich an den Treibstoffkosten. Mit größerer Höhe nimmt die Luftreibung ab und es können Treibstoff und somit Kosten eingespart werden. Da die Abstufung der Gewichtung des Treibstoffverbrauchs aber gröber ist als die Abstufung der umweltoptimalen Trajektorien, ist nur selten eine Häufung der Trajektorien zu sehen. Dies bedeutet nicht, dass die Trajektorien weniger robust sind als die umweltoptimalen. Da sich ihre Gewichtung stärker voneinander entscheidet, sind die kostenoptimalen Trajektorien auch geometrisch weiter voneinander entfernt.

Wird anhand der Pareto-Front festgestellt, dass ein Trajektoriensatz zur Gruppe 1 gehören, bedarf es kaum einer weiteren Untersuchung bezüglich Robustheit. Alle

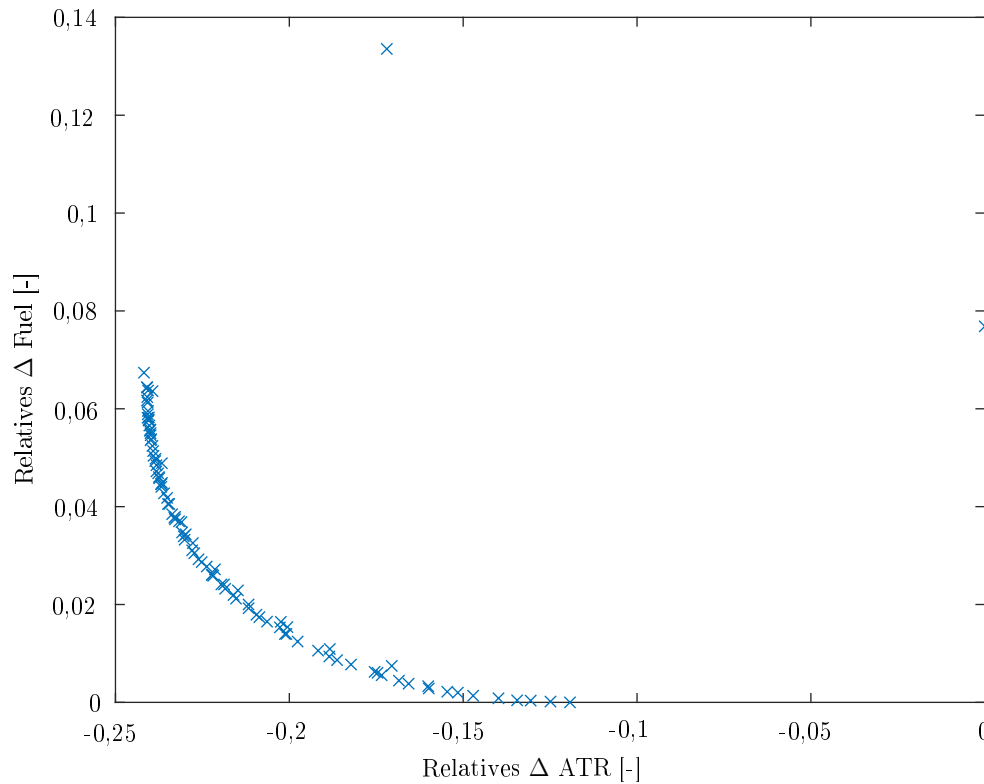


Abbildung 4.7: Pareto-Front des Trajektoriensatzes „00016\_EDDH\_GCLP“ aus ATM4E.

Trajektorien liegen innerhalb eines engen Korridors und besitzen viele Nachbartrajektorien mit kaum abweichendem Verlauf. Abhängig davon, ob Kosten oder Umweltauswirkungen eine höhere Priorität eingeräumt werden soll, unterscheiden sich die zu wählenden Trajektorien hauptsächlich in der Höhe. In einigen Fällen gibt es Abweichungen zur Seite, die dann aber auch für viele Trajektorien auftreten und stetig zunehmen. Es treten keine plötzlichen Sprünge in den Trajektorien auf. Alle Trajektorien können als robust eingestuft werden, da es viele Partner in geometrischer Nähe und mit ähnlicher Gestalt gibt. Einzelne Trajektorien für Gruppe 1 als robuster als andere einzustufen, ist kaum möglich. Alle Trajektorien besitzen gleichermaßen weitere, in ihrer Gestalt gut passende Trajektorien in ihrer direkten Umgebung. Der gesamte Trajektoriensatz besteht aus ausschließlich und gleichermaßen robusten Trajektorien.

**Gruppe 2:** Als Beispiel für Gruppe 2 wurde Datensatz „00025\_ESSA\_GCRR“ von Stockholm-Arlanda nach Lanzarote verwendet. Die zugehörigen Trajektorien und ihre Clustereinteilung sind in Abbildung 4.8 zu sehen. Obwohl die Pareto-Front nicht kontinuierlich ist (siehe Abbildung 4.2), hat dies keine Auswirkungen auf die Trajektorien. Sie sehen wie ein Beispiel für Gruppe 1 aus. Innerhalb eines engen Korridors sind alle Trajektorien zu finden. Es gibt wenn überhaupt nur kleine Lücken zwischen den Trajektorien in einer Größenordnung, die bezüglich der realen Flugplanung als vernachlässigbar klein einzustufen ist.

Es gibt aber auch andere Beispiele innerhalb dieser Gruppe. Dazu gehört der Datensatz „00080\_GCRR\_EKCH“ für einen Flug von Lanzarote nach Kopenhagen-Kastrup. In der ungefilterten Pareto-Front, die in Abbildung 4.9 zu sehen ist, treten

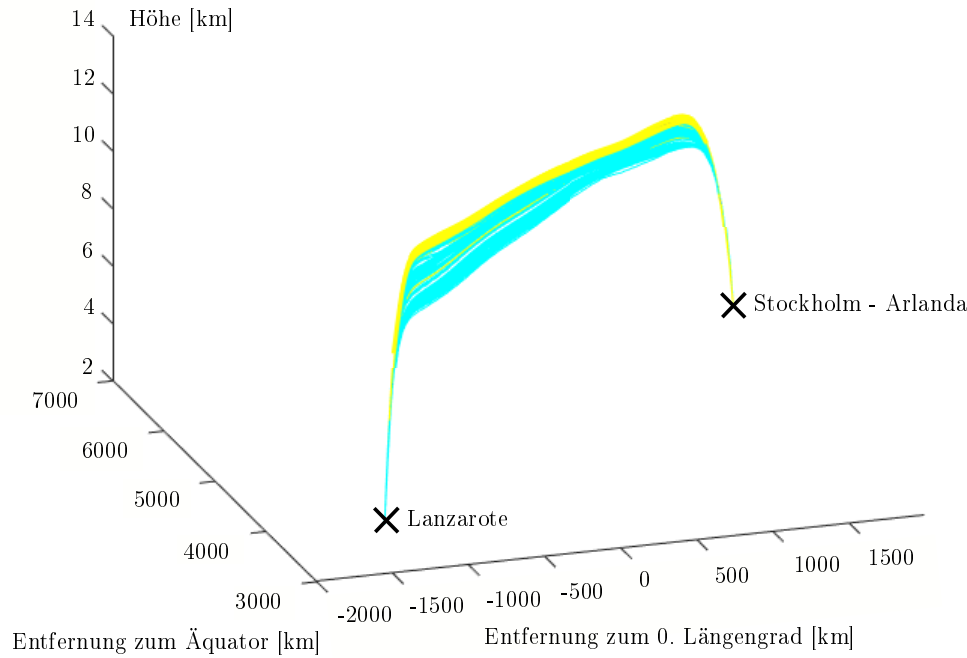


Abbildung 4.8: Endgültige Clustereinteilung für Datensatz „00025\_ESSA\_GCRR“.

wie für Gruppe 2 charakteristisch Sprünge zwischen senkrecht verteilten Häufungen auf. Dabei ist insbesondere bei diesen Daten fraglich, ob eine Reduzierung des  $\Delta$  ATR um knapp 90 % realistisch ist oder ob es sich hierbei um fehlerhafte Daten handelt.

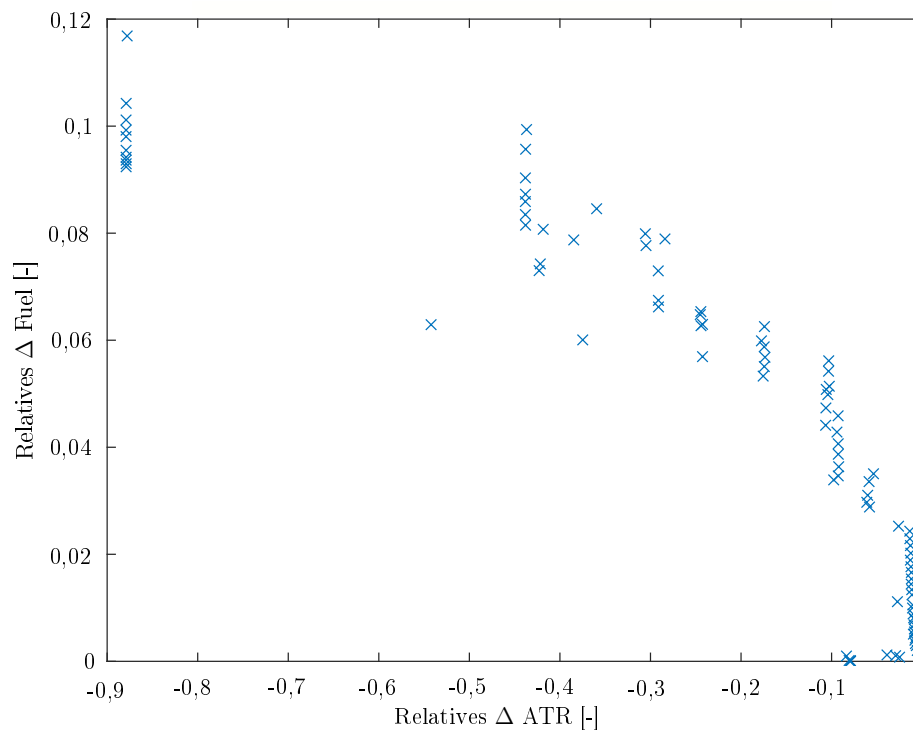


Abbildung 4.9: Pareto-Front des Trajektoriensatzes „00080\_GCRR\_EKCH“.



In den Trajektorien zu diesem Datensatz lassen sich zunächst keine Auffälligkeiten feststellen, die mögliche Fehler bei der Berechnung bestätigen würden oder bestätigen, dass die Reduzierung des  $\Delta$  ATR um 90 % nicht realistisch ist. Die zugehörigen Trajektorien zu „00080\_GCRR\_EKCH“ sind in Abbildung 4.10 geplottet. Es wurde noch keine Einteilung in Cluster vorgenommen. Es handelt sich bei der Abbildung um die aus den gegebenen Datenpunkten interpolierten Verläufe. Ein Farbverlauf zwischen schwarz und rot stellt die Gewichtung der Kosten- und Umweltfaktoren dar. Eine schwarze Trajektorie stellt das Kostenoptimal dar, eine rote eine umwelt-optimale Trajektorie. Farbliche Abstufungen ins Dunkelrote sind die entsprechenden Kompromisse. Zu sehen ist außerdem, dass die Einzeltrajektorie in geringster Höhe durch die Interpolation einige Ungenauigkeiten aufweist. Die in Abbildung 4.10 ab-

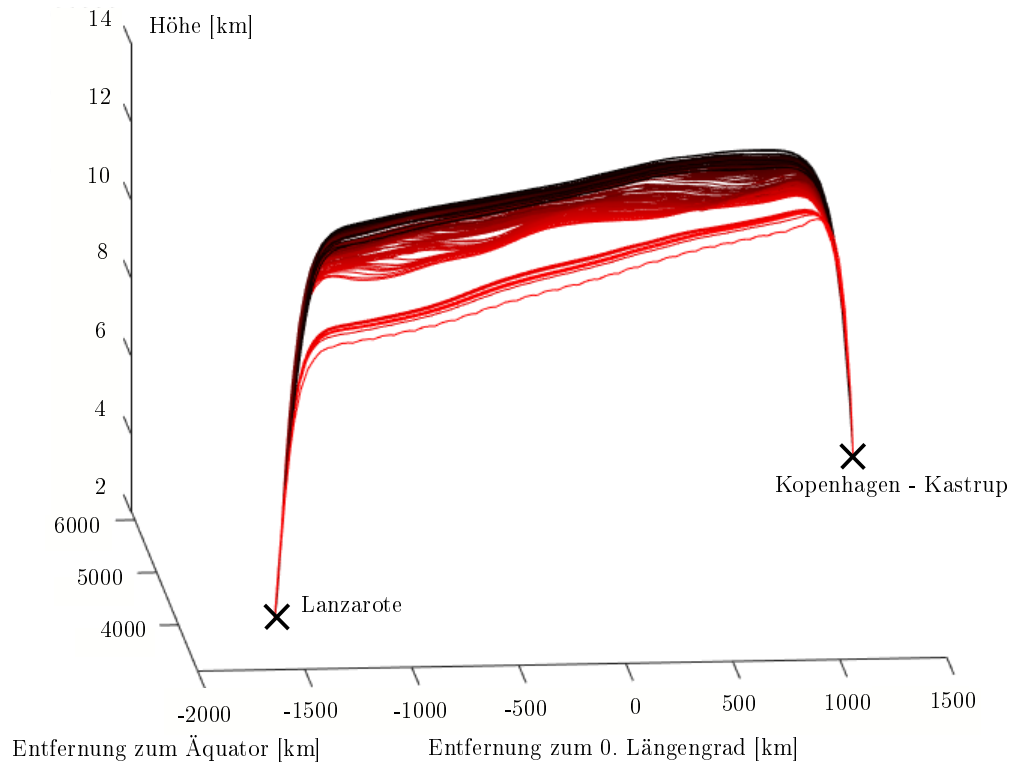


Abbildung 4.10: Interpolierte Trajektorien zu Datensatz „00080\_GCRR\_EKCH“.

gebildeten Trajektorien wurde in Abbildung 4.11 in Cluster eingeteilt. Dabei sind zwei Cluster, türkis und gelb entstanden. Fast alle Trajektorien gehören zum türkisen Cluster, während das gelbe Cluster nur einen sehr kleinen Anteil ausmacht. Vergleich man zudem Abbildung 4.10 und Abbildung 4.11, wurden eine Trajektorie in geringer Höhe gelöscht. Insgesamt funktioniert die Clusteranalyse in diesem Beispiel eher schlecht. Es stellt sich zum Beispiel die Frage, warum für die Trajektorien in geringer Höhe kein weiteres Cluster gebildet wurde, obwohl diese eine große räumliche Lücke zu den anderen Trajektorien aufweisen. Gleichzeitig wurde ein Cluster gebildet, das aus sehr wenigen Trajektorien besteht beziehungsweise von den türkisen Trajektorien verdeckt wird und sich somit in unmittelbarer Nähe befindet. Der Abstand und die daraus resultierende Clusterbildung ist relativ. In Grenzfällen können durch die in 3.11 erklärte Vorgehensweise gegebene Daten anders interpretiert werden, als die menschliche Intuition es tun würde. Im gezeigten Fall wurden die Abstände innerhalb des türkisen Clusters als sehr gering eingestuft und nur ein

Cluster gebildet. Hingegen sind die Trajektorien in geringer Höhe vergleichsweise weit entfernt voneinander, sodass sie gelöscht wurden.

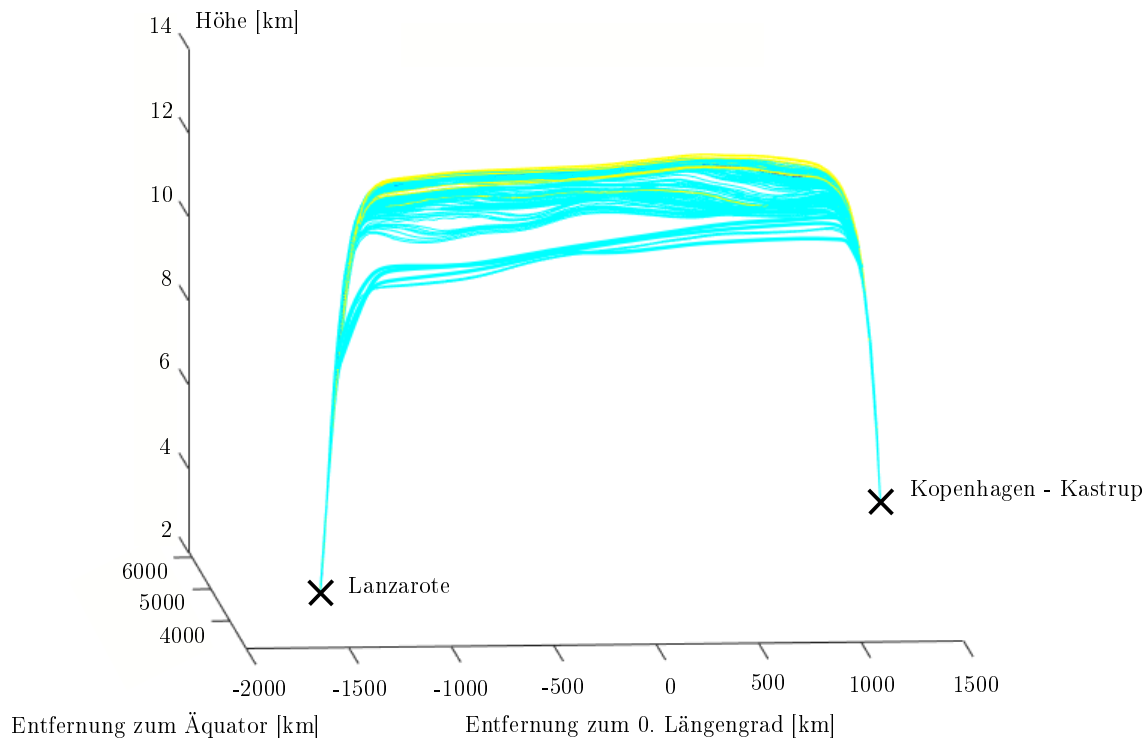


Abbildung 4.11: Endgültige Clustereinteilung für Datensatz „00080\_GCRR\_-EKCH“.

Die Clusteranalyse für Gruppe 2 erkennt fast nur robuste Trajektorien. Trotz der Sprünge innerhalb der Pareto-Front verlaufen die Trajektorien oft wie in Gruppe 1 in einem engen, nahezu kontinuierlich gefüllten Korridor. Alle Trajektorien sind in einem solchen Fall robust. Insgesamt gibt es kaum Lücken von einer relevanten Größe zwischen den Trajektorien. Es gibt aber auch Fälle, in denen Lücken zwischen den Trajektorien auftreten oder auch einzelne Trajektorien gelöscht werden. Das Löschen von Trajektorien ist der Aussage „gar nicht robust“ gleichzusetzen. Gelöschte Trajektorien besitzen keine Nachbarn in ihrer Nähe. Sobald Trajektorienparameter leicht abweichen oder anders gewichtet werden, entsteht ein völlig anderer Trajektorienverlauf. Lücken zwischen den Trajektorien werden eher schlecht erkannt. Um eine sinnvolle Aussage zur Robustheit treffen zu können, muss die Ähnlichkeitsberechnung hinzugezogen werden.

**Gruppe 3:** Erneut wird das Beispiel aus Abschnitt 4.1.1 aufgegriffen. Dafür sind in Abbildung 4.12 die zugehörigen Trajektorien zur Pareto-Front oder auch Punktwolke aus Abbildung 4.3 zunächst ohne Cluster abgebildet. Wie in allen anderen Plots entspricht eine schwarze Trajektorie einer kostenoptimalen, eine rote Trajektorie ist umweltoptimal und Mischfarben aus schwarz und rot sind entsprechend Kompromisse zwischen den beiden Optimalen. Eine Trajektorie mit völlig anderem Verlauf ist in Abbildung 4.12 zu sehen sowie einige größere Lücken zwischen den Trajektorien. In etwa lässt sich ein Korridor wie bei den anderen beiden Gruppen feststellen, jedoch ist dieser deutlich weiter gestreut und besitzt zwischen den Trajektorien größere Lücken.

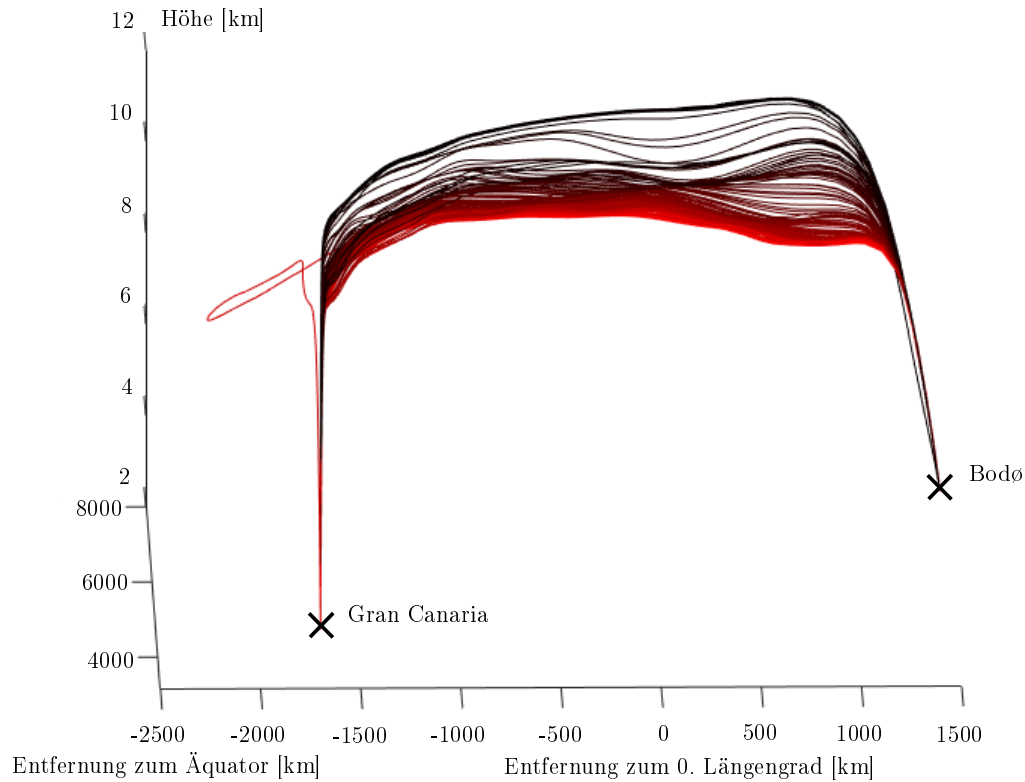


Abbildung 4.12: Interpolierte Trajektorien zu Datensatz „00045\_GCLP\_ENBO“.

Bei der folgenden Clusteranalyse, in Abbildung 4.13 zu sehen, ist für Datensatz „00045\_GCLP\_ENBO“ festzustellen, dass die Einzeltrajektorie sowie einige kostenoptimale Routen in großer Höhe gelöscht wurden. Es wurde in drei Cluster geteilt, wobei das türkise wieder einmal das größte Cluster ist und alle Trajektorien in geringer Höhe und damit die eher umweloptymalen Trajektorien beinhaltet. Wie in Gruppe 1 ist der Grund dafür, dass die umweloptymalen Trajektorien in feinerer Abstufung generiert wurden und dadurch dichter beieinander sind.

Nach dem Löschen von Trajektorien beträgt der Abstand der größten Lücke in Abbildung 4.13 etwa 500 m und ist damit als sehr klein anzusehen. Für dieses Beispiel sind die verbleibenden Trajektorien nach dem Löschen als robust anzusehen, die gelöschten Trajektorien sind nicht robust.

In Gruppe 3 treten aber auch andere Fälle auf. Die ungefilterte Pareto-Front in Abbildung 4.14 für den Datensatz „00058\_GCLP\_EKCH“ von Gran Canaria nach Kopenhagen-Kastrup lässt noch nicht auf den Trajektorienverlauf schließen. Wie die anderen Plots in dieser Gruppe sind die Gewichtungspunkte ohne Filter wie in einer Wolke verteilt. Ein klarer Verlauf der Pareto-Front ist nicht erkennbar.

Die zugehörigen Trajektorien sind in Abbildung 4.15 dargestellt. Auch hier wurde der übliche Farbverlauf verwendet. Anders als bei allen bisherigen Plots sind die Trajektorien nun nicht mehr grob nach ihrer Farbe getrennt, sondern verlaufen wild durcheinander. Obwohl eigentlich eine Grenze für die Mindestflughöhe von 10 500 m existiert, wird diese unterschritten. Es gibt keinen klaren Korridor mehr. Die Trajektorien verlaufen innerhalb eines sehr breit gefächerten Gebietes. Sowohl Höhe, Breiten- als auch Längengrad unterscheiden sich stark voneinander und sind nicht

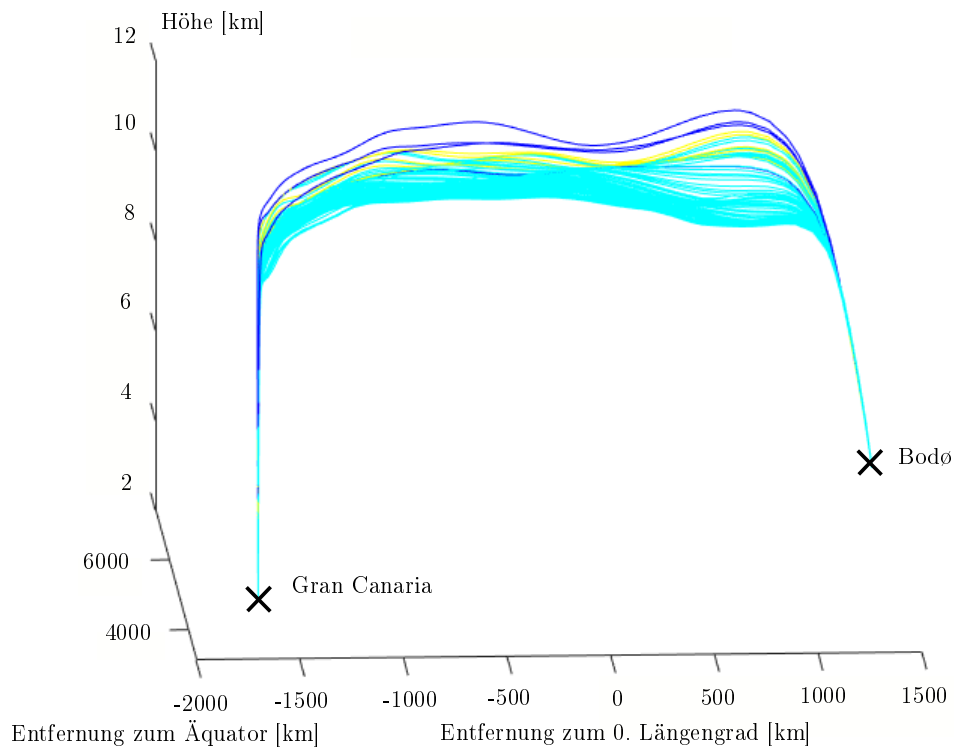


Abbildung 4.13: Endgültige Clustereinteilung für Datensatz „00045\_GCLP\_EN-BO“.

wie bei den vorherigen Beispielen in den Unterschieden so gering, dass sie als vernachlässigbar eingestuft werden können.

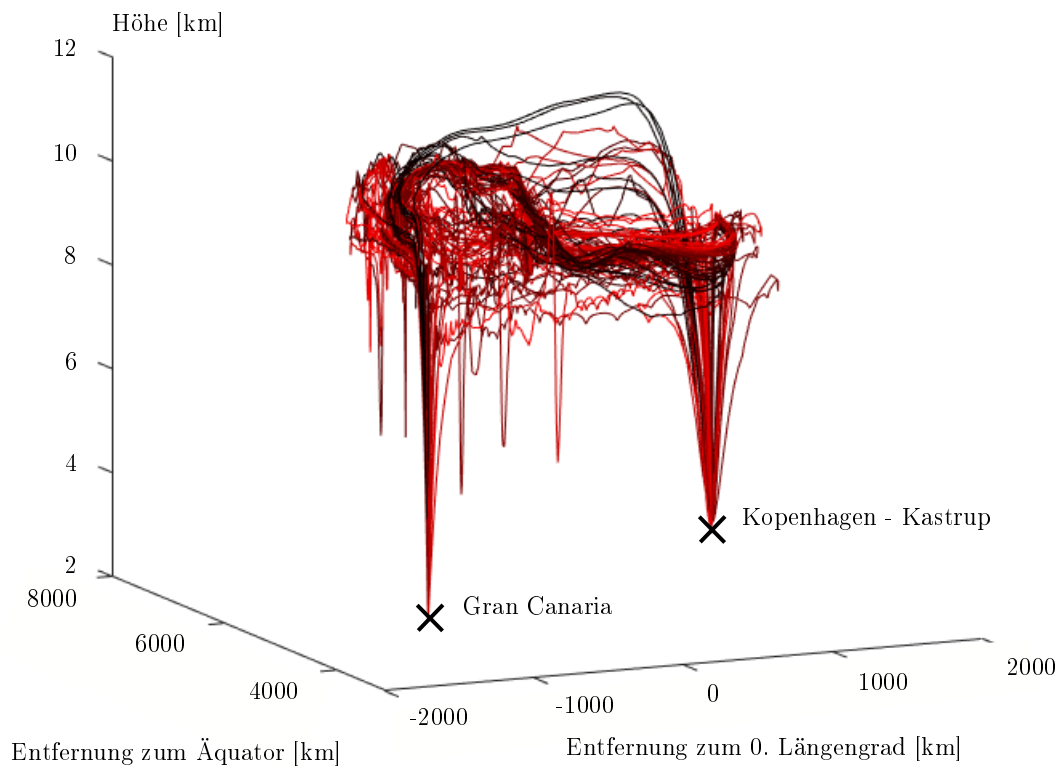


Abbildung 4.15: Interpolierte Trajektorien zu Datensatz „00058\_GCLP\_EKCH“.

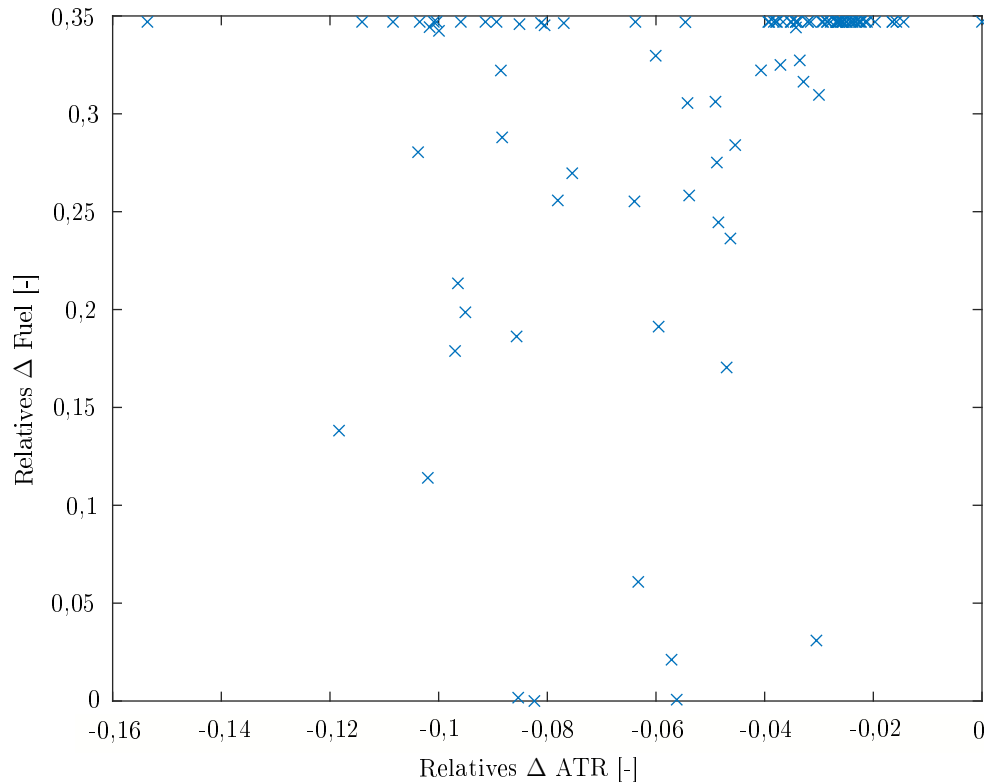


Abbildung 4.14: Pareto-Front des Trajektoriensatzes „00058\_GCLP\_EKCH“.

Die Clusteranalyse gestaltet sich entsprechend der verstreuten Trajektorienverläufe schwierig. Viele Trajektorien wurden gelöscht, der Rest wurde in sechs Cluster eingeteilt wie in Abbildung 4.16 zu sehen. Wieder ist das türkise Cluster am größten. Die Einteilung des türkisen Clusters wirkt zunächst willkürlich. Dieser Eindruck entsteht, da am Start (Gran Canaria) stark variierende Trajektorien in dasselbe Cluster eingeteilt wurden. Diese Trajektorien sind jedoch im weiteren Verlauf sehr eng beieinander, weshalb sie in das gleiche Cluster eingeteilt wurden. Ähnlich verhält es sich mit den nach oben ausreißenden Trajektorien des türkisen Clusters. Auch sie sind in anderen Teilen sehr nah bei weiteren Trajektorien, sodass sie trotz der großen Entfernung in einigen Abschnitten in das gleiche Cluster eingestuft wurden.

Bei der Clusteranalyse ist kein Vergleich zwischen verschiedenen Routen möglich. Nur innerhalb eines Trajektoriensatzes mit demselben Start und Ziel kann verglichen werden, welche Trajektorien robust, nicht robust oder robuster als andere Trajektorien sind. Der Grund dafür liegt in der relativen Messung und Einteilung der Cluster. Es wird keine Distanz vorgegeben, die die Grenze zwischen nah beieinander und weit entfernt beziffert. Stattdessen wird aufgrund verschiedener Kriterien eine Clusteranzahl festgelegt, die eine möglichst zielführende Einteilung der Trajektorien liefert. In Abhängigkeit der Daten wird die Grenze der Cluster dynamisch angepasst. So können Trajektorien innerhalb eines Clusters nur 10 m entfernt sein und werden darüber als weit voneinander entfernt eingestuft. In anderen Datensätzen werden Distanzen von 1 km als nah eingestuft, weil die übrigen Trajektorien vergleichsweise weit entfernt liegen.

In dieser Art der relativen Messung liegt auch der Grund, warum die Clusteranaly-

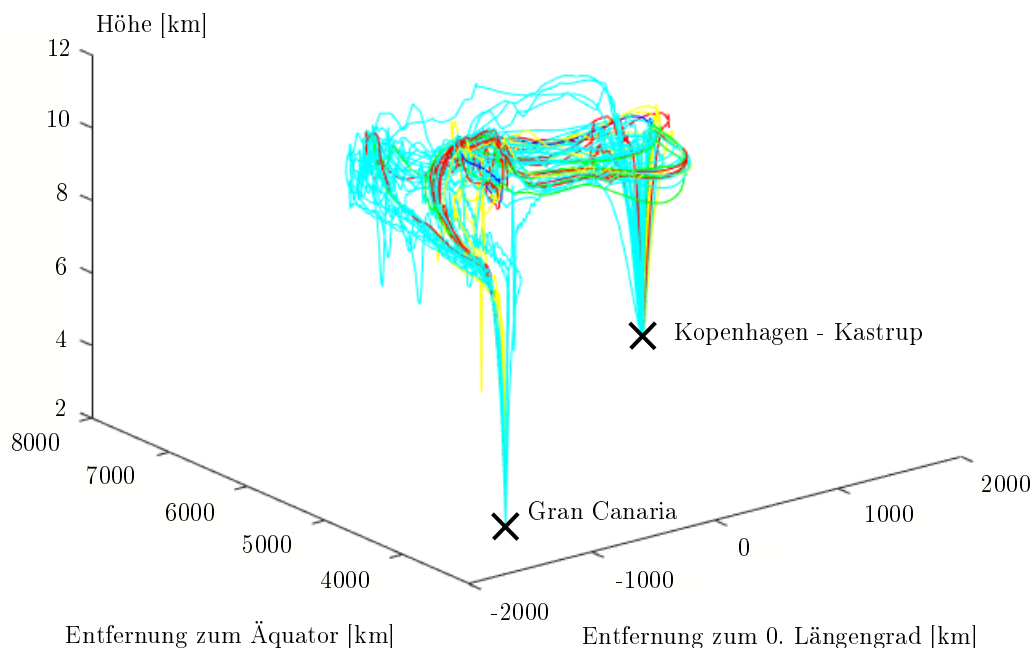


Abbildung 4.16: Endgültige Clustereinteilung für Datensatz „00058\_GCLP\_EKCH“.

se, so wie sie hier implementiert wurde, für Gruppe 2 oft Einteilungen entgegen der menschlichen Intuition vornimmt. Einerseits soll das Verfahren für eng beieinander liegende, nahezu parallel verlaufende Trajektorien gelten und diese voneinander unterscheiden. Andererseits gibt es weit gestreute Trajektorien, die sich bis zu 2000 km voneinander abweichen. Auch hier soll die Clusteranalyse gültig sein. Für die beiden Randfälle von perfekten, also nahe beieinander liegenden parallelen Trajektorien und für ein sehr streuendes Feld wird diese Aufgabe gut erfüllt. Dazwischen befinden sich Fälle, bei denen die Trajektorien einige, nicht besonders große Lücken oder abweichende Verläufe aufweisen. Genau diese Fälle treten verstärkt bei Gruppe 2 auf. Das Verfahren liefert hier eher schlechte Ergebnisse, da es keinen sinnvollen Kompromiss bezüglich der Einteilung findet. Einerseits müssen die Cluster gewisse Abweichungen hinnehmen, sodass überhaupt Gruppierungen entstehen können und nicht jede Trajektorie ihr eigenes Cluster ist. Andererseits muss die Abstufung fein genug gewählt werden und in genügend Cluster unterschieden werden. Genau dieses Finden von Feinheit in den Clustern ist für Gruppe 2 leider oft nicht gegeben. Für die anderen beiden Gruppen liefert die Clusteranalyse jedoch sinnvolle Ergebnisse, wenn sie auch manchmal der menschlichen Intuition widersprechen. Für den Übergang sind weitere Kriterien zu finden, anhand derer die Clustereinteilung vorgenommen wird, sowie Grenzwerte festzulegen, bei denen Daten gelöscht werden oder eine feinere Unterteilung vorgenommen wird.

### 4.1.3 Ergebnisse der Ähnlichkeitsberechnung

Bei der Ähnlichkeitsberechnung wird jede Trajektorie innerhalb eines Datensatzes mit allen anderen verglichen und eine Zahl bestimmt, die die Ähnlichkeit eines Paares beschreibt. In den meisten Fällen sind 101 Trajektorien innerhalb eines Datensatzes von ATM4E gegeben. Wir jede Trajektorie mit jeder anderen verglichen,

werden etwa 11 000 Vergleiche durchgeführt. Trajektorien, die mit sich selbst verglichen werden, weisen das Ergebnis „0“, was gleichbedeutend mit identisch ist, auf und sind daher nicht relevant. Außerdem ist der Vergleich symmetrisch, weshalb die Vergleichsdurchführungen auf 5 500 Vergleiche reduziert werden können. Jedes Ergebnis wird in ein Datenarray gespeichert. Dies ist eine große Datenmenge, die ein Mensch nicht spontan erfassen kann. Die vollständigen Daten sind in MATLAB in den Arrays **Ähnlichkeit...** zu finden. Für die einfache Erfassung der Daten wurden Plots der Trajektorien erstellt, auf die im folgenden eingegangen wird. Für die Plots wurde eine Mischform aus Ähnlichkeitsberechnung und Clusteranalyse verwendet. Trajektorien, die verglichen mit dem gesamten Datensatz eine hohe Ähnlichkeit zueinander aufweisen, werden in ein Cluster gruppiert. Die Anzahl der Cluster ist dynamisch und richtet sich nach dem Datensatz. Sie kann zwischen fünf und zehn Clustern betragen. Die Ergebnisse sind auch hier entsprechend der obigen Einteilung (4.1.1) in Gruppe 1, Gruppe 2 und Gruppe 3 aufgeteilt worden.

**Gruppe 1:** Datensatz „00016\_EDDH\_GCLP“ ist ein nahezu perfektes Beispiel für geradlinige, parallel zueinander verlaufende Trajektorien. Während die Clusteranalyse trotzdem eine recht grobe Einteilung der Trajektorien in vier Cluster vornimmt, teilt die Ähnlichkeitsberechnung die Trajektorien in acht verschiedene Gruppen ein. In Abbildung 4.17 ist diese Einteilung zu sehen.

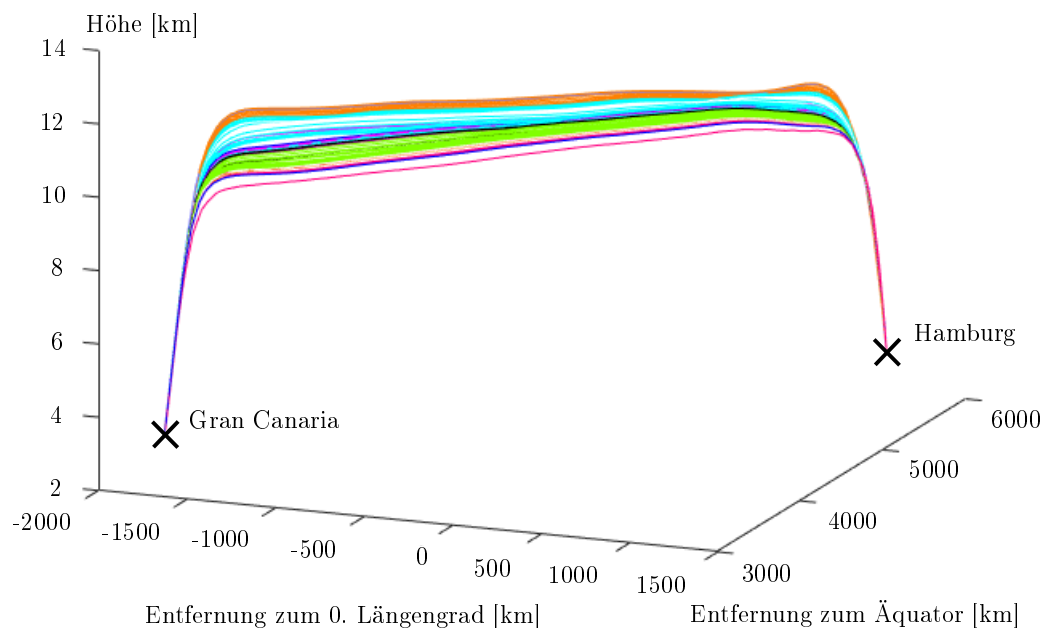


Abbildung 4.17: Berechnete Ähnlichkeit für Datensatz „00016\_EDDH\_GCLP“.

Verschiedene Farben stehen für die Zugehörigkeit zu verschiedenen Gruppen. Bei dem betrachteten Datensatz entstehen durch die Einteilung fast kontinuierliche farblich zueinander passende Streifen ohne Unterbrechung (wobei in Abbildung 4.17 eine blau eingefärbte Trajektorie eine Ausnahme bildet). Wie bei der Clusteranalyse bestätigt die Berechnung der Ähnlichkeit, dass in einem solchen Datensatz alle Trajektorien robust sind und es zwischen ihnen nur geringe Abstufungen bezüglich einer Abstufung der Robustheit gibt, da auch verschieden eingefärbte Gruppen nur gerin-

ge Lücken zwischen sich aufweisen. Datensatz „00016\_EDDH\_GCLP“ ist bezüglich der Trajektorien ein besonders geordneter Datensatz.

Typischerweise sehen die Trajektorien eher wie in Abbildung 4.18 aus. Dort ist die Ähnlichkeitsberechnung für Datensatz „00001\_UBBB\_ELLX“ zu sehen.

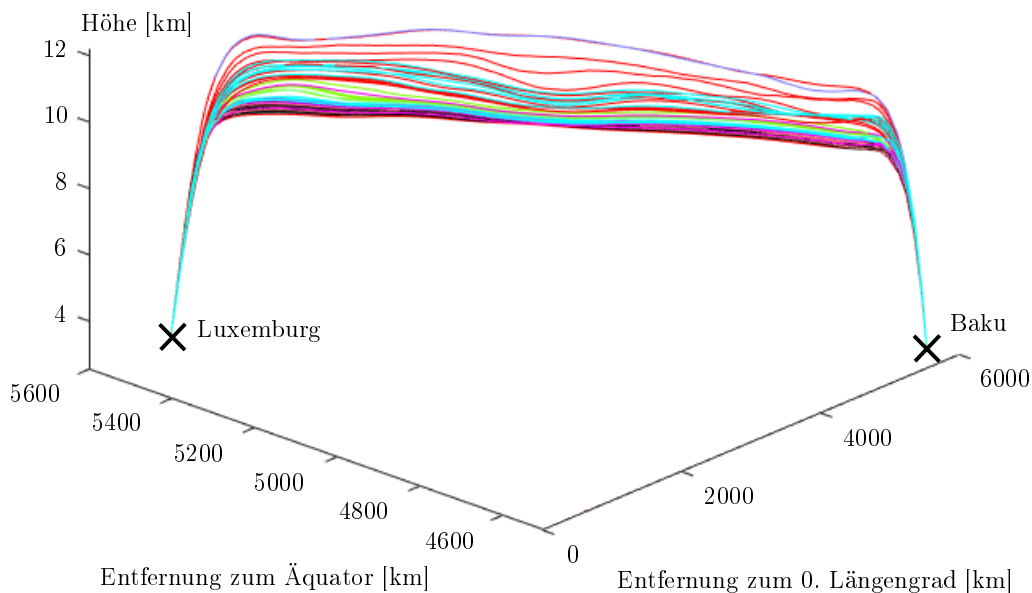


Abbildung 4.18: Berechnete Ähnlichkeit für Datensatz „00001\_UBBB\_ELLX“.

Anders als bei der Clusteranalyse sind Trajektorien mit derselben Gruppenzugehörigkeit nicht unbedingt direkt benachbart, wie es besonders an den rot eingefärbten Trajektorien deutlich wird. Insgesamt sind bei diesem Datensatz aber auch alle Gruppen der Ähnlichkeitsberechnung eng beisammen. Damit unterstützt die Ähnlichkeitsberechnung die oben getroffene Aussage, dass alle Trajektorien robust sind, sofern durch die Pareto-Kurve die Zugehörigkeit zu Gruppe 1 bestimmt werden kann.

**Gruppe 2:** Wie bei der Clusteranalyse bestimmt auch die Ähnlichkeitsberechnung, dass sich Gruppe 1 und Gruppe 2 zu großen Teilen überschneiden. Datensatz „00025\_ESSA\_GCRR“, der Sprünge in der Pareto-Front aufweist und damit zu Gruppe 2 gehört, ist wie in Abbildung 4.19 zu sehen nicht von einem Datensatz in Gruppe 1 zu unterscheiden. Die Trajektorien verlaufen parallel zueinander und die Gruppen sind durch die Ähnlichkeitsberechnung so eingeteilt, dass farblich einheitliche Bänder aus Trajektorien entstehen. Trotzdem besitzen die unterschiedlichen Gruppen keine großen Lücken zueinander, der gesamte Datensatz ist als robust anzusehen.

Es gibt aber auch Beispiele in Gruppe 2, die anders derer in Gruppe 1 sind. Datensatz „00080\_GCRR\_EKCH“, der in Abbildung 4.20 geplottet wurde, gehört dazu. Wird nur Abbildung 4.20 betrachtet, unterscheidet sich die Einteilung nicht von den bisherigen. Erst im Vergleich mit einem Plot aller Trajektorien, zum Beispiel Abbildung 4.10, ist zu erkennen, dass alle Trajektorien in geringer Höhe gelöscht wurden. Die Ähnlichkeitsberechnung ist in Fällen, bei denen mehrere Korridore mit einer deutlichen Lücke zu verarbeiten sind, recht streng darauf bedacht, den größeren von beiden Korridoren zu wählen und den kleineren Bereich zu ignorieren.



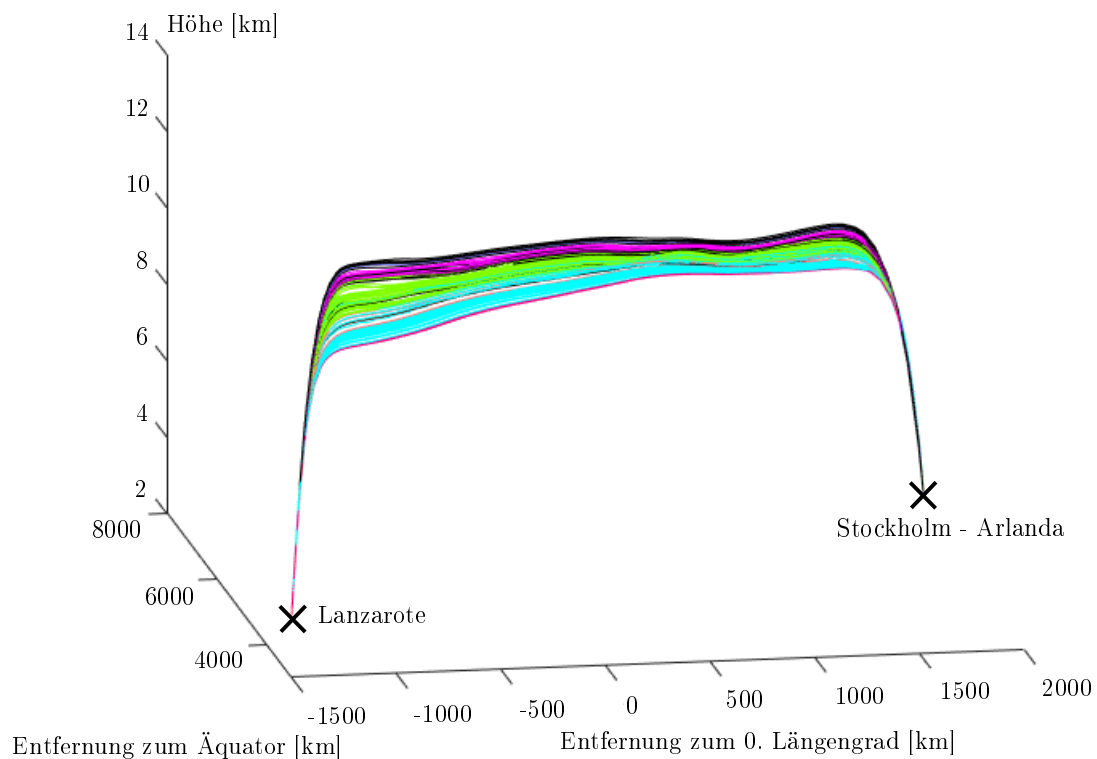


Abbildung 4.19: Berechnete Ähnlichkeit für Datensatz „00025\_ESSA\_GCRR“.

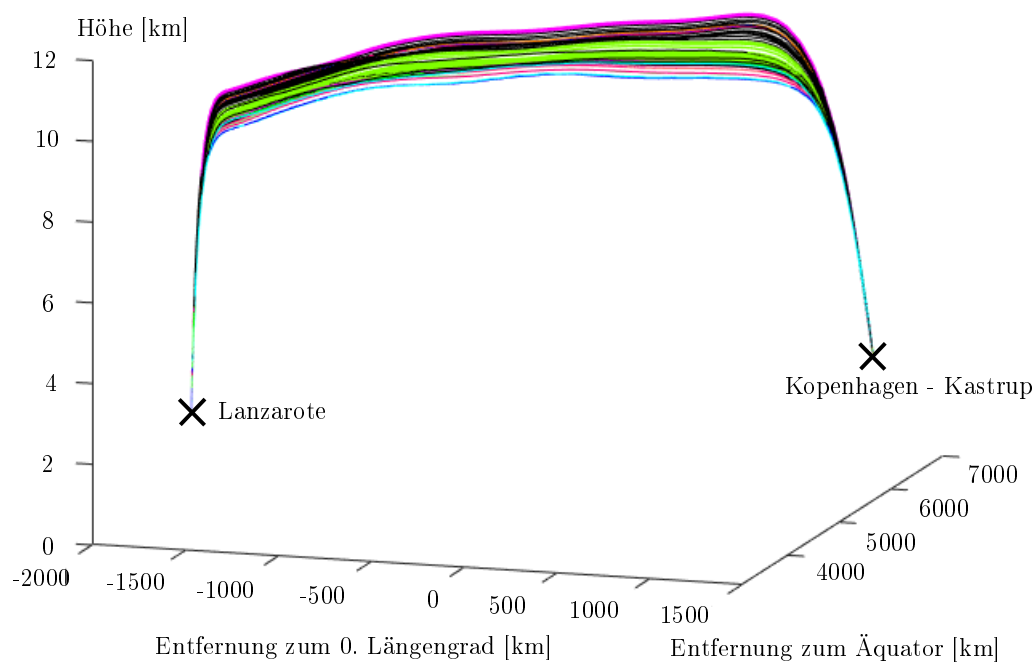


Abbildung 4.20: Berechnete Ähnlichkeit für Datensatz „00080\_GCRR\_EKCH“.

Die Ähnlichkeitsberechnung sortiert damit die Trajektorien deutlich genauer als die Clusteranalyse. Trajektorien, die bei der Clusteranalyse nicht gelöscht wurden, jedoch bei der Ähnlichkeitsberechnung aussortiert wurden, können somit als robust,

aber weniger robust als die verbleibenden Trajektorien der Ähnlichkeitsberechnung eingestuft werden.

Damit sind die Trajektorien in Gruppe 2 oft in ihrer Gesamtheit robust. In einigen Fällen ist durch die Kombination zwischen Clusteranalyse und Ähnlichkeitsberechnung nicht mehr der gesamte Datensatz als robust anzuerkennen, sondern Abstufungen zwischen einigen Trajektorien vorzunehmen.

**Gruppe 3:** Datensatz „00045\_GCLP\_ENBO“ ist ein gutes Beispiel, da bei der Ähnlichkeitsberechnung andere Trajektorien gelöscht werden als bei der Clusteranalyse. Vergleicht man Abbildung 4.13 und Abbildung 4.21, sind bei der Ähnlichkeitsberechnung die Trajektorien in großer Höhe verblieben, während sie bei der Clusteranalyse gelöscht wurden. Da Clusteranalyse und Ähnlichkeitsberechnung verschiedene Funktionsweisen besitzen, ist nicht generell zu sagen, welches Verfahren stärker zu gewichten ist. In jedem Fall ist aber klar, dass in beiden Verfahren verbleibende Trajektorien als uneingeschränkt robust gelten, einmal aussortierte Trajektorien weniger robust sind und bei beiden Verfahren gelöscht Trajektorien als nicht robust zu klassifizieren sind.

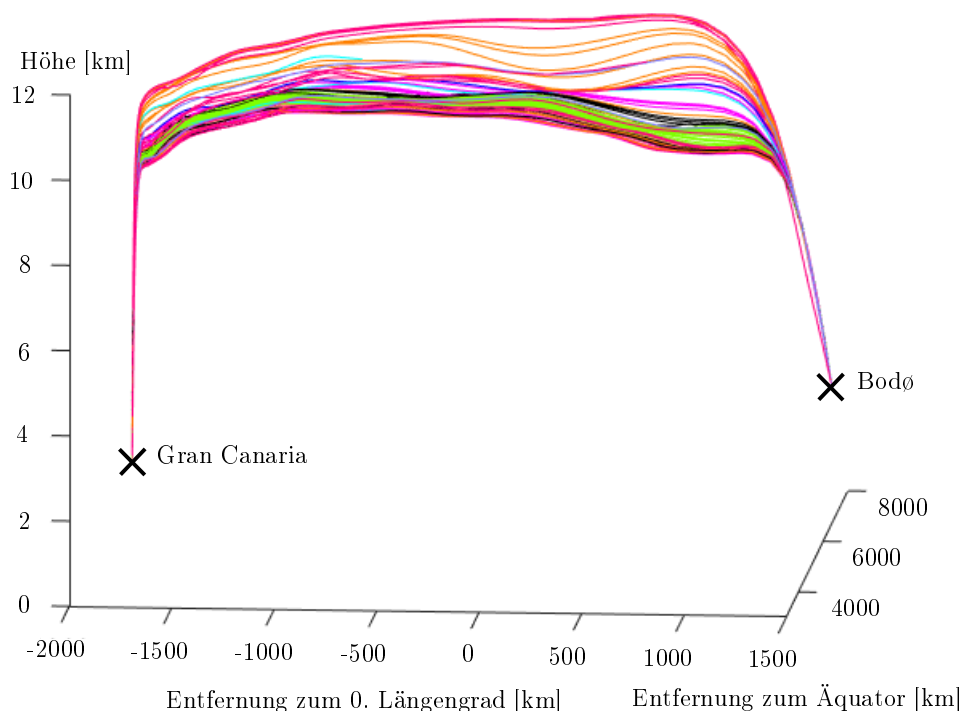


Abbildung 4.21: Berechnete Ähnlichkeit für Datensatz „00045\_GCLP\_ENBO“.

Bei Datensatz „00058\_GCLP\_EKCH“ stößt die Ähnlichkeitsberechnung an ihre Grenzen. Die Trajektorien sind zu verschieden, um durch dieses Verfahren klassifiziert werden zu können. Abbildung 4.22 zeigt das Ergebnis der Ähnlichkeitsberechnung. Nur wenige Trajektorien verbinden Start und Ziel, die meisten Trajektorien enden abrupt, obwohl der ursprüngliche Trajektoriensatz erst am Ziel endende Trajektorien besitzt. Der Grund dafür liegt im Verfahren der Ähnlichkeitsberechnung. Um für das Verfahren vergleichbar zu sein, müssen die Trajektorien ungefähr dieselbe Länge besitzen. Die Trajektorien in Abbildung 4.22 unterscheiden sich jedoch so stark voneinander, dass die Trajektorien beschnitten werden mussten, um überhaupt

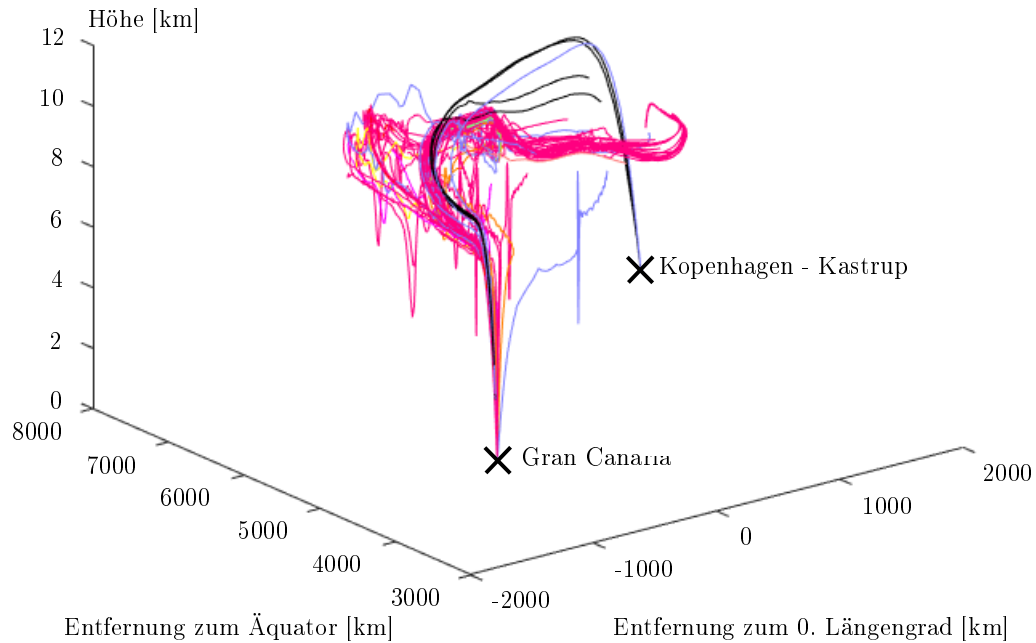


Abbildung 4.22: Berechnete Ähnlichkeit für Datensatz „00058\_GCLP\_EKCH“.

die Ähnlichkeitsberechnung anwenden zu können. Insgesamt ist die Ähnlichkeitsberechnung eher auf das Aufzeigen kleiner Unterschiede zwischen den Trajektorien ausgelegt, für grobe Unterscheidungen ist besser die Clusteranalyse einzusetzen. Da nur Teile der Trajektorien berücksichtigt wurden, ist das Ergebnis der Ähnlichkeitsberechnung sehr skeptisch zu sehen. Speziell bei diesem Datensatz könnte anhand des Plots die Aussage getroffen werden, dass die pink eingefärbten Trajektorien robuster als alle anderen sind, da sie die meisten Trajektorien beinhaltet und vergleichsweise die Trajektorien nah beieinander liegen. Diese Aussage ist aber nicht allgemein gültig zu treffen und mit Vorsicht zu betrachten, da die Gestalt auch aufgrund numerischer Fehler entstanden sein kann.

#### 4.1.4 Zusammenfassung für ATM4E

Die ungefilterten Pareto-Fronten liefern sehr gute Tendenzen bezüglich der Robustheit. Kontinuierliche Fronten sprechen für einen sehr robusten Trajektoriensatz. Mit zunehmenden Sprüngen und Ausreißern werden die Trajektorien weniger robust. Dabei können verschiedene Fälle eintreten und eine genaue Klassifikation kann erst mit der Betrachtung weiterer Daten vorgenommen werden.

Für Gruppe 1 ist die Aussage der Clusteranalyse und Ähnlichkeitsberechnung in den meisten Fällen deckungsgleich bis sehr wenig voneinander abweichend. Die Trajektorien liegen alle innerhalb eines engen Korridors, Ausreißer treten nicht auf. Quasi alle Trajektorien in einem solchen Datensatz sind auf gleiche Weise robust. Gibt es Sprünge in der ungefilterten Pareto-Front, gibt es in den Trajektorien oft Abstufungen in der Robustheit. Bereiche, die Clusteranalyse und Ähnlichkeitsberechnung beide beinhalten, sind als uneingeschränkt robust anzusehen. Neben diesen Trajektorien gibt es aber auch die, die bei einem der beiden Verfahren aussortiert wurden. Diese sind weniger robust, können aber im Gegensatz zu den durch beide Verfahren aussortierten Trajektorien noch als robust klassifiziert werden. Gibt es keine klare

Pareto-Front mehr, sind die Trajektorien oft deutlich voneinander verschieden. Die Abweichungen können so groß werden, dass die Verfahren an ihre Grenzen stoßen. In einem solchen Fall kann berechnet werden, dass der gesamte Datensatz wenig robust ist und welche Trajektorien(-gruppe) am ehesten als robust eingestuft werden kann.

## 4.2 Ergebnisse mit FlyATM4E

Anders als bei ATM4E werden für FlyATM4E zuerst die Clusteranalyse und die Ähnlichkeitsberechnung betrachtet. Die gewünschten Informationen sind nicht wie bei ATM4E robuste Trajektorien beziehungsweise eine Abstufung der Robustheit, sondern die Robustheit verschiedener Wetterlagen zu analysieren. Dafür werden die verschiedenen Wetterlagen wie ein einziger Datensatz behandelt und die Clusteranalyse auf alle Trajektorien unabhängig vom Wetter angewendet. Die Ergebnisse der Clusteranalyse sind in 4.2.1 aufgeführt. Die Ergebnisse der Ähnlichkeitsberechnung sind in 4.2.2 zu finden. Die dadurch als robust klassifizierten Trajektorien werden anschließend unterschieden. Es wird diskutiert, ob und welche Wetterlagen besonders viele robuste Trajektorien besitzen und ob es einen speziellen Gewichtungsbereich zwischen Kosten und Umwelteinflüssen gibt, der auf dieselbe Weise als robuster als andere klassifiziert werden kann. Dies geschieht in 4.2.3. Anschließend werden die Ergebnisse in 4.2.4 zusammengefasst.

Anders als bei ATM4E wurden keine Trajektorien zwischen bestimmten Flughäfen berechnet, sondern Verbindungen zwischen Bereichen bestimmt. Die Berechnung gängiger Flugverbindungen wie in ATM4E stellte einen zu großen Berechnungsaufwand dar, weshalb die Verbindung zwischen Regionen betrachtet wird, die stellvertretend für mehrere Flughäfen stehen. Abbildung 4.23 zeigt zwei Karten. Links sind alle zu betrachtenden Flüge abgebildet, rechts die Zusammenfassung mehrerer Flüge zu einer Route, die näherungsweise den Trajektorienverlauf beschreibt. Die Streckenbeschreibung erfolgt nun durch die in der Karte dargestellten Gebiete, im betrachteten Fall von CG nach FG. Dies entspricht einem Flug von der spanischen Riviera in den Großraum bei London.

### 4.2.1 Ergebnisse der Clusteranalyse

Um bei der Clusteranalyse feststellen zu können, welche Trajektorien entfernt wurden, ist zunächst ein Überblick über alle vorhandenen Trajektorien hilfreich. Dieser ist in Abbildung 4.24 dargestellt. Wie bei ATM4E existiert eine Farbskala, die die Gewichtungsfaktoren der Trajektorien angibt. Rot steht für eine umweltoptimale Trajektorie, schwarze für eine kostenoptimierte Route und entsprechende Abstufungen sind Kompromisse. In Abbildung 4.24 sind alle gegebenen Trajektorien zu sehen. Eine interessante Eigenschaft ist hier, dass umweltoptimale Trajektorien bei einer bestimmten Wetterlage und kostenoptimale Trajektorien bei einer anderen Wetterlage sich jeweils entsprechen. Werden alle Trajektorien betrachtet, sind diese nicht mehr in etwa farblich geordnet wie bei ATM4E, sondern vermischen sich.

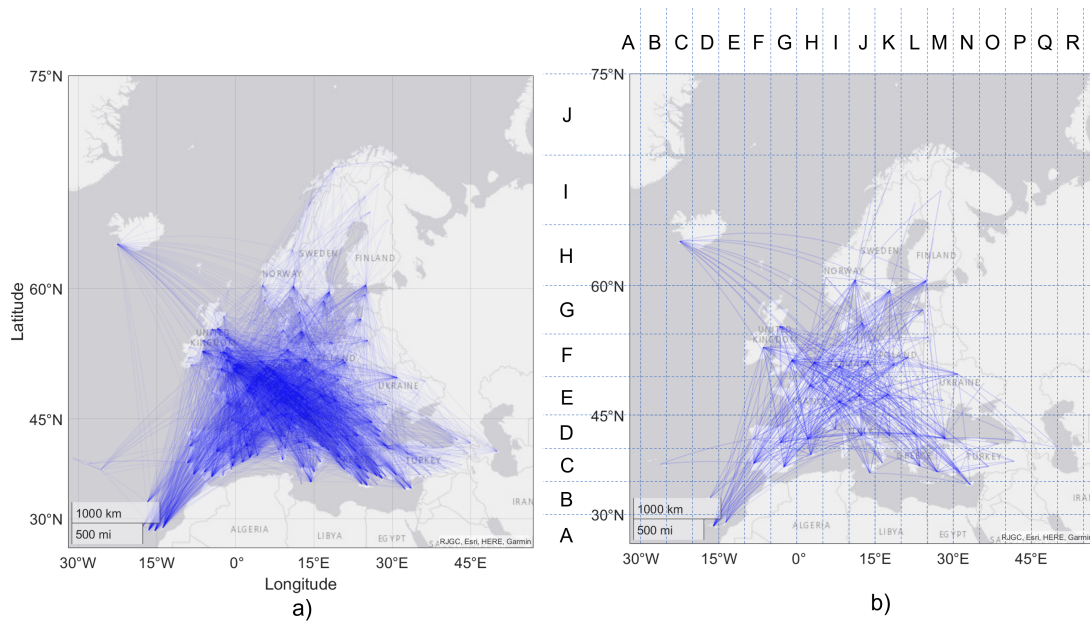


Abbildung 4.23: Aufteilung von Regionen im Projekt FlyATM4E Quelle: [53].

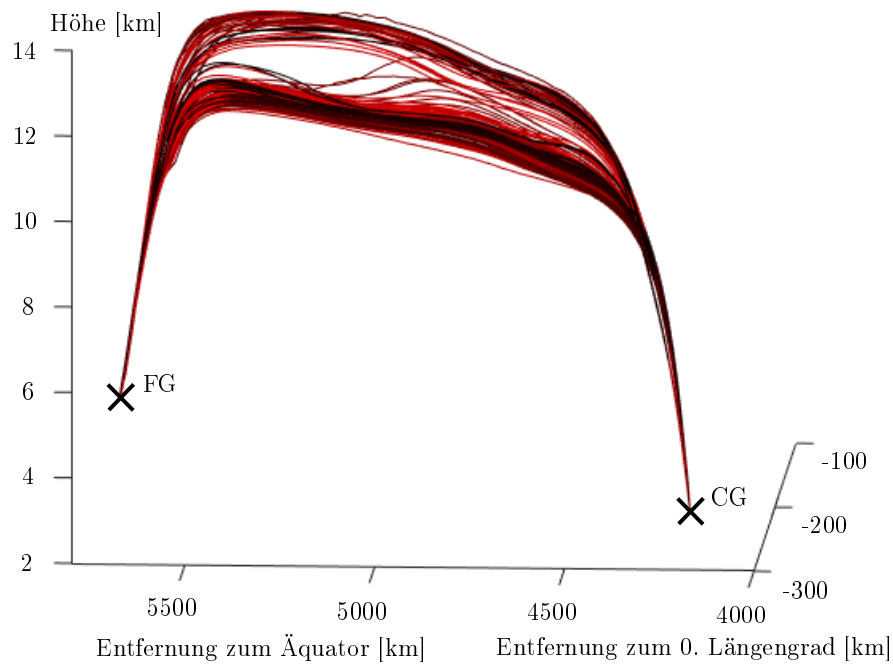


Abbildung 4.24: Interpolierte Trajektorien zu Datensatz „0001\_CG-FG\_Climate“.

Wird anschließend die Clusteranalyse in Abbildung 4.25 für diesen Datensatz betrachtet, wurde in drei Cluster aufgeteilt und alle Trajektorien in größerer Höhe gelöscht. Ausreißer sind bereits in Abbildung 4.24 nicht vorhanden und müssen entsprechend nicht gelöscht werden. Das türkise Cluster ist erneut das dominante Cluster, dunkelblau und gelb decken jeweils einen schmalen Randbereich ab.

Die Clusteranalyse kommt somit zu dem Schluss, dass die Trajektorien in geringer Höhe robuster sind als die in größerer Höhe. In Abschnitt 4.1 wurde erwähnt, dass dieses Ergebnis nicht unbedingt erwartet wurde. Es wurde davon ausgegangen,

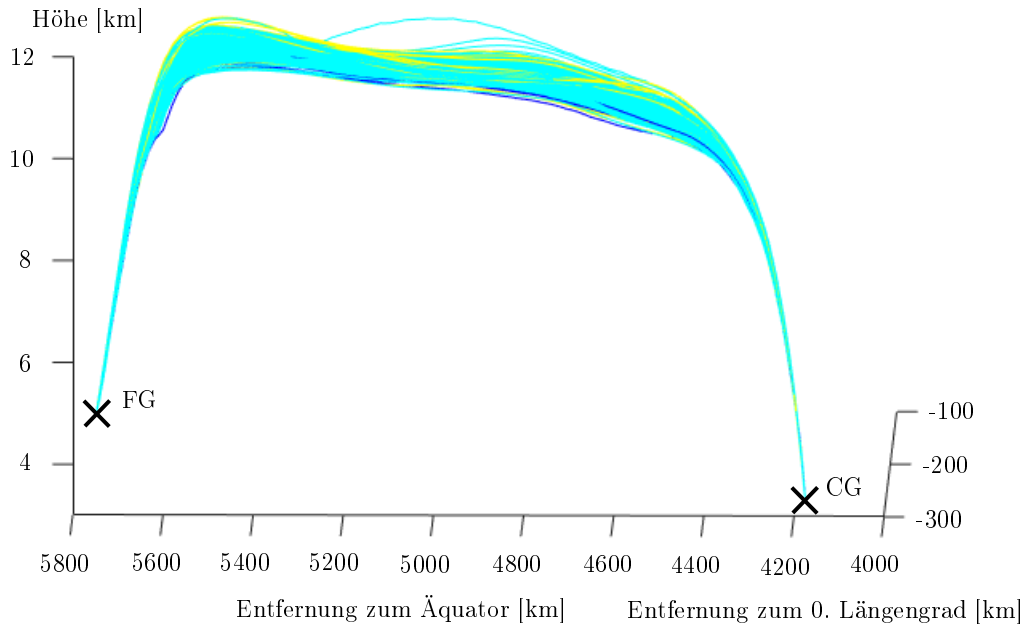


Abbildung 4.25: Endgültige Clustereinteilung für Datensatz „0001\_CG-FG\_Clima-te“.

dass die kostenoptimalen Trajektorien, die sich in größeren Höhen befinden, robuster sind, da ihr einziger Einflussfaktor die Treibstoffmenge ist, die sich in etwa mit der Höhe skaliert. Die umweltoptimalen Trajektorien befinden sich bei ATM4E in den geringsten Höhen, die zugelassen sind bei etwa 10 500 m. Da mehr Einflussfaktoren für sie gelten, müssten diese Trajektorien öfter Abweichungen erfahren und damit weniger robust sein. Die Ursache für ein Ergebnis entgegen der Erwartung lag hier in den Pareto-Kurven, die ungleichmäßig verteilt waren. Die umweltoptimalen Trajektorien wurden mit einer geringeren Abstufung in der Gewichtung untersucht als die kostenoptimalen, dadurch mehr Trajektorien erstellt, sodass die umweltoptimalen Trajektorien als robuster eingestuft wurden. Bei FlyATM4E sind es jedoch nicht ausschließlich die umweltoptimalen Trajektorien, sondern eine Mischung verschiedener Gewichtungen von Kosten und Umwelteinflüssen, die sich in geringer Höhe befinden. Warum die Clusteranalyse diese Trajektorien als robuster einstuft, muss einen anderen Grund haben.

## 4.2.2 Ergebnisse der Ähnlichkeitsberechnung

Auch die Ähnlichkeitsberechnung wurde auf alle Trajektorien aus dem gegebenen Datensatz aus FlyATM4E angewendet. In Abbildung 4.26 ist das Ergebnis zu sehen. Die Trajektorien in geringer Höhe bilden eine große Gruppe, die grün eingefärbt ist. Anders als bei der Clusteranalyse wird die Trajektorien in großer Höhe nicht gelöscht, sondern als weitgehend eine Gruppe identifiziert. In Kombination mit der Clusteranalyse lässt dies darauf schließen, dass die in der Clusteranalyse gelöschten Trajektorien nicht als vollständig nicht robust einzustufen sind, sondern nur etwas weniger robust als die Trajektorien in geringer Höhe sind. Da die Clusteranalyse ein eher grobes Verfahren zur Sortierung darstellt, befinden sich die gelöschten Trajek-

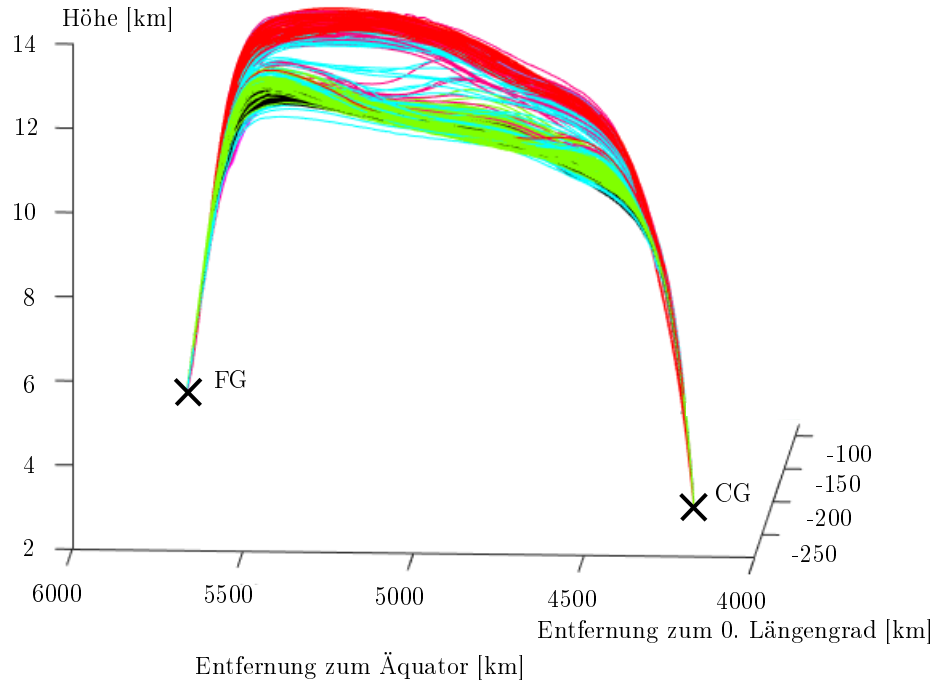


Abbildung 4.26: Berechnete Ähnlichkeit für Datensatz „0001\_CG-FG\_Climate“.

torien sehr wahrscheinlich dicht an der Grenze zwischen den Bereichen, die bei der Clusteranalyse als robust und nicht robust eingestuft werden.

### 4.2.3 Plots der Pareto-Fronten

In Abbildung 4.27 sind die Gewichtungen aller Trajektorien geplottet. Dabei wurden jeweils verschiedene Farben für unterschiedliche Wetterlagen verwendet. Auffällig an der ungefilterten Pareto-Front ist der stufenartige Verlauf. Ein solcher Verlauf war bei ATM4E bei keiner Trajektorie zu erkennen. Zwischen den  $\Delta ATR$  von -0,3 und -0,32 liegt zudem ein Bereich, in dem nur wenige Gewichtungen liegen und die Dichte der Kreuze für die jeweilige Gewichtung entsprechend gering ist. Dieser Bereich steht nicht im Zusammenhang mit der zum Beispiel in Abbildung 4.24 zu sehenden Lücke zwischen den Trajektorien. Es handelt sich hierbei um andere Trajektorien, die jedoch keinem bestimmten Bereich der Pareto-Front zuzuordnen sind. Insgesamt sind die Kreuze unterschiedlicher Farbe nur schwierig miteinander vergleichbar. Innerhalb einer Farbe sind die Trajektorien wie ein Datensatz aus ATM4E zu betrachten. Durch die unterschiedlichen Wetterlagen ändern sich jedoch die Trajektorien, sodass Trajektorien mit fast derselben Gewichtung, aber unterschiedlichen Wetterlagen einen völlig anderen Verlauf erhalten.

Insgesamt konnte keine Systematik festgestellt werden, wie die ungefilterte Pareto-Front und der Trajektorienverlauf zusammenhängen. Der Trajektorienverlauf scheint maßgeblich von den sich von Datensatz zu Datensatz verändernden Wetterdaten geprägt zu sein. Da es sich um zehn typische Wetterlagen für die gewählte Region handelt, ist zwischen den Wetterdaten kein numerischer Zusammenhang gegeben.

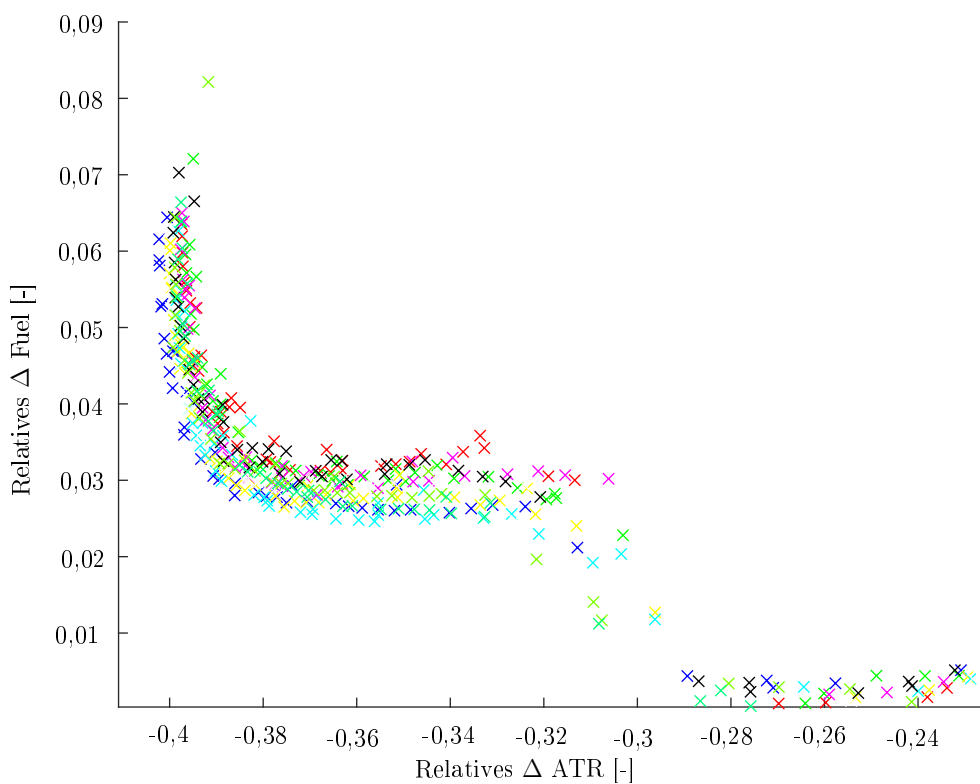


Abbildung 4.27: Paerto-Front des Trajektoriensatzes „0001\_CG-FG\_Climate“.

#### 4.2.4 Zusammenfassung für FlyATM4E

Für den gegebenen Fall wurde berechnet, dass die Trajektorien in geringerer Höhe robuster sind als die in größerer Höhe. Die Differenz bezüglich der Robustheit scheint jedoch gering zu sein, da nur die Clusteranalyse diese Daten löscht, die Ähnlichkeitsberechnung nicht. Anders als bei ATM4E ist durch die ungefilterte Pareto-Front kein Informationsgewinn gegeben. Aufgrund unterschiedlicher Wetterlagen und -daten ist die Vergleichbarkeit der Trajektorien nicht gegeben. Gehören die Gewichtungen zu verschiedenen Wetterlagen, können bei gleicher Gewichtung die Trajektorien einen deutlich unterschiedlichen Verlauf annehmen. Andererseits gehören eng beieinander liegende Trajektorien zu völlig verschiedenen Gewichtungen.

Da nur zwei vollständige Datensätze für FlyATM4E vorliegen, kann nicht getestet werden, ob die höhere Robustheit von Trajektorien in geringerer Höhe systematisch auftritt oder dies nur für den betrachteten Einzelfall gilt. Jedoch kann ausgesagt werden, dass verschiedene Wetterlagen den Zusammenhang zwischen Gewichtung und Trajektorienlage beeinflussen und die Trajektorien nicht wie bei ATM4E durch die ungefilterte Pareto-Front einer Gruppe mit bestimmten Eigenschaften zugeordnet werden können.



## 5. Zusammenfassung und Ausblick

Mit der Clusteranalyse und der Berechnung von Ähnlichkeit liegen zwei Methoden vor, um die Robustheit von Trajektorien zu klassifizieren. Die Clusteranalyse ist besonders für parallel verlaufende Trajektorien und weit verstreute Trajektorien mit sehr unterschiedlichem Verlauf geeignet. In einem Bereich zwischen weit gestreuten und eng beieinander liegenden parallelen Trajektorien weist die Clusteranalyse einige Schwächen auf. Die Berechnung von Ähnlichkeit hingegen funktioniert für parallel verlaufende Trajektorien und solche, die einige Abweichungen aufweisen. Jedoch ist die Ähnlichkeitsberechnung für sehr weit verstreute Trajektorien nicht gut anwendbar. Voraussetzung für die Ähnlichkeitsberechnung ist eine etwa gleiche Länge der Trajektorien, was in einem solchen Fall nicht gegeben ist. Damit eignen sich beide Verfahren für parallel verlaufende Trajektorien. Die Ähnlichkeitsberechnung kann die Schwächen der Clusteranalyse in dem Bereich zwischen parallelen Trajektorien und Trajektorien in beliebiger Anordnung ausgleichen. Für Trajektoriensätze mit weit verstreuten Trajektorien kann wiederum die Clusteranalyse besser als die Ähnlichkeitsberechnung eingesetzt werden. Die beiden Verfahren ergänzen sich und können somit die untersuchten Trajektoriensätze klassifizieren.

Die Klassifizierung der Robustheit der Trajektorien wurde bewusst nur sehr grob eingeteilt. Da für die Robustheit der Trajektorien die Nähe zu weiteren eine große Rolle spielt, kann eine Trajektorie für sich nicht robust sein. Es gibt lediglich Gruppen von Trajektorien, die in ihrer Gesamtheit robust sind. Gefundene Gruppen können eine Abstufung in ihrer Robustheit erhalten, wenn zum Beispiel die Abstände der Trajektorien innerhalb einer Gruppe sich deutlich von den Abständen in einer anderen Gruppe unterscheiden, sich die Trajektorienpositionen sehr unterscheiden oder wenn nicht beide Verfahren, Ähnlichkeitsberechnung und Clusteranalyse, dieselbe Gruppe bilden. Die programmierten MATLAB-Skripte lassen aber auch die scharfe Abstufung der Robustheit jeder Trajektorie zu. Eine solche Rangfolge ist jedoch nicht sinnvoll, da sie nur die gegebenen Daten berücksichtigt, jedoch nicht einbezieht, dass die Pareto-Kurven nicht nur durch die vorhandenen Punkte repräsentiert werden, sondern weitere Gewichtungen existieren, für die keine Trajektorie berechnet wurde. Daher werden nur Bereiche bestimmt, in denen sich viele Trajektorien befinden und die Bereiche somit als robust zu klassifizieren sind.

Ein bestimmte Gewichtung, die robuste Trajektorien erzeugt, ist weder bei den Daten aus ATM4E noch aus FlyATM4E festzustellen. Wo Bereiche robuster Trajektorien entstehen, ist bei jedem Trajektoriensatz anders. Jedoch können die Daten aus ATM4E anhand ihrer ungefilterten Pareto-Front in drei Gruppen aufgeteilt werden und so auch ohne Berechnung erste Aussagen zur Robustheit dieses Datensatzes getroffen werden. Besonders bei Gruppe 3 ist zu klären, ob die gegebenen Punktwolken tatsächlich so auftreten oder ob es sich dabei um Berechnungsfehler handelt.

Zur endgültigen Bestimmung der Robustheit ist es nach dem Stand dieser Arbeit notwendig, die Plots von einem fachkundigen Menschen auswerten zu lassen. Ein Mensch kann die Daten interpretieren und vermuten, wo zum Beispiel sich kontinuierliche Bereiche der Pareto-Front befinden könnten und damit verbunden ein sehr robuster Bereich liegt. Dem Computer hingegen fehlen für diese Analyse viele Informationen. Zum jetzigen Zeitpunkt ist die Kontinuität nur eine Vermutung, die nicht zutreffen muss. Es bedarf weiterer Recherche und Berechnungen, ob gewisse Bereiche als kontinuierlich angesehen werden dürfen und anhand welcher Kriterien dies zu begründen ist. Genauso ist der Grund festzustellen, warum einige Flüge sehr parallele Trajektorienverläufe aufweisen und andere Trajektorien weit verstreut liegen. Die Ursache dafür wird wahrscheinlich in den verwendeten Wetterdaten und -metriken liegen. Hier wäre eine Untersuchung zur Abhängigkeit von bestimmten Wetterlagen oder auch Regionen und der Robustheit der Trajektorien interessant. Auch weitere Untersuchungen der Daten aus FlyATM4E sind sinnvoll. Aufgrund der Fristen für FlyATM4E und der Datenmenge war im Rahmen dieser Arbeit die Analyse von zwei Datensätzen möglich. Neben der Robustheitsermittlung durch die Clusteranalyse und Ähnlichkeitsberechnung für genau diese beiden Flüge konnten keine allgemeinen Zusammenhänge zwischen den Daten hergestellt werden.

# Literaturverzeichnis

- [1] ACKERMANN, J.: Robust Flight Control System Design. In: *8th IFAC World Congress on Control Science and Technology for the Progress of Society* (1981)
- [2] AGARWAL, P. K. ; AVRAHAM, R. B. ; KAPLAN, H. ; SHARIR, M.: Computing the Discrete Frechet Distance in Subquadratic Time. In: *Cornell University* (2018)
- [3] ALT, H. W.: *Lineare Funktionalanalysis*. Springer-Verlag Berlin Heidelberg, 2012
- [4] ANDERSON, K. ; PETERS, G.: The trouble with negative emissions. In: *Science* (2016)
- [5] BAE, S.: JavaScript Data Structures and Algorithms. In: *Apress, Berkeley, CA* (2019)
- [6] BASRI, R. ; COSTA, L. ; GEIGER, D. ; JACOBS, D.: Determining the similarity of deformable shapes. In: *Vision Research* (1998)
- [7] BAUEN, A. ; BITOSS, N. ; GERMAN, L. ; HARRIS, A. ; LEOW, K.: Sustainable Aviation Fuels: Status, challenges and prospects of drop-in liquid fuels, hydrogen and electrification in aviation. In: *Johnson Matthey Technology Review* (2020)
- [8] BOPST, J. ; HERBENER, R. ; HÖLZER-SCHOPOHL, O. ; LINDMAIER, J. ; MYCK, T. ; WEISS, J.: Umweltschonender Luftverkehr lokal – national – international. In: *Redaktion Umwelt Bundesamt* (2019)
- [9] BURKHARDT, U. ; BOCK, L. ; BIER, A.: Mitigating the contrail cirrus climate impact by reducing aircraft soot number emissions. In: *Atmospheric Science* (2018)
- [10] CASTINO, F. ; YIN, F. ; GREWE, V. ; YAMASHITA, H. ; MATTHES, S. ; BAUMANN, S. ; DIETMÜLLER, S. ; SOLER, M. ; SIMORGH, A. ; LINKE, F. ; LÜHRS, B.: Seasonal Variability of Aircraft Trajectories reducing NO<sub>x</sub> -climate Impacts under a Multitude of Weather Patterns. In: *11th Sesar Innovation Days* (2022)
- [11] CAVALLO, A. ; DE MARIA, G. ; VERDE, L.: Robust flight control systems - A parameter space design. In: *Journal of Guidance Control and Dynamics* (1992)
- [12] CHIARAMONTI, D.: Sustainable Aviation Fuels: The Challenge of Decarbonization. In: *Energy Procedia* (2019)

- [13] COHEN, I. ; AYACHE, N. ; SULGER, P.: Tracking Points on Deformable Objects Using Curvature Information. In: *ECCV* (1992)
- [14] DEIDEWIG, F. ; DÖPELHEUER, A. ; LECHT, M.: Methods to Assess Aircraft Engine Emissions in Flight. In: *Proceedings of the ICA* (1996)
- [15] FRIZON, R.: *Overview: SESAR 2020 in a nutshell*. <https://www.sesarju.eu/activities>, 2022. – Aufgerufen: 23.09.2022
- [16] FRÖMMING, C. ; GREWE, V. ; BRINKOP, S. ; JÖCKEL, P.: Documentation of the EMAC submodels AIRTRAC 1.0 and CONTRAIL 1.0. In: *Geoscientific Model Development* (2014)
- [17] FRÖMMING, C. ; GREWE, V. ; JÖCKEL, P. ; BRINKOP, S. ; DIETMÜLLER, S. ; GARNY, H. ; PONATER, M. ; TSATI, E. ; MATTHES, S.: Climate cost functions as a basis for climate optimized flight trajectories. In: *Air Traffic Management Seminar* (2013)
- [18] GREEN, J.: Air Travel-Greener by Design. Mitigating the environmental impact of aviation: Opportunities and priorities. In: *Aeronaut* (2005)
- [19] GREWE, V. ; CHAMPOUGNY, T. ; MATTHES, S. ; FRÖMMING, C. ; BRINKOP, S. ; SØVDE, O. A. ; IRVINE, E. A. ; HALSCHEIDT, L.: Reduction of the air traffic's contribution to climate change: A REACT4C case study. In: *Atmospheric Environment* (2014)
- [20] GREWE, V. ; DAMERIS, M. ; FICHTER, C. ; LEE, D. S.: Impact of aircraft NO<sub>x</sub> emissions. Part 2: Effects of lowering the flight altitude. In: *Meteorologische Zeitschrift* (2002)
- [21] GREWE, V. ; FRÖMMING, C. ; MATTHES, S. ; BRINKOP, S. ; PONATER, M. ; DIETMÜLLER, S. ; JÖCKEL, P. ; GARNY, H. ; TSATI, E. ; DAHLMANN, K.: Aircraft routing with minimal climate impact: The REACT4C climate cost function modelling approach (V1.0). In: *Geoscientific Model Development* (2014)
- [22] GREWE, V. ; MATTHES, S. ; DAHLMANN, K.: The contribution of aviation NO<sub>x</sub> emissions to climate change: Are we ignoring methodological flaws. In: *Environmental Research Letters* (2019)
- [23] HARTJES, S. ; HENDRIKS, J. ; VISSER, H.: Contrail Mitigation through 3D Aircraft Trajectory Optimization. In: *16th AIAA Aviation Technology, Integration, and Operations Conference, Washington, DC, USA* (2016)
- [24] HODGES, J.L. ; LEHMANN, E.L.: Estimates of location based on rank tests. In: *The Annals of Mathematical Statistics* 34 (1963)
- [25] HOLLADAY, J. ; ABDULLAH, Z. ; HEYNE, J.: Sustainable Aviation Fuel: Review of Technical Pathways. In: *U. S. Department of Energy* (2020)
- [26] HUBER, F.J.: Robust Statistics. In: *Wiley, New York* (1981)
- [27] HUTTENLOCHER, D. P. ; KLANDERMAN, G. A. ; RUCKLIDGE, W. J.: Comparing Images Using the Hausdorff Distance. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* (1993)

- [28] IATA: *Our Commitment to Fly Net Zero by 2050*. <https://www.iata.org/en/programs/environment/flynetzero/>, 2021. – Aufgerufen: 23.09.2022
- [29] IATA: *Resolution on the Industry's Commitment to Reach Net Zero Carbon Emissions by 2050*. <https://www.iata.org/contentassets/dcd25da635cd4c3697b5d0d8ae32e159/iata-agm-resolution-on-net-zero-carbon-emissions.pdf>, 2021. – Aufgerufen: 23.09.2022
- [30] INDYK, P. ; MATOUSEK, J. ; SIDIROPOULOS, A.: Low-distortion Embeddings of Finite Metric Spaces. In: *Taylor & Francis Group* (2003)
- [31] IRVINE, E. A. ; HOSKINS, B. J. ; SHINE, K. P.: A simple framework for assessing the tradeoff between the climate impact of aviation carbon dioxide emissions and contrails for a single flight. In: *Environ* (2014)
- [32] JACOBSEN, M. Z.: Review of solutions to global warming, air pollution, and energy security. In: *Energy & Environmental Science* (2009)
- [33] JÖCKEL, P. ; KERKWEIG, A. ; POZZER, A. ; SANDER, R. ; TOST, H. ; RIEDE, H. ; BAUMGAERTNER, A. ; GROMOV, S. ; KERN, B.: Development cycle 2 of the modular earth submodel system (MESSy2). In: *Geoscientific Model Development* (2010)
- [34] JÖCKEL, P. ; TOST, H. ; POZZER, A. ; KUNZE, M. ; KIRNER, O. ; BRENNINKMEIJER, C. ; BRINKOP, S. ; CAI, D. ; DYROFF, C. ; ECKSTEIN, J.: Earth System Chemistry integrated Modelling (ESCiMo) with the Modular Earth. In: *Geoscientific Model Development* (2016)
- [35] JEKEL, C. F. ; VENTER, G. ; VENTER, M. P. ; STANDER, N. ; HAFTKA, R. T.: Similarity measures for identifying material parameters from hysteresis loops using inverse analysis. In: *International Journal of Material Forming* (2018)
- [36] KÜGLER, D.: *SESAR 2020*. <https://www.dlr.de/fl/desktopdefault.aspx/tabid-13648/>, 2022. – Aufgerufen: 23.09.2022
- [37] LAURES, G. ; SZYMIK, M.: Grundkurs Topologie. In: *Berlin, Heidelberg: Springer Berlin Heidelberg* (2015)
- [38] LEE, D. S. ; FAHEY, D. W. ; SKOWRON, A. ; ALLEN, M. R. ; BURKHARDT, U. ; CHEN, Q. ; DOHERTY, S. J. ; FREEMAN, S. ; FORSTER, P. M. ; FUGLESTVEDT, J. ; GETTELMAN, A. ; LEON, R. R. ; LIM, L. L. ; LUND, M. T. ; MILLAR, R. J. ; OWEN, B. ; PENNER, J. E. ; PITARI, G. ; PRATHER, M. J. ; SAUSEN, R. ; WILCOX, L. J.: The contribution of global aviation to anthropogenic climate forcing for 2000 to 2018. In: *Atmos. Environ.* (2021)
- [39] LÜHRS, B. ; LINKE, F. ; MATTHES, S. ; GREWE, V. ; YIN, F.: Climate Impact Mitigation Potential of European Air Traffic in a Weather Situation with Strong Contrail Formation. In: *MDPI* (2020)
- [40] LINKE, F.: Vorlesung Luftverkehr und Umwelt. In: *DLR* (2022)

- [41] MAGNI, J.-F. ; BENNANI, S. ; TERLOUW, J.: Robust Flight Control - A Design Challenge. In: *Lecture Notes in Control and Information Sciences* (1997)
- [42] MANEN, J. ; GREWE, V.: Algorithmic climate change functions for the use in eco-efficient flight planning. In: *Transportation Research Part D: Transport and Environment* (2019)
- [43] MARJANI, S. ; TESCHE, M. ; BRÄUER, P. ; SOURDEVAL, O. ; QUAAS, J.: Satellite observations of the impact of individual aircraft on ice crystal number in thin cirrus clouds. In: *Geophysical Research Letters* (2022)
- [44] MATTHES, S.: *Air Traffic Management for Environment: Partners*. <https://www.atm4e.eu/partners/>, 2016. – Aufgerufen: 23.09.2022
- [45] MATTHES, S.: *Project ATM4E*. <https://www.atm4e.eu/index.html>, 2018. – Aufgerufen: 23.09.2022
- [46] MATTHES, S.: *FlyATM4E*. <https://flyatm4e.eu/>, 2020. – Aufgerufen: 23.09.2022
- [47] MATTHES, S. ; GREWE, V. ; DAHLMANN, K. ; FRÖMMING, C. ; IRVINE, E. ; LIM, L. ; LINKE, F. ; LÜHRS, B. ; OWEN, B. ; SHINE, K. ; STROMATAS, S. ; YAMASHITA, H. ; YIN, F.: A Concept for Multi-Criteria Environmental Assessment of Aircraft Trajectories. In: *MDPI* (2017)
- [48] MATTHES, S. ; GREWE, V. ; DAHLMANN, K. ; FRÖMMING, C. ; IRVINE, E. ; LIM, L. ; LINKE, F. ; LÜHRS, B. ; OWEN, B. ; SHINE, K. P.: A Concept for Multi-Criteria Environmental Assessment of Aircraft Trajectories. In: *Aerospace* (2017)
- [49] MATTHES, S. ; GREWE, V. ; LEE, D. ; LINKE, F. ; SHINE, K. ; STROMATAS, S.: ATM4E: A concept for environmentally-optimized aircraft trajectories. In: *Greener Aviation* (2016)
- [50] MATTHES, S. ; LÜHRS, B. ; DAHLMANN, K. ; GREWE, V. ; LINKE, F. ; YIN, F. ; KLINGMAN, E. ; SHINE, K. P.: Climate-Optimized Trajectories and Robust Mitigation Potential: Flying ATM4E. In: *MDPI* (2020)
- [51] MATTHES, S. ; LIM, L. ; BURKHARDT, U. ; DAHLMANN, K. ; DIETMÜLLER, S. ; GREWE, V. ; HASLERUD, A. S. ; J. HENDRICKS AN, B. O. ; PITARI, G. ; RIGHI, M. ; SKOWRON, A.: Mitigation of Non-CO<sub>2</sub> Aviation's Climate Impact by Changing Cruise Altitudes. In: *Aerospace* (2021)
- [52] MATTHES, S. ; SCHUMANN, U. ; GREWE, V. ; FRÖMMING, C. ; DAHLMANN, K. ; KOCH, A. ; MANNSTEIN, H.: Climate Optimized Air Transport. In Atmospheric Physics: Background-Methods Trends. In: *Springer: Berlin/Heidelberg, Germany* (2012)
- [53] MEUSER, M. M. ; LÜHRS, B. ; GOLLNICK, V. ; LINKE, F. ; MATTHES, S. ; SIMONE DIETMÜLLER, S. ; BAUMANN, S. ; SOLER, M. ; SIMORGH, A. ; YIN, F. ; CASTINO, F.: Mitigation of Aviation's Climate Impact through Robust Climate Optimized Trajectories in Intra-European Airspace. In: *33rd Congress of the International Council of the Aeronautical Sciences* (2022)

- [54] MYHRE, G. ; SHINDELL, D. ; BRÉON, F. ; COLLINS, W. ; FUGLESTVEDT, J. ; HUANG, J. ; KOCH, D. ; LAMARQUE, J. ; LEE, D. S. ; MENDOZA, B.: Anthropogenic and Natural Radiative Forcing. In: *Climate Change 2013: The Physical Science* (2013)
- [55] NG, K. S. ; FAROOQ, D. ; YANGA, A.: Global biorenewable development strategies for sustainable aviation fuel production. In: *Renewable and Sustainable Energy Reviews* (2021)
- [56] ROECKNER, E. ; BÄUML, G. ; BONAVENTURA, L. ; BROKOPF, R. ; ESCH, M. ; GIORGETTA, M. ; HAGEMANN, S. ; KIRCHNER, I. ; KORNBLUEH, L. ; MANZINI, E. ; RHODINA, A. ; SCHLESE, U. ; SCHULZWEIDA, U. ; TOMPKINS, A.: The atmospheric general circulation model ECHAM 5. PART I: Model description. In: *Max-Planck-Institut für Meteorologie* (2003)
- [57] SEBASTIAN, T. B. ; KLEIN, P. N. ; KIMIA, B. B.: On aligning curves. In: *IEEE Transactions on Pattern Analysis and Machine intelligence* (2003)
- [58] SIMPSON, L. ; BASHIOUM, D. ; CARR, E.: Computer flight planning in the North Atlantic. In: *Journal of Aircraft* (1965)
- [59] TEOH, R. R. ; SCHUMANN, U. U. ; MAJUMDAR, A. A. ; STETTLER, M. E. J.: Mitigating the Climate Forcing of Aircraft Contrails by Small-Scale Diversions and Technology Adoption. In: *Environ* (2020)
- [60] THE MATHWORKS, Inc.: *Cluster Analysis - Unsupervised learning techniques to find natural groupings and patterns in data.* <https://de.mathworks.com/help/stats/cluster-analysis.html>, 2022. – Aufgerufen: 23.09.2022
- [61] THE MATHWORKS, Inc.: *Fit curve of surface to data.* <https://de.mathworks.com/help/curvefit/fit.html#bto2vuv-1-fitOptions>, 2022. – Aufgerufen: 23.09.2022
- [62] THIESSEN, S.: *Recherche zu Ansätzen für die Bemessung der Ähnlichkeit von vierdimensionalen Flugzeugtrajektorien*, TU Hamburg, Projektarbeit, 2022
- [63] VLACHOS, M. ; KOLLIOS, G. ; GUNOPULOS, D.: Discovering similar multi-dimensional trajectories. In: *Proceedings - International Conference on Data Engineering* (2002)
- [64] VOSHAAR, T.: *Höheres Krebsrisiko durch Luftverschmutzung.* <https://www.lungenaerzte-im-netz.de/news-archiv/meldung/article/hoeheres-krebsrisiko-durch-luftverschmutzung/>, 2016. – Aufgerufen: 23.09.2022
- [65] WILHELM, L. ; GIERENS, K. ; ROHS, S.: Meteorological Conditions that Promote Persistent Contrails. In: *Applied Science* (2022)
- [66] YAMASHITA, H. ; YIN, F. ; GREWE, V. ; JÖCKEL, P. ; MATTHES, S. ; KERN, B. ; DAHLMANN, K. ; FRÖMMING, C.: Various aircraft routing options for air traffic simulation in the chemistry-climate model EMAC 2.53: AirTraf 2.0. In: *Geoscientific Model Development* (2019)

- [67] YAMASHITA, H. ; YIN, F. ; GREWE, V. ; JÖCKEL, P. ; MATTHES, S. ; KERN, B. ; DAHLMANN, K. ; FRÖMMING, C.: Newly developed aircraft routing options for air traffic simulation in the chemistry–climate model EMAC 2.53: AirTraf 2.0. In: *Geoscientific Model Development* (2020)
- [68] YIN, F. ; GREWE, V. ; CASTINO, F. ; RAO, P. ; MATTHES, S. ; YAMASHITA, H. ; DAHLMANN, K. ; FRÖMMING, C. ; DIETMÜLLER, S. ; PETER, P. ; KLINGAMAN, E. ; SHINE, K. ; LÜHRS, B. ; LINKE, F.: Predicting the climate impact of aviation for en-route emissions: The algorithmic climate change function sub model ACCF 1.0 of EMAC 2.53. In: *Geoscientific Model Development* (2021)
- [69] YIN, F. ; GREWE, V. ; MANEN, J. ; MATTHES, S. ; YAMASHITA, H. ; IRVINE, E. ; SHINE, K. P. ; LÜHRS, B. ; LINKE, F.: Verification of the ozone algorithmic climate change functions for predicting the short-term NO<sub>x</sub> effects from aviation en-route. In: *International Conference on Research in Air Transportation (ICRAT)* (2018)
- [70] YOUNES, L.: Computable Elastic Distances between Shapes. In: *Society for Industrial and Applied Mathematics* (1998)



# Abbildungsverzeichnis

2.1	101 optimierte Trajektorien für einen Flug von Gran Canaria nach Trondheim. . . . .	14
2.2	510 optimierte Trajektorien für einen Flug von Granada nach London bei zehn verschiedenen Wetterlagen. . . . .	15
2.3	Hierarchie der Ordnungsstrukturen für Mengen. . . . .	17
2.4	Methode der minimalen Energie mit diskretisierten und kontinuierlichen Trajektorien. . . . .	22
2.5	Trajektoriensatz „00005_GCLP_ENVA“ mit der Unterscheidung in zehn harte Cluster. . . . .	24
2.6	Trajektoriensatz „00005_GCLP_ENVA“ mit der Unterscheidung in fünf weiche Cluster. . . . .	25
2.7	Trajektoriensatz „00005_GCLP_ENVA“ ohne Cluster. . . . .	26
2.8	Trajektoriensatz „00005_GCLP_ENVA“ mit gelöschten Elementen und der Einteilung in fünf Cluster. . . . .	27
2.9	Beispiele für Clusterbildung für eine festgelegte Anzahl von Clustern. . . . .	28
2.10	Beispiele für Clusterbildung für festgelegte maximale Abstände. . . . .	28
2.11	Beispiel für Clusterbildung ohne hierarchische Struktur. . . . .	29
2.12	Beispiel für Clusterbildung mit hierarchischer Struktur. . . . .	29
2.13	Trajektoriensatz „00001_UBBB_ELLX“ mit ausschließlich stabilen Trajektorien. . . . .	31
2.14	Trajektoriensatz „00005_GCLP_ENVA“ mit einigen instabilen Trajektorien. . . . .	32
3.1	Gewichtung von Kosten- und Umweltparametern für Datensatz „00008_EFHK_GCTS“. . . . .	37
3.2	Verknüpfung der Funktionen für Ähnlichkeitsberechnung innerhalb von MATLAB. . . . .	39
3.3	Trajektoriensatz „00115_EKYT_GCLP“ mit Überflug des 0. Längengrades. . . . .	48
3.4	Trajektorienpunkte zu Datensatz „00001_UBBB_ELLX“. . . . .	49

3.5	Interpolierte Trajektorien zu Datensatz „00001_UBBB_ELLX“.	50
3.6	Trajektorienpunkte zu Datensatz „00113_EGCC_LTBA“.	50
3.7	Interpolierte Trajektorien zu Datensatz „00113_EGCC_LTBA“.	52
3.8	Schematische Darstellung der Berechnung durch ein diskretes Verfahren.	54
3.9	Schematische Darstellung der Berechnung durch ein (nahezu) kontinuierliches Verfahren.	55
3.10	Beispielhafter Vergleich von Trajektorien mit verschiedener Lage.	56
3.11	Vergleich von Trajektorien mit unterschiedlicher Referenz.	58
3.12	Vergleich identischer Trajektorien mit unterschiedlicher Auflösung.	59
3.13	Vergleichsschritte mit der Fréchet Metrik.	60
3.14	Beispiel für Trajektorien in umgekehrter zeitlicher Reihenfolge.	61
3.15	Beispiel für Trajektorie mit einem versetzten Datenpunkt.	62
3.16	Beispiel für Trajektorie mit fehlerhaftem Datenpunkt.	63
3.17	Skalierbarkeit bei unterschiedlicher Wahl des Koordinatensystems.	65
3.18	Stauen von zwei Trajektorien.	65
3.19	Zwei Trajektorien, die durch Integration verglichen werden.	66
3.20	Körper mit verschiedenen Positionen und Geschwindigkeiten.	67
3.21	Funktionsverknüpfung für das Anpassen an einen Zeichensatz.	69
3.22	Beispiel für verschiedene Auswahl von Zeichensätzen.	71
3.23	Iterationsschritt für die Festlegung eines Zeichensatzes.	72
3.24	Vergleich von dynamischer und statischer Zeichengröße.	73
3.25	Wahl von verschiedenen Zeichengrößen.	74
3.26	Abschnittsweise Anpassung an Trajektorie.	75
3.27	Anpassung per Verfahren der kleinsten Differenz.	76
3.28	Trajektorien approximiert durch Zeichensatz mit vier Zeichen.	78
3.29	Schrittweise Verwendung von kleineren Zeichen.	79
3.30	Beispiel für Clusterbildung für gegebene Punkte.	81
3.31	Bestimmung einer geeigneten Clusteranzahl.	82
3.32	Clusterbildung am Beispiel von Datensatz „00005_GCLP_ENVA“ mit Löschen von Daten.	83
3.33	Clusterbildung am Beispiel von Datensatz „00005_GCLP_ENVA“ ohne Löschen von Daten.	84
3.34	Gleichzeitige Clusteranalyse von 10 Datensätzen aus FlyATM4E.	86

3.35	Beispiel für Einzelvergleich und anschließendes Zusammenfügen. . . .	87
3.36	Clusteranalyse für Daten mit gleichen Abständen. . . . .	88
3.37	Ablaufdiagramm der Clusteranalyse in MATLAB. . . . .	89
4.1	Pareto-Front des Trajektoriensatzes „00001_UBBB_ELLX“ aus ATM4E.	92
4.2	Pareto-Front des Trajektoriensatzes „00025_ESSA_GCRR“ aus ATM4E.	93
4.3	Pareto-Front des Trajektoriensatzes „00045_GCLP_ENBO“ aus ATM4E.	94
4.4	Endgültige Clustereinteilung für Datensatz „00001_UBBB_ELLX“. .	95
4.5	Endgültige Clustereinteilung für Datensatz „00001_UBBB_ELLX“ gedreht. . . . .	95
4.6	Endgültige Clustereinteilung für Datensatz „00016_EDDH_GCLP“. .	96
4.7	Pareto-Front des Trajektoriensatzes „00016_EDDH_GCLP“ aus ATM4E.	97
4.8	Endgültige Clustereinteilung für Datensatz „00025_ESSA_GCRR“. .	98
4.9	Pareto-Front des Trajektoriensatzes „00080_GCRR_EKCH“. . . . .	98
4.10	Interpolierte Trajektorien zu Datensatz „00080_GCRR_EKCH“. . . .	99
4.11	Endgültige Clustereinteilung für Datensatz „00080_GCRR_EKCH“. .	100
4.12	Interpolierte Trajektorien zu Datensatz „00045_GCLP_ENBO“. . . .	101
4.13	Endgültige Clustereinteilung für Datensatz „00045_GCLP_ENBO“. .	102
4.15	Interpolierte Trajektorien zu Datensatz „00058_GCLP_EKCH“. . . .	102
4.14	Pareto-Front des Trajektoriensatzes „00058_GCLP_EKCH“. . . . .	103
4.16	Endgültige Clustereinteilung für Datensatz „00058_GCLP_EKCH“. .	104
4.17	Berechnete Ähnlichkeit für Datensatz „00016_EDDH_GCLP“. . . . .	105
4.18	Berechnete Ähnlichkeit für Datensatz „00001_UBBB_ELLX“. . . . .	106
4.19	Berechnete Ähnlichkeit für Datensatz „00025_ESSA_GCRR“. . . . .	107
4.20	Berechnete Ähnlichkeit für Datensatz „00080_GCRR_EKCH“. . . . .	107
4.21	Berechnete Ähnlichkeit für Datensatz „00045_GCLP_ENBO“. . . . .	108
4.22	Berechnete Ähnlichkeit für Datensatz „00058_GCLP_EKCH“. . . . .	109
4.23	Aufteilung von Regionen im Projekt FlyATM4E Quelle: [53]. . . . .	111
4.24	Interpolierte Trajektorien zu Datensatz „0001_CG-FG_Climate“. . .	111
4.25	Endgültige Clustereinteilung für Datensatz „0001_CG-FG_Climate“. .	112
4.26	Berechnete Ähnlichkeit für Datensatz „0001_CG-FG_Climate“. . . .	113
4.27	Paerto-Front des Trajektoriensatzes „0001_CG-FG_Climate“. . . . .	114



# A. Anhang

Alle Daten und MATLAB-Skripte sind digital beigelegt. Aufgrund des Umfangs wurde auf das Abdrucken des MATLAB-Codes verzichtet.