

Institut für Parallele und Verteilte Systeme  
Universität Stuttgart  
Universitätsstraße 38  
D-70569 Stuttgart

Bachelorarbeit Nr. 2

# **Partitionierung von Modellen für räumlich verteilte Umgebungsphänomene**

Maren Tilk

<b>Studiengang:</b>	Informatik
<b>Prüfer:</b>	Prof. Dr. rer. nat. Dr. h. c. Kurt Rothermel
<b>Betreuer:</b>	Dipl.-Inf. Damian Philipp
<b>begonnen am:</b>	22. März 2012
<b>beendet am:</b>	21. September 2012
<b>CR-Klassifikation:</b>	C.2.1



## Kurzfassung

Beim *opportunistischen Public Sensing* werden mobile Geräte wie beispielsweise Smartphones benutzt, um ein räumlich verteiltes Umgebungsphänomen, wie eine Temperaturverteilung oder einen Geräuschpegel zu erfassen. Dazu werden Messpunkte festgelegt, an denen ein Wert abgefragt werden soll. Damit dies möglich ist, muss sich im Umkreis eines Messpunktes mindestens ein mobiles Gerät befinden, das den entsprechenden Wert aufnehmen kann. Durch Einsatz eines Modells kann der Aufwand der Erfassung reduziert werden. Ein solches Modell bildet die Korrelation der Werte an verschiedenen Punkten ab. Als gut geeignet für die meisten Umgebungsphänomene hat sich eine multivariate Gaußverteilung erwiesen, die auch in dieser Arbeit benutzt wird. Wenn für einen Teil der Messpunkte Werte bekannt sind, können an den anderen Stellen Werte aus dem Modell berechnet werden. Das kann dazu benutzt werden, die Zahl der angefragten Messpunkte zu reduzieren. Jedoch werden selten alle angefragten Messwerte auch erhalten. Für diesen Fall ist es möglich, auf Basis des Modells alternative Messwerte anzufragen. Diese werden auf den mobilen Geräten ausgewählt. Dadurch entsteht jedoch zusätzlicher Aufwand, da das Modell bei jeder Anfrage mitgeschickt werden muss. In dieser Arbeit wird untersucht, ob sich durch die Partitionierung des Modells und der Optimierung auf den Teilgebieten, dieser Aufwand reduzieren lässt. Damit ergibt sich eine Möglichkeit, eine Alternativplanung zu nutzen ohne den vergrößerten Nachrichtenaufwand in Kauf nehmen zu müssen.

## Abstract

*Opportunistic Public Sensing* uses mobile devices such as smartphones to measure spatially distributed phenomena like temperature or noisiness. It determines measure points at which data shall be taken. There has to be at least one mobile device near a measure point, which can gather the relevant data. Model-driven sensing reduces the costs of data collection. Such a model maps the correlation of data at different points. For most of the spatially distributed phenomena a multivariate Gaussian distribution is a suitable model and also used in this work. If values are known at some of the measure points, for the other points values can be derived from the model. This can be used to reduce the count of inquired measure points. However, usually not all of the inquired measure points actually can gather data. In such a case it is possible to select based on the model alternative points to get a value. However, this creates additional costs as the model has to be sent in each query. This work analyzes, if partitioning the model and optimization on the partitions can reduce these costs. This gives the possibility to use alternative selection without the negative aspect of the enlarged query size.



# Inhaltsverzeichnis

<b>1</b>	<b>Einleitung</b>	<b>7</b>
<b>2</b>	<b>Grundlagen</b>	<b>9</b>
2.1	Modellgetriebene Erfassung . . . . .	11
2.2	Multivariate Gaußverteilung . . . . .	11
2.3	Verwandte Arbeiten . . . . .	13
<b>3</b>	<b>Systemmodell und Anforderungen</b>	<b>15</b>
3.1	Systemmodell . . . . .	15
3.2	Anforderungen . . . . .	16
<b>4</b>	<b>Entwurf</b>	<b>19</b>
4.1	Aufbau und Ausführung einer Anfrage . . . . .	19
4.1.1	Sensorrollen . . . . .	20
4.2	Alternativplanung . . . . .	21
4.2.1	Ablauf der Alternativplanung . . . . .	21
4.2.2	Begrenzung der Neuauswahl . . . . .	22
4.3	Modellpartitionierung . . . . .	23
4.3.1	Modellpartitionierung mit Zusatzwissen . . . . .	24
4.3.2	Modellpartitionierung und Alternativplanung . . . . .	25
<b>5</b>	<b>Implementierung</b>	<b>27</b>
5.1	Alternativplanung . . . . .	27
5.2	Partitionierung des Modells . . . . .	28
5.2.1	Partitionierung ohne Zusatzwissen . . . . .	28
5.2.2	Partitionierung mit Zusatzwissen . . . . .	29
<b>6</b>	<b>Evaluation</b>	<b>31</b>
6.1	Methodologie . . . . .	31
6.2	Diskussion . . . . .	33
<b>7</b>	<b>Zusammenfassung und Ausblick</b>	<b>37</b>
	<b>Literaturverzeichnis</b>	<b>39</b>

## Abbildungsverzeichnis

---

2.1	Dichtefunktion der Gaußverteilung . . . . .	11
3.1	Darstellung des Systemmodells . . . . .	15
4.1	Nachrichtenformate . . . . .	21
6.1	Simulationsergebnisse zu Effizienz und Qualität . . . . .	32
6.2	Häufigkeit tatsächlich gelesener Sensoren . . . . .	33

## Verzeichnis der Algorithmen

---

5.1	Modellpartitionierung . . . . .	28
-----	---------------------------------	----

# 1 Einleitung

Smartphones erfreuen sich heutzutage immer größerer Beliebtheit, sodass immer mehr Menschen täglich ein solches benutzen. Zudem werden diese Geräte zunehmend leistungsfähiger. Das macht sie auch für verteilte Anwendungen interessant, die das Potenzial der Smartphones heranziehen, möglichst ohne den eigentlichen Benutzer des Gerätes zu beeinträchtigen. In dieses Gebiet fällt auch die Anwendung des *Public Sensing*. Smartphones können heutzutage eine Reihe verschiedener Sensorwerte erfassen wie zum Beispiel Temperatur, Helligkeit, Geräuschpegel, Bewegung usw. Zudem sind sie in der Lage, zum Beispiel mittels GPS, ihren Standort zu ermitteln. Wenn man nun die Position mit den Werten verbindet und zudem eine größere Menge solcher Daten erfasst und zusammenfügt, kann man in einem größeren Gebiet die Verteilung und Entwicklung einer bestimmten Größe, wie zum Beispiel der Temperatur, erfassen. Der Ansatz von *Public Sensing* besteht nun darin, sich das Vorhandensein vieler internetfähiger Smartphones zu Nutze zu machen. Anstatt in einem Gebiet Sensoren zur Werterfassung auszubringen, sollen die Smartphones, die sich in diesem Bereich befinden, ihre Sensorwerte weitergeben. Die Verwendung mobiler Geräte hat den Vorteil, dass kein zusätzlicher Aufwand durch das Verteilen fest installierter Messstationen entsteht.

Es bringt aber auch eine Reihe neuer Probleme mit sich. Für eine Messanfrage werden Punkte festgelegt, an denen ein Wert erfasst werden soll. Durch die Mobilität der Geräte kann jedoch nicht sicher gestellt werden, dass an jedem geforderten Punkt auch ein Messwert aufgenommen werden kann. Zudem verändert sich diese Situation ständig, da die Besitzer mit ihren Smartphones ständig in Bewegung sind. Das System muss sich daher auf eine häufig ändernde Platzierung der mobilen Geräte einstellen können.

Bei aller Optimierung des Messergebnisses sollte dennoch der bei der Erfassung entstehende Aufwand nicht übermäßig hoch sein. Eine Möglichkeit diesen zu minimieren besteht darin, ein Modell für das Gebiet aufzustellen und mit dessen Hilfe die Abdeckung durch gemessene Sensorwerte zu verbessern. Ein solches Modell beschreibt den Zusammenhang der Werte an verschiedenen Punkten. Auf diese Weise kann man für Punkte, an denen man keine Messwerte erhalten hat, solche möglichst annähernd berechnen. Außerdem kann man dadurch von vorn herein die Anzahl der tatsächlichen Sensorabfragen reduzieren und weitere Messwerte berechnen. Nach bestimmten Kriterien kann man so vor der Anfrage eine Auswahl an Punkten festlegen, die ausreichen, um auf deren Basis die Verteilung im gesamten Gebiet hinreichend gut zu ermitteln.

Dabei kann es jedoch immer vorkommen, dass für einige Punkte keine Messwerte empfangen werden können. In diesem Fall kann man nun versuchen, Messwerte von weiteren Punkten zu bekommen, um das Fehlen der Werte zu kompensieren und ein ähnlich gutes Ergebnis

zu bekommen. Für diese Alternativplanung kann man das gesamte Gebiet heranziehen, um weitere Messpunkte abzufragen. Dies ist jedoch immer mit einem Mehraufwand verbunden, da zusätzliche Informationen über das zugrunde liegende Modell verbreitet werden müssen.

Im Rahmen dieser Bachelorarbeit soll nun untersucht werden, ob sich durch Partitionierung des Gesamtgebiets und Optimierung auf den Teilgebieten die Alternativplanung mit dem Modell benutzen lässt, ohne dass der Nachteil durch den vergrößerten Nachrichtenaufwand auftritt. Dazu wird das Modell in Untermodelle aufgeteilt, die jeweils einen kleineren Teil des Gebiets abdecken. Das erhoffte Ziel dieser Gebietspartitionierung ist es, den Nachrichtenaufwand, der durch das Versenden des Modells entsteht, zu reduzieren ohne die Qualität des Ergebnisses zu verschlechtern. Für diese Bachelorarbeit wurde dieser Fragestellung mittels einer Simulation eines Public-Sensing-Netzwerkes nachgegangen.

Die Arbeit ist wie folgt aufgebaut. Zunächst werden in Kapitel 2 die Grundlagen des *Public Sensing* und der modellgetriebenen Erfassung vorgestellt, sowie auf einige verwandte Arbeiten eingegangen. In Kapitel 3 wird dann das Systemmodell beschrieben um anschließend die Ziele dieser Arbeit zu erläutern. Kapitel 4 umfasst den Entwurf des Systems. Hier wird zunächst der Ablauf der Alternativplanung vorgestellt sowie das daraus resultierende Problem des vergrößerten Nachrichtenaufwands. Anschließend wird die Modellpartitionierung als Lösung dieses Problems eingeführt. In Kapitel 5 werden Details zur Implementierung und Simulation vorgestellt. Kapitel 6 befasst sich mit der Auswertung der aus der Simulation gewonnenen Werten. Zuletzt enthält Kapitel 7 eine kurze Zusammenfassung der Arbeit sowie einen Ausblick auf weiterführende Untersuchungen.



## 2 Grundlagen

Beim *Public Sensing* werden mobile Geräte mit integrierten Sensoren eingesetzt, um Messwerte einzuholen, um mit diesen ein verteiltes Umgebungsphänomen zu erfassen. Darunter versteht man die Verteilung einer bestimmten Messgröße, wie zum Beispiel der Temperatur, deren Werte an verschiedenen Punkten korreliert sind. Als mobile Geräte eignen sich Smartphones, da diese heutzutage weit verbreitet sind. Die eigentlichen Benutzer können als Teil des Systems betrachtet werden, denn sie sorgen für die Mobilität der Geräte, die eine grundlegende Eigenschaft des *Public Sensing* ist. Zudem können sie in der Regel auch auf die gesammelten Daten zugreifen und haben so selbst einen Nutzen von der Bereitstellung der Sensoren ihres Smartphones. Die Anwendung läuft neben der normalen Benutzung des Geräts ohne den Nutzer zu stören. Die Herausforderungen liegen hauptsächlich darin, den richtigen Zeitpunkt für eine Erfassung zu bestimmen sowie den Benutzer des Smartphones möglichst wenig zu beeinträchtigen und die Energie des Geräts zu schonen. Zudem muss die Privatsphäre des Nutzers gewahrt werden. Beim *Public Sensing* werden jedoch hauptsächlich nicht personenbezogene Umgebungsphänomene betrachtet wie Temperatur, Geräuschpegel oder Verkehrsaufkommen. Dennoch können auch hier Informationen über den Benutzer gesammelt werden, die sich beispielsweise aus dessen Position oder Bewegung ergeben. Ein System, das die Privatsphäre des Benutzers schützt, ist *AnonySense* [CKK<sup>+</sup>08]. Die durch ein Public-Sensing-System gesammelten Informationen können für jeden von Interesse sein, daher ist ein Ziel des *Public Sensing*, diese allen zur Verfügung zu stellen.

Es gibt neben dem *Public Sensing* noch ähnliche Ansätze, die sich mit anderen Arten von Daten und deren Verwendung beschäftigen. Alle verschiedenen Ansätze werden unter dem Begriff des *People-centric Sensing* [CEL<sup>+</sup>08] zusammengefasst. Allen gemeinsam ist, dass die Daten mit den Sensoren mobiler Geräte erfasst werden, ohne den Benutzer des Geräts bei dessen Verwendung einzuschränken. Beim *Personal Sensing* werden stark auf den Nutzer bezogene Werte erfasst oder überwacht, entsprechend muss hierbei verstärkt die Vertraulichkeit der Daten beachtet werden. Ziel des *Social Sensings* ist es, Informationen die eigene Person betreffend anderen Leuten einer sozialen Gruppe zugänglich zu machen.

Bei der Art der Erfassung und Einbindung des Smartphonennutzers unterscheidet man zwischen zwei Ansätzen [LEM<sup>+</sup>08]. Bei der mitbestimmten („*participatory*“) Erfassung kann der Benutzer selbst entscheiden, ob und wann er dem System welche seiner Daten zur Verfügung stellt. Er hat die volle Kontrolle was erfasst wird und was er lieber privat hält. Dadurch werden etwaige Ergebnisse stark von der Selektivität der Teilnehmer beeinflusst. Die zweite Möglichkeit ist die opportunistische („*opportunistic*“) Erfassung. Hierbei nimmt die Anwendung selbstständig und ohne dass es vom Benutzer registriert wird Werte auf.

Wie sehr man den Benutzer des Smartphones in die Datenerfassung einbezieht, muss für jedes Einsatzgebiet einzeln entschieden werden. Anhaltspunkt dafür ist die Art der zu erfassenden Daten und einige andere Faktoren, die das erwartete Verhalten der Benutzer betreffen. Auch eine Mischform aus beiden Ansätzen ist denkbar. In [LEM<sup>+</sup>08] wird ein Modell vorgestellt, nach dem man entscheiden kann, welchen Ansatz man in einem bestimmten Fall verwendet. In dieser Arbeit wird die opportunistische Erfassung betrachtet.

Der Vorteil der Verwendung mobiler Geräte beim *Public Sensing* im Vergleich zur herkömmlichen Benutzung stationärer Sensoren an zuvor festgelegten Punkten besteht darin, dass keine Sensorgeräte ausgebracht werden müssen. Diese Art der Erfassung wird dadurch ermöglicht, dass heutzutage eine große Anzahl an mobilen Geräten, die Messwerte erfassen können, bereits vorhanden ist. Zudem ist *Public Sensing* wesentlich flexibler als ein statisches Sensornetzwerk. Bei letzterem kann man nach dem Installieren der Sensoren nur noch schwer etwas ändern und muss sich somit zuvor genau überlegen, an welchen Stellen man die Geräte anbringen soll. Es erfordert erneute Arbeit, wenn man weitere Sensoren hinzuzufügen will und besonders, wenn man die Positionen bereits installierter Sensoren nachträglich ändern will. Beim *Public Sensing* entfällt dieser Aufwand. Man kann ohne große Änderung des bestehenden Aufbaus weitere Messstellen hinzunehmen oder deren Position verändern. Man kann sogar bei Bedarf einfacher einen neuen Sensortyp hinzufügen. Man hat auch die Möglichkeit an Stellen Werte zu bekommen, an denen es nicht möglich ist einen Sensor zu montieren, zum Beispiel auf offenen Plätzen. Zudem hat es sich gezeigt, dass mobile Sensoren die Abdeckung eines Gebiets im Vergleich zu statischen Sensoren erhöhen [LBD<sup>+</sup>05].

Ein Nachteil besteht jedoch darin, dass man nicht weiß, an welcher Stelle sich ein Gerät befinden wird, da diese ihre Position ständig ändern können. Eine Möglichkeit, auf dieses Problem einzugehen, ist die Einführung sogenannter virtueller Sensoren, wie es in [PDR11] vorgestellt wird. Ein virtueller Sensor ist ein theoretisch ausgebrachter Sensor an einem festen Punkt, an dem man Messwerte eines bestimmten Typs erfassen will. Dadurch erhält man feste Punkte, an denen Erfassungen stattfinden sollen, allerdings nicht immer möglich sind. Will man nun zu diesem Punkt einen Sensorwert erhalten, so muss sich in dessen Nähe ein Smartphone oder auch eine festgelegte Anzahl an Geräten befinden, die dann den geforderten Sensorwert liefern. Welches Gerät für einen Sensor Werte liefern darf, kann am einfachsten dadurch geregelt werden, dass es innerhalb eines festgelegten Radius um den virtuellen Standpunkt liegen muss. Allgemeiner besitzt jeder virtuelle Sensor eine beliebig festgelegte Fläche, in der der Messwert aufgenommen werden muss. Die Flächen aller virtuellen Sensoren dürfen sich dabei an keiner Stelle überschneiden, so dass für jeden Standort eines Gerät eindeutig definiert ist, für welchen Sensor es Werte liefern kann. Die aufgenommenen Werte werden dann mit der Position des jeweiligen virtuellen Sensors verknüpft, in dessen Bereich sich das Gerät zum Zeitpunkt der Messung befunden hat. Im folgenden soll dieser Bereich *Erfassungsbereich* des virtuellen Sensors genannt werden.

Man klassifiziert die virtuellen Sensoren als verfügbar oder nicht verfügbar. Die Verfügbarkeit eines virtuellen Sensors hängt davon ab, ob sich in dessen Erfassungsbereich zum aktuellen Zeitpunkt ein Smartphone befindet, das den erforderlichen Messwert bereitstellen kann. Befindet sich kein passendes Gerät im Erfassungsbereich, so ist der Sensor nicht verfügbar. Durch

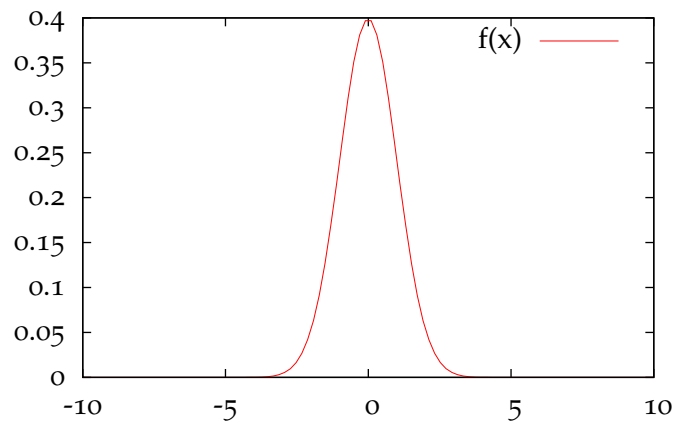


Abbildung 2.1: Dichtefunktion der Gaußverteilung mit  $\mu = 0$  und  $\sigma = 1$

die Beweglichkeit der mobilen Geräte lässt sich auch die Verfügbarkeit eines Sensors nicht vorherbestimmen. Sie zeigt sich erst durch eine erfolgreiche oder nicht erfolgreiche Abfrage. Wenn dem Sensornetzwerk nicht genügend mobile Geräte zur Verfügung stehen, ist die Wahrscheinlichkeit für einen virtuellen Sensor nicht verfügbar zu sein natürlich höher.

## 2.1 Modellgetriebene Erfassung

Um auf das Ausfallen einiger Sensorwerte zu reagieren, das aus der Nichtverfügbarkeit eines virtuellen Sensors resultiert, und nach Möglichkeit sogar die Anzahl der auszulesenden Sensorwerte zu reduzieren, bietet sich die Anwendung einer modellgetriebenen Erfassung an. Durch ein Modell wird die Korrelation der Werte an jedem Messpunkt innerhalb des untersuchten Gebiets untereinander dargestellt. Diese Korrelation muss im Vorfeld ermittelt und auf ein geeignetes Modell übertragen werden. Mit Hilfe dieses Modells ist es dann möglich, Werte an nicht erfassten Stellen aus den empfangenen Werten abzuleiten. Für unterschiedliche Arten von zu erfassenden Werten eigenen auch unterschiedliche Modelle besser oder schlechter. Ein Modell, das sich im Zusammenhang mit räumlich verteilten Umgebungsphänomenen gut eignet und zu diesem Zweck auch schon mehrfach verwendet wurde (so zum Beispiel in [DGM<sup>+</sup>04]), ist die Annahme einer multivariaten Gaußverteilung für die Verteilung der Messwerte.

## 2.2 Multivariate Gaußverteilung

Zunächst betrachtet man jeden Sensor als gaußverteilte Zufallsvariable. Die Dichtefunktion der Gauß- oder Normalverteilung ist gegeben durch die Formel 2.1.

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x - \mu)^2}{2\sigma^2}\right) \quad (2.1)$$

Das zugehörige Schaubild ist in Abbildung 2.1 dargestellt und auch als Gauß'sche Glockenkurve bekannt. Hierbei bezeichnet  $\mu$  den Mittel- oder Erwartungswert, der im Schaubild die Stelle des Maximums festlegt.  $\sigma$  steht für die Standardabweichung, die die Breite der Kurve bestimmt. Interessanter als die Standardabweichung ist  $\sigma^2$ , was als Varianz bezeichnet wird. Mit Werten für  $\mu$  und  $\sigma$  ist die Verteilungsfunktion vollständig bestimmt. Betrachtet man nun die Sensoren nicht mehr für sich allein sondern in ihrer Gesamtheit, muss man diese Darstellung erweitern. Die Verteilung der Zufallsvariablen ist nicht unabhängig voneinander. Sensoren, die nah beieinander oder in ähnlicher Umgebung liegen, können stark korrelieren. Man muss daher die Sensoren in einer gemeinsamen Verteilung modellieren. Die Erweiterung der Gaußverteilung für mehrere Variablen ist die multivariate Gaußverteilung. Aus der zweidimensionalen Glockenkurve wird in diesem Fall eine mehrdimensionale, glockenförmige Kurve, was bei mehr als zwei Variablen jedoch nicht mehr als Schaubild darstellbar ist. Das Schaubild besitzt immer eine Dimension mehr als Variablen, da der Dichtewert eine zusätzliche Dimension darstellt. Bei zwei Variablen erhält man so schon ein dreidimensionales Schaubild und so weiter. Das Maximum im Hinblick auf den Dichtewert befindet sich dann an der Stelle der Mittelwerte aller Variablen. Die multivariate Gaußverteilung besitzt also ein Mittelwertvektor, der als Einträge die Mittelwerte aller eindimensionalen Gaußverteilungen besitzt.

Die Varianz auf den mehrdimensionalen Fall zu übertragen ist dagegen aufwändiger. Um die Verteilung zu bestimmen, muss die Korrelation zwischen jedem einzelnen Variablenpaar angegeben werden. Die Abhängigkeit zwischen zwei verschiedenen Sensoren  $v$  und  $v'$  wird als Kovarianz bezeichnet und als  $\sigma_{v,v'}$  geschrieben. Sei nun die Menge aller virtuellen Sensoren als  $V = \{v_i\}$  bezeichnet, dann schreibt man die Kovarianzen aller modellierten virtuellen Sensoren in eine quadratische Matrix  $\Sigma_{V,V}$ . Als Einträge der Matrix steht in jeder Zeile  $i$  und Spalte  $j$  der Wert für  $\sigma_{v_i,v_j}$ . Auf der Hauptdiagonalen stehen die Werte  $\sigma_{v_i}^2$ , also die Varianzen der Sensoren. Der Mittelwertvektor sei mit  $M$  bezeichnet und beinhaltet entsprechend den Eintrag  $\mu_{v_i}$  an der Stelle  $i$ . Damit besteht ein multivariates Gaußmodell im Grunde aus der Kovarianzmatrix und dem Mittelwertvektor. Zu Beginn muss das Modell gelernt werden, indem die Kovarianzmatrix und der Mittelwertvektor geeignet mit Werten gefüllt werden. Solche Lernverfahren sollen aber nicht Teil dieser Bachelorarbeit sein.

Durch Messungen erhält man nun für einige der virtuellen Sensoren aktuelle Werte, mit denen man aus dem Modell Werte für die übrigen Sensoren berechnen kann. Die Menge der Sensoren, für die man tatsächlich Werte erhält sei  $W \subseteq V$ . Für die Sensoren  $U = V \setminus W$  will man nun Näherungswerte berechnen. Unter Berücksichtigung der gemessenen Werten  $P_W$  lassen sich mit den folgenden Gleichungen für jeden Sensor  $u \in U$  ein bedingter Mittelwert  $\mu_{u|P_W}$  und eine bedingte Varianz  $\sigma_{u|P_W}^2$  berechnen, durch die die Gaußverteilung eindeutig bestimmt wird.

$$\mu_{u|P_W} = \mu_u + \Sigma_{u,W} \Sigma_{W,W}^{-1} (P_W - M_W) \quad (2.2)$$

$$\sigma_{u|P_W}^2 = \sigma_u^2 - \Sigma_{u,W} \Sigma_{W,W}^{-1} \Sigma_{W,u} \quad (2.3)$$

Hierbei bezeichnet  $\Sigma_{u,W}$  den Vektor aus der Kovarianzmatrix  $\Sigma_{V,V}$ , der lediglich die Kovarianzen für den Sensor  $u$  mit den Sensoren  $w \in W$  enthält. Entsprechend ist die Matrix  $\Sigma_{W,W}$  auf die Zeilen und Spalten der Sensoren aus  $W$  reduziert.

Nachdem ein solches Modell für ein Public-Sensing-System erstellt wurde, kann es zur Reduzierung des Erfassungsaufwandes eingesetzt werden. Dazu werden nur von einer zunächst festgelegten Auswahl an virtuellen Sensoren Messwerte abgefragt. Die dabei erhaltenen Werte werden als  $P_W$  in obigen Formeln eingesetzt, um an den anderen virtuellen Sensoren ebenfalls Werte zu bekommen. Die ausgewählten Sensoren sollten nach Möglichkeit verfügbar sein, was bei der Auswahl jedoch nicht garantiert werden kann.

Für eine Anfrage möchte man eine Werteverteilung eines bestimmten Sensortyps erhalten. Um ein Ergebnis so nah wie möglich an der Realität zu bekommen könnte man einfach von sämtlichen Sensoren ein Messergebnis einholen. In diesem Fall ist aber auch der Energieaufwand maximal. Ziel ist es deshalb, eine Auswahl an Sensoren zu finden um nur von diesen Werte abzufragen. Wie bereits erläutert kann man dann auch mit dieser geringeren Anzahl an Sensorwerten und zusätzlichem Wissen über die Korrelation der Sensorpunkte für nicht durch Messung erfasste Stellen Näherungswerte berechnen. Am besten wäre es natürlich, die Sensorauswahl so klein wie möglich zu halten und dennoch ein annehmbares Ergebnis zu erhalten. Dafür benötigt man einen geeigneten Algorithmus zur Auswahl der Sensoren [KSGo8]. Hierfür eignet es sich, immer den Sensor zu finden, der eine maximale Verbesserung des Ergebnisses verspricht und dann solange weitere Sensoren hinzunehmen, bis das erwartete Ergebnis die festgelegte Qualität hat. Bei der Auswahl wird berücksichtigt, was aus bisherigen Anfragen über die Verfügbarkeit der Sensoren bekannt ist. Dennoch kann es immer vorkommen, dass unter den angefragten Sensoren nicht alle verfügbar sind. In diesem Fall wird durch das Fehlen eingeplanter Sensorwerte das Ergebnis schlechter und liegt damit nicht mehr innerhalb der festgelegten Toleranz.

Um dem Ausbleiben angefragter Sensorwerte entgegenzuwirken, gibt es den Ansatz der Alternativplanung, das bedeutet, es werden direkt auf den Knoten alternative Sensoren ausgewählt, wenn ein angefragter Sensor nicht verfügbar ist [Sta11]. Dazu wird bei einer Anfrage das Modell mitgeschickt, auf dessen Grundlage die Alternativauswahl stattfindet.

## 2.3 Verwandte Arbeiten

Auf dem Gebiet des *Public Sensing* gibt es bereits verschiedene Arbeiten. In [Eis08] wird *MetroSense* vorgestellt. Dabei handelt es sich um ein opportunistisches System, das die Benutzung von mobilen und statischen Sensoren vereint. Dabei wird besonders auf die Skalierbarkeit des Systems Wert gelegt.

*MobGeoSen* [KBP<sup>+</sup>08] benutzt ebenfalls Smartphones zur Datenerfassung. Über einen Datenlogger können weitere Sensortypen angeschlossen werden, die das Smartphone selbst nicht besitzt. In regelmäßigen Abständen werden die aktuellen Werte der Sensoren mit der aktuellen GPS-Position gespeichert. Die gesammelten Daten können anschließend auf einer Karte graphisch dargestellt werden. Zudem kann der Benutzer Bilder aufnehmen, um den Kontext der Daten zu veranschaulichen. Es werden nur die Werte eines einzelnen Geräts betrachtet. Eine räumliche Verteilung entsteht nur für den Weg, auf dem sich das Gerät während der Erfassung bewegt hat.

In [HBZ<sup>+</sup>06] wird *CarTel* vorgestellt, ein System, das opportunistisch verkehrsbezogene Daten erfasst und dazu auf Sensoren in Autos zugreift. Die Daten werden während der Fahrt gesammelt und an ein Portal gesendet, wenn eine Internetverbindung hergestellt werden kann. Dort können die Daten gefiltert oder visualisiert werden.

Mit der Erfassung von Umgebungsphänomenen befasst sich [MSN<sup>+</sup>09] in Form von Lärmverschmutzung. Hier wird durch die Teilnehmer die Aussagekraft der reinen Messwerte erhöht, da diese den Lärm genauer präzisieren können, indem sie beispielsweise die Geräuschquelle zu einem Ort bekannt geben.

In anderen Arbeiten wurden auch bereits multivariate Gaußverteilungen zur modellgetriebenen Erfassung eingesetzt, zum Beispiel in [DGM<sup>+</sup>04]. Dadurch wird es möglich, Werte an nicht erfassten Stellen zu berechnen. Allerdings werden hier nur statische Sensorsysteme behandelt.

In [KSG08] wird mit *Saturate* ein Algorithmus vorgestellt, mit dem sich effizient eine bestmögliche Teilmenge der Sensoren bestimmen lässt, von denen tatsächlich Werte abgefragt werden. Allerdings wird auch hier nur auf statische Sensornetzwerke eingegangen.

## 3 Systemmodell und Anforderungen

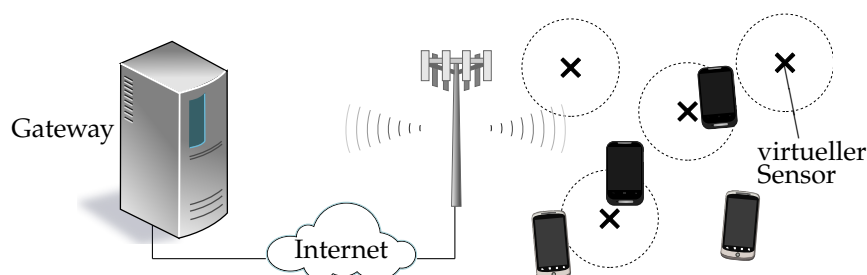
In diesem Kapitel soll nun zunächst das Systemmodell vorgestellt werden, um danach die Anforderungen, für die diese Arbeit Lösungen präsentiert, vorzustellen.

### 3.1 Systemmodell

Das Public-Sensing-System orientiert sich am Aufbau in [PSDR<sub>12</sub>] und besteht aus einem Gateway sowie mehreren mobilen Geräten wie zum Beispiel Smartphones. In Abbildung 3.1 ist dieser Aufbau dargestellt.

Das Gateway ist für ein bestimmtes Gebiet zuständig und nimmt für dieses Anfragen des Benutzers entgegen und es verwaltet und bearbeitet diese, indem es entsprechende Anfragen an das Smartphone-Netzwerk weitergibt. Die ankommenden Antworten werden gespeichert und dem Benutzer zur Verfügung gestellt.

Die Smartphones sind über das gesamte Messgebiet verteilt und können sich frei bewegen. Sie sind mit verschiedenen Sensoren sowie einem Positionierungssystem, wie zum Beispiel GPS ausgestattet. Auf diese Weise können sie sich selbst lokalisieren und Sensorwerte wie die Temperatur oder den Geräuschpegel aufnehmen. Zudem verfügen sie über eine Möglichkeit, diese Messwerte dem Gateway mitzuteilen sowie untereinander mit benachbarten Geräten zu kommunizieren. Durch geographisches Routing leiten sie die Anfragen des Gateways in Richtung der ausgewählten virtuellen Sensoren. Da Smartphones mobil sind, kann nicht vorherbestimmt werden, an welcher Position sie sich befinden und es werden darüber keine Annahmen getroffen. Die Messwerterfassung findet dann an den virtuellen Sensoren statt, wie es in den Grundlagen bereits erklärt wurde.



**Abbildung 3.1:** Schematische Darstellung des Systemmodells, das aus einem Gateway und mobilen Geräten besteht.

Der Benutzer eines Public-Sensing-Systems stellt Anfragen über einen oder mehrere Sensortypen, die für eine bestimmte Menge in einem bestimmten Gebiet bestimmten Gebiet erfasst werden sollen und legt fest, wie weit durch das Modell berechnete Messwerte von der Realität abweichen dürfen. Das Gateway bestimmt in diesem Gebiet virtuelle Sensoren und entscheidet, an welchen ein Wert erfasst werden soll. Häufig soll eine Erfassung über einen längeren Zeitraum stattfinden, sodass periodisch Messwerte eingeholt werden müssen. Dies übernimmt ebenfalls das Gateway, indem es zu den entsprechenden Zeitpunkten Anfragen sendet. Eine Anfrage an die mobilen Geräte enthält die vom Gateway ausgewählten virtuellen Sensoren sowie Informationen über deren bisherige Verfügbarkeit und deren Rollen, also ob sie als zur Berechnung der übrigen Sensoren benutzt werden sollen oder als Kontrollsensoren.

## 3.2 Anforderungen

Auch wenn das *PublicSensing* Vereinfachungen im Bereich verteilter Sensorsystem mit sich bringt, wie eine einfachere Inbetriebnahme und Verteilung von Messwerten oder das Zurückgreifen auf bereits vorhandene Ressourcen in Form von Smartphones, birgt es weiterhin Herausforderungen. Im Zuge dieser Arbeit werden Lösungen für diese Probleme vorgestellt.

**Ausgefallene Sensoren kompensieren** Ein Problem ergibt sich aus der unberechenbaren Mobilität der mobilen Geräte. Da nicht vorherbestimmt werden kann, ob sich bei einem virtuellen Sensor zu einem bestimmten Zeitpunkt ein Gerät befinden wird, kann es dazu kommen, dass ein virtueller Sensor abgefragt wird, in dessen Nähe sich gerade kein Gerät befindet, das in der Lage ist, den angefragten Sensorwert aufzunehmen. An diesem virtuellen Sensor kann kein Messwert erfasst werde; auf ein herkömmliches Sensornetz übertragen entspricht dies einem ausgefallenen Sensor. Man kann nur unter sehr günstigen Bedingungen davon ausgehen, an allen gewünschten Messpunkten einen tatsächlich gemessenen Wert zu erhalten. Selbst wenn man davon ausgeht, eine sehr hohe Dichte an mobilen Geräten im Erfassungsgebiet zu haben, erhöht dies die Zuverlässigkeit nur bedingt, da durch die alleinige Anwesenheit noch keine Messwernerfassung gewährleistet ist. Das Gerät könnte gerade benutzt werden oder sich in der Hosentasche befinden, was je nach Sensortyp das Messergebnis verfälschen kann und daher von der Erfassung ausgeschlossen werden muss.

Man muss also auf ausfallende Sensoren reagieren und den Ausfall nach Möglichkeit kompensieren. In dieser Arbeit wird gezeigt, wie dies durch Abfragen von Ersatzsensoren möglich ist, und wie man diese neuen Sensoren auswählt. Das Ergebnis soll dabei möglichst dicht an die angestrebte Qualität herankommen. Voraussetzung für die Neuauswahl von Sensoren ist das zugrunde liegende Modell, das dafür in jeder Anfrage mitgeschickt werden muss.

**Aufwand reduzieren** Das Modell wächst stark mit der Menge der betrachteten Sensoren, da im Modell jeder Sensor und die Korrelation untereinander abgebildet werden



muss. Diese Korrelation ist im Modell in einer zweidimensionalen Kovarianzmatrix abgebildet, sodass deren Größe mit steigender Sensorzahl quadratisch wächst. Daher erfordert die Verwendung des Modells zur Alternativplanung besonders bei großen Sensornetzwerken einen erheblichen Nachrichtenaufwand.

Aus diesem Grund soll ebenfalls eine Möglichkeit dargestellt werden, mit der dieser Mehraufwand, der durch das Mitsenden des Modells zustande kommt, reduziert werden kann, sodass die Alternativplanung ohne erheblichen Mehraufwand möglich wird. In dieser Arbeit soll zu diesem Zweck die Strategie der Modellpartitionierung vorgestellt werden. Der Gedanke liegt darin, das Messgebiet in mehrere Teilgebiete zu unterteilen und die Auswahl alternativer Sensoren nur für das jeweilige Teilgebiet zu treffen. Dadurch wird zudem gewährleistet, dass eine Alternativenanfrage nur innerhalb des Teilgebiets weitergeleitet werden muss. Das Modell wird auf die Teilgebiete beschränkt, wodurch der Nachrichtenaufwand verringert wird. In dieser Bachelorarbeit soll dafür zunächst gezeigt werden, wie man ein Modell aufteilen kann und wie sich damit eine Alternativplanung mit annehmbarem Aufwand realisieren lässt.



## 4 Entwurf

In diesem Kapitel wird nun vorgestellt, wie man zuverlässiger genügend Messwerte erfassen kann, ohne den Übertragungsaufwand übermäßig zu erhöhen. Dazu wird zunächst dargestellt, wie eine Anfrage aufgebaut wird und wie diese bei nicht Erreichen einen virtuellen Sensors zur Alternativplanung benutzt werden kann. Anschließend wird genauer darauf eingegangen, auf welche Weise man eine alternative Auswahl an Sensoren treffen kann. In einem nächsten Schritt wird der Einsatz der Modellpartitionierung und deren Erweiterung mit Zusatzwissen über die anderen Gebiete vorgestellt, was am Ende mit der Alternativplanung verknüpft werden soll, um so deren vergrößerten Nachrichtenaufwand zu kompensieren.

### 4.1 Aufbau und Ausführung einer Anfrage

Das Gateway verwaltet die Aufträge der Benutzer. Ein Auftrag kann aus einer einmaligen oder aus einer periodisch stattfindenden Messwerterfassung für eine bestimmte Menge virtueller Sensoren bestehen. Tritt für einen Auftrag der Zeitpunkt für eine Messwerterfassung ein, so generiert das Gateway eine entsprechende Anfrage an die mobilen Geräte.

Eine Anfrage beinhaltet in erster Linie die Menge der virtuellen Sensoren, von denen Werte erwartet werden. Diese Menge wird für jeden Abfragezeitpunkt so ausgewählt, dass das erwartete Ergebnis bei Erhalt aller angefragten Werte in einem festgelegten Toleranzbereich liegt, das bedeutet unterhalb einer bestimmten Varianzgrenze bleibt. Das Ergebnis ergibt sich aus den tatsächlich gemessenen Werten sowie wie den daraus aus dem Modell abgeleiteten Werten für die anderen Sensoren. Da in der Regel nicht an allen Stellen Werte aufgenommen werden, ergibt sich für die Gesamtverteilung immer eine abschätzbare Ungenauigkeit durch die Varianz, die bei Verwendung der Ergebnisse beachtet werden muss. Die Auswahl der abzufragenden Sensoren für einen Zeitpunkt erfolgt über einen Greedy-Algorithmus [KSGo8], der solange weitere Sensoren hinzunimmt, bis damit die gewünschte Varianz erreicht werden kann.

Neben dieser Auswahl an Sensoren für die Berechnung werden noch zusätzlich Kontrollsensoren ausgewählt, mit deren Werten man dann das Ergebnis dieser Berechnung vergleichen und beurteilen kann, wie gut das Modell noch an die Realität angepasst ist. Weichen die gemessenen Werte der Kontrollsensoren von den berechneten ab, so muss das Modell angepasst werden. Einzelne Abweichungen werden dabei noch toleriert, da diese auch durch einen fehlerhaften Messwert hervorgerufen werden können. Um das Modell während des Ablaufs anpassen zu können, wechseln sich sogenannte Lern- und Optimierungsphasen ab. In der Lernphase werden alle Sensoren abgefragt um aus den Werten ein neues Modell

erstellen zu können. Wenn noch kein Modell bekannt ist, muss mit einer Lernphase gestartet werden. Ist ein Modell erstellt, wird in die Optimierungsphase gewechselt. In dieser Phase kann dann das Modell eingesetzt werden, um nur einen Teil der Sensoren abzufragen und die anderen aus dem Modell abzuleiten. Es wird daher angestrebt, dass der Anteil der Optimierungsphasen möglichst groß ist, da nur in dieser Zeit die Vorteile der modellgetriebenen Erfassung genutzt werden. Wird nun während der Optimierungsphase durch die Kontrollsensoren festgestellt, dass das Modell nicht mehr passend ist, setzt eine neue Lernphase ein.

Das Gateway schickt nun die Anfrage über die zu erfassenden Sensoren an ein Gerät im Messgebiet, welches die Anfrage geographisch in Richtung der virtuellen Sensoren weiterleitet. Ist ein Gerät in der Lage, einen erfordernten Wert aufzunehmen, so schickt es diesen an das Gateway. Durch die Mobilität der Geräte können angefragte Werte ausfallen, falls der entsprechenden Sensor nicht verfügbar ist.

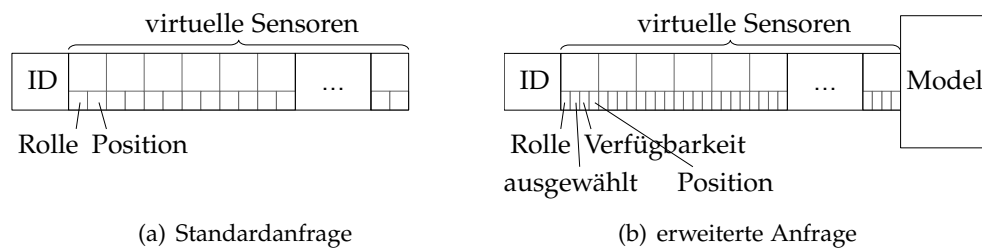
### 4.1.1 Sensorrollen

Um das Wissen über die bisherige Verfügbarkeit der Sensoren nutzen zu können, wird jeder Sensor markiert, der sich als nicht verfügbar herausgestellt hat. Zudem werden den Sensoren Rollen zugeteilt, nach denen sie zur Zeit verwendet werden. Dabei werden vier verschiedenen Typen unterschieden. In diese Rollen werden die Sensoren bei der Auswahl durch das Gateway eingeteilt.

- **unbenutzte Sensoren:** Sensoren, von denen keine Werte direkt abgefragt werden, weil sie nicht verfügbar sind oder für die Auswahl nicht benötigt werden.
- **ausgewählte Sensoren:** Sensoren, von denen Werte erfragt werden, um sie in das Modell einzusetzen und eine Werteverteilung für das Messgebiet zu berechnen.
- **Kontrollsensoren:** Sensoren, von denen Werte erfragt werden, aber nicht in die Berechnung durch das Modell einfließen. Für diese Sensoren werden zusätzlich genau wie für unbenutzte Sensoren Werte berechnet, um diese mit den tatsächlichen Werten zu vergleichen. Dadurch kann die Abweichung des Modells von der Realität abgeschätzt werden.
- **nicht modellierte Sensoren:** Sensoren, die zwar bekannt sind, aber nicht verwendet werden und auch keinen Eintrag in der Kovarianzmatrix besitzen.

Bei einer Anfrage wird versucht, von allen ausgewählten Sensoren und den Kontrollsensoren Messwerte zu erhalten. Stellt sich ein Sensor dabei als nicht verfügbar heraus, da von ihm kein Wert geliefert wird, so wird dieser fortan den unbenutzten Sensoren zugeordnet und als nicht verfügbar markiert.

Der Aufbau einer Anfrage des Gateway an die mobilen Konten ist in Abbildung 4.1(a) dargestellt. Sie beinhaltet alle Sensoren, von denen Werte angefordert werden. Zu jedem Sensor ist zudem dessen aktuelle Rolle sowie dessen Position gespeichert. Aufgenommene



**Abbildung 4.1:** Nachrichtenformat für Standardanfragen (a) und erweiterte Anfragen mit der Möglichkeit zur Partitionierung und zur Alternativbehandlung (b).

Werte werden zusammen mit diesen Informationen in einer Antwort an das Gateway gesendet.

## 4.2 Alternativplanung

Wie bereits geschildert kann es vorkommen, dass nicht an allen festgelegten Sensoren ein Wert erfasst werden kann. Dies ist der Fall, wenn sich in dem von einem Sensor überwachten Gebiet kein mobiles Gerät befindet, das einen entsprechenden Wert liefern kann. Das kann verschiedene Ursachen haben, zum Beispiel, weil sich zur Zeit einfach kein Smartphone innerhalb des Sensorgebiets befindet oder weil ein eventuell vorhandenes keine sinnvollen Werte liefern kann, beispielsweise wenn der Benutzer es gerade in der Hosentasche trägt. Man kann nun nur die übrigen Werte in das Modell einsetzen, um ein Ergebnis zu bekommen. Allerdings erhält man so ein Ergebnis, welches nicht die geforderte Varianz einhält.

Eine Möglichkeit ist nun, eine Art zweite Chance zu benutzen und zusätzliche Sensoren anzufragen, die dann einen Wert liefern können, sodass das Ergebnis mit diesen neuen Werten wieder im Rahmen der Toleranz liegt. Natürlich können ebenso hierbei Werte ausfallen. Man kann es in diesem Fall wiederum eine Auswahl neuer Werte treffen und diese anfragen. Wie oft man eine weitere Alternativauswahl durchführen will, wenn weiterhin nicht alle Sensoren verfügbar sind, kann man nach verschiedenen Strategien entscheiden.

Ist ein Sensor nicht verfügbar, so wird dies auf der Ebene der mobilen Geräte dadurch festgestellt, dass es nicht möglich ist, die Anfrage an ein Gerät innerhalb des Erfassungsgebiets eines der angefragten Sensoren zu senden, das den geforderten Messwert aufnehmen kann. In diesem Fall kann durch die mobilen Knoten eine Alternativplanung stattfinden, sodass man den langen Weg über das Gateway spart.

### 4.2.1 Ablauf der Alternativplanung

Damit eine Alternativplanung auf der Ebene der mobilen Geräte überhaupt möglich ist, sind weitere Informationen und damit ein erweitertes Nachrichtenformat nötig, das in Abbildung 4.1(b) schematisch dargestellt ist. Hier sind nicht nur die ausgewählten Sensoren enthalten,

sondern sämtliche Sensoren des Gebiets sowie ein Modell, das alle Sensoren abbildet. Ohne das Modell ist es nämlich nicht möglich eine Auswahl zu treffen, da man die Auswirkung eines Sensors auf die Varianz nicht abschätzen kann. Um nun die angeforderten von den anderen Sensoren zu unterscheiden, muss dieser Zustand zu jedem Sensor angefügt sein. Zudem ist es notwendig, die bisherige Verfügbarkeit eines Sensors bei der Alternativauswahl mit einzubeziehen, um nicht Sensoren zu wählen, die bekannterweise nicht verfügbar sind. Die aktuelle Verfügbarkeit wird ebenfalls zu jedem Sensor mitgeliefert. Wie schon in der Grundform enthält die Nachricht auch zu jedem Sensor Informationen über dessen Rolle und Position.

Der Ablauf der Alternativbehandlung sieht wie folgt aus. Bei einer Anfrage liefert nur ein Teil der Sensorauswahl tatsächlich Messwerte. Unter den bisher angefragten Sensoren müssen nun die Kategorien *Messwert gelesen*, *Kontrollwert gelesen* und *kein Wert bekommen* unterschieden werden. Diese Informationen sind für eine Neuauswahl nötig. Die Sensoren, an denen keine Wert ausgelesen werden konnte, müssen ab sofort als nicht verfügbar eingestuft werden, damit sie von einer erneuten Auswahl ausgeschlossen sind. Für die Sensoren, von denen bereits Werte vorhanden sind, muss sichergestellt werden, dass sie zur alternativen Auswahl gehören, damit nicht unnötig viele Werte neu angefragt werden. So werden nur die nicht verfügbaren durch andere Sensoren ersetzt.

Um nun eine Neuauswahl zu treffen, wählt man zunächst alle Sensoren aus, von denen man Werte bekommen hat. Anschließend nimmt man wie beim Erstellen der Anfrage solange neue verfügbare Sensoren hinzu, bis das zu erwartete Ergebnis im Toleranzbereich liegt. Aus dieser Auswahl fragt man die Sensoren ab, für die noch kein Wert vorliegt. Da die Sensoren für die man Kontrollwerte aufgenommen hat, bei der Auswahl vorläufig wie bisher nicht ausgewählte Sensoren behandelt werden, können auch diese in die Neuauswahl aufgenommen werden. Es muss dann kein neuer Wert gelesen werden, sondern einfach der bereits vorhandene umdeklariert werden. Als Ersatz können eventuell zusätzliche Kontrollsensoren ausgewählt werden.

### 4.2.2 Begrenzung der Neuauswahl

Ein Sensor, der als Alternative zu einem nicht verfügbaren gewählt wurde, kann ebenfalls nicht verfügbar sein. Wählt man auch für diesen einen alternativen Sensor, kann dieser wiederum nicht verfügbar sein. Es ist nicht klar, ob es überhaupt einen geeigneten Sensor gibt. Um den Aufwand der Alternativplanung zu begrenzen ist es sinnvoll, ein Abbruchkriterium für diese Rekursion festzulegen. Eine Möglichkeit ist die Wahl eines Limits für die Anzahl der Neuauswahlen. Wurden bereits eine festgelegte Anzahl an Wiederholungen durchgeführt und es gibt weiterhin Sensoren in der aktuellen Auswahl, die nicht verfügbar sind, bricht man die Alternativplanung ab. Das hat den Vorteil, dass man durch dieses Limit den Aufwand, den man für die Alternativplanung verwenden will, gut regulieren will. Soll nur wenig Energie auf die Alternativplanung verwendet werden, wählt man ein geringeres Limit. Durch diese Strategie hat man jedoch nur wenig Einfluss auf das Erhalten einer guten Menge tatsächlich gemessener Werte. Zwar sind die Chancen höher, bei einem größeren Limit näher an die Toleranzgrenze heranzukommen, jedoch kann es nicht garantiert werden, dass

bei einer möglichen vorhandenen Menge an verfügbaren Sensoren diese immer gefunden wird.

Eine weitere Möglichkeit für ein Abbruchkriterium ist, die Anzahl der Sensoren die ausfallen können zu begrenzen, also auf das Erhalten einer bestimmten Anzahl an erhaltenen Werten zu warten. Ist diese Zahl noch nicht erreicht und es sind noch Sensoren vorhanden, von denen noch nicht bekannt ist, dass sie nicht verfügbar sind, wird die Alternativauswahl fortgesetzt. Der Vorteil dieser Strategie besteht darin, dass man Güte der Erfassungen besser regulieren kann, da wenn möglich eine bestimmte Zahl an Messwerten garantiert wird. Man hat dabei allerdings nicht mehr unter Kontrolle, wie viel Aufwand in die Alternativplanung gesteckt wird. Das Abbruchkriterium kann schon nach wenigen Neuauswahlen eintreten oder im schlechtesten Fall werden so lange Neuauswahlen getroffen, bis alle Sensoren einmal Teil einer Auswahl waren, aber noch immer nicht von genügend virtuellen Sensoren Werte erhalten werden konnten.

Alternativ kann man auch eine weitere Toleranzgrenze einführen, die die erhaltenen Werte erfüllen müssen. Analog zur vorigen Strategie kann dann abgebrochen werden, wenn dieser Wert erreicht ist oder bis alle Sensoren einmal ausgewählt waren oder zumindest als nicht verfügbar bekannt sind. Auch bei diesem Ansatz kann man wie beim vorigen die Güte der Erfassung regulieren, hat aber keinen Einfluss darauf, wie oft die Alternativauswahl durchgeführt wird.

Denkbar ist auch, Strategien zur Begrenzung der Wiederholung mit Strategien zum Erreichen einer bestimmten Güte zu kombinieren und auf den frühesten Abbruch zu warten. Je nachdem, ob man mehr auf das Einsparen an Aufwand oder auf die Qualität des Ergebnisses achten möchte, kann man durch anpassen der jeweiligen Parameter die eine oder die andere Strategie stärker gewichten.

### 4.3 Modellpartitionierung

Durch Alternativplanung erhält man zwar mehr Messwerte und kann damit das Ergebnis der Erfassung verbessern, allerdings stellt es auch eine Steigerung des Nachrichtenaufwands dar. Den größten Anteil hat dabei, dass das Modell in jeder Anfrage mitgeschickt werden muss. Um diesen Aufwand zu verringern, muss die Größe des Modells in einer Anfrage reduziert werden. Um das Modell zu verkleinern, muss man es auf einen Teil der Sensoren reduzieren. Beschränkt man beispielsweise das Modell auf die Hälfte der Sensoren, so halbiert sich die Größe des Mittelwertvektor; da die Kovarianzmatrix zweidimensional ist, wird deren Größe sogar geviertelt. Die Kovarianzmatrix dominiert die Größe des Modells, daher wird beim Halbieren der Sensoranzahl die Größe des gesamten Modells annähernd geviertelt.

Um nun verkleinerte Modelle benutzen zu können, partitioniert man das Modell, das heißt man verteilt die Sensoren auf mehrere disjunkte Teilmengen und reduziert das Modell jeweils auf diese Teilmengen um so mehrere Teilmodelle zu erhalten. Die initiale Auswahl der Sensoren läuft dabei weiterhin gleich ab wie ohne die Partitionierung, sodass die Sensoren erst anschließend aufgeteilt werden. Wo bisher eine einzelne Anfrage gesendet

wurde, muss nun für jede der Sensorteilmengen eine Anfrage mit dem entsprechenden Teilmodell gesendet werden. Dadurch erhöht sich bei  $n$  Partitionen auch die Anzahl der Anfragen auf  $n$ . Jede Anfrage enthält jedoch nur noch  $\frac{1}{n^2}$  des ursprünglichen Modells, sodass insgesamt nur noch  $\frac{1}{n}$  des Modells übertragen wird.

Im Vergleich zum ursprünglichen Modell werden die Korrelationen zwischen Sensoren, die in verschiedenen Partitionen liegen, nicht übertragen. Dadurch können bei einer Alternativauswahl auch nur Sensoren aus der selben Partition ausgewählt werden, da nur diese bekannt sind. Aus diesem Grund ist es sinnvoll, die Sensoren räumlich aufzuteilen, das heißt das gesamte Messgebiet in Teilgebiete zu unterteilen und die Sensoren entsprechend in Untermengen aufzuteilen. Ein alternativer Sensor liegt so immer in der Nähe des ursprünglichen Sensors, was für das Routing von Vorteil ist. Die Alternativenanfrage muss über weniger Knoten geroutet werden. Zudem haben benachbarte Sensoren häufig ein ähnliches Umfeld. Es ist allerdings ebenso möglich, dass zwei Sensoren stark korreliert sind, obwohl sie sich an entfernten Punkten im Messgebiet befinden, aber ein ähnliches Umfeld besitzen. Solche Sensoren als Alternativen heranzuziehen wäre zwar für das Ergebnis von Vorteil, allerdings wird dieser Vorteil durch den vergrößerten Routingaufwand wieder kompensiert. Aus diesem Grund werden die Sensoren nach ihrer geographischen Lage aufgeteilt.

Eine einfache Art die Partitionierung für eine Anfrage festzulegen ist, das ursprüngliche Gebiet in gleich große Untergebiete zu teilen. Für jeden Sensor muss nun festgestellt werden, in welchem der entstandenen Teilgebiete er liegt, um ihn in die entsprechende Untermenge aufzunehmen. Nach diesen Untermengen wird anschließend das Modell partitioniert, indem die Kovarianzmatrix und der Mittelvektor des gesamten Modells jeweils auf die Einträge für die Sensoren im jeweiligen Untergebiet reduziert werden. Mit einer Untermenge der virtuellen Sensoren und dem zugehörigen partitionierten Modell erzeugt das Gateway eine neue Anfrage und schickt diese an ein Gerät im entsprechenden Teilgebiet.

### 4.3.1 Modellpartitionierung mit Zusatzwissen

Man hat also durch die Partitionierung den Vorteil, dass man ein wesentlich kleineres Modell übertragen muss, um bei der Alternativplanung ähnliche Sensoren wählen zu können. Verwendet man dazu allerdings die selben Kriterien wie im ungeteilten Fall kann dies möglicherweise zu Problemen führen. Jedes Untergebiet wird nun versuchen, mit seinen wenigen Sensoren die vorgegebene Varianz zu erreichen ohne zu berücksichtigen, dass aus anderen Teilgebieten ebenfalls Messergebnisse zu Verfügung stehen, die die Varianz ebenfalls reduzieren. Es ist daher davon auszugehen, dass jedes Gebiet zu viele Sensoren auswählt.

Um zu verhindern, dass in einem Teilgebiet bei der Alternativplanung zu viele Sensoren ausgewählt werden, kann man durch die Anfrage jedes Gebiet mit zusätzlichem Wissen über die anderen Teilgebiete ausstatten. Kennt jede Partition die Sensoren, die in den anderen Gebiete abgefragt werden sollten, kann sie davon ausgehen, dass von diesen oder von gleichwertigen Sensoren Werte erhalten werden. Sollte ein Sensor in einem anderen Teilgebiet nicht verfügbar sein, so findet dort ebenfalls eine Alternativplanung statt. Jedes Teilgebiet verlässt sich bei seiner Alternativplanung darauf, dass die Erfassung in den



anderen Gebieten die geforderte Qualität liefert. Damit die angefragten Daten der anderen Teilgebiete berücksichtigt werden können, kann die Anfrage allerdings nicht mehr so effizient aufgeteilt werden. Zusätzlich zu den eigenen Sensoren muss eine Anfrage an ein Gebiet so noch die ausgewählten Sensoren der anderen Gebiete umfassen. Das bedeutet, dass auch das Modell umfangreicher ist als ohne das Extrawissen. Dadurch dass sich sämtliche ausgewählten Sensoren in allen Partitionen befinden, sind die Anfragenachrichten nicht mehr disjunkt.

Auch wenn man so nicht mehr eine vollständige Partitionierung erreicht, bringt dieser Ansatz Vorteile mit sich. Im Vergleich zum Standardansatz ermöglicht er eine Alternativplanung, wodurch sich das Ergebnis einer Erfassung auf mehr tatsächliche Messwerte stützen kann. Im Vergleich zur Alternativplanung auf Basis des gesamten Modells ist das mitzuschickende Modell deutlich kleiner, auch wenn in jedes Teilgebiet eine Anfragenachricht gesendet werden muss. Zudem wird der Auswahl zu vieler Sensoren, wie sie bei der einfachen Partitionierung auftritt, entgegengewirkt.

### 4.3.2 Modellpartitionierung und Alternativplanung

Wurde das Gebiet und mit ihm das Modell partitioniert, enthält eine Anfrage nur noch Informationen zum jeweiligen Teilgebiet. Die Alternativplanung muss also nicht weiter angepasst werden, um die Auswahl neuer Sensoren auf dieses Teilgebiet zu begrenzen.

Um das Zusatzwissen über die anderen Teilgebiete miteinzubeziehen bedarf es allerdings einiger wenigen Änderungen. Die Sensoren, von denen bekannt ist, dass sie in den anderen Gebieten angefragt werden, werden in der Alternativplanung des eigenen Teilgebiets behandelt, als wären für diese bereits Messwerte vorhanden. So wird berücksichtigt, dass auch aus den anderen Teilgebieten Messwerte zur Verfügung stehen werden. Gleichzeitig wird verhindert, dass sie aus dem falschen Gebiet angefragt werden, um dessen ausgefallene Sensoren zu kompensieren.



## 5 Implementierung

In einer Simulationsumgebung sollen die Möglichkeiten der Alternativplanung und der Partitionierung untersucht werden. In dieser Vereinfachung sind keine Smartphones nötig, sondern man erhält durch das Auslesen eines Sensors Werte aus einer eingespeicherten Messreihe. Dadurch ist die Simulation wesentlich einfacher durchzuführen.

Durch den Verzicht auf tatsächliche oder simulierte mobile Geräte wird auch die Kommunikation unter den Geräten nicht simuliert. Jeder virtuelle Sensor kann direkt abgefragt werden indem der entsprechende Eintrag aus der Messreihe ausgelesen wird. Zudem kann die Alternativplanung nur global ausgeführt werden. Damit unterscheidet sich die Implementierung in einigen Punkten von den Annahmen im Entwurf. Diese Implementierung stellt jedoch eine einfache Art dar, zu untersuchen, welche Möglichkeiten die Modellpartitionierung im Zusammenspiel mit der Alternativplanung hinsichtlich der Qualität der Daten und dem Aufwand hat.

Zur Bearbeitung einer Standardanfrage wird für jeden der angefragten Sensoren der Wert aus dem Datensatz abgerufen. Ist kein Wert vorhanden, stellt dies einen nicht verfügbaren Sensor dar. Ein erhaltener Wert wird mit der Position und der Rolle des Sensors als Antwort an das Gateway geschickt.

### 5.1 Alternativplanung

Um eine Alternativauswahl durchführen zu können, muss zunächst beim Abfragen der Sensoren erfasst werden, von welchen Sensoren man einen Messwert erhält und in welcher Rolle (ausgewählter Sensor oder Kontrollsensor). Ist mindestens einer der angefragten Sensoren nicht verfügbar setzt die Alternativplanung ein. Die Verfügbarkeit aller Sensoren kann aus der Anfrage ausgelesen werden, sodass lediglich der Zustand der Sensoren, die bei der aktuellen Abfrage keinen Wert geliefert haben, als nicht verfügbar aktualisiert werden muss.

Zur eigentlichen Alternativauswahl werden nun zu den bereits erfolgreich abgefragten Sensoren weitere ausgewählt, die noch verfügbar sind, um erneut eine Auswahl zu erzeugen, die die vorgegebene Varianz erfüllen kann. Wurden von zu wenigen Kontrollsensoren Werte erhalten, werden zu diesem Zweck ebenfalls zusätzliche Sensoren ausgewählt. Die zusätzlichen ausgewählten Sensoren werden ebenfalls abgefragt und die Werte im Erfolgsfall mit der Position und der entsprechenden Rolle als Antwort an das Gateway geschickt.

Es wird darauf verzichtet, die Alternativplanung häufiger als einmal auszuführen, da davon auszugehen ist, dass jede weitere Alternativplanung weniger zusätzliche Sensorwerte Messwerte beisteuern kann. Angenommen die Wahrscheinlichkeit der Verfügbarkeit für jeden Sensor ist gleich groß und jeder noch freie Sensor wird mit gleicher Wahrscheinlichkeit in die Alternativauswahl aufgenommen. Sei die Wahrscheinlichkeit für die tatsächliche Verfügbarkeit eines Sensors  $p$  und die Anzahl der neu ausgewählten Sensoren sei  $m$ . Dann werden im Schnitt von  $p \cdot m$  Sensoren tatsächlich Werte ausgelesen. Da die Anzahl der zur Berechnung verwendeten Sensoren begrenzt ist, wird  $m$  mit jeder weiteren Alternativplanung kleiner und somit auch die Zahl der neuen Messwerte.

### 5.2 Partitionierung des Modells

Für die Partitionierung kann das Messgebiet sowohl in  $x$ - als auch in  $y$ -Richtung aufgeteilt werden. Dazu wird aus den Positionen der Sensoren die Grenzen des abgedeckten Gebiets ermittelt, um dieses anschließend in  $x$ - und  $y$ -Richtung in  $k$  bzw.  $l$  Teile zu gliedern, sodass  $k \cdot l$  Untergebiete entstehen.

#### 5.2.1 Partitionierung ohne Zusatzwissen

---

##### Algorithmus 5.1: Modellpartitionierung

---

**Require:** Modell  $(\Sigma, \mu)$ , Sensorliste  $V$ , SensorIndexMapper

**Ensure:** geteiltes Modell

```
keepvector  $\leftarrow$  false
for all  $v \in V$  do
     $i \leftarrow v.getIndex()$ 
    keepvector[ $i$ ]  $\leftarrow$  true
end for
for  $j = |V| - 1$  to 0 do
    if keepvector[ $j$ ] == false then
         $\mu.entferneEintrag(j)$ 
         $\Sigma.entferneZeile(j)$ 
         $\Sigma.entferneSpalte(j)$ 
         $SensorIndexMapper.entferneSensor(j)$ 
    end if
end for
```

---

Die Sensoren sind durch einen Sensor-Index-Mapper an das Modell gebunden. Dieser ordnet jedem Sensor die Indizes der Einträge im Mittelvektor und in der Kovarianzmatrix zu. Der Algorithmus zur Modellpartitionierung ist in 5.1 skizziert. Er erhält eine Menge  $V$  von Sensoren und reduziert den Mittelwertvektor, die Kovarianzmatrix und den Sensor-Index-Mapper auf diese Sensoren. Am Ende ist das ursprüngliche Modell auf die eingegebenen

Sensoren reduziert. Zunächst müssen die zu den Sensoren gehörenden Indizes festgestellt werden. In einem Vektor, der hier *keepvector* genannt wird, werden genau die Einträge an den Indizes auf *true* gesetzt, für die sich der Sensor in  $V$  befindet. Dieser Vektor wird später benutzt, um danach die Reduktion des Mittelwertvektors und der Kovarianzmatrix vorzunehmen. In diesen werden alle Einträge entfernt, an denen im *keepvector* *false* steht. Ebenso wird der Sensor-Index-Mapper auf den aktuellen Stand gebracht. Dazu müssen hier ebenfalls die Einträge zu den überflüssigen Sensoren gelöscht werden. Beim Löschen eines Sensor-Index-Paares werden die folgenden Indizes dekrementiert, sodass kein Sprung in der Nummerierung entsteht. Beim Löschen der Paare muss daher auf die Reihenfolge geachtet werden. Wenn man das Paar mit dem höchsten Index zuerst löscht, ändern sich die Indizes der anderen zu löschenden Paare nicht, da nur Indizes angepasst werden, die nicht gelöscht werden müssen.

### 5.2.2 Partitionierung mit Zusatzwissen

Da jedes Untergebiet keine Informationen über die Sensorauswahl der anderen Untergebiete hat, wird es bei der Alternativplanung versuchen, mit den zur Verfügung stehenden Sensoren die für das Gesamtgebiet geforderte Qualität einzuhalten. Zusammengenommen übersteigt die Qualität aus allen Untergebieten dann jedoch voraussichtlich die geforderte Qualität stark, was bedeutet, dass eine kleinere Sensorauswahl den Ansprüchen genügt hätte. Die Untergebiete wählen also mehr Sensoren aus als nötig sind, was den Nachrichtenaufwand erhöht. Um die Partitionierung mit Zusatzwissen über die anderen Teilgebiete zu realisieren, dürfen die Anfragen nicht nur auf die Sensoren des jeweiligen Gebiets beschränkt werden. Die Sensoren, die in anderen Gebieten angefragt werden sollen, müssen ebenfalls aufgenommen werden. Um diese zu erkennen, haben sie zwar die entsprechende Rolle ausgewählte Sensoren oder Kontrollsensoren, gehören aber nicht zu den abzufragenden Sensoren. Bei der Alternativplanung lassen sie sich dadurch erkennen, dass sie die einzigen Sensoren sind, die nicht angefragt werden aber dennoch in der Rolle ausgewählter Sensor oder Kontrollsensor auftauchen. Für die alternative Auswahl müssen sie wie bereits abgefragte Sensoren behandelt werden.



# 6 Evaluation

In diesem Kapitel werden die Ergebnisse aus einer Simulation des in dieser Arbeit vorgestellten Entwurfs dargestellt.

## 6.1 Methodologie

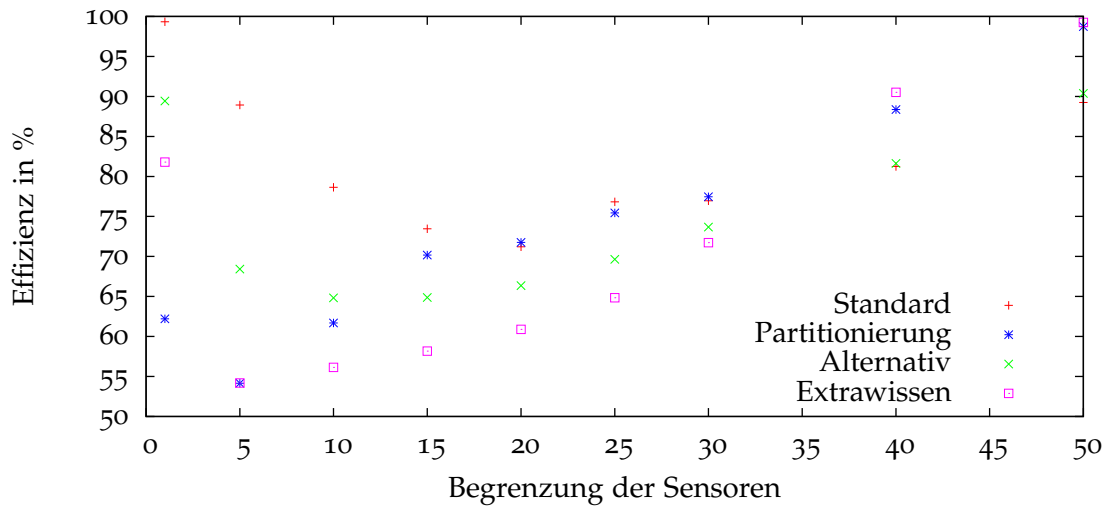
Als Grundlage für die Simulation dient eine Messreihe über eine Temperaturverteilung [DGM<sup>+</sup>04]. Daraus können Werte für die virtuellen Sensoren zu festen Zeitpunkten gelesen werden. Der Datensatz umfasst 50 Sensoren. Durch fehlende Werte im Datensatz werden nicht verfügbare Sensoren simuliert. Zusätzlich wird durch ein Limit die Anzahl der zu lesenden Sensoren bei der Auswahl begrenzt, wodurch zusätzlich eine geringere Verfügbarkeit erzwungen wird. Dieses Limit ist eine Variable der Simulation um das Verhalten der Ansätze bei unterschiedlicher Verfügbarkeit zu untersuchen.

Im Folgenden werden vier verschiedene Szenarien miteinander verglichen. Beim Standardfall werden Standardanfragen versendet und beantwortet. Für jeden enthaltenen Sensor wird der entsprechende Wert aus der Messreihe gelesen, ist dieser vorhanden, wird er dem Gateway gemeldet.

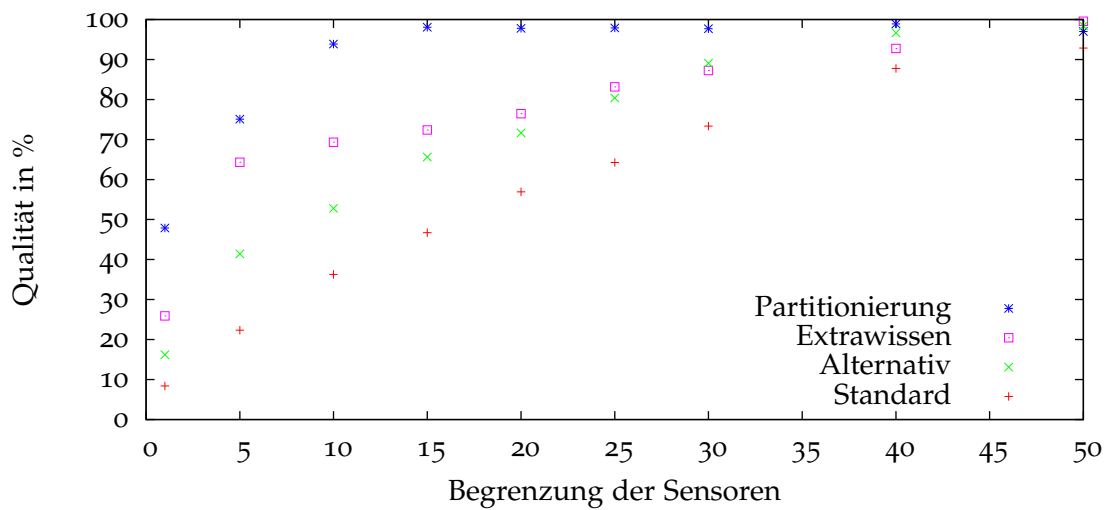
Beim einfachen Alternativplanungsansatz werden erweiterte Anfragen versendet, das heißt es sind auch die nicht angefragten Sensoren bekannt. Die Anfrage wird zunächst wie im Standardfall behandelt. Können jedoch nicht für alle Sensoren Werte ermittelt werden, wird eine neue Sensorauswahl getroffen. Dabei werden die ausgefallenen Sensoren durch neue, noch verfügbare Sensoren ersetzt. Bei diesem Ansatz ist es jedoch nicht möglich, die Gesamtzahl der zu lesenden Sensoren durch das Limit zu begrenzen. Das Limit wird lediglich auf das Teilgebiet bezogen. Für die zusätzlich ausgewählten Sensoren werden dann ebenfalls Werte aus der Messreihe gelesen und weitergeleitet.

Bei der Partitionierung wird ebenfalls die erweiterte Anfrage verwendet. Dazu wird zunächst die Anfrage auf  $2 \times 2$  Untergebiete aufgeteilt, sodass vier Anfragen mit jeweils einem Viertel der Sensoren und des Modells entstehen. Diese Anfragen werden dann jeweils wie beim Alternativplanungsansatz bearbeitet.

Die Partitionierung mit Zusatzwissen stellt eine Erweiterung zum vorigen Fall dar. Bei der Aufteilung werden aber zusätzlich sämtliche ausgewählten Sensoren in allen Teilanfragen übernommen. Allerdings sind sie nur im entsprechenden Untergebiet als abzufragende Sensoren markiert, sodass sie in den anderen Gebieten nicht gelesen werden. Für die Alternativplanung werden diese Sensoren ermittelt und wie Sensoren mit empfangenen



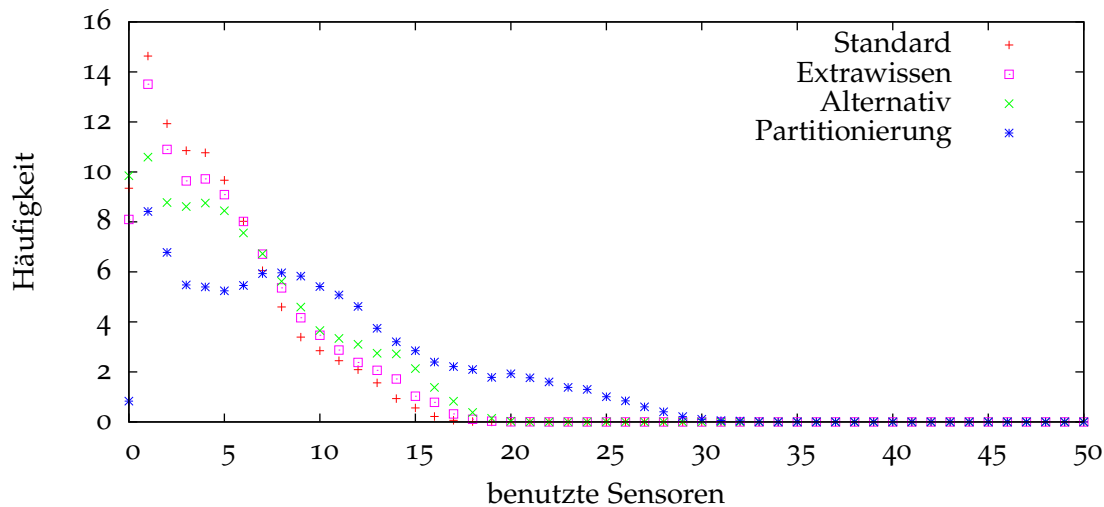
(a) Effizienz



(b) Qualität

**Abbildung 6.1:** Effizienz (a) und Qualität (b) der vier Ansätze: Standardansatz, das Benutzen einer Alternativplanung, Partitionierung mit anschließender Alternativplanung und Partitionierung mit Alternativplanung mit Extrawissen über die anderen Teilgebiete.





**Abbildung 6.2:** Verteilung der Häufigkeit tatsächlich gelesener Sensoren bei einer Begrenzung von 20 Sensoren pro Anfrage

Werten behandelt. Durch dieses Extrawissen soll verhindert werden, dass zu viele Sensoren pro Teilgebiet ausgewählt werden.

Für jeden der vier Ansätze wurde auf den selben Messwerten simuliert. Das Limit für die Sensorauswahl wurde jeweils variiert, um die Verfügbarkeit der Sensoren zu verändern. In Abbildung 6.1(a) sieht man, welche Effizienz sich für die einzelnen Ansätze bei verschiedenen Begrenzungen für die Sensorwahl ( $x$ -Achse) ergeben. Die Effizienz stellt dabei das Verhältnis zwischen den Lernphasen, in denen das Modell angepasst werden muss, und den Optimierungsphasen, in denen das Modell benutzt werden kann, dar. Je größer der Anteil der Optimierungsphasen, desto größer ist die Effizienz. Abbildung 6.1(b) zeigt die Qualität der Messwerte, das heißt die Häufigkeit, mit der die Abweichung des berechneten Ergebnisses von den Kontrollwerten innerhalb der Toleranz liegt. Dazu werden nur die Anfragen während der Optimierungsphase herangezogen.

Da meistens nicht alle Sensoren Werte liefern, entspricht die Zahl der angefragten Sensoren meist nicht der Zahl der tatsächlich erhaltenen. Wie viele Werte tatsächlich erhalten wurden, wenn die Auswahl auf 20 Sensoren beschränkt ist, ist in Abbildung 6.2 dargestellt.

## 6.2 Diskussion

Aus der Abbildung 6.1(a) erkennt man, dass für alle vier Ansätze die Effizienz zunächst fällt und dann mit steigender Sensorzahl wieder ansteigt. Dies lässt sich damit erklären, dass zunächst zu wenige Messwerte vorhanden sind, um die Genauigkeit des Modells zu überprüfen. Es kann zwar sein, dass das Modell nicht gut angepasst ist, dieser Umstand aber vom System nicht so schnell bemerkt wird wie bei mehr Sensoren. Dadurch wird weniger und

erst später in die Lernphase gewechselt. Vor dem Anstieg ist der Wert des Standardansatzes am höchsten und wird dann mit mehr zur Verfügung stehenden Sensoren von den anderen Ansätzen eingeholt und sogar überstiegen. Zudem kann man erkennen, dass der Anstieg der Kurve bei den Ansätzen mit Partitionierung aber auch der reinen Alternativplanung früher beginnt als beim Standardansatz. Wenn der Grund hierfür tatsächlich die bessere Erkennung der Abweichung des Modells ist, müssten die erweiterten Ansätze eine höhere Qualität besitzen, da sie offensichtlich über bessere Information verfügen.

Betrachtet man den Verlauf der Qualität in Abbildung 6.1(b) sieht man, dass die Kurven aller drei erweiterten Ansätze über dem Standardansatz verlaufen. Die höchste Qualität liefert dabei der Ansatz mit Partitionierung und Alternativplanung ohne Zusatzwissen über die anderen Teilgebiete. Die Partitionierung mit Extrawissen verläuft für weniger verfügbare Sensoren zunächst oberhalb der Kurve der einfachen Alternativplanung, nähert sich aber für mehr Sensoren immer stärker dieser Kurve an und fällt am Ende sogar leicht darunter.

Es wurde im Verlauf der Arbeit bereits erläutert, dass die Partitionierung ohne Extrawissen zu viele Sensoren auswählt. Bei der Alternativplanung werden zusätzliche Sensoren ausgewählt. Auch für diese Auswahl soll das Limit eingehalten werden. Beim einfachen Partitionierungsansatz ist jedoch zum Zeitpunkt der Alternativplanung von der Auswahl der anderen Gebiete bekannt, so dass das Limit nur auf die eigene Auswahl bezogen werden kann. Dadurch werden insgesamt mehr Sensoren abgefragt, als es das Limit zulassen würde. In Abbildung 6.2 kann man dies nun deutlich erkennen. Hier ist dargestellt, von wie vielen Sensoren tatsächlich ein Wert erhalten wird, wenn die Anzahl der angefragten Sensoren auf 20 begrenzt ist. Je höher das Schaubild im linken Bereich ist, desto weniger Sensoren werden benutzt. Die anderen drei Ansätze haben nie mehr Sensorwerte als das Limit erhalten. Es ist daher auch davon auszugehen, dass nie mehr als das Limit angefragt wurden. Der einfache Partitionierungsansatz hat trotz Beschränkung bei manchen Anfragen weit über 20 Messwerte erhalten. Das erklärt die bessere Qualität im Vergleich zu den anderen Ansätzen, denn wie aus der Abbildung 6.1(b) gut ablesen kann, steigt die Qualität je mehr Messwerte zur Verfügung stehen.

Zur Lösung dieses Problems wurde der Ansatz mit dem Zusatzwissen über die anderen Teilgebiete entwickelt. Betrachtet man den Verlauf der zugehörigen Kurven in Abbildung 6.2 so sieht man, dass dadurch tatsächlich das Problem der übermäßigen Sensorauswahl verhindert wird.

Bei dieser Simulation wurde in den meisten Fällen genau 20 Sensoren angefragt, sodass die Begrenzung der Sensoren durch das Limit und nicht durch das Erreichen der Varianz festgelegt wurde. Der Unterschied zwischen den Ansätzen hinsichtlich der benutzten Sensoren ergibt sich aus der unterschiedlichen Alternativplanung, da sich die Szenarien nur in der Behandlung der Anfragen unterscheiden. Da der Standardansatz keine Alternativplanung benutzt, werden hier im Vergleich am wenigsten Sensoren benutzt. Dass die einfache Partitionierung zu viele Sensoren auswählt, wurde bereits dargelegt. In Abbildung 6.2 erkennt man, dass der einfache Alternativansatz etwas mehr Sensoren verwendet als die Partitionierung mit Extrawissen. Bei der Alternativplanung unterscheiden sich diese beiden Ansätze lediglich darin, dass bei der Partitionierung für jedes Teilgebiet getrennt alternative Sensoren ausgewählt werden. Die tatsächliche Verfügbarkeit der neu gewählten Sensoren

kann aber in beiden Fällen nicht beeinflusst werden. Eine Möglichkeit, die den Unterschied erklären kann ist, dass die einfache Alternativplanung neue Sensoren aus einem anderen Teilgebiet wählt wie das, in dem die ausgefallenen Sensoren durch die Partitionierung liegen. Es kann nun sein, dass die Wahrscheinlichkeit der Verfügbarkeit von Sensoren, die weiter von den ausgefallenen Sensoren entfernt liegen, größer ist, zum Beispiel weil Sensoren flächendeckend nicht verfügbar sind. In diesem Fall kann die einfache Alternativplanung eine Neuauswahl treffen, von der mehr Werte erfasst werden können als bei der Neuauswahl der Partitionierung mit Zusatzwissen.

Aus der Simulation hat sich gezeigt, dass durch die Alternativplanung die Qualität der Ergebnisse aus der modellgetriebenen Erfassung verbessert wird. Mit dem erhöhten Aufwand bei der Nachrichtenübertragung ist diesem Vorteil der einfachen Alternativplanung jedoch ein erheblicher Nachteil gegenübergestellt. Alle Beobachtungen zusammengenommen zeigt sich hier die Partitionierung mit Zusatzwissen als guter Ansatz die Alternativplanung einzusetzen. Im Vergleich zur einfachen Alternativplanung besitzt sie den Vorteil, dass ein wesentlich kleineres Modell mitgeschickt werden kann. Zudem ist eine lokale Alternativauswahl in einer realen Umgebung sinnvoller, um Routingaufwand zu sparen. Die Qualität der Erfassung ist gut, was unter diesen Szenarien nur von der Partitionierung ohne Zusatzwissen übertroffen wurde, die jedoch wesentlich mehr Sensoren benutzt hat. Dadurch ist auch der Vorteil gegenüber diesem Ansatz gegeben.



## 7 Zusammenfassung und Ausblick

In dieser Arbeit wurde eine Methode vorgestellt, wie man die Qualität der Messergebnisse in modellgetriebenen opportunistischen Public-Sensing-Systemen verbessern kann. Bei der modellgetriebenen Messwerterfassung kann ein Teil der Messpunkte aus einem Modell abgeleitet werden, sodass weniger Messwerte tatsächlich erfasst werden müssen. Durch die Mobilität der Messgeräte kann es in einem Public-Sensing-System häufig vorkommen, dass an einem bestimmten Punkt kein Wert erfasst werden kann. Um diesem Problem entgegenzuwirken wurde die Methode der Alternativplanung vorgestellt.

In der Simulation hat sich gezeigt, dass die Qualität durch alternative Sensorauswahl tatsächlich erhöht werden kann. Um die Alternativplanung direkt auf den mobilen Knoten ausführen zu können, muss das Modell in jeder Anfrage mitgeschickt werden. Da dies mit einem erheblichen Nachrichtenmehraufwand verbunden ist, wurde in dieser Arbeit die Partitionierung des Modells zur Lösung dieses Problems vorgestellt. Dabei wird das untersuchte Gebiet in mehrere Untergebiete aufgeteilt und das Modell jeweils auf diese Untergebiete reduziert.

Bei einer Partitionierung des Modells, sind in jedem Teilgebiet nur die eigenen Sensoren bekannt. Über die anderen Gebiete liegt jeweils keinerlei Information vor. Es hat sich in der Simulation gezeigt, dass dadurch eine Alternativplanung nicht mehr effizient möglich ist. Ohne Wissen über die anderen Teilgebiete wählt jedes Gebiet dabei zu viele Sensoren aus, da versucht wird, die geforderte Qualität ohne die anderen Gebiete zu erreichen.

Um das Modell so zu partitionieren, dass bei einer Alternativplanung innerhalb eines Teilgebiets auch die anderen Gebiete berücksichtigt werden, wurde in dieser Arbeit die Partitionierung mit Zusatzwissen vorgestellt. Dabei werden jedem Untergebiet die ausgewählten Sensoren der anderen Gebiete mitgeteilt. Die Partitionierung dadurch jedoch nicht mehr disjunkt. Die Nachrichtengröße bleibt aber weiterhin kleiner als bei der Alternativplanung ohne Partitionierung. Zudem nutzt man den Vorteil, dass die Alternativplanung keine weit entfernten Sensoren auswählen kann, was den Routingaufwand im Vergleich zur einfachen Alternativplanung reduziert.

### Ausblick

In dieser Arbeit wurden nur einige der möglichen Parameter untersucht. In weiterführenden Arbeiten kann getestet werden, welche Auswirkungen eine stärker ausgeprägte Alternativplanung hat, die nicht nur einmalig Ersatzsensoren aussucht. Es wurde die Annahmen aufgestellt, dass die erste Alternativauswahl noch den größten Erfolg verspricht. Dennoch

kann es von Nutzen sein, weitere Alternativen zu wählen, falls sich die ersten als nicht verfügbar erweisen. Verschiedenen Ansätze zur Wiederholung der Alternativplanung können in weiteren Studien auf Vor- und Nachteile untersucht werden.

Der Hauptaspekt dieser Arbeit liegt darin, die Qualität der Messungen auch bei ausfallenden Sensoren hoch zu halten. Der dadurch zusätzliche entstehende Aufwand soll nach Möglichkeit gering gehalten werden. Dieser Aufwand wurde durch die Größe der Anfrage und die Zahl der angefragten Sensoren lediglich qualitativ abgeschätzt. In einem nächsten Schritt kann der tatsächliche Nachrichtenaufwand der verschiedenen Szenarien genauer untersucht werden.

## Literaturverzeichnis

- [CEL<sup>+</sup>08] A. T. Campbell, S. B. Eisenman, N. D. Lane, E. Miluzzo, R. A. Peterson, H. Lu, X. Zheng, M. Musolesi, K. Fodor, G.-S. Ahn. The Rise of People-Centric Sensing. *IEEE Internet Computing*, 12(4):12–21, 2008. doi:10.1109/MIC.2008.90. URL <http://dx.doi.org/10.1109/MIC.2008.90>. (Zitiert auf Seite 9)
- [CKK<sup>+</sup>08] C. Cornelius, A. Kapadia, D. Kotz, D. Peebles, M. Shin, N. Triandopoulos. Anonymsense: privacy-aware people-centric sensing. In *Proceedings of the 6th international conference on Mobile systems, applications, and services*, MobiSys '08, S. 211–224. ACM, New York, NY, USA, 2008. doi:10.1145/1378600.1378624. URL <http://doi.acm.org/10.1145/1378600.1378624>. (Zitiert auf Seite 9)
- [DGM<sup>+</sup>04] A. Deshpande, C. Guestrin, S. R. Madden, J. M. Hellerstein, W. Hong. Model-driven data acquisition in sensor networks. In *Proceedings of the Thirtieth international conference on Very large data bases - Volume 30*, VLDB '04, S. 588–599. VLDB Endowment, 2004. URL <http://dl.acm.org/citation.cfm?id=1316689.1316741>. (Zitiert auf den Seiten 11, 14 und 31)
- [Eis08] S. B. Eisenman. People-Centric Mobile Sensing Networks, 2008. (Zitiert auf Seite 13)
- [HBZ<sup>+</sup>06] B. Hull, V. Bychkovsky, Y. Zhang, K. Chen, M. Goraczko, A. K. Miu, E. Shih, H. Balakrishnan, S. Madden. CarTel: A Distributed Mobile Sensor Computing System. In *4th ACM SenSys*. Boulder, CO, 2006. (Zitiert auf Seite 14)
- [KBP<sup>+</sup>08] E. Kanjo, S. Benford, M. Paxton, A. Chamberlain, D. S. Fraser, D. Woodgate, D. Crellin, A. Woolard. MobGeoSen: facilitating personal geosensor data collection and visualization using mobile phones. *Personal Ubiquitous Comput.*, 12(8):599–607, 2008. doi:10.1007/s00779-007-0180-1. URL <http://dx.doi.org/10.1007/s00779-007-0180-1>. (Zitiert auf Seite 13)
- [KSGo8] A. Krause, A. Singh, C. Guestrin. Near-Optimal Sensor Placements in Gaussian Processes: Theory, Efficient Algorithms and Empirical Studies. *J. Mach. Learn. Res.*, 9:235–284, 2008. URL <http://dl.acm.org/citation.cfm?id=1390681.1390689>. (Zitiert auf den Seiten 13, 14 und 19)
- [LBD<sup>+</sup>05] B. Liu, P. Brass, O. Dousse, P. Nain, D. Towsley. Mobility improves coverage of sensor networks. In *Proceedings of the 6th ACM international symposium on Mobile ad hoc networking and computing*, MobiHoc '05, S. 300–308. ACM, New York, NY, USA, 2005. doi:10.1145/1062689.1062728. URL <http://doi.acm.org/10.1145/1062689.1062728>. (Zitiert auf Seite 10)

- [LEM<sup>+</sup>08] N. D. Lane, S. B. Eisenman, M. Musolesi, E. Miluzzo, A. T. Campbell. Urban sensing systems: opportunistic or participatory? In *Proceedings of the 9th workshop on Mobile computing systems and applications*, HotMobile '08, S. 11–16. ACM, New York, NY, USA, 2008. doi:10.1145/1411759.1411763. URL <http://doi.acm.org/10.1145/1411759.1411763>. (Zitiert auf den Seiten 9 und 10)
- [MSN<sup>+</sup>09] N. Maisonneuve, M. Stevens, M. E. Niessen, P. Hanappe, L. Steels. Citizen noise pollution monitoring. In *Proceedings of the 10th Annual International Conference on Digital Government Research: Social Networks: Making Connections between Citizens, Data and Government*, dg.o '09, S. 96–103. Digital Government Society of North America, 2009. URL <http://dl.acm.org/citation.cfm?id=1556176.1556198>. (Zitiert auf Seite 14)
- [PDR11] D. Philipp, F. Durr, K. Rothermel. A Sensor Network Abstraction for Flexible Public Sensing Systems. In *Mobile Adhoc and Sensor Systems (MASS), 2011 IEEE 8th International Conference on*, S. 460–469. 2011. doi:10.1109/MASS.2011.52. (Zitiert auf Seite 10)
- [PSDR12] D. Philipp, J. Stachowiak, F. Dürr, K. Rothermel. Towards Optimized Public Sensing Systems using Data-driven Models. Technischer Bericht Informatik 2012/05, Universität Stuttgart, Fakultät Informatik, Elektrotechnik und Informationstechnik, Germany, Universität Stuttgart, Institut für Parallele und Verteilte Systeme, Verteilte Systeme, 2012. URL [http://www2.informatik.uni-stuttgart.de/cgi-bin/NCSTR/L/NCSTR\\_view.pl?id=TR-2012-05&engl=0](http://www2.informatik.uni-stuttgart.de/cgi-bin/NCSTR/L/NCSTR_view.pl?id=TR-2012-05&engl=0). (Zitiert auf Seite 15)
- [Sta11] J. Stachowiak. *Optimized acquisition of spatially distributed phenomena in public sensing systems*. Diplomarbeit, Universität Stuttgart, Holzgartenstr. 16, 70174 Stuttgart, 2011. URL <http://elib.uni-stuttgart.de/opus/volltexte/2011/6983>. (Zitiert auf Seite 13)



## **Erklärung**

Hiermit versichere ich, diese Arbeit selbständig verfasst und nur die angegebenen Quellen benutzt zu haben.

---

(Maren Tilk)