

Vergleich von Fourier-Transformation und der Methode der kleinsten Quadrate harmonischer Funktionen zur Grundwasser-Gezeitenanalyse

(Comparison of the Fourier transform and the
harmonic least-squares method for analysing
groundwater tides)

Bachelorarbeit

von

Vincent Ried

am Geophysikalischen Institut/Institut für Angewandte
Geowissenschaften

Referent: Prof. Dr. Thomas Bohlen

Koreferent: Dr. Gabriel Rau

Bearbeitungszeit: 24.09.2019 – 17.01.2020

Erklärung zur Selbstständigkeit

Ich versichere, dass ich diese Arbeit selbstständig verfasst habe und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe, die wörtlich oder inhaltlich übernommenen Stellen als solche kenntlich gemacht und die Satzung des KIT zur Sicherung guter wissenschaftlicher Praxis in der gültigen Fassung vom 17.05.2010 beachtet habe.

Karlsruhe, den 17.01.2020, _____
Vincent Ried

Als Ansichtsexemplar genehmigt von

Karlsruhe, den 17.01.2020, _____
Prof. Dr. Thomas Bohlen

Abstract

Eine nachhaltige Nutzung von Grundwasser setzt die Kenntnis hydrogeomechanischer Parameter voraus. Diese werden bisher hauptsächlich durch aufwändige Permeametertests oder Pumpversuche bestimmt. Eine Grundwasser-Gezeitenanalyse bietet die Möglichkeit, Bodenkennwerte anhand von Grundwasserstandsdaten zeit- und kosteneffizient zu ermitteln. Hierfür findet häufig die Fourier-Transformation Anwendung. Mit der Methode der kleinsten Quadrate lassen sich die Amplituden harmonischer Schwingungen ebenfalls bestimmen. Diese Arbeit vergleicht beide Verfahren miteinander, um die Frequenzanalyse von Grundwasserdaten zu verbessern. Außerdem werden daraus Rahmenbedingungen einer Grundwassermessung hinsichtlich der Gezeitenanalyse definiert. Diese umfassen Parameter wie die Messdauer, die Quantisierung der Messdaten sowie Datenlücken. Bei synthetischen Zeitreihen und an einem realen Datensatz werden diese drei Messparameter variiert und ihr Einfluss auf die Genauigkeit der durch beide Regressionsanalyseverfahren bestimmten Amplituden betrachtet. Die Untersuchung zeigt, dass die Fourier-Transformation für die meisten Parametereinstellungen ungenauere Ergebnisse liefert als die Methode der kleinsten Quadrate. Letztere ist insbesondere bei kurzen sowie lückenhaften Zeitreihen der Fourier-Transformation vorzuziehen. Es wird gezeigt, dass eine Standard-Sensorauflösung von 1-5 cm ausreichend, um eine Gezeitenanalyse für den vorliegenden realen Datensatz durchzuführen.

A sustainable usage of groundwater presumes the knowledge of hydrogeomechanical parameters. Those are usually determined by expensive permeameter or pumping tests. Tidal analysis of groundwater enables determining soil characteristics in a time- and cost-effective manner. This is mainly done using the Fourier transform, although calculating the amplitudes of harmonics oscillations is also possible with the least-squares method. This document provides recommendations to improve groundwater measurements in respect of tidal analysis. In addition, basic conditions of groundwater measurement are defined. These are in particular the data length, the quantisation as well as data gaps. For this purpose, these three parameters are varied from synthetic time series and from a real data set. The effect on the accuracy of the determined amplitudes from tidal components is observed, using both the Fourier transform and the least-squares method. As a result, it is shown that the Fourier transform is for almost each parameter setting less accurate than the least-square method, especially for short or incomplete time series. Beside, the standard transducer resolution of 1-5 cm is satisfactory for tidal analysis of the examined real data set.

Inhaltsverzeichnis

1	Einführung	13
1.1	Hintergrund und Motivation	13
1.2	Ursachen der Grundwasser-Gezeiten	15
1.3	Aufgabenstellung	16
2	Methodik	19
2.1	Literaturrecherche zur Regressionsanalyse	19
2.1.1	Die Fourier-Analyse	19
2.1.2	Die Methode der kleinsten Quadrate	21
2.2	Synthetische Messreihen	22
2.3	Reale Messdaten	25
2.4	Einfluss verschiedener Parameter	26
2.4.1	Länge der Zeitreihe	26
2.4.2	Auflösung des Drucksensors	26
2.4.3	Datenlücken	27
2.5	Anwendung der Regressionsanalyseverfahren	29
3	Ergebnis	31
3.1	Synthetische Datensätze	31
3.1.1	Fourier-Analyse	31
3.1.2	Methode der kleinsten Quadrate	34
3.2	Realdaten	37
4	Diskussion	43
4.1	Der Einfluss der Messdauer	43
4.2	Der Einfluss der Quantisierung	46
4.3	Der Einfluss der Datenlücken	49
5	Schlussfolgerung und Ausblick	53
	Literaturverzeichnis	55
	Anhang	57

Abbildungsverzeichnis

1.1	Gezeiteneinflüsse auf eine Grundwassermessstelle (Piezometer) (McMillan et al., 2019)	14
1.2	Arbeitsschritte und Forschungsstand der Grundwasser-Gezeitenanalyse (McMillan et al., 2019)	15
1.3	Übersicht der Gezeitenkomponenten von McMillan et al. (2019) mit ihren jeweiligen Frequenzen in Zyklen pro Tag (cycles per day, cpd) .	16
2.1	Abgetastetes Signal aus drei Sinusschwingungen mit dazugehörigem Frequenzspektrum	20
2.2	Durch die Hanning-Fensterfunktion verringern sich die Sprünge zwischen letztem und ersten Messwert aufeinanderfolgender Perioden . .	21
2.3	Durch eine Modellfunktion (blau) angepasste Messwerte (rot): Die Residuen (grün) sollten durch die Wahl einer geeigneten Modellfunktion minimiert werden.	22
2.4	Nach Formel 2.9 modellierte Messreihe von stündlichen Messungen. Zur besseren Sichtbarkeit beträgt das Signal-Rausch-Verhältnis 10. . .	23
2.5	Datensatz einer Grundwassermessstelle in Baldry, Australien (bore hole, BH)(McMillan et al., 2019) (blau) und 7-Tages-Trend (orange) .	25
2.6	Vom 7-Tages Trend bereinigte Realdaten (Differenz vom orangefarbenen zum blauen Graphen aus Abbildung 2.5)	26
2.7	Quantisierung der Funktionswerte eines kontinuierlichen Signals (Rau et al., 2019): Je nach Auflösung des Messinstruments entstehen Stufen in den aufgezeichneten Daten.	27
2.8	Quantisierungen eines modellierten Ursprungssignals mit Frequenzen f_{M_2} und f_{S_2} und jeweils einer Amplitude von 10 mm.	28
2.9	Ausschnitt einer nach Formel 2.9 simulierten Messreihe mit Datenlücken	29
2.10	Beispiel des Amplitudenspektrums eines synthetischen Signals	30
3.1	Amplitudenspektrum von Zeitreihen verschiedener Längen ohne Rauschen in der Umgebung von f_{M_2} und f_{S_2}	31
3.2	Einfluss der Länge der Messreihe auf die Amplituden der Komponenten M_2 (blau) und S_2 (schwarz) bei verschiedenen SNRs bei einer Auswertung mittels FT	32
3.3	Amplitudenspektren der quantisierten Signale aus Abbildung 2.8	33
3.4	Einfluss der Quantisierung der Messwerte auf die Amplituden der Komponenten M_2 (blau) und S_2 (schwarz) bei verschiedenen SNRs bei einer Auswertung mittels FT	33

3.5	Einfluss des Lückenanteils auf die Amplituden der Komponenten M_2 (blau) und S_2 (schwarz) bei verschiedenen SNRs bei einer Auswertung mittels FT	34
3.6	Einfluss der Länge der Messreihe auf die Amplituden der Komponenten M_2 (blau) und S_2 (schwarz) bei verschiedenen SNRs bei einer Auswertung mittels der MdkQ	35
3.7	Einfluss der Quantisierung der Messwerte auf die Amplituden der Komponenten M_2 (blau) und S_2 (schwarz) bei verschiedenen SNRs bei einer Auswertung mittels MdkQ	36
3.8	Einfluss des Lückenanteils auf die Amplituden der Komponenten M_2 (blau) und S_2 (schwarz) bei verschiedenen SNRs bei einer Auswertung mittels MdkQ	36
3.9	Einfluss des Lückenanteils bei über 90 % auf die Amplituden der Komponenten M_2 (blau) und S_2 (schwarz) bei verschiedenen SNRs bei einer Auswertung mittels MdkQ	37
3.10	Frequenzspektrum der Realdaten	37
3.11	Amplituden der Komponenten M_2 und S_2 der Realdaten bei steigender Messdauer	38
3.12	Quantisierungen der Realdaten	39
3.13	Amplituden von M_2 und S_2 der Realdaten bei steigender Quantisierung bei Auswertung mit FT und MdkQ	39
3.14	Durch FT ermittelte Amplituden der Realdaten bei steigendem Lückenanteil	40
3.15	Durch MdkQ ermittelte Amplituden der Realdaten bei steigendem Lückenanteil. Die rote Linie kennzeichnet die Amplitude bei vollständigem Datensatz.	41
4.1	Amplituden von M_2 und S_2 aus FT und MdkQ für kurze Messdauern. Die Linien deuten die ermittelten Amplituden der Komponenten bei vollständigem Datensatz an.	44
4.2	Amplitudenspektren der Realdaten um f_{M_2} und f_{S_2} bei steigender Quantisierung q	48
4.3	Ein um die Quantisierungsgrenze pendelndes Signal erzeugt ein stark variierendes quantisiertes Signal	49
4.4	Amplitudenspektrum der Realdaten bei steigendem Lückenanteil	51

Tabellenverzeichnis

1	Übersicht der verwendeten Abkürzungen und Variablen	11
2.1	Übersicht der Modellparameter: Die mit * gekennzeichneten Parameter beziehen sich nicht auf die Versuchsreihe der Variation der Messdauer.	24

4.1	Mittelwerte in mm der Amplituden aus Abbildung 3.2 der Messdauern von 60 bis 120 Tagen.	45
4.2	Maximale Quantisierung in Einheiten von A_0 , mit der eine Genauigkeit von 90 % erreicht wird.	47
4.3	Durch lineare Regression ermittelte Amplituden in mm von M_2 und S_2 der synthetischen Daten nach Erhöhung des Lückenanteils	51

Abkürzungen

Tabelle 1: Übersicht der verwendeten Abkürzungen und Variablen

Abkürzung	Bedeutung	Referenz
A_i	Amplitude einer Gezeitenkomponente	Abschnitt 2.1.2
AT	Atmosphärische Gezeiten	Abschnitt 1.2
cpd	Zyklen pro Tag	Abschnitt 1.2
Δf	Frequenzauflösung des Spektrums	Puente León (2019)
$\Delta f_{M_2S_2}$	Frequenzabstand von f_{M_2} und f_{S_2}	Abschnitt 2.4.1
ET	Erdgezeiten	Abschnitt 1.2
f_i	Frequenz einer Gezeitenkomponente	Abschnitt 2.4
f_N	Nyquist-Frequenz	Puente León (2019)
FT	Fourier-Transformation	Abschnitt 2.1.1
GM	Gezeitenmethode	Abschnitt 1.1
L	Lückenanteil	Abschnitt 2.4.3
MdkQ	Methode der kleinsten Quadrate	Abschnitt 2.1.2
N	Anzahl Messwerte	Abschnitt 2.1.1
SNR	Signal-Rausch Verhältnis	Abschnitt 2.2
σ	Rauschanteil	Abschnitt 2.2
t_A	Abtastzeit	Abschnitt 2.1.1
T	Messdauer	Abschnitt 2.1.1

Kapitel 1

Einführung

1.1 Hintergrund und Motivation

Grundwasser ist für Menschen eine lebenswichtige Ressource. Es ist die Grundlage für Ökosysteme und essenziell für die Wasserversorgung von Bevölkerung und Industrie. Um eine möglichst nachhaltige Nutzung dieses Guts zu gewährleisten, ist es notwendig, die unterirdische Hydrosphäre zu erforschen (McMillan et al., 2019; Jordan and Abraham, 1995).

Grundwasserstände werden mit Piezometern gemessen. Dabei handelt es sich um ein Bohrloch, in dem die zeitliche Änderung des Grundwasserstandes durch eingelassene Drucksensoren beobachtet wird (Abbildung 1.1). Steigt der Wasserspiegel, erhöht sich der Druck p an der Messsonde. Dieser lässt sich durch die Dichte ρ_w des Wassers auf die Höhe der Wassersäule h über dem Drucksensor zurückrechnen (vgl. (Rau et al., 2019)):

$$h = \frac{p}{\rho_w g} \quad (1.1)$$

$g \approx 9,81 \text{ m/s}^2$ ist hierbei die Gravitationskonstante.

Durch die Kenntnis von Grundwasserständen wird üblicherweise die Fließrichtung bestimmt (Hiscock and Bense, 2014). Aktueller Forschungsstand ist, aus der zeitlichen Änderung des Grundwasserspiegels weitere Informationen über den Untergrund zu gewinnen (Acworth et al., 2015; McMillan et al., 2019). Dazu liegt der Fokus bei der Analyse von Grundwasserdaten auf den hochfrequenten Anteilen mit schwacher Amplitude. Diese werden meistens von langperiodischen saisonalen Anteilen überlagert.

Bodenkennwerte wie die hydraulische Leitfähigkeit oder Speicherkoeffizienten können bislang nur durch aufwändige Pumpversuche im Feld oder Permeametertests im Labor bestimmt werden. Diese sind zeit- und kostenintensiv, da sie Ausrüstung, Energieversorgung und Personal erfordern (McMillan et al., 2019). Außerdem ist die Repräsentanz des Versuchs nicht gewährleistet. Eine indirekte Methode, die die hydrogeomechanischen Eigenschaften aus Grundwasserstandsdaten zugänglich macht, ist eine effizientere Alternative zu klassischen Bestimmungsverfahren.

McMillan et al. (2019) postulieren in *Review of Geophysics*, dass aus der gezeitenbedingten Variation der Grundwasserspiegellhöhen Bodenkennwerte ermittelt werden können. Außerdem ist in der Hydrogeologie der Spannungszustand des Grundwasserleiters von großem Interesse. Diesen zu bestimmen ist jedoch für zerklüftetes

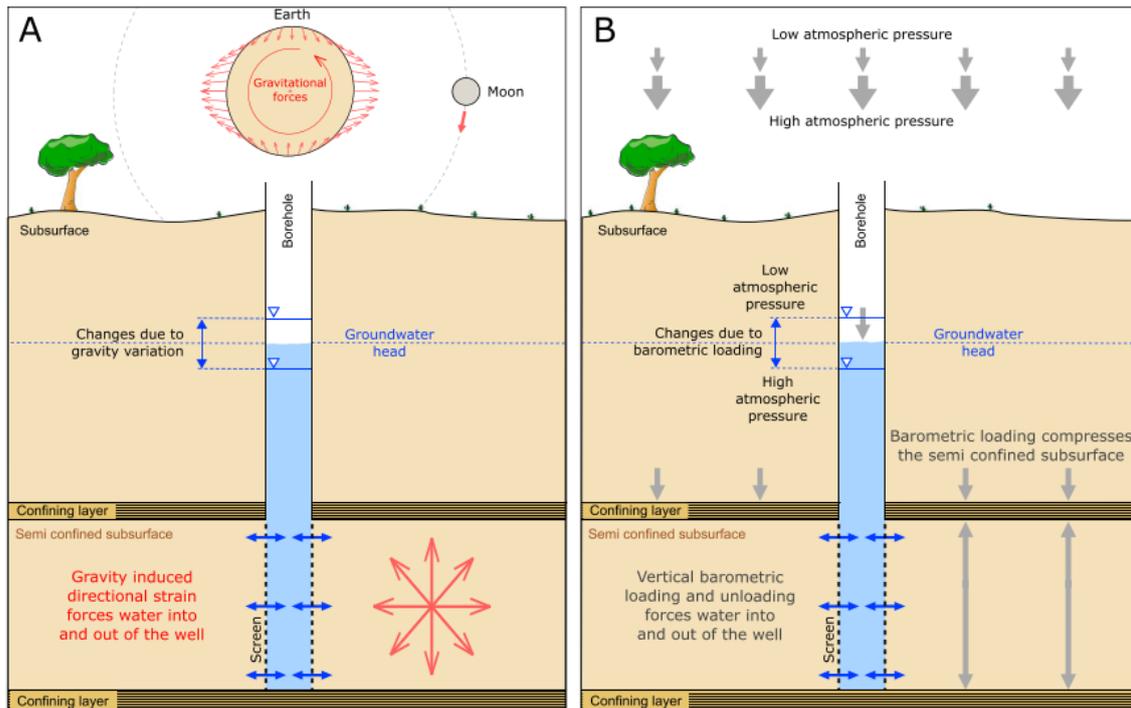


Abbildung 1.1: Gezeiteneinflüsse auf eine Grundwassermessstelle (Piezometer) (McMillan et al., 2019)

Gestein besonders schwierig (Rahi and Halihan, 2013), sodass sich eine passive Methode besonders für diese Fälle lohnt.

Zur Grundwasser-Gezeitenanalyse hat sich das in Abbildung 1.2 gezeigte Verfahren etabliert: Bisher lässt sich die Transmissivität aus Amplituden der Schwingungskomponenten ermitteln. Dies geschieht mithilfe der Diskreten Fourier-Transformation (DFT) aus Gezeiten-, Atmosphären- und Grundwasserdaten.

Basierend auf Planetenbahndaten der NASA stellte Kudryavtsev (2004) das 0,025 nGal genaue Gezeitenmodell KSM03 für den Zeitraum von 1600 bis 2200 auf. Solche Modelle können auch der Grundwasser-Gezeitenanalyse dienen, da sie neben den Amplituden der Gezeitenkomponenten auch die entsprechenden Frequenzen zur Verfügung stellen. Durch die Kenntnis dieser Frequenzen lässt sich neben der DFT auch die Methode der kleinsten Quadrate auf eine Summe harmonischer Schwingungen anwenden. Diese ermittelt genau wie die Fourier-Transformation die Amplituden der einzelnen Signalanteile.

In dieser Arbeit bezeichnet *Gezeitenmethode* (GM) das Verfahren, hydrogeomechanische Parameter aus den gezeitenbedingten Veränderungen von Grundwasserspiegelhöhen zu bestimmen. Die GM ist das übergeordnete Ziel, in das sich diese Abschlussarbeit eingliedert und die es zu optimieren gilt. Dazu werden zwei Methoden miteinander verglichen: die Analyse des Frequenzspektrums mittels Fourier-Transformation (FT) und eine Anpassung der Messdaten mit der Methode der kleinsten Quadrate (MdkQ). Es sind die Auswirkungen (1) der Länge der Messreihe, (2) der Datenquantisierung durch Digitalisierung und (3) dem Anteil an Datenlücken auf den Fehler der modellierten Funktion von Interesse. Diese Parameter sind ausschlaggebend für die Qualität von Messdaten und variieren stark zwischen Datensets.

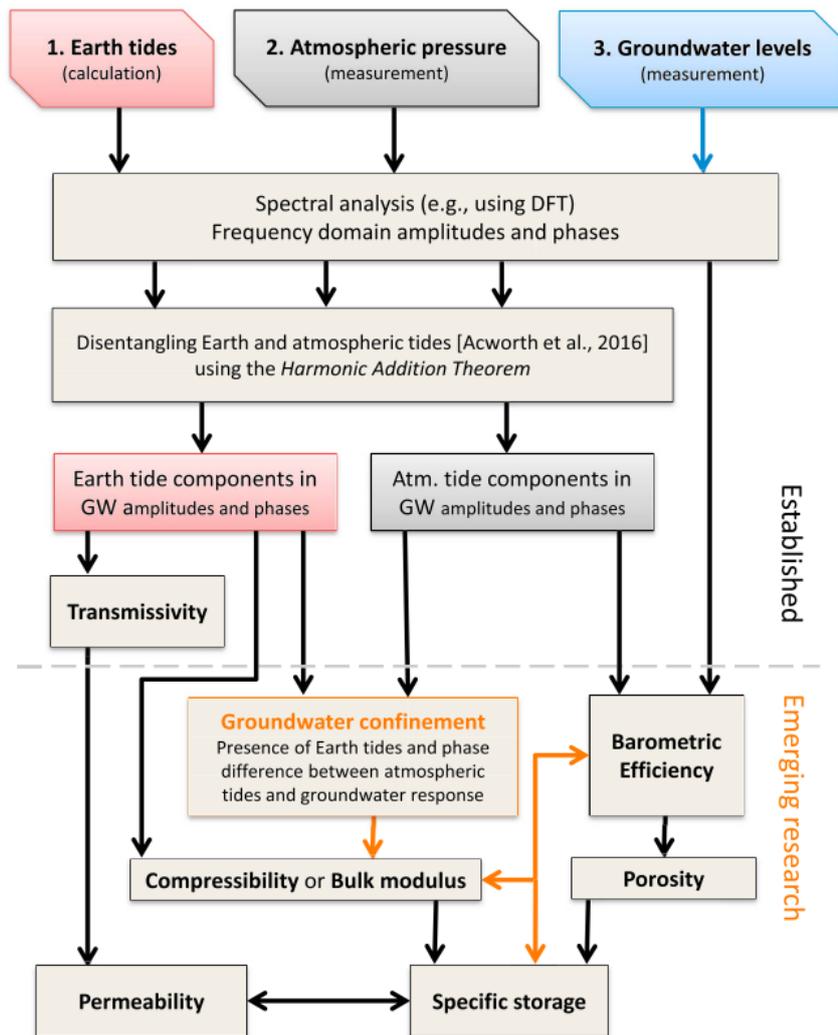


Abbildung 1.2: Arbeitsschritte und Forschungsstand der Grundwasser-Gezeitenanalyse (McMillan et al., 2019)

1.2 Ursachen der Grundwasser-Gezeiten

Gezeiten sind im klassischen Sinn periodische Änderungen des Meeresspiegels aufgrund gravitativer Einflüsse der Himmelskörper auf die Ozeane. In der Geophysik ist dieser Begriff auf die Verformung der festen Erde erweitert: Diese reagiert als elastischer Körper mit einer radialen Verformung von ± 21 cm in den mittleren Breiten auf die Anziehungskraft der Himmelskörper (Clauser, 2016).

Diese Verformungen wirken sich auch auf Grundwasserleiter aus: Gravitationsbedingte Spannungen verringern den Porenraum und zwingen Grundwasser in das Bohrloch (McMillan et al., 2019). Dies hat einen messbaren Höhenunterschied des Grundwasserspiegels zur Folge.

Auch die Atmosphäre oszilliert aufgrund der Anziehungskraft von Erde, Sonne und Mond. Nach Chapman and Malin (1970) ist aber vor allem die thermische Ausdehnung durch die Sonne Auslöser für atmosphärische Druckschwankungen, die sich auf das Grundwasser auswirken.

Man unterscheidet die Erdgezeiten (earth tides, ET) und atmosphärische Gezeiten (atmospheric tides, AT) nach ihrer Wirkungsart:

- Die ET umfassen die Effekte durch Verformung des Grundwasserleiters (Cooper et al., 1967),
- die AT ändern durch Druckänderungen in der Atmosphäre den Grundwasserstand (Haurwitz, 1964).

Zur Veranschaulichung der AT hilft die Analogie zwischen dem Grundwasserspiegel und einer Membran zwischen zwei Medien: Durch barometrische Änderungen der Atmosphäre hebt und senkt sich diese Trennebene.

Abbildung 1.3 gibt eine Übersicht der Komponenten des Frequenzspektrums für Grundwassergezeiten. Dabei weist der Index 1 oder 2 darauf hin, ob es sich um eine nahezu tägliche (diurnal) oder halbtägliche (semidiurnal) Komponente handelt.

Table 1
Table of Major Tidal Components Ordered According to Frequency in Cycles per Day (cpd)

Darwinian name	Frequency (cpd)	Tidal potential (m^2/s^2)	Tidal gravity variation (m/s^2)	Tidal dilation (-)	Description	Attribution
Diurnal						
O_1	0.929536	5.363385	8.26E-06	3.347E-08	Principal lunar diurnal	Earth
M_1	0.966446	10.286769	1.58E-05	6.419E-08	Lunar diurnal	Earth
P_1	0.997262	7.407625	1.14E-05	4.622E-08	Diurnal lunar perigee	Earth
S_1	1.000000				Principal solar atmospheric pressure (thermal)	Atmosphere
K_1	1.002738	22.924982	3.53E-05	1.431E-07	Lunar solar diurnal	Earth
Semidiurnal						
N_2	1.895982	12.963403	1.996E-05	8.089E-08	Lunar elliptic semidiurnal (variation in Moon distance)	Earth
M_2	1.932274	42.060943	6.477E-05	2.625E-07	Principal lunar semidiurnal	Earth
S_2	2.000000	19.309855	2.973E-05	1.205E-07	Principal solar semidiurnal	Atmosphere/Earth
K_2	2.005476	11.791770	1.816E-05	7.358E-08	Lunar Solar Semidiurnal	Earth

Note. Tidal gravity variations (m/s^2) and tidal dilation are calculated from the tidal potential ($V; m^2/s^2$) as $g * V/r$ and $= V/g * (\frac{2}{5}h - 3\frac{2}{5}l)/r$, respectively, where g is gravity, r is the radius of the Earth, and $\frac{2}{5}h$ and $\frac{2}{5}l$ are assumed Love numbers of 0.6 and 0.07, respectively. Note that S_1 has been included due to its large superposition effect on the other tidal components although it is not of gravitational origin. Table adapted from Darwin (1899), Munk and MacDonald (1960), and Agnew (2010).

Abbildung 1.3: Übersicht der Gezeitenkomponenten von McMillan et al. (2019) mit ihren jeweiligen Frequenzen in Zyklen pro Tag (cycles per day, cpd)

Beispielsweise verstärkt die Wärmeausdehnung der Atmosphäre die S_2 -Komponente (Haurwitz, 1964). M_2 entsteht hauptsächlich durch das Gravitationsfeld des Mondes (Chapman and Westfold, 1956).

1.3 Aufgabenstellung

Das pragmatische Ziel dieser Bachelorarbeit ist, den Einfluss von Gezeitenkräften auf den Grundwasserspiegel möglichst genau zu quantifizieren, sodass mithilfe der Gezeitenmethode mit minimalem Daten- und Messaufwand möglichst viele Erkenntnisse gewonnen werden können. Da sich so hydrogeomechanische Eigenschaften berechnen lassen, hat die Bestimmungsgenauigkeit einen großen Einfluss auf das Ergebnis.

Die Auswirkungen von Gezeitenkräften auf das Grundwasser lassen sich durch harmonische Funktionen beschreiben, die sich mit anderen oft stärkeren Umwelteinflüssen wie z.B. natürlichen Schwankungen überlagern. Dabei gilt es, die Eigenschaften der harmonischen Funktionen (Amplituden und Phasen) für definierte Frequenzen zu bestimmen. Aktuelle Forschung untersucht, wie sich aus Amplitude und Phase der Gezeitenkomponenten hydrogeomechanische Parameter ableiten lassen (McMillan et al., 2019; Rahi and Halihan, 2013; Acworth et al., 2016). Diese Arbeit

konzentriert sich auf die Bestimmung der Amplituden der Gezeitenkomponenten M_2 und S_2 in synthetischen und realen Grundwasserstandsdaten.

In der Hydrogeologie wird für die Frequenzanalyse standardmäßig die Fourier-Transformation verwendet. Diese hat einige Nachteile, zum Beispiel die Unterscheidung von naheliegenden Frequenzen im Spektrum und die Abhängigkeit der Frequenzauflösung von der Länge der Zeitreihe. Um dies zu umgehen, bietet sich die Methode der kleinsten Quadrate von harmonischen Funktionen an. Diese ermittelt die Amplituden der Einzelkomponenten, indem mit gegebenen Frequenzen eine Modellfunktion mit möglichst kleinen Fehlerquadraten an die Messdaten angepasst wird. Durch Empirie und Modellen wie dem KSM03 von Kudryavtsev (2004) sind die benötigten Frequenzen gut bekannt, sodass sich die Methode der kleinsten Quadrate als effizienteres und genaueres Bestimmungsverfahren für die Amplituden von Gezeitenkomponenten anbietet. Diese Arbeit stellt beide Regressionsanalyseverfahren gegenüber, um ihre Vor- und Nachteile im Kontext der Grundwasser-Gezeitenanalyse herauszuarbeiten.

Dazu werden beide Methoden hinsichtlich ihrer Leistungsfähigkeit bezüglich der in der hydrogeologischen Praxis üblichen Rahmenbedingungen verglichen. Dies sind insbesondere die Länge der Zeitreihe, Quantisierung der Messdaten sowie Datenlücken. Um dies zu untersuchen, werden sowohl synthetische Zeitreihen als auch ein realer Datensatz betrachtet: Deren Längen, Quantisierungen und Lückenanteile werden variiert, um den Einfluss auf die durch Fourier-Transformation und Methode der kleinsten Quadrate ermittelten Amplituden zu quantifizieren.

Ziel ist, aus den Ergebnissen dieser Bachelorarbeit (1) die Frequenzanalyse von existierenden Grundwasserdaten zu verbessern und (2) Rahmenbedingungen hinsichtlich einer auf die Gezeitenmethode abgestimmte Grundwassermessung zu definieren.

Kapitel 2

Methodik

2.1 Literaturrecherche zur Regressionsanalyse

Diese Arbeit betrachtet Signale, die als Funktion der Zeit t erfasst werden. Messinstrumente wandeln diese von einem zeitlich kontinuierlichen Signal y (analog) in eine diskrete Zeitreihe y_n (digital) mit üblicherweise konstantem Abtastintervall t_A um. Buttkus (2000) erläutert, die äquidistante Abtastung habe sich aufgrund der technisch und rechnerisch einfachen Handhabung durchgesetzt.

Regressionsanalysen zielen darauf ab, aus einer Folge von Messwerten eine möglichst stetige Funktion zu rekonstruieren, indem sie die Funktionswerte zwischen den Abtastzeitpunkten schätzen. Diese Arbeit konzentriert sich auf die beiden Regressionsverfahren Fourier-Analyse und die Methode der kleinsten Quadrate.

2.1.1 Die Fourier-Analyse

Signale lassen sich als Überlagerung von Sinusschwingungen darstellen (Fourier, 1822). Da es sich um diskrete, digitalisierte Daten handelt, findet im Kontext der Gezeitenmethode die diskrete Fourier-Transformation (DFT) mit der vereinfachten Schreibweise

$$Y_k = \sum_{n=0}^{N-1} y_n e^{-i2\pi k \frac{n}{N}} \quad (2.1)$$

Anwendung. Das Messinstrument misst üblicherweise mit einem konstanten Abtastintervall t_A ab. Dadurch entsteht die Messreihe y_n mit N Messwerten. Dies ist in Abbildung 2.1 anhand eines Beispielsignals aus drei Schwingungen veranschaulicht. Unter der Annahme, dass sich das Ursprungssignal nach Ablauf der Messdauer T periodisch fortsetzt, geht y_n durch die DFT in ein diskretes Frequenzspektrum Y_k über. Nach Puente León (2019) ist die Frequenzauflösung Δf dabei als Abstand von zwei Bins im Frequenzspektrum durch

$$\Delta f = \frac{1}{T} = \frac{1}{Nt_A} \quad (2.2)$$

gegeben. Man erkennt an Formel 2.2, dass sich die Frequenzauflösung verbessert, je länger die Messung durchgeführt wird. Im Abschnitt 2.4.1 spielt dies eine wichtige Rolle bei der Untersuchung des Einflusses der Messdauer auf die Genauigkeit der Ergebnisse.

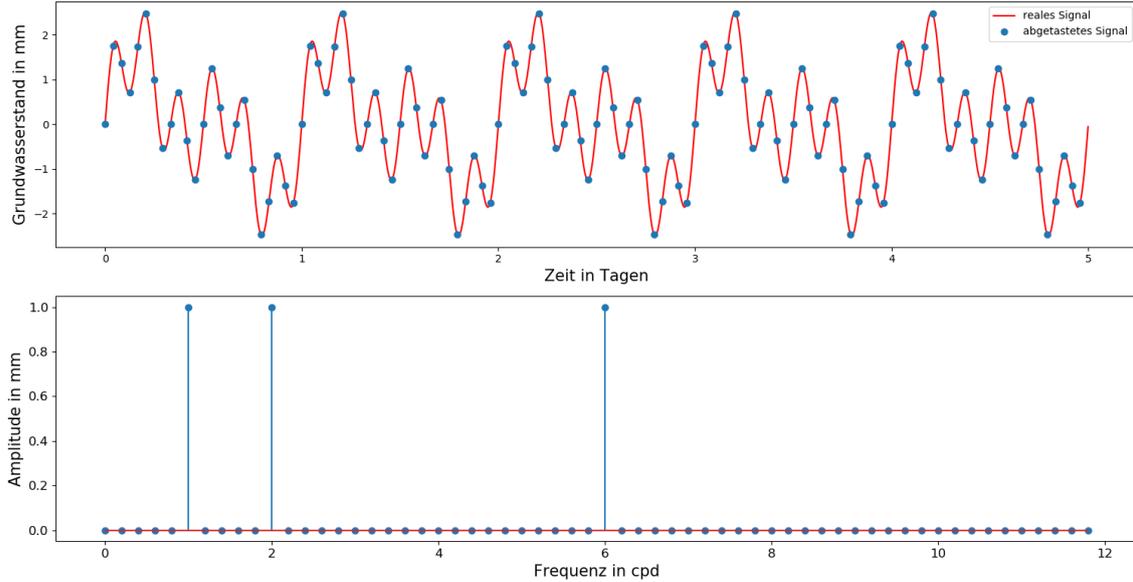


Abbildung 2.1: Abgetastetes Signal aus drei Sinusschwingungen mit dazugehörigem Frequenzspektrum

Durch die Diskretisierung der Frequenzen lassen sich Frequenzen zwischen zwei Bins im Spektrum nicht eindeutig identifizieren. Die Energie einer solchen Signal-komponente verteilt sich auf die benachbarten Frequenzbins. Dies wird als Leakage-Effekt bezeichnet (Buttkus, 2000).

Bei der Fourier-Transformierten Y_k handelt es sich um eine komplexe Zahl, aus deren Betrag sich das Amplitudenspektrum ergibt. Daraus ergeben sich schließlich die Amplituden der Komponenten des Signals. Diese lassen sich als Y-Achsen-Wert der Höchstwerte im Amplitudenspektrum ablesen. In dieser Arbeit wird der Einfluss auf die Amplitude untersucht; die Phase bleibt außer Betracht.

Die Rücktransformation der DFT lautet

$$y_n = \frac{1}{N} \sum_{k=0}^{N-1} Y_k e^{i2\pi k \frac{n}{N}}. \quad (2.3)$$

Damit lässt sich ein diskretes Messsignal y_n aus seinem Frequenzspektrum rekonstruieren. Dabei ist das Abtasttheorem zu beachten: Es besagt, dass bandbegrenzte Signale (mit Grenzwert B) mit ihrer Fourier-Transformierten

$$Y(f) = 0 \quad \text{mit} \quad |f| \geq B, \quad B > 0 \quad (2.4)$$

anhand der abgetasteten Werte $y(t_n)$ vollständig rekonstruierbar sind (Puente León, 2019). Da im Kontext der GM nur Signale mit der endlichen Anzahl Frequenzen des Gezeitenkatalogs (Abbildung 1.3) zu betrachten sind, trifft das Abtasttheorem auf die zu untersuchenden Messsignale zu.

Grundwasserstands-Messreihen sind zeitdiskrete, nicht periodische Signale einer bestimmten Länge. Außerhalb des Zeitraums T , in dem die N Abtastwerte aufgenommen wurden, wird das Signal bei der Auswertung mittels DFT periodisch fortgesetzt. Dadurch entsteht ein diskretes Frequenzspektrum. Dies ist Voraussetzung für eine weitere digitale Verarbeitung bei endlichem Speicherplatz (Puente León, 2019).

Um Sprünge zwischen zwei aufeinander folgenden Perioden zu vermeiden, wird das Signal mit einem Hanningfenster multipliziert, bevor die Transformation in den Fourierraum stattfindet (Abbildung 2.2). Dieses gewichtet die Signalwerte so, dass die Zeitreihe an ihren Rändern abflacht, um eben in die nächste Periode überzugehen. Dies geschieht mithilfe einer Kosinusfunktion, die den n -ten Wert des Signals aus M Werten durch

$$w(n) = 0,5 - 0,5 \cos\left(\frac{2\pi n}{M-1}\right) \quad 0 \leq n \leq M-1 \quad (2.5)$$

gewichtet (NumPy, 2017; Harris, 1978).

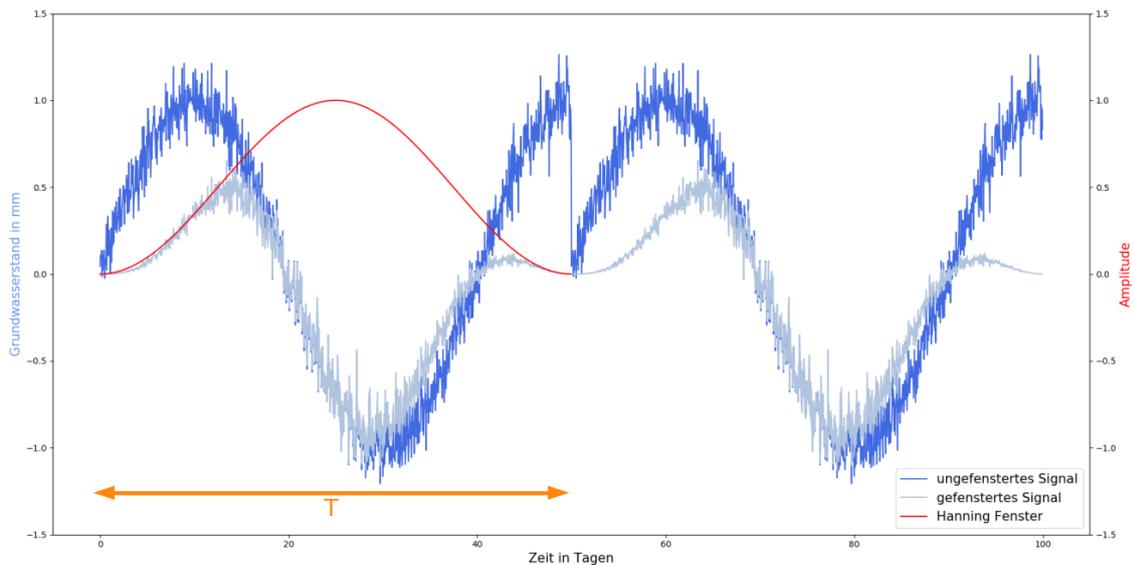


Abbildung 2.2: Durch die Hanning-Fensterfunktion verringern sich die Sprünge zwischen letztem und erstem Messwert aufeinanderfolgender Perioden

Zur Berechnung der Amplitudenspektren dient die Numpy-Funktion `np.fft.fft`. Diese berechnet nach Formel 2.1 das komplexe Frequenzspektrum einer Messreihe. Durch die Betragsfunktion `np.abs` lässt sich dieses in ein reelles Amplitudenspektrum überführen.

2.1.2 Die Methode der kleinsten Quadrate

Die Methode der kleinsten Quadrate (MdkQ) ist ein Verfahren der Ausgleichsrechnung. Ziel ist es, die Summe der Quadrate der Abstände (Residuen) zwischen Messdaten und der Anpassungsfunktion zu minimieren (Abbildung 2.3). Dazu variiert ein Algorithmus bestimmte Parameter in der modellierten Funktion, bis die bestmögliche Anpassung erreicht ist.

Je nach Fragestellung sind verschiedene Ansätze für die modellierte Funktion sinnvoll. In diesem Fall ist das Modell eine Summe von harmonischen Schwingungen mit den Frequenzen f_i des Gezeitenkatalogs (Abbildung 1.3), um der periodischen Natur der Schwingung gerecht zu werden:

$$y(t) = \sum_{i=0}^N [A_i \sin(2\pi f_i t) + A_i \cos(2\pi f_i t)]. \quad (2.6)$$

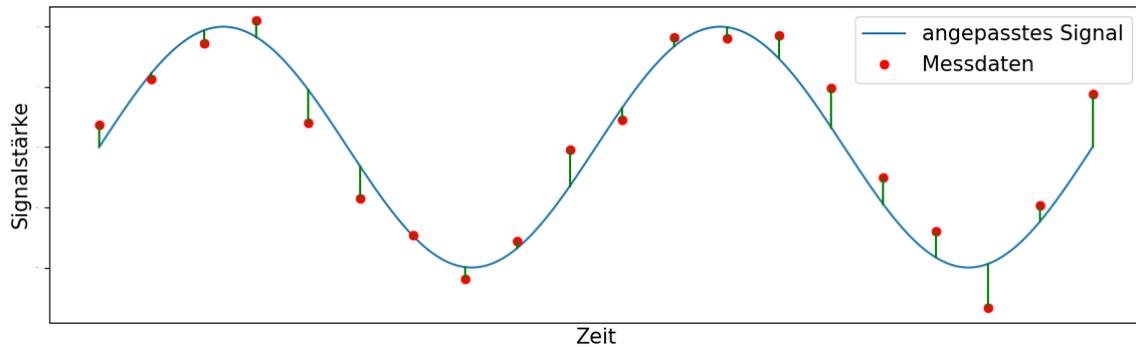


Abbildung 2.3: Durch eine Modellfunktion (blau) angepasste Messwerte (rot): Die Residuen (grün) sollten durch die Wahl einer geeigneten Modellfunktion minimiert werden.

Damit hat die MdkQ einen Freiheitsgrad weniger als die FT: Die Frequenzen sind festgelegt. Die variierenden Parameter sind nur die Amplituden A_i . Diese sind so zu wählen, dass minimale Fehlerquadrate zwischen Modell und Realdaten entstehen.

Für die Berechnungen steht die Funktion `lsp_estimate` von Prof. Peter Stoica der Uppsala University und Jonathan Eric Tuck der Stanford University zur Verfügung (siehe Anhang). Diese verwendet die Numpy-Funktion `np.linalg.lstsq`, um Amplituden von harmonischen Sinusfunktionen mit den kleinsten Residuen bezüglich einer gegebenen Messreihe zu ermitteln. Da `np.linalg.lstsq` lineare Gleichungssysteme löst, muss zunächst eine Matrix Φ aus Sinus- und Kosinustermen der Messzeitpunkte t_i und Frequenzen f_{M_2} und f_{S_2} erstellt werden:

$$\Phi = \begin{pmatrix} \sin(f_{M_2}t_0) & \cos(f_{M_2}t_0) & \sin(f_{S_2}t_0) & \cos(f_{S_2}t_0) \\ \vdots & \vdots & \vdots & \vdots \\ \sin(f_{M_2}t_{N-1}) & \cos(f_{M_2}t_{N-1}) & \sin(f_{S_2}t_{N-1}) & \cos(f_{S_2}t_{N-1}) \end{pmatrix} \quad (2.7)$$

Damit kann schließlich die lineare Gleichung

$$\Phi x = y \quad (2.8)$$

mit den Messdaten y nach x aufgelöst werden, sodass $|y - \Phi x|^2$ minimal ist.

Als Eingangsparameter benötigt die Funktion `lsp_estimate` somit eine Zeitreihe y , an die die modellierte Funktion angepasst werden soll, einen Zeitvektor t und die Frequenzen f_{M_2} und f_{S_2} . Da synthetische Messreihen nur aus Sinustermen bestehen, entfallen dort die Koeffizienten der Kosinustermine.

2.2 Synthetische Messreihen

Es zu untersuchen, wie sich Länge der Messreihe, die Auflösung des Drucksensors und Datenlücken auf das rekonstruierte Signal auswirken. Diese sind von Messreihe zu Messreihe unterschiedlich. Neben dem Signal-Rausch-Verhältnis sind sie maßgeblich für die Qualität der Daten und damit für die Genauigkeit der Ergebnisse.

Die Python 3 Skripte aus dem Anhang unter Skripte/Synthetische Daten erstellen synthetische Daten. Deren Vorteil ist, dass alle Parameter modifizierbar sind und

die Effekte der Messparameter isoliert von weiteren Einflüssen an den Gezeitenkomponenten M_2 und S_2 betrachtet werden können.

Üblicherweise werden Proben im konstanten Zeitintervall entnommen. Dies hat praktische Gründe, da dem Messgerät so neben der aktuellen Uhrzeit nur die Samplingrate als Parameter für die gesamte Messung übergeben werden muss. Um möglichst praxisnah zu bleiben, betrachtet diese Arbeit daher nur Signale mit zeitlich gleichförmiger Probennahme, also mit konstanter Samplingrate.

Zur Vereinfachung bestehen die synthetisierten Daten aus lediglich zwei harmonischen Schwingungen, deren Frequenzen nah beieinander liegen: M_2 und S_2 . Die daraus folgenden Erkenntnisse lassen sich analog auch auf Signale mit einer höheren Anzahl an Komponenten anwenden.

Die Amplituden der Komponenten S_2 und M_2 sind besonders hoch (Abbildung 1.3), was sie charakteristisch für Grundwassermesssignale macht. Jacob (1940) demonstrierte außerdem, dass sich die barometrische Effizienz aus dem Verhältnis der Amplituden von M_2 und S_2 abschätzen lässt. Dadurch sind diese Komponenten gute Repräsentanten für die zur Gezeitenmethode relevanten Signalanteile.

Die hier betrachteten synthetischen Signale bestehen aus den harmonischen Anteilen M_2 und S_2 sowie normalverteiltem Rauschen σ :

$$y(t) = A_{M_2} \sin(2\pi f_{M_2} t) + A_{S_2} \sin(2\pi f_{S_2} t) + \sigma. \quad (2.9)$$

Ein solches Signal ist in Abbildung 2.4 zu sehen.

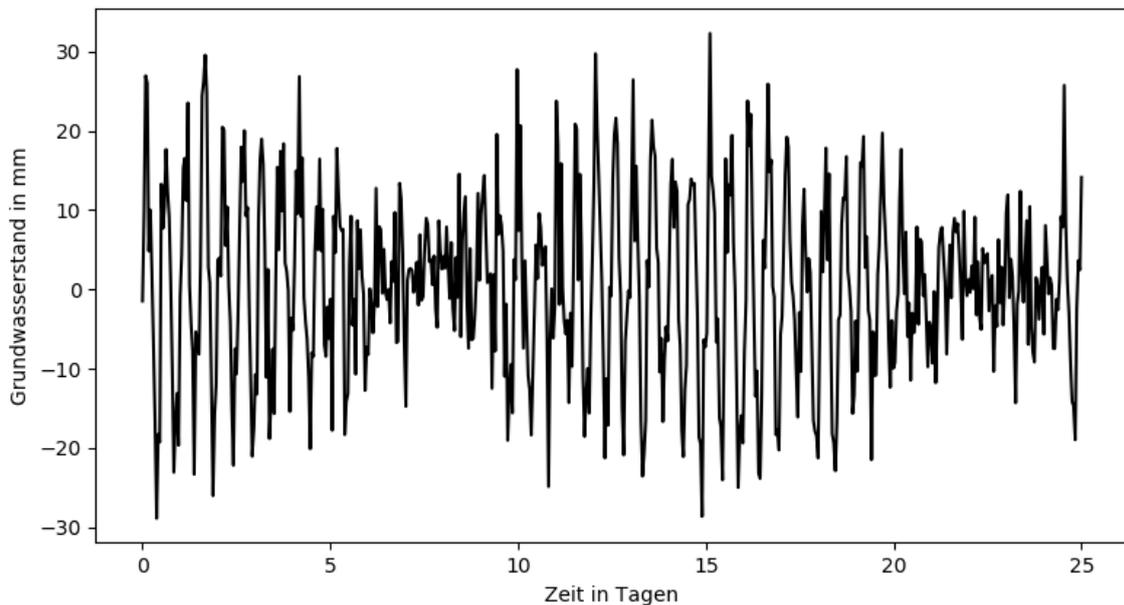


Abbildung 2.4: Nach Formel 2.9 modellierte Messreihe von stündlichen Messungen. Zur besseren Sichtbarkeit beträgt das Signal-Rausch-Verhältnis 10.

Das Rauschen ist der Teil eines Signals, der sich nicht durch eine mathematische Funktion eindeutig beschreiben lässt. Es variiert in unvorhersehbarer Weise und entsteht durch statistische Fehler in der Messung. Thermisches Rauschen, Rundungsfehler und andere äußere und innere Rauschquellen liefern einen zufälligen Beitrag zum Signal. Dies wird in den synthetischen Daten durch normalverteilte Zufallswerte σ simuliert.

Die Amplituden A_i betragen in den synthetisierten Signalen für beide Gezeitenkomponenten S_2 und M_2 10 mm. Verschiedene Einstellungen des Signal-Rausch-Verhältnisses (signal to noise ratio, SNR) sollen einen Überblick über unterschiedliche Datenvoraussetzungen geben.

In Tabelle 2.1 sind die Parameter der generierten Signale zusammengefasst. Die mit * gekennzeichneten Einträge variieren mit der Messdauer und gelten nur für die Untersuchung der Quantisierung und des Lückenanteils.

Tabelle 2.1: Übersicht der Modellparameter: Die mit * gekennzeichneten Parameter beziehen sich nicht auf die Versuchsreihe der Variation der Messdauer.

	Kürzel	S_2	M_2
Amplitude	A_i	10 mm	10 mm
Phase	φ	0 rad	0 rad
Abtastzeit	t_A	1 h	
Messdauer*	T	456 d	
Anzahl Messwerte*	N	8 760	
Signal-Rausch Verhältnis	SNR	∞ , 2 und 0,5	

Die Abtastzeit t_A beträgt eine Stunde, was ein in der Praxis gängiger Wert ist. Damit lassen sich entsprechend des Nyquist-Shannon-Theorems (vgl. (Puente León, 2019)) Frequenzen von maximal

$$f_N = \frac{1}{2t_A} = 12 \text{ cpd} \quad (2.10)$$

auffösen. Dies ist deutlich größer, als die nach Abbildung 1.3 erwarteten Komponenten, die alle unter 3 cpd liegen. Bei einem Abtastintervall von 4 Stunden (1/6 Tag), wie es als maximalwert von Acworth et al. (2015) angegeben wird, beträgt die Nyquistfrequenz $f_N = 3 \text{ cpd}$. Die FT kann bis zu dieser Frequenz Amplituden im Spektrum ermitteln. Die für die GM interessanten Frequenzen sind somit bei stündlicher Abtastung im Spektrum enthalten.

Die Messdauer des synthetischen Datensatzes beträgt 456 Tage. Da dies annähernd ein gemeinsames Vielfache von T_{M_2} und T_{S_2} ist, reduziert diese Messdauer den Leakage-Effekt für diese Komponenten. T_{M_2} und T_{S_2} bezeichnen die Perioden der Komponenten M_2 und S_2 .

Durch den Rauschanteil unterscheiden sich die synthetisierten Signale voneinander: Bei jedem Durchlauf des Programms ist der Schwingung ein neuer, zufälliger Anteil σ überlagert. Es bietet sich deshalb eine statistische Betrachtung an, da bei einer hohen Anzahl Durchläufe erkennbare Muster zu erwarten sind. Die folgende Untersuchung generiert stets 1 000 synthetisierte Signale pro Parametereinstellung. Sie betrachtet die Mittelwerte der berechneten Amplituden.

Da durch diese statistische Betrachtung sehr viele Berechnungen anfallen, wurde auf eine effiziente Softwarearchitektur geachtet: Möglichst viele Befehle sind in den erstellten Programmen außerhalb der Schleifen implementiert, um so Rechenaufwand und -zeit zu verkürzen.

Die Amplituden der rekonstruierten Signale sind das Kriterium zur Beurteilung der Qualität der Modellsignale: Der Abstand zur Amplitude des Ursprungssignals (jeweils 10 mm für M_2 und S_2) ist der Fehler des Modells. Dies gilt für die FT genauso wie für die MdkQ bei allen zu untersuchenden Parametern.

2.3 Reale Messdaten

Die Regressionsanalyseverfahren werden auch auf reale Daten angewendet. Dazu dient ein Beispieldatensatz, den bereits McMillan et al. (2019) analysierten (siehe Anhang baldry.csv, BH1). Im Folgenden wird diese Messreihe als *Realdaten* bezeichnet, um sie von den synthetisch erzeugten Daten abzugrenzen.

Die Realdaten sind in Abbildung 2.5 als Funktion der Zeit dargestellt. Die Daten wurden zwischen dem 24.10.2003 um 0:00 UTC und dem 15.02.2016 um 20:00 Uhr UTC aufgenommen. Die Messdauer beträgt mehr als 12 Jahre (4497 Tage) bei permanenter, stündlicher Probennahme. Insgesamt bestehen die Realdaten somit aus 107948 Werten. AHD ist ein australisches Höhenreferenzsystem (Australian Height Datum). Die Einheit der Funktionswerte in Abbildung 2.5 ist „Meter über AHD“. Die Auflösung des verwendeten Drucksensors beträgt 1 cm.

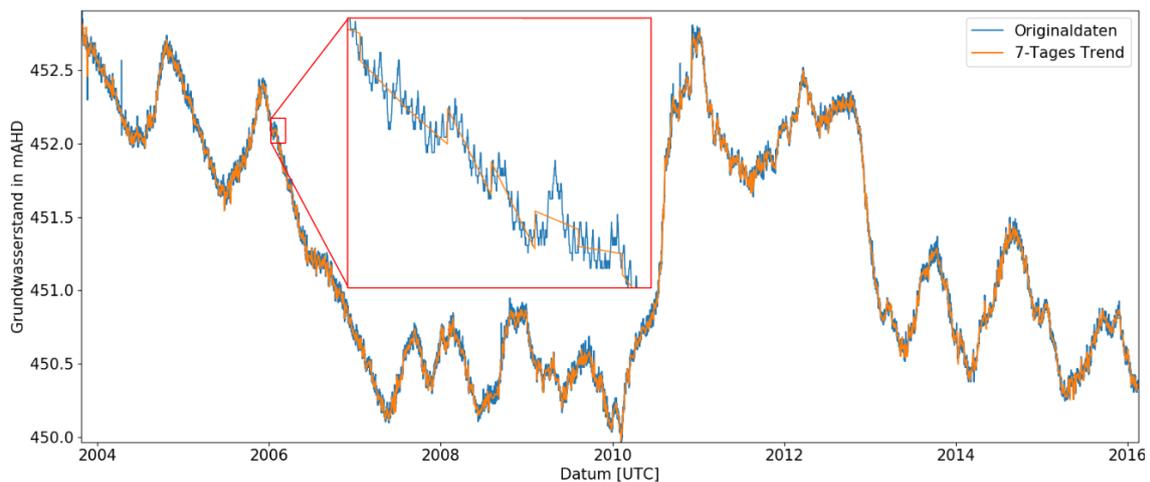


Abbildung 2.5: Datensatz einer Grundwassermessstelle in Baldry, Australien (bore hole, BH)(McMillan et al., 2019) (blau) und 7-Tages-Trend (orange)

Um langperiodische Anteile zu eliminieren, sind die Realdaten im Folgenden vom 7-Tages-Trend bereinigt. Dies wird im Folgenden als *Detrending* bezeichnet. Es entsteht dadurch die Zeitreihe aus Abbildung 2.6, auf die im Folgenden die Regressionsanalyseverfahren angewendet werden. So kann sichergestellt werden, dass hauptsächlich Signalkomponenten mit Frequenzen in der Größenordnung der Gezeitenkomponenten und höher im Spektrum zu sehen sind. Die Bereinigung vom 7-Tages-Trend hat somit einen ähnlichen Effekt wie ein Hochpassfilter.

In Abbildung 2.6 sind einige große Ausschläge des vom Trend bereinigten Grundwasserstands zum Beispiel um den Jahreswechsel 2003/2004 zu erkennen. Diese könnten aus der Aneinanderreihung der 7-Tage-Fenster entstehen: Da die Detrending-Fenster nicht ineinander übergehen, sondern die lineare Regression in aufeinander folgenden Abschnitten durchgeführt wird, entstehen Stufenartige Sprünge im Trend (vgl. orangefarbene Linie in Abbildung 2.5). Die Ausschläge aus Abbildung 2.6 repräsentieren somit nicht zwangsweise tatsächliche Schwankungen des Grundwasserspiegels, sondern könnten Artefakte des gewählten Detrending-Verfahrens sein. Da es bei dieser Arbeit nicht auf die Genauigkeit der Berechnung von Bodenparametern ankommt, kann ein Verlust an Datenqualität zugunsten von Rechenzeit in Kauf genommen werden.

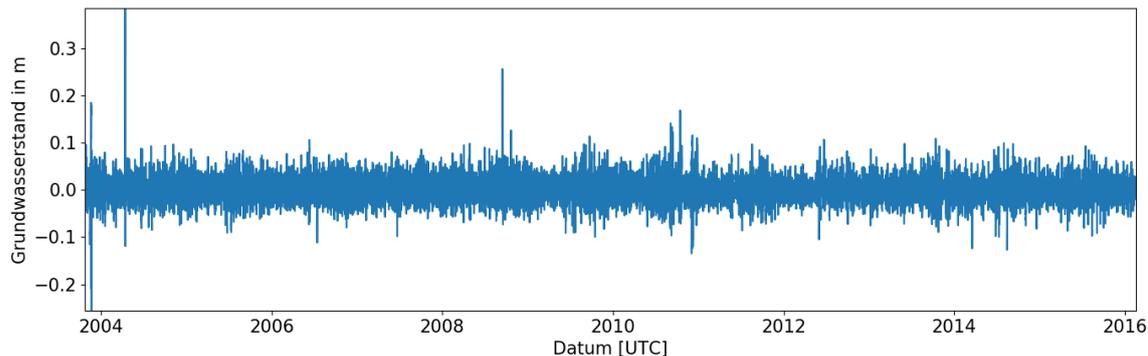


Abbildung 2.6: Vom 7-Tages Trend bereinigte Realdaten (Differenz vom orange-farbenen zum blauen Graphen aus Abbildung 2.5)

Wie bei den synthetischen Daten wird bei den Realdaten untersucht, wie sich die Amplituden von M_2 und S_2 verändern, wenn sich die (1) Messdauer ändert, (2) die Auflösung verschlechtert und (3) der Lückenanteil erhöht.

2.4 Einfluss verschiedener Parameter

2.4.1 Länge der Zeitreihe

Wie Formel 2.2 erkennen lässt, ist die Dauer T der Zeitreihe entscheidend für die Frequenzauflösung Δf bei der FT. Diese ist umgekehrt proportional zur Anzahl der Messwerte N und zum Abtastintervall t_A (vgl. Gleichung 2.2). Um M_2 und S_2 im Spektrum als Hochpunkte erkennen zu können, muss mindestens ein Wert dazwischen liegen. Die Frequenzauflösung Δf sollte also mindestens dem halben Frequenzabstand von S_2 und M_2 entsprechen:

$$\Delta f \leq \frac{1}{2} \Delta f_{M_2 S_2} = 0,033863 \text{ cpd} = \frac{1}{29,531 \text{ d}} \quad (2.11)$$

Mit wachsender Messdauer entstehen mehr Messwerte. Bei der Anwendung der MdkQ schränkt dies auch die Freiheitsgrade der modellierten Funktion ein, da mehr Datenpunkte berücksichtigt werden müssen. Der Effekt des Rauschens sollte bei einer größeren Anzahl Messwerte weniger ins Gewicht fallen, da sich dadurch der normalverteilte, simulierte Messfehler beim Anlegen der Modellfunktion herausmittelt.

Die Länge der synthetischen Messreihen lässt sich durch Erhöhen des Zeitvektors einfach verändern. Für die Realdaten werden zur Variation der Messdauer verschiedene Ausschnitte des kompletten Datensatzes betrachtet: zunächst die ersten 1000, dann jeweils zwei weitere Messwerte bis zum kompletten Datensatz. Mittels FT und der MdkQ werden daraus für jede betrachtete Länge die Amplituden von M_2 und S_2 ermittelt.

2.4.2 Auflösung des Drucksensors

Aufgrund der begrenzten Speicherkapazität von digitalen Messinstrumenten lassen sich Messwerte lediglich approximieren: Elektronische Geräte können nur endlich kleine Differenzen von Messwerten auflösen. Diese Differenz wird im Folgenden als

Auflösung oder *Quantisierung* des Signals bezeichnet. Die endliche Genauigkeit von eingespeicherten Werten führt zu Fehlern, die im Folgenden abzuschätzen sind.

Abbildung 2.7 veranschaulicht die Problematik: Bei einer Auflösung von 5 mm sind Stufen in der diskreten Zeitreihe aufgrund der Quantisierung deutlich zu erkennen. Je genauer die Quantisierung der Messwerte, desto kleiner werden die Abstände vom aufgezeichneten Messwert zur realen, physikalischen Größe und damit auch die Fehler. Dieser Effekt ist einer der Gründe, weshalb zwischen dem tatsächlichen Signal und den Messdaten zu unterscheiden ist: Kein Messgerät kann Zeitreihen fehlerfrei wiedergeben.

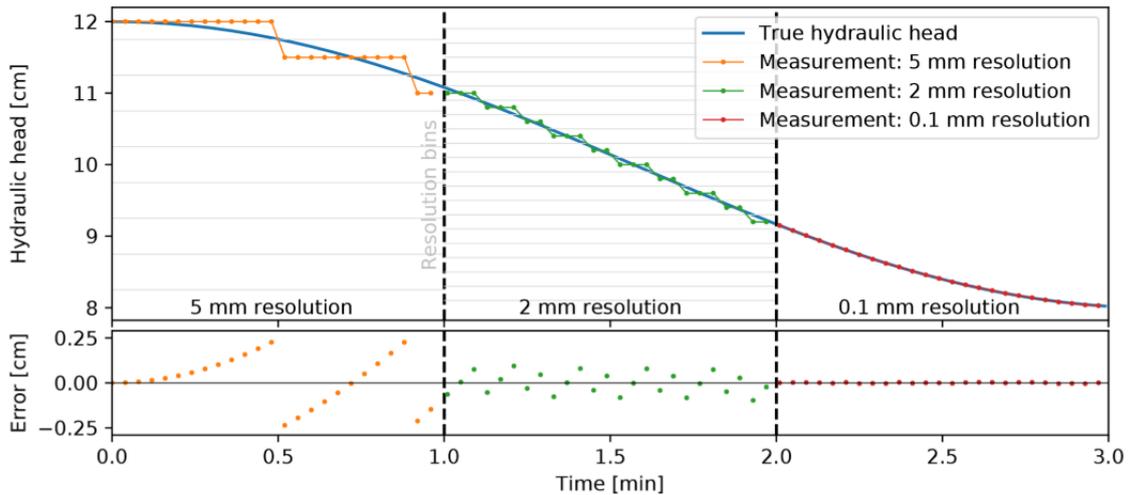


Abbildung 2.7: Quantisierung der Funktionswerte eines kontinuierlichen Signals (Rau et al., 2019): Je nach Auflösung des Messinstruments entstehen Stufen in den aufgezeichneten Daten.

Zur Beurteilung des Einflusses der Datenquantisierungen auf die Ergebnisse der Gezeitenmethode rundet das Python Programm mit der Funktion `quant` (vgl. Skript im abgedruckten Anhang) diese Signale auf Vielfache des eingestellten Quantums. Dies simuliert die beschränkte Auflösung des Messgerätes. Beispiele sind in Abbildung 2.8 gezeigt. Durch das Signalrauschen entstehen auch Werte über der Gesamtamplitude der synthetischen Daten von $A_{M_2} + A_{S_2} = 20$ mm.

Um möglichst praxisnah zu bleiben, wird die Auflösung der Realdaten vor dem Detrenden verändert. Von zunächst mit 1 cm genau gemessenen Daten wird Quantisierung der Daten erhöht und die Amplituden auch hier für jede Einstellung mit der MdkQ und der Fouriertransformation ermittelt.

Der Vergleich der Amplituden aus dem Amplitudenspektrum bzw. aus der MdkQ mit den ursprünglichen Amplituden des Modells lässt schließlich eine Beurteilung der angewendeten Methoden zu.

2.4.3 Datenlücken

Bei Wartungsarbeiten, Fehlfunktionen des Gerätes oder sonstigen Störungen können Lücken in Zeitreihen entstehen. Der in Abbildung 2.9 gezeigte synthetische Datensatz weißt konsequente Lücken auf, wie sie zum Beispiel bei Ausfall des Messgerätes entstehen. In dieser Arbeit werden wegen der einfacheren Programmierung und

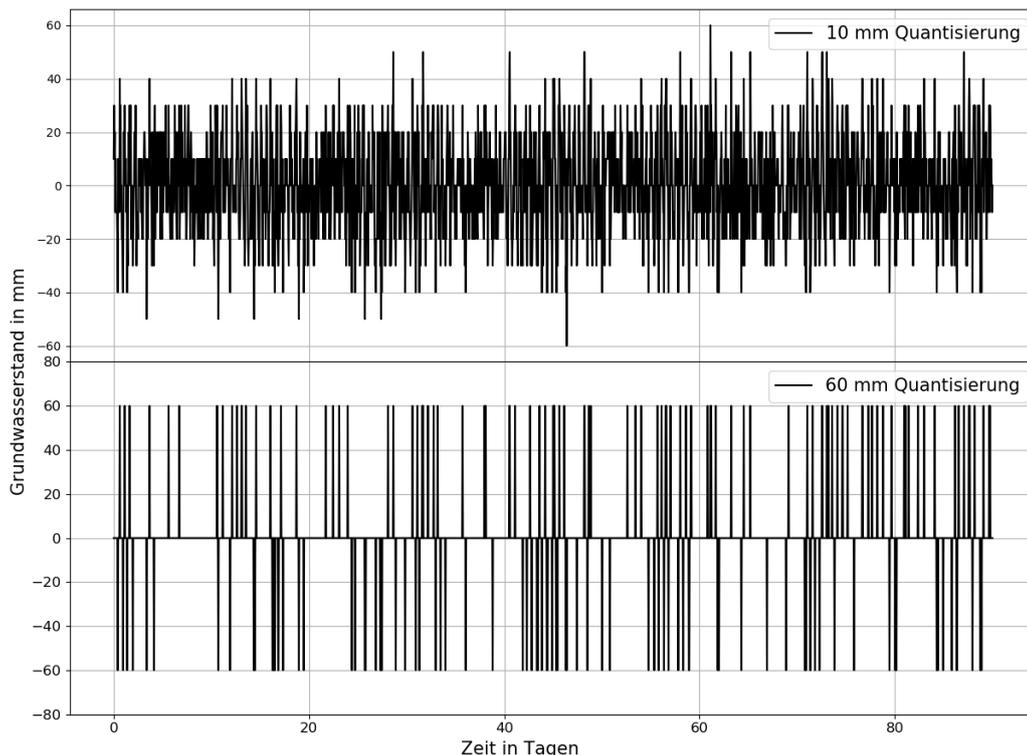


Abbildung 2.8: Quantisierungen eines modellierten Ursprungssignals mit Frequenzen f_{M_2} und f_{S_2} und jeweils einer Amplitude von 10 mm.

Handhabung Lücken an zufällig erstellten Stellen betrachtet. Diese liegen nicht unbedingt hintereinander.

Lücken stellen besonders für die DFT ein Problem dar, da dort ein konstantes Abtastintervall Voraussetzung ist (Stahn and Gizon, 2008). Für die Prozessierung der synthetischen Daten werden die Datenlücken durch Nullen ersetzt, dem Mittelwert des Signals. Auch bei den Realdaten werden Nullen für die Lücken eingesetzt, bevor die DFT angewendet wird. Dies hat den Effekt, dass die ermittelten Amplituden von M_2 und S_2 sinken und nicht durch die Interpolation mit von Null verschiedenen Werten verstärkt werden. Dadurch wird ein leicht vorhersehbares Verhalten der Komponenten mit steigendem Lückenanteil sichergestellt.

Durch das Ersetzen mit Nullen wird außerdem der bei großen Datenmengen nicht zu vernachlässigende Rechenaufwand, der z.B bei linearer Interpolation entsteht, vermieden. Das Auffüllen der Lücken mit Nullen ist somit eine einfache und schnelle Methode, den Datensatz auf die DFT vorzubereiten. Sicher sind aber auch weitere Umgangsformen mit den Datenlücken denkbar.

Für die MdkQ ist das Füllen der Datenlücken mit Werten nicht notwendig, da diese auch unvollständige Signale anpassen kann. Diesem Regressionsanalyseverfahren kann der lückenhafte Datensatz somit direkt übergeben werden.

Lücken werden mithilfe der Python-Funktion `lueckidx` erstellt: An einer einstellbaren Anzahl Indizes werden im Datensatz Lücken generiert. Zur Erstellung der Datenlücken wird die Funktion `random.sample` aus dem Paket `random` in Python importiert. Diese ermöglicht, einen beliebig langen Vektor verschiedener Einträge aus einem größeren Vektor zu entnehmen. Der entnommene Vektor gibt die Stellen

an, an denen in der Zeitreihe die Lücken entstehen.

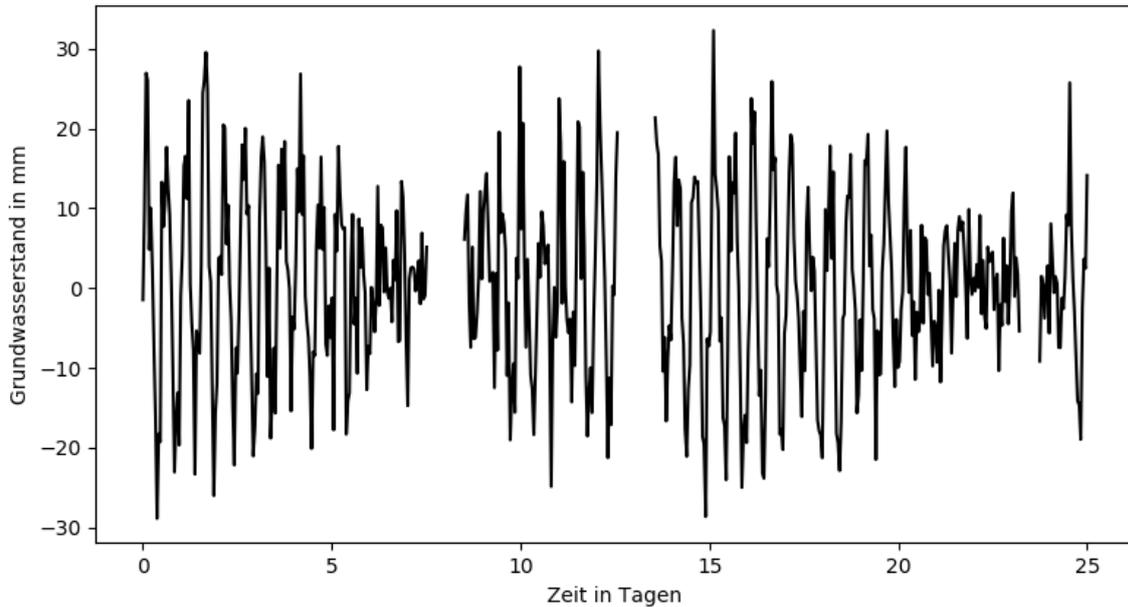


Abbildung 2.9: Ausschnitt einer nach Formel 2.9 simulierten Messreihe mit Datenlücken

Von Datenlücken kann nur gesprochen werden, wenn die Samplingrate konstant ist, also eine gleichförmige Probennahme stattfindet. Dies rechtfertigt die Beschränkung auf Signale mit konstanter Samplingrate. Außerdem darf die Lücke nicht am Anfang oder Ende der Zeitreihe stehen. Anderenfalls ist sie nicht als solche zu erkennen, da man dies auch als erhöhtes Samplingintervall bzw. verringerte Messdauer interpretieren kann.

Lücken werden in den Realdaten generiert, bevor die das Detrending durchgeführt wird. Der daraus entstandene Datensatz unterscheidet sich leicht von dem aus Abbildung 2.6, da die lineare Regression trotz Datenlücken durchgeführt wird. Auf diesen werden die FT und die MdkQ angewendet, um die Amplituden von M_2 und S_2 zu bestimmen.

Die aus FT und MdkQ gewonnenen Amplituden werden schließlich mit den Amplituden verglichen, die von den Regressionsanalyseverfahren bei vollständigem Datensatz ermittelt werden.

2.5 Anwendung der Regressionsanalyseverfahren

Für beide Regressionsanalyseverfahren generieren die Programme jeweils Messreihen nach Gleichung 2.9. Die FT überführt das synthetische Signal zunächst ins Amplitudenspektrum. Durch das Rauschen entsteht dort neben den beiden Höchstwerten bei f_{M_2} und f_{S_2} ebenfalls Rauschen entlang des kompletten auflösbaren Spektrums (Abbildung 2.10). Bei den Realdaten werden die Regressionsanalyseverfahren direkt auf die Zeitreihe angewendet.

Bei Verwendung der FT sucht das Programm im Bereich zwischen 1,85 cpd und 2,1 cpd die beiden höchsten Maxima. Durch die Modellgleichung 2.9 ist zu erwarten, dass dies bei den Frequenzen f_{M_2} und f_{S_2} der Fall ist.

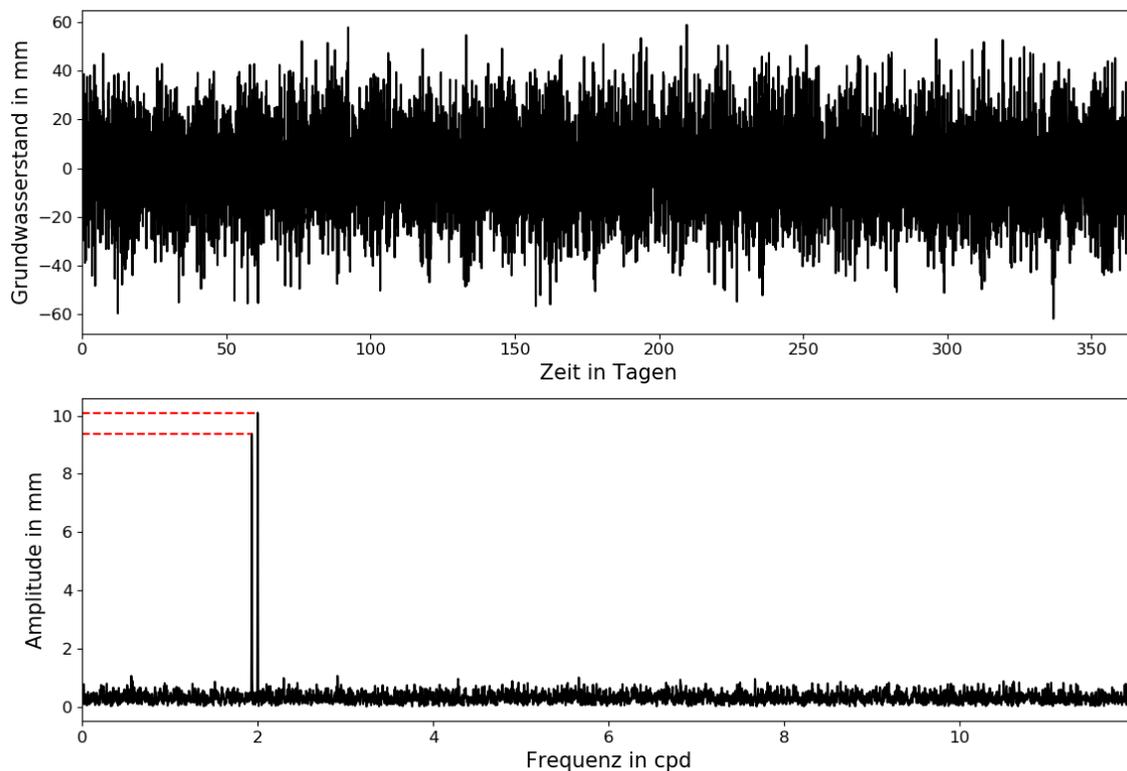


Abbildung 2.10: Beispiel des Amplitudenspektrums eines synthetischen Signals

Bei der Methode der kleinsten Quadrate gehen die Frequenzen f_{M_2} und f_{S_2} als Eingangsparameter in das Programm ein. Der Funktion `lsp_estimate` (siehe abgedruckter Anhang) von Peter Stoica und Jonathan Eric Tuck gibt unmittelbar die Amplituden aus, an denen die höchste Übereinstimmung mit dem Ursprungssignal erreicht ist.

Aufgrund des Rauschanteils im Signal werden sowohl der FT als auch die MdkQ auf die 1000 synthetisierten Signale angewendet. Der Mittelwert der resultierenden Amplituden ist schließlich mit denen des unveränderten synthetisch erzeugten Signals zu vergleichen. Bei den Realdaten liegen die Amplituden bei 1cm-Quantisierung und vollständigem Datensatz als Vergleichswert vor.

Kapitel 3

Ergebnis

3.1 Synthetische Datensätze

3.1.1 Fourier-Analyse

Länge der Zeitreihe

Nach Gleichung 2.11 ist eine Mindestmessdauer von $T_{min} = 29,531$ Tagen gefordert, um die gewünschte Frequenzauflösung von $\Delta f = 0,033863$ cpd zu erreichen. Um dies zu überprüfen, werden Messreihen verschiedener Längen zunächst ohne Signalauschen erstellt. Damit soll auch das Rauschen im Spektrum eliminiert werden, um den Einfluss der Messdaueränderung besser zu erkennen. Abbildung 3.1 zeigt die dazugehörigen Amplitudenspektren. Die Längen der modellierten Messreihen betragen 15, 30, 60 und 365 Tage.

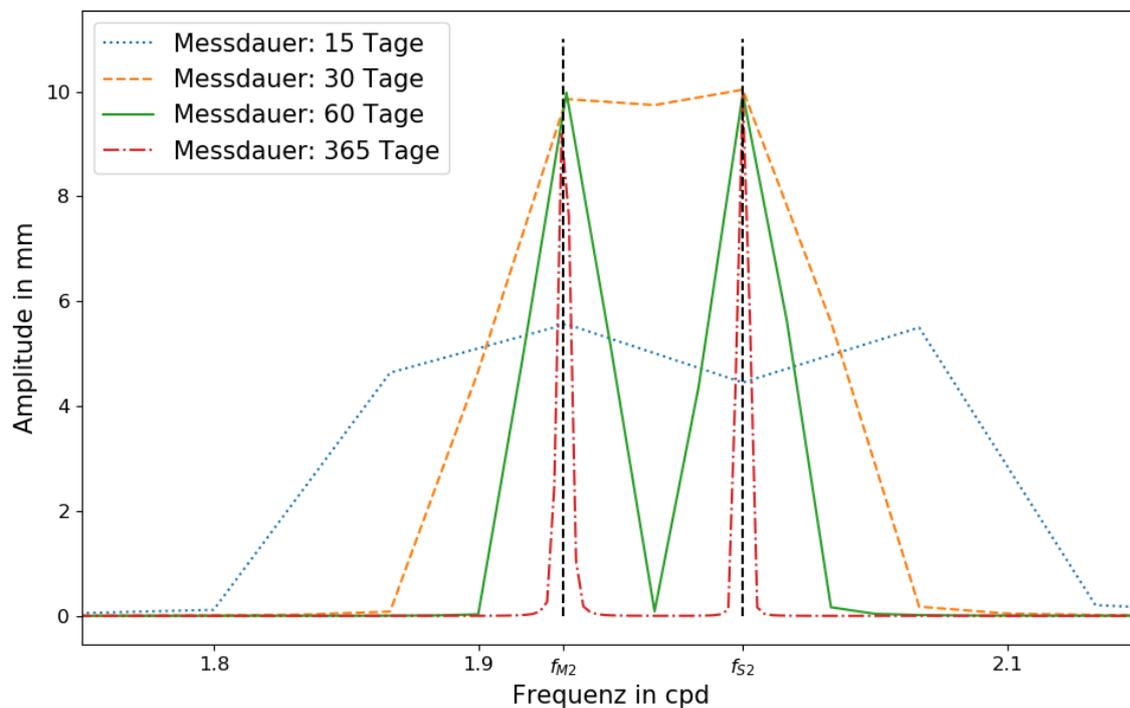


Abbildung 3.1: Amplitudenspektrum von Zeitreihen verschiedener Längen ohne Rauschen in der Umgebung von f_{M_2} und f_{S_2}

Das Ergebnis der statistischen Auswertung durch die FT ist in Abbildung 3.2 zu sehen: Dort sind die Mittelwerte Amplituden von M_2 (blau) und S_2 (schwarz) als Funktion der Messdauer aufgetragen.

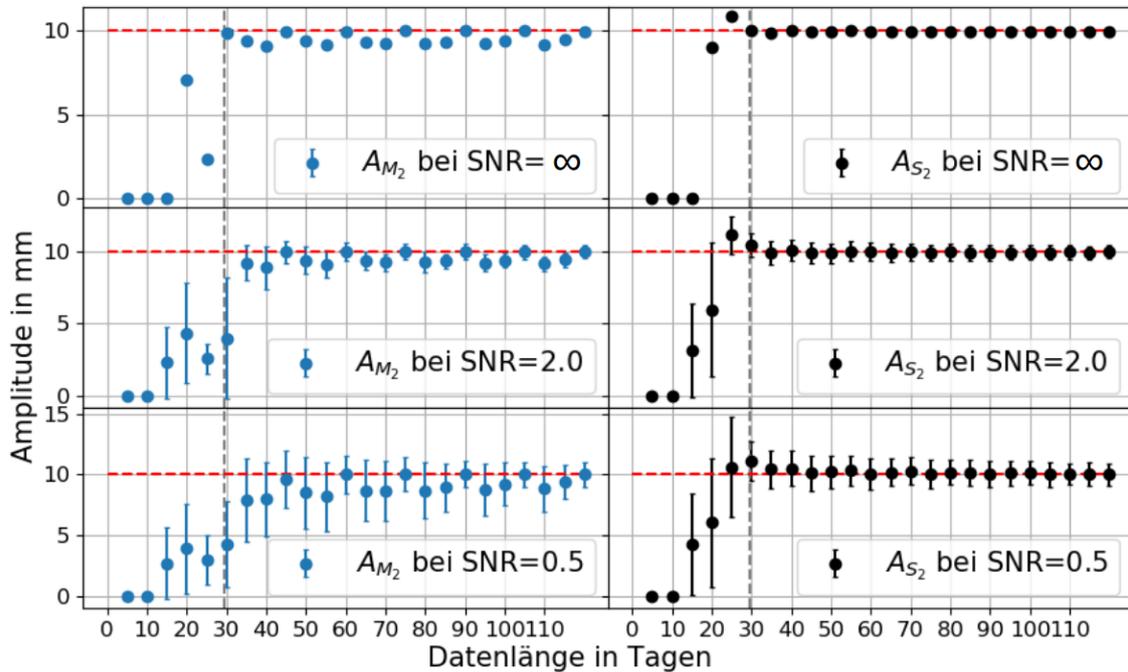


Abbildung 3.2: Einfluss der Länge der Messreihe auf die Amplituden der Komponenten M_2 (blau) und S_2 (schwarz) bei verschiedenen SNRs bei einer Auswertung mittels FT

Die Berechnungen wurden für jeweils SNR von ∞ , 2 und 0,5 durchgeführt, um den Idealfall, einen realistischen Fall und Messdaten mit hohem Rauschanteil abzudecken. Die Fehlerbalken zeigen die Standardabweichungen und die Punkte den Mittelwert der 1000 Durchläufe für jede Messdauer. Dies gilt für alle folgenden statistischen Analysen der synthetischen Daten.

Auflösung des Drucksensors

Die quantisierten Signale aus Abbildung 2.8 ergeben die in Abbildung 3.3 gezeigten Amplitudenspektren. Hierbei fällt auf, dass die Amplituden an den Frequenzen von S_2 und M_2 mit steigender Quantisierung absinken.

Wie bei der Untersuchung der Messdauer wurden für jede Quantisierung 1000 Signale generiert und jeweils das Amplitudenspektrum berechnet. Die Mittelwerte der Amplituden von M_2 (blau) und S_2 (schwarz) sind in Abbildung 3.4 gezeigt. Das Programm führt dies für SNRs von ∞ , 2 und 0,5 durch. Die Fehlerbalken zeigen die Standardabweichungen der 1000 Durchläufe.

Die Quantisierung zeigt erst bei Quanten $q > 20$ mm einen Effekt. Sobald q groß genug ist, sinken die Amplituden der Komponenten M_2 und S_2 . Hierbei scheint es, als ließen kleinere SNR höhere Quantisierungen zu.

Außerdem fällt ein Anstieg der ermittelten Amplituden bei einer Quantisierung um 20 mm kurz vor dem Abfall der Amplituden im Fall ohne Rauschanteil auf. Dieser Buckel tritt nicht für die Berechnungen mit Rauschen auf. Berechnungen mit

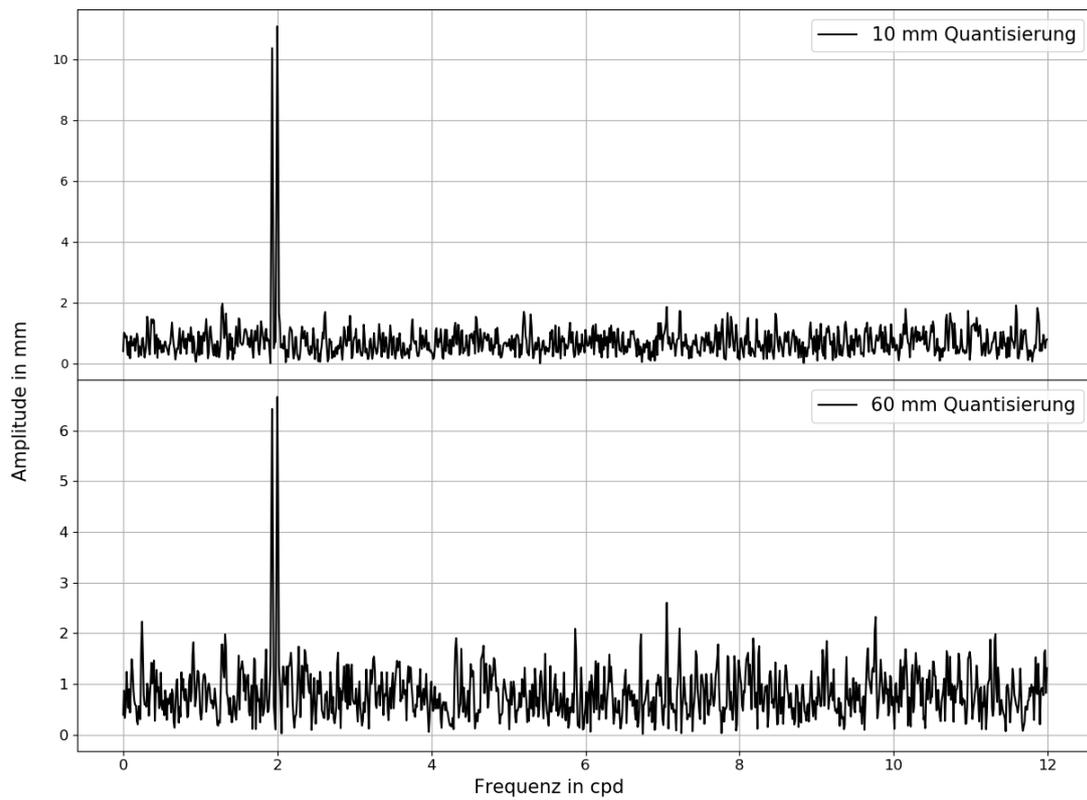


Abbildung 3.3: Amplitudenspektren der quantisierten Signale aus Abbildung 2.8

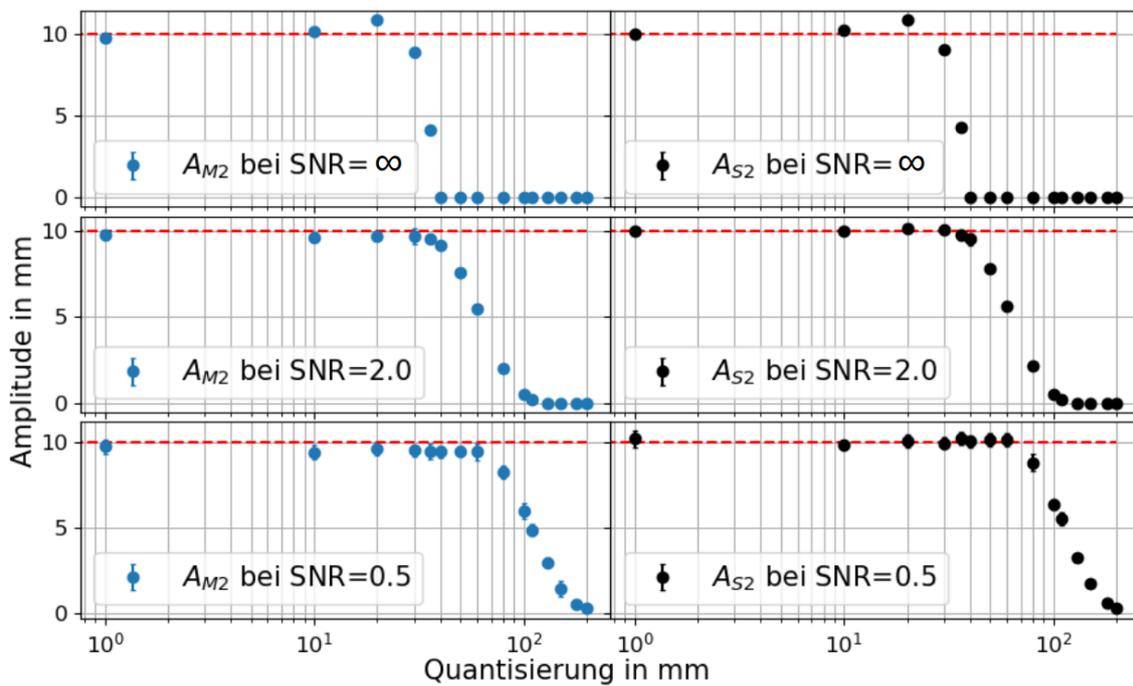


Abbildung 3.4: Einfluss der Quantisierung der Messwerte auf die Amplituden der Komponenten M_2 (blau) und S_2 (schwarz) bei verschiedenen SNRs bei einer Auswertung mittels FT

höherer Auflösung um 20 mm bestätigen, dass es sich hierbei nicht nur um einen einzelnen Ausreißer handelt.

Datenlücken

Die Untersuchung des Einflusses der Lücke ergibt Abbildung 3.5. Die Amplituden verringern sich scheinbar linear, wenn der Lückenanteil steigt. Die Standardabweichungen vergrößern sich bei schwachen SNRs.

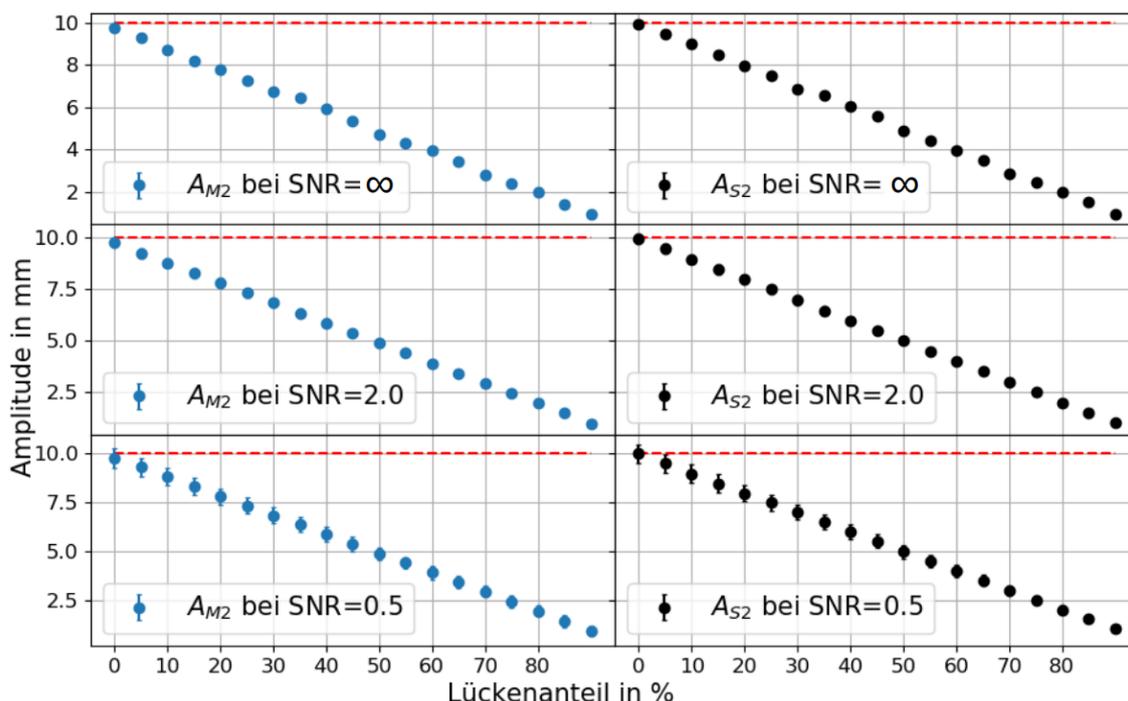


Abbildung 3.5: Einfluss des Lückenanteils auf die Amplituden der Komponenten M_2 (blau) und S_2 (schwarz) bei verschiedenen SNRs bei einer Auswertung mittels FT

3.1.2 Methode der kleinsten Quadrate

Länge der Zeitreihe

Abbildung 3.6 zeigt die Veränderung der gemittelten Amplituden der Schwingungskomponenten M_2 (blau) und S_2 (schwarz) mit der Länge der Messreihe. Die betrachteten Messdauern sind bei der MdkQ dieselben wie zuvor bei der FT in Abschnitt 3.1.1.

Die gemittelten Amplituden schwanken schon ab einer Messdauer von 10 Tagen mit nur noch $\pm 0,3$ mm um die Amplituden des synthetisierten Signals. Sogar für das geringe SNR von 0,5 sind die Unterschiede unter 3%. Die Abweichungen der Mittelwerte sowie die Standardabweichungen verringern sich mit steigender Messdauer.

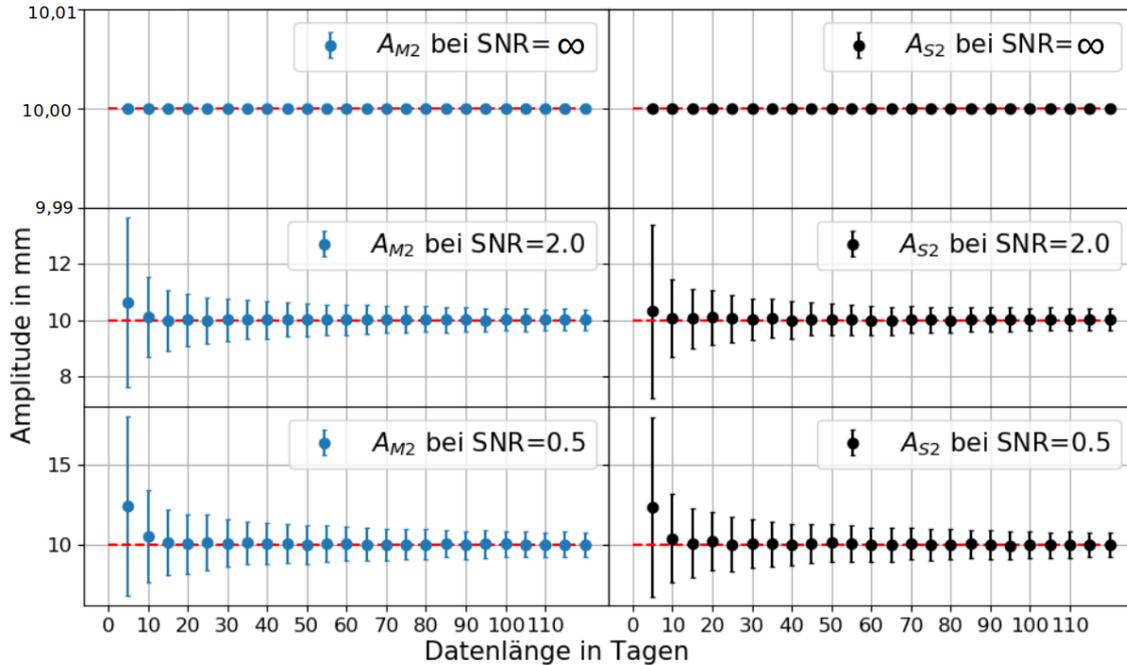


Abbildung 3.6: Einfluss der Länge der Messreihe auf die Amplituden der Komponenten M_2 (blau) und S_2 (schwarz) bei verschiedenen SNRs bei einer Auswertung mittels der MdkQ

Auflösung des Drucksensors

Das Ergebnis der Untersuchung der Datenquantisierung mittels MdkQ ist in Abbildung 3.7 zu sehen. Hier zeigt sich ein ähnlicher Verlauf wie bei der FT: Zunächst ermittelt die MdkQ für jede Quantisierung bis 1 mm eine Amplitude von $10 \pm 0,3$ mm. Bei hohen Quantisierungen sinken diese jedoch auf 0 ab. Je stärker der Rauschanteil, umso größer kann die Quantisierung sein, bevor die Amplitude verschwindet.

Auch hier ist ein leichter Anstieg der ermittelten Amplituden für $\text{SNR}=\infty$ bei 20 mm zu erkennen. Wie auch bei der FT sinken die ermittelten Amplituden danach auf 0 mm. Bei SNR verschieden von ∞ ist dieser Effekt nicht zu beobachten.

Datenlücken

In Abbildung 3.8 sind die Ergebnisse der Untersuchung des Einflusses des Lückenanteils bei Verwendung der MdkQ dargestellt. Im Gegensatz zur FT sind die Lücken nicht durch Nullen ersetzt, da die MdkQ auch lückenhaften Zeitreihen anpassen kann.

Die ermittelten Amplituden liegen für beide Komponenten bei unter 1% Abweichung bis zu einem Lückenanteil von 90%. Darüber werden im Mittel die 10 mm auch sehr gut getroffen, jedoch steigt die Standardabweichung ab 90% Lücken auf über ± 1 mm. Ab 99% Lückenanteil ist auch der Mittelwert der 1000 Durchgänge nicht mehr genau. Er weicht um mehr als 5% von den 10 mm ab.

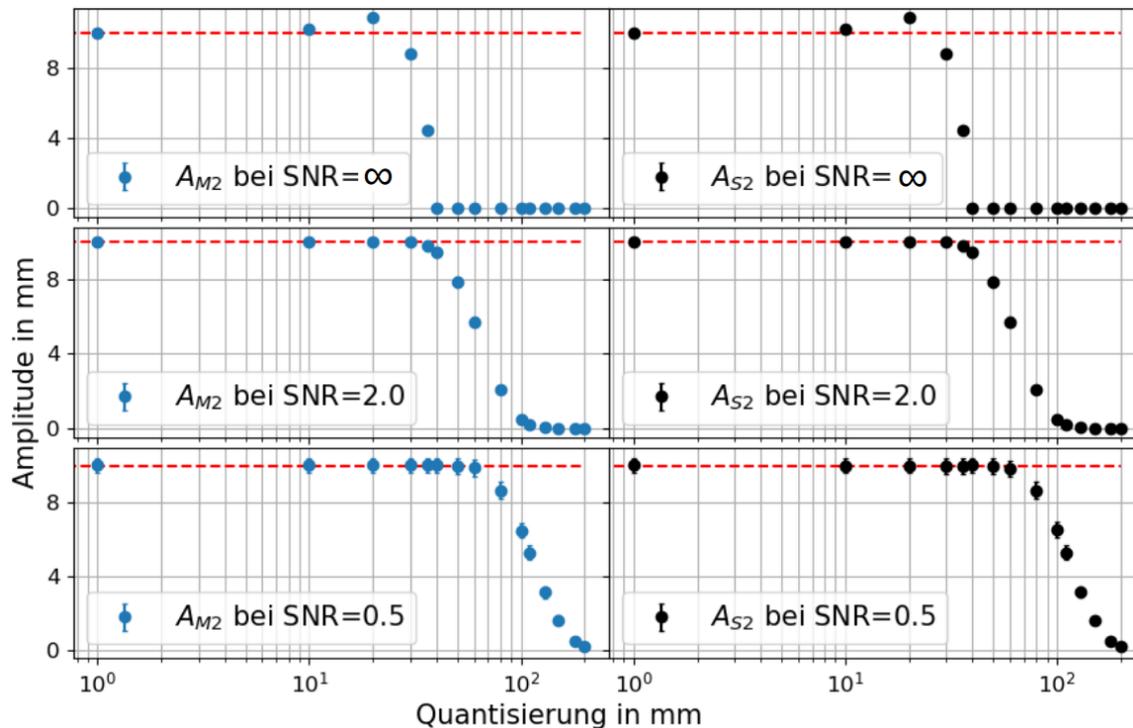


Abbildung 3.7: Einfluss der Quantisierung der Messwerte auf die Amplituden der Komponenten M_2 (blau) und S_2 (schwarz) bei verschiedenen SNRs bei einer Auswertung mittels MdkQ

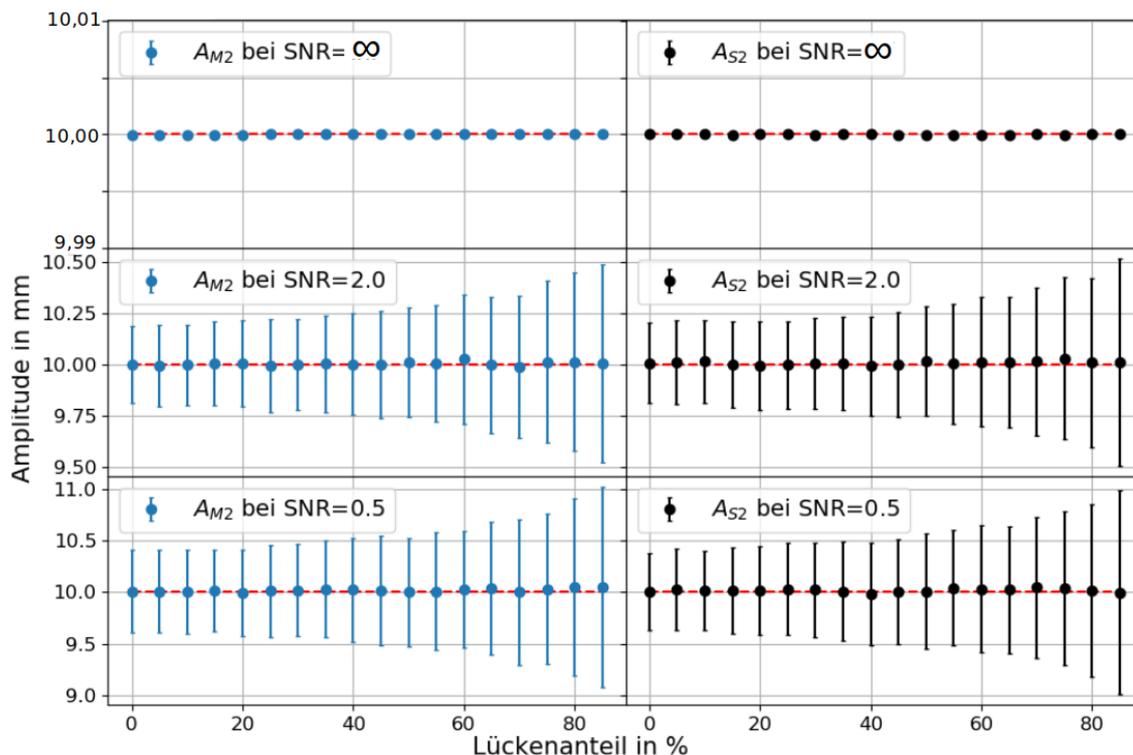


Abbildung 3.8: Einfluss des Lückenanteils auf die Amplituden der Komponenten M_2 (blau) und S_2 (schwarz) bei verschiedenen SNRs bei einer Auswertung mittels MdkQ

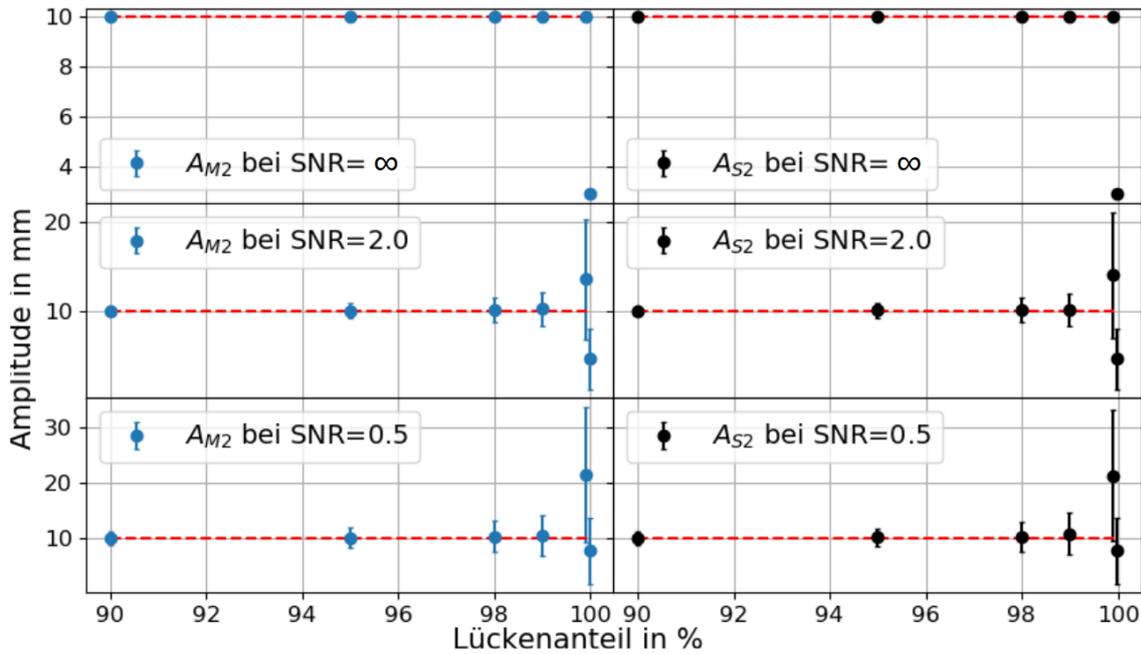


Abbildung 3.9: Einfluss des Lückenanteils bei über 90% auf die Amplituden der Komponenten M_2 (blau) und S_2 (schwarz) bei verschiedenen SNRs bei einer Auswertung mittels MdkQ

3.2 Realdaten

Das Frequenzspektrum der Realdaten (Abbildung 3.10) zeigt Hochpunkte bei 1 cpd und ganzzahligen Vielfachen davon. Die langperiodischen Anteile mit Perioden von über einer Woche sind durch das Herausrechnen des 7-Tages Trendes aus den Realdaten abgeschnitten. Diese saisonalen Signalkomponenten sind hier nicht von Interesse.

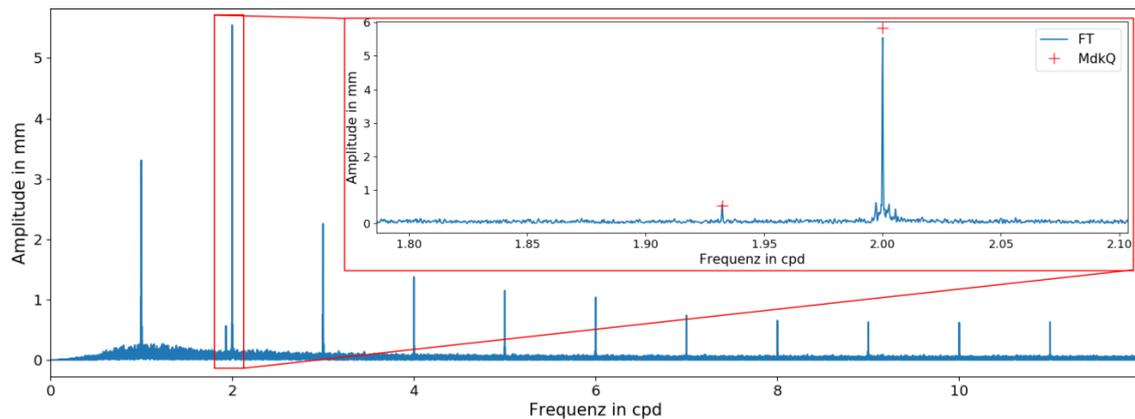


Abbildung 3.10: Frequenzspektrum der Realdaten

M_2 und S_2 sind im Spektrum gut zu erkennen: Ihre Amplituden betragen bei Berechnung mittels FT $A_{M_2} = 0,56$ mm und FT $A_{S_2} = 5,54$ mm. Dies zeigt sich auch in den SNRs: $SNR_{S_2} = 21578$ ist fast 10^2 Mal größer als $SNR_{M_2} = 223$. Die MdkQ erzielt ähnliche Ergebnisse: $A_{S_2} = 5,84$ mm ist dabei ebenfalls circa 10 Mal so hoch wie $A_{M_2} = 0,53$ mm.

Einfluss der Messdauer

Analog zur Untersuchung des Einflusses der Messdauer bei den synthetischen Daten wurden bei den Realdaten ebenfalls die Messdauer schrittweise erhöht und die Amplituden von M_2 und S_2 beobachtet. Das Ergebnis der zweistündlichen Erhöhung von 1000 Messwerten bis zum vollständigen Datensatz (107948 Werte) ist in Abbildung 3.11 zu sehen.

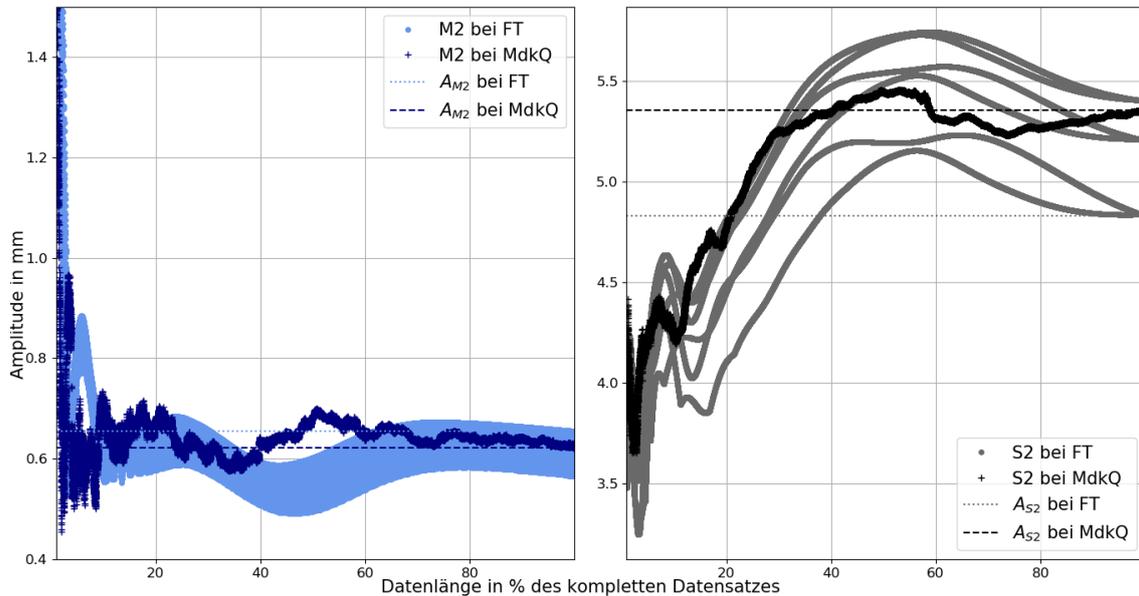


Abbildung 3.11: Amplituden der Komponenten M_2 und S_2 der Realdaten bei steigender Messdauer

Die durchgezogenen Linien stellen die ermittelten Amplituden aus FT und MdkQ dar, wenn der vollständige Datensatz verwendet wird. Bei M_2 ist ab einer Datenlänge von 10000 Messwerten (10 % des kompletten Datensatzes) die Amplitude des vollständigen Datensatzes sowohl für FT als auch MdkQ um $\pm 20\%$ genau getroffen. Insgesamt streuen die Werte aus der FT sehr stark (um $\pm 0,1$ mm ab 10000 Messdaten) und schwanken zudem noch um $\pm 0,1$ mm. Die Werte aus der MdkQ streuen hingegen nicht und schwanken ab 10 % des Datensatzes um $\pm 0,1$ mm.

Die S_2 -Komponente zeigt ähnliche Charakteristiken: Während die Ergebnisse der MdkQ keine Streuung aufweist, scheinen die Ergebnisse aus der FT sechs verschiedene Verläufe der Amplituden zu bilden.

Einfluss der Quantisierung

Abbildung 3.12 zeigt eine Abtastung des gleichen Signals mit größeren Auflösungen. Je größer das Quantum q , mit dem die Werte weiterverarbeitet werden, umso höher die Stufen in der Darstellung der Zeitreihe.

Die ermittelten Amplituden der Realdaten (Abbildung 3.13) zeigen bei steigender Quantisierung ein ähnliches Verhalten wie die der synthetischen Datensätze. Zunächst schwanken die Amplituden um den Wert ohne Quantisierung (jeweils die durchgezogene Linie), bis sie ab 10 cm Auflösung divergieren. Bei der M_2 Komponente ist dies schon ab 5 cm der Fall. Bis dahin schwanken die Amplituden um höchstens 10% um die Amplitude ohne Quantisierung.

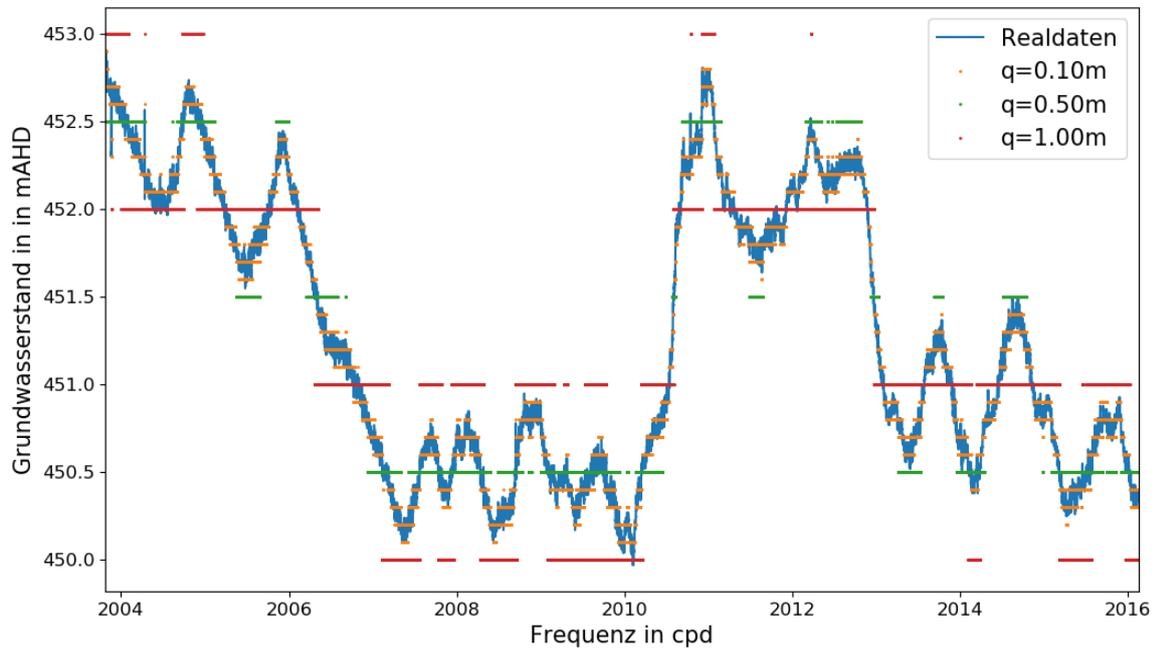


Abbildung 3.12: Quantisierungen der Realdaten

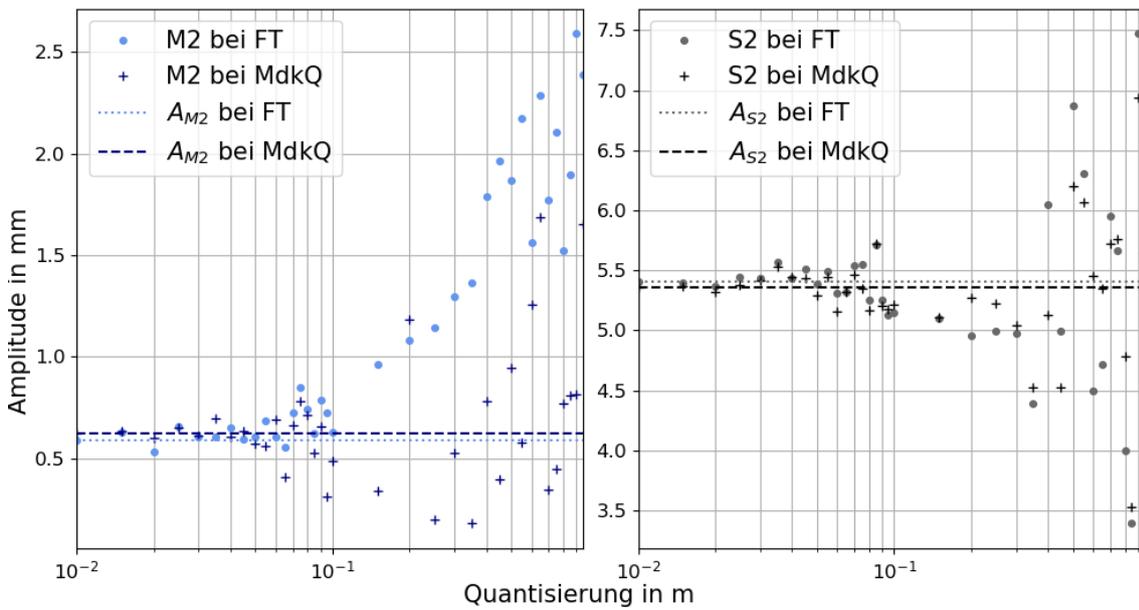


Abbildung 3.13: Amplituden von M_2 und S_2 der Realdaten bei steigender Quantisierung bei Auswertung mit FT und MdkQ

Bei den durch FT ermittelten Amplitude steigen die Amplituden von M_2 mit größerer Auflösung. Bei S_2 liegen die meisten Werte unter dem ursprünglichen Ergebnis. Die MdkQ liefert hingegen eine Verteilung mit wachsender Streuung bei steigendem q . Eine Vorzugsrichtung ist nicht zu erkennen: Es werden sowohl höhere als auch niedrigere Amplituden als A_{M_2} bzw. A_{S_2} ermittelt.

Einfluss des Lückenanteils

Wie auch bei den Realdaten werden die Lücken durch Nullen bei Auswertung mit der FT ersetzt. Es zeigt sich ein lineares Abfallen der Amplituden.

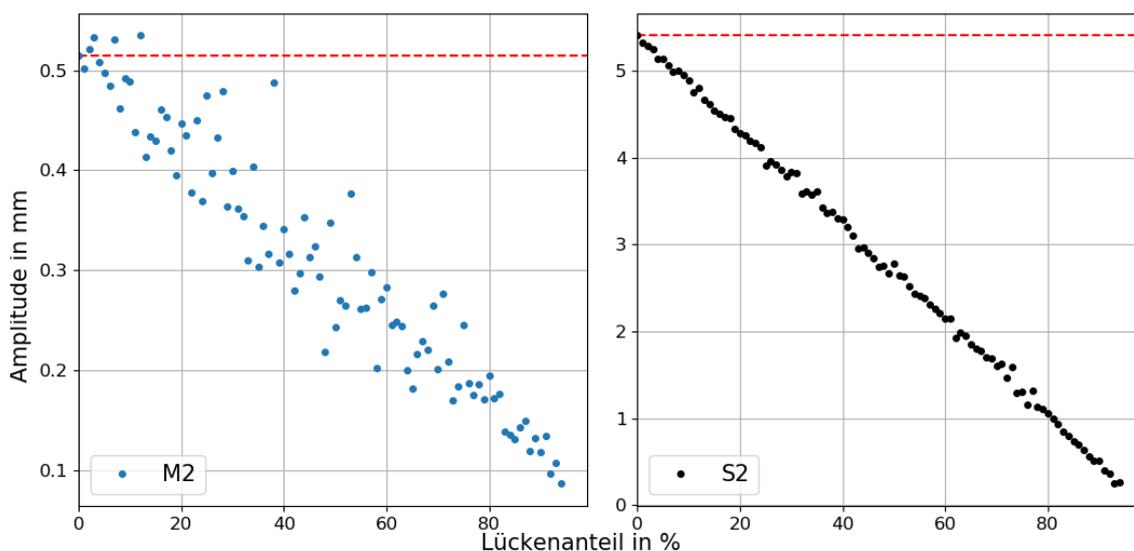


Abbildung 3.14: Durch FT ermittelte Amplituden der Realdaten bei steigendem Lückenanteil

Bei Auswertung mittels MdkQ streuen die ermittelten Amplituden zunehmend mit steigendem Lückenanteil um die Amplitude bei vollständigem Datensatz.

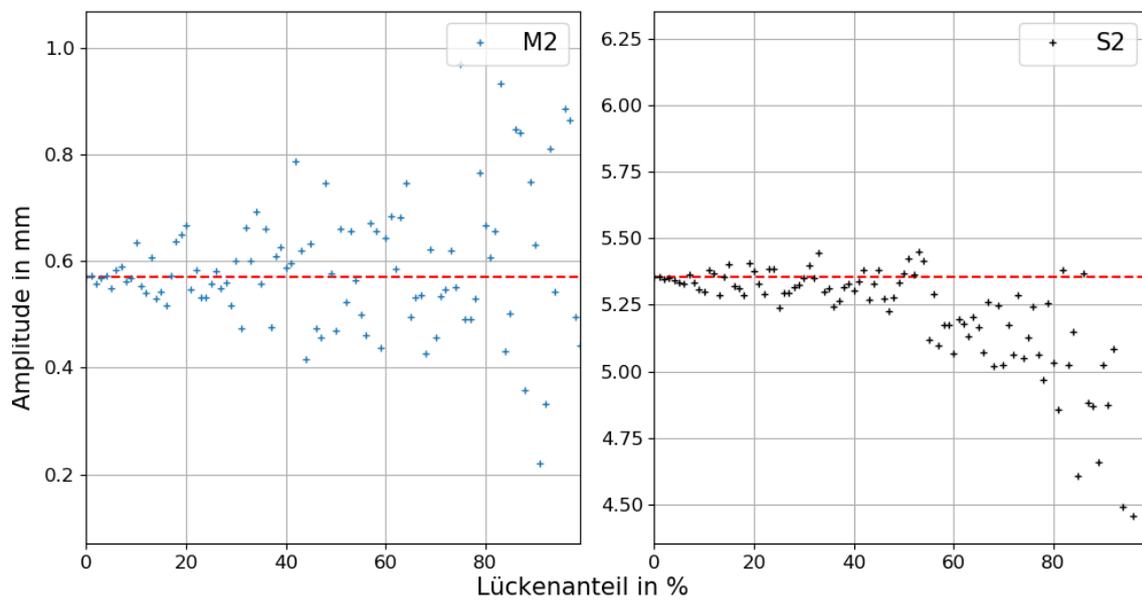


Abbildung 3.15: Durch MdkQ ermittelte Amplituden der Realdaten bei steigendem Lückenanteil. Die rote Linie kennzeichnet die Amplitude bei vollständigem Datensatz.

Kapitel 4

Diskussion

4.1 Der Einfluss der Messdauer

Für die GM sollte eine Mindestmessdauer von 30 Tagen erreicht sein.

Wie Abbildung 3.1 erkennen lässt, sind bei unzureichender Messdauer die Komponenten S_2 und M_2 des Ursprungssignals (ohne Rauschen) im Amplitudenspektrum nicht als verschiedene Hochpunkte zu erkennen. Der in Gleichung 2.11 geforderte Schwellenwert T_{min} , ab dem die Komponenten im Spektrum voneinander trennbar sind, konnte mittels synthetischer Messreihen bestätigt werden: Ist dieser nicht erreicht, verschwinden im Amplitudenspektrum die Hochpunkte für die Komponenten M_2 und S_2 (Abbildung 3.1). Verallgemeinert lässt sich daraus ableiten, dass eine Messreihe mindestens doppelt so lange sein muss wie das Inverse der gewünschten Frequenzauflösung Δf :

$$T > \frac{2}{\Delta f}. \quad (4.1)$$

Dies stimmt mit den Erkenntnissen aus Acworth et al. (2016) überein: Bei einer Messdauer von 15 Tagen entspricht die Frequenzauflösung Δf zwar ungefähr der Differenz von f_{M_2} und f_{S_2} , jedoch liegt bei Messdauern unter $T_{min} = 29,531$ Tagen kein Frequenzbin zwischen f_{M_2} und f_{S_2} , der diese als Frequenzhoch hervorhebt. Darum muss die Messdauer bei der Auswertung mit der FT die Mindestmessdauer überschreiten. Diese entspricht dem doppelten Frequenzabstand von zwei Signalanteilen.

Durch das Signalrauschen empfiehlt es sich, bei der FT längere Zeitreihen als T_{min} zu verwenden. Der zufällige Anteil lässt diesen Schwellenwert unschärfer erscheinen, da erst ab weiteren 5 Tagen für alle betrachteten SNR ein Grenzwert zu erahnen ist. Um diesen pendeln die Amplituden aufgrund des Leakage-Effekts (Puate León, 2019).

Bei der Untersuchung der Realdaten zeigt Abbildung 3.11 besonders in der S_2 -Komponente eine meist niedrigere Amplitude, als jene aus der kompletten Zeitreihe. Außerdem sind bei Verwendung der Hälfte des verfügbaren Datensatzes die Amplituden von S_2 besonders groß. Diese langperiodischen Schwankungen dieser Amplitude deuten darauf hin, dass die Amplitude von S_2 instationär ist, sich also im Laufe der Messung verändert. Diese Problematik tritt bei den synthetischen Messdaten

nicht auf und erschwert eine Aussage über die Mindestmessdauer bei den Realdaten. Die Amplitude der Komponenten bei Betrachtung des kompletten Datensatzes ist lediglich ein Mittelwert der Amplituden über den gesamten Messzeitraum. Die Betrachtung eines möglichst langem Datensatz ist somit nicht unbedingt vorteilhaft, wenn es sich um instationäre Amplituden handelt.

Zur besseren Untersuchung der Mindestmessdauer in den Realdaten sind in Abbildung 4.1 die ermittelten Amplituden von M_2 und S_2 für kurze Messdauern ab Messbeginn am 24,10,2003 gezeigt. Die Messdauer wurde schrittweise um 23 Stunden erhöht.

Es ist eine besonders hohe Abweichung der Amplituden bei kurzen Messdauern zum Sollwert (gestrichelte Linien in Abbildung 4.1) zu erkennen. Dies kann auf eine tatsächliche starke Änderung der Amplitude in den ersten 100 Tagen hindeuten. Wahrscheinlicher ist jedoch, dass wie bei den Realdaten für kurze Messdauern die Verfahren aufgrund des Leakage-Effekt und wegen des Rauschens stark von den eigentlichen Amplituden abweichen

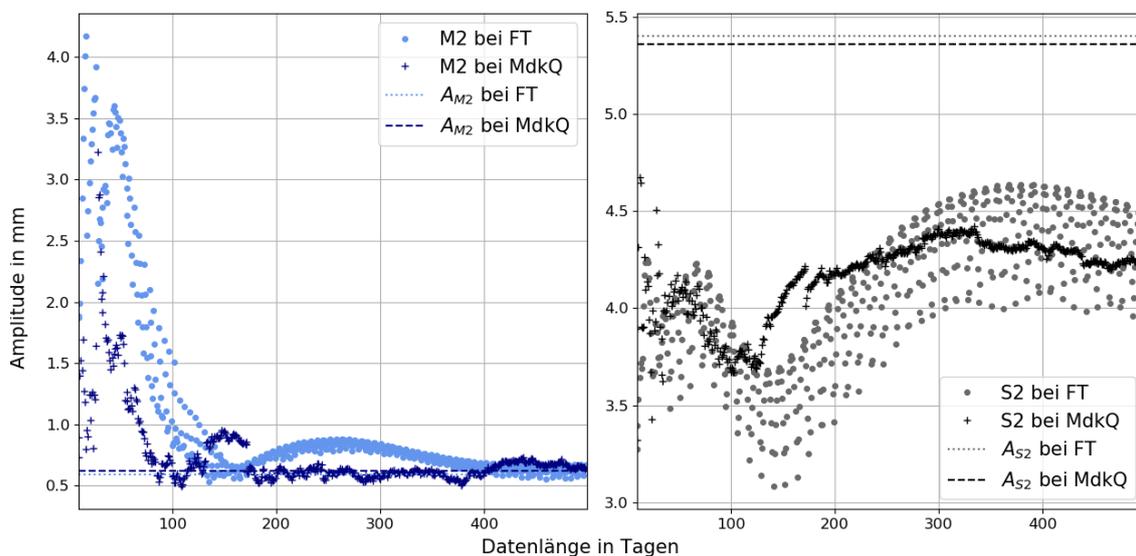


Abbildung 4.1: Amplituden von M_2 und S_2 aus FT und MdkQ für kurze Messdauern. Die Linien deuten die ermittelten Amplituden der Komponenten bei vollständigem Datensatz an.

Die betrachteten Realdaten enthalten neben M_2 und S_2 noch weitere Komponenten sowie Signalrauschen (Abbildung 3.10). Dies erschwert die Analyse des Datensatzes zusätzlich. Die starken Änderungen der Amplituden der Komponenten scheinen ab circa 200 Tagen abzunehmen.

Bei den synthetischen Daten ist aufgrund des einfachen Aufbaus des Datensatzes die Mindestmessdauer bei Auswertung mit FT zu erkennen. Für die MdkQ reicht auch schon eine geringere Messlänge, jedoch sorgt das Signalrauschen für Fehler. Bei den Realdaten liegt eine instationärer Fall für die S_2 -Komponente vor. Aufgrund der Überlagerung mit anderen Komponenten sowie Signalrauschen bei der Analyse dieses Datensatzes sollten hierbei eher Messdauern ab 200 Tagen betrachtet werden.

Der Leakage-Effekt sorgt für Fehler bei der Auswertung mittels FT

Ab T_{min} überwiegt in den synthetischen Messreihen der Fehler durch den Leakage-Effekt den Fehler des Rauschens: Die Amplituden weichen je nach SNR um bis zu 20% von der ursprünglichen Amplitude ab. Tabelle 4.1 zeigt die Mittelwerte der für 1000 Durchgänge gemittelten Amplituden für Messdauern ab der doppelten Mindestmessdauer $2T_{min}$. Der Leakage-Effekt tritt besonders stark bei der M_2 -Komponente auf. Dort ist die Abweichung zur Amplitude des generierten Signals am höchsten.

Tabelle 4.1: Mittelwerte in mm der Amplituden aus Abbildung 3.2 der Messdauern von 60 bis 120 Tagen.

Komponente	Modell	SNR= ∞	SNR=2	SNR=0,5
M_2	10	9,53	9,55	9,27
S_2	10	9,95	9,98	10,07

Dies liegt daran, dass die Frequenz $f_{M_2} = 1,932274$ cpd kaum von einem Frequenzbin im Spektrum getroffen wird. Die Messdauer erreicht aufgrund des gewählten Abtastintervalls kein ganzzahliges Vielfaches der M_2 -Periode

$$T_{M_2} = \frac{1}{\Delta f_{M_2}} = 0,518 \text{ Tage.}$$

Dadurch verteilt sich die Energie im Spektrum auf die nächstgelegenen Bins (Sundararajan, 2003). Um dies zu vermeiden, bietet sich eine Messdauer an, die sowohl Vielfache von M_2 als auch S_2 ist. Dies ist bei 456 Tagen mit 0,1% Abweichung für $881 \cdot T_{M_2}$ der Fall (vgl. Abschnitt 2.2).

In den Realdaten unterscheiden sich die Kurven der durch FT und MdkQ ermittelten Amplituden deutlich: Die Ergebnisse der FT in Abbildung 4.1 schwanken je nach gewählter Messdauer nach 200 Tagen um $\pm 0,3$ mm bei M_2 und $\pm 0,4$ bei S_2 . Diese starken Schwankungen sind auch in Abbildung 3.11 zu sehen und erklärt die scheinbar unterschiedlichen Kurven: Durch die Schrittweise Erhöhung geht im Spektrum Energie der Komponenten M_2 und S_2 auf die nächstgelegenen Bins über. Da in Abbildung 3.11 und Abbildung 4.1 verschiedene Inkremente der Messdauern gewählt wurden, entstehen durch den Leakage-Effekt andere Muster. Dies sorgt für beträchtliche Fehler auf die ermittelte Amplitude aus der FT.

Bei der MdkQ spielt der Leakage-Effekt keine Rolle, da die Amplituden nicht aus dem Spektrum ermittelt werden. Die Fehler auf die Amplituden von M_2 und S_2 liegen nach 200 Tagen bei respektive $\pm 0,1$ mm und $\pm 0,2$ mm, und damit deutlich weniger als bei der FT.

Die MdkQ liefert auch bei kurzen Messzeiträumen genauere Ergebnisse als die FT.

Die Auswertung mittels MdkQ ist auch für kurze Messdauern bei der Auswertung der synthetischen performant (Abbildung 3.6). Schon nach 10 Tagen ist bei jedem betrachteten SNR die Amplitude von 10 mm im Mittel um $\pm 0,3$ mm genau getroffen. Die MdkQ lässt die für die FT berechnete Mindestmessdauer außer Betracht, da die

Frequenzen schon vor der Berechnung der Amplituden bekannt sind. Hier spielt der Effekt des Signalrauschens bei kurzen Datenlängen eine wichtigere Rolle: Bei 5 Tagen beträgt die Standardabweichung noch maximal 5,7 mm und nimmt mit steigender Messdauer weiter ab. Bei 20 Tagen Messdauer ist sie bei allen betrachteten SNR auf unter 2 mm gesunken.

Der Leakage-Effekt ist bei der MdkQ nicht vorhanden: Für beide Komponenten M_2 und S_2 schmiegen sich die Mittelwerte der Amplituden mit wachsender Messdauer schnell an die des Modells an. Besonders für die ideale Messreihe mit $\text{SNR}=\infty$ stimmen die Amplituden sehr genau mit denen des Modells überein. Aber auch für die Signale mit Rauschanteil liefert die MdkQ gute Ergebnisse: Die Abweichung zum Modell liegt in der ersten Nachkommastelle.

Auch in den Realdaten weichen die Ergebnisse aus der MdkQ bei kurzen Messdauern weniger vom langjährigen Mittel ab als die der FT. Dies ist insbesondere in Abbildung 4.1 bei M_2 zu erkennen. Abbildung 4.1 zeigt außerdem, dass sich die Ergebnisse aus der MdkQ schon für kürzere Messreihen dem Sollwert annähern. Bei einer Untersuchung von besonders kurzen Messreihen ist die MdkQ somit der Auswertung mittels FT zu bevorzugen.

4.2 Der Einfluss der Quantisierung

Eine Standard-Sensorauflösung von 1-5 cm ist ausreichend, um die GM für den untersuchten Datensatz durchzuführen.

Die aus den Realdaten ermittelten Amplituden weichen für kleine Quantisierungen bis 5 cm um maximal 0,2 mm für M_2 und maximal 0,3 mm für S_2 von ihrem Sollwert ab (Abbildung 3.13). Letzterer wurde aus den mit 1 cm quantisierten Rohdaten ermittelt.

Die Analyse hat gezeigt, dass auch Amplituden detektiert werden können, die eine Größenordnung kleiner als die Quantisierungen sind: Die Daten wurden mit 1 cm Auflösung aufgezeichnet und es konnten Komponenten mit Amplituden im Bereich einiger Millimeter nachgewiesen werden. Der relative Fehler beträgt bis 5 cm Quantisierung weniger als 20 % für die M_2 -Komponenten und weniger als 10 % für die S_2 -Komponente.

Die untersuchten Quantisierungen sind größer als die der in der Praxis gängigen Drucksensoren. Rau et al. (2019) konnten zeigen, dass die Auflösung von belüfteten Drucksensoren (vented pressure transducer) bis 1,5 mm genau ist. Handelt es sich um unbelüftete Drucksensoren, verdoppelt sich die Quantisierung, da zwei Messgeräte nötig sind, um den Wasser- und Atmosphärendruck aufzuzeichnen.

Rahi and Halihan (2013) haben Spannungszustände mit der Gezeitenmethode an verschiedenen Messpunkten bestimmt. Die SNRs liegen dort in der gleichen Größenordnung wie die der Realdaten (ca. 10^2). Mit diesem Signal-Rausch-Verhältnis können die Autoren eindeutige Aussagen über den Spannungszustand treffen.

Signale mit geringerem SNR erzielen bei hohen Quantisierungen bessere Ergebnisse.

Tabelle 4.2 zeigt in Einheiten der ursprünglichen Amplitude $A_0 = 10$ mm der synthetischen Daten, bis zu welcher Quantisierung eine Genauigkeit von 90 % erreicht

wird. Dies ist die maximal zulässige Quantisierung, mit der die aus den Regressionsverfahren gewonnenen Amplituden über 9 mm hoch sind.

Tabelle 4.2: Maximale Quantisierung in Einheiten von A_0 , mit der eine Genauigkeit von 90 % erreicht wird.

Komponente:	FT		MdkQ	
	M_2	S_2	M_2	S_2
SNR= ∞	2	3	2	2
SNR=2	4	5	4	4
SNR=0,5	6	7	6	6

Auffällig ist, dass für die Signale mit größerem Rauschanteil höhere Quantisierungen möglich sind, bevor die Ergebnisse dispergieren. Da mit sinkendem SNR die Standardabweichung der zufälligen synthetisierten Fehler steigt, übersteigen mehr Werte der Messreihe das Quantum, mit dem die Werte abgespeichert werden. So entstehen auch bei hohem Rauschanteil Messreihen mit nicht verschwindenden Amplituden. Aus diesem Grund entsteht im Fall ohne Rauschen die abrupte Abbruchkante in Abbildungen 3.4 und 3.7. Für reale Messreihen ist dies von Vorteil, da diese immer mit einem Rauschanteil behaftet sind.

Eine Erhöhung der Quantisierung der Messdaten verringert die Energie des synthetischen Signals und bewirkt bei den Realdaten eine Zunahme der Streubreite der ermittelten Amplituden.

Durch die Quantisierung sinken die ermittelten Amplituden sowohl unter Verwendung der FT als auch der MdkQ bei den synthetischen Daten (Abbildungen 3.4 und 3.7). Dies zeigt sich auch im Frequenzspektrum der synthetischen Daten (Abbildung 3.3). Da diese 0 mm als Mittelwert haben, werden bei Erhöhung der Quantisierung mehr Werte durch Nullen ersetzt. Dadurch geht Energie des Signals verloren und die Amplituden sinken.

In den Realdaten streuen die Amplituden zunehmend mit steigender Quantisierung (Abbildung 3.12). Da die Quantisierung vor dem Detrend der Daten durchgeführt wird, ist Null zunächst nicht der Mittelwert der Zeitreihe. Durch die Auflösung des Sensors geht somit keine Energie des Signals verloren, da der tatsächliche Wert dem nächst höheren oder nächst niedrigeren Vielfachen des Quantums zugewiesen wird. Welchem davon ist abhängig vom Wert des Quantums und der Lage des Messwertes zwischen den beiden Vielfachen. Ob durch die Quantisierung mehr oder weniger Daten durch das nächst höhere Vielfache des Quantums ersetzt werden, bleibt dem Zufall überlassen. Dadurch steigt die Streubreite des Ergebnisses.

Die Leistungsfähigkeit von FT und MdkQ bei der Analyse stark quantisierter Daten ist ähnlich.

Bei der Auswertung der synthetischen Daten ermitteln die Programme bis zu einer Datenquantisierung von 10 mm bei jedem eingestellten SNR für beide Regressionsverfahren eine konstante Amplitude von $10 \pm 0,41$ mm. Sieht man von der stark vom

Leakage-Effekt betroffenen M_2 -Komponente ab, sind die Ergebnisse insgesamt mit $10 \pm 0, 15$ mm noch genauer.

Bei der MdkQ liegt die Abweichung von den Ursprungsamplituden bei einer Quantisierung unter 10 mm bei den synthetischen Signalen mit Rauschen bei höchstens 0,25%. Danach sinken die Amplituden mit steigender Quantisierung bis auf 0 ab, da nur weniger Messpunkte das Quantum überschreiten.

Auch bei den Realdaten konnte ein Zusammenhang zwischen den ermittelten Amplituden mit der Auflösung des Messinstruments festgestellt werden (vgl. Abbildung 3.12): Ab $q = 5$ cm für M_2 und $q = 10$ cm für S_2 divergieren die ermittelten Amplituden.

Die Frequenzspektren einer synthetischen Messreihe und der Realdaten (Abbildungen 3.3 und 4.2) zeigen, dass bei steigender Quantisierung das SNR deutlich absinkt: Die Amplitude des Rauschens erhöht sich und bei M_2 ist für den realen Datensatz spätestens bei $q = 10$ cm kein Hochpunkt im Spektrum mehr zu erkennen. Die ermittelten Amplituden für S_2 sind ähnlich hoch bei FT und MdkQ.

Die in Abbildung 3.13 dargestellten Amplituden aus der FT sind lediglich die beiden höchsten Werte im Bereich von 1,85 und 2,1 cpd. Deren Stellen stimmen bei starker Quantisierung nicht genau mit den gesuchten Frequenzen f_{M_2} und f_{S_2} überein.

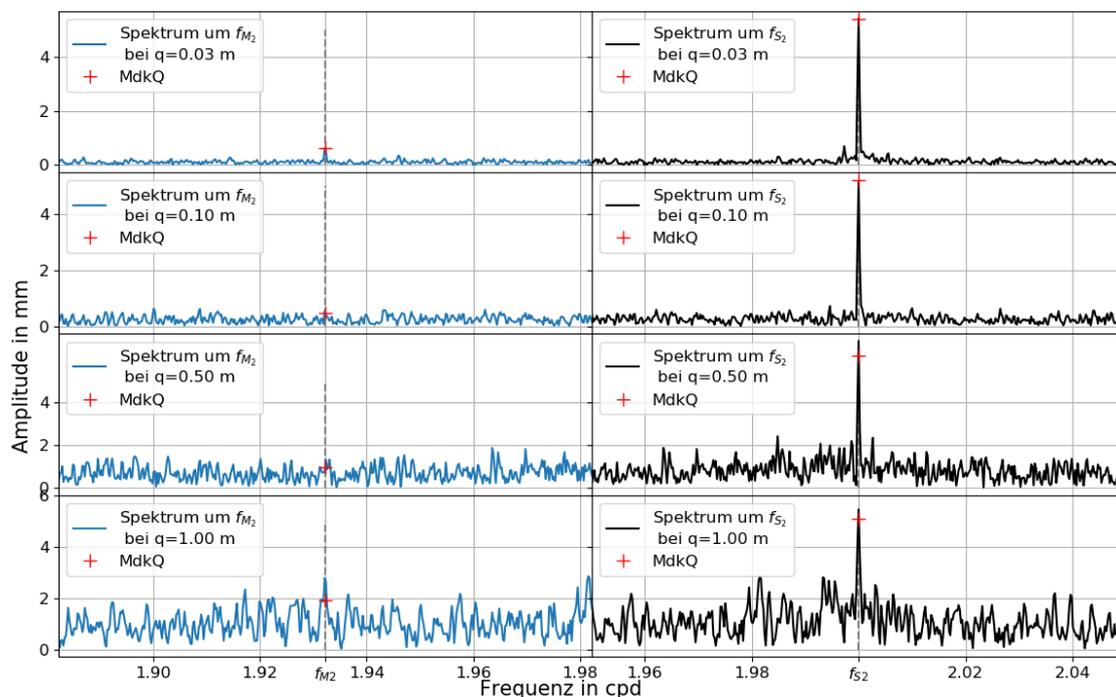


Abbildung 4.2: Amplitudenspektren der Realdaten um f_{M_2} und f_{S_2} bei steigender Quantisierung q

Beide Regressionsverfahren haben die Schwierigkeit, bei hoher Quantisierung tendenziell binäre Daten anpassen zu müssen: Pendeln die Daten um die Grenze zwischen zwei Vielfachen $(n+1)q$ und nq des Quantums q , entstehen Sprünge in der quantisierten Zeitreihe: Eine kleine Änderung des tatsächlichen Signals sorgt für Strukturbrüche in der quantisierten Zeitreihe (Abbildung 4.3).

Insgesamt hat die Auswertung gezeigt, dass kein ausschlaggebender Unterschied

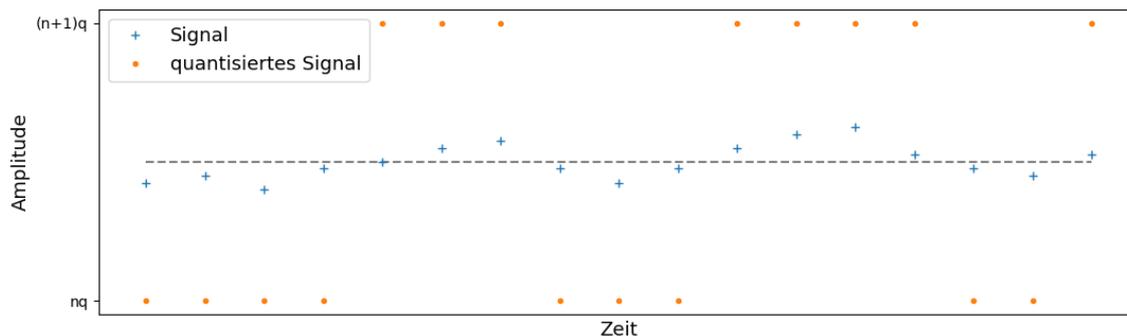


Abbildung 4.3: Ein um die Quantisierungsgrenze pendelndes Signal erzeugt ein stark variierendes quantisiertes Signal

in der Leistungsfähigkeit der beiden Regressionsanalyseverfahren bei stark quantisierten Daten festgestellt werden konnte. Beide Methoden liefern für die Realdaten bis 5 cm Auflösung für M_2 um $\pm 10\%$ genaue Ergebnisse. Bei S_2 ist diese Genauigkeit noch bis 10 cm Auflösung erreicht. Dies liegt höchstwahrscheinlich an der deutlich größeren Amplitude von S_2 (vgl. Abbildung 3.10).

4.3 Der Einfluss der Datenlücken

Die MdkQ ist der FT bei lückenhaften Datensätzen vorzuziehen.

Ein Vorteil der MdkQ ist, dass das Ergebnis im Gegensatz zur FT nicht mehr korrigiert werden muss. Die MdkQ liefert direkt die Amplituden mit den kleinsten Fehlerquadraten für den lückenhaften Datensatz.

Da die MdkQ auch mit unregelmäßig gesampelten Daten umgehen kann, brauchen die Datenlücken nicht durch Werte ersetzt werden. Dadurch lässt sich dem Programm auch ein unvollständiger Datensatz übergeben. Durch die zufällige Generierung der Lücken liegen weniger Datenpunkte vor, um ein Modell für die Messreihe zu erstellen. Dies erschwert es, den Frequenzen f_{M_2} und f_{S_2} die gleichen Energien bzw. Amplituden zuzuordnen, wie beim Signal ohne Datenlücken. Dies erklärt die zunehmende Streuung in Abbildung 3.8.

Bei sehr großen Lückenanteilen (über 90 % in Abbildung 3.9) ist dieser Effekt in ausgeprägter Form zu beobachten. Ab 99,9 % Lückenanteil, wobei der Datensatz nur noch aus 10 Werten besteht, liegen die relativen Fehler deutlich über 100 %. Während die FT jedoch aufgrund des niedrigen SNR nicht zwangsläufig die Amplituden der gesuchten Frequenzen trifft, ermittelt die MdkQ weiterhin Amplituden für die Frequenzen von M_2 und S_2 .

Insgesamt liefert die MdkQ somit auch bei hohen Lückenanteilen noch gute Ergebnisse. Während bei der FT im Spektrum bei hohen Lückenanteilen kein Hochpunkt im Spektrum mehr zu erkennen ist, lässt sich bei Verwendung der MdkQ trotzdem eine Amplitude für die entsprechende Komponente ermitteln.

Zu einem ähnlichen Schluss kommen auch He et al. (2009): Die Autoren schlagen eine Auswertung zeitlich ungleichförmig abgetasteter Daten mittels einer iterativ gewichteten Methode der kleinsten Quadrate vor.

Fehler durch Datenlücken lassen sich bei Auswertung mittels FT bedingt korrigieren.

Die Funktion `detrend` aus dem `Signal`-Paket kann lediglich gleichmäßig gesampelte Zeitreihen von ihrem Trend bereinigen. Es musste somit die Funktion `linear-window3` erstellt werden, um auch lückenhaften Daten vom linearen Trend bereinigen zu können. Diese ermittelt wegen des unterschiedlichen Algorithmus' andere Amplituden (Abbildung 3.15) für den vollständigen Datensatz, wie die aus dem Spektrum der Realdaten aus Abbildung 3.10.

Ein Interpolieren von Datenlücken ist nur bei isolierten, fehlenden Werten sinnvoll. Ein Vergleich verschiedener Interpolationsmethoden liefert Musial et al. (2011). Handelt es sich um längere Datenlücken, können diese in keiner angemessenen Weise durch geschätzte Daten ersetzt werden (Dilmaghani et al., 2007). Das Interpolieren mit Nullen ist somit eher für eine geringe Anzahl zufällig erstellter Lücken sinnvoll: Werden viele Lücken generiert, steigt die Wahrscheinlichkeit, mehrere Lücken hintereinander zu erhalten.

Sowohl für die Realdaten als auch bei den synthetischen Daten ist beim Ersetzen der Lücken durch Nullen ein linearer Zusammenhang zwischen der Amplitude A_L und dem Lückenanteil L bei der Auswertung mittels FT zu erkennen:

$$A_L = mL + A. \quad (4.2)$$

A ist die Amplitude der jeweiligen Komponente bei vollständigen Datensatz. Durch das Ersetzen der Datenlücken mit 0 (dem Mittelwert der Zeitreihe) geht Energie der Schwingung verloren. Dadurch verringern sich alle durch die FT bestimmten Amplituden. Mit Kenntnis des Lückenanteils L (in %) kann durch Umformen von Gleichung 4.2 eine korrigierte Amplitude A_0 aus A_L ermittelt werden:

$$A_0 = A_L - mL. \quad (4.3)$$

Die Steigung m lässt sich aus den Punkten $(L|A_L)$ und $(100|0)$ berechnen:

$$m = \frac{A_L}{L - 100}. \quad (4.4)$$

Setzt man m aus Gleichung 4.4 in Gleichung 4.3 ein, ergibt sich für A_0

$$A_0 = \frac{100}{100 - L} A_L \quad (4.5)$$

Mit Ausdruck 4.5 lässt sich somit eine korrigierte Amplitude berechnen. Da auch die Realdaten einen linearen Abfall mit steigendem Lückenanteil aufweisen, kann mit Formel 4.5 auch dort die ermittelte Amplitude korrigiert werden.

Setzt man beispielsweise bei 30% des Realdatensatzes Lücken ein, ergibt sich für S_2 eine Amplitude von $A_L^{S_2} = 3,90$ mm im Fourierspektrum. Durch eine Korrektur nach Formel 4.5 ist $A_0^{S_2} = 5,56$ mm. Die Abweichung zur Amplitude bei lückenlosem Datensatz beträgt damit -0,4% (vgl. Abschnitt 3.2).

Bei M_2 ist bei 30% Lückenanteil $A_L^{M_2} = 0,42$ mm im Spektrum, wodurch sich mit Korrektur $A_0^{M_2} = 0,60$ mm ergibt. Die Abweichung zur Amplitude ohne Lückenanteil beträgt dann -6,5% und ist betragsmäßig größer als die der S_2 -Komponente. Dies ist auf den Leakage-Effekt zurückzuführen, da M_2 , wie in Abschnitt 4.1 beschrieben, stärker davon betroffen ist.

Ein weiterer Vorschlag zur Korrektur von lückenhaften Datensätzen ist eine lineare Regression durch die Amplituden, die bei Erhöhung des Lückenanteils entstehen. Aus den Daten von Abbildung 3.4 berechnen sich mithilfe der `Scipy interpolate`-Funktion `polyfit` die ursprünglichen Amplituden aus Tabelle 4.3. Diese Funktion verwendet die MdkQ, um eine Gerade möglichst genau an die Messwerte anzupassen.

Tabelle 4.3: Durch lineare Regression ermittelte Amplituden in mm von M_2 und S_2 der synthetischen Daten nach Erhöhung des Lückenanteils

SNR	M_2	S_2
inf	9,79	9,99
2	9,76	9,96
0,5	9,76	9,96

Die Erhöhung des Lückenanteils verringert die Amplituden von M_2 und S_2 . Dies erschwert, die Maximas bei f_{M_2} und f_{S_2} neben dem Rauschen im Spektrum zu identifizieren. Abbildung 4.4 zeigt dies anhand des Amplitudenspektrums der Realdaten: Die Amplitude von M_2 ist bei 30 % Lücken kaum noch zu erkennen. In diesen Fällen werden, wie auch schon bei niedrigeren Lückenanteilen L , die beiden Werte im Spektrum im Bereich von 1,85 cpd bis 2,05 cpd als Amplituden für M_2 und S_2 gewählt.

Ob eine lineare Regression oder eine Korrektur nach Formel 4.2 genauer ist, hängt somit vom vorhandenen Lückenanteil ab. Eine allgemeine Empfehlung abzugeben, bei welchen Lückenanteilen welche Methode zu bevorzugen ist, setzt eine weitere Untersuchung voraus.

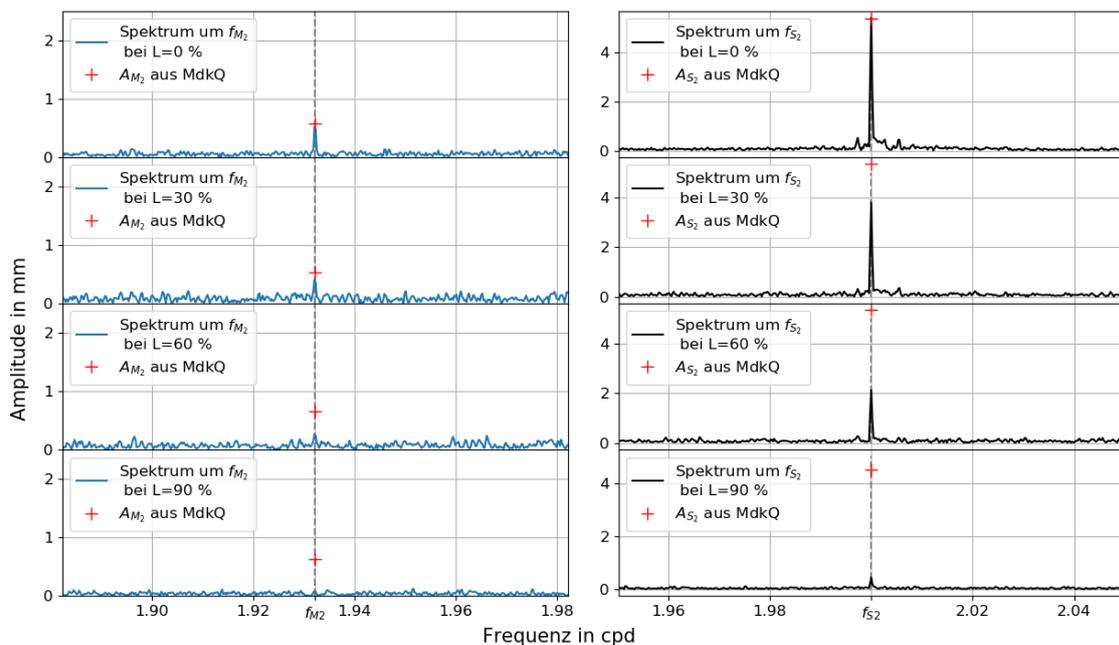


Abbildung 4.4: Amplitudenspektrum der Realdaten bei steigendem Lückenanteil

Ob der Lückenanteil einen Einfluss auf den Erwartungswert der Amplitude hat, hängt davon ab, ob diese stationär ist.

Die Untersuchung des Einflusses von Datenlücken bei den Realdaten hat gezeigt, dass mit Zunahme des Lückenanteils das Streuungsmaß zunimmt (Abbildungen 3.5, 3.8 und 3.9). Dies liegt daran, dass die Lücken zufällig im Datensatz generiert wurden. Im Fall mit konstanter Amplitude streuen die ermittelten Amplituden in beide Richtungen um den Sollwert.

Bei den Realdaten ist für S_2 in Abbildung 3.15 eine Abnahme der Amplitude mit steigendem Lückenanteil zu sehen. Im Gegensatz zu M_2 ist die Amplitude von S_2 instationär über den Messzeitraum. Wie in Abschnitt 4.1 gezeigt wurde, ist ihre Intensität für die meisten gewählten Messzeiträume geringer, als im Mittel über den kompletten vorhandenen Datensatz.

Steigt die Anzahl der zufällig verteilten Lücken, erhöht sich somit die Wahrscheinlichkeit, eine Amplitude geringer als A_{S_2} aus FT bzw. MdkQ zu erhalten. Die geringere Anzahl an Messzeiträumen mit erhöhter Intensität der S_2 -Komponente verliert somit an Gewicht gegenüber den Messzeiträumen mit stärkerer Amplitude dieser Komponente (Abbildung 3.15). Dies hat das Absinken der Amplitude der S_2 -Komponente mit dem Lückenanteil in Abbildung 3.15 zur Folge.

M_2 ist näherungsweise konstant über die Messdauer (vgl. Abschnitt 4.1). Dies eine zunehmende Streuung mit wachsendem Lückenanteil um den langjährigen Mittelwert A_{M_2} in Abbildung 3.15 zur Folge. Hier ist kein Trend der Amplitude mit zunehmender Anzahl an Lücken zu erkennen.

Für zeitlich variierende Amplituden ist somit die Auswertung eines langen Datensatz nicht besser als für einen kürzeren. Ändert sich die Amplitude einer Komponente mit der Zeit, ist es ratsam, einen eher kurzen Datensatz (mindestens jedoch 30 Tage) ab circa dem Zeitpunkt, an dem die Amplitude bestimmt werden soll, zu betrachten.

Kapitel 5

Schlussfolgerung und Ausblick

Hydrogeomechanische Parameter wie der Spannungszustand oder Speicherkoeffizienten sind von zentraler Bedeutung für die Charakterisierung von Untergründen. Die Bestimmung dieser Bodenkennwerte ist kosten- und zeitintensiv, da dazu Messungen im Feld (z.B. Pumpversuche) oder an Bodenproben im Labor (z.B. Permeameter-Test) durchgeführt werden müssen. Die Gezeitenmethode (GM) bietet Potential für eine effizientere Alternative: Anhand von gezeitenbedingten Änderungen von Grundwasserspiegelhöhen lassen sich hydrogeomechanische Eigenschaften des Aquifers berechnen.

Die Fourier-Transformation (FT) hat sich für die GM etabliert. Hierbei liefert das Frequenzspektrum die Amplituden der einzelnen Teilschwingungen. Diese lassen sich auch mithilfe der Methode der kleinsten Quadrate (MdkQ) für harmonische Schwingungen ermitteln. In dieser Arbeit werden beide Regressionsanalyseverfahren gegenübergestellt und ihre Leistungsfähigkeit in Bezug auf drei Messparameter untersucht: (1) die Messdauer, (2) die Auflösung des Messgerätes und (3) Lücken in der Zeitreihe.

Es wurden synthetische Messreihen aus den realistischen Gezeitenkomponenten M_2 und S_2 erstellt. Diese Teilschwingungen der Grundwasser-Gezeiten haben Perioden von 1,932274 cpd und 2,000000 cpd (cycles per day) und liegen im Frequenzspektrum dicht beieinander. Außerdem wurden die Komponenten M_2 und S_2 in einem Datensatz einer über 12 Jahre langen Messung mit stündlicher Probenahme nachgewiesen. Die synthetischen Zeitreihen und der reale Datensatz dienen als Grundlage der Untersuchung: Messdauer, Quantisierung der Daten und Lückenanteil wurden variiert und ihren Einfluss auf die Amplituden von M_2 und S_2 analysiert.

Die wichtigsten Erkenntnisse der Untersuchungen lauten wie folgt:

- Die Bestimmungsgenauigkeit der Methode der kleinsten Quadrate ist insbesondere bei kurzen Zeitreihen bis 30 Tagen sowie bei lückenhaften Datensätzen genauer als die der Fourier-Transformation. Dies liegt teilweise an Fehlern durch den Leakage-Effekt.
- Eine Standard-Sensorauflösung (ca. 1 cm, maximal aber von bis zu 5 cm) ist ausreichend, um die GM für den realen Datensatz durchzuführen.
- Die Voraussetzung eines regelmäßig gesampelten Datensatzes für die Fourier-Transformation verringert ihre Leistungsfähigkeit gegenüber der Methode der kleinsten Quadrate.

- Die Messdauer sollte bei stündlicher Probennahme mindestens 30 Tage betragen, um die Komponenten M_2 und S_2 im Fourierspektrum voneinander trennen zu können.
- Die Auswertung eines möglichst langem Datensatz ist bei zeitlich variierenden Amplituden problematisch.

Der Einfluss der drei Messparameter auf die Phase der Teilschwingungen wurde in dieser Arbeit nicht betrachtet. Dazu sind weitere Untersuchungen notwendig.

Außerdem wurden die Einflüsse der Quantisierung und des Lückenanteils nur für Datenlängen von mindestens 456 Tagen bei stündlicher Probennahme untersucht. Bei besonders kurzen lückenhaften Zeitreihen könnten weitere Effekte auftreten, die in dieser Untersuchung nicht betrachtet werden konnten. Eine Quantifizierung der Effekte bei kombinierter Problematik durch kurze Messdauer, schlechte Datenquantisierung und hohem Lückenanteil bedarf weiterer Studien.

Die Auswertung des Einflusses des Lückenanteils mittels Fourier-Transformation führte zu zwei Korrekturvorschlägen für die ermittelten Amplituden: eine Korrektur nach Formel 4.5 sowie durch lineare Regression bei weiterer Erhöhung des Lückenanteils. Um herauszufinden, welche Korrektur sich unter welchen Umständen besser eignet, sind weitere Untersuchungen notwendig.

In dieser Arbeit wurde der Einfluss einzelner, zufällig verteilter Datenlücken untersucht. Treten diese jedoch über längere andauernde Zeitintervalle auf, was in der Praxis eher vorkommt, lassen sich diese schlecht interpolieren. Ein Vergleich der untersuchten Regressionsanalyseverfahren bei zunehmender Lückendauer ist für weitere Untersuchungen angebracht. Hierzu sei auf Untersuchungen für zur Fourier-Transformation verwandte Regressionsanalyseverfahren von Munteanu et al. (2016) und Munteanu et al. (2016) verwiesen.

Insgesamt hat die Methode der kleinsten Quadrate aufgrund genauere Ergebnisse gegenüber der Fourier-Transformation überzeugt. Ein Nachteil der Methode der kleinsten Quadrate für harmonische Schwingungen ist, dass sie die Kenntnis der Frequenzen der Teilschwingungen voraussetzt: Sind diese unbekannt, lässt sich dieses Verfahren nicht anwenden. Empirie und Modelle wie der Gezeitenkatalog KSM03 bieten jedoch im Kontext der Gezeitenanalyse eine sehr genaue Kenntnis über die relevanten Signalanteile.

Literaturverzeichnis

- Acworth, R. I., Halloran, L. J., Rau, G. C., Cuthbert, M. O., and Bernardi, T. L. (2016). An objective frequency domain method for quantifying confined aquifer compressible storage using Earth and atmospheric tides. *Geophysical Research Letters*.
- Acworth, R. I., Rau, G. C., McCallum, A. M., Andersen, M. S., and Cuthbert, M. O. (2015). Understanding connected surface-water/groundwater systems using Fourier analysis of daily and sub-daily head fluctuations. *Hydrogeology Journal*, 23(1):143–159.
- Buttkus, B. (2000). *Spectral Analysis and Filter Theory in Applied Geophysics*. Springer-Verlag Berlin Heidelberg.
- Chapman, S. and Malin, S. R. C. (1970). Atmospheric Tides, Thermal and Gravitational: Nomenclature, Notation and New Results. *Journal of the Atmospheric Sciences*, 27(5):707–710.
- Chapman, S. and Westfold, K. C. (1956). A comparison of the annual mean solar and lunar atmospheric tides in barometric pressure, as regards their worldwide distribution of amplitude and phase. *Journal of Atmospheric and Terrestrial Physics*.
- Clauser, C. (2016). *Einführung in die Geophysik*. Springer Spektrum, Berlin, Heidelberg, 2., aktual edition.
- Cooper, H. H., Bredehoeft, J. D., and Papadopoulos, I. S. (1967). Response of a finite-diameter well to an instantaneous charge of water. *Water Resources Research*.
- Dilmaghani, S., Henry, I. C., Soonthornnonda, P., Christensen, E. R., and Henry, R. C. (2007). Harmonic analysis of environmental time series with missing data or irregular sample spacing. *Environmental Science and Technology*, 41(20):7030–7038.
- Fourier, J. B. J. (1822). *Théorie Analytique de la Chaleur*. Cambridge University Press, Cambridge.
- Harris, F. (1978). On the use of windows for harmonic analysis with the discrete Fourier transform. *Proceedings of the IEEE*, 66(1):51–83.
- Haurwitz, B. (1964). Atmospheric Tides. *Science*, 144(3625):1415–1422.

- He, H., Li, J., and Stoica, P. (2009). Spectral Analysis of Non-Uniformly Sampled Data: A New Approach Versus the Periodogram. In *2009 IEEE 13th Digital Signal Processing Workshop and 5th IEEE Signal Processing Education Workshop*, volume 57, pages 375–380. IEEE.
- Hiscock, K. M. and Bense, V. F. (2014). *Hydrogeology Principles and Practice*. Chichester [u.a.] : Wiley Blackwell.
- Jacob, C. E. (1940). On the flow of water in an elastic artesian aquifer. *Transactions, American Geophysical Union*, 21(2):574.
- Jordan, H. H. and Abraham, T., editors (1995). *Hydrogeologie : Grundlagen und Methoden; regionale Hydrogeologie: Mecklenburg-Vorpommern, Brandenburg und Berlin, Sachsen-Anhalt, Sachsen, Thüringen*. Enke, Stuttgart, 2., stark überarb. u. erw. Aufl. edition. IMD-Felder maschinell generiert (GBV).
- Kudryavtsev, S. M. (2004). Improved harmonic development of the Earth tide-generating potential. *Journal of Geodesy*.
- McMillan, T. C., Rau, G. C., Timms, W. A., and Andersen, M. S. (2019). Utilizing the Impact of Earth and Atmospheric Tides on Groundwater Systems: A Review Reveals the Future Potential. *Reviews of Geophysics*, pages 1–35.
- Munteanu, C., Negrea, C., Echim, M., and Mursula, K. (2016). Effect of data gaps: comparison of different spectral analysis methods. *Annales Geophysicae*, 34(4):437–449.
- Musial, J. P., Verstraete, M. M., and Gobron, N. (2011). Comparing the effectiveness of recent algorithms to fill and smooth incomplete and noisy time series. *Atmospheric Chemistry and Physics Discussions*.
- NumPy (2017). NumPy — NumPy.
- Puente León, F. (2019). *Signale und Systeme*. De Gruyter, Berlin, 7., überarb. edition.
- Rahi, K. A. and Halihan, T. (2013). Identifying Aquifer Type in Fractured Rock Aquifers using Harmonic Analysis. *Ground Water*, 51(1):76–82.
- Rau, G. C., Post, V. E. A., Shanafield, M., Krekeler, T., Banks, E. W., and Blum, P. (2019). Error in hydraulic head and gradient time-series measurements: a quantitative appraisal. *Hydrology and Earth System Sciences*, 23(9):3603–3629.
- Stahn, T. and Gizon, L. (2008). Fourier analysis of gapped time series: Improved estimates of solar and stellar oscillation parameters. *Solar Physics*.
- Sundararajan, D. (2003). The Discrete-Time Fourier Transform. In *Digital Signal Processing*, pages 63–76. WORLD SCIENTIFIC.

Anhang

Alle im Rahmen dieser Arbeit erstellten Programmcodes sowie die untersuchten Realdaten sind auf dem beigefügten Datenträger abgespeichert. Der Python-Code zur Erstellung von Abbildung 3.7 ist im Folgenden beispielhaft abgedruckt.

```
1 # Einfluss der Quantisierung, MdkQ
2 import matplotlib.pyplot as plt
3 import numpy as np
4
5 def sig(t, amps, phases, f,SNR):
6     harmonic = np.zeros(t.shape)
7     for i in range(len(f)):
8         harmonic += amps[i] * np.sin(2*np.pi*f[i]*t + phases[i])
9     #gaussian noise
10    noise = np.random.normal(0, np.amax(harmonic)/np.sqrt(SNR), len(t))
11    #signal plus noise
12
13    harmonic += noise
14    return harmonic
15
16 def lsp_estimates(y, tt, freqs):
17     '''
18     Inputs:
19     y - estimated output. should be an N x 1 numpy array.
20     tt - time indices. Should be an N x 1 numpy array.
21     freqs - frequencies to look for.
22     Outputs:
23     alpha_est - estimated amplitudes of the sinusoids.
24     phi_est - estimated phases of the sinusoids.
25     error_variance - variance of the error. MSE of
26     reconstructed signal compared to y.
27     theta - parameters such that ||y - Phi*theta|| is
28     minimized, where Phi is the matrix defined by
29     freqs and tt that when multiplied by theta is a
30     sum of sinusoids.
31     '''
32    N = y.shape[0]
33    f = freqs*2*np.pi
34    num_freqs = len(f)
35    Phi = np.empty((N, 2*num_freqs))
36    Sins = np.zeros((N,num_freqs))
37    Coss = np.zeros((N,num_freqs))
38    for i in range(N):
39        for j in range(num_freqs):
40            Sins[i,j] = np.sin(f[j]*tt[i])
41            Coss[i,j] = np.cos(f[j]*tt[i])
42
43    Phi[:,::2] = Sins
```

```

44 Phi[:,1::2] = Coss
45 Phi = np.hstack((Phi,np.ones((N,1))))
46 # when data is short, 'singular value' is important!
47 # 1 is perfect, larger than 10-5 or 10-6 there's a problem
48 theta = np.linalg.lstsq(Phi, y)[0]
49 y_hat = Phi@theta
50 #mean = theta[:,0]
51 theta_nomean = theta[:-1]
52 a_hat = theta_nomean[:,2]
53 b_hat = theta_nomean[1::2]
54
55 alpha_est = np.sqrt(a_hat**2 + b_hat**2)
56 #phi_est = np.arctan2(b_hat,a_hat)
57
58 #error_variance = float(np.linalg.norm(y - y_hat)**2/y.shape[0])
59 return alpha_est.flatten()#, phi_est.flatten(), error_variance, theta.flatten()
60 , y_hat
61
62 def quant(y,quantum):
63     return (np.round(y/quantum)*quantum)
64
65 #%% Allgemein
66 FREQUENCES_Complete = {'solar': 1/365.2421875, 'lunar': 1/29.530587981, 'O1':
67     0.929536, 'M1': 0.966446, 'P1': 0.997262, 'S1': 1.0, 'K1': 1.002738, 'N2':
68     1.895982, 'M2': 1.932274, 'S2': 2.0, 'K2': 2.005476}
69 FREQUENCES = {'M2':1.932274,'S2':2.0}
70 F=np.array(list(FREQUENCES.values()))
71
72 quanten=(1,10,20,30,36,40,50,60,80,100,110,130,150,180,200) # Quantisierungen
73 #quanten=np.arange(10,30,0.5)
74 ratios=(np.inf,2,0.5) # Sig to noise Ratios (genau drei angeben)
75
76 stdev_M2=np.zeros((len(quanten),len(ratios)))
77 stdev_S2=np.zeros((len(quanten),len(ratios)))
78 amp_M2=np.zeros((len(quanten),len(ratios)))
79 amp_S2=np.zeros((len(quanten),len(ratios)))
80
81 #allgemeine Signalparameter
82 samp_per_day = 24
83 amps=(10,10) # in mm
84 phases=(0,0) # in rad
85 days=456
86 N=samp_per_day*days
87 t=np.linspace(0,days,N)
88 anzahl=1000 # Anzahl an Durchlaeufer fuer statistische Aussage
89
90 idx_q=0
91 #%% gehe moegliche Quantisierungen durch
92 for q in quanten:
93     #Rechenfortschritt
94     prog=100*(idx_q/len(quanten))
95     idx_SNR=0
96     for SNR in ratios:
97         amp_f=np.zeros((2,anzahl)) # amp_f[0,i] fuer kleinere Frequenz M2
98
99         if SNR == np.inf:
100             y=sig(t,amps,phases,F,SNR)
101             amp_f[:,0]= lsp_estimates(quant(y,q), t, F)

```

```

99
100     stdev_M2[idx_q,idx_SNR]=0 #Standardabweichung von M2
101     stdev_S2[idx_q,idx_SNR]=0
102     amp_M2[idx_q,idx_SNR]=amp_f[0,0] #Mittelwert von M2
103     amp_S2[idx_q,idx_SNR]=amp_f[1,0]
104
105     else:
106         for i in range(anzahl):
107             y=sig(t,amps,phases,F,SNR)
108             amp_f[:,i]= lsp_estimates(quant(y,q), t, F)
109
110             stdev_M2[idx_q,idx_SNR]=np.std(amp_f[0,:]) #Standardabweichung von
111 M2
112             stdev_S2[idx_q,idx_SNR]=np.std(amp_f[1,:])
113             amp_M2[idx_q,idx_SNR]=np.mean(amp_f[0,:]) #Mittelwert von M2
114             amp_S2[idx_q,idx_SNR]=np.mean(amp_f[1,:])
115
116             #Rechenfortschritt anzeigen
117             prog+=100/len(quanten)*idx_SNR/len(ratios)
118             print("%.2f" % prog,"%")
119             idx_SNR+=1
120         idx_q+=1
121
122 %% Plot
123 plt.close("all")
124 fig, subplt = plt.subplots(len(ratios), len(F),sharex='col', sharey='row',
125                             gridspec_kw={'hspace': 0.05, 'wspace': 0})
126 fig.canvas.set_window_title('Einfluss der Quantisierung bei HLSQ')
127 fig.text(0.5, 0.02, 'Quantisierung in mm', ha='center', fontsize=15)
128 fig.text(0.04, 0.5, 'Amplitude in mm', va='center', rotation='vertical',fontsize
129 =15)
130 #plt.setp(subplt,xticks=quanten)
131 plt.setp(subplt,yticks=np.arange(0,11.5,4))
132 plt.setp(subplt,xticks=np.arange(10,30,5))
133 plt.tick_params(axis='both', which='major', labelsize=12)
134
135 for j in range(len(F)):
136     for i in range(len(ratios)):
137         subplt[i,j].grid(True, which='both')
138         subplt[i,j].set_xscale('log')
139         subplt[i,j].tick_params(axis='both', which='major', labelsize=12)
140
141         if j==0:
142             subplt[i,j].errorbar(quanten, amp_M2[:,i], stdev_M2[:,i], label='$f_{
143 M2}$ bei SNR=%.1f'%ratios[i], fmt='o',capsize=1.5)
144         if j==1:
145             subplt[i,j].errorbar(quanten, amp_S2[:,i], stdev_S2[:,i], label='$f_{
146 S2}$ bei SNR=%.1f'%ratios[i], fmt='ko',capsize=1.5)
147             subplt[i,j].plot((0,np.amax(quanten)), (10,10), 'r--')
148             subplt[i,j].legend(loc='lower left', fontsize=15)
149 plt.show()

```

Listing 1: Beispiel für Python Code zur Erstellung von Abbildung 3.7

Danksagung

Ein besonderer Dank gilt meinem Betreuer Dr. Gabriel Rau für die hilfreichen Anregungen und Kritiken beim Anfertigen dieser Abschlussarbeit. Danke auch an Prof. Dr. Thomas Bohlen für allgemeines Feedback zum Arbeitsfortschritt und für die hilfreichen Sprechstunden. Ich danke Daniel Schweizer, Petre Stoica und Jonathan Eric Tuck für die Unterstützung beim Programmieren und Bereitstellen von Programmcode. Ebenfalls möchte ich mich bei Dr. Thomas Forbriger und seinen Kollegen am Black Forest Observatory für die Führung durch ihre Forschungseinrichtung Anfang September 2019 bedanken. Dank sei auch den Mitarbeitern des Geophysikalischen Instituts und des Instituts für angewandte Geowissenschaften für die Lehrveranstaltungen und Beratungsgespräche, die mich auf diese Arbeit vorbereitet haben. Danke an alle Korrekturleser insbesondere meine Eltern, die mir mein Studium ermöglicht haben und mich stets mit einem offenen Ohr und Ratschlägen unterstützt haben.