Universität Stuttgart
Geodätisches Institut

Master thesis

Understanding the hydrological signature in gravity data

Philipp Schollmeier

October 14th, 2023

Supervisors: M.Sc. Clara Bützler
Dr.-Ing. Mohammad Tourian

Declaration of independence

With my signature I declare that I wrote this work independently and that I did not use sources other than those specified. I have fully identified all parts of this work whose wording, meaning or argumentation has been taken from other works (including the World Wide Web and other electronic text and data collections), stating the sources. I further declare that I wrote this work without the support of AI-based chatbots.

Stuttgart, 14.10.2023

place, date, signature

Abstract

Over the past two decades, the subsequent advancements in Superconducting Gravimeters (SGs) have ushered in a level of precision that enables the measurement of the impact of ground water and soil water on gravity. Because of the challenging nature of monitoring the total water volume and the relatively subtle amplitude of the hydrological signal, a comprehensive understanding of the precise hydrological signature in continuous gravity data remains elusive.

In this study, I use SG data in conjunction with hydrological measurements from a geoscientific observatory in Germany to find the signature of hydrological signals in gravity data. I scrutinize the various steps involved in extracting this signal, presenting new methodologies, including a technique to eliminate oscillations in gravity residuals that are likely attributed to remaining tidal signals due to an imperfect tidal model. A major contribution of this work involves constructing a data-driven model that incorporates precipitation and soil moisture measurements to elucidate gravity variations. I address critical questions such as the impact of utilizing soil moisture data on the model's performance, determining the optimal model for achieving the closest fit with gravity measurements, and assessing the applicability of computed model parameters to new epochs. Furthermore, I provide recommendations for refining the model-building process in future investigations.

Results show that a convolution of the different hydrological timeseries with one half of a Gaussian bell curve leads to a strong agreement with the gravity measurements. The use of soil moisture data significantly improves the fit, especially when the measurement stations are spatially well distributed. This fit becomes less strong when the computed parameters are applied to new events, but the approach showed promise for some of the events.

Enhancing our comprehension of the hydrological influence on gravity measurements holds promising implications, potentially positioning SGs as instruments for monitoring soil and ground water in the future. Moreover, this improved understanding could elevate the precision of analyzing other subtle signals, such as the effects of Polar Motion.

Summary in German

Eine zentrale Aufgabe der Geodäsie ist die Bestimmung der physikalischen Form der Erde. Diese wird realisiert durch die Messung ihres zeitvariablen Gravitationsfeldes. Die gegenwärtig genauesten Instrumente zur Bestimmung der Schwereänderungen an einem festen Ort sind Supraleitende Gravimeter (SG). In ihrer Messeinheit wird eine Niob-Kugel von einigen Zentimetern Durchmesser mit einem sehr stabilen Magnetfeld zum Schweben gebracht. Bei einer Positionsänderung der Kugel wird die Stärke des Magnetfeldes nachgesteuert und diese Steuerung wird auf eine Änderung der Schwere zurückgeführt.

Die gemessenen Schwereänderungen werden allerdings nicht nur durch Veränderungen der gravitationellen Anziehung der Erde ausgelöst, sondern weitere Vorgänge wie Gezeiten oder Variationen des Zentrifugalvektors an der Messposition, welche z.B. durch Polwanderung verursacht werden, beeinflussen die Messungen. Ein eher kleiner Teil der Messungen sind die Auswirkungen der Hydrologie.
Der Einfluss der Hydrologie ist aufgrund der relativ kleinen Amplitude schwierig zu bestimmen. Außerdem können die hydrologischen Vorgänge selbst nur unvollständig überwacht werden. Während bei Gezeiten die Position der verursachenden Himmelskörper aus astronomischen Katalogen genau bekannt ist, kann man das Volumen des gefallenen Regens nicht überall, sondern nur an wenigen Punkten, bestimmen. Das Ziel dieser Arbeit ist ein besseres Verständnis, wie sich Hydrologie auf die Schweremessungen auswirkt. Es soll ein daten-getriebenes Modell erstellt werden, welches mittels Bodenfeuchte- und Niederschlagsmessungen die Schwere modelliert. Um die Signatur der Hydrologie zu finden, müssen zunächst die größeren Effekte auf die Schwere so gut wie möglich aus der Zeitreihe entfernt werden.
Ein besseres Verständnis der Hydrologie in SG-Messungen könnte zu neuen Einsichten in die kleineren Signale im Schweresignal führen, wie Oszillationen des Erdkerns, Beobachtungen von vulkanischen Gefahren und, vielleicht am wichtigsten in einer Zeit, in der Trinkwasser eine kostbare Ressource wird, Rückschlüsse auf Grundwasser-Höhen.

Für die Untersuchungen stehen Messungen des Geowissenschaftlichen Gemeinschaftsobservatoriums (BFO) in Schiltach zur Verfügung. Dort steht in einer ehemaligen Silbermine, gut abgeschirmt von menschlichen Störquellen, das Supraleitende Gravimeter SG-056. Das Instrument besitzt zwei Schweresensoren, die einen vertikalen Abstand von etwa 20 cm haben, und erreicht eine Genauigkeit von etwa 0.01 µGal bei einer Rauschperiode von 1 Minute. Es stehen Beobachtungen mit einer Messrate von 60 Sekunden über das gesamte Jahr 2022 zur Verfügung. Darüber hinaus verwende ich die Daten eines Niederschlags-Sensors am Eingang der Mine sowie von vier Bodenfeuchte-Stationen (mit insgesamt sieben Sensoren), die im Laufe des Jahres 2022 im Waldboden oberhalb des SG installiert worden sind.

Im ersten Schritt müssen die weiteren Schweresignale verstanden und abgezogen werden. Die Gezeiten werden mit einer synthetischen Zeitreihe, welche auf einer Gezeitenanalyse

von Eva Schroth basiert, subtrahiert. Der Atmosphärendruck hat einen Einfluss auf die gemessene Schwere, unter anderem durch die gravitative Wirkung der zusätzlichen Masse über dem Gravimeter. Für die Korrektur werden die Messungen eines Luftdruck-Sensors des BFO mit einem festen Faktor multipliziert und von den Messungen abgezogen. Eine Kleinste-Quadrate-Ausgleichung ermittelt einen besten Wert von $-0.246\,\mu\text{Gal/hPa}$. Da die Auswirkungen der Atmosphäre aber deutlich komplizierter sind, dürfte ein einzelner Korrekturfaktor nicht ausreichen. Vor allem für schnelle Druckänderungen mit einer Periode von einigen Tagen oder weniger ist dieser Wert nicht ideal.
Nach dieser Korrektur sind in den Schwereresiduen noch übrige Oszillationen zu sehen, die nicht durch die Hydrologie zu erklären sind. Ich komme zu dem Schluss, dass diese durch Imperfektionen der Gezeiten-Korrektur verursacht werden. Die Schwereresiduen zeigen eine Korrelation mit den synthetischen Gezeiten, besonders nach einer Verzögerung von etwa eineinhalb Stunden. Mit einer Kreuzkorrelation wird der Wert dieser Verzögerung bestimmt und mit einer Kleinste-Quadrate-Ausgleichung eine Amplitude zwischen den beiden Zeitreihen berechnet. Die synthetischen Gezeiten werden dann – mit der bestimmten Verzögerung verschoben und der Amplitude multipliziert – von der Schwerereihe abgezogen. Das resultierende Signal zeigt sowohl visuell als auch in einer Frequenzanalyse deutlich weniger Oszillationen. Als letzte Korrekturen des Signals werden ein linearer Trend aus den Messungen beider Sensoren entfernt sowie eine Ausreißer-Eliminierung durchgeführt.

Auf Grundlage der Regenmessungen bestimme ich im Jahr 2022 elf bedeutende Regenereignisse, auf denen ich die Modellierung durchführe. Ich beschränke mich auf solche Ereignisse, da in diesen Zeitabschnitten ein deutlicheres hydrologisches Signal vorhanden sein dürfte, als wenn es über lange Phasen fast gar nicht regnet. Zu jedem Regen-Event werden drei Tage von Schwere-Beobachtungen zur Modellierung verwendet.
Beim Design des Modells gehe ich davon aus, dass die Eingangsgrößen – Niederschlag und Bodenfeuchte – für unterschiedliche Ereignisse die gleiche Signatur in den Schweredaten hinterlassen. Das heißt, dass z.B. ein Regenpaket sich zunächst stark auf die Schwere auswirkt, nachdem es gefallen ist, und danach die Auswirkungen kleiner werden, weil das Wasser mit der Zeit abfließt. Das bedeutet mathematisch, dass die Bodenfeuchte- und Niederschlagsmessungen mit einem Kern gefaltet werden. Die Parameter dieser Kerne, die Amplituden und die Verfallszeiten werden zunächst mit einer Gittersuche näherungsweise bestimmt und danach mit einer nicht-linearen Ausgleichung genauer. Sowohl für Regen als auch für Bodenfeuchte werden verschiedene Arten von Kernen getestet. Mit der hinteren Hälfte einer Gauß-Kurve zeigt sich jeweils die beste Übereinstimmung mit den Messungen hinsichtlich des RMSE. Für den Niederschlag werden aber sehr viel längere Verfallszeiten, teilweise über einen Monat, gefunden, als für die Bodenfeuchte.
Bei Hinzunahme von mehr Bodenfeuchte-Sensoren zum Modell wird der RMSE-Abstand zu den Messungen stetig kleiner. Die Verbesserungen lassen allerdings nach dem dritten Sensor nach und die Korrelation mit den Messungen wird nicht mehr besser. Es zeigt sich, dass ein Modell mit Sensoren, die räumlich näher zusammen liegen, eine schlechtere Übereinstimmung bringt. Die beiden Hauptmodelle sind aus diesem Grund P3S, welches Regen und drei gut verteilte Bodenfeuchte-Sensoren nutzt, sowie P7S welches alle sieben Sensoren verwendet. Die besten Modelle erreichen eine Korrelation mit den Messungen über die 33 Beobachtungstage von etwa 0.86. Diese Ergebnisse sind jeweils deutlich besser als das bisher verwendete Bouguer-basierte Modell, das nicht an die Daten angepasst wird,

sondern feste Zeitkonstanten und Amplituden besitzt. Das Bouguer-Modell erreicht über die 33 Tage nur eine Korrelation von 0.76 mit den Messungen.
Es kann keine gute physikalische Interpretation der ermittelten Parameter gefunden werden. Die Amplituden für die Bodenfeuchte haben häufig ein positives Vorzeichen, während physikalisch gesehen nur ein Negatives Sinn ergeben würde, da sich die zusätzliche Wassermasse im Boden oberhalb des Gravimeters in der Mine befindet.
In einem letzten Versuch wird ein Modell mit variablen Parametern angesetzt. Wenn der Boden feucht ist, kann er mehr von dem gefallenen Regen speichern, was sich zum Beispiel in einer höheren Amplitude für den Niederschlags-Kern ausdrücken könnte. Daher wird für die Amplitude und die Verfallszeit des Regen-Kerns eine lineare Abhängigkeit von der Bodenfeuchte vor Beginn des Regens eingeführt. Das resultierende Modell P3SV bringt aber keine wesentliche Verbesserung gegenüber den anderen beiden Bodenfeuchte-Modellen.

Schließlich teste ich die Übertragbarkeit der Modelle P3S, P3SV und P7S sowie des OnlyP-Modells, welches nur den Regen verwendet. Dazu führe ich eine Auslassungsvalidierung durch. Dabei führe ich die Bestimmung der Parameter nur auf acht der elf Ereignisse durch und beobachte die Ergebnisse dieser Modelle auf den drei übrigen Events. Die Vorteile der datengetriebenen Modelle gegenüber dem Bouguer-Modell sind nun verschwunden. Alle neu berechneten Modelle schneiden ähnlich oder schlechter ab als der Bouguer-Ansatz. In einem zweiten Schritt wende ich die Modelle auf neue Regenereignisse im Jahr 2023 an. Auch hier zeigt meist das Bouguer-Modell die besten Ergebnisse. Lediglich OnlyP kann ähnlich gute Ergebnisse wie die Baseline erzielen.

Eine Übertragung der Modelle ist also eher nicht möglich. Verschiedene Gründe dürften für dieses Resultat verantwortlich sein: Der von mir bestimmte atmosphärische Faktor ist suboptimal für schnelle Luftdruckänderung und verfälscht damit die Schwereresiduen. Aus diesem Grund sind die Ergebnisse für einige Perioden, speziell die Regenereignisse in 2023, sehr schlecht. In Zukunft sollte ein frequenzabhängiger Faktor verwendet werden.
Die Beobachtungsgleichung ist nicht ideal. Dass alle unterschiedlichen Sensoren mit verschiedenen Kernen gefaltet werden, ist physikalisch schwer zu begründen. Man könnte zum Beispiel die verschiedenen Bodenfeuchtemessungen zusammenrechnen und dann erst auf die Schwere übertragen. Das könnte auch zu Modellen mit weniger Parametern führen. In der Übertragung auf die Ereignisse in 2022 und 2023 zeigte sich, dass P7S, welches die meisten Parameter verwendet, am schlechtesten übertragbar ist.
Zudem wären mehr als 33 Tage für die Berechnung der Parameter erstrebenswert, um eine bessere Chance auf eine allgemeinere Gültigkeit der resultierenden Modelle zu erhalten. Bei der Vorgehensweise wäre es in Zukunft außerdem gut, wenn die berechneten Modelle häufig auf unabhängigen Epochen validiert werden, damit man direkt beurteilen kann, ob man das Modell in die richtige Richtung entwickelt oder nicht.
Eine perfekte Übertragbarkeit eines datengetriebenen Modells ist aber ohnehin ein sehr hoch gestecktes Ziel. Die Modelle funktionieren gut auf den Epochen, wo sie modelliert wurden. Durch Hinzunahme von Bodenfeuchtedaten wird eine deutliche Verbesserung der Übereinstimmung erreicht. Die verwendete iterative nicht-lineare Kleinste-Quadrate-Ausgleichung scheint zur Bestimmung solcher Modelle gut geeignet zu sein. Mit einigen der vorgeschlagenen Verbesserungen wie einem physikalisch realistischeren Modell sollten dann auch Einsichten in Bereiche wie Geologie oder Grundwasserbeobachtung möglich sein.

Contents

1. Introduction

One of the main tasks of geodesy is determining the physical shape of the Earth, which is realized by the measurement of its time-variable gravity field (Van Camp et al., 2017). Gravity mainly consists of the gravitation of the Earth and the centrifugal acceleration caused by its rotation (Van Camp et al., 2017). You can measure gravity by dropping a mass in a vacuum and observing the time it takes to travel a determined distance; that is the principle of a free-fall gravimeter (Van Camp et al., 2017). Other forms of gravimeters are pendulum and spring gravimeters (Faller, 1967; Lacoste, 1934). For the latter one, the basic principle is to measure the extension of a spring with a mass attached to its end (Van Camp et al., 2017). A variation of this type is the superconducting gravimeter (SG) (Van Camp et al., 2017). Instead of a mechanical spring as suspension, it uses a very stable magnetic field (Goodkind, 1999). In this way, the SG reaches a precision that is about two orders of magnitude better than that of a free-fall gravimeter (Van Camp et al., 2017).

A relative gravimeter like the SG does not measure the absolute value of gravity but rather its time differences (Van Camp et al., 2017). The recorded timeseries is the sum of different gravitational signals. The largest are caused by the gravitational attraction of the sun and the moon; these are the tides. Smaller effects arise because of highly variable masses on the Earth or close to its surface, like ocean tides, the atmosphere, or soil water (Van Camp et al., 2017). Another aspect is the change in position of the gravimeter in response to tides (Earth tide) or atmospheric pressure. The centrifugal vector can change due to Polar Motion or changes in Earth's rotation speed (length of day tides) (Van Camp et al., 2017). There are instrumental effects such as drift, outliers, or periods of lower-quality data. These can be caused, for example, by problems with the cooling. Finally, gravitational sources in Earth's interior, like differences in groundwater or mass transport in Earth's mantle, can influence the gravity readings. The following equation summarizes the signals that compose the observations (Hinderer et al., 2007):

$$g_{\text{(obs.)}} = g_{\text{(disturb.)}} + g_{\text{(tides)}} + g_{\text{(atm.)}} + g_{\text{(Polar Mot.)}} + g_{\text{(LOD Tide)}} + g_{\text{(drift)}} + g_{\text{(hydro)}} + g_{\text{(other)}}. \quad (1.1)$$

$g_{\text{(obs.)}}$ is the signal that is observed by the SG. $g_{\text{(disturb.)}}$ are disturbances in the data that have instrumental or environmental causes, like earthquakes. $g_{\text{(tides)}}$ includes the gravitational attractions of the different celestial bodies and the gravity effect of the Earth tides. $g_{\text{(atm.)}}$ are the effects of mass changes in the atmosphere. An increased mass above the gravimeter reduces the measured gravity because of its gravitational attraction. On the other hand, it forces the ground down, which increases the experienced gravity at the SG. A change in the rotation axis of the Earth relative to its surface causes a difference in the angular distance between the position vector of the measurement site and the rotation axis (Van Camp et al., 2017). This means the centrifugal acceleration due to the Earth's rotation is altered. These changes are captured in the term $g_{\text{(Polar Mot.)}}$. $g_{\text{(LOD Tide)}}$ is the signal that is caused by changes in the centrifugal vector because of the changing rotation speed of the Earth. Figure 1.1 shows how these signals may look. Superconducting gravimeters have instrumental drifts of a few µGal/yr. $g_{(\mathrm{drift})}$ contains this drift. $g_{(\mathrm{hydro})}$ is the main interest of this work. Water on top of and in the soil, as well as ground water, has a gravitational effect on the gravimeter. $g_{(\mathrm{other})}$ comprises different phenomena like glacial isostatic adjustment (GIA), slow earthquakes, or tectonics. In order to analyze the timeseries for a smaller term like $g_{(\mathrm{hydro})}$, the larger signals first have to be understood and precisely subtracted.

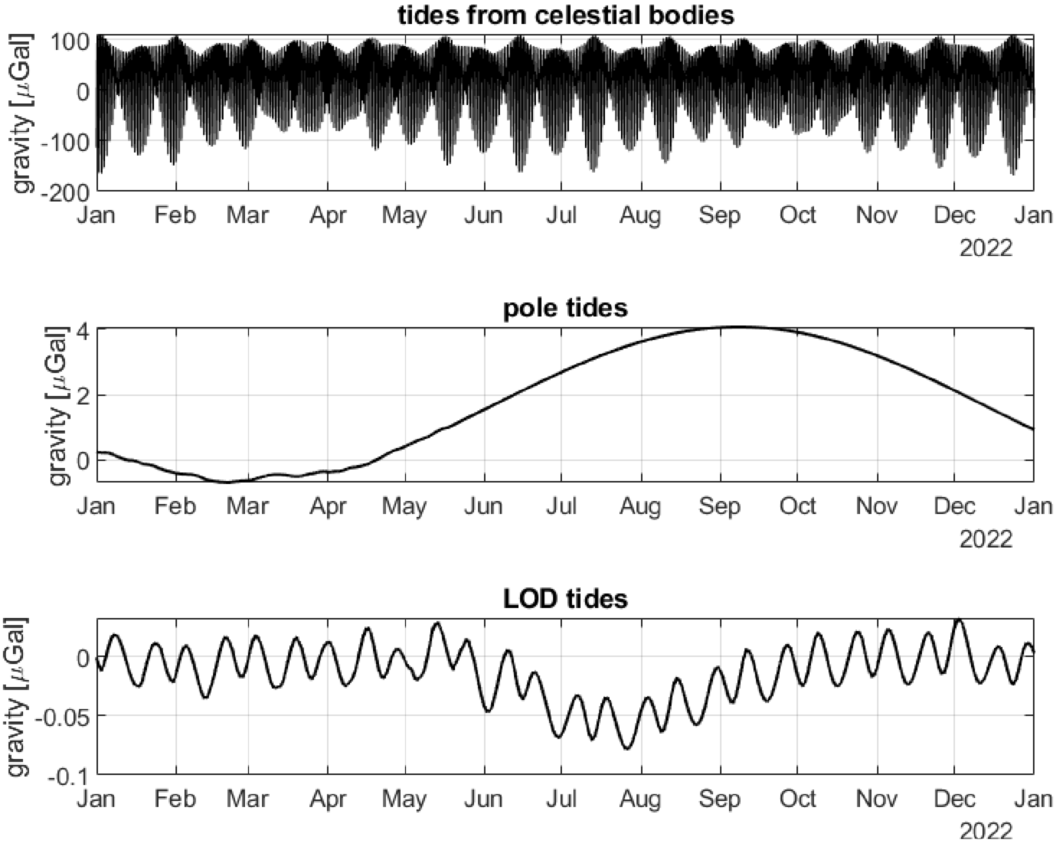

Figure 1.1.: The timeseries of tides, Pole tides, and LOD tides from a Synthetic Tides timeseries over 2022. The computation of this timeseries will be explained in detail in Section 4.2.

As the hydrological effect will only cause a very small part of the whole signal, it is difficult to identify. An approximation with a Bouguer plate and rain measurements concludes that the maximum amplitude of the signal will only be a few µGal (see Section 5.1) while the observed gravity changes reach about 250 µGal. An additional difficulty is the hydrological measurement itself. For other signals, the observation of the causative processes is easier. For example, for the tides, the positions of the celestial bodies are well known from astronomical catalogs, and the Earth tides can be surveyed by fixed GPS stations. For hydrology, the exact water masses inside the soil are very complicated. A rain gauge only measures precipitation at a single point and not throughout the region. These measurements also cannot capture the speed of dissipation or the water running away. You cannot know the water mass at more than a few metres of depth without a lot of effort in installing the sensors and significantly altering the system. Because of these challenges, the hydrological signal in gravity measurements is not well determined. In this work, I will use hydrological and gravity measurements of the Black Forest Observatory (BFO) to estimate this signal for that location. The hydrological measurements not only comprise recordings of a rain gauge but also of six soil moisture stations that were newly deployed in 2022. I will try to find a mathematical model for gravity that uses the hydrological data and fits the SG data. This model will be called *hydrological model* during this thesis. A good understanding of the hydrological signal in SG gravity series around the world can lead to new insights into smaller parts of the measurements like geological processes, oscillations of the Earth's core (Rosat and Hinderer, 2011), observation of volcanic hazards, and, perhaps most importantly, in an age where drinking water is becoming a sparse resource, groundwater levels (Van Camp et al., 2017).

The first goal of this master thesis is to investigate the different steps of the gravity corrections and find a good solution for them. These corrections are trying to accurately subtract the other signals in (1.1) of the observed signal. The final gravity series should, in the best possible way, reflect how gravity is hydrologically influenced. Then this time-series of gravity residuals will be analysed in relation to the precipitation and soil moisture measurements to build a mathematical model that explains the biggest part of the signal. So far, BFO only uses the precipitation data in a rough model that is based on the physical view of a Bouguer plate and a fixed slow disappearance of the rain influence due to dissipation (Schroth, 2013). In this thesis, I will develop a new approach that is mostly data-driven and takes physical constraints less into account. Which kind of function can best describe the gravity series? How well can the model explain the gravity measurements? Is the usage of soil moisture data beneficial for the model, or is precipitation sufficient? Which sensors should be used in the model? Is it always suitable to use as many probes as possible? Do the adjusted parameters have any physical significance? Can a model with parameters that change with soil moisture explain gravity better than one with constant parameters? These are the questions that I will try to answer.

Chapter 2 gives a detailed description of the measurement principle of superconducting gravimeters and of the SG at the BFO. Chapter 3 presents all the used data and the first pre-processing steps for them. Chapter 4 details the different corrections that have to be applied to the gravity series before any hydrological analysis can be done. For some steps, like the subtraction of atmospheric influence and the elimination of unwanted oscillations, different methods are compared. Chapter 5 first explains and evaluates the hydrological model that was established by Schroth (2013). After that, I will try to improve the results of that baseline by building my own model. This is done by successively adding soil moisture data to the adjustment while comparing different convolution kernels and combinations of sensors. In the last section of that chapter, I will develop a model that uses soil moisture-dependent parameters. Chapter 6 will compare different models that I have built in the previous chapter over the 2022 data but also apply the models to new data from 2023. I will try to quantify the improvement of my models over the baseline of Schroth (2013) and decide which model is the best to use in the future. Chapter 7 will discuss the results that I have found and give an outlook on which parts of the pre-processing and model development can be improved in the future.

2. Superconducting Gravimeter

2.1. Physical principle

The SG is the instrument of choice for long-term continuous gravity measurements (Krause et al., 2009). It is a spring-type gravimeter but without the actual spring (Goodkind, 1999). Therefore, it does not suffer from the wear of the spring and remains stable over a long time (Hinderer et al., 2007).

Instead, the connection to the test mass is realised by a magnetic field (Hinderer et al., 2007). The test mass is a hollow niobium sphere with a diameter of a few centimeters (Hinderer et al., 2007). Niobium is, like most metals, a superconducting material (Peiniger and Piel, 1985). That means at a certain critical temperature T_c, 9.2 K in this case (Peiniger and Piel, 1985), its electrical resistance suddenly drops to zero and its interior becomes free of magnetic and electrical fields (Tinkham and Emery, 1996). If an exterior magnetic field is applied to the superconductor below its critical temperature, as stated in Faraday's law, currents are induced on the surface of the sphere that compensate magnetic flux in the interior of it (Hinderer et al., 2007). The two magnetic fields repel each other, and therefore the superconducting sphere is forced upwards (Hinderer et al., 2007).

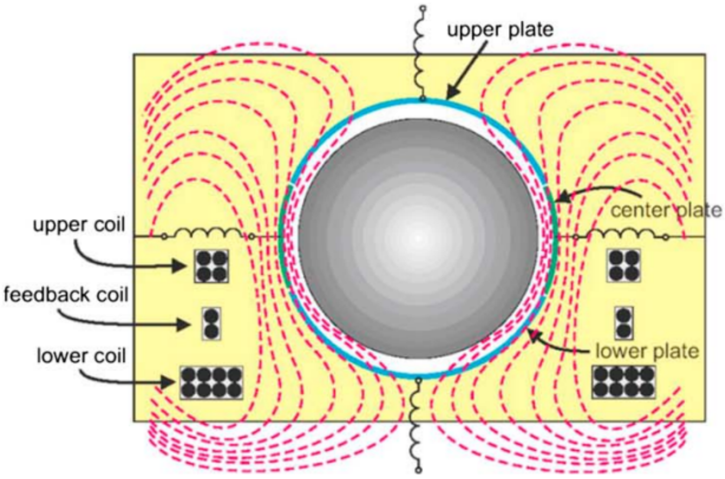

Figure 2.1.: Presentation of an SG sensor; the niobium sphere is in the middle; the three coils are shown as well as the plates that measure the position of the sphere (see Section 2.2); the magnetic field (red lines) is expelled from the interior of the sphere (Hinderer et al., 2007)

This is exactly what happens in an SG: currents in two coils produce the magnetic field; the niobium sphere, which is cooled to about 4 K by liquid helium, is forced upwards and starts to levitate (Van Camp et al., 2017). This upward force is the analogue of the restoring force of the spring in a spring gravimeter (Schroth, 2013). As gravity changes, the sphere will be forced up or down. The sphere will be kept in a stable position by changing the current in a third coil in response to the movement of the sphere (Goodkind, 1999). This needed current will be recorded as the observation quantity, and its changes are related back to gravity changes (Forbriger and Heck, 2018). This calibration can be done, for example, with measurements of absolute gravimeters at the same time (Forbriger and Heck, 2018). Figure 2.1 depicts the layout of such a sensor with the niobium sphere and the three coils.

2.2. Different parts in detail

The sphere is not the only part of the SG that is superconducting. Similarly, the coils that are producing the main magnetic field for levitation are, so no continuous power is needed for the production of the field (Goodkind, 1999). First, a heater raises the temperature above the critical temperature of the coils (Goodkind, 1999). A current is applied to the coils, and the heater is turned off (Goodkind, 1999). Once the temperature is below T_c the current is trapped inside the coils (Hinderer et al., 2007).

That means the magnetic field from the levitating coils stays extremely stable over a long time (Westerhaus et al., 2018). The control of the vertical position of the sphere therefore requires an additional feedback coil (Goodkind, 1999).

By adjusting the ratio of the currents in the upper and lower coils, the restoring force can be arbitrarily close to zero (Goodkind, 1999). That means that a very small change in current causes a big displacement (Hinderer et al., 2007). That is one of the big strengths of the SG, as tiny differences in gravity cause measurable displacements and can be accurately corrected (Hinderer et al., 2007). A schematic depiction of the gravity sensing unit can be seen in Figure 2.2.

The niobium sphere and the coils are cooled to a temperature of 4.2 K with liquid helium (Westerhaus et al., 2018). This is done with a Dewar, which is a multi-walled container, for example, made of steel (Hinderer et al., 2007). Between the walls is a vacuum to prevent heat exchange, and the inside is filled with liquid helium around the sensing unit (Hinderer et al., 2007). There will still be some heat leak from the outside, and the helium slowly evaporates (Goodkind, 1999). A refill of helium would interrupt the measurements, so modern SGs are equipped with a cold head which condensates the helium again.

The position of the sphere is measured with a capacitance bridge (Goodkind, 1999). There is about 1 mm space between the sphere and the sphere-shaped plates of the bridge (Hinderer et al., 2007). By the measurements of the bridge, the current of the feedback coil will be adjusted so the sphere stays near the zero position (Goodkind, 1999). The measured quantity is the voltage, which is proportional to the current in the feedback coil (Forbriger and Heck, 2018). The data is acquired by high-quality digital voltmeters that reach 7.5 decimal digits of resolution (Hinderer et al., 2007). The data is sampled at higher frequencies, than which will later be analysed (Goodkind, 1999). At the BFO, this is done at 20 kHz and then low-pass filtering and anti-aliasing are carried out to get rid of instrumental events and, most importantly, marine microseism (personal communication from Walter Zürn). In the case of this work, the final product is recordings with a sampling rate of 60 s.

A number of possible disturbances need to be controlled. The whole assembly of coils and the sphere is placed inside a superconducting cylinder so that no external magnetic fields influence the sphere (Hinderer et al., 2007). A mu-metal shielding outside of the vacuum

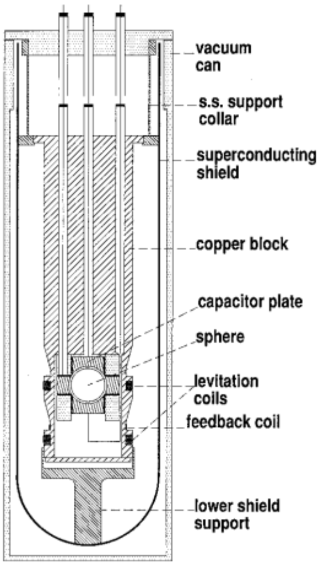

Figure 2.2.: Layout of the gravity sensing unit of an SG; the additional feedback coil is visible (Goodkind, 1999)

can further reduce the influence of Earth's magnetic field (Goodkind, 1999). Temperature is kept very stable at the mK level by the liquid helium, as small differences influence the superconducting penetration depth (Goodkind, 1999). That is the thickness of the layer where the currents in the sphere appear that compensate the magnetic field.
A tilt of the instrument would cause a movement of the levitation coils relative to the sphere; therefore, the location of the sphere in the magnetic field changes, and this can lead to a change in the levitating force (Hinderer et al., 2007). Tilt is usually controlled by two tiltmeters in orthogonal axes that are situated on the vacuum can (Goodkind, 1999). The readings of the tiltmeters are used to keep the instrument tilt stable by adjustment of micrometer screws (Goodkind, 1999). The electronics of the SG can be influenced by radio waves, so the whole assembly is placed inside a radio frequency shielding (Goodkind, 1999).

2.3. The SG-056 at the BFO

The Black Forest Observatory was founded in 1971 and since then is run by the Geophysical Institute (GPI) of the Karlsruhe Institute of Technology and the Geodetic Institute (GIS) of the University of Stuttgart (Zürn, 2014). It is located in the Heubach valley near the village of Schiltach in the Middle Black Forest in Baden-Württemberg. The location can be seen in Figure 2.3. The bedrock at Schiltach is made up of granite, and then closer to the surface, the softer „Buntsandstein“ (Zürn, 2014).
The site was chosen in the 1960s because of its remoteness and stable ground (Zürn, 2014): In contrast to the much softer and wetter ground of the Rhine valley, the granite bedrock of the Middle Black Forest ensures that the location of the instruments does not move much in response to changing atmospheric pressure (personal communication from Rudolf Widmer-Schnidrig). Such movement would not only disturb the precise magnetic and seismic measurements that are carried out, but also the gravimetric measurements. Vertical ground movement causes a gravity change of about 2 µGal/cm (Van Camp et al.,

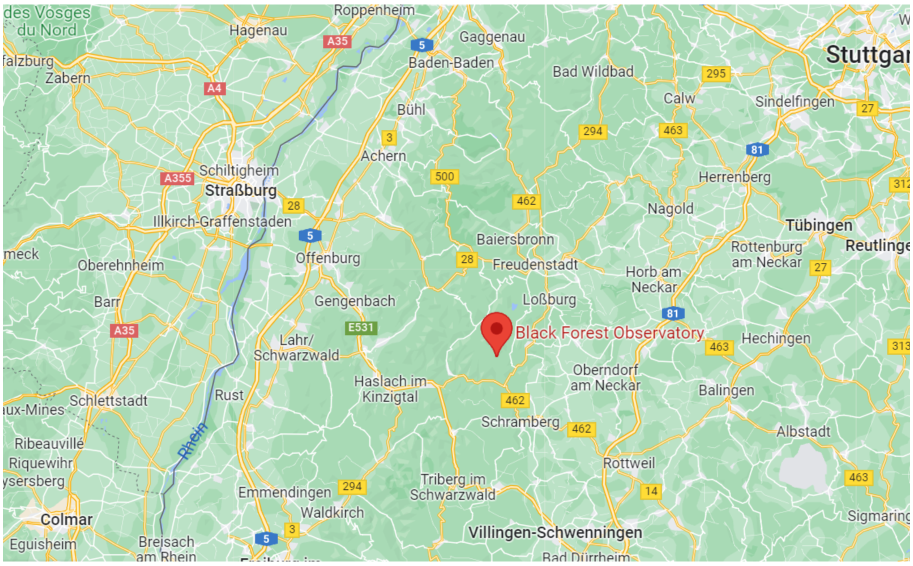

Figure 2.3.: The location of the BFO in western Baden-Württemberg (by Google Maps)

2017). The second big advantage is the surroundings. Being at the edge of a quiet valley, there are few sources of human noise, like roads or construction sites (Westerhaus et al., 2018).

Most of the instruments are set up in an old silver and cobalt mine, called the mine „Anton“. The main tunnel is 700 m long and has many side tunnels and vaults. It contains the SG that is located about 500 m inside the mountain and 100 m below the surface (Zürn, 2014). The gravimeter sits between two airlocks that are closed most of the time. They provide a thermally extremely stable environment at close to 10 °C and prevent any rapid pressure changes in the SG vault.

Other locations of the observatory include the laboratory hut. The outside pressure sensor is on top of the lab, and the precipitation gauge is a few metres away from it in an area clear of any trees or other obstructions. There is also the „amagnetic hut“ that is about 200 metres up the mountain with an additional pressure sensor and a precise GNSS antenna. Finally, six soil moisture stations were installed in the forest above the SG over the year 2022.

The first gravimetric measurements were performed in 1973, and since 2009, the BFO has owned a state-of-the-art dual-sphere superconducting gravimeter by GWR from San Diego, California (Zürn, 2014). It is surrounded by a container that keeps the SG dry. Two sensors have the advantage that instrumental issues can be identified more easily and can give insights into the uncertainties of the measurements (Kroner et al., 2005). The SG-056 is a special version as it not only has two spheres (niobium spheres of 2.54 cm diameter, with a vertical offset of about 20 cm, both placed in the same Dewar) but these have different masses at 17.7 g (G1, lower, about 32 cm above the ground) and the standard 4.34 g (G2, higher) (Forbriger and Heck, 2018; Zürn and Duffner, 2018). This was done to see whether the sphere with the higher mass suffers from less noise due to Brownian motion (Westerhaus et al., 2018). The Dewar is filled with liquid helium that keeps the sensor temperature at 4.5 K. A cold head reliquefies the evaporated helium. Another special feature of the SG is that because of the stable ground, tilt correction is turned off (Zürn, 2014).

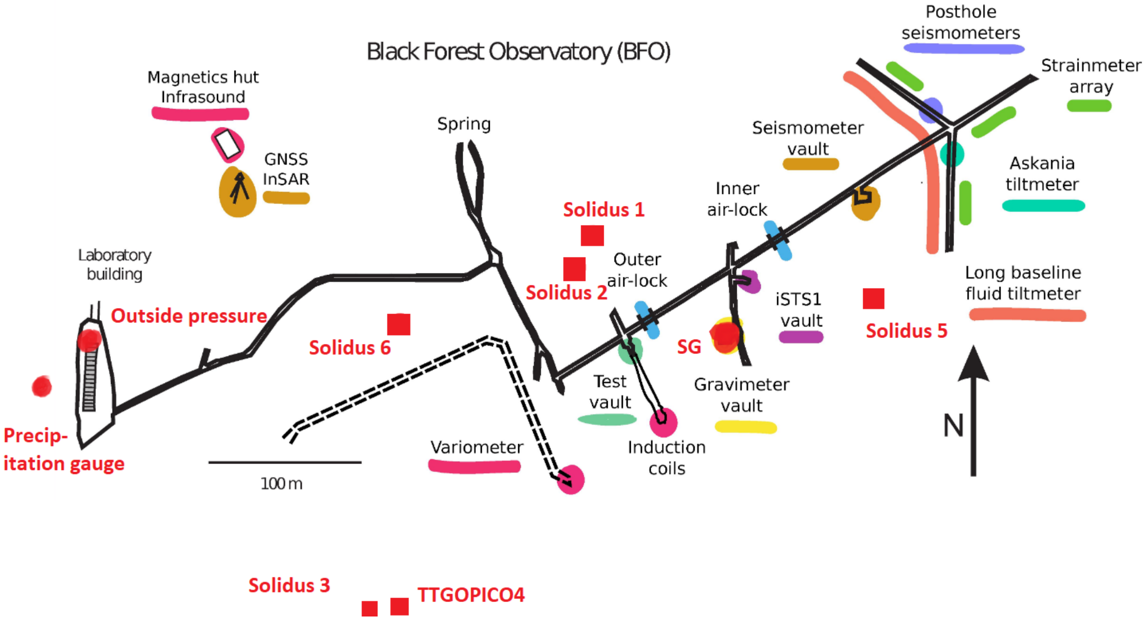

Figure 2.4.: The layout of the BFO (BFO, 2023); the important locations for this work are highlighted in red. The soil moisture stations are marked by red rectangles.

Figure 2.5.: The superconducting gravimeter of the BFO. The blue container on the left side is the Dewar. The silver cube on top is the cold head (Duffner et al., 2018)

An SG can reach an accuracy of 0.1 nm/s^2 for noise at a 1 min period (Van Camp et al., 2017). The noise is believed to be mostly produced by Brownian motion and was observed to be lower for the heavier sphere (Westerhaus et al., 2018). The sensors also experience drifts of around 1-5 µGal/yr (Hinderer et al., 2007). They are different for the two sensors and are believed to be caused by the ageing of the getters that are delivering the liquid helium. This ageing causes a slow change in pressure inside the Dewar, and this might be responsible for the observed drifts (personal communication from Walter Zürn).
Rosat and Hinderer (2011) compared thirty superconducting gravimeters over the world and concluded that the heavier sphere of SG-056 showed the best noise levels over frequencies larger than 0.1 mHz; that equals a period of just under three hours. This speaks for the favourable surroundings of the BFO location and also for the fact that raising the mass of the sphere is beneficial to noise levels.

3. Data and case studies

Table 3.1 gives an overview of the data that were used in this study. All measurements are from 2022. The main timeseries are the measurements of the two sensors of SG056, G1-F60 and G2-F60 with a sampling rate of 60 s. The other important ones are tunnel pressure (p) and outside pressure that are used for atmospheric correction; the precipitation (prec) gauge and the six soil moisture (sm) stations are used for the main model to study the hydrological influence on the gravity readings. The stations record soil moisture with one or two probes, as well as temperature at each probe. Additionally, a timeseries of synthetic tides with a sampling rate of 60 s is available for this most important correction (see Chapter 4). The acquisition and provision of the gravimeter and the precipitation gauge were done by the staff of the BFO (Black Forest Observatory (BFO), 1971).

Table 3.1.: Available data recorded at the BFO

quantity	name	time window	sampling rate	remarks; height [m]
gravity	SG056	01.01.-20.12.	1 min	2 sensors; 589
p tunnel	Br1-F60	01.01.-20.12.	1 min	590
p outside	Br2-F60	01.01.-20.12.	1 min	data outage; 591
precipitation	DL1	01.01.-20.12.	1 min	590
soil moisture	Solidus 1	24.01.-31.12.	ca. 10 min	2 probes, SMT100; 710
soil moisture	Solidus 2	30.04.-31.12.	ca. 10 min	2 probes, SMT100; 705
soil moisture	Solidus 3	30.04.-31.12.	ca. 10 min	2 probes, SMT100; 660
soil moisture	Solidus 5	26.07.-31.12.	ca. 10 min	2 probes, IMKO; 760
soil moisture	Solidus 6	...	ca. 10 min	unusable; 660
soil moisture	TTGOPICO4	14.03.-31.12.	ca. 6 min	1 probe, IMKO

In this chapter, the data that are used for the hydrological model are presented in more detail, namely precipitation, soil moisture, and gravity. The first pre-processing steps with these timeseries will be explained as well.

3.1. Precipitation

The precipitation gauge is a few metres west of the laboratory hut (see Figure 2.4). Figure 3.1 shows a photo of the gauge. The horizontal distance to the SG is about 400 m. It is a gauge of the tipping bucket style. A small bucket is emptied after 0.1 mm of rain has fallen, and a logger records how often emptying happens. This yields a count of tippings for every minute over the 365 days. There are no data gaps in the 365 × 1440 epochs, but it is possible that the logger has not picked up tippings at times and just wrote 0 there.

Because there are a lot of epochs with a count of 0 when there was no significant rain, it is almost impossible to find out whether such data outages occurred.

Figure 3.1.: BFO rain gauge, from the direction of the laboratory hut, looking west; own picture, taken in January 2023

Looking at the counts for every minute (Figure 3.2) shows how much the precipitation rate varies. The rain events are difficult to distinguish, except for some heavy rains. In summer, there are a few strong maximums, while in winter, periods of constant light rain can be seen.

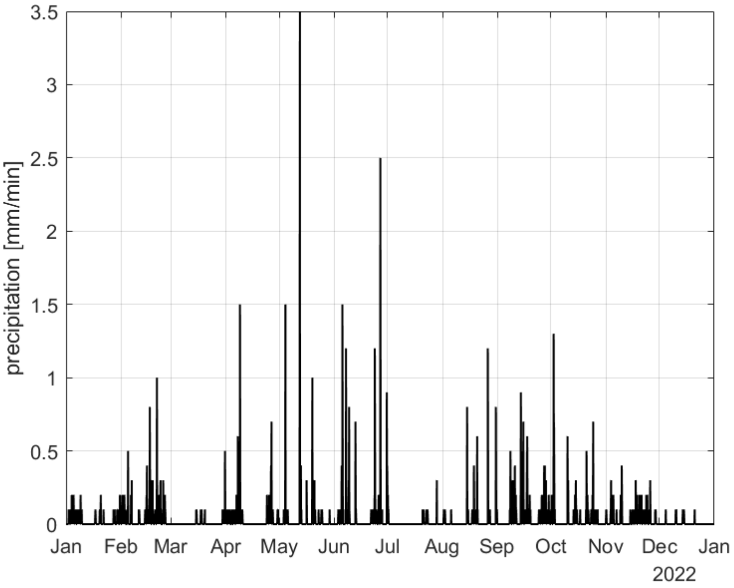

Figure 3.2.: The minute data of precipitation

A better picture of the volume of fallen rain can be seen when the counts are summed over a period of time, for example, an hour, which gives hourly precipitation (Figure 3.3). This summing shows that significant rain events or periods occur quite frequently over the whole year, although there are also some dry periods, like in July and, surprisingly, in March and December.

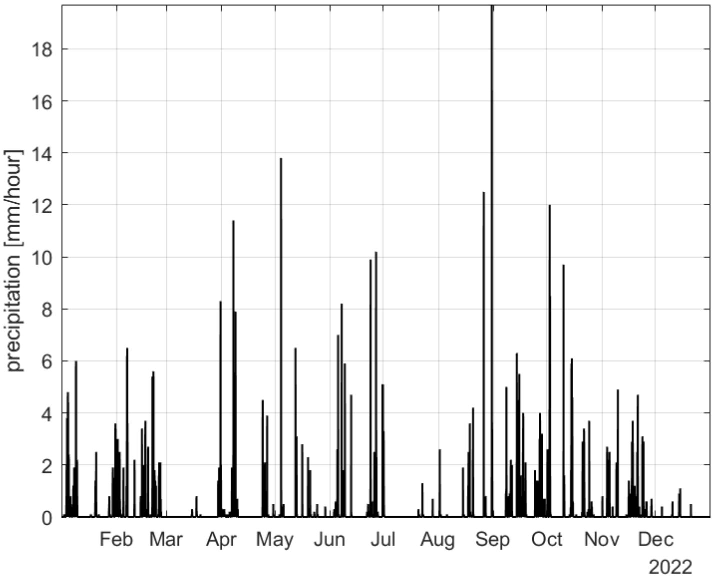

Figure 3.3.: Hourly precipitation in 2022 over BFO

The most rain recorded in one hour was 19.7 mm on the 31st of August, between 0 a.m. and 1 a.m. This corresponds to the lowest warning of heavy rain by the Deutsche Wetterdienst (15-25 L/m^2), so none of the 60-minute sums are implausible (Deutscher Wetterdienst, 2023).

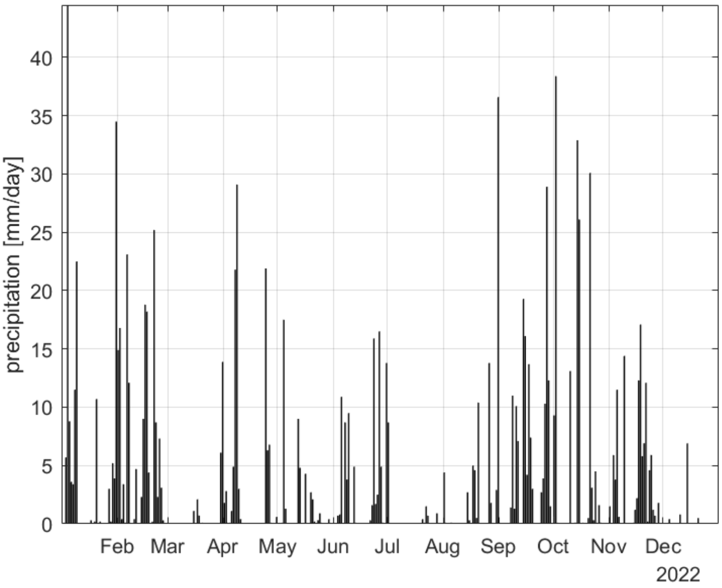

Figure 3.4.: Daily precipitation in 2022 over BFO

Summing over a whole day (Figure 3.4) shows that the rain volumes in winter and autumn can be even higher than in summer, despite the lower rain intensity. The day with the biggest amount of precipitation is January 4th with 44.5 mm. At that early stage of the year, there were no soil moisture sensors in operation (see Section 3.2). The highest precipitation volume when soil moisture data is available is the 2nd of October with 38.4 mm. This graphic will be the basis for choosing the rain events that will be analysed later in this work (see 3.3).

3.2. Soil moisture

Over the year 2022, a total of six soil moisture stations were installed at the forest surface. Good coverage was reached from May onwards. Their positions can be seen in Figure 3.5. The gravimeter lies between the stations Solidus #1 and #2 in the west and Solidus #5 in the east; there is no sensor directly above it. In late 2023, the additional station Solidus #8 was installed near the position of the SG. Its data will be beneficial for future analyses.

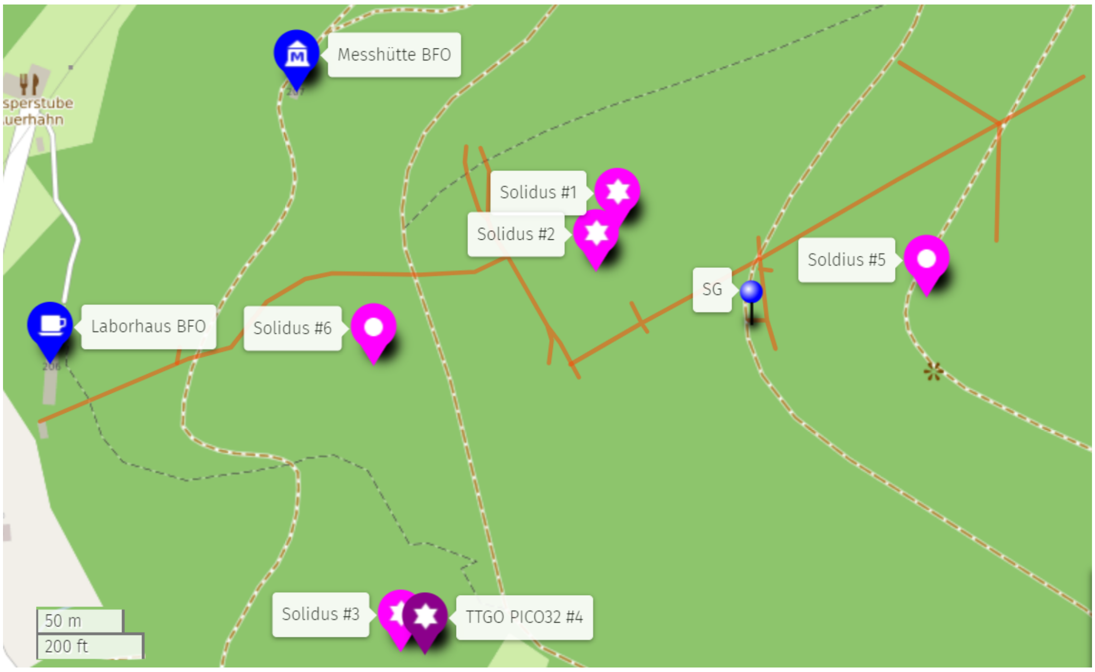

Figure 3.5.: Map of the different soil moisture stations relative to the main tunnel of the mine and the SG; Map by OpenStreetMap (SMBFO, 2023)

The Solidus stations consist of two probes, each at depths of 30 and 50 cm. They are connected by cable to a transmitter that sends the measurements back to the laboratory hut. The undulating terrain and obstructions from trees might explain some of the data outages that occur frequently (see Figure 3.6). The TTGO PICO stations have a different data acquisition system that allows only one sensor. That sensor is located 50 cm deep. The soil moisture measurements (Figure 3.6) cannot be used directly because, apart from data gaps, there are a lot of epochs with values at zero percent and others with impossible high values. Therefore, a detailed outlier filtering had to be carried out first.

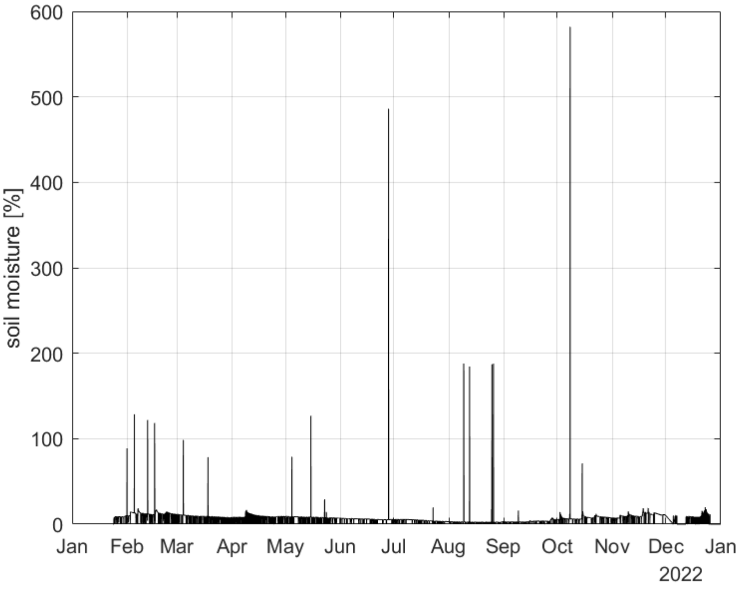

Figure 3.6.: Soil moisture readings of Solidus #1, probe 0 before any outlier filtering

The outlier filtering is done with an own function that flags outliers by finding values that are unreasonably high or low or that have too large a derivative since the last non-outlier. The derivative is considered too large if it exceeds 0.3 %·number of epochs since last non-outlier. The interval of acceptable values was manually chosen for each sensor. For example, for the sensors of Solidus #1 only values between 2 % and 20 % will be used. After this flagging, the outlier epochs are interpolated linearly unless the gap without good data is more than 24 hours. In this case, a data gap is left. Figure 3.7 shows the soil moisture timeseries for the two sensors of Solidus #1 after filtering.

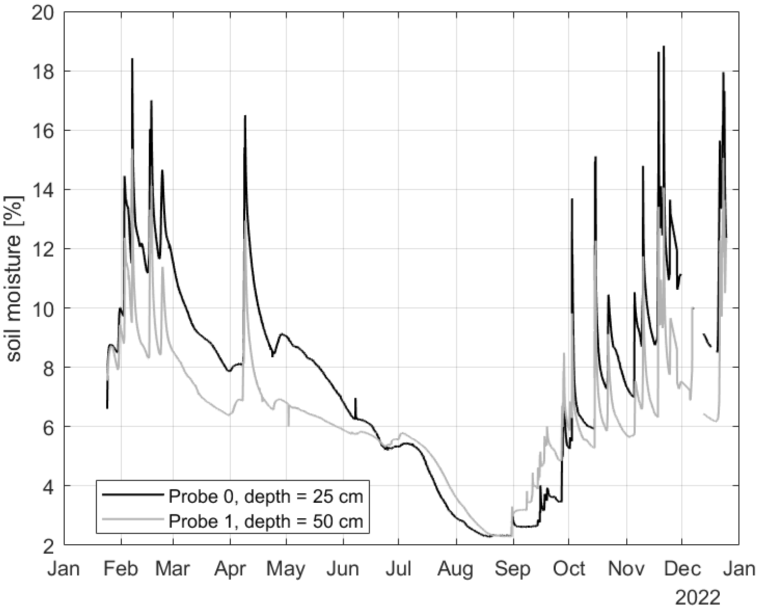

Figure 3.7.: Soil moisture readings of Solidus #1, both probes after outlier filtering

The same outlier filtering function, with different values that are considered reasonable, can be used to filter the temperature timeseries that suffer from similar problems, like the moisture data. Figure 3.8 shows the filtered temperature timeseries of the sensors of Solidus #1. In this case, all points outside of a 0.8 °C - 21 °C interval are discarded, while the limit for the maximum slope remains unchanged. The resulting series (e.g., Figure 3.8) can give a strong clue as to which probe of the two is higher by looking at the daily temperature interval.

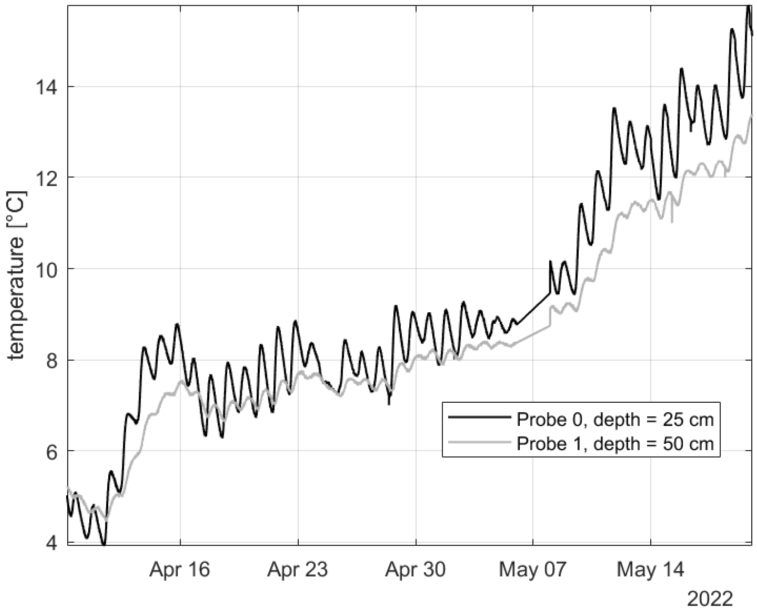

Figure 3.8.: Section of temperature timeseries of both probes of Solidus #1

Probe 0 shows a clearly larger daily variation and therefore must be closer to the surface. I assigned the depths in Figures 3.7 and 3.8 by these considerations and the information in SMBFO (2023). The same comparison can be done for the other stations that have two probes (see Table 3.2).

Station	probes	time window	number of epochs	sampling rate
Solidus #1	2, probe 0 higher	24.01.–31.12.	56316 (~54200 valid)	9.0 min
Solidus #2	2, probe 1 higher	30.04.–31.12.	31118 (~29000 valid)	12.2 min
Solidus #3	2, probe 1 higher	30.04.–31.12.	32181 (~30300 valid)	11.7 min
Solidus #5	2 (mostly only 1) not above each other	26.07.–31.12.	14862 (~14000 valid)	16.3 min
Solidus #6	2 (one not working)	28.07.–31.12.	603 (~200 valid)	1211 min
TTGOP.4	1	14.03.–31.12.	50055 (~49800 valid)	8.5 min

Table 3.2.: Overview of the soil moisture stations, data from 2022

A few of the stations are not used for further analyses. Solidus #6 has very few usable epochs. The epochs are too infrequent for any meaningful hydrological insights. Solidus #5 only has one probe for the biggest part of the year. Also, the data is more infrequent than for the other stations, and there is a big data gap in November. The seven timeseries of the probes of Solidus 1-3 and TTGOPICO4 will therefore be part of the hydrological

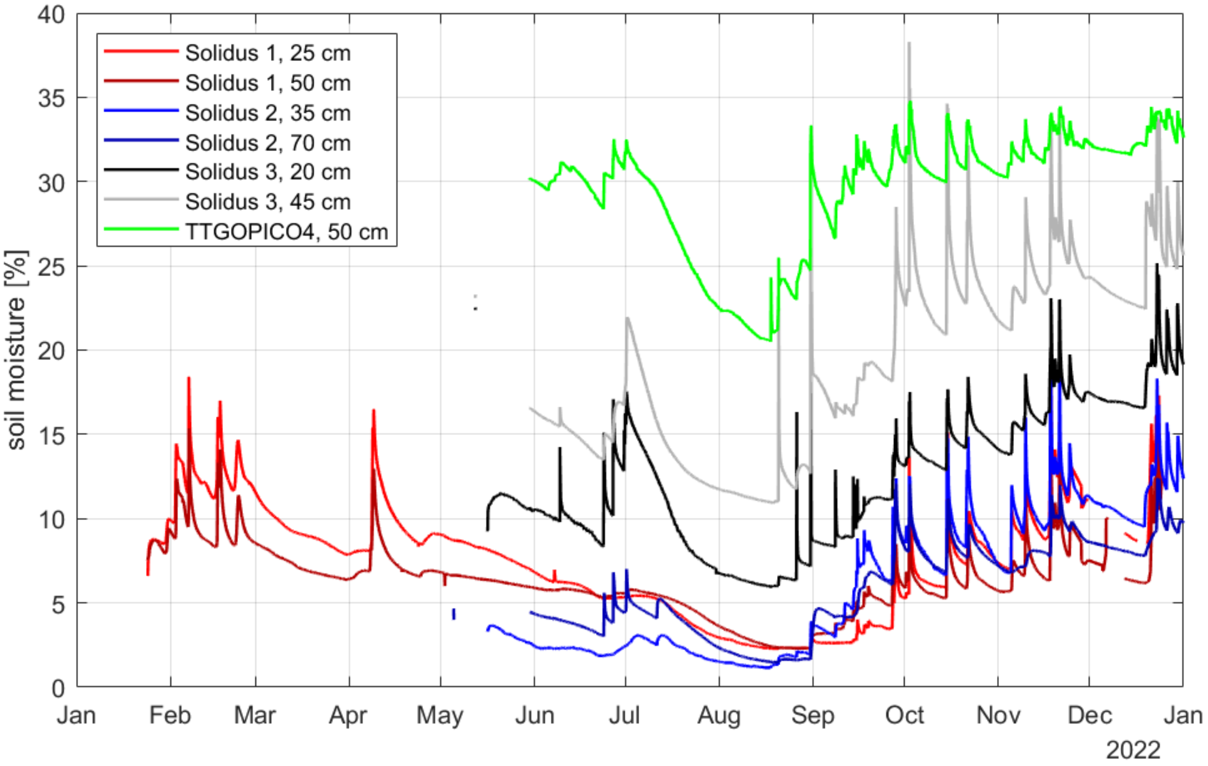

Figure 3.9.: Soil moisture data from all used stations after outlier filtering

analysis.
Figure 3.9 provides an overview of all the soil moisture timeseries that are used. As expected, the curves show similar trends, with simultaneous peaks after a rain and then a slow drop. The seasonal trend of drier soil in summer is also visible in all timeseries. All the probes have quite different levels of soil moisture; at Solidus 3, there is even a significant gap between the two probes, even though they are directly above each other. The magnitude of the peaks also varies. While Solidus 3, 45 cm shows strong variations, they are much smaller for TTGOPICO4.
Figure 3.10 gives a different overview of the measurements. It plots the measurements of the higher probe of Solidus 1 against the recordings of the other sensors that are captured simultaneously. For this, the seven timeseries were interpolated on the same time grid with a 60 s sampling rate. You can see that TTGOPICO 4 has the highest values but the lowest amplitude in the recordings. The other probes are all of the SMT100 type and seem to have similar amplitudes in their data. I use a probe of 25 cm as a reference on the x-axis, and it can be seen that the other two higher probes of Solidus 2 and 3 both show a stronger correlation with the reference timeseries. This means these point clouds are closer to a straight line and do not have a big spread in the y-direction. The lower probe of Solidus 1 is also strongly correlated, which is what you expect as the sensors are only 25 cm apart. The two lower probes of Solidus 2 and 3 are not so close to a straight line, so the correlation is lower. This is probably because the lower probes have a delayed reaction to rain. The lower probe of Solidus 3 has the lowest correlation with the reference timeseries. That makes sense, as it is both deeper in the soil and horizontally far away from the sensor.

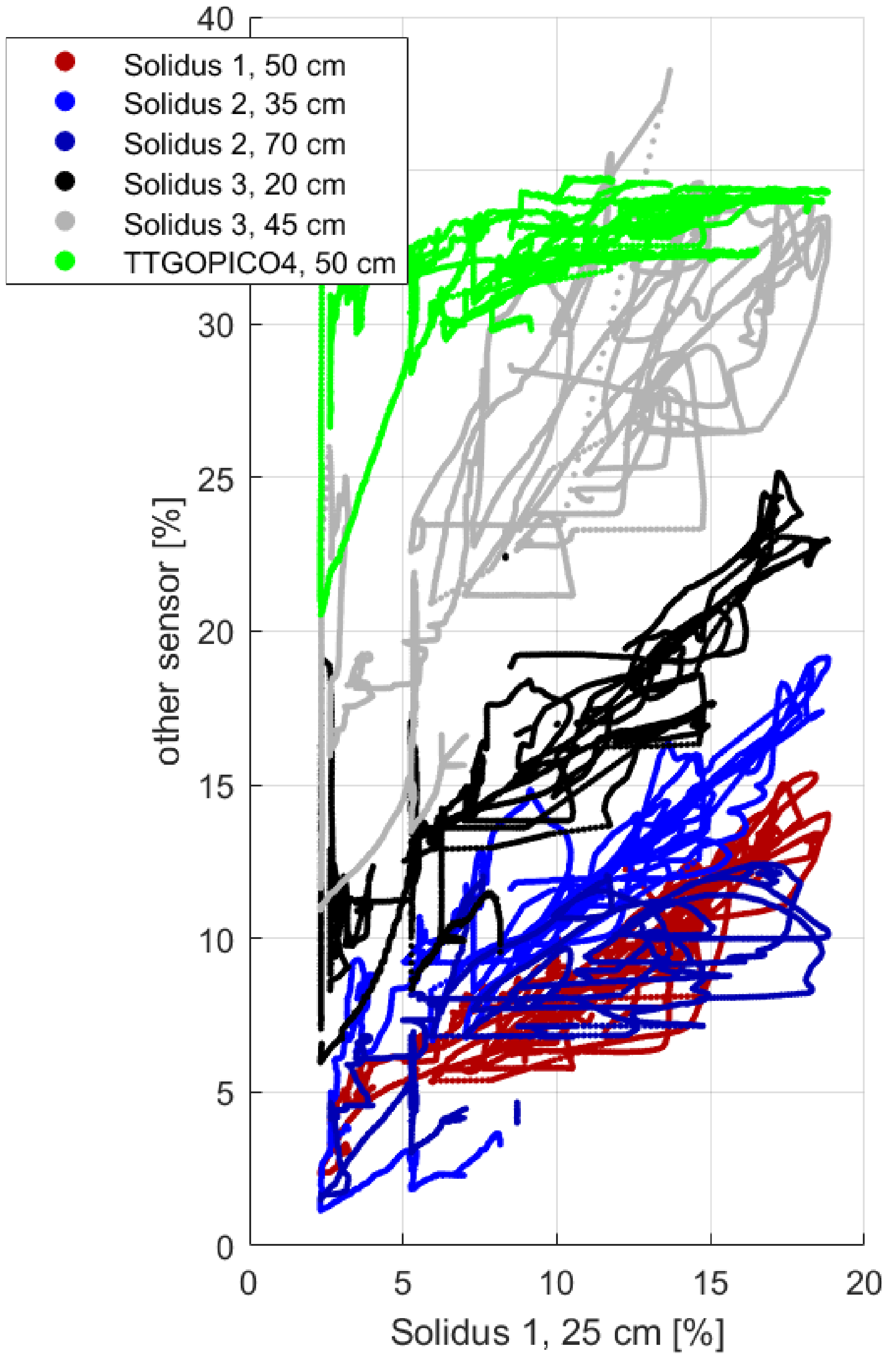

Figure 3.10.: Soil moisture data from the higher probe of Solidus 1 plotted against the measurements of the other six probes

3.3. Gravity

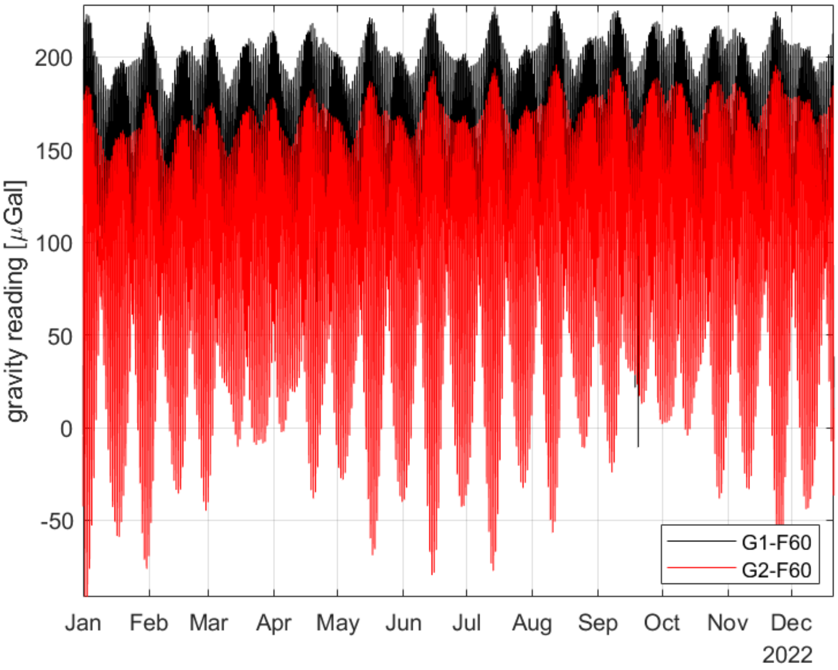

Figure 3.11.: Gravity readings of SG056 before any corrections

The gravity data are two timeseries for the two sensors at a sampling rate of 60 seconds. Because the instrument is a superconducting gravimeter, the readings are relative values and mean nothing by themselves. Rather, the course of the measurements of one sensor can be analysed.
Figure 3.11 shows the gravity series after multiplication with the given calibration factors for both sensors.
The gap between both sensors stays very constant. Only one big outlier in G1-F60 is visible in September. The pattern is a typical tide signal. Before any hydrological analysis can be done, the tide subtraction is the first of a number of corrections that have to be carried out. These corrections are detailed in Chapter 4.

3.4. Selected events for analysis

Between May 30th and November 29th, there are no longer data gaps for any of the seven probes that were chosen for the analyses (see Figure 3.9). In this portion of the year, I searched the precipitation timeseries for days with significant amounts of rain. I identified eleven rain days or rain events. During these events, you can be more confident that a large part of the gravity signal after the described corrections of Chapter 4 will be caused by soil water. Therefore, these eleven events, each with a window of three days, will be used for the later adjustments. Table 3.3 summarizes the eleven events.

Table 3.3.: Rain events in 2022 with soil moisture data from Solidus #1, probe 1. The soil moisture value before the event is an average value in the last hour before the rain starts.

date	sm before [%]	sm max [%]	prec [mm/d]
June 23rd	5.33	5.39	15.9
August 1st	3.54	3.55	4.4
August 26th	2.31	2.31	13.8
August 31st	2.3	3.31	36.6
September 14th	3.91	4.62	19.3
September 27th	4.87	6.7	28.9
October 2nd	6.32	9.82	38.4
October 14th	5.29	12.27	32.9
October 21st	5.89	9.12	30.1
November 9th	6.6	11.74	14.4
November 18th	6.31	13.41	17.1

October 2nd has the highest amount of precipitation. This event will be mostly used as a visual demonstration of the different correction models (see top left of Figure 3.13). The adjustments and the compared metrics, like the mean correlation coefficient, use the data from all eleven events, though.

The following two graphics show the gravity residuals and the precipitation and soil moisture measurements over these events. For two of the events, namely the August 1st and 26th rains, there is no increase in the soil moisture measurement of the depicted probe of Solidus 1. This is similar to the recordings of the other sensors. The two events have a relatively low amount of water, and as it is falling in a short period of time on the dry soil of late summer, the soil moisture shows no reaction. This observation means that you cannot expect too much from a gravity model for these events, as the model can only use information from the precipitation gauge. This will also negatively influence the overall result of the eleven events.

The gravity residual is my estimation of $g_{(\mathrm{hydro})}$. When it rains, you expect the value to decrease, and after the rain has stopped, it should slowly rebound. This is mostly what can be seen in the gravity series, although there are sections like the fast gravity increase on June 23rd that do not seem to be related to rain. In the next chapter, it will be discussed how to calculate these residuals from the gravity measurements. The presented timeseries of gravity residuals in the two graphics are the final results of this chapter.

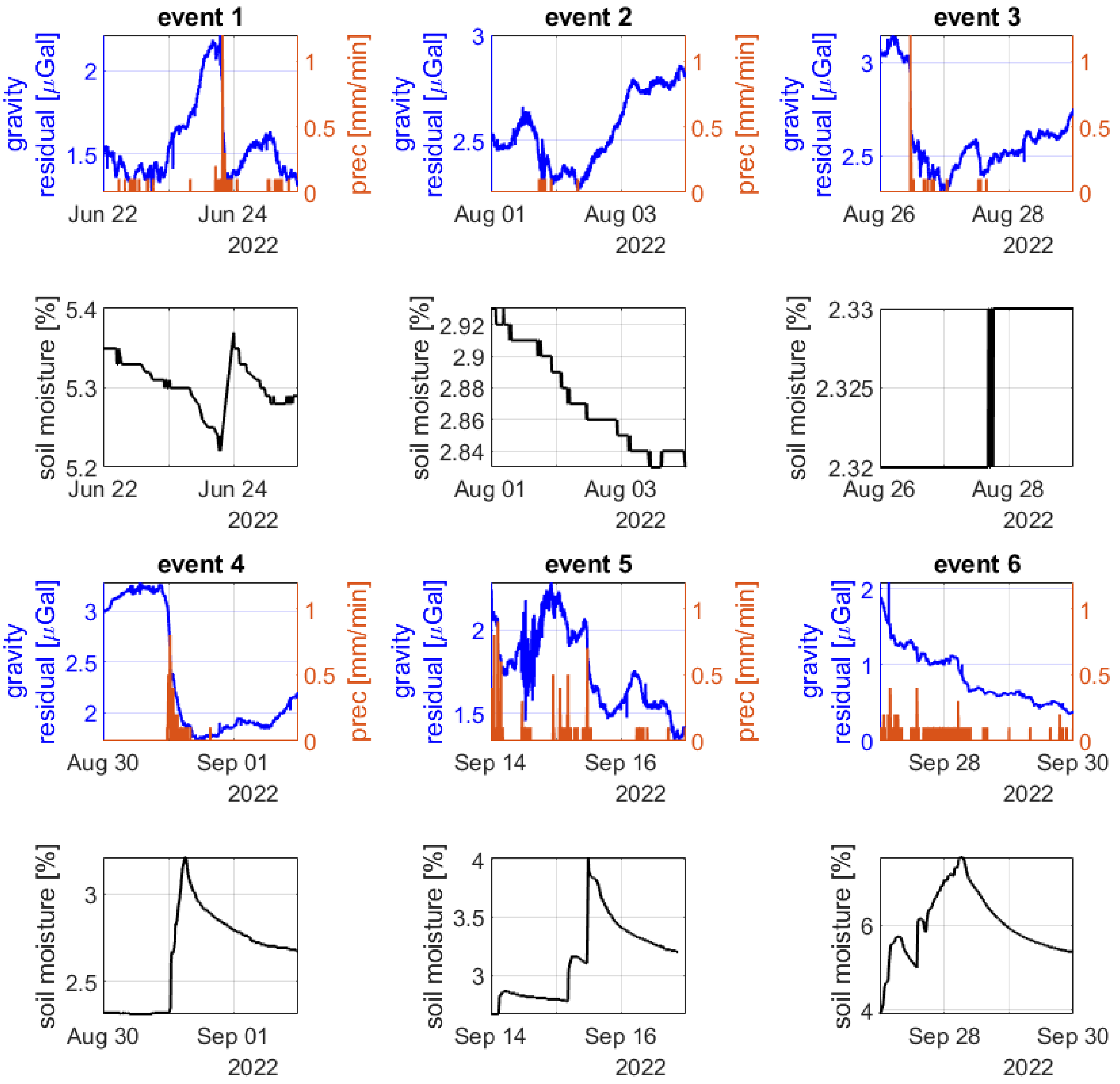

Figure 3.12.: Gravity residuals, soil moisture readings, and precipitation measurements over the first six events. The gravity residuals were calculated with the corrections of Chapter 4. For clarity, only one of the gravity sensors (the lower one) and one soil moisture probe (Solidus 1, 25 cm) are depicted.

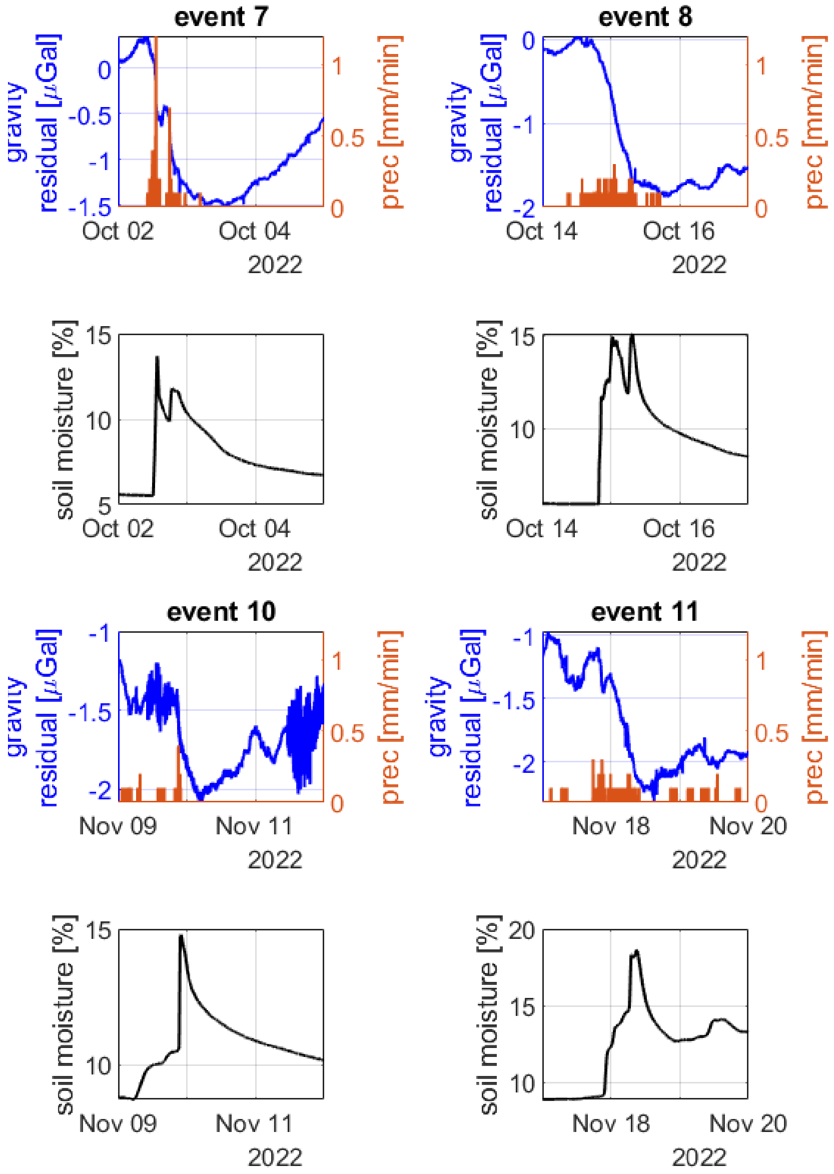

Figure 3.13.: Gravity residuals, soil moisture readings, and precipitation measurements over the final five events. The gravity residuals were calculated with the corrections of Chapter 4. Only one of the gravity sensors (the lower one) and one soil moisture probe (Solidus 1, 25 cm) are depicted.

4. Calculating the gravity residuals

This chapter describes how I get from the raw SG measurements (Figure 3.11) to a signal that can be further used for hydrological analysis. There are much larger mass transports in the environment of the gravimeter than water on top of or in the soil, which can be sensed by the gravimeter. These signals first need to be accurately subtracted. Section 4.1 deals with getting rid of large outliers. In Section 4.2 the tidal impact is corrected, which is followed by the atmospheric influence (4.3), a subtraction of residual oscillations (4.4), the removal of drift (4.5) and a final outlier filtering in both timeseries (4.6).

4.1. Outliers

The first step in processing the gravity data is to identify outliers and eliminate them. In this work, an outlier is a gravity measurement that has a sudden jump from previous measurements. The magnitude of this jump can range from about 1 µGal to over 100 µGal. Such a fast change will have no hydrological cause and therefore needs to be removed before a hydrological analysis. The most likely cause of big jumps in the data are earthquakes. The vibrations in the ground transfer to the gravimeter, and in the worst-case scenario, the niobium sphere touches one of the plates of the capacitance bridge, rendering a few epochs completely useless (personal communication from Walter Zürn). Even if that is not the case, it is easy to imagine that the feedback coil tries to hold the sphere in the center while the ground vibrates, and so the vibrations will be interpreted as gravity changes.

An outlier elimination can be done after the correction of tides (see for that correction Section 4.2) as that step will reveal outliers more clearly, but with a different approach that I am using, outliers can be detected immediately.

The method takes advantage of the fact that there are two sensors, which, apart from an offset and a small drift, should have almost identical timeseries. So if the difference between the sensors changes suddenly, there is an outlier in at least one of the sensors. The written function flags certain epochs as outliers when the difference between the two sensors is more than three standard deviations different from the median in a sliding two-hour window around the epoch. Afterwards, for all these epochs, both timeseries are linearly interpolated between the non-outliers, as the method does not determine which sensor has the outlier.

Figure 4.1 shows the results of the removal. As intended, the sudden big or small values in the difference disappear. What remains is almost a straight line that represents the different drift rates of both sensors. There are two notable departures from this in May and August. Both are instrumental issues that were registered at BFO. In May, a lightning strike caused the failure of different components. This also affected the connection to the outside pressure sensor (see Section 4.3). In August, there was an issue with the power

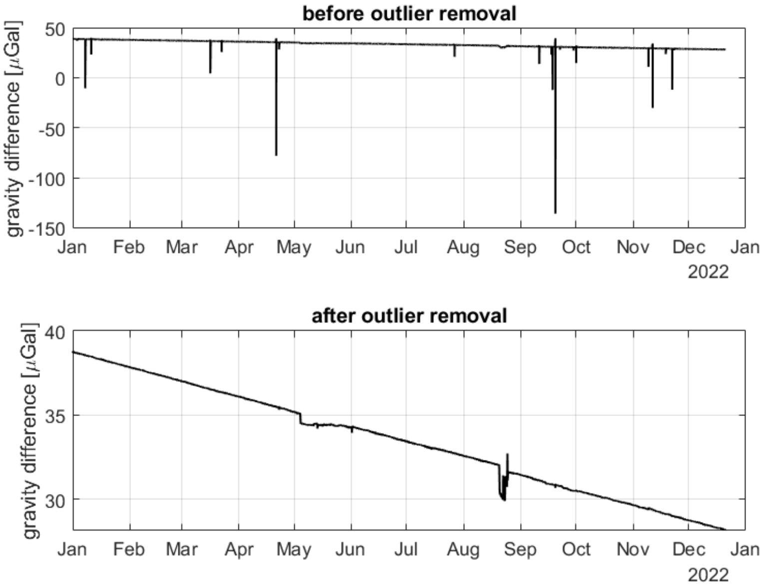

Figure 4.1.: Difference between timeseries of both sensors before and after outlier removal

supply of the cryogenic cooling system (personal communication from Thomas Forbriger). So the readings in these two sections have to be analysed with caution.

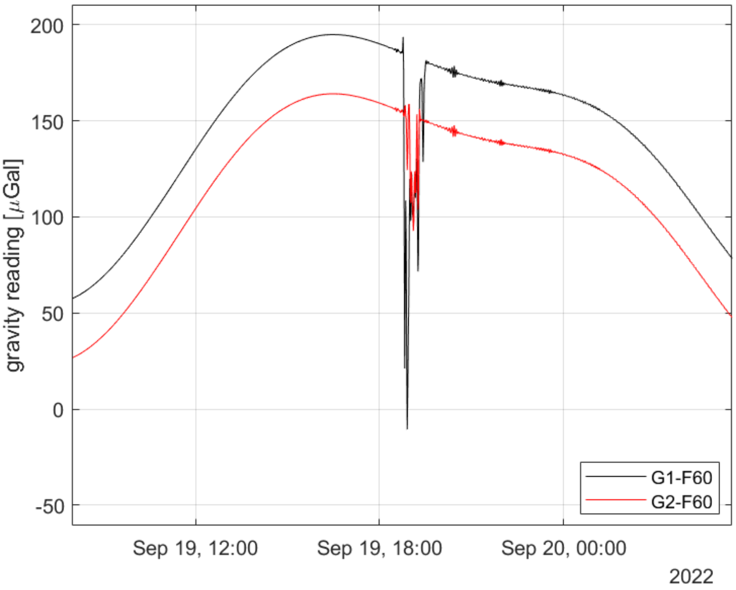

Figure 4.2.: Gravity measurements of both sensors on the afternoon of September 19th when a magnitude 7.7 M_W struck Mexico

The biggest jump in Figure 4.1 before the outlier removal is one at the end of September with an amplitude of nearly −200 µGal. This is the same outlier that was visible in the gravity timeseries over the year (see Figure 3.11) in G1-F60. Contrary to what it looked like so far, this is not one single outlier but an episode of about one hour. The episode happened on September 19th. Figure 4.2 shows the measurements of the two sensors. The first epoch that is flagged as an outlier is at 18:42, although the first signs of unrest start at 18:17. On that afternoon at 18:05 UTC the 2022 Michoacán earthquake with a magnitude of 7.7 M_W happened. The epicenter was in Mexico (GEOFON GFZ Potsdam, 2023*a*). This explains the large jumps in both sensors that evening.

4.2. Tides

The first component that has to be subtracted is the tidal signal. The masses of the sun, the moon, and other celestial bodies exert a strong gravitational pull on the spheres. Additionally, there is a solid Earth component that is caused by the change in the position of the gravimeter in the Earth-fixed system because the ground reacts to the gravitational pulls and so moves up or down (Van Camp et al., 2017). While the direct gravitational influence can be computed very precisely from the positions of the bodies at a given time, the solid Earth part is not so straightforward; the response of the Earth is a complex process and cannot be completely predicted. Rather, from previous gravity timeseries gravimetric factors and phase lags for different wave groups are estimated. They represent the movement of the Earth in response to the tidal forcing and secondary effects like ocean load (Hinderer et al., 2007). These estimations may not be perfect for a given day in 2022, though.

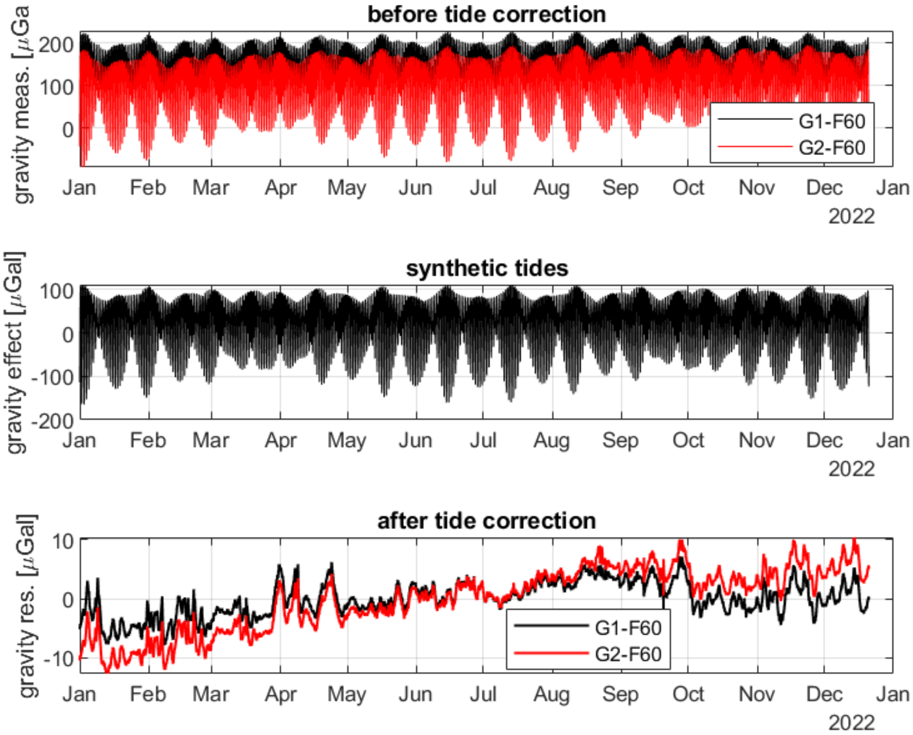

Figure 4.3.: Gravity timeseries of both sensors before and after the correction of tides; the synthetic tides that were used for correction are in the middle

As the tides are not the central interest of this work, I use a timeseries with parameters from a tidal analysis that was done by Eva Schroth. She used data from the SG056 and the program *ETERNA 3.4 analyze* for this analysis (Wenzel, 1998). This synthetic tides series also includes components for the influence of Pole tides and Length of Day tides (see Section 1). The gravity timeseries before and after tide correction can be seen in Figure 4.3.

The resulting timeseries has a smaller magnitude by a factor of more than 10 and shows no obvious oscillations like the raw signal.

4.3. Atmosphere

The second important correction is the atmospheric pressure. A higher pressure means there is an air column with more mass above the gravimeter. This mass has a gravitational effect on the gravimeter, and this effect needs to be subtracted from the signal. There is also a secondary effect: the higher pressure pushes down the ground and therefore changes the position of the gravimeter and so the measured gravity. While added gravitational pull decreases the gravity value, a lower position of the gravimeter causes an increase. The former effect will be stronger though, so higher pressure reduces the experienced gravity (Van Camp et al., 2017).
The model to correct the atmospheric influence will be a simple factor by which the pressure, which was measured by the pressure sensor on top of the laboratory hut, is multiplied. This product is then subtracted from the gravity timeseries. This model only roughly corrects the influence of the pressure as it ignores the frequency of the change. This does play an important role though, as the Earth does not respond as strongly to fast pressure oscillations, so the position-induced gravity change is reduced (see (Hinderer et al., 2007)). There are also inertial forces in the sphere to mind. Nonetheless, a factor of −0.34 µGal/hPa has previously shown to be a decent value for correcting most of the pressure influence and will be used in the beginning (Zürn, 2014).
The pressure data from the outside sensor has some major data gaps in May (Figure 4.4) which were caused by a storm that cut the connection between the sensor and the SG, which usually records both inside (tunnel) and outside pressure. These measurements are therefore lost and cannot be used for corrections.

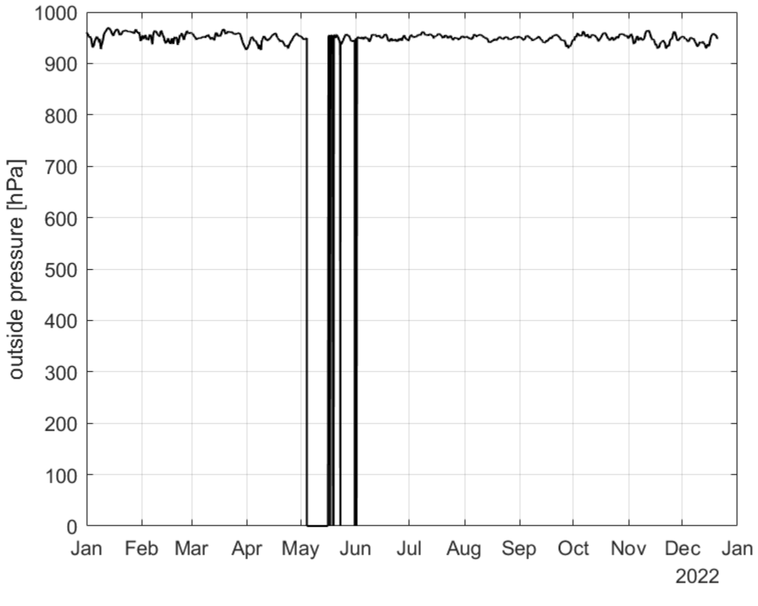

Figure 4.4.: The atmospheric pressure, recorded outside, on top of the laboratory hut; the data outages in May are visible

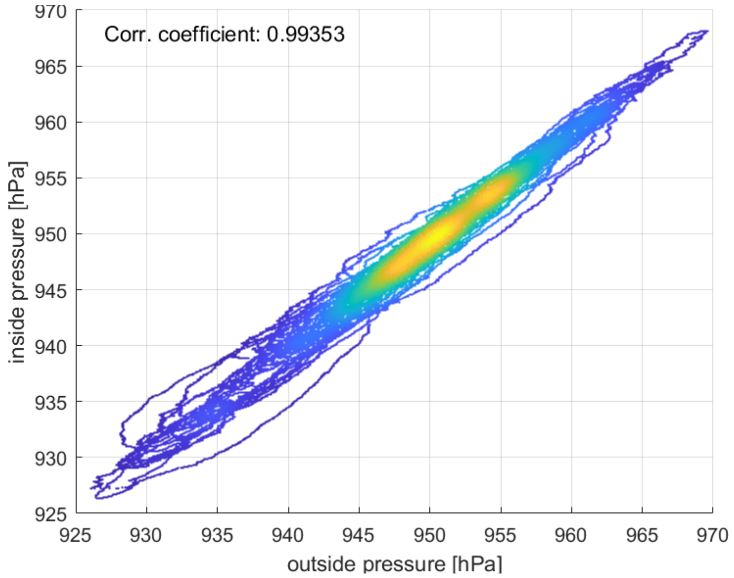

Figure 4.5.: Comparison between inside and outside pressure for the epochs where the outside pressure is valid; color is the point density; the bright yellow areas have the highest density

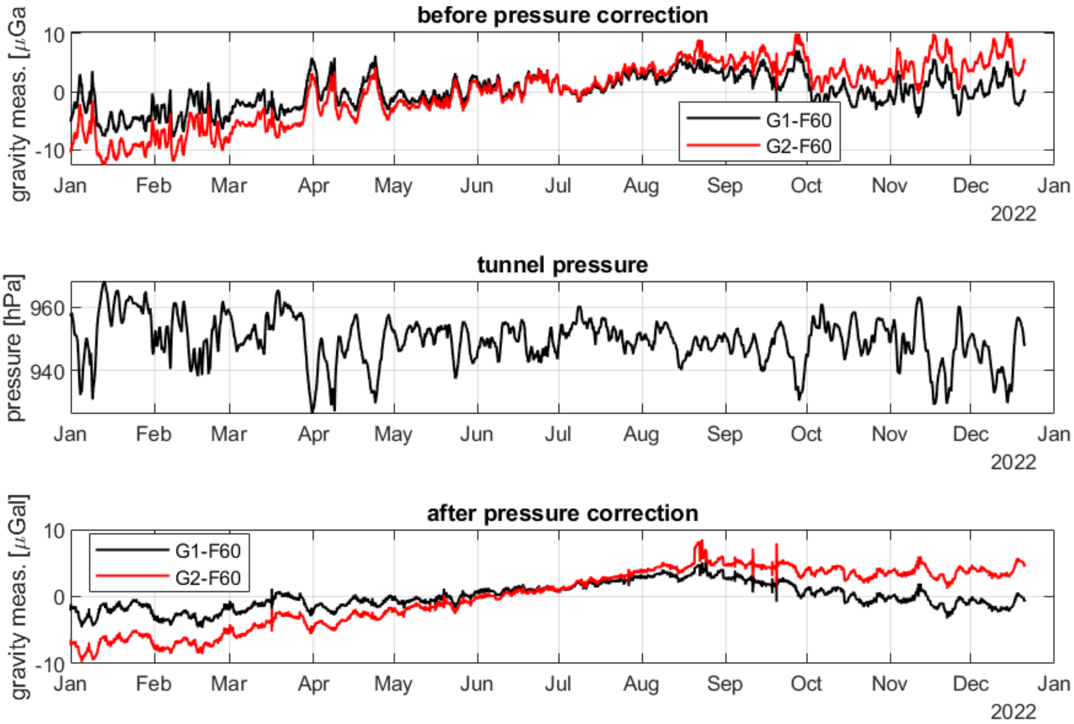

Figure 4.6.: Gravity timeseries of both sensors before and after the correction of atmospheric pressure with the factor of $-0.34\,\mu\text{Gal/hPa}$; in the middle is the timeseries of the tunnel pressure

The main analysis will be done with rain events between June and December (see Chapter 5) as this is when the soil moisture data were available. These gaps are therefore not a problem when building the hydrological model. Still, I searched for an alternative way to fill them. I use the inside pressure recordings as they have a correlation of 0.993 with the outside pressure for the epochs with usable outside measurements (Figure 4.5). Whenever outside pressure is available, it is used for the correction.

Figure 4.6 compares the gravity series before this correction with the pressure measurements over the year. The rough anti-proportional connection between them, at least on a short (a few days to weeks) period is visible and the resulting gravity timeseries has greatly reduced variations.

But is −0.34 µGal/hPa a proper factor for our purposes? The factor f can be determined by a linear adjustment of the following equation:

$$f\delta p = \delta g. \tag{4.1}$$

Here, δp are the reduced pressure measurements, which means the mean is subtracted, and δg are the reduced gravity measurements after the correction of tides. Both are vectors with a length equal to the number of epochs. The estimation is then done in the usual manner by the formula

$$\hat{f} = (\delta p^T \delta p)^{-1} \delta p^T \delta g. \tag{4.2}$$

This can either be done with the whole timeseries or with shorter windows that are moved through the year and then averaging of the found results for $\hat{f}$. The result of the latter method will represent the short-term response of gravity, depending on the length of the window. It is not influenced by any low-frequency changes in pressure. The downside is that fewer data points have the danger of producing unrealistic results if only a few points are unexpectedly high or low, for example, in a rain event. This can be mitigated, though, by the averaging at the end.

An example of how the estimation for f can be thrown off is given in Figure 4.7. There are a number of points with a δp of just over 0 and a δg quite significantly below 0. The adjustment searches for a straight line through the origin that has a minimal distance from all the points in the window. These low data points will skew the slope of the straight line to a very big negative value.

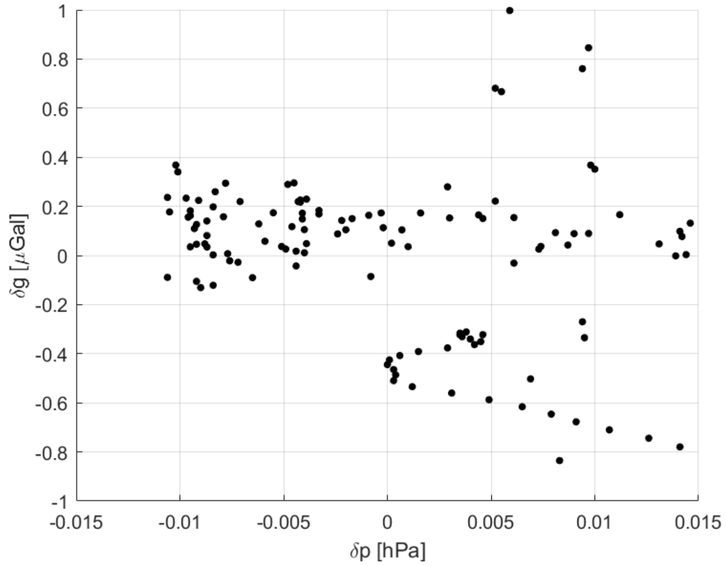

Figure 4.7.: One section of pressure and gravity data; these will produce an unrealistic value for f

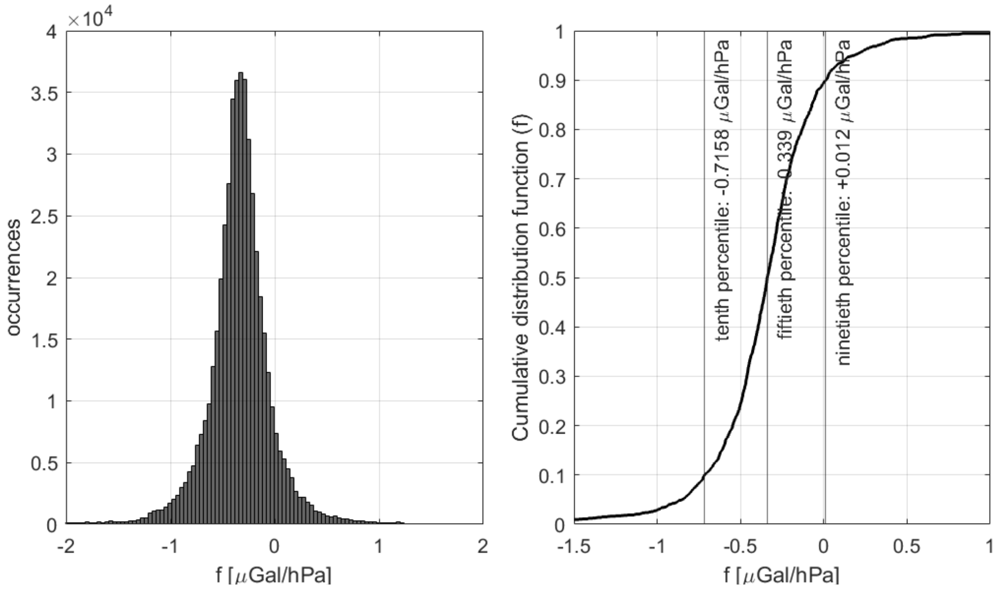

Figure 4.8.: Probability density function (left) and cumulative density function (right) of the adjusted values $\hat{f}$. A window of six hours length was moved over the year. It was repeatedly pushed one minute further. The results of the data from G2-F60 are depicted.

A window length of 6 hours produces reasonable results in terms of outliers and still captures the short-period response of the gravity signal. The described adjustment is carried out with $6 \cdot 60$ epochs, and then the window is pushed one minute further. To reduce the problem of outliers, the points are given a weight. This weight is the absolute value of δp. It was seen in the last paragraph that points with a small δp can be problematic. That is why they get a smaller weight. This weighting mitigates the described problem but does not solve it. In the end, the average as well as the median of $\hat{f}$ are computed. The latter is less influenced by outliers.

Figure 4.8 shows the distribution of results for f in this series of adjustments. The large number of unrealistic values can be seen. The majority of values are around $-0.3\,\mu\text{Gal/hPa}$, but around 18 % of values end up over $0\,\mu\text{Gal/hPa}$ or below $-0.7\,\mu\text{Gal/hPa}$. For sensor 1 of the SG, the mean value for f is $-0.3363\,\mu\text{Gal/hPa}$ with the median at $-0.3310\,\mu\text{Gal/hPa}$; when carrying out the same adjustments with the gravity measurements of sensor 2, the results are very similar at $-0.3364\,\mu\text{Gal/hPa}$ and $-0.3319\,\mu\text{Gal/hPa}$ respectively. These results suggest that the provided value of $-0.34\,\mu\text{Gal/hPa}$ is reasonable. When choosing shorter time windows, the results tend to be even closer to $-0.34\,\mu\text{Gal/hPa}$ or below it.

There are different ways to look at it, though. Running the adjustment in (4.1) for the whole year leads to very different values. For sensor 1, the adjustment finds a value of $-0.2460\,\mu\text{Gal/hPa}$ (see Figure 4.9) and $-0.2468\,\mu\text{Gal/hPa}$ for sensor 2. Further analyses also yield arguments for these values: I am doing a one-dimensional grid search by carrying out the correction of atmospheric influence with different values of f between $0\,\mu\text{Gal/hPa}$ and $-0.35\,\mu\text{Gal/hPa}$ with steps of $0.01\,\mu\text{Gal/hPa}$ between them. The atmospheric correction should reduce the variations in the timeseries. The standard deviation of the different timeseries over the whole year is analysed, and a minimum shows at

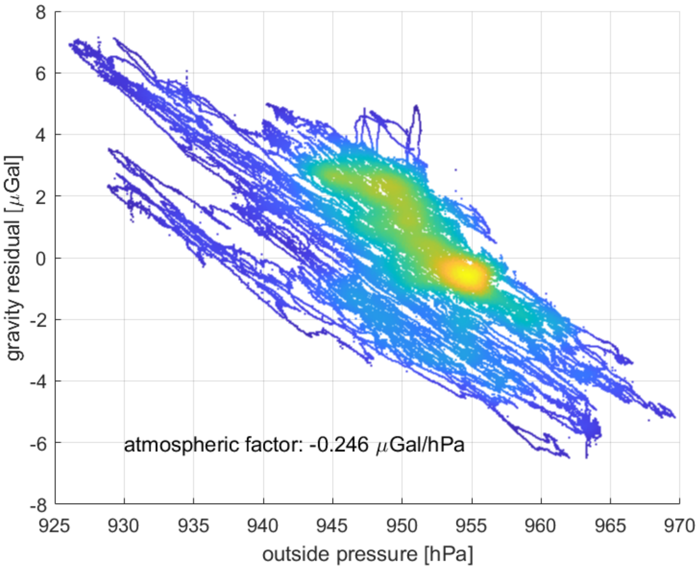

Figure 4.9.: The reduced pressure readings and gravity values of G1-F60, after the correction of tides and a removal of drift, over the year, plotted against each other. The colour is the point density, with yellow areas having the highest density.

−0.25 µGal/hPa (Figure 4.10). The standard deviation was a full 10 % better than the value with −0.34 µGal/hPa. Another metric that can be looked at is the correlation with the pressure timeseries. After the correction, the gravity timeseries should not have any remaining connection with the pressure, so the correlation should be close to 0. Computing this correlation over the year, I find the zero point between −0.24 µGal/hPa and −0.25 µGal/hPa (Figure 4.11). The timeseries with −0.34 µGal/hPa shows a correlation with the pressure of nearly 40 %.

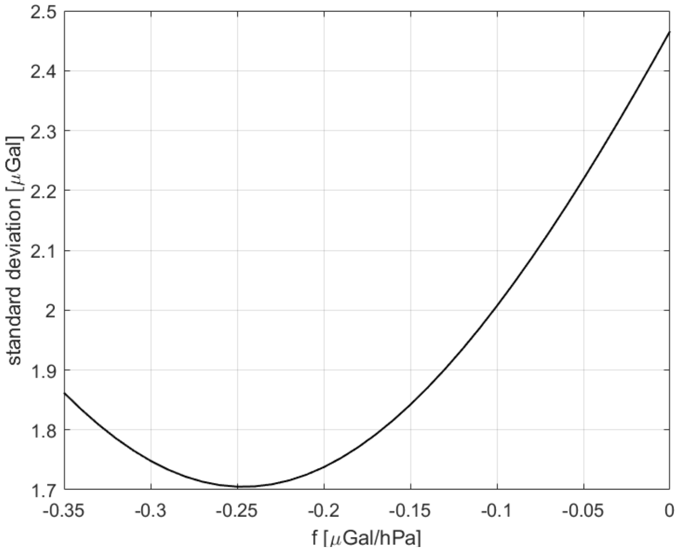

Figure 4.10.: Standard deviation over the whole year of the resulting gravity timeseries that were corrected with different values of f

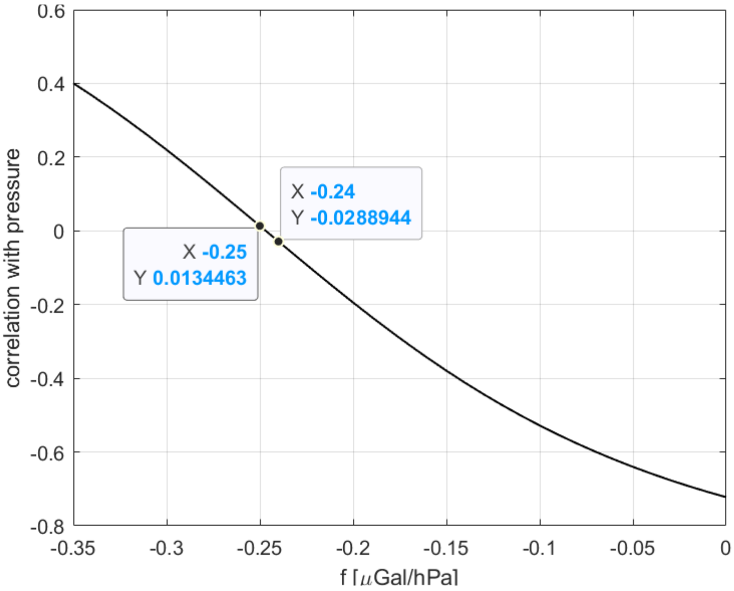

Figure 4.11.: Correlation over the whole year with outside pressure of the gravity timeseries that were corrected with different values of f

Figure 4.12 shows the disadvantages of a correction with $f = -0.34$ µGal/hPa. It shows the resulting gravity series after pressure correction with −0.34 µGal/hPa and −0.25 µGal/hPa in January alongside the measured outside pressure. The upper graph clearly shows a correlation with pressure. In fact, it is as high as 0.84. The gravity series with the modified factor has a negligibly small correlation of −0.12.

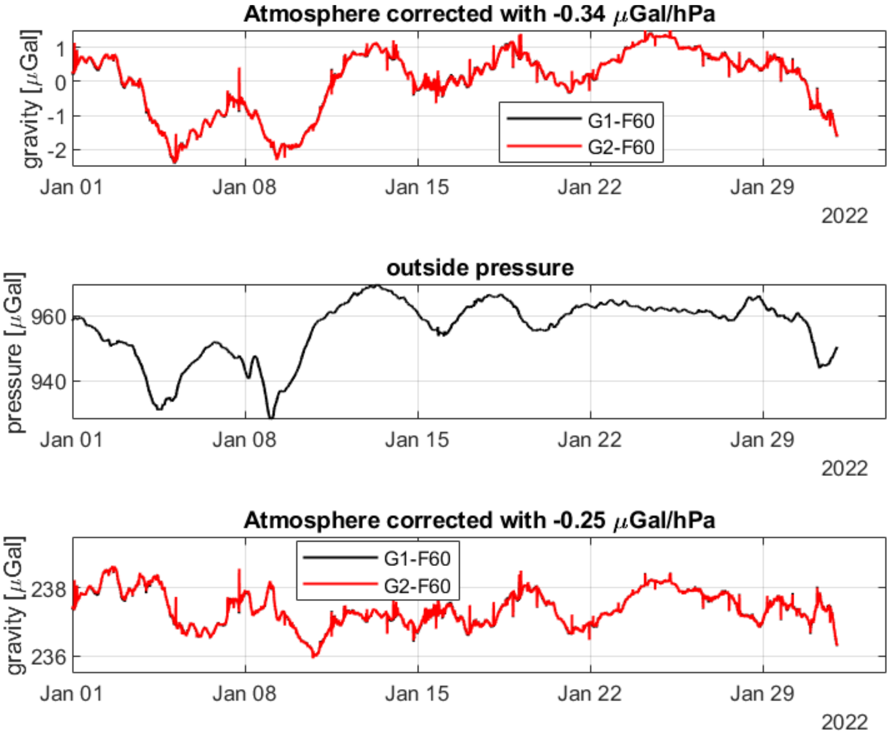

Figure 4.12.: The gravity timeseries over January after correction with two different values of f, with the pressure timeseries in between them; linear trends were subtracted from all gravity series to show the correlations more clearly

When analysing a longer timeseries of one or more months, the −0.34 µGal/hPa correction is not ideal. The short-term analysis does show the merits of that factor, though. Some fast pressure peaks, like the negative one on January 9th (Figure 4.12) translate as an inverse peak in the −0.25 µGal/hPa corrected gravity series.

One common method when analysing the gravity series of a time window of a few days is to carry out a least squares adjustment to estimate f for just the few days of the analysis. I would advise against this approach based on our data. Figure 4.13 compares the gravity residuals and the atmospheric pressure over the October 2nd event. It illustrates the problem of adjusting an atmospheric factor for such a time window: There is a natural anti-correlation between gravity and pressure. Before a rain happens, gravity is higher as the soil is still dry. Leading up to a rain, the pressure also decreases due to meteorological reasons. As the rain has passed and the weather gets friendlier, the pressure also climbs while the gravity takes a minimum at the end of the rain because of the wetter soil. Figure 4.14 shows another example of the falling pressure ahead of a rain. This common pattern has also been described by Krause et al. (2009). For the three depicted days, an adjustment would lead to a value of nearly −0.48 µGal/hPa which is unreasonable and makes the gravity series really flat. Therefore, it is better to use longer data series of at least a month to estimate f so that this anti-correlation does not create a bias in f.

Figure 4.13 also illustrates the uncertainty in this atmospheric correction. As the rain and the pressure changes generally occur together, you cannot be sure how much of the gravity signal is caused by the atmosphere and how much by soil moisture. Using a different factor in the pre-processing will also lead to vastly different hydrological models that are computed.

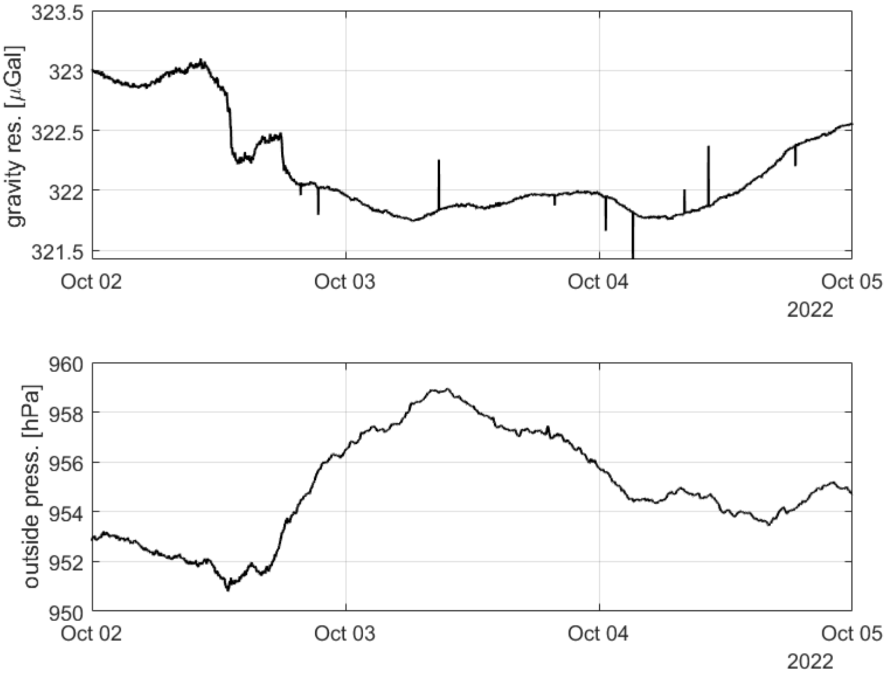

Figure 4.13.: The gravity residuals after correcting the atmospheric pressure with f = −0.34 µGal/hPa and the outside pressure over the October 2nd event

Because of the high uncertainty in the short-period adjustments of f (Figure 4.8) and because the −0.25 µGal/hPa correction produces visually quite reasonable results (see end

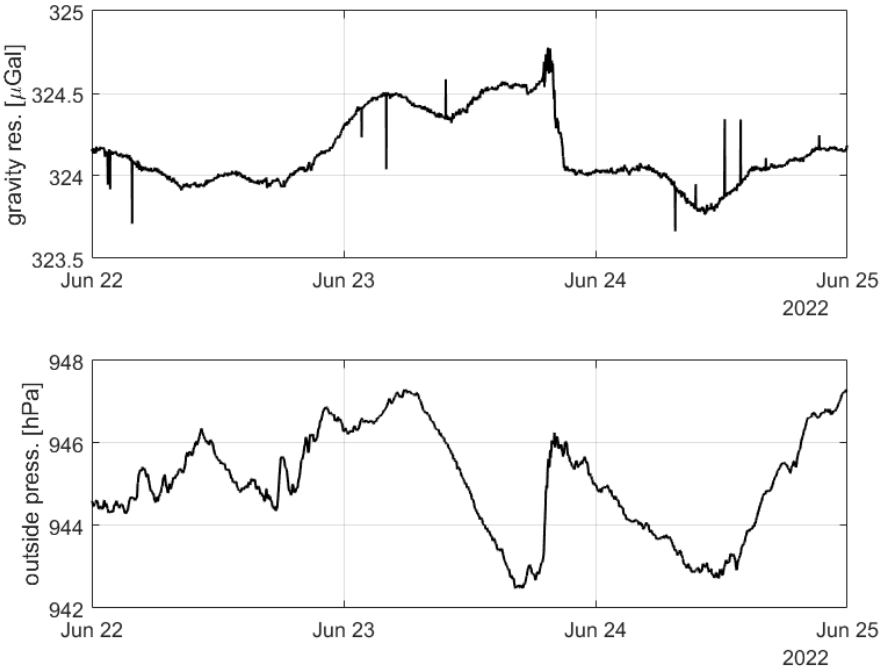

Figure 4.14.: The gravity residuals with $f = -0.34\,\mu\mathrm{Gal/hPa}$ and the outside pressure over the June 23rd event. The rain happens on the evening of the 23rd

of chapter), I decide to use the result of the full-year adjustment of $-0.246\,\mu\mathrm{Gal/hPa}$ (Figure 4.9). Some of the high-frequency pressure will not be perfectly corrected by this, though. Hinderer et al. (2007) has a detailed analysis of the higher atmospheric factor for higher frequency changes. So really, a frequency-dependent factor, as suggested by them, needs to be aimed for in the future.

4.4. Remaining tidal signals

After correcting the atmospheric influence on the gravity series with the value of $-0.246\,\mu\mathrm{Gal/hPa}$, even smaller signals become visible. Looking at the timeseries over a smaller time scalem like a month, shows clear, small oscillations with periods of about half a day. The oscillations look very similar to tides. When laying the synthetic tides over the timeseries (Figure 4.15) you can see that the timings of the maxima and minima of both timeseries are almost matched. It seems that the gravity signal sometimes has a small delay of about one or two hours, as was highlighted in Figure 4.15. This could suggest that there are some errors in the phase lags of the response of the solid Earth to tides. Because of the binding forces in the Earth and inertia, the Earth does not respond instantly to tidal forcing but has different delays depending on the frequency of the forcing. These phase lags have been estimated for the synthetic tides, but these estimations used data from previous years and not 2022. These phase lags change over time because of loading tides (Schroth, 2013). These are load changes of the ocean or the atmosphere that are caused by tidal forcing.

These imperfect gravimetric factors and phase delays are probably the cause of the observed oscillations.

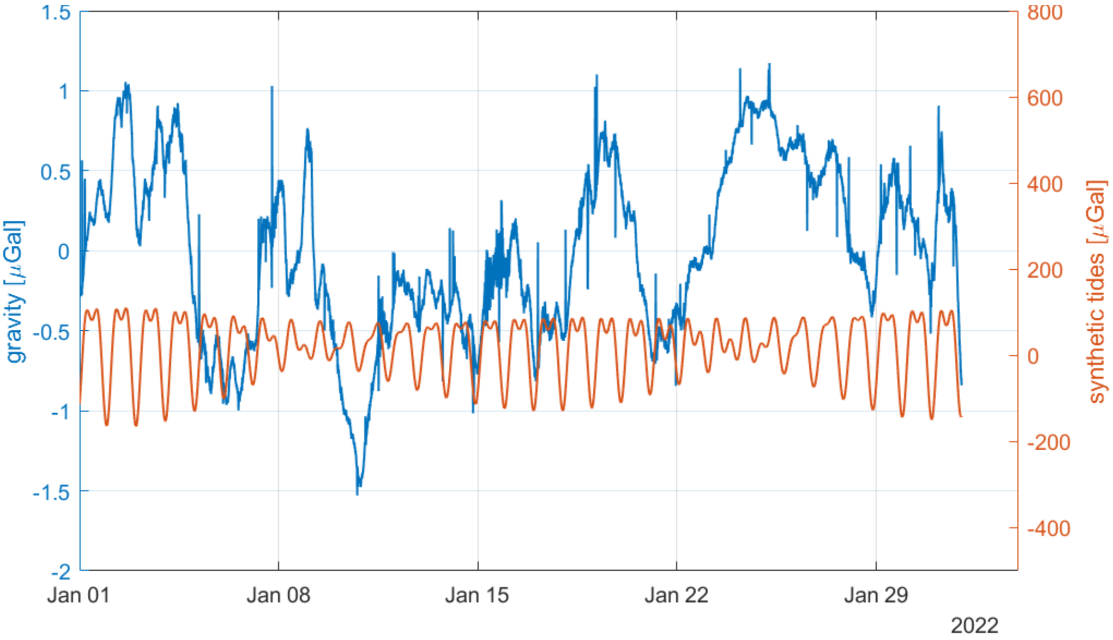

Figure 4.15.: The gravity timeseries of sensor 1 after correction of atmospheric pressure, with the synthetic tides plotted above it

As a first possible strategy to get rid of these unwanted oscillations, I take the main tidal frequencies, fit a wave with these frequencies to the timeseries, and subtract them. If only a handful of frequencies are fitted, than it is likely that only the unwanted tidal remains are eliminated and not any hydrological signal is destroyed. The four main tidal components, namely the periods of 1 solar day, 1 lunar day (24 hours and 50 minutes), half a solar day, and half a lunar day, are chosen for elimination.

The problem can be adjusted linearly, despite the fact that intuitively, an amplitude and a phase need to be adjusted for every frequency. The oscillation equation can be rewritten in this way:

$$g(t) = \sum_{i=1}^{4} \left(a_i \sin(f_i t + \phi_i)\right) + c_0 + e(t) = \sum_{i=1}^{4} \left(s_i \sin(f_i t) + c_i \cos(f_i t)\right) + c_0 + e(t). \tag{4.3}$$

e are again the residuals or inconsistencies. This e is the signal that remains after the subtraction of these oscillations. c_0 is an offset that is generally the mean value of g. f_i are the four subtracted frequencies; a_i, s_i and c_i are real-valued amplitudes; and ϕ_i is a phase shift between 0 and 2π. This phase shift is the reason why the left side of the equation cannot be adjusted with linear least-squares; but because of this identity, you can just estimate s_i and c_i, and so the adjustment can be done linearly after all. The parametric model for the adjustment is thus:

$$g(t) = \begin{pmatrix} \sin(f_1 t) & \cos(f_1 t) & \sin(f_2 t) & \cdots \end{pmatrix} \begin{pmatrix} s_1 \\ c_1 \\ s_2 \\ \cdots \end{pmatrix} + e(t). \tag{4.4}$$

This equation has m rows as it includes all the observation epochs. The adjustment is carried out as described in Section A.2. All the epochs are assumed to have the same precision, so no weighting matrix P is needed. This general concept is called *Least-Squares Spectral Analysis (LSSA)* (Craymer, 1998).

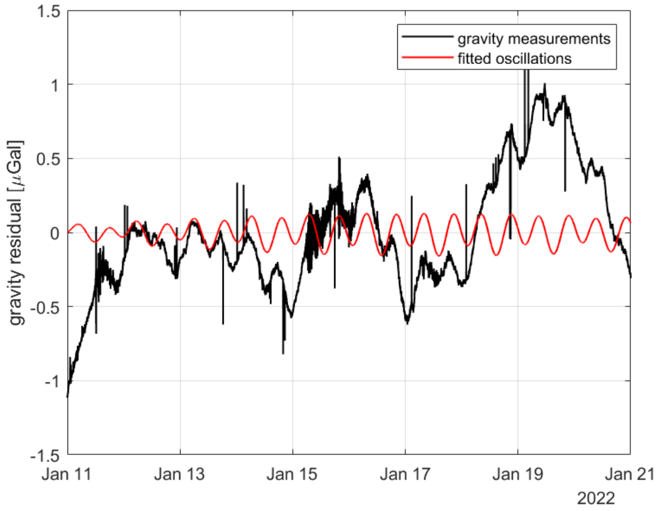

Figure 4.16.: Comparison of the gravity residuals and the fitted oscillations over 10 days in January; the estimation of the amplitudes was done for the whole year, not just these 10 days

Figure 4.16 shows that the adjustment finds a reasonable fit, and in Figure 4.17 it can be seen that the oscillations in the signal significantly decrease. The standard deviation of the signal over the month of January slightly decreases with this correction, from 0.5165 µGal to 0.5153 µGal.

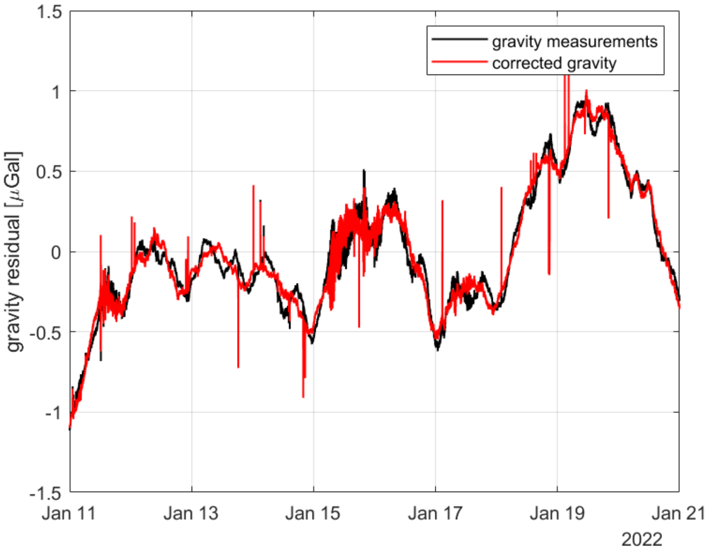

Figure 4.17.: Comparison of the gravity residuals before and after subtracting the oscillations over a few days in January

A few oscillations, for example, between January 11th and January 15th, remain though, and I found no explanation in the soil moisture data for them. A frequency analysis of the corrected signal (Figure 4.18) shows that only a small part of the frequencies have been removed. For example, the power for the frequency of 1 cpd is exactly 0 after LSSA. But there was an even higher peak at 1.003 cpd that is not altered by LSSA at all. To get rid of all these peaks with an LSSA would require the elimination of at least 20-30 frequencies. Removing so many frequencies would entail the risk of destroying some hydrological signal. Therefore, I will explore an entirely different approach and compare its results with the previously presented LSSA.

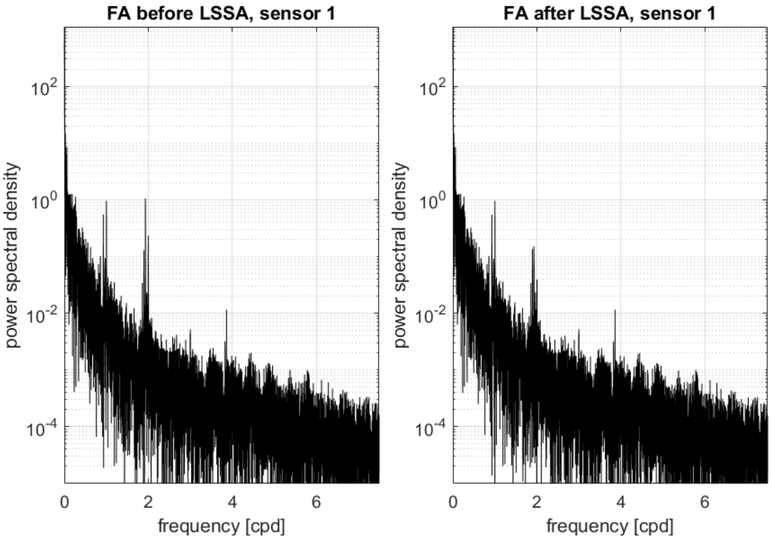

Figure 4.18.: Lomb-scargle periodograms of the signal with and without LSSA for the whole year

For the second approach to eliminating the oscillations, I return to the original observation: the oscillations are slightly delayed from the tides. How large is this delay?
This can be computed by doing a cross-correlation between the gravity signal and the synthetic tides. That is done by shifting the gravity series forward and backward by different amounts and computing the correlation coefficient with the tides. Figure 4.19 shows the results of this cross-correlation over January. The correlation for the unshifted signal is about 0.15. It takes a maximum at a positive shift of 87 minutes at a little over 0.17. A correlation of under 0.2 might not seem significant, but it is a correlation over a long time period, and the tides are very periodic. It is not expected that $g_{(\text{hydro})}$ should have such a periodic character. So it is desired that the correlation between tides and gravity residuals be a lot closer to zero.
The cross-correlation has shown that the observed oscillations have a delay from the tide signal of just under one and a half hours. This delay is not constant, which is not surprising as it was said that the phase lags of the wave groups are time-dependent. For February, the delay increases to 93 minutes, and if the cross-correlation is carried out for the full year (Figure 4.20) the result is 113 minutes.
If there is a strong correlation between the gravity signal and the tides at a certain delay, then the tide signal, delayed by the found amount of time, could be subtracted from the gravity series after estimating an amplitude by a least-squares adjustment. The adjustment is:

$$g(t) = g_{st}(t + \Delta t) \cdot a_t + e(t). \tag{4.5}$$

g_{st} is the timeseries of the synthetic tides, Δt is the found delay of the gravity residuals relative to the tides, and a_t is the amplitude that is estimated. I will call this method *Delayed Tides Adjustment* (DTA).

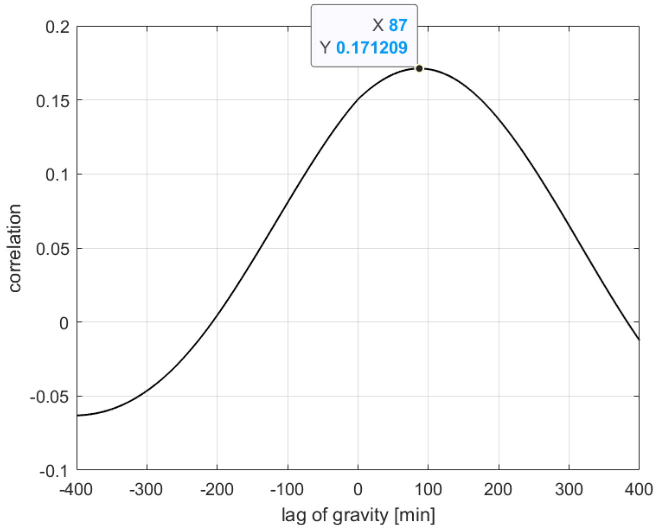

Figure 4.19.: Cross correlation between the synthetic tides and the gravity residuals after atmospheric correction; a positive value corresponds to a positive delay of gravity to the tides; the highest correlation value was highlighted

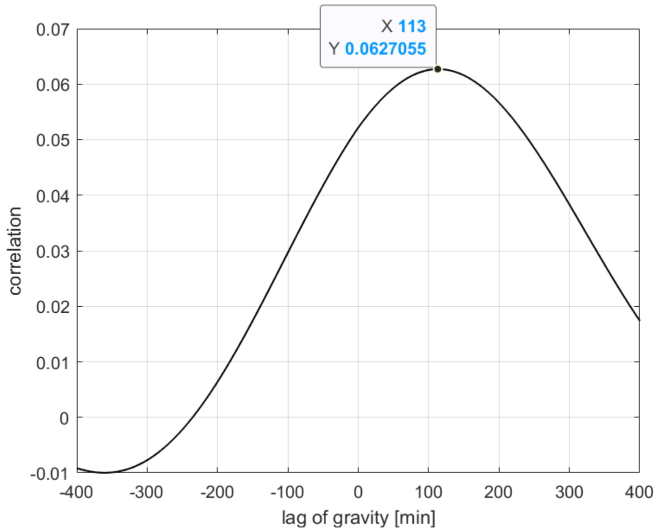

Figure 4.20.: Cross correlation between the synthetic tides and the gravity residuals after atmospheric correction for the full year

As was seen, the exact value of the delay is not clear; there are many choices. If the delay that was found for January (87 minutes) is used and the adjustment of (4.5) is carried out with just the epochs of January, there are some promising results: The standard deviation of the resulting signal has a much stronger reduction than with the method of frequency fitting; the value falls to 0.5083 µGal. If the timeseries is corrected over the whole year with this delay and amplitude, the std actually shows a significant worsening relative to before the correction (see Table 4.1). So the results of January cannot just be applied to the full year.

Rather, a full adjustment with the data and the found delay (113 minutes) of the whole year is necessary. Carrying this out shows the desired results. The std for January is, as expected, not quite as good as when only the data from January was used, but the value of 0.5094 µGal is at least not too far off the January result. The std over the whole year shows at least some improvement over LSSA. The full results are listed in Table 4.1.

Figure 4.21 shows the fitted signal over 10 days in January. The resulting gravity series can be seen in Figure 4.22 next to the LSSA corrected timeseries. The correction is not perfect, but some oscillations of the LSSA approach, like on January 28th, are eliminated or reduced.

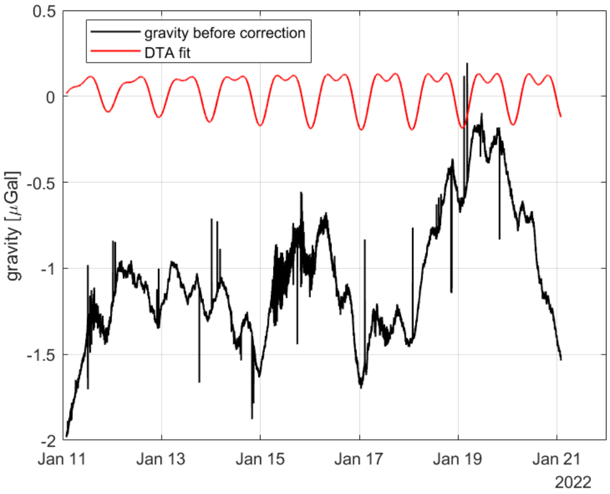

Figure 4.21.: Gravity residuals and fitted delayed tides over a few days in January; the tides were fitted with data from the whole year

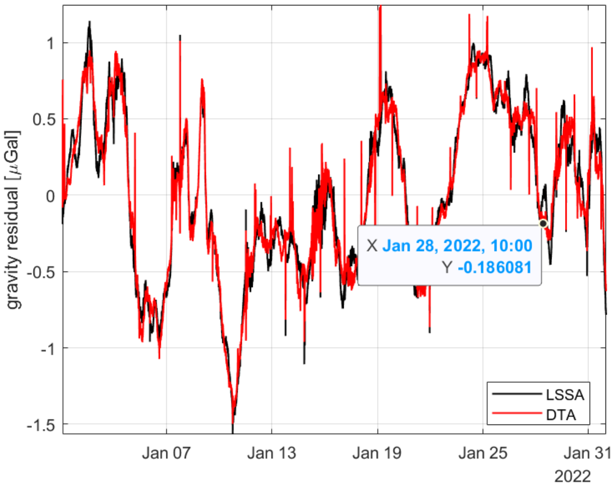

Figure 4.22.: Comparison between the results of the LSSA and delayed tides fitting to eliminate the observed oscillations

As the approach was conceived by looking at the cross-correlation between the gravity residual and the synthetic tides, it is not surprising that a repeated analysis with this metric shows a big improvement. The cross-correlation over the whole year now has a maximum of just 0.017 (Figure 4.23), down from 0.063 before the correction (Figure 4.20), while the LSSA only improved the maximum cross-correlation to 0.049.

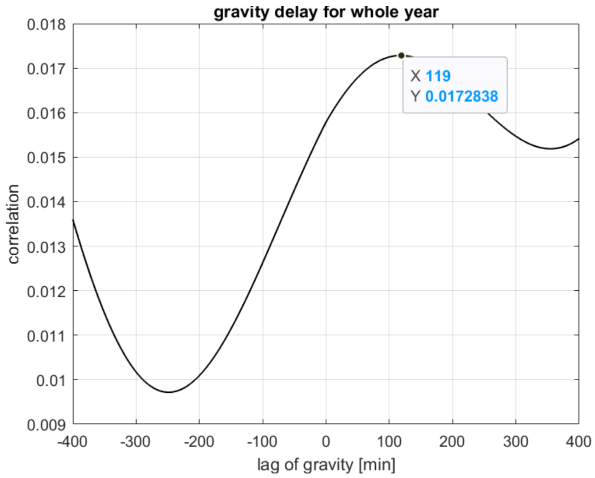

Figure 4.23.: Cross-correlation between the synthetic tides and the gravity residuals after subtracting the delayed tides for the whole year

Table 4.1.: Comparison of different methods for eliminating the remaining oscillations

method	std over Jan. [µGal]	std over year [µGal]	max. cross-corr.
no correction	0.5165	1.6264	0.0627
LSSA	0.5153	1.6254	0.0487
DTA with Jan.	0.5083	1.6551	0.0265
DTA with full year	0.5094	1.6233	0.0173

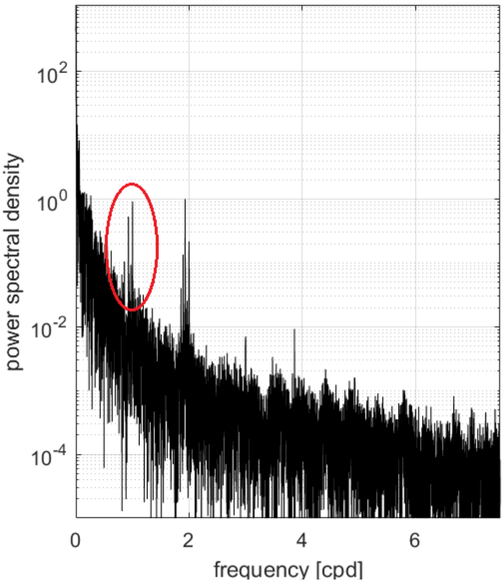

Figure 4.24.: Lomb-scargle periodogram of the gravity residuals of G1-F60 over the whole year after the atmosphere correction. The area where the DTA shows the clearest improvement is highlighted with a red ellipse.

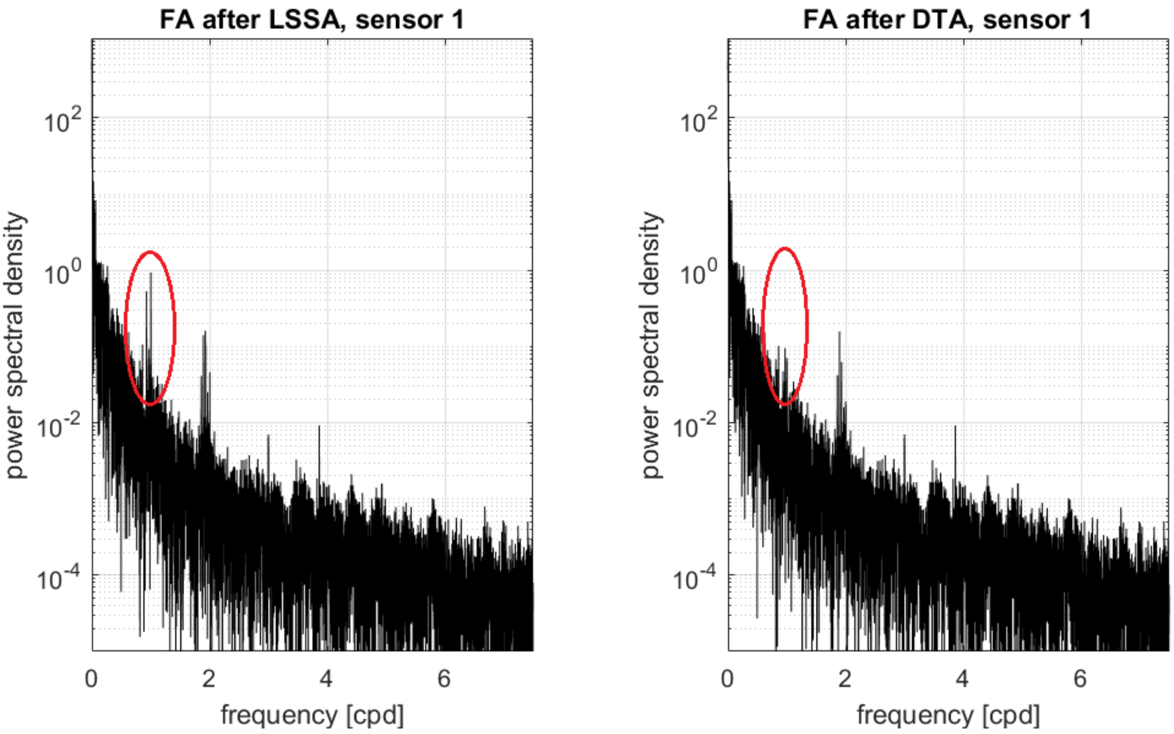

Figure 4.25.: Comparison of the frequencies in the gravity signal after subtraction of the fitted sines and the delayed tide signal. The area where the DTA shows the clearest improvement is highlighted with a red ellipse.

The spectral analysis confirms that the DTA approach achieves better results. Figure 4.24 shows the Lomb-Scargle periodogram of the gravity series without any corrections of the oscillations. Figure 4.25 shows the same analysis for the gravity series after subtracting the oscillations of the four frequencies and after subtracting the delayed tide signal. For the latter method, the peaks at 1 cpd are almost completely gone. There are some peaks in the half-daily area, but they are much fewer and lower than for LSSA. The LSSA only specifically changes the amplitudes of the four frequencies that are eliminated. The periodogram looks similar to the analysis without any correction, apart from the biggest one or two peaks.
Based on these metrics, I conclude that the fitting of the delayed tide signal leads to better results, so the resulting timeseries will be used in all further steps.

4.5. Removal of drift

Like all other relative gravimeters, the SG suffers from an instrumental drift that cannot be accurately predicted (Hinderer et al., 2007). Any changes in the gravity series in a time frame of a year or longer are meaningless because of that, and a removal of drift over the whole year is needed. This needs to be done for both sensors separately because, as was seen earlier in Figure 4.1, the drift rates of both sensors are quite different. The same figure also shows that the difference between the two sensors is almost exactly a straight line, so a linear trend for both sensors will be subtracted. This is done by fitting and subtracting a polynomial of degree 1 for both series separately. Figure 4.26 depicts both timeseries with the adjusted linear trends. Figure 4.27 shows the resulting timeseries of the sensors.

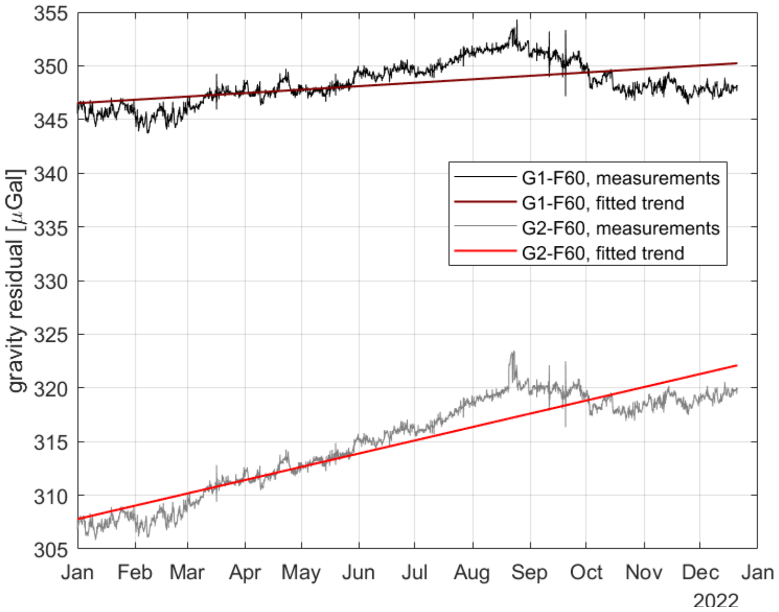

Figure 4.26.: The gravity residuals of both sensors after DTA and the fitted linear trends

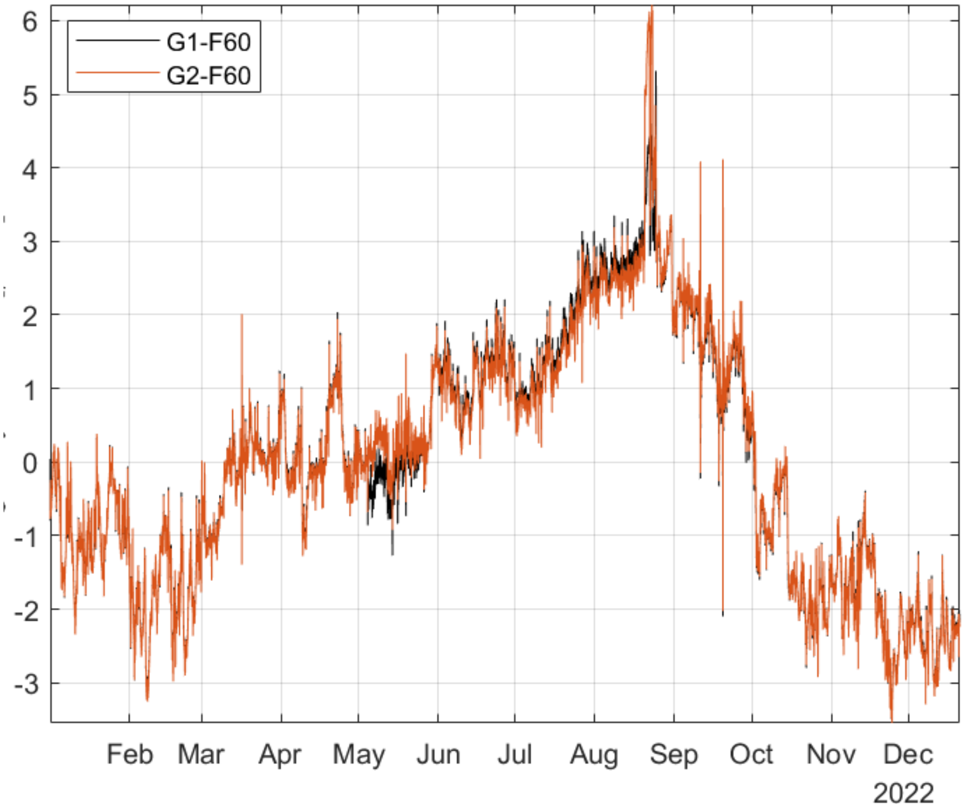

Figure 4.27.: Gravity timeseries after removal of drifts

4.6. Final outlier removal

After performing all the above corrections, it is evident that a few outliers are still in the data. The rough outlier filtering in Section 4.1 was not able to correct all of them. The most likely cause of the peaks is earthquakes.

How exactly does the signature of an earthquake look in the gravity series? The earthquake that was felt strongest by humans in south-western Germany in the year 2022 happened on September 10th at 15:58 UTC (Landesamt für Geologie, Rohstoffe und Bergbau, 2022); it had a magnitude of 4.8 and its epicenter was near Mulhouse, France, about 100 km southwest of the observatory. The earthquake can be seen in the gravity residuals as a short peak (Figure 4.28). Within 10 minutes, the gravity value climbs by over 0.5 µGal. The maximum value of the peak happens at 16:04.

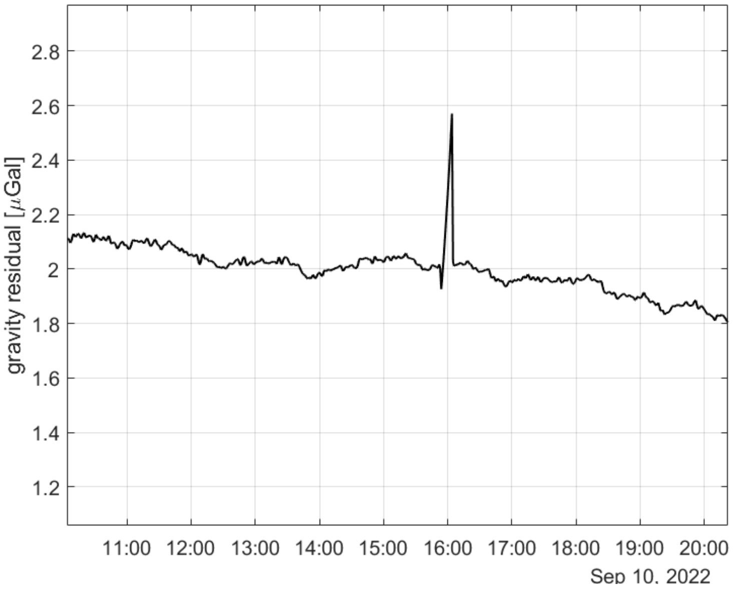

Figure 4.28.: Gravity residuals of G1-F60 over September 10th, where an earthquake happened at 15:58 UTC

Different from such short peaks, there are frequently periods of high noise in the timeseries. If you move only a little bit further from the September 10th event, this peak is absolutely dwarfed by the signal that appears at the beginning of September 11th (see Figure 4.29). It is a day-long period of unrest. At that time, the second-strongest earthquake of 2022 occurred, a magnitude 7.5 earthquake in Papua New Guinea (GEOFON GFZ Potsdam, 2023*b*). It happened at 23:47 UTC on the 10th, while the period of unrest in the SG started at almost exactly 00:00, showing how quickly the fastest seismic waves travel around the globe. This unrest continued for over one day. Because of this length, the data of this day is nearly unusable for hydrological analysis. At least for the weaker events, a final outlier filtering can lead to cleaner data.

As the other distracting signals are gone, both timeseries can now be filtered individually, and the outliers are eliminated. I use a moving window that flags every epoch with a value that is more than three standard deviations away from the median in this window. These epochs are then discarded, and they are linearly interpolated.

It is a sliding window, so it is repeatedly pushed one epoch further. A window length of 5

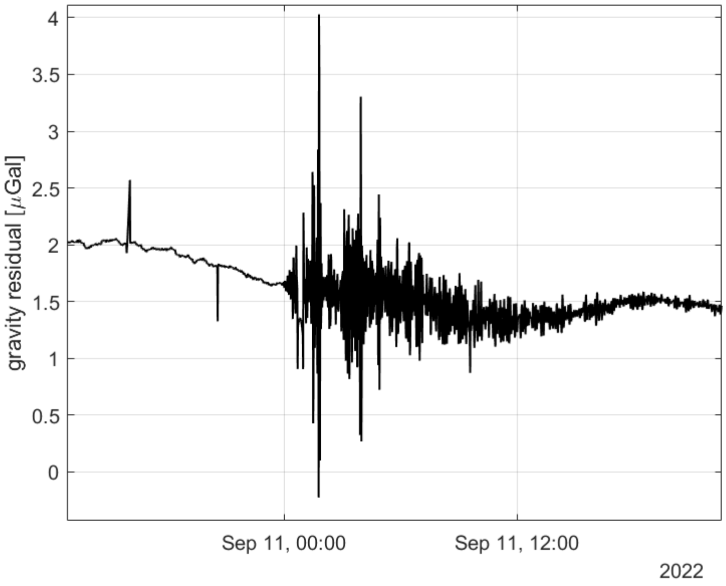

Figure 4.29.: Gravity residuals of G1-F60 over September 10th to 12th; the Mulhouse earthquake can be seen on the left

hours is found to produce satisfactory results. The limit of three standard deviations was also found by experimenting with different values. An elimination with this limit will not remove every epoch that is influenced by something other than hydrology. For example, in Figure 4.29 only the biggest peaks will be flagged. You could also use two standard deviations as the criterion for an outlier, but discarding too many epochs will also not improve the final result as you would just linearly interpolate over a long period of time. Figure 4.30 shows which outliers are found for sensor 1 in January, and Figure 4.31 shows the final result after the outliers are discarded.

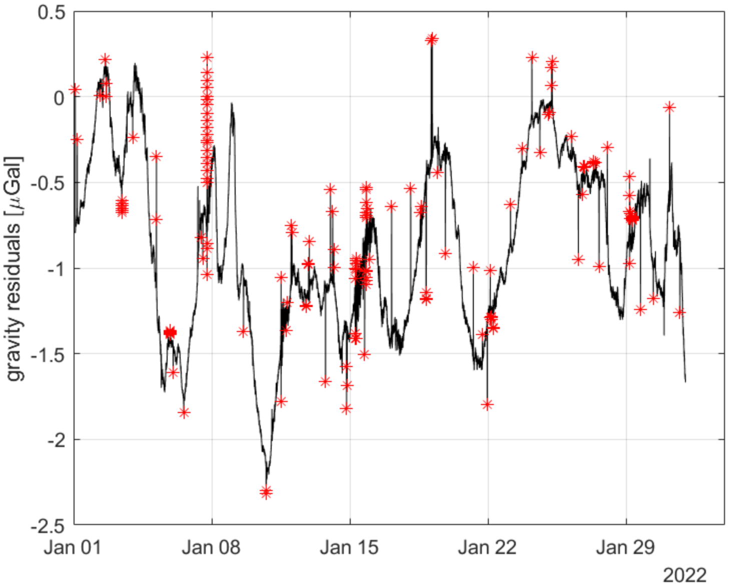

Figure 4.30.: Gravity in January for G1-F60, outliers are highlighted with red stars

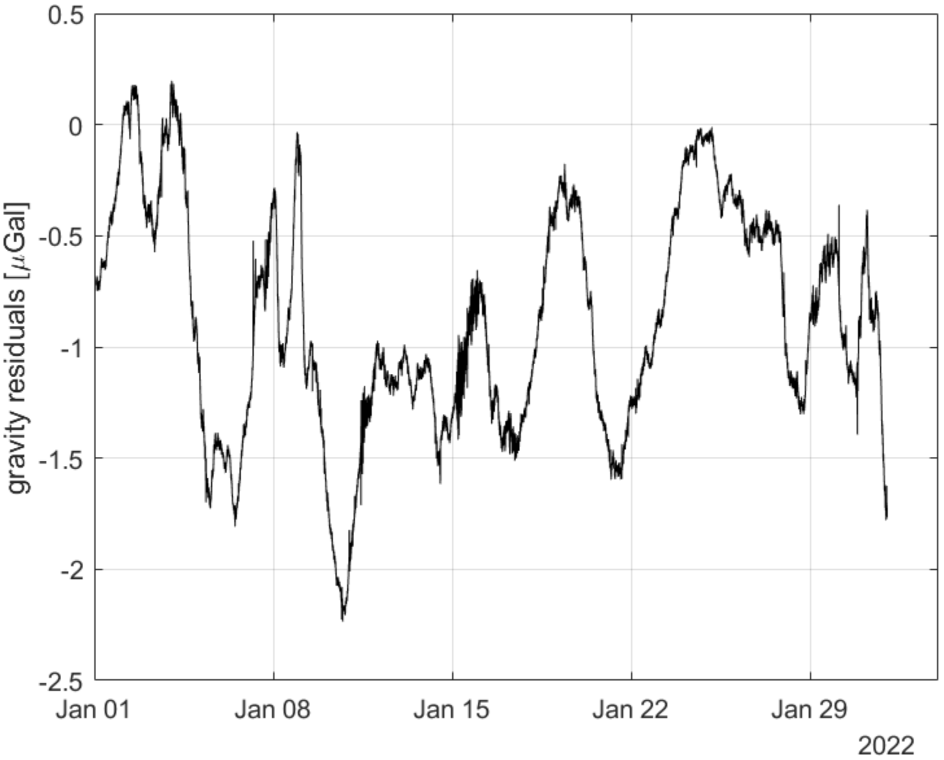

Figure 4.31.: Gravity series of G1-F60 in January after the final outlier removal

In conclusion, three steps in the gravity processing were newly introduced in this chapter: the atmospheric factor was reduced to −0.246 µGal/hPa. The disposal of unwanted oscillations was done, instead of fitting sine and cosine curves of four frequencies, by fitting delayed synthetic tides to the timeseries and subtracting them. Finally, an outlier filtering after the step of drift removal was done for both timeseries.
Figure 4.32 shows the different gravity timeseries with the original processing, referred to as *baseline*, and the refinements of this chapter, referred to as *refined processing*, over the course of the October 2nd rain event. The lower diagrams contain the course of air pressure and precipitation to give a better understanding of the event. There are two separate rains, one in the afternoon and one in the evening of October 2nd. After that, there is almost no precipitation. The original gravity series showed a notable decrease at the end of October 3rd, which cannot be explained by rain. In the new gravity series, this minimum is gone, and there is a more regular recovery of gravity, as one expects with the drying of the soil after a rain. Such unreasonable minima occurred frequently in the earlier signal. They are partly due to the tidal remains, but Figure 4.32 shows that the pressure also plays a role. It takes a maximum around noon on October 3rd. When one uses an atmospheric factor that is too big, this maximum shows up in the gravity series. Both changes on their own, the different atmospheric factor and the tidal adjustment, cannot produce a curve with such a smooth gravity recovery. This only works with both processing changes. There are other events, like the one on August 2nd, where a different atmospheric factor seems to be better suited though (see Section 5.2.4).

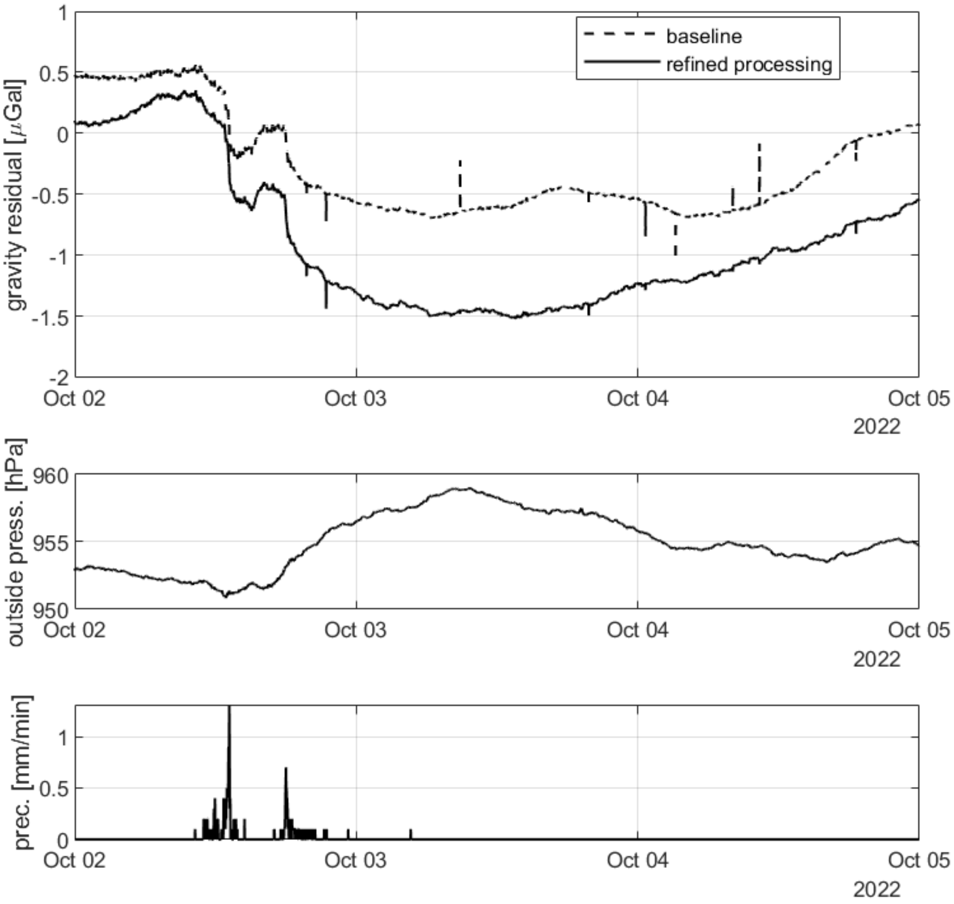

Figure 4.32.: Comparisons of the final gravity results of G1-F60 over the October 2nd event. The differences are different atmospheric factors (baseline: −0.34 µGal/hPa, refined: −0.246 µGal/hPa), a different removal of unwanted oscillations (baseline: LSSA, refined: DTA) and a final outlier filtering in the refined gravity series. The lower diagrams show air pressure and precipitation over the same period of time.

5. Hydrological analysis of gravity data

As discussed in Section 3.2, the soil moisture stations at the BFO have just been added in 2022. In the past, the modeling of the hydrological influence on the gravity measurements was done by using only the data from the precipitation gauge. The aim of this chapter is to build a model that uses the information from the soil moisture stations and that reaches a better fit with the measurements than the Bouguer approach that has been used so far. The model parameters will be adjusted over the eleven selected rain events (see Table 3.3), each with a time window of three days to understand the gravity rebound after the rain. The main goal in this chapter is that the model reaches a small RMSE with regard to the gravity residuals that I computed in Chapter 4, for these 33 days of observation. As I am using a data-driven approach, one cannot expect that the adjusted parameters are going to be ideal for new observation epochs that were not part of the adjustment. I will check in Section 6.3 whether an application of the parameters for new events is possible.

5.1. Modeling through Bouguer approach

The previous correction model that only uses precipitation is based on the work of Schroth (2013). The used observation equation was introduced by Harnisch and Harnisch (2006) and uses the assumption of a Bouguer plate. The impact of a package of rain is not assumed to be immediate, rather there is a gradual increase that is being justified with the infiltration of the water in the ground:

$$\Delta g_i = 2\pi G\rho \sum_{j=1}^{i-1} r_j (1 - \exp(-\frac{i-j}{\tau_1})) \exp(-\frac{i-j}{\tau_2}). \tag{5.1}$$

G is the gravitation constant, and ρ is the density of water. r_j is the precipitation height over a certain sampling rate. Schroth (2013) did her analysis based on hourly precipitation and gravity data, but I am using minutely data as most of the observations have that sampling rate. τ_1 is a time constant describing the infiltration speed of the water, while τ_2 describes the water running away. The result Δg_i is the impact on gravity at epoch i. Harnisch and Harnisch (2006) used $\tau_1 = 1\,\text{h}$ and $\tau_2 = 720\,\text{h}$. Schroth (2013) carried out a grid search with about two years of gravity and precipitation data, trying values between 0.2 h and 10.2 h for τ_1 and 20 h and 720 h for τ_2. She compared the values of RMSE for the resulting timeseries and found $\tau_1 = 4.2\,\text{h}$ and $\tau_2 = 400\,\text{h}$ to show the best results. This is the baseline correction model with which I will compare my results.

As the water height for the Bouguer approximation, it uses a convolution of the precipitation measurements and a kernel that describes the infiltration and disappearance of water. The kernel can be seen in Figure 5.1.

It shows a rise from no impact to a water height of more than 0.9 mm at around 19 hours.

Afterwards, the influence slowly fades away in an exponential decline. After two months, there is still nearly 3 % of the original water quantity left.

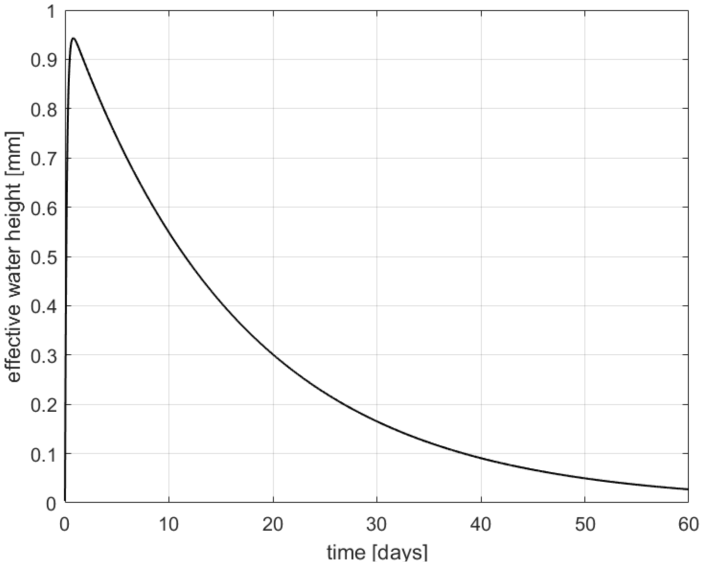

Figure 5.1.: The signature of a rain package of 1 mm height on gravity according to the Bouguer approach

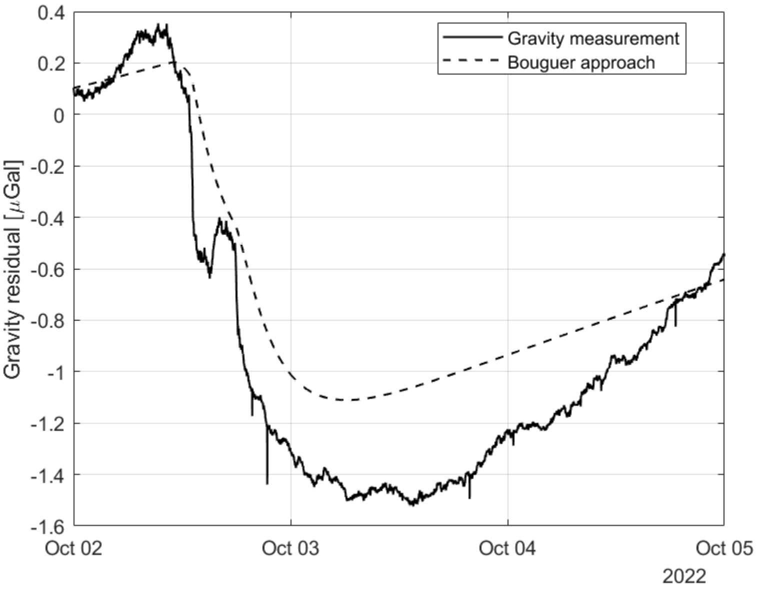

Figure 5.2.: Result of the Bouguer approach for the October 2nd event; the Bouguer graph was shifted in y-direction, so that both timeseries start at the same value

Figure 5.2 shows the result of the Bouguer approach for the October 2nd event. In this case, the correction recreates the overall trend of gravity well, although details like the small jump in gravity between the two rains are not visible and the overall amplitude of gravity decrease is slightly underestimated. For these three days, this correction reaches a strong correlation to the gravity signal of 0.975.

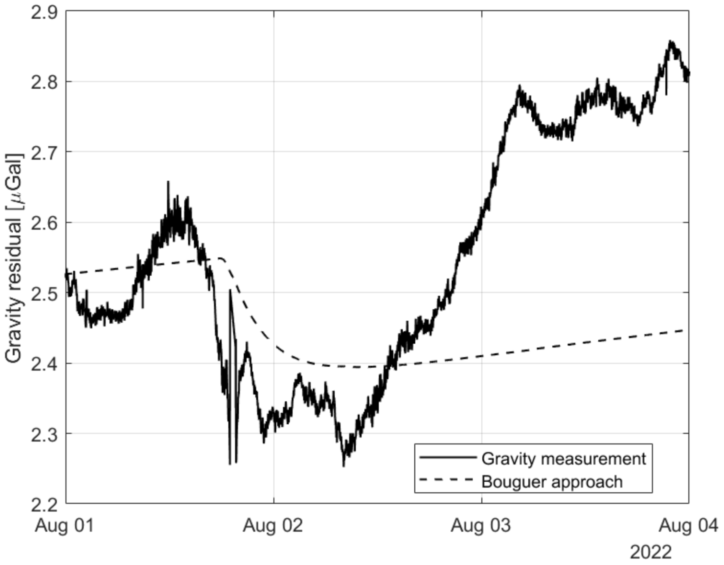

Figure 5.3.: Result of the Bouguer approach for the August 1st event

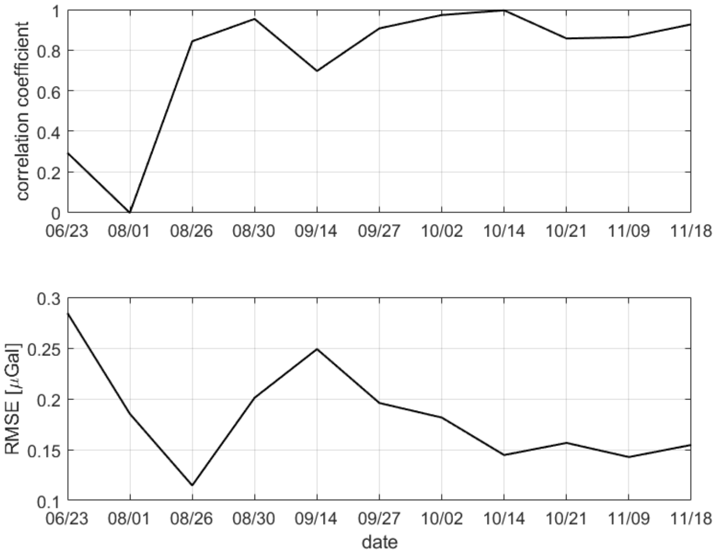

Figure 5.4.: Correlation coefficients and RMSEs of the Bouguer approach over the eleven rain events

There are other events where this approach struggles, though. For example, for the August 1st event, the magnitude of gravity change is overestimated, and the resulting correlation between the two timeseries is only about 0.015 (Figure 5.3). It makes sense that a model that always assumes the same impact of a water package over the whole year will overestimate the gravity decrease in summer and underestimate it in winter. A rain of the same intensity in summer has a lesser influence on gravity as the water is draining off quickly because the soil cannot take up that much water when it is dry. Especially for these summer events, it could be beneficial to use soil moisture measurements, as these do capture the different responses of the soil in summer and winter, in contrast to the precipitation recordings. A smaller change in soil moisture for a summer rain could

translate into a smaller gravity change in the model. The building of a model with soil moisture readings is done over Section 5.2.
The average correlation of the Bouguer approach over the eleven events is 0.761 and the RMSE of the resulting timeseries over the 33 days is 0.186 µGal. Figure 5.4 shows the different correlations and RMSEs over the different events. It is clear that the approach has more problems in describing the gravity variation for the summer events.

5.2. Data-driven models

This section describes the newly introduced building of data-driven models that use precipitation and soil moisture. I will start by using only precipitation, like the Bouguer approach. Then soil moisture sensors are successively added to the model. The determination of the parameters is done with grid searches and linearized least-squares (LLS) adjustments over the $3 \cdot 11$ days of observation that were described in Section 3.4. The early models that use only a few soil moisture sensors are used for comparison and as starting values for the later grid searches and LLS iterations.

5.2.1. Data-driven model with just precipitation

How can this model be improved? As this work aims at a data-driven model, first the main rain event of October 2nd is studied alongside the precipitation data (Figure 5.5).

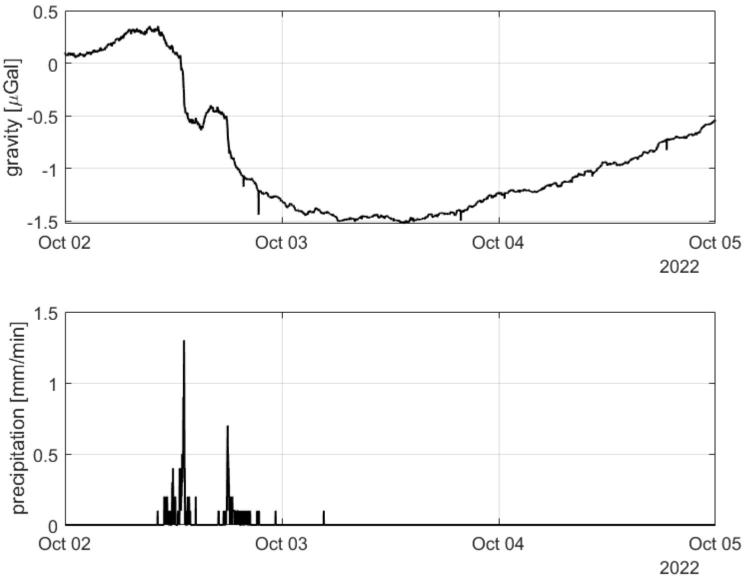

Figure 5.5.: Gravity residual of G1-F60 and precipitation rate for the October 2nd event

For both rain events, the maximum negative derivative of gravity occurs when the precipitation rate is at its maximum. Therefore, it does not seem logical to use a convolution kernel that starts at impact 0. Rather, a simple exponential decay is taken into consideration:

$$\Delta g_i = P_0 \sum_{j=i-N}^{i} r_j \exp\left(-\frac{i-j}{\tau}\right). \tag{5.2}$$

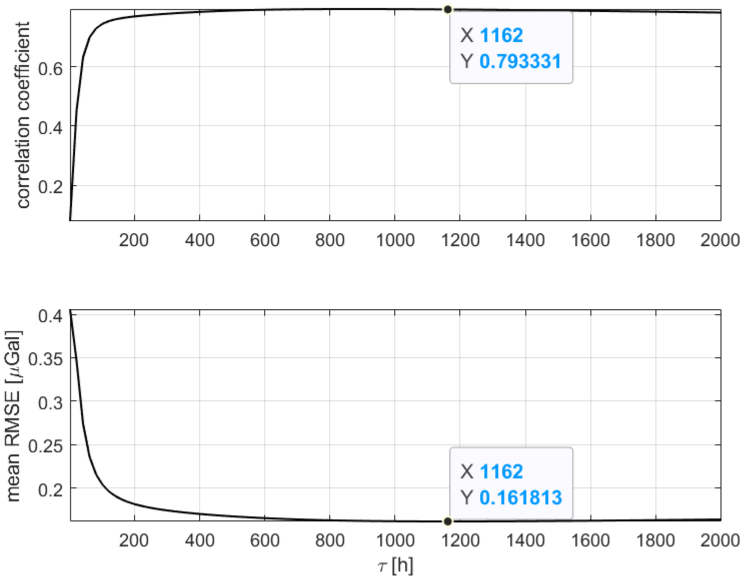

Figure 5.6.: Result of the grid search for the decay time, only using precipitation data with an exponential decay; the best time for RMSE is highlighted

The differences from the formula of the Bouguer approach (5.1) are that the summation is now over a limited amount of epochs to restrict computation time; for precipitation, $N = 60\,\mathrm{d}$ is chosen so that decay times of up to 800 h, about 33 days, are possible without the kernel suddenly dropping to zero from a relatively high point. The infiltration part is gone. The decay time τ will not be fixed but adjusted to the data, and the amplitude P_0 will also be adjusted because it was seen in Section 5.1 that the Bouguer plate value does not explain the data ideally.

To get a first idea of possible results, a grid search is carried out using the data from the eleven events together. Three days of gravity data are used for each event. The list of events can be seen in Table 3.3. Values for τ between 2 and 2002 hours with a step size of 20 h are tested. The different decay times are tried successively, and a best-fitting P_0 is adjusted by a least-squares adjustment. The observation model is that of Equation (5.2) only with an added offset c_k for the different events, as the observation epochs are not continuous. Figure 5.6 shows the RMSE over the eleven events for the different choices of τ.

The search finds better results for long decay times. The RMSE takes a minimum at 1162 h. Both the RMSE value and the mean correlation coefficient of the model with this decay time are better than the result that the Bouguer approach reached. The maximum value for the mean correlation coefficient (cc) appears earlier, at around 900 h.

I repeat the grid search with different kernels. One other possibility is the latter half of a Gaussian kernel. It has a derivative of zero at the start. A higher falling rate appears later, in contrast to the exponential decay. The computation of the kernel is done by

$$k_G(d) = \exp\left(-\frac{d^2}{2\sigma^2}\right). \tag{5.3}$$

The usual factor of $\frac{1}{\sqrt{2\pi\tau}}$ is omitted. If you inserted k_G with the prefactor into Equation (5.2) you would create a direct correlation between P_0 and τ. This is undesirable when analysing the results of the parameters.

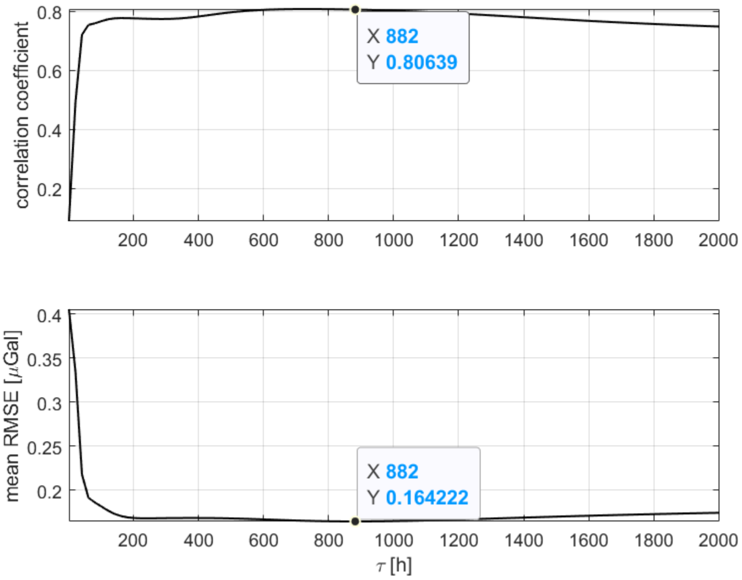

Figure 5.7.: Result of the grid search for the decay time, only using precipitation data with a Gaussian decay, the best time for RMSE is highlighted

The grid search for an optimal decay time with this kernel shows some mean correlation coefficients that are higher than with the exponential decay. But the RMSE values are higher than with the exponential kernel. Figure 5.7 shows the results of this grid search. You could also increase the exponent of d in Equation (5.3) further to three or four, but no good results for the RMSE were found.
The grid search only finds an approximate solution for the parameters, depending on the step size. These results can be used as starting values for an LLS adjustment, though. The iteration is done as described in Section A.3 and uses the gravity measurements of the same $3 \cdot 11$ days. The exponential decay kernel is used along with the starting values of $\tau = 1162\,\mathrm{h}$, the amplitude $-0.034\,\mu\mathrm{Gal/mm}$ and the computed offsets. They were all found in the grid search. The results of the search can be seen in Figure 5.6.
Figure 5.8 shows the response over the October 2nd event of the model that results from the LLS adjustment. In comparison to the Bouguer approach (Figure 5.2) the shape of the second maximum is now better recreated, but the gravity recovery of the final two days fits less well. The amplitudes of the gravity decrease are still too low. Figure 5.9 shows how the RMSE and the changes in the parameter vector between the iteration steps change. As the results of the grid search are used, the iteration already starts from a very good point, and the parameters are only changed slightly. The iteration behaves as intended. The RMSE gets consistently lower and arrives at a similar point as the results of the grid search. It starts from a slightly higher point than the grid search result, as some starting values are only taken approximately. The final mean correlation coefficient value of 0.7933 is also essentially the same as from the grid search. The iteration finds a value for the decay time of about 1170 h. In this case, the LLS adjustment has little benefit as the results from the grid search are equally good. With more parameters that are going to be adjusted later, this refining might become more important.
Figure 5.10 compares the correlation coefficients between the LLS-adjusted precipitation model with the exponential kernel on the one hand and the Bouguer baseline on the other hand. The newly adjusted model performs equally or better on all events except 7 and

10.

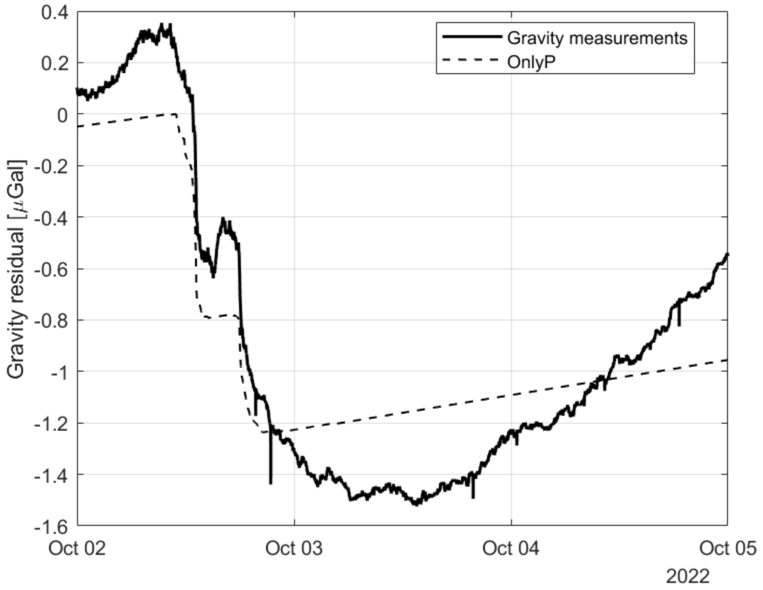

Figure 5.8.: The resulting model response over the October 2nd event of the LLS adjustment that uses precipitation with the exponential kernel

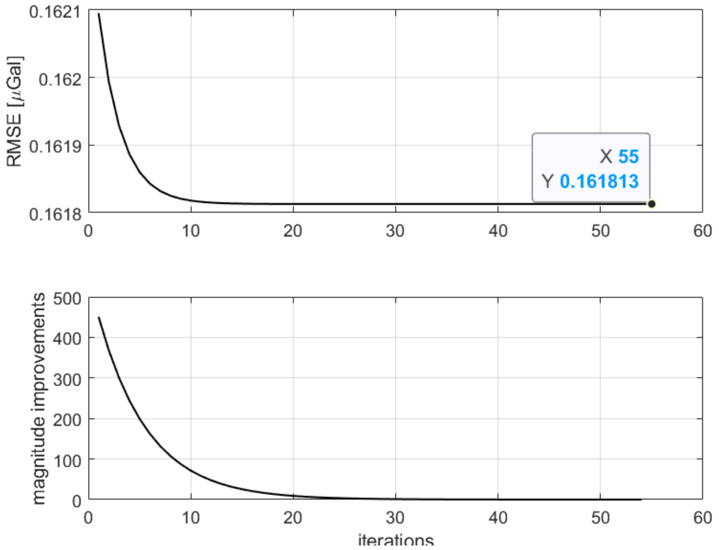

Figure 5.9.: The course of the overall RMSE and the magnitude of improvements between the iterations of the LLS adjustment that uses precipitation with the exponential kernel. The improvement is not a physical quantity, as parameters with different units are added. It is rather a quantity to monitor the iteration.

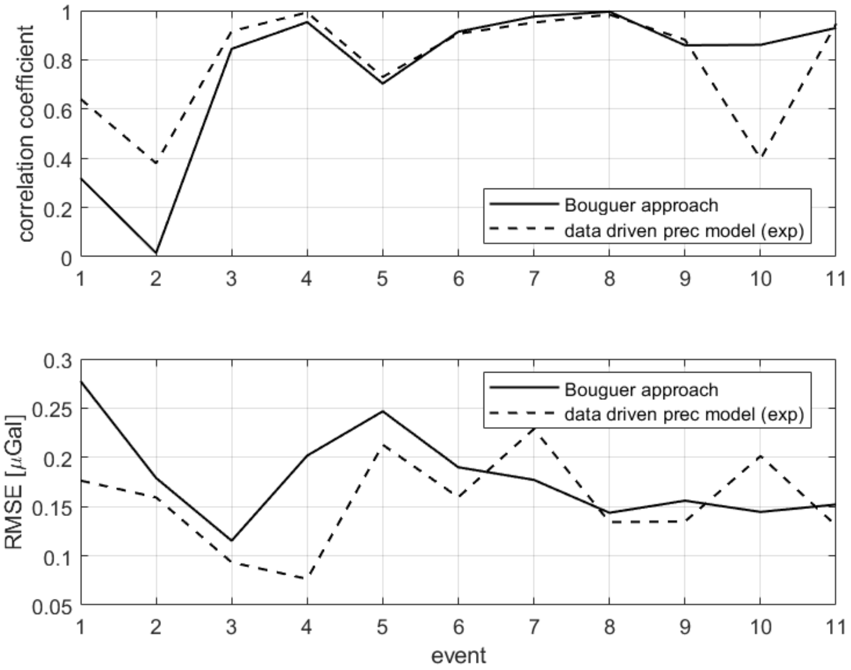

Figure 5.10.: Comparison of the correlation coefficients and RMSE over the eleven events of the Bouguer approach and the new precipitation-based model (with exponential kernel) resulting from the LLS-adjustment

To improve on this, in the next section, the soil moisture data will be used.

5.2.2. Data-driven model with one soil moisture sensor

If soil moisture data are added to the model, at least two non-linear parameters have to be adjusted. From a physical point of view, the probes represent a more long-term memory of the fallen rain, but they still cannot be taken directly into an adjustment and also need a kernel with which they are convoluted. The reason for this is that the gravimeter is influenced by the water mass in the full 100 metres of soil above the gravimeter, while the probes are only at the surface. That means that the water might have drained away from the soil near the probe but is still in the region and only slowly disappears. One possible choice for such a behaviour is the latter half of a Gaussian kernel, as its derivative at time 0 is 0 and it only becomes gradually higher (negatively). The new observation equation for a correction with one soil moisture probe therefore reads:

$$\Delta g_i = P_0 \sum_{j=i-N_1}^{i} r_j \exp\left(-\frac{i-j}{\tau_p}\right) + P_1 \sum_{j=i-N_2}^{i} s_j \exp\left(-\frac{(i-j)^2}{2\tau^2}\right) + c_k. \tag{5.4}$$

s_j is the soil moisture reading at time j. I use probe 0 of Solidus #1 as the first soil moisture timeseries. The readings need to be interpolated to the times of the precipitation measurements so that the same kind of convolution can be done. This is done linearly with the existing Matlab function *interp1*. There are two different integration times: N_1 and N_2. As the soil moisture in itself has a longer memory span of previous rain, it is not necessary to integrate it over a period of months. For N_2, I choose a period of 72 h, but I have to be careful that τ does not become bigger than about 24 h. Otherwise, artificial

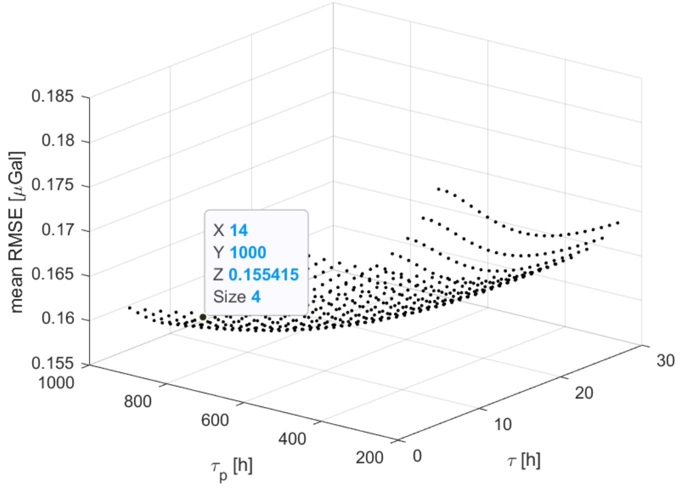

Figure 5.11.: RMSE results of the grid search with the exponential kernel for precipitation and the Gaussian kernel for one soil moisture probe (Solidus #1, probe 0)

jumps, due to the kernel ending, will appear in the simulated time series. c_k is once again the offset for every different event.

Just like in the last section, I am doing a grid search to get an idea of what kind of τ produces good results. This time $\tau_p = 200, 240, \ldots, 960, 1000\,\mathrm{h}$ and $\tau = 5, 6, \ldots, 26, 27\,\mathrm{h}$ are tried, and P_0, P_1, and c_k are adjusted with a least-squares adjustment.

Figure 5.11 shows the results of the search with the exponential kernel for precipitation. The best RMSE result is the 0.1554 µGal down from the about 0.1618 µGal without soil moisture.

Figure 5.12 shows how this result compares with the found result of the grid search without soil moisture. You can see that the model does not improve for every event. For three of the events, there is a visible worsening in the RMSE and correlation coefficient. Only the overall RMSE is guaranteed to be the same or better than without soil moisture. It seems like with the inclusion of soil moisture, the results become slightly more consistent. Some events, like event 4, that were extremely well described by the precipitation-only model have suffered slight setbacks, while the worst results of the previous model, like events 7 or 10, have clearly improved.

As the goal is to explain the gravity measurements as closely as possible, I repeat the grid search with different kernels for precipitation and soil moisture. It will show that convoluting precipitation with the Gaussian kernel has some advantages over the exponential kernel for the resulting correlation (Table 5.1). The Gaussian decay has a degree two for the time difference between the gravity measurement epoch and the time of fallen rain (see (5.3)). That is why I test some higher degrees as well, like the following Degree Three kernel:

$$k_T(d) = \exp\left(-\frac{d^3}{3\sigma^3}\right). \tag{5.5}$$

The exponential and Gaussian kernels are tried for soil moisture. For each combination, the best values for RMSE are chosen, and the RMSE and mean correlation coefficient values are shown in Table 5.1.

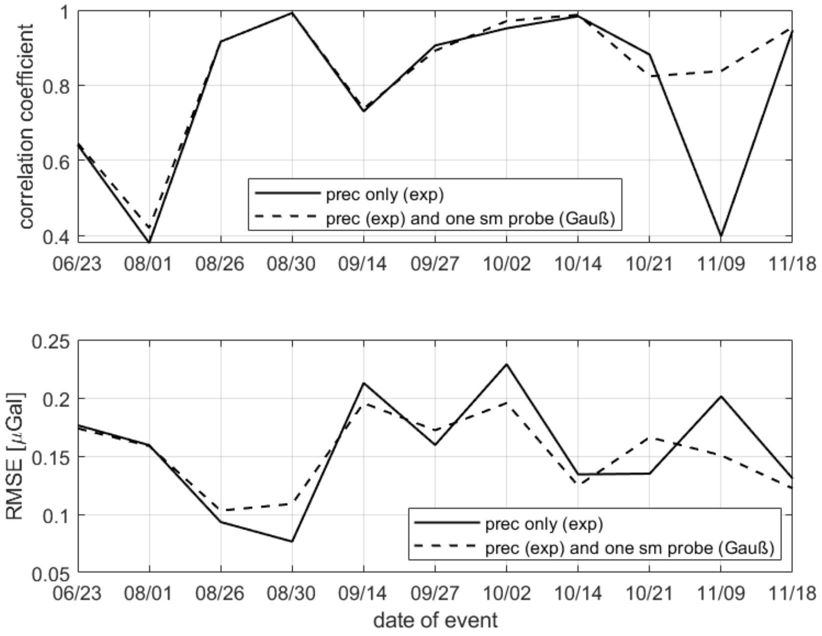

Figure 5.12.: Comparison of the RMSEs and cc for the eleven events for the precipitation-only model, with the exponential kernel; and the model using precipitation and one soil moisture probe, with the exponential kernel for prec and the Gaussian kernel for sm, the results are from the grid searches.

Table 5.1.: Comparison of different kernels with precipitation and one soil moisture sensor; these are the best found results of the grid searches

kernels	RMSE [µGal]	mean cc
only precipitation, exp	0.1618	0.7933
prec, exp; sm, Gaussian	0.1554	0.8313
prec, Gaussian; sm, Gaussian	0.1549	0.8542
prec, Degree Three; sm, Gaussian	0.1551	0.8628
prec, exp; sm, exp	0.1561	0.8318
prec, Gaussian; sm, exp	0.1558	0.8536
prec, Degree Three; sm, exp	0.1561	0.8629

The results look similar for the different kernel combinations. For soil moisture, the exponential kernel performs slightly worse in terms of the RMSE. I will use the Gaussian kernel for all future soil moisture series. For precipitation, the kernels show different strengths. The mean correlation coefficients strongly improve with the higher degree kernels. Some even higher degrees were also tested, and the tendency is that the mean correlation coefficient improves while the RMSE worsens (results not shown). Because of these ambiguous results, I leave the question open about whether the exponential, the Gaussian, or the Degree Three kernel is best suited until more tests are carried out with an increased number of soil moisture sensors. For the exponential kernel, the grid search found the best result for the highest tried value of τ_p, so the ideal value is likely to be above that. That is why I also need to compare the results for these kernels after an LLS adjustment. The results are presented in the remainder of this section.

The LLS adjustment with the exponential kernel finds a precipitation decay time of 1154 h, and the final RMSE is at 0.1551 µGal, which is about how the other kernels performed in the grid search. Figure 5.13 shows the response of the model over the October 2nd event after the adjustment alongside two models without soil moisture. Compared to the data-driven precipitation-only model, this looks smoother, and the gravity minimum after the rain has passed on October 3rd is better matched to the measurements.

After the LLS adjustment, there is nothing to choose between the three kernels in terms of RMSE (Table 5.2). The exponential kernel cannot reach the levels of the mean correlation coefficient of the other two models.

Table 5.2 shows the full results of the LLS adjustments, including the adjusted amplitudes. For reference, the Bouguer approach has an amplitude of −0.042 µGal/mm.

Table 5.2.: Comparison of different kernels with precipitation and one soil moisture sensor after LLS adjustment

kernels	τ_p [h] P_0 [µGal/mm]	τ [h] P_1 [µGal/%]	RMSE [µGal]	mean cc
only precipitation, exp	1170 −0.034	- -	0.1618	0.7933
prec, exp; sm, Gaussian	1154 −0.030	12.9 $-5.96 \cdot 10^{-5}$	0.1551	0.8342
prec, Gaussian; sm, Gaussian	867 −0.028	12.1 $-7.25 \cdot 10^{-5}$	0.1549	0.8538
Degree Three; sm, Gaussian	880 −0.027	11.9 $-7.84 \cdot 10^{-5}$	0.1551	0.8628

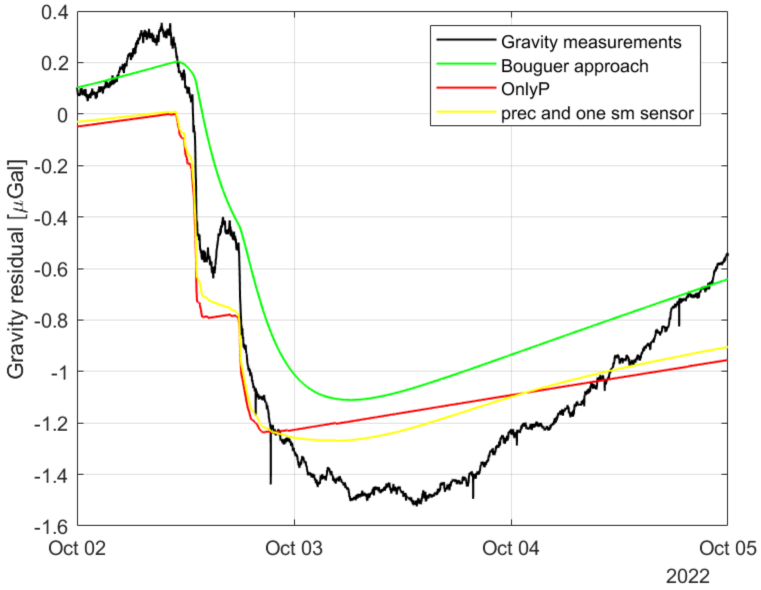

Figure 5.13.: The resulting model of the LLS adjustment using precipitation with the exponential kernel and data from Solidus #1, probe 0 with the Gaussian kernel. It is compared with the Bouguer baseline and the model that used only precipitation with an exponential kernel.

5.2.3. Data-driven model with two soil moisture sensors

Using two soil moisture probes gives a better picture of the overall soil situation around the gravimeter if they are not close to each other. In this section, the data from the first probe of Solidus #2 will be used in addition to the previously used probe 0 of Solidus #1. The observation equation becomes:

$$\Delta g_i = P_0 \sum_{j=i-N_1}^{i} r_j \exp\left(-\frac{i-j}{\tau_p}\right) + \sum_{h=1}^{2} P_h \sum_{j=i-N_2}^{i} s_{j,h} \exp\left(-\frac{(i-j)^2}{2\tau_h^2}\right) + c_k. \tag{5.6}$$

Here $s_{j,h}$ is the soil moisture reading of the h-th timeseries (Solidus #1, 0 or Solidus #2, 0) at epoch j.

There are now three different decay times that need to be established, which makes a grid search more difficult. As a first step, I vary τ_p and τ_1, while τ_2 has a fixed ratio to τ_1. Like in the last section, τ_p values between 280 h and 1000 h are tried with 40 h steps between them. τ_1 runs from 5 h to 20 h with a step size of one hour.

For example, Figure 5.14 shows the result when τ_2 is half as big as τ_1. The best RMSE in this case is the 0.1516 µGal that is highlighted.

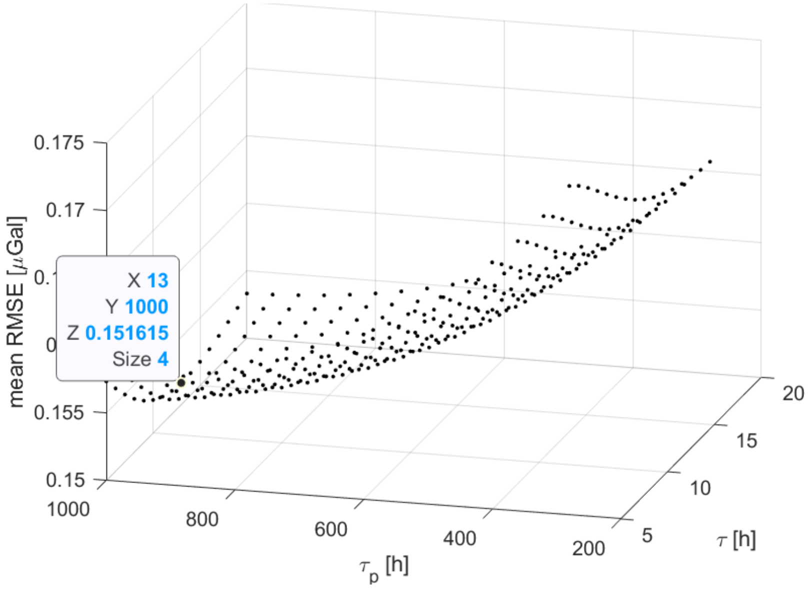

Figure 5.14.: The RMSE values of the grid search with two soil moisture probes. Precipitation is convoluted with the exponential kernel; τ_2 is fixed to be half as big as τ_1

When precipitation is convoluted with the exponential kernel, the best RMSE results always occur with τ_p = 1000 h (e.g., Figure 5.14). This remains true for different ratios between τ_1 and τ_2. Because of this, the grid search is repeated with τ_p fixed at 1000 h and τ_1 and τ_2 varying independently between 5 h and 20 h. Figure 5.15 shows the results of this grid search. The best RMSE this time reaches 0.1508 µGal. These are good starting values for an LLS iteration. The grid search already showed the tendency of τ_2 to become very low. This is also showing in the iteration. To avoid τ_2 trending towards a value of only a few minutes, I am increasing the weight of the starting observation τ_2 = 5 h. A result with such a small τ_2 has a slightly lower overall RMSE, but the resulting model

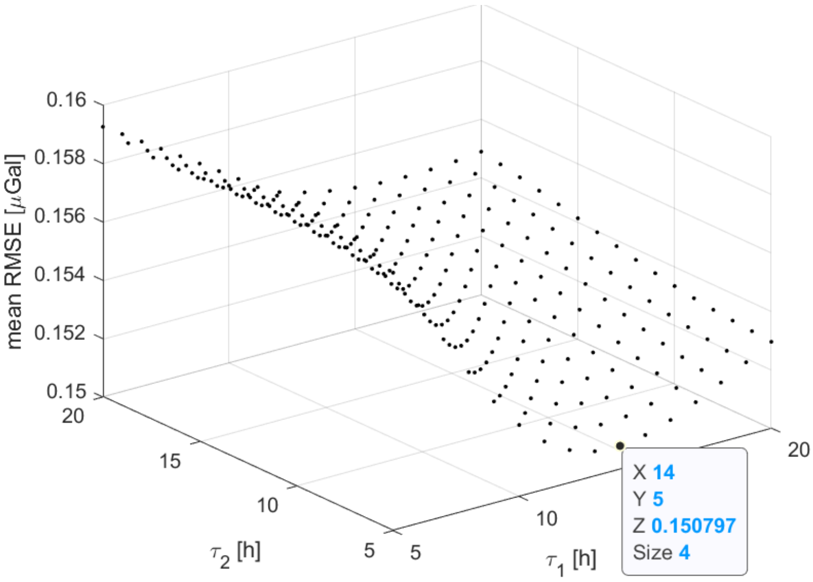

Figure 5.15.: The RMSE values of the grid search with two soil moisture probes. Precipitation is convoluted with the exponential kernel; τ_p is fixed while τ_1 and τ_2 are varied

visually does not fit the data well (Figure 5.17), and it produces much worse mean correlation coefficient results. To determine exactly how big the weight has to be, an L-Curve analysis can be carried out (Hansen, 2000). In this case, I choose a weight of $5 \cdot 10^{-4}$.
With the LLS adjustment, the RMSE improves to 0.1497 µGal. The short increase in gravity between the two rains is now better captured, as well as the timing of the minimum after the end of the rain (see Figure 5.16).
The misfortune of this result is the fact that it comes with amplitudes that go against the physical expectation. It is expected that higher soil moisture always decreases gravity, and therefore all the amplitudes have to be negative. As the least-squares adjustment only adjusts the timeseries so that it fits best, it does not consider this restriction. The estimated amplitudes with these found decay times are −0.033 µGal/mm, $-9.45 \cdot 10^{-5}$ µGal/% for Solidus #1 and $+3.59 \cdot 10^{-4}$ µGal/% for Solidus #2, so one amplitude is positive. This underlines that this work does not build a full-fledged physical correction model but a data-driven model in which the parameters are not necessarily physically representative. As the fit of the resulting model is quite strong (see Figure 5.16) and it is not straightforward to design an adjustment that only finds negative amplitudes, I will use these results for the further analyses.
The same sequence of analyses — grid search for τ_p and τ_1 with τ_2 fixed to it, choosing a value for τ_p and searching for the best τ_1 and τ_2 values, and finally the LLS adjustment — is repeated for the Gaussian kernel and Degree Three kernel for precipitation. Table 5.3 shows the results of the iterations. Compared with the results with one sm probe, the mean cc have not improved while the RMSE value has decreased by about 0.05 µGal.

5.2.4. Data-driven model with three soil moisture sensors

With three soil moisture probes, it is more difficult to find the best decay times by a grid search and to present the results in a clear way to understand the patterns.

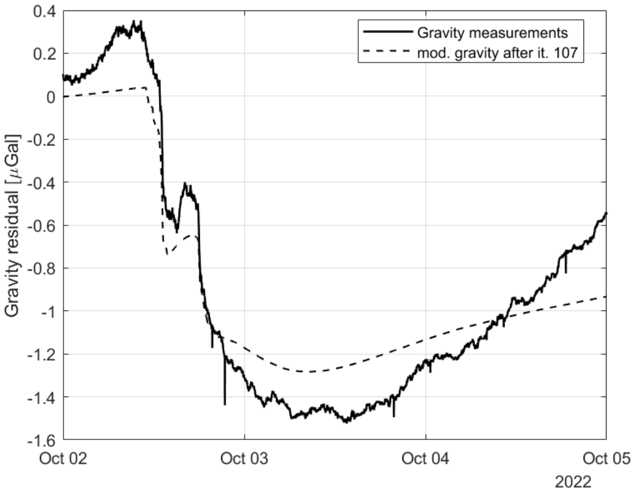

Figure 5.16.: The resulting model of the LLS adjustment using precipitation with the exponential kernel and data from two sm probes with the Gaussian kernel

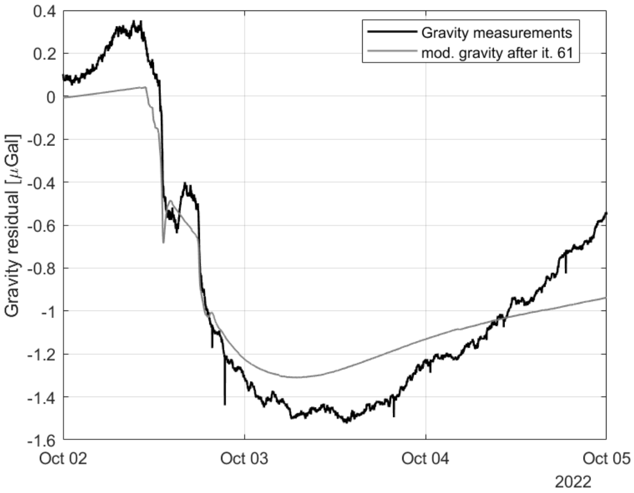

Figure 5.17.: The response of the model with an exponential kernel with precipitation and with two sm probes after the LLS adjustment with a smaller weight for τ_2

Table 5.3.: Comparison of different kernels with precipitation and two soil moisture sensors after LLS adjustment; the units of the amplitudes are omitted for space reason; it is P_0 [µGal/mm], P_1, P_2 [µGal/%]

kernels	τ_p [h] P_0	τ_1 [h] P_1	τ_2 [h] P_2	RMSE [µGal] mean cc
prec, exp; sm, Gaussian	1047	12.5	3.0	0.1497
	-0.033	$-9.45 \cdot 10^{-5}$	$3.59 \cdot 10^{-4}$	0.8324
prec, Gaussian; sm, Gaussian	802	12.2	3.4	0.1490
	-0.031	$-1.1 \cdot 10^{-4}$	$3.33 \cdot 10^{-4}$	0.8524
prec, Degree Three; sm, Gaussian	820	12.2	3.6	0.1492
	-0.030	$-1.16 \cdot 10^{-4}$	$3.13 \cdot 10^{-4}$	0.8633

I carry out a grid search with τ_p fixed at 1000 h and τ_1, τ_2 and τ_3 being tried between 4 and 12 h with 90 min steps in between. That makes a total of $9^3 = 729$ combinations of decay times. In the first search, precipitation is convoluted with the exponential kernel, while probe 1 of Solidus #1 is added to the two previous sensors. The resulting RMSEs are depicted successively (Figure 5.18), as a plot of all four dimensions is impossible without any simplifications.

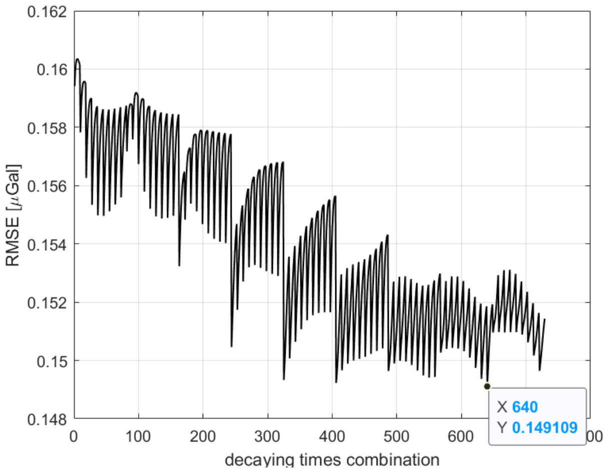

Figure 5.18.: RMSE results of the grid search with three sm sensors (both probes of Solidus #1, p0 of Solidus #2) and precipitation with the exponential kernel

The best found result shows an RMSE of 0.1491 µGal, which is only slightly better than the 0.1508 µGal that was found in the grid search without Solidus #1,1. This is not a strong improvement, although it has to be noted that the grid search with three decay times is coarser so better solutions might be possible with an LLS adjustment. Meanwhile, the mean correlation coefficient degrades from 0.8317 to 0.8159.

Now I do the same grid search with a different set of soil moisture sensors. Instead of using another Solidus #1 probe, the first of Solidus #3 is used. This station is further away from the SG and about 250 m southwest from Solidus #1. It might provide a different picture of the soil moisture in the vicinity, as the probes that have been used so far are all quite close to each other.

The results of this grid search are much improved. The best found RMSE improves to 0.138 µGal (Figure 5.19). The improvement from using two sensors is 7.5 times as big as when adding the second probe of Solidus #1 to the data. The mean cc of this result is with 0.8406 also slightly higher than the one the two-sensor grid search found.

As a refinement, I carry out a second grid search around the best results of the first one (results not shown). This is followed by the usual LLS adjustment, which finds an RMSE of 0.1377 µGal. Figure 5.20 shows the response of the resulting model over the October 2nd event.

The same searches and adjustments are repeated with the other two precipitation kernels. Table 5.3 shows the results.

Just like in the earlier comparisons, the three kernels perform extremely similarly but are different in the mean cc value. So where exactly are the differences between them? Figure 5.21 shows the RMSE and correlation values of the three models over the eleven events. The biggest difference is that the Gaussian and Degree Three models have much better

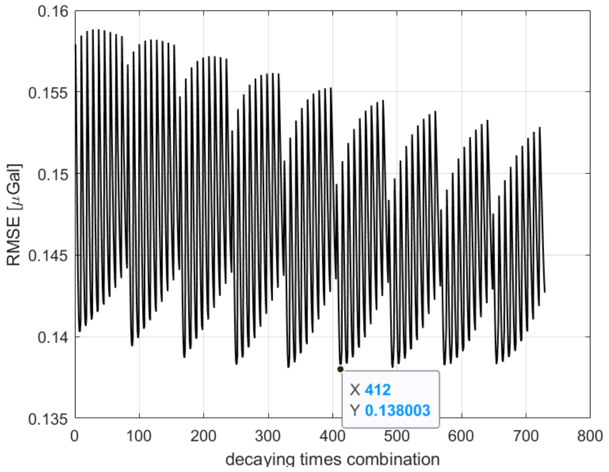

Figure 5.19.: RMSE results of the grid search with three sm sensors (first probes of Solidus #1-3) and precipitation with the exponential kernel

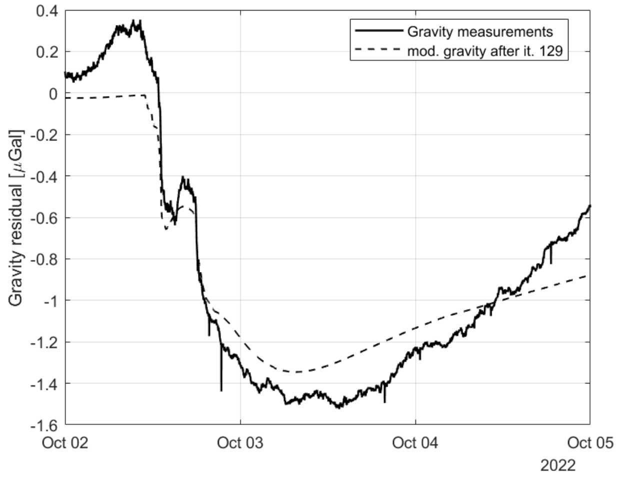

Figure 5.20.: Response of the resulting model of the LLS adjustment; the first probes of Solidus #1-3 are used and prec is convoluted with the exponential kernel; the units of the amplitudes are the same as usual

Table 5.4.: Comparison of different kernels with precipitation and three soil moisture sensors (probes 0 of Solidus #1-3) after LLS adjustment

prec kernel	τ_p [h] P_0	τ_1 [h] P_1	τ_2 [h] P_2	τ_3 [h] P_3	RMSE [µGal] mean cc
exp	1049	9.7	2.1	12.5	0.1377
	-0.027	$-6.43 \cdot 10^{-5}$	$7.32 \cdot 10^{-4}$	$-7.58 \cdot 10^{-5}$	0.8422
Gaussian	824	10.9	3.0	11.9	0.1375
	-0.025	$-6.99 \cdot 10^{-5}$	$5.13 \cdot 10^{-5}$	$-7.98 \cdot 10^{-5}$	0.8663
Degree Three	851	11.8	3.8	11.9	0.1378
	-0.025	$-7.03 \cdot 10^{-5}$	$4.07 \cdot 10^{-5}$	$-8.16 \cdot 10^{-5}$	0.8803

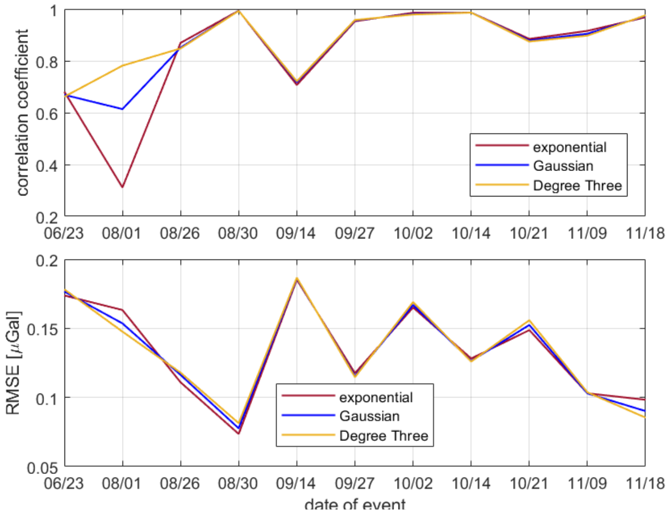

Figure 5.21.: The RMSE and cc values of the LLS-adjusted models with the three different precipitation kernels

results for the August 1st event. For all other events, the three models are very similar. On some of them, the exponential model is the best fit, while, for example, the November 18th also shows a significant separation in favor of the other two kernels.

What is the problem with the August 1st event? It is a summer rain with the least amount of precipitation of all eleven events, so it is clear that this event will be more difficult for fitting a model. Figure 5.22 shows the gravity residuals for the event, along with the three model responses and the pressure. The gravity residuals do not show a course like the October 2nd event, where there is a clear decrease. After a small fall, the gravity quickly rises to a higher value than at the start. It is influenced by other factors, most likely the atmospheric pressure. There is an anti-correlation between the gravity residuals and the pressure. The pressure at the end of this timeseries is lower than at the beginning, so an imperfect correction of the atmosphere can explain why gravity rises to such a high value. If you want to model such an event in the correct way, more attention needs to be spent on the atmospheric correction. Neither of the three models can explain the gravity timeseries properly, even though the Degree Three model shows a higher correlation with the measurements.

As this event shows the biggest differences in the correlation coefficients of the models, it has a big influence on the mean cc values. It is unwise to choose a model based mainly on an event that is influenced only by a small part of soil water. Therefore, I will carry on with the Gaussian kernel, as this looks like a good compromise in Figure 5.21 and the overall RMSE is the lowest of the models (Table 5.4). The resulting model will be called P3S, and its observation equation is

$$\Delta g_i = P_0 \sum_{j=i-N_1}^{i} r_j \exp\left(-\frac{(i-j)^2}{2\tau_p^2}\right) + \sum_{h=1}^{3} P_h \sum_{j=i-N_2}^{i} s_{j,h} \exp\left(-\frac{(i-j)^2}{2\tau_h^2}\right) + c_k. \tag{5.7}$$

During many points of the model building process, one may ask oneself why to choose the best results with regard to RMSE as opposed to the correlation. At this point, I could choose a Degree Three or Four kernel and reach much higher correlations. The answer is given in Figure 5.23. It shows the result of a model that goes along the lines of the

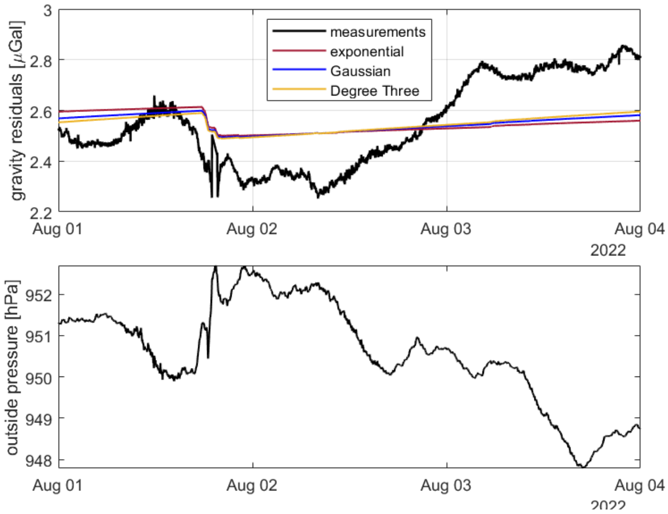

Figure 5.22.: The first graph shows the gravity residuals together with the responses of the three models over the August 1st event. The models use the three described prec kernels. The lower plot depicts the outside pressure measurements during that event.

best correlation values. One ends up with a model that does have a high correlation with the measurements but does not recreate the correct amplitudes in the gravity changes. The RMSE-optimized model delivers a more convincing timeseries. The building of this model is briefly described in Appendix C.

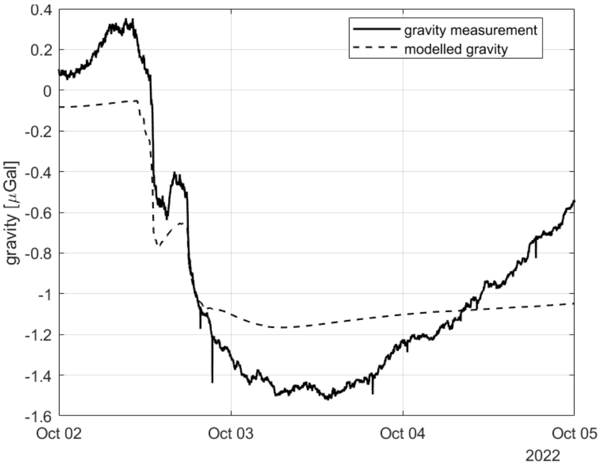

Figure 5.23.: Best result found in the grid search with regard to mean cc over the October 2nd event; it uses the Degree Three kernel for precipitation and the Gaussian kernel with three sm probes

5.2.5. Data-driven model with four soil moisture sensors

To do a grid search with four soil moisture sensors, my approach is to fix the decay time of the first sensor to a previously found value. τ_p is also fixed, like in the last section. The remaining decay times are determined by a three-dimensional grid search.

It was seen in the last section that adding data from a new station can be beneficial to the results. That is why I first try to add the final remaining soil moisture station, TTGOPICO 4. This station only has one sensor and is located near Solidus #3. For the grid search, τ_p and τ_1 are fixed to 832 h and 11 h, respectively, which were the results for these parameters of the LLS adjustment with three soil moisture sensors. Values between 4 and 16 hours are tried for τ_2, τ_3, and τ_4. Figure 5.24 shows the resulting RMSE values of the search. The best RMSE value is 0.1372 µGal, which is only a mini improvement over the results of the three-sensor adjustment (see Table 5.5).

This search and the subsequent refinement by LLS adjustment are repeated for the other three available sensors. For some choices, the RMSE does improve more significantly, for example, with probe 1 of Solidus #2. The full results are shown in Table 5.5. Figure 5.25 shows the model response over the October 2nd event with Solidus 2,1. The model is now closer to the measurements in the morning of October 2nd, but the shape of the smaller maximum is less accurate than with the three-sensor model. This reflects the results of the metrics. Using a fourth probe can improve the RMSE, but the mean correlation is not improved and sometimes even decreases.

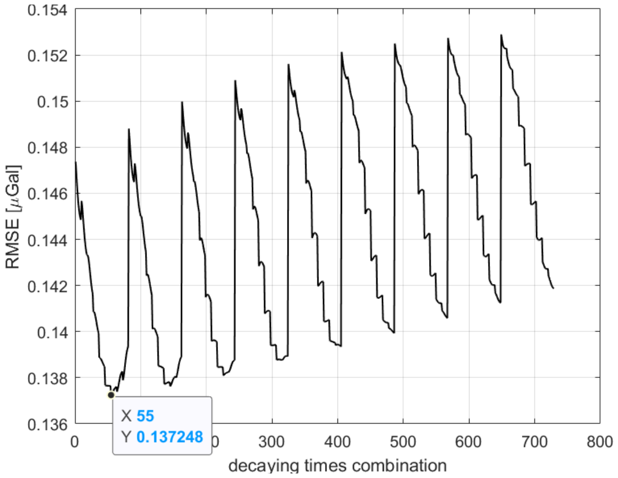

Figure 5.24.: RMSE values of the grid search with four soil moisture sensors. τ_1 is kept at a fixed value. The fourth sensor is TTGOPICO 4.

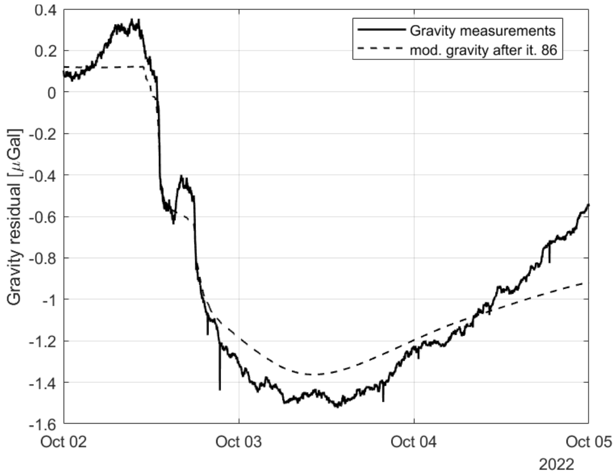

Figure 5.25.: Response for the October 2nd event of the LLS adjusted model that uses probe 1 of Solidus #2 as the fourth sensor

In an additional experiment, I add a soil moisture timeseries to the data that was already used before. So Solidus 1,0 is now convoluted with two different Gaussian kernels, and there are only three different soil moisture timeseries used in the adjustment. The resulting model (see last row of Table 5.5) has an RMSE that is among the best of the real 4-sensor models, and the mean correlation is also similar. So the addition of the fourth sensor has not brought a strong improvement in our case. To judge whether the usage of three soil moisture sensors is generally sufficient, you need more data, maybe also a wider variety of the locations of the soil moisture stations. The magnitude of the improvements certainly becomes smaller after the first few probes.

Table 5.5.: Comparison of the 4-sensor LLS adjustments with different choices of the added fourth sensor; the units of the amplitudes are as normal; the last line describes a model that re-uses an sm series that was already part of the model

fourth	τ_p [h]	τ_1 [h]	τ_2 [h]	τ_3 [h]	τ_4 [h]	RMSE [µGal]
sensor	P_0	P_1	P_2	P_3	P_4	mean cc [%]
none	824	10.9	3.0	11.9	-	0.1375
	-0.025	$-6.99 \cdot 10^{-5}$	$5.13 \cdot 10^{-5}$	$-7.98 \cdot 10^{-5}$	-	0.8663
TTGO-	868	10.5	2.1	12.6	3.2	0.1369
PICO4	-0.024	$-7.83 \cdot 10^{-5}$	$6.96 \cdot 10^{-4}$	$-7.02 \cdot 10^{-5}$	$-9.30 \cdot 10^{-5}$	0.8639
Solidus 3	878	10.4	1.7	12.5	3.6	0.1367
p1	-0.024	$-6.25 \cdot 10^{-5}$	$9.18 \cdot 10^{-4}$	$-7.63 \cdot 10^{-5}$	$-9.76 \cdot 10^{-5}$	0.8672
Solidus 1	829	16.3	7.3	13.0	17.2	0.1348
p1	-0.023	$-1.24 \cdot 10^{-4}$	$1.97 \cdot 10^{-4}$	$-8.87 \cdot 10^{-5}$	$1.11 \cdot 10^{-4}$	0.8475
Solidus 2	866	15.5	6.2	12.9	19.6	0.1341
p1	-0.025	$-8.87 \cdot 10^{-5}$	$2.27 \cdot 10^{-4}$	$-9.36 \cdot 10^{-5}$	$5.99 \cdot 10^{-5}$	0.8627
Solidus 1	824	5.8	4.6	12.9	2.4	0.1346
p0	-0.027	$-3.68 \cdot 10^{-4}$	$3.88 \cdot 10^{-4}$	$-7.22 \cdot 10^{-5}$	$6.15 \cdot 10^{-4}$	0.8649

5.2.6. Data-driven model with seven soil moisture sensors

As a final comparison, I am doing the adjustment with all seven available sensors. The grid search is done by taking the results with four sensors, which used Solidus 3,1 as the final probe, and fixing these five decay times. The remaining three times are determined with a grid search between 4 and 16 h. The best combination of times reaches an RMSE of 0.1337 µGal, which is a small improvement over the adjusted results with four sensors. The LLS adjustment this time does make a difference. It finds a set of parameters with a resulting RMSE of 0.1305 µGal. Figure 5.26 shows the model response for the main event. I will call the resulting model P7S.

The found RMSE is, as expected, the best of all the models that were tried. The improvement over the model that used just three sensors is only 5 % though. The mean correlation coefficient value is with 0.8579 slightly lower than most of the results that were found for the models with three or four sensors (Table 5.5). In Figure 5.26, you can see that details like the gravity increase between the rains are captured less well than in some previous models. In my case, with the locations of the stations and the available data, using more than three soil moisture sensors has a slight benefit.

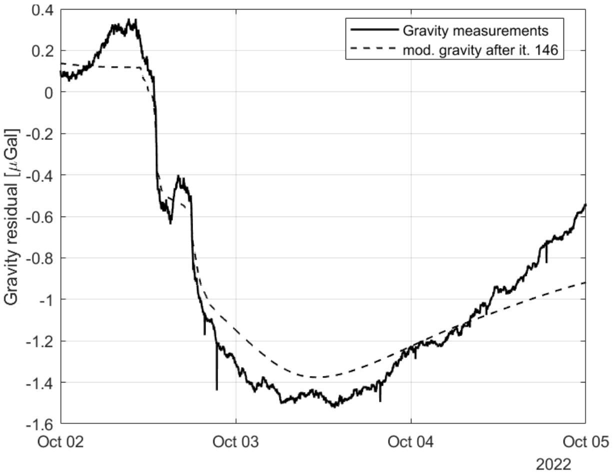

Figure 5.26.: Response for the October 2nd event of the LLS adjusted model that uses all seven sm probes

5.3. Model with soil moisture-dependant parameters

Using the same set of parameters for every event might not be ideal. It is expected that the soil reacts differently to a rain event when it is dry than when it is more moist. In the former case, the ground cannot take as much water. The rain water should drain off faster, which would translate into a shorter decay time τ_p and a lower amplitude P_0.

I first take a step back to get an idea of what the relationship with soil moisture might look like. So far, the adjustment has always used the data from all eleven events together to estimate the parameters plus the offsets. You can also apply the algorithm to the events on their own. The resulting parameters might represent the soil moisture dependency of the response of gravity to rain.

An adjustment for one single event will fit the measurements much closer than the model of the eleven events together. Figure 5.27 shows the modeled gravity for the August 1st event of the model that only uses the data of this event. Previously, the modeled gravity did not follow the measurements very closely, and I concluded that most of the signal, especially the gravity increase on August 2nd, is not hydrologically caused. With no different data to mind, this model now closely complies with the trend of the gravity series. To achieve the substantial gravity increase, the model chooses an amplitude P_2 for the second soil moisture probe that has the biggest positive value of all amplitudes of the eleven events at 0.022 µGal/%. This observation already makes it questionable whether any parameters of the single event adjustments can be meaningfully analysed.

Figure 5.28 shows the estimated precipitation decay times for the eleven events plotted against the soil moisture values before the events. The results seem to be random. There is no trend or clear relationship visible. The picture for other parameters, like the precipitation amplitude, is similar.

It seems that the moisture-dependent reaction of the soil is more complicated than I imagined. There are other factors, like air temperature, the duration, and the intensity of the rain, that could have an impact on the water storage properties of an event. As the single event adjustment wants to explain some parts of the gravity signal that are unrelated to soil water, the resulting parameters are also disturbed.

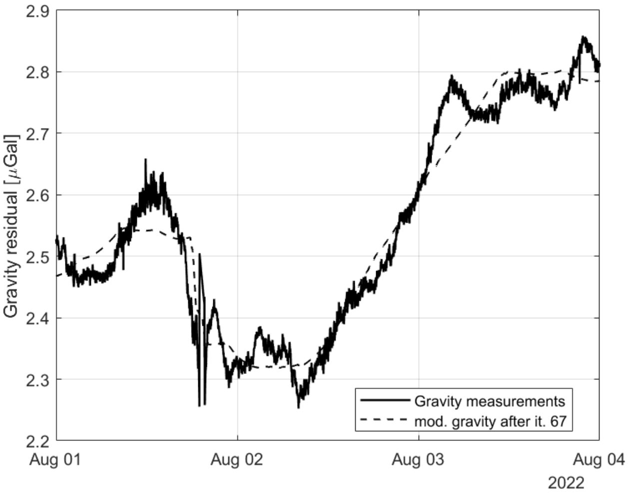

Figure 5.27.: Response of the single event model with three sm sensors for August 1st.

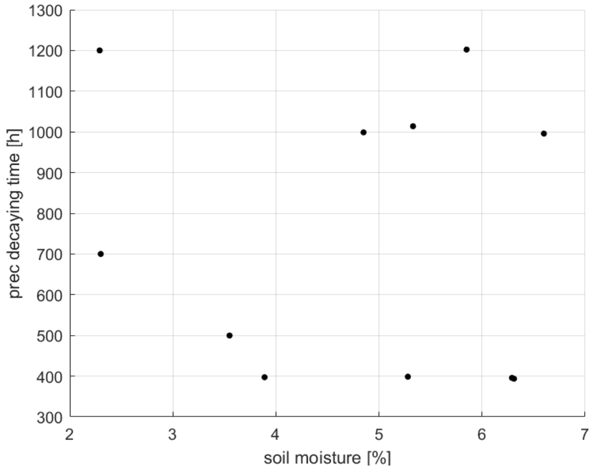

Figure 5.28.: Resulting precipitation decay times for the different events plotted against the soil moisture before the start of the rain; the sm values are from Sol1,1

I add a linear relationship between the soil moisture at the start of the event, which I will call s_k, and the precipitation amplitudes and decay times to the model. s_k is the average soil moisture in the last hour before the rain. For this, a minimum measurement of probe 1 of Solidus #1 is chosen. The exact values can be seen in Table 3.3. The new observation equation is:

$$\Delta g_i = (o_1 + l_1 s_k) \sum_{j=i-N_1}^{i} r_j \exp\left(-\frac{(i-j)^2}{2(o_2 + l_2 s_k)^2}\right) + \sum_{h=1}^{3} P_h \sum_{j=i-N_2}^{i} s_{j,h} \exp\left(-\frac{(i-j)^2}{2\tau_h^2}\right) + c_k. \quad (5.8)$$

o_1 and o_2 are the offsets for the precipitation amplitude and decay time. l_1 and l_2 are the soil moisture slopes of these quantities.

For the LLS adjustment to find a good solution, again, starting values for the two offsets and slopes are needed. Two separate grid searches are used for this.

The first search focuses on the decay time. I need to find an offset and a slope:

$$\tau_p(k) = o_2 + l_2 s_k. \quad (5.9)$$

For o_2 times between 600 h and 1200 h are tried. The slope is varied between −1000 min/% and 1000 min/% with a step size of 50 min/%. For the three soil moisture probes, the decay times of the previous P3S model are fixed, and only the offset and slope of τ_p are varied. The grid search iterates through the combinations and computes the RMSEs.
Figure 5.29 shows the result of the search. The best RMSE is reached with $o_2 = 900$ h and $s_l = -700$ min/%, with a value of 0.1375 µGal that is an improvement over the P3S model only in the fifth digit.

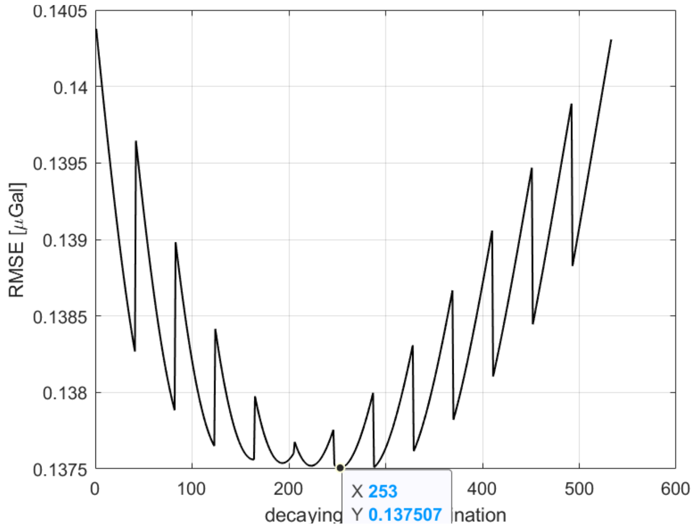

Figure 5.29.: The RMSE values of the grid search that compares different offsets and soil moisture slopes for the precipitation decay time

The second grid search investigates the soil moisture dependence of P_0. The variability of the decay time is left out. τ_p is fixed at 832 h. I add this dependence to the search by multiplying the kernel $\exp\left(-\frac{(i-j)^2}{2\tau_p^2}\right)$ with the factor $\tilde{o}_1 + \tilde{l}_1 s_k$. The least-squares adjustment of a single step of the grid search will determine an amplitude $\tilde{P}_0$ for $(\tilde{o}_1 + \tilde{l}_1 s_k)\sum_{j=i-N_1}^{i} r_j \exp\left(-\frac{(i-j)^2}{2(o_2+l_2 s_k)^2}\right)$. The product $\tilde{P}_0 \cdot (\tilde{o}_1 + \tilde{l}_1 s_k)$ then equals $P_0(k)$.
I try values for $\tilde{o}_1$ between 0.5 and 1.5 with a step size of 1.5 and for $\tilde{l}_1$ between −0.02/% and 0.02/% with 0.002/% steps. As $(\tilde{o}_1 + \tilde{l}_1 s_k)$ is afterwards multiplied with the computed amplitude of the least-squares adjustment, $\tilde{o}_1$ and $\tilde{l}_1$ are relative values. Some combinations of values are almost equivalent. What really matters is the quotient between $\tilde{o}_1$ and $\tilde{l}_1$. This can be seen in the results of the search (Figure 5.30), where the grid search finds multiple best solutions. I choose $\tilde{o}_1 = 0.5$ and $\tilde{l}_1 = -0.006$/% as the starting points for the iteration. The improvement of the best RMSE is slightly higher than when the τ_p was sm varied.
These are again the starting values of the LLS adjustment. The adjustment uses (5.8), so a soil moisture dependence of the amplitude and decay time of precipitation. That means there are two additional parameters to the P3S model which might lead to an improvement in RMSE. The adjustment with four soil moisture sensors also had eight adjusted parameters apart from the event offsets. The best sensor choices there reached RMSEs of around 0.134 µGal (see Table 5.5), so that is the benchmark for this model.
The results fall short of these expectations. The final RMSE is at 0.1374 µGal, so the

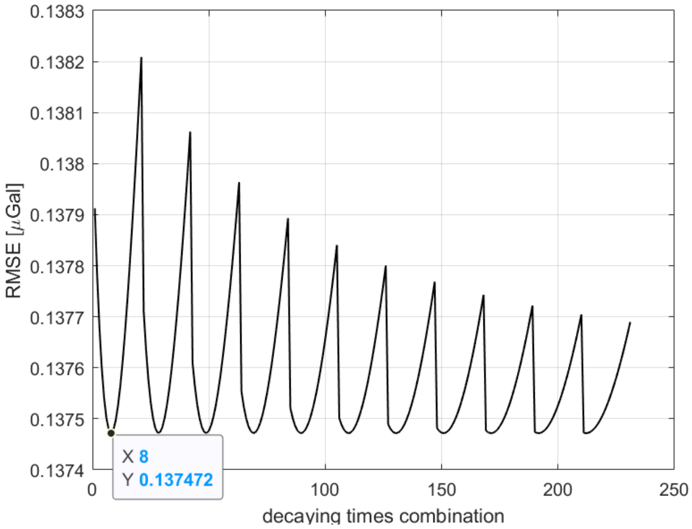

Figure 5.30.: The RMSE values of the grid search that compares different offsets and soil moisture slopes for the precipitation amplitude

improvement over the constant parameter model P3S is barely noticeable. Figure 5.31 shows the response of the model for the October 2nd event.
The adjustment finds a solution of 920 h for o_2 and -687 min/% for l_2, which agrees with the findings of the grid search. The results for offset and slope of the amplitude cannot be directly compared with the grid search as the $\tilde{P}_0$ is between them. l_1 has a positive value, though. As the precipitation amplitude is negative, this means that the absolute value of it decreases with higher soil moisture. The negative l_2 means that the decay is lower for wetter soil. From the physical standpoint, both slopes should have had the opposite sign. This confirms that for the data and the pre-processing that I used, the soil moisture dependence of the precipitation parameters has no significant benefits. I will compare this model again in Section 6, where the model will be called P3SV for its soil moisture variability.

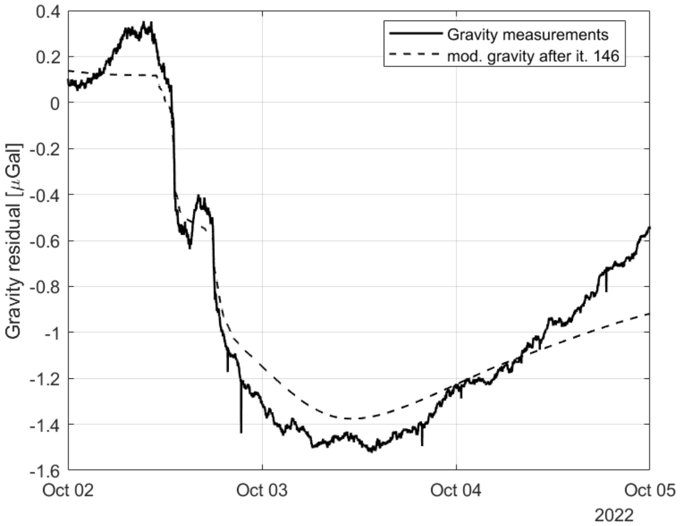

Figure 5.31.: Response for the October 2nd event of the LLS-adjusted model that uses three sm probes and two soil moisture-dependent parameters (P3SV)

6. Evaluation and discussion

I have built mainly three different soil moisture models: P3S uses the data from three soil moisture sensors as well as precipitation. P3SV uses the same data, but the precipitation parameters are linearly dependent on soil moisture values before the rain. P7S uses data from all seven useful soil moisture sensors that were available, but with constant parameters. In this chapter, I will compare the performance of these models in terms of RMSE and correlation coefficient, for data from the rain events that have fed the models as well as for new data from 2023. I will also compare them with two baselines that use only precipitation data: the Bouguer approach that was introduced by Schroth (2013) and the OnlyP model that I developed in Section 5.2.1. Table 6.1 repeats the findings for these different models for the 2022 events.

Table 6.1.: Comparison of the two baselines and the three major gravity models, the number of parameters does not include offsets of the different events

model	parameters	mean cc	RMSE [µGal]
Bouguer	2	0.761	0.186
OnlyP	2	0.8064	0.1642
P3S	8	0.8663	0.1375
P7S	16	0.8579	0.1305
P3SV	10	0.8684	0.1374

6.1. Number of sensors

6.1.1. Comparison of models with and without soil moisture

The results of Chapter 5 already suggest that adding soil moisture observations to the gravity model improves the results. The OnlyP model cannot reach the levels in RMSE and correlation that the models with soil moisture can. Table 6.1 also points out that the soil moisture models have used more parameters than OnlyP. It is possible that these improvements are only due to the increased degrees of freedom of the models. That is why I am also trying a model that uses the precipitation data repeatedly with different Gaussian kernels so that the number of parameters is the same as in P3S.

I call this model P4, as it uses precipitation convoluted with four different Gaussian kernels. In a grid search, I find that, in contrast with the earlier precipitation decay times of around 800 - 1000 h, these times are now lower. They do not get as low as with the soil moisture data.

Figure 6.1 shows a visual comparison between the results after an LLS adjustment, the basic model with only one precipitation timeseries and the model with soil moisture. P4 is only slightly different from OnlyP, while with soil moisture, the details like the location of the minimum on October 3rd are much better explained. In Figure 6.2, the RMSEs and correlations for the eleven events are compared. It confirms that the P3S is the superior model for almost all winter events. The first three events are in summer and late summer, and they show less variation in the sm probes. For these events, the models without soil moisture are sometimes better. In the overall RMSE and mean correlation coefficient, P4 cannot reach the same values as the model with soil moisture. The improvement in RMSE over OnlyP is only about 33 % compared with P3S.

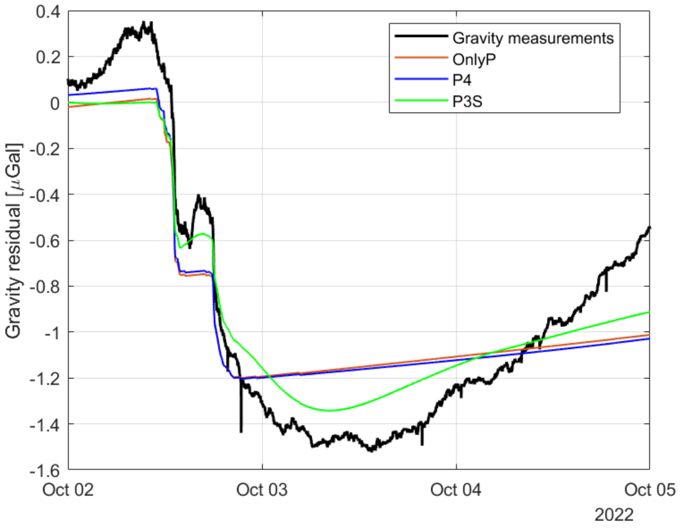

Figure 6.1.: Responses for the October 2nd event of the LLS adjusted models that use precipitation with one Gaussian kernel (OnlyP), precipitation with four Gaussian kernels (P4), or precipitation plus three soil moisture stations

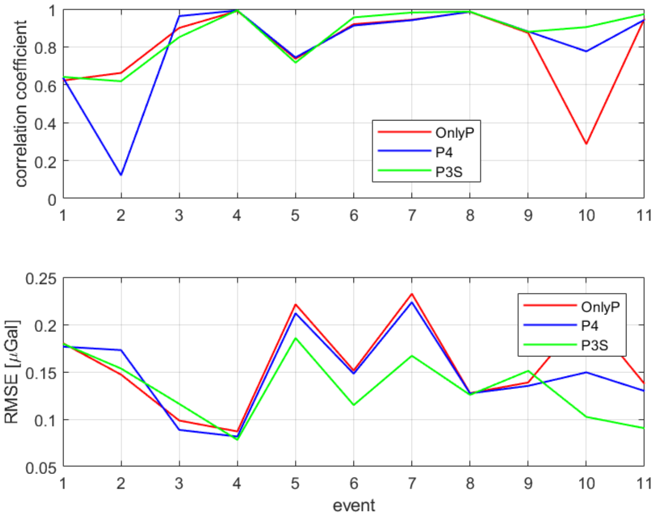

Figure 6.2.: Comparison of the RMSE and cc values for the eleven events of P4 with two other models

The results of this study suggest that using soil moisture as additional information is beneficial. The improvements in results are far greater than have to be attributed to the increase in degrees of freedom.

6.1.2. Comparison of models with three and less soil moisture sensors

The next question is whether one soil moisture sensor is enough or whether it is advantageous to use more than one. To check this, I adjust a model with precipitation and one soil moisture probe that is used three times by convoluting it with different Gaussian kernels. It uses the sm data from Solidus 3,0, and I call it PS3. This model again has eight parameters, so it can be easily compared with P4 and P3S. The overall RMSE improves over the model that only uses precipitation but is also far away from P3S in terms of RMSE. Interestingly, the mean correlation coefficient is very close to P3S. The full results can be seen in Table 6.2.

I am doing the same comparison with a model that uses two soil moisture sensors but also has eight parameters. It uses Solidus 1,0 with one Gaussian kernel, while Solidus 3,0 is convoluted with two different kernels. This model is called P2S3. For this choice of sensors, the adjustment finds quite large decay times for all three sm sensors. There is no obvious physical explanation for that. The RMSE only marginally improves and remains far behind the value of the model that uses three different sm series. From this, I conclude that the model with three sm probes is better than the ones with one or two probes. This is only true with regards to the RMSE, as the mean correlation coefficient value is similar across all three models. Figure 6.4 compares the model responses over the October 2nd event. With more available information, the models are able to explain more details of the gravity timeseries. Figure 6.3 shows the correlation coefficients and RMSEs of the three models over all the events. P3S has the best RMSE over five out of the eleven events, including most of the winter rains. In Section 5.2.5, I have seen that using a fourth sm probe does not produce significantly better results than just adding a second Solidus 1,0 series to the data. So it seems that three soil moisture probes is a good choice for the model, and additional series mostly lead to a risk of overfitting. It has to be noted, though, that this result likely depends on the selection of epochs. If I had chosen the days for the adjustment differently, the best number of sensors might be another one.

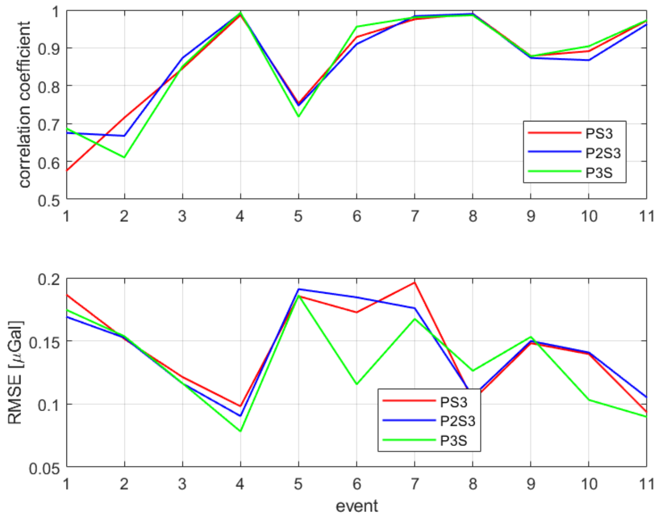

Figure 6.3.: Comparison of the RMSE and cc values for the models using one, two, or three sm probes. All models have eight adjusted parameters.

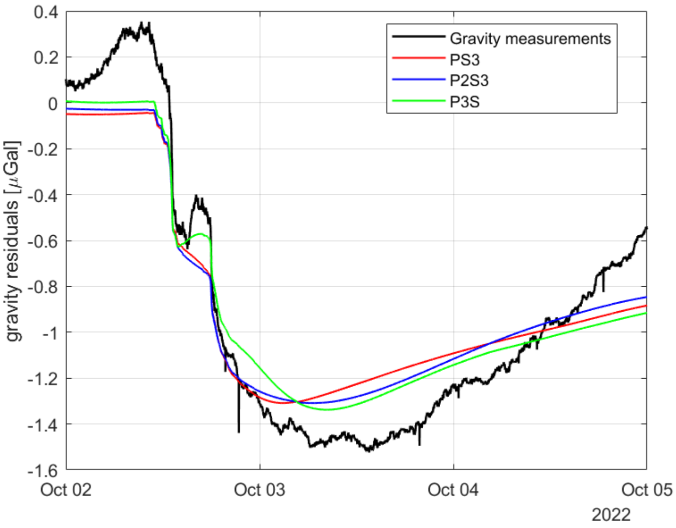

Figure 6.4.: Responses for the October 2nd event of the LLS adjusted models that use precipitation and one sm sensor (PS3), precipitation and two sm sensors (P2S3), or precipitation plus three soil moisture stations (P3S). All models have eight adjusted parameters.

Table 6.2.: Comparison of the P3S model with different approaches that use the same data repeatedly, for example, four times the precipitation data. The τ_i of the models are not necessarily related to the same observation data.

model	τ_p [h]	τ_1 [h]	τ_2 [h]	τ_3 [h]	RMSE [µGal]	mean cc
OnlyP	884	-	-	-	0.1642	0.8064
P4	618	82	298	299	0.1555	0.8112
PS3	879	3.9	4.0	12.6	0.1495	0.8648
P2S3	997	28.4	21.7	22.1	0.1476	0.8674
P3S	832	11.0	3.0	11.9	0.1375	0.8663

One reason why adding soil moisture series to the data does not lead to big improvements is the correlation between the sensors. Figure 3.10 depicts how different soil moisture series are strongly or more weakly correlated. I concluded that probes that are closer to each other are more strongly correlated. Probes that are positioned at a similar height also have a higher correlation. The P3S model uses the probes 0 of Solidus 1-3. It includes both probes at a higher (30 cm) and lower (50 cm) depth. It also uses information from the northern cluster of sm stations (Solidus 1 and 2) and the southern region (Solidus 3 and TTGOPICO 4, see Figure 3.5). That is why additional probes only lead to a small improvement of the model. One could achieve this by using one higher probe in the northern region and a lower one in the south. This is exactly what P2S3 does, but as was seen in Table 6.2, the model with the additional Solidus 2,0 is better. So the information of a lower probe close to the higher probe, Solidus 1,0, seems to have helped the model. But every other of the four probes is either from a station that is already used or it is from a station close to and at the same height as a used sensor. That is why the P3S model is a good compromise between a limited number of parameters and using most of the available information.

6.1.3. Investigating the benefits of using all soil moisture data

The final question I want to address is what the benefits of using the soil moisture measurements from all seven sensors are. For that purpose, I adjust a model with 16 parameters that uses the same three sm sensors of P3S as input. Solidus 1,0 and Solidus 2,0 are each convoluted twice with Gaussian kernels, while Solidus 3,0 is convoluted three times. I call this model P3S7.
Figure 6.5 shows the responses of P3S, P7S, and P3S7 over the October 2nd event. P3S7 describes the intermediate maximum better than P7S. The latter model has a course that is closer to the data in the morning of October 2nd, and in general, it shows a compromise where the difference to the measurements is never too big but the small details are not really recreated. This explains the decrease in the mean correlation coefficient (see Table 6.3). Figure 6.6 compares the RMSE and correlation values of the three models over the events. While P7S has a lower RMSE than P3S on almost every event, the comparison with P3S7 is much closer. Both models show a better result for about an equal number of events.
The overall RMSE of P3S7 is 0.1337 µGal. This is not quite as good as the P7S model. The improvement over P3S is about 54 % of when using the new four probes (see Table 6.3). So in this case, there is value in adding sm data, but it seems that the benefits of adding become smaller. I will see in Section 6.3 whether this model is only better for the data that it was fed with or whether it also produces better results on new data.

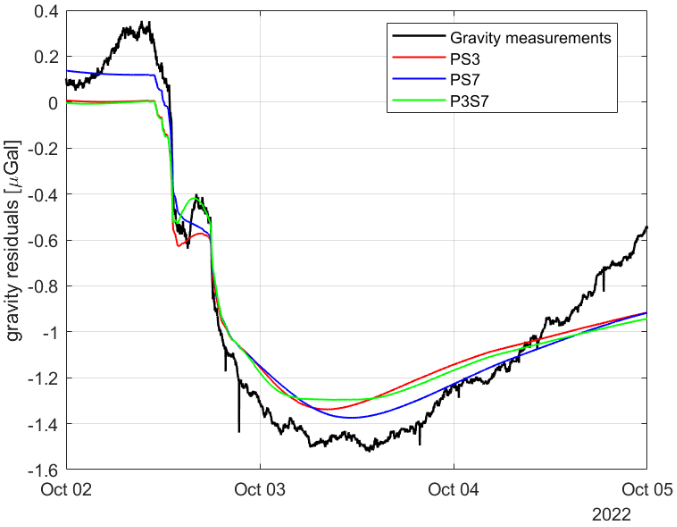

Figure 6.5.: Responses for the October 2nd event of the LLS adjusted models that use precipitation and three sm sensors (PS3), precipitation and three sm sensors with repeated convolutions (P3S7), or precipitation plus seven soil moisture stations (P7S).

The model responses like those in Figure 6.5 and mean correlations of around 0.85 show that finding the hydrological signature by using precipitation and soil moisture data is possible. Most of the details in the variations in Figure 6.5 are recreated by the models, even though the model was adjusted with ten more events. The exact modeling of the amplitude of the gravity decrease and the rate of recovery after the rain cannot quite be achieved, and the fit is not as strong for all the events.

You can never be certain how much of a gravity signal is hydrologically caused. Especially the atmosphere always has to be taken into account as a possible signal source. But especially the gravity variations in the afternoon of October 2nd show strong signs of the hydrological influence. The stopping of the gravity drop and then a second drop is very closely explained by PS3, so it is likely that this gravity change mainly has hydrological causes.

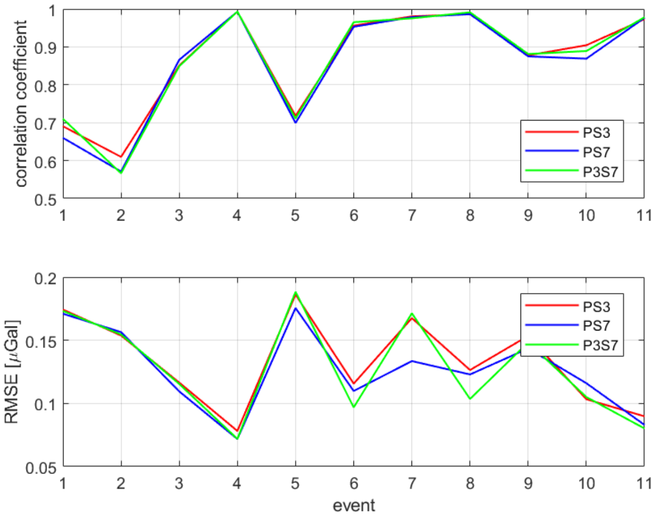

Figure 6.6.: Comparison of the RMSE and cc values for the models using three or all seven sm probes. The latter two models both have sixteen adjusted parameters.

Table 6.3.: Comparison of the models using three and all seven sm probes. P3S7 has the three probes multiply convoluted so that the model has the same number of parameters as P7S

model	mean cc	RMSE [µGal]
P3S	0.8663	0.1375
P3S7	0.8645	0.1337
P7S	0.8579	0.1305

6.2. Leave-out validation

One important question in the evaluation of the models is how well they can be applied to time windows that were not part of the modeling. Because very little physical consideration has gone into the building of the models, their applicability is very uncertain. If they show a good performance in predicting the gravity variations of such events, that means that the found parameters hold information about the general trends of gravity variations following a rain.

In this section, such an analysis is done by leaving out three of the eleven events in the

modeling and assessing the performance of these variations of the models on the three events that were left out. This is called leave-out validation. The models will be compared amongst themselves and with the Bouguer approach.
For the analysis, the three final rains, which happen in October and November, are omitted from the modeling. Starting values are determined for the different models by carrying out grid searches on the eight modeling events. Just using the starting values from Chapter 5, which resulted from grid searches with all eleven events, would induct information from the final events into the models and thus falsify the results. The non-linear least-squares adjustment is then done as described in Section A.3. Figure 6.7 compares the results of this validation for the two models that use only precipitation, the Bouguer approach and the data-driven model OnlyP, over the first left-out events. It is clear that the advantage of the data-driven model in Section 6.1 has disappeared. The Bouguer approach follows the overall variations of the gravity measurements better. Over these three days, the correlation coefficient of the Bouguer approach with the measurements is 0.86 while OnlyP is at 0.81.
Figure 6.8 shows the correlation coefficients and RMSE of the Bouguer and the OnlyP models over the three events that were left out of the modeling. The OnlyP model is much more inconsistent over these three events than the Bouguer model. It has a very bad result for the second event. The overall result of the Bouguer model is better (see also Table 6.4).

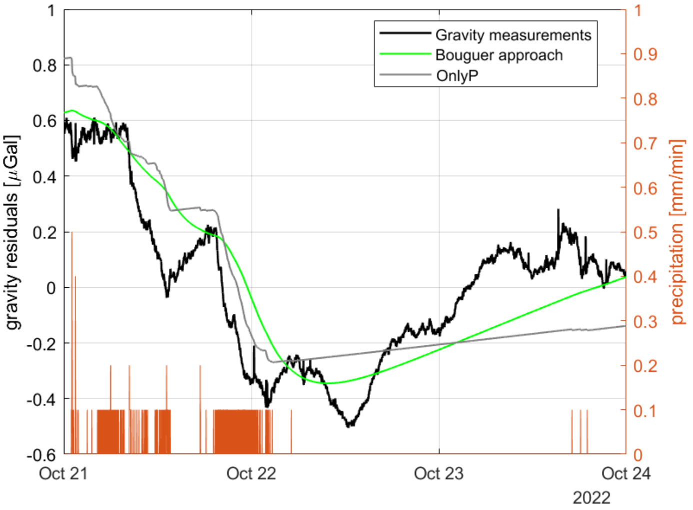

Figure 6.7.: Comparison of the gravity response of the Bouguer approach and the OnlyP model over the October 21st event during the leave-out validation. This is one of three events that are left out of the modeling of OnlyP.

A comparison of the Bouguer-based model with the three models that use soil moisture shows similar results. The data-driven models do not show significant improvements over the Bouguer approach and even perform worse on many occasions. Figure 6.9 shows the gravity responses of these models over the October 21st model, and Figure 6.10 compares the overall results of all three events. The P3S and P3SV models have comparable results to the Bouguer one, while the P7S model is the worst of the four models on two of the three events.

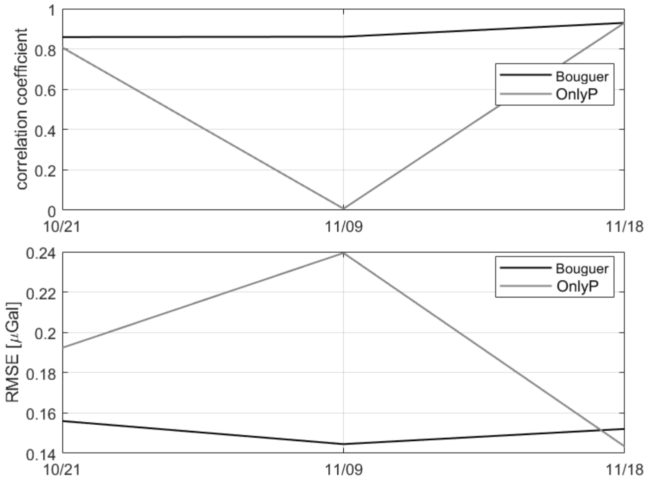

Figure 6.8.: Comparison of the performance of the Bouguer approach and the OnlyP model over three events. The latter model was determined without these three events.

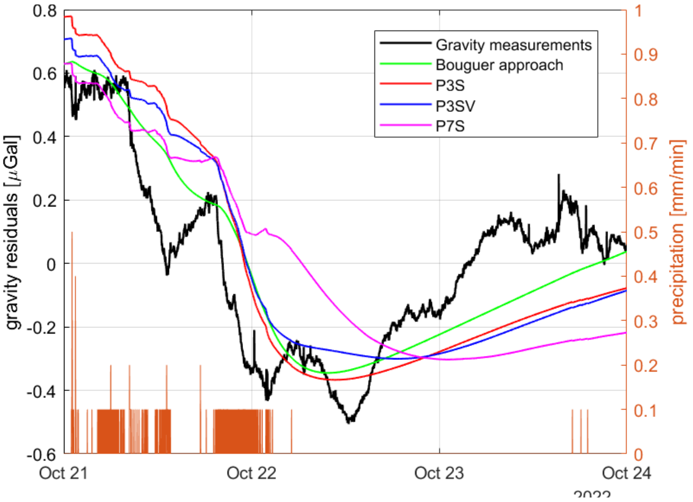

Figure 6.9.: Comparison of the gravity response of the Bouguer approach and three models that use soil moisture over the October 21st event during the leave-out validation. This is one of three events that are left out of the modeling.

Table 6.4 summarizes the results of the different models over the three left-out events and compares them with the full models that included all eleven events. None of the models represents a significant improvement over the Bouguer approach. P3S and P3SV are at least very close to the Bouguer model, and P3S is slightly better in the overall RMSE. As expected, the results are far behind the models that included these three events. The gap is particularly strong for the OnlyP and P7S models.

So these results suggest that the application of models to new events is difficult. The Bouguer baseline performs at least as well as the new models. To assess whether this is true for the full models that include all 33 days for the modeling, I will also apply these models to completely new rain events.

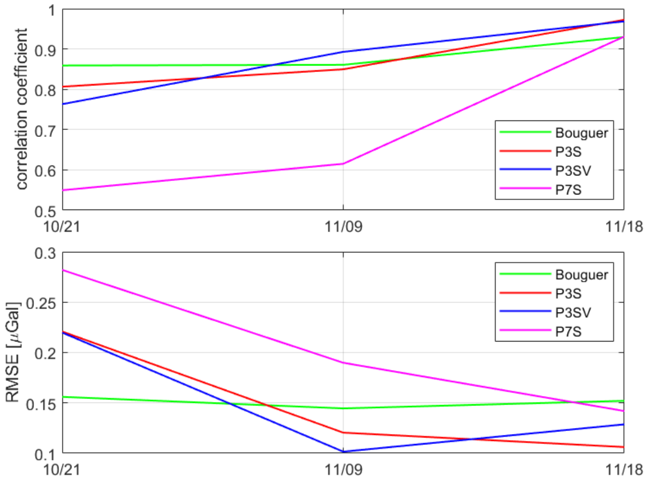

Figure 6.10.: Comparison of the performance of the Bouguer approach and three models that use soil moisture over three events. These models were determined without these three events.

Table 6.4.: Comparison of the two baselines and the three soil moisture-included gravity models being applied to the three 2022 rain events that were left out of the modeling. The final two columns show the results of the full models over these three events.

	leave-out validation		full model	
model	mean cc	RMSE [µGal]	mean cc	RMSE [µGal]
Bouguer	0.8831	0.1509		
OnlyP	0.5522	0.1988	0.7032	0.1665
P3S	0.8761	0.1427	0.9187	0.1182
P7S	0.6985	0.2127	0.9063	0.1169
P3SV	0.8748	0.1583	0.9171	0.1194

6.3. Application to new data

To obtain new rain events, data at the beginning of 2023 is used. The challenge of finding a period to compare the models is that all the input data, the different soil moisture sensors, precipitation, and gravity, need to be available without interruptions or high noise levels. Figure 6.11 shows the soil moisture measurements of Solidus #1 from the end of my previous analyses (2022, December 20th) until the point up to which the data are available at the time of this evaluation (May 2023). The station had already experienced some data gaps in the end of 2022. There is a big gap between 2022, December 24th, and 2023, January 27th. In this period, both probes only reported values of 0 % soil moisture. As Solidus #1 is part of all of my gravity models that use soil moisture, I am using rain events from February and March of 2023. One long event is visible on February 3rd and 4th. In March, there are at least three more peaks in soil moisture.

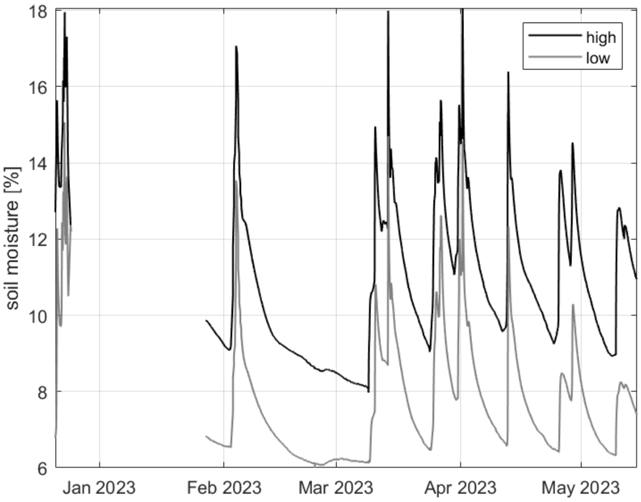

Figure 6.11.: Outlier filtered soil moisture measurements of both probes of Solidus #1 in 2023 and the end of 2022. You see the long data gap during most of January.

I am choosing the rain events by the daily amounts of precipitation (Figure 6.12). Four events are chosen (Table 6.5). The pre-processing of gravity is done as described in Chapter 4. The atmospheric factor of the previous adjustment of $f = -0.246\,\mu\text{Gal/hPa}$ is kept. The removal of oscillations is done with the Delayed Tide Adjustment, where the delay and amplitude of the correction are adjusted with the data from the whole time period between January 2022 and March 2023. The timeseries of gravity residuals (Figure 6.13) shows another restriction in choosing the period of comparison: There are big outliers in gravity at the beginning of February. The cause is the 7.8 $\mathrm{M_W}$ earthquake on February 6th that occurred near the border between Turkey and Syria (GEOFON GFZ Potsdam, 2023*c*). The spikes in the measurements are so big that even after the outlier filtering, strong peaks remain. This supports the understanding that earthquakes that are closer to the SG cause bigger and longer disturbances compared to earthquakes of the same magnitude that are further away. Because of this, the rain event at the beginning of February can only be analysed until February 5th.

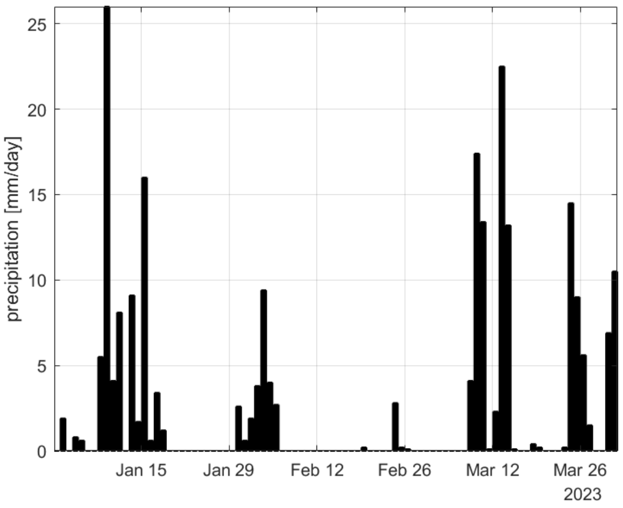

Figure 6.12.: Daily precipitation during the first three months of 2023

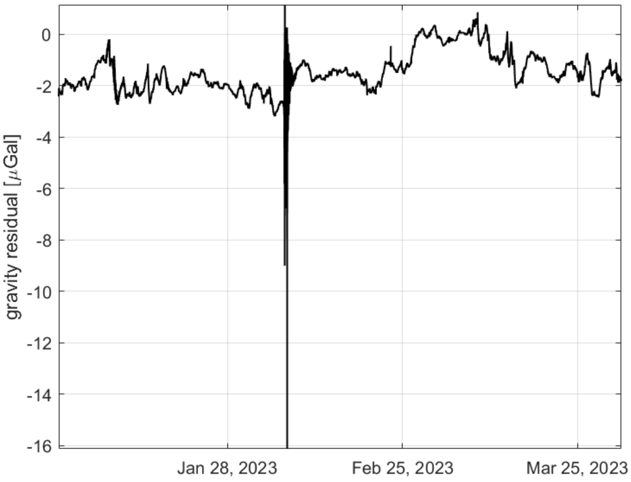

Figure 6.13.: Gravity residuals of G1-F60 over the first three months of 2023 after all the corrections, including a removal of linear trend and a final outlier filtering, have been carried out

Table 6.5.: Rain events in 2023 with soil moisture data from Solidus #1, probe 1

day	sm before [%]	sm max [%]	prec [mm/d]
February 3rd	6.53	13.54	9.4
March 9th	6.12	10.81	17.4
March 13th	8.68	14.7	22.5
March 24th	6.45	10.62	14.5

First, I am comparing the two baselines that only use precipitation as input. The first model is the Bouguer approach that was established by Schroth (2013). The second one is the OnlyP model that I have developed in Section 5.2.1. Figure 6.14 shows the responses of the two models for the February 3rd event. Both models have a much too small decrease in gravity when the rain happens. The fast rebound after the rain is only somewhat captured. The models have very similar results for the four events (Figure 6.15). The mean correlation coefficient for both is around 62 % (see Table 6.6). Even for the Bouguer approach, that is lower than the results of the eleven events of 2022, where they were in the high 70s. The reason for this is the bad results for event 4. That event includes a fast gravity increase of about 1 µGal during the third day (Figure 6.16). None of the available models are able to recreate this increase. It is most likely caused by a decrease in pressure ahead of a second rain on the afternoon of March 26th.

The problem of the modeled gravity decrease being too small remains for the three models that use soil moisture. None of them are able to recreate the course of gravity any better than the baseline model (Figure 6.17). This supports the notion of Section 6.2 that these data-driven models cannot really be applied to new events. Apart from event 2, the Bouguer baseline has better fits with the measurements (Figure 6.18). The results of the P3S model and the model with soil moisture-dependent precipitation parameters are almost the same. P7S performs slightly better than the other two models but still cannot reach the same accuracy as the baselines (see Table 6.6).

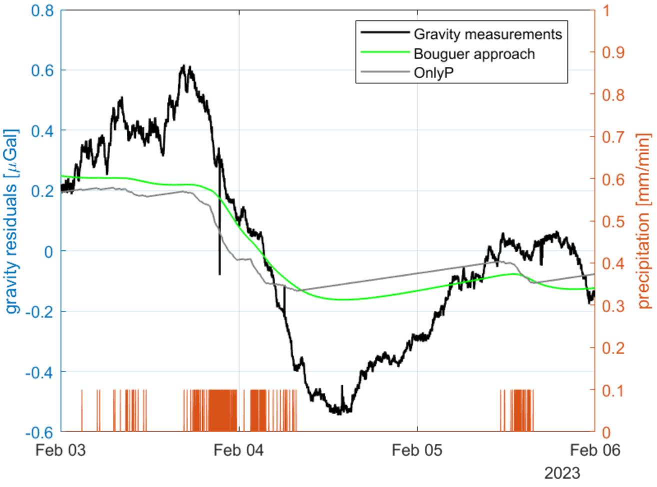

Figure 6.14.: Gravity responses of the two baseline models over the February 3rd event

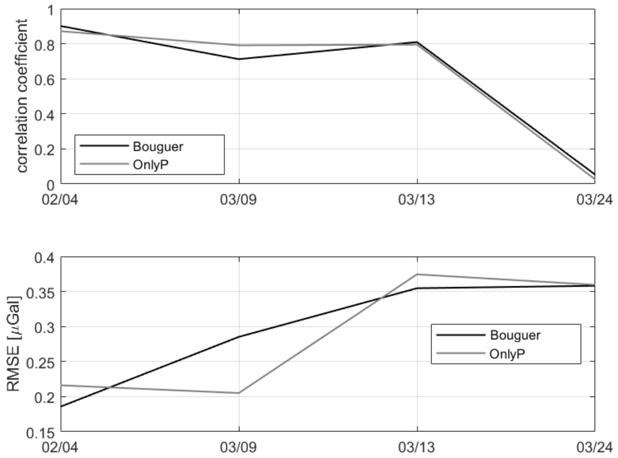

Figure 6.15.: The correlation coefficients and RMSE for the two baseline models over the four 2023 rain events

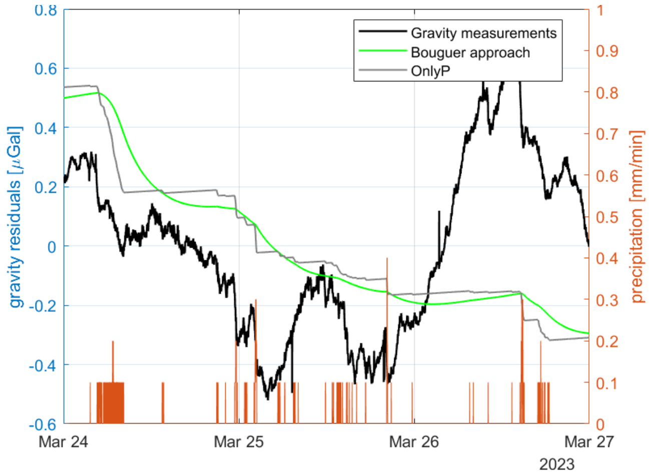

Figure 6.16.: Gravity responses of the two baseline models over the problematic March 24th event

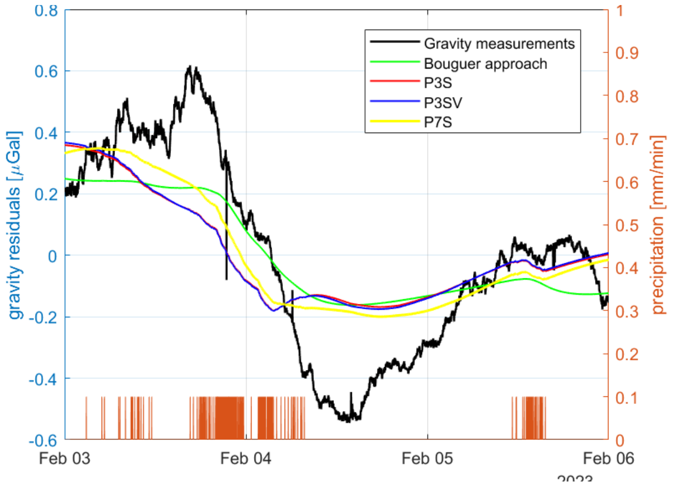

Figure 6.17.: Gravity responses of the three models that use soil moisture and the Bouguer baseline over the February 3rd event

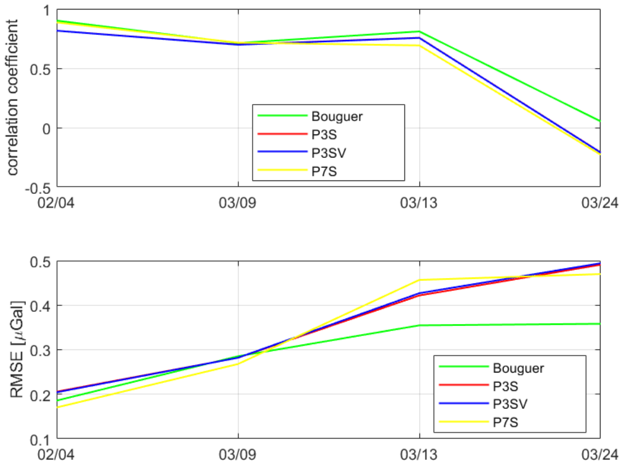

Figure 6.18.: Correlation coefficients and RMSE for the three models that are using soil moisture and the Bouguer baseline over the four 2023 rain events

Table 6.6.: Comparison of the two baselines and the three soil moisture included gravity models being applied to the four 2023 rain events

model	mean cc	RMSE [µGal]
Bouguer	0.6188	0.1834
OnlyP	0.6208	0.1805
P3S	0.5154	0.2217
P7S	0.5146	0.2231
P3SV	0.5161	0.2195

7. Conclusion

I have built three main data-driven gravity models that use the data from the precipitation gauge and different soil moisture sensors. The linearized least-squares algorithm is well suited for solving this non-linear problem. The model responses describe the gravity well for the epochs of the adjustment. The mean correlation coefficients over eleven rain events reach about 0.86 while the RMSE of P3S is about 26 % lower than that of the Bouguer baseline. Using a convolution with the latter half of a Gaussian kernel for both the precipitation measurements and the soil moisture data produces a good match with the measurements. Different kernels, especially for precipitation, are almost equally good, though.

Using the soil moisture data strongly improves the fit to the gravity series. These results cannot be recreated by reusing precipitation measurements. More soil moisture sensors improve the RMSE values, but these improvements become smaller and smaller. Especially after the third sensor, the results of adding a new sensor can almost be achieved by using another sensor for a second time. The correlation is also not improved after the third probe. Using data from soil moisture stations in a different location showed a strong benefit. This suggests that it could be advantageous to deploy the soil moisture stations more evenly in the forest as opposed to the more cluster-like layout of 2022. Using the new station, Solidus #8, which is almost directly above the SG, will probably yield a significant improvement.

Despite a good modeling performance, I did not find physical significance in the parameters. Some of the amplitudes always have a not-expected sign, and there is no trend in the parameters over the year when they are adjusted for each rain event individually. That is why the modeling with variable parameters is not able to significantly improve the results. It is likely that with different pre-processing and perhaps a different gravity model, one could come to a different conclusion.

Using the adjusted models on new events in 2023 without changing any parameters does not yield good results. The RMSE and correlation values of the models that use soil moisture lag significantly behind the Bouguer baseline. A sample size of four events is not large, but the trend is clear. These models are able to explain the events for which they were adjusted, but they do not fit these new events well. I hoped that by using so many events, the resulting models would hold general trends of the gravity response to a rain, but the analysis of the 2023 events did not confirm these hopes.

The leave-out validation confirmed that the data-driven models are usually slightly worse than the Bouguer baseline in application to new data. The correlation was higher than in the 2023 events, which is probably due to the nature of the rain events. The 2023 events have lower amounts of rain and are therefore more difficult to model. In both the leave-out validation and application to the 2023 data, the P7S performed the worst of all the models, so just adding more soil moisture data does not result in a more universally applicable model.

A number of steps in my procedure have probably contributed to the failure of the transmission of the models. The choice of the atmospheric factor of −0.246 µGal/hPa is unfortunate for this work as I am always analysing 72 hours of gravity data. For such a time window, higher frequencies are more important, and they need a factor closer to −0.3 µGal/hPa. Ideally, a frequency-dependent atmospheric factor should be used (Hinderer et al., 2007). Because of this, pressure changes with periods of hours to a few days leave big disturbing signals in the gravity residuals, and some events, like the one on March 24th, 2023, become completely unusable.

The choice of the observation equation is probably not ideal. The different timeseries were convoluted with different kernels and added together. What physically happens is that there is a certain overall moisture of the soil, which has a gravimetric impact. The precipitation and soil moisture measurements are all direct or indirect observations of this quantity. One could first compute these observations together, for example, by averaging the soil moisture values and then correcting gravity according to this average. A model that uses less adjusted parameters would probably be better applicable. For a smaller number of parameters, it is also more likely that these are physically meaningful, or you could force them to have the sign that you would expect.

I recommend that in a future development of such a model, one should also always check the interim models with outside epochs to make sure that one does not develop a model that is overfitted for the adjustment period.

The size of the adjustment epochs is probably also not sufficient. I applied the models to rain events in spring, while the models were adjusted for rain events mostly in autumn. If there are more events throughout the year, an adjustment is more likely to pick up on overall behaviour in the gravity response and does not just fit for a few events. If the pre-processing is done more cleanly, it is also imaginable to use continuous time periods for the adjustment.

Some results, like the mean correlation of more than 0.85 of the P3S model in the leave-out validation, show that the approach has promise. This gives hope that when the mentioned improvements are carried out and the data are of high enough quality, it is possible that the gravity residuals can lead to new insights into uplift or subsidence of the ground, into geological processes, or into ground water levels. As I have only analysed short time windows of gravity and have always adjusted an offset for these events, there are no insights possible into ground water level. With a continuous investigation period, this could be possible and could become one of the most important applications of an SG. An instrument, for example, in southern Europe or California, could find many insights into an issue that is rapidly becoming a central one of the modern world.

Bibliography

BFO (2023), ‘Virtual tour of the Anton mine’, `https://www.black-forest-observatory.de/95.php`. Accessed: 2023-09-22.

Black Forest Observatory (BFO) (1971), ‘Black Forest Observatory Data. GFZ Data Series’, `https://doi.org/10.5880/BFO`. Accessed: 2023-10-06.

Craymer, M. (1998), The least squares spectrum, its inverse transform and autocorrelation function, theory and some applications in geodesy, PhD thesis, University of Toronto.

Deutscher Wetterdienst (2023), ‘Starkregen’, `https://www.dwd.de/DE/service/lexikon/begriffe/S/Starkregen.html`. Accessed: 2023-08-29.

Faller, J. E. (1967), ‘Precision Measurement of the Acceleration of Gravity’, *Science* **158**(3797), 60–67.

Forbriger, T. and Heck, A. (2018), Frequency response of the superconducting gravimeter SG056, *in* A. Heck, K. Seitz, T. Grombein, M. Mayer, J. Stövhase, H. Sumaya, M. Wampach, M. Westerhaus, L. Dalheimer and P. Senger, eds, ‘Schw(Ehre), wem (Schw)Ehre gebührt’, KIT Scientific Publishing, pp. 57–67.

Forster, O. (2016), *Analysis 1. Differential- und Integralrechnung einer Veränderlichen*, Springer Spektrum, Wiesbaden, Germany.

GEOFON GFZ Potsdam (2023*a*), `http://geofon.gfz-potsdam.de/eqinfo/event.php?id=gfz2022skgc`. Accessed: 2023-10-03.

GEOFON GFZ Potsdam (2023*b*), `http://geofon.gfz-potsdam.de/eqinfo/list.php?datemin=2022-01-01&datemax=2022-12-31&latmax=&lonmin=&lonmax=&latmin=&magmin=7&fmt=html&nmax=`. Accessed: 2023-10-03.

GEOFON GFZ Potsdam (2023*c*), `http://geofon.gfz-potsdam.de/eqinfo/event.php?id=gfz2023cnwr`. Accessed: 2023-10-12.

Goodkind, J. (1999), ‘The superconducting gravimeter’, *Review of Scientific Instruments* **70**.

Großmann, W. (1969), *Grundzüge der Ausgleichungsrechnung*, Springer-Verlag Berlin Heidelberg GmbH, Berlin/Heidelberg, Germany.

Hansen, P. (2000), The L-curve and its use in the numerical treatment of inverse problems, *in* ‘InviteComputational Inverse Problems in Electrocardiology’, WIT Press. InviteComputational Inverse Problems in Electrocardiology ; Conference date: 01-01-2000.

Harnisch, G. and Harnisch, M. (2006), ‘Hydrological influences in long gravimetric data series’, *Journal of Geodynamics* **41**, 276–287.

Hinderer, J., Crossley, D. and Warburton, R. (2007), *Gravimetric Methods – Superconducting Gravity Meters*, Vol. 3, Elsevier B.V., pp. 65–122.

Krause, P., Naujoks, M., Fink, M. and Kroner, C. (2009), 'The impact of soil moisture changes on gravity residuals obtained with a superconducting gravimeter', *Journal of Hydrology* **373**(1), 151–163.

Kroner, C., Dierks, O., Neumeyer, J. and Wilmes, H. (2005), 'Analysis of observations with dual sensor superconducting gravimeters', *Physics of the Earth and Planetary Interiors* **153**, 210–219.

Lacoste, L. J. B. (1934), 'A New Type Long Period Vertical Seismograph', *Journal of Applied Physics* **5**, 178–180.

Landesamt für Geologie, Rohstoffe und Bergbau (2022), 'Erdbeben bei Sierentz, Dep. Haut-Rhin, F', `https://erdbeben.led-bw.de/erdbeben/220910_1557`. Accessed: 2023-09-14.

Lomb, N. (1976), 'Least-Squares Frequency Analysis of Unequally Spaced Data', *Astrophysics and Space Science* **39**(2), 447–462.

Marquardt, D. (1963), 'An algorithm for least-squares estimation of nonlinear parameters', *Journal of the Society for Industrial and Applied Mathematics* **11**(2), 431–441.

Pearson, K. (1896), 'VII. Mathematical contributions to the theory of evolution.—III. Regression, heredity, and panmixia', *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character* **187**, 253–318.

Peiniger, M. and Piel, H. (1985), 'A Superconducting Nb3Sn Coated Multicell Accelerating Cavity', *IEEE Transactions on Nuclear Science* **32**(5), 3610–3612.

Rosat, S. and Hinderer, J. (2011), 'Noise levels of superconducting gravimeters: Updated comparison and time stability', *Bulletin of the Seismological Society of America* **101**, 1233–1241.

Schroth, E. (2013), Analyse von Gezeitenregistrierungen des Supraleitenden Gravimeters SG-056, Diplomarbeit, Karlsruhe Institute of Technology.

Senthilnathan, S. (2019), 'Usefulness of correlation analysis', *SSRN Electronic Journal* .

SMBFO (2023), 'Soil moisture activity at the Black Forest Observatory', `http://smbfo.gis.uni-stuttgart.de/`. Accessed: 2023-08-20.

Tinkham, M. and Emery, V. (1996), 'Introduction to superconductivity', *Physics Today* **49**, 74–.

Van Camp, M., De Viron, O., Watlet, A., Meurers, B., Francis, O. and Caudron, C. (2017), 'Geophysics from terrestrial time-variable gravity measurements', *Reviews of Geophysics* **55**.

Wenzel, H. (1998), *Earth tide data processing package ETERNA 3.30: the nanogal software*, Ducarme, B., pp. 487–494.

Westerhaus, M., Duffner, P., Forbriger, T., Heck, B., Widmer-Schnidrig, R. and Zürn, W. (2018), Das Geowissenschaftliche Gemeinschaftsobservatorium Schiltach (BFO), *in* K. Seitz, ed., 'Festschrift zur 150-Jahr-Feier des Geodätischen Instituts (1868 - 2018)', Geodätisches Institut, pp. 255–274.

Wikipedia contributors (2023*a*), 'Hyperparameter optimization — Wikipedia, the free encyclopedia', `https://en.wikipedia.org/w/index.php?title=Hyperparameter_optimization&oldid=1177947506`. [Online; accessed 3-October-2023].

Wikipedia contributors (2023*b*), 'Ridge regression — Wikipedia, the free encyclopedia', `https://en.wikipedia.org/w/index.php?title=Ridge_regression&oldid=1172832572`. [Online; accessed 3-October-2023].

Willmott, C., Ackleson, S., Davis, R., Feddema, J., Klink, K., Legates, D., O'Donnell, J. and Rowe, C. (1985), 'Statistics for the Evaluation and Comparison of Models', *Journal of Geophysical Research* .

Zürn, W. (2014), 'Listening to the Earth - The Schiltach Observatory BFO', *50 Years Geophysical Institute Karlsruhe, 1964 to 2014 - Expectations and Surprises* pp. 285–310.

Zürn, W. and Duffner, P. (2018), Spezielle Störung in Registrierungen des Supraleitenden Gravimeters SG-056 am BFO, *in* A. Heck, K. Seitz, T. Grombein, M. Mayer, J. Stövhase, H. Sumaya, M. Wampach, M. Westerhaus, L. Dalheimer and P. Senger, eds, 'Schw(Ehre), wem (Schw)Ehre gebührt', KIT Scientific Publishing, pp. 303–308.

A. Used methods of adjustment theory

The central goals of this work are all connected to estimating numeric parameters, e.g., what is the best value for the atmospheric factor in the correction of this influence? What is the amplitude and decay time of a best-fit between a precipitation and gravity timeseries? Sometimes a lot of parameters that are connected have to be estimated simultaneously; sometimes it is only one parameter. Three different approaches will be used for this estimation.

A.1. Grid search

The easiest way to search for parameters in a parametric optimization problem with no more than a handful of parameters is the grid search. You try different values of a set of parameters and choose the best values of the parameters by comparison of a certain metric (Wikipedia contributors, 2023*a*).
This method only works when the range of values for the parameters is roughly known and when there are only one or two parameters to be estimated simultaneously. With more dimensions, grid search quickly becomes very slow. The task is usually to estimate a float number and not a number from a discrete set of possibilities (except for the delay in Section 4.4). This is, in principle, an issue for grid search, as it always tries out a finite number of solutions. A solution for this is an iterative approach; for example, if one parameter is searched between 0 and 10, then first of all every integer (0, 1, 2, …, 10) is checked. If the two best results were 5 and 6, then the next iteration checks 5.0, 5.1, …, 6, and so on.
The advantages of grid search are that it is easy to understand, and (unlike the iterative least-squares adjustment), it is very unlikely that within an adjustment procedure one finds a local minimum that is not the global minimum, as basically all possibilities are tried. Grid searches will be applied in most situations when no more than two parameters are searched.

There are a wide range of metrics that are used to determine the best values of the parameters. Of course, they will also be used in the other methods to assess the quality of the results that were found.
The most obvious metric is the **RMSE** (root-mean-square-error). For two timeseries $(a_i)_{i=1}^n$, $(b_i)_{i=1}^n$ the RMSE is (Willmott et al., 1985)

$$\mathrm{RMSE}(a,b) = \sqrt{\frac{\sum_{i=1}^n (a_i - b_i)^2}{n}}. \tag{A.1}$$

This is the difference between the timeseries and, for example, gives an idea of the magnitude of how far the fitted model is away from the data. Sometimes one sets b as the mean value of a. In this case, RMSE(a,$\bar{a}$) is equivalent to the standard deviation of a, which in some situations is desirable to minimise. One disadvantage of the RMSE is that it can sometimes be misleading. For example, if one timeseries shows the same trends as the other but is consistently higher, this will result in a big RMSE despite both variables being very similar. That is why it is very important to also look at the correlation.

The **correlation** tells about the statistical dependence of two timeseries (Senthilnathan, 2019). It answers the questions of how much information a holds about b. In this work, the quantity that will be analysed is the correlation coefficient, as it is not dependent on the magnitudes of the timeseries. It is calculated by (Pearson, 1896)

$$\rho(a,b) = \frac{\mathrm{cov}(a,b)}{\sigma(a)\sigma(b)} = \frac{1}{n-1}\frac{\sum_{i=1}^{n}(a_i-\bar{a})(b_i-\bar{b})}{\sigma(a)\sigma(b)}, \tag{A.2}$$

where $\mathrm{cov}(a,b)$ is the covariance between a and b, $\sigma(a)$ is the standard deviation of a, and $\bar{a}$ is its mean value. This always gives a value between -1 and 1, where 0 means that both timeseries are statistically independent and 1 means they are completely linearly dependent. A negative value means there is an anticorrelation, meaning a higher value in a more often than not coincides with a lower value in b. The correlation coefficient is therefore very straightforward to interpret. Together with the RMSE, these two metrics usually give a very good idea of which timeseries is a good fit and which ones are not.

One analysis that can be a good supplementary assessment is the **spectral analysis**. A signal can be decomposed into an infinite series of sinusoid oscillations with different frequencies. A spectral analysis finds the amplitudes of the different frequencies and therefore allows a conclusion about which periods mainly make up the signal. The usual method of choice here is the Fourier transform, but this requires equally spaced points. As this is not always guaranteed in the available data, for example, in the soil moisture measurements, the Lomb-Scargle periodogram is used instead. As this will be used for all spectral analysis, the comparability of the different analyses is ensured. The Lomb-Scargle periodogram is defined as (Lomb, 1976)

$$P_{LS}(f) = \frac{1}{2\sigma^2}\left(\frac{\left[\sum_{k=1}^{N}(a_k-\bar{a})\cos(2\pi f(t_k-\tau))\right]^2}{\sum_{k=1}^{N}\cos^2(2\pi f(t_k-\tau))} + \frac{\left[\sum_{k=1}^{N}(a_k-\bar{a})\sin(2\pi f(t_k-\tau))\right]^2}{\sum_{k=1}^{N}\sin^2(2\pi f(t_k-\tau))}\right). \tag{A.3}$$

The time t_k is the observation time of a_k, σ is the standard deviation, and the delay τ has to satisfy

$$\tan(2(2\pi f)\tau) = \frac{\sum_{k=1}^{N}\sin(2(2\pi f)t_k)}{\sum_{k=1}^{N}\cos(2(2\pi f)t_k)}. \tag{A.4}$$

The Lomb-Scargle periodogram can be used to compare whether two timeseries show similar peaks. Sometimes it is also desired that certain peaks disappear or become lower, so one might use the periodogram and choose the value that has the lowest or fewest peaks visible in a certain section.

A.2. Least-squares adjustment

The following section is based on Großmann (1969).

The most important tool of geodesists when estimating parameters is, of course, the least-squares adjustment. It is used when the unknown parameters x are believed to have a linear relationship with the data y:

$$y = Ax + e, \tag{A.5}$$

where y, x, and e are vectors, while A can be a matrix. The vector e holds the inconsistencies. Usually (and reasonably) there are more measurements (let the number of measurements be called m) than unknown parameters (let the number of parameters be called n), and therefore the linear equation system $Ax = y$ cannot be solved without the inconsistencies e. The aim is to find a set of parameters that minimizes $\|e\|$ or equivalently $e^T e = \sum_{i=1}^{n} e_i^2$, hence the name „least-squares“. In a more general approach, the observations can be weighted with a quadratic matrix P, then the weighted least-squares $e^T Pe$ is minimised. If the weight matrix P is diagonal, that means that the values on the diagonal represent a level of confidence in their respective observation (the value p_{ii} corresponds to the weight of the observation y_i). If there are also non-zero values away from the diagonal, this models correlations between the observations. For example, if y is a vector of gravity measurements, one might assume that a too high value in the first epoch can mean that the second measurement will be too high as well, maybe due to drift. However, such modeling is beyond the scope of this work, and drifts are usually eliminated before any parameter estimation, so P will always remain a diagonal matrix.
The estimation is done by the formula

$$\hat{x} = (A^T PA)^{-1} A^T Py. \tag{A.6}$$

This estimation ensures that $e^T Pe$ becomes minimal. It is also an unbiased estimator of x, meaning the expected value of $\hat{x}$ is x. If one assumes that the inconsistencies e are random mistakes in the observations that are distributed with a normal distribution around 0 (values closer to zero have a higher probability) and P^{-1} as covariance matrix, the estimator actually finds the parameters that maximise the probability density function.

The result will satisfy the orthogonality condition $A^T P\hat{e} = 0$, where $\hat{e} = y - \hat{y} = y - A\hat{x}$. This condition can be used to make sure that the computations were done correctly.

The matrix $A^T PA$ is a quadratic matrix with n rows, and as in this work, only a handful of parameters are estimated at the same time, the inversion and so the adjustment can be carried out very quickly.

This method basically always finds the globally best set of parameters. Unfortunately, it cannot be used most of the time, as real-world problems very rarely show a linear relationship between the parameters and the data. Usually this relationship is more complicated; for example, an exponential decay will be assumed between the measurements of the rain gauge and the gravity data. This cannot be written in the form of A.5, and therefore a different approach is needed. This is the iterative least-squares adjustment.

A.3. Non-linear least-squares adjustment

The following section is based on Marquardt (1963).

For this method, a non-linear relationship between the parameters and the observations is assumed:

$$y = f(x). \tag{A.7}$$

If a good approximate value x_0 is available, the difference between the function values at x_0 and x almost solely depends on the linear part of f and so on the partial derivatives of f:

$$f(x) - f(x_0) \approx \begin{pmatrix} \frac{\partial f}{\partial x_1}(x) & \cdots & \frac{\partial f}{\partial x_n} \end{pmatrix}_{x=x_0} \begin{pmatrix} \Delta x_1 \\ \cdots \\ \Delta x_n \end{pmatrix}, \tag{A.8}$$

where $\Delta x := x - x_0$. The error of the mistake in the aprroximation is of the magnitude $O(\Delta x^2)$, which underlines the importance of good approximate values.
The equation of A.8 is now very similar to the case of the linear relationship. The unknowns are now Δx, which are the improvements of the approximate value x_0. The measurements are now the reduced observations: $\Delta y := f(x) - f(x_0) = y - f(x_0)$. Finally, the design matrix A is now the partial derivatives of f:

$$A := \begin{pmatrix} \frac{\partial f}{\partial x_1}(x) & \cdots & \frac{\partial f}{\partial x_n} \end{pmatrix}_{x=x_0}. \tag{A.9}$$

With these designations, (A.8) can be written as

$$\Delta y = A \Delta x. \tag{A.10}$$

So, Δx can be estimated with a normal least-squares adjustment: $\widehat{\Delta x} = (A^T P A)^{-1} A^T P \Delta y$. The resulting Δx is, as mentioned, the improvement of the approximate values x_0: $\hat{x} = x_0 + \widehat{\Delta x}$. As the relationship f between parameters and observation was linearized, one such adjustment is usually not enough to find a good estimation of x. Rather, the found $\hat{x}$ is believed or hoped to be a better approximation of x than x_0. So $\hat{x}$ can be used as the new approximate value for another adjustment. This therefore becomes an iterative approach that continues until the improvements $\widehat{\Delta x}$ become smaller than some small value, at which point the iteration is stopped and one hopes that one has arrived at a good result for $\hat{x}$. To try to verify this, the main check $A^T P \hat{e} = 0$? (see Figure A.1) is carried out. Figure A.1 shows a flowchart that summarizes the iterative computations.

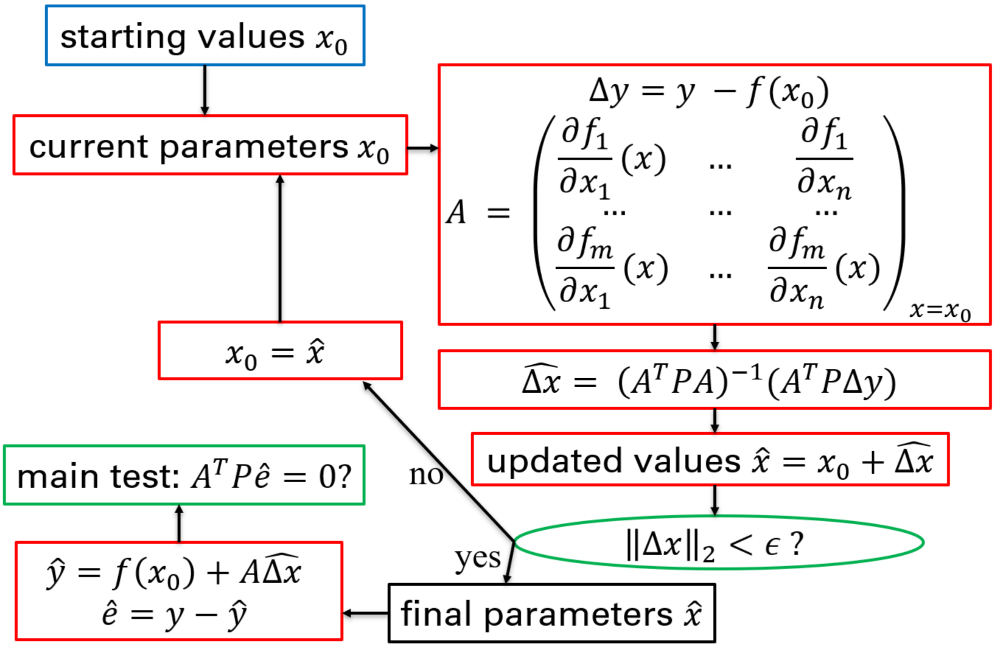

Figure A.1.: The sequence of computations for the non-linear least-squares adjustment (own diagram)

There is one important addition to this algorithm that is often used in this thesis: the iterative least-squares adjustment is not a very stable procedure, meaning it is not unlikely that the iteration heads off in a completely wrong direction and gets stuck there with unreasonable results. This is especially the case because, in most situations, it is difficult to find good approximate values for all parameters. One also does not want to try and execute the iteration many times in the hope that, at some point, good results are found. Instead, the adjustment is given some a priori information. So far, y was always a vector that just included all the measurements. Now there will be additional rows that include the starting values. For example, the $m+1$st row of the observation equation A.7 might read:

$$a_{1,0} = a_1. \tag{A.11}$$

Here, a_1 is an amplitude, one of the parameters that is estimated, and $a_{1,0}$ is its approximate value right at the beginning of the iteration. This value represents a somewhat reasonable result, and the line prevents the iteration from going completely off with a value for a_1 that is far too high or too low. This artificial observation also needs a weight in P so that the strain is neither too strong so that the iteration never moves away from $a_{1,0}$ nor too weak, so there is the danger that the iteration goes off to an unrealistic result for a_1. This is a common method, for example, in the adjustments of photogrammetry when the attitudes of a camera are estimated alongside the measured points from a measurement campaign.

This method can also be found in adjustment theory, though. The usual approach for regularisation, which means stabilising a badly conditioned system of equations, i.e., A^TPA is close to being singular, is the Tikhonov regularisation. In its generalized form, where one wants to minimise the distance between the x and x_0, weighted by the matrix Q, in addition to solving the linear system of equation $Ax = b$ in a least-squares sense with weights P, the Tikhonov regularisation reads as follows (Wikipedia contributors, 2023*b*):

$$x^* = (A^TPA + Q)^{-1}(A^TPb + Qx_0). \tag{A.12}$$

This is the case of a linear adjustment, so for a comparison, the corresponding values have to be exchanged: x^* is now $\Delta x^* = x^* - x_k$ (in this work the estimation is indicated by a hat instead of a star); here, x_k means the estimated parameters after k iterations, to be able to distinguish this from the start values x_0 at the very beginning of the iteration. b becomes Δb (it was so far called Δy), and A now includes the partial derivatives of the observation equation. In the linear case, $\|x - x_0\|_Q$ was intended to be minimised. In the non-linear case, the update Δx is the unknown, and therefore $\|x - x_0\|_Q = \|(x - x_k) - (x_0 - x_k)\|_Q$ is now the minimised quantity. This means that the a-priori information is $x_0 - x_k$. The formula (A.12) in a non-linear approach thus reads:

$$\Delta x^* = (A^TPA + Q)^{-1}(A^TP\Delta b + Q(x_0 - x_k)). \tag{A.13}$$

Why is this equivalent to the formulas that were presented earlier? The goal is now to rewrite the update step of the algorithm in Figure A.1 to (A.13). In the remainder of the chapter, all the variables from the notation of the Wikipedia article will be underlined, while the previous designations will not be.

The observation equation (A.7) in this work consists of two parts. First, there are the observations $\tilde{y}$. The final few lines of the equations are the parameters with their respective starting values:

$$\begin{aligned} y_{m+1} &: p_{1,0} = p_1, \\ &\dots \\ y_{m+n} &: p_{n,0} = p_n. \end{aligned} \tag{A.14}$$

p_1 describes the first unknown parameter, $p_{1,0}$ is its starting value, and so on. These equations have to be partially derived after the parameters, which is trivial: the $m+1$st line of A has a 1 in the first column and zeros everywhere else, the $m+2$nd line has the one in the second column, etc. That means A will have the partial derivatives of the observations in the first m lines, as if there were no stabilisation. This matrix will now be called $\tilde{A}$. The final n rows of A are simply a unit matrix:

$$A = \begin{pmatrix} \tilde{A} \\ I_n \end{pmatrix}. \tag{A.15}$$

The weight matrix P includes the normal weights of the observations in the matrix $\tilde{P}$ and then the weights for the stabilisation, which I call Q:

$$P = \begin{pmatrix} \tilde{P} & 0_{m\times n} \\ 0_{n\times m} & Q \end{pmatrix}. \tag{A.16}$$

The zeros away from the diagonal appear because, as mentioned earlier, the weight matrices are always diagonal.
Finally, the reduced observation $\Delta y = y - f(x_k)$ needs to be described. The first m lines are the usual difference between the observations (designated $\tilde{y}$) and the model of the so far estimated parameters ($\tilde{f}(x_k)$). The final n lines are because of Equation (A.14) the differences between the starting values x_0 and momentary parameters x_k. This means:

$$\Delta y = \begin{pmatrix} \tilde{y} - \tilde{f}(x_k) \\ x_0 - x_k \end{pmatrix}. \tag{A.17}$$

With these designations, the update step (see Figure A.1) reads as

$$\begin{aligned} \hat{\Delta x} &= (A^T P A)^{-1} A^T P \Delta y \\ &= \left(\begin{pmatrix} \tilde{A}^T & I_n \end{pmatrix} \begin{pmatrix} \tilde{P} & 0_{m\times n} \\ 0_{n\times m} & Q \end{pmatrix} \begin{pmatrix} \tilde{A} \\ I_n \end{pmatrix} \right)^{-1} \begin{pmatrix} \tilde{A}^T & I_n \end{pmatrix} \begin{pmatrix} \tilde{P} & 0_{m\times n} \\ 0_{n\times m} & Q \end{pmatrix} \begin{pmatrix} \tilde{y} - \tilde{f}(x_k) \\ x_0 - x_k \end{pmatrix} \\ &= (\tilde{A}^T \tilde{P} \tilde{A} + Q)^{-1} (\tilde{A}^T \tilde{P} (\tilde{y} - \tilde{f}(x_k)) + Q(x_0 - x_k)). \end{aligned}$$

Now this just needs to be brought into the same designation as Equation (A.13). $\tilde{A}$ is just the normal design matrix, so it is $\underline{A}$. $\tilde{P}$ is the same as $\underline{P}$ because it is the weights of the observations; $\tilde{y} - \tilde{f}(x_k)$ is the reduced observation without the stabilisation, so it is $\underline{\Delta b}$. So the final result is:

$$\underline{\Delta x^*} = (\underline{A}^T \underline{P}\,\underline{A} + \underline{Q})^{-1} (\underline{A}\,\underline{P}\,\underline{\Delta b} + \underline{Q}(\underline{x_0} - \underline{x_k})). \tag{A.18}$$

This is exactly what was to be shown (Equation (A.13)). So the approach used is equivalent to a generalized Tikhonov regularization in the linearised case.

B. Proof of Equation 4.3

In this section, it will be proven that you can always find real-valued parameters s_i and c_i so that

$$\sum_{i=1}^{4} (a_i \sin(f_i t + \phi_i)) + c_0 + e(t) = \sum_{i=1}^{4} (s_i \sin(f_i t) + c_i \cos(f_i t)) + c_0 + e(t). \tag{B.1}$$

holds. This equation was used in Section 4.4 for the Least-Squares Spectral Analysis to eliminate residual oscillations in the gravity data.
First of all, I will deduce what the new amplitudes s_i and c_i must be, and then I will show why this choice indeed satisfies (B.1):
I will show that for an arbitrary index i, with given values of a_i, ϕ_i, and f_i, there exist real-valued amplitudes s_i and c_i such that

$$a_i \sin(f_i t + \phi_i) = s_i \sin(f_i t) + c_i \cos(f_i t) \tag{B.2}$$

holds. First, certain values for t are tried so that some components of this equation disappear. For $t = 0$, this yields:

$$a_i \sin(\phi_i) \overset{(B.2)}{=} s_i \sin(0) + c_i \cos(0) = c_i.$$

This is the constraint for c_i. The second constraint is found by choosing $t = \pi/(2f_i)$:

$$a_i \sin(\pi/2 + \phi_i) = a_i \cos(\phi_i) \overset{(B.2)}{=} s_i \sin(\pi/2) + c_i \cos(\pi/2) = s_i.$$

These are the searched-for choices for the new amplitudes. They will now be inserted into the right side of (B.2), and the identity will be shown:

$$\begin{aligned} r.s. &= s_i \sin(f_i t) + c_i \cos(f_i t) \\ &= a_i \cos(\phi_i) \sin(f_i t) + a_i \sin(\phi_i) \cos(f_i t) \\ &= a_i (\cos(\phi_i) \sin(f_i t) + \sin(\phi_i) \cos(f_i t)) \\ &= a_i \sin(f_i t + \phi_i) = l.s. \end{aligned}$$

The penultimate equality results from the first addition theorem (Forster, 2016). So the equality was shown, and the adjustment can be done linearly.

C. Model based on the correlation coefficient

A model can also be developed with a focus on correlation. In this section, I will briefly present the analyses that I did with this goal. The steps are analogous to the model building in Chapter 5.

C.1. Data-driven model with just precipitation

A simple exponential decay will be taken into consideration:

$$\Delta g_i = P_0 \sum_{j=i-N}^{i} r_j \exp\left(-\frac{i-j}{\tau}\right). \tag{C.1}$$

To get a first idea of possible results, a grid search is carried out, testing values for τ between 2 and 2002 hours. The different values are tested successively, and a best-fitting P_0 is adjusted by a least-squares adjustment. The observation epochs are the eleven events with three days of gravity data each. The observation model is that of (C.1) only with an added offset c_k for the different events, as the observation epochs are not continuous. Figure C.1 shows the mean RMSE over the eleven events for the different choices of τ.

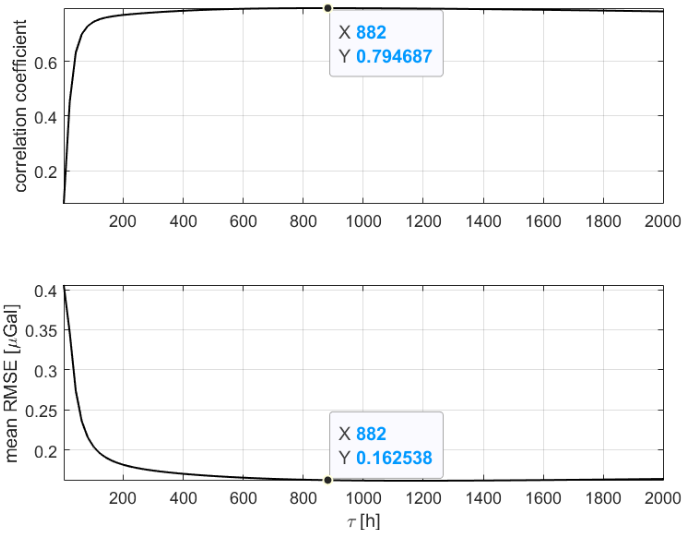

Figure C.1.: Result of the grid search for the decay time, only using precipitation data, the best time for correlation coefficient is highlighted

The RMSE and the mean correlation coefficient (cc) show the best values for very high decay times, around 800 to 1200 h. The best mean correlation coefficient reaches about

0.7947.
A latter half of a Gaussian kernel is also tried. It is computed by

$$k_G(d) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{d^2}{2\sigma^2}\right). \tag{C.2}$$

A grid search with this kernel finds a slightly higher maximum correlation of 0.8082 at a similar decay time. Going further in this direction, the exponent of d could be increased further. Another grid search is done with the following Degree-Three kernel:

$$k_T(d) = \exp\left(-\frac{d^3}{3\sigma^3}\right). \tag{C.3}$$

The prefactor is left out as it would only influence the value of the found amplitudes, so it is irrelevant. Figure C.2 shows the results with a slightly improved maximum correlation at 0.8156.

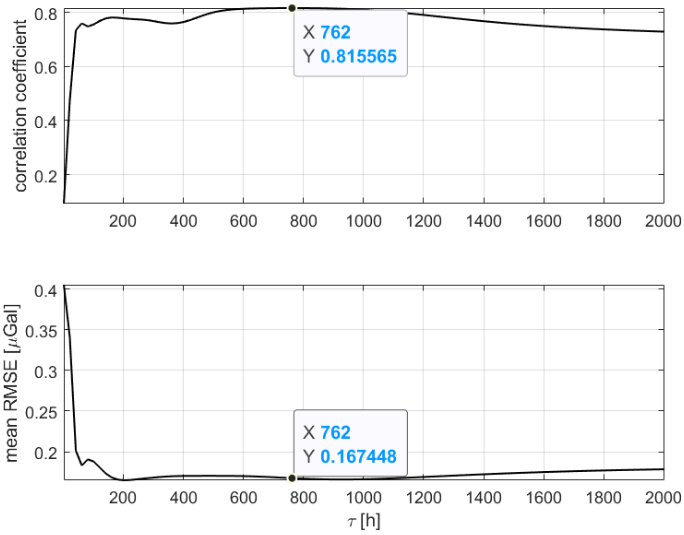

Figure C.2.: Result of the grid search for the decay time, only using precipitation data, with the Degree-Three kernel

Any further increase of the exponent of d only marginally improves the correlation coefficient, while the mean RMSE gets slightly worse. So after these tests, it remains to be seen whether the exponential decay, the Gaussian kernel, or the Degree-Three kernel is ideal for precipitation. For the rest of the section, the exponential kernel is used.
An LLS adjustment is carried out with the results of the grid search as starting values. The iteration is done as described in Section A.3.
In the iteration, τ rises further until it converges at 926 h. The iteration finds a similar value for the amplitude in $\hat{a} = -0.032\,\mu\text{Gal/mm}$. Figure C.3 shows the course of the iteration over the October 2nd event; every twentieth iteration is displayed. At the end of the iteration, after 70 loops, the RMSE stands at 0.161 85 µGal. The magnitude of improvements also constantly decreases, as intended. It is simply the case that, with only precipitation data, the iteration cannot find any better results.

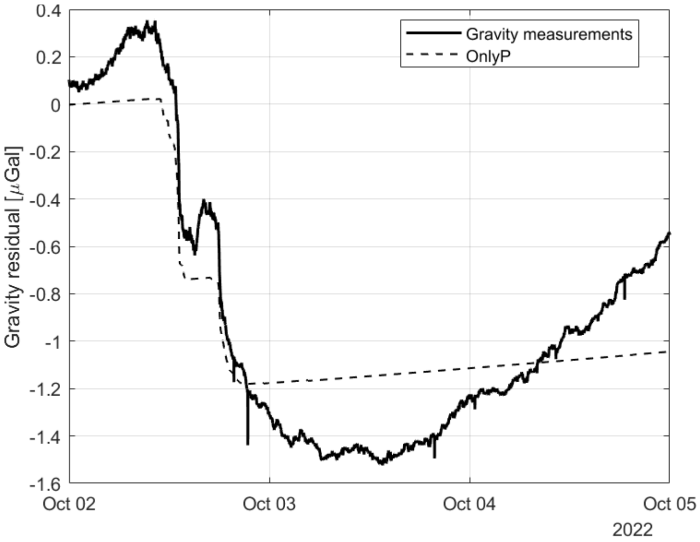

Figure C.3.: The resulting timeseries of the LLS adjustment with only precipitation used with the Degree-Three kernel

The mean correlation coefficient of the events has very slightly decreased to 0.7938 from the 0.7947 of the grid search. The reason for this is that least-squares adjustment always means optimizing parameters with regard to the RMSE. That is not equivalent to the correlation coefficient, and in this case, the grid search also found lower RMSE values for decay times around 1000 h (see Figure C.1). Still, the mean correlation is higher than in the Bouguer approach. Figure C.4 compares the correlation coefficients between these two corrections. The newly adjusted model performs equally or better on all events except 7 and 10.

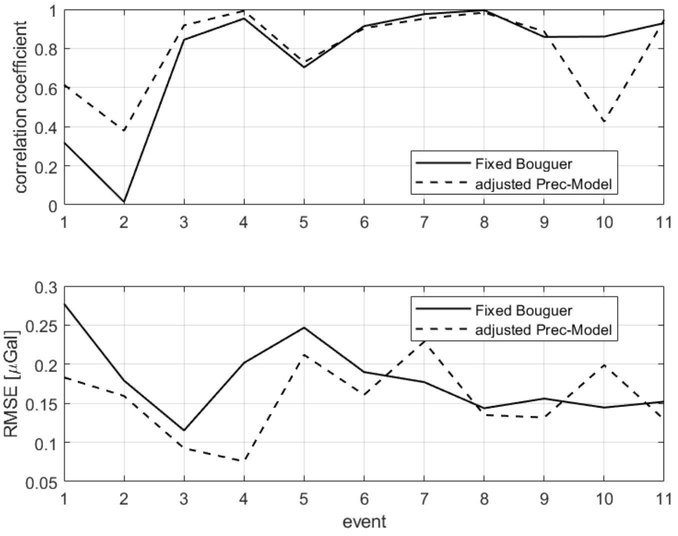

Figure C.4.: Comparison of the correlation coefficients and RMSE over the events of the Bouguer approach and the new precipitation correction resulting from the LLS-adjustment

To improve on this, in the next chapter, the soil moisture data will be used.

C.2. Data-driven model with one soil moisture sensor

The new observation equation for a correction with one soil moisture probe reads:

$$\Delta g_i = P_0 \sum_{j=i-N_1}^{i} r_j \exp\left(-\frac{i-j}{\tau_p}\right) + P_1 \sum_{j=i-N_2}^{i} s_j \exp\left(-\frac{(i-j)^2}{2\tau^2}\right) + c_k. \tag{C.4}$$

Now s_j is the soil moisture reading at time j.
Just like in the last section, I will do a grid search to get an idea of what kind of τ produces good results. This time, $\tau_p = 200, 240, \ldots, 1000\,\text{h}$ and $\tau = 6, 7, \ldots, 27\,\text{h}$ are tried, and P_0, P_1, and c_k are adjusted with a least-squares adjustment.
The results of this grid search again heavily favour long decay periods for τ_p, where the best results all occurred at 1000 hours. For the soil moisture, the best values for RMSE and mean correlation coefficient appear for periods between 10 and 16 hours, with the highest mean correlation coefficient reaching 0.8316. Figure C.5 shows all the computed values of the mean correlation coefficient. With longer times for τ_p, the mean correlation coefficient rises up to about 0.84 at 2000 h, but the RMSE value worsens, so only the best result until 1000 h is used.

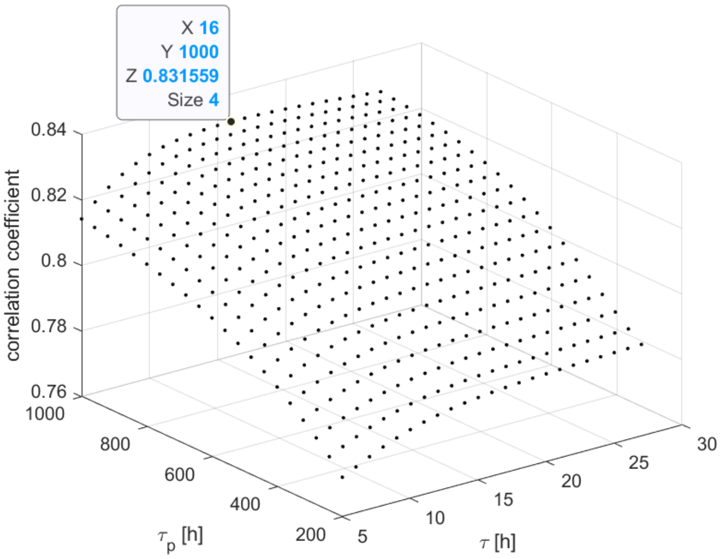

Figure C.5.: CC values of the grid search using precipitation and one soil moisture probe

Another comparison of the kernels again shows slight advantages for the other presented alternatives. Using the Gaussian kernel for precipitation in the same analysis yields a maximum correlation coefficient of 0.8569 while the Degree Three kernel reaches 0.8664. The full results of the latter grid search are shown in Figure C.6. Figure C.7 shows a comparison of the RMSE values and the correlation coefficients of the best results of the three grid searches. You can see that there is no clear best solution, but the results vary for the different events. The Degree Three kernel has its advantages in the two summer events but does not perform so well for the events 10 and 11.
Because of the higher mean value of the correlation, I choose the last kernel to use for precipitation in the further analyses. So the new observation equation is:

$$\Delta g_i = P_0 \sum_{j=i-N_1}^{i} r_j \exp\left(-\frac{(i-j)^3}{3\tau_p^3}\right) + P_1 \sum_{j=i-N_2}^{i} s_j \exp\left(-\frac{(i-j)^2}{2\tau^2}\right) + c_k. \tag{C.5}$$

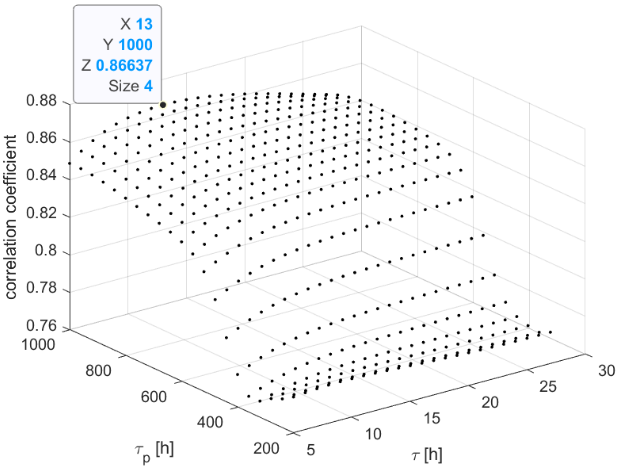

Figure C.6.: CC values of the grid search using precipitation with the Degree Three kernel and one soil moisture probe

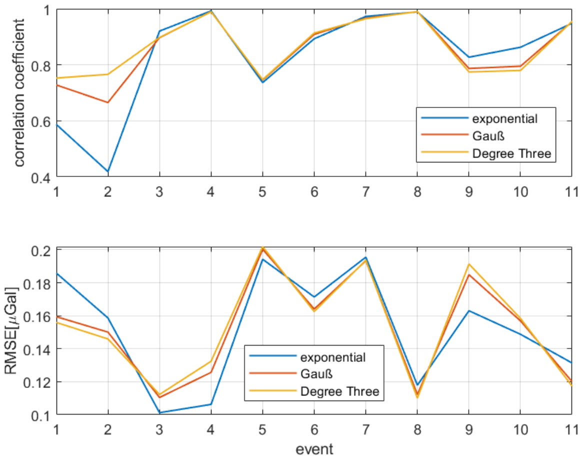

Figure C.7.: RMSE and cc values over the eleven events of the models with one soil moisture probe and precipitation with different kernels

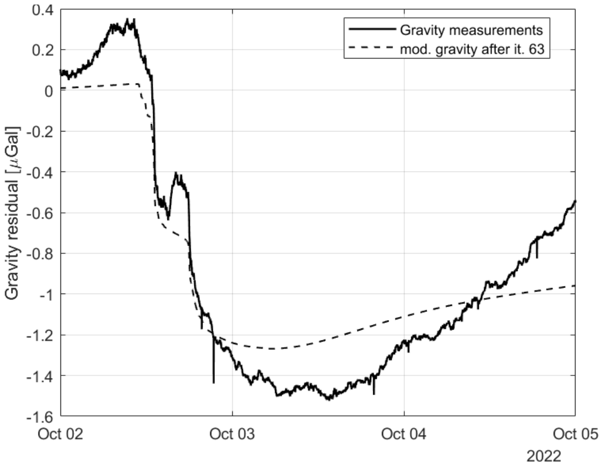

Figure C.8.: Modeled gravity with the result of the LLS-Adjustment with precipitation with the Degree Three kernel, one soil moisture probe

In the LLS adjustment, the decay times both become slightly lower, with τ at 11.8 h and τ_p at 888 h. The mean correlation coefficient again slightly decreases over the starting values to 0.8631. The resulting model over October 2nd (Figure C.8) looks slightly smoother than with only precipitation (Figure C.3) and resembles the true signal more closely.

C.3. Data-driven model with two soil moisture sensors

In this section the data from the first probe of Solidus #2 will be used in addition to the previously used probe 0 of Solidus #1. The observation equation becomes:

$$\Delta g_i = P_0 \sum_{j=i-N_1}^{i} r_j \exp\left(-\frac{(i-j)^3}{3\tau_p^3}\right) + \sum_{h=1}^{2} P_h \sum_{j=i-N_2}^{i} s_{j,h} \exp\left(-\frac{(i-j)^2}{2\tau_h^2}\right) + c_k. \tag{C.6}$$

Here $s_{j,h}$ is the soil moisture reading of the h-th timeseries (Solidus #1, 0 or Solidus #2, 0) at epoch j.

There are now three different decay times that need to be established, which makes a grid search much more difficult. As a first step, I vary τ_p and τ_1, while τ_2 has a fixed ratio to τ_1. For example, Figure C.9 shows the result when τ_2 is half as big as τ_1. The best mean correlation coefficient in this case is the 0.8707 that is highlighted.

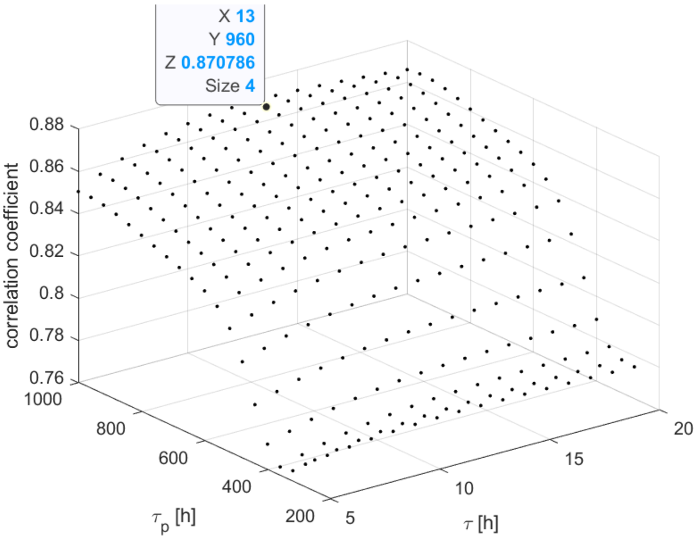

Figure C.9.: The mean correlation coefficient between gravity signal and modeled signal with different values of τ_p and τ_1 with τ_2 being fixed at half as long as τ_1

In general, the longest precipitation decay times at or around 1000 h show the best results, even when the ratio between τ_1 and τ_2 is changed. Because of that, I decide to do another grid search with τ_p fixed at 1000 h and τ_1 and τ_2 variable. The best correlation coefficient is reached with $\tau_1 = 13$ h and $\tau_2 = 5$ h, with a result of 0.8711, which is a slight improvement over the value of the fixed ratio search.

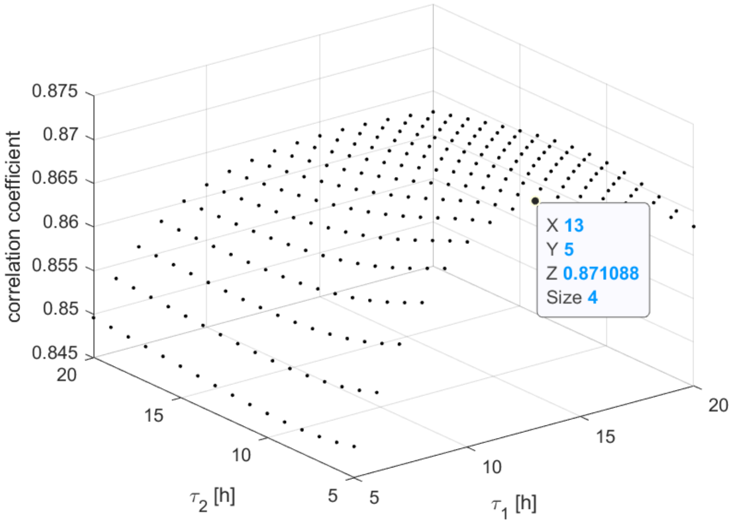

Figure C.10.: The mean correlation coefficient of the grid search with τ_p fixed at 1000 h and τ_1 and τ_2 variable

Using the found decay times and amplitudes as starting values for another LLS-Iteration gives a result with an average correlation coefficient of 0.8626. Figure C.11 shows the resulting model for the October 2nd event.

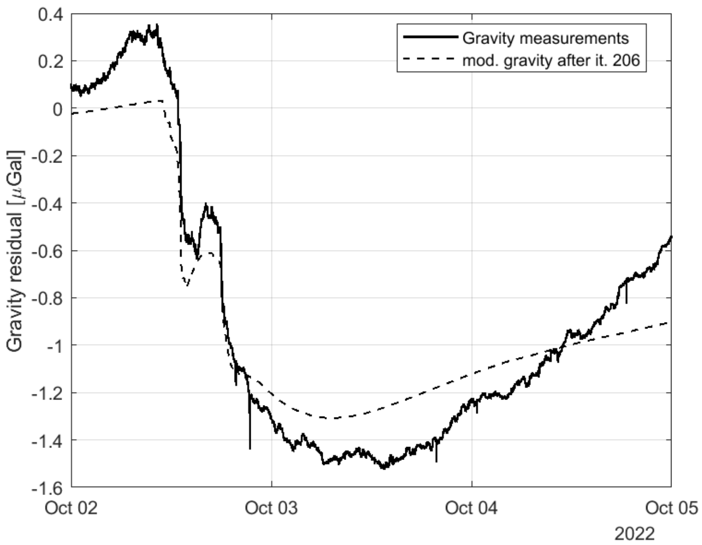

Figure C.11.: The modeled gravity series over the October 2nd event with the result from the LLS-Iteration with precipitation and two soil moisture probes

C.4. Data-driven model with three soil moisture sensors

I carry out a grid search with τ_p fixed at 1000 h and τ_1, τ_2 and τ_3 being tried between 4 and 12 h with 90 min steps in between. The resulting mean correlation coefficients are depicted successively in Figure C.12.

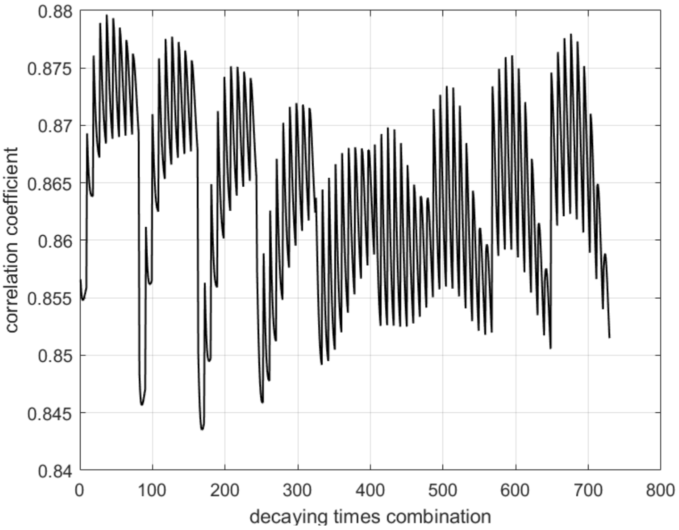

Figure C.12.: The resulting mean correlation coefficients for different soil moisture decay times with precipitation (Degree Three kernel) and three soil moisture probes

The best combination is number 37 with a mean correlation coefficient of 0.8796, which is an improvement over the best value of 0.8711 with only two SM probes. A more detailed search around the first best result yield shorter times for τ_1 and τ_3 with a correlation of 0.8810.

Figure C.13 shows the best result of this search for the October 2nd event.

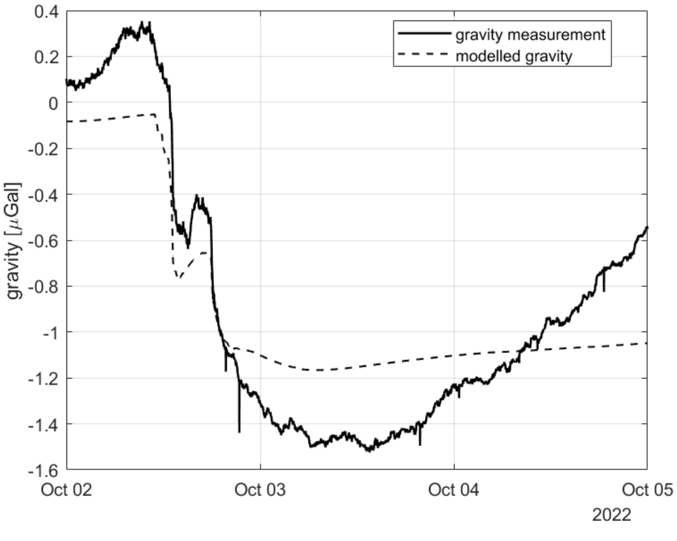

Figure C.13.: Best result found in the grid search with regard to mean cc over the October 2nd event

As the response of the model over the October 2nd event simulates a smaller and smaller gravity decrease, as the correlation values improve, this building was dropped in favor of an RMSE-based search.