



HOCHSCHULE HEILBRONN

Bachelorthesis
Angewandte Informatik

Evaluierung der Genauigkeit eines Vorhersagemodells zur Schätzung von Chlorophyll-a Werten basierend auf Daten kosteneffizienter IoT-Sensoren

Name: Robert Kessler
Matrikelnummer: 205015

26. Juli 2023

Eingereicht bei:
Referent: Prof. Dr.-Ing. Wolfgang Heß
Korreferent: Dr. Christoph Maier

Abstract

In der vorliegenden Arbeit wird die Anwendung des Internet of Things (IoT) zur Erkennung und Prognose von Algenblüten durch die Schätzung des Chlorophyll-a-Gehalts untersucht. Es erfolgt die Entwicklung eines IoT-Systems, welches relevante Messwerte wie Temperatur, pH-Wert und gelöste Feststoffe (TDS) erfasst, um daraus den Chlorophyll-a-Gehalt abzuleiten. Dieser Prozess basiert auf einer multiplen Regressionsanalyse, die auf Daten der Landesanstalt für Umwelt, Messungen und Naturschutz trainiert wurde. Zusätzlich wird eine benutzerfreundliche Anwendung entwickelt, die es ermöglicht, die gesammelten Daten in Echtzeit zu visualisieren. Die Genauigkeit des Vorhersagemodells wird sowohl durch eine Validierung mithilfe von Testdaten, als auch mit eigen gemessenen Werten des IoT-Systems überprüft. Auch wenn einige Abweichungen festgestellt wurden, zeigt die Arbeit die Möglichkeit auf, durch IoT-Technologie eine frühzeitige Erkennung von Algenblüten zu ermöglichen. Die Arbeit schließt mit einer Diskussion über mögliche Erweiterungen und Verbesserungen für zukünftige Forschungen ab.

Keywords— Internet of Things (IoT), Algenblüten, Chlorophyll-a, Umweltüberwachung, Sensoren, Temperatur, pH, TDS, Vorhersagemodell, Validierung, Datenvisualisierung, Datenanalyse, Regression

Inhaltsverzeichnis

Abkürzungsverzeichnis	vi
Glossar	vii
Abbildungsverzeichnis	ix
Tabellenverzeichnis	xii
1 Einleitung	1
1.1 Hintergrund und Motivation	1
1.2 Problemstellung und Zielsetzung	2
1.3 Aufbau der Arbeit	2
2 Theoretische Grundlagen	5
2.1 Internet of Things (IoT)	5
2.1.1 Definition von IoT	5
2.1.2 Anwendung in der Umweltüberwachung	6
2.2 Algenblüten	8
2.2.1 Definition und Merkmale einer Algenblüte	8
2.2.2 Auswirkungen	9
2.2.3 Algen als menschliche Ressource	11
2.3 Chlorophyll-a	13
2.3.1 Eigenschaften und Funktionen	13
2.3.2 Indikator für Algenwachstum	13
2.3.3 Gängige Methoden zur Messung von Chlorophyll-a	15
2.4 Sensorik	18
2.4.1 Temperatur	18
2.4.2 pH	22
2.4.3 TDS	25
2.5 ThingSpeak	28
2.5.1 Was ist ThingSpeak?	28
2.5.2 Funktionsweise	28
2.5.3 Nutzung von ThingSpeak	28
3 Methodik und Vorgehensweise	30
3.1 Datenquellen und Datenerfassung	30
3.1.1 Messung anhand des Internet of Things (IoT)-Systems	30
3.1.2 Messwerte zum Training der Vorhersage	31
3.2 Schätzung der Chlorophyll-a Werte	32
3.3 Validierung der Vorhersagen	32
3.4 Implementierung des IoT-Systems	33

3.5	Implementierung der App	33
4	Datensammlung- und vorbereitung	35
4.1	Sammeln der Trainingsdaten	35
4.2	Datenreinigung	36
5	Entwicklung des Vorhersagemodells	38
5.1	Erstellung der multiplen Regression	38
5.1.1	Teilung der Daten	39
5.1.2	Anwendung der multiplen Regression	40
5.2	Validierung des Vorhersagemodells	45
6	Entwicklung und Implementierung des IoT-Systems	48
6.1	Auswahl der Technologien	48
6.2	Systemarchitektur	49
6.3	Sensorintegration	50
6.3.1	Temperatur	50
6.3.2	PH	52
6.3.3	TDS	55
6.4	Verarbeitung der Sensordaten	57
6.5	ThingSpeak	58
6.5.1	Datenspeicherung	58
6.5.2	Datenübertragung	59
7	Entwicklung der App als Nutzerschnittstelle	62
7.1	Anforderungsanalyse	62
7.1.1	Funktionale Anforderungen	62
7.1.2	Nicht-Funktionale Anforderungen	63
7.2	Auswahl der Technologien	64
7.3	Echtzeitübertragung	64
7.4	Datenvisualisierung	66
7.4.1	Seitenmenü	67
7.4.2	Startseite (Dashboard)	68
7.4.3	Messwert-Details	71
8	Evaluierung der Genauigkeit	74
8.1	Test des Vorhersagemodells anhand eigen-gemessener Werte	74
8.1.1	Schätzung der stündlichen Chlorophyll-a Werte	75
8.1.2	Validierung der Schätzung	75
8.2	Mögliche Ursachen für eine Abweichung der Schätzungen	76
9	Mögliche Erweiterungen	80
9.1	Entfernen von Ausreißern	80
9.2	Entwicklung des Generalized Additive Model (GAM)	83
9.2.1	Anwendung des GAM	83

9.2.2 Vergleich der Modelle	84
9.3 Weitere Messwerte	86
10 Diskussion	88
10.1 Vergleich mit der Zielsetzung	88
10.1.1 Zusammenfassung der Zielsetzung	88
10.1.2 Vergleich der erreichten Ergebnisse	89
10.2 Limitationen der Arbeit und zukünftige Empfehlungen	90
11 Fazit	92
Literatur	93

Abkürzungsverzeichnis

IoT Internet of Things

HAEDAT Harmful Algal Event Database

UNESCO United Nations Educational, Scientific and Cultural Organization

PSP Paralytic Shellfish Poisoning

HAB Harmful Algae Bloom

PTC Positive Temperature Coefficient

NTC Negative Temperature Coefficient

O.D. Optical Density

EDAB Ecosystem Disruptive Algal Blooms

NIR Near Infrared

SPAD Soil Plant Analysis Development

CI Chlorophyll Index

HTTP Hypertext Transfer Protocol

RSE Residual Standard Error

MAE Middle Absolute Error

GAM Generalized Additive Model

IQR Interquartile Range

Glossar

Cloud Bezieht sich auf die Bereitstellung von Ressourcen und Anbindungen meistens von externen Unternehmen.

Erdschreiber's Medium Bezieht sich auf eine Lösung bestehend aus gefiltertes Salzwasser und vielen Nährstoffen, um aquatische Organismen zu züchten.

Eukaryoten Organismen, die einen Zellkern besitzen, wie Tiere, Pflanzen und Pilze.

extrapolieren Bezieht sich auf das Vorhersagen von Werten, die außerhalb des beobachteten Bereichs liegen.

Gravimetrische Analyse Verfahren zur Bestimmung der Menge eines Stoffes in einer Probe, indem die Substanz einer chemischen Reaktion oder Trennungstechnik unterzogen wird.

Halbleiter Material, dessen elektrische Leitfähigkeit, zwischen einem Leiter und einem Nichtleiter liegt.

Ion-Swapping Beschreibt den Austausch von Ionen in einer Lösung gegen Ionen mit ähnlichen Eigenschaften.

Karbonat-Bikarbonat-System Bezeichnet ein chemisches Gleichgewichtssystem in wässrigen Lösungen, welches die Wechselwirkung zwischen Kohlensäure, Bikarbonationen und Carbonationen steuert und den pH-Wert regelt.

Korrosion Bezeichnet die Zerstörung oder Schädigung eines Materials durch chemische oder elektrochemische Reaktionen mit seiner Umgebung, wie zum Beispiel Rost.

Nernst-Gleichung Beschreibt die Konzentrationsabhängigkeit des Elektronenpotentials von zwei chemischen Substanzen.

Phytoplankton Phytoplankton sind winzige, pflanzliche aquatische Organismen, die durch Photosynthese organische Verbindungen produzieren, die von anderen Organismen als Nahrung genutzt werden können.

Potentialdifferenz Beschreibt die Spannung zwischen zwei Punkten.

Prokaryoten Organismen, die keinen Zellkern besitzen, wie Bakterien.

Q₁₀ Beschreibt den Faktor, um den die biologische Rate (hier: Wachstumsrate) bei einer Temperaturerhöhung um 10°C ansteigt.

Redoxpotential Bezieht sich auf die Fähigkeit, Elektronen zu gewinnen oder zu verlieren.

Referenzzustand Bezieht sich auf einen Zustand, der als Vergleichswert zu einem Anderen dient.

smart Bezieht sich im Kontext des IoT auf Geräte, die über Sensoren, Aktoren und Mikroprozessoren verfügen und miteinander sowie mit dem Internet kommunizieren können.

Smart Home Beschreibt die Vernetzung von verschiedenen Geräten und Systemen, die miteinander kommunizieren und sich gegenseitig steuern.

sorbieren Einen Stoff aufnehmen und an die eigene Oberfläche binden.

Technologisches Paradigma Beschreibt eine Zeitperiode, in der sich die Technologie verändert hat und dadurch die Art und Weise beeinflusst hat, wie wir Produkte und Dienstleistungen entwickeln und nutzen.

Abbildungsverzeichnis

2.1	Gemessener Zusammenhang zwischen Chlorophyll-a und der Biomasse von Algen in fünf unterschiedlichen Seen	14
2.2	Jährlicher Durchschnittswert des gemessenen Chlorophyll-a und der Größe der Biomasse in den gemessenen fünf Seen	15
2.3	Vergleich der Messung des Soil Plant Analysis Development (SPAD)-502 und der Masse an Chlorophyll	16
2.4	Vergleich der Messung der SPAD Werte und die der Chlorophyll Index (CI) _{red-edge} im Bezug auf Genauigkeit und Rauschtoleranz	17
2.5	Beispiel für eine 1-Wire Anbindung zum Mikrocontroller mit DS18B20-Baustein	19
2.6	Charakteristische Thermistor-Temperaturkurve, anhand der zwei Varianten PTC und NTC	20
2.7	Algenwachstumsrate in Abhängigkeit der Temperatur anhand eines Beispiels der Algenart <i>Staurastrum Cingulum</i>	21
2.8	Algenpopulation in Zusammenhang mit dem pH-Wert und der Temperatur	25
2.9	Wachsende Algenpopulation in Erdschreiber's Medium anhand der Algenart <i>Dunaliella tertiolecta</i>	27
4.1	Rohe Datensätze des Exports der Messwerte der Messstation Mannheim am Neckar	36
4.2	Ausschnitt des verwendeten Datensatzes als CSV Datei	37
5.1	Eingelesene Spaltennamen in R aus dem importierten Datensatz	38
5.2	Teilung des Datensatzes in ein Trainings- und ein Testdatensatz mithilfe von zufällig ausgewählten Reihen im Datensatz	39
5.3	Aufteilung des Datensatzes in Trainings- und Testdatensatz und Ausgabe der Anzahl an Spalten der jeweiligen Teile	40
5.4	Implementierung des Vorhersagemodells anhand der multiplen Regression mit Ausgabe der geschätzten Vorhersageergebnisse	40
5.5	Durchführung der linearen Regression mit Ausgabe der Vorhersageergebnisse	41
5.6	Die Residuen der Ergebnisse der multiplen Regression	42
5.7	Die Regressionskoeffizienten, die sich aus der multiplen Regression ergeben und die Gewichtungen der einzelnen Messwerte beschreiben	43
5.8	Variabilität der entwickelten multiplen Regression	44
5.9	Validierung der Vorhersage anhand des Testdatensatzes	46
5.10	Berechnung des MAE anhand der Validierung zum Testdatensatz mit Ausgabe der Validierungsergebnisse	46
5.11	Visualisierung der geschätzten Chlorophyll-a Werte (in rot) mit den tatsächlichen Werten (in blau)	47

6.1	Geplante IoT-Systemarchitektur in Schichten dargestellt	50
6.2	Schaltplan des DS18B20 mit Anbindung zum Arduino im Single External Power Mode	51
6.3	Verbaute pH-Sonde mit separatem BNC-Connector	53
6.4	Lineare Gleichungsdarstellung der Kalibrierung von pH-Werten nach analogen Spannungswerten	54
6.5	Methode zur Umrechnung der Spannungswerte in einen pH-Bereich	54
6.6	Algorithmus zum Ausgleichen der Temperaturschwankungen in den Messwerten	56
6.7	Umrechnung der Spannungswerte in einen TDS-Bereich anhand der Koeffizienten	56
6.8	Kurve des gleitenden Durchschnitts über die jeweiligen 100 Messungen im Vergleich zu den gemessenen pH-Werten, die Ausreißer enthalten	58
6.9	Erstellter Kanal mit knapp 20.000 Einträgen von gemessenen Temperatur-, pH- und TDS-Werten	59
6.10	Methode zum Setzen der ThingSpeak-Felder im angelegten Kanal und das Senden dieser Daten	60
7.1	Implementierung einer streambasierten Datenübertragung anhand von HTTP	65
7.2	Parsing der HTTP-erhaltenen Daten in Objekte, die für die Visualisierung verwendet werden können	66
7.3	Menü zum Durchschalten der jeweiligen Seiten, geöffnet durch das Menü-Icon oder per Swipe	68
7.4	Bei App-Start angezeigtes Dashboard zeigt Messwerte und Eigenschaften an	70
7.5	Methode zur Schätzung der Chlorophyll-a Biomasse anhand der Koeffizienten aus dem entwickeltem Vorhersagemodell	71
7.6	Detaillierte Messwertansicht	73
8.1	Vergleich der gemessenen stündlichen Messwerte mit den Sensordaten der LUBW	76
8.2	Ausreißer in den Chlorophyll-a Messungen des verwendeten Datensatzes .	78
8.3	Beziehungen zwischen dem Chlorophyll-a Wert und den unabhängigen Variablen	79
9.1	IQR-Methode zum Entfernen von Ausreißern in einem gegebenen Datensatz	81
9.2	Vergleich der Chlorophyll-a Daten mit (links) den originalen Daten und (rechts) ohne Ausreißern nach Anwendung der Interquartile Range (IQR)-Methode	82
9.3	Residuen aus der multiplen Regression auf einen Datensatz ohne Ausreißer angewandt	82
9.4	Implementierung des GAM in R mit Ausgabe der geschätzten Vorhersagewerten	83

9.5	Ausgabe der Ergebnisse aus dem implementierten GAM	84
9.6	Berechnung und Ausgabe des MAE des GAM durch Validierung mithilfe des Testdatensatzes	85
9.7	Ranking aller Messwertkombinationen sortiert nach MAE mit Ausgabe der Vorhersageergebnisse	87

Tabellenverzeichnis

2.1	Algenwachstumsrate in Zellteilungen pro Tag in Abhängigkeit der Temperatur anhand von sieben beobachteten Algenarten	21
2.2	Abhängigkeit der pH-Skala mit der Temperatur der Lösung	23
3.1	Messwerte der Messstation in Mannheim anhand der gefilterten Werte . .	32
4.1	Mögliche Datensatz-Struktur der Messwerte nach Bereinigung	36
6.1	ThingSpeak Resultcodes nach Senden über die das multi-field Update der offiziellen Bibliothek	61
8.1	Stündliche gemessene Durchschnittswerte mit Schätzung des Chlorophyll-a Gehaltes jeder Stunde	75
8.2	Stündliche gemessene Durchschnittswerte mit Schätzung des Chlorophyll-a Gehaltes jeder Stunde	76
9.1	Vergleich der Modellergebnisse	85

1 Einleitung

Der folgende Teil dient dazu, das Thema vorzustellen und einen Überblick über den Inhalt und die Ziele der Arbeit zu geben.

1.1 Hintergrund und Motivation

Schon seit dem 17. Jahrhundert sorgen Algenblüten auf der gesamten Erde für negative Auswirkungen auf das umliegende Ökosystem, die Wirtschaft und unsere Gesellschaft (Karlson et al., 2021).

Die United Nations Educational, Scientific and Cultural Organization ([UNESCO](#)) pflegt eine umfangreiche Datenbank namens Harmful Algal Event Database ([HAEDAT](#)), welche historische Ereignisse von Algenblüten katalogisiert. Seit dem Jahr 1770 sind in dieser Datenbank mehr als 13.000 Einträge verzeichnet, darunter 84 nur an Deutschlands Nordküste (“Harmful Algal Information System”, 2023).

Das vermehrte Vorkommen von Algenblüten kann in unterschiedlichen Art und Weisen Konsequenzen nach sich ziehen. Die [HAEDAT](#) beschreibt unter anderem viele Fälle von Paralytic Shellfish Poisoning ([PSP](#)) in Deutschland. Dies ist eine Vergiftung, hervorgerufen durch das Essen von Schalentieren, welche aus kontaminierten Becken gefischt wurden.

Einige der [Phytoplankton](#), die Algenblüten bilden, stoßen Toxine aus, die von Hallegraeff et al. (1995) beschrieben, entweder direkte und fatale, als auch indirekte und langanhaltende Folgen auf Mensch und Umwelt haben können.

Algen müssen allerdings nicht immer negative Effekte auf das umliegende Ökosystem haben. Algen haben sich als wichtige Nutzorganismen für Menschen erwiesen und die Forschung diesbezüglich nimmt immer weiter zu.

Als vorrangiges Beispiel steht Mehta und Gaur (2005) mit der Beschreibung, wie Algen benutzt werden können, um Schwermetalle gezielt aus dem Wasser zu [sorbieren](#).

Eine weitere Studie zeigt auf, wie rote Algen genutzt werden können, um Papier nachhaltig herzustellen (Seo et al., 2010).

Die Langzeitüberwachung von schädlichen Algenblüten, ab sofort mit Harmful Algae Bloom ([HAB](#)) abgekürzt, lässt sich auch anderweitig einsetzen.

Algen, im Generellen, sind nicht nur von hoher Wichtigkeit in einem funktionierenden

Ökosystem, sie können ebenfalls als Indikator für Ökosystemveränderungen dienen, die zum Beispiel durch Umweltverschmutzung verursacht werden. Die Forschung von McCormick und Cairns (1994) beschreibt die Möglichkeiten, Algen in vielen Beispielen als Indikator für Veränderungen zu nutzen, da diese sehr schnell und zuverlässig auf einen großen Umfang von Schadstoffen reagieren.

Der Drang, größere Blüten zu erkennen und zu verhindern wird somit immer größer.

1.2 Problemstellung und Zielsetzung

Da die Überwachung von Algenblüten in einem breiten Spektrum Anwendung findet, ist ein System nötig, welches sich auch in den unterschiedlichsten Gebieten einsetzen lässt. Dieses System nennt sich **IoT**, welches sich in den letzten Jahrzehnten als **Technologisches Paradigma** erwiesen hat (Lee & Lee, 2015).

Die Kosten der einzelnen Geräte sinken, währenddessen ihre Popularität immer weiter zunimmt. Da viele Menschen allerdings nicht bereit sind, viel Geld für einzelne Sensoren auszugeben, stellt sich die Frage, wie IoT kostengünstig genutzt werden kann, um Algenblüten zu erkennen und vorherzusagen.

Das Ziel dieser Arbeit ist es, ein Sensorsystem mithilfe von handelsüblichen Sensoren aufzubauen, welches Gewässer auf mögliche Indikatoren und Auftreten von Algenblüten untersucht. Diese Sensordaten werden analysiert und anhand dessen können Aussagen gemacht werden, wie hoch die Chlorophyll-a Konzentration in der gemessenen Substanz ist. Eine App soll dafür sorgen, dass die gemessenen Werte immer von überall aus einsehbar sind und der Nutzer benachrichtigt wird, wenn sich ein Wasserparameter auffällig verhält.

Zum Abschluss der Arbeit soll geklärt werden, ob ein IoT-System, das wenig kostet, für diesen spezifischen Anwendungsfall geeignet ist, oder es doch mehr Equipment verlangt.

1.3 Aufbau der Arbeit

In der vorliegenden Arbeit werden nach der Einleitung die theoretischen Grundlagen, auf die die Arbeit aufbaut, behandelt. Dazu wird das **IoT** definiert und seine Anwendungen in der Umweltüberwachung, um die Relevanz noch einmal zu betonen. Danach wird das Phänomen der Algenblüten, ihre Merkmale, Auswirkungen und mögliche Nutzen weiter betrachtet. In diesem Zusammenhang wird näher auf das Molekül Chlorophyll-a eingegangen, das als Indikator für das Algenwachstum dient. Diesem Kapitel folgt eine

detaillierte Beschreibung der Sensorik, die zur Messung von Temperatur, pH und TDS eingesetzt wird. Dabei wird darauf eingegangen, wie sich diese Messwerte auf chemische Ebene erfassen lassen und welchen Einfluss die jeweiligen Messwerte auf die Algenpopulationen haben. Abschließend wird ThingSpeak vorgestellt, eine Cloud-Anwendung, die zur Speicherung und Auslesung der Messwerte verwendet wird.

Im darauf folgenden Methodik-Kapitel wird beschrieben, woher die Daten abstammen, die zur Entwicklung des Vorhersagemodells benötigt werden. Darüber hinaus wird erläutert, wie die Schätzung an Chlorophyll-a stattfindet und wie die Validierung der Vorhersagen aussehen soll. Ebenso wird die Implementierung des IoT-Systems und der App zur genauen Datenvisualisierung näher erklärt, um einen groben Rahmen zu bilden.

Nach der Diskussion der Methodik folgt die Praxis und somit das Kapitel zur Datensammlung und -vorbereitung, in dem erläutert wird, wie die Trainingsdaten gesammelt und die Daten gereinigt werden. Dies bildet gleichzeitig die Grundlage für das darauf folgende Kapitel zur Entwicklung des Vorhersagemodells, in dem das Verfahren der multiplen Regression verwendet und dessen Validität überprüft wird. Hier wird ebenso Code gezeigt, um darzustellen, wie die Datenanalyse abläuft und detailliert, wie das Vorhersagemodell entwickelt wird.

Im folgenden Praxisteil wird die Entwicklung und Implementierung des IoT-Systems näher beschrieben. Dabei wird die Auswahl der Technologien, die Systemarchitektur, die Integration der Sensoren und dessen Datenverarbeitung von großer Bedeutung sein. Hier spielt ebenfalls ThingSpeak und dessen Nutzung eine Rolle.

Auf die Entwicklung des IoT-Systems folgt wie Entwicklung der App als Nutzerschnittstelle. Hier wird eine Anforderungsanalyse erstellt, um die jeweiligen Funktionen zu beschreiben. Es wird darauf eingegangen, welche Technologien verwendet werden, wie die Echtzeitübertragung und dessen Aktualisierungen gestaltet ist und es wird anhand Screenshots gezeigt, wie die Nutzerschnittstelle aussieht.

Daraufhin folgt eine gründliche Evaluierung der Genauigkeit des Systems und des Modells. Hier werden die Vorhersagen mit eigenen Messwerten verglichen und mögliche Ursachen für große Abweichungen erörtert.

Der letzte Teil des Hauptteils beschäftigt sich mit den Überlegungen, wie das System weiter verbessert werden könnte, was dazu nötig ist und eine Erklärung, weshalb diese Erweiterungen nicht in die Arbeit einfließen.

Im Kapitel Diskussion und Interpretation der Ergebnisse des Schlussteils werden die Ergebnisse dieser Arbeit reflektiert und interpretiert. Hier wird darauf eingegangen, welche potenzielle Verbesserungen und Herausforderungen bei der Implementierung des Systems in der Praxis vorliegen.

Das wesentliche Ziel dieser Arbeit ist es, ein effektives und effizientes IoT-basiertes System zur Überwachung und Prognose von Algenblüten bereitzustellen und damit einen Beitrag zur Lösung des Problems der Algenblüten und ihrer Auswirkungen auf die Umwelt und die menschliche Gesundheit zu leisten.

2 Theoretische Grundlagen

Dieses Kapitel setzt den Rahmen für die vorliegende Arbeit und stellt die wichtigen Konzepte und Bereiche vor, die für das Verständnis der Arbeit von Bedeutung sind. Zunächst wird das Internet der Dinge (IoT) definiert und dessen Anwendung in der Umweltüberwachung erörtert. Danach wird das Phänomen der Algenblüten diskutiert, einschließlich ihrer Merkmale, Auswirkungen und ihrer Rolle als menschliche Ressource. Im Zusammenhang mit den Algenblüten wird der Fokus auf Chlorophyll-a gerichtet, einem wichtigen Indikator für das Algenwachstum. Hier werden die Eigenschaften und Funktionen sowie gängige Methoden zur Messung erläutert. Der Sensorikabschnitt befasst sich mit der Bedeutung der Temperatur-, pH- und TDS-Messungen für die Beobachtung und Vorhersage des Algenwachstums. Abschließend wird die Plattform ThingSpeak vorgestellt, die in dieser Arbeit verwendet wird, um IoT-Daten zu sammeln, zu analysieren und darzustellen. Es werden dessen Funktionen und die spezifische Nutzung innerhalb dieses Projekts dargestellt. Dieses Kapitel bildet damit die Grundlage für die darauf folgenden Teile der Arbeit, in denen das entwickelte IoT-System, das Vorhersagemodell und die App präsentiert werden.

2.1 Internet of Things (IoT)

Im nachfolgenden Kapitel soll die Definition des [IoT](#), auf der die vorliegende Arbeit basiert, erörtert sowie dessen Bedeutung in der heutigen Welt dargelegt werden.

2.1.1 Definition von IoT

Die Definition des Internet der Dinge ([IoT](#)) ist nicht universell festgelegt und befindet sich kontinuierlich in Entwicklung und Anpassung an die fortgeschrittene Technik. Im Groben bezeichnet [IoT](#) allerdings die Vernetzung von herkömmlichen Geräten mit dem Internet durch Hardware- und Softwarekomponenten, die es ermöglichen, mit anderen Geräten und Menschen zu kommunizieren und zu interagieren (Mouha, [2021](#)).

Dies bringt einen technologischen Wandel in unseren Alltag, auch ohne, dass wir wissen, dass ein Gerät Teil eines Netzwerkes ist. Ein Beispiel hierfür ist das Konzept des [Smart Home](#), das inzwischen weite Verbreitung gefunden hat. Dies ist ein Teil des [IoT](#) und bezeichnet die Vernetzung von Sensoren und Geräten, die das Ziel haben das eigene Heim,

mithilfe von Sensoren und Aktoren zu automatisieren, um den Alltag sicherer, einfacher und energieeffizienter zu machen.

Solch Vernetzung findet auch in der Fitness-Industrie statt, indem Smart Watches die eigenen Vitaldaten überprüfen, an eine Datenbank senden und extern aufrufen lassen.

Ein Gerät oder ein System ist [smart](#), wenn es über Sensoren, Aktoren und Mikroprozessoren verfügt, um untereinander oder mit dem Internet kommunizieren zu können. Ein weiteres Beispiel hierfür wären Smart Cities, die unter anderem mithilfe von Sensoren Ampeln schalten und Fahrzeugen Signale geben, um Wartezeiten zu minimieren und Energie einzusparen.

Es wird erwartet, dass das [IoT](#) in den kommenden zehn Jahren weiterhin bedeutenden Einfluss auf das Internet und die Wirtschaft haben wird. Huawei schätzt, dass im Jahre 2025 bis zu 100 Milliarden Vernetzungen stattfinden und dies eine wirtschaftliche Auswirkung von bis zu 11 Trillionen Dollar jährlich mit sich zieht (Manyika et al., [2015](#)). Dies ist auf geringere Gerätekosten, erhöhten Cloud-Speicher sowie verbesserte Geschwindigkeiten und Leistungen zurückzuführen.

In dieser Arbeit wird das [IoT](#) ebenso als die Vernetzung von preisgünstigen Geräten definiert, wobei darauf geachtet wird, dass die verwendete Hardware erschwinglich ist und keinen hohen Kosten erfordert. Jede Teilkomponente sollte unter 50€ kosten und keine monatlichen Gebühren mit sich ziehen.

2.1.2 Anwendung in der Umweltüberwachung

[IoT](#) wird häufig eingesetzt, um Ländereien zu überwachen, die von negativen Umweltauswirkungen betroffen sind. Ein weiteres Anwendungsbeispiel, welches dadurch entstand, nennt sich die Präzisionslandwirtschaft. Dies ist die Kombination von herkömmlicher Landwirtschaft mit der Technologie, um die einzelnen Naturvariablen, die eine Ernte beeinflussen können, zu identifizieren, zu analysieren und zu verwalten. Das Resultat ist mehr Nachhaltigkeit, Rentabilität und besseren Umweltschutz (Araby et al., [2019](#)). Hierbei werden durch Sensoren die natürlichen Umwelteinflüsse überwacht, verstanden und geregelt. Das Projekt von Araby et al. ([2019](#)) beschäftigt sich mit der Überwachung von Ländern, um effizient Nährstoffe zu verabreichen und die Umwelt zu schützen, indem weniger Pestiziden versprüht werden, als benötigt. Die Arbeit beinhaltet auch die Anwendung von Machine-Learning-Algorithmen, die auf Basis von Messdaten Krankheiten, wie den weit verbreiteten Kraut- und Knollenfäule, identifizieren und entsprechende Maßnahmen einleiten können.

Im Bereich der Wasserüberwachung gibt es ebenfalls viele Anwendungsfälle. Zum einen können mit Hilfe von Wasserstandssensoren Fluten verhindert und Warnmeldungen ausgegeben werden.

Zum Anderen wird IoT verwendet, um die allgemeine Wasserqualität zu überwachen. Hierbei werden anhand von Sensoren, Wasserwerte ausgelesen, die mithilfe eines Prozessors verarbeitet und extern visualisiert werden. Zusätzlich können bei Auffälligkeiten direkt Maßnahmen mithilfe von Motoren ergriffen werden.

2.2 Algenblüten

Im folgenden Abschnitt wird erläutert, wie eine Algenblüte definiert ist und aus welchen Typen sich diese abbilden. Dies ist wichtig zu verstehen, da jeder Algentyp unterschiedliche Merkmale und Anpassungsfähigkeiten hat. Es wird ebenfalls darauf eingegangen, welche Folgen Algenblüten auf die Menschheit haben und wie diese genutzt werden können. Zusätzlich wird aufgezeigt, wie relevant die Überwachung von Algen in der heutigen Zeit ist.

2.2.1 Definition und Merkmale einer Algenblüte

Als Algenblüte bezeichnet man eine plötzliche und massenhafte Vermehrung von Algen in einem Gewässer, unabhängig, ob dieses steht oder fließt. Ursachen hierfür können Umweltverschmutzungen, Klimaveränderungen und zu viele Nährstoffe, insbesondere Phosphat sein (mehr zu den Nährstoffen in Kapitel 2.4.3: TDS).

Besondere Merkmale einer Algenblüte können die Trübung des Gewässers oder plötzliches Fischsterben sein. Eine vermehrte Algenpopulation zeigt sich oft durch eine grüne, braune oder auch rote Färbung der Gewässer an Küsten und in Aquarien.

In der folgenden Arbeit werden zwei weltweit vorkommende Organismen vorgestellt.

Algen (Eukaryoten)

Unter gewöhnlichen Algen verstehen die meisten Menschen die grünen Algen, die in Aquarien oder an Küsten vorkommen. Allerdings gibt es eine Vielzahl an Algenarten, die sich vom Aussehen und der Charakteristik unterscheiden:

Grünalgen (Chlorophyta) Es gibt geschätzt rund 6.000 - 8.000 Arten von Grünalgen, wobei der Großteil dieser Art zum Seetang gehört und sich überwiegend in Süßwasser befindet (Chapman, 2013). Diese typischen Grünalgen sind invasiv, allerdings nicht aus eigenem Antrieb heraus. Sie werden meist durch menschlichen Transport in neue Ökosysteme eingeführt.

Rotalgen (Rhodophyta) Von Rotalgen gibt es geschätzt rund 4.000 - 5.000 unterschiedliche Arten, welche sich zu 90% in Salzwasser aufhalten. Viele Rotalgen werden an steilen Küsten beobachtet und produzieren Calciumcarbonat, welches essentiell für das Wachstum von Korallenriffen ist (Chapman, 2013).

Braunalgen (Phaeophyta) Diese Algenart wird nur auf rund 1.500 - 2.000 geschätzt, sind kleiner aber robuster. Wie auch bei den Rotalgen sind diese im Salzwasser heimisch

und werden häufig bei steinigen Küsten beobachtet. Sie sind häufig der Ebbe ausgesetzt und haben gelernt bei starken Trockenzuständen zu überleben.

Phytoplankton Obwohl diese wesentlich kleiner sind als die Algenarten, die zum Seetang gehören, können diese bei massenhafter Ansammlung eine Blüte bilden, die ganze Küsten einnehmen können. Dies nennt man dann **HAB** und können ebenfalls wie beim Seetang in unterschiedlichen Farben auftreten. Bei massivem Aufkommen nennt man solch Phänomen *Green Tide*, *Red Tide* oder *Brown Tide*. **Phytoplankton** stellen ebenfalls den Großteil der Nahrung für Lebewesen dar und generieren rund die Hälfte des Sauerstoffs auf dieser Erde (Chapman, 2013).

Kieselalgen (*Bacillariophyta*) Die Artenvielfalt der Kieselalgen, auch als Diatomeen bekannt, wird von Forschern auf eine Spanne von 60.000 bis 600.000 verschiedenen Arten geschätzt. Diese mikroskopisch kleinen Algen bauen sich eine eigene Wand aus transparentem Silikat und sind häufig in Süßwasser zu finden, beispielsweise in Aquarien. Da Kieselalgen sehr empfindlich auf Parameter- und Nährstoffveränderungen reagieren, eignen sich diese perfekt als Indikator für saubere und gesunde Wasserqualität (Sabater, 2009).

Cyanobakterien (Prokaryoten)

Wie andere Bakterien sind Cyanobakterien in nahezu allen Gewässern vertreten, unabhängig davon, ob es sich um ein Aquarium, Meer, See, Fluss oder Teich handelt. Diese blau-grünen Bakterien (Name abstammend von der Farbe *Cyan*) sind geschätzt schon um die 3.5 Milliarden Jahre alt und gehören zu den ältesten bekannten Formen des Lebens auf der Erde. Cyanobakterien spielten laut Forschern eine große Rolle bei der Formung der Erdatmosphäre, da diese eine der ersten Organismen waren, die durch Photosynthese Sauerstoff produzierten (Chapman, 2013). Darüber hinaus können diese ebenfalls Stickstoff fixieren und für andere Organismen zugänglich machen. Das erklärt auch, warum sich Cyanobakterien problemfrei in schadhaften Gewässern aufhalten können. Wie auch andere **Phytoplankton** können diese zu einer **HAB** heranwachsen. Cyanobakterien geben allerdings durch den Stoffwechsel Toxine ab, die für Mensch und Umwelt gefährlich werden können.

2.2.2 Auswirkungen

Wie schon in Kapitel 1.1 und in der Definition einer Algenblüte angesprochen, haben Algenblüten in größerem Ausmaß weitreichende Auswirkungen auf Mensch und Umwelt.

Im folgenden Abschnitt werden die relevantesten Folgen dargestellt.

Paralytic Shellfish Poisoning (PSP)

PSP ist eine Art der Lebensmittelvergiftung, welche durch das Essen von kontaminierten Muscheln hervorgerufen wird. Die produzierten Giftstoffe ähneln stark dem potenten Nervengift Tetrodotoxin, welches auch in Kugelfischen und dem bekannten Blauring-Oktopus vorkommt (Falconer, 1993). Die Vorfälle an PSP steigen seit 1970 stetig an und kann sich überall auf der Welt verbreiten, unabhängig davon, ob diese in der jeweiligen Region bekannt sind, oder nicht.

Ökosystem

Alleine durch den Klimawandel erwärmen sich die freien Gewässer, werden saurer und verlieren an Sauerstoff (Griffith & Gobler, 2020). Zusätzlich wird das Wachstum von Algen durch diese Veränderungen beschleunigt, was ebenfalls negative Auswirkungen auf das umliegende Ökosystem hat. Eine der häufig genannten Auswirkungen ist, wie schon mehrfach beschrieben, das Produzieren von Giftstoffen, welche die einheimische Nahrungskette und die umgebende Luft vergiftet. Darüber hinaus führt die Algenblüte dazu, dass der gelöste Sauerstoff verbraucht wird, der von anderen Pflanzen und Tieren benötigt wird. Ebenso wird das Kiemengewebe von Fischen angegriffen und schwer beschädigt (Anderson et al., 2015). Solche Algenblüten werden auch Ecosystem Disruptive Algal Blooms (EDAB) genannt und umfassen alle Algenarten, die negative Auswirkungen auf das Ökosystem haben. Bei größeren EDAB ist ebenfalls ein weitflächiges Vogelsterben zu beobachten (Anderson et al., 2015).

Wirtschaft

Der Wirtschaftsaspekt richtet sich nach Veränderungen in den Mustern menschlicher Aktivitäten oder der Ressourcennutzung, die als Folge von Algenblüten auftreten können (Hoagland & Scatasta, 2006). Sobald Algenblüten auftreten, ist es Fischereien sowohl in der EU als auch in den Vereinigten Staaten verboten, gefangene Meereslebewesen zu verkaufen. Der Staat muss dann eine Abwägung treffen, ob die Kosten für die Maßnahmen zur Beseitigung der Algen höher oder niedriger sind, als die Schäden, die sie verursachen. Auch der Tourismus leidet stark darunter, wenn Seen und Meere nicht zugänglich sind und viele Läden nur einen eingeschränkten Handel haben.

2.2.3 Algen als menschliche Ressource

Obwohl Algen meist negativ assoziiert werden, haben sich diese in vielen Gebieten als sehr nützlich erwiesen. Wie schon in Kapitel 2.2 beschrieben, waren Cyanobakterien eine der Hauptausschlaggeber für den Sauerstoff, den wir heute atmen. Algen im Allgemeinen produzieren rund 50% des Sauerstoffs dieser Erde (Chapman, 2013). Im folgenden Teil wird beschrieben, welche Nutzen Algen zur heutigen Zeit für uns Menschen haben und warum es nützlich sein kann, diese zu überwachen und zu züchten:

Nahrung

Algen bieten eine emissionsarme Alternative zum handelsüblichen Tierfutter, da diese beispielsweise bei Kühen zu einer Reduzierung an Methanemissionen führen kann. Darüber hinaus sind Algen auch für Menschen eine gute Alternative, da diese reich an Vitamine, Proteine, Nährstoffe und Antioxidantien sind. Das Review von Kovač Blagojević et al. (2013) beschreibt im Detail die Vorteile einer Ergänzung der Ernährung mit Algen und weist auf, inwiefern diese genutzt werden, um Emissionen zu verringern und Krankheiten vorzubeugen.

Darüber hinaus sind Algen die Hauptnahrungsquelle für viele Meeresbewohner, die in der menschlichen Ernährung von großer Bedeutung sind. Braune Algen werden ebenfalls benutzt, um die Konsistenz von Nahrungsmitteln zu beeinflussen.

Konsumgüter

Rote Algen sind unter anderem eine Hauptquelle für Carrageen. Dies sind natürliche Kohlenhydrate und werden sowohl in Essen, als auch in Pflegeprodukten angewandt (Necas & Bartosikova, 2013). Beispiele hierfür wären Eis, Bier, Shampoo, Soja Milch und Tierfutter.

Rote Algen werden auch zur Herstellung von Papier verwendet, um das Abholzen von Wäldern zu vermeiden (Seo et al., 2010).

Medizin

Die Forschung an roten Algen ist nach wie vor aktiv, da sie Substanzen enthalten, die antibakteriell wirken und das Potenzial zur Krebsprävention haben, die schon in der experimentellen Medizin und pharmazeutisch angewendet werden (Necas & Bartosikova, 2013). Zusätzlich können Algen (vor allem die roten, grünen und braunen) und

Cyanobakterien pharmazeutische Verbindungen eingehen, die gegen virale Infektionen verabreicht werden (wie beispielsweise Herpes oder AIDS) (Chapman, 2013).

Weitere Nutzen

Wie bereits in Kapitel 1.1 erwähnt, besitzen Algen die Fähigkeit, Schwermetalle zu [sor-bieren](#). Dies können wir uns zu Nutze machen, um Abwasser zu reinigen und eine gute Wasserqualität zu gewährleisten (2005). Kieselalgen, auch Diatomee genannt reagieren sehr empfindlich auf Veränderungen der Wasserqualität und können so hervorragend als Indikator dienen (Sabater, 2009). Nebenbei sind Algen ein Hauptbestandteil des erneuerbaren Biotreibstoffes, da diese effiziente Photosynthese betreiben und keine Energie für Blüten oder Stämme aufwenden müssen. Sie können ebenfalls sowohl in Süß-, Salz- oder auch in Brackwasser gehalten werden. Ein weiterer Vorteil hiervon ist, dass diese auch wichtige Nährstoffe aus Abwasser ziehen und CO₂ verwenden, um zu wachsen, was diese nicht nur leicht zu züchten, sondern auch sehr umweltfreundlich machen (Chapman, 2013).

2.3 Chlorophyll-a

In der vorliegenden Arbeit wird die Masse einer Algenblüte anhand der Anzahl von Chlorophyll-a Pigmenten geschätzt. Dafür soll in folgendem Kapitel erläutert werden, welche Rolle Chlorophyll-a beim Wachstum von Algen spielt und welche Messmethoden sich bereits erweisen haben.

2.3.1 Eigenschaften und Funktionen

Chlorophyll-a, ein essenzieller Bestandteil der Photosynthese bei Pflanzen, fungiert als Katalysator bei der Umwandlung von Lichtenergie in chemische Energie. Sie ist als einzige der Chlorophyll Familie in allen Organismen enthalten, die Photosynthese betreiben, von Cyanobakterienzellen bis hin zu den größten Bäumen dieser Erde (Björn et al., 2009). Chlorophyll-a absorbiert primär blaues und rotes Licht und reflektiert und überträgt grünes Licht, was der Pflanze ihre charakteristische grüne Farbe verleiht (Virtanen et al., 2022).

2.3.2 Indikator für Algenwachstum

Die Konzentration von Chlorophyll-a im Wasser dient oft als Anzeiger oder [Referenzzustand](#) für die Biomasse der darin vorhandenen Algen (Poikāne et al., 2010). Die Studie von Desortová (1981) beschreibt den Zusammenhang von Chlorophyll-a Werten mit der Biomasse der Algen. Weitere Untersuchungen haben gezeigt, dass die Chlorophyllkonzentration von mehreren Algenarten in einer Blüte abhängt und sich diese im Laufe des Wachstums verändern kann. Zusätzlich werden große Schwankungen der Konzentration festgestellt, was allerdings mit der Nährstoffkonzentration in Verbindung steht.

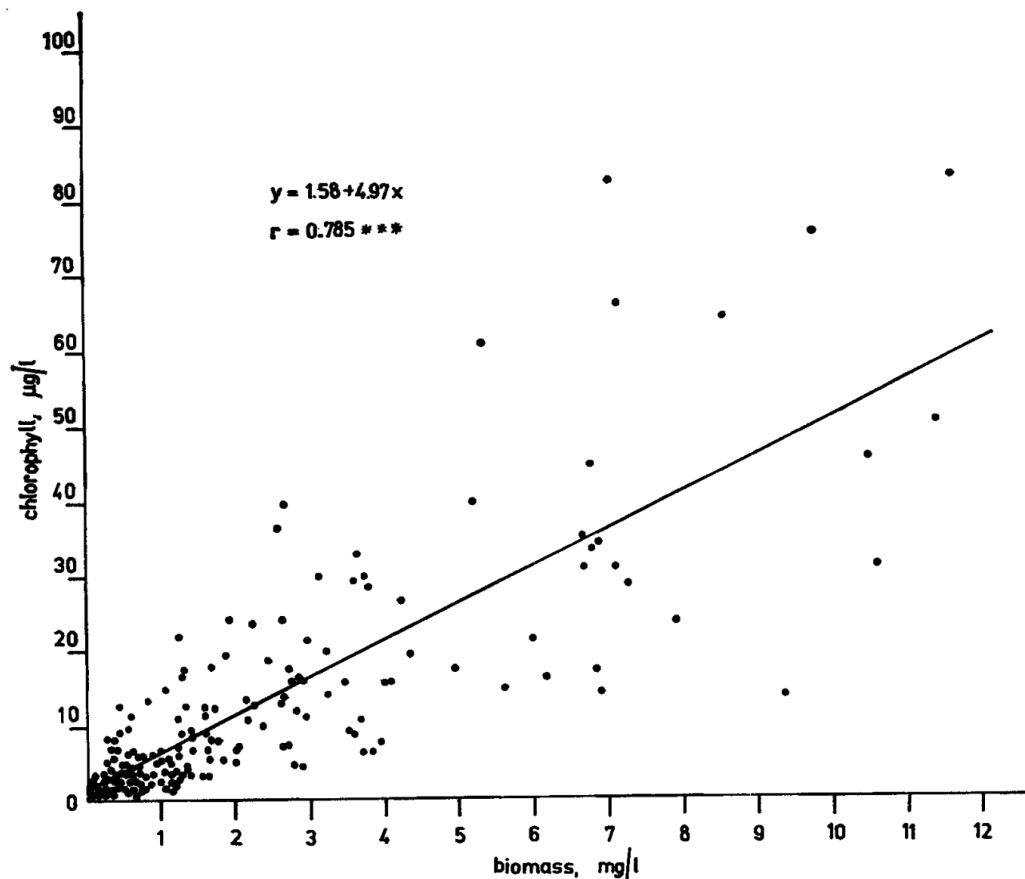


Abbildung 2.1: Gemessener Zusammenhang zwischen Chlorophyll-a und der Biomasse von Algen in fünf unterschiedlichen Seen
Quelle: Poikāne et al. (2010)

In Abbildung 2.1 wird aufgezeigt, wie sich bei steigender Chlorophyll-a-Konzentration die Biomasse der Algen vergrößert. Diese Messungen wurden in fünf unterschiedlichen Seen in Tschechien vorgenommen und visualisiert. Wenn allerdings der jährliche Mittelwert der einzelnen Seen in Betracht gezogen wird, entfallen die Extremen und es kann eine geeignetere Schlussfolgerung gezogen werden. Dieser jährlicher Mittelwert wird in Abbildung 2.2 dargestellt und zeigt einen engeren Zusammenhang ohne Ausreißer oder größerer Streuung.

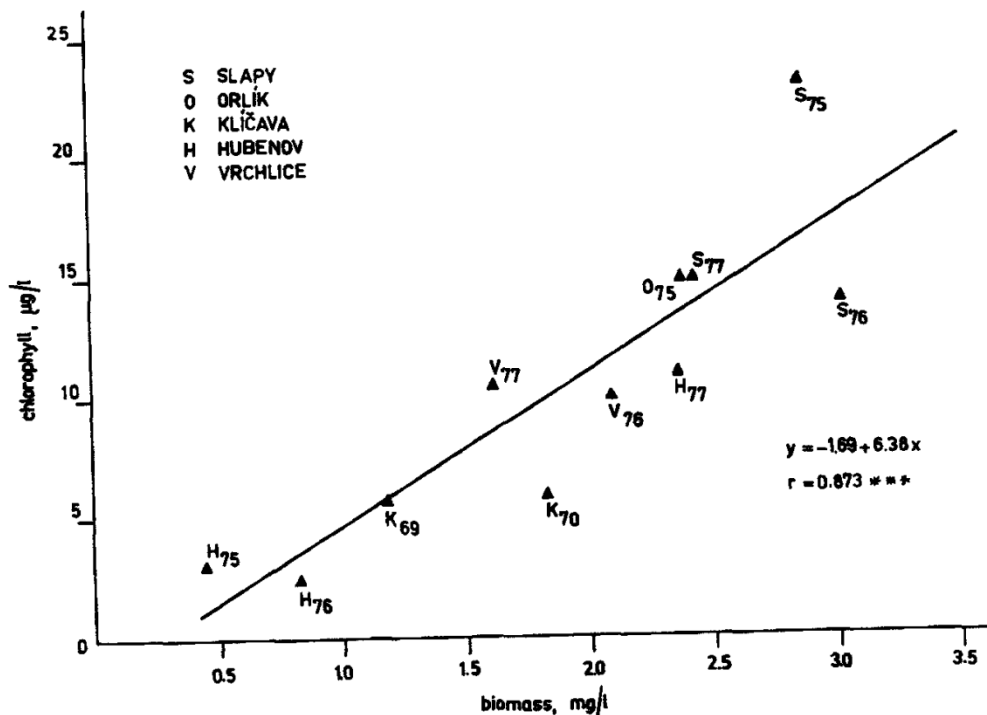


Abbildung 2.2: Jährlicher Durchschnittswert des gemessenen Chlorophyll-a und der Größe der Biomasse in den gemessenen fünf Seen
 Quelle: Poikāne et al. (2010)

2.3.3 Gängige Methoden zur Messung von Chlorophyll-a

Die herkömmliche Methode zur Messung der Algenbiomasse besteht in der mikroskopischen Zählung und Messung jeder einzelnen Algenzelle. Da dies aber viel Zeit benötigt und technisch schwierig umzusetzen ist, wurde ein Verfahren entwickelt, welches direkt den Chlorophyll-a Anteil mithilfe der Spektrophotometrie misst (Desortová, 1981). Ein Spektrophotometer, ein Gerät, das in solchen Verfahren verwendet wird, erfasst die Anzahl der Photonen (Lichtintensität), die durch eine bestimmte Lösung hindurchdringen. Dazu wird ein Licht, das dem Tageslicht entspricht, durch die Lösung gegeben und das reflektierte und durchdrungene Licht wird in die einzelnen Wellenlängen aufgeteilt. Jedoch wird dazu eine Lösung hinzugegeben, um die Absorption des zugefügten Lichtes messen zu können,

Da allerdings zur Messung der Lichtabsorption eine Lösung beigefügt werden muss, ist diese Messmöglichkeit nicht nur präzise, sondern leider auch sehr destruktiv (Steele et al., 2008).

Eine weitere, weniger destruktive Methode zur Messung der Algenbiomasse ist die Ver-

wendung eines **SPAD**-Meters.. Dies ist ein Chl-Meter, welcher ähnlich zum Spektrophotometer die Durchlässigkeit misst, allerdings nur von zwei Wellenlängen (rot zur Messung der Absorption und Near Infrared (**NIR**) als Referenz). Die Studie von Steele et al. (2008) beschäftigt sich mit der Genauigkeit des **SPAD** und inwiefern eine gute Messung an Chlorophyll möglich ist. In Abbildung 2.3 ist dargestellt, wie die Messung des **SPAD** und der Chlorophyll-Werte zusammenhängen.

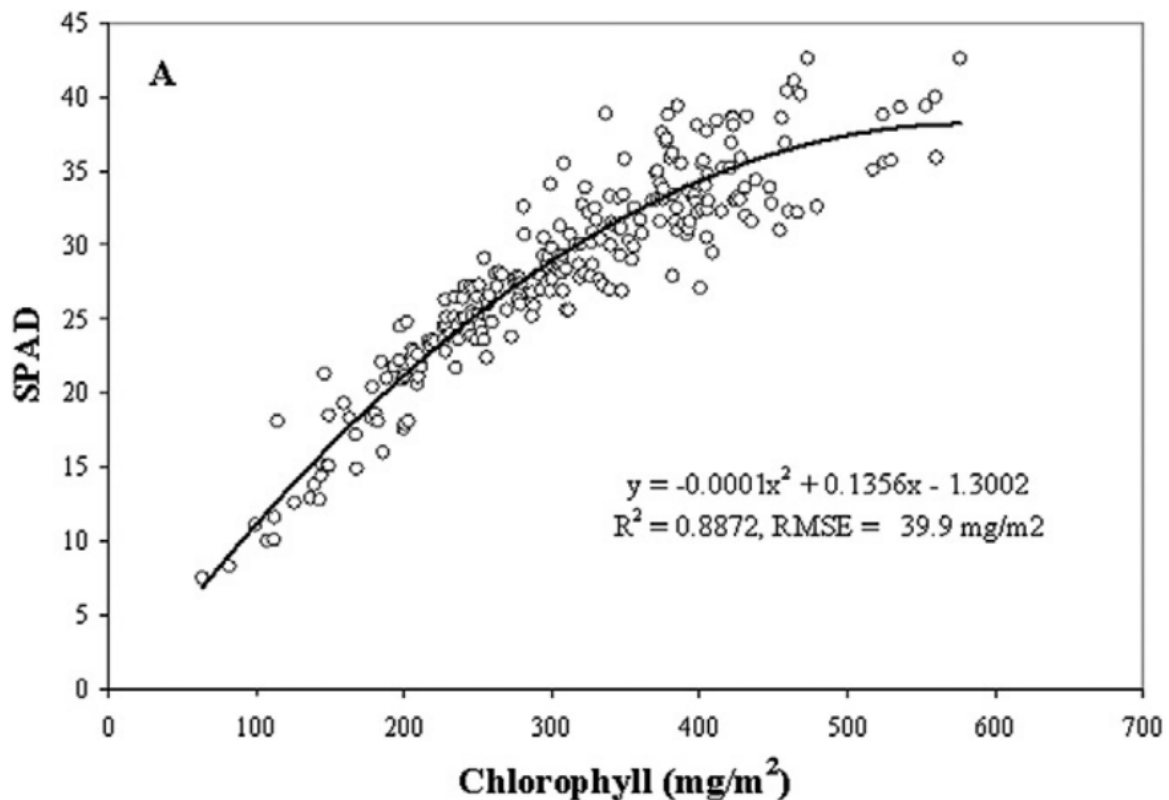


Abbildung 2.3: Vergleich der Messung des **SPAD**-502 und der Masse an Chlorophyll
Quelle: Steele et al. (2008)

Der SPAD zeigt eine adequate Sensitivität zum Chlorophyll, allerdings wird dieser über einem Wert von 300 mg/m² weniger zuverlässig.

Wenn allerdings die Laborwerte des **CI_{red-edge}** verwendet werden, die zusätzlich einen Index für die Reflektierung im roten und **NIR** Spektrum mitführt, kann man linear ohne Rauschen einen besseren Zusammenhang feststellen und auch genau auf den Gesundheitsstatus einer Pflanze schließen.

Die Unterschiede zwischen den **SPAD**-Werten und denen der **CI_{red-edge}** aus dem Labor werden in Abbildung 2.4 dargestellt.

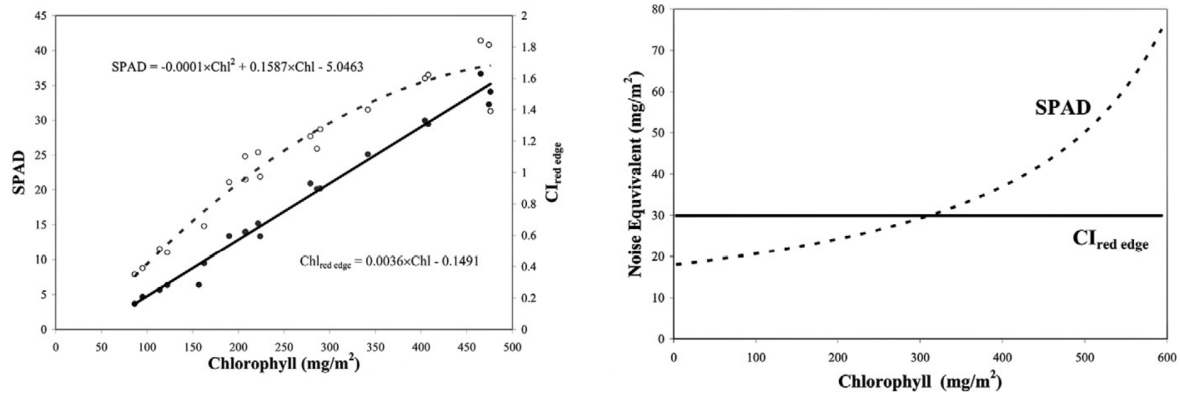


Abbildung 2.4: Vergleich der Messung der SPAD Werte und die der $CI_{red-edge}$ im Bezug auf Genauigkeit und Rauschtoleranz
Quelle: Steele et al. (2008)

Da allerdings sowohl die Messmöglichkeit des Spektrophotometer, des SPADs, des $CI_{red-edge}$ oder der Zählung der Zellen sehr teuer und zeitaufwändig ist, ist es von hoher Wichtigkeit eine Alternative zu finden, welche es ermöglicht automatisiert und kostengünstig Chlorophyll Werte zu messen oder zu schätzen.

Die folgende Arbeit beschäftigt sich deshalb mit der Frage, ob IoT Sensoren geeignet sind, um mithilfe von externen Messdaten die Masse an Chlorophyll-a genau zu schätzen und somit eine Alternative zu bieten.

2.4 Sensorik

Im folgenden Kapitel wird die Sensorik erklärt, die in der vorliegenden Arbeit benutzt wird. Dabei soll darauf eingegangen werden, wie die Sensoren funktionieren, was diese messen und welchen Mehrwert diese dem Projekt bieten.

2.4.1 Temperatur

Im praktischen Teil dieser Arbeit wird zur Messung der Wassertemperatur der DS18B20 verwendet. Es gibt verschiedene Varianten des Sensors, die sich in Gehäuseform, Kabellänge, Genauigkeit und Auflösung unterscheiden. Für diese Arbeit ist jedoch derjenige Sensor am besten geeignet, der ein Edelstahlgehäuse besitzt. Somit ist dieser wasserdicht und stellt keine Gefahr vor [Korrosion](#) dar.

Um die Anbindung des DS18B20 zu verstehen soll kurz erklärt werden, was ein 1-Wire Bus ist, da dieser darauf aufbaut.

1-Wire Bus

Bei Wetterstationen ist es üblich, dass mit jedem Sensor immer mehr Kabel und Schnittstellen aufkommen. Bei stetig wachsender Anzahl an Sensoren, um eine breite Masse an Werten abzudecken, wird dies schnell kompliziert und unübersichtlich. Dieses Problem versucht die 1-Wire Schnittstelle, von der damaligen Firma Dallas Semiconductor entwickelt, zu lösen, in dem es nur einen Kabel für die Daten mehrerer Sensoren benötigt. Der Name 1-Wire ist eventuell irreführend, da neben der Datenleitung zusätzlich eine Leitung für Ground und eine für die positive Spannungsversorgung vorliegen muss (wenn der angebundene 1-Wire Baustein nicht selbst über eine eigene Stromversorgung verfügt).

Die Funktionsweise liegt darin, dass jeder angeschlossene 1-Wire Baustein eine 16-stellige Registrierungsnummer enthält, welche für die spätere Zuordnung nötig ist. Die 1-Wire Bausteine werden mithilfe eines Masters (hier der Mikrocontroller) gesteuert, um Daten über eine Verbindung an alle Teilnehmer zu senden und zu empfangen. Dazu muss ein Pull-Up-Widerstand vor den Slaves geschaltet werden, um einen stabilen HIGH-Pegel zu halten, wenn sich der Sensor im Leerlauf befindet. Ansonsten kann es vorkommen, dass der Sensor unregelmäßig und unerklärlich durch aufkommende Störungen beeinflusst wird.

In Abbildung [2.5](#) ist ein Beispiel dargestellt, wie ein 1-Wire Bus mit Pull-Up Widerstand aussehen kann.

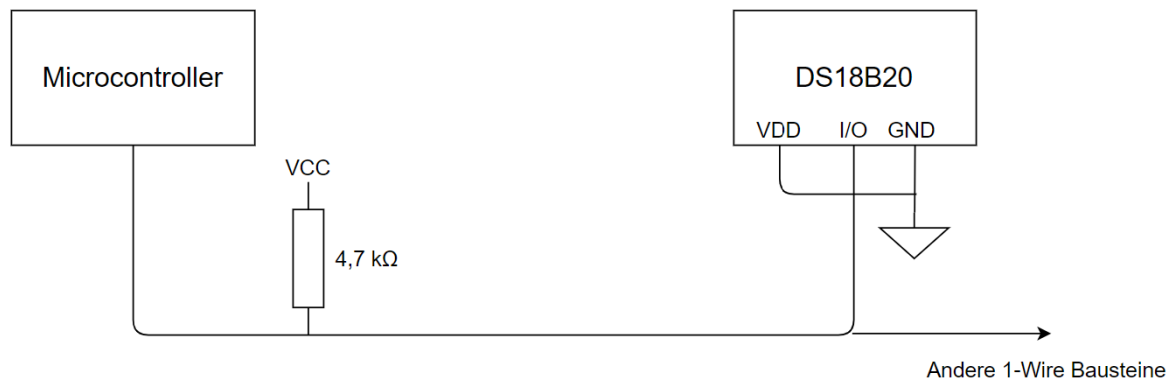


Abbildung 2.5: Beispiel für eine 1-Wire Anbindung zum Mikrocontroller mit DS18B20-Baustein

Dementsprechend wird der DS18B20 im praktischen Teil dieser Arbeit wie oben beschrieben per 1-Wire verbunden und angesteuert.

Funktionsweise

Der DS18B20 enthält zur Messung der Temperatur einen Thermistor.

Ein Thermistor (abstammend vom Englischen: THERMally sensitive ResISTOR) ist ein [Halbleiter](#), der wie ein wärmeempfindlicher Widerstand agiert. Dazu gibt es zwei Varianten: den Positive Temperature Coefficient ([PTC](#)), bei der der Widerstand mit steigender Temperatur ansteigt, und den Negative Temperature Coefficient ([NTC](#)), bei der der Widerstand mit zunehmender Temperatur abfällt (Esenowo Jack et al., [2016](#)).

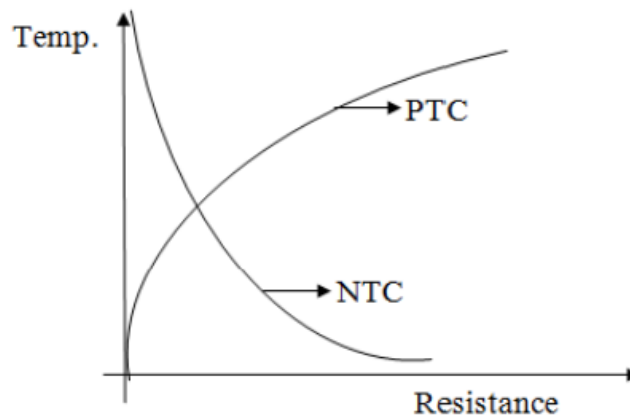


Abbildung 2.6: Charakteristische Thermistor-Temperaturkurve, anhand der zwei Varianten PTC und NTC

Quelle: Esenowo Jack et al. (2016)

Abbildung 2.6 zeigt die charakteristische Temperaturkurve eines Thermistors und dessen Widerstand in Abhängigkeit zur Temperatur. Der Widerstand fällt oder steigt somit nicht linear, sondern anhand einer Kurve. Wie diese Kurve aussieht hängt allerdings stark von dem Material des Thermistors ab. Während Thermistoren, deren Anwendungen in niedrigen Temperaturen (-55 bis 70°C) stattfinden, niedrigere Widerstände benutzen (2252 bis 10.000 Ω), können Thermistoren bei höheren Temperaturen weitaus mehr als 10.000 Ω an Widerstand bieten.

Einfluss auf die Algenpopulation

Die Temperatur des Gewässers ist ein wichtiger Umweltfaktor, der die Algenpopulationen stark beeinflusst (Teoh et al., 2010).

Der Zusammenhang zwischen der Temperatur und dem Algenwachstum wird häufig mithilfe eines Temperatur-Koeffizienten Q_{10} oder mithilfe der Arrhenius-Funktion beschrieben. Die Arrhenius-Funktion stammt aus der Physikalischen Chemie und beschreibt eine exponentielle Abhängigkeit zwischen Reaktionsgeschwindigkeit und der Temperatur. Sie beruht auf der Annahme, dass eine Reaktion nur dann abläuft, wenn die beteiligten Teilchen eine ausreichende Aktivierungsenergie aufweisen. Es wird ebenso vermutet, dass das Wachstum von Algen bis zu einem Optimum hin steigt und danach bei höheren Temperaturen wieder abfällt. Man kann ebenso beobachten, dass die optimale Temperatur eine erhöhte Stoffwechselrate begünstigt und das Wachstum stärkt (Teoh et al., 2010).

Die Studie von Butterwick et al. (2005) beobachtet das Wachstum von acht unterschiedlichen Algenarten in Relation mit steigenden Temperaturwerten.

In Abbildung 2.7 kann anhand des Beispiels der Algenart *Staurastrum Cingulum* eine erhöhte Wachstumsrate bei erhöhter Temperatur beobachtet werden. Die Studien von Teoh et al. (2010) resultierten ebenfalls in der Annahme, dass das Optimum der meisten Algenarten bei über 20°C liegt.

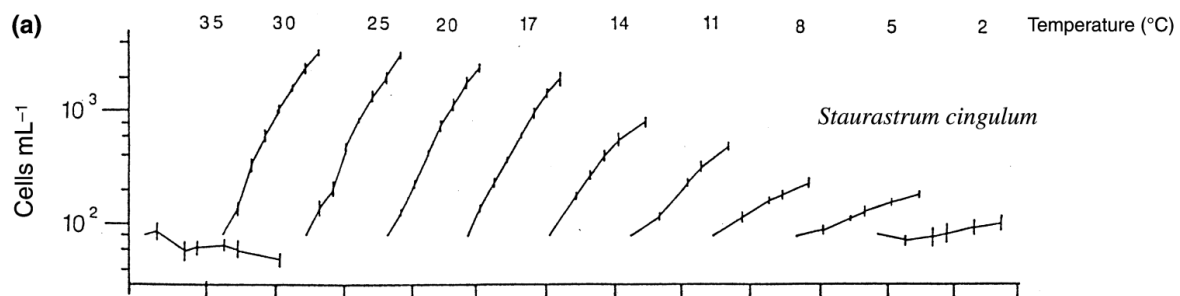


Abbildung 2.7: Algenwachstumsrate in Abhängigkeit der Temperatur anhand eines Beispiels der Algenart *Staurastrum Cingulum*
Quelle: Butterwick et al. (2005)

	Temperature (°C)										
Species	2	5	8	11	14	17	20	25	30	35	
<i>Asterionella formosa</i>	0.40 (0.01)	0.61 (0.02)	0.77 (0.01)	1.05 (0.01)	1.44 (0.02)	1.68 (0.15)					
<i>Cryptomonas marssonii</i>	0.21 (0.02)	0.38 (0.03)	0.62 (0.06)	0.81 (0.05)	0.75 (0.13)	1.51 (0.04)	1.55 (0.04)	1.34 (0.03)	0		
<i>Dinobryon divergens</i>	0.20 (0.03)	0.29 (0.02)	0.44 (0.01)	0.50 (0.01)	0.66 (0.01)	0.79 (0.11)	0.77 (0.05)	0.72 (0.05)	0.59 (0.07)	0	
<i>Ceratium furcoides</i>	0	0	0	0.06 (0.02)	0.13 (0.01)	0.72 (0.03)	0.67 (0.02)	0.56 (0.02)	0.37 (0.10)	0	
<i>Tribonema</i> sp.	0.06 (0.01)	0.15 (0.02)	0.24 (0.08)	0.41 (0.01)	0.55 (0.04)	0.23 (0.01)	0.29 (0.02)	0.29 (0.01)	0.28 (0.02)	0	
<i>Staurastrum cingulum</i>	0.08 (0.04)	0.17 (0.03)	0.24 (0.04)	0.42 (0.01)	0.58 (0.04)	0.31 (0.11)	0.45 (0.13)	0.45 (0.13)	0.53 (0.07)	0	
<i>Tychonema bourrellyi</i>	0	n.s.	0.29 (0.07)	0.46 (0.06)	0.70 (0.10)	0.53 (0.05)	0.52 (0.03)	0.53 (0.07)	0		
						0.63 (0.01)	0.71 (0.01)	0.81 (0.02)	0.87 (0.01)	0.95 (0.59)	0
						0.76 (0.08)	1.07 (0.05)	0.95 (0.06)	0.91 (0.08)	n.s.	0

Tabelle 2.1: Algenwachstumsrate in Abhängigkeit der Temperatur anhand von sieben beobachteten Algenarten
Quelle: Butterwick et al. (2005)

Die Tabelle 2.1 zeigt, wie viele Zellteilungen die beobachteten Algenarten innerhalb eines Tages unter unterschiedlichen Temperaturen von 2-35°C ausgeübt haben.

Es muss allerdings erwähnt werden, dass so wie bei anderen Werten, jede Algenart ihr

eigenes Optimum besitzt und somit ein breites Spektrum an Toleranz und Temperatur-Veränderungspotenzial aufweist.

2.4.2 pH

Schon seit 1920 wird der Säuregehalt einer Lösung in den unterschiedlichsten Gebieten mithilfe des pH-Wertes bestimmt (Mesmer & Holmes, 1992).

Bevor auf die Funktionsweise der Sensoren eingegangen wird, soll die Definition von pH erklärt werden, um zu verstehen, welche Prozesse ablaufen, die diesen Wert messen und inwiefern diese einen Einfluss auf die Algenpopulation haben.

Definition von pH

Das **P** in PH steht für 'Potenz' (im Sinne von 'Maß der Kraft') im Deutschen, im Englischen wird es als 'Power' interpretiert. Das **H** ist das chemische Symbol für Wasserstoff. PH beschreibt also das Potential an Wasserstoffionen und lässt auf den Säuregehalt einer Lösung schließen.

Ebenso, wie bestimmtes Wasser eine Temperatur von 30°C hat, hat beispielsweise Orangensaft einen pH-Wert von 4,0. Somit lassen sich genaue Aussagen über den Säuregehalt treffen.

pH-Skala

Die pH-Skala wurde entwickelt, um einen Weg zu schaffen, den Säuregehalt und die Basizität (oder Alkalität) einer Lösung genau zu bestimmen.

Die Wasserstoff-Ionen (H^+ , Ionen mit positiver Ladung) wirken als starke Säure und die Hydroxid-Ionen (OH^- , Ionen mit negativer Ladung) als starke Base. Sie reagieren miteinander und bilden Wasser (H_2O). Somit befindet sich in reinem Wasser bei 25°C weder eine übermäßige Anzahl an Wasserstoff-, noch Hydroxidionen - das Wasser ist neutral, hat also einen pH-Wert von 7.

Da allerdings die meisten Lösungen weniger als genau 1 Mol pro Liter an H^+ oder OH^- besitzen, wurde die pH-Skala von 0 - 14 erschaffen (Westcott, 2012). Bei sehr sauren oder basischen Lösungen ist es rein-theoretisch möglich, einen pH-Wert von unter 0, oder über 14 zu haben, allerdings ist dies in der Praxis sehr schwierig zu reproduzieren.

Ein wichtiger Aspekt ist ebenfalls, dass der pH-Wert von externen Faktoren beeinflusst werden kann. Für diese Arbeit ist es wichtig, den Zusammenhang zwischen pH und

Temperatur zu verstehen. Wie bereits erwähnt, gelten die Normalbedingungen für Ionenverhalten nur bei genau 25°C.

In Tabelle 2.2 soll gezeigt werden, wie sich die pH-Skala in Abhängigkeit der Temperatur verschiebt. Somit hat das reine Wasser, welches einen pH-Wert von 7 unter 25°C haben sollte einen anderen pH-Wert bei anderen Temperaturen. Die Abweichung vom Wert 7, der als neutral gilt, nimmt immer weiter zu, desto mehr Abweichung sich von den 25°C ergeben, wobei sich die zwei Extremen weiterhin symmetrisch um die Neutralität bewegen.

Neutral or Basic Solutions versus Temperature

Solution	Temperature (°C)		
	25°	0°	60°
Neutral			
(H ⁺)	10^{-7}	3.3×10^{-8}	3.1×10^{-7}
(OH ⁻)	10^{-7}	3.3×10^{-8}	3.1×10^{-7}
pH	7	7.47	6.51
Basic			
(H ⁺)	10^{-14}	1.14×10^{-15}	9.6×10^{-14}
(OH ⁻)	10^0	10^0	10^0
pH	14	14.943	13.017

Tabelle 2.2: Abhängigkeit der pH-Skala mit der Temperatur der Lösung
Quelle: Westcott (2012)

Messverfahren

Da im praktischen Teil dieser Arbeit eine Glaselektrode benutzt wird, ist im folgenden Text nur diese Messmethode beschrieben.

Die pH-Sonde besteht aus zwei Elektroden (einer Sensorelektrode und einer Referenzelektrode), welche die Veränderungen der Wasserstoffionen-Konzentration messen. Die Glaselektrode beinhaltet eine Lösung mit einem pH-Wert von genau 7, also hat diese Lösung eine bestimmte Anzahl an Wasserstoffionen. Somit misst die Glaselektrode die Differenz der beiden pH-Werten von der zu testenden Lösung und der Lösung mit pH-Wert 7. Dies geschieht durch die Differenz der Spannung, die die jeweiligen Wasserstoffionen der einzelnen Lösung produzieren (“How Does A pH Probe Work?”, 2021).

Die Glaselektrode enthält eine Wasserstoffionen-sensible Glasschicht, die auf die Konzentration der Wasserstoffionen in der Lösung reagiert. Bei der Reaktion wechseln die Wasserstoffionen in der Probe und die Natriumionen im Glas der Elektrode die Plätze, wodurch eine elektrische Spannung erzeugt wird. Dieser Ionenaustauschprozess ist das Herzstück der pH-Messung mit einer Glaselektrode. Dieses Verfahren nennt man in der Elektrochemie auch [Ion-Swapping](#) und beschreibt das Hauptprinzip der pH-Sonde.

Jetzt, da nach dem Austausch der Ionen eine Differenz des Säuregehaltes zwischen der Lösung und der Elektrode entsteht, resultiert es ebenfalls in einer Differenzspannung zwischen den Lösungen (auch [Potentialdifferenz](#) genannt). Desto höher die [Potentialdifferenz](#), desto größer ist die Differenz an Wasserstoffionen und vice versa. Diese Differenzspannung wird nun benutzt, um den pH-Wert mithilfe der [Nernst-Gleichung](#) auszurechnen.

Einfluss auf die Algenpopulation

Die meisten Algenarten wachsen bei pH-Werten von 5 - 12 am besten, da viele ihrer biochemischen Prozesse, einschließlich der Photosynthese, am effizientesten in diesem pH-Bereich ablaufen. Währenddessen der pH-Wert Auswirkungen auf das Algenwachstum hat, scheint das auch andersrum der Fall zu sein. Diese können schon in mehreren Wochen den pH-Wert regulieren und zu niedrige Werte erhöhen und zu hohe Werte verringern. So bleibt das Wasser mit hoher Algenpopulation oft bei einem pH-Wert von 8.5 - 11.0. Bei sterilen Überwachungen, also bei Sicherstellung, dass keine Verunreinigungen oder andere Mikroorganismen das Ergebnis beeinflussen, liegt der Bereich bei pH 4.0 - 10.0 (Dubinsky & Rotem, [1974](#)).

Der pH-Wert ist einer von vielen Umweltfaktoren, die ein Ökosystem stark beeinflussen können, darunter deren [Redoxpotential](#), das [Karbonat-Bikarbonat-System](#) und die Verfügbarkeit von Nährstoffen. Menschliche Aktivitäten, wie die industrielle Verschmutzung und die Versauerung von Regenwasser durch erhöhte CO₂- und SO₂-Konzentrationen, können den pH-Wert von Gewässern erheblich verändern. Diese Veränderungen können das Algenwachstum beeinflussen, indem sie den pH-Wert außerhalb des für Algen optimalen Bereichs verschieben. Darüber hinaus können solche Veränderungen die Verfügbarkeit von Nährstoffen beeinflussen und so indirekt das Algenwachstum beeinflussen. Weniger bekannt und verstanden sind die langfristigen pH-Veränderungen, die möglicherweise mit der ungleichen Aufnahme von Nährstoffen zusammenhängen.

Da, wie in Tabelle [2.2](#) gezeigt, der pH-Wert stark von der Temperatur abhängt, wird in der Beobachtung von Algenwachstum unter pH-Werten ebenfalls die Temperatur be-

achtet.

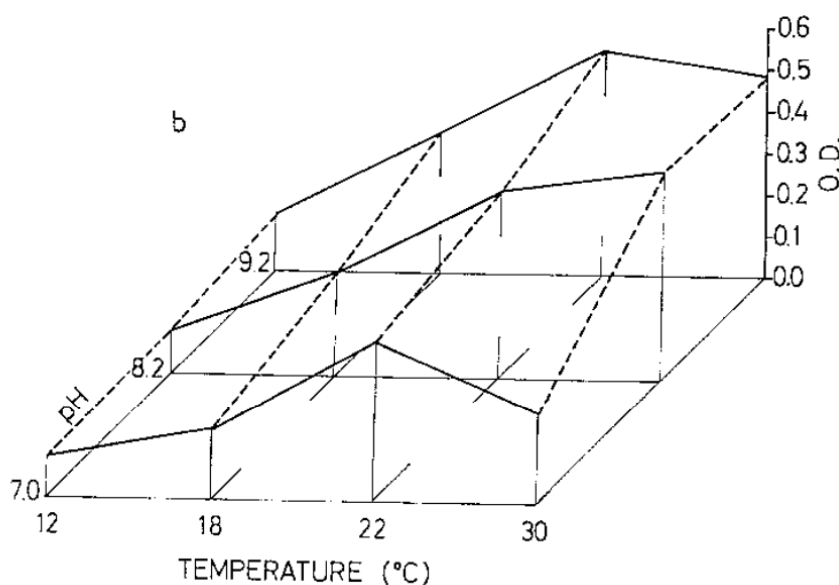


Abbildung 2.8: Algenpopulation in Zusammenhang mit dem pH-Wert und der Temperatur

Quelle: Dubinsky und Rotem (1974)

In Abbildung 2.8 sieht man die gemessenen Algen, in Abhängigkeit von der Temperatur und dem pH-Wert. Die Größe der Population wird in Optical Density (O.D.), angegeben, welche die Absorption von Licht durch die jeweilige Algenart beschreibt und als Messgröße für die Konzentration der Algenzellen dient.

Nach der Studie von Dubinsky und Rotem (1974) liegt der größte gemessene Wert bei pH 9.2 und 24°C.

2.4.3 TDS

Für die Bestimmung der Nährstoffkonzentration im Wasser ist die Messgröße TDS (Totally Dissolved Solids) weit verbreitet. TDS bezeichnet sowohl die Konzentration an anorganischen Substanzen, wie Salz, als auch einen kleinen Teil an organischen Stoffen.

Dieser Wert wird entweder in ppm (parts per million) oder mg/L (Milligramm pro Liter) angegeben und dient beispielsweise als Indikator, ob unser Wasser zum Trinken geeignet ist (Zaman Zad Ghavidel & Montaseri, 2014). Der Großteil, der den Wert ausmacht, besteht meistens aus den sechs Hauptionen Kalzium, Magnesium, Natrium, Bicarbonate,

Chloride und Sulfate.

Messverfahren

Es gibt zwei gängige Methoden zur Messung des TDS-Levels:

Die erste Möglichkeit ist ein manuelles Verfahren, welches sich [Gravimetrische Analyse](#) nennt. Dazu wird die zu messende Lösung verdampft, sodass nur noch feste Rückstände übrig bleiben. Diese werden dann gewogen, analysiert und verrechnet. Diese Methode ist die beste, um den TDS-Wert präzise zu bestimmen, ist allerdings sowohl zeit-, als auch kostenaufwendig. Die zweite Messmöglichkeit misst die Leitfähigkeit der Substanz. Diese Methode ist bei Lösungen mit einer Mehrheit an anorganischen Salzen geeigneter und errechnet sich den TDS Wert aus der Messung.

Dazu wird die Messung der Leitfähigkeit, angegeben in mS/cm (Mikrosiemens pro Centimeter), direkt auf einen TDS Wert, angegeben in ppm oder mg/L, geschlossen. Das ist die bekannteste und einfachste Methode. Jedoch existieren viele Studien, die meinen, dass trotz der Anwendung in den unterschiedlichsten Gebieten mit einer linearen Umwandlung von mS/cm zu ppm, sich dieser Umwandlungsfaktor von der Art des Gewässers unterscheidet.

Die Studie von Thirumalini und Joseph (2009) beschäftigt sich mit dieser Frage, wie dieser Umwandlungsfaktor aufgebaut und anwendbar ist. Das Resultat zeigt auf, dass sich der Faktor bei vielen Tests von 0.59 bis 0.93 befand. Allerdings wird davon ausgegangen, dass in Süßwasser eine lineare Umwandlung mit Faktor 0.65 geeignet ist, währenddessen in Salzwasser diese Korrelation nicht-linear verläuft.

Es ist wichtig zu beachten, dass der Umwandlungsfaktor von 0.65 hauptsächlich für Gewässer ohne Verunreinigungen gilt. Bei Gewässern mit höherem Verschmutzungsanteil kann dieser Faktor zwischen 0.53 und 0.93 variieren.

Einfluss auf die Algenpopulation

Algen benötigen Nährstoffe, um optimales Wachstum erzielen zu können. Zu diesen Nährstoffen zählen Kobalt, Eisen, Magnesium, Nitrat und noch viele weitere Substanzen. Die Studie von Chen et al. (2011) bezog sich auf die Auswirkungen der einzelnen Substanzen auf die Algenpopulation. Dazu wurde, wie in Abschnitt [pH](#), die Algenpopulation mithilfe der [O.D.](#) gemessen und beobachtet. In der genannten Studie wird eine bestimmte Algenart, welche ebenfalls für andere Arten repräsentativ ist, in eine Flasche mit [Erd-schreiber's Medium](#) gefüllt. Die Lösung und die Algenpopulation wird täglich auf die Werte [O.D.](#), pH und Temperatur überprüft.

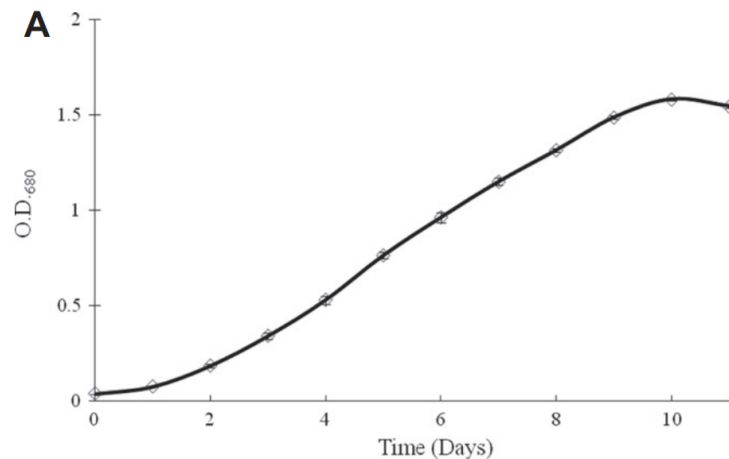


Abbildung 2.9: Wachsende Algenpopulation in [Erdschreiber's Medium](#) anhand der Algenart *Dunaliella tertiolecta*
 Quelle: Chen et al. (2011)

In der Abbildung 2.9 wird die Wachstumsrate der Algenpopulation in [Erdschreiber's Medium](#) veranschaulicht. Diese zeigt, dass die Wachstumsrate dieser Algenart an Tag 5 am Höchsten ist und ab Tag 10 und 11 wieder abnimmt. Unter allen Nährstoffen wurde Phosphat mit 84% Reduktion am Schnellsten abgebaut. Phosphat wird in sehr vielen Stoffwechselprozessen benötigt und ist einer der bedeutsamsten Minerale für Organismen, welches ebenfalls in einer TDS-Überwachung gemessen wird.

Daraufhin hat man versucht, die einzelnen Nähstoffe einzeln beseitigt, um somit eine Wachstumsdifferenz zu beobachten. Darin wurde sichtbar, dass das Weglassen von Eisen die meiste Reduzierung der Wachstumsrate zur Folge hatte, gefolgt von Kobalt und Magnesium. Dass Phosphat nicht in der Liste auftaucht, obwohl vorhin als sehr wichtig beschrieben, erklärt sich Chen et al. (2011), dass die Algenart während der Kultivierung Phosphor abgespeichert hatte, welches ja zuvor verabreicht und am Schnellsten verbraucht wurde.

Daher lässt sich schlussfolgern, dass die Nährstoffe, wie sie durch den TDS-Wert gemessen werden, direkte Auswirkungen auf das Algenwachstum haben und in der Lage sind, dieses zu beschleunigen.

2.5 ThingSpeak

Mit der zunehmenden Digitalisierung und dem Wachstum des [IoT](#) ist die Erfassung, Verarbeitung und Speicherung von Daten essentiell. Um dies so einfach wie möglich zu gestalten hat sich ThingSpeak etabliert, um den Kunden viel Arbeit abzunehmen.

2.5.1 Was ist ThingSpeak?

ThingSpeak ist eine webbasierte [IoT](#)-Plattform und wurde 2010 von MathWorks, einem Unternehmen für mathematische Berechnungssoftware, speziell für die Erfassung, Analyse und Visualisierung von Echtzeitdaten entwickelt. Es basiert auf einem offenen Quellcode und bietet eine Vielzahl von Funktionen und Erweiterungen für die Datenverarbeitung.

2.5.2 Funktionsweise

Die zentrale Speicherungsmöglichkeit von ThingSpeak sind sogenannte "Channels". Ein Channel ist wie ein Projekt aufgebaut, in das die Daten abgespeichert und visualisiert werden. Jeder Kanal kann verschiedene Felder enthalten, in denen spezifische Datenpunkte gespeichert werden. Benutzer können über eine REST-API Daten in ihre Kanäle schreiben oder aus ihnen lesen. Eine API, kurz für Application Programming Interface, definiert eine Schnittstelle zwischen Software-Komponenten und ermöglicht es ihnen, Funktionen und Dienste von anderen Software-Systemen effizient zu nutzen, ohne dass detailliertes Wissen über die interne Funktionsweise des jeweiligen Systems notwendig ist. Die Plattform unterstützt ebenfalls Echtzeit-Datenströme und erlaubt die Integration von externen Diensten über Webhooks.

2.5.3 Nutzung von ThingSpeak

ThingSpeak bietet schon in der kostenfreien Version viele Möglichkeiten an.

Datenvisualisierung

ThingSpeak bietet viele Möglichkeiten zur Visualisierung und Analyse von Daten. Der Nutzer kann in der Weboberfläche direkt Diagramme, Graphen und Dashboards erstellen, um die Daten zu visualisieren. Zu den Diagrammen zählen "Line, Bar, Column, Spline und StepDiagramme.

IoT-Datenüberwachung

ThingSpeak eignet sich ideal für die Überwachung von IoT-Geräten und deren Sensorik. Diese können sich nämlich nahtlos mit ThingSpeak verbinden, um Echtzeitdaten zu versenden. Zusammen mit der Datenvisualisierung ist eine Echtzeit-Überwachung der Messwerte möglich.

Skalierbarkeit

ThingSpeak lässt sich nahtlos mit anderen Anwendungen verbinden. Neue Messwerte können durch Hinzufügen weiterer Felder angelegt werden.

3 Methodik und Vorgehensweise

In diesem Kapitel wird die Methodik der Arbeit präsentiert, die sich auf mehrere Schlüsselbereiche konzentriert. Jeder dieser Bereiche wird in einem separaten Abschnitt behandelt, beginnend mit der Darstellung der verwendeten Datenquellen und der Vorgehensweise bei der Datenerfassung.

Im darauffolgenden Abschnitt wird der Prozess zur Schätzung des Chlorophyll-a-Gehalts detailliert beschrieben. Diese Schätzung ist von zentraler Bedeutung für die Zielsetzung dieser Arbeit, da sie das Algenwachstum indirekt abbildet und somit eine Grundlage für die Entwicklung eines Vorhersagemodells bietet. Im dritten Abschnitt wird die Validierung dieses Vorhersagemodells diskutiert und gezeigt, wie die Genauigkeit und Verlässlichkeit der Schätzungen überprüft werden kann. Anschließend wird die Planung und Implementierung des IoT-Systems und der App vorgestellt. Dieses Kapitel legt damit die methodischen Grundlagen der folgenden Arbeit und bietet einen Einblick in deren praktische Umsetzung.

3.1 Datenquellen und Datenerfassung

In diesem Abschnitt wird kurz erläutert, woher die Daten stammen, die sowohl für das Training des Vorhersagemodells, als auch für die Validierung verwendet werden.

3.1.1 Messung anhand des IoT-Systems

Die Daten, die zur Vorhersage der Chlorophyll-a Werte benötigt werden, stammen aus einem zu entwickelnden IoT-System. Es sollen mithilfe von Sensoren, die Temperatur-, pH- und Leitfähigkeitswerte gemessen und abgespeichert werden. Genauere Details der Sensoren befinden sich im Abschnitt 2.4.

Die für die Prognose der Chlorophyll-a-Werte benötigten Daten stammen aus einem zu entwickelnden IoT-System. Mit diesem System werden spezielle Sensoren eingesetzt, um die Temperatur-, pH- und Leitfähigkeitswerte zu erfassen und zu speichern. Eine ausführliche Beschreibung dieser Sensoren und ihrer Funktionsweise wurde im Abschnitt 2.4 detailliert beschrieben.

Die Messungen für diese Werte finden am gleichen Ort statt, an dem auch die Messungen von “Daten- Und Kartendienst Der LUBW” (2023) zur Validierung erhoben werden. Dieser Prozess wurde jedoch nur probeweise über einen Zeitraum von fünf Stunden durchgeführt.

3.1.2 Messwerte zum Training der Vorhersage

Um Vorhersagen über Chlorophyll-a-Werte zu treffen, sind Trainingsdaten erforderlich. Mithilfe von maschinellem Lernen können diese Daten dazu verwendet werden, Aussagen über den Chlorophyll-a-Gehalt auf der Grundlage abhängiger Werte zu treffen.

Diese Trainingsdaten basieren auf der Grundlage der *Daten aus dem Umweltinformationssystem (UIS) der LUBW Landesanstalt für Umwelt Baden-Württemberg* (“Daten- Und Kartendienst Der LUBW”, 2023). Diese Datenbank umfasst neben der Wasserqualität auch interessante Werte bezüglich Schadstoffen, dem Ausbau erneuerbaren Energien, Lärm, Geologie und Radioaktivität. Bezüglich der Wasserqualität sind in Baden-Württemberg an 26 Messstellen Sensoren platziert, um folgende Werte zu überwachen:

- T (Temperatur in Grad Celsius)
- pH (pH-Wert bei x Grad Celsius mit dimensionsloser Einheit)
- LF (Elektrische Leitfähigkeit bei x Grad Celsius in Mikrosiemens pro Zentimeter)
- O2 (Sauerstoff in Milligramm pro Liter)
- Trübungsquantifizierung (Trübung in Formazin-Nephelometrieinheit¹)
- Chlorophyll-a (Chlorophyll-a in Mikrogramm pro Liter)

In dieser Arbeit werden lediglich die Werte für Temperatur, pH, Leitfähigkeit und Chlorophyll-a-Masse berücksichtigt.

Eine Messstation, die genau diese 4 Messwerte erfasst, befindet sich in Mannheim am Neckar. Nach einer Filterung nach den benötigten Werten wird ein Datensatz zurückgegeben, welcher beispielsweise in Tabelle 3.1 dargestellt wird.

¹Nephelometrie beschreibt ein Verfahren zur Messung der Lichtstreuung durch Partikel in einer Flüssigkeit. Formazin dient derweil als Substanz zur Kalibrierung dessen.

Messstelle ↕	Ost ↕	Nord ↕	Parameter ↕	Datum ↕	Tagesmittelw... ↕	Dimensi... ↕
Mannheim, Neckar	461526	5482542	Chlorophyll a	29.05.2023	1,3	µg/l
Mannheim, Neckar	461526	5482542	Elektrische Leitfähigkeit bei ...Grad C...	29.05.2023	744	µS/cm
Mannheim, Neckar	461526	5482542	pH-Wert bei ...Grad Celsius	29.05.2023	8,1	-
Mannheim, Neckar	461526	5482542	Temperatur	29.05.2023	19	°C

Tabelle 3.1: Messwerte der Messstation in Mannheim anhand der gefilterten Werte
Quelle: “Daten- Und Kartendienst Der LUBW” (2023)

Diese Daten lassen sich auch aufwandslos als CSV exportieren und weiterverwenden. Die Nutzungsvereinbarung der “Daten- Und Kartendienst Der LUBW” (2023) schreibt lediglich eine Erwähnung der Quelle vor und bei manchen Datensätzen wird eine kommerzielle Nutzung nur per Anfrage genehmigt, welches sich allerdings nicht auf die Datensätze der Wasserqualität bezieht.

3.2 Schätzung der Chlorophyll-a Werte

Anhand der Trainingsdaten aus Kapitel 3.1.2 lässt sich die Masse an Chlorophyll-a grob schätzen.

Es wird eine multiple Regression verwendet, die aus den Trainingsdaten eine Funktion generiert, um eine abhängige Variable basierend auf unabhängigen Variablen (auch als Prädiktoren bezeichnet) zu ermitteln. Dazu werden die einzelnen Messwerte mit Regressionsgewichten verrechnet und miteinander verknüpft. Durch Einfügen der dieser Messwerte in die erhaltene Funktion kann ein Chlorophyll-a Wert abgeschätzt werden. Die errechneten Regressionsgewichte machen ebenfalls ersichtlich, wie groß die individuellen Beiträge der einzelnen Variablen sind und welche Rolle diese spielen.

3.3 Validierung der Vorhersagen

Um die Vorhersagegenauigkeit bewerten zu können soll der gesamte Datensatz in zwei separate Datensätze aufgeteilt werden. Mithilfe des Trainingsdatensatzes, welcher $\frac{4}{5}$ der gesamten Daten enthält, kann die multiple Regression entwickelt werden. Hierbei werden die Beziehungen und Gewichtungen der einzelnen unabhängigen Variablen (in diesem Fall: Temperatur, pH und Leitfähigkeit) festgelegt und die Bedeutung dieser Werte ermittelt.

Die Ergebnisse dieser Regression werden daraufhin mit dem Testdatensatz, welcher $\frac{1}{5}$ der gesamten Daten enthält, verglichen. Dazu werden die tatsächlichen Werte des Test-

datensatzes verwendet, um diese mit der Vorhersage zu vergleichen und dadurch die Leistung des Vorhersagemodells zu beurteilen.

Zusätzlich wird ein Bestimmtheitsmaß, auch R^2 genannt, ermittelt, um zu verstehen, wie groß die Einflüsse der einzelnen unabhängigen Variablen sind. Diese R^2 Werte werden daraufhin verwendet, um die Modelle zu vergleichen.

Sowohl für die Regression, als auch für die Validierung und die Darstellung wird die Programmiersprache **R** benutzt. Dies ist eine Open-Source-Programmiersprache, welche häufig für statistische Zwecke verwendet wird und großen Gebrauch in der Datenanalyse findet. Somit ist es möglich mit Skripten die vorhandenen Daten zu analysieren, anzupassen und darzustellen.

Als Entwicklungs-Plattform wird hier [RStudio](#) verwendet.

3.4 Implementierung des IoT-Systems

Vor der Implementierung des [IoT](#)-Systems muss zunächst ermittelt werden, welche Sensoren geeignet sind. Dieser Teil der Anforderungsanalyse wurde in Kapitel [2.4](#) ausführlich erläutert.

Darüber hinaus sollen die Sensoren mit dem Mikrocontroller per Breadboard verbunden und mithilfe geschriebener Software ausgelesen werden. Dabei wird berücksichtigt, welche Bibliotheken geeignet sind und welche Protokolle und Bussysteme verwendet werden. Die gemessenen Daten werden zyklisch per Hypertext Transfer Protocol ([HTTP](#)) an eine ausgewählte [Cloud](#)-Plattform gesendet und dort abgespeichert.

Die Software wird in einer vereinfachten Version der Programmiersprache C++ geschrieben. Als Entwicklungsplattform wird die [Arduino IDE 2.1.0](#) verwendet, um den Quellcode auf den Mikrocontroller zu laden.

3.5 Implementierung der App

Bevor es zur Programmierung der App geht, wird zuerst ein Softwaredesign mit Richtlinien erstellt, um eine saubere Datenhaltung und Übersichtlichkeit zu gewährleisten.

Nach der Ausarbeitung des Designs werden die einzelnen Seiten der App erstellt und mit Inhalten gefüllt. In diesem Zusammenhang soll die Anbindung an den [Cloud](#)-Dienst erörtert und implementiert werden, um die gemessenen Sensorwerte in der App darstellen zu können. Daraufhin folgt das State-Management, welches dafür sorgt, dass die Daten in der App aktuell gehalten werden und sich aktualisieren, wenn ein neuer Datensatz

verfügbar ist. In jeder Funktion der App wird auf die Usability geachtet und erläutert, wie die Nutzung der App so intuitiv und einfach wie möglich gestaltet werden kann.

Zur Implementierung wird das von Google verwaltete Framework [Flutter](#) verwendet, welches auf der Programmiersprache [Dart](#) basiert. Als Entwicklungsplattform wird [Visual Studio Code](#) verwendet.

4 Datensammlung- und vorbereitung

Dieses Kapitel konzentriert sich auf zwei entscheidende Aspekte der Arbeit: die Sammlung der Daten und dessen Vorbereitung. Diese grundlegenden Schritte sind das Fundament jeder Datenanalyse und stellen sicher, dass diese auf soliden, zuverlässigen und sauberen Daten basieren.

4.1 Sammeln der Trainingsdaten

Die Daten, die in der folgenden Arbeit zum Training des Vorhersagemodells verwendet werden, stammen aus der öffentlich zugänglichen Datenbank von “Daten- Und Kartendienst Der LUBW” ([2023](#)).

Die erforderlichen Messerwerte sind wie folgt zu finden -> *Wasser -> Oberflächengewässer -> Phys.-chem. Gewässergütedaten -> Daten der Messstationen*. Durch Eingabe der erforderlichen Parameter, wie Temperatur, pH, Leitfähigkeit und Chlorophyll-a lässt sich herausfiltern, welche Messstationen diese Werte zeitgleich liefern. Von allen 26 Messstationen ist die in Mannheim am Neckar als Einzige in der Lage alle vier Messwerte auszulesen. Sie beinhaltet auch den Sauerstoffgehalt, aber diese Messung lässt sich nicht kostengünstig in ein IoT-System integrieren und wird daher ausgelassen. Als Zeitraum wird der 01.06.2020 - 01.06.2023 genommen, um die letzten drei Jahre zu analysieren und anhand dessen die Vorhersage zu entwickeln.

Der Datensatz, welcher aus 1.095 Tagen besteht, kann entweder im .xlsx- oder im .csv-Format exportiert werden. Die Entscheidung fiel hier auf das klassische Excel-Format .xlsx, welches das Vorbereiten des Datensatzes vorerst einfacher macht. Es wird nun eine Datei ausgegeben, die inhaltlich beispielsweise in Abbildung [4.1](#) dargestellt wird.

Messstelle	=	Mannheim, Neckar CYY003 2855						
Parameter	=	758 Chloro-a Chlorophyll a Mikrogramm pro Liter						
	=	14 LF Elektrische Leitfähigkeit bei ...Grad Celsius Mikrosiemens pro						
	=	15 pH pH-Wert bei ...Grad Celsius dimensionslos						
	=	53 O2 Sauerstoff Milligramm pro Liter						
	=	4 T Temperatur Grad Celsius						
	=	8 Tr.quant Trübung, quantitativ Formazin-Nephelometrieinheiten						
Zeitraum	=	Mon Jun 01 00:00:00 CEST 2020 Thu Jun 01 23:59:59 CEST 2023						
Gewässer	GCODE	Messstelle Ost	Nord	Parameter	Datum	Tagesmitte	Dimension	
Neckar	CYY003	Mannheim	461526	5482542 Chlorophyl	01.06.2020		1.8 µg/l	
Neckar	CYY003	Mannheim	461526	5482542 Chlorophyl	02.06.2020		2.1 µg/l	
Neckar	CYY003	Mannheim	461526	5482542 Chlorophyl	03.06.2020		2.3 µg/l	
Neckar	CYY003	Mannheim	461526	5482542 Chlorophyl	04.06.2020		2.5 µg/l	
Neckar	CYY003	Mannheim	461526	5482542 Chlorophyl	05.06.2020		3.1 µg/l	
Neckar	CYY003	Mannheim	461526	5482542 Chlorophyl	06.06.2020		3.1 µg/l	

Abbildung 4.1: Rohe Datensätze des Exports der Messwerte der Messstation Mannheim am Neckar
Daten erhalten aus: “Daten- Und Kartendienst Der LUBW” (2023)

Es ist offensichtlich, dass der rohe Datensatz noch nicht für die folgende Analyse geeignet ist. Im folgenden Abschnitt wird erläutert, wie dieser Datensatz formatiert wird, um eine einfache Analyse zu gewährleisten.

4.2 Datenreinigung

Bevor es zur Formatierung der Daten geht sollte erst erkannt werden, welche Änderungen notwendig sind und wie diese umgesetzt werden können.

Es fällt auf, dass mehr Spalten vorhanden sind, als benötigt. Spalten, wie Gewässer, GCODE, Ost, Nord und mehr werden zur weiteren Analyse nicht von Bedeutung sein und sollten daher entfernt werden.

Der Datensatz ist ebenfalls so aufgebaut, dass die einzelnen Messwerte iterativ-inkrementell aufgebaut sind. In diesem Beispiel werden für jeden individuellen Messwert 1.095 Zeilen für jeden Tag angelegt. Das Endprodukt sollte allerdings so aussehen, dass lediglich 1.095 Zeilen existieren mit den jeweiligen Messwerten horizontal aufgestellt.

Um das zu erreichen werden die Spalten angelegt, die für die nachfolgende Arbeit von Bedeutung sind. Die folgende Tabelle sollte nun eine Struktur, wie in Tabelle 4.1 dargestellt, haben:

Datum	Temperatur	pH	Leitfähigkeit	Chlorophyll-a
-------	------------	----	---------------	---------------

Tabelle 4.1: Mögliche Datensatz-Struktur der Messwerte nach Bereinigung

Sobald die nicht notwendigen Eigenschaften gelöscht sind, werden die Daten in die Spalten eingetragen. Hier wurde entschieden, kein Skript zu schreiben, da der Aufwand, sich

in das Skript einzulesen, größer wäre als der Nutzen. Somit werden die einzelnen 1.095 Zeilen an Messwerten manuell kopiert und in die geeignete Spalte aus Tabelle 4.1 eingefügt.

Der Datensatz enthält allerdings immer noch Zeilen mit teilweise fehlenden Werten. Diese werden nicht separat entfernt, sondern in der Programmiersprache R bei der multiplen Regression nicht berücksichtigt. Das bedeutet, dass beispielsweise bei fehlender Temperatur eine Regression mithilfe dieser Zeile für die anderen unabhängigen Variablen weiterhin durchgeführt werden kann.

Als letzten Schritt wird der Datensatz schließlich in eine .csv-Datei umgewandelt, welche in Abbildung 4.2 veranschaulicht wird.

Datum,Temperatur,PH,Leitfaehigkeit,Chlorophyll.a				
6/1/2020	19.7	7.8	843	1.8
6/2/2020	20.4	7.8	849	1.2
6/3/2020	20.8	7.9	856	2.3
6/4/2020	20.4	8	873	2.5
6/5/2020	19.6	8.1	866	3.1
6/6/2020	18.9	8.2	862	3.1

Abbildung 4.2: Ausschnitt des verwendeten Datensatzes als CSV Datei

5 Entwicklung des Vorhersagemodells

Zur Vorhersage der Masse an Chlorophyll-a wird in dieser Arbeit eine multiple Regression verwendet.

Das Grundprinzip der multiplen Regression besteht darin, die Beziehung zwischen einer abhängigen Variablen (hier: Chlorophyll-a Masse) und mehreren unabhängigen Variablen, in diesem Fall: Temperatur, pH-Wert und Leitfähigkeit, zu modellieren.

Das Ziel dieser Modellierung ist es, eine mathematische Funktion zu ermitteln, die die beste Anpassung der Chlorophyll-a Masse an die anderen Messwerte liefert. Die Funktion wird als lineare Gleichung dargestellt, in der jedes Merkmal (unabhängige Variable) mit einem Koeffizienten (oder Gewicht) multipliziert und dann summiert wird. Die Schätzung dieser Koeffizienten erfolgt normalerweise durch Minimierung der Residuenquadratsumme, eine Methode, die auch als Methode der kleinsten Quadrate bekannt ist. Das Ergebnis der multiplen Regression ist ein Modell, das eine Vorhersage der Chlorophyll-a Masse basierend auf den gewichteten Messwerten ermöglicht. Es erlaubt auch die Bewertung der Signifikanz und Stärke des Einflusses jeder Messwerte auf die Chlorophyll-a Masse.

Im folgenden Kapitel wird die multiple Regression auf den vorbereiteten Datensatz aus Kapitel 4 angewandt. Im Folgenden wird dargestellt, wie eine multiple Regression in der Programmiersprache R angewendet wird und es wird die Bedeutung der Gewichtungen der einzelnen Merkmale hervorgehoben.

5.1 Erstellung der multiplen Regression

Der verwendete Datensatz ist mit Namen *FilledData_ThreeYears* abgespeichert und wird dementsprechend so importiert. Die Ausgabe aus Abbildung 5.1 zeigt, dass der Datensatz korrekt aufgebaut und bereit zur Analyse ist.

```
-> colnames(FilledData_ThreeYears)
[1] "Datum"           "Temperatur"      "pH"
     ↪ "Leitfaehigkeit" "Chlorophyll.a"
```

Abbildung 5.1: Eingelesene Spaltennamen in R aus dem importierten Datensatz

5.1.1 Teilung der Daten

Nun wird der Datensatz in zwei individuelle Datensätze aufgeteilt. Einmal dem Trainingsdatensatz, der verwendet wird, um das Vorhersagemodell zu trainieren und einmal dem Testdatensatz, um das erstellte Modell zu testen. Dafür wird zuerst ein Seed gesetzt, der von Zufallsgeneratoren als Startpunkt verwendet wird, um eine Reproduzierbarkeit der Analyse zu gewährleisten. Daraufhin werden zufällig Zeilen aus dem Datensatz ausgewählt und den verschiedenen Teildatensätzen zugeordnet. Somit teilen sich die Daten zu $\frac{4}{5}$ Trainingsdaten und $\frac{1}{5}$ Testdaten.

Die Teilung ist in Abbildung 5.2 dargestellt.

```
# Set the seed for replication purposes
set.seed(123)

# Then, randomly sampling from the full dataset to generate a
  ↳ training set and a test set

rs <- sample(1:dim(FilledData_ThreeYears)[1],size=round(dim(
  ↳ FilledData_ThreeYears)[1]/5),replace=F)

# Dividing data into two datasets

training.set <- FilledData_ThreeYears[-rs,] # Bigger to train
test.set <- FilledData_ThreeYears[rs,] # Smaller to test

# Sanity check: verifying the dimensions of both datasets

dim(training.set) # 4/5 of the total dataset
dim(test.set) # 1/5 of the total dataset
```

Abbildung 5.2: Teilung des Datensatzes in ein Trainings- und ein Testdatensatz mithilfe von zufällig ausgewählten Reihen im Datensatz

Nach der Teilung der Daten werden die jeweiligen Dimensionen in Abbildung 5.3 als Output ausgegeben:

```
> dim(training.set) # 4/5 of the total dataset
[1] 876    5
> dim(test.set) # 1/5 of the total dataset
[1] 219    5
```

Abbildung 5.3: Aufteilung des Datensatzes in Trainings- und Testdatensatz und Ausgabe der Anzahl an Spalten der jeweiligen Teile

Somit besteht der Trainingsdatensatz aus 876 und der Testdatensatz aus 219 zufällig ausgewählten Einträgen.

5.1.2 Anwendung der multiplen Regression

Um nun anhand der Trainingsdaten eine lineare Funktion zu ermitteln, werden die Daten in eine Funktion `lm()` für linear model eingegeben. Die Entwicklung des Vorhersagemodells wird in Abbildung 5.4 dargestellt.

```
# Perform linear regression analysis
resultsWithTemppHLeit <- lm(Chlorophyll.a ~ Temperatur + pH +
  ↳ Leitfaehigkeit, data = training.set)

# View the summary of the regression model
summary(resultsWithTemppHLeit)
```

Abbildung 5.4: Implementierung des Vorhersagemodells anhand der multiplen Regression mit Ausgabe der geschätzten Vorhersageergebnisse

Der Code erstellt ein lineares Regressionsmodell namens *resultsWithTemppHLeit* und beinhaltet die Verbindungen der Messwerte mit der Chlorophyll-a Masse. Dazu wurden die Regressionsgewichte für jeden einzelnen Messwert bestimmt und können nun für die weitere Analyse verwendet werden.

Die Ausgabe aus Abbildung 5.5 erfolgt mithilfe des `summary()` Commands.

```

Call:
lm(formula = Chlorophyll.a ~ Temperatur + pH +
    ↳ Leitfaehigkeit,
    data = trainingAllParameter.set)

Residuals:
      Min       1Q   Median       3Q      Max
-0.83185 -0.27508 -0.01782  0.26963  0.87664

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -1.053e+01  7.601e-01 -13.858  < 2e-16 *
    ↳ **
Temperatur    5.727e-02  2.152e-03  26.606  < 2e-16 *
    ↳ **
pH            1.329e+00  9.289e-02  14.312  < 2e-16 *
    ↳ **
Leitfaehigkeit 4.046e-04  1.361e-04   2.973  0.00306 *
    ↳ *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.'
    ↳ 0.1 ' ' 1

Residual standard error: 0.3559 on 687 degrees of
    ↳ freedom
Multiple R-squared:  0.5364,      Adjusted R-squared:
    ↳ 0.5344
F-statistic: 264.9 on 3 and 687 DF,  p-value: < 2.2e-16

```

Abbildung 5.5: Durchführung der linearen Regression mit Ausgabe der Vorhersageergebnisse

Der erste Teil der Ausgabe bezieht sich auf die Residuen und wird in [Abbildung 5.6](#) erneut dargestellt.

Residuals :

Min	1Q	Median	3Q	Max
-0.83185	-0.27508	-0.01782	0.26963	0.87664

Abbildung 5.6: Die Residuen der Ergebnisse der multiplen Regression

Residuen bezeichnen die Differenzen zwischen den beobachteten abhängigen Variablenwerten und den vom Regressionsmodell vorhergesagten Werten, also zeigen diese die Genauigkeit des Modells auf. Wenn das Modell perfekt passt und die Datensätze keinerlei Streuung zeigen ist hier ein Wert um 0 anzustreben. In der Praxis ist dies jedoch selten der Fall, und es wird immer eine gewisse Menge an Residuen erwartet.

Die ersten Werte, die für das Verständnis relevant sind, sind die **Min**- und **Max**-Werte. Diese stellen nämlich dar, in welchem Bereich die Vorhersage Fehler aufgezeigt hat. Die erstellte Funktion zeigt eine Abweichung von -1.83185 und 0.87664. Wenn der Datensatz nach Chlorophyll-a Masse gefiltert wird, sieht man, dass diese sich von 0.5 bis 4.2 µg/L bewegen. Also ist bei einer Varianz von 4.7 eine Abweichung von bis zu 0.8 relativ akzeptabel. Außerdem müssen für die komplette Analyse vor allem die **1Q**, **Median** und **3Q** Werte in Betrachtung gezogen werden. Diese sind ebenfalls wichtige statistische Eigenschaften und zeigen die Verteilung der Abweichungen und somit die Streuung der Daten.

Der **1Q**-Wert von -0.27508 zeigt, dass 25% der Residuen oder Abweichungen kleiner sind als -0.27508. Diese negative Abweichung deutet darauf hin, dass das Modell die Massen eher überschätzt.

Der **3Q**-Wert von 0.26963 zeigt, dass 75% der Residuen oder Abweichungen kleiner sind als 0.26963. Die positive Abweichung hingegen deutet darauf hin, dass das Modell die Masse eher unterschätzt.

Der **Median**-Wert stellt den mittleren Wert der Abweichungen dar, also in diesem Falle, dass rund 50% der Abweichungen einen Wert kleiner als -0.01782 hat und die andere Hälfte einen Wert größer als 0.01782. Der Median sollte bei einem geeigneten Modell dem Nullwert nahe sein. Da der Median in diesem Modell sehr nahe bei 0 liegt, kann dies darauf hinweisen, dass das Modell im Durchschnitt die Chlorophyll-a Masse gut vorhersagt.

Jetzt stellt sich die Frage, wieso die **Min**- und **Max**-Werte solche Streuung zeigen. Um diese Frage zu beantworten, soll in Kapitel [8.2](#) weiter ermittelt werden, wie es zu

solchen Abweichungen kommt und wie diese verringert werden könnten.

Der zweite Teil der Analyseergebnisse bezieht sich auf die Regressionskoeffizienten, die die Gewichtung jedes Merkmals in der Vorhersagegleichung repräsentieren und in Abbildung 5.7 dargestellt werden.

```

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -1.053e+01  7.601e-01 -13.858  < 2e-16 *
↪ **
Temperatur   5.727e-02  2.152e-03  26.606  < 2e-16 *
↪ **
pH           1.329e+00  9.289e-02  14.312  < 2e-16 *
↪ **
Leitfaehigkeit 4.046e-04  1.361e-04   2.973  0.00306 *
↪ *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.'
                 0.1 ' ' 1

```

Abbildung 5.7: Die Regressionskoeffizienten, die sich aus der multiplen Regression ergeben und die Gewichtungen der einzelnen Messwerte beschreiben

Die erste Zeile bezieht sich auf den Intercept, im Deutschen den y-Achsenabschnitt. Dieser wird hier auf -10.53 geschätzt und zeigt auf, dass wenn alle unabhängigen Variablen den Wert 0 haben, dass die Chlorophyll-a Masse auf -10.53 $\mu\text{g/L}$ geschätzt wird. Dies ist somit der Ausgangspunkt der Regressionsgeraden und beeinflusst die Koeffizienten. Es ist hier allerdings zu beachten, dass der Intercept nur eine Schätzung ist und nicht unbedingt eine direkte physikalische Bedeutung hat, zumal diese in diesem Falle auch sehr unrealistisch erscheint. Da die verwendeten Daten nur innerhalb eines Bereiches liegen, beispielsweise bei Temperatur von 2.5 bis 26 Grad Celcius, wird das Modell außerhalb des Bereiches über [extrapolieren](#) und unrealistische Vorhersagen machen.

Der Koeffizient der Temperatur zeigt an, dass bei einer Erhöhung der Temperatur um eine Einheit, ein Anstieg von 0.05727 in der geschätzten Chlorophyll-a Masse zu erwarten ist. Dieser Wert repräsentiert das Regressionsgewicht, das in der Berechnung der linearen

Regression mit dem Temperaturwert multipliziert wird. Der **t-Wert** von 26.606 deutet darauf hin, dass die Temperatur einen signifikanten Einfluss auf die Chlorophyll-a Masse hat.

Der **Std.Error** steht für Standardfehler und gibt an, wie stark die geschätzten Koeffizienten im Modell von den tatsächlichen Koeffizienten abweichen können. Ein kleinerer Standardfehler deutet darauf hin, dass der geschätzte Koeffizient genauer und zuverlässiger ist.

Der **Pr(< | t |)** Wert gibt an, ob der geschätzte Koeffizient statistisch signifikant von Null abweicht. Ein kleinerer p-Wert deutet darauf hin, dass der geschätzte Koeffizient wahrscheinlich nicht auf Zufall beruht und dass dieser eine signifikante Beziehung mit der Chlorophyll-a Masse aufweist. Der p-Wert wird hier mit einem Signifikanzniveau verglichen und mit Sternen bewertet.

3 Sterne zeigen, dass sich der p-Wert nahe 0 bewegt und deshalb eine wichtige Beziehung zur Chlorophyll-a Masse aufweist.

Nach den ermittelten Koeffizienten kann die lineare Gleichung nun aufgestellt werden.

$$Chlo = -10.53 + 0.05727 * Temp + 1.329 * pH + 0.0004046 * Leitf$$

Diese Gleichung kann verwendet werden, um die Chlorophyll-a Masse anhand der Messwerte beliebig zu schätzen.

Der letzte Part der Ausgabe in Abbildung 5.8 bezieht sich auf das Gesamtmodell und zeigt die Variabilität.

```
Residual standard error: 0.3559 on 687 degrees of
  ↪ freedom
Multiple R-squared:  0.5364,      Adjusted R-squared:
  ↪ 0.5344
F-statistic: 264.9 on 3 and 687 DF,  p-value: < 2.2e-16
```

Abbildung 5.8: Variabilität der entwickelten multiplen Regression

Der **Residual Standard Error (RSE)** gibt die durchschnittliche Abweichung der Werte an und sollte niedrig gehalten werden. Im Vergleich zu den Chlorophyll-a Werten im

Datensatz, die zwischen 0.5 und 4.2 liegen, ist ein **RSE** von 0.3559 relativ gering, was darauf hinweist, dass das Modell die Chlorophyll-a Masse gut vorhersagt.

Zusätzlich wurden 16 Zeilen nicht beachtet, da diese Lücken enthielten.

Der **R²** Wert gibt an, wie gut die verwendeten Messwerte die Chlorophyll-a Masse vorhersagen können. Dieser Wert wird im weiteren Verlauf zur Beurteilung der Signifikanz der einzelnen Messwerte verwendet.

Ein **R²** Wert von 0.5364 bedeutet, dass etwa 53.64% der Variabilität der Chlorophyll-a Masse durch die angegebenen Werte erklärt werden kann. Ein größerer **R²** Prozentwert ist besser für die Vorhersage und zeigt, dass das Modell gut an die Daten angepasst ist und diese wenig Streuung aufweisen. Allerdings ist ein Wert von 1 in der Praxis schwer nicht künstlich zu erreichen.

Ein Modell mit einem niedrigeren **R²**-Wert ist jedoch nicht zwangsläufig schlecht. Es zeigt lediglich eine gewisse Unvorhersehbarkeit in den Messwerten, was angesichts der Vielzahl an biologischen und chemischen Einflüssen, die die Chlorophyll-a Masse beeinflussen, durchaus sinnvoll ist.

Viele Studien können mit einem geringeren **R²** Wert hervorragend Variablen vorhersagen, solange die Einflusswerte eine signifikante Beziehung zur abhängigen Variable haben, was bei der weiter oben angegebenen Auswertung bewiesen wurde.

Die **F-Statistik** bewertet die Gesamteffizienz des Modells und überprüft ein weiteres Mal die Einflüsse der Messwerte auf die Chlorophyll-a Masse. In diesem Fall beträgt die F-Statistik 264.9 mit 3 unabhängigen Variablen und 856 Datenpunkten. Ein größerer Wert wird erreicht, wenn alle unabhängigen Variablen einen signifikanten Einfluss auf die abhängige Variable haben und zeigt auf, dass das Modell gute geeignet ist. In diesem Fall zeigt der F-Statistik Wert von 256.3, dass das Modell statistisch signifikant ist und die gegebenen Messwerte einen großen Einfluss auf die Chlorophyll-a Masse ausüben. Der p-Wert in diesem Kontext deutet ebenfalls darauf hin, dass die Einflüsse nicht auf Zufällen beruhen.

5.2 Validierung des Vorhersagemodells

Das Regressionsmodell wird nun, wie in Abbildung 5.9 gezeigt, auf den zweiten Datensatz angewandt, um zu überprüfen, wie akkurat die Schätzungen sind ob sich diese eignen.

```
# Predicting the DV in the test set using the intercept and
  ↳ coefficients from the training set
pout <- predict(resultsWithTemppHLeit,newdata=test.set)
```

Abbildung 5.9: Validierung der Vorhersage anhand des Testdatensatzes

Der Befehl `predict` verwendet das geschätzte Regressionsmodell, um den Chlorophyll-a Gehalt anhand der unabhängigen Variablen im Testdatensatz zu schätzen. Das Resultat wird in `pout` gespeichert und weiterverwendet.

Nun werden die tatsächlichen Werten von den vorhergesagten Werte subtrahiert, um eine Differenz zu erhalten. Diese Metrik wird auch Middle Absolute Error ([MAE](#)) genannt und sie gibt an, wie weit die durchschnittliche Vorhersage von den tatsächlichen Werten entfernt ist.

```
# Calculate the mean absolute error (MAE) between the predicted
  ↳ values (pout) and the actual values
mean(abs(pout-test.set$Chlorophyll.a),na.rm=T)
```

```
[1] 0.3120205
```

Abbildung 5.10: Berechnung des MAE anhand der Validierung zum Testdatensatz mit Ausgabe der Validierungsergebnisse

Als Output wird 0.3120205 ausgegeben, was den durchschnittlichen [MAE](#) darstellt. Ein MAE von 0.3120205 bedeutet, dass die durchschnittliche absolute Abweichung zwischen den vorhergesagten und den tatsächlichen Chlorophyll-a Werten etwa 0.31µg/L beträgt. Bei einer Reichweite von 0.5 bis 4.2 ist ein durchschnittlicher Fehler von 0.3120205 relativ gering.

In Kapitel [8.2](#) soll allerdings erläutert werden, wie eine solche Abweichung zustande kommt und wie diese verbessert werden kann.

In [Abbildung 5.11](#) wird veranschaulicht, wie das Vorhersagemodell in rot dargestellt, die Chlorophyll-a Werte anhand der einzelnen Messwerte schätzt. In Blau wird gezeigt, wie die Werte tatsächlich verlaufen.

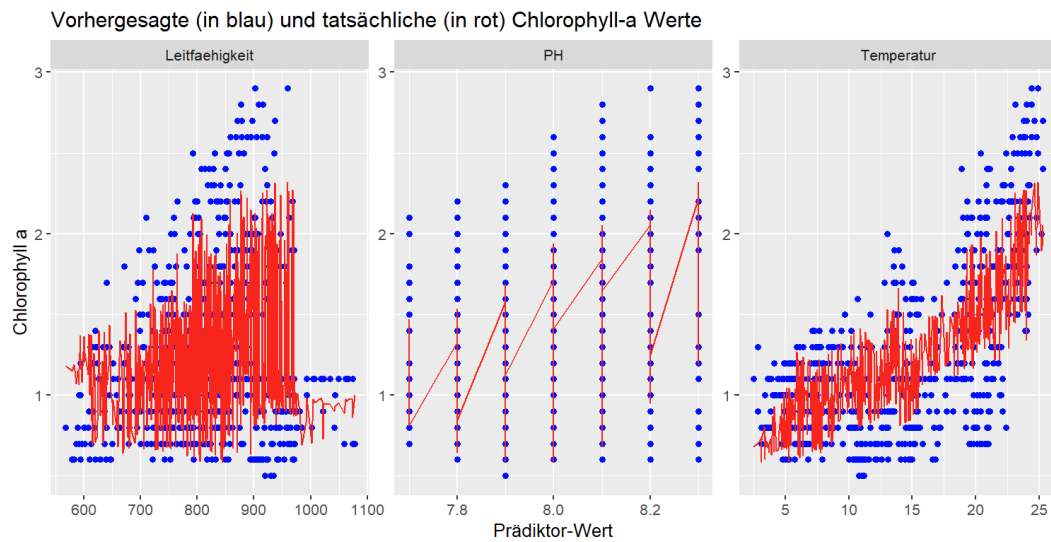


Abbildung 5.11: Visualisierung der geschätzten Chlorophyll-a Werte (in rot) mit den tatsächlichen Werten (in blau)

6 Entwicklung und Implementierung des IoT-Systems

Diese Arbeit konzentriert sich auf die Entwicklung eines IoT-Systems, das Messwerte aus einem Gewässer ausliest und mithilfe dieser Werte den Chlorophyll-a-Anteil schätzt. IoT bietet eine Möglichkeit Sensoren zu integrieren, die Daten zu verarbeiten und abzuspeichern, sodass eine kontinuierliche Überwachung der Wasserwerte von überall aus möglich ist. Diese Überwachung hilft auch dabei, potenzielle Probleme zu erkennen und zu entsprechenden Maßnahmen aufzurufen. Das Vorhersagemodell der multiplen Regression, das in Kapitel 5 entwickelt wurde, wird zum Schluss auf die gemessenen Wasserparameter angewandt.

Im Folgenden werden die einzelnen Schritte zur Entwicklung und Implementierung des IoT-Systems detailliert beschrieben. Dazu gehört die Anbindung der ausgewählten Sensoren, die Datenverarbeitung und -Speicherung und die Integration von Bibliotheken und Protokollen.

6.1 Auswahl der Technologien

In diesem Teil der Arbeit werden die drei Sensoren der Messwerte aus Kapitel 5 implementiert und für die weitere Verarbeitung verwendet.

Als Prozessor wird ein Arduino Nano 33IoT verwendet. Dieser wurde speziell für die IoT-Entwicklung hergestellt und beinhaltet ein Wi-Fi-Modul, welches eine mühelose Internetanbindung ermöglicht. Dieser Arduino ist ebenfalls sehr kompakt gehalten und lässt sich in kleinere Gehäuse einbauen.

Zur Entwicklung der Software wurde sich für die offizielle Arduino IDE 2.0 entschieden, was das Hochladen von Quellcode auf den Arduino in vereinfachtem C++ möglich macht.

Die Speicherung der Daten erfolgt durch den Cloud-Service [ThingSpeak](#). Dieser Dienst ermöglicht das Senden, Speichern und Analysieren von Messwerten, die ebenfalls zu jeder Zeit abrufbar sein sollen.

6.2 Systemarchitektur

Die Gesamtarchitektur ist in Schichten organisiert, wie in Abbildung 6.1 dargestellt.

In der untersten Schicht befinden sich die Sensoren, die mithilfe eines Breadboards und Kabeln mit dem Mikrocontroller verbunden sind. Diese Sensoren werden mithilfe von Arduino-Code ausgelesen. Dabei ist bei jedem Sensor ein anderes Protokoll oder ein anderer Bus implementiert. Die Hersteller der Sensoren bieten hierfür meist eine Bibliothek, die die Anbindung einfacher macht. Unter dem Kapitel 6.3 wird im Detail beschrieben, wie die Sensoren, sowohl angeschlossen, als auch per Software verbunden sind.

Die mittlere Schicht besteht aus der Mikrocontrollerlogik. Hier werden die Algorithmen ausgeführt, die die Messwerte auslesen und verarbeiten. Wie diese Algorithmen aussehen und welchen Nutzen diese haben werden ebenfalls in diesem Kapitel erklärt.

Als Backend wird ein Cloud-Service angebunden. Dieser nennt sich ThingSpeak und ist von den selben Herstellern von MatLab. Es ist speziell für IoT-Projekte entwickelt worden, die sich wenig mit Backend und Datenbanken auseinandersetzen. Hier ist es möglich bis zu 3 Millionen Nachrichten im Jahr zu senden. Diese Daten der für dieses Projekt kostenfreie Datenbank lassen sich ebenfalls gut analysieren und wieder auslesen, was die Anbindung zur entwickelten App aus Kapitel 7 leichter macht. In diesem Kapitel wird beschrieben, wie die ausgelesenen Daten aus dem Mikrocontroller an ThingSpeak gesendet und abgespeichert werden.

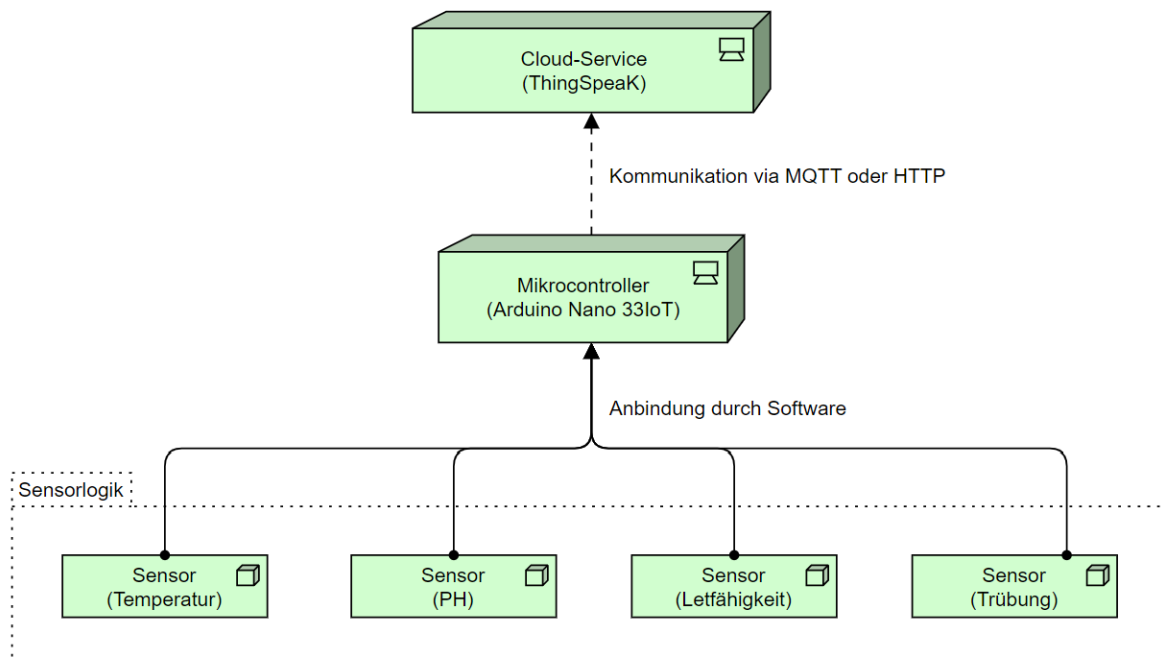


Abbildung 6.1: Geplante IoT-Systemarchitektur in Schichten dargestellt

6.3 Sensorintegration

In diesem Kapitel wird erklärt, wie die Sensoren an den Mikrocontroller angebunden sind und wie sich diese per Software ansteuern lassen.

6.3.1 Temperatur

Für die Messung der Temperatur wird ein DS18B20 Temperaturfühler verwendet, der auf die Funktionsweise eines Thermistors aufbaut und mithilfe des 1-Wire-Buss angesteuert wird.

Hardware

Abbildung 6.2 zeigt, wie der Temperatursensor angeschlossen wird. Der Sensor unterstützt mehrere Modi, die je nach Anwendungsfall unterschiedlich implementiert werden können.

Einmal dem **External Power Mode**, indem der Sensor mit 3 Anschlüssen versorgt wird: VDD(Spannung), DQ(Datenleitung) und GND(Ground). Hier wird die Stromversorgung über den VDD-Anschluss bereitgestellt, wobei über die DQ-Leitung lediglich

die Datenkommunikation erfolgt.

Im **Parasitic Mode** hingegen wird der Sensor über die Datenleitung mit Strom versorgt. Somit braucht diese nur zwei Leitungen für DQ(Datenleitung) und GND(Ground).

Für die folgende Arbeit wird allerdings der External Power Mode bevorzugt, da im parasitären Modus die Spannung variieren kann, insbesondere wenn mehrere Sensoren angeschlossen sind. Der Mikrocontroller wird vorerst per externer Stromquelle betrieben, soll aber in der Testphase auch in Batteriebetrieb laufen können. Zumal ist der External Power Mode Standard und kann ohne weiteren Quellcode und Anpassungen genutzt werden.

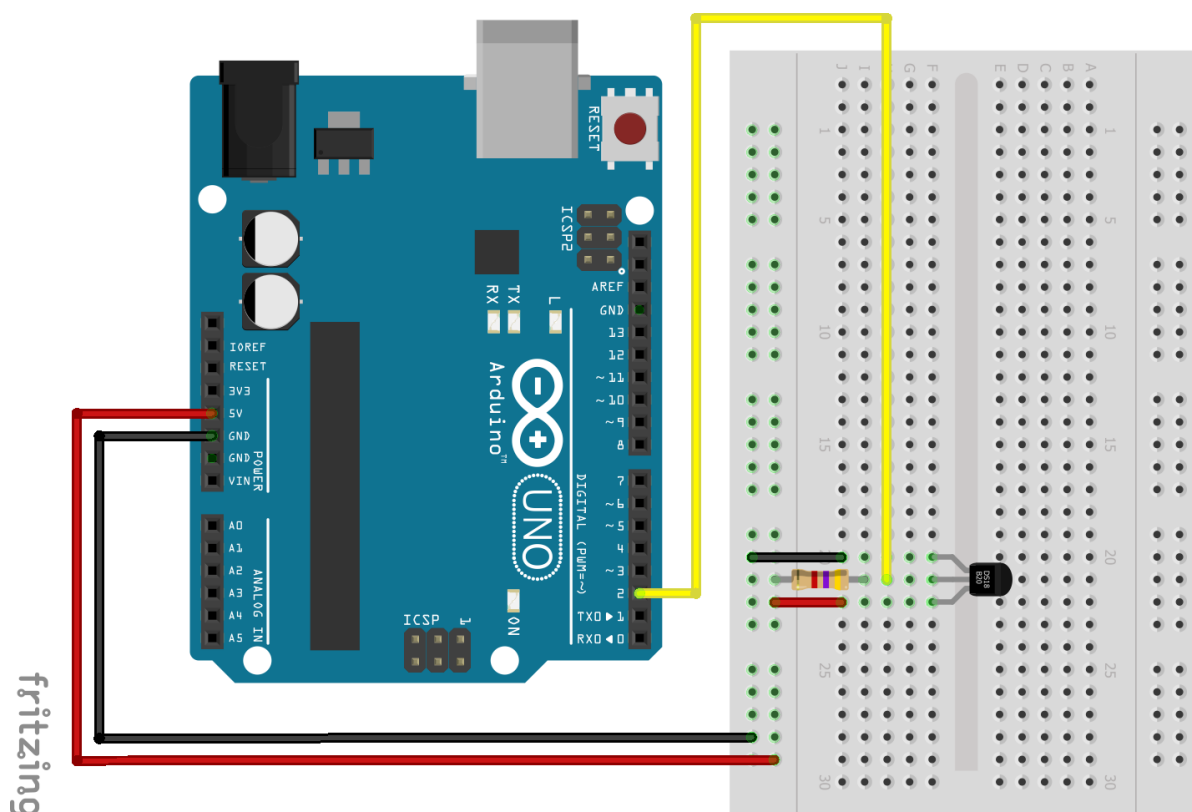


Abbildung 6.2: Schaltplan des DS18B20 mit Anbindung zum Arduino im Single External Power Mode

Quelle: Hansen (2023)

In Kapitel 2.4.1 der theoretischen Grundlagen wurde bereits erläutert, wieso der DS18B20 einen Pullup-Widerstand benötigt. Dieser besitzt einen Widerstand von $4.7\text{ k}\Omega$ und wird vor dem Sensor parallel geschaltet.

Software

Bevor die Temperatur ausgelesen werden kann, müssen die erforderlichen Bibliotheken geladen und Objekte zur weiteren Verwendung erzeugt werden.

Die **OneWire.h**-Bibliothek ermöglicht die Kommunikation mit dem DS18B20-Sensor über den oben genannten One-Wire-Bus.

Die **DallasTemperature.h**-Bibliothek ermöglicht wiederum das Auslesen der Daten des DS18B20-Sensors.

In allen folgenden Initialisierungen werden Konstanten für statische Werte benutzt, um dessen Wiederverwendbarkeit zu gewährleisten und Redundanzen zu vermeiden.

Die Setup-Methode wird einmal bei Systemstart aufgerufen und enthält alle Initialisierungen, die der Mikrocontroller in seinen Zyklen braucht.

Nach der Initialisierung des Sensors kann die Temperatur in mehreren Einheiten, wie Fahrenheit, Celsius oder auch als Spannungswert ausgelesen werden.

Der gemessene Wert kann nun in eine Variable abgespeichert und für weitere Algorithmen verwendet werden.

6.3.2 PH

Die pH-Sonde ist der teuerste von allen verwendeten Sensoren, da diese zusätzlich einen BNC-Connector benötigt, um die Werte auslesen zu können.

Hardware

Die gesamte Hardware besteht aus zwei Teilen. Einmal der pH-Sonde, die die Spannungsdifferenz misst und weitergibt und einmal der BNC-Anschluss mit zugehöriger Platine, um die gemessenen Werte umzuwandeln und auslesen zu lassen. Die Sonde wird auf den BNC-Anschluss aufgeschraubt, sodass nur die Anbindung zur BNC-Platine zu implementieren ist.

Die erworbene BNC-Platine hat die Anschlüsse schon so zusammengefasst, dass diese nur 3 Verbindungen braucht: VDD(Spannung), DQ(Datenleitung) und GND(Ground). In Abbildung 6.3 ist gezeigt, wie die pH-Sonde aussieht und ebenfalls die drei benötigten Verbindungen.

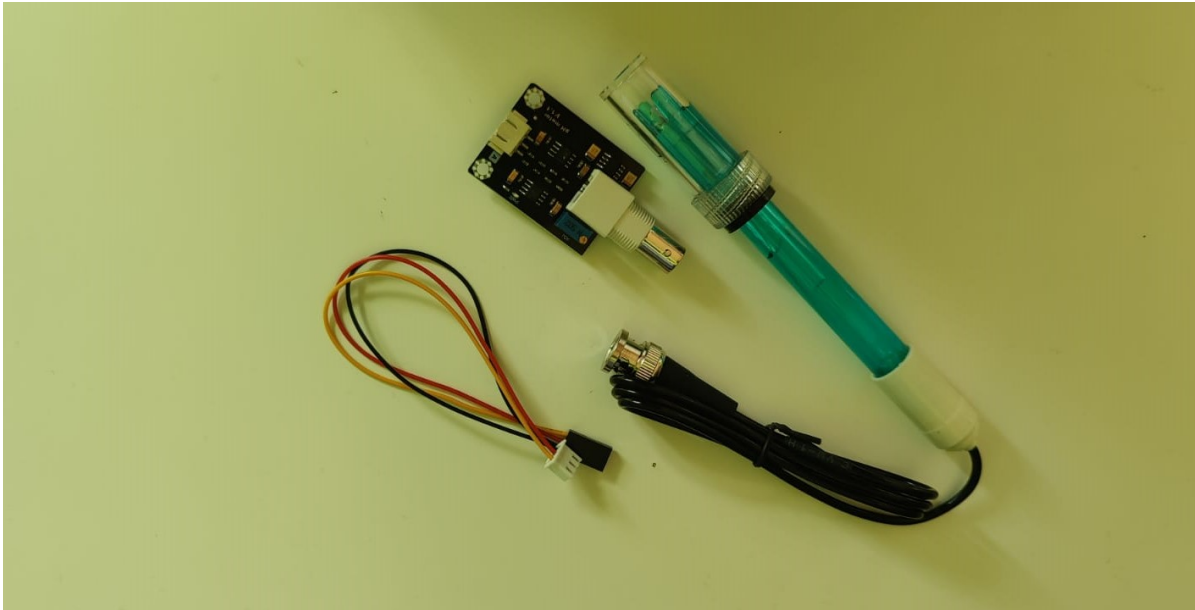


Abbildung 6.3: Verbaute pH-Sonde mit separatem BNC-Connector

Nachdem die Datenleitung an PIN 0 des Arduinos und die Spannungsversorgung angeschlossen wurden, kann nun der pH-Wert ausgelesen werden.

Software

Da die meiste Umwandlung und Verarbeitung der Messwerte schon auf der BNC-Platine stattfindet, ist das Auslesen durch Arduino-Code mühelos. Wie auch bei der Temperatur wird zuerst eine PIN-Konstante festgelegt und diese in der setup-Methode als INPUT definiert. Was noch bleibt ist das zyklische analoge Auslesen des PINS und den gemessenen Wert für die Weiterverarbeitung abzuspeichern.

Kalibrierung

Die pH-Sonde misst die elektrische Spannung, die durch den Unterschied in der Wasserstoffionenkonzentration zwischen der Probe und der internen Referenzelektrode erzeugt wird. Diese Spannung wird dann in den jeweiligen pH-Wert umgerechnet. Mithilfe einer Kalibrierung kann sichergestellt werden, dass diese Werte so genau und zuverlässig wie möglich sind.

Hierfür werden drei Referenzlösungen mit bekannten pH-Werten benötigt. Zuerst wird die pH-Sonde in eine Lösung mit einem pH-Wert von 4 eingetaucht und der analoge Spannungswert notiert. Dies wird nun für die Referenzlösung mit pH-Wert 9.18 wiederholt. Wenn nun die beiden Spannungswerte notiert wurden, werden sie linear skaliert,

sodass sich eine lineare Funktion bildet. In Abbildung 6.4 ist die lineare Gleichung dargestellt, die sich aus den beiden gemessenen Werten erschließt.

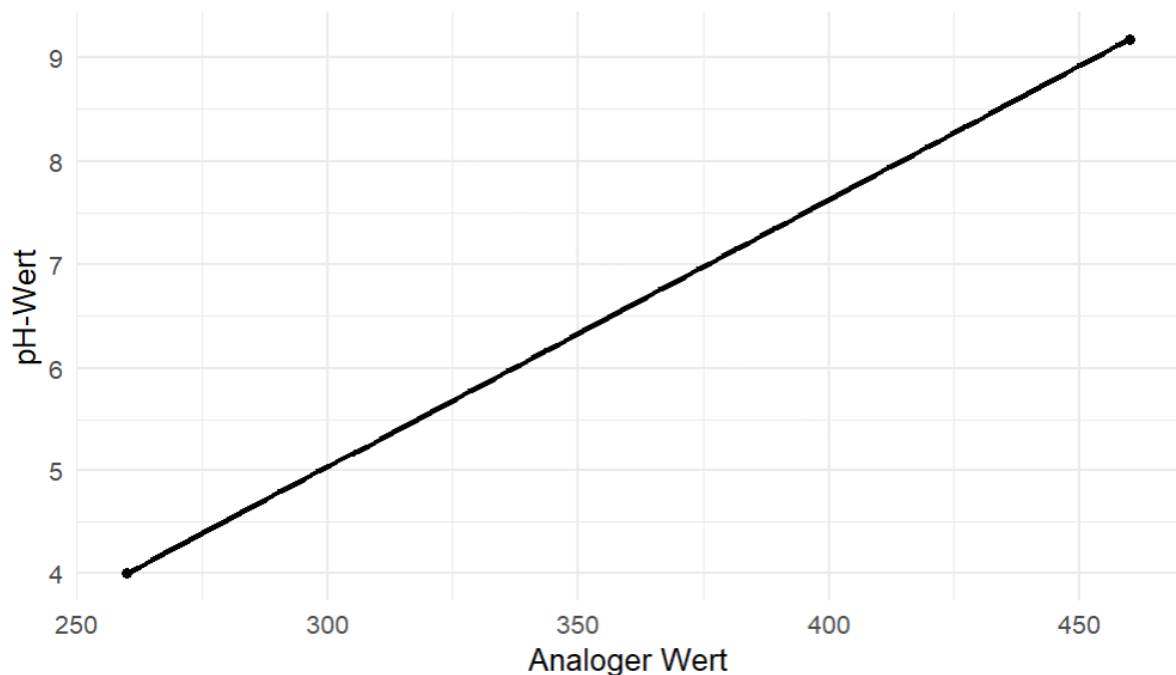


Abbildung 6.4: Lineare Gleichungsdarstellung der Kalibrierung von pH-Werten nach analogen Spannungswerten

Durch Einsetzen von neuen Spannungswerten kann somit der pH-Wert ermittelt werden. Dieser Test wird auf eine Referenzlösung mit pH-Wert 6 angewandt. Das Ergebnis ist ein Spannungswert von 337.2201, welcher ziemlich genau dem aus der linearen Gleichung entspricht.

Um nun jeden gemessenen Wert zu remappen, wird eine Funktion, wie in Abbildung 6.5 benötigt, welche genau dies ausrechnet:

```
float mapFloat(float x, float in_min, float in_max, float out_min,
    ↪ float out_max)
{
    return (x - in_min) * (out_max - out_min) / (in_max - in_min) +
        ↪ out_min;
}
```

Abbildung 6.5: Methode zur Umrechnung der Spannungswerte in einen pH-Bereich

Die Funktion **mapFloat** ist eine Erweiterung der Standardfunktion **map**, welche die Umwandlung von Gleitkommazahlen ermöglicht. Sie verwendet eine lineare Interpolation, um den Wert auf den angegebenen Ausgangsbereich abzubilden. Im folgenden Teil soll erläutert werden, wie sich die genannte Formel aufbaut.

- $(x - in_{min})$: Berechnet den Abstand zwischen dem Eingangswert x und dem unteren Wert des analogen Spannungsbereiches
- $(out_{max} - out_{min})$: Berechnet den maximalen Unterschied zwischen dem kleinsten und größten Wert des pH-Bereiches
- $(in_{max} - in_{min})$: Berechnet den maximalen Unterschied zwischen dem kleinsten und größten Wert des Spannungsbereiches
- $(x - in_{min}) * (out_{max} - out_{min})$: Skaliert den verschobenen Wert anhand der beiden Wertebereichen
- $(x - in_{min}) * (out_{max} - out_{min}) / (in_{max} - in_{min})$: Normalisiert den verschobenen und nun skaliert Wert auf einen Bereich von 0 bis 1
- $(x - in_{min}) * (out_{max} - out_{min}) / (in_{max} - in_{min}) + out_{min}$: Zum Schluss wird der kleinste Wert des pH-Bereiches aufaddiert, um den Wert auf den gewünschten Ausgangsbereich zu skalieren

6.3.3 TDS

Für die Messung der TDS-Werte wird ein gewöhnlicher TDS-Meter mit zwei Elektroden verwendet.

Hardware

Die Anbindung der Hardware erfolgt üblicherweise über drei Kabelverbindungen: VDD(Spannung), DQ(Datenleitung) und GND(Ground). Über die Datenleitung lässt sich ein analoger Spannungswert auslesen, welcher in einen TDS-Wert umgewandelt wird.

Software

Um einen korrekten TDS-Wert zu erhalten, muss der gemessene analoge Spannungswert an die Temperatur angepasst werden. Da die Temperatur schwanken kann, wird ein Algorithmus benötigt, welcher diese Schwankungen ausgleicht und die Werte anpasst.

Zur Angleichung werden 24°C eingesetzt, welche sich allerdings auch durch einen eingebauten Temperatursensor ersetzen lassen. Dieser Algorithmus wird in Abbildung 6.6 dargestellt.

```
// Read the analog value and convert to voltage value
averageVoltage = analogRead(TDS_PIN) * (float)VREF / 1024.0;

// Temperature compensation formula: FinalResult(25C) = FinalResult(
    ↳ current)/(1.0+0.02*(fTP-25.0));
float compensationCoefficient = 1.0+0.02*(temperature-25.0);
float compensationVoltage=averageVoltage/compensationCoefficient;
```

Abbildung 6.6: Algorithmus zum Ausgleichen der Temperaturschwankungen in den Messwerten

Die erste Code-Zeile rechnet den analogen Spannungswert auf eine Voltspannung um. Daraufhin wird der Kompensationskoeffizient, basierend auf der aktuellen Temperatur (hier: 24°C) berechnet, welcher verwendet wird, um den gemessenen Wert zu korrigieren und abzugleichen.

Nun kann der Spannungswert, wie in Abbildung 6.7 dargestellt, in den TDS-Gehalt umgewandelt werden. Dafür wird die Formel samt Koeffizienten aus der Anleitung von “Arduino with TDS Sensor” (2023) verwendet.

```
// Convert voltage value to TDS value
tdsValue=(133.42*compensationVoltage*compensationVoltage*
    ↳ compensationVoltage - 255.86*compensationVoltage*
    ↳ compensationVoltage + 857.39*compensationVoltage)*0.5;
```

Abbildung 6.7: Umrechnung der Spannungswerte in einen TDS-Bereich anhand der Koeffizienten

Zur Überprüfung der Genauigkeit wurden die gemessenen und ausgerechneten Werte mit einem handelsüblichen TDS-Apparat verglichen. Die Werte zeigen eine maximale Abweichung von 10 Einheiten zum tatsächlichen Wert auf, was angesichts einer Spanne von 50 bis 500 eine gute Genauigkeit aufweist.

6.4 Verarbeitung der Sensordaten

Die im vorherigen Kapitel gemessenen Daten werden nun verarbeitet und an das Backend gesendet.

Es kann jedoch vorkommen, dass die Sensoren Fehlmessungen liefern, was zu Ausreißern führt, die beispielsweise physikalisch unmöglich sind. Um das zu verhindern, wird ein Algorithmus verwendet, welcher nicht alle 30 Minuten die Daten ausliest und diese weitersendet, sondern alle 30 Minuten ein Durchschnittswert aus 100 einzelnen Messungen berechnet und diesen Durchschnitt für jeden Messwert versendet.

Dieser Algorithmus berechnet den **gleitenden Durchschnitt**. Dieser Durchschnittswert berechnet sich aus einer bestimmten Anzahl an Datenpunkten über einen bestimmten Zeitraum. Dabei wird kontinuierlich ein neuer Durchschnitt berechnet, indem der älteste Wert entfernt und der neueste Wert hinzugefügt wird. In diesem Fall verläuft das Verfahren wie folgt:

1. Messung 1 bis Messung 100: Durchschnittsberechnung dieser 100 Werte
2. Messung 2 bis Messung 101: Verschiebung des Messfensters um eine Messung nach hinten und erneute Berechnung der 100 Werte
3. Messung 3 bis Messung 103: Wie Schritt 2
4. Messung 4 bis Messung 104: Wie Schritt 3
5. ...

Somit glättet sich der Durchschnittswert über einen gewissen Zeitraum. In [Abbildung 6.8](#) wird gezeigt, wie sich der gleitende Durchschnitt im Verlauf der jeweiligen 100 Messungen bewegt. Die Kurve des gleitenden Durchschnitts ist geglättet und verhindert große Ausreißer in den Datenmengen.

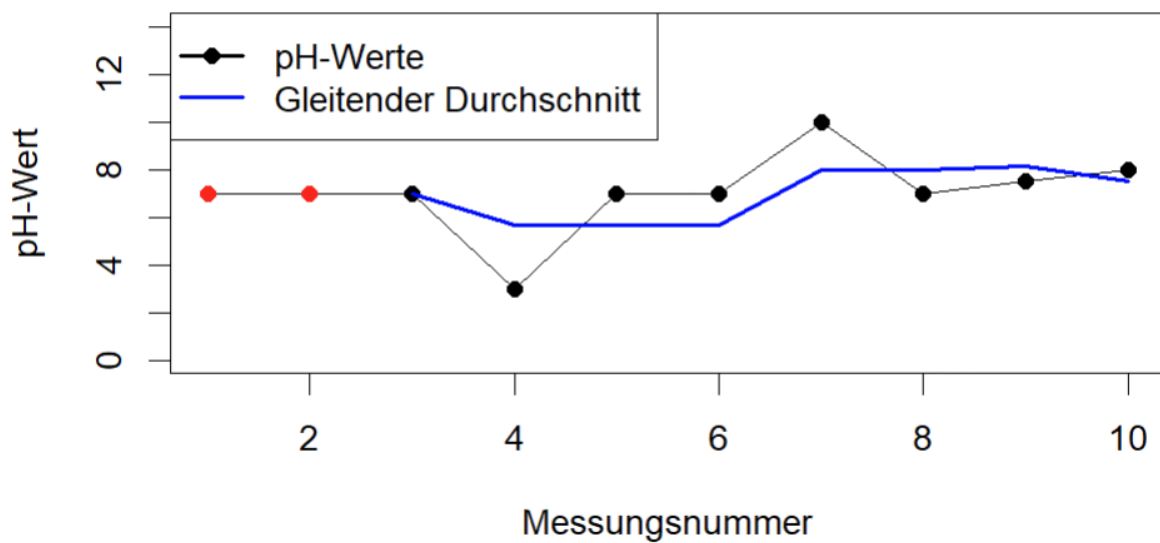


Abbildung 6.8: Kurve des gleitenden Durchschnitts über die jeweiligen 100 Messungen im Vergleich zu den gemessenen pH-Werten, die Ausreißer enthalten

6.5 ThingSpeak

Zur Datenspeicherung wurde ThingSpeak ausgewählt, da diese Cloud-Plattform kostenfrei die Speicherung und Verarbeitung der Messwerte ermöglicht und bereits positive Erfahrungen damit gemacht wurden.

6.5.1 Datenspeicherung

Um die Messwerte abzuspeichern, muss erst ein ThingSpeak-Account erstellt werden. Wenn dies getan ist gibt es die Möglichkeit Kanäle zu erstellen, die als Projekte dienen, die mehrere Messwerte abspeichern können. Hier können die einzelnen Messwerte ebenfalls visualisiert und per Weboberfläche angezeigt werden.

Channel Stats

Created: [3 months ago](#)
 Last entry: [8 minutes ago](#)
 Entries: 19994

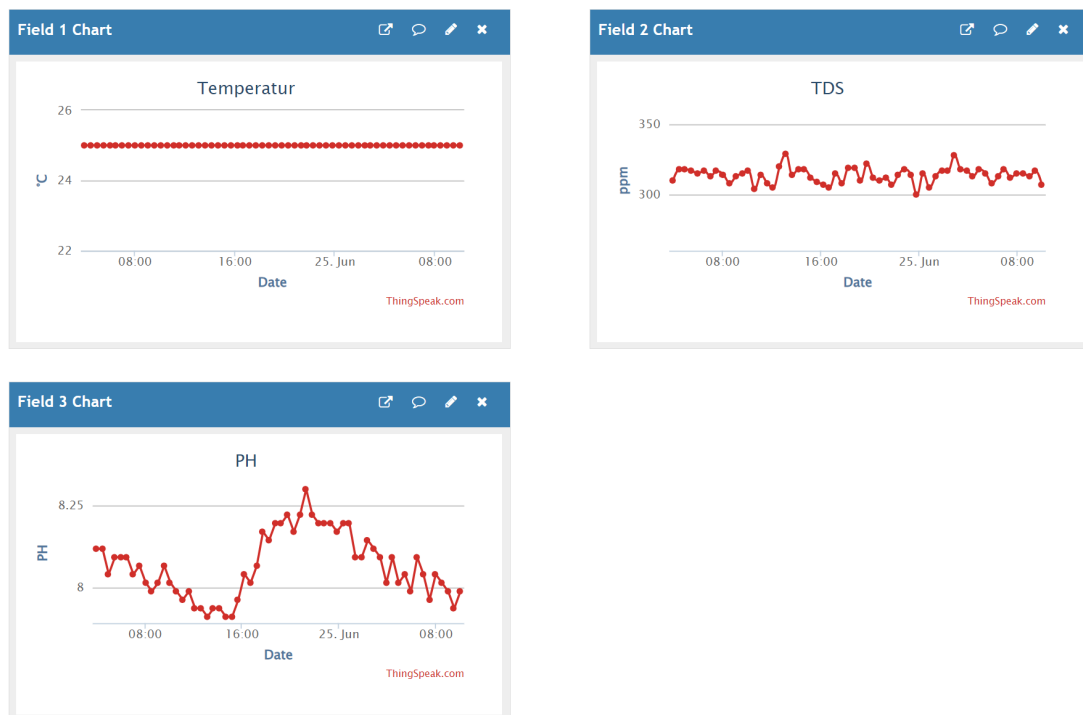


Abbildung 6.9: Erstellter Kanal mit knapp 20.000 Einträgen von gemessenen Temperatur-, pH- und TDS-Werten

Der erstellte Kanal in 6.9 beinhaltet knapp 20.000 Einträge für bereits gemessene Temperatur-, pH- und TDS-Werten. Diese Werte können unterschiedlich visualisiert werden, wobei hier die klassischen Spline-Diagramme verwendet werden. Da hier allerdings eine eigene App für die Visualisierung der Daten entwickelt wird, wird die Weboberfläche nicht weiter verwendet. Die gespeicherten Daten können ebenfalls wieder per API extern abgerufen werden.

6.5.2 Datenübertragung

ThingSpeak unterstützt Datenübertragungsprotokolle wie HTTP oder MQTT, um die gemessenen Daten zu senden. Für die Implementierung der Datenübertragung wird HTTP verwendet, da die offizielle Bibliothek der Hersteller ("ThingSpeak Communication Library for Arduino, ESP8266 and ESP32", 2023) eine einfache Schnittstelle hierfür anbietet. Jeder ThingSpeak-Kanal besitzt einen API-Key zum Schreiben und Lesen der

Daten. Sowohl die Kanalnummer, als die API-Keys werden in einer separaten Datei abgespeichert und als Header importiert. Zur Initialisierung wird ebenfalls der WifiClient benötigt, um eine Internetverbindung herzustellen.

Das Senden der Daten erfolgt über ein Multi-Field-Update, bei dem die einzelnen Felder festgelegt und dann als gesamtes Paket versandt werden. Somit reicht eine API-Anfrage aus und die Daten werden ebenfalls gleichzeitig aufgenommen. Die Implementierung für das Schreiben und Abschicken der Felder wird in Abbildung 6.10 gezeigt.

```
// Set read data to their respective ThingSpeak field
ThingSpeak.setField(1, getMedianNum(temp, READINGSCOUNT));
ThingSpeak.setField(2, getMedianNum(tds, READINGSCOUNT));
ThingSpeak.setField(3, mapFloat(getMedianNum(ph, READINGSCOUNT),
    ↪ 260.0, 460.0, 4.0, 9.18));

// Send all set fields to ThingSpeak and look out if it succeeded
int resultCode = ThingSpeak.writeFields(myChannelNumber,
    ↪ myWriteAPIKey);
```

Abbildung 6.10: Methode zum Setzen der ThingSpeak-Felder im angelegten Kanal und das Senden dieser Daten

Die Felder werden von 1 auf inkrementiert und spiegeln die Felder im ThingSpeak-Kanal wieder. Nach dem Setzen der Messwerte in die jeweiligen Felder können diese nun als Einzelpaket abgeschickt werden. Hierfür wird die Kanalnummer und der Schreib-API-Key benötigt.

Als Rückgabe erfolgt ein Resultcode, welcher wie in Tabelle 6.1 gezeigt aussehen kann und den Status des Sendens beschreibt.

Value	Meaning
200	OK / Success
404	Incorrect API key (or invalid ThingSpeak server address)
-101	Value is out of range or string is too long (> 255 characters)
-201	Invalid field number specified
-210	setField() was not called before writeFields()
-301	Failed to connect to ThingSpeak
-302	Unexpected failure during write to ThingSpeak
-303	Unable to parse response
-304	Timeout waiting for server to respond
-401	Point was not inserted (most probable cause is the rate limit of once every 15 seconds)
0	Other error

Tabelle 6.1: ThingSpeak Resultcodes nach Senden über die das multi-field Update der offiziellen Bibliothek

Quelle: “ThingSpeak Communication Library for Arduino, ESP8266 and ESP32” (2023)

7 Entwicklung der App als Nutzerschnittstelle

Dieses Kapitel konzentriert sich auf die Entwicklung einer Anwendung, die zur Überwachung von Algenblüten dient. Die Anwendung ist so konzipiert, dass sie sowohl die erfassten Daten in Echtzeit anzeigt, als auch die Ergebnisse der Vorhersage darstellt. Dieses Kapitel beschreibt die Schritte, die unternommen wurden, um die Anwendung von der Konzeption bis zur Implementierung zu entwickeln.

Der erste Abschnitt dieses Kapitels befasst sich mit der Anforderungsanalyse. Durch eine detaillierte Anforderungsanalyse wird festgelegt, welche Funktionen die Anwendung beinhalten sollte und wie sie aufgebaut sein sollte, um den Anforderungen der Benutzer gerecht zu werden.

Der nächste Abschnitt widmet sich der Auswahl der Technologien und der Planung. Im folgenden Abschnitt wird die Echtzeitanzeige der Anwendung vorgestellt. Diese Funktion ermöglicht es den Benutzern, die erfassten Daten in Echtzeit zu sehen, was für die Überwachung und die rechtzeitige Identifizierung potenzieller Probleme von relevanter Bedeutung ist. Der letzte Abschnitt dieses Kapitels befasst sich mit der Datenvisualisierung in der Anwendung.

7.1 Anforderungsanalyse

Die folgenden Anforderungen unterscheiden sich zwischen funktionalen und nicht-funktionalen Anforderungen. Diese sollen beschreiben, welche Funktion die Applikation in dieser Arbeit erfüllt.

7.1.1 Funktionale Anforderungen

Funktionale Anforderungen beziehen sich hier auf die spezifischen Funktionen der App und wie diese agieren soll, um die Funktionen zu erfüllen.

Echtzeitdaten-Anzeige

Die App muss in der Lage sein, die Daten in Echtzeit von ThingSpeak zu erhalten und anzuzeigen. Dazu soll eine Funktion zur Anzeige der Aktualität der Daten bereitgestellt werden. Hierfür wird angezeigt, wie lange die letzte Aktualisierung her ist und ebenfalls, wie viele Einträge bislang existieren.

Datenvisualisierung

Die Messwerte sollen auf einer zusätzlichen Seite in höherem Detail visualisiert und analysiert werden. Diese Seite hat die Aufgabe, weitere Trends und Muster in den Daten zu erkennen und aufzuzeigen.

Vorhersagemodell

Das entwickelte [GAM](#)-Vorhersagemodell aus Kapitel 5 wird hier verwendet, um anhand der erhaltenen Messwerte Schätzungen von der Biomasse an Chlorophyll-a zu erhalten. Diese Vorhersage soll in einer zusätzlichen Seite der App intuitiv und leicht verständlich angezeigt werden.

7.1.2 Nicht-Funktionale Anforderungen

Nicht-funktionale Anforderungen hingegen beziehen sich auf das allgemeine Verhalten, wie beispielsweise die Benutzerfreundlichkeit, Leistung und Verfügbarkeit. Diese Anforderungen sollen sicherstellen, dass die App ihre Funktionen mit der erwarteten Leistung erfüllt.

Benutzerfreundlichkeit

Die App soll eine intuitive Benutzeroberfläche haben, um eine einfache Navigation und Verwendung zu gewährleisten. Alle relevanten Informationen sollten direkt und auf einen Blick über die Startseite sichtbar sein.

Performance

Die App soll trotz großer Menge an Daten diese verarbeiten und darstellen, ohne dass es zu Verzögerungen oder größeren Ladezeiten kommt. Die Echtzeitanzeige und Aktualisierung der Daten soll reibungslos und ohne zusätzliche Ladesequenzen verlaufen.

Zuverlässigkeit

Die App sollte robust sein und in der Lage sein, den kontinuierlichen Fluss von Daten zu verarbeiten und korrekte Vorhersagen zu liefern.

7.2 Auswahl der Technologien

Für die Entwicklung der App wird das Flutter-Framework mit Programmiersprache Dart verwendet.

Zum Einen bietet Flutter durch die Hot-Reload Funktion eine höhere Produktivität und Entwicklungsgeschwindigkeit. Dadurch können Änderungen am Code vorgenommen und gleichzeitig das Resultat in der App angezeigt werden, ohne diese neu zu starten.

Zum Anderen bietet Flutter eine umfangreiche Sammlung an Widgets für die Erstellung der Benutzeroberfläche. Diese Widgets können zu jeder Zeit importiert werden und ermöglichen die intuitive und ansprechende Bedienung der Applikation.

Flutter ist ebenfalls ein Cross-Plattform Framework, was bedeutet, dass sich die Applikationen sowohl auf Android, als auch auf IOS laden lässt und mit den bereitgestellten Designs immer nativ aussieht.

Zuletzt besitzt Flutter eine große und immer-wachsende Community mit vielen Ressourcen, die bei der Entwicklung helfen können.

7.3 Echtzeitübertragung

Um alle Daten dauerhaft aktuell zu halten wird eine streambasierte Datenübertragung implementiert. Im Gegenzug zur paketbasierten Datenübertragung werden die Daten als kontinuierlichen Stream, also fließend und ohne Unterbrechung übertragen. Somit müssen nicht immer eine große Anzahl an Paketen übertragen und nach Aktualität überprüft werden.

Im Code-Ausschnitt aus Abbildung 7.1 wird gezeigt, wie eine streambasierte Übertragung in Flutter verläuft. Dazu wird die **Stream.periodic**-Funktion verwendet, die per Parameter jede Sekunde das Netzwerk nach Neuerungen überprüft. Daraufhin wird per HTTP eine REST-Anfrage an den erstellten ThingSpeak-Kanal mit dem Write-API-Schlüssel gestellt. Wenn die Anfrage erfolgreich ist (HTTP-Statuscode 200), wird das JSON-Objekt aus dem body der Server-Antwort umgewandelt und in die nächste Methode gegeben, welche die Daten weiterverarbeitet.

```

Stream<List> fetchMeasurements() =>
    Stream.periodic(Duration(seconds: 1)).asyncMap(
        (event) async {
            final response = await http.get(Uri.parse(
                'https://api.thingspeak.com/channels/[Channelnumber]/
                ↪ feeds.json?api_key=[Write-API-Key]&results=10'
            ))
            );

            if (response.statusCode == 200) {
                // If the server did return a 200 OK response,
                // then parse the JSON.
                return getData(jsonDecode(response.body));
            } else {
                // If the server did not return a 200 OK response,
                // then throw an exception.
                throw Exception('Failed to load data');
            }
        },
    );

```

Abbildung 7.1: Implementierung einer streambasierten Datenübertragung anhand von HTTP

Nachdem die Daten kontinuierlich nach Aktualität überprüft und wenn nötig neu übertragen werden, müssen diese nun für die Anzeige weiterverarbeitet werden. Neben ein paar Eigenschaften, wie in `json["channel"]` der Kanalname, das Erstell- und Updatedatum des Kanals und dessen Daten, werden im JSON-Objekt auch unter `json["feeds"]` die Daten weitergegeben. Aus diesen Daten wird eine Liste an Messwerten erstellt, die nun problemlos für die Datenanzeige verwendet werden kann.

```
List getData(Map<String, dynamic> json) {
    List<dynamic> feeds = json['feeds'];
    List<SingleMeasurement> data = <SingleMeasurement>[];
    ChannelInfo info = ChannelInfo(
        json["channel"]["name"],
        json["channel"]["created_at"],
        json["channel"]["updated_at"],
        json["channel"]["last_entry_id"]);
    for (Map map in feeds) {
        data.add(
            SingleMeasurement(
                double.parse(map["field1"]),
                double.parse(map["field2"]),
                double.parse(map["field3"]),
                map["created_at"],
                map["entry_id"],
            ),
        );
    }
    return [
        data,
        info,
    ];
}
```

Abbildung 7.2: Parsing der HTTP-erhaltenen Daten in Objekte, die für die Visualisierung verwendet werden können

7.4 Datenvisualisierung

In diesem Abschnitt wird anhand von Beschreibungen und Screenshots gezeigt, wie die Daten, einfach zu verstehen, visualisiert werden.

7.4.1 Seitenmenü

Die jeweiligen Seiten lassen sich über ein Menü steuern, welches entweder durch das Menü-Icon oben-links oder Streichen nach rechts öffnen lässt. Das Menü enthält die Seiten:

- Dashboard: Anzeige der notwendigsten Informationen auf einen Blick
- Measurements: Detaillierte Ansicht der einzelnen Messwerte mit zusätzlicher Diagramme über benutzerdefinierte Zeitspanne
- Chlorophyll-a: Anzeige der geschätzten Chlorophyll-a Werten anhand des Vorhersagemodells
- Info: Weitere Anzeige von grundlegenden Informationen
- Notifications: Einstellung von Benachrichtigungen bei Überschreiten von Schwellenwerten
- Settings: App-Einstellungen

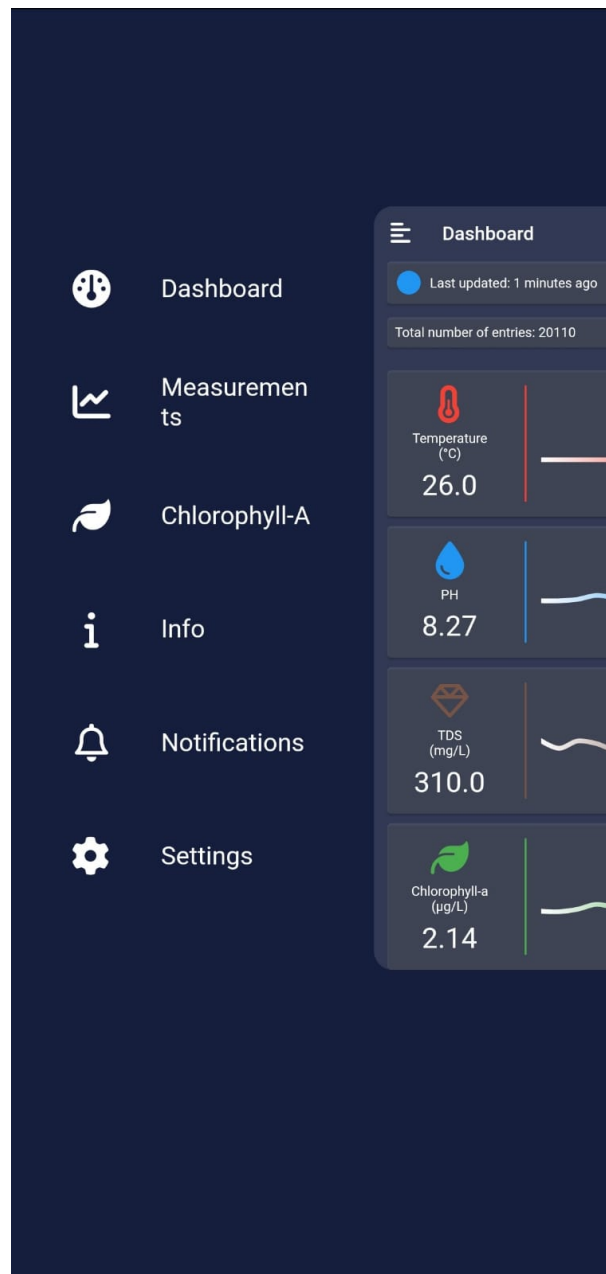


Abbildung 7.3: Menü zum Durchschalten der jeweiligen Seiten, geöffnet durch das Menü-Icon oder per Swipe

7.4.2 Startseite (Dashboard)

Das Dashboard der App ist so gestaltet, dass sie auf effiziente und übersichtliche Weise alle nötigen Informationen über die Messwerte liefert. Diese Daten werden in Echtzeit und im historischen Verlauf der letzten Stunden angezeigt, wodurch Benutzer die Möglichkeit

haben, Trends und Muster in den Messwerten leicht und auf einen Blick zu erkennen. Die Darstellung der Messwerte erfolgt in Form eines intuitiven Spline-Diagramms. Am oberen Rand der Startseite finden die Benutzer zusätzliche wichtige Informationen. Eine davon ist der Status der Datenübertragung. Hier wird durch die Verwendung von Farbcodes der Zustand des Systems dargestellt: Blau bedeutet, dass das System einwandfrei läuft und Daten ordnungsgemäß empfängt, während Rot darauf hinweist, dass das System aktuell keine neuen Daten erhält. Zusätzlich zur Übertragungsstatusanzeige wird auch die Aktualisierungszeit der letzten Daten angezeigt. Dies ermöglicht den Benutzern, das Zeitintervall seit der letzten Aktualisierung auf einen Blick zu erfassen und eine bessere Kontrolle über die Aktualität der erhobenen Daten zu haben. Ein weiteres nützliches Feature auf der Startseite ist ein speziell dafür vorgesehener Bereich, der die Gesamtzahl der bisher aufgenommenen Messwerte anzeigt. Dieses Element gibt den Benutzern einen Überblick über die Menge an gesammelten Daten und bietet so einen Mehrwert durch den Kontext der bisherigen Datenerfassung.

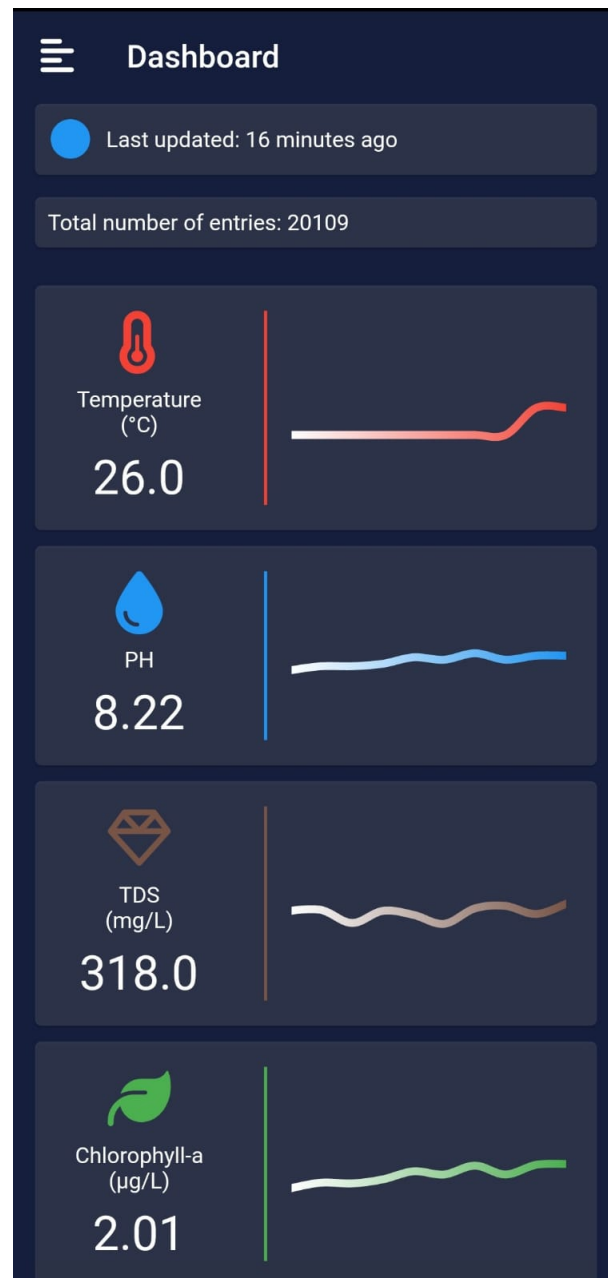


Abbildung 7.4: Bei App-Start angezeigtes Dashboard zeigt Messwerte und Eigenschaften an

Als letzten Block wird die Vorhersage aus der multiplen Regression angewandt, um anhand der einzelnen Messwerte direkt eine Schätzung an Chlorophyll-a Biomasse zu erstellen. Hierfür werden die Koeffizienten aus der Vorhersage genommen und mit den zugehörigen Messwerten multipliziert.

Um TDS in mg/L in $\mu\text{S}/\text{cm}$ umzuwandeln, wird der Umwandlungsfaktor von 1.56 ver-

wendet.

```
double estimateChlorophyllABiomass(  
    double temperature, double pH, double conductivity) {  
    double estimatedBiomass = -12.19 +  
        0.06515 * temperature +  
        1.629 * pH +  
        0.0005526 * 1.56 * conductivity;  
  
    return double.parse(estimatedBiomass.toStringAsFixed(2));  
}
```

Abbildung 7.5: Methode zur Schätzung der Chlorophyll-a Biomasse anhand der Koeffizienten aus dem entwickeltem Vorhersagemodell

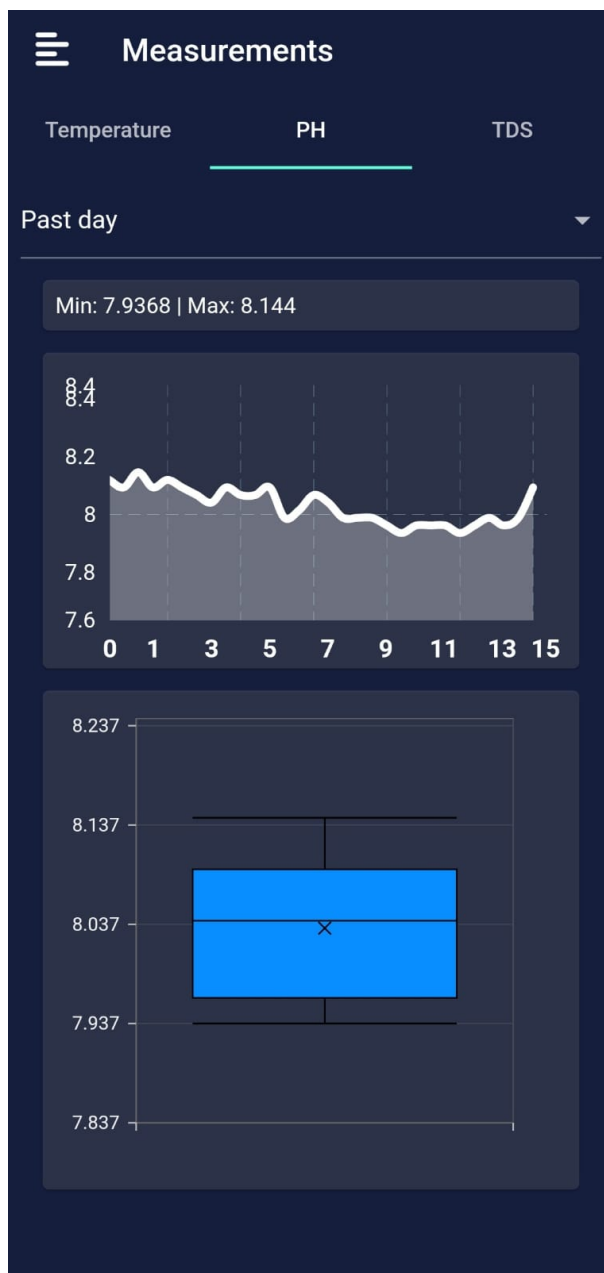
7.4.3 Messwert-Details

Die separate Detail-Seite der App ist speziell darauf ausgerichtet, die Messwerte in einem detaillierteren Kontext zu präsentieren. Sie erlaubt den Benutzern, detaillierte Eigenschaften der Daten zu gewinnen, die auf der Startseite möglicherweise nicht sofort ersichtlich sind.

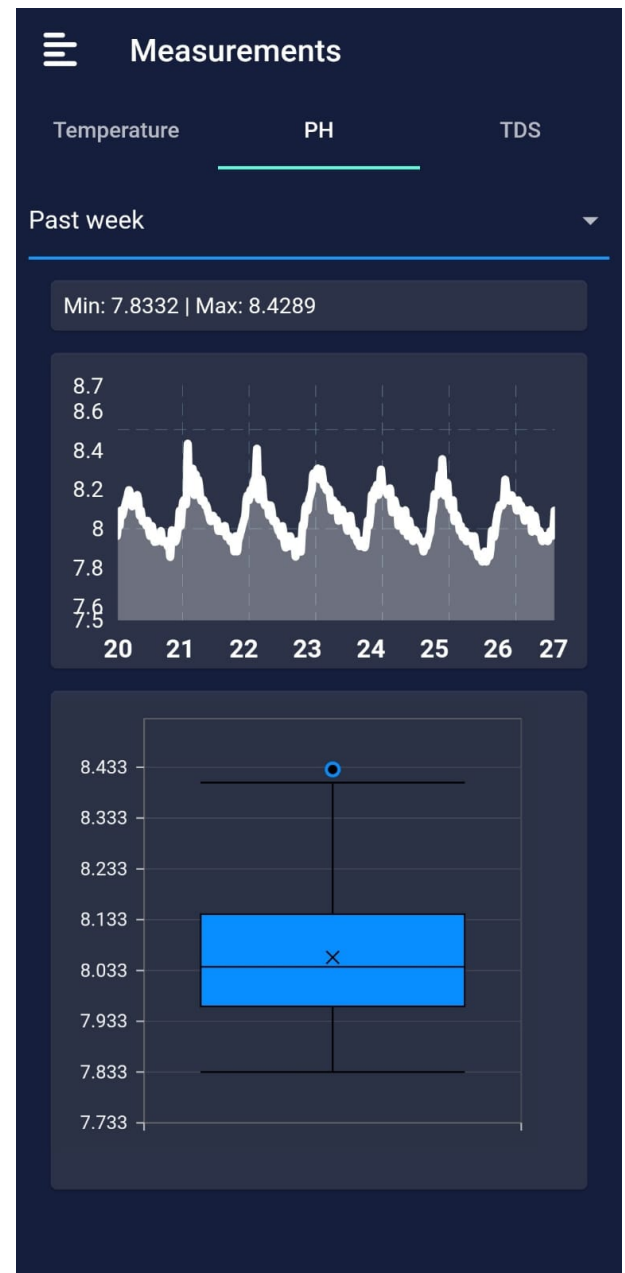
Oben auf der Seite gibt es eine Funktion, die den Nutzern die Möglichkeit bietet, den Zeitrahmen für die Visualisierung der Daten anzupassen. Dies kann entweder der aktuelle Tag oder ein Zeitraum der letzten sieben Tage sein. Diese Anpassbarkeit ermöglicht eine größere Flexibilität und detailliertere Kontrolle über die Darstellung der Daten.

Die Darstellung der Daten beginnt mit der Anzeige der minimalen und maximalen Werte des jeweiligen Messwerts. Diese Werte bieten einen Überblick über die Spannweite der Daten und liefern wichtige Eckpunkte zur Bewertung der Datenlage. Im Anschluss daran wird ein Spline-Diagramm angezeigt, das die Messwerte innerhalb des definierten Zeitraums visualisiert. Dieses Diagramm bietet eine visuelle Repräsentation der Daten, die Trends, Muster und Abweichungen auf einen Blick sichtbar macht. Unterhalb des Spline-Diagramms werden die Daten in Form eines Boxplots dargestellt. Ein Boxplot, auch bekannt als Box-Whisker-Plot, bietet eine visuelle Darstellung der statistischen Fünf-Punkte-Zusammenfassung der Daten: Minimum, erstes Quartil (unteres Quartil), Median (Zweites Quartil), drittes Quartil (oberes Quartil) und Maximum. Das untere und obere Quartil des Boxplots repräsentieren jeweils 25% der Daten, und der Kasten in

der Mitte des Plots enthält 50% der Daten - das sogenannte Interquartilbereich. Der Median, dargestellt durch eine Linie innerhalb des Kastens, teilt die Daten in zwei Hälften. Der Whisker oder die Linien, die sich von der Box erstrecken, zeigen die Variabilität außerhalb des Interquartilbereichs, und mögliche Ausreißer werden oft als einzelne Punkte dargestellt. Durch die Visualisierung dieser statistischen Merkmale ermöglicht der Boxplot eine umfassende und kompakte Ansicht der Verteilung der Daten. Dadurch können Informationen über die Zentralität, Streuung und Symmetrie der Daten sowie mögliche Ausreißer gewonnen werden.



(a) pH Messwerte des jetzigen Tages



(b) pH Messwerte der jetzigen 7 Tage

Abbildung 7.6: Detaillierte Messwertansicht

8 Evaluierung der Genauigkeit

Dieses Kapitel zielt darauf ab, die Wirksamkeit und Zuverlässigkeit des entwickelten Vorhersagemodells und des implementierten IoT-Systems zu beurteilen. Die Evaluierung ist ein wichtiger Schritt, um sicherzustellen, dass die Lösung effektiv funktioniert und die gestellten Ziele erreicht. Zuerst wird der Test des Vorhersagemodells mit Hilfe von eigen-gemessenen Werten vorgestellt. Dabei wird die Fähigkeit des Modells, den Chlorophyll-a Gehalt zu schätzen, anhand des entwickelten IoT-System bewertet. Nach der Schätzung wird die Validierung der Ergebnisse durchgeführt. Dies ist ein wesentlicher Schritt, um die Genauigkeit und Verlässlichkeit der vom Vorhersagemodell gelieferten Ergebnisse zu bestätigen. Hierbei werden die geschätzten Werte mit den tatsächlich gemessenen Werten verglichen, um zu prüfen, ob das Modell die Realität korrekt abbildet. Schließlich wird das Kapitel mit einer Diskussion über mögliche Ursachen für eine Abweichung der Schätzungen abgeschlossen.

8.1 Test des Vorhersagemodells anhand eigen-gemessener Werte

Dieser Teil konzentriert sich auf einen weiteren Test des Vorhersagemodells zur Schätzung des Chlorophyll-a Wertes. Diesmal wird das entwickelte IoT-System aus Kapitel 6 verwendet, um Messwerte zu sammeln. Die gemessenen Parameter umfassen die Wassertemperatur, den pH-Wert und den TDS-Wert. Die Messungen erfolgten an einem Samstag über einen Zeitraum von fünf Stunden hinweg und befanden sich in umliegender Reichweite zu den Sensoren der “Daten- Und Kartendienst Der LUBW” (2023). Die Messwerte wurden alle 20 Minuten erfasst und zu stündlichen Durchschnittswerten verrechnet.

Um eine Schätzung des Chlorophyll-a-Gehalts für jede Stunde zu erhalten, wird die mathematische lineare Gleichung aus Kapitel 5 verwendet:

$$Chlo = -10.53 + 0.05727 * Temp + 1.329 * pH + 0.0004046 * TDS$$

8.1.1 Schätzung der stündlichen Chlorophyll-a Werte

Durch die Eingabe der gemessenen Stundenwerte in die Formel wird nun die Schätzung durchgeführt. Als Beispiel für 15:00 Uhr wird folgende Berechnung durchgeführt:

Für den Term der Temperatur wird $0.05727 * 24.0 (= 1.37448)$ berechnet.

Für den Term des pH-Wertes wird $1.329 * 7.7 (= 10.2333)$ berechnet.

Für den Term der TDS wird $0.0004046 * 790 (= 0.319638)$ berechnet.

Durch Addition der einzelnen Terme mit dem Intercept wird eine Chlorophyll-a Schätzung von 1.397418 µg/L erhalten.

Diese Berechnung wird nun ebenfalls auf die Messwerte der anderen vier Durchschnittswerten angewendet und in Abbildung 8.1 dargestellt.

Uhrzeit	Temperatur	PH	TDS	Chlorophyll-a
15:00	24.0	7.7	790	1.397418
16:00	23.9	7.8	812	1.5335902
17:00	23.9	7.8	813	1.5340988
18:00	23.9	7.8	817	1.5356262
19:00	23.9	7.8	832	1.5416642

Tabelle 8.1: Stündliche gemessene Durchschnittswerte mit Schätzung des Chlorophyll-a Gehaltes jeder Stunde

8.1.2 Validierung der Schätzung

Zur Validierung der Schätzung werden die Daten des Messtages aus “Daten- Und Kartendienst Der LUBW” (2023) mit den Messungen verglichen. Die Messwerte, die zur Validierung verwendet werden, wurden ebenfalls im Zeitintervall von 15:00 - 19:00 Uhr erfasst. In Abbildung 8.1 werden die Messunterschiede der einzelnen Parameter und der Schätzung aufgezeigt.

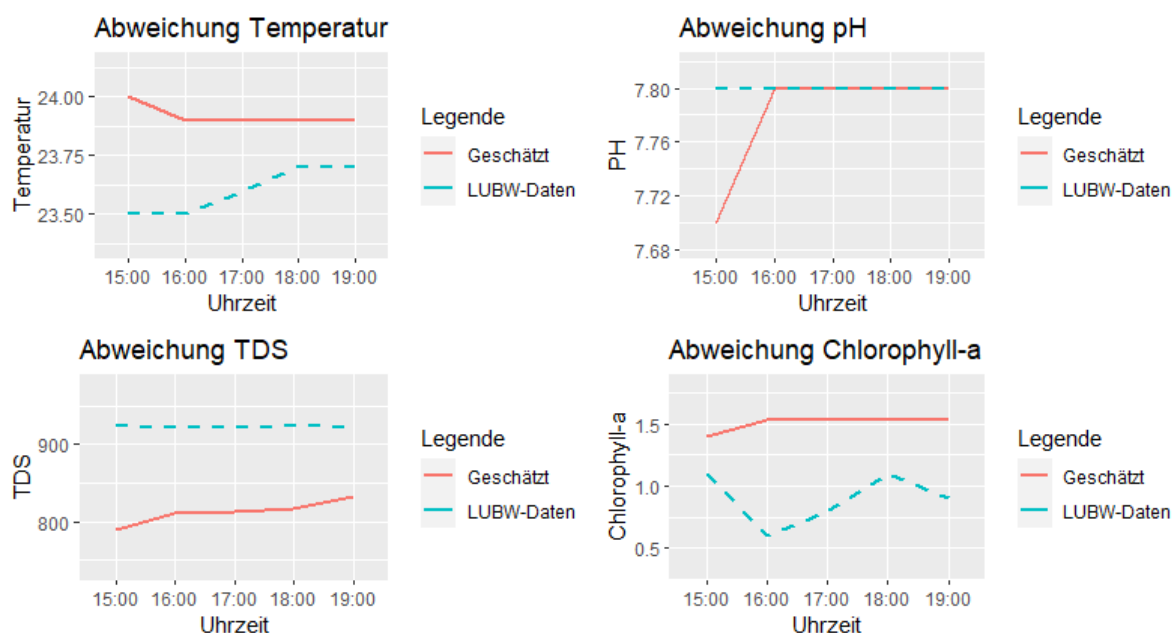


Abbildung 8.1: Vergleich der gemessenen stündlichen Messwerte mit den Sensordaten der LUBW

In Tabelle 8.2 wird dargestellt, wie sich die gemessenen Werte von IoT-Systemen von den Sensordaten der “Daten- Und Kartendienst Der LUBW” (2023) unterscheiden.

	Temperatur	PH	TDS	Chlorophyll-a
Durchschnittliche Abweichung	0.32	0.02	111	0.61
Maximale Abweichung	0.5	0.1	135	0.93

Tabelle 8.2: Stündliche gemessene Durchschnittswerte mit Schätzung des Chlorophyll-a Gehaltes jeder Stunde

Nun stellt sich die Frage, wieso die Schätzungen des Chlorophyll-a Gehaltes eine durchschnittliche Abweichung von 0.61 $\mu\text{g/L}$ aufwiesen. Diese Frage soll im folgenden Kapitel behandelt werden.

8.2 Mögliche Ursachen für eine Abweichung der Schätzungen

Eine größere Abweichung kann mehrere Ursachen haben, die in diesem Kapitel weiter erklärt werden sollen.

Größere Komplexität

Die multiple Regression hat gezeigt, dass die Variabilität der verwendeten Werten bei 47.32% liegt. Es ist anzunehmen, dass mehr chemische und biologische Prozesse ablaufen, die das Wachstum der Chlorophyll-a Masse beeinflussen, als gemessen werden kann.

Nicht berücksichtigte Messwerte

Das Vorhersagemodell wäre genauer, wenn sich entweder mehr Faktoren auslesen lassen würden (im kostengünstigen **IoT**-Bereich), oder wenn der verwendete Datensatz mehr Faktoren bereitstellen würde. Der Datensatz der Messstelle Mannheim am Neckar beinhaltet neben den drei Werten noch den Sauerstoffgehalt und die Trübung. Im Kapitel [9.3](#) wird behandelt, inwiefern eine Erweiterung der Messwerte zur Verbesserung der Vorhersage beiträgt.

Ausreißer

Ausreißer, sowohl in den Chlorophyll-a, als auch in den anderen Messwerten können sich stark negativ auf die lineare Funktion auswirken und somit eine genaue Vorhersage verhindern. In [Abbildung 8.2](#) ist dargestellt, wie sich der Chlorophyll-a Wert in den beobachteten drei Jahren bewegt.

Es kann festgestellt werden, dass es vereinzelt Ausreißer in den Messungen gibt. Solche kleinen Ausreißer sind ebenfalls stark in den Messungen des pH-Wertes zu beobachten.

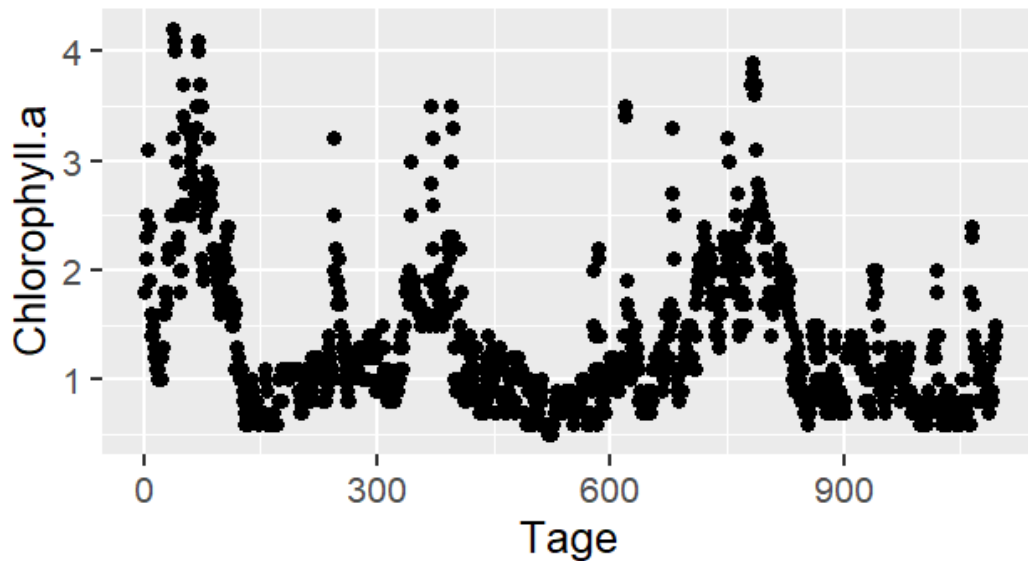


Abbildung 8.2: Ausreißer in den Chlorophyll-a Messungen des verwendeten Datensatzes

Da Ausreißer vermehrt in mehreren Positionen vorkommen, wird angenommen, dass Ausreißer hier tatsächliche Werte und seltene Ereignisse repräsentieren und somit ein wichtiger Teil der Daten darstellen.

Trotz allem wird in Kapitel 9.1 zusätzlich analysiert, wie die multiple Regression mithilfe von Trainingsdaten ohne Ausreißer abschneidet.

Nichtlineare Zusammenhänge

Ein lineares Modell macht dann Sinn, wenn die Beziehungen zwischen den Messwerten und den Chlorophyll-a Werten linear sind. Ein Blick in Abbildung 8.3 zeigt, dass die Beziehungen zwischen den Variablen und den Chlorophyll-a Werten größtenteils nicht linear erscheinen. So ist die Anwendung einer linearen Regression möglicherweise nicht angebracht.

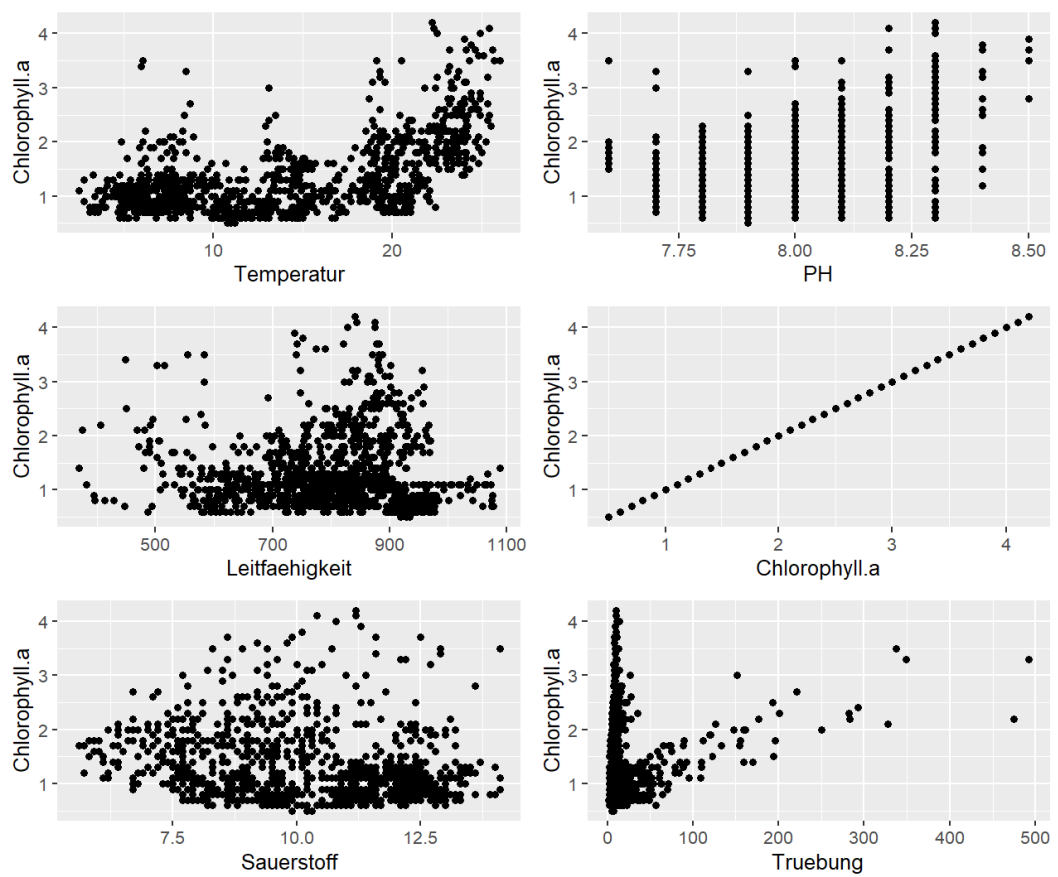


Abbildung 8.3: Beziehungen zwischen dem Chlorophyll-a Wert und den unabhängigen Variablen

Somit wird in Kapitel 9.2 untersucht, wie mithilfe von [GAM](#), anstatt einer multiplen Regression, die Chlorophyll-a Masse vorhergesagt werden kann und ob ein nicht-lineares Modell besser abschneidet.

9 Mögliche Erweiterungen

Dieses Kapitel bietet eine Vorstellung davon, wie die in dieser Arbeit präsentierten Ansätze und Methoden zukünftig verbessert und erweitert werden könnten. Es ist wichtig zu betonen, dass die in dieser Arbeit entwickelte Lösung zwar effektiv und zuverlässig ist, aber dennoch Raum für Verbesserungen und weitere Forschung bietet.

9.1 Entfernen von Ausreißern

Die Bereinigung des Datensatzes erfolgt mithilfe der [IQR](#)-Methode. Sie identifiziert Ausreißer, indem sie den Bereich berechnet, in dem die meisten Daten im Datensatz liegen und dann Datenpunkte außerhalb dieses Bereichs als Ausreißer betrachtet und entfernt. Die R-Implementierung der [IQR](#)-Methode wird in [Abbildung 9.1](#) vorgestellt.

```

# Remove spikes based of IQR
selected_vars <- c("Chlorophyll.a", "Temperatur", "pH", "
  ↳ Leitfaehigkeit", "Truebung")
data_without_outliers <- FilledDataWithO2Truebung_ThreeYears
for (column_name in selected_vars) {
  Q1 <- quantile(data_without_outliers[, column_name], 0.25, na.rm
    ↳ = TRUE)
  Q3 <- quantile(data_without_outliers[, column_name], 0.75, na.rm
    ↳ = TRUE)
  IQR <- Q3 - Q1
  lower_threshold <- Q1 - 1.5 * IQR
  upper_threshold <- Q3 + 1.5 * IQR
  data_without_outliers <- data_without_outliers[!is.na(data_
    ↳ without_outliers[, column_name]) & data_without_outliers[,
    ↳ column_name] >= lower_threshold & data_without_outliers[,
    ↳ column_name] <= upper_threshold, ]
}

# Save the cleansed dataset
write.csv(data_without_outliers, "bereinigter_datensatz.csv", row.
  ↳ names = FALSE)

```

Abbildung 9.1: IQR-Methode zum Entfernen von Ausreißern in einem gegebenen Datensatz

Der bereinigte Datensatz wird separat abgespeichert und kann nun verglichen werden. Abbildung 9.2 zeigt den Datensatz nach der Anwendung der IQR-Methode zur Entfernung der Ausreißer.

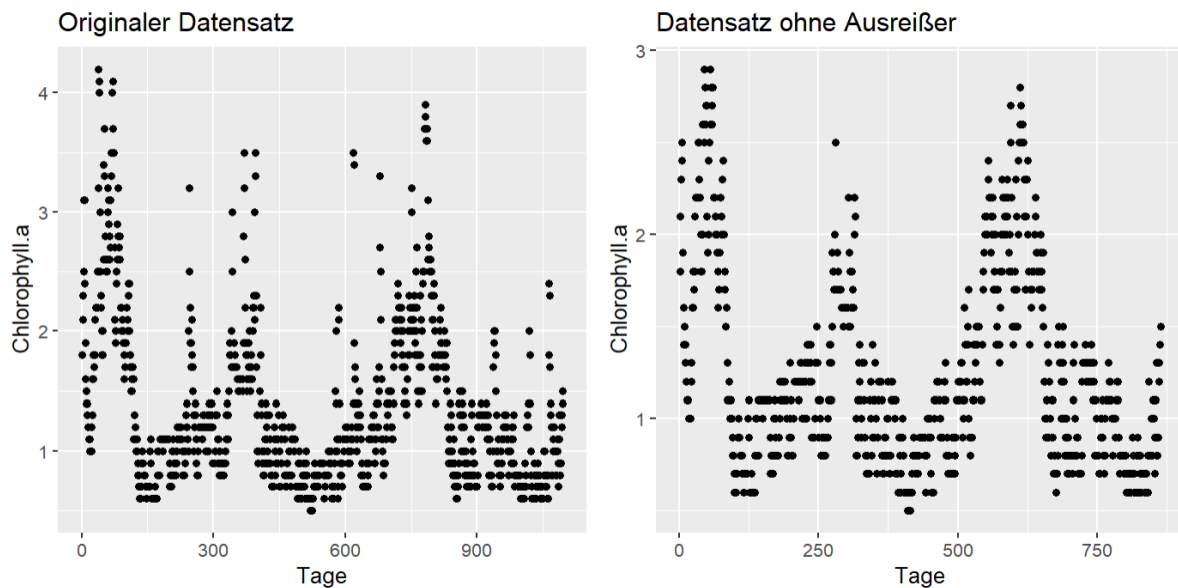


Abbildung 9.2: Vergleich der Chlorophyll-a Daten mit (links) den originalen Daten und (rechts) ohne Ausreißern nach Anwendung der [IQR-Methode](#)

Anstatt den ursprünglichen 1.095 Daten sind es noch 864, die ebenfalls in Trainings- und Testdaten unterteilt und analysiert werden. In [Abbildung 9.3](#) wurde die multiple Regression angewendet, allerdings auf den Datensatz ohne Ausreißer.

```
> print(paste("Min Residuen:", min(residuals(gam_model)
↪ )))
[1] "Min Residuen: -0.774710114597862"
> print(paste("Max Residuen:", max(residuals(gam_model)
↪ )))
[1] "Max Residuen: 0.944902069972424"
> print(paste("Median Residuen:", median(residuals(gam_
↪ model))))
[1] "Median Residuen: -0.0173757944632752"
```

Abbildung 9.3: Residuen aus der multiplen Regression auf einen Datensatz ohne Ausreißer angewandt

Dadurch konnte die Min und Max Werte der Abweichungen verbessert werden.

9.2 Entwicklung des GAM

Das GAM ist eine Modellierungstechnik, die der linearen Regression ähnelt und die Analyse nichtlinearer Beziehungen ermöglicht. Somit kann die Masse an Chlorophyll-a anhand von Messwerten vorhergesagt werden, die nicht in linearer Beziehung stehen.

In dieser Arbeit wurde jedoch die multiple Regression verwendet, da sie sich durch die Verwendung der berechneten Koeffizienten leicht in anderen Umgebungen, wie beispielsweise in der Applikation, implementieren lässt. In einem GAM werden viele Faktoren, wie beispielsweise die Glättungsfunktionen aufgrund von komplexeren Algorithmen erzeugt. Das macht es in der Regel schwer bis unmöglich, eine Schätzung in der App anhand der berechneten Werte durchzuführen. Es ist also praktikabel, die Werte wie folgt in der Programmiersprache R zu berechnen und mithilfe selbiger Funktionen auf andere Daten zu testen.

9.2.1 Anwendung des GAM

Die Implementierung des GAM ist nicht groß verschieden zu der linearen Regression. Zuerst wird das additive Modell angewandt und die Ergebnisse ausgegeben. Die Implementierung und die Ausgabe der geschätzten Vorhersagewerten werden in Abbildung 9.4 dargestellt.

```
# Implementation of the GAM
gam_model <- gam(Chlorophyll.a ~ s(Temperatur) + s(pH) + s(
  ↳ Leitfaehigkeit) + s(Truebung), data = trainingAllParameter.set
  ↳ )

# View the summary of the gam model
summary(gam_model)
```

Abbildung 9.4: Implementierung des GAM in R mit Ausgabe der geschätzten Vorhersagewerten

Die Ausgabe aus Abbildung 9.5 ist ebenfalls ähnlich der linearen Regression und lässt direkt die Werte vergleichen.

```

Parametric coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.23227     0.01186   103.9   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.'
                 0.1 ' ' 1

Approximate significance of smooth terms:
              edf Ref.df      F  p-value
s(Temperatur)    8.410   8.904 104.563 < 2e-16 ***
s(pH)             5.031   5.687  24.088 < 2e-16 ***
s(Leitfaehigkeit) 8.021   8.750   6.088 2.12e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.'
                 0.1 ' ' 1

R-sq.(adj) =  0.643   Deviance explained = 65.4%
GCV = 0.10039   Scale est. = 0.097127   n = 691

```

Abbildung 9.5: Ausgabe der Ergebnisse aus dem implementierten GAM

9.2.2 Vergleich der Modelle

Der folgende Vergleich bezieht sich auf die lineare Regression und vergleicht die Werte des GAM.

Der adjusted R^2 -Wert, also die Variabilität entsprach in der multiplen Regression 53.44% und wurde durch eine nicht-lineare Vorhersage auf 64.3% erhöht.

Durch Validierung des GAM lässt sich auch die MAE vergleichen.


```
# Predicting the DV in the test set using the intercept and
  ↳ coefficients from the training set
pout <- predict(gam_model,newdata=testAllParameter.set)

# Calculate the mean absolute error (MAE) between the predicted
  ↳ values (pout) and the actual values
mean(abs(pout-testAllParameter.set$Chlorophyll.a),na.rm=T)
```

```
[1] 0.2548194
```

Abbildung 9.6: Berechnung und Ausgabe des MAE des GAM durch Validierung mithilfe des Testdatensatzes

Somit hat sich die durchschnittliche Abweichung der linearen Regression von 0.3120205 auf eine durchschnittliche Abweichung von 0.2548194 unter nicht-linearer Analyse verbessert.

In Tabelle 9.1 wird dargestellt, wie sich die Werte im Laufe der Arbeit in den unterschiedlichen angewendeten Modellen unterscheiden haben.

Tabelle 9.1: Vergleich der Modellergebnisse

Metrik	Lineare Regression	LR ohne Ausreißer	GAM
R^2	0.53	0.52	0.64
MAE	0.31	0.31	0.25
Abw. Median	-0.02	-0.02	-0.01
Abw. Min	-0.83	-0.89	-0.74
Abw. Max	0.87	0.9	0.96

Wenn nun keine zusätzliche App benötigt werden würde und sich die einzelnen Vorhersagen eventuell manuell oder per externer API erstellt werden würden, so könnte das **GAM** eine gute Alternative zur multiplen Regression darstellen.

9.3 Weitere Messwerte

Die Rohdaten enthalten neben der Temperatur, pH und TDS auch noch den Sauerstoffgehalt und die Trübung.

Zum Schluss dieses Kapitels wird in Abbildung 9.7 noch ein Ranking durchgeführt, welches jede Messwertkombination durchgeht und ermittelt, welche Kombination den besten MAE-Wert zu diesem Datensatz liefert. Im Folgenden werden die Top-4 Kombinationen dargestellt.

Die in dieser Arbeit verwendete Messwertkombination mit Temperatur, pH und Leitfähigkeit befindet sich auf Platz 12 mit einem mittleren absoluten Fehler (MAE) von 0.312020450559308. Dies ist im Vergleich zu Platz 1 nur eine durchschnittliche Abweichung von 0.015 schlechter. Es wäre überlegenswert, andere Faktoren miteinzurechnen, allerdings wird sich die Vorhersage anhand der multiplen Regression wahrscheinlich nicht wesentlich verbessern.

```

[1] "Rank 1 : Temperatur, pH, Sauerstoff, Truebung"
[1] "Adjusted R-squared: 0.552235202648589"
[1] "MAE: 0.299271854489923"
[1] "Residuals:"
      Min.   1st Qu.   Median     Mean   3rd Qu.
      ↪      Max.
-0.836882 -0.259407  0.006691  0.000000  0.280058
      ↪ 0.903159

[1] "-----"
[1] "Rank 2 : Temperatur, pH, Sauerstoff"
[1] "Adjusted R-squared: 0.552186719772256"
[1] "MAE: 0.29937358723959"
[1] "Residuals:"
      Min.   1st Qu.   Median     Mean   3rd Qu.
      ↪      Max.
-0.836787 -0.256674  0.005106  0.000000  0.280951
      ↪ 0.903739

[1] "-----"
[1] "Rank 3 : Temperatur, pH, Leitfaehigkeit,
      ↪ Sauerstoff"
[1] "Adjusted R-squared: 0.565748160424915"
[1] "MAE: 0.302076368320751"
[1] "Residuals:"
      Min.   1st Qu.   Median     Mean   3rd Qu.
      ↪      Max.
-0.83224 -0.26843 -0.01393  0.00000  0.28002
      ↪ 0.89915

[1] "-----"
[1] "Rank 4 : Temperatur, pH, Leitfaehigkeit,
      ↪ Sauerstoff, Truebung"
[1] "Adjusted R-squared: 0.567778903728379"
[1] "MAE: 0.302590291643271"
[1] "Residuals:"
      Min.   1st Qu.   Median     Mean   3rd Qu.
      ↪      Max.
-0.83077 -0.26025 -0.01772  0.00000  0.27205
      ↪ 0.88901

```

Abbildung 9.7: Ranking aller Messwertkombinationen sortiert nach MAE mit Ausgabe der Vorhersageergebnisse

10 Diskussion

Zum Schluss dieser Arbeit wird noch einmal evaluiert, ob die definierten Ziele erreicht wurden und wie sich die Ergebnisse interpretieren lassen.

10.1 Vergleich mit der Zielsetzung

Dieses Kapitel widmet sich dem Vergleich der in der Zielsetzung der Bachelorarbeit formulierten Ziele mit den tatsächlich erreichten Ergebnissen und der Erfüllung des angestrebten Anwendungsfalls. Es soll zur Reflexion über den Fortschritt und der Ergebnisse der Arbeit dienen und die Frage beantworten, ob das entwickelte Sensorsystem unter Verwendung kostengünstiger IoT-Technologie zur Erkennung und Vorhersage von Algenblüten geeignet ist.

10.1.1 Zusammenfassung der Zielsetzung

In diesem Abschnitt werden die Ziele und Erwartungen, wie sie in der Einleitung definiert wurden, noch einmal zusammengefasst, um einen groben Überblick über die geplanten Ergebnisse zu schaffen.

Einsatz kostengünstiger IoT-Sensoren

Das Ziel dieser Arbeit ist es, ein IoT-System zu schaffen, welches kostengünstige Sensoren verwendet, um Wasserparameter zu messen, die dann zur weiteren Verarbeitung und zur Schätzung von Chlorophyll-a Werten verwendet werden. Hier wird die Frage gestellt, ob die Wasserparameter, die sich aus kostengünstigen Sensoren auslesen lassen, dazu ausreichen, um eine angemessene Schätzung durchzuführen.

Entwicklung eines Vorhersagemodells

Die erfassten Wasserparameter dienen als Grundlage für ein Vorhersagemodell, um eine Abschätzung der Biomasse des Chlorophyll-a zu ermöglichen. Dazu sollte ein geeignetes Modell gewählt werden, welches sich praktikabel einsetzen lässt und eine angemessene Schätzung liefern kann.

Entwicklung einer App zur Datenüberwachung

Als Nebenziel sollte eine App entwickelt werden, welche die gemessenen Daten in Echtzeit anzeigen lässt und dem Nutzer in unterschiedlicher Art und Weise die Daten intuitiv und leicht zu verstehen visualisiert. Hier wird ebenfalls die Vorhersage aus dem Modell angewandt und die Schätzung angezeigt.

10.1.2 Vergleich der erreichten Ergebnisse

Im nächsten Schritt werden die tatsächlich erzielten Ergebnisse der Bachelorarbeit präsentiert und mit den zuvor definierten Zielen verglichen.

IoT-System

In dieser Arbeit wurde erläutert, welche Sensoren verwendet wurden, wie diese funktionieren und weshalb sie benutzt wurden. Dabei wurde detailliert erklärt, wie die Anbindung stattfand und wie die Messwerte ausgelesen werden konnten. Es wurde für jeden einzelnen Messwert untersucht, welche Auswirkungen diese auf die Entwicklung der Algenpopulationen haben. Darüber hinaus wurden die Gründe erläutert, warum die Messwerte für Temperatur, pH-Wert und Leitfähigkeit in der Analyse verwendet wurden.

Die verwendeten Sensoren sind kostengünstig und lassen sich ohne großen Aufwand in ein IoT-System einbinden.

Vorhersagemodell

In der Arbeit wurde eine multiple Regression angewandt, welches anhand der Messwerte eine Schätzung an Chlorophyll-a tätigen kann. Dazu wurden Datensätze von “Daten-Und Kartendienst Der LUBW” (2023) verwendet und das Modell aufgrund dessen trainiert. Es wurde eine detaillierte Datenanalyse betrieben, welche Relevanz die jeweiligen Messparameter in der Vorhersage haben und diese beeinflussen können. Neben der Entwicklung des Modells und der genauen Datenanalyse wurde ebenfalls ermittelt, wie gewisse Abweichungen Zustände kommen und wie diese verhindert werden könnten. Das entwickelte Modell wurde validiert und es wurde visualisiert, welche Stärken und Schwächen dieses Modell aufzeigt.

Zum Schluss wurde das entwickelte IoT-System verwendet, um Messwerte zu sammeln und anhand des Vorhersagemodells eine Schätzung von Chlorophyll-a zu erhalten. Die

Ergebnisse des Tests zeigen, dass das angewendete Vorhersagemodell eine nützliche Methode zur Schätzung des Chlorophyll-a-Gehalts auf Basis der gemessenen Umweltparameter darstellt. Somit wurde das Ziel, ein Vorhersagemodell zu erstellen, welches angemessene Schätzungen liefert, erfüllt.

Entwicklung der App

Um die Zielsetzung der App zu vergleichen wird die Anforderungsanalyse aus Kapitel 7.1 mit der entwickelten App verglichen. Die entwickelte App implementiert eine Datenvisualisierung, welche intuitiv auf einen Blick die wichtigsten Messwerte und Ergebnisse der Schätzungen darstellt. Dazu wurden weitere Applikationsseiten entwickelt, welche mit mehreren Diagrammen die nötigen Informationen darstellt. Die App enthält eine Echtzeitdaten-Anzeige, die synchron bei neuen Messwerten eine Aktualisierung durchführt, um den Nutzer immer auf den neuesten Stand zu halten. Dabei wird die App oder Teile hiervon nicht neu geladen, sondern die Daten werden beim Erhalt stream-orientiert an den vorhandenen Daten angehängen. Neben der Darstellung der Messwerte wird hier ebenfalls die Schätzung der multiplen Regression angewandt, indem die erhaltenen Werte mit den abgespeicherten Koeffizienten aus dem Vorhersagemodell verrechnet werden. Das Ergebnis wird ebenfalls wie bei den Messwerten immer aktualisiert und in Echtzeit dargestellt. Die App schafft es, größere Mengen an Daten zu laden und zu visualisieren, ohne dass es zu Ladezeiten oder Verzögerungen kommt.

10.2 Limitationen der Arbeit und zukünftige Empfehlungen

In dieser Arbeit wurden wesentliche Fortschritte bei der Überwachung und Vorhersage von Algenblüten gemacht. Dennoch gibt es einige Limitationen und Unklarheiten, die beachtet werden sollten.

Validierung

Die Validierung der in dieser Arbeit entwickelten Methoden erfolgte anhand von Messungen, die an einem einzigen Ort und über einen kurzen Zeitraum von fünf Stunden durchgeführt wurden. Daher sind die erhaltenen Ergebnisse lediglich Stichproben und repräsentieren möglicherweise nicht die gesamte Umgebung. Zukünftige Arbeiten könnten diese Methoden an verschiedenen Orten und über längere Zeiträume validieren, um

ihre Anwendbarkeit und Genauigkeit weiter zu behandeln und zu überprüfen.

Allgemeingültigkeit des Modells

Die entwickelten Modelle wurden auf Grundlage von Daten erstellt, die an einem bestimmten Ort und zu bestimmten Zeiten gesammelt wurden. Deshalb kann nicht garantiert werden, dass die Modelle unter anderen Bedingungen oder in anderen Gewässern ebenso effektiv sind. Möglicherweise ist es notwendig für verschiedene Standorte und Umgebungsbedingungen individuelle Modelle zu entwickeln und auszutesten.

Messdifferenzen

Die während der Validierung festgestellten Messdifferenzen könnten auf mehrere Faktoren zurückzuführen sein. Eine Möglichkeit ist, dass die IoT Sensoren eine gewisse Ungenauigkeit zu den eigentlichen Werten aufweisen. Eine weitere Möglichkeit besteht darin, dass geringfügige Unterschiede in den Standorten zu unterschiedlichen Messwerten führen. Zukünftige Untersuchungen können diese Faktoren genauer untersuchen, um die Genauigkeit der erfassten Daten zu erhöhen.

Diese Limitationen verringern nicht die wichtigen Beiträge dieser Arbeit, aber sie weisen auf die Notwendigkeit hin, die entwickelten Methoden weiter zu verfeinern und zu verbessern.

Weitere Empfehlungen für zukünftige Arbeiten

Für weitere Arbeiten, die eine Verbesserung der Vorhersage oder eine Erweiterung als Ziel haben, wird empfohlen, weitere Messgrößen wie Nitrat, Phosphor, Trübung oder Sauerstoff zu berücksichtigen. Hierfür ist ein Vorhersagemodell erforderlich, das auf diese zusätzlichen Messwerten trainiert ist. Durch die Messung und Sammlung von Chlorophyll-a-Werten kann jedes Gewässer individuell anhand dieser Größe geschätzt werden, und durch die Integration weiterer, wie oben erwähnten Messgrößen und eine größere Datenmenge kann die Vorhersagegenauigkeit weiter verbessert werden. Darüber hinaus könnten durch die Anwendung anderer Modelle, die in dieser Arbeit nicht behandelt wurden, möglicherweise bessere Ergebnisse erzielt werden.

Insgesamt bieten die in dieser Arbeit präsentierten Ergebnisse wertvolle Anhaltspunkte für die weitere Forschung in diesem Bereich.

11 Fazit

Diese Arbeit hat bedeutende Fortschritte im Bereich der Algenblütenüberwachung und Vorhersage durch die Integration von IoT erzielt. Indem kostengünstige Sensoren mit einem effektiven mehrdimensionalen Vorhersagemodell kombiniert wurden, wurde ein leistungsfähiges Überwachungssystem geschaffen, das in Echtzeit akzeptable Vorhersagen über den Chlorophyll-a-Gehalt liefern kann.

Die Entwicklung und Implementierung einer benutzerfreundlichen App zur Visualisierung und Echtzeitüberwachung der erfassten Daten trägt dazu bei, die Auswirkungen der Wasserparameter auf den Algengehalt besser zu verstehen und wenn nötig rechtzeitig Maßnahmen zu ergreifen.

Es ist wichtig zu betonen, dass die Entstehung von Algenblüten komplexer ist und von vielen anderen Faktoren beeinflusst wird. Diese Arbeit stellt nur einen ersten Schritt in Richtung eines Ansatzes zur Überwachung und Vorhersage von Algenblüten dar. Weitere Untersuchungen und die Integration zusätzlicher Parameter oder alternativer Vorhersagemodellen könnten zu noch präziseren Vorhersagen führen.

Darüber hinaus bieten die Ergebnisse auch wertvolle Anhaltspunkte für die weitere Forschung. Es wurden ebenfalls Möglichkeiten zur Modellverbesserung für zukünftige Arbeiten geboten, um einen Anreiz für weitere Forschung zu geben.

Insgesamt demonstriert diese Arbeit das große Potenzial von IoT und Vorhersagemodellen zur Überwachung und Vorhersage von Umweltphänomenen. Dieses System kann ebenfalls außerhalb der Aquaristik angewendet werden, um andere Ökosysteme zu überwachen, zu verstehen und zu optimieren.

Literatur

- Anderson, C. R., Moore, S. K., Tomlinson, M. C., Silke, J., & Cusack, C. K. (2015). Chapter 17 - Living with Harmful Algal Blooms in a Changing World: Strategies for Modeling and Mitigating Their Effects in Coastal Marine Ecosystems. In J. F. Shroder, J. T. Ellis & D. J. Sherman (Hrsg.), *Coastal and Marine Hazards, Risks, and Disasters* (S. 495–561). Elsevier. <https://doi.org/10.1016/B978-0-12-396483-0.00017-0>
- Araby, A. A., Abd Elhameed, M. M., Magdy, N. M., Said, L. A., Abdelaal, N., Abd Allah, Y. T., Darweesh, M. S., Fahim, M. A., & Mostafa, H. (2019). Smart IoT Monitoring System for Agriculture with Predictive Analysis. *2019 8th International Conference on Modern Circuits and Systems Technologies (MOCASST)*, 1–4. <https://doi.org/10.1109/MOCASST.2019.8741794>
- Arduino with TDS Sensor. (2023). Verfügbar 24. Juni 2023 unter <https://randomnerdtutorials.com/arduino-tds-water-quality-sensor/>
- Björn, L. O., Papageorgiou, G. C., Blankenship, R. E., & Govindjee. (2009). A viewpoint: Why chlorophyll a? *Photosynthesis Research*, 99(2), 85–98. <https://doi.org/10.1007/s11120-008-9395-x>
- Butterwick, C., Heaney, S. I., & Talling, J. F. (2005). Diversity in the influence of temperature on the growth rates of freshwater algae, and its ecological relevance. *Freshwater Biology*, 50(2), 291–300. <https://doi.org/10.1111/j.1365-2427.2004.01317.x>
- Chapman, R. L. (2013). Algae: The world’s most important “plants”—an introduction. *Mitigation and Adaptation Strategies for Global Change*, 18(1), 5–12. <https://doi.org/10.1007/s11027-010-9255-9>
- Chen, M., Tang, H., Ma, H., Holland, T. C., Ng, K. Y. S., & Salley, S. O. (2011). Effect of nutrients on growth and lipid accumulation in the green algae *Dunaliella tertiolecta*. *Bioresource Technology*, 102(2), 1649–1655. <https://doi.org/10.1016/j.biortech.2010.09.062>
- Daten- Und Kartendienst Der LUBW. (2023). Verfügbar 30. Mai 2023 unter <https://udo.lubw.baden-wuerttemberg.de/public/index.xhtml>
- Desortová, B. (1981). Relationship between Chlorophyll- α Concentration and Phytoplankton Biomass in Several Reservoirs in Czechoslovakia. *Internationale Revue*

- der gesamten Hydrobiologie und Hydrographie*, 66(2), 153–169. <https://doi.org/10.1002/iroh.19810660202>
- Dubinsky, Z., & Rotem, J. (1974). Relations between algal populations and the pH of their media. *Oecologia*, 16(1), 53–60. <https://doi.org/10.1007/BF00345087>
- Esenowo Jack, K., Nwangwu, E. O., Agwu Etu, I., & Osuagwu, E. U. (2016). A Simple Thermistor Design for Industrial Temperature Measurement. *IOSR Journal of Electrical and Electronics Engineering*, 11(05), 57–66. <https://doi.org/10.9790/1676-1105035766>
- Falconer, I. R. (1993). Algal Toxins in Seafood and Drinking Water.
- Griffith, A. W., & Gobler, C. J. (2020). Harmful algal blooms: A climate change co-stressor in marine and freshwater ecosystems. *Harmful Algae*, 91, 101590. <https://doi.org/10.1016/j.hal.2019.03.008>
- Hallegraeff, G., Anderson, D., & Cembella, A. (1995). IOC Manuals and Guides No.33. *Manual on harmful marine microalgae*.
- Hansen, M. M. (2023). DS18B20. Verfügbar 4. Juni 2023 unter <https://github.com/matmunk/DS18B20>
- Harmful Algal Information System. (2023). Verfügbar 7. Mai 2023 unter <http://haedat.iode.org/index.php>
- Hoagland, P., & Scatasta, S. (2006). The Economic Effects of Harmful Algal Blooms. In E. Granéli & J. T. Turner (Hrsg.), *Ecology of Harmful Algae* (S. 391–402). Springer. https://doi.org/10.1007/978-3-540-32210-8_30
- How Does A pH Probe Work? (2021). Verfügbar 14. Mai 2023 unter <https://atlas-scientific.com/blog/how-does-a-ph-probe-work/>
- Karlson, B., Andersen, P., Arneborg, L., Cembella, A., Eikrem, W., John, U., West, J. J., Klemm, K., Kobos, J., Lehtinen, S., Lundholm, N., Mazur-Marzec, H., Naustvoll, L., Poelman, M., Provoost, P., De Rijcke, M., & Suikkanen, S. (2021). Harmful algal blooms and their effects in coastal seas of Northern Europe. *Harmful Algae*, 102, 101989. <https://doi.org/10.1016/j.hal.2021.101989>
- Kovač Blagojević, D., Simeunović, J., Babić, O., Mišan, A., & Milovanović, I. (2013). Algae in Food and Feed. *Food and Feed Research*, 58226, 641–1.
- Lee, I., & Lee, K. (2015). The Internet of Things (IoT): Applications, investments, and challenges for enterprises. *Business Horizons*, 58(4), 431–440. <https://doi.org/10.1016/j.bushor.2015.03.008>
- Manyika, J., Chui, M., Bisson, P., Woetzel, J. R., Dobbs, R., Bughin, J., & Aharon, D. (2015). The Internet of Things: Mapping the Value beyond the Hy-

- pe. Verfügbar 11. Mai 2023 unter <https://www.semanticscholar.org/paper/The-internet-of-things%3A-mapping-the-value-beyond-Manyika-Chui/2cf73f5c0de26be03205e5ca27da39a626404a97>
- McCormick, P. V., & Cairns, J. (1994). Algae as indicators of environmental change. *Journal of Applied Phycology*, 6(5), 509–526. <https://doi.org/10.1007/BF02182405>
- Mehta, S. K., & Gaur, J. P. (2005). Use of Algae for Removing Heavy Metal Ions From Wastewater: Progress and Prospects. *Critical Reviews in Biotechnology*, 25(3), 113–152. <https://doi.org/10.1080/07388550500248571>
- Mesmer, R. E., & Holmes, H. F. (1992). pH, Definition and measurement at high temperatures. *Journal of Solution Chemistry*, 21(8), 725–744. <https://doi.org/10.1007/BF00651506>
- Mouha, R. A. (2021). Internet of Things (IoT). *Journal of Data Analysis and Information Processing*, 9(2), 77–101. <https://doi.org/10.4236/jdaip.2021.92006>
- Necas, J., & Bartosikova, L. (2013). Carrageenan: A review. *Veterinárni medicína*, 58(4), 187–205. <https://doi.org/10.17221/6758-VETMED>
- Poikāne, S., Alves, M. H., Argillier, C., van den Berg, M., Buzzi, F., Hoehn, E., de Hoyos, C., Karottki, I., Laplace-Tretyure, C., Solheim, A. L., Ortiz-Casas, J., Ott, I., Phillips, G., Pilke, A., Pádúa, J., Remec-Rekar, S., Riedmüller, U., Schaumburg, J., Serrano, M. L., ... Wolfram, G. (2010). Defining Chlorophyll-a Reference Conditions in European Lakes. *Environmental Management*, 45(6), 1286–1298. <https://doi.org/10.1007/s00267-010-9484-4>
- Sabater, S. (2009). Diatoms. In G. E. Likens (Hrsg.), *Encyclopedia of Inland Waters* (S. 149–156). Academic Press. <https://doi.org/10.1016/B978-012370626-3.00135-6>
- Seo, Y.-B., Lee, Y.-W., Lee, C.-H., & You, H.-C. (2010). Red algae and their use in papermaking. *Bioresource Technology*, 101(7), 2549–2553. <https://doi.org/10.1016/j.biortech.2009.11.088>
- Steele, M. R., Gitelson, A. A., & Rundquist, D. C. (2008). A Comparison of Two Techniques for Nondestructive Measurement of Chlorophyll Content in Grapevine Leaves. *Agronomy Journal*, 100(3), agj2agronj20070254n. <https://doi.org/10.2134/agronj2007.0254N>
- Teoh, M.-L., Chu, W.-L., & Phang, S.-M. (2010). Effect of temperature change on physiology and biochemistry of algae: A review. *Malaysian Journal of Science*, 29(2), 82–97. <https://doi.org/10.22452/mjs.vol29no2.1>

-
- ThingSpeak Communication Library for Arduino, ESP8266 and ESP32. (2023). Verfügbar 25. Juni 2023 unter <https://github.com/mathworks/thingspeak-arduino>
- Thirumalini, S., & Joseph, K. (2009). Correlation between Electrical Conductivity and Total Dissolved Solids in Natural Waters. *Malaysian Journal of Science*, 28(1), 55–61. <https://doi.org/10.22452/mjs.vol28no1.7>
- Virtanen, O., Constantinidou, E., & Tyystjärvi, E. (2022). Chlorophyll Does Not Reflect Green Light – How to Correct a Misconception. *Journal of Biological Education*, 56(5), 552–559. <https://doi.org/10.1080/00219266.2020.1858930>
- Westcott, C. (2012). *Ph Measurements*. Elsevier.
- Zaman Zad Ghavidel, S., & Montaseri, M. (2014). Application of different data-driven methods for the prediction of total dissolved solids in the Zarinehroud basin. *Stochastic Environmental Research and Risk Assessment*, 28(8), 2101–2118. <https://doi.org/10.1007/s00477-014-0899-y>

Ehrenwörtliche Erklärung

Hiermit versichere ich, die vorliegende Arbeit ohne fremde Hilfe und nur unter Verwendung der angegebenen Hilfsmittel selbstständig verfasst zu haben. Alle Stellen, die wörtlich oder sinngemäß aus veröffentlichten oder nicht veröffentlichten Arbeiten anderer entnommen sind, habe ich kenntlich gemacht.

Heilbronn, den 26. Juli 2023

Robert Kessler