



Hochschule für Angewandte Wissenschaften Hamburg  
*Hamburg University of Applied Sciences*

# Masterarbeit

Deniz Himstedt

## Datenbasierte Analyse von Schiffswellen in der Außenweser mittels Machine Learning

*Fakultät Technik und Informatik  
Department Maschinenbau und Produktion*

*Faculty of Engineering and Computer Science  
Department of Mechanical Engineering and  
Production Management*



**Deniz Himstedt**

**Datenbasierte Analyse von Schiffswellen in der  
Außenweser mittels Machine Learning**

Masterarbeit eingereicht im Rahmen der Masterprüfung

im Studiengang Konstruktionstechnik und Produktentwicklung  
am Department Maschinenbau und Produktion  
der Fakultät Technik und Informatik  
der Hochschule für Angewandte Wissenschaften Hamburg

in Zusammenarbeit mit:  
Ramboll Deutschland GmbH  
Jürgen-Töpfer-Straße 48  
22763 Hamburg

Erstprüfer\*In: Prof. Dr. Sarah Hallerberg  
Zweitprüfer\*In : Dr. Karsten Peters

Industrielle Betreuer\*In: Dr. Karsten Peters

Abgabedatum: 07.11.2022



# Zusammenfassung

**Deniz Himstedt**

## **Thema der Masterarbeit**

Datenbasierte Analyse von Schiffswellen in der Außenweser mittels Machine Learning

## **Stichworte**

Machine Learning, Künstliche neuronale Netze, Schiffswellen

## **Kurzzusammenfassung**

Aus Erfahrungswerten von Messkampagnen werden Vorhersagemodelle aus neuronalen Netzen entwickelt, welche Schiffswellenparameter vorhersagen. Die Einflussparameter auf die Schiffswellen werden untersucht. Modelle zur einfachen Klassifizierung hoher und niedriger Wellen sowie Modelle zur Bestimmung der Wellenhöhe mit Werten werden entwickelt und beurteilt.



**Deniz Himstedt**

**Master's Thesis title**

Data based analysis of ship waves in the Außenweser by means of machine learning

**Keywords**

Machine learning, artificial neuronal networks, ship waves, ship wakes

**Abstract**

From the experience of measuring campaigns neuronal networks will be developed to predict parameters of shipwaves. The influencing parameters on the waves of ships will be analyzed. Models to classify high and low waves as well as models to predict wave heights by value will be developed and evaluated.





# Aufgabenstellung

Wellen stellen eine Belastung für Kanäle und anliegende Gebäude dar. Diese Wellen setzen sich aus natürlichen Seegangswellen und Schiffswellen zusammen. Um die Wellenbelastung auf den Kanal zu optimieren, ist ein Kenntnis der Schiffswellen und deren Einflussparameter notwendig. Aus einem Datensatz von Meeres- und Schiffspassagedaten sollen Schiffswellen von natürlichem Seegang unterschieden werden. Die Einflussparameter zur Entstehung von großen Schiffswellen werden analysiert. Basierend auf dieser Analyse wird ein Modell entwickelt, um Schiffswellen vorherzusagen und so die Wellenbelastung am Kanal einzuschätzen.



# Inhaltsverzeichnis

<b>Zusammenfassung</b>	<b>I</b>
<b>Aufgabenstellung</b>	<b>V</b>
<b>1 Einleitung</b>	<b>1</b>
<b>2 Ausgangslage</b>	<b>3</b>
<b>3 Methoden</b>	<b>7</b>
3.1 Korrelationsmatrix . . . . .	7
3.2 Künstliches neuronales Netz . . . . .	7
3.2.1 ANN mit Dense Layer . . . . .	8
3.2.2 Aktivierungsfunktion . . . . .	8
3.2.3 ANN Training . . . . .	9
3.2.4 ANN Testen . . . . .	10
3.2.5 Overfitting . . . . .	10
3.2.6 Dropout Layer . . . . .	10
3.2.7 Convolutional Layer . . . . .	10
3.2.8 Max Pooling Layer . . . . .	10
3.2.9 Flatten Layer . . . . .	11
<b>4 Numerische Experimente und Ergebnisse</b>	<b>13</b>
4.1 Künstliches neuronales Netz (ANN) mit binärer Klassifizierung . . . . .	13
4.1.1 Convolutional Neuronal Network (CNN) . . . . .	14
4.1.2 ANN mit Dense Layer (Ein Dense Layer, 32 Knoten) . . . . .	17
4.1.3 ANN mit Dense Layer (Zwei Dense Layer, je 64 Knoten) . . . . .	18
4.1.4 Variation der Wellenhöhe . . . . .	19
4.2 ANN zur Regressionsanalyse . . . . .	21
4.2.1 Großer Datensatz . . . . .	21
4.2.2 Reduzierter Datensatz . . . . .	27
4.2.3 Einzelanalyse Datensatz 2014 . . . . .	31
4.2.4 Einzelanalyse Datensatz 2021 . . . . .	33
4.2.5 Einzelanalyse reduzierter Datensatz 2021 . . . . .	35
4.2.6 Einzelanalyse Datensatz mit Primärwellen über 10 cm . . . . .	37
4.2.7 Einzelanalyse Datensatz mit Primärwellen unter 10 cm . . . . .	39
4.2.8 Datensatz reduziert durch Undersampling . . . . .	41
4.2.9 Nachträgliche Trendkorrektur . . . . .	45
4.2.10 Weitere Tests zur Optimierung der Vorhersage der Primärwelle . . . . .	49
4.2.11 Vorhersage des Absunks . . . . .	50
4.2.12 Vorhersage der Sekundärwelle . . . . .	52
4.3 Korrelationsmatrix . . . . .	54
4.3.1 Einflussparameter Primärwellenhöhe . . . . .	54
4.3.2 Einflussparameter Sekundärwellenhöhe . . . . .	55
<b>5 Software zur Prognose von Primärwellen</b>	<b>57</b>

<b>6 Zusammenfassung und Ausblick</b>	<b>59</b>
<b>Formelzeichen</b>	<b>61</b>

# 1 Einleitung

Die Belastung von Schiffswellen auf den Kanal ist für die Betreiber interessant, so dass die Belastungen untersucht werden. Schiffswellen werden beeinflusst durch die Schiffsgewichte, Schiffsgeschwindigkeit, Strömungsgeschwindigkeit, das verdrängte Wasservolumen und die Umgebung, in die das Wasser verdrängt wird. Natürliche Seegangswellen werden hauptsächlich erzeugt durch Tide und Wind. Die Firma Ramboll GmbH hat für die Wasserstraßen- und Schifffahrtsverwaltung des Bundes (WSV) im Jahre 2014 und 2021 in der Außenweser Messdaten erhoben und Schiffswellenparameter der passierenden Schiffe analytisch ausgewertet. Die Berechnung der Schiffswellen wurde bereits mit 3D Stömungssimulation [20] für einzelne Schiffe sowie mit 2D Modellen [18] mit Messdaten durchgeführt. Die Berechnung der Schiffswellen ist mit diesen Methoden nur mit einer aufwendigen Modellierung des Kanals und einzelner Schiffe oder durch ungenaue Vereinfachungen möglich. Mit den Methoden des überwachten Lernens [29] des Machine Learning ist es möglich, ein künstliches neuronales Netz anhand von Beispielen und Gesetzmäßigkeiten anzutrainieren und nachzubilden. Die variablen Parameter (Schiffs-, meteorologische und hydrologische Daten) sind durch Messungen bekannt. Die Kanalgeometrie wird dem künstlichen neuronalen Netz zwar nicht weiter mitgeteilt, bleibt über den betrachteten Zeitraum aber nahezu konstant. Ein Datensatz an Ergebnissen für Schiffswellenparameter liegt durch die vorhergehende Analyse vor. Dadurch ist es möglich, ein Modell zu entwickeln, welches die Gesetzmäßigkeiten zwischen den Daten erkennt und so Vorhersagen über die Schiffswellenparameter von passierenden Schiffen zu erzeugen. Viele Daten über Schiffe sind auch nach den einzelnen Messkampagnen verfügbar, so dass die Schiffswellenbelastung auf den Kanal auch außerhalb der Messkampagnen betrachtet werden kann.



## 2 Ausgangslage

Die Wellen in der Außenweser sind eine Überlagerung von natürlichen Seegangswellen und Schiffswellen. Die natürlichen Seegangswellen können zu Zeiten ohne Schiffspassage ausgewertet werden. Die Wellen der Schiffspassagen können nicht ohne die natürlichen Seegangswellen betrachtet werden. Um die Schiffswellen zu untersuchen, wurden die Wellen unterteilt analysiert im langwelligen Primärwellensystem und kurzwelligen Sekundärwellensystem. Die Primärwelle wird hauptsächlich dadurch erzeugt, dass das fahrende Schiff Wasser am Bug verdrängt und dieses am Heck wieder zusammenfließt. Die Analyse des Primärwellensystems erfolgt dadurch, dass im ersten Schritt das gesamte Signal der Welle mit einem Filter oberhalb einer Periode von 12 Minuten 30 Sekunden gefiltert wird, um den tidebedingten Wellenanteil zu filtern. Im zweiten Schritt wird das Signal unterhalb einer Periodendauer von 15 Sekunden gefiltert. Die Sekundärwellen entstehen bei der Umströmung des Schiffes und sind unter anderem abhängig von der Schiffsform und der Geschwindigkeit. Das Sekundärwellensystem wurde analysiert dadurch, dass aus dem Seegangssignal alle Signale oberhalb einer Periodendauer von 15 Sekunden gefiltert wurden. Dieser Frequenzbereich überlagert sich mit dem Frequenzbereich von Seegangswellen erzeugt durch Wind. Aus dem Primärwellensystem und Sekundärwellensystem wurden Kennwerte ermittelt. Auf Basis dieser Daten und analytisch ermittelter Kennwerte sollen Modelle zur Prognose von Schiffswellen entwickelt und die Qualität dieser Modelle beurteilt werden. Die Datensätze zu den Schiffswellenmesskampagnen 2014 und 2021 in der Außenweser dienen dieser Arbeit als Grundlage. In den Messkampagnen wurden folgende Daten erhoben.

Für die Messkampagnen gemessene Stationsdaten an den Stationen A, B und C:

- Wasserstand gemessen mit Wellendraht
- Wasserstand gemessen mit Drucksensor
- Luftdruck

Meteorologische Daten vom WSV:

- Windgeschwindigkeit
- Windrichtung

Übermittelte AIS (Automatisches Identifikationssystem) Schiffspassagedaten:

- Schiffslänge
- Schiffsbreite
- Schiffstiefe
- Geschwindigkeit relativ zum Grund
- Schiffskurs
- Abstand zum Messpfahl
- Koordinaten bei Passage der Messstationen

Pegelstände des WSV an den Stationen:

- Dwarsgat
- Robbensüdsteert
- Bremerhaven

Der Wellengang wird bei einer Schiffspassage analysiert auf die Schiffswellenparameter:

- Primärwellenhöhe

## 2 Ausgangslage

- Sekundärwellenhöhe
- Eintrittszeit
- Absunk
- Periodendauer der Primärwelle
- Periodendauer der Sekundärwelle
- Sunkzeit
- Stiegzeit

In der Messkampagne des Jahres 2021 wurden nur Primärwellen größer als 5 cm ausgewertet. Kleinere Primärwellen wurden nicht ausgewertet, da diese sich nicht deutlich genug von den natürlichen Seegangswellen unterscheiden. In der Messkampagne von 2014 wurden auch kleinere Primärwellen untersucht.

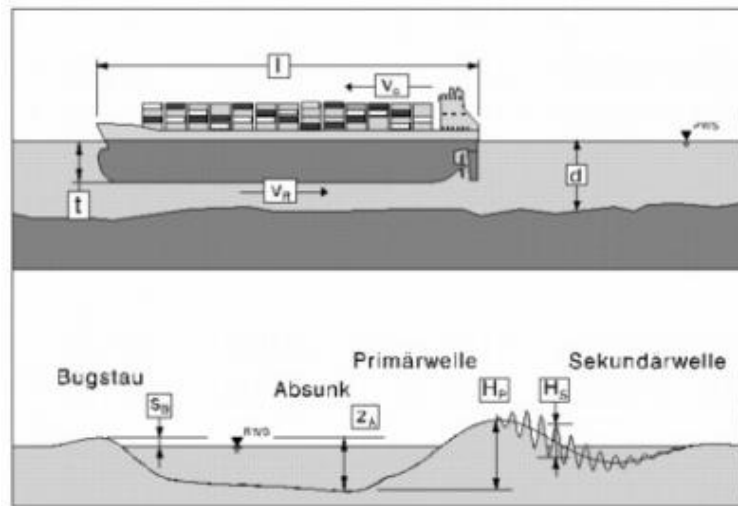


Abbildung 2.0.0.1: Schiffswellenkenngrößen Bugstau, Absunk, Primärwelle und Sekundärwelle nach [3]



A Robbennordsteert, Ereignis: 17.10.2021 15:18:00, Passage: NORTHERN-JUBILEE

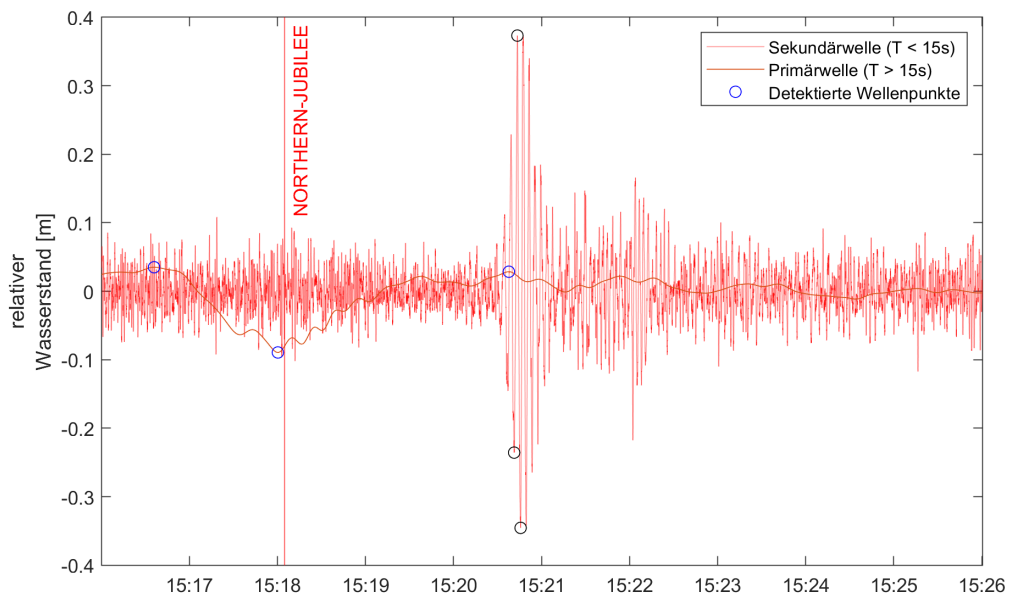
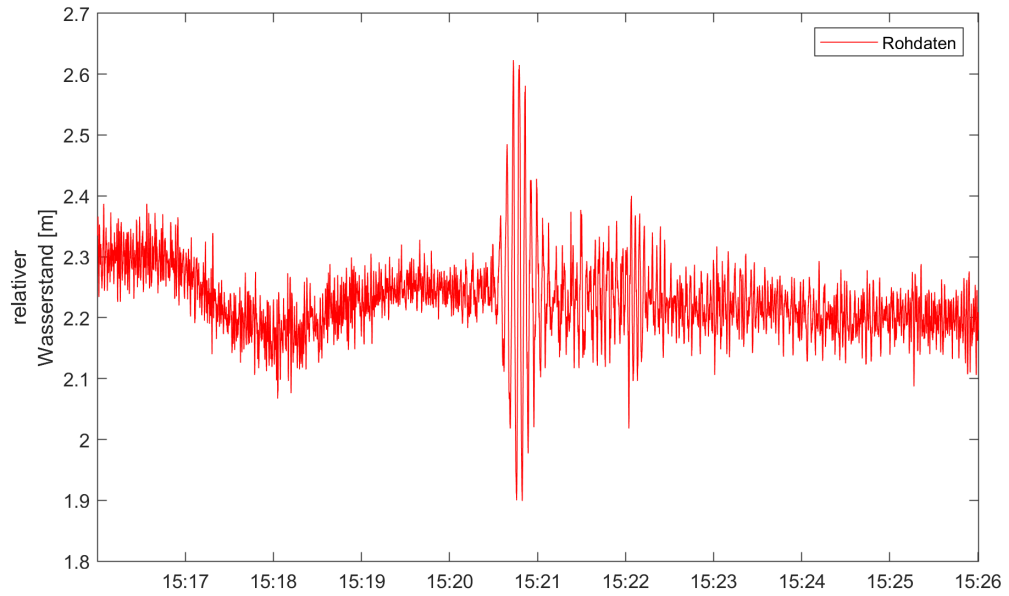


Abbildung 2.0.0.2: Schiffswellenkenngrößen ermittelt aus Messdaten



Abbildung 2.0.0.3: Positionen der Messstationen A, B und C

[17]

Messlokation	Weser km	UTM Koordinaten	
A Robbennordsteert	90,412	456005	5950933
B Langlutjen	87,694	457826	5947455
C Buhne 16	90,412	465087	5942529

Tabelle 2.0.0.1: Positionen der Messstationen

# 3 Methoden

## 3.1 Korrelationsmatrix

Der lineare Zusammenhang zwischen zwei gemessenen Größen kann durch einen Korrelationskoeffizienten [24] beschrieben werden. Dabei entspricht der Wert 1 einem vollständig linearen Zusammenhang zwischen den Variablen. Beispielsweise haben Schiffe mit einer großen Länge meist eine große Breite. Bei diesem Datensatz ist also zu erwarten, dass der Korrelationskoeffizient zwischen Schiffslänge und Schiffsbreite nahe 1 ist. Die Schiffslänge und der Wasserstand haben keinen Zusammenhang, daher ist ein Wert nahe 0 zu erwarten. Sollte der Wasserstand allerdings soweit sinken, dass lange Schiffe den Kanal nicht mehr passieren können, zeigt sich diese Korrelation in dem Koeffizienten. Bei einer negativen Korrelation fällt der Koeffizient negativ aus. In den folgenden Gleichungen finden sich alle Größen, die in die Berechnung des Korrelationskoeffizienten eingehen,

$$\mu = \frac{1}{n} \sum_{i=1}^n x_i, \quad (3.1)$$

$$\text{Var}(X) = \frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2, \quad (3.2)$$

$$\text{Cov}(X, Y) = \frac{1}{n} \sum_{i=1}^n (x_i - \mu_X)(y_i - \mu_Y), \quad (3.3)$$

$$\text{Korr}(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)}\sqrt{\text{Var}(Y)}}, \quad (3.4)$$

mit

$\mu$  Erwartungswert [23],

$x_i$  Messwert einer Zeitreihe  $\{x\}_t$  zum Zeitpunkt  $t = i$ ,

$y_i$  Messwert einer zweiten Zeitreihe  $\{y\}_t$  zum Zeitpunkt  $t = i$ ,

$X$  Zufallsvariable von der die Zeitreihe  $\{x\}_t$  als Stichprobe gesampled wurde,

$Y$  Zufallsvariable von der die Zeitreihe  $\{y\}_t$  als Stichprobe gesampled wurde,

Var Varianz [30],

Cov Kovarianz [26],

Korr Korrelationskoeffizient.

In einer Korrelationsmatrix [25] werden ein jeweiliger Parameter in der Reihe und ein jeweiliger Parameter in der Spalte aufgetragen in identischer Reihenfolge. In den Matrixelementen werden die Korrelationskoeffizienten der beteiligten Parameter eingetragen. Auf der Diagonalen wird der lineare Zusammenhang zwischen dem identischen Wert hergestellt, weshalb in diesen Elementen der Wert 1 ist. Die Korrelationsmatrix dient dazu, die linearen Zusammenhänge zwischen zwei Variablen aus vielen Variablen besser zu erkennen.

## 3.2 Künstliches neuronales Netz

Künstliche neuronale Netze [22] (ANNs - artificial neuronal networks) bieten Methoden des maschinellen Lernens, die Ähnlichkeiten mit der Funktionsweise natürlicher Neuronen in einem

Gehirn aufweisen. Grundsätzliches Funktionsziel eines ANNs ist, auf Grundlage eines Eingangsdatensatzes eine Antwort zu geben [15] [16]. Beispielsweise können die Eingangsdaten sein: Ein Schiff mit 400 m Länge und einer Geschwindigkeit von 23 Knoten passiert die Station A der Außenweser. Die zu beantwortende Frage kann sein: Ist die erzeugte Primärwelle dieses Schiffs größer als 15 cm? Eine Antwort könnte dann mit Ja oder Nein ausfallen. Im Folgenden wird auf einen typischen Aufbau von ANNs eingegangen, der nicht grundsätzlich für alle ANNs gilt.

### 3.2.1 ANN mit Dense Layer

Ein ANN besteht aus einem Input Layer, mindestens einem Hidden Layer und einem Output Layer. Der Input Layer besteht z.B. aus den Eingangsdaten wie Schiffslänge, Geschwindigkeit etc.. Diese Informationen werden auf einen Wertebereich von 0 bis 1 normiert und in die Knoten (Nodes oder Neuronen) des Input Layers eingefügt. Hinter dem Input Layer kommt ein Hidden Layer. Typischerweise ist jeder Knoten des vorhergehenden (hier Input-) Layer mit jedem Knoten des darauf folgenden Layers verbunden. Es gibt unterschiedliche Arten von Hidden Layer. Die beschriebene Arbeitsweise ist die eines Dense Layers. Wahl der Art der Hidden Layer und die Anzahl der Knoten beeinflussen die Funktionalität des ANNs. Nach dem letzten Hidden Layer kommt der Output Layer, der die Antwort in Wahrscheinlichkeit ausgedrückt enthält.

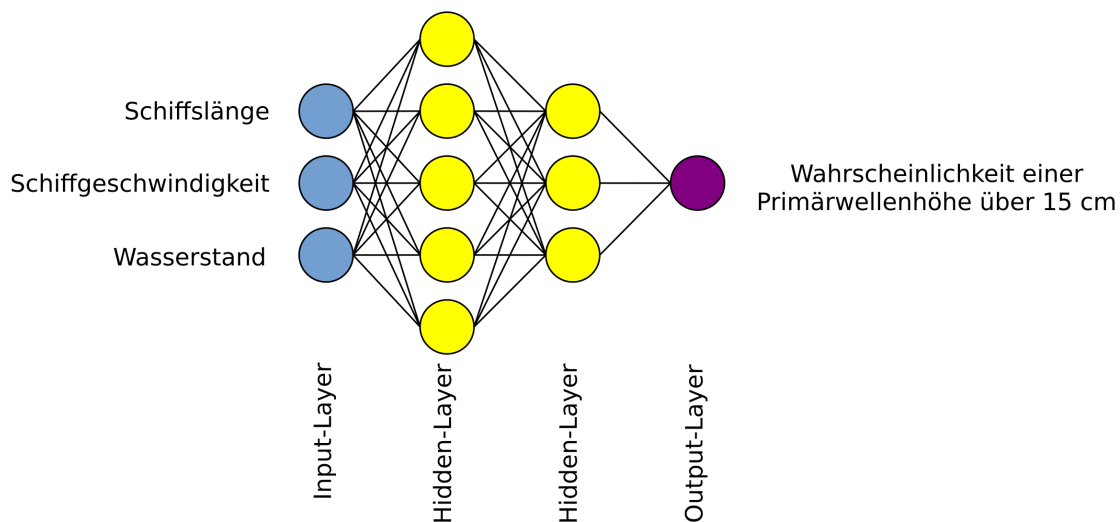


Abbildung 3.2.1.1: Beispiel eines Künstlichen neuronalen Netz

[6]

### 3.2.2 Aktivierungsfunktion

Ein einzelner Knoten [28] führt bei der Vorhersage eine Aktivierungsfunktion [5] (activation function) aus. In den Knoten gehen als Eingangsparameter die Daten jeden Knotens des vorherigen Layers. Diese Parameter ( $x$ ) werden jeweils mit einer zugehörigen Gewichtung ( $w$ ) multipliziert und anschließend summiert mit einem zusätzlichem Bias Wert ( $b$ ) [7]. So ergibt sich

$$v = \sum_{i=1}^n x_i w_i + b. \quad (3.5)$$

Diese Summe ( $v$ ) wird anschließend in die Aktivierungsfunktion ( $\varphi$ ) eingefügt. Es gibt viele unterschiedliche Aktivierungsfunktionen. Eine davon ist die Schwellwertfunktion mit der Formel



Abbildung 3.2.2.1: Eingangsdaten eines Knotens im ersten Layer

$$\varphi^{hlim}(v) = \begin{cases} 1 & \text{wenn } v \geq 0 \\ 0 & \text{wenn } v < 0 \end{cases} \quad (3.6)$$

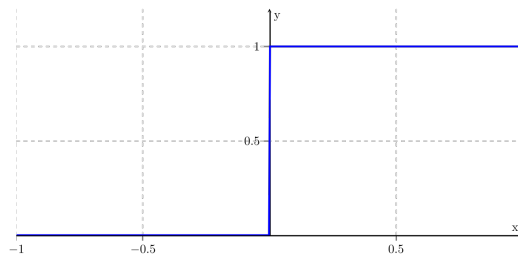


Abbildung 3.2.2.2: Schwellwertfunktion[28]

Eine weitere ist die Sigmoid Aktivierungsfunktion.

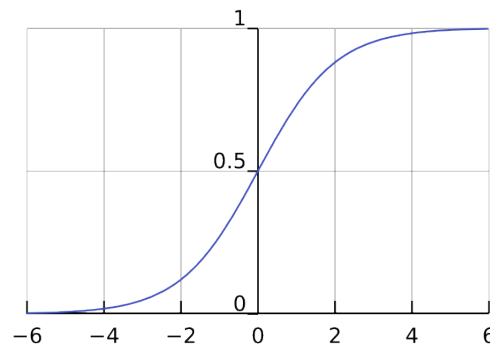


Abbildung 3.2.2.3: Sigmoid Aktivierungsfunktion[5]

Das Ergebnis der Aktivierungsfunktion geht als Dateneingang in jeden Knoten des darauf folgenden Layer.

### 3.2.3 ANN Training

Nachdem ein ANN aufgebaut wurde und die Layer mit ihrer Anzahl der Knoten und Aktivierungsfunktionen gewählt wurden, muss dieses trainiert werden, damit es entsprechende Antworten liefern kann. Dazu wird im Trainingsprozess das ANN mit Eingangsdaten und der zur erwarteten Antwort trainiert. Vor dem Training werden zufällige Werte in den einzelnen Gewichtungen ( $w$ ) und den Bias Werten ( $b$ ) der einzelnen Neuronen zugewiesen. Eine Wahl, worauf das ANN optimiert wird, ist ebenfalls notwendig. Meist wird die Loss-Funktion [10] minimiert. Loss ist die Differenz zwischen vorhergesagtem Wahrscheinlichkeitswert und der Antwort im Testdatensatz. Oftmals wird die mittlere quadratische Abweichung aller Loss Werte minimiert, so dass große Abweichungen stärker gewichtet werden als kleinere. Die Gewichtungen

und Bias-Werte werden im Trainingsprozess iterativ verändert durch einen Optimierungsprozess genannt Backpropagation [4]. Der Trainingsdatensatz braucht eine gewisse Größe, damit dass ANN der Aufgabenstellung entsprechend optimiert werden kann.

#### 3.2.4 ANN Testen

Der Loss-Wert basierend auf dem Trainingsdatensatz ist für eine Beurteilung der Qualität der Vorhersagen nicht hinreichend. Bei einem ANN, dem eine identische Frage gestellt wird wie im Training, würde die Qualität der Antwort zu günstig ausfallen. Um das ANN besser zu beurteilen, ist ein Testdatensatz notwendig. Die Daten, die zum Testen verwendet werden, stehen damit nicht mehr dem Training zur Verfügung. Falls genügend Daten vorhanden sind, bietet sich eine Aufteilung in 50% Test- und 50% Trainingsdaten an. Der Testdatensatz kann aber auch reduziert werden auf bspw. 20%.

#### 3.2.5 Overfitting

Overfitting [21] [12] ist ein Problem, das sich einstellt, wenn ein ANN zu sehr auf den Trainingsdatensatz optimiert wird und bei den Trainingsdaten wesentlich schlechtere Vorhersagen liefert. Im ANN soll der generelle Problemlösungsweg hinterlegt werden und nicht spezielle Antworten auf den Trainingsdatensatz. Die Aufteilung spezieller Merkmale eines Datensatzes kann ebenfalls zu Overfitting führen. Wenn bspw. Schiffspassagen mit extrem hohen Primärwellen in den Testdatensatz fallen, kann das ANN nicht hinreichend für diesen Fall trainiert werden.

#### 3.2.6 Dropout Layer

Der Dropout Layer [12] sorgt dafür, dass beim Training einige Knoten zufällig ignoriert werden. Diese Knoten beeinflussen dann das Ergebnis nicht mehr. Das Ziel dieser Methode ist es, Overfitting zu vermeiden.

#### 3.2.7 Convolutional Layer

Ein Convolutional Layer [8] ist eine Layerart, die häufig für Mustererkennung in Bildern eingesetzt wird. In einem ANN für Bilderkennung werden die Bilder in einem bestimmten Format aus Auflösung und Farbwerten als Eingangswerte übergeben. In den Knoten eines Convolutional Layers können bestimmte Muster wie Kanten, Formen und Farben in Bildern erkannt werden. Diese ANNs enthalten oft weitere Typen von Layern neben den Convolutional Layern.

#### 3.2.8 Max Pooling Layer

Da ein Bild eine relativ hohe Informationsdichte durch Auflösung und Farbwerte hat und oft weniger komplizierte Antworten/Muster gesucht werden, etwa das Erkennen von Ziffern mit lediglich 10 Zeichen, ist eine Dimensionsreduktion im ANN sinnvoll. Max Pooling [11] ist eine Methode, bei der aus einer großen Matrix eine kleinere Matrix erstellt wird. Dabei wird die Eingangsmatrix in viele kleine Matrizen zerlegt. Aus den kleineren Matrizen wird nur der maximale Wert aller Elemente in eine neue Matrix übertragen. Die neue Matrix hat damit eine Größe, die durch die Anzahl der kleinen Matrizen, die in die Eingangsmatrix passen, definiert ist. Ob diese Art der Dimensionsreduktion sinnvoll ist, hängt von der gegebenen Aufgabenstellung ab. Eine andere Pooling-Methode ist Average-Pooling, bei der der Mittelwert der kleinen Matrizenelemente übertragen wird.

### 3.2.9 Flatten Layer

Flatten [9] ist eine Operation, die einen höheren Tensor in einen eindimensionalen Tensor überführt. Dies ist notwendig, wenn ein ANN aus Eingangsdaten mit höheren Dimensionen (Bspw. Bilder oder Tensoren aus Convolutional Layer) einfache Antworten im Zahlenbereich 0 bis 1 geben soll.

#### Leave-One-Out-Kreuzvalidierung (LOO-CV)

Die Methode Leave-One-Out-Kreuzvalidierung [27] bietet sich bei Datensätzen an, bei denen kaum Daten vom Trainingsdatensatz entfernt werden können, um ein gutes ANN zu trainieren. Hier besteht der Testdatensatz aus einem einzelnen Dateneintrag. Für das Training stehen dann alle anderen Daten zur Verfügung. Der Testdatensatz mit nur einem Eintrag wird nach dem Training getestet und das Resultat gespeichert. Nachdem dieser Datensatz mit dem trainierten ANN getestet wurde, wird ein anderes ANN trainiert. Bei dem nächsten Training wird ein anderer Dateneintrag zum Testen verwendet und der Rest des Datensatzes steht für das Training zur Verfügung. Um die Variation der trainierten ANN auf den Trainings- und Testdatensatz zu beschränken, werden identische Startwerte für die trainierbaren Parameter verwendet. Mit dieser Methode ist es möglich, für jede Prognose für jeden Dateneintrag eine realistische Beurteilung zu treffen. Nachteil dabei ist der hohe Rechenaufwand. Da die Methode sich allerdings an kleine Datensätze richtet, ist der Rechenaufwand durch die Größe des Datensatzes schon beschränkt.

#### Undersampling

Ein ANN erzeugt schlechte Vorhersageergebnisse, wenn bestimmte Informationen im Datensatz weniger vertreten sind als andere. Der Optimierungsalgorithmus passt das Netzwerk so an, dass der Loss im gesamten Datensatz möglichst niedrig ist. Ein häufig auftretendes Ereignis wird dabei verstärkt vorhergesagt, da der Algorithmus mit der häufiger trainierten Antwort öfter richtig liegt. Undersampling [1] ist eine Methode, um ungleichmäßig verteilte Daten einander anzugleichen. Dabei werden Daten mit der häufiger vertretenen Eigenschaft soweit entfernt, bis die Eigenschaften gleichmäßig vertreten sind. Dadurch reduziert sich die Datenmenge.

#### True Positive Rate (TPR) und False Positive Rate (FPR)

Die True Positive Rate (TPR) beschreibt das Verhältnis von korrekt positiven Klassifizierungen (TP) zur Anzahl an Klassifizierungen ( $n$ ). Als Formel

$$TPR = \frac{TP}{n}. \quad (3.7)$$

Die False Positive Rate (FPR) beschreibt das Verhältnis von falsch positiven Klassifizierungen (FP) zur Anzahl an Klassifizierungen ( $n$ ). Als Formel

$$FPR = \frac{FP}{n}. \quad (3.8)$$

[13]

#### Receiver Operating Characteristic (ROC) Kurve

Die ROC Kurve dient der Bewertung eines Klassifizierungsmodells zu unterschiedlichen Schwellwerten bezüglich der True Positive Rate (TPR) und der False Positive Rate (FPR). Je nach Aufgabenstellung kann eine falsch positive Klassifizierung schwere Folgen haben, etwa bei Medikamentenverabreichungen.

[13]

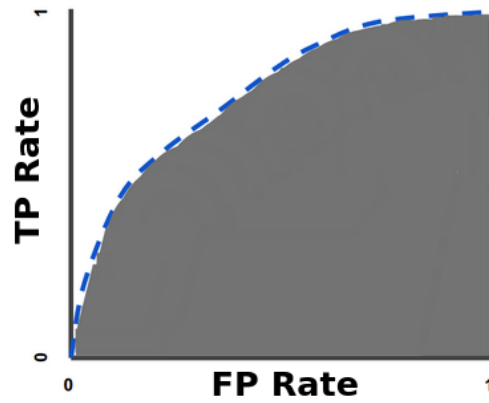


Abbildung 3.2.9.1: ROC Kurve[13]

### Area Under the ROC Curve (AUC)

Der AUC Wert ist die Fläche unter der ROC Kurve und dient der Bewertung eines Klassifizierungsmodells über den gesamten Bereich an Schwellwerten. [13]

### Software

Die Programmierung findet in der interpretierten Programmiersprache Python statt. Die neuronalen Netzwerke werden mit TensorFlow sowie Keras aufgebaut. Einzelne Methoden für das maschinelle Lernen werden mit Scikit-learn ergänzt. Das pandas Modul wird für die Datenverarbeitung verwendet. Zur numerischen Berechnung dient das Modul NumPy. Die Darstellung von Diagrammen wird mit dem Modul Matplotlib erzeugt.



# 4 Numerische Experimente und Ergebnisse

## 4.1 Künstliches neuronales Netz (ANN) mit binärer Klassifizierung

Zur Vorhersage, ob die Primärwellen von Schiffen größer oder kleiner als 15 cm sind, werden ANNs trainiert. Dabei dienen als Eingangsdaten

- Schiffslänge
- Schiffsgeschwindigkeit relativ zum Grund (SOG)
- Schiffskurs (COG)
- Wasserstand nach Wellendrahtsensor
- Wasserströmungsgeschwindigkeit
- Windgeschwindigkeit
- Abstand zum Messpfahl

Es wird der Datensatz der Station A von 2021 verwendet. Um nur Passagen zu verwenden, die nicht durch andere Schiffe beeinflusst sind, wird der Datensatz reduziert um die Fälle bei denen dieser Verdacht besteht. Zunächst werden nur Schiffe, die kürzer als 150 m sind und 7 Minuten nach oder vor einem anderen Schiff passierten heraus gefiltert. Im zweiten Schritt werden alle verbliebenen Schiffe heraus gefiltert, die 7 Minuten nach oder vor einem Schiff dieses Datensatzes passierten. Der Grund für diese Filterung ist, dass es wenig Passagen mit sehr hohen Primärwellen gibt. Von diesen wenigen Passagen dürfen nicht zu viele gefiltert werden. Es verbleiben 1736 Dateneinträge. Der Einfluss der kurzen Schiffe auf die Primärwelle der Größeren wird nach Sichtung der Zeitreihen wie bspw. in Abbildung 4.1.0.1 als vernachlässigbar angenommen.

Der Datensatz wird zufällig aufgeteilt mit einem Teil von 80% der zum Training und einem Teil von 20% der zum Testen verwendet wird. Es werden 100 ANNs mit unterschiedlichen zufälligen Startparametern trainiert, getestet und verglichen. Dies dient dazu herauszufinden, ob die trainierbaren Parameter mit unterschiedlichen Startwerten zu ähnlich sinnvollen Lösungen trainiert werden, oder ob es speziell abweichende Startwerte gibt.

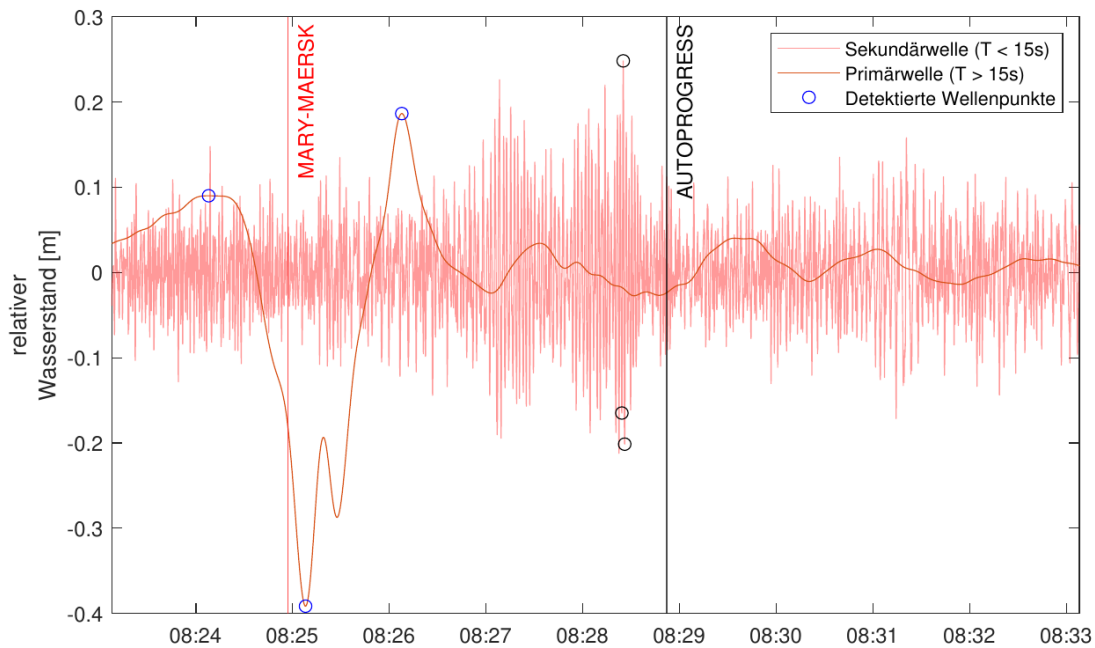


Abbildung 4.1.0.1: Der Einfluss der Autoprogress (Schiffslänge 127 m) auf die Primärwelle der Mary-Maersk (Schiffslänge 399 m) ist nicht erkennbar

### 4.1.1 Convolutional Neuronal Network (CNN)

**Aufbau** Das CNN wird zusammengesetzt aus den Layern

- Input Layer, 7 Eingangsdaten
- Convolutional Layer 2D (Nr. 1), 64 Filter
- Convolutional Layer 2D (Nr. 2), 64 Filter
- Max Pooling Layer, Poolgröße  $1 \times 7$
- Dropout, Rate 0,25
- Flatten
- Dense Layer, 64 Knoten, Aktivierungsfunktion selu
- Dense Layer, 2 Knoten, Aktivierungsfunktion softmax

Die Anzahl der trainierbaren Parameter sind 33.986 zusammengesetzt aus:

- Convolutional Layer 2D (Nr. 1), 64 Filter mit 7 Gewichtungen und einem Bias  $64(7+1) = 512$  Parameter
- Convolutional Layer 2D (Nr. 2), 64 Filter mit 7 Gewichtungen und einem Bias  $64(7+1) = 512$  Parameter
- Max Pooling Layer, kein Parameter
- Dropout 0,25, kein Parameter
- Flatten, kein Parameter
- Dense Layer, 64 Knoten mit 512 Gewichtungen und einem Bias,  $64(512 + 1) = 32832$  Parameter
- Dense Layer, 2 Knoten mit 64 Gewichtungen und einem Bias,  $2(64+1) = 130$  Parameter

### Ergebnisse

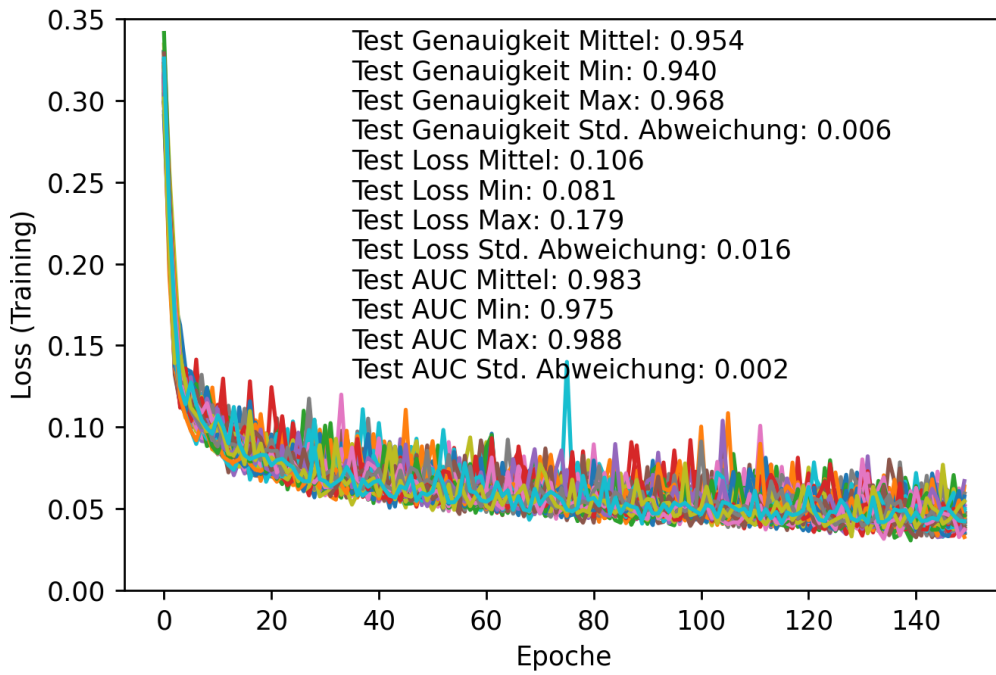


Abbildung 4.1.1.1: Optimierungsprozess Loss über Epochen, Convolutional Neuronal Network

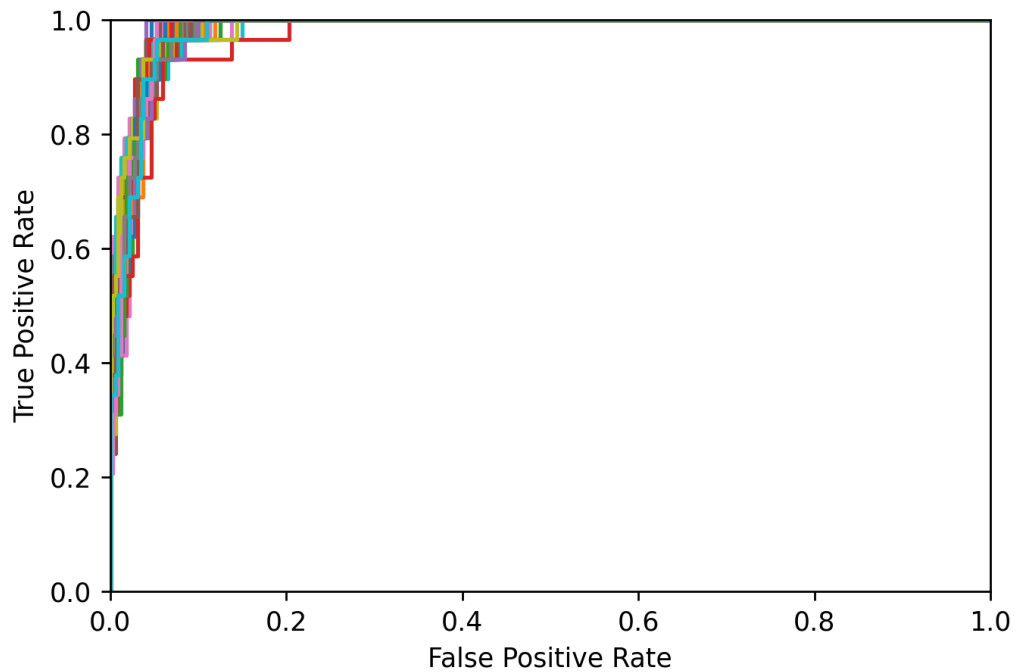


Abbildung 4.1.1.2: ROC Kurven der 100 trainierten CNNs, Convolutional Neuronal Network

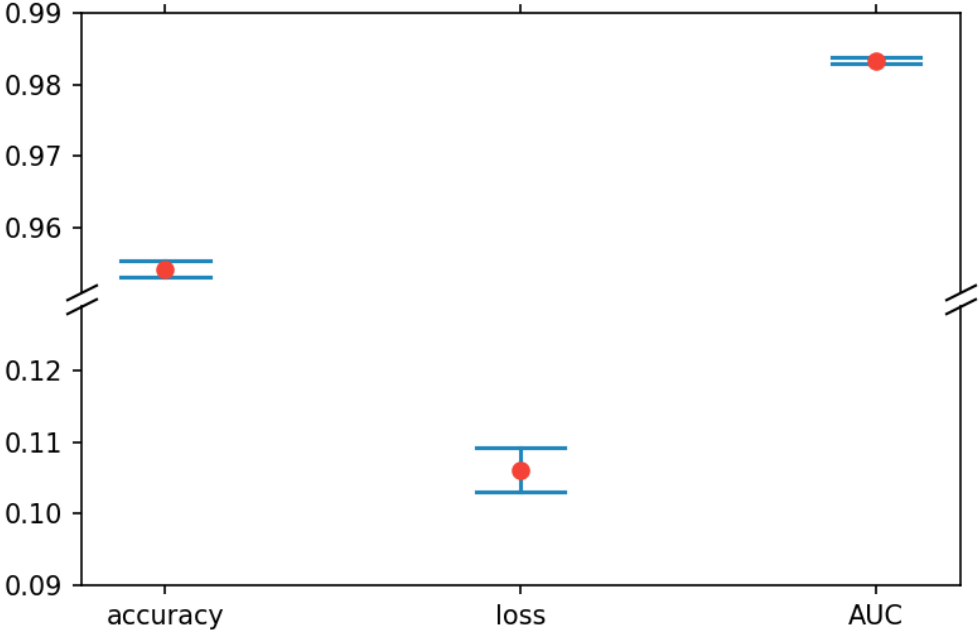


Abbildung 4.1.1.3: Konfidenzintervalle Genauigkeit, Loss und AUC, Convolutional Neuronal Network

**Diskussion** Von den 100 trainierten CNNs liegen die Vorhersagen, ob die Primärwelle größer oder kleiner als 15 cm ist, im Testdatensatz im Mittel 95,4% korrekt. Das CNN mit der schlechtesten Vorhersage hat immer noch eine Genauigkeit von 94%. Das Beste liegt bei 96,8%. Dabei ist die Spannweite 2,8%. Die mittlere Standardabweichung der Vorhersagegenauigkeiten ist 0,6%. Der Loss im Testdatensatz geht von minimal 0,081 bis maximal 0,179 mit einer Standardabweichung von 0,016. Der Loss hat damit die größte Streuung der betrachteten Kennwerte, ist aber immer noch ein sehr gutes Ergebnis. Der Loss im Trainingsdatensatz ist mit ca. 0,05 unter dem des Testdatensatzes. Die Abweichung spricht noch nicht für ein Overfitting, aber ein Training mit mehr als den 150 Epochen sollte den Loss im Testdatensatz sonderlich verbessern. Die ROC Kurven in Abbildung 4.1.1.2 zeigen einen steilen Anstieg. Lässt man eine False Positive Rate von 0,2 zu, werden in allen CNNs alle Schiffe mit einer Primärwelle über 15 cm korrekt erkannt. Bei einer FPR von 0,1 werden bei allen CNNs noch 95% der Schiffe mit einer Primärwelle über 15 cm korrekt erkannt. Die 100 AUCs spannen sich von 0,975 bis 0,988. Die AUC hat eine Standardabweichung von nur 0,002. Die Vorhersagen der CNNs sind damit sehr ähnlich. Die zufälligen Gewichtungen und Biaswerte haben keinen großen Einfluss auf die Vorhersagequalität des ANNs.

Das Convolutional neuronal Network liefert bereits erste gute Ergebnisse.

#### 4.1.2 ANN mit Dense Layer (Ein Dense Layer, 32 Knoten)

Ein einfacheres ANN soll mit einem Dense Layer statt Convolutional Layer zum Vergleich trainiert werden.

**Aufbau** Die Eingangsdaten und die Aufteilung in Test- und Trainingsdatensatz ist identisch zu 4.1.1. Der Aufbau des ANNs ist

- Input Layer, 7 Eingangsdaten
- Dense Layer, 32 Knoten, Aktivierungsfunktion relu
- Output (Dense) Layer, 1 Knoten, Aktivierungsfunktion sigmoid

Die Anzahl der trainierbaren Parameter sind 289 zusammengesetzt aus:

- Input Layer, keine Parameter
- Dense Layer, 32 Knoten mit 7 Gewichtungen und einem Bias  $32(7+1) = 256$  Parameter
- Dense Layer, 1 Knoten mit 32 Gewichtungen und einem Bias,  $1(32+1) = 33$  Parameter

[14]

**Ergebnisse** Die mittlere Genauigkeit ist identisch mit 4.1.1 bei 95,4%. Das ANN mit der schlechtesten Genauigkeit liegt bei 94,5% und damit besser als bei 4.1.1. Das ANN mit der höchsten Genauigkeit liegt bei 96,3% und liegt damit unter dem Wert von 4.1.1. Die Spannweite liegt bei 1,8% und ist damit niedriger als bei 4.1.1. Die mittlere Standardabweichung der Vorhersagegenauigkeiten liegt bei 0,3% und ist damit halb so groß wie bei 4.1.1.

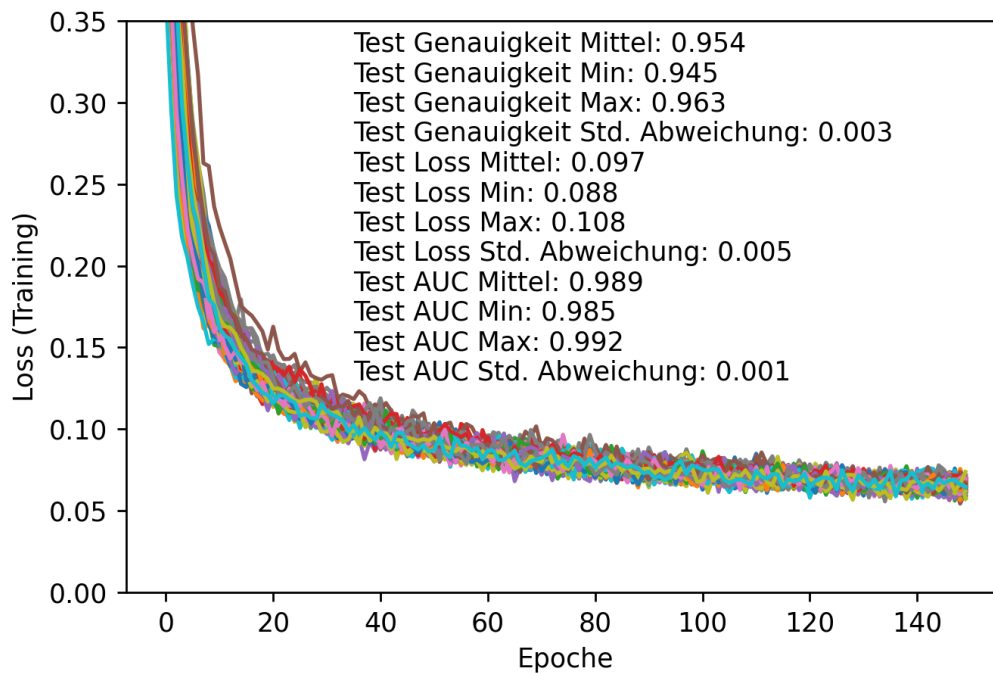


Abbildung 4.1.2.1: Optimierungsprozess Loss über Epochen, ein Dense Layer, 32 Knoten

**Diskussion** Die Ergebnisse sind sehr ähnlich wie die aus 4.1.1. Die Komplexität des ANNs konnte jedoch drastisch reduziert werden von 33.986 Parametern auf 289, wodurch das ANN aus Dense Layer dem CNN einen Vorteil hat.

### 4.1.3 ANN mit Dense Layer (Zwei Dense Layer, je 64 Knoten)

Da das ANN mit dem einfachen Dense Layer vergleichbar gute Ergebnisse lieferte wie das ANN mit Convolutional Layer, soll zum Vergleich ein komplexeres ANN mit zwei Dense Layern trainiert werden.

**Aufbau** Der Aufbau des ANNs ist

- Input Layer, 7 Eingangsdaten
- Dense Layer, 64 Knoten, Aktivierungsfunktion relu
- Dense Layer, 64 Knoten, Aktivierungsfunktion relu
- Dropout, Rate 0,5
- Output (Dense) Layer, 1 Knoten, Aktivierungsfunktion sigmoid

Die Anzahl der trainierbaren Parameter sind 4.758 zusammengesetzt aus:

- Input Layer, keine Parameter
- Dense Layer, 64 Knoten mit 7 Gewichtungen und einem Bias  $64(7+1) = 512$  Parameter
- Dense Layer, 64 Knoten mit 64 Gewichtungen und einem Bias,  $64(64+1) = 4160$  Parameter
- Dropout Layer, kein Parameter
- Output (Dense) Layer, 1 Knoten mit 64 Gewichtungen und einem Bias,  $1(64+1) = 65$  Parameter

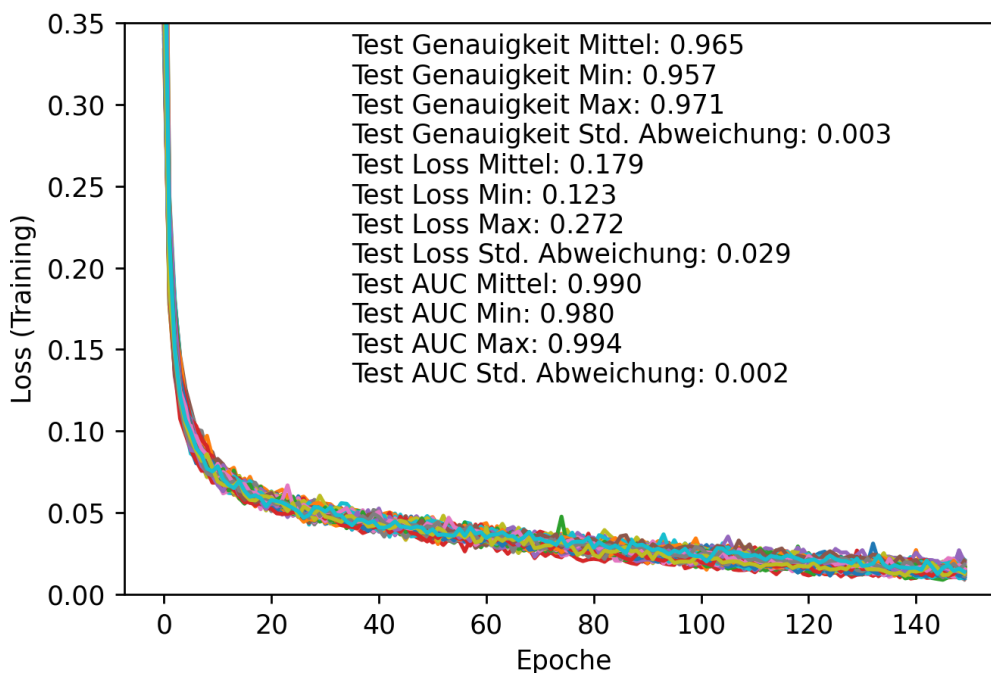


Abbildung 4.1.3.1: Optimierungsprozess Loss über Epochen, zwei Dense Layer, je 64 Knoten

**Ergebnisse** Die mittlere Genauigkeit liegt bei 96,5% und damit höher als bei 4.1.1 und 4.1.2. Das ANN mit der schlechtesten Genauigkeit liegt bei 95,7% damit besser als die mittlere Genauigkeit von 4.1.1 und 4.1.2. Das ANN mit der höchsten Genauigkeit liegt bei 97,1%. Die Spannweite liegt bei 1,4% und damit unter 4.1.1 und 4.1.2. Die mittlere Standardabweichung der Vorhersagegenauigkeiten liegt bei 0,3% und wie bei 4.1.2.

**Diskussion** Die Vorhersageergebnisse haben sich verbessert. Die Komplexität des ANNs ist mit 4.758 Parametern noch weit unter 4.1.1. Dense Layer scheinen für diese Aufgabenstellung geeigneter zu sein als Convolutional Layer.

#### 4.1.4 Variation der Wellenhöhe

Im Folgenden wird analysiert, wie die Vorhersagequalität der Klassifizierung von hohen und niedrigen Wellen bei unterschiedlichen Schwellwerten ausfällt.

**Aufbau** Die Struktur des ANNs aus 4.1.3 wird verwendet. Der Schwellwert wird von 5 cm in 1 cm Schrittweite erhöht bis auf 29 cm. Für jeden Schwellwert werden 100 ANNs mit unterschiedlichen Startwerten trainiert.

#### Ergebnisse

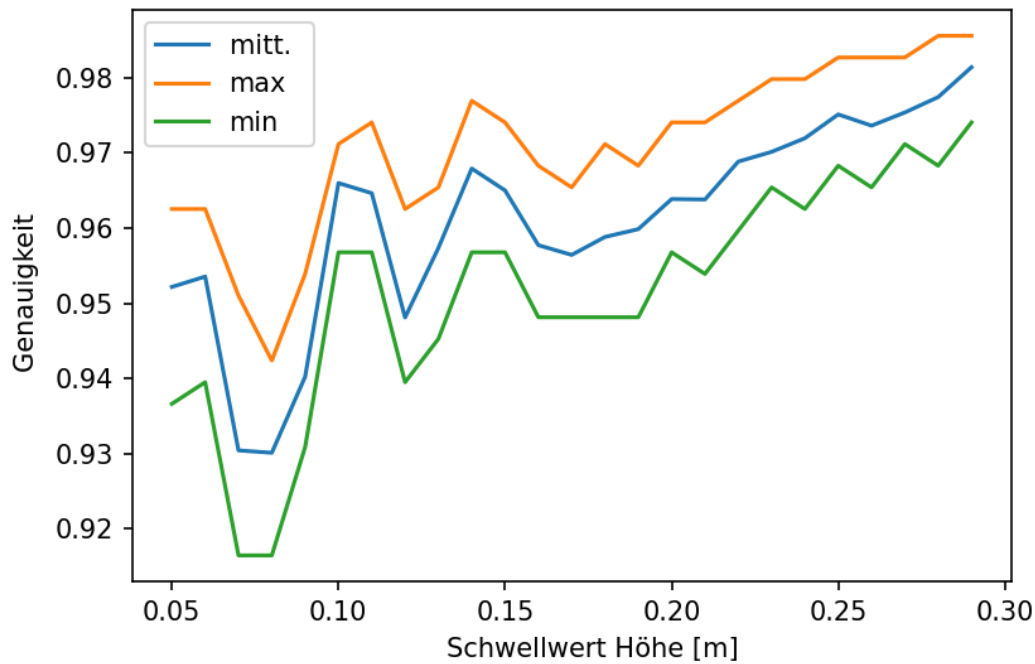


Abbildung 4.1.4.1: Vorhersagegenauigkeit bei unterschiedlichen Schwellwerten zur Klassifizierung von hohen und niedrigen Wellen

**Diskussion** In 4.1.4.1 ist zu erkennen, dass die Vorhersagegenauigkeit ab 17 cm mit dem Schwellwert nur noch steigt. Einen Einfluss auf das Ergebnis hat, dass insgesamt wenig sehr hohe Primärwellen im Datensatz vorhanden sind und damit ein ANN mit der Vorhersage einer niedrigen Primärwelle öfter richtig liegt.

Zwei lokale Minima sind zu erkennen bei einem Schwellwert von 8 cm und bei 12 cm. Die Verläufe der jeweilig besten, mittleren und schlechtesten Vorhersagegenauigkeit sind ähnlich. Ihre Differenz nimmt jedoch mit der Höhe des Schwellwertes etwas ab. Bei einem Schwellwert von 5 cm ist die Differenz zwischen höchster und niedrigster Genauigkeit 2,5%. Bei einem Schwellwert von 29 cm ist die Differenz auf 1,2 gesunken.



## 4.2 ANN zur Regressionsanalyse

ANNs können neben Klassifikationsaufgaben auch zur Regressionsanalyse verwendet werden, um die Zusammenhänge mehrerer Parameter zu erkennen und daraus kontinuierliche Werte zu bestimmen. Im Trainingsprozess ist die Antwort dann kein kategorischer Wert, sondern bspw. die Primärwellenhöhe in Meter. Wichtig hierbei ist, dass die letzte Aktivierungsfunktion auch in der Lage ist, Antworten in dem gesuchten Wertebereich zu erzeugen.

### 4.2.1 Großer Datensatz

**Aufbau** Es werden die Daten aus 2014 und 2021 verwendet, um einen möglichst großen Datensatz zu erhalten.

Als Eingangsdaten werden 3910 Passagen verwendet mit den Daten:

- Schiffslänge
- Schiffsbreite
- Tiefgang
- Schiffsgeschwindigkeit relativ zum Grund (SOG)
- Schiffskurs (COG)
- Wasserstand Station Dwarsgat
- Wasserstand Station Robbensüdsteert
- Wasserstand Station Bremerhaven
- Abstand zum Messpfahl
- Jahr der Messkampagne

Die Daten der Wasserstandsmessstationen wurden verwendet, da diese dauerhaft zur Verfügung stehen und ein ANN diese Daten auch nach der Messkampagne zur Verfügung hätte. Die Daten über den Wind wurden nicht verwendet, da diese Lücken aufweisen. Die Windgeschwindigkeit hat mit der Primärwellenhöhe keinen hohen Korrelationskoeffizienten wie noch in 4.3 gezeigt wird. Da in der Kampagne 2021 nur Primärwellen ausgewertet wurden, die größer als 5 cm sind, liegen für diese Passagen keine konkreten Werte vor, allein die Information, dass die Primärwelle unter 5 cm ist. Diese Datenlücken werden mit zufälligen Werten zwischen 0 cm und 5 cm aufgefüllt. Das Jahr der Messkampagne wird als kategorische Information in die Eingangsdaten eingegeben, so dass im ANN die Datensätze unterschiedlich behandelt werden können, falls dies dabei hilft, genauere Prognosen zu generieren.

Der Aufbau des ANNs ist

- Input Layer, 7 Eingangsdaten
- Dense Layer, 64 Knoten, Aktivierungsfunktion relu
- Dense Layer, 64 Knoten, Aktivierungsfunktion relu
- Dropout, Rate 0,5
- Output (Dense) Layer, 1 Knoten, keine Aktivierungsfunktion

Für eine Vorhersage der Primärwellenhöhe mit konkreten Werten wird im letzten Output-Layer keine Aktivierungsfunktion angegeben. Dadurch können Werte im gesamten Bereich der erwarteten Ausgabe erzeugt werden.

[2]

Die Leave-One-Out-Kreuzvalidierungs-Methode wird verwendet, um viele Daten für das Training zu nutzen und jede Schiffspassage auswerten zu können. Für jedes ANN, das zur Auswertung einer Schiffspassage trainiert wird, werden die trainierbaren Parameter mit den selben Zufallszahlen initialisiert. Das ANN Training wird beendet, wenn sich nach 20 Epochen keine weitere Verbesserung im Loss einstellt.

### Ergebnisse

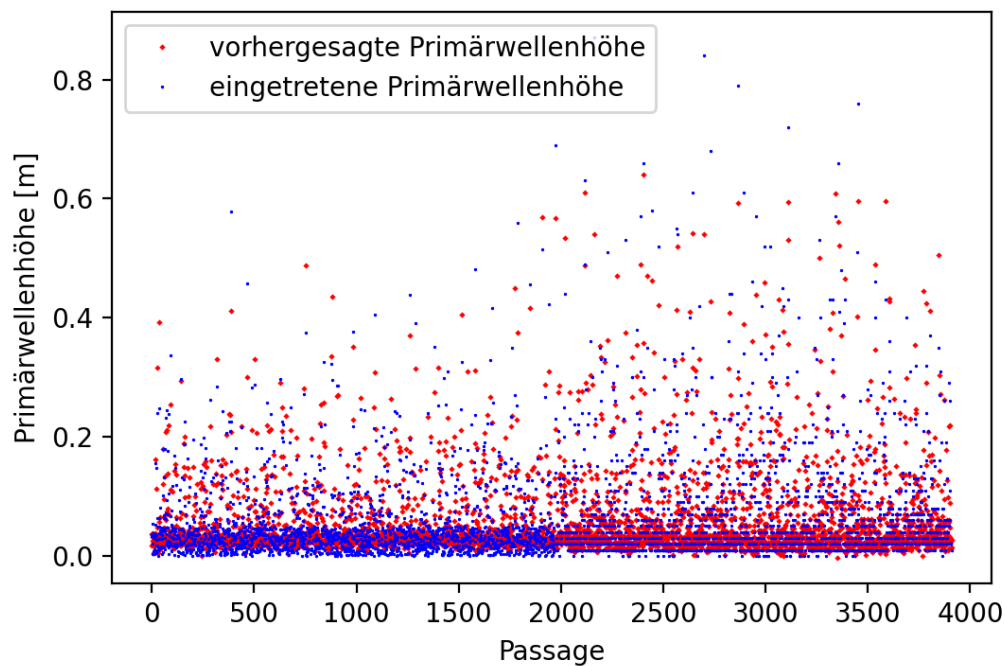


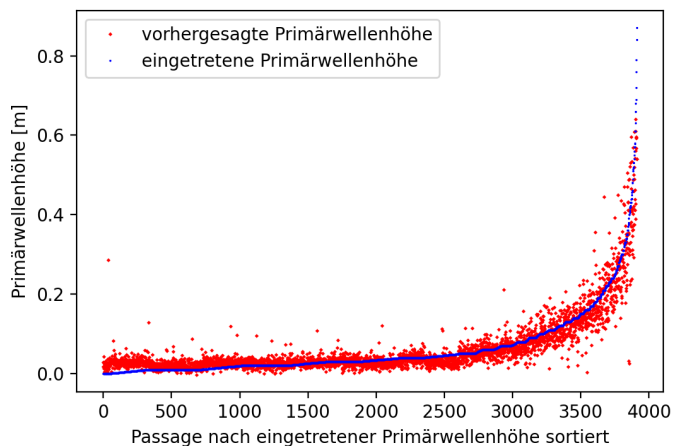
Abbildung 4.2.1.1: Eintretene Primärwellenhöhen verglichen mit vorhergesagter Primärwellenhöhe, 2014 und 2021

Differenz zwischen vorhergesagter und eingetretener Primärwellenhöhe	
Niedrigste Differenz	-37,45 cm
Größte Differenz	28,51 cm
Mittlere Differenz	-4,7 mm
Mittelwert der absoluten Differenz	1,80 cm
Standardabweichung der Differenzen	3,05 cm

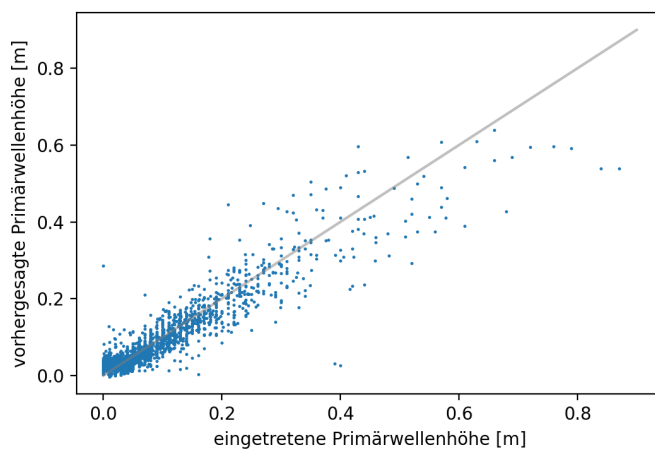
Tabelle 4.2.1.1: Vergleich der Differenzen zwischen vorhergesagter und eingetretener Primärwellenhöhe

Differenz zwischen vorhergesagter und eingetretener Primärwellenhöhe	
Niedrigste Differenz	-37,45 cm
Größte Differenz	16,62 cm
Mittlere Differenz	-6,09 cm
Mittelwert der absoluten Differenz	9,6 cm
Standardabweichung der Differenzen	10,68 cm

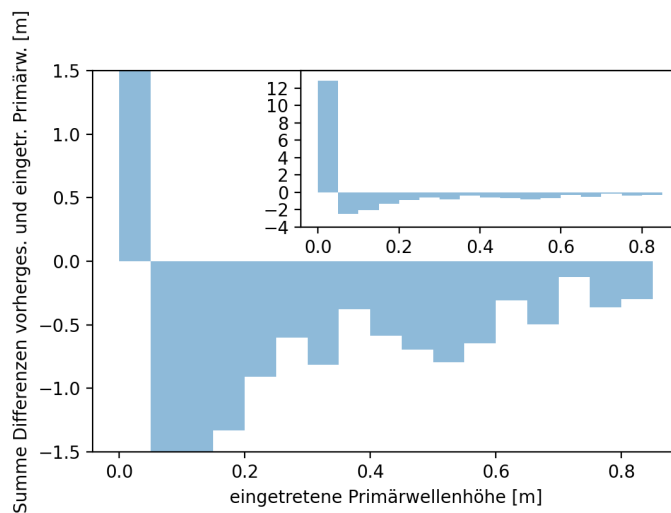
Tabelle 4.2.1.2: Vergleich der Differenzen zwischen vorhergesagter und eingetretener Primärwellenhöhe bei eingetretener Primärwellenhöhen über 30 cm



(a) Rangliste der eingetretenen Primärwellenhöhen verglichen mit Vorhersagen der Primärwellenhöhen



(b) Vorhersage Primärwellenhöhe über der eingetretenen Primärwellenhöhe



(c) Summe der Differenzen zwischen vorhergesagter und eingetretener Primärwellenhöhe in einzelnen Bereichen der eingetretenen Primärwellenhöhe

Abbildung 4.2.1.2: Ergebnisse, ANN mit großem Datensatz

Nummer	Differenz von vorhersage zu eingetretener Primärwellenhöhe	Primärwellenhöhe	Eintrittszeit	Name	Länge	Breite	Tiefgang	SOG	COG	Wasserstand Dwarsgat
-	[cm]	[cm]	-	-	[m]	[m]	[m]	[kn]	[°]	[m NHN]
1	28.51	0	21.10.2014 15:37	FINNWAVE	188	32	7.0	20.4	318	-0.78
2	23.46	21	11.12.2014 22:37	MAERSK LAGUNA	299	45	9.1	16.0	307	0.00
3	17.95	27	09.12.2021 10:42	SONGA PUMA	225	31	10.3	18.5	312.1	-1.86
4	17.62	18	23.11.2014 19:57	CMA CGM MARCO POLO	396	54	11.8	14.2	122	-1.64
5	16.62	43	07.12.2014 07:26	EVELYN MAERSK	398	56	11.4	13.3	312	-1.67
6	15.52	35	14.12.2014 01:17	MSC KRYSTAL	277	40	12.0	15.6	318	0.01
7	15.00	32	29.10.2014 22:42	COSCO NAGOYA	260	32	11.3	17.1	316	-1.49
8	14.39	25	15.10.2021 13:59	MAERSK LUZ	300	46	12.6	15.2	306.6	-0.71
9	14.09	30	11.11.2021 00:02	AIDACOSMA	344	54	8.9	15.5	307.1	-0.79
10	14.06	7	28.11.2014 23:01	RIO CADIZ	260	32	10.2	17.6	121	-2.20
11	13.12	18	02.12.2021 03:47	JAZAN	306	40	13.3	17.9	306.8	-0.29
12	12.38	23	10.12.2014 20:35	YM VANCOUVER	260	32	10.6	18.6	311	-0.15
13	12.10	35	03.11.2014 00:53	YORKTOWN EXPRESS	243	32	11.2	19.3	315	-0.65
14	11.85	nan	04.12.2021 06:34	MSC JORDAN	237	32	10.6	15.3	303.5	-1.72
15	11.75	31	07.12.2014 16:25	SEAGO BREMERHAVEN	294	32	11.9	19.8	319	0.01
16	11.21	38	06.11.2021 20:13	MAGLEBY MAERSK	399	60	11.6	15.5	306.4	-1.60
17	11.05	41	01.12.2014 01:07	WASHINGTON EXPRESS	243	32	10.1	18.5	313	-1.42
18	10.43	32	12.12.2014 09:23	GUDRUN MAERSK	367	43	12.1	15.6	314	-1.27
19	10.23	nan	24.10.2021 05:14	UMOE FIRMUS	27	10	3.2	33.4	306.9	-0.12
20	10.00	43	23.11.2014 19:18	MSC ILONA	300	40	10.8	17.2	313	-1.78
21	9.33	18	02.11.2014 15:54	SAFMARINE HIGHVELD	270	43	10.7	19.5	122	-0.55
22	9.31	44	22.10.2014 18:29	MAERSK PEG	210	32	9.2	20.4	319	-0.89
23	8.92	40	05.12.2014 16:59	MSC JULIA R.	281	32	9.4	19.0	316	-1.28
24	8.56	nan	20.10.2021 23:48	OOCL SEOUL	323	43	12.6	14.1	303.1	1.90
25	8.22	15	03.11.2021 15:02	MAERSK PENANG	210	32	9.9	20.4	317.3	-0.55
26	8.18	nan	12.12.2021 04:36	MSC ARUSHI R	281	32	12.4	13.8	299.5	1.00
27	8.13	nan	02.11.2021 07:39	XIN CHANG SHU	280	40	14.0	13.1	303.9	0.99
28	8.07	11	09.12.2014 20:49	RUTH	134	23	8.2	18.3	319	-2.05
29	8.01	32	01.12.2021 04:43	COSCO PHILIPPINES	334	43	12.3	14.7	306.9	-0.90
30	7.90	14	09.11.2021 09:44	PROMETHEUS LEADER	190	28	8.5	17.0	312.1	-2.00

Tabelle 4.2.1.3: Größte Differenzen zwischen vorhergesagter und eingetretener Primärwelle

Nummer	Differenz von vorhersage zu eingetretener Primärwellenhöhe	Primärwellenhöhe	Eintrittszeit	Name	Länge	Breite	Tiefgang	SOG	COG	Wasserstand Dwarsgat
-	[cm]	[cm]	-	-	[m]	[m]	[m]	[kn]	[°]	[m NHN]
1	-37.45	40	25.11.2014 06:00	CERES	140	23	7.3	13.4	125	-0.83
2	-35.85	39	01.12.2014 21:15	MV SONDERBORG STRAIT	158	24	7.7	13.9	124	0.49
3	-33.01	87	26.10.2014 20:15	MSC KRYSTAL	277	40	11.0	18.9	319	-1.39
4	-30.03	84	10.11.2014 22:19	MARIBO MAERSK	399	59	11.5	15.2	315	-0.90
5	-25.31	68	11.11.2014 18:53	CHOAPA TRADER	294	32	11.0	20.7	319	-0.75
6	-22.73	52	20.11.2014 18:11	MSC ANGELA	265	32	7.7	18.3	318	-1.57
7	-22.06	61	17.11.2014 03:23	MSC SORAYA	277	40	10.1	16.3	302	-0.48
8	-20.26	44	14.11.2014 18:20	CSAV LARAQUETE	261	32	10.6	19.2	317	0.48
9	-19.82	79	15.11.2014 23:41	HS BACH	246	32	11.3	19.3	318	-1.16
10	-19.17	42	06.12.2021 04:16	COSCO JAPAN	334	43	12.6	16.4	129.0	-0.49
11	-18.71	42	19.11.2014 07:25	CMA CGM MUSCA	347	45	11.7	15.3	316	0.51
12	-18.56	56	09.12.2021 22:50	MAERSK PALERMO	210	32	9.6	18.9	306.3	-1.76
13	-17.05	48	03.12.2021 04:08	MAERSK BULAN	223	32	11.7	18.9	306.7	-1.28
14	-16.71	58	26.10.2021 08:24	MARY MAERSK	399	60	12.3	16.8	309.0	-0.99
15	-16.39	76	03.12.2014 16:02	CMA CGM ANDROMEDA	363	46	11.8	16.6	319	-1.80
16	-15.82	46	28.10.2021 22:29	MSC HAMBURG	399	54	11.4	16.7	304.2	-1.18
17	-15.66	16	25.11.2014 19:17	FREDO	82	11	2.4	12.4	311	-1.42
18	-15.48	53	30.10.2014 22:42	MSC ARUSHI R.	281	32	11.6	17.4	310	-1.90
19	-15.23	33	06.11.2014 16:40	HOEGH NEW YORK	199	32	9.6	18.0	319	-1.32
20	-14.99	29	18.11.2014 06:09	CMA CGM MUSCA	347	45	12.2	18.2	126	0.38
21	-14.80	51	28.10.2014 17:02	QUANTUM OF THE SEAS	347	41	8.6	21.0	316	0.87
22	-13.68	55	07.11.2014 04:39	MAERSK LANCO	299	45	10.6	18.3	319	-1.28
23	-13.25	34	19.11.2014 06:34	MAERSK PEG	210	32	9.0	17.6	317	-0.02
24	-13.18	57	18.11.2014 15:21	MSC DIEGO	259	32	9.2	18.9	317	-1.74
25	-13.11	25	22.10.2014 11:30	CHRISTOPHER	170	26	8.2	12.4	121	3.49
26	-12.55	72	23.11.2014 18:45	SEAGO ANTWERP	293	32	11.6	18.9	319	-1.68
27	-12.21	69	21.10.2014 05:59	MSC METHONI	304	40	9.9	17.9	317	-0.98
28	-12.09	30	27.11.2021 20:37	COSCO PHILIPPINES	334	43	11.0	17.7	126.5	0.02
29	-11.98	14	05.12.2014 18:53	LEV TAIFUN	41	8	2.0	7.8	123	-1.30
30	-11.78	34	17.11.2014 03:13	MAERSK EFFINGHAM	366	48	11.1	19.1	123	-0.64

Tabelle 4.2.1.4: Kleinste Differenzen zwischen vorhergesagter und eingetretener Primärwelle

**Diskussion** In Abbildung 4.2.1.1 nahe Passage 2000 ist die Zusammenführung der Datensätze aus 2021 und 2014 zu erkennen. Die Primärwellenhöhen sind 2021 deutlich kleiner. Im Datensatz von 2014 ist in den eingetretenen Primärwellenhöhen die Genauigkeit von 1 cm Schrittweite in den Punkten zu erkennen. In Abbildung 4.2.1.2a wird eine Abschätzung der Qualität der Vorhersageergebnisse besser deutlich. Die Vorhersagen sind bereits sehr genau mit einer Standardabweichung von nur 3,05 cm. In Abbildung 4.2.1.2a ist zu sehen, dass die Steigung der eingetretenen Primärwellenhöhe bis 5 cm linear verläuft, was durch die gleichmäßig aufgefüllten Zufallswerte von 0 cm bis 5 cm kommt. In den Vorhersagen ist ebenfalls ein Unterschied zwischen diesen beiden Abschnitten durch den sprunghaften Anstieg bei einer Primärwellenhöhe von 5 cm zu erkennen. Die Vorhersagen für eingetretene Primärwellenhöhen unter 5 cm fallen in einem engeren Band aus mit nur wenigen Abweichungen. Bei den Vorhersagen für eingetretene Primärwellenhöhen über 5 cm ist eine größere Streuung zu erkennen. Man erkennt auch, dass die Primärwellenhöhe eher zu niedrig prognostiziert wird. In Abbildung 4.2.1.2b sieht man bei den größeren Primärwellen (über 5 cm), dass die Vorhersage öfter zu niedrig ausfällt als zu hoch. Die Differenz von Vorhersage zu eingetretene Wert ist im Mittel nur bei -5,7 mm. Betrachtet man jedoch nur die eingetretenen Primärwellen über 30 cm, so ist die Differenz im Mittel bei -6,09 cm. Das bedeutet auch, dass niedrigere Primärwellen höher vorhergesagt werden. Man sieht, dass die Primärwellen über 60 cm alle unterschätzt werden. In Abbildung 4.2.1.2c sieht man die summierten Differenzen zwischen eingetretener und vorhergesagter Primärwellenhöhe in 5 cm Bereichen. Es ist zu erkennen, dass die Primärwellen mit einer eingetretenen Höhe von 5 cm oft überschätzt werden. Die Primärwellenhöhen über 5 cm werden meistens unterschätzt. In Tabelle 4.2.1.3 ist die größte Differenz zwischen vorhergesagter und eingetretener Primärwelle die FINNWAVE mit einer extrem kleinen Primärwellenhöhe von 0 cm. Die Geschwindigkeit (SOG) mit 20,4 kn und die Länge lassen eine hohe Primärwelle vermuten. Der Tiefgang mit 7 m ist jedoch relativ gering. Ein Fehler in der Datenbank liegt nicht vor, da die FINNWAVE am selben Tag eine Passage mit einer Primärwellenhöhe von 2 cm hatte mit einer Geschwindigkeit (SOG) von 12 kn. Es kann sein, dass die FINNWAVE mit sehr geringer Ladung gefahren ist und dieser Fall selten im Datensatz auftritt. Das ANN kann dann dieses Muster nicht erkennen. In den 30 größten Differenzen sind 12 eingetretene Primärwellenhöhen über 30 cm und damit überproportional vertreten. Bei den 30 kleinsten Differenzen sind 26 eingetretene Primärwellenhöhen über 30 cm. Dies zeigt, dass höheren Primärwellen deutlich ungenauer vorhergesagt werden.

### 4.2.2 Reduzierter Datensatz

Die Datensätze aus 2014 und 2021 werden verwendet. Die Dateneinträge aus der Kampagne 2021 mit Primärwellenhöhen unter 5 cm werden weggelassen. Es soll untersucht werden, welchen Einfluss das Auffüllen der nicht ausgewerteten Primärwellenhöhen in der Messkampagne 2021 hat.

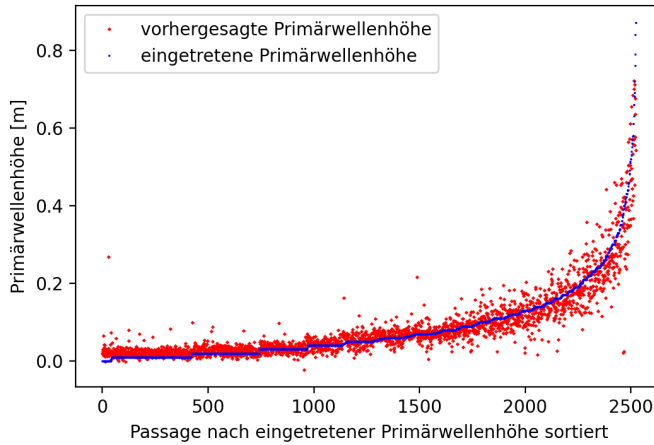
**Aufbau** Es werden 2523 Passagen in der Leave-One-Out-Kreuzvalidierungs-Methode untersucht. Bis auf die Anzahl und Wahl der Passagen entspricht der Aufbau dem in 4.2.1.

**Diskussion** Die Vorhersagegenauigkeiten aus Tabelle 4.2.2.1 sind vergleichbar mit denen in Tabelle 4.2.1.1. In Abbildung 4.2.2.1a ist zu sehen, dass die Vorhersagewerte im Bereich unter 5 cm leicht besser mit den eingetretenen Primärwellenhöhen übereinstimmen als in Abbildung 4.2.1.2a. Der Sprung bei 5 cm Primärwellenhöhe entfällt. In Abbildung 4.2.2.1c ist zu erkennen, dass die Primärwellen zwischen 5 cm und 10 cm im Durchschnitt überschätzt werden im Gegensatz zu 4.2.1.2c. Die Primärwellen werden wie in 4.2.1 bei hohen Werten tendenziell zu niedrig vorhergesagt. In Tabelle 4.2.2.2 sind größtenteils dieselben Schiffe vertreten wie in 4.2.1.3. Die FINNWAVE ist wieder das Schiff mit der schlechtesten Prognose in der Tabelle. Die darauf folgenden Plätze haben sich leicht geändert. Generell sind die Differenzen von vorhergesagter und eingetretener Primärwelle in Tabelle 4.2.2.2 kleiner als in Tabelle 4.2.1.3. In Tabelle 4.2.2.3 sind in den ersten drei Plätze mit den ungenauesten Prognosen dieselben Schiffe vertreten wie in Tabelle 4.2.1.4. Die nachfolgenden Ränge haben nur leicht gewechselt. Die Differenzen zwischen vorhergesagter und eingetretener Primärwelle sind nahezu identisch. Da der Datensatz verhältnismäßig viele Passagen mit einer Primärwelle unter 5 cm hat, kann auf das Auffüllen der nicht ausgewerteten Primärwellenhöhen verzichtet werden.

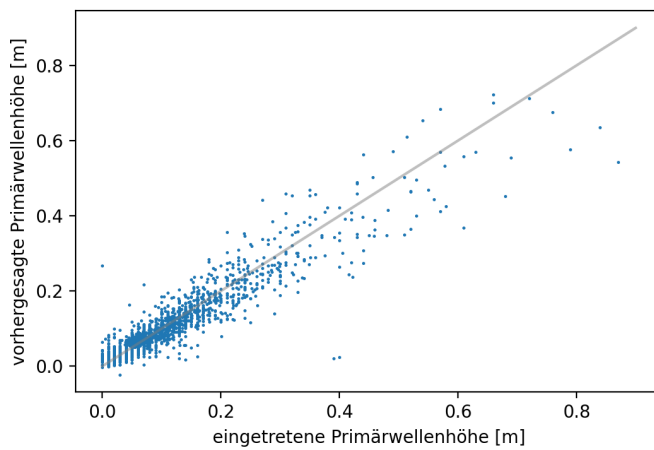
### Ergebnisse

Differenz zwischen vorhergesagter und eingetretener Primärwellenhöhe	
Niedrigste Differenz	-37,51 cm
Größte Differenz	26,75 cm
Mittlere Differenz	0,13 mm
Mittelwert der absoluten Differenz	2,00 cm
Standardabweichung der Differenzen	3,36 cm

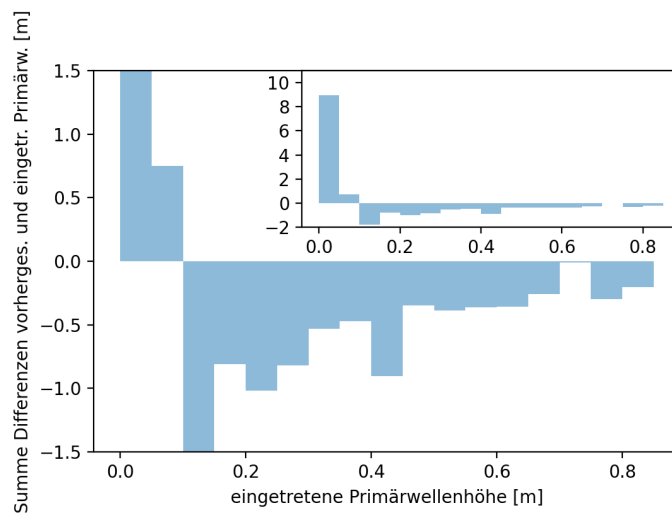
Tabelle 4.2.2.1: Vergleich der Differenzen zwischen vorhergesagter und eingetretener Primärwellenhöhe



(a) Rangliste der eingetretenen Primärwellenhöhen verglichen mit Vorhersagen der Primärwellenhöhen



(b) Vorhersage Primärwellenhöhe über der eingetretenen Primärwellenhöhe



(c) Summe der Differenzen zwischen vorhergesagter und eingetretener Primärwellenhöhe in einzelnen Bereichen der eingetretenen Primärwellenhöhe

Abbildung 4.2.2.1: Ergebnisse, ANN mit reduziertem Datensatz



Nummer	Differenz von vorhersage zu eingetretener Primärwellenhöhe	Primärwellenhöhe	Eintrittszeit	Name	Länge	Breite	Tiefgang	SOG	COG	Wasserstand Dwarsgat
-	[cm]	[cm]	-	-	[m]	[m]	[m]	[kn]	[°]	[m NHN]
1	26.75	0	21.10.2014 15:37	FINNWAVE	188	32	7.0	20.4	318	-0.78
2	17.25	27	09.12.2021 10:42	SONGA PUMA	225	31	10.3	18.5	312.1	-1.86
3	14.90	31	07.12.2014 16:25	SEAGO BREMERHAVEN	294	32	11.9	19.8	319	0.01
4	14.74	21	11.12.2014 22:37	MAERSK LAGUNA	299	45	9.1	16.0	307	0.00
5	14.67	7	28.11.2014 23:01	RIO CADIZ	260	32	10.2	17.6	121	-2.20
6	13.35	32	29.10.2014 22:42	COSCO NAGOYA	260	32	11.3	17.1	316	-1.49
7	12.30	44	22.10.2014 18:29	MAERSK PEG	210	32	9.2	20.4	319	-0.89
8	11.87	35	14.12.2014 01:17	MSC KRYSTAL	277	40	12.0	15.6	318	0.01
9	11.81	4	07.11.2021 15:02	VISURGIS	19	6	1.9	27.9	303.8	2.79
10	11.47	54	07.11.2014 07:01	JPO CANOPUS	264	32	10.8	18.0	320	-2.20
11	11.38	24	15.10.2021 08:43	COSCO VIETNAM	334	43	12.8	19.0	305.7	1.56
12	11.33	57	02.11.2014 01:28	BUXCOAST	300	40	12.3	17.1	315	-1.22
13	10.61	24	01.11.2014 13:49	SEAGO BREMERHAVEN	293	32	12.8	18.9	123	-1.12
14	10.41	18	17.11.2014 16:30	MOL PRESENCE	293	40	10.8	14.6	316	-0.46
15	10.32	35	03.11.2014 00:53	YORKTOWN EXPRESS	243	32	11.2	19.3	315	-0.65
16	9.79	18	23.11.2014 19:57	CMA CGM MARCO POLO	396	54	11.8	14.2	122	-1.64
17	9.75	36	07.12.2014 18:57	NYK DIANA	295	32	11.4	17.6	315	-1.18
18	9.74	29	29.10.2021 22:08	HUNGARY	298	48	10.5	18.6	306.6	-0.45
19	9.65	31	24.11.2014 23:28	CMA CGM MARCO POLO	396	54	11.4	14.8	318	0.95
20	9.60	51	14.12.2021 01:21	MSC ELODIE	300	48	12.5	17.9	308.5	-0.86
21	9.22	21	21.10.2014 13:48	MSC VAISHNAVI R.	281	32	11.6	19.4	319	0.16
22	8.55	11	09.12.2014 20:49	RUTH	134	23	8.2	18.3	319	-2.05
23	8.48	10	02.12.2014 21:22	MAERSK NIMES	210	30	10.6	17.4	308	1.45
24	8.21	16	15.10.2021 01:11	TORONTO	200	32	9.0	17.8	311.0	-0.59
25	8.21	49	25.10.2014 07:42	NYK NEBULA	294	32	10.9	17.8	319	-1.72
26	8.21	10	03.11.2021 00:23	BRUSSELS	278	40	13.5	17.2	309.6	1.07
27	8.01	15	26.11.2014 00:49	MSC METHONI	304	40	11.0	15.3	318	0.99
28	7.93	2	23.10.2014 21:27	AMERDIJK	168	26	9.5	14.2	319	0.62
29	7.88	23	10.12.2014 20:35	YM VANCOUVER	260	32	10.6	18.6	311	-0.15
30	7.44	7	26.10.2014 09:52	MAERSK NIAMEY	210	30	9.8	13.4	314	-0.24

Tabelle 4.2.2.2: Größte Differenzen von vorhergesagter und eingetretener Primärwelle

#### 4 Numerische Experimente und Ergebnisse

Nummer	Differenz von vorhersage zu eingetretener Primärwellenhöhe	Primärwellenhöhe	Eintrittszeit	Name	Länge	Breite	Tiefgang	SOG	COG	Wasserstand Dwarsgat
-	[cm]	[cm]	-	-	[m]	[m]	[m]	[kn]	[°]	[m NHN]
1	-37.51	40	25.11.2014 06:00	CERES	140	23	7.3	13.4	125	-0.83
2	-36.85	39	01.12.2014 21:15	MV SONDERBORG STRAIT	158	24	7.7	13.9	124	0.49
3	-32.76	87	26.10.2014 20:15	MSC KRYSTAL	277	40	11.0	18.9	319	-1.39
4	-24.23	61	17.11.2014 03:23	MSC SORAYA	277	40	10.1	16.3	302	-0.48
5	-22.71	68	11.11.2014 18:53	CHOAPA TRADER	294	32	11.0	20.7	319	-0.75
6	-21.31	79	15.11.2014 23:41	HS BACH	246	32	11.3	19.3	318	-1.16
7	-20.49	84	10.11.2014 22:19	MARIBO MAERSK	399	59	11.5	15.2	315	-0.90
8	-18.44	42	15.12.2021 03:34	VERA D	179	27	8.7	20.6	311.2	-1.36
9	-17.24	42	06.12.2021 04:16	COSCO JAPAN	334	43	12.6	16.4	129.0	-0.49
10	-16.57	44	14.11.2014 18:20	CSAV LARAQUETE	261	32	10.6	19.2	317	0.48
11	-16.06	51	28.10.2014 17:02	QUANTUM OF THE SEAS	347	41	8.6	21.0	316	0.87
12	-15.73	57	18.11.2014 15:21	MSC DIEGO	259	32	9.2	18.9	317	-1.74
13	-15.51	58	03.11.2014 17:19	SAFMARINE HIGHVELD	270	43	11.9	15.6	315	-0.11
14	-15.45	52	04.11.2014 15:43	RIO THOMPSON	210	30	9.3	20.6	319	-1.37
15	-15.03	29	18.11.2014 06:09	CMA CGM MUSCA	347	45	12.2	18.2	126	0.38
16	-14.37	42	19.11.2014 07:25	CMA CGM MUSCA	347	45	11.7	15.3	316	0.51
17	-13.66	40	09.12.2014 21:07	MAERSK KAMPALA	300	43	12.4	16.1	123	-2.00
18	-13.46	69	21.10.2014 05:59	MSC METHONI	304	40	9.9	17.9	317	-0.98
19	-13.44	25	22.10.2014 11:30	CHRISTOPHER	170	26	8.2	12.4	121	3.49
20	-13.34	16	25.11.2014 19:17	FREDO	82	11	2.4	12.4	311	-1.42
21	-13.28	44	14.11.2014 11:46	CMA CGM LAPEROUSE	365	51	12.2	15.2	121	-1.90
22	-13.26	33	06.11.2014 16:40	HOEGH NEW YORK	199	32	9.6	18.0	319	-1.32
23	-13.22	48	01.12.2014 10:05	MSC MATILDE	294	32	9.8	20.1	318	-0.23
24	-12.93	53	30.10.2014 22:42	MSC ARUSHI R.	281	32	11.6	17.4	310	-1.90
25	-12.31	14	05.12.2014 18:53	LEV TAIFUN	41	8	2.0	7.8	123	-1.30
26	-12.06	20	05.11.2021 03:36	LOTUS A	336	43	13.0	12.9	129.2	-0.11
27	-11.67	56	09.12.2021 22:50	MAERSK PALERMO	210	32	9.6	18.9	306.3	-1.76
28	-11.24	24	18.11.2014 01:27	LUCKY SUNDAY	229	32	8.5	15.3	316	-1.02
29	-11.18	46	28.10.2021 22:29	MSC HAMBURG	399	54	11.4	16.7	304.2	-1.18
30	-11.16	30	11.11.2014 21:39	RIO CARDIFF	262	32	9.3	17.2	126	-2.15

Tabelle 4.2.2.3: Niedrigste Differenzen von vorhergesagter und eingetretener Primärwelle

### 4.2.3 Einzelanalyse Datensatz 2014

Der Datensatz aus 2014 wird getrennt analysiert, um Unterschiede zwischen den Datensätzen zu erkennen.

**Aufbau** Es werden 1943 Passagen in der Leave-One-Out-Kreuzvalidierungs-Methode untersucht. Bis auf die Anzahl und Wahl der Passagen entspricht der Aufbau dem in 4.2.1.

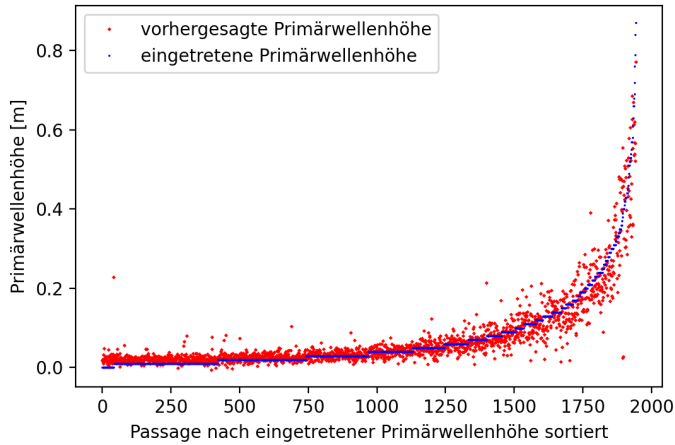
**Diskussion** Die Differenz zwischen vorhergesagten und eingetretenen Werten ist bei der niedrigsten Differenz, der Standardabweichung der Differenzen und dem Mittelwert der absoluten Differenz vergleichbar mit 4.2.1 und 4.2.2. Die größte Differenz ist um 3,9 cm im Vergleich zu 4.2.2 verbessert. Die mittlere Differenz ist bei nur noch 0,06 mm und damit niedriger als bei 4.2.1 und 4.2.2.

In Abbildung 4.2.3.1c ist zu erkennen, dass die Primärwellen zwischen 5 cm und 10 cm im Durchschnitt unterschätzt werden wie in 4.2.1.2c und nicht wie in 4.2.2.1c.

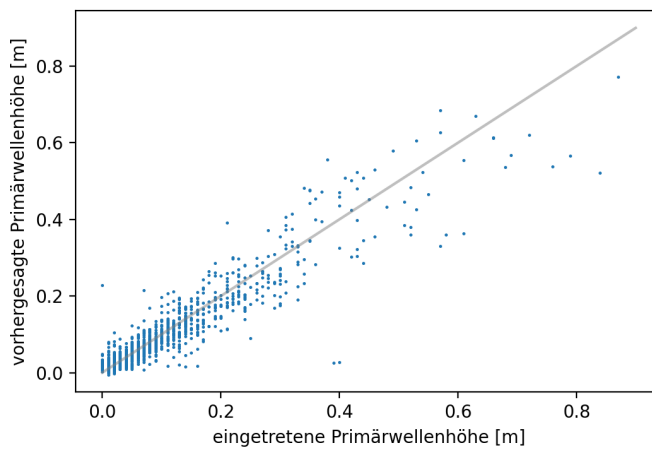
### Ergebnisse

Differenz zwischen vorhergesagter und eingetretener Primärwellenhöhe	
Niedrigste Differenz	-37,21 cm
Größte Differenz	22,82 cm
Mittlere Differenz	0,06 mm
Mittelwert der absoluten Differenz	1,92 cm
Standardabweichung der Differenzen	3,46 cm

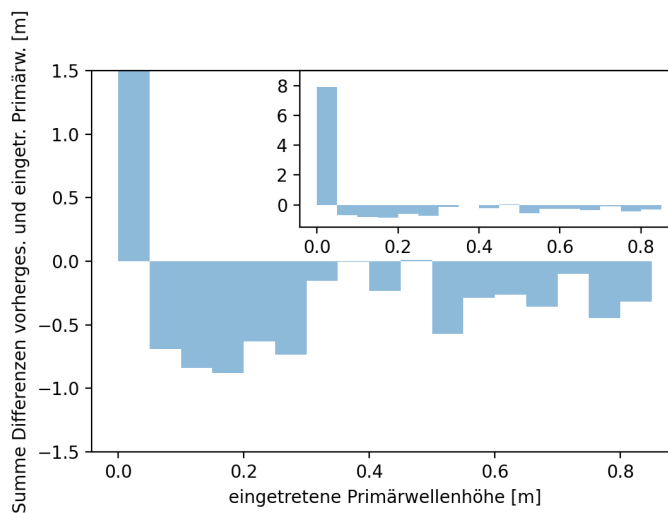
Tabelle 4.2.3.1: Vergleich der Differenzen zwischen vorhergesagter und eingetretener Primärwellenhöhe



(a) Rangliste der eingetretenen Primärwellenhöhen verglichen mit Vorhersagen der Primärwellenhöhen



(b) Vorhersage Primärwellenhöhe über der eingetretenen Primärwellenhöhe



(c) Summe der Differenzen zwischen vorhergesagter und eingetretener Primärwellenhöhe in einzelnen Bereichen der eingetretenen Primärwellenhöhe

Abbildung 4.2.3.1: Ergebnisse, Einzelanalyse Datensatz 2014

#### 4.2.4 Einzelanalyse Datensatz 2021

Der Datensatz aus 2021 wird getrennt analysiert, um Unterschiede zwischen den Datensätzen zu erkennen.

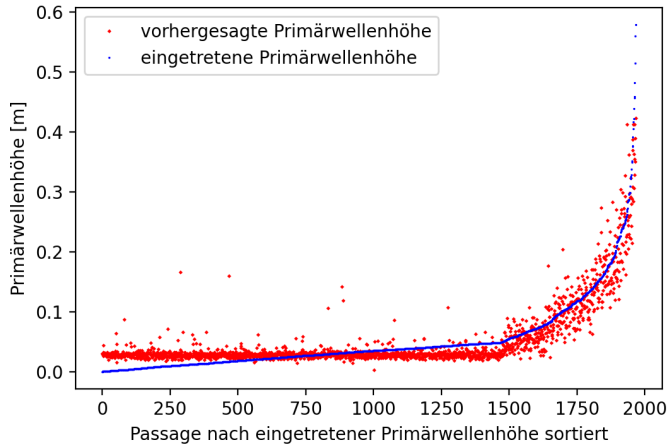
**Aufbau** Es werden 1967 Passagen in der Leave-One-Out-Kreuzvalidierungs-Methode untersucht. Bis auf die Anzahl und Wahl der Passagen entspricht der Aufbau dem in 4.2.1.

**Diskussion** Die niedrigste und größte Differenz zwischen vorhergesagter und eingetretener Primärwellenhöhe sind weniger stark ausgeprägt als in allen Untersuchungen zuvor. Hauptgrund dafür ist, dass sowohl die meist überschätzte, als auch die meist unterschätzte Primärwellenhöhe aus dem Datensatz von 2014 stammt. Die Standardabweichung ist mit 2,36 cm deutlich geringer als in 4.2.3.1 mit 3,46 cm. Ein Grund dafür sind die geringeren Primärwellenhöhen, die dadurch kommen, dass die Schiffe 2021 langsamer fahren. Das ANN hat einen kleineren Vorhersagebereich auf das es trainiert werden muss und erlangt dadurch auch bessere Ergebnisse.

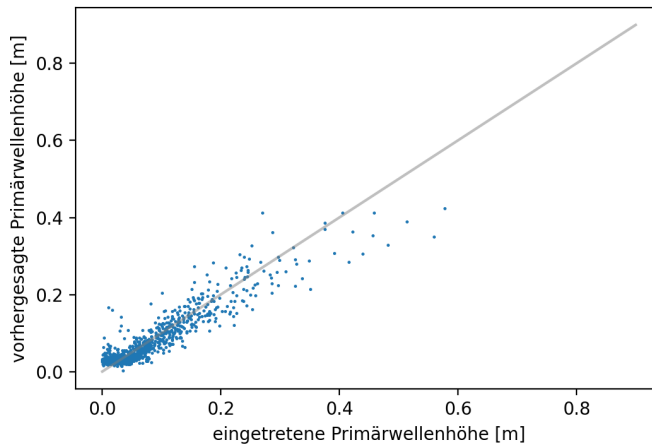
#### Ergebnisse

Differenz zwischen vorhergesagter und eingetretener Primärwellenhöhe	
Niedrigste Differenz	-21,03 cm
Größte Differenz	15,57 cm
Mittlere Differenz	0,03 mm
Mittelwert der absoluten Differenz	1,65 cm
Standardabweichung der Differenzen	2,36 cm

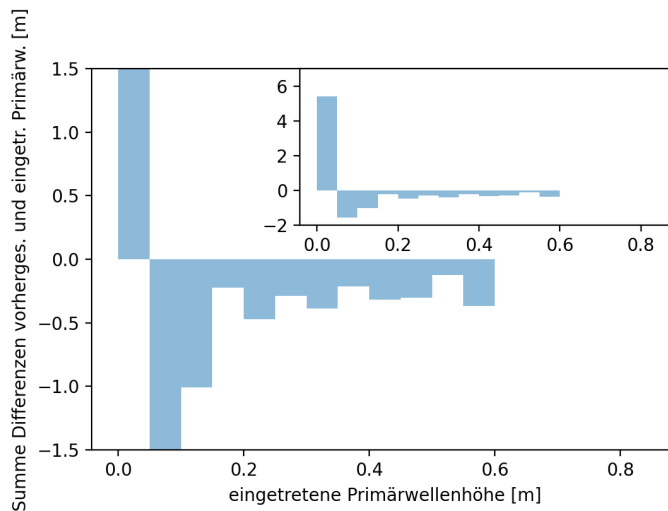
Tabelle 4.2.4.1: Vergleich der Differenzen zwischen vorhergesagter und eingetretener Primärwellenhöhe



(a) Rangliste der eingetretenen Primärwellenhöhen verglichen mit Vorhersagen der Primärwellenhöhen



(b) Vorhersage Primärwellenhöhe über der eingetretenen Primärwellenhöhe



(c) Summe der Differenzen zwischen vorhergesagter und eingetretener Primärwellenhöhe in einzelnen Bereichen der eingetretenen Primärwellenhöhe

Abbildung 4.2.4.1: Ergebnisse, Einzelanalyse Datensatz 2021

### 4.2.5 Einzelanalyse reduzierter Datensatz 2021

Der reduzierte Datensatz aus 2021 wird getrennt analysiert, um Unterschiede zwischen den Datensätzen zu erkennen.

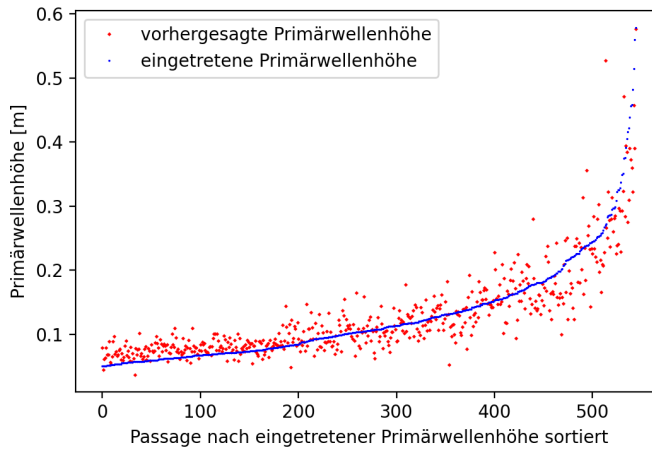
**Aufbau** Es werden 545 Passagen in der Leave-One-Out-Kreuzvalidierungs-Methode untersucht. Bis auf die Anzahl und Wahl der Passagen entspricht der Aufbau dem in 4.2.1.

**Diskussion** Vergleicht man Tabelle 4.2.4.1 und Tabelle 4.2.5.1 sieht man, dass die niedrigste Differenz in Tabelle 4.2.4.1 stärker ausgeprägt ist als in Tabelle 4.2.5.1. Bei der größten Differenz ist es anders herum. Dies kommt daher, dass die aufgefüllten Werte im Bereich 0 cm bis 5 cm die Vorhersageergebnisse nach unten beeinflussen. Die Standardabweichung der Differenzen zwischen vorhergesagter und eingetretener Primärwelle ist in Tabelle 4.2.4.1 mit 2,36 cm niedriger als in dieser Untersuchung mit 3,37 cm. Denkbar ist, dass der kleinere Datensatz die Qualität der Vorhersageergebnisse senkt. Da die Zufallswerte nur über einen Bereich von 5 cm streuen, sollten die Prognosen eine geringere Standardabweichung haben als 3,37 cm. Die Abweichung ist bei den größeren Wellen stärker ausgeprägt. Durch den geringeren Datensatz werden diese in der Berechnung der gesamten Standardabweichung stärker gewichtet. In Abbildung 4.2.5.1c sieht man, dass der Bereich von 5 cm bis 10 cm überschätzt wird, der Bereich ab 10 cm wird unterschätzt. Die Verteilung ist ähnlich der vorhergehenden Untersuchungen nur um 5 cm versetzt dadurch, dass der Bereich mit 0 cm bis 5 cm entfällt.

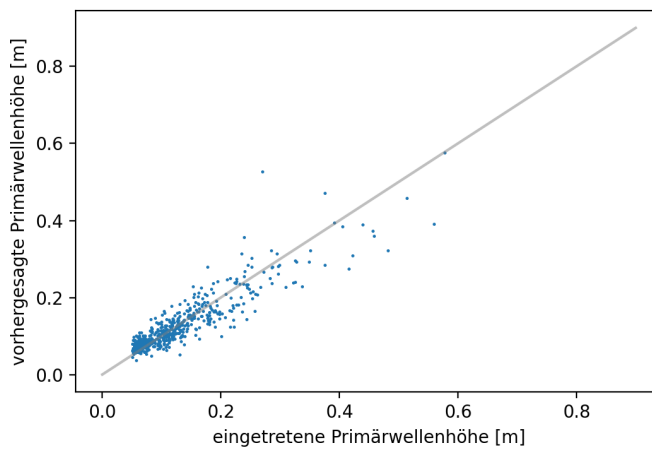
### Ergebnisse

Differenz zwischen vorhergesagter und eingetretener Primärwellenhöhe	
Niedrigste Differenz	-16,92 cm
Größte Differenz	25,73 cm
Mittlere Differenz	0,02 mm
Mittelwert der absoluten Differenz	2,28 cm
Standardabweichung der Differenzen	3,37 cm

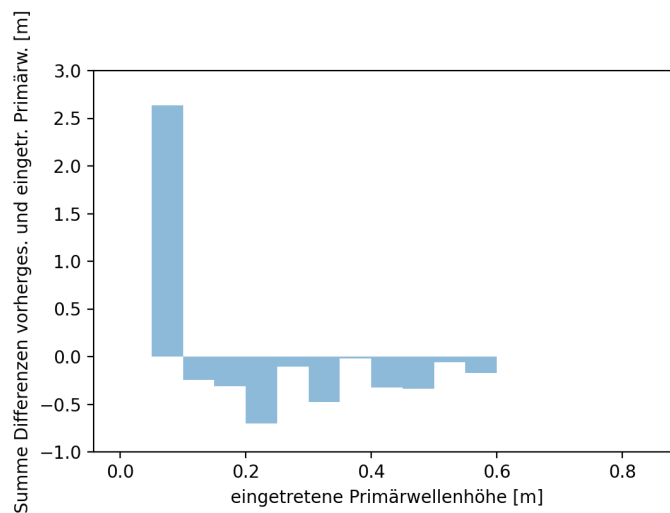
Tabelle 4.2.5.1: Vergleich der Differenzen zwischen vorhergesagter und eingetretener Primärwellenhöhe



(a) Rangliste der eingetretenen Primärwellenhöhen verglichen mit Vorhersagen der Primärwellenhöhen



(b) Vorhersage Primärwellenhöhe über der eingetretenen Primärwellenhöhe



(c) Summe der Differenzen zwischen vorhergesagter und eingetretener Primärwellenhöhe in einzelnen Bereichen der eingetretenen Primärwellenhöhe

Abbildung 4.2.5.1: Ergebnisse, Einzelanalyse reduzierter Datensatz 2021



### 4.2.6 Einzelanalyse Datensatz mit Primärwellen über 10 cm

Die Datensätze aus 2021 und 2014 werden verwendet, wobei nur Passagen mit einer Primärwellenhöhe über 10 cm ausgewertet werden.

**Aufbau** Es werden 760 Passagen in der Leave-One-Out-Kreuzvalidierungs-Methode untersucht. Bis auf die Anzahl und Wahl der Passagen entspricht der Aufbau dem in 4.2.1.

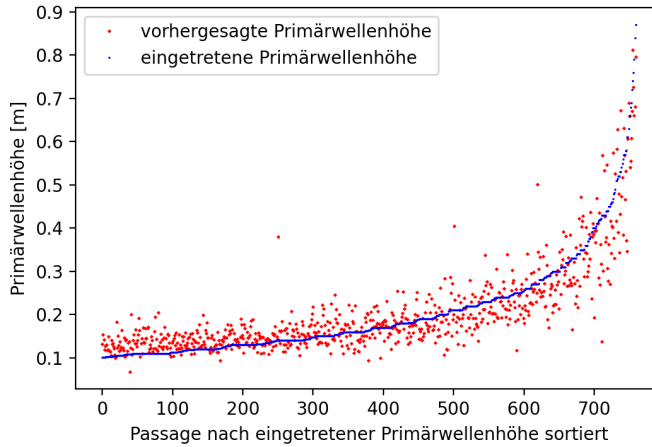
**Diskussion** In Abbildung 4.2.6.1c sieht man ähnlich wie bei 4.2.5.1c, dass die verhältnismäßig kleinen Primärwellen überschätzt und die größeren Primärwellen unterschätzt werden. Die Vorhersagegenauigkeiten in Tabelle 4.2.6.1 haben sich durch den kleineren Trainingsdatensatz im Vergleich zu allen vorhergehenden Untersuchungen verschlechtert.

#### Ergebnisse

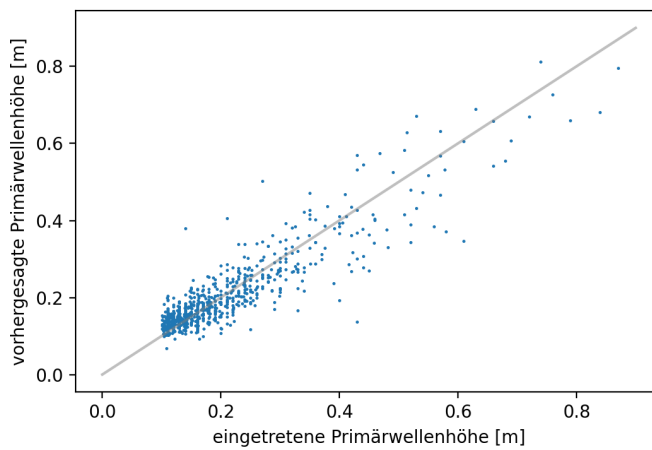
Differenz zwischen vorhergesagter und eingetretener Primärwellenhöhe	
Niedrigste Differenz	-29,27 cm
Größte Differenz	23,96 cm
Mittlere Differenz	0,07 mm
Mittelwert der absoluten Differenz	3,48 cm
Standardabweichung der Differenzen	5,01 cm

Tabelle 4.2.6.1: Vergleich der Differenzen zwischen vorhergesagter und eingetretener Primärwellenhöhe

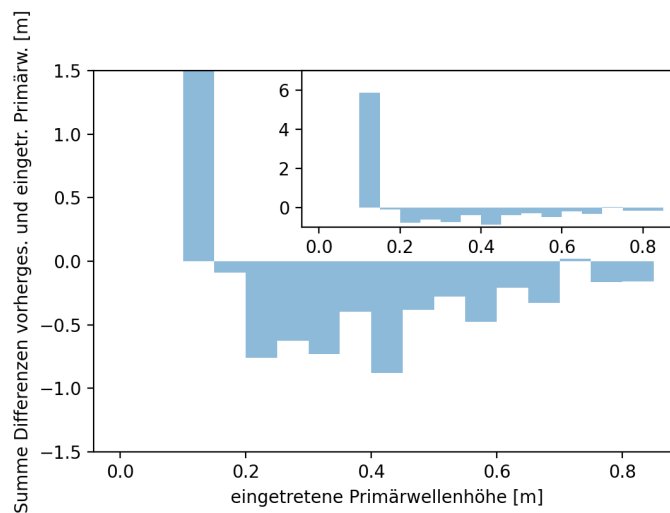
#### 4 Numerische Experimente und Ergebnisse



(a) Rangliste der eingetretenen Primärwellenhöhen verglichen mit Vorhersagen der Primärwellenhöhen



(b) Vorhersage Primärwellenhöhe über der eingetretenen Primärwellenhöhe



(c) Summe der Differenzen zwischen vorhergesagter und eingetretener Primärwellenhöhe in einzelnen Bereichen der eingetretenen Primärwellenhöhe

Abbildung 4.2.6.1: Ergebnisse, Einzelanalyse Datensatz mit Primärwellen über 10 cm

### 4.2.7 Einzelanalyse Datensatz mit Primärwellen unter 10 cm

Die Datensätze aus 2021 und 2014 werden verwendet, wobei nur Passagen mit einer Primärwellenhöhe unter 10 cm ausgewertet werden.

**Aufbau** Es werden 1902 Passagen in der Leave-One-Out-Kreuzvalidierungs-Methode untersucht. Bis auf die Anzahl und Wahl der Passagen entspricht der Aufbau dem in 4.2.1.

**Diskussion** In Abbildung 4.2.6.1c sieht man, dass auch bei der Untersuchung kleiner Primärwellen, die kleinen Primärwellen des Untersuchungsbereiches überschätzt und die größeren des Untersuchungsbereiches unterschätzt werden.

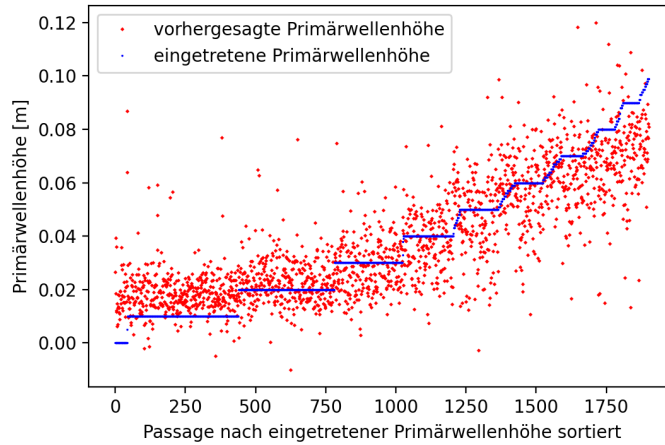
Die Genauigkeiten der Vorhersagen sind besser als die aller vorhergehenden Untersuchungen. Der Vorhersagebereich ist allerdings auch der kleinste und es liegt mit 1902 Passagen noch eine Datensatzgröße vor, mit der ein genaues ANN trainiert werden kann.

### Ergebnisse

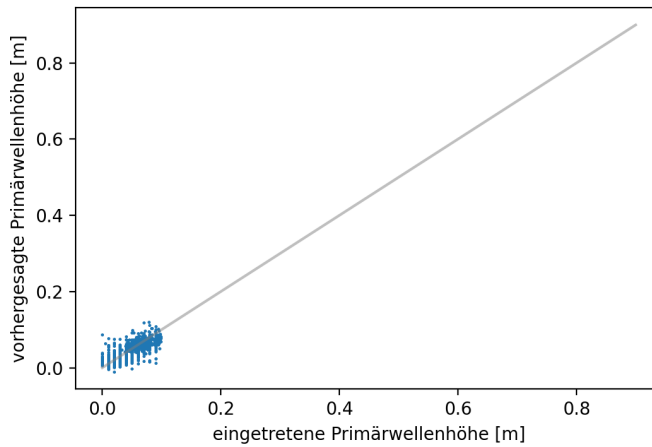
Differenz zwischen vorhergesagter und eingetretener Primärwellenhöhe	
Niedrigste Differenz	-7,67 cm
Größte Differenz	8,69 cm
Mittlere Differenz	0,02 mm
Mittelwert der absoluten Differenz	1,04 cm
Standardabweichung der Differenzen	1,42 cm

Tabelle 4.2.7.1: Vergleich der Differenzen zwischen vorhergesagter und eingetretener Primärwellenhöhe

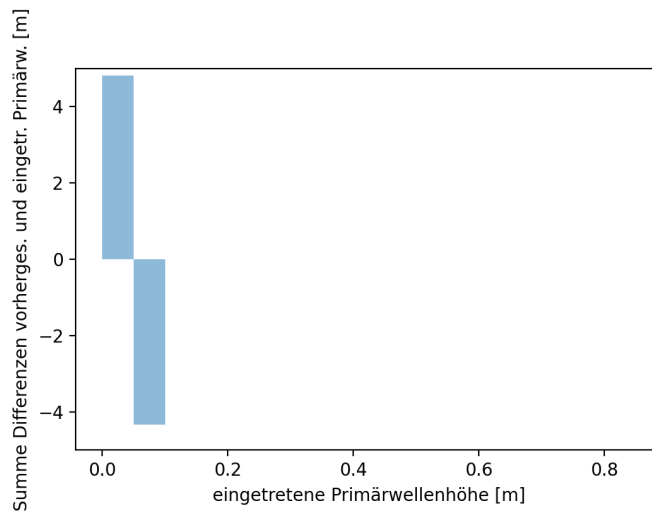
#### 4 Numerische Experimente und Ergebnisse



(a) Rangliste der eingetretenen Primärwellenhöhen verglichen mit Vorhersagen der Primärwellenhöhen



(b) Vorhersage Primärwellenhöhe über der eingetretenen Primärwellenhöhe



(c) Summe der Differenzen zwischen vorhergesagter und eingetretener Primärwellenhöhe in einzelnen Bereichen der eingetretenen Primärwellenhöhe

Abbildung 4.2.7.1: Ergebnisse, Einzelanalyse Datensatz mit Primärwellen unter 10 cm

### 4.2.8 Datensatz reduziert durch Undersampling

Die Datensätze aus 2014 und 2021 werden verwendet. Die Dateneinträge aus der Kampagne 2021 mit Primärwellenhöhen unter 5 cm werden weggelassen. Die Dateneinträge für den Trainingsdatensatz werden in den Bereichen, in denen die Primärwellenhöhen häufiger repräsentiert sind, reduziert, um so eine gleichmäßige Verteilung zu erhalten. Im Bereich von einer Primärwellenhöhe von 0 cm bis 5 cm werden zufällige Dateneinträge entfernt, bis 50 Dateneinträge in diesem Bereich übrig bleiben. Danach wird der Bereich um 5 cm versetzt, bis der gesamte Datensatz durchlaufen wurde. Im Bereich sehr hoher Primärwellen (über 40 cm) sind weniger Dateneinträge, so dass hier die gleichmäßige Repräsentation nicht gegeben ist. In Abbildung 4.2.8.1 sieht man eine gleichmäßig lineare Verteilung der Primärwellenhöhe bis 40 cm und ab dann sieht man die mit steigender Primärwellenhöhe immer seltener vertretenen Passagen.

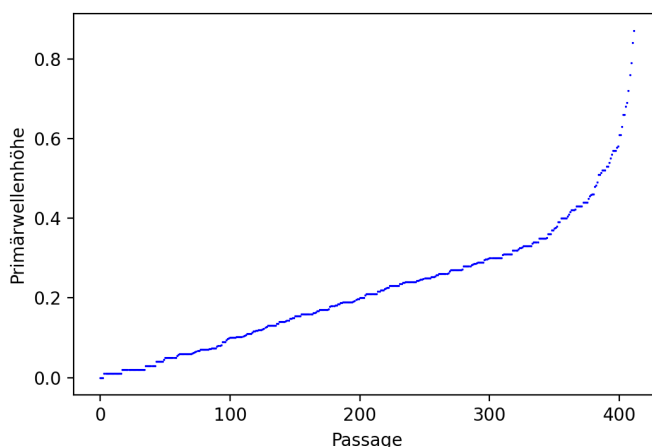


Abbildung 4.2.8.1: Eingetretene Primärwellenhöhen des Trainingsdatensatz sortiert nach der Primärwellenhöhe

**Aufbau** Insgesamt werden 2523 Passagen untersucht. Die 412 Passagen aus dem Trainingsdatensatz werden in der Leave-One-Out-Kreuzvalidierungs-Methode untersucht. Für die anderen Daten wurden alle Trainingsdaten beim Training verwendet. Bis auf die Anzahl und Wahl der Passagen entspricht der Aufbau dem in 4.2.1.

**Diskussion** In Abbildung 4.2.8.2c sieht man, dass die Primärwellenhöhen nun bis 25 cm überschätzt werden und ab 25 cm unterschätzt.

Der Mittelwert der absoluten Differenz und die Standardabweichung der Vorhersagegenauigkeiten sind in dieser Variante etwas schlechter als in 4.2.2. Die Reduzierung des Trainingsdatensatz hat sich auf die Vorhersagegenauigkeit insgesamt schädlich ausgewirkt.

### Ergebnisse

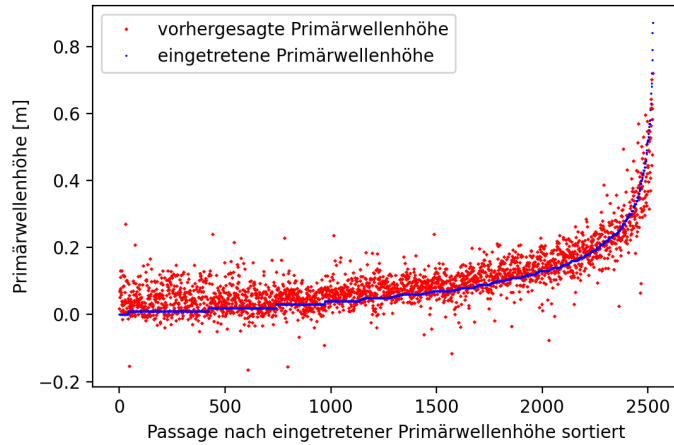
#### 4 Numerische Experimente und Ergebnisse

Differenz zwischen vorhergesagter und eingetretener Primärwellenhöhe	
Niedrigste Differenz	-32,47 cm
Größte Differenz	27,02 cm
Mittlere Differenz	1,72 mm
Mittelwert der absoluten Differenz	3,56 cm
Standardabweichung der Differenzen	4,68 cm

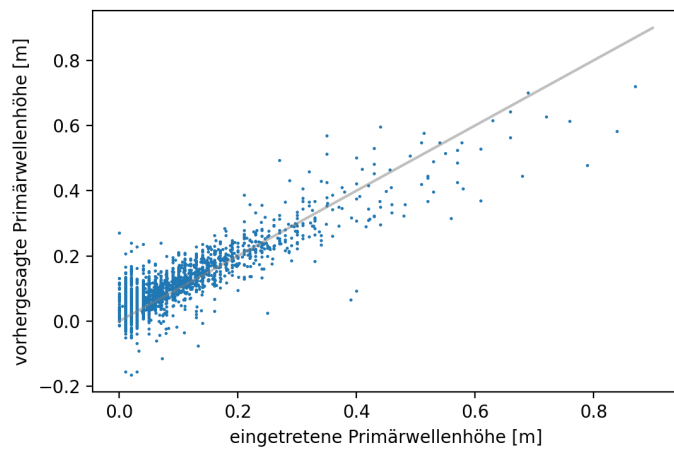
Tabelle 4.2.8.1: Vergleich der Differenzen zwischen vorhergesagter und eingetretener Primärwellenhöhe

Nummer	Differenz von vorhersage zu eingetretener Primärwellenhöhe	Primärwellenhöhe	Eintrittszeit	Name	Länge	Breite	Tiefgang	SOG	COG	Wasserstand Dwarsgat
-	[cm]	[cm]	-	-	[m]	[m]	[m]	[kn]	[°]	[m NHN]
1	27.02	0	21.10.2014 15:37	FINNWAVE	188	32	7.0	20.4	318	-0.78
2	22.38	27	09.12.2021 10:42	SONGA PUMA	225	31	10.3	18.5	312.1	-1.86
3	22.05	2	30.10.2014 20:13	MAERSK FLENSBURG	134	23	8.6	13.5	121	-0.80
4	22.00	35	14.12.2014 01:17	MSC KRYSAL	277	40	12.0	15.6	318	0.01
5	19.81	3	23.11.2014 05:22	ANNE SIBUM	151	22	6.8	13.3	122	-1.27
6	19.76	1	11.11.2014 17:44	MV SONDERBORG STRAIT	158	24	7.7	12.1	122	0.08
7	19.58	4	19.11.2014 01:13	CAPRICORNUS LEADER	199	32	7.9	14.4	123	-0.56
8	19.47	2	30.11.2014 21:20	HARMONY ACE	180	32	7.7	12.1	125	-0.12
9	17.71	21	11.12.2014 22:37	MAERSK LAGUNA	299	45	9.1	16.0	307	0.00
10	17.05	7	28.11.2014 23:01	RIO CADIZ	260	32	10.2	17.6	121	-2.20
11	16.30	35	03.11.2014 00:53	YORKTOWN EXPRESS	243	32	11.2	19.3	315	-0.65
12	15.70	1	21.10.2014 04:04	WEST STREAM	80	13	4.8	10.9	125	-1.09
13	15.62	44	22.10.2014 18:29	MAERSK PEG	210	32	9.2	20.4	319	-0.89
14	14.91	2	26.11.2014 15:48	CORAL LEADER	176	31	8.6	15.8	124	1.06
15	14.47	29	29.10.2021 22:08	HUNGARY	298	48	10.5	18.6	306.6	-0.45
16	14.24	2	10.11.2014 06:16	CCNI BILBAO	172	22	8.8	10.1	124	-0.61
17	13.96	1	27.10.2014 07:56	OLYMPIC ELECTRA	79	16	4.2	10.7	122	-1.42
18	13.88	32	29.10.2014 22:42	COSCO NAGOYA	260	32	11.3	17.1	316	-1.49
19	13.83	2	02.12.2014 12:13	MAERSK PALERMO	210	32	9.2	10.4	125	-0.45
20	13.82	1	15.11.2014 06:36	ABIS BERGEN	89	14	5.4	11.0	123	0.84
21	13.74	22	24.10.2014 05:56	MSC ROSSELLA	243	32	7.1	19.9	317	-1.42
22	13.68	1	02.12.2014 00:32	KUGELBAKE	79	19	2.9	9.4	122	-1.30
23	13.62	1	09.11.2014 18:04	BEAUTRADER	118	16	4.8	11.1	122	-0.81
24	13.56	2	05.11.2014 15:33	SONDERBORG STRAIT	158	24	8.0	12.9	120	-1.00
25	13.52	5	04.12.2014 01:41	MORNING CAMILLA	200	32	7.6	13.6	119	-0.68
26	13.49	2	23.10.2014 03:45	STEFAN SIBUM	152	24	6.8	13.3	121	-0.99
27	13.39	5	12.12.2014 19:59	SVENDBORG STRAIT	157	24	8.3	14.9	120	0.18
28	13.33	2	27.11.2014 19:17	ELISABETH	118	18	6.9	10.8	130	-0.84
29	13.30	3	29.11.2014 07:00	WYBELSUM	161	25	8.1	14.4	127	0.17
30	13.20	5	17.11.2014 12:41	ALEXANDER B	154	24	8.8	13.4	132	-1.04

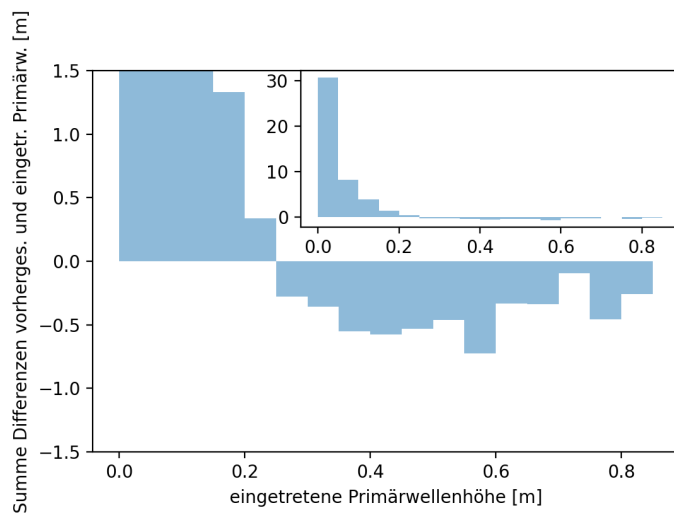
Tabelle 4.2.8.2: Größte Differenzen zwischen vorhergesagter und eingetretener Primärwelle



(a) Rangliste der eingetretenen Primärwellenhöhen verglichen mit Vorhersagen der Primärwellenhöhen



(b) Vorhersage Primärwellenhöhe über der eingetretenen Primärwellenhöhe



(c) Summe der Differenzen zwischen vorhergesagter und eingetretener Primärwellenhöhe in einzelnen Bereichen der eingetretenen Primärwellenhöhe

Abbildung 4.2.8.2: Ergebnisse, Datensatz reduziert durch Undersampling

Nummer	Differenz von vorhersage zu eingetretener Primärwellenhöhe	Primärwellenhöhe	Eintrittszeit	Name	Länge	Breite	Tiefgang	SOG	COG	Wasserstand Dwarsgat
-	[cm]	[cm]	-	-	[m]	[m]	[m]	[kn]	[°]	[m NHN]
1	-32.47	39	01.12.2014 21:15	MV SONDERBORG STRAIT	158	24	7.7	13.9	124	0.49
2	-31.25	79	15.11.2014 23:41	HS BACH	246	32	11.3	19.3	318	-1.16
3	-30.68	40	25.11.2014 06:00	CERES	140	23	7.3	13.4	125	-0.83
4	-25.68	84	10.11.2014 22:19	MARIBO MAERSK	399	59	11.5	15.2	315	-0.90
5	-24.53	56	09.12.2021 22:50	MAERSK PALERMO	210	32	9.6	18.9	306.3	-1.76
6	-24.03	61	17.11.2014 03:23	MSC SORAYA	277	40	10.1	16.3	302	-0.48
7	-23.54	68	11.11.2014 18:53	CHOAPA TRADER	294	32	11.0	20.7	319	-0.75
8	-22.43	25	22.10.2014 11:30	CHRISTOPHER	170	26	8.2	12.4	121	3.49
9	-20.91	13	20.10.2021 23:47	JANA	70	12	3.7	6.2	126.5	1.89
10	-18.73	7	01.12.2021 19:32	PILOTT.WANGEROOG	25	13	2.8	0.4	301.7	1.92
11	-18.57	3	25.11.2014 15:08	TIDE	30	8	2.3	0.0	0	1.44
12	-18.42	2	18.11.2014 09:30	TIDE	30	8	2.2	0.0	0	1.15
13	-17.34	58	03.11.2014 17:19	SAFMARINE HIGHVELD	270	43	11.9	15.6	315	-0.11
14	-16.47	1	27.11.2014 09:48	TIDE	30	8	2.0	0.0	0	-2.28
15	-15.60	48	01.12.2014 10:05	MSC MATILDE	294	32	9.8	20.1	318	-0.23
16	-15.15	52	20.11.2014 18:11	MSC ANGELA	265	32	7.7	18.3	318	-1.57
17	-14.93	87	26.10.2014 20:15	MSC KRYSTAL	277	40	11.0	18.9	319	-1.39
18	-14.64	33	06.11.2014 16:40	HOEGH NEW YORK	199	32	9.6	18.0	319	-1.32
19	-14.59	14	05.12.2014 18:53	LEV TAIFUN	41	8	2.0	7.8	123	-1.30
20	-14.53	76	03.12.2014 16:02	CMA CGM ANDROMEDA	363	46	11.8	16.6	319	-1.80
21	-14.47	57	18.11.2014 15:21	MSC DIEGO	259	32	9.2	18.9	317	-1.74
22	-14.40	11	30.10.2021 01:57	ZILTBORG	118	12	7.2	13.2	127.6	0.28
23	-14.32	44	14.11.2014 18:20	CSAV LARAQUETE	261	32	10.6	19.2	317	0.48
24	-13.40	53	30.10.2014 22:42	MSC ARUSHI R.	281	32	11.6	17.4	310	-1.90
25	-12.87	16	25.11.2014 19:17	FREDO	82	11	2.4	12.4	311	-1.42
26	-12.62	46	05.12.2014 18:45	CMA CGM ALEX. V. HUMBOLDT	396	54	12.4	17.1	120	-1.45
27	-12.62	42	06.12.2021 04:16	COSCO JAPAN	334	43	12.6	16.4	129.0	-0.49
28	-12.50	3	07.11.2021 10:39	CLAUDIA	69	10	3.7	13.3	124.7	1.00
29	-12.32	48	03.12.2021 04:08	MAERSK BULAN	223	32	11.7	18.9	306.7	-1.28
30	-12.21	6	24.10.2021 02:11	VB GEESTE	28	11	6.0	9.5	122.5	1.62

Tabelle 4.2.8.3: Kleinste Differenzen zwischen vorhergesagter und eingetretener Primärwelle



### 4.2.9 Nachträgliche Trendkorrektur

Dem Trend, dass die kleineren Primärwellenhöhen überschätzt und die größeren Primärwellen unterschätzt werden, soll nachträglich entgegen gewirkt werden. Die Daten aus 4.2.2 werden als Ausgangslage verwendet und weiter nachbearbeitet. Die Werte einer Prognose des ANNs werden im ersten Schritt nach unten hin auf 0 cm begrenzt, da keine negativen Wellenhöhen existieren. Anschließend wird die vorhergesagte Primärwellenhöhe relativ zu den Primärwellenhöhen aus dem Trainingsdatensatz mit der Formel

$$\tilde{P}_j = P_j - f \sum_{\substack{i=1 \\ i \neq j}}^n P_i - P_j \quad (4.1)$$

mit

$\tilde{P}_j$  vorhergesagte Primärwellenhöhe  
 $P_j$  Primärwellenhöhe vorhergesagt durch das ANN  
 $P_i$  Primärwellenhöhe des Trainingsdatensatzes  
 $f = 2,02447 \cdot 10^{-5}$  Korrekturfaktor

korrigiert. Der Wert des Faktors wurde dadurch bestimmt, dass die Summe der Beträge der Abweichungen von Vorhersage und eingetretener Primärwellenhöhe ein Minimum erreicht.

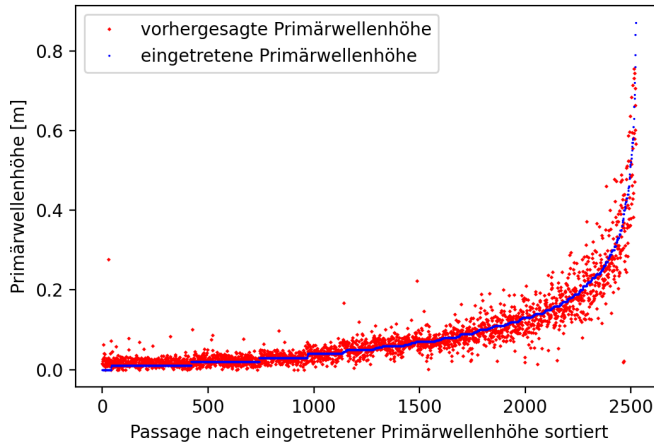
**Diskussion** In Tabelle 4.2.9.1 ist zu erkennen, dass sich die maximale und minimale Differenz zwischen vorhergesagter und eingetretener Primärwellenhöhe verschlechtert haben.

Die mittlere Differenz, der Mittelwert der absoluten Differenzen sowie die Standardabweichung haben sich leicht verbessert. Die Verbesserungen sind allerdings alle unter einem Millimeter. Die deutlichste Verbesserung ist erkennbar in Abbildung 4.2.9.1c. Die Verbesserung in den Summen der Differenzen in einzelnen Bereichen lässt sich in den Werten über den gesamten Bereich aus 4.2.9.1 nicht erkennen. Durch den Korrekturfaktor konnte der systematische Fehler durch ungleichmäßige Datenverteilung zum Teil kompensiert werden. Durch das Bestimmen des Korrekturfaktors wurde das Prognosemodell durch Informationen aus dem Test beeinflusst. Die Aussagekraft der Testergebnisse wurde damit leicht reduziert zu Gunsten besserer Vorhersageergebnisse. Dadurch, dass nur eine Gewichtung zur Kompensation einer systematischen Abweichungen hinzugefügt wurde, bleibt das resultierende Modell geeignet für den Einsatz.

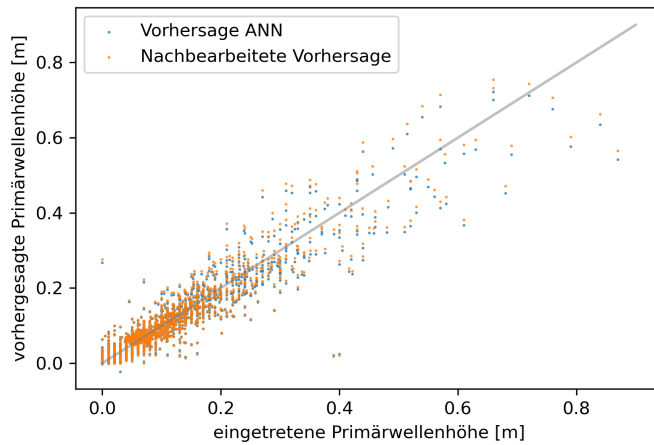
### Ergebnisse

Differenz zwischen vorhergesagter und eingetretener Primärwellenhöhe	Antwort ANN	Antwort nachbearbeitet
Niedrigste Differenz	-37,51 cm	-37,82 cm
Größte Differenz	26,75 cm	27,68 cm
Mittlere Differenz	0,13 mm	0,11 mm
Mittelwert der absoluten Differenz	1,96 cm	1,93 cm
Standardabweichung der Differenzen	3,36 cm	3,35 cm

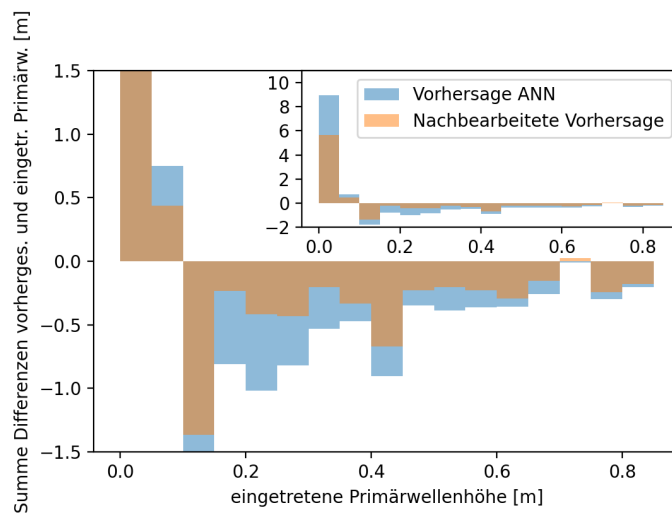
Tabelle 4.2.9.1: Vergleich der Differenzen zwischen vorhergesagter und eingetretener Primärwellenhöhe



(a) Rangliste der eingetretenen Primärwellenhöhen verglichen mit Vorhersagen der Primärwellenhöhen



(b) Vorhersage Primärwellenhöhe über der eingetretenen Primärwellenhöhe



(c) Summe der Differenzen zwischen vorhergesagter und eingetretener Primärwellenhöhe in einzelnen Bereichen der eingetretenen Primärwellenhöhe

Abbildung 4.2.9.1: Ergebnisse, Trendkorrektur

Nummer	Differenz von vorhersage zu eingetretener Primärwellenhöhe	Primärwellenhöhe	Eintrittszeit	Name	Länge	Breite	Tiefgang	SOG	COG	Wasserstand Dwarsgat
-	[cm]	[cm]	-	-	[m]	[m]	[m]	[kn]	[°]	[m NHN]
1	27.68	0	21.10.2014 15:37	FINNWAVE	188	32	7.0	20.4	318	-0.78
2	19.07	27	09.12.2021 10:42	SONGA PUMA	225	31	10.3	18.5	312.1	-1.86
3	16.81	31	07.12.2014 16:25	SEAGO BREMERHAVEN	294	32	11.9	19.8	319	0.01
4	16.13	21	11.12.2014 22:37	MAERSK LAGUNA	299	45	9.1	16.0	307	0.00
5	15.34	7	28.11.2014 23:01	RIO CADIZ	260	32	10.2	17.6	121	-2.20
6	15.22	32	29.10.2014 22:42	COSCO NAGOYA	260	32	11.3	17.1	316	-1.49
7	14.74	44	22.10.2014 18:29	MAERSK PEG	210	32	9.2	20.4	319	-0.89
8	14.39	57	02.11.2014 01:28	BUXCOAST	300	40	12.3	17.1	315	-1.22
9	14.37	54	07.11.2014 07:01	JPO CANOPUS	264	32	10.8	18.0	320	-2.20
10	13.83	35	14.12.2014 01:17	MSC KRYSTAL	277	40	12.0	15.6	318	0.01
11	12.74	24	15.10.2021 08:43	COSCO VIETNAM	334	43	12.8	19.0	305.7	1.56
12	12.28	51	14.12.2021 01:21	MSC ELODIE	300	48	12.5	17.9	308.5	-0.86
13	12.20	4	07.11.2021 15:02	VISURGIS	19	6	1.9	27.9	303.8	2.79
14	12.20	35	03.11.2014 00:53	YORKTOWN EXPRESS	243	32	11.2	19.3	315	-0.65
15	11.94	24	01.11.2014 13:49	SEAGO BREMERHAVEN	293	32	12.8	18.9	123	-1.12
16	11.65	36	07.12.2014 18:57	NYK DIANA	295	32	11.4	17.6	315	-1.18
17	11.42	18	17.11.2014 16:30	MOL PRESENCE	293	40	10.8	14.6	316	-0.46
18	11.28	31	24.11.2014 23:28	CMA CGM MARCO POLO	396	54	11.4	14.8	318	0.95
19	11.27	29	29.10.2021 22:08	HUNGARY	298	48	10.5	18.6	306.6	-0.45
20	10.78	18	23.11.2014 19:57	CMA CGM MARCO POLO	396	54	11.8	14.2	122	-1.64
21	10.69	49	25.10.2014 07:42	NYK NEBULA	294	32	10.9	17.8	319	-1.72
22	10.32	21	21.10.2014 13:48	MSC VAISHNAVI R.	281	32	11.6	19.4	319	0.16
23	9.45	66	02.11.2014 13:33	CHASTINE MAERSK	346	43	12.5	16.9	318	-1.41
24	9.11	11	09.12.2014 20:49	RUTH	134	23	8.2	18.3	319	-2.05
25	9.02	23	10.12.2014 20:35	YM VANCOUVER	260	32	10.6	18.6	311	-0.15
26	8.99	16	15.10.2021 01:11	TORONTO	200	32	9.0	17.8	311.0	-0.59
27	8.99	10	02.12.2014 21:22	MAERSK NIMES	210	30	10.6	17.4	308	1.45
28	8.75	15	26.11.2014 00:49	MSC METHONI	304	40	11.0	15.3	318	0.99
29	8.72	10	03.11.2021 00:23	BRUSSELS	278	40	13.5	17.2	309.6	1.07
30	8.63	25	21.11.2014 00:10	A. P. MOLLER	346	43	13.1	16.2	315	1.29

Tabelle 4.2.9.2: Größte Differenzen zwischen vorhergesagter und eingetretener Primärwelle

#### 4 Numerische Experimente und Ergebnisse

Nummer	Differenz von vorhersage zu eingetretener Primärwellenhöhe	Primärwellenhöhe	Eintrittszeit	Name	Länge	Breite	Tiefgang	SOG	COG	Wasserstand Dwarsgat
-	[cm]	[cm]	-	-	[m]	[m]	[m]	[kn]	[°]	[m NHN]
1	-37.82	40	25.11.2014 06:00	CERES	140	23	7.3	13.4	125	-0.83
2	-37.18	39	01.12.2014 21:15	MV SONDERBORG STRAIT	158	24	7.7	13.9	124	0.49
3	-30.43	87	26.10.2014 20:15	MSC KRYSTAL	277	40	11.0	18.9	319	-1.39
4	-22.79	61	17.11.2014 03:23	MSC SORAYA	277	40	10.1	16.3	302	-0.48
5	-20.84	68	11.11.2014 18:53	CHOAPA TRADER	294	32	11.0	20.7	319	-0.75
6	-18.80	79	15.11.2014 23:41	HS BACH	246	32	11.3	19.3	318	-1.16
7	-17.68	84	10.11.2014 22:19	MARIBO MAERSK	399	59	11.5	15.2	315	-0.90
8	-17.66	42	15.12.2021 03:34	VERA D	179	27	8.7	20.6	311.2	-1.36
9	-16.44	42	06.12.2021 04:16	COSCO JAPAN	334	43	12.6	16.4	129.0	-0.49
10	-15.61	44	14.11.2014 18:20	CSAV LARAQUETE	261	32	10.6	19.2	317	0.48
11	-14.76	29	18.11.2014 06:09	CMA CGM MUSCA	347	45	12.2	18.2	126	0.38
12	-14.71	51	28.10.2014 17:02	QUANTUM OF THE SEAS	347	41	8.6	21.0	316	0.87
13	-14.06	57	18.11.2014 15:21	MSC DIEGO	259	32	9.2	18.9	317	-1.74
14	-14.02	52	04.11.2014 15:43	RIO THOMPSON	210	30	9.3	20.6	319	-1.37
15	-13.78	58	03.11.2014 17:19	SAFMARINE HIGHVELD	270	43	11.9	15.6	315	-0.11
16	-13.65	16	25.11.2014 19:17	FREDO	82	11	2.4	12.4	311	-1.42
17	-13.39	42	19.11.2014 07:25	CMA CGM MUSCA	347	45	11.7	15.3	316	0.51
18	-13.28	25	22.10.2014 11:30	CHRISTOPHER	170	26	8.2	12.4	121	3.49
19	-12.75	40	09.12.2014 21:07	MAERSK KAMPALA	300	43	12.4	16.1	123	-2.00
20	-12.69	33	06.11.2014 16:40	HOEGH NEW YORK	199	32	9.6	18.0	319	-1.32
21	-12.67	14	05.12.2014 18:53	LEV TAIFUN	41	8	2.0	7.8	123	-1.30
22	-12.15	44	14.11.2014 11:46	CMA CGM LAPEROUSE	365	51	12.2	15.2	121	-1.90
23	-12.09	20	05.11.2021 03:36	LOTUS A	336	43	13.0	12.9	129.2	-0.11
24	-11.88	48	01.12.2014 10:05	MSC MATILDE	294	32	9.8	20.1	318	-0.23
25	-11.32	53	30.10.2014 22:42	MSC ARUSHI R.	281	32	11.6	17.4	310	-1.90
26	-11.17	13	20.11.2014 14:54	CFL PRUDENCE	118	12	3.7	12.9	316	-0.83
27	-11.06	69	21.10.2014 05:59	MSC METHONI	304	40	9.9	17.9	317	-0.98
28	-11.02	24	18.11.2014 01:27	LUCKY SUNDAY	229	32	8.5	15.3	316	-1.02
29	-10.63	30	11.11.2014 21:39	RIO CARDIFF	262	32	9.3	17.2	126	-2.15
30	-10.62	16	06.11.2014 07:03	AKERDIJK	168	26	1.3	16.6	315	-1.32

Tabelle 4.2.9.3: Kleinste Differenzen zwischen vorhergesagter und eingetretener Primärwelle

### 4.2.10 Weitere Tests zur Optimierung der Vorhersage der Primärwelle

**Vergroößerung der Neuronenanzahl** Die Anzahl der Neuronen in einem Layer wurde in dieser Variante auf 128 pro Layer erhöht. Die Layeranzahl wurde erhöht, so dass statt zwei sequenzieller Dense Layer wie in 4.2.2 nun 6 sequenzielle Dense Layer verwendet wurden. Die Testergebnisse waren jedoch sehr ähnlich. Die unnötigen Neuronen tragen zum Ergebnis weder dazu bei, dass die Informationen besser verarbeitet werden können, noch verschlechtern sie das Ergebnis durch Overfitting. Es ist davon auszugehen, dass die Knoten die zur Datenverarbeitung nicht notwendig sind, durch das Setzen auf 0 von trainierbaren Parametern deaktiviert werden.

**Einbeziehung weiterer Daten** In dieser Untersuchung wurde getestet, ob durch das Einbeziehen weiterer Variablen in dem Datensatz die Vorhersagegenauigkeit gesteigert werden kann. Zu den Daten aus 4.2.2 wurden die Koordinaten mit Rechtswert und Hochwert hinzugefügt, sowie der Schiffstyp als kategorische Information. Die Testergebnisse waren nahezu identisch zu 4.2.2, so dass davon auszugehen ist, dass diese Informationen im ANN keine Gewichtung erhalten.

**Leave-One-Out-Kreuzvalidierung mit weiter verwendeten trainierbaren Parametern** In dieser Untersuchung wurde nach dem Testen eines Dateneintrages in der Leave-One-Out-Kreuzvalidierung die trainierbaren Parameter abgespeichert und vor dem Training des nächsten Dateneintrages wieder geladen. Das ANN wurde dann trainiert, bis sich im Loss nach 20 Epochen keine weitere Verbesserung einstellte. Das erste ANN benötigte 140 Epochen, die darauf folgenden ANNs benötigten nur noch ca. 60 Epochen, um das Kriterium zur Beendigung zu erfüllen. Die Trainingszeit konnte damit erheblich reduziert werden. Die Testergebnisse waren wesentlich besser als alle anderen. Dies liegt daran, dass in den trainierbaren Parametern Optimierungen für die Testdaten hinterlegt werden und das ANN im Trainingsprozess nur optimiert wird, wo sich Differenzen im Loss zum aktuellen Trainingsdatensatz einstellen. Ein ANN im hinteren Bereich der Leave-One-Out-Kreuzvalidierung wurde entsprechend bereits oft auf den Testdateneintrag trainiert. Die Testergebnisse sind damit verfälscht.

**Trainingsdatensatz dynamisch beschränken** In einer ersten Prognose wird die Primärwelle bestimmt wie in 4.2.2. Die Trainingsdaten werden für eine zweite Prognose ausgewählt, indem 1300 Schiffspassagen mit einer nächst größeren und 1300 Schiffspassagen mit einer nächst kleineren eingetretenen Primärwellenhöhe verwendet wurden. Ausgenommen ist der vorherzusagende Dateneintrag. Dadurch wurden keine Daten für das Training verwendet, die weit von der Prognose entfernt liegen. Die Standardabweichung hat sich schon bei dieser geringen Reduktion des Datensatzes verschlechtert. Dies lässt darauf schließen, dass die Prognosequalität durch die geringe Datenmenge beschränkt ist, so dass jede weitere Reduktion des Datensatzes sich negativ auf Prognosen auswirkt.

### 4.2.11 Vorhersage des Absunks

Der Absunk soll vorhergesagt werden mit dem ANN Aufbau aus 4.2.2 und anschließend nachbearbeitet werden wie in 4.2.9, da hier die besten Ergebnisse erzielt werden konnten. Die Datensätze aus 2014 und 2021 werden verwendet. Die Dateneinträge aus der Kampagne 2021 mit Primärwellenhöhen unter 5 cm werden weggelassen.

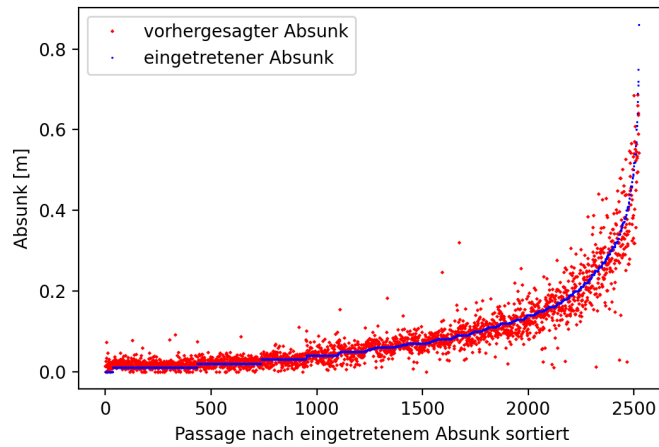
**Aufbau** Es werden 2523 Passagen in der Leave-One-Out-Kreuzvalidierungs-Methode untersucht. Die Anzahl und Wahl der Passagen entspricht dem Aufbau 4.2.2. Das Trainingsziel ist lediglich im Gegensatz zu 4.2.2 der Absunk. Die Vorhersagen des ANNs werden im Anschluss korrigiert wie in 4.2.9. Der Korrekturfaktor beträgt hier  $f = 2,13692 \cdot 10^{-5}$ .

**Diskussion** Die Standardabweichung der Differenzen zwischen vorhergesagtem und eingetretenem Wert ist bei der Vorhersage des Absunks 0,5 mm größer als bei 4.2.9. Die mittlere Differenz und der Mittelwert der absoluten Differenzen liefern ebenfalls leicht schlechtere Ergebnisse. In der Summe der Differenzen zwischen vorhergesagtem und eingetretenem Wert in einzelnen Bereichen zeigt sich generell ebenfalls die Verteilung, dass kleinere Werte überschätzt werden und größere unterschätzt werden im Mittelwert. Ab 10 cm ist der Wechsel zwischen überschätzten und unterschätzten Werten im Mittel. Eine Ausnahme ist der Bereich von 20 cm bis 25 cm bei dem die Prognosen etwas zu niedrig ausfallen. Diese Ausnahme ist auf statistische Schwankungen zurückzuführen und eine ähnliche Ausnahme ist in 4.2.9 in einem anderen Bereich sichtbar. Die Vorhersageergebnisse des Absunks sind ähnlich genau wie die Vorhersageergebnisse der Primärwellenhöhe in 4.2.9. Die Ähnlichkeit ist plausibel, da der Absunk eine Kenngröße der Primärwelle ist.

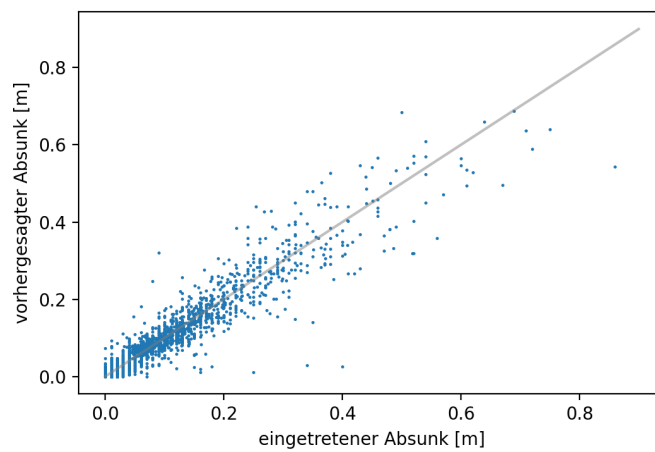
### Ergebnisse

Differenz zwischen vorhergesagtem und eingetretenem Absunk	Antwort ANN	Antwort nachbearbeitet
Niedrigste Differenz	-37,20 cm	-37,53 cm
Größte Differenz	21,81 cm	22,99 cm
Mittlere Differenz	0,06 mm	0,13 mm
Mittelwert der absoluten Differenz	2,01 cm	2,00 cm
Standardabweichung der Differenzen	3,40 cm	3,40 cm

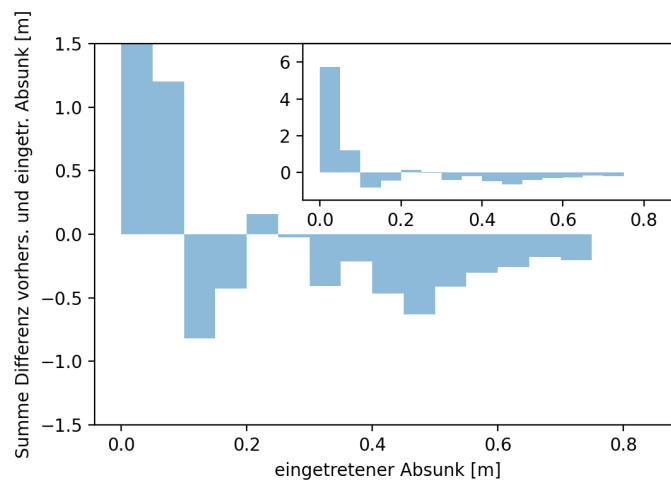
Tabelle 4.2.11.1: Vergleich der Differenzen zwischen vorhergesagtem und eingetretenem Absunk



(a) Rangliste des eingetretenem Absunk verglichen mit den Vorhersagen



(b) Vorhersage Absunk über der eingetretenem Absunk



(c) Summe der Differenzen zwischen vorhergesagtem und eingetretenem Absunk in einzelnen Bereichen des eingetretenem Absunk

Abbildung 4.2.11.1: Ergebnisse, ANN mit reduziertem Datensatz zur Vorhersage des Absunk mit Trendkorrektur

### 4.2.12 Vorhersage der Sekundärwelle

Die Sekundärwelle soll vorhergesagt werden mit dem ANN Aufbau aus 4.2.2. Die Datensätze aus 2014 und 2021 werden verwendet. Die Dateneinträge aus der Kampagne 2021 mit Primärwellenhöhen unter 5 cm werden weggelassen.

**Aufbau** Es werden 2523 Passagen in der Leave-One-Out-Kreuzvalidierungs-Methode untersucht. Die Anzahl und Wahl der Passagen entspricht dem Aufbau 4.2.2. Das Trainingsziel ist lediglich im Gegensatz zu 4.2.2 die Sekundärwelle.

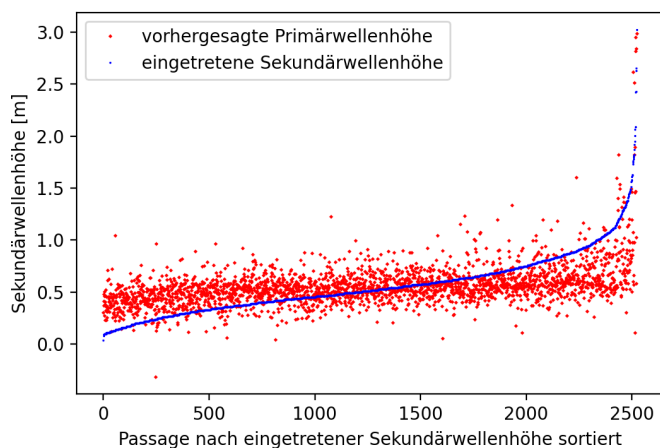
**Diskussion** Die Vorhersage der Sekundärwelle mit einem ANN und den gewählten Eingangsdaten erzielt wesentlich schlechtere Ergebnisse als die Vorhersagen der Primärwelle. In Abbildung 4.2.12.1a und Abbildung 4.2.12.1b ist zu sehen, dass bei eingetretenen sehr niedrigen und sehr hohen Sekundärwellen die Prognosen tendenziell in die richtige Richtung gehen. Im Bereich von eingetretenen Sekundärwellen von 25 cm bis 75 cm streuen die Prognoseergebnisse relativ gleichmäßig. Die Standardabweichung der Differenzen zwischen vorhergesagter und eingetretener Sekundärwelle ist mit 24,74 cm hoch. Die Sekundärwelle ist aus den gewählten Eingangsdaten schwer abzuleiten, wie sich auch in 4.3 zeigen wird.

### Ergebnisse

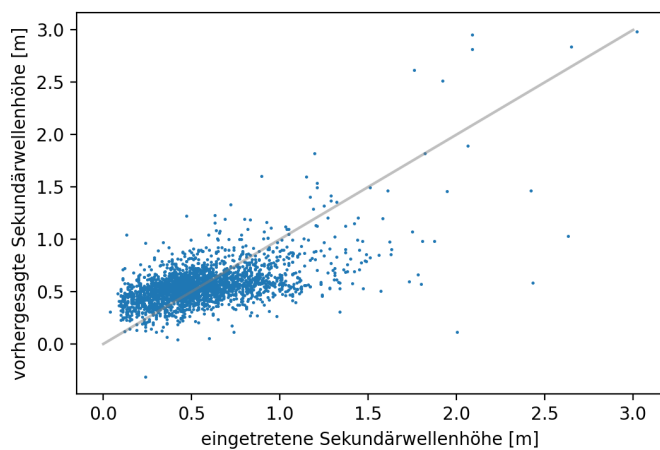
Differenz zwischen vorhergesagter und eingetretener Sekundärwellenhöhe	
Niedrigste Differenz	-1,89 m
Größte Differenz	91,49 cm
Mittlere Differenz	1,99 mm
Mittelwert der absoluten Differenz	18,48 cm
Standardabweichung der Differenzen	24,74 cm

Tabelle 4.2.12.1: Vergleich der Differenzen vorhergesagter und eingetretener Sekundärwellenhöhe

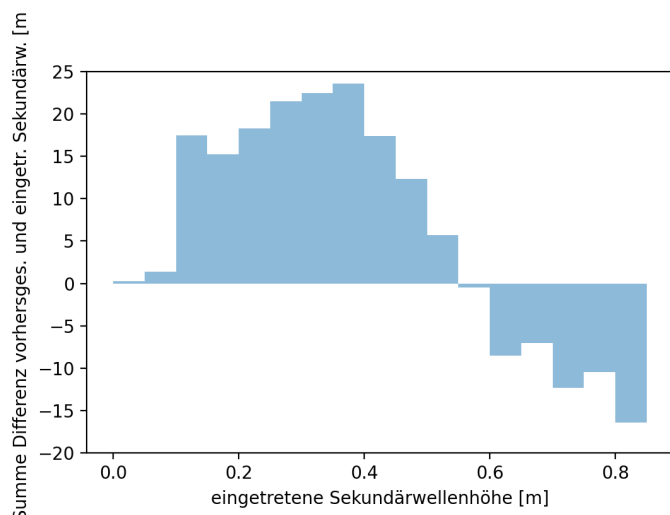




(a) Rangliste der eingetretenen Sekundärwellenhöhen verglichen mit Vorhersagen der Sekundärwellenhöhen



(b) Vorhersage Sekundärwellenhöhe über der eingetretenen Sekundärwellenhöhe



(c) Summe der Differenzen zwischen vorhergesagter und eingetretener Sekundärwellenhöhe in einzelnen Bereichen der eingetretenen Sekundärwellenhöhe

Abbildung 4.2.12.1: Ergebnisse, ANN mit reduziertem Datensatz

### 4.3 Korrelationsmatrix

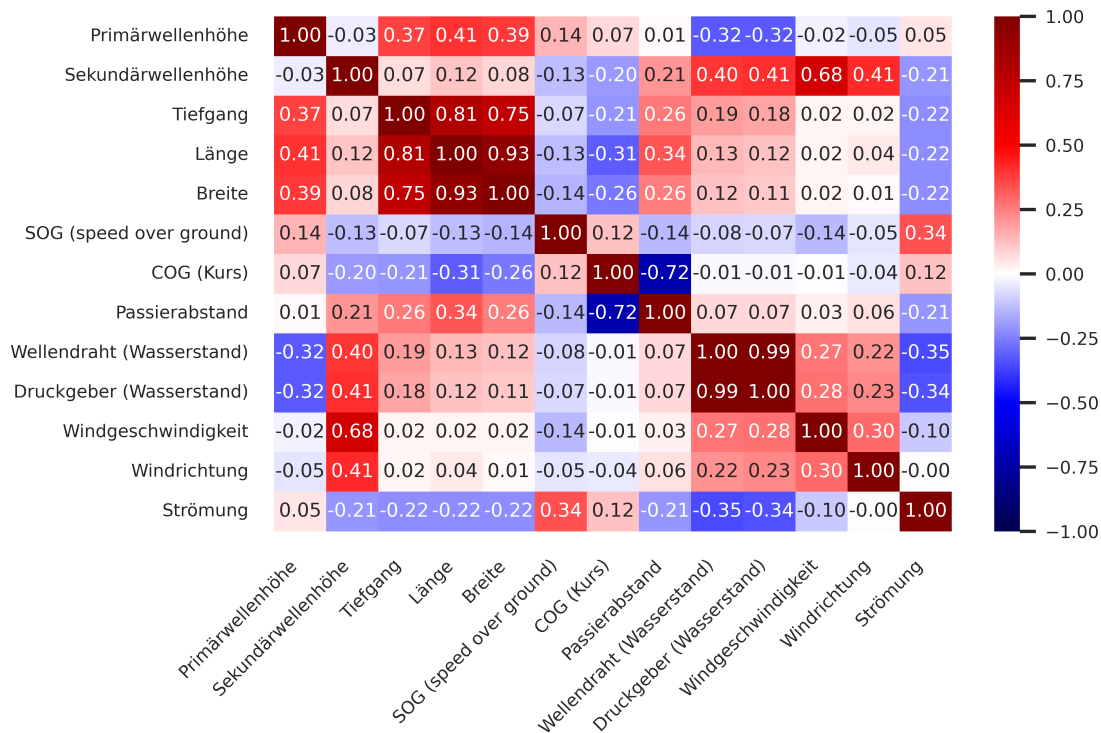


Abbildung 4.3.0.1: Korrelationsmatrix Station A 2021

In der Korrelationsmatrix der Station A 2021 Abbildung 4.3.0.1 lässt sich (wie in 3.1 bereits vermutet) erkennen, dass die Schiffslänge und die Schiffsbreite einen hohen Korrelationskoeffizienten haben, da lange Schiffe auch meist breit sind.

Der Wasserstand gemessen mit Wellendraht und der Wasserstand gemessen mit Druckgeber haben einen Korrelationskoeffizienten nahe 1, da sie die identische Größe nur mit einer anderen Methode messen.

#### 4.3.1 Einflussparameter Primärwellenhöhe

Die Primärwellenhöhe hat in Abbildung 4.3.0.1 hohe (0,37 bis 0,41) Korrelationskoeffizienten mit der Schiffslänge, Breite und dem Tiefgang. In Abbildung 4.3.0.2 und Abbildung 4.3.0.3 sind die Korrelationskoeffizienten etwas niedriger (0,33 bis 0,39) aber immer noch sehr hoch. Dies lässt den Schluss zu, dass mit der Schiffsgröße die Primärwelle steigt. Dies ist auch plausibel wenn man bedenkt, dass die Primärwelle durch das Auffüllen des verdrängten Wasservolumens des Schiffes entsteht.

Der Korrelationskoeffizient der Primärwellenhöhe und des Wasserstandes (Druckgeber sowie Wellendraht) ist bei den drei Stationen im Bereich  $-0,18$  bis  $-0,32$ . Mit steigendem Wasserstand nimmt die Primärwellenhöhe ab. Dies hängt mit der Wasser verdrängenden Querschnittsfläche und dem wasserführenden Kanalquerschnitt zusammen.

Der Korrelationskoeffizient der Primärwellenhöhe und der Geschwindigkeit über Grund (SOG) ist in Abbildung 4.3.0.1  $0,14$ , in Abbildung 4.3.0.2  $0,05$  und in Abbildung 4.3.0.3  $-0,13$ . Die Korrelation ist unerwartet niedrig. Ein erklärbarer Faktor ist, dass die Geschwindigkeit relativ zum Grund (SOG) gemessen ist und die Primärwellenhöhe eher abhängig von der Geschwindigkeit relativ zum Wasser ist. Macht man zwei Prognosen des ANNs aus 4.2.2 mit zwei unterschiedlichen Geschwindigkeiten und ansonsten identischen Parametern, so ist die Prognose

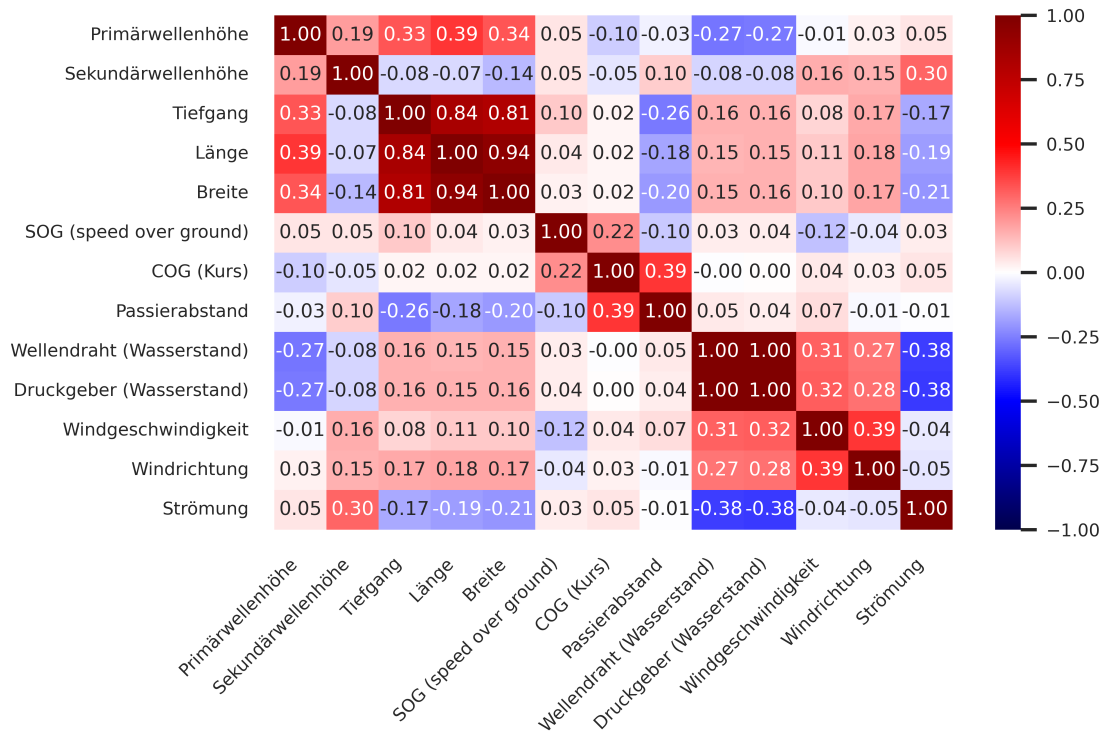


Abbildung 4.3.0.2: Korrelationsmatrix Station B 2021

der Primärwelle mit der höheren Geschwindigkeit auch höher. Ein weiterer Faktor ist, dass der Korrelationskoeffizient zwischen SOG und Schiffslänge bei Station C stark negativ (-0.43) korrelieren. Das heißt die langen Schiffe die zu großen Primärwellen neigen fahren bereits langsamer. Bei Station A ist der Korrelationskoeffizient zwischen SOG und der Schiffslänge -0,13 und bei Station B 0,04. Die Korrelation der Geschwindigkeit mit der linearen Korrelation des Korrelationskoeffizienten lässt sich nicht ordentlich abbilden und die Geschwindigkeit gewinnt erst an Bedeutung für Prognosen der Primärwellenhöhe bei Einbeziehung weiterer Parameter.

### 4.3.2 Einflussparameter Sekundärwellenhöhe

Die Sekundärwellenhöhe hat an Station A und B den höchsten (0,16 bis 0,68) Korrelationskoeffizienten mit der Windgeschwindigkeit. Diese Korrelation ist eine ungewollte, da die Wellen erzeugt durch den Wind in einem ähnlichen Frequenzbereich liegen wie die Sekundärwellen. Der Auswertalgorithmus hat die Wellen durch Wind und Sekundärwellen der Schiffe nicht komplett trennen können. Möchte man die Sekundärwellen genauer analysieren, ist eine längere Messkampagne förderlich, bei der die Zeiten mit hohen Windgeschwindigkeiten heraus gefiltert werden. Die landnahe Station C weicht hier ab und hat eine negative Korrelation mit -0,15.

Die Windrichtung hat ebenfalls sehr hohe Korrelationskoeffizienten mit der Sekundärwellenhöhe an allen drei Stationen. An Station A und B ist dieser Korrelationskoeffizient größer als der von Sekundärwellenhöhe und Windgeschwindigkeit. An Station C ist der Korrelationskoeffizient der Sekundärwellenhöhe und Windgeschwindigkeit kleiner als der von Sekundärwellenhöhe und Windrichtung. Die Windrichtung nahe des Festlandes hat Einfluss darauf, ob Wellen stärker oder weniger erzeugt werden. Die hohe Korrelation von Windrichtung und Sekundärwellenhöhe hängt auch damit zusammen, dass die Windgeschwindigkeit eine sehr hohe Korrelation mit der Windrichtung an allen drei Stationen hat.

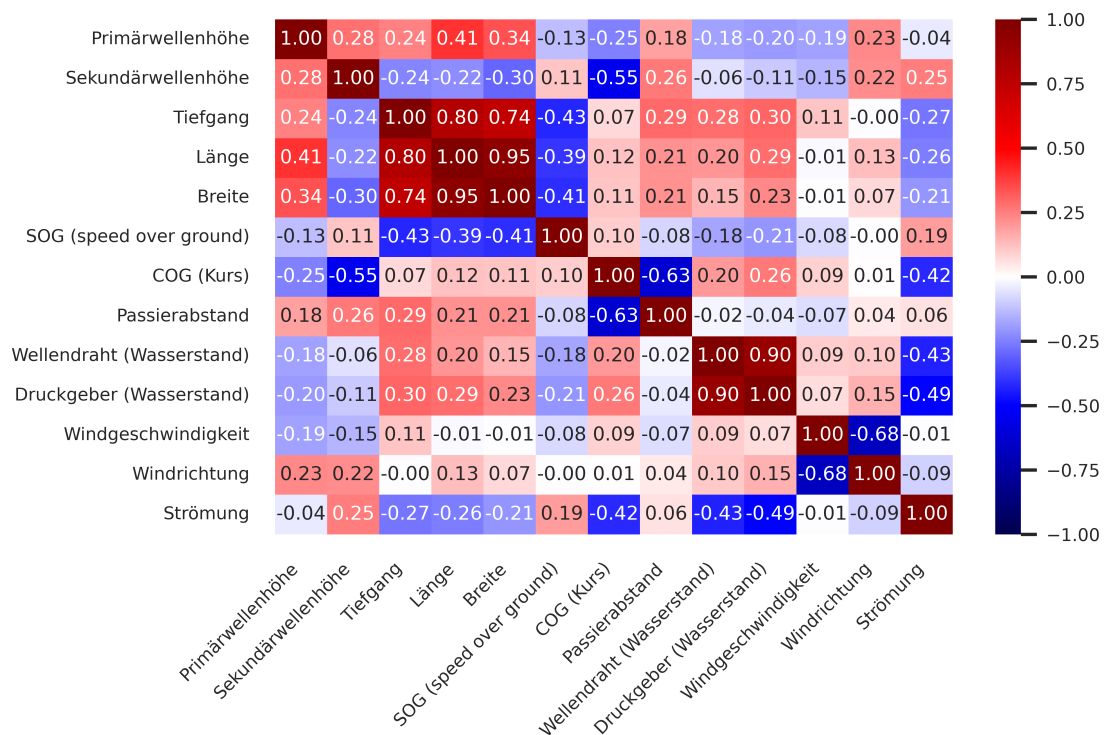
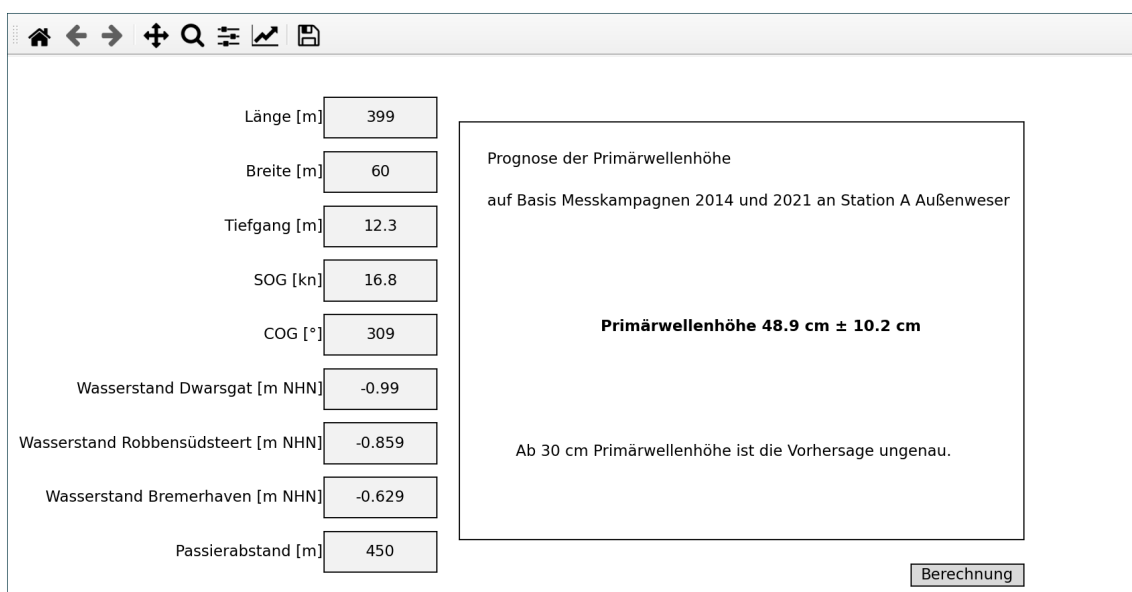


Abbildung 4.3.0.3: Korrelationsmatrix Station C 2021

## 5 Software zur Prognose von Primärwellen

Zur Berechnung von aktuell fahrenden Schiffen wurde ein Programm geschrieben, mit welchem man die Primärwellen vorhersagen kann. Das hinter liegende ANN entspricht dem aus 4.2.2, wobei alle Daten für das Training verwendet wurden. Die Prognose des ANN wird anschließend wie in 4.2.9 nachträglich um den bekannten Trend korrigiert. Mit dem Programm ist es beispielsweise möglich, die Geschwindigkeit eines Schiffes so zu verändern, dass eine bestimmte Primärwellenhöhe erreicht wird. In den Textfeldern auf der linken Seite werden die Parameter eines Schiffes eingegeben. Mit dem Button "Berechnung" wird eine Prognose ausgeführt. Das Programm benötigt zur Ausführung einen Pythoncompiler mit den Modulen Matplotlib, TensorFlow, Keras, NumPy und pandas.



The screenshot shows a web-based interface for wave prediction. On the left side, there are several input fields with their respective values:

Parameter	Value
Länge [m]	399
Breite [m]	60
Tiefgang [m]	12.3
SOG [kn]	16.8
COG [°]	309
Wasserstand Dwarsgat [m NHN]	-0.99
Wasserstand Robbensüdsteert [m NHN]	-0.859
Wasserstand Bremerhaven [m NHN]	-0.629
Passierabstand [m]	450

The main display area on the right contains the following text:

Prognose der Primärwellenhöhe  
auf Basis Messkampagnen 2014 und 2021 an Station A Außenweser

**Primärwellenhöhe 48.9 cm ± 10.2 cm**

Ab 30 cm Primärwellenhöhe ist die Vorhersage ungenau.

At the bottom right of the interface is a button labeled "Berechnung".

Abbildung 5.0.0.1: Grafische Oberfläche der Prognosesoftware



## 6 Zusammenfassung und Ausblick

Für die Beobachtung der Wellenbelastung von Primärwellen auf Kanäle nach einer Messkampagne können gute Prognosen mit neuronalen Netzen erstellt werden. Eine einfache Klassifizierung zwischen hohen und kleinen Primärwellen kann mit einer Genauigkeit von über 91% erfolgen, ab einem Schwellwert von über 15 cm sogar über 95%. Eine Prognose der Primärwellenhöhe kann mit einer Standardabweichung von 3,35 cm über den gesamten Bereich betrachtet erfolgen. Diese ist bei der Betrachtung kleiner Wellen geringer und bei Betrachtung großer Wellen größer. Der Absink wurde mit einer ähnlichen Genauigkeit vorhergesagt wie die Primärwellenhöhe. Die Sekundärwellen können wesentlich schwerer aus dem Datensatz vorhergesagt werden. Optimierungspotential für die Prognosen der Sekundärwellen ist zwar noch vorhanden, an die Qualität der Prognosen der Primärwellen ist mit diesen Methoden schwer heran zu kommen. Die Sekundärwellen hängen von der genauen umströmten Schiffsgeometrie ab, die dem Datensatz nicht zugrunde liegt. Diese könnte ein neuronales Netz erst über eine Identifikation von individuellen Schiffen machen. Eine stärkere Einbeziehung der Winddaten könnte ebenfalls helfen. Die Daten über den Wind wurden hier nicht einbezogen, da diese Lücken aufweisen. Ein Herausfiltern der lückenhaften Daten würde die Anzahl der auswertbaren Passagen reduzieren.

Die hier verwendeten Methoden sind für den Einsatz zur Wellenberechnung neu. In [20] wurde ein Schiff aufwändig mit 3D Modellierung und CFD-Analyse simuliert. Die Schiffswellen konnten genauer bestimmt werden. In [18] wurden zwei Schiffe mit insgesamt 20 Dateneinträgen betrachtet und mittels einer 2D Simulation berechnet. Die Primärwellenhöhen wurden berechnet, wobei diese dort in zwei Größen aufgeteilt dokumentiert wurde. Die Streuung dieser getrennten Variablen ähnelt optisch der Streuung, die in 4.2.9.1b zu sehen ist. In [19] wurden unterschiedliche Formeln verwendet um Schiffswellenkennwerte zu ermitteln und mit Messwerten zu vergleichen. Untersucht wurden 39 einzelne Schiffspassagen. Die verwendeten Parameter entsprechen auch den hier verwendeten. Die hohen Schiffswellen konnten in dieser Untersuchung besser berechnet werden als für die kleineren. Eine Ergänzung der Methoden aus dem hier verwendeten Ansatz mittels ANNs und der Methode aus [19] ist denkbar. In [19] wurde aber auch auf die starke lokale Anpassung der Formeln eingegangen.

Es zeigte sich in den einzelnen Varianten der neuronalen Netze, dass schon eine leichte Reduktion der Datenmenge die Qualität der Prognosen senkt. Dies spricht dafür, dass die Prognosen bei längeren Messkampagnen oder dem Hinzufügen von Daten aus folgenden Messkampagnen weiter gesteigert werden können. Gerade Daten von Schiffen mit hohen Primärwellen sind im vorhandenen Datensatz selten. Eine Verbesserung wäre es, wenn die Primärwellenhöhen ausgewogen im Datensatz vertreten wären. Das Hinzufügen weiterer Daten aus zukünftigen Messkampagnen könnte den Datensatz insgesamt soweit erhöhen, dass über eine Auswahl ein ausgewogener Datensatz erstellt werden könnte. Unter der Voraussetzung, dass mehr Daten zur Verfügung stünden, wäre eine Aufteilung in Kategorien möglich, für die getrennte ANNs trainiert werden. Beispielsweise könnten die Primärwellenhöhen nach einer vorhergehenden Berechnung noch einmal genauer berechnet werden durch ein ANN, das speziell für diesen Primärwellenhöhenbereich optimiert werden müsste. Leerfahrten, wie die der FINNWAVE am 21.10.2014 könnten auch getrennt behandelt werden. Dies würde die Prognosen der Leerfahrten und die der Anderen verbessern. Ein bereits trainiertes Modell kann auf andere Orte durch ein neues Training übertragen werden. An der neuen Lokation sind dann weniger Daten notwendig für Vorhersagen mit einer ähnlich guten Qualität. Bei den vorliegenden Daten macht eine Übertragung eines ANNs von Lokation A auf Lokation B keinen Sinn, da hier

dieselben Schiffe passieren und die Ergebnisse dem Overfitting entsprechen würden. Generell ist dieses Vorgehen aber attraktiv ein aufwendiges Modell zu entwickeln und anschließen auf andere Orte zu übertragen. Anhand der Korrelationsmatrix konnten die relevanten Parameter bestimmt werden, um die Primärwellen und Sekundärwellen vorherzusagen. Es wurde hier deutlich, dass in den Sekundärwellen ein deutlicher Anteil an natürlichen Wellen durch Wind enthalten ist. Für bessere Analysen der Sekundärwellen ist es auch vorteilhafter, Daten über längere Messkampagnen zu sammeln, um dann die Sekundärwellen nur in Zeiten mit geringen Windgeschwindigkeiten auszuwerten.



# Formelzeichen, Abkürzungen

## Formelzeichen

TPR		true positive rate (Richtig-positiv-Rate)
FPR		false positive rate (Falsch-negativ-Rate)
TP		true positive (richtig positiv)
FP		false positive (falsch positiv)
$v$		Übertragungsfunktion
$\varphi$		Aktivierungsfunktion
$w$		Gewichtung
$b$		Bias
$\mu$		Erwartungswert
$x$		diskreter Wert
$y$		diskreter Wert
$X$		Zufallsvariable
$Y$		Zufallsvariable
$x_i$		Messwert einer Zeitreihe $\{x\}_t$ zum Zeitpunkt $t = i$
$y_i$		Messwert einer zweiten Zeitreihe $\{y\}_t$ zum Zeitpunkt $t = i$
$X$		Zufallsvariable von der die Zeitreihe $\{x\}_t$ als Stichprobe gesampled wurde
$Y$		Zufallsvariable von der die Zeitreihe $\{y\}_t$ als Stichprobe gesampled wurde
Var		Varianz
Cov		Kovarianz
Korr		Korrelationskoeffizient
$\tilde{P}_j$	m	vorhergesagte Primärwellenhöhe
$P_j$	m	Primärwellenhöhe vorhergesagt durch das ANN
$P_i$	m	Primärwellenhöhe des Trainingsdatensatz
$f$		Korrekturfaktor

## Abkürzungen

AIS	Automatisches Identifikationssystem
ANN	Künstliches neuronale Netz (artificial neuronal network)
AUC	Area Under the ROC Curve
CNN	Convolutional Neuronal Network
LOO-CV	Leave-One-Out-Kreuzvalidierung
ROC	Receiver Operating Characteristic
WSV	Wasserstraßen- und Schifffahrtsverwaltung des Bundes



# Literatur

- [1] 2U. *Undersampling*. [Online; Stand 15. September 2022]. 2022. URL: <https://www.mastersindatascience.org/learning/statistics-data-science/undersampling>.
- [2] Kumar A. *Keras Neural Network for Regression Problem*. [Online; Stand 21. Juli 2022]. 2020. URL: <https://vitalflux.com/keras-neural-network-for-regression-problem/>.
- [3] BAW. *Definition der Kenngrößen*. [Online; Stand 12. Juli 2022]. 2022. URL: [https://wiki.baw.de/de/index.php/Definition\\_der\\_Kenngr%C3%B6%C3%9Fen](https://wiki.baw.de/de/index.php/Definition_der_Kenngr%C3%B6%C3%9Fen).
- [4] Rumelhart D.E. et al. „Learning representations by back-propagating errors“. In: *Nature* 323 (1986), S. 533–536. DOI: <https://doi.org/10.1038/323533a0>.
- [5] deeplizard. *Activation functions in a neural network*. [Online; Stand 7. Juli 2022]. 2017. URL: <https://deeplizard.com/learn/video/mOpILLfpXWE>.
- [6] deeplizard. *Artificial Neural Networks explained*. [Online; Stand 23. Juni 2022]. 2017. URL: [https://deeplizard.com/learn/video/hfK\\_dvC-avg](https://deeplizard.com/learn/video/hfK_dvC-avg).
- [7] deeplizard. *Bias in an Artificial Neural Network explained*. [Online; Stand 14. September 2022]. 2017. URL: <https://deeplizard.com/learn/video/HetFihsXSys>.
- [8] deeplizard. *Deep learning with convolutional neural networks*. [Online; Stand 16. Juli 2022]. 2017. URL: [https://deeplizard.com/learn/video/YRhxdVk\\_sIs](https://deeplizard.com/learn/video/YRhxdVk_sIs).
- [9] deeplizard. *Flatten, Reshape, and Squeeze Explained*. [Online; Stand 20. Juli 2022]. 2018. URL: <https://deeplizard.com/learn/video/fCVuiW9AFzY>.
- [10] deeplizard. *Loss in a Neural Network explained*. [Online; Stand 16. Juli 2022]. 2017. URL: <https://deeplizard.com/learn/video/Skc8nqJirJg>.
- [11] deeplizard. *Max Pooling in Convolutional Neural Networks explained*. [Online; Stand 16. Juli 2022]. 2018. URL: [https://deeplizard.com/learn/video/YRhxdVk\\_sIs](https://deeplizard.com/learn/video/YRhxdVk_sIs).
- [12] deeplizard. *Overfitting in a neural network*. [Online; Stand 20. Juli 2022]. 2017. URL: <https://deeplizard.com/learn/video/DEMmkFC6IGM>.
- [13] Google Machine Learning Education. *Classification: ROC Curve and AUC*. [Online; Stand 14. September 2022]. 2022. URL: <https://developers.google.com/machine-learning/crash-course/classification/roc-and-auc>.
- [14] Chollet F. *Structured data classification from scratch*. [Online; Stand 14. September 2022]. 2020. URL: [https://keras.io/examples/structured\\_data/structured\\_data\\_classification\\_from\\_scratch](https://keras.io/examples/structured_data/structured_data_classification_from_scratch).
- [15] Rosenblatt F. „The perceptron: A probabilistic model for information storage and organization in the brain“. In: *Psychological Review* 65 (1958), S. 386–408. DOI: <https://doi.org/10.1037/h0042519>.
- [16] Hopfield J.J. und Tank D.W. „Neural computation of decisions in optimization problems“. In: *Biol. Cybern.* 52 (1985), S. 141–152. DOI: <http://doi.org/10.1007/BF00339943>.
- [17] Peters K. und Bauer T. „Fahrrinnenanpassung Weser Schiffswellenmessung in der Außenweser 2014 Zusammenfassung der Ergebnisse“. In: (Aug. 2016).

- [18] Dam K.T. et al. „Investigation of ship waves in a narrow channel“. In: *Journal of Marine Science and Technology* 13 (2008), S. 223–230. DOI: <https://doi.org/10.1007/s00773-008-0005-6>.
- [19] Mao L., Chen Y. und Li X. „Characterizing ship-induced hydrodynamics in a heavy shipping traffic waterway via intensified field measurements“. In: *Water Science and Engineering* 13.4 (2020), S. 329–338. ISSN: 1674-2370. DOI: <https://doi.org/10.1016/j.wse.2020.11.001>. URL: <https://www.sciencedirect.com/science/article/pii/S1674237020300934>.
- [20] Ji Sheng S. et al. „NUMERICAL PREDICTION OF SHIP-GENERATED WAVES IN RESTRICTED CHANNEL“. In: *Marine 2011 Computational Methods in Marine Engineering IV* (2011).
- [21] Mitchell T.M. *Machine Learning*. McGraw-Hill, 1997. ISBN: 0070428077.
- [22] McCulloch W.S. und Pitts W. „A Logical Calculus of the Ideas Immanent in Nervous Activity“. In: *The bulletin of mathematical biophysics volume 5* (1943), S. 115–113. DOI: <https://doi.org/10.1007/BF02478259>.
- [23] Wikipedia. *Erwartungswert* — *Wikipedia, die freie Enzyklopädie*. [Online; Stand 8. Juli 2022]. 2022. URL: <https://de.wikipedia.org/w/index.php?title=Erwartungswert&oldid=219804439>.
- [24] Wikipedia. *Korrelationskoeffizient* — *Wikipedia, die freie Enzyklopädie*. [Online; Stand 7. Juli 2022]. 2022. URL: <https://de.wikipedia.org/w/index.php?title=Korrelationskoeffizient&oldid=220483300>.
- [25] Wikipedia. *Korrelationsmatrix* — *Wikipedia, die freie Enzyklopädie*. [Online; Stand 7. Juli 2022]. 2022. URL: <https://de.wikipedia.org/w/index.php?title=Korrelationsmatrix&oldid=220493980>.
- [26] Wikipedia. *Kovarianz (Stochastik)* — *Wikipedia, die freie Enzyklopädie*. [Online; Stand 22. August 2022]. 2022. URL: [https://de.wikipedia.org/w/index.php?title=Kovarianz\\_\(Stochastik\)&oldid=224390849](https://de.wikipedia.org/w/index.php?title=Kovarianz_(Stochastik)&oldid=224390849).
- [27] Wikipedia. *Kreuzvalidierungsverfahren* — *Wikipedia, die freie Enzyklopädie*. [Online; Stand 20. Juli 2022]. 2022. URL: <https://de.wikipedia.org/w/index.php?title=Kreuzvalidierungsverfahren&oldid=223744044>.
- [28] Wikipedia. *Künstliches Neuron* — *Wikipedia, die freie Enzyklopädie*. [Online; Stand 20. August 2022]. 2022. URL: [https://de.wikipedia.org/w/index.php?title=K%C3%BCnstliches\\_Neuron&oldid=221281731](https://de.wikipedia.org/w/index.php?title=K%C3%BCnstliches_Neuron&oldid=221281731).
- [29] Wikipedia. *Überwachtes Lernen* — *Wikipedia, die freie Enzyklopädie*. [Online; Stand 26. Juli 2022]. 2022. URL: [https://de.wikipedia.org/w/index.php?title=%C3%9Cberwachtes\\_Lernen&oldid=223982748](https://de.wikipedia.org/w/index.php?title=%C3%9Cberwachtes_Lernen&oldid=223982748).
- [30] Wikipedia. *Varianz* — *Wikipedia, die freie Enzyklopädie*. [Online; Stand 22. August 2022]. 2022. URL: <https://de.wikipedia.org/w/index.php?title=Varianz&oldid=225427319>.

# Erklärung zur selbstständigen Bearbeitung der Arbeit

Hiermit versichere ich,

*Deniz Himstedt*

dass ich die vorliegende Masterarbeit - bzw. bei einer Gruppenarbeit die entsprechend gekennzeichneten Teile der Arbeit - mit dem Thema:

*Datenbasierte Analyse von Schiffswellen in der Außenweser mittels Machine Learning*

ohne fremde Hilfe selbständig verfasst und nur die angegebenen Quellen und Hilfsmittel benutzt habe. Wörtlich oder dem Sinn nach aus anderen Werken entnommene Stellen sind unter Angabe der Quellen kenntlich gemacht.

- die folgende Aussage ist bei Gruppenarbeiten auszufüllen und entfällt bei Einzelarbeiten -

Die Kennzeichnung der von mir erstellten und verantworteten Teile der Masterarbeit ist erfolgt durch:

---

Ort

Datum

Unterschrift