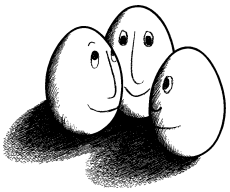
Studienarbeit

Zusammenhangsanalyse von Materialeigenschaften und Röntgenbeugung (XRD) für ternäre Systeme

David Arnu



Studienarbeit
am Fachbereich Informatik
der TU Dortmund

Dortmund, 22. Oktober 2013

Betreuer:

Prof. Dr. Katharina Morik
Dipl.-Inform. Hendrik Blom

Inhaltsverzeichnis

Abbildungsverzeichnis

Tabellenverzeichnis

1. Einleitung und Problembeschreibung

Die Anzahl der möglichen Mischungsverhältnisse von verschiedenen Stoffen ist eine der größten Herausforderungen in der Materialforschung. Neben theoretischen Vorüberlegungen und Berechnungen, sind physikalische Untersuchungen an tatsächlichen Materialproben ein wichtiges Hilfsmittel zum Entdecken neuer, nützlicher Materialien. Das Problem hierbei ist die Vielzahl möglicher Kombinationen, welche eine vollständige Erfassung und Verarbeitung aller Daten erschwert.

Charakteristische Eigenschaften, wie z.B. Materialhärte oder Formgedächtniseigenschaften, treten oft nur in bestimmten Bereichen des Mischungsverhältnisses auf, wobei die Übergänge meist relativ exakt entlang von Kennlinien auftreten. Eines der bekanntesten Beispiele dürfte das Eisen-Kohlenstoff-Diagramm [1] sein, dass bereits die Komplexität der Phasenübergänge bei nur zwei Materialkomponenten verdeutlicht.

Einen sehr guten Überblick über das Arbeitsgebiet der kombinatorischen Materialforschung und die Anwendungsmöglichkeiten von Data-Mining bietet der Artikel von Rajan mit dem Titel *"Combinatorial Materials Science*: *Experimental Strategies for Accelerated Knowledge Discovery*" [2].

In der vorliegenden Arbeit soll untersucht werden, ob bestimmte Methoden der Datenanalyse und Techniken des maschinellen Lernens geeignet sind, die Materialforschung bei der Auswertung und Analyse von Messdaten zu unterstützen. Dabei liegt das Augenmerk auf der Kategorisierung von Materialmischungen mit ähnlichen Eigenschaften und der Zusammenhangsanalyse zwischen verschiedenen physikalischen Messgrößen.

Hierfür werden im folgenden Kapitel zunächst die physikalischen Grundlagen genauer erläutert und die Merkmale der untersuchten Datensätze beschrieben. Anschließend werden die verwendeten Analyse- und Modellierungsmethoden, insbesondere Dynamic-Time-Warping (DTW) und die Regressions-SVM, erläutert (Kapitel 3). Im ersten analytischen Kapitel 4 werden Regressionsmodelle und Entscheidungsbäume angewendet um den elektrischen Widerstand zu bestimmen. In Abschnitt 5 werden die durch Röntgenbeugung gewonnenen Messreihen näher betrachtet, insbesondere wird eine neue Methode angewendet um Messpunkte anhand der DTW-Distanz ihrer XRD-Spektren zu gruppieren. Darauf folgend werden in Kapitel 6 lineare Regression und Clustering eingesetzt, um den Zusammenhang zwischen XRD-Spektren und Materialeigenschaften zu untersuchen. Die Ergebnisse werden im abschließenden Kapitel 7 zusammengefasst und diskutiert.

Ziel dieser Arbeit ist es zu Untersuchen, ob grundlegende Zusammenhänge zwischen verschiedenen Materialeigenschaften erkennbar sind. Dabei wird auf den Einfluss von Expertenwissen über die physikalischen Grundlagen verzichtet. Auch eine effiziente Anordnung der Untersuchungsmethoden steht zunächst nicht im Vordergrund. Stattdessen wird vorerst versucht aus deutlich aufwendiger zu gewinnenden, aber detaillierteren, Informationen (XRD-Spektren) auf einfacher zu ermittelnde Größen (Widerstand) zu schließen. Dies ist motiviert durch zwei Argumente: die Vermutung, dass bei einem ge-

funden Zusammenhang auch die Rückrichtung zu folgern ist und, dass sich das Verfahren möglicherweise auch auf aufwendigere, oder die Materialprobe zerstörende, Messverfahren übertragen lässt.

2. Daten

Die Strukturierung von Mischungsverhältnissen in der Materialforschung bedarf einiger erklärender Worte, da hier, im Gegensatz zu anderen mehrdimensionalen Daten einige Besonderheiten berücksichtigt werden müssen. Außerdem soll in diesem Kapitel die Struktur der untersuchten Datensätze und die verwendeten Messmethoden erläutert werden.

2.1. Ternäre Systeme

Unter der vereinfachten Annahme, man könne das Mischungsverhältnis von Materialien nur in ganzen 1%-Schritten messen, ergeben sich für zwei Stoffe folglich 100 verschiedene Kombinationen. Für drei Stoffe sind es aufgrund des exponentiellen Wachstums bereits 5050 ($\binom{101}{3}$) mögliche Kombinationen. Zusätzliche Einflussfaktoren bei der Verarbeitung, wie Druck oder Temperatur, können den Suchraum noch zusätzlich vergrößern. Die große Anzahl möglicher Materialien motiviert die Suche nach Möglichkeiten der effektiven Darstellungsmethoden und raschen systematischen Verarbeitung neuer Proben, wie z.B. in der Arbeit von Long et al. [3] beschrieben.

Eine sehr nützliche Darstellung für das Mischungsverhältnis von drei Materialien sind ternäre Dreiecke, die eine zweidimensionale Projektion der Mischung dreier Stoffe sind. An jeder Seite des Dreiecks wird die Konzentration einer Komponente eingetragen. Jeder Punkt innerhalb des Dreiecks gibt genau ein Mischungsverhältnis an, wobei die Nebenbedingung, dass an jedem Punkt die Summe aller Anteile 100% ergibt, eingehalten wird. Die Abbildung 2.1 zeigt[1], wie das Mischungsverhältnis abgelesen werden kann. Für die auf der unteren Seite eingetragene Komponente lässt sich die Konzentration anhand der Linien parallel zur linken Seite des Dreiecks ablesen. Für die auf der rechten und linken Seite eingetragenen Werte ist das Verfahren analog.

In dieser Arbeit werden ternäre Dreiecke jedoch nicht weiter verwendet, denn die untersuchten Daten haben bereits eine Anordnung in X/Y-Koordinaten. Dies ist begründet in der Art wie Materialproben erzeugt werden, welches im folgenden Abschnitt erklärt wird.

2.2. Praktische Erzeugung von ternären Systemen

Für die praktische Untersuchung werden Materialsammlungen auf Waferscheiben erzeugt. Dabei werden zunächst die drei Elemente, mittels Ionisation, auf eine Trägerscheibe aus Silizium aufgedampft. Einzelne Bereiche der Scheibe werden dabei unterschiedlich lange

[1] Quelle: https://en.wikipedia.org/wiki/File:Ternary_plot_1.png

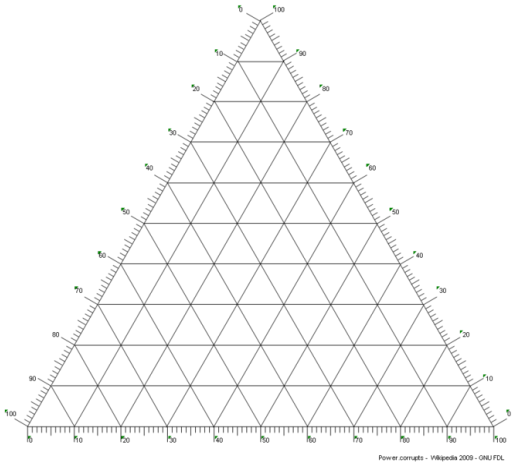


Abbildung 2.1.: Beispiel eines ternären Diagramms mit Koordinatenlinien

exponiert, so dass für jedes einzelne Material ein Gradient mit unterschiedlicher Schichtdicke entsteht. Da diese Gradienten jeweils verschiedene Richtungen haben, bildet jeder Punkt auf der Waferscheibe ein bestimmtes Mischungsverhältnis der aufgetragenen Elemente ab. Mit Hilfe dieses Verfahrens lassen sich in recht kurzer Zeit Materialbibliotheken erstellen und neue Materialkombinationen auf ihre Eigenschaften untersuchen. Da es sich um ein stetiges System handelt, ist für eine systematische Erfassung und Messung eine gewisse Diskretisierung notwendig. Hierfür wird ein Gitter von Messpunkten über den Wafer gelegt, wobei die Feinheit des Gitters und somit die Anzahl der Messungen variieren kann. In den vorliegenden Proben wurde ein Raster mit 342 Messpunkten verwendet, jedoch sind auch feinere Auflösungen möglich, wie ein Datensatz mit 4489 Punkten zeigt. In der Praxis werden meist keine vollständigen, ternären Systeme, von 0 bis 100 Prozent, erstellt; die Gradienten können ein beliebiges Intervall abdecken, somit kann der zu untersuchende Raum auf interessante Gebiete eingeschränkt werden.

2.3. Untersuchte Materialproben

Für die Untersuchung stehen Messdaten von drei verschiedenen ternären Systemen zur Verfügung: Kupfer-Nickel-Zink (CuNiZn), Nickel-Chrom-Rhenium (NiCrRe) und Titan-Cobalt-Wolfram (TiCoW). Diese Daten wurden ausgewählt, weil sie möglichst vollständig sind. Weitere Datensätze weisen oftmals deutlich mehr fehlende Werte auf. Die Gründe hierfür sind zum Beispiel Fehler in der Materialprobe, oder Probleme bei der bei der automatischen Messung. Für den TiCoW-Datensatz liegen nur die Ergebnisse der Röntgenbeugungsanalyse vor, weswegen er auch nur in der Analyse des räumlichen Zusammenhangs in Kapitel 5 verwendet wird.

2.4. Untersuchte Merkmale

Für die drei Materialproben liegen die Ergebnisse für mehrere Messungen vor. Die Messverfahren, sowie die Ergebnisse werden nun kurz erörtert und Abbildung 2.2 gibt einen Überblick über die Verteilung der bekannten Messwerte.

Chemische Zusammensetzung Die chemische Zusammensetzung der Systeme ist die eindeutigste Beschreibung für einen Datenpunkt, da sie die Zusammensetzung einer Materialprobe exakt definiert (abgesehen von hier nicht betrachteten äußeren Umständen wie Druck oder Temperatur bei der Verarbeitung). Andererseits ist die Kenntnis der Zusammensetzung alleine nicht ausreichend um andere Materialeigenschaften abzuleiten, was gerade die Schwierigkeit und Motivation der kombinatorischen Materialforschung ausmacht.

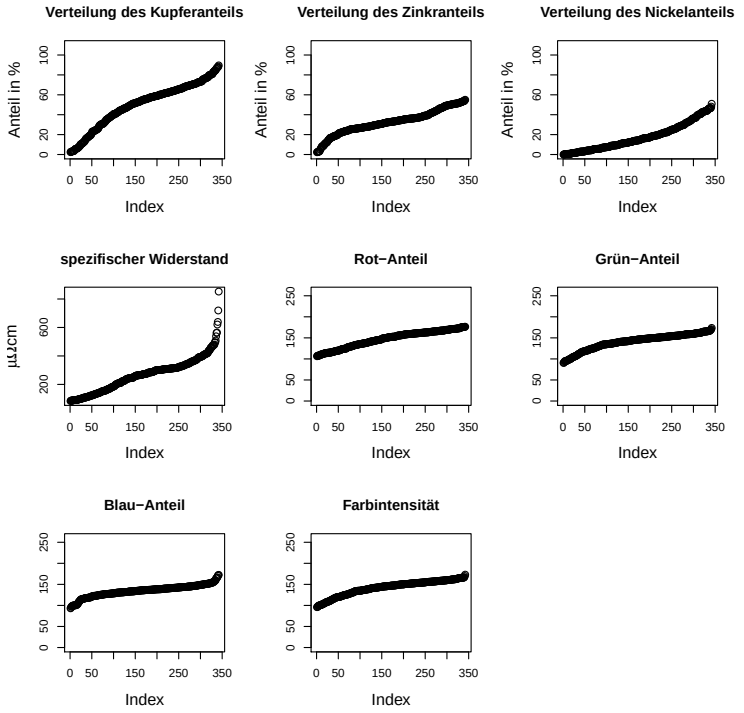
Farbwerte des reflektierten Lichts Eine weitere Prüfgrößen ist die Farbzusammensetzung des reflektierten Lichts von ein Messpunkt. Hierfür werden die Farbeanteile für Rot, Grün und Blau mit Werten (in der zwischen 0 und 255 (8 Bits) gemessen. Neben den einzelnen Werten gibt es auch eine aggregierte Größe, die die drei Werte anteilig zusammenfasst. Die Verteilung der Farbspektren ist in Abbildung 2.2 zu sehen.

Elektrischer Widerstand Der elektrischen Widerstand. Dieser liegt in zwei Größen vor, einmal als direkt gemessener Widerstand und als spezifischer Widerstand, der als Materialkonstante unabhängig von der Schichtdicke an der gemessen Stelle ist. Für den CuNiZn-Datensatz wird für die folgenden Untersuchungen der spezifische Widerstand verwendet, da er die genauere Größe darstellt. Für den NiCrRe-Datensatz liegt dieser Wert jedoch nicht vor, weswegen hier der elektrische Widerstand verwendet wird. Diese beiden Werte sind jedoch eng miteinander verbunden und im Folgenden wird zumeist nur von (elektrischen) Widerstand gesprochen, unabhängig davon welche Größe genau verwendet wird.

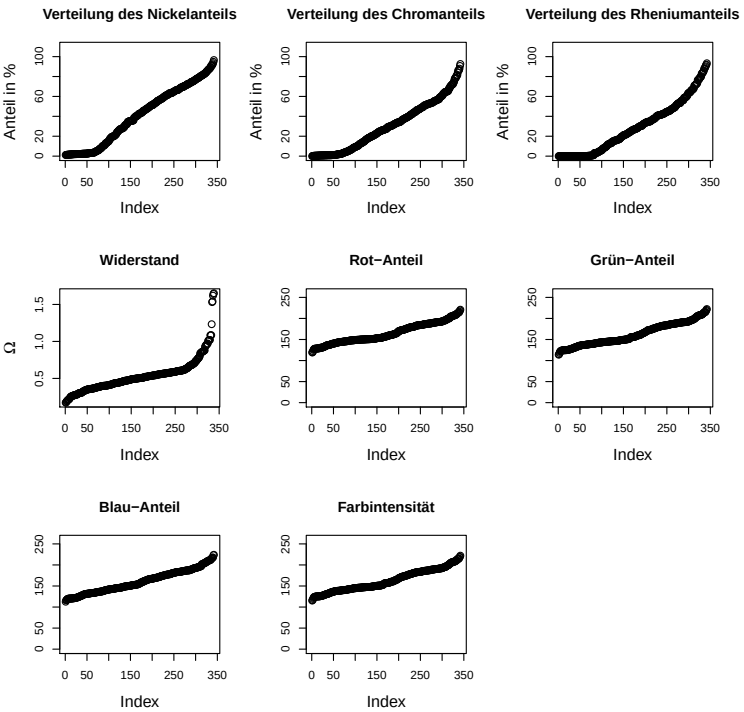
Datenaufbereitung der Widerstandsmessungen Bei der Untersuchung der CuNiZn-Daten fielen zwei Widerstandsmessungen am Rand des Wafers, mit Widerstandswerten über $700\mu\Omega$, als deutliche Ausreißer auf, welche als Messfehler identifiziert werden konnten. Außer bei der Breitensuche in Abschnitt 5.4, werden diese beiden Punkte bei allen Untersuchungen im Vorfeld entfernt, um die Ergebnisse nicht zu verzerren. Bei dem NiCrRe-Datensatz fehlen vier Widerstandswerte; die entsprechenden Messpunkte werden, sofern der Widerstand betrachtet wird, ebenfalls nicht berücksichtigt.

Spitzenwerte (*peaks*) Auf die genaue Struktur der gemessenen XRD-Spektren wird erst später eingegangen, da sowohl der physikalische Hintergrund, als auch die Verteilung der Ergebnisse näher erläutert wird. Es ist jedoch ist bereits hier zu vermerken, dass ihre interessantesten Merkmale die ausgeprägten Spitzenwerte sind, die die Messungen bei bestimmten Winkeln aufweisen (siehe Abb. 5.1).

Um die Spitzenwerte – auch Gipfel oder *peaks* genannt – in den Messreihen zu identifizieren, genügt es die gemessenen lokalen Extremwerte zu extrahieren und ihrer Größe



(a) CuNiZn



(b) NiCrRe

Abbildung 2.2.: Verteilung der erhobenen Messwerte

nach zu sortieren. Um ein lokales Maximum zu identifizieren, werden je die Intensitäten der beiden nächst kleineren und nächst größeren Winkel mit dem betrachteten Wert x an der Position i verglichen, wenn die Relation

$$x_{i-2} < x_i \wedge x_{i-1} < x_i \wedge x_{i+1} < x_i \wedge x_{i+2} < x_i$$

gilt, dann kann davon ausgegangen werden, dass es um sich ein lokales Maximum *(peak)* handelt. Dabei werden bewusst zwei links und rechts liegende Werte betrachtet, um kleinere lokale Gipfel, die beispielsweise durch Messungenauigkeiten entstehen können, auszuschließen. Versuche haben gezeigt, dass diese Annäherung für die stark ausgeprägten Spitzenwerte der XRD-Spektren ausreicht und keine genaueren Verfahren, wie zum Beispiel Glättung oder Extremwertberechnung mittels Differentialrechnung, nötig sind.

Die Automatisierung stellt bereits eine deutliche Vereinfachung des allgemeinen Arbeitsprozesses zur Analyse der XRD-Spektren dar; so beschreiben Takeuchi *et al.* [4] einen ganz ähnlichen Prozess, der jedoch per Hand durchgeführt wird und etwa drei Stunden für die Daten eines einzelnen Wafers in Anspruch nimmt.

Die Winkel der gefundenen Gipfel werden ihrer Größe nach gespeichert, wobei die Anzahl der gespeicherten Werte vorher festgelegt wird. Mit Hilfe dieser Operation lässt sich eine gesamte Spektralreihe von mehreren tausend Messungen auf einige wenige charakteristische Werte reduzieren. Ausgehend von der Position der Gipfel lassen sich weitere Merkmale ableiten. Derzeit werden neben der Position des Gipfels auch die folgenden abgeleiteten Werte berechnet:

1. Gemessene Intensität am Gipfel (n Werte)
2. Absolute Differenz zwischen der Intensität eines Gipfels und der Intensität des größten Gipfels ($n - 1$ Werte)
3. Absolute Differenz zwischen der Intensität eines Punktes und der Intensität des nächst größeren Gipfels ($n - 1$ Werte)
4. Absolute Differenz zwischen der Position eines Gipfels und der Position des größten Gipfels ($n - 1$ Werte)
5. Absolute Differenz zwischen der Position eines Punktes und der Position des nächst größeren Gipfels ($n - 1$ Werte).

Ob eine bestimmte Auswahl dieser Merkmale und Anzahl der betrachteten Gipfel für bestimmte physikalische Eigenschaften einer Materialprobe charakteristisch ist, wird in Kapitel 6 untersucht.

X/Y-Koordinaten auf der Waferscheibe Für einige Untersuchungen ist ein Zusammenführen der XRD-Spektren mit den übrigen Messungen notwendig, denn die Koordinaten der Messpunkte auf dem Wafer sind nicht in den Spektraldatensätzen enthalten. Diese weisen nur eine eindeutige ID auf. Durch dieses Zusammenfügen ist es möglich jeder der Spektraldaten auch eine X/Y-Koordinate auf dem Wafer zuzuweisen. Für den TiCoW-Datensatz liegen die Daten nur in Form der genaueren Messungen mit 4489 Werten vor, aus denen aber die X/Y-Koordinaten extrahiert werden konnten.

Die Koordinaten werden in Werten von 0 bis 21 angegeben, wobei aufgrund der runden Waferscheibe nicht alle Wertepaare vorkommen, wie zum Beispiel die Abbildung 5.5 zeigt.

3. Methoden

Dieser Abschnitt gibt einen Überblick über die verwendeten Methoden. Ziel ist es die zugrunde liegenden Techniken aufzuzeigen und eine eigenständige Interpretation der Resultate zu ermöglichen. Für ausführliche Beschreibungen sei auf die angegebene Literatur verwiesen.

3.1. Dynamic-Time-Warping (DTW) als Distanzmaß für Zeit- oder Messreihen

Es gibt viele verschiedene Ansätze, um die Ähnlichkeit zweier Mess- oder Zeitreihen zu bestimmen. Die Begriffe Zeit- und Messreihe können in diesem Zusammenhang analog verwendet werden, da lediglich Entscheidend ist, dass es eine kontinuierliche Folge von Beobachtungen gibt, ob diese Folge nun durch eine Zeitachse, oder wie im folgenden Fall durch aufsteigende Winkel der Röntgenmessung definiert ist, ist für die verwendeten Verfahren nicht von Belang.

Eine einfache Möglichkeit ist den Abstand zwischen den zu einem Zeitpunkt t gemessenen Werten zu berechnen und zu summieren. Ein Nachteil dieser Technik ist, dass die allgemeine Ähnlichkeit der Reihen nicht berücksichtigt wird. Aufgrund des bekannten Problems, dass die Ausschläge gleicher Materialphasen in den XRD-Spektren für verschiedene Messpunkte leicht verschoben sein können, wird stattdessen das Verfahren des Dynamic-Time-Warping (DTW) [5] genutzt.

Beim Dynamic-Time-Warping (DTW) handelt es sich um ein Verfahren, welches eine optimale Anordnung von zwei Zeitreihen berechnet. Formal werden zwei Zeitreihen $X := (x_1, \ldots, x_n)$ mit $n \in \mathbb{N}$ und $Y := (y_1, \ldots, y_m)$ mit $m \in \mathbb{N}$ zu einem Bildraum $\mathcal{F} = X \times Y$ zusammengefügt. Mit Hilfe einer Kosten- oder Distanzfunktion $c : \mathcal{F} \times \mathcal{F} \to \mathbb{R}_{\geq 0}$, z.B. der Manhattan-Metrik oder dem euklidischen Abstand, wird jedem Paar des Bildraumes ein Wert zugewiesen, der als Maß für die Ähnlichkeit zwischen diesen zwei Punkten dient. Als Resultat erhält man eine Kostenmatrix $C \in \mathbb{R}^{n \times m}$, in der jeder Punkt $C(i, j) = c(x_i, y_j)$ der Ähnlichkeit zwischen den Punkten x_i und y_j entspricht.

Eine optimale Anordnung der Zeitreihen entspricht dann einem Pfad mit dem Ursprung im Punkt $p_1 = (x_1, y_1)$ bis zum Punkt $p_L = (x_n, y_m)$, welcher die aufsummierten Kosten aller besuchten Punkte minimiert. Dabei gilt neben der bereits erwähnten Einschränkung des Start- und Endpunktes auch, dass der Pfad nicht streng monoton und stetig wächst, also weder Punkte übersprungen, noch Umwege gegangen werden.

Die Summe der Kosten des optimalen Pfades ist die DTW-Distanz, welche als Maß für die Ähnlichkeit zweier Zeitreihen dient. Zu beachten ist, dass im Allgemeinen die DTW-Distanz nicht die Dreiecksungleichung erfüllt, selbst wenn dies für die Kostenfunktion c gilt. Auch kann es mehrere optimale Pfade geben, welche jedoch die gleiche DTW-Distanz haben. Für den einfachen Fall, dass lediglich die Ähnlichkeit von zwei

oder mehr ähnlich aufgebauten Reihen untersucht werden soll, ist dies jedoch keine gravierende Einschränkung. Ein Vorteile der DTW-Distanz isr, dass die beiden Reihen nicht zwingend die gleiche Länge haben müssen, was vor allem bei sehr langen, physikalischen Messungen leicht vorkommen kann. Außerdem ist das Verfahren relativ robust gegenüber zeitlichen Verschiebungen des Signals, da der optimale Pfad mit minimalen Kosten nicht zwingend in beiden Achsen gleich stark ansteigen muss, sondern, wie der Name nahelegt, dynamisch auch ähnliche Punkte mit geringer Kostenfunktion in der Nachbarschaft findet. Der gefundene Pfad kann zwar dadurch länger werden, aber in seinen Gesamtkosten dennoch günstiger sein als die Summe der Kosten des kürzesten (geraden) Pfades.

$y_5 = 1$	3	2	1	3	***1***
$y_4 = 3$	1	0	1	***1***	1
$y_3 = 4$	0	1	2	***0***	2
$y_2 = 1$	3	2	***1***	3	1
$y_1 = 4$	***0***	***1***	2	0	2
	$x_1 = 4$	$x_2 = 3$	$x_3 = 2$	$x_4 = 4$	$x_5 = 2$

Abbildung 3.1.: Beispiel eines optimalen DTW-Pfades. Die Tabelleneinträge sind der Abstand zwischen den Wertepaaren (x_i, y_j) und die markierten Einträge sind die Zellen des günstigens DTW-Pfades von (x_1, y_1) nach (x_5, y_5). Der so gefundene Pfad hat eine summierte Abweichung von 4, während der kürzeste Pfad eine Abweichung von 6 hätte.

Die Berechnung des optimalen Pfades lässt sich mittels dynamischer Programmierung in einer Laufzeitkomplexität von $\mathcal{O}(NM)$ realisieren. Es gibt verschiedene Ansätze die Laufzeit weiter zu senken [6, 7], aber für die Größenordnung der hier untersuchten XRD-Spektren erweist sich der einfache Ansatz als schnell genug (die Berechnung aller 116.964 möglichen DTW-Distanzen bei 342 Messpunkten benötigt nur wenig Sekunden).

3.2. Korrelationskoeffizient nach Bravais-Pearson

Als Maß für den den linearen Zusammenhang zwischen zwei Größen kann der Korrelationskoeffizient nach Bravais-Pearson [8] verwendet werden. Der Korrelationskoeffizient ergibt sich durch die empirische Kovarianz, normiert durch das Produkt der Standardabweichungen

$$r = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2 \sum_{i=1}^{n}(y_i - \overline{y})^2}}$$

und besitzt folgende Eigenschaften:

- der Wertebereich ist normiert auf: $-1 \leq r \leq 1$
- Ist $r = \pm 1$, so liegen alle Punkte auf einer Geraden mit Steigung entsprechend dem Vorzeichen.

3.3. Lineare Regression

Das Ziel einer linearen Regression ist es einen Zusammenhang zwischen einer abhängigen Zielgröße y und einer oder mehreren unabhängigen Einflussgrößen $x_1, \ldots, x_m$ zu bestimmen. Die Regression lässt sich als Gleichung

$$Y = X\beta + \epsilon$$

auffassen, ausgeschrieben entspricht dies den n Gleichungen $y_i = \beta_0 + \beta_1 x_{1,i} + \ldots + \beta_m x_{m,i} + \epsilon_i$ mit $i = 1 \ldots n$. Dabei ist $Y \in \mathbb{R}^n$ der Vektor der abhängigen Variablen y_i, $X \in \mathbb{R}^{n \times (m+1)}$ enthält n Ausprägungen der Einflussgrößen $x_1, \ldots, x_m$ und $\beta \in \mathbb{R}^{m+1}$ ist der unbekannte Parametervektor, den es zu bestimmen gilt; und $\epsilon \in \mathbb{R}^n$ stellt den Fehlervektor dar. Aufgabe der Regression ist es eine Lösung für β zu finden, so dass der Fehlerterm ϵ minimiert wird. Unter der statistischen Annahme, dass die einzelnen Elemente ϵ_i unabhängig identisch normalverteilt sind und den Erwartungswert 0 haben – kurz $\epsilon_i \sim \mathcal{N}(0, \sigma^2)$ – liefert die Methode der kleinsten Quadrate [8] eine optimale Lösung für den Parametervektor β.

Das adjustierte Bestimmtheitsmaß R^2 [9] ist eine sinnvolle Beurteilungsgröße für die Güte eines linearen Modells. Es gibt das Verhältnis der durch das Modell erklärten Varianz zur gesamten Fehlervarianz an. Nimmt R^2 den Wert 1 an, besteht ein vollständiger linearer Zusammenhang und die Zielgröße Y wird vollständig durch die Regressoren erklärt; bei einem Wert von 0 gibt es keinen linearen Zusammenhang. Ein weiterer Vorteil ist, dass das adjustierte Bestimmtheitsmaß zu komplexe lineare Modelle mit sehr vielen Einflussgrößen bestraft. Denn eine Anomalie des einfachen Bestimmtheitmaßes ist, dass das Hinzufügen weiterer Einflussgrößen (selbst nicht relevante, wie reine Zufallszahlen) dessen Wert verbessert.

3.4. Regressions-Support-Vector-Machine (SVM)

Eine Support-Vector-Machine (SVM) ist ein Klassifikationsverfahren bei dem versucht wird, eine Hyperebene durch den Merkmalsraum zu legen um zwei Klassen von Beispielen optimal zu separieren. Die Hyperebene soll dabei so positioniert werden, dass der Abstand zu den am nächsten gelegenen Punkten maximal wird; dieser möglichst breite Rand soll auch für unbekannte Punkte eine zuverlässige Klassifikation gewährleisten. Die Punkte entlang des Randes bilden die namensgebenden Stützvektoren der SVM.

Formal lässt sich das Lösungsverfahren der SVM folgendermaßen zusammenfassen. Für eine Trainingsmenge

$$T = \{(x_1, y_1), \ldots, (x_n, y_n) | \, x_i \in \mathbb{R}^n, \, y_i \in \{-1, +1\}\}$$

lässt sich die Lösung des Optimierungsproblems auf die Maximierung der Karush-Kuhn-Tacker-Bedingung

$$\sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j < x_i, x_j > \tag{3.1}$$

zurückführen, wobei die α_i für n Lagrange-Multiplikatoren der Nebenbedingung, dass jeder Punkt korrekt klassifiziert wird, steht.

Die strikte Forderung, dass alle Beispiele korrekt klassifiziert werden, kann gelockert werden in dem zusätzlich Schlupfvariablen werden; diese bestrafen die Verletzung der Nebenbedingung. Die Summe der Verletzung wird mit einer Konstanten C gewichtet und dem Minimierungsproblem 3.1 hinzugefügt. Je größer die Gewichtung durch C ist, desto stärker wird eine saubere, lineare Lösung angestrebt.

Um eine nichtlineare Seperation zu ermöglichen, können die Daten in einen höher dimensionalen Raum transformiert werden. Die Verwendung eines Kernels, zum Beispiel eine Polynomfunktion, erleichtert dieses Verfahren dadurch, dass die Transformation lediglich implizit erfolgt und das Minimierungsproblem selbst nicht in dem höherdimensionalen Raum gelöst werden muss.

Um mit einer SVM ein Regressionsproblem zu lösen, ist die Beurteilungsgröße nicht mehr die korrekte Klassifikation, sondern der Abstand der Daten zu der gesuchten Hyperebene. Daraus ergibt sich die doppelte Menge von Nebenbedingungen, da für jeden Punkt der Abstand ober- und unterhalb der Hyperebene berücksichtigt werden muss.

Als Implementierung wird die ν-SVM [10] aus der LibSVM-Bibliothek [11] verwendet.

Im Gegenzug zu linearen Modellen kann die Regressions-SVM durch die Verwendung eines Kernels auch kompliziertere, nichtlineare, Wechselwirkungen zwischen den Einflussgrößen modellieren. Ein Nachteil ist, dass die gefundenen Parametereinstellungen und Stützvektoren schwer zu interpretieren sind – im Gegensatz zum Parametervektor β für lineare Modelle. Außerdem hängt die Vorhersagegüte einer SVM stark von den Voreinstellungen z.B. für C oder der Kernelfunktion ab. Eine Optimierung dieser Einstellungen ist also zwingend notwendig um repräsentative Ergebnisse zu erhalten.

3.5. k-means-Clustering

Das Ziel des Clustering ist es, ähnliche Merkmale zu Gruppen zusammen zu fassen. Eines der bekanntesten und zugleich einfachsten Verfahren hierfür ist das *k-means-Clustering* [12], auch bekannt als Lloyd-Algorithmus. Der Algorithmus beruht auf den iterativen Schritten die Menge der Beobachtungen zunächst in k Gruppen zu unterteilen. Ein Punkt wird der Klasse zugeteilt, zu deren Zentrum er am nächsten liegt. Beim Start der nächsten Iteration werden die Klassenzentren so verschoben, dass sie im Schwerpunkt aller ihrer zugehörigen Punkte liegen.

1. Initialisiere Klassenzentren
2. Weise Punkte dem jeweils nächsten Klassenzentren neu zu
3. Setze Schwerpunkte der Klassen als neue Klassenzentren
4. Prüfe Abbruchkriterium und beginne mit neuer Iteration

Als Abbruchkriterium kommt eine minimale Veränderung der Klassenzentren oder eine maximale Anzahl von Iterationen in Frage. Je nach ursprünglichen Startpunkten für die Klassenzentren kann das Endergebnis variieren.

3.6. Entscheidungsbäume

Eine andere Möglichkeit der Klassifikation bieten Lernverfahren mit Entscheidungsregeln. Der Vorteil liegt hierbei daran, dass die Ergebnisse in der Regel leicht zu interpretieren sind. Eine Regel könnte zum Beispiel die Form haben:

> *"Wenn der Kupferanteil größer ist als 40% und der Nickelanteil kleiner als 20%, dann ist der spezifische Widerstand kleiner als 200$\mu\Omega$."*

Entscheidungsbäume sind eine Form des Regelbasiertenlernens, bei der die Regeln durch innere Knoten eines Baumes dargestellt werden. Die Blattknoten stellen dann die Entscheidung für eine Kategorie dar. Für das obige fiktive Beispiel gäbe es somit einen Knoten der nach dem Kupferanteil unterscheidet, danach einen Knoten der den Nickelanteil untersucht und zuletzt einen Blattknoten, der die Entscheidung für einen Widerstand kleiner 200$\mu\Omega$ repräsentiert.

Für eine gegebene Menge von Beispielen lässt sich ein Entscheidungsbaum finden, der diese Menge perfekt in ihre Klassen einteilt, dieser Baum wäre jedoch exakt an diese Beispiele angepasst und würde für neue, unbekannte Beispiele mit großer Wahrscheinlichkeit sehr schlechte Vorhersagen treffen. Man muss einen Mittelweg finden zwischen der Überanpassung an bekannte Daten und einer sehr groben Unterteilung, durch nur sehr wenige Regeln.

Der verwendete Algorithmus ähnelt dem C4.5-Algorithmus von Quinlan [13]. Dieser teilt an jedem Knotenpunkt die Daten derart auf, dass das gewählte Kriterium (Informationsgewinn, *Accuracy*) maximiert wird. Eine Überanpassung wird am Ende dadurch verhindert, dass der Algorithmus versucht zu kleine Verzweigungen durch entfernen der entsprechenden Regeln den Entscheidungsbaum wieder zu verkleinern, ein Vorgang der *pruning* [14] genannt wird. Ein Vorteil dieses Verfahrens ist, dass sowohl nominelle, als auch stetige Attribute verwendet werden können.

Die Güte eines Entscheidungsbaumes kann durch die Angabe der *Precision*- und *Recall*-Werte beurteilt werden. Der *Precision*-Wert ist der Anteil zwischen den richtigen Vorhersagen für eine Klasse und der Gesamtzahl der Vorhersagen. Der *Recall*-Wert ist ist das Verhältnis zwischen korrekten Vorhersagen und der Anzahl aller Elemente in dieser Klasse. Beide Werten haben den Vorteil. dass sie auch berechnet werden können, wenn die Grundgesamtheit der Beobachtungen unbekannt ist. Sind alle möglichen Beobachtungen hingegen bekannt kann auch die *Accuracy*, dass Verhältnis aller richtigen Vorhersagen zur Gesamtzahl der Beobachtungen, berechnet werden.

3.7. Kreuzvalidierung

Ein Problem bei komplexen Modellen im maschinellen Lernen ist die Überanpassung an vorhandene Daten, was auf die Vorhersagegüte für neue, unbekannte, Daten verschlechtert. Ein Beispiel wäre zum Beispiel, wenn ein vorhandenes Grundrauschen von Sensoren mit in die Vorhersage einfließen würde. Diese Anfälligkeit von Methoden mit vielen Parametern ist einer der Gründe, warum oft auf einfachere Modelle, wie die lineare Regression, verwendet werden.

Eine Überanpassung kann durch eine Kreuzvalidierung verhindert werden. Dabei werden die Daten in eine Trainings- und eine Testmenge unterteilt. Mit Hilfe der Trainingsmenge wird dann das Vorhersagemodell erstellt und dessen Qualität auf der Testmenge überprüft. Damit kann sichergestellt werden, dass das Methode auch auf neuen, unbekannten Daten zuverlässige Ergebnisse liefert.

Um eine zufällig besonders günstige Auswahl der Trainings- und Testmenge zu verhindern, wird häufig eine k-fache Kreuzvalidierung verwendet. Dabei werden die Daten in k gleich große Teilmengen unterteilt. Anschließend wird in k Durchläufen jeweils eine der Mengen als Testmenge ausgewählt und die übrigen k-1 als Trainingsmengen. Wichtig ist bei der anschließenden Beurteilung, dass die durchschnittliche Leistung aller Validierungsläufe betrachtet wird und nicht nur der beste Durchgang.

4. Direkte Modellierung des elektrischen Widerstandes

In diesem Abschnitt soll untersucht werden, ob der elektrische Widerstand anhand anderer Messgrößen vorhergesagt werden kann. Zwei Anwendungsfälle motivieren diesen Ansatz: die Modellierung ersetzt entweder eine vollständige Messung, oder aber der Vergleich des Modells mit den tatsächlichen Werten offenbart unvorhergessene Abweichung, die auf unerwartete Materialphasen hindeuten können.

4.1. Regression mittels SVM

Die erste verwendete Methode ist die Regressions-SVM. Mit ihrer Hilfe wird nach einem funktionalen Zusammenhang zwischen dem Widerstand, als Zielgröße, und den übrigen Messwerten, als Einflussvariablen, gesucht. Dieser Zusammenhang soll möglichst allgemein sein und nicht nur für die getätigten Messungen gelten, weswegen die Ergebnisse mittels einer fünffachen Kreuzvalidierung überprüft werden.

Bei der Verwendung einer SVM ist zu beachten, dass der Wertebereich der Variablen normiert wird; außerdem bedarf es es einer genauen Einstellung der Paramter. Eine Gittersuche über die möglichen Parameterkombiniationen ergab die in Tabelle 4.1 angegebenen Werte für die untersuchten Datensätze.

Bei der Kreuzvalidierung ergibt sich mit diesen Einstellungen eine durchschnittlicher Fehlerrate von $4,93\%$ für CuNiZn und $10,89\%$ für NiCrRe auf. Im Vergleich dazu, ist ein einfaches lineares Modell für die gleichen Attribute mit einer Fehlerrate von etwa 22%, bzw. 30% deutlich schlechter.

Betrachtet man die Verlaufskurve (Abb. 4.1) der tatsächlichen Widerstandswerte und der modellierten Werte, so fällt auf, dass der Modellfehler mit zunehmendem Widerstand größer wird, bei einem fehlerfreien Modell, würden die Punkte exakt auf einer Geraden liegen. Für die NiCrRe-Daten ist diese Streuung noch stärker ausgeprägt und der relative Fehler entsprechend groß. Die Untersuchungen legen nahe, dass eine Erhöhung des Gamma-Wertes die Ergebnisse für die NiCrRe-Daten noch minimal verbessern könnten. Allerdings führt dies zu einer deutlichen Erhöhung der Rechenzeit (von einigen Sekun-

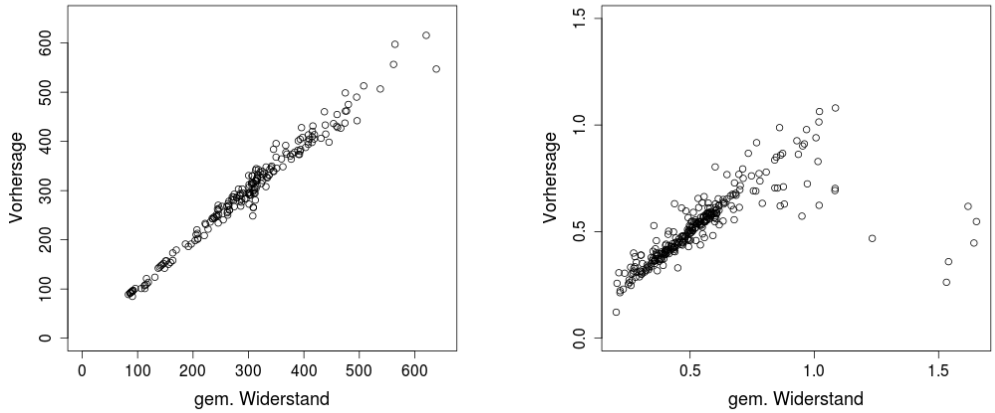
γ	Polynom Grad	C	ϵ	ν
$1,5$	5	40	$0,5$	$0,8$

(a) CuNiZn

γ	Polynom Grad	C	ϵ	ν
1	5	20	$0,01$	$0,8$

(b) NiCrRe

Tabelle 4.1.: Parametereinstellung für die $\nu - SVM$



(a) CuNiZn (b) NiCrRe

Abbildung 4.1.: Scatterplot zwischen gemessenen und modellierten Widerständen der Regressions-SVM

den auf mehrere Minuten pro Durchgang der Kreuzvalidierung), ohne dass die extremen Ausreißer im oberen Bereich positiv beeinflusst werden.

4.2. Klassifikation mittels Entscheidungsbäumen

Für die Klassifikation durch Entscheidungsbäume ist eine Einteilung der Zielgröße in Klassen notwendig. In diesem Fall wurden die Widerstandswerte in drei gleich große Klassen aufgeteilt; es ergeben sich Klassengrenzen von $(-\infty - 209)$; $[209 - 309)$; $[309 - \infty)$ für CuNiZn und von $(-\infty - 0,43)$; $[0,43 - 0.56)$; $[0,56 - \infty)$ für NiCrRe.

Es zeigt sich, dass die Güte des Lernverfahren sehr stark von der Parameterwahl abhängt. Für die CuNiZn-Daten haben einfache Entscheidungsbäume haben Probleme einzelne Bereiche scharf zu trennen, die größte Fehlerrate liegt in der mittleren Klasse. Eine sehr gute Kategorisierung mit einer hohen Genauigkeit von über 90% erzeugt im Gegenzug einen sehr komplexen Entscheidungsbaum. Die Abbildungen 4.2 und 4.3 zeigen Beispiele dieser beiden extremen Fälle für die CuNiZn-Daten.

Es scheint also keine einfache Menge an Regeln zu geben, welche die vorliegenden Daten gut klassifizieren können. Auffällig ist jedoch das erste Blatt im Entscheidungsbaum in Abbildung 4.2, das alle Beispiele mit mit einem Nickel-Anteil größer als $9,68\%$ der Kategorie des größten Widerstandes zuordnet, dabei aber in etwa der Hälfte der Beispiele falsch liegt. Diese Besonderheit wird im nächsten Abschnitt näher untersucht.

Für die NiCrRe-Daten findet sich keine einfache Regelmenge, die einen kleinen Entscheidungsbaum mit wenigen Knoten erstellt. Die Suche nach einem Entscheidungsbaum mit möglichst hoher Accuracy führt zu einem ähnlich komplexen Modell (Abb. 4.4) wie bei den CuNiZn-Daten, im Gegensatz zu diesen bleibt die Accuracy jedoch niedrig.

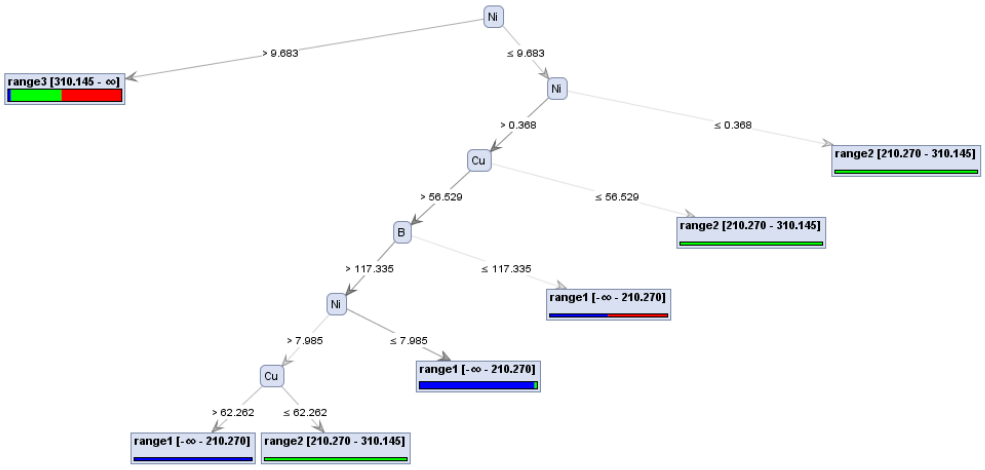


Abbildung 4.2.: Kleiner aber ungenauer Entscheidungsbaum für die CuNiZn-Daten. Range 1-3 in den Blattknoten gibt die zugeordnete Klasse an; die Farben entsprechen den tatsächlichen Kategorien der Beispiele.

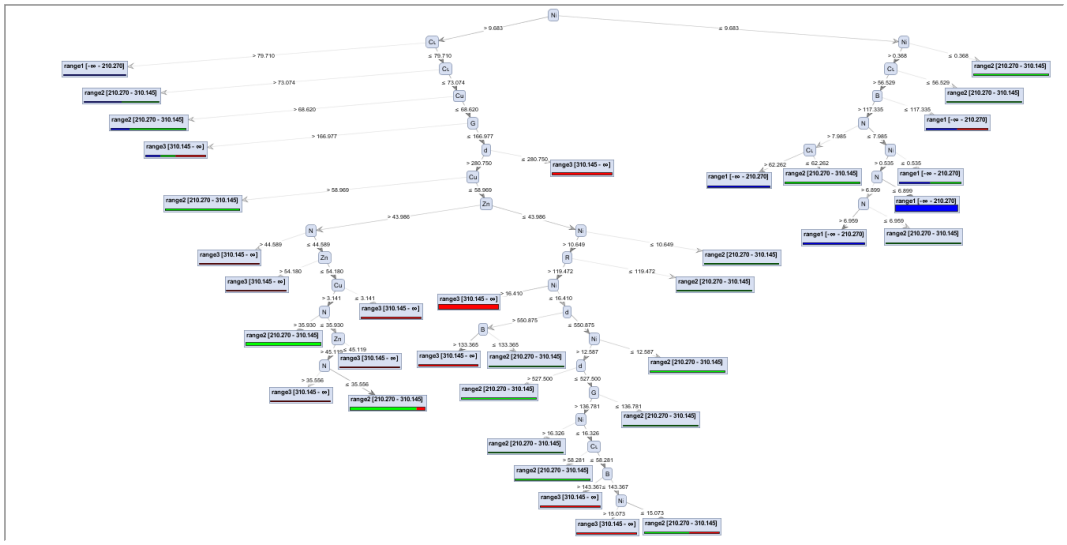


Abbildung 4.3.: Komplexer, über angepasster Entscheidungsbaum (CuNiZn)

Vorhersage/ wahre Klasse	w. Kl. 1 $[-\infty - 210)$	w. Kl. 2 $[210 - 310)$	w. Kl. 3 $[310 - \infty)$	Precision
vorh. Kl. 1 $[-\infty - 210)$	97	7	1	92%
vorh. Kl. 2 $[210 - 310)$	6	12	0	66%
vorh. Kl. 3 $[310 - \infty)$	11	95	113	51%
Recall	85%	10%	99%	
	Accuracy: 65%			

Tabelle 4.2.: Vorhersage Verteilung zum Entscheidungsbaum aus Abbildung 4.2

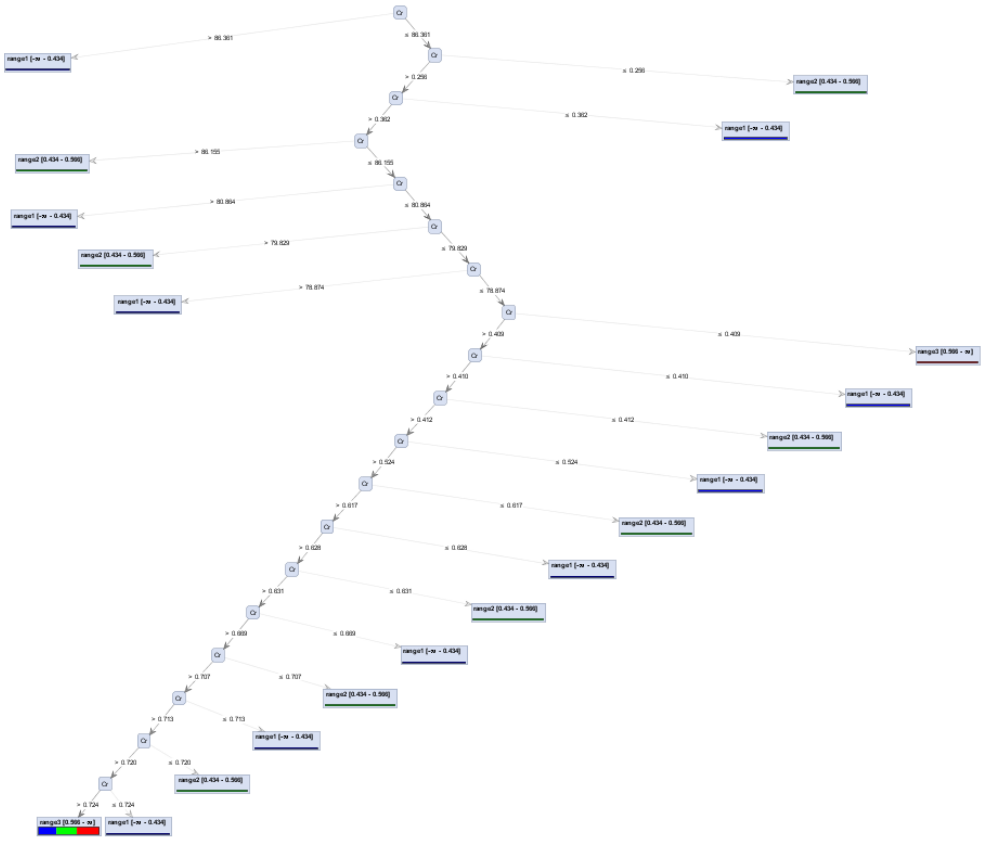


Abbildung 4.4.: Komplexer, über angepasster Entscheidungsbaum (NiCrRe)

Vorhersage/ wahre Klasse	w. Kl. 1 $[-\infty - 210)$	w. Kl. 2 $[210 - 310)$	w. Kl. 3 $[310 - \infty)$	Precision
vorh. Kl. 1 $[-\infty - 210)$	105	5	1	94%
vorh. Kl. 2 $[210 - 310)$	8	87	19	76%
vorh. Kl. 3 $[310 - \infty)$	1	22	94	80%
Recall	92%	76%	82%	
	Accuracy: 83%			

(a) CuNiZn

Vorhersage/ wahre Klasse	w. Kl. 1 $[-\infty - 0,43)$	w. Kl. 2 $[0,43 - 0,5)]$	w. Kl. 3 $[0,56 - \infty)$	Precision
vorh. Kl. 1 $[-\infty - 0,43)$	93	24	8	74%
vorh. Kl. 2 $[0,43 - 0,56)$	15	72	19	67%
vorh. Kl. 3 $[0,56 - \infty$	4	17	86	80%
Recall	83%	63%	76%	
	Accuracy: 74%			

(b) NiCrRe

Tabelle 4.3.: Vorhersagegüte der komplexen Entscheidungsbäumen aus Abbildung 4.3 und 4.4

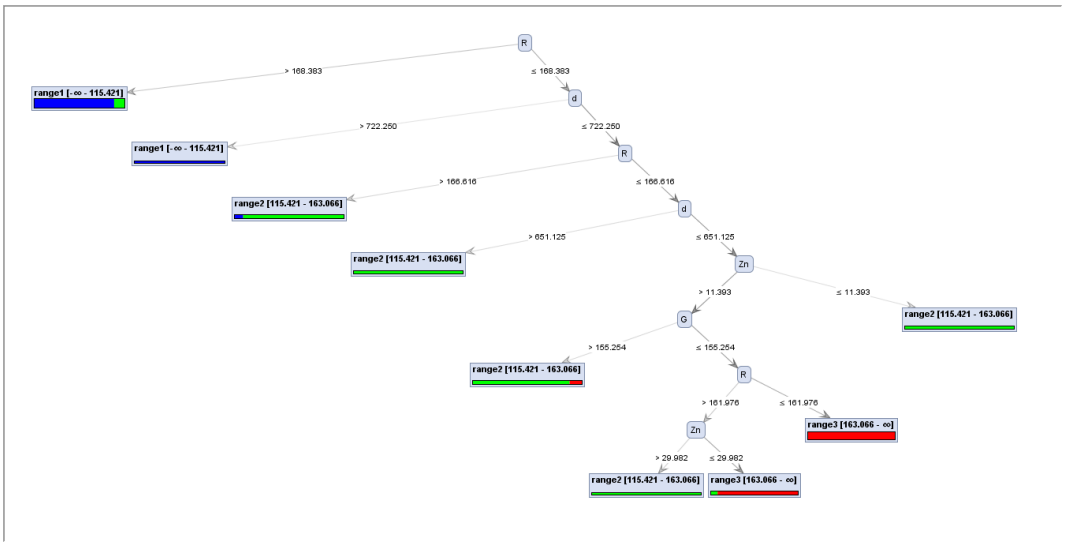


Abbildung 4.5.: Entscheidungsbaum für Messwerte mit einem Nickelanteil $\leq 9,683\%$

4.3. Kombination von Regression und Entscheidungsbäumen

Das Ergebnis der Auswertung der Entscheidungsbäume zeigt, dass ohne eine Überanpassung an die vorliegenden Daten es anscheinend keine einfache Regelmenge zur Kategorisierung des Widerstandes gibt. Der einfachere Entscheidungsbaum aus Abbildung 4.2 ist insofern jedoch interessant, da ein großer Bereich der Daten, nämlich die Messungen für einem Nickelanteil größer als $9,683\%$ nicht korrekt getrennt werden können. Ein möglicher Ansatz ist nun, die Daten anhand dieser Regel aufzuteilen und für den einen Bereich eine Vorhersage mittels Regression zu versuchen und den anderen Bereich weiterhin mittels eines Entscheidungsbaumes zu kategorisieren.

Für die Regression mittels der SVM müssen neben der Vorfilterung der Werte keine weiteren Besonderheiten beachtet werden. Für die 212 Messungen mit einem Nickelanteil größer als $9,68\%$ ergibt sich mit den in Abschnitt 4.1 gefunden Parametereinstellungen ein Vorhersagefehler von $5,10\%$, welcher etwas schlechter ist der Fehler über alle Daten.

Für die Entscheidungsbäume ist es notwendig die Klassen neu einzuteilen, da für Messungen mit einem Nickelanteil $< 9,68\%$ nur ein einziges Beispiel in der Klasse für einen spezifischen Widerstand größer als $310\mu\Omega$ verbleibt und so die Auswertung verzerrt wäre. Für die verbleibenden 127 Beispiele ergibt sind eine neue Einteilung der Grenzen $(-\infty - 115)\,;\,[115 - 163)\,;\,[163 - \infty)$ und der daraus resultierende Entscheidungsbaum ist in Abbildung 4.5 zu sehen und die Fehlerverteilung in Tabelle 4.4, wobei die Parametereinstellung unverändert übernommen wurden. Mit einer *Accuracy* von 80% ist dieser Baum deutlich besser geeignet als der vorherige, ohne eine viel größere Komplexität aufzuweisen.

Vorhersage/ wahre Klasse	w. Kl. 1 $[-\infty - 115)$	w. Kl. 2 $[115 - 163)$	w. Kl. 3 $[163 - \infty)$	Precision
vorh. Kl. 1 $[-\infty - 115)$	35	4	0	89%
vorh. Kl. 2 $[115 - 163)$	6	29	5	72%
vorh. Kl. 3 $[163 - \infty)$	1	9	38	79%
Recall	83%	69%	88%	
	Accuracy: 80%			

Tabelle 4.4.: Vorhersage Verteilung zum Entscheidungsbaum aus Abbildung 4.5

5. Einbeziehung von XRD-Spektren in die Modellierung

Neben den zuvor untersuchten Materialeigenschaften, stehen als Informationsquelle noch die aus der Röntgenbeugungsanalye gewonnenen XRD-Spektren zur Verfügung. Bei diesen gibt es jedoch zu beachten, dass diese nicht als einfache numerische Werte vorliegen, sondern als Messreihe über eine Vielzahl von gemessenen Winkeln.

Zunächst erfolgt eine Beschreibung der physikalischen Grundlage und der resultierenden Daten. Werden die Ergebnisse mittels Dynamic-Time-Waping zusammengefasst und überprüft ob diese Methode eine sinnvolle Aggregation der XRD-Spektren ermöglicht.

5.1. Röntgenbeugungsanalyse

Eine besonders genaue Strukturanalyse der Materialprobe ist die Analyse mittels Röntgenbeugung (engl. *X-Ray Diffraction*). Dabei wird ein Punkt der Probe aus verschiedenen Winkeln geröntgt und die reflektierte Strahlung gemessen. Die physikalische Grundlage dieses Verfahrens beruht darauf, dass die Wellenlänge von Röntgenstrahlung zwischen 1 nm ($10^{-9}\,m$) und 1 pm ($10^{-12}\,m$) liegt, was in etwa dem Abstand von Atomen in einer Kristallstruktur entspricht. Je nach Abstand und Anordnung der Atome entstehen so bei unterschiedlichen Winkeln unterschiedlich starke Interferenzen. Durch diese Interferenzen unterscheidet sich wiederum die Intensität der reflektierten Strahlung, was Rückschlüsse auf die atomare Struktur der Materialprobe erlaubt. Das Verfahren hat zwei Nachteile: die Analyse eines vollständigen Wafers kann mehrere Tage benötigen und es entsteht eine sehr große Datenmenge. Verschiedene Visualisierungsmöglichkeiten der aus den Messungen gewonnenen Spektren zeigen Takeuchi et al. [4] und gehen dabei auch auf die Schwierigkeiten der Analyse näher ein.

5.2. Datengrundlage XRD-Spektren

Die XRD-Spektren werden für einen Punkt auf dem Wafer als eine Reihe von Messungen in unterschiedlichen Winkeln θ gemessen. Für die betrachteten Datensätze liegen diese Winkel in Bereichen zwischen $30 - 85°$ (NiCrRe, TiCoW) bzw. $26 - 80°$ (CuNiZn) wobei zwischen 2030 (CuNiZn) und 4189 (NiCrRe, TiCoW) verschiedene Winkel gemessen werden. Die Abstufung der Winkel ist also mit Schrittweiten von etwa einem hundertstel Grad sehr genau, wobei die θ-Werte selbst bis auf acht Nachkommastellen genau angegeben sind.

Die meisten Werte der gemessenen Intensität liegen in Bereichen zwischen 0 und 1000. Spitzenwerte hingegen können eine Intensität von über 1.000.000 haben, liegen aber im

Mittel zwischen 10.000 und 100.000. Diese Spitzenwerte sind in der Regel auf nur wenige, nebeneinander liegende, Winkel beschränkt und somit als deutliche Spitzen in den Messreihen zu erkennen.

Für die Analyse der Materialien gilt es, diese Spektren mit denen bekannter Materialien zu vergleichen um auf bekannte Materialphasen zu schließen. Dabei treten zwei verschiedene Probleme auf. Zum einen sind die Peaks, also die Ausschläge der Messungen bei bestimmten Winkeln, alleine als Merkmal nicht ausreichend, um bestimmte Eigenschaften zu folgern. Sie sind lediglich ein Indiz dafür, dass eine ähnliche Materialphase vorliegen kann. Zum anderen kann die Lage der Spitzen durch Messeffekte, wie zum Beispiel Überlagerungen der Signale nebeneinander liegender Proben, um einige Grad verschoben sein. Die Abbildung 5.1 zeigt typische Beispiele für die Form der gemessenen Spektren.

5.3. Auswertung des Dynamic-Time-Warping für die vorliegenden Spektraldaten

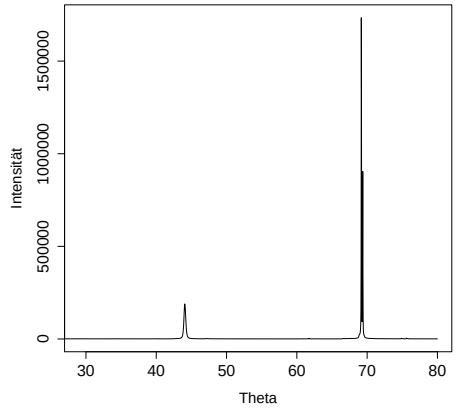
Um einen Überblick über die Anwendung des DTW auf die XRD-Spektren zu bekommen, werden nun die Ergebnisse der drei vorliegenden Messungen einzeln analysiert. Für diese Analyse werden alle Kombinationen von Punkten betrachtet.

Der Vergleich der drei Histogramme aus Abbildung 5.2 zeigt deutliche Unterschiede der verschiedenen Systeme. Zunächst sind die Größenordnungen der gemessenen Distanzen sehr stark. Während beim TiCoW-System die maximale Distanz 54.252 beträgt, liegt sie bei CuNiZn bei 3.531.199 und bei NiCrRe sogar bei 3.940.893. Dabei fällt insbesondere die extreme Schiefe des Histogramms bei NiCrRe auf; während bei den anderen beiden Systemen Median und Mittelwert relativ nah beieinander liegen, ist dort der Mittelwert beinahe acht mal so groß. Alleine diese großen Unterschiede, sowohl in den einzelnen Systemen, als auch zwischen den Punkten innerhalb eines Systems, zeigen, dass eine allgemeine Klassifizierung anhand der DTW-Distanz schwierig ist.

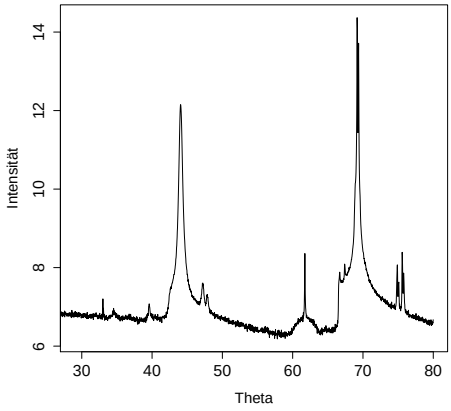
5.4. Aggregation mittels Breitensuche und DTW-Distanzen

Trotz der starken Unterschiede der gemessenen Spektralreihen, zeigt ein einfacher Vergleich, dass viele Messreihen einander ähneln. Es stellt sich die Frage, ob ähnliche Muster auch in ähnlichen Bereichen des Wafers angesiedelt sind. Damit ließe sich zum Beispiel überprüfen, ob sich der Verlauf einer Materialphase auch in den XRD-Spektren wiederfinden lässt.

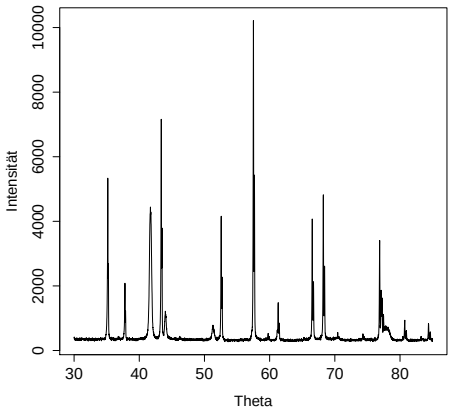
Hierfür wird die Nachbarschaft eines gegebenen Startpunktes in schrittweise größer werdenden Kreisen auf ihre DTW-Distanz bezüglich des Startpunktes untersucht. Um den Suchraum einzuschränken, wird die Suche in eine bestimmte Richtung abgebrochen, falls die Distanz über einem festgelegten Schwellwert liegt.



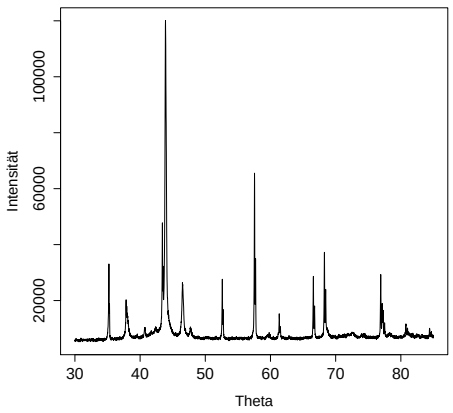
(a) CuNiZn



(b) CuNiZn mit logarithmierten Intensitätswerten

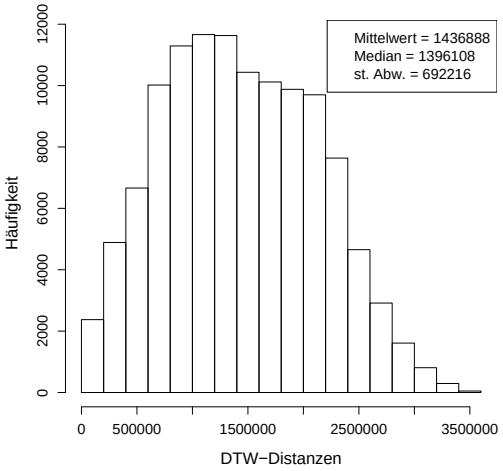


(c) NiCrRe

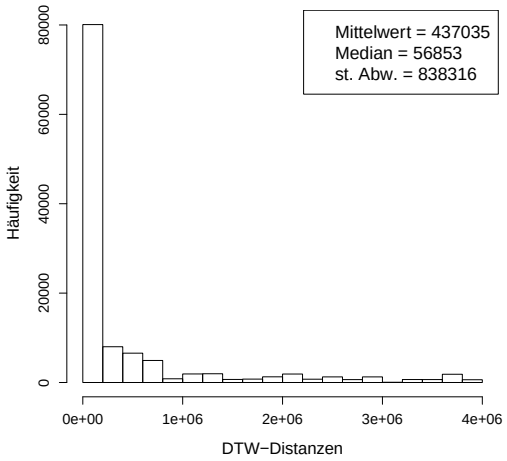


(d) TiCoW

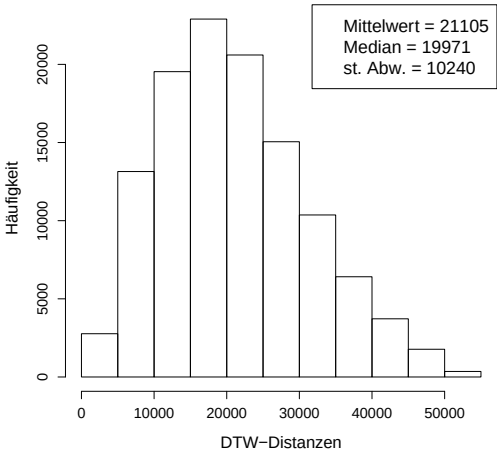
Abbildung 5.1.: Beispiele für gemessene XRD-Spektren der unterschiedlichen Materialproben



(a) DTW-Distanzen im CuNiZn-System



(b) DTW-Distanzen im NiCrRe-System



(c) DTW-Distanzen im TiCoW-System

Abbildung 5.2.: Histogramme der DTW-Distanzen zwischen jeweils allen Punkten der einzelnen tertiären Systeme

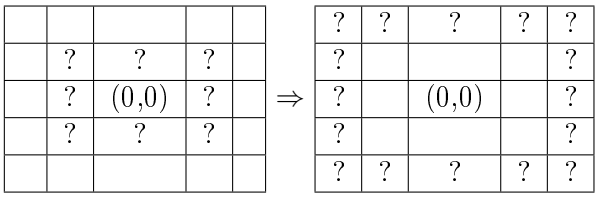


Abbildung 5.3.: Untersuchte Punkte im ersten & zweiten Iterationsschritt

5.5. Funktionsbeschreibung und Implementierung

Als Startparameter für die DTW-Breitensuche werden zwei Angaben benötigt: die Startposition der Suche auf dem Wafer als X/Y-Koordinaten und die maximale DTW-Distanz, die zwischen diesem Startwert und einem weiteren Punkt liegen darf, damit dessen Spektralmuster noch als ähnlich genug angesehen wird.

Für das Zusammenfassen ähnlicher Messwerte wird eine Breitensuche verwendet. Ausgehend vom Startwert werden in jeder Iteration alle Punkte mit einer Distanz, bezüglich ihrer Lage auf dem Wafer, gleich der Iterationszahl, auf ihre DTW-Distanz untersucht. Im ersten Schritt werden die acht Punkte der Moore-Nachbarschaft untersucht, im zweiten die 16 Zellen mit einem Abstand von zwei zum Ursprung, wie in Abbildung 5.3 zu sehen ist. Die Anzahl der zu untersuchenden Punkte steigt mit jedem Iterationsschritt um acht.

Das Verfahren ist in RapidMiner [15] als eigenständiger Operator implementiert worden. Mit Operatoren werden in RapidMiner einzelne, modulare, Programmteile bezeichnet, welche zu einem fertigen Programm zusammengefügt werden können. Die Ein- und Ausgabeformate der Operatoren sind festgelegt und erlauben es so auch komplizierte Programmabläufe zu realisieren. So erwartet der Operator für die DTW-Breitensuche eine Beispielmenge von XRD-Spektren, welche durch die X/Y-Koordinaten eindeutig identifizierbar sind. Als Parameter für die Suche hat der Operator die Startwertkoordinaten und die maximale DTW-Distanz. Die Ausgabe ist die gleiche vollständige Beispielmenge der Eingabe, mit den zusätzlichen Informationen über die Kategorisierung aller Punkte (diese ermöglicht z.B die folgenden Abbildungen), sowie die berechneten DTW-Distanzen.

5.6. Verwendete Metriken zur Begrenzung des Suchraumes

Um die Suche effizienter zu gestalten, werden Richtungen in denen zuvor Punkte gefunden worden sind, deren DTW-Distanz über dem Schwellwert lag, nicht weiter betrachtet. Um festzulegen welche Punkte nicht mehr weiter untersucht werden, werden zwei verschiedene Metriken implementiert:

Direkter Vorgänger Diese einfache Metrik berücksichtigt jeweils nur die Klassifikation des, vom Startpunkt aus gesehenen, direkten Vorgängers, um zu entscheiden, ob die DTW-Distanz eines Punktes berechnet werden soll. Die Abbildung 5.4 zeigt dieses Verfahren. Für Eckpunkte wird die Distanz in beiden Koordinaten verringert, für die übrigen Punkte lediglich in einer. Dieses Verfahren hat den Nachteil, dass einzelne Punkte mit einer großen DTW-Distanz den Suchraum sehr stark einschränken können.

?	?	-	?	?
?	×	×	×	-
?		(0,0)		?
?			×	-
?	?	?	?	-

-	?	-	?	-
?	×	×	×	?
?	×	(0,0)		?
?			×	?
?	?	?	?	-

Abbildung 5.4.: Ausgewertete Punkte im zweiten Iterationsschritt anhand des direkten Vorgängers (links) und der Nachbarschaft (rechts). Ein × kennzeichnet Punkte deren Distanz zu groß ist.

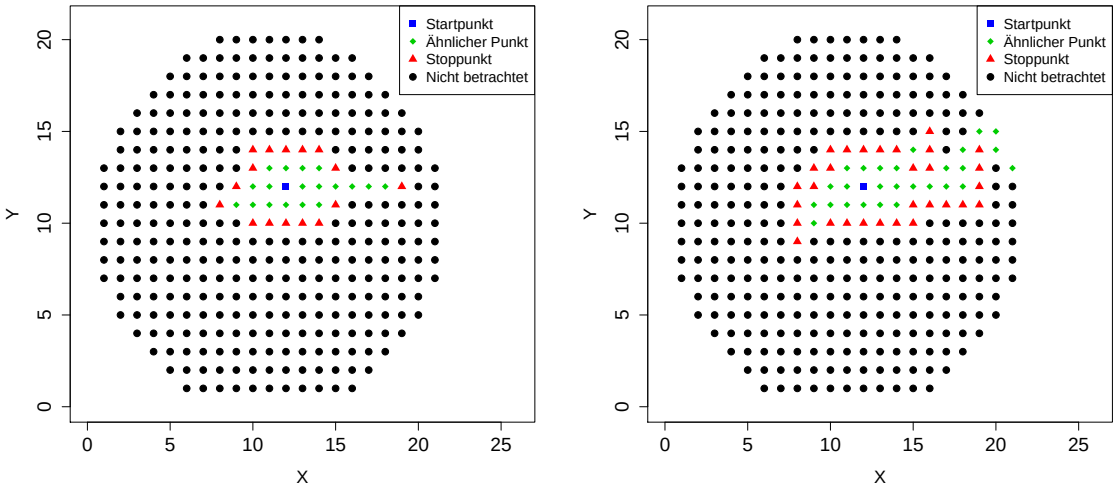


Abbildung 5.5.: Vergleich der direkten Vorgänger- (links) und der Nachbarschafts- (rechts) Suchmetriken auf dem NiCrRe-Wafer mit Schwellwert 1000

Nachbarschaftsmetrik Um diesen Nachteil auszugleichen, betrachtet die zweite implementierte Metrik alle Vorgänger eines Punktes. In diesem Fall wird ein Punkt nur dann nicht ausgewertet, wenn alle drei Vorgänger entweder ein Stopp-Punkt sind oder zuvor nicht ausgewertet wurden. Ausnahmen bilden die Eckpunkte, bei denen wie zuvor nur der direkte Vorgänger betrachtet wird, was daran liegt, dass die restliche Nachbarschaft selbst im noch nicht ausgewerteten Bereich liegt.

Wie zu erwarten ist, bewirkt das Einbeziehen von mehr Vorgängern, dass der erkundete Suchraum größer ist, während die einfache Metrik sehr große Bereiche aufgrund einzelner Punkte nicht weiter betrachtet. Dieser Unterschied ist in Abbildung 5.5 gut zu sehen. Aufgrund der besseren Abdeckung wird im Folgenden nur noch die Nachbarschaftsmetrik betrachtet, jedoch mag es auch Anwendungsszenarien für eine deutlich restriktivere Erkundung geben, wie sie die Methode des direkten Vorgängers ermöglicht. Bei beiden verwendeten Metriken endet die Suche, wenn keine neuen Suchpunkte mehr in einer Iteration zur Verfügung stehen, weil zuvor alle Richtungen blockiert wurden oder alle Punkte untersucht wurden. Für die Ausgabe werden die Punkte entsprechend ihrer Kategorisierung markiert: entweder als ähnlicher Punkt, als Punkt über dem Schwellwert

oder als nicht betrachtet. Für alle untersuchten Punkte wird außerdem die berechnete DTW-Distanz gespeichert.

5.7. Anzahl gefundener Punkte mit maximaler DTW-Distanz

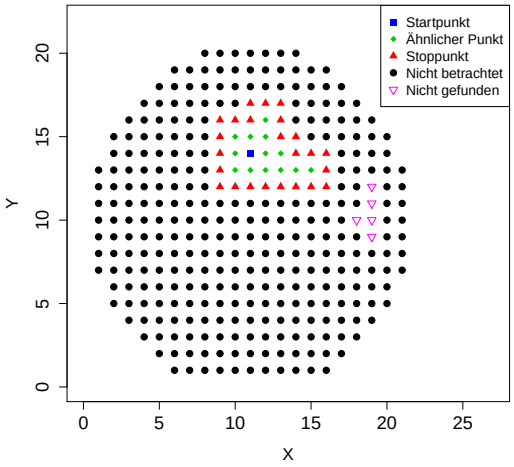
In einem ersten Schritt wird überprüft wie stark die lokalen Abhängigkeiten der Distanzen ausgeprägt sind. Für diesen Zweck werden alle Distanzen zu einem bestimmten Punkt berechnet. Anschließend kann die tatsächliche Anzahl der Punkte mit einer DTW-Distanz kleiner der maximalen Distanz d und die Anzahl der durch den Operator gefundenen Punkte verglichen werden.

Die Abbildung 5.6 zeigt vier Ergebnisse, die verdeutlichen sollen welche Fälle je nach Startpunkt und gewählter maximaler Distanz auftreten können. Die beiden Ergebnisse a) und b) zeigen typische Fälle für eine erfolgreiche Suche; der Großteil der ähnlichen Punkte liegt in einem zusammenhängenden Gebiet um den Startpunkt herum und selbst die nicht gefunden Punkte bilden einen zusammenhängenden Bereich. Außerdem zeigt sich anhand der unterschiedlichen maximalen DTW-Distanzen noch einmal deutlich wie unterschiedlich die Strukturen der XRD-Spektren sind. Das Ergebnis in c) hingegen zeigt einen Fall in dem die Zusammenfassung mittels der DTW-Distanz nicht gut funktioniert. Die ähnlichen Punkte liegen deutlich abseits des Startpunktes und verteilen sich stärker über den gesamten Wafer. Dieses Beispiel ist in sofern auch interessant, als dass es deutlich weniger ähnliche Punkte gibt, obwohl die maximale DTW-Distanz zehnmal größer ist als in Beispiel b). Zuletzt zeigt der Fall d) auf dem TiCoW-Wafer, dass die Verteilung der ähnlichen Punkte auch deutlich weniger kompakt sein kann als in den anderen Beispielen. Außerdem zeigt sich hier eine Schwäche in der Struktur der konzentrischen Suchstrategie, da ähnliche Punkte nicht gefunden werden, obwohl sie direkt an gefundene grenzen.

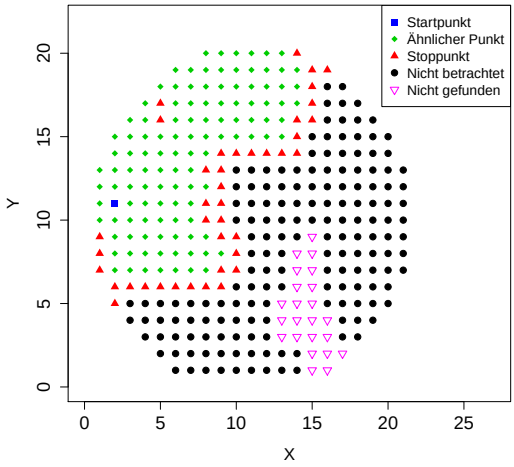
Für einen allgemeinen Überblick werden nun mehrere Startpunkte mit verschiedenen maximalen DTW-Distanzen systematisch getestet. Abbildung 5.7 zeigt die gewählten 40 Startpunkte. Für die maximale Distanz d werden die Werte 500, 1000, 5000, 10.000 untersucht. Aufgrund der durchschnittlich deutlich niedrigeren Distanzen (vgl. Abb. 5.2) wird zusätzlich für den NiCrRe-Datensatz noch der Wert 100 betrachtet. Somit ergeben sich insgesamt 520 Kombinationen.

5.8. Auswertung der Aggregation

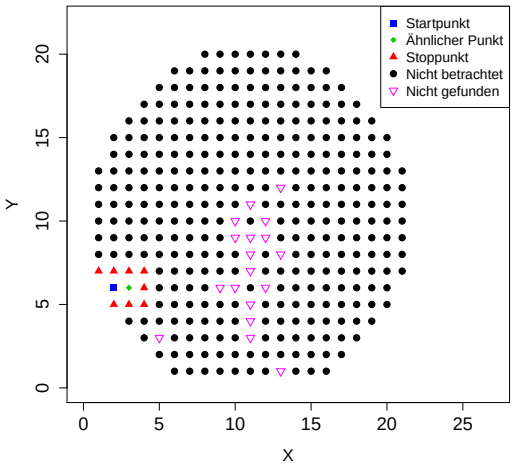
In Tabelle 5.1 sind die durch die DTW-Breitensuche nicht gefunden Messpunkte für die oben erwähnten 520 Durchgänge aufgelistet. Im Durchschnitt werden rund zwei Drittel aller ähnlichen Punkte durch die Breitensuche gefunden. Dabei spielt allerdings die gewählte maximale Distanz eine recht große Rolle. Ist sie zu klein gewählt, können unähnliche Punkte dafür sorgen, dass die Suche zu früh abbricht. Ist die Distanz hingegen verhältnismäßig groß, so werden fast alle Punkte als ähnliche betrachtet und die Suche liefert kaum relevante Informationen. Es erscheint daher sinnvoll den Schwellwert anhand vorheriger Untersuchungen der Spektren auf den jeweiligen Datensatz anzupassen.



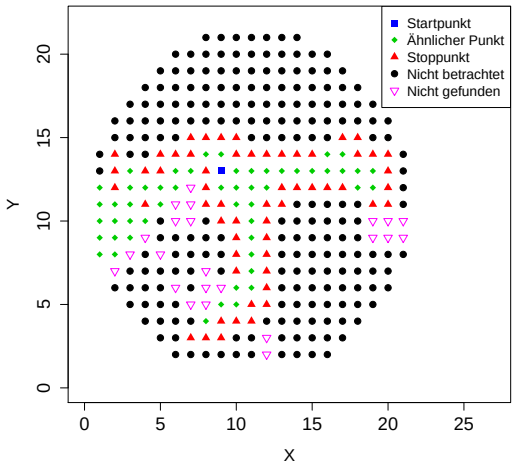
(a) CuNiZn, max. Distanz 5000



(b) NiCrRe, max. Distanz 100



(c) NiCrRe, max Distanz 1000



(d) TiCoW, max Distanz 10.000

Abbildung 5.6.: Beispiel Ergebnisse der DTW-Breitensuche

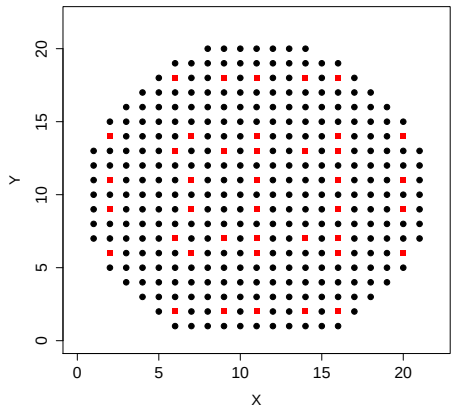


Abbildung 5.7.: Gewählte Startpunkte der DTW-Breitensuche

	Mittlw.	Max.	Relativ	rel. 100	rel. 500	rel. 1000	rel. 5000	rel. 10000
CuNiZn	25,6	221	0,31	-	0,15	0,26	0,43	0,40
NiCrRe	49,5	184	0,30	0,36	0,46	0,43	0,17	0,05
TiCoW	28,2	187	0,36	-	0,06	0,20	0,61	0,58
Gesamt	35,6	221	0,32	-	0,23	0,30	0,40	0,34

Tabelle 5.1.: Übersicht der nicht gefundenen Punkte durch das DTW Clustering

Abgesehen davon zeigen die Ergebnisse, dass die DTW-Distanz im Ansatz als Maß für Ähnlichkeit von Materialproben geeignet scheint. Wäre dies nicht der Fall, so würden sich die, nach diesem Maß, ähnlichen Punkte stärker über den gesamten Wafer verteilen und es würden mit der vorgeschlagenen Suchstrategie deutlich weniger Punkte gefunden.

Eine interessante Fragestellung wäre, ob Gebiete ähnlicher Punkte mit Gebieten bekannter Materialphasen übereinstimmen oder derartige Gebiete vielleicht als bislang unbekannte Phasenübergänge identifiziert werden können. Die gewählte Strategie der Breitensuche weist auch noch einige Schwächen auf und kann sicherlich verfeinert werden, um die Anzahl gefundener Punkte zu erhöhen. In Hinblick auf die aufwendige Bestimmung der XRD-Spektren wäre auch eine Untersuchung des Verhältnis von betrachteten Punkten zu gefundenen Punkten interessant. Ein mögliches Verfahren, das dies berücksichtigt, wird im letzten Abschnitt in Ausblick gestellt.

6. Zusammenhang von XRD-Spektren und physikalischen Größen

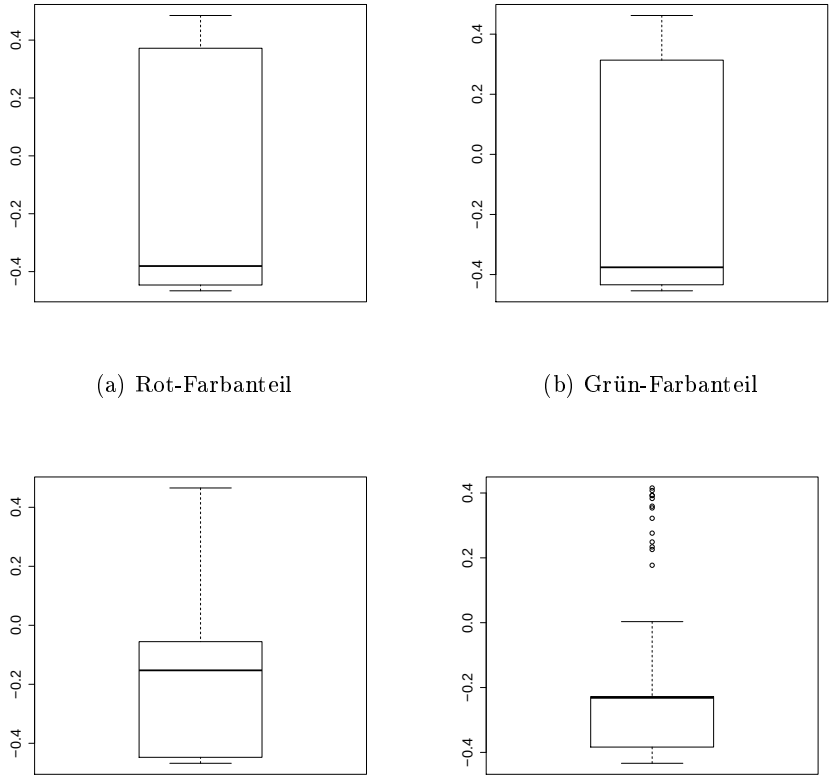
Ein Nachteil der DTW-Distanzen ist, dass sie nur zwischen zwei Referenzpunkt berechnet werden. So wird bei der im vorherigen Abschnitt beschriebenen Breitensuche die Distanz immer bezüglich eines vorgegebenen Startwertes berechnet. Für einen vollständigen Überblick der Ähnlichkeiten aller Punkte zueinander müsste demzufolge die Distanz zwischen allen Kombinationen von Punkten berechnet werden, was deutlich mehr Rechenzeit erfordern würde und außerdem die Komplexität der Daten erhöht, da zu jedem der n Messpunkte $n-1$ weitere Merkmale gespeichert werden müssten. Aus diesem Grund erfolgt nun eine Reduktion der gesamten Spektralreihe auf einige wenige Merkmale, die unabhängig von den Messungen der übrigen Punkte sind. Anschließend wird untersucht, ob sich aus diesen reduzierten Daten Materialeigenschaften ableiten lassen.

6.1. Korrelation zwischen DTW-Distanzen und weiteren Messgrößen

In Hinblick auf weiterführende Untersuchungen wäre es auch von Interesse, ob die DTW-Distanz direkt mit den Materialeigenschaften korreliert ist. Die untersuchte Hypothese ist, dass falls zwei Punkte eine große DTW-Distanz aufweisen, ihre XRD-Spektren sich also stark unterscheiden, sich auch die Materialeigenschaften deutlich unterscheiden. Um dies zu überprüfen, wurde der Korrelationskoeffizient zwischen der DTW-Distanz und den erfassten Materialeigenschaften (Widerstand, Materialkonzentration, RGB) für jeden der zuvor untersuchten Messpunkte berechnet. Für diese Untersuchungen werden nur die Daten der CuNiZn- und den NiCrRe-Datensätze berücksichtigt, da für den TiCoW-Wafer wegen fehlender Annotation der Messwerte eine Zuordnung zwischen XRD-Spektren und Materialeigenschaften nicht möglich ist.

Die stärkste gefundene Korrelation liegt lediglich bei $-0,63$ zwischen der DTW-Distanz und der Rheniumkonzentration im NiCrRe-System. Hierbei ist zu beachten, dass dieser Punkt am Randbereich des Wafers mit der stärksten Rheniumkonzentration liegt und es somit leicht zu erklären ist, dass Punkte mit einer größeren DTW-Distanz auch weiter entfernt auf dem Wafer liegen und somit eine geringere Rheniumkonzentration aufweisen. Betrachtet man die Verteilung der Korrelationswerte der anderen Materialeigenschaften (Abb. 6.1), so zeigt sich auch hier insgesamt eine Tendenz einer leicht negativen Korrelation, wobei es in allen Fällen auch etliche Punkte gibt, bei denen es eine positive Korrelation gibt.

Die DTW-Distanz als alleiniges Merkmal scheint als Erklärungsgröße für die Unterschiede von Materialmischungen nicht auszureichen. Eine naheliegende Vermutung dabei ist, dass durch die Reduktion der ganzen Messreihe auf eine einzelne Größe zu viele Infor-



(a) Rot-Farbanteil

(b) Grün-Farbanteil

(c) Blau-Farbanteil

(d) elektrischer Widerstand

Abbildung 6.1.: Überblick der Korrelationen zwischen DTW-Distanz und Materialeigenschaften

mationen über die genaue Form des XRD-Spektrums verloren gehen. Außerdem hängen die Distanzen vom gewählten Startpunkt ab, was eine Verallgemeinerung zusätzlich erschwert. Aus diesen Ergebnissen lässt sich der folgende Ansatz, den Merkmalsraum der XRD-Spektren kompakter zu erfassen und in diesem Raum nach Zusammenhängen zwischen XRD-Spektren und physikalischen Eigenschaften zu suchen, folgern.

6.2. Regressionsmodelle mithilfe der XRD-Spektren

In Abschnitt 3.1 zeigte sich, dass sich der elektrische Widerstand gut anhand anderer Materialeigenschaften modellieren lässt. Nun soll ein ähnlicher Versuchsaufbau mithilfe der extrahierten Spitzenwerte der XRD-Spektren betrachtet werden. Um zusätzliche Variabilität durch Parameter, wie die Wahl des Kernels einer SVM, zu vermeiden, wird auf die zuvor verwendete ν-SVM [10] verzichtet und eine einfache lineare Regression verwendet. Ein vergleichbares Vorgehen nutzten Dell'Ann et al. [16] für Vorhersagen der Materialkonzentrationen in einem Titan-Nickel-Kupfer-System. Dabei wurden als Größen jedoch die Spitzenwerte aus den Messungen eines Flugzeitmassenspektrometers (ToF-SIMS) [17] verwendet.

Für das allgemeine lineare Modell der Form werden zunächst als Einflussgrößen die vollständigen Informationen der 15 größten Peaks verwendet; damit ist die Anzahl der Regressoren n in dieser Gleichung 86. Als Zielgrößen werden alle verfügbaren Materialeigenschaften verwendet.

Die Beurteilungsgrößen für die linearen Modelle sind in Tabelle 6.1 zusammengefasst. Besonders auffällig ist, dass der elektrische Widerstand sehr schlechte Werte aufweist, weswegen diese beiden Modelle genauer betrachtet werden.

Zunächst wird die Größe der Modelle mittels Variablenselektion reduziert. Bei der zweistufigen Variablenselektion wird zunächst dem Modell jeweils iterativ eine weitere Einflussgröße hinzugefügt bis ein Hinzufügen weiterer Variablen das Modell nicht weiter verbessert. In einem zweiten Schritt wird von einem vollständigen Modell ausgehend in jeder Iteration eine Variable entfernt bis das Abbruchkriterium erfüllt ist. Das beste Modell aus diesen beiden Schritten ist dann das Ergebnis der Variablenselektion. Je nach Bewertungskriterium und Auswahlstrategie kann es hierbei zu unterschiedlichen Modellen kommen.

Das verwendete R-Paket MASS [18, 19] benutzt als Gütekriterium das *Akaike's Information Criterion* (AIC) [20], welches sowohl die Anpassungsgüte, als auch die Komplexität des Modells berücksichtigt. Mit dieser Selektion reduziert sich die Anzahl der Variablen auf 13 (CuNiZn) und 11 (NiCrRe), wobei das Bestimmtheitsmaß sich im Vergleich zum vollständigen Modell nicht wesentlich ändert.

Durch die Einbeziehung von Wechselwirkungen zwischen den Einflussgrößen, beispielsweise dem Produkt der Intensitäten, oder durch nicht lineare Transformationen ließe sich im Einzelnen die Modellgüte sicherlich weiter verbessern. Dies würde jedoch mit einer deutlichen Überanpassung der Modelle einhergehen und dem Ziel einer möglichst einfachen Erklärung der Zusammenhänge zwischen XRD-Spektren und Materialeigenschaften widersprechen.

Positiv zu vermerken ist, dass die Anpassungsgüte der Materialkonzentration für beide ternäre Systeme ähnlich gut ist, wie bei der zitierten Untersuchung der ToF-SIM-Daten

Zielgröße	R^2
Cu	0,64
Ni	0,71
Zn	0,44
R	0,69
G	0,62
B	0,40
Int	0,63
Rho	0,44

(a) CuNiZn-System

Zielgröße	R^2
Ni	0,74
Cr	0,50
Re	0,89
R	0,65
G	0,63
B	0,60
Int	0,64
Wid	0,13

(b) NiCrRe-System

Tabelle 6.1.: Übersicht R^2-Indikator für verschiedene vollständige lineare Modelle

des TiCuNi-Datensatzes. Dort ergab sich für die untersuchten Modelle zur Nickelkonzentration ein adj. R^2 von $0,81$. Lediglich die Zink- und Chromkonzentrationen weisen deutlich niedrigere Ergebnisse auf.

6.3. Korrelation von Materialeigenschaften bei ähnlichen *top-n-peaks*-Merkmalen

Eine weitere Möglichkeit den Zusammenhang zwischen den Spitzenwerten der XRD-Spektren und der Materialeigenschaften zu untersuchen ist zu prüfen, ob Punkte, deren Spektralreihen ähnliche Merkmale aufweisen, auch Ähnlichkeiten in den Materialeigenschaften haben. Die extrahierten Werte der Spektren werden hierfür zu ähnlichen Gruppen zusammengefasst und anschließend wird überprüft, ob diese Ähnlichkeit sich auch in den relevanten Materialeigenschaften widerspiegelt. Um die Ähnlichkeit zu bestimmen, wird ein einfaches *k-means*-Clustering auf die extrahierten Werte angewendet. Es werden verschiedene Werte für die Anzahl k der Cluster ausprobiert; ebenso variiert die Art und Anzahl der extrahierten Merkmale. Für die Untersuchung werden alle $k \in \{2, \ldots, 5\}$ und $n \in \{1, \ldots, 10\}$ betrachtet, sowie alle Kombinationen von zusätzlichen Informationen des *top-n-peaks*-Operators, wodurch sich insgesamt 8960 betrachtete Kombinationen je Datensatz ergeben. Für jeden dieser Datenpunkte wird die Korrelation zwischen dem Widerstand und den anderen Materialeigenschaften berechnet.

Zum Vergleich zeigt die Tabelle 6.2 zunächst die Korrelationskoeffizienten zwischen dem elektrischen Widerstand und den übrigen Materialeigenschaften an. Die sehr ausgeprägten Korrelationen zwischen dem Widerstand bezüglich Nickel und Kupfer lassen sich damit erklären, dass Kupfer ein sehr guter elektrischer Leiter ist und demzufolge ein hoher Kupferanteil oft mit einem niedrigen Widerstand einhergeht. Die übrigen Messgrößen weisen hingegen keine sehr ausgeprägten Korrelationen auf.

Nun kann betrachtet werden, ob sich für bestimmte Partitionierungen durch Clustering bessere Korrelationen ergeben. Die gewonnen Daten müssen aufgrund einiger Extremwerte zunächst bereinigt werden. Problematisch sind solche Fälle, in denen der Clustering-Algorithmus einem Cluster nur sehr wenige Werte zuweist, wodurch die Korrelation für die Werte stark verzerrt wird. Es gibt Fälle in denen einem Cluster nur zwei Messpunk-

	Widerstand
Cu	-0,59
Ni	0,70
Zn	-0,30
R	-0,63
G	-0,47
B	-0,03
Int	-0,50

(a) CuNiZn-Wafer

	Widerstand
Ni	0,02
Cr	0,15
Re	-0,17
R	0,26
G	0,27
B	0,25
Int	0,27

(b) NiCrRe-Wafer

Tabelle 6.2.: Korrelationskoeffizienten für den elektrischen Widerstand aller Messpunkte

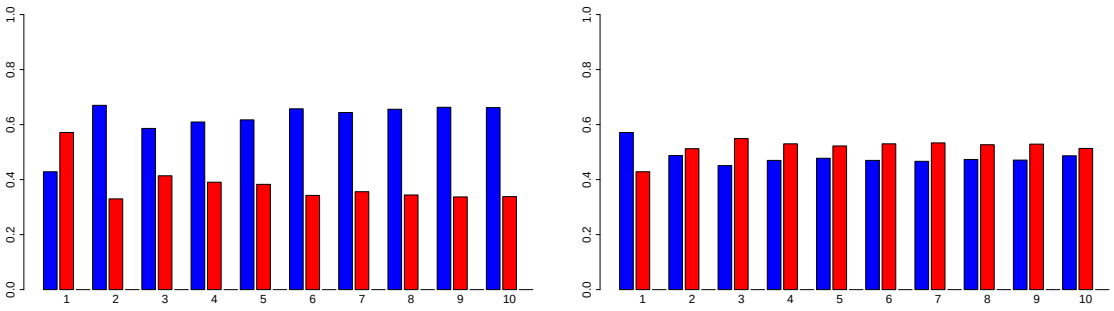
	Widerstand
Cu	-0,62
Ni	0,74
Zn	-0,38
R	-0,67
G	-0,55
B	-0,15
Int	-0,56

(a) CuNiZn-Wafer

	Widerstand
Ni	0,25
Cr	0,29
Re	-0,42
R	0,13
G	0,03
B	0,11
Int	0,06

(b) NiCrRe-Wafer

Tabelle 6.3.: Mittlere Korrelationskoeffizienten für den elektrischen Widerstand aller Messpunkte. Es wird zunächst über die jeweiligen Cluster gemittelt und dann der Durchschnitt für alle Beispiele berechnet



(a) CuNiZn-Wafer (b) NiCrRe-Wafer

Abbildung 6.2.: Verhältnis der gefundenen Cluster mit einem durchschnittlich besseren (blau) oder schlechteren (rot) Korrelationskoeffizienten, in Bezug auf die Anzahl der betrachteten Spitzenwerte $(1 \ldots 10)$

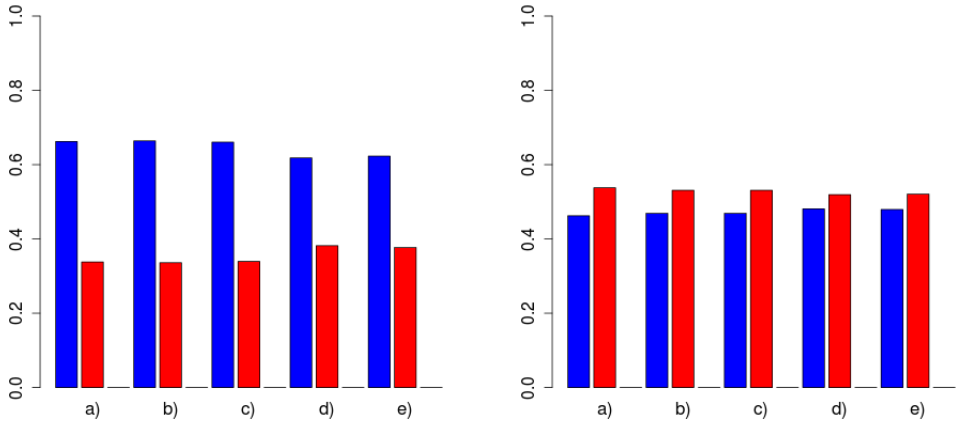
te zugewiesen werden und für diese dann ein Korrelationskoeffizient von ± 1 berechnet wird. Diese offensichtlichen Ausreißer werden entfernt, wodurch 348 (CuNiZn) bzw. 136 (NiCrRe) Punkte wegfallen.

In Tabelle 6.3 werden zunächst die Korrelationen für alle Parameterkombinationen des *top-n-peaks*-Operators betrachtet. Um die durchschnittliche Korrelation zu bestimmen, wird zunächst für jede Parameterkombination die Korrelation für einzelne Cluster gemittelt und nochmals der Mittelwert über alle Beispiele berechnet. Für alle Materialkonzentrationen verbessert sich dadurch die Korrelation sichtbar; für die Farbwerte sind die Ergebnisse teilweise besser, teilweise schlechter. Am auffälligsten ist die Verbesserung des Korrelationskoeffizienten für die Nickel-Konzentration im NiCrRe-System die von $0,02$ auf $0,25$ ansteigt. Die beobachteten Verbesserungen liegen vermutlich darin begründet, dass die gemessenen XRD-Spektren von der Materialzusammensetzung abhängig sind und diese Abhängigkeit sich bereits in wenigen Merkmalen widerspiegelt.

Für die Beurteilung, ob ein ausgewähltes Merkmal eine mögliche Auswirkung auf die Güte der Korrelation hat, wird für jede Partitionierung betrachtet, ob die mittlere Korrelation innerhalb der gefundenen Cluster besser ist als die Werte aus Tabelle 6.2. Würde ein Merkmal die Korrelation positiv beeinflussen, so wäre zu vermuten, dass wenn dieses Merkmal mit einfließt, sich auch die beobachteten Korrelationen verbessern würden.

Zunächst wird betrachtet, ob die Anzahl Spitzenwerte die Korrelation beeinflusst (Abb. 6.2). Hier ist jedoch kein Einfluss zu erkennen; das Verhältnis zwischen besseren und schlechteren Korrelationskoeffizienten entspricht bei beiden Datensätzen in etwa dem Gesamtverhältnis aller Beobachtungen. Lediglich wenn nur der größte Ausschlag betrachtet wird, unterscheiden sich die Anteile, jedoch mit unterschiedlichen Resultaten.

Der Einfluss der weiteren Merkmale auf den Korrelationskoeffizienten ist in Abbildung 6.3 zu sehen. Jedoch ist auch hier kein sichtlicher Unterschied zwischen den Merkmalen zu erkennen.



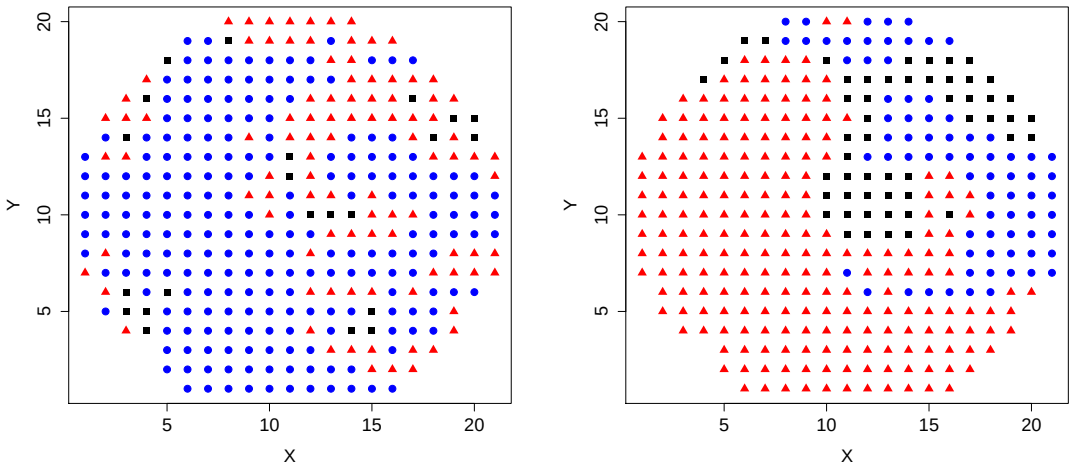
(a) CuNiZn-Wafer (b) NiCrRe-Wafer

Abbildung 6.3.: Verhältnis der gefundenen Cluster mit einem durchschnittlich besseren (blau) oder schlechteren (rot) Korrelationskoeffizienten, für die *top-n*-Parameter: a) Intensität, b) Distanz zum Spitzenwert, c) Distanz zum Vorgänger, d) Differenz zum Spitzenwert, e) Differenz zum Vorgänger

6.4. Räumlicher Zusammenhang zwischen Clustering und Materialzusammensetzung

Ebenfalls von Interesse könnte die räumliche Lage von ähnlichen Spektralreihen auf einem Wafer sein. Die Abbildung 6.4 zeigt als Beispiel wie unterschiedlich sich die gefundenen Cluster auf dem Wafer verteilen, wenn unterschiedliche Merkmale der XRD-Spektren berücksichtigt werden.

Die Schwierigkeit hierbei ist es ein geeignetes Beurteilungskriterium für die gefundenen Cluster zu definieren. Die Frage ist, welche Information aus den XRD-Spektren sind für ein bestimmtes Kriterium relevant und welche sorgen nur für zusätzliches Rauschen, welches die gesuchte Information überlagert. So sind wahrscheinlich große zusammenhängende Gebiete positiv zu beurteilen, wie in Abbildung 6.4a zu sehen, insbesondere, wenn sie sich mit der Verteilung von bekannten Materialphasen überdecken. Andererseits könnten die zusätzlichen Informationen, die zu einer stärkeren Streuung der Cluster führen (vgl. Abb. 6.4b), Hinweise für Phasenübergänge beinhalten. Für eine gründliche Beurteilung des Nutzens dieser Aufbereitung der Daten ist es notwendig ein für die Art der Untersuchung geeignetes Dichtekriterium zu definieren, welches die Streuung der einzelnen Cluster bewertet und dabei den diskreten Merkmalsraum der Waferkoordinaten berücksichtigt.



(a) Top-5-peaks

(b) Top-5-peaks und sämtliche weiteren Informationen

Abbildung 6.4.: Räumliche Anordnung der gefundenen Cluster ($k = 3$) für unterschiedliche Merkmale der XRD-Spektren (CuNiZn).

7. Zusammenfassung und Ausblick

Die gezeigten Ergebnisse bestätigen den Eindruck, dass mit Hilfe von Datenanalyse und Data-Mining Zusammenhänge zwischen verschiedenen Materialeigenschaften aufgezeigt werden können, die ansonsten in der großen Menge von Informationen verborgen bleiben. Die aufgeworfenen Fragen konnten keinesfalls erschöpfend bearbeitet werden. Die Resultate geben aber eine Richtungen an, in die weiter geforscht werden könnte, oder zeigen Methoden auf, die für kommende Untersuchungen von Nutze sein können.

Für die Kategorisierung der XRD-Spektren anhand ihrer DTW-Distanz konnte gezeigt werden, dass es einen erkennbarer Zusammenhang zwischen der Verteilung der Materialzusammensetzung (gemessenen anhand der Verteilung der Messpunkte auf dem Wafer) und der Ähnlichkeit der Spektralreihen gibt. Offene Fragen dabei bleiben die Wahl der geeigneten maximalen DTW-Distanz zwischen ähnlichen Punkten und die des Startpunktes. Dennoch könnte mithilfe dieses Verfahrens beispielsweise die Anzahl von Datenbankabfragen nach ähnlichen Mustern deutlich reduziert werden, indem zunächst in der Probe ähnliche Bereiche zusammengefasst werden und dann nur Referenzwerte aus den Datenbanken betrachtet werden, deren DTW-Distanz nicht zu groß ist. Auch die Einbeziehung der DTW-Distanzen zur Verdeutlichung von Phasenübergängen oder Ausreißern könnte sowohl in der automatischen Verarbeitung, als auch für die visuelle Analyse durch geeignete bildgebende Verfahren genutzt werden. Eine einfache Möglichkeit wäre die DTW-Distanz zu allen anderen Punkten für einen ausgewählten Punkt darzustellen, um so bestimmte Eigenheiten der Verteilung manuell zu erkennen.

Hingegen scheint die direkte Modellierung von Materialeigenschaften allein anhand der DTW-Distanz eine weniger geeignete Methode zu sein. Hier ist der Informationsverlust durch die Reduktion der Messreihe auf einen Wert offensichtlich zu groß.

Der Ansatz die XRD-Spektren zunächst auf wenige, möglichst relevante, Größen zu reduzieren scheint vielversprechend. Die Güte der aus den extrahierten Werten erstellten linearen Modelle bekräftigt die Vermutung, dass sich ein allgemeiner Zusammenhang aus den XRD-Spektren und verschiedenen Materialeigenschaften herleiten lässt. Hier könnte die Verwendung von nicht linearen Methoden, wie der Regressions-SVM, die gezeigten Ergebnisse weiter verbessern.

Die Partitionierung mittels Clustering über die Merkmale der Spitzenwerte hingegen scheint keinen Einfluss auf die Korrelation der Materialeigenschaften zu haben. Zwar steigt die durchschnittliche Korrelation innerhalb der Cluster leicht an, der Effekt kann jedoch auf keines der Merkmale zurückgeführt werden. Vielleicht führt hier die Wahl eines anderen Clustering-Algorithmus oder eine modifizierte Fragestellung zu besseren Ergebnissen.

Ein weiterer möglicher Einflussfaktor könnte sein, dass Ausschläge des Grundmaterials, in diesem Fall also das Silizium der Waferscheibe, die gemessenen XRD-Spektren stark

dominieren und somit die Messung verzerren. In diesem Fall könnte eine Filterung der in Frage kommenden Winkel hilfreich sein.

Insgesamt zeigen alle bisher untersuchten Methoden tendenziell gute Ergebnisse. Allerdings lassen sich alleine anhand der bisher gemessenen Werte noch keine zufriedenstellenden Vorhersagemodelle erstellen. Für allgemeinere Aussagen wäre es insbesondere notwendig die getesteten Verfahren auf mehr Materialproben anzuwenden um die Robustheit der Verfahren zu testen. Ebenso von Interesse, wäre zu überprüfen ob bekannte Anomalien oder Materialphasen identifiziert werden können, dies würde entsprechend annotierte Datensätze erfordern.

Einer der größten Nachteile des Verfahrens bleibt jedoch bestehen, nämlich, dass die XRD-Analyse sehr zeitaufwendig ist. Interessant bezüglich einer schnellen Erfassung und Auswertung neuer Materialproben wäre es, wenn nicht mehr die gesamte Waferscheibe geröntgt werden müsste, sondern aufgrund der hier aufgezeigten Beziehungen eine gezielte Stichprobe ausreichen würde, um ein möglichst genaues Bild über die Verteilung der XRD-Spektren zu modellieren.

Ein mögliches Szenario wäre, mit Hilfe der einfacher zu erfassenden Messungen, wie elektrischer Widerstand oder RGB-Werte, die DTW-Distanz benachbarter Punkte zu prognostizieren und dies zu nutzen, um für ausgewählte Punkte eine vollständige Messung der XRD-Spektren durchzuführen. Die Modelle könnten dann mit jeder weiteren Messung kontinuierlich verbessert werden, um auf die Besonderheiten einer unbekannten Materialprobe einzugehen. Letztendlich könnte ein solches Modell sinnvolle Aussagen über die XRD-Spektren des gesamten Wafers machen, ohne dass eine vollständige und zeitaufwendige Messung nötig wird.

Literaturverzeichnis

[1] Wolfgang Weißbach and Uwe Bleyer. *Werkstoffkunde und Werkstoffprüfung*. Vieweg, 1988.

[2] Krishna Rajan. Combinatorial materials sciences: Experimental strategies for accelerated knowledge discovery. *Annu. Rev. Mater. Res.*, 38:299–322, 2008.

[3] C.J. Long, J. Hattrick-Simpers, M. Murakami, R.C. Srivastava, I. Takeuchi, V.L. Karen, and X. Li. Rapid structural mapping of ternary metallic alloy systems using the combinatorial approach and cluster analysis. *Rev Sci Instrum*, 78(7):072217, 2007.

[4] I. Takeuchi, C. J. Long, O. O. Famodu, M. Murakami, J. Hattrick-Simpers, G. W. Rubloff, M. Stukowski, and K. Rajan. Data management and visualization of x-ray diffraction spectra from thin film ternary composition spreads. *Review of Scientific Instruments*, 76(6):062223, 2005.

[5] M. Müller. *Information Retrieval for Music and Motion*, chapter 4: Dynamic Time Warping, pages 69–83. Springer, 2007.

[6] Eamonn J Keogh and Michael J Pazzani. Scaling up dynamic time warping for datamining applications. In *Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 285–289. ACM, 2000.

[7] Yasushi Sakurai, Masatoshi Yoshikawa, and Christos Faloutsos. Ftw: fast similarity search under the time warping distance. In *Proceedings of the twenty-fourth ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems*, pages 326–337. ACM, 2005.

[8] Joachim Hartung, Bärbel Elpelt, and Karl-Heinz Klösener. *Statistik: Lehr-und Handbuch der angewandten Statistik; mit zahlreichen, vollständig durchgerechneten Beispielen*. Oldenbourg Verlag, 2005.

[9] Martina Mittlböck, Michael Schemper, et al. Explained variation for logistic regression. *Statistics in medicine*, 15(19):1987–1997, 1996.

[10] Bernhard Schölkopf, Alex J. Smola, Robert C. Williamson, and Peter L. Bartlett. New support vector algorithms. *Neural Comput.*, 12(5):1207–1245, May 2000.

[11] Chih-Chung Chang and Chih-Jen Lin. LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, 2:27:1–27:27, 2011.

[12] T. Hastie, R. Tibshirani, and J. H. Friedman. *The Elements of Statistical Learning*. Springer, july 2003.

[13] J. Ross Quinlan. *C4.5: programs for machine learning*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1993.

[14] Leonard A Breslow and David W Aha. Simplifying decision trees: A survey. *The Knowledge Engineering Review*, 12(01):1–40, 1997.

[15] Ingo Mierswa, Michael Wurst, Ralf Klinkenberg, Martin Scholz, and Timm Euler. Yale: Rapid prototyping for complex data mining tasks. In Lyle Ungar, Mark Craven, Dimitrios Gunopulos, and Tina Eliassi-Rad, editors, *KDD '06: Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 935–940, New York, NY, USA, August 2006. ACM.

[16] Rossana Dell'Anna, Paolo Lazzeri, Roberto Canteri, Christian J. Long, Jason Hattrick-Simpers, Ichiro Takeuchi, and Mariano Anderle. Data analysis in combinatorial experiments: Applying supervised principal component technique to investigate the relationship between tof-sims spectra and the composition distribution of ternary metallic alloy thin films. *QSAR & Combinatorial Science*, 27(2):171–178, 2008.

[17] John C Vickerman and David Briggs. *ToF-SIMS: surface analysis by mass spectrometry*. IM publications Chichester, 2001.

[18] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2013.

[19] W. N. Venables and B. D. Ripley. *Modern Applied Statistics with S*. Springer, New York, fourth edition, 2002. ISBN 0-387-95457-0.

[20] Hamparsum Bozdogan. Model selection and akaike's information criterion (aic): The general theory and its analytical extensions. *Psychometrika*, 52(3):345–370, 1987.