



MAGISTERARBEIT / MASTER'S THESIS

Titel der Magisterarbeit / Title of the Master's Thesis

„Angewandte Statistische Methoden im Condition Monitoring und
Predictive Maintenance für Industrieanlagen“

verfasst von / submitted by

Peter Holzer Bakk.rer.soc.oec.

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of

Magister der Sozial- und Wirtschaftswissenschaften (Mag. rer. soc. oec.)

Magister of social and economic sciences (Mag. rer. soc. oec.)

Wien, 2017 / Vienna 2017

Studienkennzahl lt. Studienblatt /
degree programme code as it appears on
the student record sheet:

A 066 951

Studienrichtung lt. Studienblatt /
degree programme as it appears on
the student record sheet:

Magisterstudium Statistik UG2002

Betreut von / Supervisor:

Univ.-Prof. Mag. Dr. Walter Gutjahr

INHALTSVERZEICHNIS

1	Einführung.....	4
2	Fallstudie.....	7
2.1	Datenbeschreibung	7
2.2	Datenaufbereitung	8
2.2.1	Festlegung der minimal notwendigen Aufzeichnungsrate	10
2.3	Deskriptive Statistiken, Datenübersicht	10
2.4	Abgeleitete Signale (Features).....	14
2.5	Fehlende Signale.....	16
2.6	Festlegung des Events und des Remaining Useful Life Countdowns	16
2.7	Festlegung der Remaining Useful Life (RUL)	16
2.8	Gewichtung des relevanten Zeitraums.....	21
2.9	Modellvergleich Allgemein	21
3	Methodik.....	22
3.1	Das simple Modell	26
3.1.1	Ergebnisse des Simplen Modells	27
3.2	Das multiple lineare Regressionsmodell	28
	Implementierung	28
	Variablenselektion	29
3.2.1	Beschreibung und Ergebnisse der linearen Modelle	30
3.2.2	Selektion des Modelles mit der besten Vorhersagegüte.....	38
3.3	Das Regression Tree Modell	43
	Implementierung	43
	Variablenselektion	43
3.3.1	Die Regression Tree Modelle im Detail	43
3.3.2	Selektion des Modelles mit der besten Vorhersagegüte.....	46
3.4	Das Random Forest Regression Modell.....	49
	Variablenselektion	49
3.4.1	Die Random Forest Regression Modelle im Detail	49
3.4.2	Selektion des Modelles mit der besten Vorhersagegüte.....	50
3.5	General Additive Models (GAM)	55
	Implementierung	55

Variablenselektion	56
3.5.1 Die GAM Regression Modelle im Detail	56
3.5.2 Selektion des Modelles mit der besten Vorhersagegüte	57
3.6 Das Support Vector Regression Modell	62
3.6.1 Die SVR Regression Modelle im Detail	66
3.6.2 Selektion des Modelles mit der besten Vorhersagegüte	67
3.7 Vergleich der Ergebnisse der Regressionsmethoden	71
3.7.1 Vergleich der Mean Event Time Deviations: p-Werte des Wilcoxon Rangsummentests	72
3.7.2 Vergleich RUL-Time-Deviation: p-Werte des Wilcoxon Rangsummentests	73
3.7.3 Vergleich RUL Time Deviation auf positive RULs: p-Werte des Wilcoxon Rangsummentests	73
3.8 Bestimmung des größtmöglichen Zeitintervalls ohne Informationsverlust für die Modelle	74
4 Betriebswirtschaftliche Simulation	76
5 Generalisierbarkeit der Modelle	80
5.1.1 Generalisierbarkeit der Modelle	80
6 Ausblick	84
6.1 Rückschluss der Variablenwichtigkeiten auf die Fehlerrsache	84
6.2 Neuronale Netzwerke	84
7 Conclusio	85
8 Appendix	87
9 Literaturverzeichnis	88
Abstract	91

1 EINFÜHRUNG

“I was not predicting the future, I was trying to prevent it.”

— Ray Bradbury (1920 – 2012)

In dieser Arbeit werden statistische und Data Mining Methoden erarbeitet und gegenübergestellt, die für Industrieanlagen Aufschluss geben sollen, in welchem Zustand sich eine Maschine befindet und wie hoch die Wahrscheinlichkeiten sind, dass gewisse Bauteile oder Baugruppen einen Bruch erleiden. Des Weiteren soll aufgrund der Maschinenzustandsdaten und der Sensorik eine Schätzung für eine fehlerfreie Laufzeit der Anlagen ermittelt werden. Dieses Vorgehen von Nutzung der Maschinendaten für eine Vorausschauende Wartungsstrategie wird in der Industrie *Predictive Maintenance* (Mobley R. , 1990) genannt, und ist ein großer Hoffnungsträger für eine effizientere Nutzung von Ressourcen. Durch strategisch geplante Wartungen können Materialien und Arbeitskräfte gezielt eingesetzt werden, und somit Kosten für Personal, Ersatzmaterial und Reparaturen eingespart werden.

Überblick Instandhaltungskonzepte:

- **Reparatur nach Ausfall:**

Wird vor allem bei Maschinen durchgeführt für die keine oder nur wenig Zustandsmessungen vorhanden sind, bzw. wenn deren Ausfall keine hohen Kosten verursacht. (Beispiel: Tausch von Glühbirne oder Druckerpatrone)

- **Vorbeugende Wartung:**

Anwendung findet sich Häufig bei Verschleißteilen, deren Überwachung schwierig oder teuer ist. Ebenso werden vom Hersteller oft Wartungsintervalle festgelegt die für eine Produktgarantie einzuhalten sind. (Beispiel: Tausch von Zahnriemen oder Ölwechsel beim Auto, Tausch von Stanzwerkzeugen, Reinigung von Ablagerungen durch Produktion etc.)

- **Zustandsorientierte Wartung oder Condition Based Maintenance (CBM):**

Durch Messungen werden Bauteile überwacht, und nur bei Auftreten von Mängeln entsprechende Wartungstätigkeiten durchgeführt. Diese Messungen können sowohl automatisiert durch die Sensorik oder von einem Techniker durchgeführt werden.

- **Predictive Maintenance (PM) oder Vorausschauende Instandhaltung:**

Darunter versteht man im Allgemeinen ein technisches System, das den Zustand und den Wartungsbedarf von Industrieanlagen oder Maschinen voraussagen kann (Jardine, 2006). Durch zunehmende technische Aufrüstung und Modernisierung solcher Anlagen werden immer mehr und sensiblere Sensoren eingesetzt. Diese liefern immer mehr Information über den aktuellen Maschinenzustand, und geben im Idealfall sogar über zukünftiges Fehlverhalten Aufschluss. PM steht für das modernste Konzept der Instandhaltung und ist ein Schlüsselement in der sogenannten Industrie 4.0.

Related Works

Die einfachsten Vorausschauenden Instandhaltungssysteme beschäftigen sich meist nur mit einer Lebenszeitanalyse. Diese werden oft als Badewannenverteilung oder mit einem Betriebsstundenzähler modelliert. Diese Methoden können aber keine tatsächliche Abnutzung oder Zustandsänderungen der Anlage abschätzen, da sie nie die aktuellen Messwerte betrachten, sondern nur die historisch erfasste Lebenszeit einer Komponente. Die moderne PM basiert auf den tatsächlich gemessenen Daten einer Anlage und kann im besten Fall in Echtzeit nicht nur den aktuellen Maschinenzustand bewerten, sondern auch den zukünftigen Wartungs- oder Reparaturaufwand und wann dieser notwendig wird abschätzen. Einen guten Überblick über angewandte PM Methoden bietet Okoh et al. (Okoh, Roy, Mehnen, & Redding, 2014) oder Hashemian & Bean (Hashemian & Wendell, 2011). Xiao-Sheng et al. bietet eine gute Metastudie über Anwendungsfälle im industriellen Umfeld (Si, Wang, Hu, & Zhou, 2011). Weiters trifft man häufig auf Vibrationsanalysen, bei denen gezielt zusätzliche Sensorik verbaut wird, welche durch Schwingungsmessungen einen generellen Gesundheitszustand der Anlage schätzen soll (Mobley R. K., 2002). Ausgehend davon wird die Restlebenszeit extrapoliert und die Verlässlichkeit mit einem Konfidenzintervall angegeben. Jedoch sind diese Methoden meist nur im speziellen Umfeld anwendbar, da insbesondere auf die Maschineneigenheiten eingegangen werden muss. Die folgende Fallstudie soll eine generelle Herangehensweise zu einer PM Lösung bieten. Dies soll dadurch erreicht werden, dass Ereignisse definiert werden, welche aus den Daten selbst erzeugt werden können. Dadurch wird es möglich den Zeitreihen der Sensorik eine Restlebenszeit zuzuordnen (Khanh, Fouladirad, Barros, Levrat, & Benoit, 2013). Diese kann wiederum, ganz allgemein, von jeder Regressionsmethode als zu erklärende Variable verwendet werden. Die folgende Methodik und Simulation baut auf einer anderen Fallstudie von mir und Dipl.-Ing. Dr. Georg Pözlbauer auf (Holzer, Peter; Pözlbauer Georg, 2014), bei der ebenfalls eine Restlebenszeit definiert wurde und anhand der Messwerte zu jedem Zeitpunkt im Echtzeitbetrieb eine Prognose zur verbleibenden Zeit bis zum definierten Ereignis erstellt wurde.

Aufbau der Fallstudie

Im Kapitel 2 wird die Fallstudie beschrieben und die Datenaufbereitung erklärt. Dazu gehört das Filtern auf die relevanten Datenzyklen und die Erstellung von unabhängigen Zeitreihen.

Des Weiteren wird die zu erklärende Variable „Remaining Useful Life“ (Xiongzi, Jinsong, Diyin, & Yingxun, 2011), die Restlebenszeit eingeführt, welche im Echtzeitbetrieb anhand der Messwerte geschätzt werden soll. Im Kapitel 3 werden fünf Regressionsmethoden vorgestellt und auf den aufbereiteten Datensätzen der Fallstudie angewandt. Dabei werden jene Modelle selektiert, welche die besten Prognosegüten erreichen. Diese werden dann im Kapitel 4 in einer Betriebswirtschaftlichen Simulation verwendet, um die Auswirkungen einer Anwendung von Predictive Maintenance Methoden und der aktuellen Wartungsmethode vergleichbar zu machen. Abschließend wird im Kapitel 5 die Generalisierbarkeit der PM Modelle diskutiert, indem die Prognosegüten auf dem Datensatz einer baugleichen Anlage geschätzt werden.

2 FALLSTUDIE

Eine Stranggießanlage produziert Stahl aus geschmolzenen Rohstoffen. Beim aktiven Abkühlen des Stahls mit Wasser entsteht Wärme und Feuchtigkeit, welche durch zwei Dampfabsauganlagen abgefördert wird. Die Überwachung und die vorausschauende Instandhaltung dieser Absauganlage, mittels prädiktiver statistischer Methoden, ist das Hauptaugenmerk dieser Fallstudie. Es stellt sich die Frage ob durch automatisiertes Auswerten der Maschinendaten die Wartungszyklen optimiert werden können. Dazu werden die Wartungsstrategien, die sich durch die Handlungsempfehlungen der erstellten statistischen Modelle ergeben, mit der sich im Einsatz befindlichen Wartungsstrategie verglichen.

Die Dampfabsauganlagen sorgen für eine kontinuierliche Absaugung von heißem Dampf, der bei der Kühlung der Stranggießanlage im Stahlgusswerk anfällt. Durch den Dampf lagert sich nach und nach Schmutz aus der Abluft an den Ventilatorschaufeln ab. Ein Vibrationsmessgerät das an der Außenseite der Laufschaufeln angebracht ist überwacht den Zustand der Anlage. Die am Lager der Antriebswelle gemessenen Schwingungen werden in elektrische Signale umgewandelt, welche kontinuierlich aufgezeichnet werden. Diese Werte werden über einen kurzen Zeitraum (weniger als eine Sekunde) vom Schwingungsmessgerät selbst gemittelt. Durch Vibrationsmessung kann ein Zusammenhang zwischen Verschmutzungsgrad, Lagerschäden und Anlagenzustand geschaffen werden. Die Absauganlage Ost und West sind Baugleich und haben dieselben Überwachungselemente.

2.1 DATENBESCHREIBUNG

In der Fallstudie werden Daten aus dem Zeitraum von Juni 2012 bis August 2015 ausgewertet. Insgesamt werden 19 Basissignale als mögliche Einflussvariablen behandelt. 9 Signale pro Anlage und ein Signal (Gießgeschwindigkeit) stammt aus der vorgelagerten Anlage. Die folgende Tabelle zeigt eine Übersicht der vorhandenen Messwerte.

Bezeichnung Anlage Ost	Einheit / Typ	Beschreibung
Drehmoment Ost Drehmoment West	Newton [N] Kontinuierlich gemessen	Drehmomentmessung des Motors, gemessen mit Frequenzumrichter.
Leistung Ost Leistung West	Watt [W] Kontinuierlich gemessen	Leistungsmessung des Motors, gemessen mit Frequenzumrichter.
Motorstrom Ost Motorstrom West	Ampere [A] Kontinuierlich gemessen	Stromaufnahme des Motors.
Gerätetemperatur Ost Gerätetemperatur West	Celsius [°C] Kontinuierlich gemessen	Temperatur des Frequenzumrichters.
Motortemperatur Ost Motortemperatur West	Celsius [°C] Kontinuierlich gemessen	Temperatur des Motors.
Drehzahl Ost Drehzahl West	Umdrehungen pro Minute [U/min] Kontinuierlich gemessen	Umdrehungen des Ventilators.
Vibration Ost Vibration West	Schwinggeschwindigkeit [mm/s] Kontinuierlich gemessen	Schwingungsmessung am Lager der Antriebswelle.
Gießgeschwindigkeit	Geschwindigkeit Kontinuierlich gemessen für Ost und West	Gießgeschwindigkeit der Stranggussanlage.
Temperatur Ost Temperatur West	Celsius [°C] Kontinuierlich gemessen	Umgebungs-temperatur der Anlage
Zustand Ost Zustand West	Kontinuierlich gemessen als Faktor	Betriebsart der Anlage

Tabelle 1: Übersicht der Messwerte.

2.2 DATENAUFBEREITUNG

Um einen effizienten Einsatz des Speicherplatzes zu garantieren erfolgt die Datenaufzeichnung zu vorgegebenen diskreten Zeitpunkten oder mittels Datenfilter. Es kommen 3 Aufzeichnungstypen zum Einsatz:

- a) Ohne Filter. Alle Messwerte werden ohne Filterung aufgezeichnet. Entweder erfolgt dies zu fix vorgegebenen Intervallen oder dann, wenn der Messwert auftritt. Es wird ein Zeitstempel mit dem entsprechenden Messwert generiert.
- b) Totbandfilter. Es werden nur Messwerte aufgezeichnet, die vom vorangehenden aufgezeichneten Messwert um eine vorgegebene Toleranzschwelle abweichen. (Speichereffizient bei gering variierenden Signalen)

- c) Vektorbandfilter. Es werden alle Messwerte weggelassen, die auf der Verbindungsstrecke zwischen zwei aufgezeichneten Messwerten die vorgegebene Toleranzschwelle nicht überschreiten. (Speichereffizient bei linear verlaufenden Signalen)

Zu jedem aufgezeichneten Messwert wird ein Zeitstempel im Millisekundenbereich erstellt. Durch diese Aufzeichnungsstrategie werden die Signale asynchron aufgezeichnet und müssen zur statistischen Auswertung diskretisiert werden. Die Entwickler und Techniker der Dampfabsauganlage setzen die Toleranzschwelle aufgrund ihrer Erfahrungswerte. Dadurch ist davon auszugehen, dass keine relevante Information durch die Filterung verloren geht. Die nachfolgende Skizze zeigt wie die Daten diskretisiert werden. Dabei werden die zuletzt ankommenden Messwerte auf den nachfolgenden Intervallzeitpunkt übertragen. Wenn kein neuer Messwert innerhalb des angegebenen Intervalls ankommt, dann wird der zuletzt gemessene Wert auf den aktuellen Zeitpunkt übertragen. So ist garantiert, dass zu jedem diskreten Zeitpunkt auch alle Messwerte von allen Variablen zur Auswertung zu Verfügung stehen. Diese Vorgangsweise wird auch als Last Observation Carried Forward (LOCF) bezeichnet.

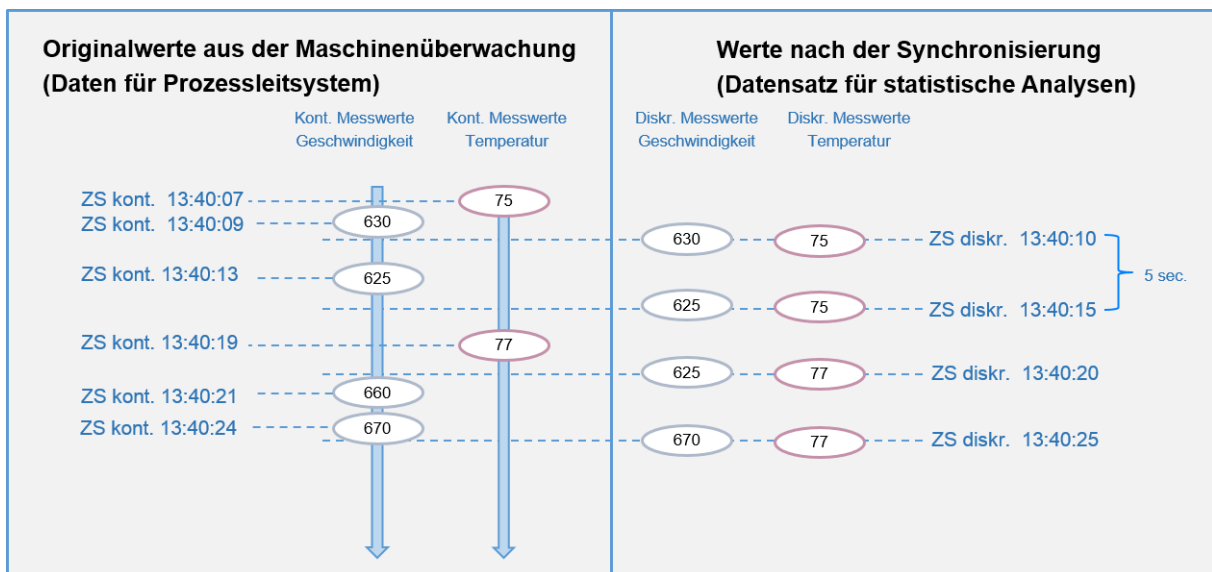


Abbildung 1: Beispiel Skizze der Zeitstrahlen von ankommenden Daten und Diskretisierung in 5 Sekunden Beobachtungsintervalle.

In dieser Arbeit werden nur Daten nach der Diskretisierung betrachtet. Grundsätzlich gibt es noch andere Strategien um die Messwerte mit diskreten Zeitstempel zu versehen. Es kann zum Beispiel zwischen den ankommenden Messwerten eine lineare Interpolation vorgenommen werden, oder im Falle eines fehlenden zukünftigen Messwertes könnte dieser als lineare Extrapolation der beiden zuletzt beobachteten Werte berechnet werden. Für die Dampfabsauganlage gibt es keine hochfrequente Datenaufzeichnung (die hochfrequente

Schwingungsmessung wird vom Sensor selbst über einen Zeitraum gemittelt und als Vibrationsmesswert ausgegeben), weshalb man durchgehend die Signale mit den zuletzt ankommenden Messwerten diskretisieren kann, ohne Informationsverluste einbüßen zu müssen.

2.2.1 Festlegung der minimal notwendigen Aufzeichnungsrate

Die Aufzeichnungsrate der Signale liegt zwischen mehreren Sekunden und mehreren Stunden, je nachdem wie stark und schnell sich diese verändern können. Zunächst wurden die Daten auf 5 Sekunden Intervalle diskretisiert, da nur in wenigen Fällen mehrere Messungen für ein Signal innerhalb dieses Intervalls stattfanden. Die Anlagenverschmutzung sowie etwaige Verschleißprobleme der Maschine bauen sich über einem langen Zeitraum auf. Es ist davon auszugehen, dass die Beobachtungen in unmittelbarer zeitlicher Nähe sehr stark miteinander korrelieren. Dadurch sollte es ohne Informationsverlust möglich sein die Daten in einem gewissen Zeitraum zusammenzufassen und als einzelne Beobachtung anzusehen. Dies wurde gemacht, indem das gleitende Minimum für ein 1-Minuten Intervall errechnet wurde. Dabei wurde darauf geachtet, dass nur die vergangenen 12 Beobachtungen (12x 5 Sekunden) gemittelt werden. Somit können die Daten ebenso im Echtzeitbetrieb aufbereitet und ausgewertet werden. Durch das gleitende Minimum erhält man eine geglättete Version jeder Variable, auf denen dann jede 12te Beobachtung ausgewählt wird und den reduzierten Datensatz definiert. Ob es dadurch tatsächlich zu keinem relevanten Informationsverlust kommt, wurde nach Erstellung der Modelle geprüft, indem Modelle mit demselben Setting und auf demselben Datensatz nur mit gröberen und feinerem Auswahlintervall, erstellt und miteinander verglichen wurden. Dies wird im Kapitel 3.8 detailliert ausgeführt. Die Beobachtungen werden mit dem gleitenden Minimum zusammengefasst und nicht, wie oft üblich mit dem gleitenden Mittelwert, da bei den Vibrationsmessungen gelegentlich für kurze Zeit Spitzenwerte auftreten können, die durch externe Einflüsse oder Resonanzbereiche verursacht werden. Diese Werte können zu Verzerrungen in der Zustandsüberwachung und in der Eventdefinition führen. (siehe Kapitel 2.6)

2.3 DESKRIPTIVE STATISTIKEN, DATENÜBERSICHT

Die auf 1-Minuten Intervalle diskretisierten Daten ergeben eine Beobachtungsanzahl von 1.669.469 Datenpunkten. Alle Variablen bis auf „Zustand“ sind kontinuierliche Werte. Die Variable „Zustand“ gibt die Betriebsart der Anlage an. Die relevanten Produktionsbetriebe sind 50 (Gießen) und 60, 61, 62, 63 (Ausfahren, Durchbruch, Ausfördern, Niederkühlen). Hier muss die Dampfabsauganlage laufen um den heißen Dampf vom produzierten Strangguss abzufördern.

Betriebsart	Beschreibung
10	WARTUNG
20	KALTSTRANG EINFAHREN TOPFEEDING (TF)
21	KALTSTRANG EINFAHREN BOTTOMFEEDING (BF)
30	TIPPEN
40	RÜSTEN
50	GIEßEN
60	AUSFAHREN
61	DURCHBRUCH
62	AUSFÖRDERN
63	NIEDERKÜHLEN
80	GIEßSPALTMESSEN

Tabelle 2: Beschreibung der Betriebsarten der Dampfabsauganlage

	TIME	Drehmoment OST	Drehzahl OST	Leistung OST
nbr.val	1669469	1668954	1656142	1668954
nbr.null	0	7754	8439	6836
nbr.na	0	515	13327	515
min	2012-06-01 02:05:00	-99.69	-39.2	-85.14
max	2015-08-04 10:30:00	63.88	739.8	84.97
range	1159 Tage	163.6	779	170.1
sum	-	50554640	921282247	55930583
median	-	30.96	591.9	32.87
mean	-	30.29	556.3	33.51
SE.mean	-	0.01	0.12	0.01
CI.mean.0.95	-	0.02	0.23	0.03
var	-	161.2	22879	363.3
std.dev	-	12.7	151.3	19.06
coef.var	-	0.42	0.27	0.57

Tabelle 3: Deskriptive Statistiken der kontinuierlichen Daten. Nbr.val steht für die Anzahl an beobachteten Werten. Nbr.null steht für die Anzahl an fehlenden Werten. Nbr.na ist die Anzahl an „NA“ Werten. SE.mean ist die Standardabweichung des Mittelwertes. CI.mean.0.95 ist das 95% Konfidenzintervall des Mittelwertes. Coef.var ist die durch den Mittelwert korrigierte Standardabweichung.

	Motortemp OST	Vibration OST	Temp OST	Motorstrom OST	Gerätetemp OST
nbr.val	1668774	1668954	1668947	1668954	1665979
nbr.null	942	4	4	8439	4
nbr.na	695	515	522	515	3490
min	0	0	0	0	0
max	62.05	9.94	81.99	145.5	91.4
range	62.05	9.94	81.99	145.5	91.4
sum	55403579	2600467	67499348	175693078	80556210
median	32.62	1.15	38.74	108.7	48.12
Mean	33.2	1.56	40.44	105.3	48.35
SE.mean	0.01	0	0.01	0.02	0
CI.mean.0.95	0.01	0	0.03	0.04	0
Var	66.07	1.32	307.2	536	8.05
std.dev	8.13	1.15	17.53	23.15	2.84
coef.var	0.24	0.74	0.43	0.22	0.06

Tabelle 4: Deskriptive Statistiken der kontinuierlichen Daten (Fortsetzung).

	Betriebszeit in Prozent
Zustand 10	13,01
Zustand 20	2,31
Zustand 21	0,01
Zustand 30	0,85
Zustand 40	1,86
Zustand 50	74,90
Zustand 60	6,46
Zustand 61	0,01
Zustand 62	0,02
Zustand 63	0,06
Zustand 80	0,47
Zustand NA's	0,03

Tabelle 5: Betriebsarten der Anlage und die Betriebszeiten in Prozent von der Gesamtlaufzeit.

Zeitpunkte der Anlagenreinigungen Ost und West:

Wenn keine schwerwiegenden Probleme mit der Anlage auftreten, wird diese im Intervall von ca. sechs Monaten gereinigt. Rückwirkend bedeutete diese simple Wartungsstrategie,

dass die Absauganlage oft in einem schädlichen Betriebszustand betrieben werden musste. Ein vollständiges Abschalten der Anlage war kurzfristig nicht möglich, da sonst die gesamte Stranggießanlage außer Betrieb gesetzt hätte werden müssen, was wiederum zu kostspielig gewesen wäre. Um trotzdem noch Stahl zu produzieren, wurden die Dampfabsauganlagen, welche akut zu hohe Vibrationswerte aufwiesen mit reduzierter Last betrieben. Dadurch konnte aber nicht mehr das gesamte Dampfvolument abgesaugt werden, und es kann (verursacht durch die Übermenge an heißem Dampf) zu Betriebsstörungen und Schädigungen an der gesamten Anlage kommen. Des Weiteren könnte die Qualität des Stahls unter der unzureichenden Kühlung leiden. Die Reinigungsarbeiten werden immer an beiden Anlagen zugleich durchgeführt, um die Zeiten des Stillstands möglichst gering zu halten. Die folgenden Abbildungen zeigen wie sich die Vibrationsmessungen an den beiden Dampfabsaugungen über die Zeit aufbauen, bis sie wieder gereinigt werden.

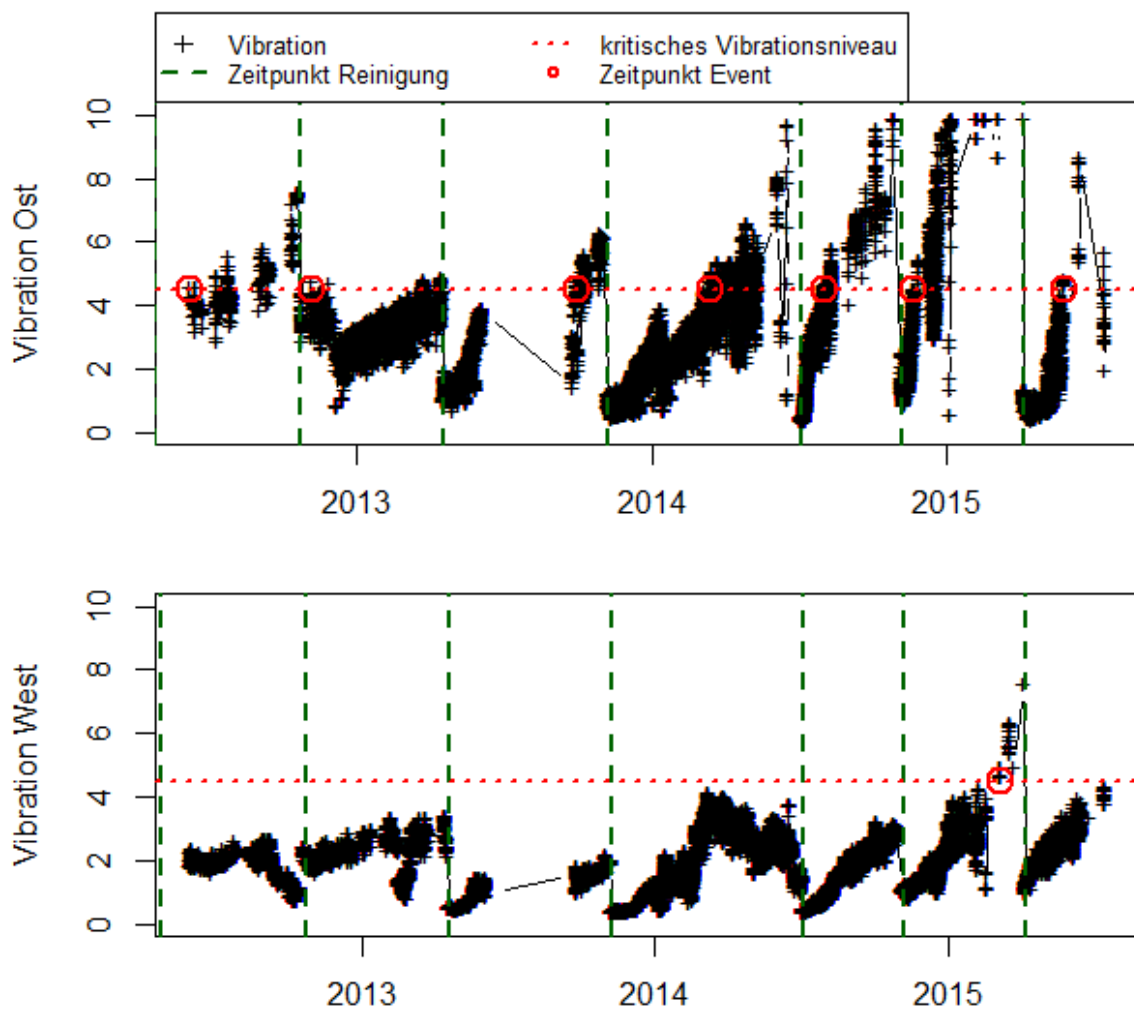


Abbildung 2: Die Vibrationsmessungen der Anlage Ost und West. Die Zeitpunkte der Reinigung sind als grüne Vertikalen eingezeichnet. Die roten Kreise sind die Zeitpunkte der Events, also die Zeitpunkte an denen die Anlage erstmalig in einem schädlichen Zustand betrieben wird (siehe Kapitel 2.6). Die schwarzen Linien werden eingezeichnet, wenn die Maschine nicht im Automatikbetrieb läuft und daher die Daten nicht für die Modellbildung

verwendet werden können (siehe Kapitel 2.7). Die Anlage West läuft zumeist mit niedrigeren Vibrationswerten als die Anlage Ost und verursacht nur ein einziges Event.

2.4 ABGELEITETE SIGNALE (FEATURES)

Bei zeitabhängigen Daten wie sie in dieser Fallstudie vorliegen, kann viel Information von den Modellen erst verwendet werden, wenn entsprechende Ableitungen der Signale, in weiterer Folge Features genannt, berechnet werden (Widodo, et al., 2009). Oft wird der Zeithorizont für eine Ableitung auf mehrere Ableitungstypen zugleich angewendet, und im Modell auf Signifikanz bewertet. Dies hat den Vorteil, dass gleiche Zeithorizonte Signalübergreifend von den Technikern leichter interpretiert werden können. Für die Techniker der Anlage ist es von Bedeutung in welchem Zeitraum man das Signal betrachtet. So können zum Beispiel durch eine Analyse der *Steepest Ascend/Descend* der Drehzahl, Rückschlüsse gezogen werden wann die Anlage zu schnell hochgefahren bzw. abgebremst wird und daher zu hohe Belastungsspitzen auftreten. Solche Fehlsteuerungen der Anlage können sehr gut mit dieser Art von abgeleiteten Signalwerten festgestellt und angepasst werden. Jedoch gilt auch, je weniger Signale im Modell verwendet werden, desto einfacher ist es generalisierbare Regeln zu interpretieren. Deshalb wird in der Fallstudie darauf geachtet, nicht zu viele verschiedene Ableitungen zu generieren und in den Modellen mitaufzunehmen. Features beziehen sich auf eine Historie und beinhalten mehr als einen Datenpunkt, da sie eine Zeitreihe modellieren. Bei der Datenaufbereitung wird immer so vorgegangen, dass es im Livebetrieb möglich ist, dieselbe Aufbereitungsmethode anzuwenden, denn dann können die Modelle eine Prognose auch in Echtzeit liefern.

Aus oben genannten Gründen wird bei der Erstellung der abgeleiteten Signale wie folgt vorgegangen. Ab dem Zeitpunkt, an dem die Anlage in den Betriebsmodus 60 kommt, werden die Daten in einem Datenblock zusammengefasst, bis der Betriebsmodus 60 wieder verlassen wird. Auf diesen Datenblock werden dann die Ableitungen berechnet. Im Median beträgt die Dauer dieser Datenblöcke 6 Minuten und beinhalten daher 6 Datenpunkte. Insgesamt gibt es 1874 solche Datenblöcke. 79 Blöcke davon dauern weniger als zwei Minuten an. Auf diesen können keine Ableitungen berechnet werden, da zumindest zwei Datenpunkte notwendig sind. Diese Datenblöcke müssen bei Modellen, welche die abgeleiteten Signale als Einflussvariablen verwenden, ausgeschlossen werden.

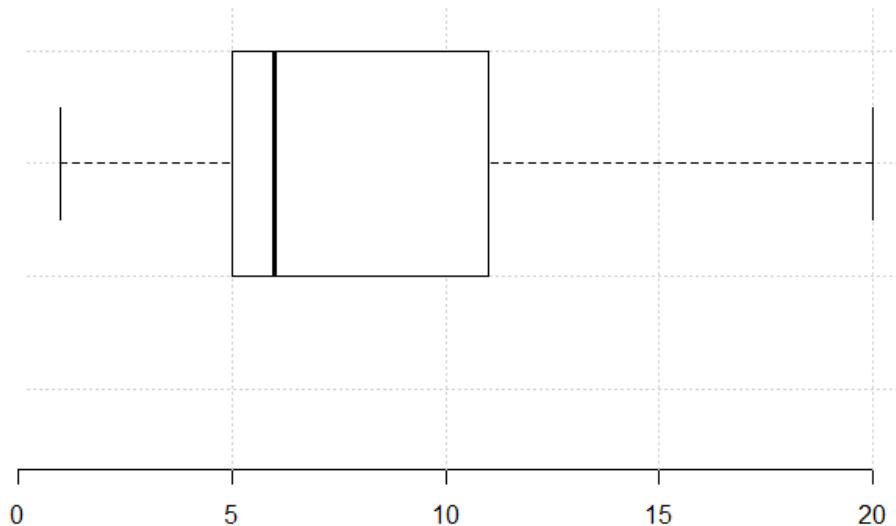


Abbildung 3: Boxplot der Dauer in der sich die Anlage durchgehend im Betriebsmodus 60 befindet. Im Median läuft die Anlage 6 Minuten in diesem Modus. Der Mittelwert der Dauer beträgt 7,77.

Signal	Bezeichnung im Modell	Beschreibung
Steilster Anstieg im Datenblock	<i>Steepest Ascend 1</i> bzw. <i>sAsc1</i>	Auf dem Signal werden die Differenzen von einer Minute zur nächsten berechnet. Das Maximum dieser Werte (in diesem Datenblock) definiert den steilsten Anstieg innerhalb einer Minute. (Maximum des 1 Minuten Lags)
Steilster Abfall im Datenblock	<i>Steepest Descend 1</i> bzw. <i>sDesc1</i>	Wird wie oben berechnet, jedoch das Minimum.
Standardabweichung	<i>sd</i>	Die Standardabweichung des Signales innerhalb des Datenblocks.
Kurtosis	<i>kurtosis</i>	Die Wölbung des Signales innerhalb des Datenblocks
Schiefte	<i>skewness</i>	Die Schiefe des Signales innerhalb des Datenblocks

Tabelle 6: Übersicht der abgeleiteten Signale (Features).

Ein weiteres wichtiges Signal ist die *Laufzeit seit der letzten Reinigung*. Dieses Signal wird für jene Modelle ausgeschlossen, welche nur auf einem Reinigungsintervall trainiert werden, da

sonst mittels dieses Signales ein perfekter Fit entsteht. Dadurch würden Signale aus dem Modell eliminiert und das Modell für Prognosen somit unbrauchbar.

2.5 FEHLENDE SIGNALE

Durch die erfolgte diskretisierende Datenaufbereitung, bei welcher fehlende Beobachtungen mit den zuletzt bekannten Beobachtungen ersetzt werden (LOCF), kommt es nur zu Beginn eines Datensatzes zu Datenlücken. Nämlich nur dann, wenn noch gar keine Beobachtung vorliegt. In der Fallstudie sind dies sehr wenige Stunden am Beginn der Datenaufzeichnung und daher bei einem Zeithorizont von 1159 Tagen vernachlässigbar. Aus diesem Grund wurden die Datenzeitpunkte, an denen noch nicht für alle Variablen ein Wert vorhanden war, ausgeschlossen.

2.6 FESTLEGUNG DES EVENTS UND DES REMAINING USEFUL LIFE COUNTDOWNS

Mittels Schwingungsmessung an den Antriebswellen des Motors Ost und West wird der Verschmutzungsgrad der Laufschaufeln des Ventilators indirekt geschätzt. Wird ein Vibrationssignal von 4,5mm/s über mehrere Sekunden hinweg gemessen, löst der Vibrationssensor ein Warnsignal aus. Dieses Warnsignal kann an der Anlagensteuerung quittiert werden. Allerdings können bei längerem Betrieb im hohen Vibrationsbereich erhebliche Schäden an Antrieb und Lagern entstehen. Deshalb wird in diesem Fall häufig die Drehzahl etwas gesenkt, womit sich auch die Vibrationswerte senken lassen. Nach einer Phase mit geringerer Drehzahl wird versucht die Anlage wieder in Vollbetrieb zu nehmen. Lässt sich dies aber nicht mehr gewährleisten ohne zu hohe Vibrationswerte zu erhalten, dann wird ein Wartungsauftrag für die Anlage ausgelöst. Die Wartungsarbeiten werden an eine externe Firma vergeben. Die Reaktionszeiten dafür reichen von mehreren Tagen bis zu mehreren Wochen. Die Reparaturarbeiten sind meist Reinigungstätigkeiten der Laufschaufeln, jedoch werden dafür ein bis zwei Schichten benötigt. Wenn dies ungeplant geschieht, sind die Ausfallkosten sehr hoch, da die Stahlproduktion unterbrochen wird und die Stranggießlinie abgeschaltet werden muss. In dieser Fallstudie soll festgestellt werden, ob und wann ein frühzeitiges Erkennen des schädlichen Zustandes möglich ist. Dadurch soll der Wartungsauftrag vorzeitig und automatisiert ausgelöst werden können. Somit würde die Planbarkeit für Material und Personal erhöht werden, und die Wahrscheinlichkeit für ein technisches Gebrechen durch den Betrieb auf schädlichen Vibrationsniveau reduziert werden.

Das Event welches es vorherzusagen gilt, wird definiert als der Zeitpunkt an dem ein Wartungsauftrag ausgelöst werden soll.

2.7 FESTLEGUNG DER REMAINING USEFUL LIFE (RUL)

Der Zeitpunkt für ein Event wird anhand der Daten durch ein Regelwerk festgelegt. Dieses wird wie folgt definiert:

1. Die Anlage befindet sich im Betriebsmodus 60 (Ausfahren). Damit wird gewährleistet, dass vergleichbare Rahmenbedingungen für die Anlage gelten, wie zum Beispiel die Fördermenge an Dampf.
2. Die Anlage läuft mit einer Drehzahl von zumindest 700 Umdrehungen pro Minute. Damit wird gewährleistet, dass sich die Anlage nicht in einem Resonanzbereich befindet an dem hohe Vibrationswerte ein normales Verhalten wären. Des Weiteren ist bei einer hohen Drehzahl sichergestellt, dass die Anlage unter Last läuft, was wiederum vergleichbare Rahmenbedingungen für die Vibrationsmessung schafft.
3. Die Vibrationsmessung überschreitet zum ersten Mal, nach einer Reinigung der Schaufelräder, das Niveau von 4,5mm/s. Dieser Schwellwert wurde von den Anlagenbetreibern festgelegt und gilt als jenes Niveau das bei dauerhaften Überschreiten zu schwerwiegenderen Anlagenstörungen führen kann. Das erstmalige Überschreiten dieses Wertes soll festlegen, dass sich die Anlage ab diesem Zeitpunkt in einem verunreinigten Zustand befindet. Das bedeutet also auch, dass es innerhalb eines Reinigungsintervalls nur ein einziges Event geben kann.

Bemerkung: Die Vibrationsmessungen wurden wie oben beschrieben, über einen Zeitraum von einer Minute gesammelt. Deren Minimum wird dem Datensatz übergeben. Damit sollen kurzfristige Vibrationsspitzen die sich durch Verunreinigungen die sofort wieder abfallen oder durch externe Schocks ergeben können ausgeschlossen werden.

Sind alle 3 Bedingungen erfüllt, wird dieser Zeitpunkt als Eventzeitpunkt definiert. Ausgehend vom Zeitpunkt an dem die Anlage von einer Reinigung kommend wieder in Betrieb genommen wird, bis zu dem Eventzeitpunkt wird ein Countdown berechnet und als „Remaining Useful Life“ (RUL) definiert. Die zeitliche Distanz von einer Beobachtung zur nächsten ist eine Betriebsminute (siehe Kapitel 2.2.1). Die RUL wird so festgelegt, dass die erste Beobachtung im Reinigungsintervall die zeitliche Differenz zwischen Zeitpunkt der Reinigung und Zeitpunkt des Events ist. Die nächste Beobachtung erhält den selben Wert abzüglich der einen Minute die von einer zur nächsten Beobachtung vergehen. Dies wird so fortgeführt bis es zu einer neuen Reinigung kommt.

Beispiel:

Am 18. April 2013 geht die Anlage nach dem zweiten Reinigungsintervall wieder in Betrieb. Am 29. September 2013 tritt das Event auf. Rückwirkend von diesem Zeitpunkt kann die RUL gesetzt werden. Die Laufzeit des Vollbetriebs bis zum Event ergibt den ersten Wert des Countdowns. Danach wird immer eine Minute runtergezählt, wenn sich die Anlage im Vollbetrieb befindet. Nach dem Eventzeitpunkt wird der Countdown weitergezählt, bis es zur Reinigung der Anlage kommt. Danach beginnt ein neuer Datenblock.

Zeitstempel	RUL	Vollbetrieb	Bemerkung
2013-04-18 22:00:00	1131	Ja	Reinigungsarbeiten des 2. Intervalls abgeschlossen
2013-04-18 22:01:00	1130	Ja	
...	...	Ja	
2013-04-18 22:11:00	1120	Ja	
2013-04-18 22:12:00	1120	Nein	Wenn die Anlage nicht im Vollbetrieb läuft, wird nicht runtergezählt
...	1120	Nein	
2013-04-19 07:02:00	1119	Ja	
2013-04-19 07:03:00	1118	Ja	
...	...	Ja	
2013-04-19 07:20:00	1101	Ja	
2013-04-19 07:21:00	1101	Nein	
...	
2013-09-29 21:55:00	1	Ja	
2013-09-29 21:56:00	0	Ja	Zeitpunkt des Events
2013-09-29 21:57:00	-1	Ja	
...	
2013-11-02 01:20:00	-341	Ja	Reinigungsarbeiten des 3. Intervalls werden begonnen

Tabelle 7: Die RUL ist ein Countdown der in 1-Minuten Inkrementen runtergezählt wird. Beginnend mit dem Zeitpunkt einer Reinigung wird für jede Minute im Vollbetrieb ein Wert abgezogen, bis es zu einer neuen Reinigung kommt. Die RUL Werte nach dem Event sind negativ und markieren den Zeitraum an dem die Anlage nicht mehr in Bestimmungsgemäßen Zustand betrieben wird.

Definition der Remaining Useful Life

Indizes: $m \in M = \{1, \dots, 7\}$ *Reinigungen*

$j \in J = \{1, \dots, 7\}$ *Events*

Parameter: t_{R_m} *Zeitpunkt der m-ten Reinigung*

$t_{R_{m+1}} = t_{R_m} - 1$	eine Minute nach dem Zeitpunkt der m-ten Reinigung
t_{E_j}	Zeitpunkt des j-ten Events
$L_{t_{E_j}}$	RUL zum Zeitpunkt des j-ten Events (wird gleich 0 gesetzt)
$L_{t_{R_m}}$	Remaining Useful Life zum Zeitpunkt der m-ten Reinigung
\hat{L}_t	geschätzte Remaining Useful Life zum Zeitpunkt t
T_m	Menge der Beobachtungszeitpunkte im Intervall m
n_{T_m}	Anzahl an Beobachtungen im Intervall m : $t \in T_m$

Der Vektor der RUL für den ersten Datenblock wird wie folgt indiziert.

$$L_{t_{R_1}} = t_{E_1} - t_{R_1} \quad (1)$$

$$L_{t_{R_1+1}} = t_{E_1} - t_{R_1+1} = t_{E_1} - t_{R_1} - 1 \quad (2)$$

...

$$L_{t_{E_1}} = t_{E_1} - t_{R_1} - (t_{E_1} - t_{R_1}) = 0 \quad (3)$$

$$L_{t_{E_1+1}} = t_{E_1} - t_{R_1} - (t_{E_1} - t_{R_1}) - 1 = -1 \quad (4)$$

...

$$L_{t_{R_2-1}} = t_{E_1} - t_{R_2-1} = t_{E_1} - t_{R_2} - 1 \quad (5)$$

Der erste Wert des RUL Vektors (1) ist die Differenz vom Zeitpunkt des Events und dem Zeitpunkt der Reinigung in Minuten. Der zweite Eintrag (2) ergibt sich aus dem ersten Wert abzüglich einer Minute. Dies wird so fortgeführt, bis zum Zeitpunkt der nächsten Reinigung. Dadurch sind alle RUL Werte vor einem Event positiv, und danach negativ. Gibt es zwischen zwei Reinigungen kein Event, so kann mit dieser Struktur keine RUL erstellt werden. In diesem Fall werden die Daten nicht zur Modellerstellung verwendet.

Block Nr	Beginn des Zeitblocks	Eventzeitpunkt Ost	Zeit von Beginn bis zum Event Ost [TAGE]	Eventzeitpunkt West	Zeit von Beginn bis zum Event West
1	2012-04-24	2012-06-09	46	-	-
2	2012-10-23	2013-11-05	13	-	-
3	2013-04-18	2013-09-29	164	-	-
4	2013-11-07	2014-03-14	127	-	-
5	2014-07-04	2014-08-02	29	-	-
6	2014-11-06	2014-11-22	16	2015-03-07	121
7	2015-04-07	2015-05-27	50	-	-

Tabelle 8 - Reinigungen: Die Tabelle zeigt die Zeitpunkte der Reinigungen welche auf der Ost und Westanlage zeitgleich durchgeführt werden und die Zeitpunkte der Events. Bei der Ost Anlage gibt es in jedem Reinigungsintervall ein Event – ab diesem Zeitpunkt läuft die Anlage in einem schädlichen Betriebszustand bis sie gereinigt wird. Die West Anlage läuft nur im sechsten Reinigungsintervall im schädlichen Zustand.

Die Remaining Useful Life wird so definiert, dass nur dann runtergezählt wird, wenn die Anlage im Vollbetrieb läuft. Dies ist der Fall, wenn sich die Anlage im Betriebszustand 60 (Ausfahren) befindet, während die Drehzahl über 700U/min gehalten wird. Dieser Zustand wird auch als **Verschleißmodus** bezeichnet. Für die Modellbildung werden nur jene Daten verwendet, bei denen sich die Anlage im Verschleißmodus befindet. Die restlichen Daten werden aus den prädiktiven Analysen ausgeschlossen, da diese zu keiner Verminderung der RUL führen. Eine Voranalyse der Daten hat gezeigt, dass die Vibrationswerte auf dem gleichen Niveau bleiben, wenn sich die Anlage im Ruhezustand befindet, oder nur gering belastet wird. Durch diese Datenreduktion, ergibt sich ein Datensatz mit 14554 Beobachtungen auf 7 Zeitblöcken für die Anlage Ost, und 2944 Beobachtungen auf einem Zeitblock für die Westanlage.

Bemerkung: Im ersten Reinigungsintervall vergehen 46 Tage vom Beginn bis zum Event, jedoch wird die Anlage nur einmal im Vollbetrieb gefahren. Nämlich zwei Stunden vor dem Event. Wegen dieser kurzen Zeit im Verschleißmodus, wird der Datenblock des ersten Reinigungsintervalls aus der Modellbewertung ausgeschlossen.

2.8 GEWICHTUNG DES RELEVANTEN ZEITRAUMS

Für den Maschinenbetreiber ist es meist nicht von großem Interesse eine exakte RUL zu kennen, wenn diese noch sehr hoch geschätzt wird. Je näher der Zeitpunkt rückt, an dem eine Wartung durchzuführen ist, umso genauer sollten die Prognosen werden. Um den Fokus der Modelle auf diesen relevanten Zeitraum zu legen, sollten Abweichungen der Schätzungen auch höher bestraft werden je näher das Event rückt. Dies wird in der Modellerstellung berücksichtigt, indem die Beobachtungen je nach zeitlicher Distanz vom Eventzeitpunkt gewichtet werden. Die Methoden zur Gewichtung müssen je nach Regressionsmethode unterschiedlich implementiert werden, und sind in der Methodenbeschreibung der einzelnen Modelle beschrieben.

2.9 MODELLVERGLEICH ALLGEMEIN

Der gesamte Datensatz wird in Datenblöcke unterteilt, deren Zeitspannen sich jeweils vom Zeitpunkt einer Reinigung bis zur nächstfolgenden Reinigung erstrecken. Innerhalb eines Datenblocks bilden die Messwerte der Einflussvariablen eine hochgradig abhängige Zeitreihe. Es wird angenommen, dass die Messwerte der Einflussvariablen zwischen den Datenblöcken unabhängig sind. Das heißt, dass der erste gemessene Wert einer beliebigen Variable nach der Reinigung unabhängig ist von allen Messwerten derselben Variable aus einem anderem Reinigungsintervall (Datenblock). Wegen der Abhängigkeiten innerhalb eines Datenblockes, werden für sämtliche Validierungsverfahren die Trainings-, Test- und Validierungssets immer blockweise ausgewählt. Ein Testset darf also keine Datenzeile aus einem im Trainingsset verwendeten Datenblock beinhalten und umgekehrt. Dementsprechend darf auch im Validierungsset keine Datenzeile verwendet werden, deren Datenblock schon für das Trainings- oder Testset verwendet wurde.

Formale Definition eines Datenblocks:

1. Die Zeitspanne eines Datenblockes wird vom Zeitpunkt einer Reinigung bis zum Zeitpunkt der nächsten Reinigung festgelegt.
2. Innerhalb der Zeitspanne muss ein Event auftreten. Tritt keines auf, dann kann der Datensatz dieses Reinigungsintervalls nicht für eine Modellbildung verwendet werden, da keine RUL festgelegt werden kann.

Da an der Anlage West nur ein einziges Event auftritt, kann zwar ein Modell auf dem entsprechenden Datenblock trainiert, aber nicht getestet oder validiert werden. Deshalb werden die Daten aus der Anlage West zunächst aus den Analysen ausgeschlossen. Im Kapitel 5 wird darauf eingegangen ob und in welchem Ausmaß man die anhand der Ost-Datenblöcke erstellten Modelle für eine Vorhersage an der West-Anlage in Betracht ziehen kann. Wenn dies mit einer zufriedenstellenden Güte gelinge, dann könnte man die erstellten Modelle generalisierbar für baugleiche Dampfabsauganlagen verwenden.

3 METHODIK

Zur Analyse der Fallstudie werden Lineare Modelle, Regression Trees, Random Forests, General Additive Models und Support Vector Regressions erstellt. Für jede dieser Methoden werden verschiedene Modelle mit unterschiedlichen Einflussvariablen bzw. Variablenselektionsverfahren sowie unterschiedliche Gewichtungen für Beobachtungen erstellt (Vapnik, The Nature of Statistical Learning Theory., 1995).

Zuerst werden nur die Basissignale als Einflussvariablen für die verschiedenen Methoden verwendet. Dann werden die Modelle nur auf den Daten mit positiven RUL Werten trainiert. Danach werden die Modelle mit den abgeleiteten Signalen erweitert und es erfolgt eine Variablenselektion für jeden Modelltyp. Des Weiteren wird geprüft ob eine höhere Gewichtung der Daten in der zeitlichen Nähe zum Event zu besseren Prognoseergebnisse führt. Zuletzt werden jene Modelle mit den besten Prognosegüten miteinander kombiniert.

Es werden 3 Kategorien der Prognosegüten festgelegt:

Event Time Deviation (ETD)

Die Abweichung vom Zeitpunkt, an dem das erste Mal eine RUL von null oder niedriger prognostiziert wird (2.Bedingung in (2)) und dem Zeitpunkt des Events (t_{Ei}), also dem Zeitpunkt an dem die tatsächliche RUL gleich null ist, wird als *Event Time Deviation* bezeichnet (3). Wird im gesamten Reinigungsintervall keine RUL von null oder niedriger prognostiziert, dann wird die *Event Time Deviation* auf den Wert festgelegt, der sich bei der Wartungsstrategie des Betreibers ergäbe (siehe Kapitel 3.1 Das Simple Modell). In diesem Fall wird der geschätzte Eventzeitpunkt auf die Startzeit des Zeitblocks plus sechs Monate festgelegt (1.Bedingung in (2)). Die Einheit wird in Tagen angegeben. Die Menge aller (diskreten) Zeitpunkte im Reinigungsintervall i wird durch T_m definiert (1).

$$T_m = \{t_{R_m}, t_{R_{m+1}}, \dots, t_{R_{m+1}-1}\} \quad (1)$$

$$\widehat{t}_{E_m} = \begin{cases} t_{R_m} + 180 & , \nexists \widehat{L}_t < 0 \\ \min\{t\} & , \widehat{L}_t < 0 \end{cases} \quad \text{für } t \in T_m \quad (2)$$

$$ETD_m = t_{E_m} - \widehat{t}_{E_m} \quad (3)$$

Die Event Time Deviation kann durch zufällige Einflüsse sehr stark schwanken. Wenn die RUL Prognosen stärker streuen, dann ist ein Wert von null oder niedriger möglicherweise schon früh erreicht, obwohl die Prognosen im Mittel der tatsächlichen RUL entsprechen würden. Um den Effekt dieses (zufälligen) Unterschreitens der Nulllinie abzufangen, werden die RUL Prognosen für die Event Time Deviation geglättet. Die Definition des Zeitpunktes der erstmaligen Unterschreitung ändert sich dann zu:

$$\widehat{t}_{E_m} = \begin{cases} t_{R_m} + 180 & , \nexists s^{(n)}(\widehat{L}_t) < 0 \\ \min\{t\} & , s^{(n)}(\widehat{L}_t) < 0 \end{cases} \quad \text{für } t \in T_m \quad (4)$$

Wobei die „Smoothing Funktion“ $s^{(n)}$ ein ungewichteter gleitender Mittelwert n -ter Ordnung ist. Der Glättungsparameter k wird für jedes Modell individuell optimiert. Um diesen Wert festzulegen, wird auf jedem Reinigungsintervall ein Modell trainiert und die ETD auf den Testintervallen mit verschiedenen Glättungsparametern berechnet. Der optimierte Glättungsparameter ist derjenige, bei dem die durchschnittliche ETD am niedrigsten ist.

Zusätzlich wird die Glättung der RUL (5) auch um die vergangene Zeit (innerhalb der zu glättenden RUL Werte) korrigiert. Das heißt, dass von den prognostizierten RUL-Werten die Zeitdifferenz zum Glättungszeitpunkt abgezogen wird. Dadurch wird verhindert, dass es bedingt durch die Glättung zu einem zeitlichen Offset kommt. Würde man nicht zeitlich korrigieren, dann würde es bei einem perfektem Fit zu einer Diskrepanz zwischen der tatsächlichen RUL und der geglätteten prognostizierten RUL kommen.

$$s^{(n)}(\widehat{L}_t) = \frac{1}{n+1} \sum_{i=0}^n \widehat{L}_{t_0-i} - \frac{n-1}{2} \quad (5)$$

Bemerkung: Der Mittelwert der ETD (6) wird auf den absoluten Abweichungen berechnet. Dies bedeutet, dass der Prognosefehler gleich schwer wiegt, unabhängig davon ob das prognostizierte Event vor oder nach dem tatsächlichen Event liegt.

$$\overline{ETD}_m = \sum_{m \in M} |ETD_m| \quad (6)$$

RUL Time Deviation (RTD)

Für jeden Beobachtungszeitpunkt ist die tatsächliche RUL bekannt, sofern es ein Event im Datenblock gibt. Ebenso kann für jeden Beobachtungszeitpunkt aus den Maschinendaten eine RUL durch ein Modell geschätzt werden. Die Absolutwerte der Differenzen dieser beiden Werte ergibt die RUL Time Deviation (7). Der Mittelwert der RTD für alle Reinigungsintervalle ergibt die *Mean RUL Time Deviation* (8). Die Zeiteinheiten werden in Minuten bemessen.

$$RTD_m = \frac{1}{n_{T_m}} \sum_{t \in T_m} |L_t - \widehat{L}_t| \quad (7)$$

$$\overline{RTD}_m = \sum_{m \in M} \frac{1}{n_{T_m}} \sum_{t \in T_m} |L_t - \hat{L}_t| \quad (8)$$

RUL Time Deviation auf positiven RULs

Die Mean RUL Time Deviation auf positiven RUL (9) wird aus denselben Differenzen wie in der RUL Time Deviation (10) gebildet, jedoch nur für Beobachtungen deren tatsächliche Restlaufzeit positiv war – also nur bis zum Zeitpunkt des Events.

$$T_m^{pos} = \{t_{R_m}, t_{R_m+1}, \dots, t_{E_m}\} \quad (9)$$

$$RTD_m^{pos} = \frac{1}{n_{T_m}^{pos}} \sum_{t \in T_m^{pos}} |L_t - \hat{L}_t| \quad (10)$$

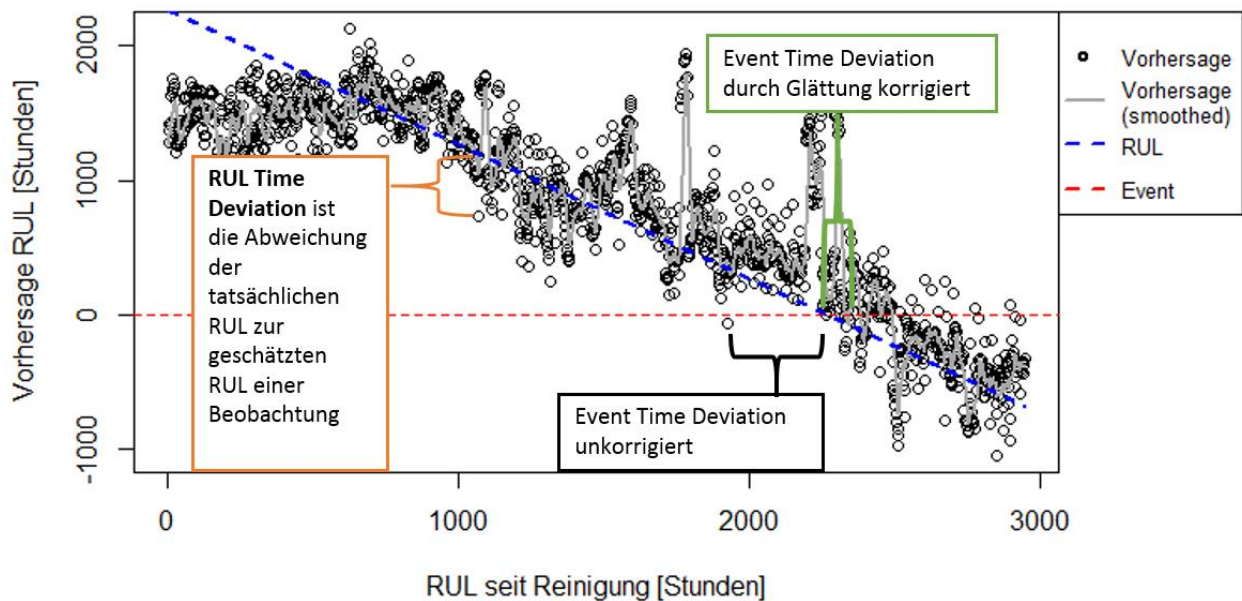


Abbildung 4: Beispielplot der Vorhersagen auf dem dritten Reinigungsintervall. Die Werte stammen aus dem linearen Modell, welches auf dem fünften Intervall trainiert wurde. Das Gütekriterium RUL Time Deviation wird durch die Abweichung der geschätzten RUL zur tatsächlichen RUL bestimmt. Das Gütekriterium Event Time Deviation ist die Zeitspanne vom erstmaligen negativ geschätzten RUL-Wert und dem Zeitpunkt des Events. Für jedes Modell werden die RUL geglättet, diese Werte bestimmen das Gütekriterium der geglätteten Event Time Deviation.

In den nachfolgenden Kapiteln werden die Modelle, ihre Umsetzung und Ergebnisse beschrieben. Im Kapitel 3.7 wird für jede Regressionsmethode das beste Modell ausgewählt

und deren Prognosegüten werden auf jedem Datenblock miteinander und mit dem *simplem Modell* verglichen.

3.1 DAS SIMPLE MODELL

Viele Industrieanlagen werden nach fixen Zeitintervallen gewartet. Wie lange diese Zeitintervalle sein sollen, wird häufig von Technikern und Experten aufgrund von Erfahrungswerten festgelegt. Herstellervorgaben können dabei eine wichtige Rolle spielen, vor allem dann wenn eine Gewährleistung der Anlagenteile nur dann besteht, wenn gewissen Wartungszeiten und Arbeiten eingehalten werden. In der Fallstudie besteht eine Mischform an Instandhaltungskonzepten. Grundsätzlich erfolgt eine **Vorbeugende Wartung**, welche besagt, dass alle 6 Monate die Reinigungsarbeiten an den Laufrädern durchgeführt werden soll. Die Anlage OST verursachte in 7 Wartungsintervallen aber 7-mal vorzeitig so starke Vibrationsprobleme, dass der Wartungsauftrag schon vorzeitig ausgelöst werden musste. Es wurde also überwiegend mit der Strategie **Reparatur nach Ausfall** vorgegangen. Diese tatsächliche Herangehensweise bezeichnen wir nachfolgend als **Das Simple Modell**, deren Kennzahlen zugleich als Basiswerte dienen, an deren die anderen Modelle bewertet werden. Die Event Time Deviation (1) des simplen Modelles wird berechnet indem vom Startzeitpunkt des Intervalls 180 Tage hinzugezählt werden und der Zeitpunkt des Events abgezogen wird. Die *RUL Time Deviation* hat zu jedem Beobachtungszeitpunkt eines Intervalls denselben Wert, da der prognostizierte Eventzeitpunkt zu Beginn eines Intervalls festgelegt und nicht geändert wird. Somit entspricht der Mittelwert der Prognoseabweichungen genau der Distanz einer einzelnen Abweichung (2).

$$ETD_m = t_{R_m} + 180 - t_{E_m} \quad \text{für alle } m \in M \quad (1)$$

$$\begin{aligned} RTD_m &= \frac{1}{n_{T_m}} \sum_{t \in T_m} |L_t - \hat{L}_t| = \\ &= \frac{1}{n_{T_m}} n_{T_m} |L_t - \hat{L}_t| = \\ &= |L_t - \hat{L}_t| \end{aligned} \quad (2)$$

Da in diesem Fall die geschätzte Restlaufzeit und die tatsächliche Restlaufzeit parallel verlaufen, ist auch die *ETD* gleich der *RTD*. Die *RTD auf den positiven Restlaufzeiten* muss ebenso gleich der *RTD* sein, da ja zu jedem Zeitpunkt dieselbe Abweichung vorliegt.

3.1.1 Ergebnisse des Simplen Modells

Reinigungs- Intervall	Event Time Deviation [Tage]	Mean RUL Time Deviation [Tage]	Mean RUL Time Deviation auf positiven RULs [Tage]
1	-134.7	134.7	134.7
2	-167	167	167
3	-16	16	16
4	-53.6	53.6	53.6
5	-151.3	151.3	151.3
6	-164.5	164.5	164.5
7	-130.8	130.8	130.8
<i>Abs. MW</i>	<i>130.8</i>	<i>130.8</i>	<i>130.8</i>
Absoluter MW ohne Intervall 1	116.8	116.8	116.8

Tabelle 9: Ergebnisse der Prädiktiven Gütekriterien des Simplen Modells. Diese Werte sollen durch die verwendeten Regressionsmethoden unterboten werden. Wenn dies gelingt, dann können die Reinigungsarbeiten zu gezielten Zeitpunkten festgelegt werden. Die Kennwerte aus dem ersten Reinigungsintervall sollen nicht in den Modellvergleich miteinfließen, da sich die Anlage nur sehr kurz im relevanten Betrieb befand. Diese Werte und die Mittelwerte welche das erste Intervall inkludieren, sind in den Tabellen in grau gehalten, und dienen nur der Veranschaulichung.

3.2 DAS MULTIPLE LINEARE REGRESSIONSMODELL

Das lineare Regressionsmodell geht davon aus, dass ein linearer Zusammenhang zwischen einer erklärenden Variable (Regressor) und der zu regressierenden (Response) Variable besteht. Das multiple Regressionsmodell beinhaltet mehrere Regressoren für die ein linearer Bezug zur Responsevariable durch geeignete Methoden geschätzt wird.

Implementierung

Für jede Beobachtung i (Datenzeile im diskretisierten Datensatz) lässt sich das lineare Modell wie folgt anschreiben:

$$L_t = \beta_0 + \beta_1 * X_{t1} + \beta_2 * X_{t2} + \dots + \beta_k * X_{tk} + e_t$$

L_t repräsentiert die Responsevariable und ist in der Fallstudie die generierte Remaining Useful Life (RUL). X_{t1} ist die Beobachtung der Messvariable 1 zum Zeitpunkt t , X_{t2} die von der Messvariable 2 usw. Die β_* sind die Regressionskoeffizienten welche es zu schätzen gilt, indem sie so gewählt werden, dass die Residuen e_t (Abweichungen oder auch Prognosefehler) minimiert werden. Durch Anwendung der Kleinstquadrate-Methode werden die β_* so bestimmt, dass die Summe der quadratischen Abweichungen minimiert wird:

$$\sum_{i=1}^n e_{t_i} = \sum_{i=1}^n (L_{t_i} - \beta_0 - \beta_1 * X_{t_i,1} - \dots - \beta_k * X_{t_i,k})^2$$

Lineare Regressionsmodelle können gute Prognoseergebnisse liefern, wenn folgende Grundannahmen erfüllt sind.

Additivität:

Die erklärenden Variablen sind nicht signifikant voneinander beeinflusst (abhängig). Dadurch können die einzelnen Variablen mit den Koeffizienten β multipliziert, und aufsummiert werden.

Linearität:

Die Responsevariable kann durch eine lineare Funktion der erklärenden Variablen und den Fehlertermen abgebildet werden.

Varianzhomogenität:

Die Fehlerterme haben dieselbe Varianz unabhängig von den Werten der erklärenden Variablen.

Variablenselektion

Für das lineare Modell gibt es eine Vielzahl an Variablenselektionsverfahren. In der Fachliteratur übliche Verfahren sind das korrigierte Bestimmtheitsmaß, Mellow's Cp, Akaike Information Criterion, sowie das Bayes Information Criterion, LASSO und LARS. In der Fallstudie werden die zwei gängigsten Methoden AIC und BIC verwendet, sowie ein heuristisches Selektionsverfahren das Prognosegüte anstatt eines Bestimmtheitsmaßes zur Bewertung der Variablen verwendet. Dieses Verfahren wird auch Kreuzvalidierungsmethode zur Variablenselektion genannt (CV-Methode), und kann grundsätzlich für jede Regressionsmethode angewandt werden (Stone, 1974). Dabei wird wie folgt vorgegangen:

1. Der Datensatz wird in k Subdatensätze zerlegt, wobei ein Subdatensatz genau einem Datenblock entspricht.
2. Es werden alle Subdatensätze bis auf den k -ten zur Erstellung des Modells verwendet. Auf dem k -ten Subdatensatz werden mittels des soeben berechneten Modells die Restlaufzeiten prognostiziert, und die absoluten Prognosefehler aufsummiert.¹
3. Schritt 2 wird mit jedem Subdatensatz k durchgeführt. Es werden alle absoluten Prognosefehler aufsummiert.
4. Schritt 1 bis 3 wird für jede gewünschte Modellparametrisierung bzw. Variablenauswahl durchgeführt. Das Modellsetting mit der geringsten Summe der absoluten Prognoseabweichungen wird ausgewählt und mit dessen Setting wird auf den gewünschten Datensatz das für den Betrieb zu verwendende Modell erstellt.

Sind keine weiteren Informationen über die Relevanz der Variablen verfügbar, dann ist der naive Ansatz alle Kombinationen von Variablen mit der CV-Methode zu testen. Dies führt bei m Variablen zu $2^m - 1$ Variablensettings. Man nennt dies die *Vollständige Modellselektion (All-Subset-Selection)*. Die Zahl an möglichen Settings steigt mit jedem zusätzlichen Signal also sehr schnell an. Dies kann insbesondere bei rechenintensiveren Regressionsmethoden wie zum Beispiel den Random Forest's zu langen Berechnungszeiten führen, und somit ein „durchprobieren“ aller Möglichkeiten praktisch unmöglich machen. Ansätze die Rechenzeiten in überschaubaren Maß zu halten sind:

- Vorwärts Selektion (Forward Selection)
Die Vorwärts Selektion kann nicht für jede Regressionsmethode verwendet werden, da manche Methoden eine Mindestanzahl an Variablen benötigen um ein Modell erstellen zu können. Dies betrifft zum Beispiel die Random Forests.
- Rückwärts Selektion (Backward Selection)

¹ In der Fachliteratur wird meist der quadrierte Prognosefehler aufsummiert. Für die Fallstudie sind aber die absoluten Prognoseabweichungen leichter zu interpretieren. Es kann somit aufgrund der durchschnittlichen absoluten Abweichung ein zeitlicher Sicherheitspuffer geschätzt werden, der notwendig ist um einen vorrausschauenden Wartungsauftrag früh genug auszulösen.

Die Rückwärts Selektion hat den entscheidenden Nachteil, dass Variablen die einmal ausgeschlossen worden sind, nicht wieder in ein Modell aufgenommen werden können. Dies kann bei stark korrelierten Variablen zu schlechten Ergebnissen führen.

- Schrittweise Selektion (Stepwise)

In dieser Arbeit wird die schrittweise Selektion mittels der CV Methode auf den Prognosefehlern, zur Variablenselektion verwendet. Diese Methode kann unabhängig von der Art der Regressionsmethode angewandt werden, und die Berechnungszeiten bleiben akzeptabel lange.

Die CV Methode wird ohne den Datensatz des Block 1 angewendet, da dieser Datensatz nur wenige Messpunkte mit positiver RUL hat und zu starken Ausreißern in der Validierung führt.

3.2.1 Beschreibung und Ergebnisse der linearen Modelle

3.2.1.1 Lineares Modell mit allen Basissignalen (LM)

Dieses lineare Modell verwendet alle Basissignale die aus der Maschinensteuerung kommen. Es wird weder eine Variablenselektion vorgenommen, noch wird eine Gewichtung für die Beobachtungen festgelegt. Die RUL (L_i) wird durch die Formel

$$L_i = \beta_0 + \beta_1 * X_{i1} + \beta_2 * X_{i2} + \dots + \beta_k * X_{i9} + e_i \quad \text{für } i = 1, \dots, n \text{ Beobachtungen}$$

bestimmt, wobei X_i die 9 Basissignale (ohne das Signal „Zustand“, da nur Daten verwendet werden bei denen sich die Anlage auf Zustand „Ausfahren“ befindet) der i -ten Beobachtung sind.

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	-4.18e+04	1.83e+03	-22.87	< 2e-16	***
Drehmoment_OST	-1.47e+00	3.92e+01	-0.04	0.9702	
Drehzahl_OST	1.18e+01	2.27e+00	5.17	2.3e-07	***
Leistung_OST	-7.57e+01	2.93e+01	-2.59	0.0097	**
Motorstrom_OST	1.60e+02	6.38e+00	25.13	< 2e-16	***
Gerätetemp_OST	3.99e+02	1.10e+01	36.39	< 2e-16	***
Motortemp_OST	-3.89e+01	3.52e+00	-11.06	< 2e-16	***
Giessgeschwindigkeit	-1.72e+02	1.88e+01	-9.17	< 2e-16	***
Vibration_OST	-4.42e+02	4.80e+00	-92.04	< 2e-16	***
Temp_OST	2.35e+01	8.97e-01	26.26	< 2e-16	***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1					
Residual standard error: 932 on 14544 degrees of freedom					
Multiple R-squared: 0.485, Adjusted R-squared: 0.484					
F-statistic: 1.52e+03 on 9 and 14544 DF, p-value: <2e-16					

Modellübersicht 1: Es sind alle Basissignale bis auf das Drehmoment statistisch signifikant und können in das Modell aufgenommen werden. Dies ist bei einer sehr hohen Anzahl an Beobachtungen (14544) auch zu erwarten. Dass das Drehmoment eine nicht signifikante Einfluss Variable ist lässt sich vielleicht mit einer sehr hohen Korrelation zur Leistung erklären.

Auch die Signifikanz der Leistung scheint unter der Korrelation zum Drehmoment abgeschwächt zu sein. Die Modellgüte (Adjusted R-squared) ist mit 0.484 ein mittelmäßiges Resultat.

3.2.1.2 Lineares Modell mit AIC stepwise selektierten Signalen (LM AIC)

Dieses lineare Modell verwendet alle Basissignale die aus der Maschinensteuerung kommen und wendet das Akaike Information Criterion (AIC) zur Variablenselektion an.

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	-4.19e+04	1.31e+03	-32.01	<2e-16	***
Vibration_OST	-4.42e+02	4.79e+00	-92.24	<2e-16	***
Gerätetemp_OST	3.99e+02	1.09e+01	36.45	<2e-16	***
Temp_OST	2.36e+01	8.74e-01	26.95	<2e-16	***
Motorstrom_OST	1.60e+02	6.25e+00	25.65	<2e-16	***
Motortemp_OST	-3.89e+01	3.48e+00	-11.18	<2e-16	***
Giessgeschwindigkeit	-1.72e+02	1.85e+01	-9.32	<2e-16	***
Drehzahl_OST	1.18e+01	1.42e+00	8.33	<2e-16	***
Leistung_OST	-7.68e+01	6.46e+00	-11.90	<2e-16	***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1					

Residual standard error: 932 on 14545 degrees of freedom
 Multiple R-squared: 0.485, Adjusted R-squared: 0.485
 F-statistic: 1.71e+03 on 8 and 14545 DF, p-value: <2e-16

Modellübersicht 2: Das AIC Selektionsverfahren entfernt lediglich die nicht signifikante Drehmoment Variable aus dem vorigen Modell. Es führt aber zu keiner nennenswerten Verbesserung des linearen Modells.

3.2.1.3 Lineares Modell mit BIC stepwise selektierten Signalen (LM BIC)

Dieses lineare Modell verwendet alle Basissignale die aus der Maschinensteuerung kommen und wendet das Bayesian Information Criterion (BIC) zur Variablenselektion an. Bei diesem Verfahren hängt der Strafterm zur Aufnahme neuer Variablen von der Anzahl der Beobachtungen ab. In dieser Anwendung wird das BIC Verfahren eine striktere Auswahl an Variablen treffen, da eine große Beobachtungsanzahl von 14554 zu einem großen Strafterm führt.

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	-3.80e+04	1.05e+03	-36.21	< 2e-16	***
Vibration_OST	-4.42e+02	4.80e+00	-92.17	< 2e-16	***
Gerätetemp_OST	3.98e+02	1.09e+01	36.30	< 2e-16	***
Temp_OST	2.31e+01	8.79e-01	26.27	< 2e-16	***
Motorstrom_OST	1.60e+02	6.38e+00	25.13	< 2e-16	***
Motortemp_OST	-4.03e+01	3.48e+00	-11.60	< 2e-16	***
Drehmoment_OST	-1.00e+02	8.65e+00	-11.61	< 2e-16	***
Giessgeschwindigkeit	-1.82e+02	1.84e+01	-9.86	< 2e-16	***

```

Drehzahl_OST          6.60e+00  1.09e+00  6.07  1.3e-09 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 932 on 14545 degrees of freedom
Multiple R-squared:  0.485, Adjusted R-squared:  0.484
F-statistic: 1.71e+03 on 8 and 14545 DF,  p-value: <2e-16

```

Modellübersicht 3: Das BIC Selektionsverfahren verhält sich trotz des größeren Strafterms ebenso wie das AIC Verfahren und entfernt lediglich die nicht signifikante Drehmoment Variable aus dem Modell 3.2.1.1, ohne eine nennenswerte Verbesserung des Modells.

3.2.1.4 Modell nur auf positiven RUL (LM positive RUL)

Dieses Modell wird nur auf Beobachtungen trainiert, die zeitlich vor dem Auftreten des Events liegen ($i = 1, \dots, n$). Somit tragen Datenzeilen mit negativen RUL nicht zur Modellbildung bei. Es soll somit überprüft werden, ob beim Weglassen der Daten, bei denen sich die Anlage in einem Schadhafte Zustand befindet, die Prognosegüte erhöht.

$L_{t_i} = \beta_0 + \beta_1 * X_{t_i,1} + \beta_2 * X_{t_i,2} + \dots + \beta_9 * X_{t_i,9} + e_{t_i}$ für $i = 1, \dots, n$ Beobachtungen

```

                Estimate Std. Error t value Pr(>|t|)
(Intercept)    -2.08e+04  1.81e+03 -11.47 < 2e-16 ***
Drehmoment_OST -3.29e+02  3.90e+01  -8.45 < 2e-16 ***
Drehzahl_OST    7.82e+00  2.29e+00   3.42 0.00064 ***
Leistung_OST    1.68e+02  2.93e+01   5.75 9.3e-09 ***
Motorstrom_OST  3.35e+01  6.41e+00   5.22 1.8e-07 ***
Gerätetemp_OST  4.16e+02  9.81e+00  42.38 < 2e-16 ***
Motortemp_OST  -6.08e+01  3.22e+00 -18.92 < 2e-16 ***
Giessgeschwindigkeit 1.12e+02  1.82e+01   6.18 6.9e-10 ***
Vibration_OST  -5.38e+02  7.50e+00 -71.71 < 2e-16 ***
Temp_OST        -6.37e+00  8.77e-01  -7.26 4.2e-13 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 674 on 8211 degrees of freedom
Multiple R-squared:  0.459, Adjusted R-squared:  0.458
F-statistic: 773 on 9 and 8211 DF,  p-value: <2e-16

```

Modellübersicht 4: Die Modellgüte (Adjusted R-squared) kann nicht erhöht werden, indem die Daten nach einem Event weggelassen werden. Das heißt, dass es wahrscheinlich besser ist, alle Beobachtungen mit in die Modellgestaltung aufzunehmen.

3.2.1.5 Modelle inklusive abgeleiteten Signalen (LM mit Features)

Dieses Modell beinhaltet die Basissignale ($X_{\cdot 1}, \dots, X_{\cdot 9}$) sowie die zusätzlich erstellten abgeleiteten Signale ($Z_{\cdot 1}, \dots, Z_{\cdot p}$).

$$L_{t_i} = \beta_0 + \beta_1 * X_{t_i,1} + \beta_2 * X_{t_i,2} + \dots + \beta_9 * X_{t_i,9} + \beta_{10} * Z_{t_i,1} + \dots + \beta_{9+p} * Z_{t_i,p} + e_{t_i}$$

für $i = 1, \dots, n$ Beobachtungen

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	-32803.45	2091.50	-15.68	< 2e-16	***
Drehmoment_OST	92.27	46.56	1.98	0.04755	*
Drehzahl_OST	12.48	2.54	4.91	9.1e-07	***
Leistung_OST	-114.93	35.04	-3.28	0.00104	**
Motorstrom_OST	69.28	8.16	8.50	< 2e-16	***
Gerätetemp_OST	331.37	12.34	26.84	< 2e-16	***
Motortemp_OST	33.08	5.28	6.26	4.0e-10	***
Giessgeschwindigkeit	-34.24	21.34	-1.60	0.10868	
Vibration_OST	-408.66	10.64	-38.42	< 2e-16	***
Temp_OST	31.19	1.24	25.21	< 2e-16	***
Drehmoment_OST_sDes1	512.19	50.01	10.24	< 2e-16	***
Drehmoment_OST_sAsc1	-138.70	48.01	-2.89	0.00387	**
Drehmoment_OST_sd	-564.64	124.68	-4.53	6.0e-06	***
Drehmoment_OST_skewness	-460.42	95.72	-4.81	1.5e-06	***
Drehmoment_OST_kurtosis	-230.94	42.00	-5.50	3.9e-08	***
Drehzahl_OST_sDes1	9.50	2.89	3.29	0.00099	***
Drehzahl_OST_sAsc1	-25.63	3.36	-7.63	2.5e-14	***
Drehzahl_OST_sd	-50.56	8.62	-5.86	4.6e-09	***
Drehzahl_OST_skewness	-17.32	22.83	-0.76	0.44818	
Drehzahl_OST_kurtosis	26.94	4.77	5.65	1.6e-08	***
Leistung_OST_sDes1	-333.15	40.09	-8.31	< 2e-16	***
Leistung_OST_sAsc1	131.24	32.88	3.99	6.6e-05	***
Leistung_OST_sd	631.65	87.37	7.23	5.1e-13	***
Leistung_OST_skewness	246.03	91.30	2.69	0.00705	**
Leistung_OST_kurtosis	36.75	36.13	1.02	0.30906	
Motorstrom_OST_sDes1	-42.16	12.90	-3.27	0.00109	**
Motorstrom_OST_sAsc1	11.57	16.50	0.70	0.48300	
Motorstrom_OST_sd	-227.41	32.71	-6.95	3.8e-12	***
Motorstrom_OST_skewness	-111.02	34.21	-3.24	0.00118	**
Motorstrom_OST_kurtosis	19.09	17.26	1.11	0.26877	
Gerätetemp_OST_sDes1	-369.61	65.37	-5.65	1.6e-08	***
Gerätetemp_OST_sAsc1	-310.60	80.77	-3.85	0.00012	***
Gerätetemp_OST_sd	231.62	82.00	2.82	0.00474	**
Gerätetemp_OST_skewness	1.09	14.63	0.07	0.94054	
Gerätetemp_OST_kurtosis	-31.35	7.43	-4.22	2.5e-05	***
Motortemp_OST_sDes1	527.94	53.32	9.90	< 2e-16	***
Motortemp_OST_sAsc1	-437.54	73.11	-5.99	2.2e-09	***
Motortemp_OST_sd	511.33	42.78	11.95	< 2e-16	***
Motortemp_OST_skewness	-12.73	36.17	-0.35	0.72496	
Motortemp_OST_kurtosis	203.60	25.76	7.90	2.9e-15	***
Giessgeschwindigkeit_sDes1	-178.87	59.02	-3.03	0.00244	**
Giessgeschwindigkeit_sAsc1	-722.62	151.69	-4.76	1.9e-06	***
Giessgeschwindigkeit_sd	-480.52	96.80	-4.96	7.0e-07	***

Giessgeschwindigkeit_skewness	115.94	20.22	5.73	1.0e-08	***
Giessgeschwindigkeit_kurtosis	-20.71	13.04	-1.59	0.11222	
Vibration_OST_sDes1	63.16	60.89	1.04	0.29966	
Vibration_OST_sAsc1	-464.71	82.38	-5.64	1.7e-08	***
Vibration_OST_sd	-212.39	57.87	-3.67	0.00024	***
Vibration_OST_skewness	67.96	25.40	2.68	0.00747	**
Vibration_OST_kurtosis	57.94	11.89	4.87	1.1e-06	***
Temp_OST_sDes1	8.62	17.55	0.49	0.62313	
Temp_OST_sAsc1	-10.75	11.39	-0.94	0.34556	
Temp_OST_sd	6.60	7.51	0.88	0.37961	
Temp_OST_skewness	-431.77	30.54	-14.14	< 2e-16	***
Temp_OST_kurtosis	-96.69	24.82	-3.90	9.9e-05	***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 871 on 12801 degrees of freedom
(1698 observations deleted due to missingness)

Multiple R-squared: 0.582, Adjusted R-squared: 0.58

F-statistic: 330 on 54 and 12801 DF, p-value: <2e-16

Modellübersicht 5: Durch die vielen zusätzlichen Signale hat dieses Modell eine stark erhöhte Modellgüte von 0.58 bekommen. Es gibt aber noch sehr viele Variablen die nicht signifikant sind, und eventuell aus dem Modell entfernt werden sollten. Die Gefahr eines Overfittings ist hier sehr groß. Wie stark betroffen dieses Modell davon ist, wird sich in den Analysen der prädiktiven Modellgüte zeigen.

3.2.1.6 Modelle inklusive abgeleiteten Signalen und AIC Variablenselektion (LM mit Features AIC)

Dieses Modell beinhaltet die Basissignale (X_1, \dots, X_9) sowie die zusätzlich erstellten abgeleiteten Signale (Z_1, \dots, Z_p). Die Variablenselektion mit der AIC Methode führt zu einem um 10 Signale verringertes Modell (gegenüber dem Modell 3.2.1.5), wobei die Modellgüte bei 0,58 bleibt.

3.2.1.7 Modelle inklusive abgeleiteten Signalen und BIC Variablenselektion (LM mit Features BIC)

Die Variablenselektion mit der BIC Methode führt zu einem um 19 Signale verringerten Modell (gegenüber dem Modell 3.2.1.5), wobei die Modellgüte bei 0,579 annähernd gleichbleibt.

3.2.1.8 Modelle inklusive abgeleiteten Signalen und Variablenselektion mit der CV-Methode (LM mit Features CV)

Die Variablenselektion mit der CV Methode führt zu einer sehr stark reduzierten Auswahl an Signalen.

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-8136.292	571.067	-14.2	<2e-16 ***
Vibration_OST	-505.624	6.227	-81.2	<2e-16 ***
Motortemp_OST_kurtosis	-145.996	7.080	-20.6	<2e-16 ***
Gerätetemp_OST_sd	993.030	57.504	17.3	<2e-16 ***
Drehzahl_OST	13.623	0.789	17.3	<2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1				

Residual standard error: 1060 on 12209 degrees of freedom

Multiple R-squared: 0.393, Adjusted R-squared: 0.393

F-statistic: 1.98e+03 on 4 and 12209 DF, p-value: <2e-16

Modellübersicht 6: Mit vier Einflussvariablen ist dieses Modell sehr schlank. Daher geht auch die Modellgüte stark zurück auf 0,393. Eine gute prädiktive Prognosegüte würde bedeuten, dass man durch die einfache Zusammensetzung des Modells eine gute Interpretationsmöglichkeit bekommt.

3.2.1.9 Lineares Modell mit verstärktem Lernen gegen Ende der Events (Gewichtetes LM)

Je länger die Anlage im Betrieb ist, desto höher ist der Verschmutzungsgrad, und desto besser lässt sich dieser feststellen. Theoretisch sollte sich dies auch in den Messwerten widerspiegeln, indem zum Beispiel das Vibrationsniveau steigt. Folgt man diesen Ansatz, kann es sinnvoll erscheinen die Messwerte je nach Anlagenlaufzeit zu gewichten. Im Idealfall sind die Gewichte für die Messwerte rund um das Event hoch, und nehmen ab je größer die Distanz zum Event wird. Eine einfache Form dies für lineare Modelle umzusetzen, sind lineare Gewichtungen in Bezug auf die RUL. Deshalb wurden die Gewichte so gewählt, dass die (innerhalb eines Reinigungsintervalls) zeitlich am weitesten entfernte Beobachtung das geringste Gewicht für das Modell erhält. Bis zum Event steigt die Gewichtung linear an, und nimmt ab diesem Zeitpunkt wieder linear ab. Damit sollen die Modelle eine höhere Prognosegenauigkeit bekommen, wenn sich die Anlage dem kritischen Zustand nähert. Die Anlage wird meistens auch nach Auftreten des Events weiter betrieben, da die laufende Stahlproduktion nicht unterbrochen werden soll. Aus diesem Grund sind die Daten nach einem Event auch von Bedeutung, und sollten mit ins Modell aufgenommen werden.

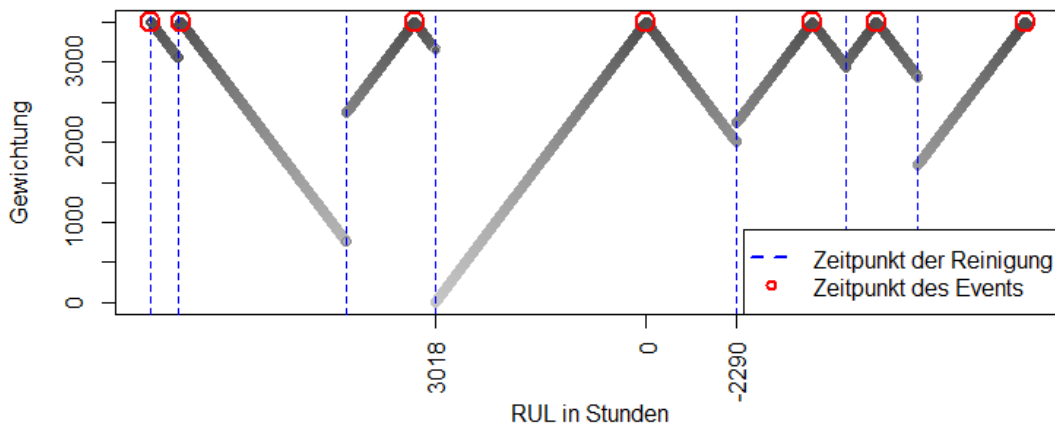


Abbildung 5: Gewichte der Beobachtungen für die 7 Zeitblöcke. Die roten Kreise markieren die Zeitpunkte der Events. Die blauen Vertikalen markieren die Zeitpunkte der Reinigungen.

Das Modell mit den Gewichtungen kann wie folgt umgeschrieben werden.

$$L_{t_i}^* = \beta_0 + \beta_1 * X_{t_{i,1}}^* + \beta_2 * X_{t_{i,2}}^* + \dots + \beta_k * X_{t_{i,9}}^* + e_{t_i}^* \quad \text{für } i = 1, \dots, n^* \text{ Beobachtungen}$$

mit

$$L_{t_i}^* := \sqrt{w_{t_i}} L_{t_i}$$

$$X_{t_i}^* := \sqrt{w_{t_i}} X_{t_i}$$

$$e_{t_i}^* := \sqrt{w_{t_i}} e_{t_i}$$

$$w_{t_i} := \max(L_T) - |L_{t_i}|$$

Wobei L_T die Restlaufzeit im gesamten Beobachtungszeitraum ist, also aller Reinigungsintervalle.

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	-2.05e+04	1.43e+03	-14.31	< 2e-16	***
Drehmoment_OST	-4.89e+01	3.05e+01	-1.60	0.11	
Drehzahl_OST	2.30e+00	1.79e+00	1.28	0.20	
Leistung_OST	9.00e+00	2.28e+01	0.39	0.69	
Motorstrom_OST	1.03e+02	4.93e+00	20.87	< 2e-16	***
Gerätetemp_OST	1.77e+02	8.22e+00	21.50	< 2e-16	***
Motortemp_OST	-2.21e+01	2.76e+00	-7.98	1.5e-15	***
Giessgeschwindigkeit	-1.12e+02	1.49e+01	-7.54	4.9e-14	***
Vibration_OST	-3.14e+02	3.74e+00	-83.90	< 2e-16	***
Temp_OST	1.70e+01	7.09e-01	23.98	< 2e-16	***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 727 on 14543 degrees of freedom

Multiple R-squared: 0.408, Adjusted R-squared: 0.408
 F-statistic: 1.11e+03 on 9 and 14543 DF, p-value: <2e-16

Modellübersicht 7: Laut Modellgüte ist das Modell mit gewichteten Beobachtungen marginal schlechter als das Basismodell. Dies ist aber zu erwarten, da das Modell Abweichungen in der Nähe des Events stärker bestraft, und größere Fehler bei weit entfernten Beobachtungen in Kauf genommen werden. Dadurch sinkt die Modellgüte etwas ab, was jedoch nicht bedeuten muss, dass sich die prädiktiven Eigenschaften des Modells verschlechtern. Vor allem sollte die Event Time Deviation von gewichteten Modellen besser prognostiziert werden können, da sich die Prognosen um das Event herum konzentrieren.

3.2.1.10 Gewichtetes lineares Modell mit abgeleiteten Signalen (gewichtetes LM mit Features)

Bei diesem Modell werden sowohl die abgeleiteten Signale integriert, als auch die Gewichtung der Beobachtungen berücksichtigt. Es handelt sich also um eine Mischung aus dem Modell 3.2.1.5 und dem Modell 3.2.1.9. Die Hinzunahme der abgeleiteten Signale verbessern die Modellgüte, wohingegen die Gewichtung die Modellgüte etwas schlechter ausfallen lässt. Dadurch ergibt sich eine mittelmäßige Güte von 0.49.

3.2.1.11 Gewichtetes lineares Modell mit abgeleiteten Signalen und AIC Variablenselektionsverfahren (gewichtetes LM mit Features AIC)

Dieses Modell basiert auf dem Modell 3.2.1.10, auf dem eine AIC Variablenselektion durchgeführt wurde. Die Modellgüte ist mit einem Wert von 0.491 im Mittelmaß. (Details im Anhang)

3.2.1.12 Gewichtetes lineares Modell mit abgeleiteten Signalen und BIC Variablenselektionsverfahren (gewichtetes LM mit Features BIC)

Dieses Modell basiert auf dem Modell 3.2.1.10, auf dem eine BIC Variablenselektion durchgeführt wurde. Die Modellgüte ist mit einem Wert von 0.479 im Mittelmaß.

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	-1.23e+04	4.69e+02	-26.29	< 2e-16	***
Vibration_OST	-2.96e+02	7.20e+00	-41.08	< 2e-16	***
Motortemp_OST_sd	5.21e+02	2.63e+01	19.83	< 2e-16	***
Gerätetemp_OST	1.59e+02	8.64e+00	18.38	< 2e-16	***
Temp_OST_skewness	-3.86e+02	2.40e+01	-16.11	< 2e-16	***
Drehzahl_OST_sAsc1	-3.35e+01	1.98e+00	-16.90	< 2e-16	***
Temp_OST	2.14e+01	9.12e-01	23.47	< 2e-16	***
Motorstrom_OST	2.67e+01	2.95e+00	9.06	< 2e-16	***
Giessgeschwindigkeit_sAsc1	-6.62e+02	9.59e+01	-6.90	5.4e-12	***
Giessgeschwindigkeit_skewness	1.21e+02	1.21e+01	9.99	< 2e-16	***
Leistung_OST_sAsc1	2.12e+02	2.18e+01	9.73	< 2e-16	***
Drehzahl_OST_kurtosis	2.80e+01	2.01e+00	13.89	< 2e-16	***
Gerätetemp_OST_kurtosis	-5.84e+01	4.91e+00	-11.91	< 2e-16	***
Motortemp_OST_skewness	-1.14e+02	2.44e+01	-4.67	3.1e-06	***
Vibration_OST_sd	-2.51e+02	4.32e+01	-5.81	6.3e-09	***
Drehmoment_OST_kurtosis	-1.55e+02	1.11e+01	-13.89	< 2e-16	***
Drehmoment_OST_sAsc1	-2.17e+02	2.90e+01	-7.49	7.6e-14	***
Vibration_OST_kurtosis	6.47e+01	7.66e+00	8.45	< 2e-16	***
Temp_OST_sDes1	6.62e+01	7.57e+00	8.74	< 2e-16	***

Drehmoment_OST_skewness	-1.70e+02	2.11e+01	-8.06	8.4e-16	***
Vibration_OST_skewness	9.66e+01	1.74e+01	5.55	2.9e-08	***
Motortemp_OST_sDes1	5.86e+02	3.67e+01	15.96	< 2e-16	***
Motortemp_OST_sAsc1	-5.99e+02	4.87e+01	-12.32	< 2e-16	***
Motortemp_OST	2.89e+01	3.97e+00	7.29	3.2e-13	***
Gerätetemp_OST_sDes1	-2.75e+02	4.43e+01	-6.21	5.4e-10	***
Motortemp_OST_kurtosis	1.26e+02	1.90e+01	6.64	3.2e-11	***
Temp_OST_kurtosis	-9.42e+01	1.60e+01	-5.88	4.1e-09	***
Motorstrom_OST_kurtosis	3.35e+01	8.76e+00	3.82	0.00013	***
Gerätetemp_OST_skewness	-4.32e+01	1.13e+01	-3.84	0.00012	***
Vibration_OST_sAsc1	-2.17e+02	5.93e+01	-3.66	0.00025	***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 691 on 12825 degrees of freedom
Multiple R-squared: 0.481, Adjusted R-squared: 0.479
F-statistic: 409 on 29 and 12825 DF, p-value: <2e-16

3.2.1.13 Gewichtetes lineares Modell mit abgeleiteten Signalen und der CV-Methode zur Variablenselektion (gewichtetes LM mit Features CV)

Bei diesem Modell werden sowohl die abgeleiteten Signale integriert, als auch die Gewichtung der Beobachtungen berücksichtigt. Das Auswahlkriterium der CV-Methode reagiert sehr viel strenger als die Verfahren die auf der Modellgüte selbst basieren.

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	-5670.060	483.036	-11.74	<2e-16	***
Vibration_OST	-336.998	5.899	-57.13	<2e-16	***
Gerätetemp_OST_kurtosis	-35.628	3.170	-11.24	<2e-16	***
Drehzahl_OST	9.677	0.666	14.53	<2e-16	***
Vibration_OST_sd	-331.734	39.776	-8.34	<2e-16	***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 865 on 12209 degrees of freedom
Multiple R-squared: 0.294, Adjusted R-squared: 0.294
F-statistic: 1.27e+03 on 4 and 12209 DF, p-value: <2e-16

Modellübersicht 8: Durch das strenge Auswahlverfahren der CV-Methode bleiben nur vier Einflussvariablen übrig, und die Modellgüte sinkt stark ab auf 0.294.

3.2.2 Selektion des Modelles mit der besten Vorhersagegüte

Es sollen die Modelle anhand der Vorhersagegenauigkeit auf unabhängigen Testsets bewertet werden. Jedes Modell wird auf je einem Datenblock trainiert und auf den übrigen 6 Datenblöcke werden die Prognostizierten RUL Zeiten ausgewertet. Es wird jenes Modell als das Beste ausgewählt, welches in der Kategorie RUL Time Deviation auf den positiven Restlaufzeiten den niedrigsten Wert hat.

Modell	Geglättete Mean Event Time Deviation [Tage]	Mean RUL Time Deviation [Tage]	Mean RUL Time Deviation auf positiven RULs [Tage]
gewichtetes LM mit Features BIC (3.2.1.12)	37.45	62.22	44.39
gewichtetes LM mit Features CV	41.01	64.87	46.76
LM mit Features CV	34.86	68.23	47.19
gewichtetes LM mit Features AIC	46.77	64.36	47.87
LM mit Features BIC	48.92	67.49	48.57
gewichtetes LM	32.86	67.36	49.28
LM	30.86	70.75	50.48
LM BIC	34.03	70.18	50.59
LM AIC	34.42	70.43	50.74
gewichtetes LM mit Features	41.71	68.22	51.55
LM mit Features AIC	48.87	70.46	52.03
LM positive RUL	36.89	54.24	54.24
LM mit Features	47.72	73.79	55.4

Tabelle 10: Die Prognosegüten sind ohne die Ergebnisse des 1. Intervalls aggregiert. Der Grund dafür war die sehr kurze Laufzeit im Betriebsrelevanten Zustand. Das gewichtete LM mit Features BIC wird als bestes Modelle der Linearen Regressionsmethode ausgewählt, da es die geringsten Abweichungen in der Kategorie Mean RUL Time Deviations auf positiven RULs hat.

Einige Trainingssets eignen sich nicht dazu, eine zufriedenstellende Prognosegüte für die Testsets zu erreichen. Modelle die auf dem ersten Datenblock trainiert werden, haben eine sehr schlechte Prognosegüte für alle anderen Datenblöcke. Dies kann bedeuten, dass die Anlage in der ersten Aufzeichnungsperiode maßgeblich anders betrieben wurde als in den übrigen Perioden. Um ein erfolgreiches Modell für eine Echtzeitprognose zu erstellen sollte in Betracht gezogen werden diese Datenperiode nicht in das Trainingsset aufzunehmen. Es ist daher auch sinnvoll die Prognosegüten der auf dem ersten Datenblock trainierten Modelle von sämtlichen Evaluierungen auszuschließen.

Das Modell mit der besten Prognosegüte im Detail:

Das *gewichtete Modell mit Features und BIC Selektion (3.2.1.12)* hat die beste Prognosegüte in der Kategorie Mean RUL Time Deviation auf positiven RUL, und damit für den Anwendungsfall die relevanteste Voraussetzung für eine gute Langzeitprognose im

Echtzeitbetrieb. Zur Bestimmung der Event Time Deviation wurden die prognostizierten RULs mit einer Breite von 22 Stunden geglättet.

Trainset	Test 1	Test 2	Test 3	Test 4	Test 5	Test 6	Test 7	MW	Absoluter MW ohne Test 1
1	-	16.49	55.34	147.5	15.74	61.23	125	70.22	
2	0	-	84.5	303.4	40.41	25.88	9.02	77.2	92.64
3	-33	11.83	-	8.26	0	0.75	0.5	9.058	4.26
4	0	11.27	0.46	-	15.93	-0.33	0.75	4.79	5.74
5	0	-4.7	-9.8	64.75	-	-1.37	1.05	13.61	16.33
6	0	-0.22	133.3	30.78	7.05	-	-142	52.26	62.71
7	0	11.82	124.8	64.78	11.97	1.71	-	35.84	43.01
Abs. MW	5.502	9.38	68.03	103.2	15.18	15.21	46.43	37.57	
Absoluter MW ohne Trainset 1		7.96	70.57	94.39	15.07	6.01	30.71		<u>37.45</u>

Tabelle 11: Event Time Deviation des Modells mit Features und BIC Selektion (3.2.1.12). Der Mittelwert aller Prognoseabweichungen ist mit 37,45 viel besser als beim Simplen Modell (116,8).

Trainset	Test 1	Test 2	Test 3	Test 4	Test 5	Test 6	Test 7	MW	MW ohne Test 1
1	-	420.8	171.2	509.8	147.1	112.4	244.9	267.7	
2	58.01	-	74.29	120	55.93	67.18	51.58	71.17	73.8
3	30.44	217.5	-	152.2	33.45	35.95	67.18	89.45	101.3
4	51.06	80.85	21.68	-	33.04	54.51	28.71	44.98	43.76
5	60.66	101.4	16.47	70.43	-	23.48	24.57	49.51	47.27
6	6.72	74.02	19.74	69.19	18.55	-	28.52	36.12	42
7	84.76	110.8	35.51	77.06	41.23	61.61	-	68.5	65.25
MW	48.61	167.6	56.49	166.4	54.89	59.19	74.25	89.63	
MW ohne Trainset 1		116.9	33.54	97.79	36.44	48.55	40.11		<u>62.22</u>

Tabelle 12: RUL Time Deviation des Modells mit Features und BIC Selektion (3.2.1.12). Der Mittelwert aller Prognoseabweichungen ist mit 62,22 besser als beim Simplen Modell (116,8).

Trainset	Test 1	Test 2	Test 3	Test 4	Test 5	Test 6	Test 7	MW	MW ohne Testset 1
1	-	67.83	194.9	628.7	173.9	80.34	246.3	232	
2	65.56	-	72.84	143.3	49.6	28.98	51.62	68.6	69.27
3	5.85	30.14	-	176.1	37.76	27.59	67.46	57.4	67.81
4	79.73	6.86	19.27	-	24.04	50.57	28.81	34.8	25.91
5	16.21	23.05	17.02	78.86	-	25.69	24.68	30.9	33.68
6	14.17	8.69	24.47	83.15	23.95	-	28.69	30.5	33.79
7	88.27	14.77	23.92	90.58	23.82	25.39	-	44.4	35.7
MW	44.97	25.22	58.73	200.1	55.52	39.76	74.6	71.2	
MW ohne Trainset 1		16.7	31.5	114.4	31.83	31.64	40.25		<u>44.39</u>

Tabelle 13: RUL Time Deviation auf positiven RUL des gewichteten Linearen Modells mit Features und BIC Selektion (3.2.1.12). Der gesamte (ungewichtete) Mittelwert beträgt 44,39 und ist deutlich besser als beim Simplen Modell (116,8). Die Kennwerte vom ersten Reinigungsintervall und die daraus resultierenden Mittelwerte sind in grau gehalten.

Die Ergebnisse in der Tabelle 13 zeigen relativ robuste Prognoseabweichungen, unabhängig davon auf welchem Trainset trainiert wurde. Alleine am zweiten und dritten Intervall zeigen sich mit 69,27 bzw. 67,81 etwas größere Abweichungen im Mittel.

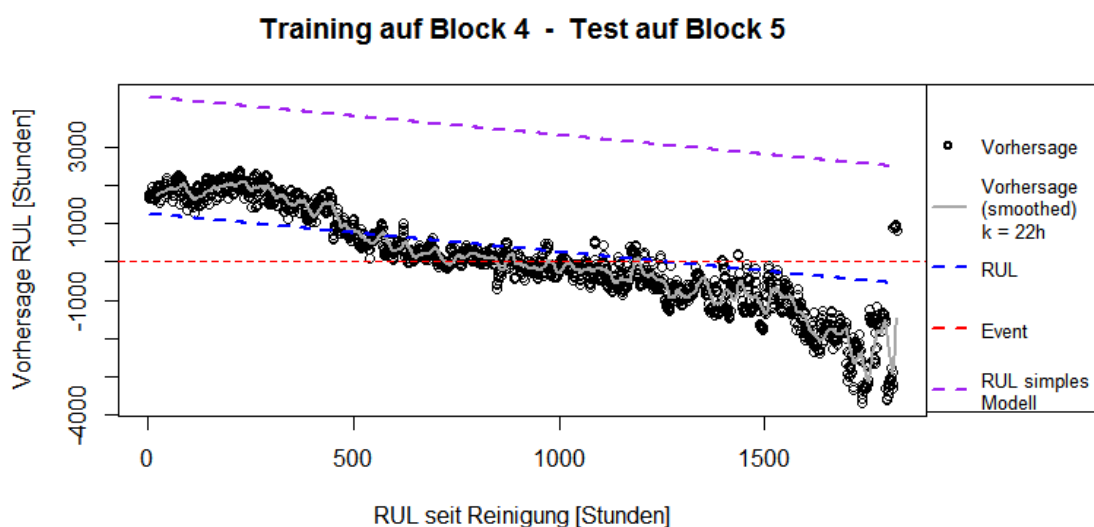


Abbildung 6: Visualisierung der Prognosegüte des gewichteten linearen Modells mit Features und BIC Selektion (3.2.1.12)

Die Abbildung 6 zeigt die Prognosen am fünften Intervall des Modells (3.2.1.12) das auf dem vierten Intervall trainiert wurde. Die tatsächliche RUL ist die blaue Linie. Der Zeitpunkt des Events ist erreicht, wenn die RUL gleich null ist. Dieser Zeitpunkt ist mit der roten Linie markiert. Die optimierte Fensterbreite für die geglätteten RUL Werte wurde mit 22 Stunden berechnet. Im Vergleich zur prognostizierten RUL des simplen Modells (violette Linie) sind die Vorhersagen des linearen Modells wesentlich genauer.

3.3 DAS REGRESSION TREE MODELL

(Breiman, Classification and regression trees, 1984)

Regression Trees versuchen komplexe Zusammenhänge in kleine Partitionen des Datensets aufzuteilen, in denen die Interaktionen handhabbarer werden. Diese Partitionen werden in immer kleinere Subpartitionen zerlegt (Rekursive Partitionierung) (Breiman, Classification and regression trees, 1984), bis man innerhalb der Subpartitionen mit simplen Methoden gute Regressionsmodelle erstellen kann. Durch die hierarchisch aufgebauten Entscheidungsregeln eignen sich Decision Trees hervorragend dazu, die Maschinenzustände besser zu verstehen und zu interpretieren.

Implementierung

Regression Trees (RT) können so viele Entscheidungsknoten erstellen, bis jede Beobachtung in eine eigene Partition eingeteilt wird. Jedoch würde dies in den meisten Anwendungsfällen zum Overfit führen. Der RT Algorithmus wurde für jedes Setup in der Fallstudie wie folgt umgesetzt:

Zunächst wurde ein sehr komplexer Tree berechnet, mit einem *complexity parameter* (*cp*) von 0,001. Dieser Parameter gibt an, um wieviel sich die Güte eines Trees mindestens verbessern muss wenn ein zusätzlicher Entscheidungsknoten erstellt wird. Er wird so niedrig gewählt, dass garantiert wird, dass der Tree mit der niedrigsten kreuzvalidierten Fehlerrate enthalten ist. Danach erfolgt das stutzen des Baumes (Pruning), bei dem Schritt für Schritt wieder Entscheidungsknoten entfernt werden, bis der ermittelte optimale *complexity parameter* erreicht wird. Der optimale *cp* ist die niedrigste beobachtete kreuzvalidierte Fehlerrate beim schrittweisen erstellen des Trees, plus der Standardabweichung der kreuzvalidierten Fehlerrate. Durch dieses Vorgehen soll gewährleistet werden, dass jener Baum entsteht, der die besten Prognoseeigenschaften besitzt.

Verwendete Software und Pakete:

Zum Berechnen der RTs wird die *rpart* Funktion in der R-Library „*rpart*“ verwendet. Zum Pruning der Trees wird die integrierte *prune* Funktion verwendet, wobei der *complexity parameter* (*cp*) wie oben beschrieben gewählt wird.

Variablenselektion

Eine eigene Variablenselektion muss bei Regression Trees in der Regel nicht durchgeführt werden, da in jedem Knotenpunkt alle zur Verfügung stehenden Variablen auf ihren Informationsgehalt geprüft werden.

3.3.1 Die Regression Tree Modelle im Detail

3.3.1.1 Einfaches Regression Tree Modell (RT)

Dieses Modell wird nur mit den Basissignalen trainiert.

3.3.1.2 RT Modell nur auf positiven RUL Zeiten (RT auf positiven RUL)

Dieses Modell wird nur auf Beobachtungen trainiert, die zeitlich vor dem Auftreten des Events liegen. Somit tragen Datenzeilen mit negativen RUL nicht zur Modellbildung bei. Problematisch ist, dass der Regression Tree keine negativen RUL Zeiten prognostizieren kann, wenn er im Trainingsdatensatz keine negativen RUL Werte hat. Deshalb werden die *Event Time Deviation* nie vor dem Event liegen, wenn die RUL Grenze auf null gesetzt wird.

Dieses Modell ist trotzdem sinnvoll, da diese RUL Grenze auch auf positive Werte gesetzt werden kann, um eine Vorlaufzeit für die Wartungsarbeiten zu gewährleisten. Dies wird im Kapitel 1 *Simulation* beschrieben.

3.3.1.3 RT Modell mit verstärktem Lernen gegen Ende der Events (Gewichtetes RT Modell)

Die Gewichtung der einzelnen Beobachtungen erfolgt mit denselben Gewichten wie sie in den linearen Modellen verwendet wurden. Diese lassen sich direkt in den Algorithmus der Regression Trees einpflegen. Die Fehlerabweichungen der Beobachtungen werden mit den Gewichten multipliziert, wodurch ein gezielteres Lernen im zeitlichen Bereich rund um das Event ermöglicht wird. Dieses Modell zeigt als Regressionsmethode der RT die beste Prognosegüten.

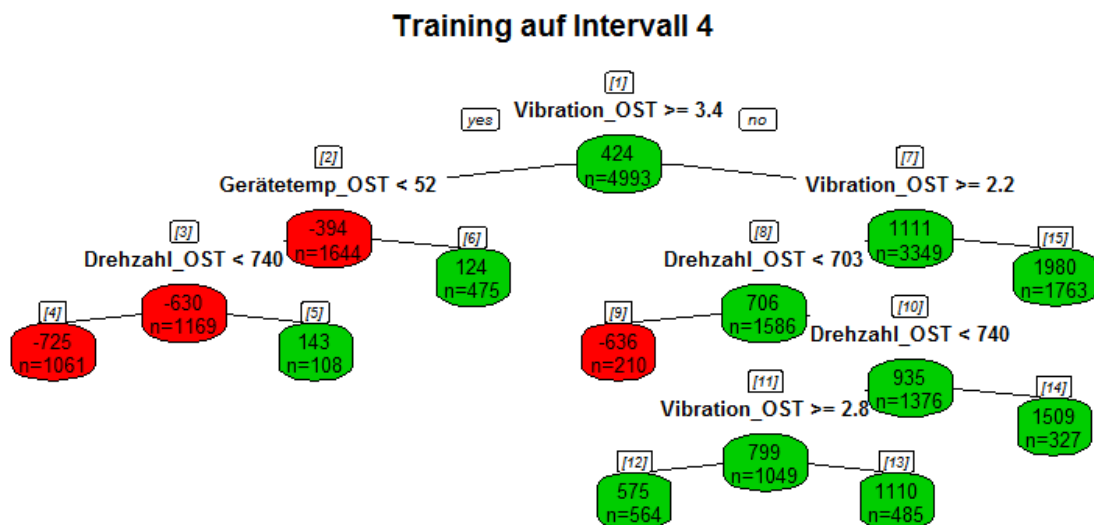


Abbildung 7: Regressionsbaum nach dem Pruning. In jedem Knotenpunkt wird eine Variable auf die beschriebene Bedingung geprüft. Ist diese erfüllt, führt die Entscheidungsregel nach links, sonst nach rechts. Die obere Zahl in den Boxen ist das Prognoseergebnis für die RUL wenn man den Baum bis zu diesem Knoten folgt. Die untere Zahl gibt an, wie viele Beobachtungen beim Anlernen des Baumes in diese Klasse gefallen sind. Die Zahl in der weißen Box steht für den Knotenpunkt. Je wichtiger der Knotenpunkt bei der Konstruktion des Baumes war, desto niedriger ist diese Zahl.

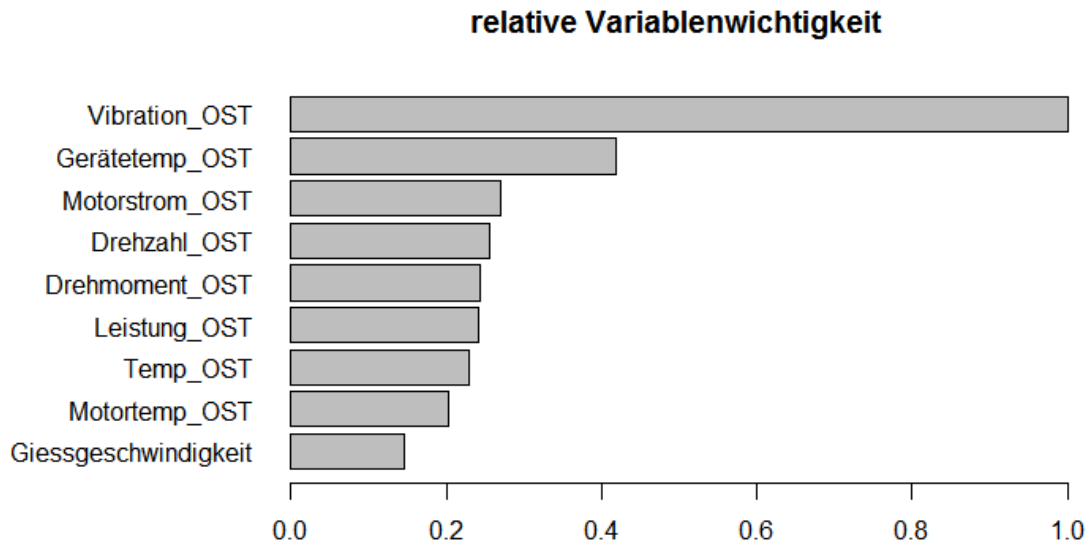


Abbildung 8: Die relative Variablenwichtigkeit des gewichteten Regression Trees zeigt, dass das Signal Vibration_OST bei der Erstellung des Entscheidungsbaumes den größten Einfluss ausübt.

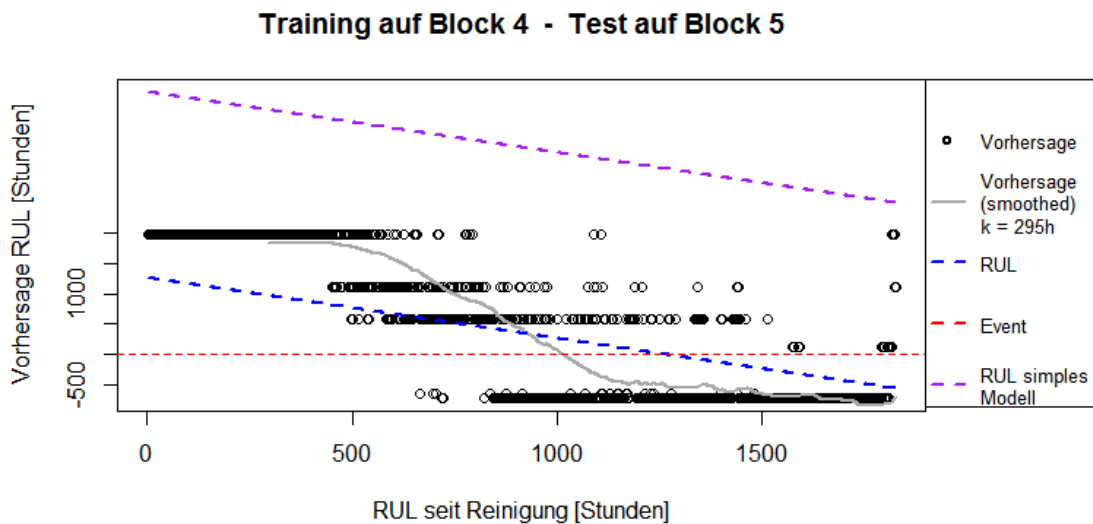


Abbildung 9: Die Auswertung der Prognosen vom auf den 4. Reinigungsintervall trainierten gewichteten RT Modell. Wie schon in der Abbildung ersichtlich ist, prognostiziert dieser Modelltyp die RUL in groben Sprüngen. Dies ist für ein RT Modell typisch, da die Regressionswerte in grobe Klassen eingeteilt werden. Durch starkes Glätten der Vorhersagen entsteht ein kontinuierlicher Verlauf der Prognosen (graue Linie), der für den Anlagenbetreiber wiederum brauchbar wird um den Wartungszeitpunkt abschätzen zu können.

3.3.1.4 RT Modell inklusive abgeleiteten Signalen (RT mit Features)

Dieses Modell beinhaltet alle Signale inklusive der Abgeleiteten. Eine spezielle Variablenselektion wird trotz zusätzlicher abgeleiteten Signalen nicht durchgeführt, da dies ohnehin implizit im RT Algorithmus geschieht.

3.3.1.5 Gewichtetes RT Modell inklusive abgeleiteten Signalen (gewichteter RT mit Features)

Dieses Modell kombiniert das Modell 3.3.1.3 und Modell 3.3.1.4.

CV Modelle

Auf den 5 oben genannten Modelltypen wird je ein Modell erstellt, bei dem eine Variablenselektion nach der beschriebenen CV Methoden durchgeführt wird.

3.3.2 Selektion des Modelles mit der besten Vorhersagegüte

Modell	Geglättete Mean Event Time Deviation [Tage]	Mean RUL Time Deviation [Tage]	Mean RUL Time Deviation auf positiven RULs [Tage]
gewichteter RT (3.3.1.3)	38.21	68.19	54.94
gewichteter RT mit Features CV	41.61	67.52	54.98
gewichteter RT CV	41.73	67.87	55.12
RT auf positiven RUL CV	92.22	56.16	56.19
gewichteter RT mit Features	39.43	69.32	56.48
RT auf positiven RUL	92.22	56.78	56.81
RT	38.01	71.09	58.13
RT mit Features CV	41.71	72.84	58.35
RT CV	37.87	73.03	58.85
RT mit Features	34.22	73.21	59.82

Tabelle 14: Aggregierte Prognoseabweichungen zur Modellwahl.

Das Modell mit der besten Prognosegüte im Detail:

Das Modell *gewichteter RT* (3.3.1.3) hat die beste Prognosegüte in der Kategorie Mean RUL Time Deviation auf positiven RUL, und damit für den Anwendungsfall die relevanteste Voraussetzung für eine gute Langzeitprognose im Livebetrieb. Der optimierte Glättungszeitraum für die Event Time Deviation betrug 295 Stunden. Die Tabelle bis zeigen die Details der Prognosegüte dieses Modells.

Trainset	<i>Test</i> 1	Test 2	Test 3	Test 4	Test 5	Test 6	Test 7	<i>MW</i>	Absoluter MW ohne Test 1
1	-	16.49	61.42	267.4	92.85	97.03	146.8	113.7	
2	0	-	86.05	301.4	49.47	15.09	126	96.33	115.6
3	0	11.83	-	8.43	6.39	-4.19	67.8	16.44	19.73
4	0	11.83	-9.8	-	9.64	-6.6	-107	24.2	29.04
5	0	11.83	-10.9	2.25	-	-2.75	9.99	6.292	7.55
6	0	11.83	122.7	65.78	12.59	-	5.26	36.37	43.64
7	0	11.83	-29.4	0.95	2.52	-23.7	-	11.42	13.7
Abs. MW	0	12.61	53.4	107.7	28.91	24.9	77.2	43.53	
Absoluter MW ohne Trainset 1		11.83	51.8	75.75	16.12	10.47	63.27		<u>38.21</u>

Tabelle 15: Event Time Deviation des gewichteten RT Modells (3.3.1.3)

Trainset	<i>Test</i> 1	Test 2	Test 3	Test 4	Test 5	Test 6	Test 7	<i>MW</i>	MW ohne Test 1
1	-	338.8	178.2	461.8	181	93.49	327.2	263.4	
2	58.62	-	81.6	124.1	68.23	52.5	83.97	78.17	82.08
3	19.64	202.2	-	152.9	32.71	32.88	57.8	83.02	95.69
4	21.14	81	38.07	-	30.05	35.96	39.54	40.96	44.92
5	8.31	108.2	11.27	69.83	-	19.74	21.82	39.86	46.17
6	17.89	165.3	44.17	161.6	52.07	-	78.04	86.52	100.2
7	19.39	83.44	16.01	64.78	16.64	19.16	-	36.57	40.01
MW	24.16	163.2	61.55	172.5	63.46	42.29	101.4	89.79	
MW ohne Trainset 1		128	38.22	114.6	39.94	32.05	56.23		<u>68.19</u>

Tabelle 16: RUL Time Deviation des gewichteten RT Modells (3.3.1.3)

Trainset	<i>Test 1</i>	Test 2	Test 3	Test 4	Test 5	Test 6	Test 7	<i>MW</i>	MW ohne Test 1
1	-	61.53	220.6	591	243.4	147.1	335.8	266.6	
2	139.5	-	80.15	153.3	73.44	52.11	84.76	97.21	88.75
3	16.1	18.67	-	175.5	37.12	35.9	58.91	57.03	65.21
4	32.17	28.08	43.72	-	33.6	56.34	40.16	39.01	40.38
5	4.66	11.06	11.69	79.03	-	27	22.03	25.91	30.16
6	42.93	13.91	52.4	196.9	66.62	-	79.5	75.38	81.86
7	7.11	5.82	15.74	71.65	13.89	9.21	-	20.57	23.26
MW	40.41	23.18	70.72	211.2	78.01	54.61	103.5	83.1	
MW ohne Trainset 1		15.51	40.74	135.3	44.93	36.11	57.07		<u>54.94</u>

Tabelle 17: RUL Time Deviation auf positiven RUL des gewichteten RT Modells (3.3.1.3)

Die Regression Trees sind sehr leicht zu Parametrisieren und zeigen ein Regelwerk wie sie die Prognosewerte berechnen. Dies macht sie sehr praktikabel, um die Zusammenhänge der Daten besser zu verstehen. Zudem können nichtlineare und komplexe Regressionen modelliert werden. Weiters ist es nicht unbedingt notwendig eine Variablenselektion durchzuführen, da dies implizit im Algorithmus geschieht. Einen kleinen Wehrmutstropfen liefern die ungewünschten Sprunghaften Prognosewerte, welche erst durch eine geeignete Glättung im Echtzeitbetrieb brauchbar werden (siehe Abbildung 9).

3.4 DAS RANDOM FOREST REGRESSION MODELL

(Breiman, Random Forests, 2001)

Die Random Forest Regression ist den Regression Trees ähnlich. Mit dem Unterschied, dass beliebig viele Regression Trees erstellt werden und wie es bei den Random Forests üblich ist wird in jedem Entscheidungsknoten ein zufälliger Pool von Variablen ausgewählt. Von diesen Variablen wird jene für den Entscheidungsknoten herangezogen, welche die Varianz für die Zielvariable am besten verringern kann. Dies wird so fortgeführt, bis keine Trennung mehr möglich ist. Die Anzahl wieviele Variablen zufällig gewählt werden, (in jedem Knoten kommen alle Variablen mit gleicher Gewichtung in Frage) hängt von der Anzahl der Einflussvariablen ab. Es wurde mit Rücksicht auf einfache Parametrisierung der Vorschlag von Breiman (Breiman, Classification and regression trees, 1984) mit $p/3$ bei Regressionsproblemen gewählt (wenn p die Anzahl der Einflussvariablen ist).

Um eine Reproduzierbarkeit zu gewährleisten wurde die `set.seed()` Funktion in R auf 123 gesetzt. Dadurch wird beim Trainieren des RF immer der gleiche Zufallsalgorithmus mit demselben Startwert verwendet. Somit kann für jedes Setup der exakt gleiche RF wiederhergestellt werden.

Verwendete Software und Pakete:

Die Random Forests werden mit der R-Library „*ranger*“ erstellt. In der *ranger* Funktion ist eine Option integriert, die eine direkte Gewichtung für jede Beobachtung zulässt (dies ist in der Standardlibrary „*randomForest*“ nicht möglich). Des Weiteren benötigt dieses Paket weniger Berechnungszeit als die *randomForest* Funktion.

Variablenselektion

Eine eigene Variablenselektion ist nicht zwingend erforderlich, da dies implizit durch den Algorithmus geschieht, indem in jedem Knotenpunkt die beste Variable zur Klassifizierung gewählt wird.

3.4.1 Die Random Forest Regression Modelle im Detail

Das Setup der Random Forests erfolgt analog zu den Regression Trees.

3.4.1.1 *Random Forest Modell nur mit den Basissignalen (RF)*

Dieses Modell wird nur mit den Basissignalen trainiert.

3.4.1.2 *Random Forest Modell nur auf positiven RUL (RF auf positiven RUL)*

Dieses Modell wird nur auf Beobachtungen trainiert, die zeitlich vor dem Auftreten des Events liegen.

3.4.1.3 Gewichtetes Random Forest Modell (gewichteter RF)

Die Gewichtung der einzelnen Beobachtungen erfolgt mit denselben Gewichten wie sie in den linearen Modellen verwendet wurden. Diese können als Option direkt in den *ranger* Algorithmus eingegeben werden.

3.4.1.4 Random Forest Modell inklusive abgeleitete Signale (RF mit Features)

Dieses Modell beinhaltet alle Signale inklusive der Abgeleiteten.

3.4.1.5 Gewichtetes Random Forest Modell inklusive abgeleiteten Signalen (gewichteter RF mit Features)

Dieses Modell kombiniert das Modell 3.4.1.3 und Modell 3.4.1.4.

Random Forest CV Modelle

Auf den 5 oben genannten Modelltypen wird je ein Modell erstellt, bei dem eine Variablenselektion nach der beschriebenen CV Methoden durchgeführt wird.

3.4.2 Selektion des Modelles mit der besten Vorhersagegüte

Modell	Geglättete Mean Event Time Deviation [Tage]	Mean RUL Time Deviation [Tage]	Mean RUL Time Deviation auf positiven RULs [Tage]
gewichteter RF CV (3.4.1.3 CV)	56.3	68.8	56.51
gewichteter RF mit Features CV	57.96	69.73	56.91
RF auf positiven RUL	91	57.25	57.27
RF auf positiven RUL CV	88.31	57.69	57.71
gewichteter RF	44.99	71.28	57.98
RF CV	58.52	71.11	58.26
gewichteter RF mit Features	59.9	73.7	58.68
RF mit Features CV	63.25	72.67	58.87
RF	51.08	73.76	60.16
RF mit Features	60.42	75.66	60.58

Tabelle 18: Aggregierte Prognoseabweichungen zur Modellwahl.

Das gewichtete Random Forest Modell auf dem eine Variablenselektion mit der CV-Methode durchgeführt wurde, hat den niedrigsten Wert in der Kategorie Mean RUL Time Deviation auf positiven RULs, und gilt damit als das am besten geeignete Modell zur Vorhersage des Reinigungszeitpunktes.

Trainset	Test 1	Test 2	Test 3	Test 4	Test 5	Test 6	Test 7	MW	Absoluter MW ohne Test 1
1	-	16.49	61.42	267.4	92.85	97.03	146.8	113.7	
2	0	-	86.05	301.4	49.47	15.09	126	96.33	115.6
3	0	11.83	-	8.75	0	-3.07	-0.94	4.098	4.91
4	0	11.83	120.4	-	13.9	-3.07	-0.99	25.03	30.04
5	0	11.83	116.9	9.61	-	-0.02	-11.5	24.97	29.96
6	0	11.83	132.7	65.78	13.84	-	7.68	38.64	46.37
7	-180	-168	-16	-54.2	-151	-164	-	122.4	110.9
<i>Abs.MW</i>	30	38.66	88.91	117.9	53.57	47.17	48.99	60.74	
Absoluter MW ohne Trainset 1		43.1	94.4	87.95	45.71	37.19	29.42		<u>56.3</u>

Tabelle 19: Event Time Deviation des gewichteten RF Modells mit CV Selektion (3.4.1.3 CV)

Trainset	Test 1	Test 2	Test 3	Test 4	Test 5	Test 6	Test 7	MW	MW ohne Test 1
1	-	352.3	171.6	448.4	171.2	83.37	308.7	255.9	
2	57.66	-	88.26	125.4	71.6	49.87	87.43	80.04	84.51
3	31.71	203.8	-	162.5	36.89	31.68	61.52	88.03	99.29
4	26.99	67.84	22.71	-	25.24	33.43	35.38	35.27	36.92
5	12.06	100.2	9.53	83.59	-	15.78	28.75	41.65	47.56
6	18.96	163.1	50.94	164	56.4	-	81.63	89.19	103.2
7	22.61	87.25	15.73	63.89	15.63	23.87	-	38.16	41.27
MW	28.33	162.4	59.79	174.6	62.83	39.67	100.6	89.75	
MW ohne Trainset 1		124.5	37.43	119.9	41.15	30.93	58.94		<u>68.8</u>

Tabelle 20: RUL Time Deviation des gewichteten RF Modells mit CV Selektion (3.4.1.3 CV)

Trainset	Test 1	Test 2	Test 3	Test 4	Test 5	Test 6	Test 7	MW	MW ohne Test 1
1	-	58.18	212.6	569.4	232.7	126.8	316.6	252.7	
2	109.5	-	87.86	155.6	78.96	52.1	88.25	95.37	92.55
3	4.3	13.05	-	187	39.26	24.4	63.13	55.18	65.36
4	43.8	18.04	24.47	-	26.63	60.01	35.6	34.76	32.95
5	1.48	16.17	10.18	100.6	-	19.03	29.37	29.47	35.07
6	46.24	18.66	61.93	202.4	73.66	-	83.52	81.08	88.04
7	6.42	18.42	15.74	68.8	11.75	10.83	-	21.99	25.11
MW	35.29	23.75	68.79	214	77.15	48.86	102.7	81.51	
MW ohne Trainset 1		16.87	40.04	142.9	46.05	33.27	59.97		<u>56.51</u>

Tabelle 21: RUL Time Deviation auf den positiven RULs des gewichteten RF Modells mit CV Selektion (3.4.1.3 CV)

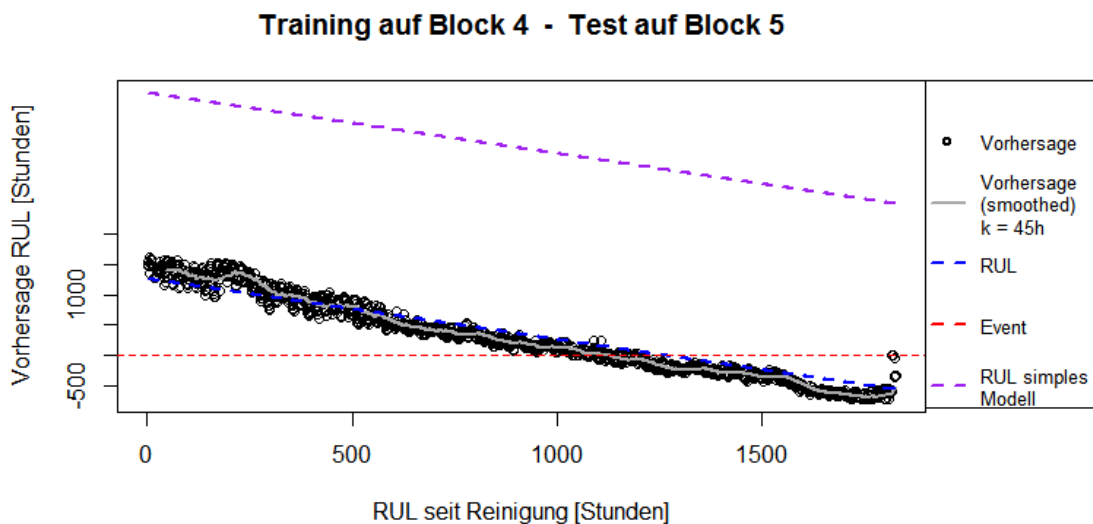


Abbildung 10: Prognosegüte des Random Forest Modells (3.4.1.3 CV).

Die Prognosen des besten Random Forest Modells zeigen, dass sie stark der tatsächlichen RUL folgen. Die Schwankungen der RUL können zwar mehrere hundert Stunden variieren, dies kann aber durch eine geeignete Glättung (hier gleitender Mittelwert von 45 Stunden) ausgeglichen werden. Die Prognoseabweichungen auf dem zweiten, dritten und sechsten

Intervall sind relativ groß. Auf den übrigen Intervallen weichen diese bei weitem nicht so stark ab. Dies könnte bedeuten, dass die RF Modelle nicht besonders robust für diese Art von RUL Schätzungen sind.

Das Random Forest Modell (3.4.1.3 CV) im Detail:

Type:	Regression
Number of trees:	500
Sample size:	14554
Number of independent variables:	8
Mtry:	2
Target node size:	5
Variable importance mode:	impurity
OOB prediction error:	243385
R squared:	0.8554

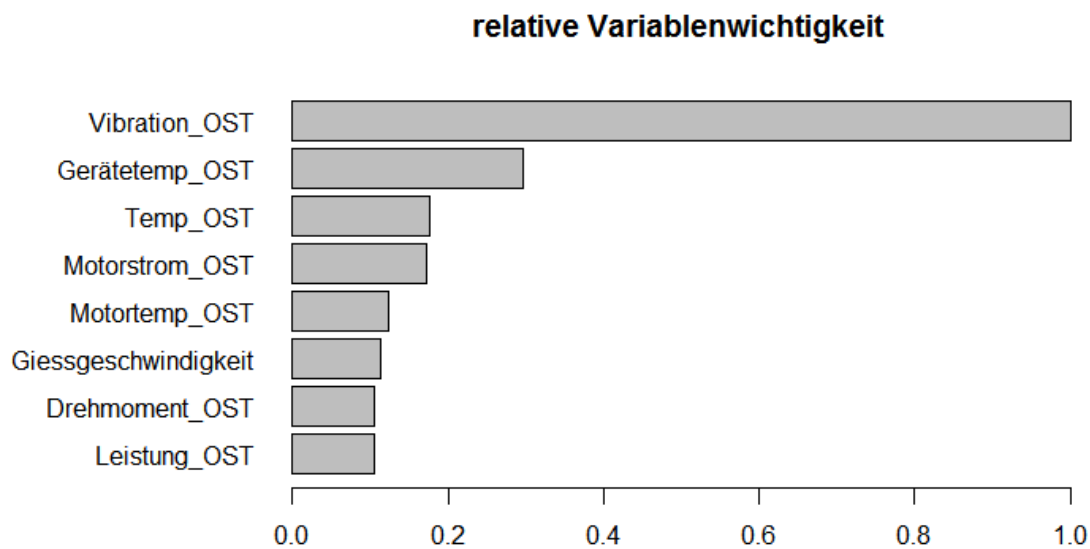


Abbildung 11: Die relative Variablenwichtigkeit² des Random Forest Modells (3.4.1.3 CV) zeigt ein ähnliches Bild wie die Regression Trees. Auch hier wird das Regelwerk maßgeblich von den Vibrationswerten bestimmt.

Die Random Forest Modelle sind relative einfach zu parametrisieren, da sie sehr robust gegenüber den Parametern sind. Außerdem können sie sehr komplexe Zusammenhänge modellieren ohne eine gezielte Variablenselektion durchführen zu müssen. Es wurde die Anzahl an Trees mit 500 gewählt, da die Parameterplots gezeigt haben, dass sich bei den meisten Modellen zwischen 100 und 200 Trees schon ein Sättigungseffekt des Lernens eingestellt hatte. Wenn mehr als die optimale Anzahl an Trees trainiert werden, kommt es

² Die Bewertung der Variablenwichtigkeit erfolgte nach dem Kriterium der Impurity, also der durchschnittlichen Reduktion der Varianz beim Verwenden einer Variable (Details siehe R Package „ranger“).

zwar zu keinem Overfit, jedoch werden keine neuen Informationen erlernt. Dies führt also nur zu einer erhöhten Rechenzeit. Durch die hohe Anzahl an trainierten Trees, sind die erstellten Modelle als Blackbox einzustufen, was eine direkte Interpretationsmöglichkeit sehr schwierig machen kann.

3.5 GENERAL ADDITIVE MODELS (GAM)

(Hastie & Tibshirani, General Additive Models, 1986) (O'Sullivan, 1986)

Mit den GAM wird die Response Variable modelliert als Kombination von Glättungsfunktionen die von den Regressoren abhängig sind. (Ruppert, 2003) Als Glättungsfunktionen werden zumeist penalisierte polynomiale Splines oder Smoothing Splines (Reinsch, 1967) (Silverman, 1985) (Fahrmeier, 2009) verwendet.

Implementierung

Für jede Beobachtung i (Datenzeile im diskretisierten Datensatz) lässt sich das GA Modell wie folgt anschreiben:

$$L_i = s_1(X_{i1}) + s_2(X_{i2}) + \dots + s_k(X_{ik}) + e_i$$

Für jede Variable X wird eine eigene Glättungsfunktion $s(\cdot)$ geschätzt. Die linearen Kombinationen dieser geschätzten Funktionen dienen wie beim linearen Modell als Einflussvariablen für die Zielvariable. In dieser Studie kommen sechs Glättungsfunktionen zum Einsatz:

Cubic Regression Splines (cr), *Cubic Shrinkage Regression Splines (cs)*, *Thin Plate Regression Splines (tp)*, *Thin Plate Shrinkage Regression Splines (ts)*, *Duchon Splines (ds)*, *P-Splines (ps)* (Wood, Generalized additive models: an introduction with R, 2006)

Die Glättungsfunktionen können bei den erstellten Modellen mittels eines Komplexitätsparameters gesteuert werden. Durch ein integriertes Cross Validation Verfahren in der *gam* Funktion der *mgcv* Library, kann man den Splinetypen und den Glättungsparameter auch automatisiert auswählen lassen. Für jeden Spline Typen wurden die Modelle mit folgenden Glättungsparameter K berechnet: 3, 5, 7, 10, 15 sowie einer automatischen Parameterwahl. In den Ergebnissen werden jene Modelle aufgelistet, welche die besten Prognoseergebnisse in der Kategorie RUL Time Deviation auf positiven RUL lieferten.

Verwendete Software und Pakete:

Die GAM werden mit der R-Library „mgcv“ (Wood, Modelling and Smoothing Parameter Estimation with Multiple Quadratic Penalties, 2000) erstellt. In der *gam* Funktion ist eine Option integriert, die eine direkte Gewichtung für jede Beobachtung zulässt. Des Weiteren sind die oben genannten Glättungsfunktionen bereits implementiert.

Variablenselektion

Für GAMs gibt es Selektionsverfahren, die den Stepwise-Methoden bei den linearen Modellen entsprechen. Jedoch muss für jeden Glättungsterm $s(\cdot)$ der Freiheitsgrad bzw. die Komplexität festgelegt werden. Sonst kann eine Variable durch zunehmende Komplexität einen so guten Fit erzeugen, dass diese zur wichtigsten Variable im Modell wird. Da oben festgelegt wurde, dass das Modell selbst die optimale Komplexität festlegen soll bzw. von mir vorgegeben wird, kann also diese Stepwise-Methode nicht angewandt werden. Daher wird die CV-Methode, welche die Variablenwichtigkeit anhand der Validierungssets festlegt angewandt. Denn diese Methode greift nicht in die Bestimmung der Komplexität des Modells ein, sondern evaluiert lediglich die Prognosegüte.

3.5.1 Die GAM Regression Modelle im Detail

Die GAM Modelle wurden zunächst mit derselben Struktur trainiert, wie sie in den drei bisherigen Methoden beschrieben wurde. Dabei wurde die Parametrisierung so festgelegt, dass die jeweilige Modellkomplexität (bzw. Glättungsgrad) vom implementierten Algorithmus selbst gewählt wird. Zusätzlich wurden durch *Grid Search* Modelle erstellt bei denen die Parametrisierung manuell vorgenommen wurde. Es wurden sechs verschiedene Glättungsmethoden mit je fünf Komplexitätsgraden verwendet. Für jede dieser 31 Kombinationen (inklusive Modell mit automatischer Parameterwahl) wurden die Modelle 3.5.1.1 bis 3.5.1.5 sowohl mit als auch ohne CV-Variablenselektion berechnet und ausgewertet. Somit gibt es für jede Parameterwahl je zehn Modelle, wobei jeweils jenes Modell in die Ergebnisliste aufgenommen wird, welches in der entsprechenden Kategorie die niedrigste Abweichung für die „RUL Time Deviation auf den positiven RULs“ hatte.

3.5.1.1 *GAM Modell nur mit den Basissignalen (GAM)*

In dieser Kategorie wird das Modell nur mit den Basissignalen trainiert.

3.5.1.2 *GAM Modell nur auf positiven RUL (GAM auf positiven RUL)*

Dieses Modell wird nur auf Beobachtungen trainiert, die zeitlich vor dem Auftreten des Events liegen.

3.5.1.3 *Gewichtetes GAM Modell (gewichtetes GAM)*

Die Gewichtung der einzelnen Beobachtungen erfolgt mit denselben Gewichten wie sie in den linearen Modellen verwendet wurden. Diese können als Option direkt in den *gam* Algorithmus eingegeben werden.

3.5.1.4 *GAM Modell inklusive abgeleitete Signale (GAM mit Features)*

In dieser Kategorie beinhaltet das Modell alle Signale inklusive der Abgeleiteten.

3.5.1.5 *Gewichtetes GAM Modell inklusive abgeleiteten Signalen (gewichtetes GAM mit Features)*

Dieses Modell kombiniert das Modell 3.5.1.3 und Modell 3.5.1.4.

GAM CV Modelle

Auf den 5 oben genannten Modelltypen wird je ein Modell erstellt, bei dem eine Variablenselektion nach der beschriebenen CV Methoden durchgeführt wird.

3.5.2 Selektion des Modelles mit der besten Vorhersagegüte

Für jede der zehn Kategorien wird jenes Modell gewählt welches den niedrigsten Wert in der *Mean RUL Time Deviation auf positiven RUL* hat, und in der Tabelle 22 mit den Modellen aus den anderen Kategorien verglichen.

Modell	Geglättete Event Time Deviation [Tage]	Mean RUL Time Deviation [Tage]	Mean RUL Time Deviation auf positiven RULs [Tage]	K	Spline
gewichtetes GAM mit Features CV (3.5.1.5 CV)	30.6	63.22	46.94	3	'cs'
gewichtetes GAM	37.66	61.82	48.99	3	'cs'
gewichtetes GAM CV	30.68	67.57	49.64	3	'cr'
GAM	34.53	66.61	50.53	3	'ts'
GAM auf positiven RUL	52.93	32.36	52.41	7	'ts'
GAM CV	38.71	71.53	54.38	3	'cs'
GAM auf positiven RUL CV	NA	55.59	55.59	3	'ts'
GAM mit Features CV	44.88	71.51	58.01	3	'cr'
gewichtetes GAM mit Features	55.87	73.77	55.14	3	'ts'
GAM mit Features	50.34	88.7	69.52	3	'cr'

Tabelle 22: Aggregierte Prognoseabweichungen zur Modellwahl.

Das Modell mit den besten Prognoseergebnissen auf den Zeiträumen vor dem Event ist das *gewichtete GAM mit Features CV*. Die Parametrisierung dieses Modells benutzt *Cubic Shrinkage Regression Splines* mit einem Komplexitätswert von $k=3$.

Das *Gewichtete SVR mit Features CV* Modell (3.5.1.5 CV) im Detail (trainiert am gesamten Datensatz Ost inklusive Features):

Family: gaussian
Link function: identity

Parametric coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	188.46	6.46	29.2	<2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Approximate significance of smooth terms:

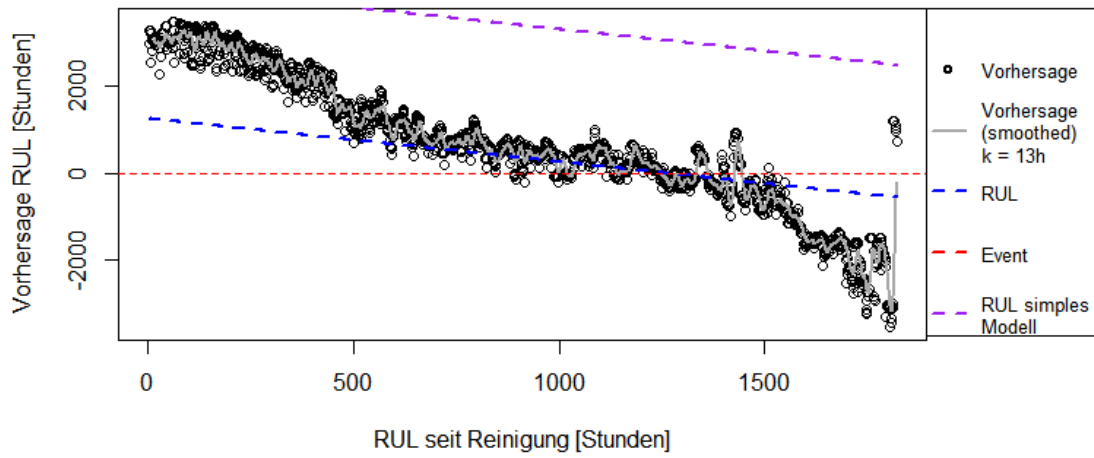
	edf	Ref.df	F	p-value
s(Vibration_OST)	2.00	2.00	4060.4	<2e-16 ***
s(Leistung_OST_kurtosis)	1.06	1.24	11.8	3e-04 ***
s(Gerätetemp_OST_sd)	1.96	2.00	99.7	<2e-16 ***
s(Motortemp_OST_sDes1)	1.99	2.00	405.3	<2e-16 ***
s(Drehzahl_OST)	1.98	2.00	194.0	<2e-16 ***
s(Motorstrom_OST_kurtosis)	2.00	2.00	69.7	<2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R-sq.(adj) = 0.441 Deviance explained = 44.3%
GCV = 5.1127e+05 Scale est. = 5.1079e+05 n = 12856

Die besten drei GAM Modelle verwenden die Gewichtung für den relevanten Zeitraum. Dies zeigt, dass dies bei den GAM Modellen zu besseren Ergebnissen führen wird. Zudem zeigt die Variablenselektion mit der CV Methode eine wesentlich bessere Prognosegüte beim *gewichteten GAM mit Features CV* als bei dem Pendant ohne CV Selektion. Dies zeigt wiederum, dass GAM mit vielen Variablen zu Overfitting neigen. Die Modelle mit dem niedrigsten Komplexitätsgrad von 3, liefern in allen bis auf einer Kategorie die besten Prognosen. Dies deutet darauf hin, dass es bei höheren Komplexitäten bereits zum Übertrainieren kommt.

Training auf Block 4 - Test auf Block 5



Trainset	Test 1	Test 2	Test 3	Test 4	Test 5	Test 6	Test 7	MW	Absoluter MW ohne Test 1
1	-	0.23	17.18	74.73	16.36	14.65	4.04	21.2	
2	-3.04	-	85.4	301.3	50.2	25.86	13.3	79.85	95.21
3	-3.04	11.82	-	8.77	18.43	-0.85	0.79	7.283	8.132
4	-3.04	11.82	0	-	9.63	4.9	0.91	5.05	5.452
5	-3.04	11.82	0	-3.87	-	-0.83	-113	22.11	25.93
6	-3.04	11.26	122.3	59.82	4.96	-	-0.68	33.68	39.81
7	-37.3	10.92	-9.8	16.09	1.33	-7.22	-	13.77	9.072
<i>Abs.MW</i>	8.74	9.645	39.12	77.43	16.82	9.052	22.14	26.14	
Absoluter MW ohne Trainset 1		11.53	43.5	77.97	16.91	7.932	25.76		<u>30.6</u>

Tabelle 23: Event Time Deviation des gewichteten GAM mit Features CV (3.5.1.5 CV).

Trainset	<i>Test 1</i>	Test 2	Test 3	Test 4	Test 5	Test 6	Test 7	<i>MW</i>	MW ohne Test 1
1	-	484.9	135.6	473.5	142.8	52.41	241.7	255.2	
2	109	-	88.53	124.5	96.81	103.6	81.78	100.7	99.04
3	19.53	201.3	-	143.3	20.23	30.91	40.99	76.04	87.34
4	15.61	77.8	38.44	-	37.85	53.74	55.18	46.44	52.6
5	12.68	106.1	7.81	68.98	-	16.67	19.21	38.58	43.75
6	7.23	75.48	15.71	69.64	17.18	-	33.48	36.45	42.3
7	28.16	115.5	17.03	87.51	23.31	27.91	-	49.91	54.26
MW	32.04	176.8	50.53	161.2	56.37	47.54	78.73	86.18	
MW ohne Trainset 1		115.2	33.5	98.79	39.08	46.56	46.13		<u>63.22</u>

Tabelle 24: RUL Time Deviation des gewichteten GAM mit Features CV (3.5.1.5 CV).

Trainset	<i>Test 1</i>	Test 2	Test 3	Test 4	Test 5	Test 6	Test 7	<i>MW</i>	MW ohne Test 1
1	-	35.81	154.1	575.4	186	41.31	242.7	205.9	
2	105.1	-	76.72	145.5	82.31	52.41	81.73	90.64	87.74
3	5.64	22.3	-	163.1	22.31	43.17	41.21	49.63	58.42
4	35.44	20.53	40.95	-	37.25	61.78	55.48	41.91	43.2
5	5.78	5.89	8.82	77.46	-	25.26	19.31	23.75	27.35
6	3.3	5.49	19.46	82.41	22.43	-	33.64	27.79	32.69
7	5.62	15.33	12.8	96.38	18.68	18.03	-	27.81	32.24
MW	26.82	17.56	52.14	190	61.5	40.33	79.01	66.77	
MW ohne Trainset 1		13.91	31.75	113	36.6	40.13	46.27		<u>46.94</u>

Tabelle 25: RUL Time Deviation auf den positiven RULs des gewichteten GAM mit Features CV (3.5.1.5 CV).

Durch die Glättungsterme können nichtlineare Zusammenhänge gut modelliert werden. Jedoch zeigt sich auch, dass es bei schlechter Parametrisierung schnell zu einem Overfit kommen kann. Die Resultate zeigen bis auf dem zweiten Reinigungsintervall recht zuverlässige Prognosen. Dies deutet darauf hin, dass wenn die Glättungsterme richtig gewählt werden, robuste Modelle möglich sind. Die Interpretation dieser Terme ist im Allgemeinen schwierig, da sich diese nicht in menschlich verständliche Formeln fassen

lassen. Der hohe Parametrisierungsaufwand und die langen Berechnungszeiten sind ein weiterer Nachteil.

3.6 DAS SUPPORT VECTOR REGRESSION MODELL

(Drucker, Burges, Kaufman, Smola, & Vapnik, 1997) (Schölkopf, Burges, & Vapnik, Incorporating Invariances in support vector learning machines, 1996) (Vapnik, Golowich, & Smola)

Die Grundidee der SVR ist, die Daten mittels einer nichtlinearen Abbildung in einen höherdimensionalen Raum zu transformieren und auf diesen eine Lineare Regression durchzuführen (Müller, et al., 1997). Somit entspricht die lineare Regression im höherdimensionalen Raum einer nichtlinearen Regression im ursprünglichen Raum.

Für die SV-Regression ergibt sich das Minimierungsproblem (Meyer, 2015):

$$\min_{\alpha, \alpha^*} \frac{1}{2} (\alpha - \alpha^*)' Q (\alpha - \alpha^*) + \epsilon \sum_{i=1}^l (\alpha_i + \alpha_i^*) + \sum_{i=1}^l y_i (\alpha_i + \alpha_i^*) \quad (1)$$

$$\text{sodass } 0 \leq \alpha_i, \alpha_i^* \leq C, i = 1, \dots, l \quad (2)$$

$$\text{und } \sum_{i=1}^l (\alpha_i - \alpha_i^*) = 0 \quad (3)$$

$$\text{wobei: } Q_{ij} = y_i y_j K(x_i, x_j) \quad (4)$$

Dabei ist Q eine positive semidefinite $l \times l$ Matrix und $K(x_i, x_j)$ der Kernel. C ist die obere Schranke für die Verlustfunktion. Dieser *cost* Parameter muss größer als 0 gewählt werden. Das Minimierungsproblem wird lösbar, indem der „Kernel Trick“ (Schölkopf & Smola, Learning with kernels: support vector machines, regularization, optimization, and beyond, 2002) angewandt wird. Die Berechnungszeiten der SV-Regression sind im Vergleich zu den bisherigen Methoden um ein Vielfaches länger. Eine effiziente Umsetzung des Algorithmus ist in der R-library e1071 (Meyer, 2015) enthalten, welche auch die folgenden Kernels integriert hat.

Kernel	Formel
Linear	$u'v$
Polynomial	$\gamma(u'v + c_0)^d$
Radial Basis	$e^{-\gamma u-v ^2}$

Zunächst wurde eine Voranalyse auf 10 Prozent der Daten gemacht, welche zufällig ausgewählt wurden. Ziel dieser Voranalyse war es, ein sinnvolles Parameterset für ein *Grid Search Tuning* erstellen zu können (Widodo & Yang, Support vector machine in machine condition monitoring and fault diagnosis, 2007). Mit dieser Datenstichprobe wurden Modelle mit folgenden Setting erstellt:

Paramter	Wertebereich	Verwendung bei Kernel
Cost C	4, 8, 16, 32, 64, 128, 256, 512	Linear, Polynomial, Radial
ε	0, 0.02, 0.04, 0.06, 0.08, 0.1, 0.12, 0.14, 0.16, 0.18, 0.2	Linear, Polynomial, Radial
γ	0.01, 0.11, 0.2, 0.8	Polynomial, Radial
degree	1, 2, 3, 4, 5, 7, 15, 20	Polynomial
Coef0	±0.1, ±0.5, ±2, ±10	Polynomial

Dadurch ergaben sich $3 * 8 * 11 * 4 = 1056$ Modelle. Als sinnvolles Parametersetting wurden jene Modelle betrachtet, welche auf dem nicht zum Training verwendeten Datensampling eine möglichst gute Prognosegüte vorweisen konnten. Um eine vertretbare Durchlaufzeit zu gewährleisten wurden nur die besten 12 Settings für die weitere Modellerstellung zugelassen. Wie in Abbildung 12 ersichtlich ist, sollte ein „*radial kernel*“ gewählt werden. Außerdem tendieren Modelle mit *Cost*-Werten von 4 bis 64 und *Epsilon* Werten von 0 bis 0.1 zu guten Prognoseergebnissen. Abbildung 13 zeigt, dass der *Gamma*-Parameter um 0.2 herum gewählt werden sollte. Beim *polynomialen* Kernel wurden zunächst der *degree* Parameter auf 3 gesetzt, und der *coef0* auf 0 (Standardeinstellung in e1071 library). Dies führte zu konstant schlechten *mean absolute error (mae)* Raten von ca. 730 Stunden. Das Tuning speziell für diese beiden Parameter ergab im besten Fall eine *mae* von 578 Stunden. Somit wurde der *polynomial* Kernel, egal mit welcher Parametrisierung, vom *Radial Basis* Kernel dominiert. Die SV-Regressionen mit *linearem* Kernel werden nur mittels *Cost* und *Epsilon* parametrisiert, und erzielen unabhängig davon relativ ähnliche Prognoseabweichungen, welche alle vom *Radial Basis* Kernel dominiert werden. Die Voranalyse macht deutlich, dass der *Radial Basis* Kernel der am besten geeignete für die Modelle ist.

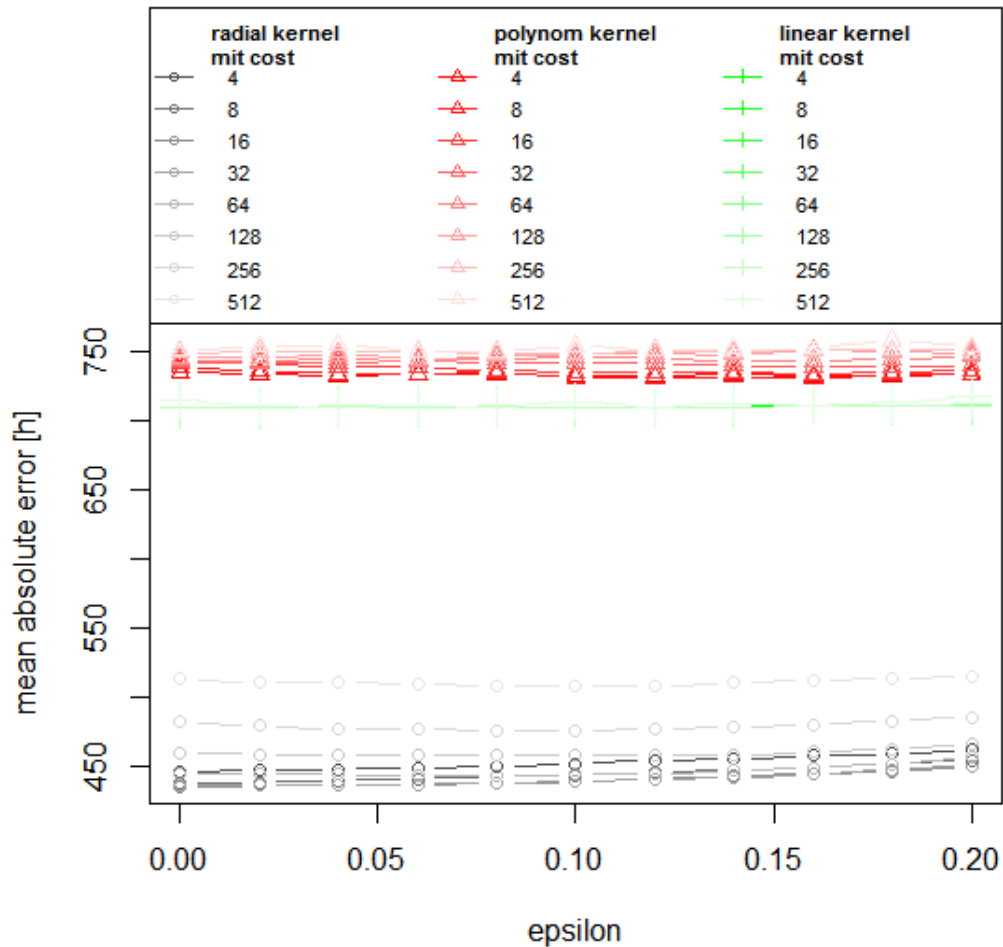


Abbildung 12: Absolute Abweichungen der Prognosen auf dem Validierungsset (nicht zum trainieren verwendeten Datensatz). Die Voranalyse zeigt, dass eine SVR mit radialem Kernel die besten Prognosen liefern sollte. Zudem sollte der optimale Cost-Parameter zwischen 4 und 128 liegen. Die Abhängigkeit vom Epsilon-Parameter ist sehr gering, sollte aber niedriger als 0,15 gewählt werden, da ab diesem Schwellwert die Prognoseabweichungen zu steigen beginnen.

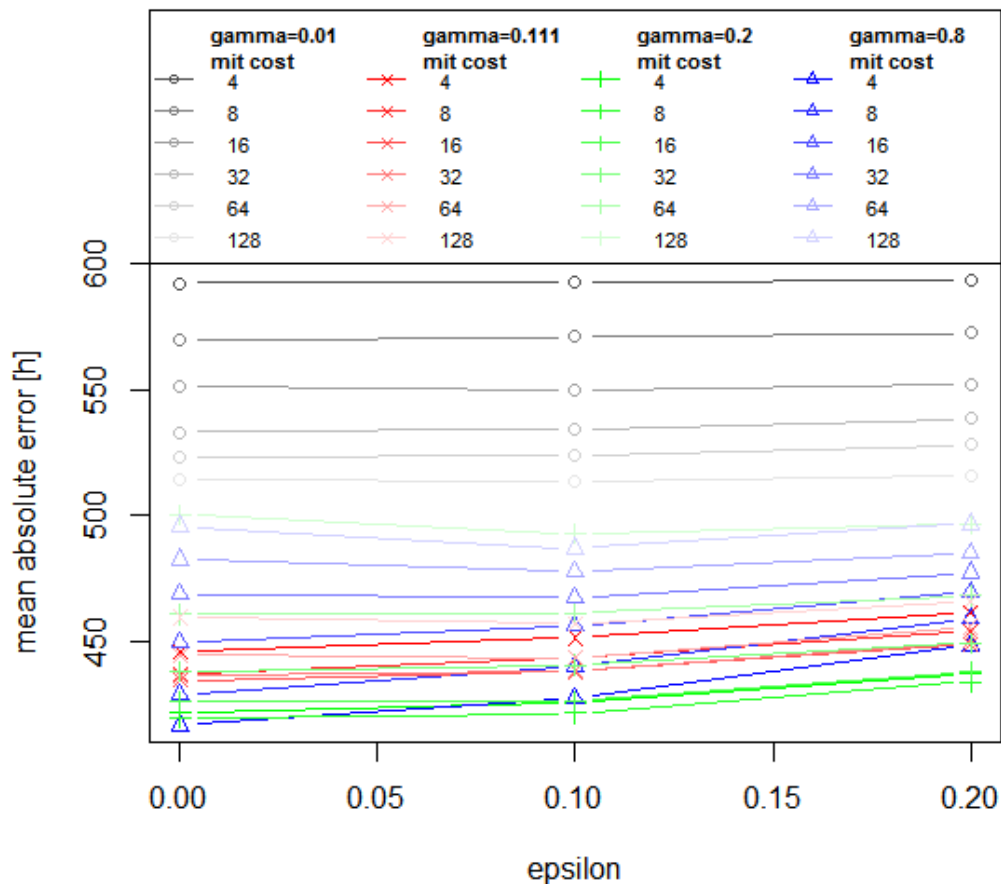


Abbildung 13: Absolute Abweichungen der Prognosen auf dem Validierungsset (nicht zum trainieren verwendeten Datensatz). Diese Voranalyse zeigt die Abhängigkeit der Prognoseabweichungen vom Parameter Gamma, zu unterschiedlichen Cost und Epsilon Konditionen. Modelle mit einem Gamma von 0.2 erzielen die genauesten Prognosen, und sind robust gegenüber Änderungen der übrigen Parameter.

Verwendete Software und Pakete:

e1071 (Meyer, 2015)

In der für die Berechnung der Modelle verwendeten Library e1071, gibt es keine Option um die Gewichtung für die Beobachtungen bei einem Regressionsproblem eingeben zu können. Damit trotzdem diejenigen Beobachtungen, die nahe an den Ereignissen liegen eine höhere Bedeutung zukommt, kann man durch gezielte Datenvervielfältigung eine Gewichtung für die Modellerstellung erzwingen. Zunächst wird derselbe Gewichtungsvektor wie in Abbildung 5 erstellt. Danach wird dieser auf ganzzahlige Einheiten von eins bis zehn normiert. Nun wird das gewichtete Datenset erstellt, indem jede Datenzeile entsprechend ihres ermittelten Gewichtungswertes vervielfältigt wird. Dadurch ergibt sich zwar ein

größeres Datenset als zuvor, jedoch können die SVR gezielter auf den gewünschten Ereignishorizont trainiert werden.

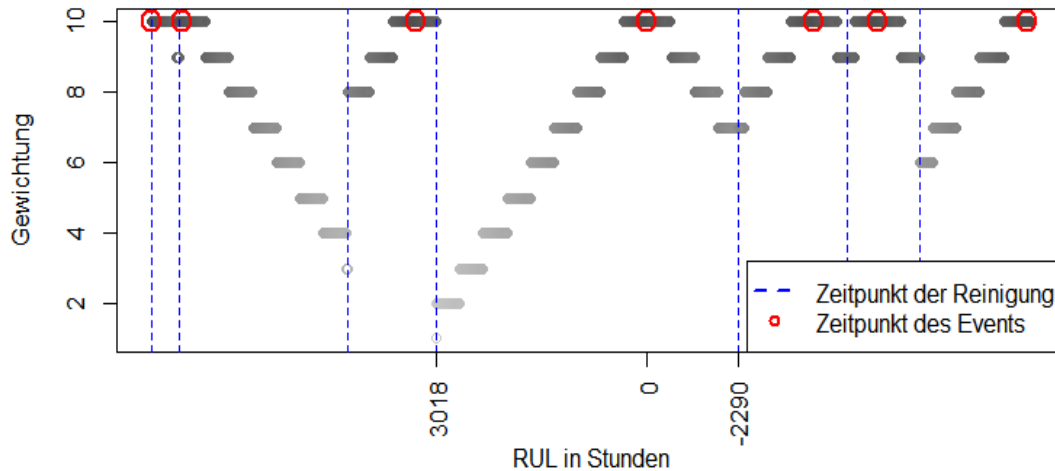


Abbildung 14: Gewichtung für SVR. Die Beobachtungen werden entsprechend ihrer Gewichtung vervielfältigt, um den Fokus der Modellerstellung gezielt auf die relevanten Zeitbereiche zu legen.

Variablenselektion (Rakotomamonjy, 2003) (Zhang & Wu, Variable selection for support vector machines in moderately high dimensions, 2016) (Müller, et al., 1997)

Es gibt einige Selektionsverfahren für die Support Vector Regression Methode, welche auf dem zum Modelltraining verwendeten Daten ein Ranking für die Variablenwichtigkeit erstellt. Um mit den Selektionsverfahren in der Fallstudie konsistent zu bleiben, wird jedoch wieder die CV-Methode zur Auswahl der geeigneten Variablen verwendet.

3.6.1 Die SVR Regression Modelle im Detail

Aufgrund der Voranalyse wurden die folgenden zehn Modellstrukturen mit jeder der Parameterkombination in Tabelle 26 erstellt. Dadurch ergeben sich genau 12 Modelle pro Kategorie.

Paramter	Wertebereich
Cost C	4, 16, 64
ϵ	0, 0.05
γ	0.2, 0.4
Kernel	Radial

Tabelle 26: Parameterkombinationen für die SVR.

3.6.1.1 SVR Modell nur mit den Basissignalen (SVR)

Dieses Modell wird nur mit den Basissignalen trainiert.

3.6.1.2 SVR Modell nur auf positiven RUL (SVR auf positiven RUL)

Dieses Modell wird nur auf Beobachtungen trainiert, die zeitlich vor dem Auftreten des Events liegen.

3.6.1.3 Gewichtetes SVR Modell (gewichtete SVR)

Die Gewichtung der einzelnen Beobachtungen erfolgt mit der zuvor beschriebenen indirekten Gewichtungsmethode.

3.6.1.4 SVR Modell inklusive abgeleitete Signale (SVR mit Features)

Dieses Modell beinhaltet alle Signale inklusive der Abgeleiteten.

3.6.1.5 Gewichtetes SVR Modell inklusive abgeleiteten Signalen (gewichtete SVR mit Features)

Dieses Modell kombiniert das Modell 3.6.1.3 und Modell 3.6.1.4.

SVR CV Modelle

Auf den 5 oben genannten Modelltypen wird je ein Modell erstellt, bei dem eine Variablenselektion nach der beschriebenen CV Methoden durchgeführt wird.

3.6.2 Selektion des Modelles mit der besten Vorhersagegüte

Für jede der zehn Kategorien wird jenes Modell gewählt welches den niedrigsten Wert in der *Mean RUL Time Deviation auf positiven RUL* hat, und in der Tabelle 27 mit den Modellen aus den anderen Kategorien verglichen.

Modell	Geglättete Event Time Deviation [Tage]	Mean RUL Time Deviation [Tage]	Mean RUL Time Deviation auf positiven RULs [Tage]	Cost	Epsilon	Gamma
Gewichtete SVR mit Features CV (3.6.1.5 CV)	30.34	59.38	44.59	16	0	0.4
Gewichtete SVR CV (3.6.1.3)	27.16	59.87	46.71	4	0	0.2
Gewichtete SVR	46.5	62.7	51.44	4	0	0.2
SVR mit Features CV	40.79	66.07	50.16	4	0	0.2
SVR CV	32.01	68.97	52.56	4	0	0.2
SVR	45.97	70.85	56.89	4	0.05	0.2
Gewichtete SVR mit Features	99.22	74.92	58.89	4	0	0.2
SVR auf positiven RUL	-	58.58	58.57	4	0.05	0.2
SVR mit Features	94.93	82.8	64.54	4	0	0.2
SVR auf positiven RUL CV	-	61.79	61.8	64	0	0.4

Tabelle 27: Aggregierte Prognoseabweichungen zur Modellwahl.

Das Modell mit den besten Prognoseergebnissen auf den Zeiträumen vor dem Event ist das *Gewichtete SVR mit Features CV* Modell (3.6.1.5 CV). Wegen der Variablenselektion mit der CV Methode entspricht dieses Modell im Wesentlichen dem *gewichteten SVR CV* Modell (3.6.1.3), da in beiden Fällen nur die Variable „Vibration_OST“ selektiert wird. Die leichten Unterschiede in den Prognosegüten entstehen dadurch, dass beim Datensatz mit Features einige Datenzeilen unvollständig sind (bedingt durch die Erstellung der Features) und aus dem Trainingsdatensatz ausgeschlossen werden.

Das Gewichtete SVR mit Features CV Modell (3.6.1.5 CV) im Detail (trainiert am gesamten Datensatz Ost inklusive Features):

Parameters:

SVM-Type: eps-regression
 SVM-Kernel: radial
 cost: 16
 gamma: 0.4
 epsilon: 0

Number of Support Vectors: 43313

10-fold cross-validation on training data:

Total Mean Squared Error: 875690

Squared Correlation Coefficient: 0.4483

Mean Squared Errors:

868313 909810 880244 856622 894847 848759 868365 897499 871210 861229

Trainset	Test 1	Test 2	Test 3	Test 4	Test 5	Test 6	Test 7	MW	Absoluter MW ohne Test 1
1	-	46.69	65.32	151.2	80.44	68.87	141.7	92.36	
2	-3.03	-	56.56	370.1	21.51	2.17	29.37	80.46	95.94
3	-3.03	11.83	-	12.93	1.19	0	-0.46	4.907	5.282
4	-3.03	11.83	117.6	-	9.62	1.07	6.12	24.89	29.26
5	-3.03	11.83	2.45	7.61	-	0	0	4.153	4.378
6	-3.03	11.83	127.2	67.26	14.26	-	5.24	38.15	45.17
7	-33.01	0	0	-0.86	-2.06	-7.22	-	7.192	2.028
<i>Abs.MW</i>	8.027	15.67	61.54	101.7	21.51	13.22	30.48	36.01	
Absoluter MW ohne Trainset 1		9.464	60.78	91.75	9.728	2.092	8.238		30.34

Tabelle 28: Event Time Deviation des gewichteten SVR mit Features CV Modells (3.6.1.5).

Trainset	<i>Test 1</i>	Test 2	Test 3	Test 4	Test 5	Test 6	Test 7	<i>MW</i>	MW ohne Test 1
1	-	365	161.3	413.6	165.6	84.86	348.6	256.5	
2	129	-	95.99	124	93.18	77.63	81.44	100.2	94.44
3	23.64	179.4	-	116.8	28.66	46.71	41.53	72.79	82.62
4	12.2	79.44	44.55	-	34.73	50.75	64.21	47.65	54.74
5	10.81	99.07	10.35	52.65	-	20.78	21.22	35.81	40.81
6	5.54	59.99	13.44	54.52	17.21	-	30.51	30.2	35.13
7	23.71	103.3	15.21	68.71	23.97	31.68	-	44.42	48.57
MW	34.15	147.7	56.8	138.4	60.56	52.07	97.91	83.94	
MW ohne Trainset 1		104.2	35.91	83.33	39.55	45.51	47.78		59.38

Tabelle 29: RUL Time Deviation des gewichteten SVR mit Features CV Modells (3.6.1.5).

Trainset	<i>Test 1</i>	Test 2	Test 3	Test 4	Test 5	Test 6	Test 7	<i>MW</i>	MW ohne Test 1
1	-	49.26	205	529	237.9	113.3	351.3	247.6	
2	141.3	-	75.25	135.7	89.52	51.09	81.02	95.65	86.52
3	0.56	13.13	-	130.6	25.57	43.64	41.83	42.55	50.95
4	14.64	13.62	50.25	-	36.85	70.78	64.71	41.81	47.24
5	1.61	11.59	11.67	57.5	-	30.08	21.34	22.3	26.44
6	8.18	6.73	17.22	66.94	23.63	-	30.73	25.57	29.05
7	16.23	14.13	14.55	75.58	17.48	15.03	-	25.5	27.35
MW	30.42	18.08	62.32	165.9	71.83	53.99	98.49	71.57	
MW ohne Trainset 1		11.84	33.79	93.27	38.61	42.12	47.93		44.59

Tabelle 30: RUL Time Deviation auf den positiven RULs des gewichteten SVR mit Features CV Modells (3.6.1.5).

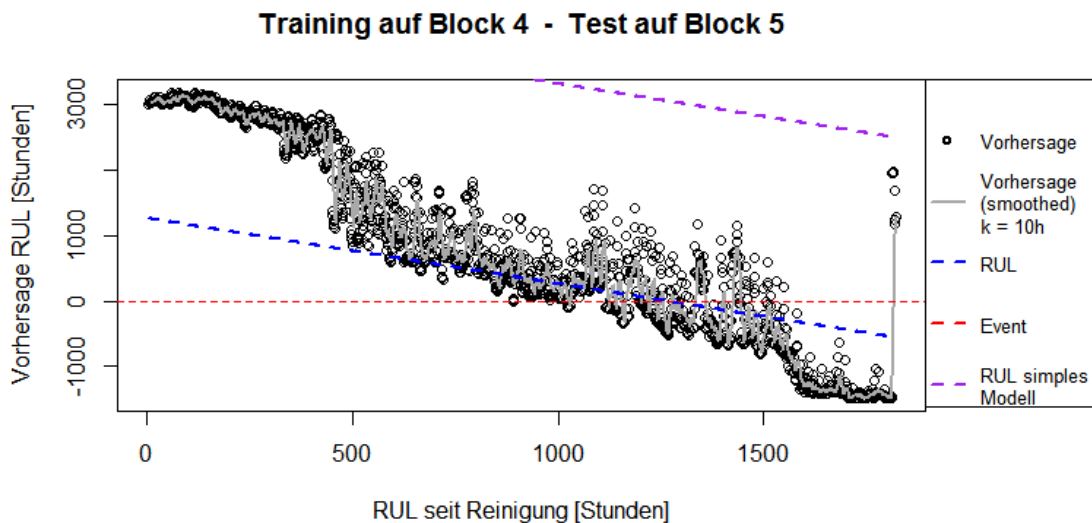


Abbildung 15: Dieser Prognoseplot zeigt, dass vor allem zu Beginn eines Intervalls, (hier bis ca. 500 Stunden) die RUL Schätzungen noch deutlich danebenliegen. Im Zeitraum vor dem Event streuen die Prognosen zwar stark, aber durch eine Glättung von 10 Stunden, folgen sie der tatsächlichen RUL schon viel besser. Dadurch wurde das Event 9,62 Tage vor dem Auftreten prognostiziert (siehe Tabelle 28).

Der Parametrisierungsaufwand der SVR ist erheblich, und die Berechnungszeiten sehr hoch. Die Prognosegüten sind bei allen Trainingssets (bis auf dem zweiten) relativ gut. Durch die Hochdimensionierung des Raumes lässt sich weder eine Formel noch ein menschlich interpretierbares Regelwerk aus dem Modell ableiten. Zugute kommt dem Modell, dass es alleine durch die Auswahl des Vibrationssignales gute Vorhersagen liefern kann. Dies könnte in weiterer Folge bedeuten, dass es sich gut generalisieren, also auf anderen Anlagen des gleichen Bautyps übertagen lässt, ohne an Prognosegüte einzubüßen. Vorausgesetzt natürlich, dass die Vibrationsmessung an den anderen Anlagen vergleichbar vorgenommen wird.

3.7 VERGLEICH DER ERGEBNISSE DER REGRESSIONSMETHODEN

Es werden die Prognosekennzahlen der besten Modelle für jede Regressionsmethode mit denen des Simples Modells verglichen. Für jedes Trainingsset können sechs unabhängige Prognosen auf den Testsets erstellt werden. Der Mittelwert davon soll mit dem Mittelwert des Simples Modells auf signifikanten Unterschied getestet werden. Die Prognoseergebnisse in den Kategorien Event Time Deviation, RUL Deviation und RUL Positive Deviation zeigen eine nicht normale Verteilung (Bei nahezu jedem Modell und Testset). Weshalb ein t-Test zur Bestimmung eines signifikanten Unterschiedes des Lagewertes nicht zulässig ist. Daher wird auf den nichtparametrischen Wilcoxon Rangsummentest zurückgegriffen, auch wenn dieser eine geringere Power hat. Die Voraussetzungen für den Test sind erfüllt:

- Unabhängigkeit der Stichproben – Die Prognosewerte des Simplen Modells (Tabelle 9) stehen in keinem Zusammenhang mit den Werten aus irgendeinem Modell, und umgekehrt.
- Unabhängigkeit innerhalb der Stichprobe – Die 6 Testsets sind untereinander unabhängig, da diese jeweils eine neue Zeitreihe definieren, und nicht für die Modellerstellung genutzt wurden.

Die Hypothese dieser Arbeit lautet:

Können die Wartungszyklen mittels der Regressionsmethoden statistisch gesehen optimiert werden, und ist das Resultat signifikant besser als die Strategie mit fixen Wartungsintervallen?

Die Antwort auf diese Frage liefert der Hypothesentest nach Wilcoxon für die drei Prognosekriterien, wenn das simple Modell mit dem jeweils besten Modell aus jeder Regressionsmethode verglichen wird. Es genügt ein einseitiger Test, da nur die Frage beantwortet werden soll, ob die Prognoseabweichungen der Regressionsmethoden niedriger sind als jene des simplen Modells.

Vergleich der besten Modelle mit dem simplen Modell.

3.7.1 Vergleich der Mean Event Time Deviations: p-Werte des Wilcoxon Rangsummentests

Intervall	Lineares Modell gewichtetes LM mit Features BIC (3.2.1.12)	Regression Tree gewichteter RT (3.3.1.3)	Random Forest gewichteter RF CV (3.4.1.3 CV)	GAM gewichtetes GAM mit Features CV (3.5.1.5 CV)	SVR Gewichtete SVR mit Features CV (3.6.1.5 CV)
1	0.12	0.409	0.409	0.013	0.242
2	0.12	0.155	0.155	0.12	0.12
3	0.001	0.002	0.002	0.001	0.001
4	0.001	0.002	0.002	0.001	0.002
5	0.004	0.001	0.004	0.004	0.002
6	0.066	0.013	0.021	0.013	0.013
7	0.013	0.004	0.803	0.004	0.004

Tabelle 31: Die Testergebnisse der Mean Event Time Deviation: Diese belegen eindeutig, dass durch jede der 3 Methoden die Event Time Deviations gegenüber dem Simplen Modell signifikant verbessert werden können.

3.7.2 Vergleich RUL-Time-Deviation: p-Werte des Wilcoxon Rangsummentests

Intervall	Lineares Modell gewichtetes LM mit Features BIC (3.2.1.12)	Regression Tree gewichteter RT (3.3.1.3)	Random Forest gewichteter RF CV (3.4.1.3 CV)	GAM gewichtetes GAM mit Features CV (3.5.1.5 CV)	SVR Gewichtete SVR mit Features CV (3.6.1.5 CV)
1	0.996	0.998	0.998	0.992	0.996
2	0.845	0.845	0.845	0.845	0.88
3	0.91	0.88	0.91	0.758	0.758
4	0.758	0.591	0.469	0.803	0.706
5	0.531	0.197	0.197	0.197	0.197
6	0.155	0.758	0.845	0.197	0.197
7	0.758	0.409	0.409	0.409	0.409

Tabelle 32: Der Wilcoxon Test bestätigt nicht, dass es einen signifikanten Unterschied zwischen den Prognosewerten der besten Regressionsmethoden und dem Simplen Modell gibt.

3.7.3 Vergleich RUL Time Deviation auf positive RULs: p-Werte des Wilcoxon Rangsummentests

Intervall	Lineares Modell gewichtetes LM mit Features BIC (3.2.1.12)	Regression Tree gewichteter RT (3.3.1.3)	Random Forest gewichteter RF CV (3.4.1.3 CV)	GAM gewichtetes GAM mit Features CV (3.5.1.5 CV)	SVR Gewichtete SVR mit Features CV (3.6.1.5 CV)
1	0.992	0.996	0.996	0.968	0.992
2	0.803	0.91	0.88	0.88	0.91
3	0.591	0.706	0.469	0.409	0.35
4	0.469	0.591	0.469	0.706	0.65
5	0.294	0.066	0.12	0.066	0.066
6	0.066	0.65	0.758	0.066	0.066
7	0.35	0.047	0.066	0.09	0.155

Tabelle 33: Auf den Prognoseabweichungen der positiven RUL können die Wilcoxon Tests keinen signifikanten Unterschied zwischen dem Simplen Modell und den Modellen der Regressionsmethoden ausmachen. Lediglich die Regression Trees, welche am siebten

Intervall trainiert wurden zeigen mit 0.047 einen signifikanten Eintrag. Dies ist aber bei 6x5 Tests kein aussagekräftiges Resultat.

Die Event Time Deviation kann, durch jede der gezeigten Regressionsmethoden, gegenüber dem Simplen Modell wesentlich verbessert werden. Die Tests auf den RUL Time Deviations können keinen Unterschied zwischen den Simplen Modell und den Anderen bestätigen. Wie in den Prognoseplots ersichtlich ist, weichen die Prognosewerte in vielen Zeitbereichen zu stark von der tatsächlichen RLZ ab. Daher sollten die RUL Prognosen mit einem geeigneten Glättungsfaktor auf einem entsprechend gut gewähltem Sliding Window für die Echtzeitprognosen geglättet werden.

3.8 BESTIMMUNG DES GRÖßTMÖGLICHEN ZEITINTERVALLS OHNE INFORMATIONSVERLUST FÜR DIE MODELLE.

Zu Beginn der Datenaufbereitung wurde das Aufzeichnungsintervall auf 1 Minute festgelegt. Einige Signale der Anlage werden aber viel häufiger gemessen. Zum Beispiel werden die Vibrationswerte annähernd einmal pro Sekunde aufgezeichnet. In diesem Kapitel wird überprüft ab welchem zeitlichen Diskretisierungsintervall die Modelle keine besseren Prognosen mehr liefern werden. Um zu testen ob es zu einem Informationsverlust kommt, wenn man die Datenaufzeichnung in gröbere Intervalle zusammenfasst, werden Modelle mit demselben Setting erstellt und die Kennzahlen und Prognosewerte gegenübergestellt.

Modellgüte	Kriterium	15 Sekunden	30 Sekunden	1 Minute	2 Minuten
Lineares Modell	Adj. R ²	0.5051	0.4793	0.479	0.4459
Regression Trees	1-Xerror ³	0.7038	0.678	0.6836	0.682
Random Forests	oob MSE ⁴	0.7661	0.8554	0.9029	0.9359
GAM	dev.expl ⁵	0.4494	0.4429	0.4547	0.4127
SVR	scorrcoef ⁶	0.4999	0.4483	0.48	0.4585

Tabelle 34: Vergleich des Informationsgehaltes der Datensets mit unterschiedlicher Aufzeichnungsrate anhand der Modellgüte.

³ Xerror ist die 10-fach Kreuzvalidierungsfehlerrate die beim Erstellen des Regression Trees berechnet wird und Werte zwischen 0 und 1 annimmt. Je höher der Wert 1-Xerror ist, desto mehr Varianz kann mit dem RT Modell erklärt werden.

⁴ oob MSE ist der „out of bag Mean Squared Error“ der beim Erstellen des Random Forests berechnet wird.

⁵ Dev.expl ist die durch das erstellte GAM Modell erklärte Varianz.

⁶ Scorrcoef ist der Squared Correlation Coefficient der beim berechnen der SVR erstellt wird.

Prognosegüte	15Sekunden	30Sekunden	1Minute	2Minuten
Lineares Modell	61.26	59.97	62.22	57.2
Regression Trees	66.71	58.68	68.19	66.6
Random Forests	62.28	66.47	68.8	65.51
GAM	73.23	46.94	63.22	54.89
SVR	60.05	54.59	59.38	58.9

Tabelle 35: Vergleich des Informationsgehaltes der Datensets mit unterschiedlicher Aufzeichnungsrate anhand der Prognosegüte (Mean RUL Time Deviation ohne Trainingsset 1).

In Tabelle 34 ist ersichtlich, dass das Gütekriterium für alle Regressionsmethoden in etwa gleichbleibt, unabhängig davon wie das Aufzeichnungsintervall gewählt wurde. Die Tabelle 35 stellt fest, dass es auch auf die Prognosegüten kaum einen Einfluß gibt, wenn das Intervall verkürzt oder verlängert wird. Insbesondere ist kein Trend erkennbar. Hätte die Aufzeichnungsrate einen signifikanten Einfluss auf die Prognosegüte, dann müsste es deutliche Anzeichen einer abfallenden *Mean RUL Time Deviation* geben, je hochfrequenter das Intervall wird.

4 BETRIEBSWIRTSCHAFTLICHE SIMULATION

Es kann auf den historischen Daten eine Simulation für die Kosten und Nutzen bzw. zu den Laufzeiten der Anlage erstellt werden.

Hierzu werden 3 Anlagenzustände definiert:

- *Vollbetrieb*: Die Anlage steht uneingeschränkt zur Verfügung
- *Reduzierter Betrieb*: Die Anlage ist nur im eingeschränkten Drehzahlbereich verfügbar
- *Wartung*: Die Anlage wird außer Betrieb gesetzt und Reinigungsarbeiten werden durchgeführt. Die Dauer dieser Arbeit wird auf durchschnittlich 24 Stunden geschätzt.

Es wird angenommen, dass die Anlage im Vollbetrieb produktiver ist als im reduzierten Betrieb, da im Vollbetrieb schneller und qualitativ hochwertiger produziert werden kann. Wieviel Kosten pro Stunde durch den reduzierten Betrieb entstehen ist nicht bekannt. Es können lediglich die durchschnittlichen Wartungskosten beziffert werden. Deshalb wird die Simulation nicht auf die Kosten und Nutzen optimiert, sondern nur auf die Laufzeit in Vollbetrieb bei möglichst geringen Wartungsaufwand. Die Vorgangsweise der Simulation ist dann wie folgt:

Die Anlage befindet sich von Beginn eines Reinigungsintervalls so lange im *Vollbetrieb*, bis:

- **vom Modell eine Wartung beauftragt wird.** Dies wird von der geglätteten RUL Prognose ausgelöst, wenn diese das erste Mal kleiner als ein zu optimierender Schwellwert ist. Wenn dieser Schwellwert niedrig gewählt wird, dann bedeutet dies, dass die Anlage länger im Vollbetrieb laufen soll. Umgekehrt bedeutet ein hoher Schwellwert, kürzere Laufzeiten im Vollbetrieb, jedoch wird somit auch das Risiko einen Bruch zu erleiden reduziert. Die Auswahl des Schwellwertes lässt sich mit einem linearen Optimierungsproblem lösen, wenn dazu die Betriebskosten und Nutzen der verschiedenen Anlagenzustände bekannt sind. (Holzer, Peter; Pözlbauer Georg, 2014) Da keine Kosten und Nutzen beziffert werden können, wird der Schwellwert zunächst auf Null gesetzt, damit eine Laufzeitsimulation möglich wird. Der Wert Null bedeutet in der Praxis, dass ohne Vorwarnzeit die Anlage heruntergefahren und gewartet werden kann.
- **das Event eintritt.**

Die Anlage befindet sich im *reduzierten Betrieb*, wenn das Event eintritt, bevor das Modell eine Wartung in Auftrag gegeben hat. Die Anlage bleibt solange in diesem Zustand, bis eine Wartung durchgeführt wird, also bis zum Beginn eines neuen Reinigungsintervalls.

Die Anlage erhält eine 24 Stunden dauernde *Wartung*, wenn:

- ein Wartungsauftrag vom Modell beauftragt wird. (RUL Prognose kleiner als Null).

- oder am Ende eines Reinigungsintervalls, wenn bis dahin keine Wartung ausgelöst wurde.

Damit man die simulierten Betriebszeiten der verschiedenen Regressionsmethoden miteinander vergleichen kann, werden die Zeiten in Prozent der Gesamtlaufzeit angegeben.

Die Dauer für den Vollbetrieb (2), ergibt sich aus der Differenz der Event Time Deviation (ETD_m) und den Laufzeiten (LZ_m) bis zum Event (1). Das erste Reinigungsintervall wird wieder aufgrund der geringen Laufzeit aus den Berechnungen ausgeschlossen. Daher läuft der Index m von 2 bis 7.

$$LZ_m = t_{R_{m+1}} - t_{E_m} \quad (1)$$

mit

t_{R_m} Zeitpunkt der m-ten Reinigung

t_{E_m} Zeitpunkt des m-ten Events

und

$$D_{Voll_m} = \begin{cases} LZ_m - ETD_m & , ETD_m > 0 \\ LZ_m & , ETD_m \leq 0 \end{cases} \quad \text{für } m \in \{2, \dots, 7\} \quad (2)$$

$$ETD_m = t_{E_m} - \widehat{t}_{E_m} \quad (3)$$

mit

\widehat{t}_{E_m} Prognostizierter Zeitpunkt des m-ten Events

Die Dauer für den reduzierten Betrieb (4), ergibt sich aus der Differenz zwischen dem Eventzeitpunkt und der dem prognostizierten Eventzeitpunkt, wenn dieser erst nach dem Eventzeitpunkt liegt. Bzw. ist die Dauer des reduzierten Betriebes gleich Null, wenn das Event rechtzeitig (also vor dem tatsächlichen Eventzeitpunkt) prognostiziert wurde.

$$D_{red_m} = \begin{cases} -ETD_m & , ETD_m < 0 \\ 0 & , sonst \end{cases} \quad \text{für } m \in \{2, \dots, 7\} \quad (4)$$

Die Dauer des Vollbetriebs in Prozent (5) errechnet sich aus der Summe der Dauer im Vollbetrieb für alle Testintervalle, dividiert durch die simulierte Gesamtdauer.

$$D_{\%Voll} = \frac{D_{voll}}{D_{voll} + D_{red} + D_{war}} * 100 \quad (5)$$

mit

$D_{war} = 24h$ Die Dauer einer Wartung wird mit 24h festgelegt.

$$D_{voll} = \sum_{m=2}^7 D_{voll_m}$$

$$D_{red} = \sum_{m=2}^7 D_{red_m}$$

Die Dauer des reduzierten Betriebs in Prozent (6) errechnet sich analog zu oben, aus der Summe der Dauer im reduzierten Betrieb für alle Testintervalle, dividiert durch die simulierte Gesamtdauer.

$$D_{\%red} = \frac{D_{red}}{D_{voll} + D_{red} + D_{war}} * 100 \quad (6)$$

und analog

$$D_{\%war} = \frac{D_{war}}{D_{voll} + D_{red} + D_{war}} * 100 \quad (7)$$

Beispiel für das beste lineare Modell für die Anlage Ost: Aus der Tabelle 8 – *Reinigungen* und der Tabelle 11 – *Event Time Deviation* kann man die notwendigen Werte entnehmen. Es wird jenes Modell ausgewählt, welches die niedrigsten Werte in der Kategorie *RUL Time Deviation auf positiven RUL* hat. Dies ist laut *Tabelle 13* das Modell, das auf dem 4. Reinigungsintervall trainiert wurde.

$$\begin{aligned} D_{voll} &= (13 - 11,27) + (164 - 0,46) + (29 - 15,93) + (16 - 0) + (50 - 0,75) \\ &= 243,6 \text{ Tage} \end{aligned}$$

$$D_{red} = 0 + 0 + 0 + 0,33 + 0 = 0,33 \text{ Tage}$$

$$D_{war} = 5 \text{ Tage}$$

Und somit:

$$D_{\%voll} = \frac{243,6}{243,6 + 0,33 + 5} * 100 = 97,86\%$$

$$D_{\%red} = \frac{0,33}{243,6 + 0,33 + 5} * 100 = 0,13\%$$

$$D_{\%war} = \frac{5}{243,6 + 0,33 + 5} * 100 = 2,01\%$$

	% in Vollbetrieb	% im reduzierten Betrieb	% in Wartung
Simple Modell	30,01	69,44	0,55
Lineares Modell mit Features und BIC Selektion (3.2.1.12)	97,86	0,13	2,01
Gewichtetes Regression Tree Modell (3.3.1.3)	66,03	32,65	1,32
Gewichtetes Random Forest Modell mit CV Selektion (3.4.1.3 CV)	93,28	3,01	3,71
gewichtetes GAM mit Features CV (3.5.1.5 CV)	98	0	2
Gewichtete SVR mit Features CV (3.6.1.5 CV)	96,18	0	3,82

Tabelle 36 Simulationsergebnis mit den geglätteten Event Time Deviation und dem Schwellwert von 0 Tagen

Das simple Modell ist zwar nur zu 30% im Vollbetrieb, aber dafür wird nur zu 0,55% der Zeit gewartet. Mit dem Linearen Modell wäre die Anlage zu 97,86 im Vollbetrieb gewesen und nur zu 0,13% im reduzierten Betrieb. Dafür müsste man die Anlage zu 2,01% der Gesamtzeit für die Wartungsarbeiten außer Betrieb setzen. Das Regression Tree Modell ist ein Mittelweg aus den Beiden zuerst genannten Modellen. Hier ist die Anlage immerhin zu 66% im Vollbetrieb und die Wartungszeiten sind mit 1,32% noch moderat hoch. Keine gute Option scheint das Random Forest Modell zu sein, da zwar 93% Vollbetrieb erreicht werden, jedoch die Wartungszeit mit 3,7% viel zu hoch wäre. Hier wären das Lineare Modell und das GAM in beiden Kategorien überlegen. Generell dominiert das GAM Modell in dieser Simulation alle anderen Modelle in den Kategorien „Vollbetrieb“ und „reduzierter Betrieb“.

Bemerkung:

Wenn die tatsächlichen Kosten und Nutzen der Anlage bekannt sind, dann kann die RUL Nulllinie zur Optimierung nach oben oder unten verschoben werden, je nachdem wo sich für die 6 Testintervalle der größte Gewinn erzielen lässt. Dann könnten auch gezielt die Betriebswirtschaftlichen Kennzahlen simuliert werden, und es ließe sich das Einsparungspotential besser abschätzen.

5 GENERALISIERBARKEIT DER MODELLE

5.1.1 Generalisierbarkeit der Modelle

In der produzierenden Industrie sind häufig baugleiche oder bauähnliche Anlagen beim selben Betreiber vorzufinden. In der Fallstudie ist die Anlage „West“ in dem Sinne zur Anlage „Ost“ baugleich, als dieselben Bauteile und Elektronik sowie dieselben Steuerungen verwendet werden. Nun stellt sich die Frage ob die Im Zuge der Studie erstellten Modelle auf der Ostanlage auch auf der Westanlage entsprechend gute Prognoseergebnisse liefern. Die nächste Frage, die sich aufdrängt ist ob die besten Modelle mit einem „Retraining“ auf dem gesamten Datensatz Ost bessere Prognosen liefern, als jene Modelle die nur auf den einzelnen Reinigungsintervallen erstellt wurden. Beim bisherigen Vorgehen zur Modellerstellung wurden immer nur Daten von einem Reinigungsintervall zum Trainieren der Modelle verwendet. Dies hatte den Hintergrund, dass man so sechs unabhängige Datensätze zur Modellevaluierung erhält (für die übrigen 6 Intervalle). Verwendet man nun mehr als ein Reinigungsintervall, und somit mehr als nur eine Zeitreihe mit den RUL Zeiten, dann ist es möglich die Betriebslaufzeit als Einflussvariable einzuführen, ohne dadurch sofort einen perfekten Modellfit zu bekommen. Somit stellt sich die dritte Frage, ob dieser „Betriebsstundenzähler“ einen signifikanten Einfluss auf die Prognosegüte hat, und daher für Modelle in der Echtzeitanwendung integriert werden sollte.

Um diese Fragen zu klären werden in diesem Kapitel die besten Modelle mit demselben Setup auf dem gesamten Datensatz Ost trainiert, einmal mit und einmal ohne den „Betriebsstundenzähler“. Diese Modelle werden dann mit den Modellen verglichen die auf dem zweiten bis siebten Reinigungsintervall der Ostanlage erstellt wurden. Die Daten auf der Anlage West werden als unabhängiges Testset zur Bestimmung der Prognosegüte der Modelle verwendet. Da nur im Reinigungsintervall sechs das Ereignis zum Reinigungsauftrag auftrat, kann nur auf diesen einem Datenblock die RUL Zeiten erstellt werden. Dadurch können zwar keine Teststatistiken auf den Prognosegüten erstellt werden, dennoch sollte es möglich sein eine qualitative Aussage treffen zu können, welche Modelle robuste Prognosen liefern, und ob die Variable für die Betriebszeit ein wichtiges Signal zur RUL Vorhersage sein kann.

Lineares Modell	Regression Tree	Random Forest	GAM	SVR
237.7	90.54	49.71	507.3	32.09

Tabelle 37: RUL Time Deviation auf dem Block 6 der Anlage West (ohne Betriebsstundenzähler)

Die RUL Time Deviations zeigen, dass die Random Forests und SVR Modelle robust genug sind und in etwa dieselben Prognoseabweichungen wie auf der Ostanlage zu erwarten sind. Die anderen Modelle zeigen weniger zufriedenstellende Resultate. Vor allem die GAM Modelle zeigen sehr große Ausreißer in den Prognosen. Die Regression Trees haben durch ihr relativ einfaches Regelwerk noch immer akzeptablere Ergebnisse als das Lineare Modell, welches wohl nur für baugleiche Anlagen geeignet sein wird.

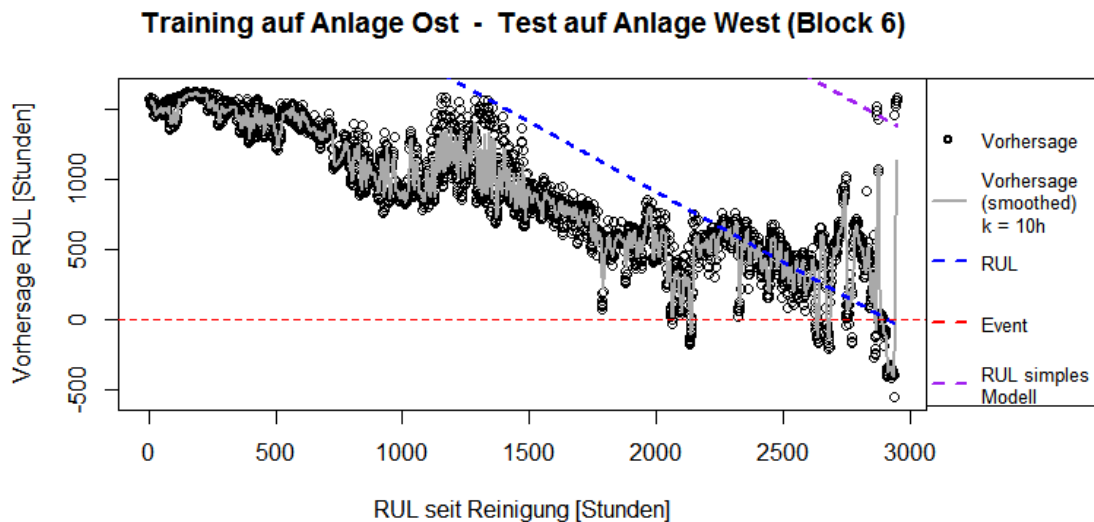


Abbildung 16: Vorhersage des SVR Modells auf der sechsten Reinigungsperiode der Westanlage (ohne Betriebsstundenzähler).

Die Variable Betriebsstundenzähler (1) wird definiert, als Zeit die seit der zuletzt getätigten Reinigung vergangen ist, und wird wieder bei null initiiert, wenn eine neue Reinigung erfolgt.

$$Z_{m_i} = t_i - t_{R_m} \quad \text{für alle } m \in M \quad (1)$$

für $i = 1, \dots, n_m$ Beobachtungen und n_m der Anzahl an Beobachtungen im Intervall m

Mit dem Betriebsstundenzähler als zusätzliche Einflussvariable werden die Modelle noch einmal neu auf dem gesamten Datensatz Ost trainiert und an der Westanlage getestet.

Lineares Modell	Regression Tree	Random Forest	GAM	SVR
273.35	76.1	47.41	496.51	34.64

Tabelle 38: RUL Time Deviation auf dem Block 6 der Anlage West mit Betriebsstundenzähler.

Die Tabelle zeigt keine massiven Unterschiede zwischen den Prognosegüten der Modelle mit und ohne den Betriebsstundenzähler. Das heißt, dass man für die Generalisierbarkeit der Modelle in dieser Fallstudie keinen solchen Einflussparameter definieren muss, um gute Prognosen zu erhalten. Außerdem könnte dies bedeuten, dass die Restlaufzeit alleine anhand der gegebenen Sensorik geschätzt werden kann, und kein manueller Input notwendig ist. Das heißt, man müsste den Modellen auch nicht mitteilen, wann eine Reinigung erfolgt ist und wann somit eine neue Zeitreihe beginnt. Durch den gleitenden Mittelwert auf den Prognoseergebnissen würde man nur abwarten müssen, bis alle Prognosen auf dem neuen Reinigungsintervall erfolgt sind, und hätte ab diesen Zeitpunkt wieder eine gültige Schätzung der RUL. Die Support Vector Regression zeigt mit einer durchschnittlichen Abweichung von 34,64 Tagen die beste Prognosegüte an der Anlage West. Ähnlich wie beim SVR Modell ohne Betriebsstundenzähler, sieht man in der Abbildung 17 dass die Prognosen zu Beginn des Intervalls noch keine guten Schätzungen für die RUL liefern. Ab dem Zeitpunkt 1200 Stunden beginnt sich ein linearer Trend einzustellen, welcher noch stark geglättet werden muss, damit die Restlaufzeit zuverlässig vorhergesagt werden kann.

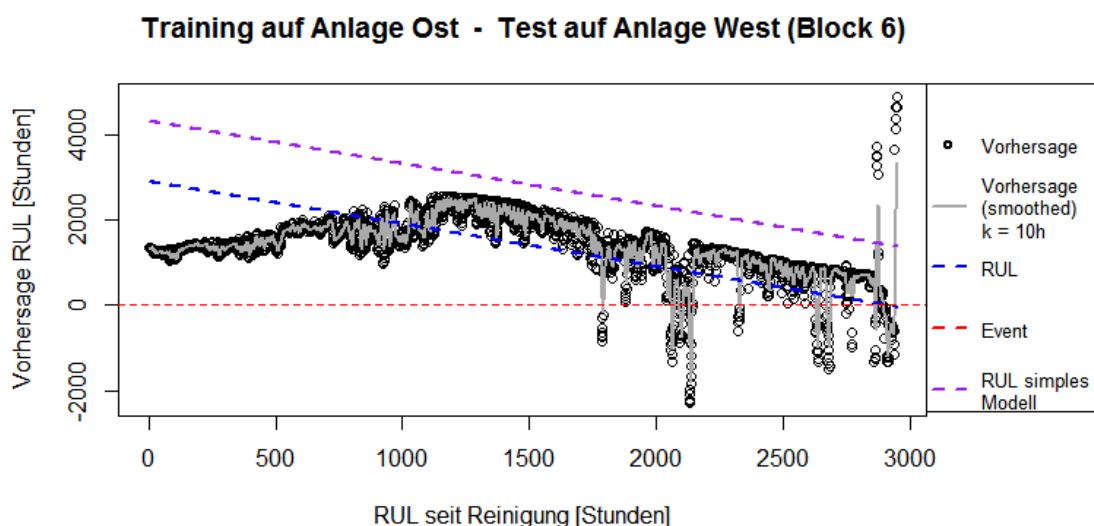


Abbildung 17: Vorhersage des SVR Modells auf der sechsten Reinigungsperiode der Westanlage mit Betriebsstundenzähler.

Das SVR Modell mit Betriebsstundenzähler im Detail:

Parameters:
 SVM-Type: eps-regression
 SVM-Kernel: radial
 cost: 16
 gamma: 0.4
 epsilon: 0

Number of Support Vectors: 43296

10-fold cross-validation on training data:

Total Mean Squared Error: 590777

Squared Correlation Coefficient: 0.6046

Mean Squared Errors:

584887 560331 588001 637459 630899 567653 564102 582778 608274 583388

Die in der Tabelle 39 dargestellten Ergebnisse stammen von den Modellen, welche auf einzelnen Reinigungsintervallen trainiert wurden. Für jede Regressionsmethode ergeben sich damit sechs Werte für die Prognoseabweichung.

Modelltraining auf dem Intervall	Lineares Modell	Regression Tree	Random Forest	GAM	SVR
2	136.3	101.3	96.02	394.5	100.8
3	179.4	64.25	58.12	35.2	46.13
4	156.9	31.57	23.34	211	23.39
5	42.83	44.47	60.13	37.66	41.92
6	42.12	60.71	77.19	73.6	61.34
7	205	33.28	30.63	51.45	51.84
MW	127.1	55.93	57.57	133.9	54.24
Median	146.6	52.59	59.13	62.53	48.99

Tabelle 39: RUL Time Deviation auf dem Block 6 der Anlage West ohne Betriebsstundenzähler. Die Modelle wurden auf je einem Reinigungsintervall der Ost Anlage trainiert.

Die Prognoseabweichungen der Random Forests und der SVR sind im Mittel höher als bei den Modellen die am gesamten Datensatz trainiert wurden. Die Resultate zeigen sehr unterschiedliche Prognoseabweichungen in Abhängigkeit vom Trainingsintervall. Die generalisierten Modelle sollten daher auf dem gesamten Datensatz trainiert werden. Jedoch scheinen die linearen und die GAM Modelle dafür ungeeignet zu sein, wie man in der Tabelle 37 und Tabelle 38 sehen kann.

6 AUSBLICK

6.1 RÜCKSCHLUSS DER VARIABLENWICHTIGKEITEN AUF DIE FEHLERRSACHE

Die linearen Modelle geben klaren Aufschluss über die Korrelationen von Einflussvariablen zur Restlaufzeit. Es lässt sich daraus ablesen um wieviel die Restlaufzeit verkürzt wird, wenn zum Beispiel die Vibration um eine Einheit zunimmt. Bei Regression Trees und Random Forests ist diese Information nicht direkt aus dem Modell ablesbar. Aber man kann immer noch die aktuellen Daten fixieren und nur die Vibrationswerte um die gewünschte Einheit verändern, und darauf mittels des Modells eine Prognose erstellen. Dies kann man natürlich mit jeder Einflussvariable simulieren, jedoch sollte beachtet werden, dass zwischen einigen Einflussgrößen starke Bindungen herrschen. Zum Beispiel wirkt sich eine Erhöhung der Drehzahl auch immer auf die Leistung aus und umgekehrt. Also sollten – bei sonst gleichbleibenden Betriebsbedingungen – je nach Zusammenhang der Variablen, immer auch andere Einflussvariablen für die Simulation der Restlaufzeit geändert werden.

Jedes statistische Modell kann eine Auflistung der wichtigsten (einflussreichsten) Variablen erstellen. Dies bietet einen guten Ausgangspunkt um ein tieferes Verständnis der Anlage und ihrer Zusammenhänge aufzubauen. Eine gezielte Ursachenanalyse kann nur mit den Technikern der Anlage durchgeführt werden. Durch die Modelle kann zwar eine mehr oder weniger verlässliche Aussage getroffen werden, wie sich die Einflussvariablen auf die Restlaufzeit auswirken, jedoch ist durch die Modelle keinerlei Kausalität abgebildet. Diese sollte mittels technischem Verständnis der Anlage und statistischen Wissen erforscht werden.

6.2 NEURONALE NETZWERKE

Ein großer Vorteil von gewissen Neuronalen Netzwerken ist, dass diese während des Betriebes dazulernen können (Online-Learning). Dies ist vor allem dann interessant, wenn sich die Fehlermuster ändern, bzw. neue Muster erkannt werden sollen. Aber auch wenn sich Schwellwerte ändern oder Sensormesswerte einen Trend (zB durch Verschmutzung ausgelöst) haben. Einen guten Überblick zur Anwendung dieser Methode bietet Umair, Masdi und Karim (Umair, Masdi, & Karim, 2014).

7 CONCLUSIO

Diese Arbeit habe ich mit dem Zitat von Ray Bradbury begonnen:

“I was not predicting the future, I was trying to prevent it.”

Ray Bradbury (1920 – 2012)

Er hat diese Aussage wohl auf sein berühmtes Buch Fahrenheit 451 bezogen. Er wollte die Zukunft verändern, indem er eine mögliche Katastrophe vorhersagt. Nicht um recht zu behalten, sondern um die Menschen vor Torheit zu bewahren.

Predictive Maintenance versucht im industriellen Umfeld dasselbe Ziel zu erreichen. Es sollen durch Vorhersagen teilweise katastrophale Maschinenschäden vermieden werden. Doch wenn man ein solches System verwendet und mögliche Katastrophen abwenden kann, bleibt immer die Frage bestehen, wie die Zukunft der Maschine ausgesehen hätte, hätte man die Vorhersagen nicht beachtet. Ein PM System ist kein reiner Selbstzweck, um sich selbst zu beweisen, sondern sollte eine bessere Zukunft ermöglichen, indem es die vorhandenen Ressourcen effizienter einsetzen soll. Wenn aber jede Katastrophe rechtzeitig abgewendet werden kann, dann sinkt wiederum die Glaubwürdigkeit der Vorhersagen, da ja angezweifelt werden kann ob es überhaupt noch zu Katastrophen käme, würde man dem System nicht vertrauen. Gewisse Zweifel werden immer bestehen bleiben, da man die alternative Zukunft nicht kennt.

In der Fallstudie wurde gezeigt, wie die Prozess- und Signaldaten einer Industrieanlage zu Zeitreihen aufbereitet und zusammengefasst werden. Es wurden diverse Signalverarbeitungen durchgeführt, die für viele Modelle wichtige erklärende Variablen erzeugten. Durch gezielte Gewichtung der relevanten Zeiträume konnten ebenso Verbesserungen in der Prognosegüte erreicht werden. Die Regressionsmethoden wurden so umgesetzt, dass sie im Echtzeitbetrieb zu jedem Zeitpunkt (im Minutentakt) eine Prognose über die zu erwartende Restlaufzeit (RUL) liefern. Die Betriebswirtschaftliche Simulation zeigt, dass es dadurch möglich ist, die Anlagenverfügbarkeit massiv zu erhöhen, wenn man den geschätzten Wartungszeitpunkten der Regressionsmethoden Folge leistet. Dies verdeutlicht der Vergleich zur aktuell angewandten Wartungsstrategie, bei dem kein PM System eingesetzt wird, sondern nach fixen Intervallen gewartet wird. Die Maschinenverfügbarkeit unter Vollbetrieb liegt beim *GAM* Modell bei 98% und beim *Linearen Modell* bei 97,86% während sie beim Fehlen eines PM Systems bei ca. 30% liegt. Die Hypothesentests zeigten, dass der Ereigniszeitpunkt für eine Wartung hoch signifikant besser prognostiziert wird als durch fixe Wartungszeitpunkte. Die Prognosegenauigkeit nimmt zu je näher das zu prognostizierende Ereignis rückt. Dadurch ermöglichen die Regressionsmodelle schon Tage oder Wochen im Voraus eine Wartung einzuplanen und dies mit zunehmender Präzision. Des Weiteren wurden die erstellten Modelle auf einem Datensatz einer baugleichen Anlage getestet. Dadurch wurde festgestellt, dass sich die

Random Forestes und die *Support Vector Regression* sehr gut für eine generalisierte Anwendung bei baugleichen Maschinen eignen, wohingegen sich die anderen Methoden nur auf der Maschine selbst anwenden lassen, auf deren Datensatz sie angelernt wurden.

8 APPENDIX

Verwendete Software

R version 3.1.1 (2014-07-10) -- "Sock it to Me"

Platform: x86_64-w64-mingw32/x64 (64-bit)

Verwendete Pakete

lawstat	Levene Test	Version 3.0
rpart	Regression Trees Modelle	Version 4.1-8
rpart.plot	Regression Trees Visualisierung	Version 1.5.3
randomForests	Random Forest Modelle	Version 4.6-12
ranger	gewichtete Random Forest Modelle	Version 0.4.0
mgcv	GAM Modelle	Version 1.8-0
e1071	Support Vector Regression	Version 1.6-7
caTools	gleitende Mittelwerte etc.	Version 1.17
zoo	Datenaufbereitung	Version 1.7-13
pander	konfigurierbarer R Output	Version 0.6.0

9 LITERATURVERZEICHNIS

- Breiman, L. (1984). *Classification and regression trees*. Monterey, USA, California: Wadsworth & Brooks/Cole Advanced Books & Software.
- Breiman, L. (2001). Random Forests. *Machine Learning* 45, S. 5-32.
- Cox, D. (1972). Regression models and life tables. *Journal of the Royal Statistical Society*(34), S. 187 - 220.
- Drucker, H., Burges, C., Kaufman, L., Smola, A., & Vapnik, V. (1997). Linear Support Vector Regression Machines. NIPS'96.
- Efron, B. (1979). Bootstrap Methods: Another Look at the Jackknife. *The Annals of Statistics*(1), S. 1-26.
- Fahrmeier, L. (2009). *Regression - Modelle, Methoden und Anwendungen*. Heidelberg Dordrecht London New York: Springer.
- Hashemian, M., & Wendell, B. C. (2011). State-of-the-Art Predictive Maintenance Techniques. *IEEE TRANSACTIONS ON INSTRUMENTATION AND MEASUREMENT, VOL. 60*, S. 3480-3492.
- Hastie, T., & Tibshirani, R. (1986). General Additive Models. *Statistical Science*, S. 297-318.
- Hastie, T., Tibshirani, R., & Friedman. (2009). *The Elements of Statistical Learning*. Springer.
- Hastie, Trevor . (1990). *Generalized Additive Models*. Chapman & Hall/CRC.
- Holzer, Peter; Pözlbauer Georg. (2014). Data Mining im Online Condition Monitoring. *AKIDA 2014 - Data Mining im Online Condition Monitoring*. Aachen: Institut für Maschinentechnik der Rohstoffindustrie, RWTH Aachen.
- Jardine, A. K. (2006). A review on machinery diagnostics and prognostics implementing condition-based maintenance. In *Mechanical Systems and Signal Processing*.
- Jardine, Andrew K.S. . (2006). A review on machinery diagnostics and prognostics implementing condition-based maintenance. *Mechanical Systems and Signal Processing*(20), S. 1483-1510.
- Khanh, L., Fouladirad, M., Barros, A., Levrat, E., & Benoit, L. (2013). Remaining useful life estimation based on stochastic deterioration models: A comparative study. *Reliability Engineering and System Safety* 112, S. 165-175.

- Li, Z., Zhou, S., Choubey, S., & Sievenpiper, C. (2007). Failure event prediction using the Cox proportional hazard model driven by frequent failure signatures. *IIE Transactions*(39), S. 303-315.
- Meyer, D. (August 2015). Support Vector Machines. FH Technikum Wien.
- Mobley, R. (1990). *An Introduction to Predictive Maintenance*. New York,: Van Nostrand Reinhold.
- Mobley, R. K. (2002). *An introduction to predictive maintenance*. Butterworth-Heinemann.
- Müller, K.-R., Smola, A., Rätsch, G., Schölkopf, B., Kohlmorgen, J., & Vapnik, V. (1997). Predicting time series with support vector machines. *Lecture Notes in Computer Science Vol.1327*, S. 999-1004.
- Okoh, C., Roy, R., Mehnen, J., & Redding, L. (2014). Overview of Remaining Useful Life Prediction Techniques in Through-Life Engineering Services. *Procedia CIRP 16*, S. 158-163.
- O'Sullivan, Y. a. (März 1986). Automatic smoothing of regression functions in generalized linear models. *Journal of the American Statistical Association*, S. 96-103.
- Rakotomamonjy, A. (2003). Variable Selection Using SVM-based Criteria. *Journal of Machine Learning Research* 3.
- Reinsch, C. (3. February 1967). Smoothing by spline functions. *Numerische Mathematik*, S. 177-183.
- Ruppert, D. (2003). *Semiparametric Regression*. Cambridge: Cambridge University Press.
- Schölkopf, B., & Smola, A. (2002). *Learning with kernels: support vector machines, regularization, optimization, and beyond*. MIT press.
- Schölkopf, B., Burges, C., & Vapnik, V. (1996). Incorporating Invariances in support vector learning machines. *Artificial Neural Networks ICANN*, S. 47-52.
- Si, X.-S., Wang, W., Hu, C.-H., & Zhou, D.-H. (2011). Remaining useful life estimation – A review on the statistical data driven approaches. *European Journal of Operational Research* 213, S. 1-14.
- Silverman, B. (1985). Some Aspects of the Spline Smoothing Approach to Non-Parametric Regression Curve Fitting. *Journal of the Royal Statistical Society. Series B (Methodological)*, S. 1-52.
- Stone, M. (1974). Cross-Validatory Choice and Assessment of Statistical Predictions. *Cross-Validatory Choice and Assessment of Statistical Predictions*, S. 111-147.

- Umair, S., Masdi, B. M., & Karim, Z. A. (2014). Time series method for machine performance prediction using condition monitoring data. *International Conference on Computer Communications and Control Technology (I4CT)* , (S. 394-398).
- Vapnik, V. (1995). *The Nature of Statistical Learning Theory*. Springer NY.
- Vapnik, V., Golowich, S., & Smola, A. (kein Datum). Support Vector Method for function approximation, regression estimation and signal processing. NIPS'96.
- Widodo, A., & Yang, B.-S. (12. 1 2007). Support vector machine in machine condition monitoring and fault diagnosis. Korea: Pukyong National University.
- Widodo, A., Kim, E. Y., Son, J.-D., Yang, B.-S., Tan, A. C., & Gu, D.-S. (2009). Fault diagnosis of low speed bearing based on relevance vector. *Expert Systems with Applications*, S. 7252–7261.
- Wood, S. (2000). Modelling and Smoothing Parameter Estimation with Multiple Quadratic Penalties. *J.R.Statist.Soc.B 62(2)*;, S. 413-428.
- Wood, S. (27. 2 2006). *Generalized additive models: an introduction with R*. CRC Press.
- Xiongzi, C., Jinsong, Y., Diyin, T., & Yingxun, W. (2011). Remaining useful life prognostic estimation for aircraft subsystems or components: A review. *IEEE International Conference on Electronic Measurement & Instruments (ICEMI)*.
- Zhang, X., & Wu, Y. (2016). Variable selection for support vector machines in moderately high dimensions. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, S. 53 - 76.
- Zhang, X., & Wu, Y. (kein Datum). *Variable Selection for Support Vector Machines in High Dimensions*. North Carolina State University.

ABSTRACT

Anhand einer Industrieanlage wird gezeigt, wie Daten vom Messsystem zu Zeitreihen aufbereitet und effizient zusammengefasst werden. Auf diesen standardisierten Datensätzen werden die Regressionsmethoden Lineare Modelle, Regression Trees, Random Forests, GAM und Support Vector Regressions erstellt. Diese prognostizieren anhand der Messwert eine Restlebenszeit für die Anlage. Dies wird so umgesetzt, dass im Echtzeitbetrieb zu jedem Zeitpunkt eine Schätzung über den Anlagenzustand vorhanden ist. Dadurch können Wartungsmaßnahmen rechtzeitig geplant und eingeleitet werden. Weiters wird eine Simulation durchgeführt, bei der die Lauf- und Wartungszeiten der Anlage berechnet werden, falls man den Wartungsvorschlägen der Modelle folgt. Dadurch wird gezeigt, dass im Vergleich zur aktuellen Wartungsstrategie, durch die Predictive Maintenance Praktik eine erhöhte Anlagenverfügbarkeit zu erwarten ist.