

Schlussbericht vom 15.03.2023

zu IGF-Vorhaben Nr. 21425 N

Thema

ProCheck - Verifikation innerbetrieblicher Materialflussprozesse durch CPS-basiertes Process Mining

Berichtszeitraum

01.11.2020 bis 30.11.2022

Forschungsvereinigung

Bundesvereinigung Logistik (BVL) e.V.

Forschungseinrichtung(en)

1 Fraunhofer-Gesellschaft e.V., Fraunhofer-Institut für Integrierte Schaltungen IIS, Nordostpark 84, 90411 Nürnberg

2 Ludwig-Maximilians-Universität München (LMU), Lehrstuhl für Datenbanksysteme und Data Mining, Oettingenstr. 67, 80538 München

Gefördert durch:

Inhaltsverzeichnis

Thema.....	1
Berichtszeitraum.....	1
Forschungsvereinigung	1
Forschungseinrichtung(en).....	1
1. Forschungsziel und Lösungsweg	7
1.1. Forschungsziel.....	7
1.1.1. Motivation und Ziel des Projekts ProCheck	7
1.1.2. Wissenschaftlich-technische und wirtschaftliche Problemstellung	8
1.1.3. Innovativer Beitrag der Forschungsergebnisse und wirtschaftliche Bedeutung der erzielten Ergebnisse für KMU	11
1.2. Arbeitsplan	12
2. Forschungsergebnisse entlang der Arbeitspakete.....	15
2.1. Arbeitspaket 1 – Prozessaufnahme und -analyse, Identifikation relevanter und repräsentativer Prozesse	16
2.1.1. Interviewstudie mit den Anwendungspartnern	17
2.1.2. Prozessaufnahme vor Ort	18
2.1.3. Prozessanalyse.....	19
2.2. Arbeitspaket 2 – Modellierung maschinenlesbarer Soll-Prozessmodelle	21
2.2.1. Anforderungsdefinition an maschinenlesbare Prozessmodellierung	21
2.2.2. Erarbeitung von Soll-Prozessmodellen.....	22
2.2.3. Maschinenlesbare Modellierung von Soll-Prozessen.....	23
2.3. Arbeitspaket 3 – Aufbau eines CPS zur Erhebung prozessbezogener Rohdaten	23
2.3.1. Input.....	24
2.3.2. Basistechnologien	25
2.3.3. Anforderungserhebung.....	26
2.3.4. Technologieauswahl und Architektur.....	28
2.3.5. Entwicklung für die Systemkomponenten	30
2.3.6. Aufbau eines Demonstrators im Betrieb	31
2.3.7. Betrieb eines Demonstrators, Herausforderungen und Anpassungen	31
2.4. Arbeitspaket 4 – Entwicklung der automatisierten Extraktion von Event-Logs aus CPS- Rohdaten.....	34
2.4.1. Datenvorauswahl.....	34
2.4.2. Erkennung von Biegevorgängen mittels Kreuzkorrelation	36
2.4.3. Erkennung von Schweißvorgängen.....	38

2.4.4. Bauteilerkennung	39
2.5. Arbeitspaket 5 – Entwicklung von Verfahren zum Abgleich von Event-Logs mit Soll-Prozessmodellen	40
2.5.1. Input.....	40
2.5.2. Granularität und Trace Clustering	41
2.5.3. Basistechniken von Process Mining	41
2.5.4. Event Data Korrelation in Manufacturing	43
2.5.5. Experimentelle Ergebnisse.....	46
2.5.6. Fazit.....	53
2.6. Arbeitspaket 6 – Technische und prozessbezogene Evaluation	53
2.7. Arbeitspaket 7 – Dokumentation und Ableitung des Handlungsleitfadens für CPS-basiertes PM.....	54
3. Ergebniszusammenfassung und Ausblick	55
4. Verwendung der Zuwendung	57
4.1. Forschungsstelle 1: Fraunhofer IIS.....	57
4.2. Forschungsstelle 2: Ludwig-Maximilians-Universität München.....	57
5. Notwendigkeit und Angemessenheit der geleisteten Arbeit	57
6. Wissenschaftlich-technischer und wirtschaftlicher Nutzen insbesondere für KMU	58
7. Durchführende Forschungsstellen.....	62
8. Anhang.....	64
8.1. Befragte Experten im Rahmen von AP 1.....	64
8.2. Im Projekt involvierte Unternehmen während der Projektlaufzeit.....	64

Abbildungsverzeichnis

Abbildung 1: Der Forschungsansatz ProCheck im Überblick.....	8
Abbildung 2: Innovations- und Verwertungspotenzial durch ProCheck als Ökosystem.....	10
Abbildung 3: Schematischer Prozessverlauf des FP 1 nach Abteilungen.....	17
Abbildung 4: Schematischer Prozessverlauf des FP 2 nach Abteilungen.....	18
Abbildung 5: Beispielprozess modelliert mithilfe von Camunda.....	22
Abbildung 6: Hallenlayout FP 1 (schematisch).....	24
Abbildung 7: Hallenlayout FP 2 Transportgeräte (schematisch).....	25
Abbildung 8: Schematische Architektur des CPS: Sensorknoten (blau), Mini-PC zur Verarbeitung (rot), Wifi-Access Point zur Kommunikation (orange), Datenbank zur Speicherung (grün).....	29
Abbildung 9: Sensorknoten zur Stromstärkemessung (geöffnet).....	30
Abbildung 10: Sensorknoten an einer TRUMPF Biegepresse zur jeweiligen Positionsbestimmung des Biegebalkens bei FP 1.....	31
Abbildung 11: Erfassung von Position und Tiefenanschlag der Biegepresse und entstehende Sensordaten.....	32
Abbildung 12: Schweißzelle - kontinuierliche Erfassung von Lichtinformationen und Stromverbräuchen.....	32
Abbildung 13: Rohdatenstrom mehrerer Sensoren am Beispiel Biegepresse: 3-Achsen-Beschleunigungssensor (oben), Näherungssensor vertikal (Mitte), Näherungssensor horizontal (unten).....	35
Abbildung 14: Typischer Datenstrom beim Betrieb der Biegepresse: Distanzwerte des Pressbalkens (oben) mit charakteristischen Merkmalen und zugehörige Beschleunigungsdaten (unten); rot, grün, blau: Beschleunigung in verschiedenen Raumachsen; schwarz: Vektorbetrag der Gesamtbeschleunigung.....	35
Abbildung 15: Pressbalken lässt Pressvorgänge mit Distanzsensoren gut erkennen (oben); Tiefenanschlag gibt Werkstück-typische Maßinformationen (unten).....	36
Abbildung 16: Schematischer Ablauf der Kreuzkorrelationsberechnung auf den Daten.....	37
Abbildung 17: Beispielplot der Kreuzkorrelationsberechnung auf Biegepressedaten.....	37
Abbildung 18: Stromverbrauch (oben) und Lichtentwicklung (unten) bei unterschiedlich langen Schweißvorgängen.....	38
Abbildung 19: Werkstücke in der Biegepresse ergeben ein jeweils typisches Signalmuster des Distanzsensors.....	39
Abbildung 20: Lichtintensität des Lichtsensor in der Schweißzelle - Bauteil-bezogene Charakteristika und Wiederholungen.....	39
Abbildung 21: Schematische Darstellung für Ereignisdaten aus verschiedenen Systemen und mit unterschiedlichen Granularitätsebenen für Identifikatoren der Prozessausführung.....	43
Abbildung 22: Automatische Ermittlung zweier Prozessmodelle mithilfe algorithmischer Verfahren zur Prozesserkennung.....	45
Abbildung 23: Whiteboard-Aufzeichnung der Hierarchie innerhalb eines Artikels, angefertigt während eines Workshops bei FP 1.....	45
Abbildung 24: Beispiel für hierarchisch verwandte Objekte - Auftragsnummer und Artikel - ergeben mehrere mögliche Granularitätsebenen. Ähnliche Farben stehen für ähnliche Artikel, die ähnliche Varianten ausführen. (FP 1).....	46

Abbildung 25: Abstandsverteilungen nach angewendeten Distanzmetriken	47
Abbildung 26: DBSCAN Ergebnisse bei verschiedenen ϵ -Werten	47
Abbildung 27: Metriken pro Cluster für verschiedene Distanzen nach Verfahren.....	49
Abbildung 28: Granularitätsanalyse hinsichtlich der Anzahl der Ist-Prozessspuren (links), Anzahl an Artikeln (Mitte) und Auftragsnummern (rechts) per Cluster für DBSCAN	50
Abbildung 29: Häufigste Varianten eines jeden Clusters	51
Abbildung 30: Clustering-Ergebnisse mit DSCAN und mengenbasiertem Editierabstand. Nach dem Trace Clustering sind die Modelle weniger komplex als in Abbildung 22 (c)	52
Abbildung 31: Projektübersicht und Inhalte.....	55

Tabellenverzeichnis

Tabelle 1: Arbeitsplan zum Projekt ProCheck.....	15
Tabelle 2: Gegenüberstellung Prozessanalyse der Anwendungspartner	20
Tabelle 3: Veröffentlichungen zu ProCheck.....	62
Tabelle 4: Befragte Experten im Rahmen von AP 1.....	64
Tabelle 5: Im Projekt involvierte Unternehmen während der Projektlaufzeit.....	64

1. Forschungsziel und Lösungsweg

1.1. Forschungsziel

1.1.1. Motivation und Ziel des Projekts ProCheck

Hohe Variantenvielfalt, kleine Losgrößen und Flexibilität: Aktuelle Marktwünsche an produzierende Unternehmen erhöhen nicht nur die Anforderungen an die Produktion, sondern setzen transparente und optimierte Materialflussprozesse voraus. Die Aufgaben der innerbetrieblichen Logistik werden mit zunehmender Individualisierung der Produktion daher immer komplexer. Referenzprojekte und Literaturquellen (vgl. Lerch et al., 2017) zeigen, dass in der Materialflusssteuerung zwar ein großes Optimierungspotenzial steckt, betroffene physische Prozesse aber oftmals bei KMU nicht digital abgebildet sind. Ineffiziente Prozessabläufe, Verzögerungen, ungenügende Auslastung oder mangelnde Qualität sind die Folge. Ebenfalls wird die notwendige Optimierung der Materialflüsse aufgrund fehlender Transparenz über tatsächlich ausgeführte Prozesse zur Herausforderung.

Das Ziel von ProCheck ist es, eine automatisierte und kontinuierliche Analyse, Optimierung und Überprüfung von Materialflussprozessen in kleinteiligen Produktionsabläufen zu ermöglichen. Dabei fokussiert sich ProCheck auf räumlich verteilte Verarbeitungs- oder Montageschritte individueller Aufträge. Mehrere KMU aus der Branche verarbeitender Unternehmen sind im Projektbegleitenden Ausschuss (PA) vertreten. Diese Unternehmen sind Hersteller von Metalltransportbehältern, Antriebstechnik, technischer Keramik sowie Stanzwerkzeuge. Die dort vorliegenden Materialflussprozesse zeichnen sich durch räumlich verteilte Maschinen- oder Montageplätze aus, die von den Aufträgen individuell durchlaufen werden und eine auftragsbezogene Material- oder Werkzeugbereitstellung erfordern. Bei der geplanten Umsetzung setzt die Innovation des Forschungsvorhabens an: In ProCheck werden Methoden des Process Mining auch für physische Materialflussprozesse nutzbar gemacht. Process Mining (PM) umschreibt Analyse-Techniken, die das Erheben, Überprüfen und Optimieren von Prozessen anhand von Datenspuren, die bei der Prozessausführung in IT-Systemen entstehen, ermöglichen (vgl. Lu et al., 2017: S. 419). Zum Einsatz von PM für Materialflussprozesse werden als Datenbasis prozesskennzeichnende Informationen über deren tatsächliche Abläufe benötigt. Diese Datenbasis ist in den IT-Systemen jedoch für eine kleinteilige manuelle Produktion nicht vorhanden. Daher sollen Cyber-Physische Systeme (CPS) zur Erhebung von sensorischen Daten in den Prozessen genutzt werden, um über eine neuartige Datenverarbeitung die Datenbasis - so genannte Event-Logs - für PM zu schaffen. Cyber-Physische Systeme sind Netzwerke kleiner mit Sensoren und Aktoren ausgestatteter Computer, die als sogenannte Eingebettete Systeme in Materialien, Gegenstände, Geräte und Maschinenteile eingebaut und über das Internet miteinander verbunden werden. In einem derartigen Internet der Dinge verbinden sich die physische und die digitale Welt (vgl. Rajkumar et al., 2010). Event-Logs sind digitale Spuren (Traces), die Prozesse bei ihrer Durchführung in IT-Systemen hinterlassen. Folgende Informationen sind dabei Bestandteil von Event-Logs: Prozessschritt bzw. Event (z. B. Bearbeiten), Prozessinstanz (z. B. Auftrags-Nr.), Zeitpunkt (z. B. 13:45 Uhr). CPS helfen somit in ProCheck, Event-Logs über Materialflussprozesse aus der Realität in die IT-Systeme abzubilden. Wissenschaftliches Ziel des Projekts ist es, sensorische Rohdaten durch den Einsatz von CPS direkt in der Produktionsumgebung zu erheben und aus den Rohdaten automatisiert prozesskennzeichnende Informationen als Event-Logs zu generieren. Darauf aufbauend sollen geeignete PM-Verfahren für das Anwendungsfeld der Materialflussprozesse produzierender KMU zum Abgleich von Ist-Prozessausführungen mit Soll-Prozessmodellen - so genanntes Conformance Checking - entwickelt werden. Conformance Checking, als

eine Technik des PM, ermöglicht über den Abgleich der Event-Logs mit Prozessmodellen die schnelle und automatisierte Erkennung von Abweichungen und Verbesserungspotenzialen (vgl. Dumas et al., 2018).

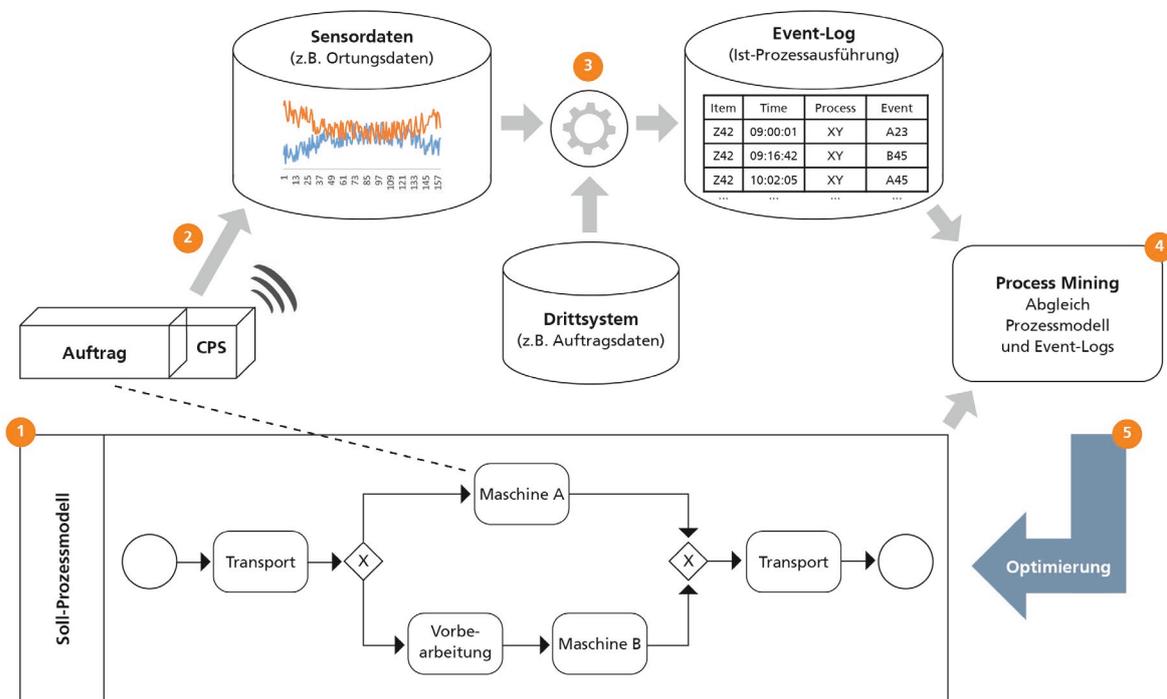


Abbildung 1: Der Forschungsansatz ProCheck im Überblick

1.1.2. Wissenschaftlich-technische und wirtschaftliche Problemstellung

Bei KMU übliche kleinteilige manuelle Produktionsabläufe bestehen aus vielen räumlich verteilten Prozessschritten und stellen hohe Anforderungen an die Materialflusssteuerung zur Bereitstellung von Teilen in der Produktion und zum Transport von Gütern von Maschine zu Maschine. Durch ineffiziente Prozessabläufe kommt es häufig zu Verzögerungen, ungenügender Auslastung von Ressourcen oder mangelnder Qualität (vgl. Arnold, 2006). Referenzprojekte mit KMU bestätigen, dass in der Materialflusssteuerung ein großes Optimierungspotenzial steckt (vgl. u. a. Förderprojekt Inwipro). Die wirtschaftliche Problemstellung liegt also in der Frage, wie Prozesse im Produktionsablauf optimiert werden können, um Kosten zu sparen, die Qualität zu verbessern oder die Durchlaufzeit zu verkürzen.

In der Intralogistik werden heute Prozessanalysen oftmals aufwändig manuell durchgeführt, da in der Praxis keine oder unzureichende Daten über die Prozesse existieren (vgl. Lieb et al., 2017). Betroffene physische Prozesse wurden häufig noch nicht im Rahmen der Digitalisierung adressiert. Insbesondere KMU setzen CPS-nahe Prozesse deutlich seltener ein als große Betriebe (vgl. Lerch et al., 2017). Dabei können CPS heute bereits Daten in nichtdigitalisierten Prozessen erheben (vgl. Ganschar, 2013; Pflaum et al., 2014; Hanak et al., 2018 & Forschungsprojekt R2D), liefern aber nicht die für PM notwendigen, prozesskennzeichnenden Event-Logs, sondern reine Sensordaten. In Bezug auf die wissenschaftlich-technische Problemstellung muss daher die Frage nach geeigneten und effizienten Auswertelgorithmen zur automatisierten Generierung prozesskennzeichnender Event-Logs auf Basis der durch CPS erhobenen Sensordaten gelöst werden.

PM wird aktuell vor allem für die Analyse von Geschäftsprozessen z. B. im Gesundheitswesen und der Finanzbranche eingesetzt, da in den Bereichen die Prozessabläufe in den IT-Systemen bereits größtenteils digitalisiert sind (vgl. Dakic et al., 2018). Hinsichtlich intralogistischer Prozesse ist hingegen zu erforschen, welche Konfiguration von PM-Verfahren und Werkzeugen geeignet ist, um die CPS-basierten prozesskennzeichnenden Informationen entsprechend zu verarbeiten und anschließend zu nutzen. Um eine effiziente Optimierung zu realisieren, soll zudem ein auf Materialflüsse spezialisierter, automatisierter Abgleich der Event-Logs mit den definierten Prozessmodellen geschaffen werden.

Die wirtschaftliche Bedeutung gut funktionierender Prozesse ist für produzierende KMU existenziell und ein wesentlicher Baustein zur Wettbewerbsfähigkeit auf dem Weltmarkt.

Das Innovationspotenzial von ProCheck liegt in der datenbasierten Analyse und Optimierung bisher nicht digitalisierter logistischer Prozesse produzierender KMU mittels CPS-technologiebasierter PM-Methoden. Die Intralogistik koordiniert in einer kleinteiligen manuellen Produktion die bedarfsgerechte Bereitstellung des Materials sowie den Transport der Güter von Maschine zu Maschine. In vielen Unternehmen des in ProCheck adressierten, verarbeitenden Gewerbes ist ein solches Produktionsumfeld mit räumlich verteilten Verarbeitungs- oder Montageschritten individueller Aufträge vorhanden. Darüber hinaus ist eine Übertragung der Ergebnisse in weitere Wirtschaftszweige bzw. Branchen möglich – u. a. Pharma & Healthcare und Logistikdienstleistungen (vgl. Pfohl, 2018). Oftmals handelt es sich um hochspezialisierte KMU, die im internationalen Wettbewerb stehen. Durch die in ProCheck angestrebte Zusammenarbeit mit mehreren produzierenden Unternehmen des PA wird ein Vorgehen für CPS-basiertes Process Mining von Materialflüssen entwickelt und die Ergebnisse branchenübergreifend an einen breiten potenziellen Nutzerkreis zurückgegeben.

Produzierende KMU erhalten die Möglichkeit, existierende CPS-Technologien flexibel und temporär einzusetzen, ihre spezifischen Produktionsprozesse digital zu erfassen und mittels PM zu analysieren und zu optimieren. Projektbeispiele des Unternehmens Capgemini zeigen, dass mit Process Mining Einsparungen der Prozesskosten von bis zu 20% möglich sind (Vgl. Eißel, 2019). Basierend auf Gesprächen mit Unternehmen des PA ergeben sich Nutzenpotenziale durch:

- die Vermeidung nicht-wertschöpfender Tätigkeiten im Prozess (bspw. Warte- und Suchzeiten, manuelle Eingaben in IT-Systeme). Lassen sich bspw. bei 100 Aufträgen pro Woche nur 10 Minuten pro Auftrag einsparen, können sich jährliche Einsparungen von ca. 30.000 € ergeben.
- geringere Kapitalbindung aufgrund kürzerer Durchlaufzeiten (bspw. Optimierung der Bearbeitungsreihenfolgen, weniger Rüstzeiten). In einem geschilderten Fall entfallen von ca. 8 - 13 Wochen Produktionszeiten etwa nur ca. 4 - 6 Wochen auf effektive Fertigungszeiten, der Rest sind Dispositions-, Wartezeiten und Sicherheitsreserven. Eingsparte ca. 4 - 7 Wochen würden die Kapitalbindung hierbei um rund 2 Mio. € senken und eine schnellere Lieferfähigkeit bringen.
- Fehlervermeidung und weniger Ausschuss aufgrund Qualitätsmängel durch falsche Abläufe. Ein Unternehmen schätzt, dass aufgrund von Prozessfehlern ca. 1 Charge im Wert von ca. 20.000 € pro Monat so fehlerhaft gefertigt wird, dass die Produktion wiederholt werden muss. Gelingt es, diesen Wert zu halbieren, können jährliche Einsparungen in Höhe von 120.000 € realisiert werden.

Die Wettbewerbsfähigkeit der KMU wird aufgrund zu erwartender kürzerer Reaktionszeiten auf individuelle Kundenwünsche, kleinerer Losgrößen und gesteigerter Prozesseffizienz und -effektivität gestärkt. Darüber hinaus können die KMU auch ihre externen Stakeholder durch eine höhere Transparenz in ihrer Produktion mit besseren Informationen bzgl. Lieferzeiten, Auftragsstatus etc. versorgen und somit das Vertrauen in die Zusammenarbeit stärken. Aber auch weitere Teilnehmer des adressierten Ökosystems (vgl. Abbildung 2) profitieren von ProCheck durch die Möglichkeit, ihre Produkte und Dienstleistungen zu verbessern und neue Geschäftsfelder zu entwickeln:

- System-/Software-Häuser können die geschaffenen Modelle, Algorithmen und Verfahren in ihre Planungs- und Steuerungssoftware für den Einsatz bei KMU integrieren.
- Berater können unter Nutzung des Handlungsleitfadens und technischer Lösungskomponenten eine weiterentwickelte Methodik zur datenbasierten Prozessberatung aufbauen.
- Hersteller von CPS-Komponenten können Produkte für PM anbieten und so neue Geschäftsfelder aufgreifen.

Bei den oben genannten Partnern im Ökosystem handelt es sich in vielen Fällen um KMU. Entsprechende Unternehmen finden sich im Projektbegleitenden Ausschuss.

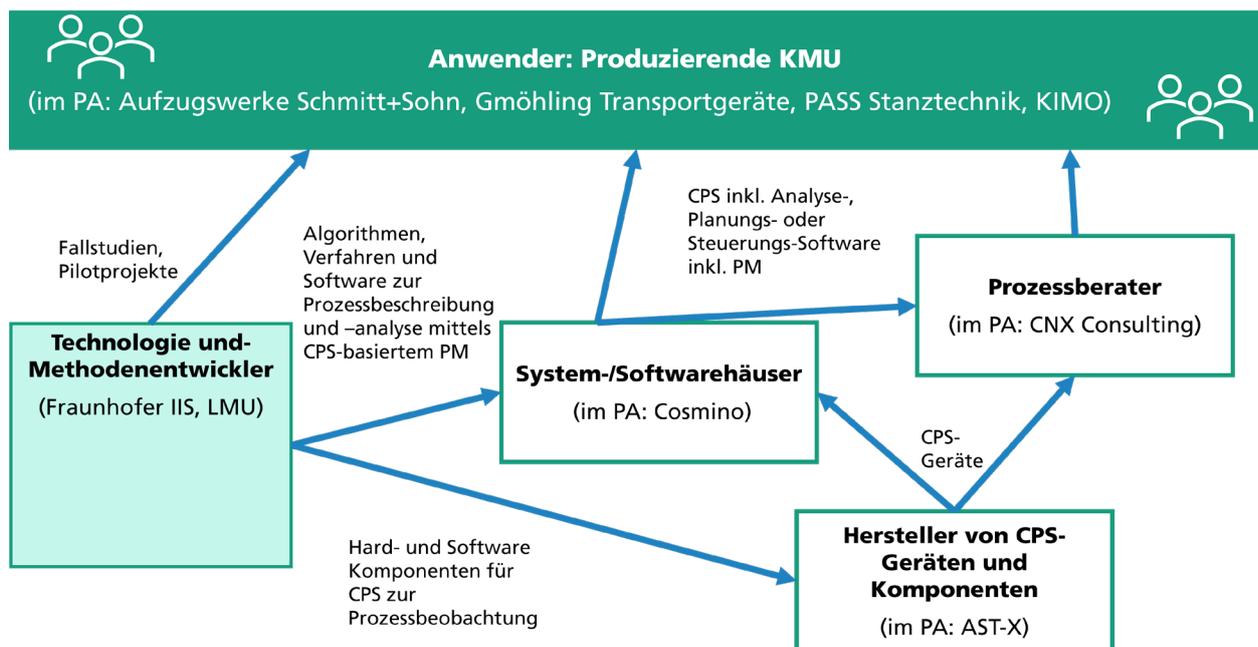


Abbildung 2: Innovations- und Verwertungspotenzial durch ProCheck als Ökosystem

Die in ProCheck entstehenden Lösungskomponenten sind daten- und modellbasiert und daher nicht spezifisch auf ein Unternehmen ausgelegt. Zudem verbleiben die Lösungskomponenten bei den Forschungseinrichtungen und können in Kooperation mit interessierten Unternehmen verwertet werden. Das Innovationspotenzial von ProCheck zeigt so nicht nur den adressierten Unternehmen des Projektbegleitenden Ausschusses Potenziale durch den Einsatz von CPS-basiertem PM für ihre Materialflussprozesse in der Produktion, sondern trägt durch seine breite Anwendbarkeit gesamtwirtschaftlich zur Steigerung der Ressourceneffizienz bei: Die Auslastung in Produktionsbetrieben wird erhöht, Durchlaufzeiten werden beschleunigt, Verschwendung wird durch optimierte Materialflüsse vermieden. Da Materialkosten (inkl. Energieverbrauch) mit rund

46% den größten Kostenblock des verarbeitenden Gewerbes ausmachen, kann davon ausgegangen werden, dass Effizienzsteigerungen in den Abläufen einen bedeutsamen Hebel zur Kostensenkung darstellen (Vgl. Statistisches Bundesamt, 2018).

1.1.3. Innovativer Beitrag der Forschungsergebnisse und wirtschaftliche Bedeutung der erzielten Ergebnisse für KMU

Das Innovationspotenzial von ProCheck aus Forschungssicht liegt in der Verknüpfung dreier, bisher weitgehend unabhängig betrachteter Forschungsbereiche:

1. Cyber-Physische Systeme zur Erfassung einzelner, bisher nicht digital abgebildeter Arbeitsschritte in der Produktion.
2. Erzeugung von Event-Logs zur gezielten Ableitung von Informationen aus den CPS-Sensordaten.
3. Process Mining, mit dem die erfassten Materialflussprozesse erhoben, analysiert und optimiert werden sollen.

Erzeugung von Event-Logs aus Sensordaten

Für die Erkennung von Ereignissen in den Rohdaten gibt es je nach eingesetzter Technologie sehr unterschiedliche Ansätze und Methoden. Diese werden für technisch verwandte Problemstellungen erfolgreich eingesetzt. So ist es möglich, bei der Auswertung von Positionsdaten aus Lokalisierungssystemen mittels Geofencing das Betreten oder Verlassen von Gebieten als Ereignis zu erkennen. Betrachtet man Signalstärkemessungen, wie sie als Rohdaten in signalstärkebasierten Lokalisierungssystemen anfallen, so lassen sich mittels Fingerprinting beispielsweise Räume aufgrund von unterschiedlichen Eigenschaften in der Signalausbreitung unterscheiden (vgl. Pahl, 2000 & Miclo et al., 2015). Hidden-Markov-Modelle oder der Einsatz von Partikelfiltern erlauben es, die Genauigkeit durch die Einbindung von Bewegungsmodellen zu steigern (Carrera et al., 2018). Solche Bewegungsmodelle könnten aus Sollprozessen abgeleitet werden, dabei darf aber die Empfindlichkeit des Systems zur Erkennung von Prozessabweichungen nicht leiden. Daten aus MEMS-Sensoren bilden oft die Grundlage für Algorithmen zur Bewegungsklassifikation oder Schritterkennung (vgl. Feldhorst, 2016). Methoden des maschinellen Lernens wie Clustering und Unsupervised Learning zeigen Strukturen auf, die anschließend durch Techniken wie Event Classification zur Aggregation neuer Muster genutzt werden. Im e-cosmos Projekt (vgl. Vieritz et al., 2013) bspw. wurde erfolgreich eine Gestenanalyse durch die Anwendung von hierarchischem Clustering durchgeführt, was zeigt, dass Methoden des Data Minings und maschinellen Lernens breit einsetzbar sind und in verschiedenen Feldern auf unterschiedlichste Weise eingesetzt werden können (vgl. Feldhorst, 2016 & Carrera et al., 2018).

Process Mining (PM)

PM (vgl. Van der Aalst, 2016) setzt sich aus drei großen Bereichen zusammen: In der Process Discovery geht es primär darum, aus Event-Log-Daten eine komprimierte Repräsentation des Prozessflusses zu generieren. Üblicherweise werden die Prozesse durch Petri-Netze oder äquivalente Netzstrukturen dargestellt. Im Conformance Checking greift man auf ein bestehendes oder evtl. durch eine vorangegangene Process Discovery ermitteltes Modell zurück, um weitere Prozessausführungen zu validieren. Gegebenenfalls werden in dieser Anomalieerkennungsmethode Abweichungen lokalisiert und im Modell hervorgehoben. Der dritte Bereich, das Model Enhancement, ist sowohl Optimierung des Modells mittels weiterer Prozessanalysen als auch die

Anreicherung mit weiteren Informationen wie Ressourcen, komplexen Operatoren oder weiteren Perspektiven. Für industrielle Produktionsprozesse geht es vor allem darum, die Prozessanalyse zur Ausführungszeit des Prozesses durchzuführen und mit den Ergebnissen direkt wieder in den Prozess einzugreifen. Techniken der Datenstromanalyse werden im PM dazu verwendet, statt Event-Logs Ströme aus Ereignissen zu verarbeiten und dynamische Modelle zu erstellen. Aktuelle Forschungsarbeiten (vgl. Burratin et al., 2014) benutzen Häufigkeitsanalysen der Ereignisrelationen und nutzen diese als Eingabe für statische Discovery-Algorithmen wie dem Heuristics Miner (vgl. Weijters et al., 2006). Durch verbesserte Speichernutzung konnten Effizienzeigenschaften verbessert werden (vgl. Hassani et al., 2015). Insgesamt ist die Umsetzung von PM-Techniken mit Ereignisströmen noch ein sehr junger Forschungsbereich (vgl. Burratin, 2018).

Kombinierte Ansätze für CPS-basiertes Process Mining

Auch wenn viele Methoden und Basistechnologien existieren, die in verwandten Anwendungsfeldern für sich bereits erfolgreich eingesetzt werden, ist die Generierung von Event-Logs aus CPS-Rohdaten speziell für PM noch nicht ausreichend erforscht. Einige kombinierte Ansätze nutzen Daten aus Lokalisierungssystemen und Geofencing für die Analyse von Prozessen. Dazu zählen eigene Vorarbeiten (vgl. Hanak et al., 2018) sowie die Anbindung von RTLS an Process Mining Software (vgl. Miclo et al., 2015). Diese Ansätze stoßen jedoch an ihre Grenzen, insbesondere dann, wenn mehrere Prozessschritte an demselben Ort stattfinden.

Da CPS-Rohdaten aus verrauschten Messungen der physischen Welt entstehen, unterliegen auch die daraus generierten Event-Logs unter Umständen noch gewissen Fehlern wie beispielsweise Sprünge zwischen Zuständen. Diese Fehler lassen sich durch Anpassung der Algorithmen speziell für PM reduzieren. Dabei ist eine enge Abstimmung zu den angewandten PM-Methoden notwendig, z. B. was die Stärke der Filterung betrifft. Weiterhin ist nicht jede Aktion, die in der physischen Welt erfasst wurde, aus Sicht der Prozessanalyse ein bedeutendes Ereignis. Die Entwicklung von Algorithmen und Methoden, die mit diesen zu erwartenden Phänomenen umgehen können, ist daher wichtiger Bestandteil des Projektes.

1.2. Arbeitsplan

Der Arbeitsplan ist in der nachfolgenden Tabelle 1 in gekürzter Form aufgeführt. Dabei werden jeweils pro Arbeitspaket (AP) die benötigten Inputs, eine kurze Arbeitspaketbeschreibung, die verwendende Methodik sowie die erwarteten Outputs aufgeführt.

AP 1	Prozessaufnahme und -analyse, Identifikation relevanter und repräsentativer Prozesse
Input	Projektstart. Bisherige Arbeiten zu den Themen Materialflussprozesse, CPS und PM (vgl. Literaturliste und Vorarbeiten der Forschungsstellen).
Beschreibung	<p>Erfassung manueller Prozesse, die mittels CPS nachverfolgt werden können und sich zur Anwendung von Conformance Checking mittels PM eignen.</p> <p>Beschreibung der Arbeitsschritte:</p> <ul style="list-style-type: none"> - Anpassung von Konzepten für Workshops und Fragebögen zur Identifizierung und Aufnahme von relevanten Prozessen - Auswahl potenziell geeigneter Prozesse in Workshops bei produzierenden Unternehmen (sowohl aus dem PA als auch außerhalb des PA) - Aufnahme und formelle Beschreibung der potenziell geeigneten Prozesse auf Basis von in Unternehmen vorhandener

	<p>Dokumentationen sowie durch Begehungen und persönliche Interviews mit verschiedenen Akteuren und Stakeholdern der betroffenen Prozesse in den jeweiligen Unternehmen</p> <ul style="list-style-type: none"> - Gemeinschaftliche Bewertung der erfassten Prozesse durch alle beteiligten Forschungseinrichtungen in Kooperation mit Praxispartnern hinsichtlich ihrer jeweiligen Eignung für das Forschungsprojekt; darauf basierend Auswahl der im weiteren Verlauf zu verwendenden Prozesse
Methodik	<ul style="list-style-type: none"> - Begehungen von Produktionsumgebungen, teilnehmende Beobachtung, Prozessmodellierung und -visualisierung (BPMN), leitfadengestützte Interviews, qualitative Inhaltsanalyse.
Output	<ul style="list-style-type: none"> - Abbildung geeigneter Materialflussprozesse bei 2 KMU
AP 2	Modellierung maschinenlesbarer Soll-Prozessmodelle
Input	<ul style="list-style-type: none"> - Repräsentative Prozesse aus der Fertigung beteiligter Unternehmen (AP 1).
Beschreibung	<ul style="list-style-type: none"> - Referenzprozesse für Conformance Checking (Analyse der Übereinstimmung von Event-Logs und Soll-Prozessen) im Rahmen von PM - Erfassung der Anforderungen für die Maschinenlesbarkeit von Prozessnotationen und -modellen - Erarbeitung von Soll-Prozessen auf Basis der ausgewählten Prozesse mit Vertretern der jeweiligen Unternehmen, insbesondere den Prozessverantwortlichen - Modellierung der maschinenlesbaren Soll-Prozesse unter Berücksichtigung der identifizierten Anforderungen und den Vorgaben der Unternehmen
Methodik	<ul style="list-style-type: none"> - Leitfadengestützte Interviews, Durchführung von Workshops, Prozessmodellierung in maschinenlesbarer Notation (bspw. BPMN)
Output	<ul style="list-style-type: none"> - Repräsentative, maschinenlesbare Soll-Prozesse von Materialflüssen produzierender KMU
AP 3	Aufbau eines CPS zur Erhebung prozessbezogener Rohdaten
Input	<ul style="list-style-type: none"> - Informationen über die zu untersuchenden Prozesse und deren räumliche Verteilung aus AP 2
Beschreibung	<ul style="list-style-type: none"> - Systeme zur Erhebung von CPS-Rohdaten sollen für den Einsatz bei produzierenden KMU zur Verfügung stehen. Dabei soll bereits bestehende CPS-Technologie des Fraunhofer IIS zum Einsatz kommen. Die CPS-Rohdaten sollen Informationen über die Umgebung bzw. den Aufenthaltsort enthalten, bspw. Identitätsnachrichten, Messungen der Signalstärken, Signallaufzeiten von Nachrichten von ortsfesten Ankerknoten oder Lage- bzw. Bewegungsinformationen. Diese können je nach Anforderung mit weiteren Daten, z. B. von einem NFC-Leser, erweitert werden. - Beschreibung der Arbeitsschritte: - Ableitung der wichtigsten Anforderungen aus den zu untersuchenden Prozessen, insbesondere die Anzahl der benötigten Geräte sowie Art und Frequenz der auszuleitenden Sensordaten; anschließende Anschaffung neuer oder Bereitstellung bestehender Hardware - Anpassung der Firmware der Geräte zur Erfassung und Ausleitung der vorher festgelegten Daten; Aufspielen der Firmware auf die Hardware und Konfiguration der Tags - Installation und Konfiguration weiterer Systemkomponenten wie

	<p>z. B. Gateway und Datenbankserver</p> <ul style="list-style-type: none"> - Test des Gesamtsystems unter Laborbedingungen - Installation und Betrieb des Systems bei zwei ausgewählten, produzierenden Unternehmen aus dem PA (min. 50 % KMU). Dabei ist vorgesehen, dass das CPS frühzeitig installiert und in den Produktionsprozess integriert wird, um eine Datenbasis für die nachfolgenden Arbeitspakete zu generieren. Das CPS wird temporär betrieben und bis zur Evaluationsphase auf einem aktuellen Software-Stand gehalten, um so Fortschritte in der Algorithmenentwicklung zu testen und auszuwerten.
Methodik	<ul style="list-style-type: none"> - Anforderungserhebung, Softwareentwicklung, Systemtest, Dokumentation
Output	<ul style="list-style-type: none"> - Ein in der Zielumgebung einsetzbares CPS inklusive Dokumentation zur Bedienung sowie zu den ausgeleiteten Daten. Prototypische Installation eines CPS in dem in AP 1 ausgewählten, aussagekräftigen Teilbereich der Produktion bei zwei beteiligten KMU sowie damit erfasste Rohdaten zu den Produktionsprozessen.
AP 4	Entwicklung der automatisierten Extraktion von Event-Logs aus CPS-Rohdaten
Input	<ul style="list-style-type: none"> - Rohdaten, welche durch das CPS erfasst werden (AP 3), sowie maschinenlesbare Referenzprozesse (AP 2)
Beschreibung	<p>Algorithmische Verfahren und prototypische Software zur automatisierten Erzeugung von Event-Logs als Basis für PM-Methoden. Hierbei werden die CPS-Rohdaten aus dem System von AP 3 zu Event-Logs verarbeitet.</p> <p>Beschreibung Arbeitsschritte:</p> <ul style="list-style-type: none"> - Recherche geeigneter Algorithmen und Verfahren zur Eventdetektion (z. B. Fingerprinting, Geofencing, Hidden-Markov-Modelle) - Festlegung auf Datenaustauschformat (z. B. XES) für die Event-Logs und Erhebung weiterer Anforderungen an eine Softwarekomponente zur Extraktion von Event-Logs aus Rohdaten. - Definition benötigter Softwarekomponenten und Funktionen - Implementierung der Softwarebausteine und Algorithmen - Tests der Funktionen der Software mit parallel erfassten Rohdaten aus AP 3
Methodik	<ul style="list-style-type: none"> - Literaturrecherche, Rapid Prototyping, Algorithmenentwicklung auf Basis von Realdaten, Softwareentwicklung, Unit-Tests
Output	<ul style="list-style-type: none"> - Software zur Erzeugung von Event-Logs aus CPS-Rohdaten inklusive Dokumentation
AP 5	Entwicklung von Verfahren zum Abgleich von Event-Logs mit Soll-Prozessmodellen
Input	<ul style="list-style-type: none"> - Event-Logs von realen logistischen Abläufen (AP 3 und AP 4), vordefinierte Prozessbeschreibungen (Soll-Prozesse) (AP 2)
Beschreibung	<ul style="list-style-type: none"> - Vergleichende Bewertung von Methoden zur Erkennung von Anomalien auf Ereignisebene und Prozessebene.
Methodik	<ul style="list-style-type: none"> - Konzeptentwicklung, Prototypentwicklung, empirische Evaluierung mit Realdaten.
Output	<ul style="list-style-type: none"> - Verfahren zum Abgleich von Event-Logs und den daraus resultierenden, detektierten Ist-Prozessen mit Soll-Prozessen.
AP 6	Technische und prozessbezogene Evaluation
Input	<ul style="list-style-type: none"> - Installation der CPS-Demonstratoren samt erfasster Daten (AP 3), Algorithmen zur Extraktion der Event-Logs (AP 4), Verfahren zum

	Abgleich mit Soll-Prozessen (AP 5)
Beschreibung	<ul style="list-style-type: none"> - Evaluation, inwiefern sich mit der eingesetzten CPS-Technologie der installierten Demonstratoren geeignete Daten zur Analyse der damit erfassten Prozesse erheben lassen. - Evaluation der automatisierten Algorithmen zur Generierung von für PM geeigneten Event-Logs hinsichtlich der Vollständigkeit und der inhaltlichen Qualität der entstehenden Event-Logs. Bewertung, inwiefern sich die Kombination aus eingesetzter CPS-Technologie und der darauf aufbauenden Algorithmen eignet, die untersuchten Prozesse vollständig zu erfassen. - Evaluierung der Erkennung von Abweichungen auf Prozess- und Ereignisebene sowie des Optimierungspotenzials für Materialflussprozesse. - Ganzheitliche Evaluierung des in ProCheck eingesetzten Verfahrens in Bezug auf das Zusammenspiel der einzelnen Lösungskomponenten sowie der Eignung des Gesamtsystems zur digitalen Abbildung und kontinuierlichen Überwachung der untersuchten Materialflussprozesse. - Ableitung des quantitativen Nutzenpotenzials.
Methodik	- Workshop mit dem PA zur Evaluation des Gesamtsystems, qualitative Analysen
Output	- Vollständige Evaluation des Gesamtsystems und der einzelnen Lösungskomponenten
AP 7	Dokumentation und Ableitung des Handlungsleitfadens für CPS-basiertes PM
Input	<ul style="list-style-type: none"> - Die Ergebnisse aller vorhergegangenen Arbeitspakete (insb. AP 4, AP 5, AP 6) - Ziel: Transfer der Projektergebnisse an interessierte KMU, Software-/Systemhäuser, Prozessberatungen und Hersteller von CPS-Technologien
Beschreibung	<ul style="list-style-type: none"> - Erstellung einer vollständigen Projektdokumentation auf Basis der Arbeiten bezüglich CPS, Erzeugung von Event-Logs, Online-Auswertung dieser mittels PM sowie der jeweiligen und übergreifenden Evaluation - Entwicklung eines praxisnahen Handlungsleitfadens für KMUs in Kooperation mit dem PA - Konzeptionierung eines Workshops zur Vermittlung der Projektergebnisse in Ergänzung zum Handlungsleitfaden und Erstellung von Material zur Durchführung des Workshops bei interessierten Parteien
Methodik	- Erstellung von Informationsmaterial
Output	- Projektdokumentation, Handlungsleitfaden, Material für Workshops

Tabelle 1: Arbeitsplan zum Projekt ProCheck

2. Forschungsergebnisse entlang der Arbeitspakete

Nachfolgend werden entlang des in Tabelle 1 dargestellten Arbeitsplans die erzielten Forschungsergebnisse detailliert beschrieben. In sieben Arbeitspaketen wurden unterschiedliche Erkenntnisse und Forschungsinhalte erarbeitet. Im Fall von Abweichungen zwischen dem im Antrag formulierten Vorgehen und den im Projekt angewendeten Methoden werden Hinweise darauf an den jeweiligen Stellen aufgeführt und die Unterschiede diskutiert.

2.1. Arbeitspaket 1 – Prozessaufnahme und -analyse, Identifikation relevanter und repräsentativer Prozesse

Die Inhalte des AP 1 wurden überwiegend von der FE 1 (Fraunhofer Arbeitsgruppe für Supply Chain Services SCS) erarbeitet. FP 1 und FP 2 beteiligten sich durch die Teilnahme an Experteninterviews und stellten ihre Räumlichkeiten für die Prozessaufnahme zur Verfügung. Die Ergebnisse wurden gemeinsam mit den Anwendungspartnern und dem Projektkonsortium bewertet. Auf diese Weise sind gemeinsame Ergebnisse entstanden.

Zunächst wurde im Rahmen einer Literaturrecherche eine aktuelle Wissensbasis über die Zusammenhänge und Verflechtungen der Nutzung von CPS und Process Mining im fertigen Gewerbe durchgeführt. Hierzu wurden auch die bereits bestehenden Erkenntnisse des Forschungsprojektes "AI4Pro - Methoden der Künstlichen Intelligenz zur Prognoseerstellung und Prozessoptimierung mittels Process Mining basierend auf Cyber-Physischen Systemen in der Produktion" des FuE-Programms "Informations- und Kommunikationstechnik des Freistaates Bayern genutzt.

Des Weiteren wurde der Forschungsgegenstand in einem Kickoff-Workshop mit Vertretern aus Industrie und Forschung diskutiert, relevante Fragen adressiert und potenzielle Anwendungspartner identifiziert. Zu den teilnehmenden Akteuren zählten Unternehmen des fertigen Gewerbes, Softwarehersteller, Hersteller von Hardware (insb. CPS), Universitäten sowie andere Interessierte (s. Anhang Kapitel 8.2). Den Teilnehmenden wurden dabei Fragen verschiedener Kategorien zur Beantwortung gestellt. Die Fragen zielten u. a. auf Themen wie

- dem Umgang mit typischen Prozessabweichungen und deren Identifikation,
- den eingesetzten Mitteln zur Erfassung, Beobachtung und Steuerung der physischen Materialflussprozesse,
- den Praxisbezug des Projektvorhabens und die Generalisierbarkeit der zu erzielenden Ergebnisse,
- weitere Anregungen an das Forschungskonsortium in Bezug auf das Forschungsvorhaben.

Aufgrund der damals einsetzenden Corona-Pandemie waren die beteiligten Unternehmen nur sehr eingeschränkt in der Lage, Ressourcen für Forschungsaktivitäten zur Verfügung zu stellen. Unter Berücksichtigung dieser Einschränkungen und basierend auf den Eindrücken und Diskussionen aus dem Workshop konnten zwei Unternehmen als potenzielle Anwendungspartner ausgewählt werden. Beide erwiesen sich aufgrund ihrer Fertigungsprozesse, ihrer Unternehmensstruktur und dem Grad der Digitalisierung für das Forschungsvorhaben am geeignetsten.

Erstens, die Aufzugswerke Schmitt+Sohn GmbH & Co. KG aus Nürnberg (im weiteren Verlauf dieses Berichts als Forschungspartner 1 oder kurz FP 1 bezeichnet), welche als fertiges Unternehmen von Aufzügen tätig ist.

Zweitens, die Gmöhling Transportgeräte GmbH aus Fürth (im weiteren Verlauf als Forschungspartner 2 oder kurz FP 2 bezeichnet), die sich auf die Produktion von Metallbehältnissen für die Bereiche Gesundheitswesen, E-Commerce, Entsorgung und weitere spezialisiert hat.

2.1.1. Interviewstudie mit den Anwendungspartnern

Mit beiden Anwendungspartnern wurden persönliche Interviews durchgeführt, um ein tieferes Verständnis für die Unternehmenscharakteristika sowie die physischen Materialflussprozesse zu erhalten. Den Unternehmen wurden dabei Fragen verschiedener Kategorien zur Beantwortung gestellt. Die Oberkategorien der Interviews enthielten jeweils Fragen

- zum Unternehmen und dem Interviewpartner allgemein,
- zu den jeweiligen Produktions- und Montageprozessen, deren Steuerung und Überwachung sowie die damit verbunden IT-Systeme (Grad der Digitalisierung),
- zur detaillierten Betrachtung der Produktionsprozesse wie der chronologische Ablauf, mögliche Abweichungen im Prozess, Dauer eines Prozessdurchlaufs, Dokumentation der Fertigungsprozesse, etc.,
- zu den Anforderungen an das Thema Process Mining und den potenziellen Nutzen für das Unternehmen,
- zu noch offenen Punkten.

Die Gesprächsleitfäden wurden auf die jeweilige Rolle des Unternehmens bzw. des Interviewpartners angepasst. Die Gesprächsinhalte wurden dann für die Arbeitspakete zur Modellierung maschinenlesbarer Soll-Prozessmodelle, zum Aufbau eines CPS zur Erhebung prozessbezogener Rohdaten und zur Entwicklung von Verfahren zum Abgleich von Event-Logs mit Soll-Prozessmodellen genutzt.

Im Anschluss wurden die Interviewergebnisse analysiert und gemeinschaftlich im Hinblick auf das Forschungsvorhaben bewertet. Aufgrund der vorherrschenden Unternehmens- und Prozesscharakteristika wurden in Abstimmung mit den Anwendungspartnern für das Forschungsvorhaben passende Materialflussprozesse ausgewählt. Entscheidend für die Auswahl der Prozesse waren neben Kriterien wie vorhandene, digitale Prozessinformationen auch die Art der Bearbeitungsschritte sowie realistische Optimierungspotenziale. Außerdem mussten die Prozesse einen gewissen Grad an Generalisierbarkeit aufweisen, um die wissenschaftliche Übertragbarkeit der zu entwickelnden Methoden sicherstellen zu können. Für Forschungspartner 1 (FP 1) wurde ein sehr linearer Prozess gewählt, welcher sechs Abteilungen umfasst und sich über zwei Fertigungsbereiche spannt. Die Bearbeitung erfolgt hauptsächlich händisch. Zu Beginn werden einzelne Rohmaterialien durch Lasern, Kanten, Bohren und Sägen bearbeitet, um diese innerhalb der Schweißerei zu Unterbaugruppen zu fertigen. Diese Unterbaugruppen werden am Prozessende zu einer Baugruppe zusammengefügt. Abbildung 3 skizziert den Prozessverlauf schematisch.

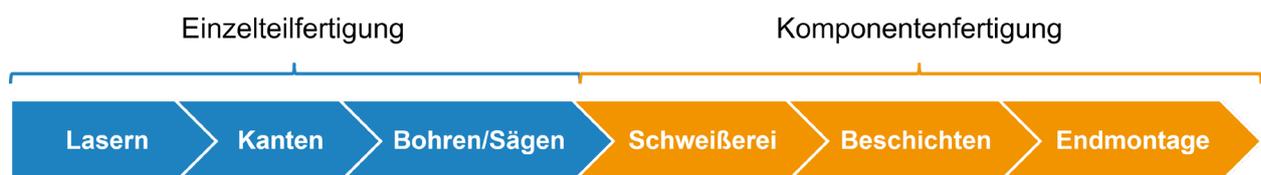


Abbildung 3: Schematischer Prozessverlauf des FP 1 nach Abteilungen

Beim Forschungspartner 2 (FP 2) ist der Prozess auch linear, jedoch finden sich zwei parallel verlaufende Linien in der Vorproduktion. Insgesamt werden sieben Abteilungen in zwei Fertigungsbereichen betrachtet. In der Vorproduktion werden die Bleche gelasert sowie gestanzt und zeitgleich die Profile bearbeitet. Beide Linien werden in der Endmontage zusammengeführt und durchlaufen dort die Abteilungen Werkzeug, Musterbau und Mechanik, Magazin und zuletzt eine finale Endmontage. In den jeweiligen Abteilungen finden sich diverse Prozessschritte, welche sowohl händische Bearbeitung von z. B. Blechen als auch Transport zwischen den Arbeitsplätzen beinhalten.

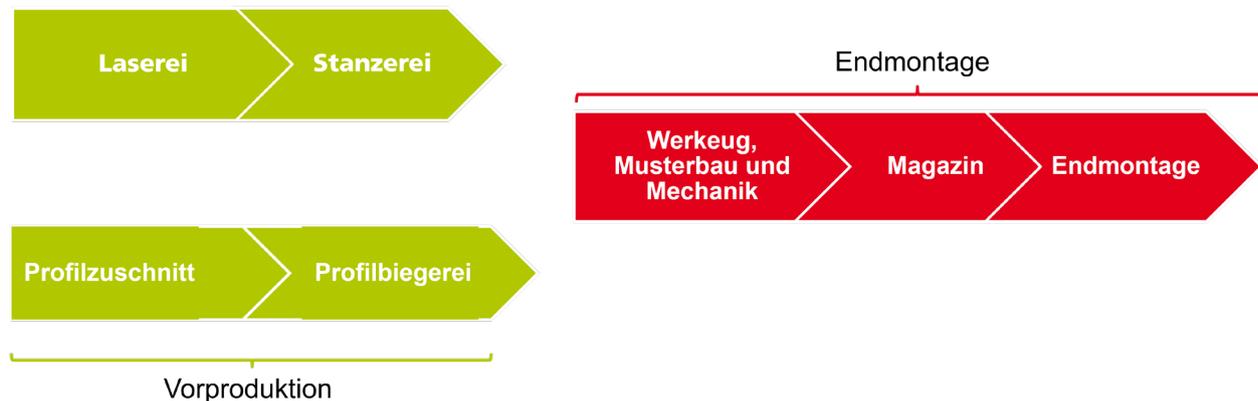


Abbildung 4: Schematischer Prozessverlauf des FP 2 nach Abteilungen

Auf die in ProCheck betrachteten Prozesse, deren Aufnahme und Analyse wird in den nachfolgenden Unterkapiteln detaillierter eingegangen. Die firmeninternen Prozesse enthalten sensible Unternehmensinformationen der Anwendungspartner und werden deshalb nicht detaillierter dargestellt. Die Abbildungen vermitteln einen schematischen Eindruck der im Prozess durchlaufenen Abteilungen und durchgeführten Bearbeitungsschritte.

2.1.2. Prozessaufnahme vor Ort

Anhand von Recherchen und der Experteninterviews im Rahmen von AP 1 wurden die für das Forschungsvorhaben passenden Fertigungsprozesse ermittelt. Für FP 1 wurde der Fertigungsprozess der Maschinenträger am Standort in Nürnberg (Kornburg) ausgewählt. Bei FP 2 wurde die Produktion des so genannten D-Wagens betrachtet. Bei der Auswahl geeigneter Produktionsprozesse arbeiteten beide Forschungseinrichtungen eng zusammen und es wurden nur gemeinsame Ergebnisse produziert. Insbesondere die Einschätzung der CPS-ExpertInnen des Fraunhofer IIS und der Process Mining-SpezialistInnen der LMU waren für den erfolgreichen Abschluss von AP 1 entscheidend.

Die tatsächlichen gegenwärtigen Materialflüsse innerhalb der Fertigung der Anwendungspartner wurden mithilfe einer Vor-Ort-Begehung erarbeitet. Hierzu wurden die Erkenntnisse aus den Interviews genutzt und die beiden Fertigungsprozesse in ihrem Ablauf Schritt für Schritt begleitet. Neben den physischen Prozessabläufen wurde ein Workshop zu den Datenbankenstrukturen durchgeführt, um relevante Prozessinformationen zu identifizieren. Von Interesse waren neben den Datenbanken auch die Art der Prozessinformationen und deren Nutzbarkeit für das Projektvorhaben. Da für Process Mining die Generierung von Event-Logs essenziell ist, musste erarbeitet werden, für welche Prozessschritte Informationen (z. B. Beginn der Tätigkeit, Ende der Tätigkeit, Dauer der Tätigkeit, Art der Tätigkeit, etc.) vorliegen und für welche nicht. Diese Erkenntnisse

waren nicht nur für die Modellierung der Fertigungsprozesse relevant, sondern lieferten auch Informationen für den Aufbau der CPS-Umgebung. Aus diesem Grund waren an den Workshops auch beide Forschungseinrichtungen beteiligt. Als Ergebnis sind Prozessmodelle der physischen Materialflussprozesse bei den Anwendungspartnern entstanden. Die modellierten Prozesse umfassen die entsprechenden Akteure in den jeweiligen Fachabteilungen und deren Tätigkeiten während der Fertigung im BPMN-Format. Der Grad der Prozessabdeckung – also die Anzahl an verfügbaren Datenquellen im Prozess – ist für das Forschungsvorhaben von großer Relevanz. Finden sich bereits nutzbare Prozessinformationen in Datenbanken, so müssen diese nicht zusätzlich über Sensorik erhoben werden. Auf der anderen Seite bieten die “Datenlücken” im Prozess eine passende Möglichkeit, relevante Informationen mithilfe von CPS zu erheben und somit ein ganzheitliches, digitales Abbild des Prozesses zu bekommen. Solche fehlenden Daten sind in der Praxis, v. a. bei KMU, aus diversen Gründen recht häufig. Es war demnach entscheidend, die vorhandenen Prozesse auf den Grad der Prozessabdeckung zu bewerten und somit die Grundlage für das weitere Forschungsvorhaben zu legen. Aus diesem Grund wurden die erfassten Prozesse um diese Informationen erweitert. Neben den physischen Prozessabläufen sind (auch unter Berücksichtigung der Aussagen aus den Interviews) ebenfalls die jeweiligen Informations- und Datenflüsse (Datenbanken, Qualität der Daten, etc.) aufgeführt. Diese Information-Lane in den Prozessmodellen ist für die Arbeiten in späteren APs essenziell und gibt Auskunft, an welchen Bearbeitungsschritten bereits Daten erhoben werden. Es galt, diese Informationen für den CPS-Aufbau, als auch für die Anwendung von Process Mining zu erfassen und bereitzustellen.

2.1.3. Prozessanalyse

Die Analyse der Experteninterviews und der Vor-Ort-Begehungen ergab, dass die Fertigungsprozesse der Anwendungspartner sowohl in Bezug auf die durchgeführten Tätigkeiten als auch auf den damit verbundenen Grad der Digitalisierung sehr heterogen sind. Tabelle 2 gibt einen Überblick über die Ergebnisse der Prozessanalyse auf Grundlage der Interviews und der Vor-Ort-Begehung. Auf der einen Seite findet sich mit FP 1 ein Unternehmen, welches viele manuelle, wenig maschinengetriebene Prozesse aufweist. Ebenso ist der Prozess, auch aufgrund der Fertigungslosgröße 1, sehr variantenreich und es ist demnach auch kein einheitliches Soll-Prozessmodell vorhanden. Insgesamt lässt sich ein sehr sequenzieller Materialflussprozess durch die Fertigung verzeichnen und die einzelnen Bearbeitungsschritte folgen einer Chronologie. So werden die entsprechenden Fachabteilungen der Einzelteillfertigung nacheinander durchlaufen, bis aus den bearbeiteten Blechen in der Komponentenfertigung erst Unterbaugruppen und dann ganze Baugruppen gefertigt werden. Zum intralogistischen Transport nutzt das Unternehmen Kommissionierwägen. Da die Prozesse einen geringen Digitalisierungsgrad aufweisen, finden sich wenig relevante Prozessinformationen (lediglich über Buchungen an den Arbeitsplätzen) und das Thema Process Mining wurde bisher nicht in Erwägung gezogen. Bei FP 2 zeigen sich zwar Parallelen in Bezug auf die Fertigungstiefe und die Prozessabfolge ist ebenfalls dynamisch, je doch ist der Grad der Digitalisierung deutlich ausgeprägter. Die Prozessschritte werden digital erfasst und es finden sich für Process Mining benötigte Informationen wie eine Eingangs- bzw. Ausgangsbuchung je Prozessschritt. Die gefertigte Losgröße variiert stark und es wird in nicht-wertschöpfende bzw. wertschöpfende Tätigkeiten differenziert. Die möglichen Optimierungspotenziale stellten aus wissenschaftlicher Sicht Kongruenz zum Forschungsvorhaben dar und deckten sich mit den in den PA-Veranstaltungen diskutierten Themen.

	Gmöhling Transportgeräte GmbH	Aufzugswerke Schmitt+ Sohn GmbH & Co. KG
Produktion	<p>Fertigung von Transportgeräten (Betrachtung von Typ D-Wagen)</p> <ul style="list-style-type: none"> • Sehr hohe Fertigungstiefe • Eine Fertigungslinie für Sonder- und Standardteile • Bearbeitung mit Sondermaschinen und handgeführten Maschinen • Fokus auf Erfüllung der zugesagten Termine 	<p>Fertigung von Aufzügen (Betrachtung der Maschinenträger)</p> <ul style="list-style-type: none"> • Sehr hohe Fertigungstiefe • Wenig maschinelle Bearbeitung • Fokus stark auf Endtermin gerichtet • Drei Wertschöpfungsstufen
Prozessablauf	<p>Parallele Bearbeitung von Blechen und Profilen, welche in der Endmontage zusammenlaufen</p> <ul style="list-style-type: none"> • Dynamisch, mit gewisser Grundordnung • Losgröße sehr unterschiedlich • Viel händisches Eingreifen in den Prozessablauf nötig • Unterscheidung in (nicht-) wertschöpfende Tätigkeiten 	<p>Chronologischer Prozessablauf, jedoch sehr variantenreich</p> <ul style="list-style-type: none"> • Dynamisch • Losgröße i. d. R. stets 1 (Ausnahme: Behälterartikel) • Prozess durchläuft Abteilungen der Bereiche Einzel- und Komponentenfertigung
Digitalisierung	<p>Monitoring über APS von etagis sowie ERP-System von abas</p> <ul style="list-style-type: none"> • Eingangs- sowie Ausgangsbuchung zu jedem Prozessschritt vorhanden • Erfassung jedes Prozessschrittes digital, aktuell und live • Prozessdaten detailliert vorhanden 	<p>Monitoring über Navision Business Central 365</p> <ul style="list-style-type: none"> • Keine Eingangs- und Ausgangsbuchung • Informationen im Auftragsterminplan über Termin-codes (ERP-System) • Buchung von Arbeitsgängen • Geringe Informationen über den Prozess
Optimierungspotenziale durch Process Mining	<ul style="list-style-type: none"> • Ortung von einzelnen Baugruppen im Lager • Conformance Checking/ Anomalieerkennung • Durchlaufzeiten ggf. Prognose 	<ul style="list-style-type: none"> • Prozesstransparenz • Conformance Checking/ Anomalieerkennung • Auftragsnachverfolgung • Identifikation von Liegezeiten • Quantifizieren von Durchlaufzeiten ggf. Prognose

Tabelle 2: Gegenüberstellung Prozessanalyse der Anwendungspartner

2.2. Arbeitspaket 2 – Modellierung maschinenlesbarer Soll-Prozessmodelle

Für AP 2 wurden zunächst Anforderungen für die Maschinenlesbarkeit von Prozessmodellen erfasst. Hier spielten neben Erfahrungen aus bereits vergangenen Forschungsprojekten hauptsächlich Literaturrecherchen eine entscheidende Rolle. In einem nächsten Schritt wurden die verfügbaren Prozessdaten in einem Workshop erarbeitet. Mithilfe der Ergebnisse aus AP 1 wurden die Fertigungsprozesse in Camunda überführt. Hierfür lag der Fokus primär auf den einzelnen Fertigungsschritten der physischen Materialflüsse. Informationen wie die zu den Prozessschritten gehörenden Datenflüsse und deren Art wurden in diesem AP bewusst nicht in die Modelle integriert, um den späteren maschinenlesbaren Abgleich zu gewährleisten. Dies wurde bereits während der Anforderungsdefinition festgelegt. Mit beiden Anwendungspartnern wurde iterativ und mittels mehrerer digitaler Austauschtreffen ein repräsentativer Stand der Prozessmodelle erarbeitet. So sind als Ergebnis Prozessmodelle im BPMN-Format der Soll-Prozesse von Materialflüssen der Anwendungspartner entstanden, welche durch XML-Code reproduziert und für Process Mining (z. B. für Conformance Checking) genutzt werden können. Auch für das AP 2 war FE 1 verantwortlich und wurde von den weiteren Projektbeteiligten unterstützt. Aufgrund des inhaltlichen Aufbaus der Forschungsprojektes war die Mitarbeit aller Personen im Projekt für eine erfolgreiche Bearbeitung wichtig.

2.2.1. Anforderungsdefinition an maschinenlesbare Prozessmodellierung

Konkret wurden folgende Anforderungen für die maschinenlesbare Modellierung festgelegt:

1. Prozessbezogene Anforderungen (Konzentration auf physische Materialflüsse und deren digitales Abbild): Die in AP 1 erarbeiteten Prozessmodelle enthalten eine Vielzahl an Informationen über Akteure, Datenbanken, Informationsflüsse, etc. Für den späteren maschinenlesbaren Abgleich sind diese zusätzlichen Informationen jedoch eher hinderlich und sorgen für fehlerhafte Ergebnisse. Aus diesem Grund wurde sich in AP 2 auf die physischen Prozessschritte fokussiert und alle zusätzlichen Informationen entfernt. Außerdem können nur solche Bearbeitungsschritte mittels Process Mining auf Konformität geprüft werden, die auch digital (z. B. über Scans, Auftragsdaten, CPS, etc.) erfasst werden.
2. Technische Anforderungen (Zusammenspiel mit Process Mining-Bibliotheken in z. B. Python und/oder R): Da der spätere Abgleich der Soll-Prozessmodelle mit den Ist-Prozessen datenbasiert mithilfe von Process Mining stattfinden soll, mussten vorab entsprechende Software-Bibliotheken und -pakete auf die Kompatibilität mit Code-Formaten (insb. XML) und Prozessnotationen (z. B. BPMN) bewertet werden.
3. Wissenschaftliche Anforderungen (Reproduzierbarkeit des Codes): Im Sinne der wissenschaftlichen Transparenz und Übertragbarkeit war es wichtig, die Reproduzierbarkeit der Ergebnisse zu gewährleisten. Demnach ist es entscheidend, für die maschinenlesbare Generierung gängige Code-Formate wie XML zu verwenden.
4. Wirtschaftliche Anforderungen (Einfachheit im Umgang mit der Software zur Prozessmodellierung): Da das Modellieren von Prozessen mithilfe einer Programmiersprache zeitaufwändig und unübersichtlich sein kann, wurde nach einem Tool gesucht, welches gängige Prozessnotationen unterstützt (z. B. BPMN) und automatisch Code (z. B. XML) ge-

neriert. So werden lediglich Kenntnisse über Modellierungssprachen benötigt und trotzdem Maschinenlesbarkeit gewährleistet. Um allen Projektbeteiligten Zugang zu den Ergebnissen gewähren zu können, wurde außerdem nach einem frei zugänglichen Tool gesucht. Letzteres war für das iterative Arbeiten an den Modellen von Vorteil.

Unter Berücksichtigung der definierten Anforderungen für die Maschinenlesbarkeit von Prozessmodellen wurde ein geeignetes Tool für die maschinenlesbare Modellierung von Prozessen identifiziert und auf Eignung bewertet. Entscheidend für die Auswahl waren neben weichen Faktoren wie dem Zugang zu den erarbeiteten Ergebnissen, auch das Zusammenspiel mit späteren Arbeitspaketen (insb. AP 5). Anschließend wurde sich für den Camunda Modeler entschieden. Camunda¹ ist eine frei zugängliche Software, mit deren Hilfe sich Prozesse in der Modellierungssprache BPMN darstellen lassen. Weiterhin wird im Hintergrund Code (XML) generiert, welcher wiederum für einen maschinenlesbaren Abgleich verwendet werden kann. Durch die nicht-lizenzbedürftige Software wurde der Zugang zu den Ergebnissen für alle Projektbeteiligten gewährleistet. Der XML-Code ermöglicht die Prozesse in digitaler Form zu verwerten und z. B. Conformance Checking durchführen zu können. Hierbei entsteht eine für KMU leicht handhabbare Prozessrepräsentation, welche keine spezifischen Kenntnisse in Programmiersprachen erfordert. Darüber hinaus besteht die Möglichkeit, die vorhandenen Informationen in den Prozessmodellen einfach abzurufen und für weiterführende Arbeiten zu nutzen. Somit erfüllt der Camunda Modeler alle vorab definierten Anforderungen. Abbildung 5 zeigt exemplarisch einen in Camunda abgebildeten Prozessverlauf.

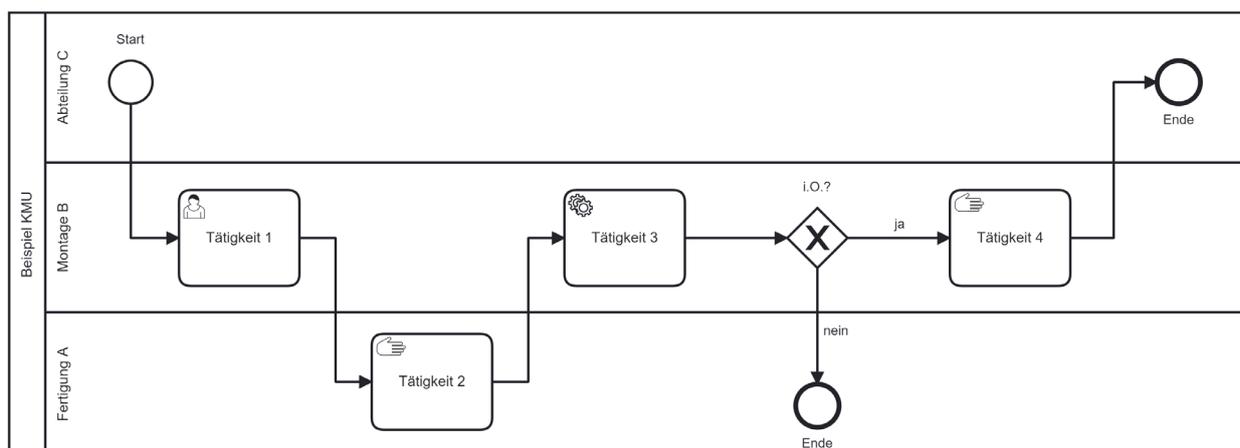


Abbildung 5: Beispielprozess modelliert mithilfe von Camunda

2.2.2. Erarbeitung von Soll-Prozessmodellen

Grundlage für die in AP 2 verwendeten Soll-Prozessmodelle bildeten die Arbeiten aus dem ersten Arbeitspaket. In AP 1 wurden bereits Soll-Prozesse festgelegt, deren Datenflüsse erarbeitet und Prozessmodelle erstellt. Es lagen demnach schon relevante Ergebnisse vor, welche gemeinsam mit den Prozessverantwortlichen der Anwendungspartner genutzt wurden, um maschinenlesbare Soll-Prozessmodelle erstellen zu können. Der Fokus lag demnach an dieser Stelle darin, die bisherigen Ergebnisse zu strukturieren und an den derzeitigen Forschungsstand anzupassen. Zum einen war es entscheidend, lediglich die Prozessschritte zu modellieren, welche auch in digitaler Form (z. B. als Buchungen in ERP) vorliegen, um überhaupt einen Abgleich mithilfe von Process

¹ Detaillierte Informationen finden sich unter: <https://camunda.com/de/platform/modeler/> (Letzter Abruf: 02.03.2023)

Mining durchführen zu können. Zum anderen waren die vordefinierten Anforderungen an die Maschinenlesbarkeit relevant. Entscheidend für das weitere Projektvorhaben und insb. für die Anwendung von Process Mining war es, lediglich die Prozessschritte zu modellieren, welche auch digital abgebildet werden können. Hierfür wurden zunächst der komplette Prozessdurchlauf erfasst und es gab demnach in Bezug auf den physischen Ablauf der Bearbeitungsschritte primär keine Veränderungen zu den in AP 1 festgelegten Prozessen. Im Sinne der iterativen Weiterentwicklung des Forschungsvorhabens galt es, diesen Grad der digitalen Prozessabdeckung regelmäßig neu zu bewerten und die Prozessmodelle entsprechend anzupassen. Insbesondere die Fachexpertise der KollegInnen zum Aufbau eines geeigneten CPS in der Zielumgebung galt es abzuwarten. Je nach Umsetzung, Art der generierten Rohdaten und v. a. nach auf diesem Weg zusätzlich abgebildeten Prozessschritten, wurde die Soll-Prozessmodelle editiert.

2.2.3.Maschinenlesbare Modellierung von Soll-Prozessen

Die maschinenlesbare Modellierung der betrachteten Fertigungsprozesse erfolgte in drei Schritten. Zunächst mussten die Prozessmodelle aus AP 1 mit den definierten Anforderungen aus AP 2 abgeglichen werden. In bereits vergangenen Projekten zeigte sich, dass ein Überfluss an Informationen zu Fehlern bei der späteren, maschinellen Verarbeitung führt. Da in den bestehenden Modellen eine Vielzahl an Informationen (Datenbanken, Akteure, Informationsflüsse, etc.) für das Forschungsvorhaben erarbeitet wurden, mussten die Modelle auf die physischen Prozessschritte der Materialflüsse limitiert werden. In einem zweiten Schritt wurden die betrachteten Fertigungsprozesse der Anwendungspartner in maschinenlesbarer Form modelliert. Hierbei wurde sich aufgrund der bereits erwähnten Gründe auf den physischen Prozessablauf fokussiert. Die in AP 2 entstandenen Modelle sind demnach weniger informativ, dadurch jedoch für die Anwendung von Process Mining-Methoden geeignet. Auch bei diesen Arbeiten wurde eng mit den Anwendungspartnern zusammengearbeitet und die Prozessmodelle über Expertengespräche erarbeitet. Weiterhin waren auch hier beide Forschungseinrichtungen beteiligt und es wurden gemeinsame Ergebnisse geschaffen. In einem letzten Schritt wurde die Maschinenlesbarkeit getestet. Hierzu wurden die erarbeiteten Prozessmodelle gemeinsam mit den Forschenden der LMU in Python hochgeladen und auf Basis des XML-Codes reproduziert. Dieser Schritt war entscheidend, um das AP 2 vernünftig abzuschließen. Es konnte gezeigt werden, dass Materialflussprozesse bei KMU des fertigen Gewerbes maschinenlesbar modelliert und beispielhaft für den Einsatz von Process Mining-Verfahren genutzt werden können. Im Sinne der iterativen Arbeit wurden die Prozesse regelmäßig an den Projektstand angepasst. Dieses Vorgehen war notwendig,

2.3. Arbeitspaket 3 – Aufbau eines CPS zur Erhebung prozessbezogener Rohdaten

Die Bearbeitung von Arbeitspaket 3 lag in erster Linie bei FE 1 (Bereich für Lokalisierung und Vernetzung) und erfolgte in Rücksprache mit den restlichen im Projekt involvierten Parteien. FP 1 und FP 2 stellten ihre Räumlichkeiten für Analysen und den Aufbau eines CPS zur Verfügung.

Für AP 3 war es zentral, für die Arbeitsweise des Anwendungspartners geeignete Daten zu erheben. Für die Rohdatenerhebung wurde dazu zusätzliche Sensorik genutzt, um physikalisch repräsentative Effekte einbeziehen zu können, die üblicherweise außer Acht gelassen werden und nicht erhoben werden. Dieser Ansatz trägt der Tatsache Rechnung, dass nur ausgewählte Anlagen und Maschinen direkt Informationen liefern können, diese Anlagen sind aber nicht typisch für die Fertigung von kleineren und mittleren Unternehmen. Es ist daher relevant, Informationslücken zu schließen und prozessrelevante Daten zu erheben, die typischerweise nicht und

keinesfalls flächendeckend vorhanden, gleichzeitig zur Zielerreichung aber absolut notwendig sind.

2.3.1. Input

Die wichtigste Basis für Entscheidungen und Untersuchungen zur Rohdatenerhebung bildeten die zu observierenden Prozesse.

Für die Abbildung der abstrakten Prozesse in die reale Produktionswelt, den Shopfloor, waren die Hallenpläne eine wichtige Grundlage. Typischerweise sind die Produktionen von kleinen und mittleren Unternehmen in funktionale Abteilungen untergliedert. Dies ist insbesondere der Fall, wenn, wie hier, eine werkstattähnliche Produktion mit hoher Variantenvielfalt und vielen manuellen Arbeits- und Transportschritten vorzufinden ist.

Forschungspartner 1

Das Hallenlayout von FP 1 ist in Abbildung 6 dargestellt.

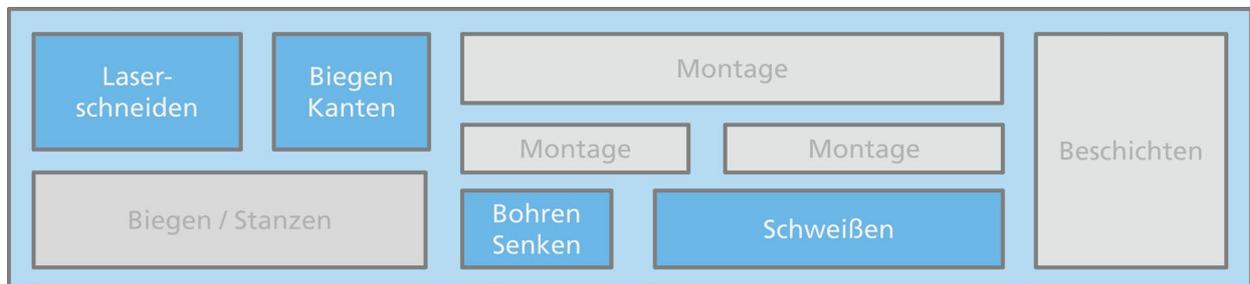


Abbildung 6: Hallenlayout FP 1 (schematisch)

Wie in AP 1 beschrieben beginnt ein typischer Prozessdurchlauf links oben mit der Anlieferung der Rohbleche, aus denen dann die nötigen Formen per Laser ausgeschnitten werden. Die ausgeschnittenen Blechstücke werden dann mit Wagen der Reihe nach zu den Bereichen Biegen/Kanten, Bohren/Senken und Schweißen transportiert. Der mögliche Einsatz eines CPS endet vor dem Beschichten, da der Einsatz von Farben und Lacken den Einsatz von elektronischen Geräten - insbesondere von provisorischen Sensorknoten - unmöglich macht. Außerdem ist eine Trennung der Bauteile nach Fertigungsaufträgen beim Beschichten so gut wie nicht vorhanden. Die sensorische Überwachung der Montage erschien durch die Varianz an Bearbeitungsschritten und dem Mangel an klar zuordenbaren Sensorsignalen als nicht vielversprechend. Das Laserschneiden erfolgt durch eine moderne Maschine, die detaillierte Daten inklusive Zeitstempel aufzeichnen kann. Für die restlichen Bearbeitungsschritte sind nur tägliche Buchungen im ERP-System vorhanden.

Forschungspartner 2

Das Hallenlayout von FP 2 ist hier schematisch dargestellt.

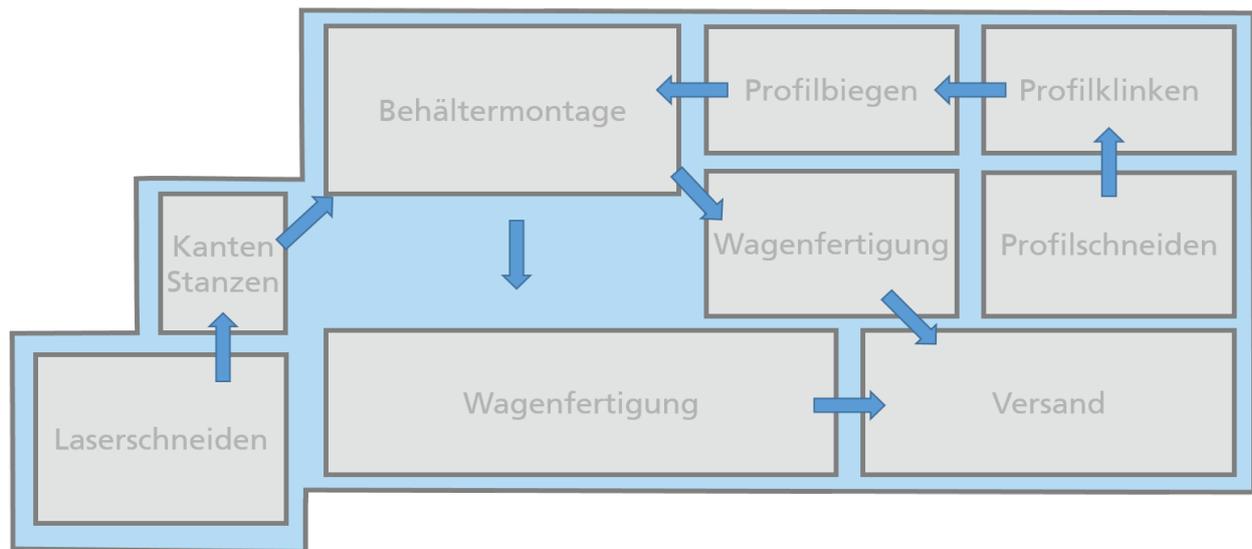


Abbildung 7: Hallenlayout FP 2 Transportgeräte (schematisch)

Ein typischer Prozessablauf beginnt zweigleisig beim Laserschneiden und beim Profilschneiden. In der Behältermontage laufen die getrennt gefertigten Teile zusammen und werden dann in der entsprechenden Wagenfertigung fertiggestellt. Abgeschlossene Arbeitsschritte werden mittels Barcode-Scannern gebucht, allerdings findet dies oft nicht zeitnah genug statt.

2.3.2. Basistechnologien

Unter einem Cyber-Physischen System (CPS) versteht man die Verbindung von eingebetteten Systemen, die physikalische Größen mittels Sensoren messen oder diese über Aktoren verändern, mit einem Kommunikationsnetz (vgl. Sanislav et al., 2012). Über dieses Kommunikationsnetz tauschen die eingebetteten Systeme untereinander oder auch mit anderen vernetzten Computersystemen Daten aus. Solche zusätzlichen – meist zentralen – Computersysteme kommen immer dann zum Einsatz, wenn Rechenleistung oder Hardware-Ressourcen auf den eingebetteten Systemen selbst nicht für eine Verarbeitung der gemessenen Sensordaten ausreichen oder die Daten mehrerer verteilter Sensoren miteinander verknüpft werden sollen.

Die Varianz von eingebetteten Systemen in CPS reicht von Kleinstgeräten mit häufig nur einem Sensor und einer Speicherausstattung von wenigen Kilobytes bis hin zu Mini-PCs wie z. B. dem Raspberry Pi mit Gigabytes an Speicher, und mehreren Sensoren und Netzwerktechnologien. Die Größe des Gerätes hat über die Leistung und Hardware-Ausstattung hinaus natürlich auch Einfluss auf die Kosten und den Stromverbrauch, welcher insbesondere wichtig wird, falls das eingebettete System mit einer Batterie oder sogar über Energy Harvesting betrieben werden soll.

An Sensoren gibt es ebenso eine Vielzahl verschiedener Typen, welche beispielsweise physikalische Größen wie Beschleunigung, Drehbeschleunigung, elektrische Felder, Magnetfelder, Licht, Lautstärke, Feuchtigkeit, Luftdruck, aber auch komplexe Signale wie z. B. Bilder oder Geräusche erfassen können. Die Anbindung der Sensoren erfolgt dabei über standardisierte Hardware-Schnittstellen, wobei komplexe Signale aufgrund des erhöhten Datenaufkommens andere Schnittstellen (bspw. I2S) erfordern als einfache Signale (z. B. TTL, I2C, SPI, UART).

Die Kommunikation in einem CPS kann für stationäre Sensorknoten auch drahtgebunden (z. B. Ethernet) erfolgen, aus Flexibilitätsgründen werden jedoch meist drahtlose Funktechnologien (z. B. WiFi, Bluetooth, LPWAN, ZigBee, Thread) verwendet. Die Funktechnologie muss dabei einen Kompromiss aus vielen erstrebenswerten, aber konkurrierenden Eigenschaften eingehen, wie z. B. Datenrate, Latenz, Reichweite, Energieeffizienz, Netzgröße, usw.

Die externe Verarbeitung von gesammelten Sensordaten kann entweder auf herkömmlicher PC-Hardware aber auch in der Cloud stattfinden. Hierzu können verschiedenste Algorithmen wie z. B. aus der Künstlichen Intelligenz verwendet werden. Entscheidend sind dabei üblicherweise die Rechenleistung und der Speicherplatz des Systems.

2.3.3. Anforderungserhebung

Die Anforderungen für die Konzeption eines CPS im industriellen Umfeld werden zum Teil in Interviews mit den Industriepartnern, aber auch durch Begehungen, Laborexperimente oder Tests vor Ort ermittelt. Dabei spielen die vorhandene Infrastruktur (z. B. Stromversorgung, Kommunikation), betriebsbedingte Ausschlusskriterien (z. B. Anforderungen bzgl. Staubschutz und Spritzwasserschutz oder auch Datenschutzerfordernungen), operationelle und organisatorische Aspekte (z. B. wie können oder sollen die Mitarbeiter mit dem CPS interagieren?), sowie die Anforderungen an die Qualität, Quantität und Frequenz der zu erfassenden Daten eine Rolle. Dabei können die vorhandene Infrastruktur sowie die Ausschlusskriterien üblicherweise grob in Interviews erhoben werden, aber häufig müssen Details bei einer Begehung geklärt oder sogar durch Tests vor Ort ermittelt werden. Unter operationellen Aspekten gilt es zu erfassen, welche Änderungen im Betriebsablauf für den Betrieb des CPS akzeptabel sind, ggf. kann sich hier ein Zielkonflikt bzgl. des Umfangs und der Häufigkeit der zu erhebenden Daten ergeben. Kandidaten für vielversprechende Sensoriken sowie die voraussichtlich nötige Datenmenge und -Qualität lassen sich bei Begehungen vor Ort und Beobachtung des Fertigungsbetriebs ermitteln, müssen aber in Laborexperimenten sowie Vor-Ort-Tests verifiziert werden.

Generelle Anforderungen

Bei beiden Anwendungspartnern ergaben sich folgende grundlegenden Rahmenbedingungen und Anforderungen:

- Ein Einsatz von mobiler Sensorik an den Bauteilen wurde bei FP 1 als unnötig erkannt, da die Einführung und der Einsatz von Barcode-Scannern zur Verfolgung von Bauteilen geplant waren. Bei FP 2 waren Barcode Scanner im Einsatz, trotzdem kann mobile Sensorik an Bauteilen von Vorteil sein.
- Genügend Stromquellen für die stationäre Sensorik waren vorhanden, so dass Batteriebetrieb nur bei mobiler Sensorik nötig ist.
- Ein firmeneigenes WLAN war ebenfalls vorhanden, sollte aber aus Sicherheitsgründen nicht für die Anbindung von CPS verwendet werden; ein kabelgebundenes Netzwerk schied aus praktischen Gründen aus.
- Etliche Vorgänge, insbesondere Biege- oder Schweißvorgänge spielen sich im einstelligen Sekundenbereich ab, daher war eine Sub-Sekunden-Auflösung der Datenerfassung nötig.
- Für die Biegevorgänge wurde die genauere Erfassung der Bewegung des Pressbalkens als interessant erkannt (Abstand zu Referenzpunkten, Vibrationen, Geräusche).

- Für die Schweißvorgänge wurde die sensorische Erfassung des Lichtbogens als relevant erkannt (Lichtintensität, Stromstärke, Geräusche).

Anforderungen im Detail

Insbesondere für den FP 2 ergaben sich folgende weiteren Rahmenbedingungen und Anforderungen:

- Das Operationalisieren der Materialflussprozesse ist relevant, um variantenreiche, kleinteilige und flexible Produktion möglichst kleiner Losgrößen zu realisieren.
- Eine automatisierte, schnelle und kontinuierliche Analyse der Materialflussprozesse ist die Grundlage für die Ableitung von Verbesserungen.
- Es soll eruiert werden, inwieweit aktuell verfügbare CPS-Technologien hier Daten generieren können aufgrund derer man auf die tatsächliche Prozessausführung schließen kann.
- Die Produktion umfasst ca. 10.000 verschiedene Artikel p.a. mit 15-30 verschiedenen Fertigungsstationen. Die Produkte bestehen aus (meist) großvolumigen Behältern, welche aus Aluminium gefertigt werden.
- Die Kundenstruktur (Vertriebsgesellschaften etc.) führt zu Aufträgen, die mit Stückzahlspitzen erhebliche Herausforderungen an terminliche und räumliche Kapazitäten stellen.
- Wichtig ist zu wissen, wo sich die Komponenten eines Auftrags gerade befindet (im Prozess und insbesondere räumlich, vor dem Hintergrund teils kreativen Lagerung wegen akuter Platzprobleme).
- Zentrale Herausforderungen entstehen durch die begrenzten Kapazitäten und deren Handhabung: limitierter Platz → Art der (Zwischen-) Lagerung, begrenzte Arbeitszeitkapazität → Optimierung im 2-Schicht-Betrieb, kurzfristige Reaktion auf Lieferanforderungen → Umpriorisierung und Anforderung zusätzlichen, nicht etabliert eingearbeiteten (Leih-) Personals
- Es sind keine festen Warenträger vorhanden.
- Es gibt eine große Bandbreite von Halbzeugen großer Fertigungstiefe von Kleinteilen zu Profilen, Rahmen und insb. Behältermänteln.
- Der größte Teil der Werkstücke besteht aus Aluminium, so dass keine magnetische Tag-Montage möglich ist.
- Eine Einzelstückkennzeichnung ist während der Fertigung bis zur Montage i. A. nicht notwendig oder zielführend.
- Insbesondere die Mäntel und natürlich (teil-) montierte Behälter konsumieren durch ihr Volumen enorme Flächen in Lagerung und Produktion; hier ist das Auffinden und Zuordnen von Auftragsposten die wichtigste Herausforderung.
- Die Situation aus stark begrenzten räumlichen Kapazitäten und hoher Personalfuktuation verstärkt diese Herausforderung, da u. a. Aufträge nur verteilt gelagert und z. B. am nächsten Tag nicht mehr identifiziert und zugeordnet werden können.

- Häufig können Aufträge nicht in einer Schicht vollständig bearbeitet und abgeschlossen werden, so dass nicht nachvollzogen werden, welchen Status ein Auftrag hat und wo alle bereits gefertigten Teile lagern.
- Im vorliegenden System werden bereits mittels Scannern Auftragsbezogene Daten erhoben. Konkret wird die Fertigstellung eines Arbeitsschrittes in der jeweiligen Abteilung verbucht. Diese Buchung erfolgt aber weitestgehend nicht zeitnah zur Tätigkeit, sodass anderweitig Informationen näher an der Echtzeit des Vorgangs erhoben werden sollten.
- Weiter ist keine zeitgerechte Information im System vorhanden über Lagerort, Fertigstellungsgrad von noch nicht vollständig bearbeiteten Aufträgen oder Materialzufluss vorhanden.
- Manuelle Tätigkeiten sollen nicht zusätzlich entstehen oder sogar reduziert oder ersetzt werden, wie z. B. das Scannen von Tätigkeiten oder das Informieren von Abteilung zu Abteilung durch das Planungspersonal.
- Für die nicht eingearbeiteten oft temporär Angestellten sollen vor allem Automatismen Fehler vermeiden und Prozess-organisatorische Tätigkeiten erspart werden.
- Eine technische Anbindung eines CPS an das vorhandene Planungssystem soll erfolgen
- Bei kleinen und insbesondere Zukaufteilen, die in Warenträgern – z. B. Gitterboxpaletten – transportiert werden, ist eine Mitgabe eines Tags mit dem Warenbegleitschein absolut denkbar.
- Aktive Tags mit Ladeinfrastruktur ist kaum praktisch umsetzbar. Langlaufende, batteriebetriebene Tags sind aber denkbar.
- Wünschenswert wäre dennoch eine Bereichsüberwachung mit Identifikation der Werkstücke, um die individuelle Kennzeichnung (insb. von Mänteln und Profilen) überflüssig zu machen.
- Kombinationen aus mobilen Tags für Warenträger und Bereichsüberwachung wären akzeptabel.
- Eine flächendeckende Installation ist denkbar und tatsächlich wünschenswert. Insb. ist dies der Fall, wenn dadurch eine flächendeckende (Bereichs-) Ortung möglich wird.
- Eine Investition – auch im sechsstelligen Bereich – ist denkbar für ein solches Tracking, da die Nutzwerte diese aufwiegen können.

2.3.4. Technologieauswahl und Architektur

Sensorknoten

Da FP 2 pandemiebedingt den Betrieb eines CPS-Demonstrators nicht unterstützen konnte, lag der Schwerpunkt der Entwicklung des CPS auf den Anforderungen von FP 1. Als zentraler Mikrocontroller für die Sensorknoten wurde der Arduino Nano RP2040 Connect ausgewählt. Dieser ist von Haus aus mit einem eingebauten WiFi-Modul, einem PDM-Mikrofon und einem Beschleunigungssensor ausgestattet. Außerdem bietet er eine umfangreiche und leicht zu verwendende Sammlung an Software-Bibliotheken. Der Vorteil von WiFi gegenüber anderen Funktechnologien ist die hohe Verbreitung, die sich über eine erhöhte Verfügbarkeit von Hard- und Software bemerkbar macht, sowie eine hohe Datenrate. Der Hauptnachteil – der relativ hohe Stromverbrauch – kommt bei den stationären Sensorknoten nicht zum Tragen. Als Lichtsensor wurde der APDS-

2.3.5. Entwicklung für die Systemkomponenten

Sensorknoten

Für die Sensorknoten wurde ein Layout für eine Leiterplatte erstellt, um Mikrocontroller und Sensoren so in einem Standardkunststoffgehäuse unterzubringen, dass die Sensoren leicht über Löcher im Gehäuse zugänglich sind. Für die Sensoren und das Micro-USB-Kabel zur Stromversorgung wurden Löcher ins Gehäuse gebohrt. Zur Anbringung des Sensorknotens in der Fertigung wurden Magnete an einer Gehäusesseite befestigt. Softwareseitig wurde eine Firmware auf Basis von verfügbaren Hardware-Treibern aus dem Arduino-Umfeld entwickelt, die die vorhandenen Sensoren in periodischen Abständen abfragen und als JSON-Nachricht über MQTT versenden. Für Konfigurationsänderungen auf den Sensorknoten wurde auch der Empfang von JSON-Nachrichten über MQTT in der Firmware implementiert.

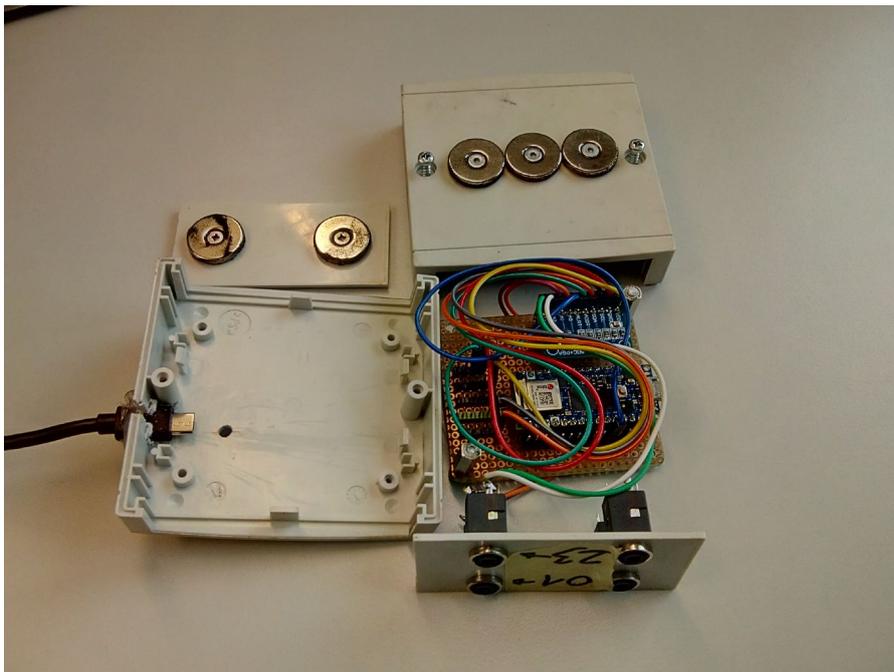


Abbildung 9: Sensorknoten zur Stromstärkemessung (geöffnet)

Da bei der Übertragung von Mikrofonaufnahmen Probleme mit der Größe der MQTT-Pakete, der Stabilität der WiFi-Verbindung und dem Rechenaufwand bei der Serialisierung der Daten auftreten, wurde für die Mikrofondaten eigens ein binäres TCP-basiertes Übertragungsprotokoll konzipiert und implementiert. Die Umwandlung in das eigentlich gewünschte Format (entweder eine einzelne WAV-Datei oder eine Serie von MQTT-Nachrichten mit WAV-Samples) erfolgte auf der Datensenke durch einen separaten Service.

Datensenke

Auf der Datensenke werden die MQTT-Nachrichten mit Metainformationen in einem Protobuf-basierten Binärformat in einer Datei gespeichert. Dabei kommt ein dem HDLC-Protokoll entlehntes Framing zum Einsatz, damit unvollständige Schreibvorgänge die Datei nicht unlesbar machen können. Ein entsprechender Service, der diese Speicherung vornimmt, sowie Werkzeuge zum Verarbeiten/Konvertieren des Nachrichtenarchivs wurden implementiert. Jeweils nach Abholen der Datei beim Forschungspartner wurden die Sensordaten im JSON-Format automatisiert extra-

hiert und zur besseren Verarbeitung nach Sensorwerten getrennt in eine SQLite-Datenbank geschrieben. Ein entsprechendes Werkzeug wurde entwickelt.

2.3.6. Aufbau eines Demonstrators im Betrieb

Bei FP 1 wurde ein Access-Point zentral erhöht und drei WLAN-Repeater in der Nähe einiger relevanter Bearbeitungsplätze angebracht und mit Strom versorgt. Der Intel NUC Mini-PC als Datensinke fand in einem Büroareal in der Mitte der Werkshalle Platz.

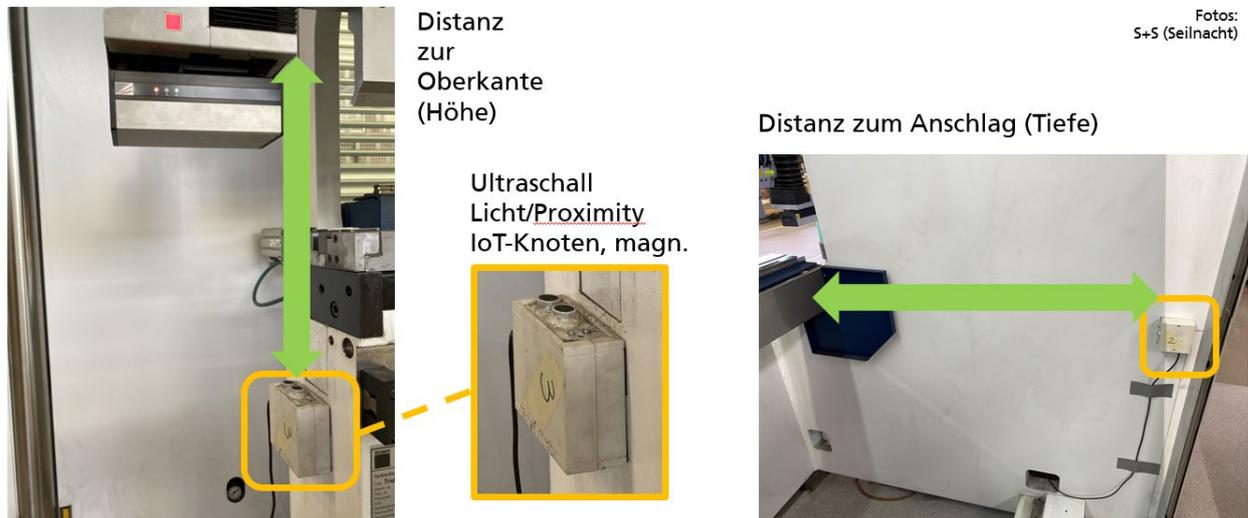


Abbildung 10: Sensorknoten an einer TRUMPF Biegepresse zur jeweiligen Positionsbestimmung des Biegebalkens bei FP 1

Eine TRUMPF Biegepresse und eine Schweißarbeitszelle wurde mit jeweils zwei Sensorknoten ausgestattet.

Bei der Biegepresse überwachte ein Sensorknoten die vorwärts/rückwärts-Bewegung (Abbildung 10, rechtes Bild) und der zweite die Pressbewegung (hoch/runter, linkes Bild). Dabei wurde zuerst Beschleunigung und Nähe zu einem Referenzpunkt vom Pressbalken ausgemessen, später jedoch per Ultraschall von einem festen Punkt aus die Entfernung zum Pressbalken.

In einer Schweißarbeitszelle wurde der Schweißvorgang ebenfalls mit Sensoren erfasst (siehe Abbildung 12). Dabei kam ein Sensorknoten zum Einsatz, der über einen Lichtsensor die Lichtintensität des Schweißbogens beim Schweißen misst. Zusätzlich wurde ein zweiter Sensorknoten installiert, der über eine Stromzange den Stromverbrauch des Schweißgeräts aufzeichnete. Dazu wurde ein Leitungsteiler eingesetzt, um die einzelnen Phasen des Kabels getrennt zugänglich zu machen. Im Rahmen von dedizierten Messkampagnen wurden zeitweise auch Mikrofonaufnahmen der Schweißvorgänge und des Betriebs einer weiteren Schwenkbiegemaschine erstellt.

2.3.7. Betrieb eines Demonstrators, Herausforderungen und Anpassungen

Das CPS hat über einen Zeitraum von über sieben Monaten nahezu durchgängig Sensordaten aufgezeichnet. Dabei konnte eine große Menge an Daten und wertvolle praktische Erfahrungen gewonnen werden.

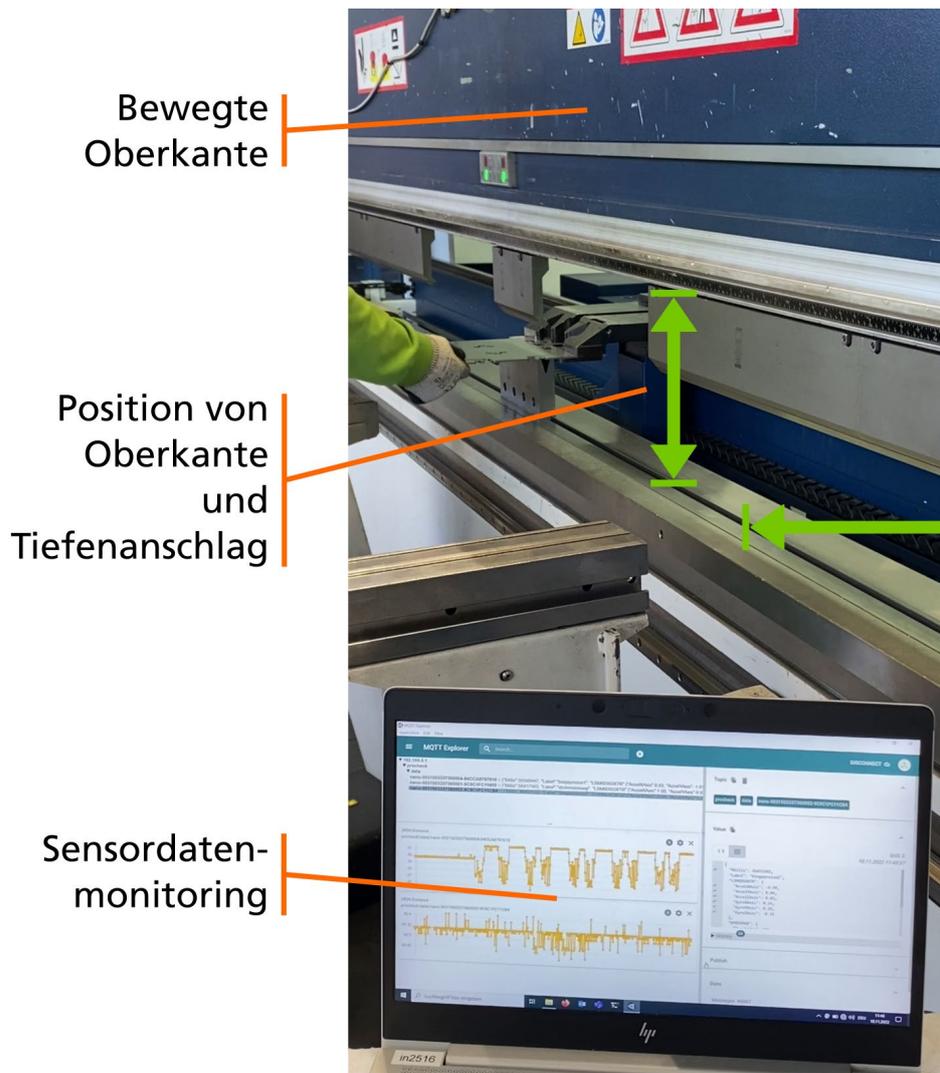
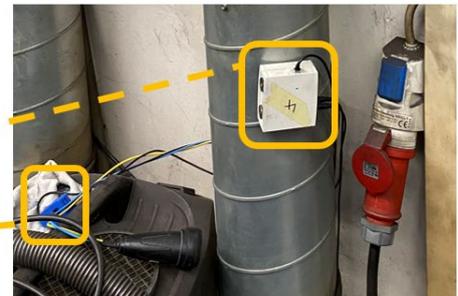


Abbildung 11: Erfassung von Position und Tiefenanschlag der Biegepresse und entstehende Sensordaten



Helligkeit
und
Farbwerte



Indirekte Strommessung
1-phasiger Verbraucher
3-phasiger Verbraucher

Fotos:
S+5 (Seilnacht)

Abbildung 12: Schweißzelle - kontinuierliche Erfassung von Lichtinformationen und Stromverbräuchen

Technische Herausforderungen

Der Aufbau der Sensorknoten sah Öffnungen in den Gehäusen vor, um den Sensoren die Erfassung der Signale zu ermöglichen. Ein wachsendes Problem stellten dabei Metallspäne dar, die zum einen durch die Öffnungen des Gehäuses zur Elektronik vordringen und dort Ausfälle verursachen konnten. Zum anderen setzten sich die Metallspäne und auch Metallstaub an den Magneten fest, die zur Befestigung der Sensorknoten benutzt wurden. In manchen Bereichen (z. B. Lackiererei) konnte ein solch provisorischer Aufbau eines Sensorknotens mit Löchern im Gehäuse gar nicht eingesetzt werden. Bei einem späteren Produkt kann diesen Herausforderungen durch geeignete, angepasste Gehäuse und Befestigungen leicht begegnet werden.

Berichten der Mitarbeiter zufolge führte der Einsatz des Leitungsteilers für die Stromstärkemessung in der Schweißzelle zur Reduzierung der maximalen Stromstärke, die dem Schweißgerät zur Verfügung stand, und damit zu einer Beeinträchtigung der Fertigung. Die Strommessung wurde daraufhin eingestellt, allerdings stellt auch dies das Konzept nicht in Frage, da es sich durch die Wahl eines anderen Leitungsteilers, der im Rahmen des Projekts jedoch nicht mehr beschafft werden konnte, beheben lässt.

Übertragung und Speicherung der Sensordaten

Die hochfrequente Aufzeichnung der Sensordaten stieß teilweise an die verfügbare Bandbreite des Funknetzes, insbesondere bei der Aufzeichnung von Mikrofondaten. Dies wurde möglicherweise durch Störungen des Funksignals durch die Vielzahl an Metallgegenstände in der Fertigungshalle verschärft. Um einen größeren Datenverlust zu vermeiden, wurde bei einem Verbindungsabbruch zeitnah durch einen Hardware-Watchdog ein Reset auf den betroffenen Sensorknoten ausgeführt. Durch ein besonderes Feature des MQTT-Protokolls (retained messages) stehen einem Sensorknoten nach einem Neustart sofort die aktuellen Konfigurationen zur Verfügung, ohne dass eine Speicherung auf dem Mikrocontroller stattfinden muss. Ebenso kann eine Konfigurationsänderung vorgenommen werden, auch wenn der entsprechende Mikrocontroller gerade nicht erreichbar ist.

Eine längerfristige hochfrequente Aufnahme von Sensordaten führte zu einem beträchtlichen Datenaufkommen. Dieses Problem verschärfte sich natürlich mit wachsender Anzahl von Sensorknoten und insbesondere mit komplexen Sensordaten wie z. B. Mikrofonaufnahmen. Auch wenn im konkreten Fall die Größe der Festplatte des als Datensinke verwendeten Mini-PCs auch die Länge der ununterbrochenen Datenaufzeichnung limitiert hat, so stellte die Datenmenge generell eher für die Auswertung ein Problem dar. So überstieg eine Aufnahme von wenigen Wochen schon die Größe des Arbeitsspeichers eines typischen Entwickler-PCs. Hier bot sich ein strukturiertes Speicherformat an, das komplexe Vorfilterungen auf den Rohdaten erlaubt, ohne die gesamte Aufnahme in den Arbeitsspeicher laden oder auch nur komplett einlesen zu müssen. Im Projekt wurde hierfür auf SQLite-Datenbanken zurückgegriffen.

Auf den Einsatz von Server-basierten Datenbanken oder spezialisierten Zeitreihendatenbanken wie z. B. InfluxDB mit eigener Abfragesprache wurde verzichtet, allerdings böten diese eventuell noch Potential für eine weitere Effizienzsteigerung bei der Analyse der Datenmengen.

Akzeptanz bei den Mitarbeitenden

Im Betrieb musste das USB-Netzteil bestimmter Sensorknoten trotz eindeutiger Beschriftung erstaunlich häufig einem Handyladekabel weichen. Der Aspekt, dass hier die Sorge vor einer möglichen Überwachung des Arbeitsverhaltens eine Rolle gespielt hat, kann nicht prinzipiell ausgeschlossen werden. Insgesamt ist aber sowohl das CPS selbst als auch die ermittelten Daten, die mit Hilfe der MQTT-Schnittstelle live demonstriert werden konnten, bei den Mitarbeitenden auf Neugier und Interesse gestoßen. Da bei allen Bestandteilen des CPS in der Fertigungshalle ein temporärer Ausfall der Stromversorgung nicht ausgeschlossen werden konnte, wurden jegliche Hard- und Software so konzipiert und konfiguriert, dass die Sensordatenaufzeichnung mit Wiederherstellung der Versorgungsspannung automatisch fortgesetzt wurde, wodurch die Mitarbeiter vor Ort den Sensordatenausfall minimieren und selbständig beheben konnten.

2.4. Arbeitspaket 4 – Entwicklung der automatisierten Extraktion von Event-Logs aus CPS-Rohdaten

Wie schon AP 3, wurde auch das AP 4 von FE 1 (Bereich Lokalisierung und Vernetzung) begleitet und inhaltlich bearbeitet. Die Inhalte bauen auf den in AP 3 erzielten Ergebnisse auf.

2.4.1. Datenvorauswahl

Eine explorative Datenanalyse ist besonders herausfordernd, wenn sowohl die Datenmenge eine effiziente Verarbeitung am PC erschwert als auch der überwachte Zeitraum eine manuelle Sichtung der Sensorsignale verbietet. Da beide erschwerende Bedingungen hier vorlagen, wurden die aufgezeichneten Sensornachrichten zuerst in ihre Einzelsignale zerlegt und strukturiert in einer SQLite-Datenbank abgelegt.

Dazu wurde ein entsprechendes Software-Werkzeug geschaffen. Dies ermöglichte über die Abfragesprache SQL Untersuchungen und Filterungen der Daten, ohne diese dazu komplett in den Arbeitsspeicher laden zu müssen. Um in dem großen aufgezeichneten Zeitintervall die Bereiche zu identifizieren, an denen eine Bearbeitung in den Sensorsignalen zu erkennen war, wurde über SQL die gesamte Aufnahme in kleine feste Zeitfenster unterteilt, innerhalb derer sich das betrachtete Sensorsignal im Vergleich zum Durchschnittswert signifikant veränderte.

Dabei wurde sowohl der absolute als auch der relative Wertebereich berücksichtigt und aneinandergrenzende relevante Zeitintervalle zusammengefasst. Die so identifizierten Zeitintervalle wurden maschinenlesbar in einer JSON-Datei protokolliert und zur besseren Visualisierung als Diagramm geplottet. Neben dem eigentlich untersuchten Sensorsignal konnten zusätzlich noch weitere Sensorquellen dargestellt werden, um Korrelationen erkennen zu können. Um die Funktionsweise des Vorauswahlverfahrens beurteilen zu können, wurden die Zeitintervalle ohne relevante Sensoraktivität ebenso untersucht. Beispiele für Plots der identifizierten Zeitintervalle sind unten zu sehen. Ein typischer Durchlauf dieser Voranalyse auf einer Datenaufzeichnung von vier Wochen dauerte trotz des beschriebenen Vorgehens immer noch mehr als eine Stunde.

Im Folgenden (Abbildung 13) sind zunächst die synchronisierten Sensoraufzeichnungen der ersten Sensorausstattung mit Beschleunigungssensoren und Näherungssensoren am Pressbalken einer Biegepresse exemplarisch dargestellt.

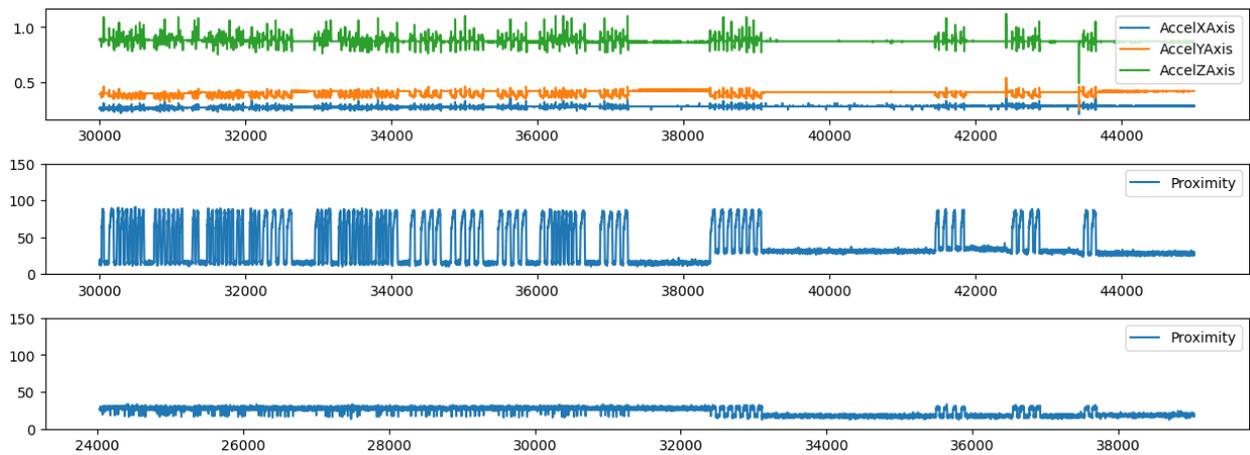


Abbildung 13: Rohdatenstrom mehrerer Sensoren am Beispiel Biegepresse: 3-Achsen-Beschleunigungssensor (oben), Näherungssensor vertikal (Mitte), Näherungssensor horizontal (unten)

Insbesondere bei den Näherungsdaten ist die typischen Phasen der Bewegung des Pressbalkens zu erkennen (Abbildung 14).

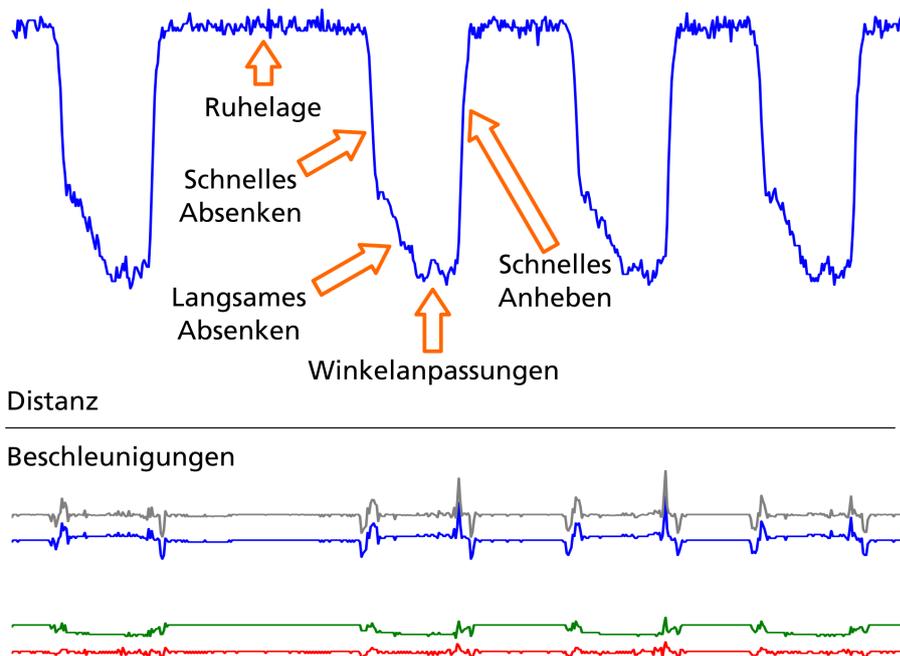


Abbildung 14: Typischer Datenstrom beim Betrieb der Biegepresse: Distanzwerte des Pressbalkens (oben) mit charakteristischen Merkmalen und zugehörige Beschleunigungsdaten (unten); rot, grün, blau: Beschleunigung in verschiedenen Raumachsen; schwarz: Vektorbetrag der Gesamtbeschleunigung

Die Anbringung der Sensorknoten an bewegten Teilen der Maschinen erwies sich durch die kabelgebundene Stromversorgung als auf Dauer nicht betriebssicher, weshalb bei der finalen Sensorausstattung auf die Beschleunigungswerte verzichtet und der Näherungssensor durch einen Ultraschallsensor ersetzt wurde. Ein automatisch identifizierter Teilausschnitt der aufgezeichneten Daten mit zeitlich synchronisierter Tiefeninformation ist in Abbildung 15 zu sehen.

2.4.2. Erkennung von Biegevorgängen mittels Kreuzkorrelation

Zur Beurteilung von Ähnlichkeit zweier Signale existiert in der Signalverarbeitung die sogenannte Kreuzkorrelation.

Dabei bildet die Kreuzkorrelation zwei gleichgroße Signalintervalle auf einen Wert zwischen 1 (Signale sind identisch) und -1 (Signale sind genau entgegengesetzt) ab. Dieses Verfahren wurde im Projekt bei der Auswertung der Abstandsdaten der Biegepresse angewendet. Dafür wurden zuerst verschiedene Referenzsignalverläufe unterschiedlicher Länge aus historischen Daten ausgewählt und anschließend an jeder Stelle eines als relevant identifizierten Datenintervalls mit den Messdaten verglichen.

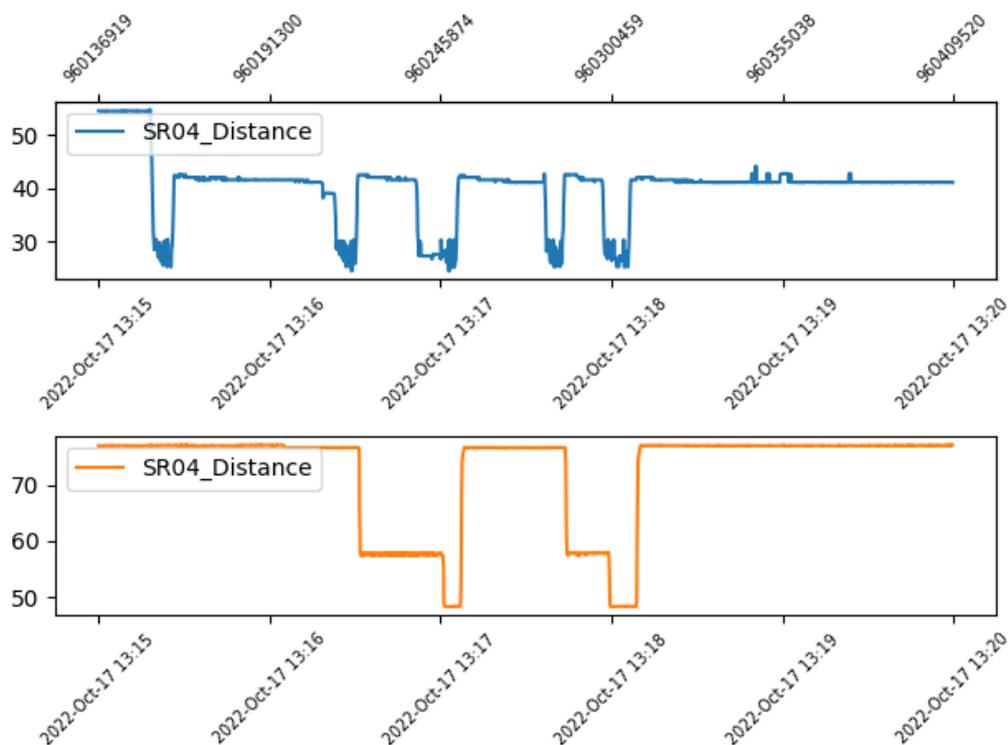


Abbildung 15: Pressbalken lässt Pressvorgänge mit Distanzsensor gut erkennen (oben); Tiefenanschlag gibt Werkstück-typische Maßinformationen (unten)

Das Maximum dieser Kreuzkorrelationen der Referenzsignale mit den Messdaten gab einen Hinweis darauf, zu welchem Zeitpunkt im Messsignal vermutlich ein Pressvorgang stattgefunden hat, und auch welchem Referenzsignal dieser Pressvorgang am meisten ähnelt. Zum Festlegen eines genauen Zeitpunktes des erkannten Pressvorgangs wurden lokale Maxima jeweils innerhalb eines bestimmten Zeitintervalls identifiziert, die darüber hinaus auch noch über einem gewissen Schwellwert liegen mussten. Das Zeitintervall entsprach dabei der Länge des kürzesten verwendeten Referenzsignals. Der Ablauf ist schematisch in Abbildung 16 dargestellt.

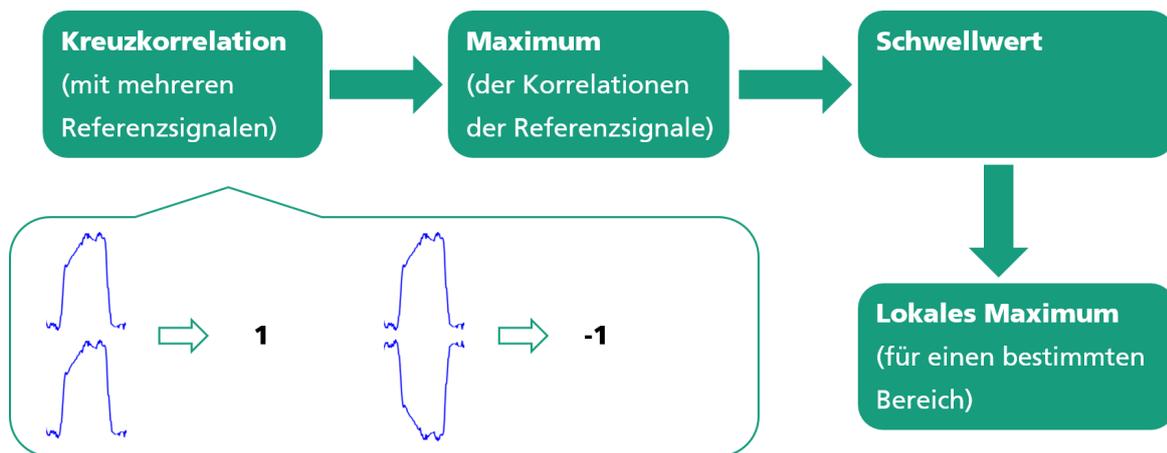


Abbildung 16: Schematischer Ablauf der Kreuzkorrelationsberechnung auf den Daten

Sowohl die Festlegung des Schwellwertes als auch die Auswahl der Referenzsignale erfolgte hierbei iterativ, und der visuelle Abgleich des Datensignals mit den Kreuzkorrelationsmaxima gab einen deutlichen Hinweis darauf, welche Pressvorgänge mit den bisherigen Einstellungen noch nicht ausreichend erkannt wurden. Einen graphischen Plot eines typischen Ergebnisses dieser Analyse ist in Abbildung 17 zu sehen.

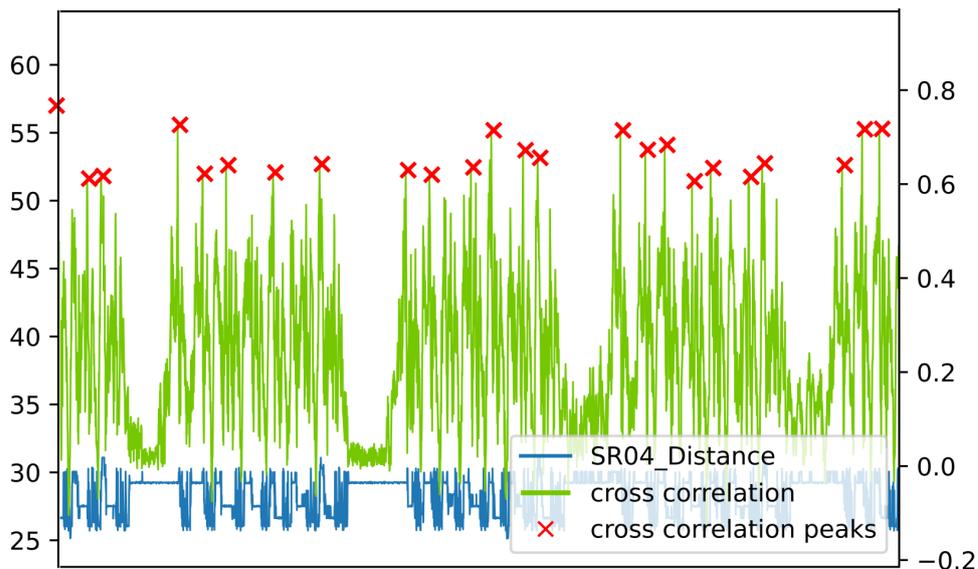


Abbildung 17: Beispielplot der Kreuzkorrelationsberechnung auf Biegepressdaten

Diese Verfahren eignete sich besonders für die Analyse der Abstandsdaten der Biegepresse, da hier der Arbeitsschritt zwar manuell vorbereitet und ausgelöst wird, der eigentliche Pressvorgang aber von der Maschine und damit gut reproduzierbar ausgeführt wird. Es zeigte sich darüber hinaus, dass insbesondere der Absolutwert des zweiten Abstandssignals (Tiefe) entscheidend dazu beitragen kann, Biegevorgänge zu differenzieren (siehe Abbildung 15).

Dieses Vorgehen lässt sich natürlich nicht nur auf Einzelpressungen als Referenzsignal anwenden, sondern auch auf Folgen von Biegevorgängen für ein gesamtes Bauteil. Die Ermittlung von Referenzsignalen für verschiedene Bauteile ist mit Aufwand verbunden, allerdings ist für die Suche nach der Bearbeitung von einigen wenigen bestimmten Bauteilen im Datenstrom dieser Ansatz vielversprechend. Leider zeigte sich, dass insbesondere auch kleinere Pausen zwischen den

Einzelbiegungen die Kreuzkorrelation stark beeinflussen und reduzieren können. Hier spielen die manuellen Anteile des Arbeitsschrittes eine entscheidende Rolle.

2.4.3. Erkennung von Schweißvorgängen

Die Schweißvorgänge in einer Schweißkabine wurden zum einen über eine Stromstärkemessung an der Stromversorgung des Schweißgerätes und zum anderen über einen Lichtsensor aufgezeichnet (Abbildung 18). Als besonders offensichtlich stellten sich die Aufzeichnungen der Stromstärken heraus. So konnten z. B. der Einschaltvorgang durch eine leichte Erhöhung des Ruhestroms und einzelne Schweißvorgänge durch starke Stromstärke-Peaks nachgewiesen werden. Zur Unterscheidung von verschiedenen Schweißvorgängen wurden unterschiedliche Zeitdauern herangezogen (Schweißpunkt vs. Schweißnaht) und mit dem Abstand von steigenden und fallenden Flanken des Messsignals verglichen.

Bei den Lichtdaten waren meist ebenfalls deutliche Helligkeitsänderungen bei Schweißvorgängen zu beobachten. Allerdings fanden auch häufig Verdeckungen oder Abschattungen des Lichtbogens – entweder durch den Mitarbeiter oder das bearbeitete Bauteil – statt, sodass in den Messdaten dann nur indirekte Reflektionen zu erkennen waren. Eine analytische Erkennung und Unterscheidung von Schweißvorgängen wurden dadurch erschwert. Hier wäre die Kombination von mehreren räumlich verteilten Lichtsensoren sinnvoll.

Durch die Möglichkeit, die Stromstärkemessung als Referenzwert für die Lichtdaten zu benutzen und den entsprechenden Schweißvorgängen automatisch Labels zuzuweisen (siehe Abbildung 20), bieten sich für die Erkennung von Schweißvorgängen Methoden des überwachten maschinellen Lernens an. Aufgrund des Mangels an Messdaten durch den Abbruch der Aufzeichnung und eines anderen Fokus in diesem Projekt, wurde dieser Ansatz aber nicht weiterverfolgt.

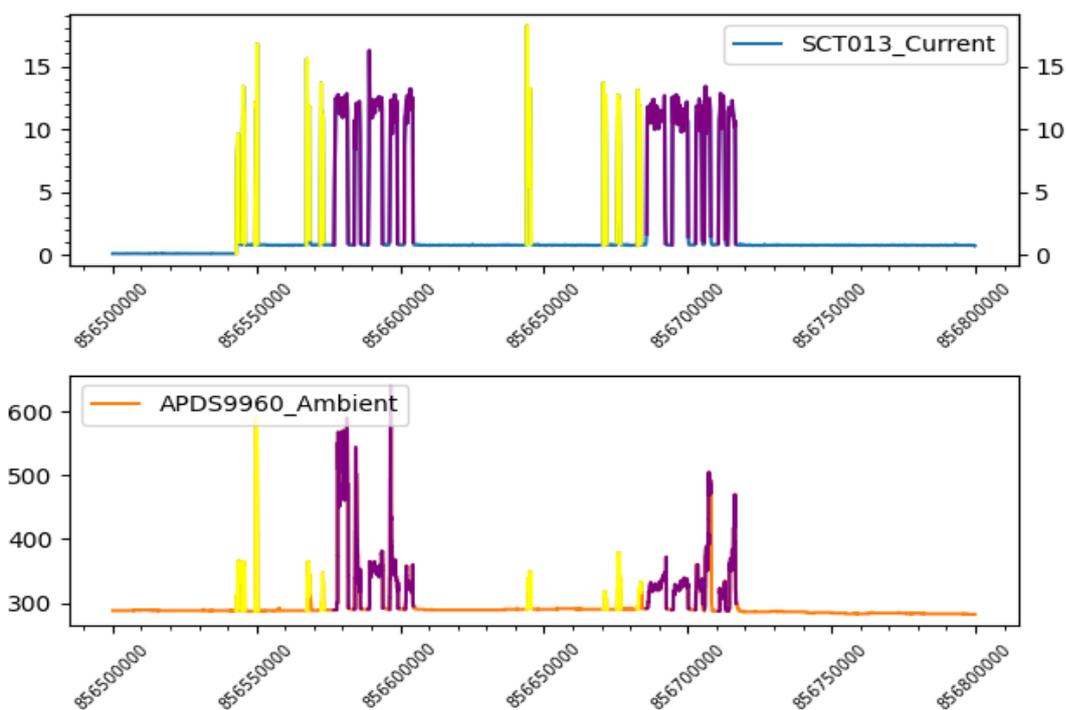


Abbildung 18: Stromverbrauch (oben) und Lichtentwicklung (unten) bei unterschiedlich langen Schweißvorgängen

2.4.4. Bauteilerkennung

Wie in Abbildung 19 zu sehen ist, kam es vor, dass mehrere gleichartige Bauteile hintereinander an der Maschine bearbeitet wurden. In der Abfolge von Pressvorgängen, die z. B. über die Kreuzkorrelation und den Tiefenabstand identifiziert wurden, und zum Teil auch den zwischenliegenden Pausen konnten sich lokal wiederholende Muster identifiziert werden.

Eine Möglichkeit, dies algorithmisch zu erreichen, besteht darin, ein gleitendes Fenster über die identifizierten Bearbeitungsaktionen zu legen und die Aktivitätenfolge im Fenster mit mehreren gleichlangen Mustern zu vergleichen, die aus Wiederholungen von Aktionsfolgen unterschiedlicher Länge bestehen. Das nach einem bestimmten Abstandsmaß am besten passende Muster wird ausgewählt. Ein mögliches Abstandsmaß, das auch Einfügungen berücksichtigen kann, ist z. B. die Levenshtein-Distanz. Bei Übergängen zwischen zwei am besten passenden Mustern können entweder die Pausen berücksichtigt werden, da beim Wechsel zu einem anderen Bauteil die Maschine neu programmiert werden muss und so immer eine größere Pause zu verzeichnen ist, oder der Verlauf der berechneten Distanzen untersucht werden.

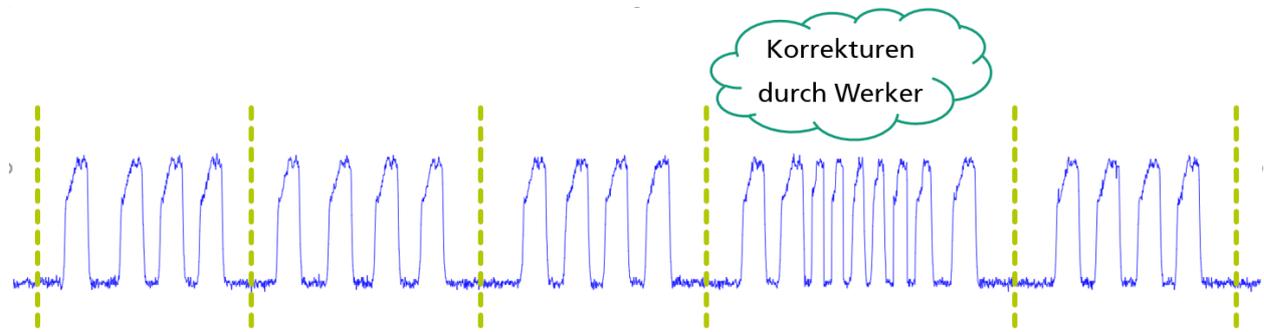


Abbildung 19: Werkstücke in der Biegepresse ergeben ein jeweils typisches Signalmuster des Distanzsensors

Ein ähnliches Vorgehen wäre auch beim Schweißarbeitsplatz denkbar, allerdings kommt es hier aufgrund der stärkeren manuellen Tätigkeit zu größeren Variationen wie z. B. Vertauschungen oder abgebrochenen/fortgesetzten Schweißnähten. Eine so reguläre Folge von Sensorsignalen wie in Abbildung 20 ist leider die Ausnahme.

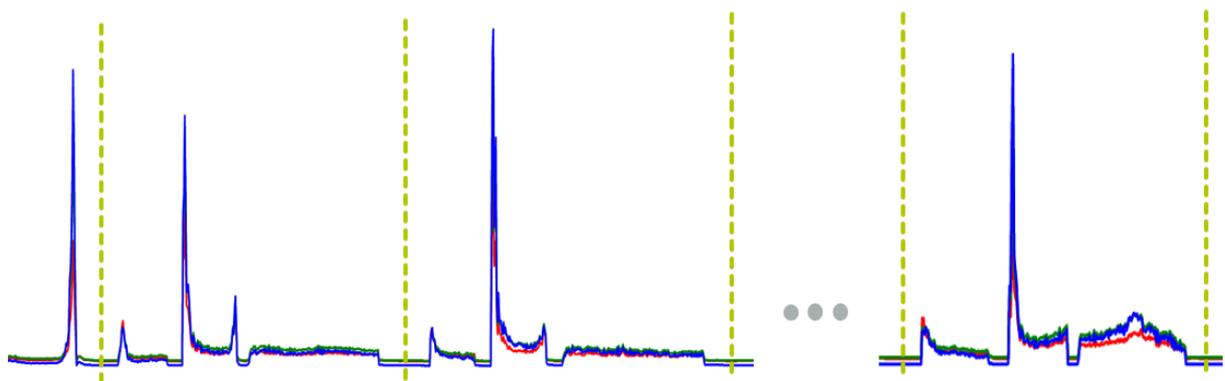


Abbildung 20: Lichtintensität des Lichtsensor in der Schweißzelle - Bauteil-bezogene Charakteristika und Wiederholungen

Die Erkennung von gleichartigen Sequenzen von Aktivitäten ist eine Möglichkeit, im Nachhinein einen Auftragsbezug zu erkannten Arbeitsschritten herzustellen, indem man die Anzahl der bearbeiteten Werkstücke und eventuell bauteilspezifische Zusatzinformationen (z. B. Anzahl von

Kanten) mit aktuellen Buchungsinformationen abgleicht. Dies hat Aussicht auf Erfolg, wenn

- die Buchungsinformationen für die relevanten Zeitabschnitte zur Auswertung verfügbar sind,
- alle Maschinen an denen ein Auftrag (zeitgleich) bearbeitet werden kann, auch mit dem CPS ausgerüstet sind,
- es genügend gleichartige Teile in einem Fertigungsauftrag gibt, um aufeinanderfolgende gleichartige Sequenzen identifizieren zu können.

Auch wenn alle genannten Voraussetzungen erfüllt sind, ist mit Erkennungsfehlern zu rechnen.

Da ohne Barcode-Scanner im Projekt kein Auftragsbezug der aufgezeichneten Sensordaten hergestellt werden konnte, war auch die geplante Präzisierung der Buchungszeiten für die Event-Logs nicht möglich.

2.5. Arbeitspaket 5 – Entwicklung von Verfahren zum Abgleich von Event-Logs mit Soll-Prozessmodellen

Federführend im AP 5 war der Lehrstuhl für Datenbanksysteme und Data Mining der LMU München (FE 2). FE 1 war unterstützend und beratend tätig. Der FP 1 lieferte die notwendige Datengrundlage und stand außerdem für Rückfragen zur Verfügung.

2.5.1. Input

Als Input für dieses Arbeitspaket dienten sowohl die in AP 1 festgelegten und in AP 2 modellierten Prozessmodelle sowie Daten aus dem ERP-System von FP 1. Da FP 2 aufgrund der vorherrschenden pandemischen Situation an der Entwicklung der Process Mining-Verfahren nicht teilnehmen konnte, lag der Schwerpunkt auf den Produktionsdaten von FP 1. Wie in Kapitel 2.4 bereits dargelegt wurde, konnte über die eingesetzte Sensorik dort der gewünschte Auftragsbezug und damit die Generierung von Event-Logs noch nicht zufriedenstellend gelöst werden. Aus diesem Grund wurde die Anwendung der Process Mining-Verfahren auf ERP-Daten aus dem Buchungssystem von FP 1 durchgeführt. Für ein besseres Verständnis der Daten wurde vor Ort in Nürnberg-Kornburg ein Workshop durchgeführt und gemeinsam mit Verantwortlichen des FP 1 ein Datenmodell erarbeitet, welches die verschiedenen Hierarchieebenen der Daten abbildet. Da einzelne Bearbeitungsschritte in diesem Prozess hochgradig manuell sind, waren nur wenig automatisch generierte Prozessinformationen verfügbar. Aus diesem Grund war es entscheidend, die vorhandenen Datenquellen zu bewerten und aus den identifizierten ERP-Daten Event-Logs zu erzeugen. Aufgrund der verschiedenen Datenquellen und der Qualität der Daten war diese so genannte Event-Daten-Korrelation herausfordernd, konnte jedoch erfolgreich umgesetzt werden. Dieser Ansatz bietet auf Umwegen die Möglichkeit, die Prozessdaten des FP 1 für Process Mining-Verfahren dennoch nutzbar zu machen. Im Anschluss konnten die ERP-Daten mithilfe von Process Mining-Verfahren untersucht werden. Diese Analysen auf den Buchungsdaten sollten zum einen die Praxistauglichkeit der eingesetzten Process Mining-Techniken für Produktionsumgebungen unter Beweis stellen und zum anderen die Grundlage für weiterführende Forschungsarbeiten, insb. die Verwendung von Sensordaten, schaffen. Von großem Interesse ist das Zusammenspiel mit den in AP 3 bzw. AP 4 entstandenen Ansätzen bildet Anknüpfungspunkte für weiterführende Forschungsaktivitäten, weshalb die in AP 5 entwickelten Ansätze für Event-Logs aus Sensordaten ebenfalls anwendbar sein sollten.

2.5.2. Granularität und Trace Clustering

Vor der Anwendung von Process Discovery oder Conformance Checking, insbesondere in realen Szenarien, ist eine Vorverarbeitung der Daten notwendig, um die Ereignisdaten in sogenannte Event-Logs zu transformieren. Diese Ereignisdaten finden sich häufig in Quellen wie z. B. Geschäftsprozessmanagementsystemen (BPMS) (vgl. Pourmirza et al., 2019) oder Enterprise Resource Planning Systemen (ERP) (vgl. De Weerd & Wynn, 2022). Im fertigen Gewerbe, insbesondere bei KMUs, sind relevante Daten oft über mehrere Quellen und in einer Vielzahl von Formen verstreut (vgl. ebd.). Daher findet man in den seltensten Fällen einen eindeutigen, durchgängigen Prozessausführungsidentifikator. In der betrachteten Fallstudie haben die Kandidaten für Prozessausführungskennungen eine hierarchische Struktur mit verschiedenen Granularitätsebenen. Um intrinsische Strukturen in einem unbeschrifteten Datensatz zu entdecken, wurden Trace-Clustering-Ansätze verwendet. Beim Trace Clustering wird das Ereignisprotokoll in Teilmengen aufgeteilt, um das Verständnis und die Analyse jeder Teilmenge zu erleichtern, z. B. durch die Erstellung eines Prozessmodells für jede Teilmenge (vgl. Marin-Castro & Tello-Leal, 2014).

2.5.3. Basistechniken von Process Mining

Beim Process Mining ist ein *Event-Log* L ein Datensatz, der aus mehreren Ereignissen besteht und den Ablauf eines Prozesses beschreibt. Ein *Event* ist die kleinste Einheit eines Event-Logs. Es beschreibt eine einzelne Ausführung eines Schrittes in einem Prozess und besteht aus mindestens drei Komponenten: *Case ID* c , *Activity ID* a und einem *Zeitstempel* t , üblicherweise ausgedrückt als $e = (c, a, t)$. Die Case ID c gibt an, zu welcher Prozessausführung ein Event gehört. Eine Activity ID a beschreibt die Art des Schrittes, bei dem dieses Event aufgetreten ist. Schließlich gibt der Zeitstempel t eines Events den Zeitpunkt der Durchführung dieses Ereignisses an. Zeitstempel können ordinale Werte für Events oder andere intervallbezogene Maße wie Datum und Uhrzeit enthalten. Im letzteren Fall sind Abstände zwischen Ereignissen und Vergleiche der Dauer möglich. Aggregiert ist eine Spur die Sammlung von Ereignissen mit der gleichen Case ID und eine Variante ist die Folge von Aktivitäten mit der gleichen Case ID, bezeichnet als $V_k = [a_i, \dots, a_j]$ und mit der Länge $|V_k|$.

Process Discovery Techniken (vgl. Van der Aalst, 2016) erzeugen *Prozessmodelle*, die einem bestimmten Ereignisprotokoll entsprechen. Modelle auf niedrigerer Ebene, wie Directly Follows Graphs, sind einfach zu verstehen und verwenden Vorrangbeziehungen zwischen Ereignissen zur Darstellung eines Ereignisprotokolls. Dennoch führt dieser Ansatz zu fehlerhaft modellierten Prozesssequenzen, wenn die Startreihenfolge zweier Aktivitäten nicht festgelegt ist. Im Gegensatz dazu verwenden Modelle auf höherer Ebene, wie BPMN oder Petri-Netze, Splits und Joins, um parallele Kompositionen und Entscheidungen in einem Prozess darzustellen (vgl. Van der Aalst, 2022). Um Prozessmodelle automatisch zu konstruieren, werden Process Discovery-Algorithmen verwendet. Diese lassen sich in Bottom-Up- und Top-Down-Ansätze unterteilen (vgl. Van der Aalst, 2018): Bottom-up-Ansätze, wie z. B. Alpha Miner (vgl. Van der Aalst, 2004) und Heuristics Miner (vgl. Weijters & Van der Aalst, 2003), berücksichtigen zunächst lokale Prozessmuster, um das Modell zu erstellen. Auf der anderen Seite untersuchen Top-Down-Ansätze, wie z. B. der Inductive Miner (vgl. Leemans et al., 2013), die gemeinsamen Teile des Event-Logs und zerlegen diese in Unterprotokolle (sublogs), bis ein Basisfall gefunden wird (vgl. Van der Aalst, 2022). Aufgrund der hohen Varianz in den Prozessspuren des Produktionsprozesses bei FP 1 und der hohen Varianz in der Artikelhäufigkeit wurde der Inductive Miner als Beispiel für einen Top-down-Ansatz ausgewählt.

Sobald ein Prozessmodell vorhanden ist, wird die Konformitätsprüfung (*Conformance Checking*) verwendet, um nicht-konforme Prozessspuren zu detektieren. Nicht-konforme Prozessspuren sind Ausführungen, die nicht mit dem Soll-Modell übereinstimmen. Zwei der am weitesten verbreiteten Techniken zur Konformitätsprüfung sind Alignments (vgl. Baier et al., 2014) und Token-based Replay (vgl. Rozinat & Van der Aalst, 2008). Token-based Replay reproduziert Prozessspuren auf entsprechenden Petri-Netzen und zählt fehlende Token, wenn an einem Eingabeknoten ein zusätzliches Token fehlt, um die nächste benötigte Transition zu aktivieren, sowie verbleibende Token, die nie verbraucht werden. Andererseits bilden Alignments Ist-Prozessspuren auf ihr ähnlichstes Verhalten im Prozessmodell ab und markieren Abweichungen mit „übersprungen“ oder „eingefügt“.

Im Rahmen dieses Projekts wurde ein Analysedokument zum Stand der Technik der Conformance Checking-Methoden, deren Gegenüberstellung und Eignung für KMU des fertigen Gewerbes verfasst.

Bekanntes Clustering-Verfahren

Die richtige Granularität für Ereignisse in einem Datensatz zu finden, ist eine wichtige Aufgabe vor der Anwendung von Process Mining. Um intrinsische Strukturen in einem unbenannten (*unlabeled*) Datensatz zu entdecken, wurden bei den in diesem Kapitel vorgestellten Analysen Clustering-Ansätze angewendet. Beim Trace Clustering wird das Ereignisprotokoll in Teilmengen aufgeteilt, um das Verständnis und die Analyse jeder Teilmenge zu erleichtern, z. B. durch die Erstellung eines Prozessmodells für jede Teilmenge (vgl. Marin-Castro & Tello-Leal, 2014). Bei den Arbeiten in ProCheck wurde insbesondere auf hierarchische und dichte-basierte Ansätze fokussiert. Diese berücksichtigen die hierarchische Beziehung zwischen Auftrag, Produkten und Artikeln in der Fertigung und vermeiden darüber hinaus jegliche Annahmen über die Verteilung von Varianten unter verschiedenen Artikeln oder von nicht-konformen Ist-Spuren. Hierarchisches Clustering (vgl. Nielsen, 2016) offenbart komplexe, mehrstufige, verschachtelte Strukturen in Datensätzen. Die Ergebnisse des hierarchischen Clustering werden häufig in Erreichbarkeitsdiagrammen oder Dendrogrammen visualisiert. Insbesondere Agglomerative Hierarchische Clustering (AggloCl) Ansätze arbeiten bottom-up (vgl. Nielsen, 2016): Ausgehend von einem binären Merge-Baum werden die Datenelemente durch dessen Blätter dargestellt. Als nächstes werden die ähnlichsten Teilmengen, einschließlich Singleton-Teilmengen, paarweise zusammengeführt. Dieser Vorgang wird so lange wiederholt, bis alle Elemente des Datensatzes in der Baumstruktur enthalten sind. Der Abstand zwischen den Teilmengen wird als Verknüpfungsdistanz (Linkage Distance) bezeichnet.

Der Pionieralgorithmus DBSCAN (vgl. Ester et al., 1996) ist ein dichte-basierter Bottom-up-Clustering-Ansatz. Unter Verwendung einer beliebigen Ähnlichkeitsdistanz zwischen Elementen und dichte-basierten Begriffen werden die Elemente im Datensatz entweder als Kernpunkte, Randpunkte oder Ausreißer klassifiziert. Unabhängig davon, ob es sich um eine Verknüpfungsdistanz für hierarchisches Clustering oder eine Ähnlichkeitsdistanz für DBSCAN handelt, wurden im Forschungsprojekt verschiedene Edit Distanzen berücksichtigt, die in den experimentellen Ergebnissen dieses APs näher betrachtet werden: Mengendifferenzgröße, Levenshtein-Distanz, Längste Gemeinsame Teilfolge (LCS), Damerau-Levenshtein-Distanz und Jaro-Distanz.

Die einfache Editierdistanz zwischen Sets zählt Einfügungen und Löschungen, um ein Set in ein anderes zu transformieren. Die LCS-Distanz (Longest Common Sub-Sequence) zählt ebenfalls

Einfügungen und Löschungen, berücksichtigt jedoch die Reihenfolge der Elemente. Zum besseren Verständnis soll folgendes Beispiel einer LCS-Abstandsrechnung für zwei mögliche Varianten dienen:

$V1 = [„Lasern“, „Kanten-Biegen“, „Grundieren“, „Montieren“]$ und $V2 = [„Lasern“, „Kanten-Biegen“, „Senken“, „Senken“, „Montieren“]$. Ihre längste gemeinsame Teilsequenz ist $[„Lasern“, „Kanten-Biegen“, „Montieren“]$, mit der Länge drei $|LCS(V1, V2)| = 3$. Daher ist der LCS-Abstand definiert als $(|V1| - |LCS(V1, V2)|) + (|V2| - |LCS(V1, V2)|)$, hier $(4 - 3) + (5 - 3) = 3$. Die LCS-Distanz zwischen $V1$ und $V2$ ist 3.

Die Levenshtein-Distanz (Lev) verwendet zusätzlich Substitutionen und die Damerau-Levenshtein-Distanz (Damlev) erlaubt auch Transpositionen. Die Jaro-Winkler-Distanz (Jaro) schließlich erlaubt und zählt nur Transpositionen. Für einen fairen Vergleich wurden in den Analysen die Distanzmetriken normalisiert.

2.5.4. Event Data Korrelation in Manufacturing

Umwandlung von Fertigungsdaten in ein Event-Log

Ereignisprotokolleinträge, wie sie für Process Mining verwendet werden, erfordern mindestens die folgenden Elemente:

1. Fall-ID (Case ID),
2. Aktivitäts-ID (Activity ID) und einen
3. Zeitstempel.

Benchmark-Datensätze enthalten diese Elemente immer. In diesem Arbeitspaket erfolgte eine Analyse solcher Datensätze und der Aufbau einer Datenbank „Event Logs Benchmark Insights“, die Ergebnisse dieser Analyse automatisiert gespeichert hat. Anders als bei der Verwendung von

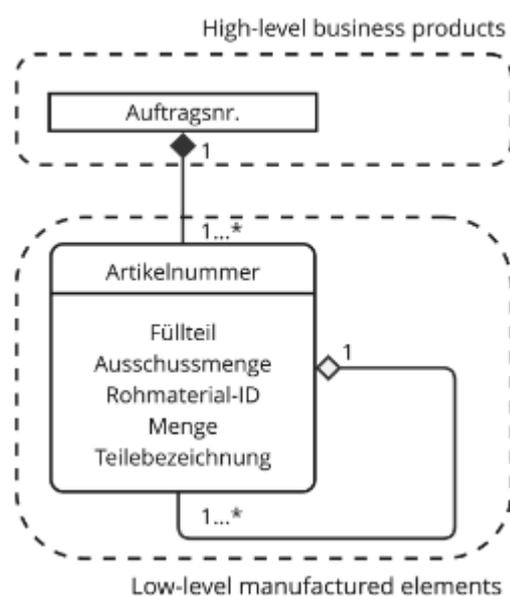


Abbildung 21: Schematische Darstellung für Ereignisdaten aus verschiedenen Systemen und mit unterschiedlichen Granularitätsebenen für Identifikatoren der Prozessausführung

Benchmark-Daten waren jedoch bei der Verwendung der Produktionsdaten des FP 1 weder die Case-ID, die Aktivitäts-ID noch der Zeitstempel in dem Datensatz eindeutig gekennzeichnet. Dieser Herausforderungen galt es mit Event Data Korrelation zu begegnen. Die Buchungseinträge enthalten u. a. die Spalten: *Produktionsschritt*, *die erwartete Dauer des Produktionsschrittes*, *die produzierte „Artikelnummer“*, *benutztes Rohmaterial*, *das Datum der Produktion* und die *„Auftragsnummer“*.

Abbildung 21 zeigt die semantische Beziehung zwischen den Spalten. Ein Buchungseintrag enthält einen Zeitstempel (generiert aus dem Produktionsdatum) und eine Aktivität (Produktionsschritt), die direkt als Ereignisattribut übernommen werden können. Um eine geeignete Case ID zu übernehmen, ist es wichtig, die semantische Beziehung zwischen den übrigen, oben genannten Spalten zu verstehen. Ein Auftrag („Auftragsnummer“) ist ein Identifikator und setzt sich aus mindestens

einem Artikel („Artikelnummer“) zusammen, der eine Klasse bezeichnet und die Aggregation weiterer Artikel sein kann oder auch nicht. Daraus ergibt sich eine hierarchische Struktur von Artikeln, die zum selben Produktauftrag gehören. Die implizite Kapselung von Artikeln, Unterprodukten und Aufträgen im Event-Log verhindert eine direkt lesbare Identifizierung einer eindeutigen Case ID durch eine vollständige Prozessausführung und wirkt sich direkt auf die Länge der Ist-Prozessspuren und damit auf die Komplexität des Prozesses aus. Darüber hinaus enthält ein und dasselbe Ereignisprotokoll eine hohe Varianz an produzierten Artikeln, die in einer unüberwachten (*unsupervised*) Umgebung mit demselben Auftrag verbunden sein können oder auch nicht.

Nachdem ein umfassender Eindruck über die verfügbaren Prozessdaten und deren Zusammenspiel geschaffen werden konnte, wurden die Daten ausgiebig vorverarbeitet. Da die in den Daten angegebenen Zeitstempel bzw. Kalenderdaten für die meisten Aktivitäten nicht geeignet sind, um die Reihenfolge der Ereignisse in einem Fall abzuleiten, wurde ihre Vorverarbeitung in zwei Teile aufgeteilt:

- Erstens wurden die mit Zeitstempeln versehenen Ereignisse an der richtigen Stelle eingefügt. Hier wurden die Ergebnisse der in AP 1 durchgeführten Experteninterviews genutzt, um die Prozessreihenfolge zu rekonstruieren.
- Zweitens wurde in enger Abstimmung mit den Prozessverantwortlichen des FP 1 die Dauer jedes Ereignisses entsprechend der von ihnen erwarteten Aktivitätsdauer quantifiziert. Auf diesem Weg konnten die Aktivitäten innerhalb einer Prozessausführung in der besprochenen Reihenfolge angeordnet werden. Die Zuarbeit des Anwendungspartners war an dieser Stelle essenziell, um valide Aussagen über die Dauer der entsprechenden Prozessaktivitäten treffen zu können.

Um interaktiv zu untersuchen, welche Spalten im Datensatz sich bestmöglich für die Analyse eignen, wurde ein Start2PM-Dashboard erstellt. Dieses Tool lieferte mithilfe einer statistischen Analyse passende „Kandidaten“ in Form von Werten für die Elemente des Ereignisprotokolls, die in Abbildung 22 dargestellt sind. Die Abbildung zeigt zwei Prozessmodelle, die automatisch ermittelt wurden, d. h. mit Hilfe von Algorithmen zur Prozesserkennung aus Fallstudien Daten. Prozessmodell (a) wurde unter Verwendung einer feinkörnigen Perspektive, Prozessmodell (b) unter einer grobkörnigen Perspektive und Prozessmodell (c) unter Verwendung derselben Daten, jedoch abstrahiert unter Verwendung einer gemischten granularen Ansicht der Daten, ermittelt. Alle repräsentieren die Produktion von 4043 Artikeln, die sich auf 1988 Aufträge verteilen. Die Ergebnisse dieses Dashboards schlugen „Beschreibung“ als Activity ID (18 Aktivitäten) und eine Kombination der beiden Spalten „Auftragsnummer“ (1988 Aufträge, 435 Varianten) und „Artikelnummer“ (4043 Artikel, 397 Varianten) als geeignete Case ID vor. Dies resultiert in 13104 Spuren (Traces) und 75 Varianten. Da die Elemente im vorverarbeiteten Datensatz die festgelegten Kriterien für ein Event-Log erfüllen, werden diese von nun an als Event-Log bezeichnet.

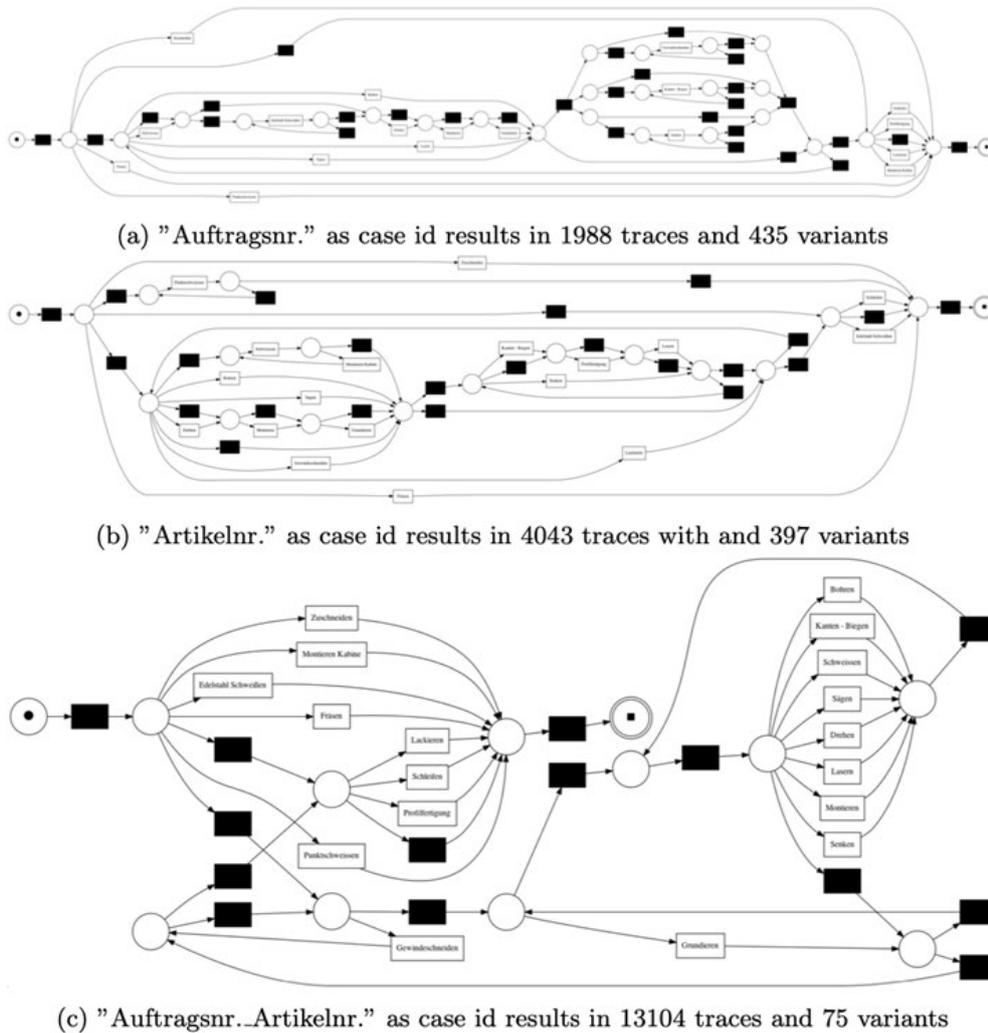


Abbildung 22: Automatische Ermittlung zweier Prozessmodelle mithilfe algorithmischer Verfahren zur Prozesserkennung

Produktionsperspektive und Granularität der Case ID

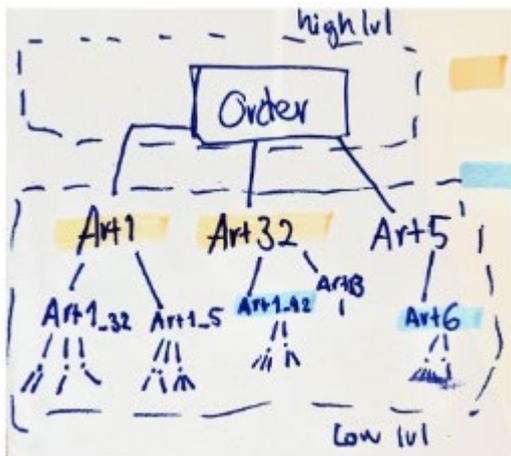


Abbildung 23: Whiteboard-Aufzeichnung der Hierarchie innerhalb eines Artikels, angefertigt während eines Workshops bei FP 1

Die Abbildung 23 zeigt ein Beispiel für die Hierarchie der Artikel innerhalb eines Auftrags beim FP 1, die in einem Workshop erarbeitet wurde, wobei gleiche Farben für ähnliche Artikel stehen. Zum besseren Verständnis der Strukturen wurden die Ergebnisse des Vor-Ort-Workshops nach Rücksprache mit den Verantwortlichen des FP 1 in Abbildung 24 erfasst. Auf der obersten Ebene befinden sich Auftragsnummern (Identifiers) und auf der niedrigste Ebene Artikelnummern (Klasse). Die Kapselung von Artikeln beeinträchtigt die Prozessermittlung nicht, da Aktivitäten nur für den Ausgangsartikel gezählt werden. Die Kapselung von Artikeln beeinträchtigt nicht die Konformitätsprüfung, obwohl nicht-konforme Ist-Prozessspuren von verschiedenen Artikeln und/oder Aufträgen aggregiert werden. Das Clustering von Prozessspuren ergibt Gruppen

ähnlicher solcher Spuren in Bezug auf die Schrittfolgen der Prozessausführung, d. h. Prozessvarianten. Obwohl davon ausgegangen wird, dass Instanzen gleicher Artikel und sogar ähnlicher Artikeltypen ähnliche Prozessspuren aufweisen, bedeutet Ähnlichkeit nicht, dass auch ähnliche Artikel ausgegeben werden. In den Ergebnissen der angewendeten Verfahren zeigt sich, dass Produktfamilien anhand der Ähnlichkeit der Prozessspuren identifiziert werden können. Es ist zu beachten, dass sowohl Ähnlichkeiten zwischen Prozessspuren innerhalb desselben Auftrags auftreten können als auch zwischen Prozessspuren aus verschiedenen Aufträgen.

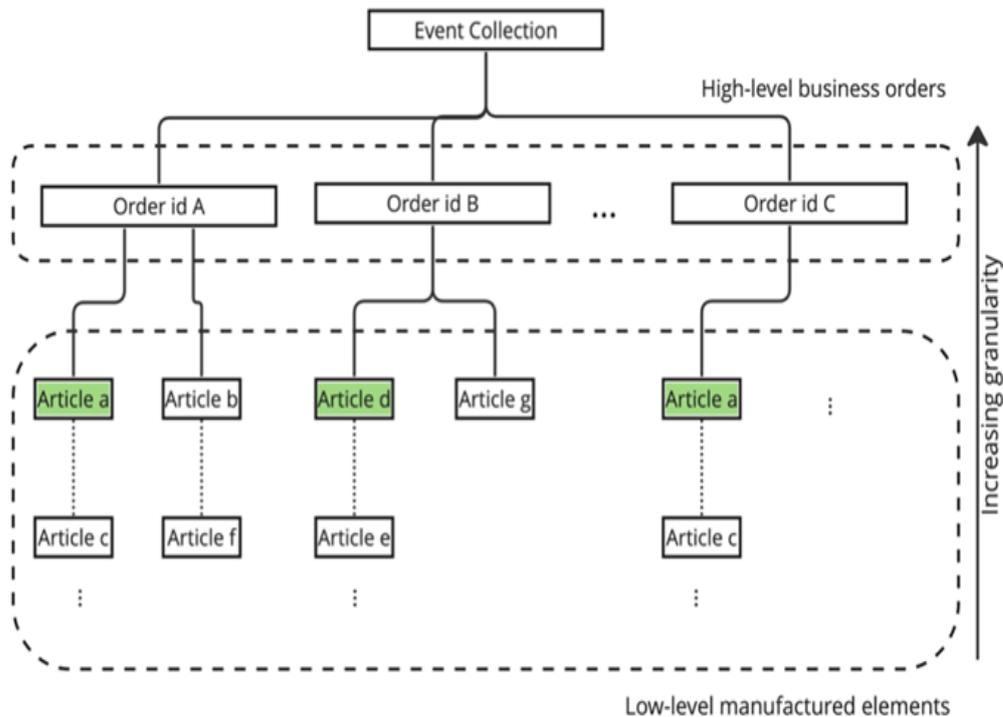


Abbildung 24: Beispiel für hierarchisch verwandte Objekte - Auftragsnummer und Artikel - ergeben mehrere mögliche Granularitätsebenen. Ähnliche Farben stehen für ähnliche Artikel, die ähnliche Varianten ausführen. (FP 1)

2.5.5. Experimentelle Ergebnisse

Unter Verwendung der Auftragsnummer_Artikelnummer als Case ID wurde die Abfolge der Aktivitäten, d. h. die Varianten, zum Clustern von Prozessspuren verwendet. Durch die Kombination verschiedener Distanzen innerhalb der Prozesse wurden mithilfe der Clustering-Algorithmen DBSCAN (vgl. Ester et al., 1996) und AggloCl (vgl. Nielsen, 2016) die entsprechenden Prozessspuren geclustert. In diesem Abschnitt werden die experimentellen Ergebnisse für diese Fallstudie vorgestellt. Es werden die Metriken der Datenperspektive sowie die Ergebnisse des Process Mining vor und nach dem Clustering verglichen. Dieser Schritt war entscheidend, um die angewendeten Techniken final zu evaluieren. Die verschiedenen Cluster-Algorithmen wurden in verschiedenen Kombinationen durchgeführt und zusätzlich die Ergebnisse des Process Mining vor und nach dem Clustering verglichen. AggloCl wurde unter Verwendung von Single und Complete Linkage evaluiert und die Anzahl der Cluster wurde auf $N_{clusters} = 7$ gesetzt. Für DBSCAN wurde $MinPts = 2$ festgesetzt und $\epsilon \in [0.1, 0.2, 0.3, 0.4, 0.5]$ abgestimmt. Das in ProCheck entwickelte Framework ist in Python implementiert. Es wurden Clustering-Ansätze von sklearn angewendet und Prozessmodelle mit pm4py generiert (vgl. Berti et. al, 2019).

Qualitative Analyse zu Trace Clustering

Zunächst galt es die Verteilungen der Abstände, die durch die verschiedenen Distanzmetriken erzeugt wurden, zu bewerten. Abbildung 25 zeigt die Gegenüberstellung der Abstandsverteilungen der in AP 5 verwendeten Distanzmetriken. Es lässt sich festhalten, dass die Abstandsverteilungen multimodal sind, was darauf hindeutet, dass Cluster in den Daten gefunden werden können. Gestrichelte Linien zeigen Quartile und durchgezogene Linien die Mittelwerte für die Abstände zwischen den Spuren. Relativ hohe Mittelwerte für Distanzen weisen auf eine hohe Sparsamkeit in den Daten hin, was den ϵ -Parameter für DBSCAN entspannt.

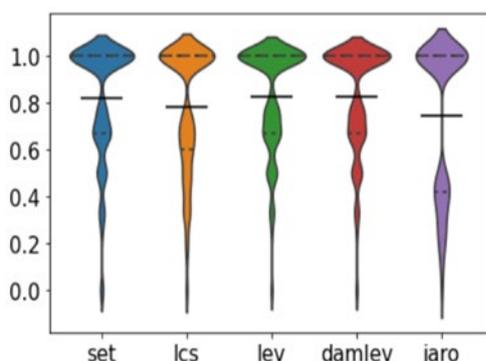


Abbildung 25: Abstandsverteilungen nach angewendeten Distanzmetriken

Abbildung 26 zeigt die Auswirkungen verschiedener Werte für ϵ auf vier verschiedene Metriken: Das oberste Diagramm zeigt die resultierende Anzahl von Clustern. Idealerweise lassen sich so viele Cluster erwarten, wie es Gruppen ähnlicher Artikel in der Produktion von FP 1 gibt – d. h. Artikelbereiche. Da jedoch die genaue hierarchische Beziehung zwischen den Artikeln eine Unbekannte ist, wurde die tatsächliche Anzahl der Artikel-Sortimente ignoriert. Daher kann geschlossen werden, dass eine detaillierte bzw. umfassendere Sicht auf den Datenbereich die Identifikation von unterschiedlichen Prozessmodellen verbessert, was zu mehr Clustern führt.

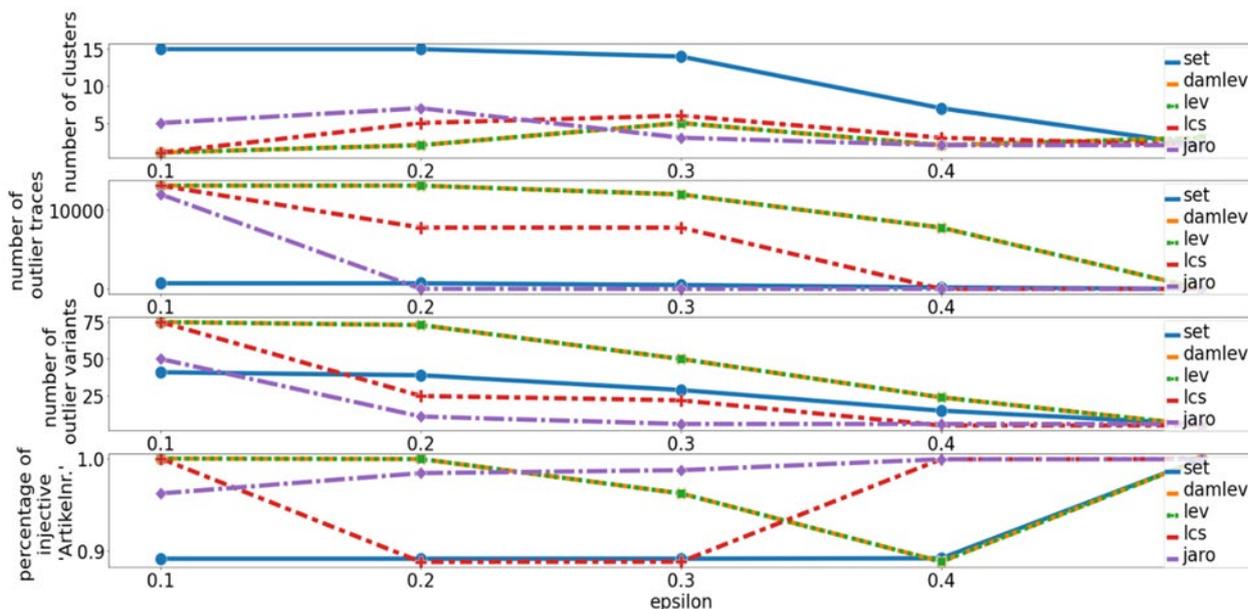


Abbildung 26: DBSCAN Ergebnisse bei verschiedenen ϵ -Werten

Wie bereits angeführt, klassifiziert DBSCAN verrauschte Datenpunkte als Ausreißer. Idealerweise sollten die meisten Spuren und Varianten zu irgendeinem Cluster gehören. Das bedeutet, dass eine geringere Anzahl von Ausreißer-Spuren und -Varianten für das zweite und dritte Diagramm bevorzugt wird. Das unterste Diagramm zeigt den Prozentsatz der injektiven Artikelnummer. Es stellt dar, wie viele Artikel eindeutigen Clustern zugeordnet sind. In Anbetracht der Tatsache, dass wir nicht erwarten, dass die Produktion desselben Artikels zwischen den Instanzen stark variiert, sollte dieser Prozentsatz idealerweise hoch sein. Es konnte gezeigt werden, dass ein Kompromiss zwischen diesen vier Metriken unvermeidlich ist. Dennoch zeigt die Set-Distanz mit $\epsilon = 0,3$ die vielversprechendsten Ergebnisse. Diese Kombination ergibt sich aus 14 Clustern, 514 Ausreißerspuren, 29 Ausreißervarianten und 89 % injektiven Artikeln. Obwohl der beste Kompromiss zu fast 40 % Ausreißervarianten führt, machen diese nur weniger als 4 % der Ausreißerspuren aus, was an der ungleichmäßigen Verteilung der Spuren auf die Varianten liegt.

Im weiteren Verlauf der Arbeit untersuchen wir die DBSCAN-Ergebnisse unter Verwendung des entsprechenden optimalen ϵ für jeden Abstand. Darüber wurde der entwickelte Ansatz auch anhand prozessbezogener Metriken wie der Anzahl der Varianten pro Cluster bewertet. Die Verwendung einer Clustermethode wird präferiert, wenn die Anzahl der Varianten pro Cluster gleichmäßiger verteilt ist. Abbildung 27 zeigt die zuvor vorgestellten Metriken pro Cluster für verschiedene Distanzen sowie die beiden Verfahren AggloCl auf der linken und DBSCAN auf der rechten Seite. Für AggloCl wurde $N_{\text{clusters}} = 7$ als Parameter gesetzt und für DBSCAN variiert die Anzahl der Cluster zwischen vier und 14. Bei den meisten Distanzmaßen in beiden Diagrammen ist zu beobachten, dass ein oder zwei Cluster eine deutlich höhere Anzahl von Varianten aufweisen als der Rest. Dies deutet darauf hin, dass es meist ein oder zwei Cluster gibt, die die meisten der weniger ähnlichen Varianten enthalten. Insbesondere bei Verwendung von AggloCl scheinen die Cluster 0 und 2 den Rest der Varianten aus dem Ereignisprotokoll zu enthalten, die nicht in andere Cluster passen. Im Allgemeinen sind die DBSCAN-Ergebnisse gleichmäßiger verteilt als die AggloCl-Ergebnisse. Der Grund dafür ist, dass DBSCAN im Gegensatz zu AggloCl im Rahmen der Clusterbildung Ausreißer herausfiltert. DBSCAN-Methoden mit Distanzen, die mehr Ausreißer gefunden haben, weisen insgesamt eine geringe Anzahl von Varianten auf. Dies zeigt sich ebenfalls in Abbildung 27. Betrachtet man die Ergebnisse zwischen den Abstandsmetriken, so hebt sich der festgelegte Abstand bei beiden Clustermethoden vom Rest ab. Da sowohl die Reihenfolge als auch die Wiederholung von Aktivitäten in den Spuren bei der Verwendung von Sets vernachlässigt werden, ermöglicht die Set-Distanz eine stärker auf Aktivitäten fokussierte Clusterung und ist weniger empfindlich gegenüber der Spurenlänge. Darüber hinaus zeigt die Verwendung von DBSCAN eine signifikant höhere Anzahl von Varianten in Cluster 0 als der Rest bei Verwendung von LCS- und Jaro-Distanzen. Der Grund dafür ist, dass sie mehr verschiedene Varianten in diesem Cluster enthalten, die sonst ausgeschlossen werden. Für die weitere Analyse wurde sich deshalb hauptsächlich auf DBSCAN mit der eingestellten Distanz $\text{MinPts} = 2$ und $\epsilon = 0,3$ konzentriert.

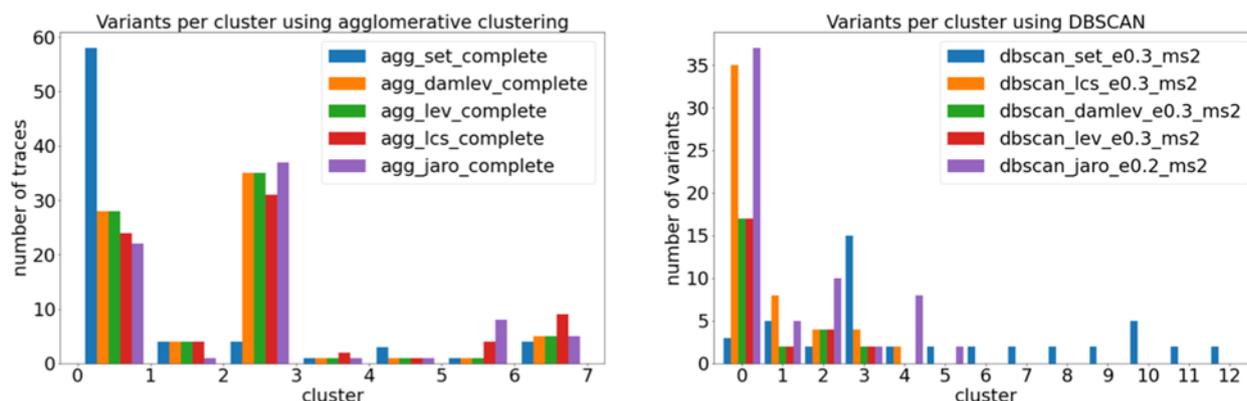


Abbildung 27: Metriken pro Cluster für verschiedene Distanzen nach Verfahren

Darüber hinaus wurde auch die Anzahl der Spuren pro Cluster, die Anzahl der Varianten pro Cluster, die Anzahl der Artikel pro Cluster und die Anzahl der Auftrags-IDs pro Cluster ausgewertet. Einerseits zeigt die Anzahl der Spuren pro Cluster, unabhängig von der Granularität, die tatsächliche Population jedes Clusters und ist ein Indikator für die Häufigkeit von Varianten in einem bestimmten Cluster. Andererseits sind die Anzahl der Artikel und Bestellnummern pro Cluster Indikatoren für die Granularität jedes Clusters. Geht man davon aus, dass die Herstellung von Artikeln der unteren Ebene weniger vielfältig ist und dass Artikel der unteren Ebene in mehr Bestellnummern und häufiger verwendet werden als Artikel der oberen Ebene, so repräsentieren Spuren in Clustern mit einer hohen Anzahl von Artikeln pro Cluster die Granularität der unteren Ebene. Ebenso repräsentieren Spuren in Clustern mit einer hohen Anzahl von Bestellungen pro Cluster eine niedrige Granularität. Umgekehrt sind Artikel mit hoher Granularität mit einer geringen Anzahl von Artikeln und Bestellnummern pro Cluster verbunden. Abbildung 28 zeigt die Anzahl der Spuren pro Cluster auf der linken Seite für die entsprechende Anzahl von Varianten pro Cluster auf der rechten Seite von Abbildung 27. Die Varianten für DBSCAN und Set Distance sind relativ gleichmäßig verteilt, mit Ausnahme einer Spitze in Cluster 3. Betrachtet man das entsprechende Prozessmodell in Abbildung 30 (c), so stellt man fest, dass diese Cluster Spuren enthält, die mit „Lasern“ beginnen und mit anderen 5 Aktivitäten in verschiedenen Kombinationen, aber in unveränderter Reihenfolge folgen. Betrachtet man die hohe Anzahl von Spuren für Cluster 3, so stellt man fest, dass die Varianten nicht nur relativ lang, sondern auch recht häufig sind. Als Konsequenz kann festgehalten werden, dass die Anzahl der Spuren die Ergebnisse der Clusterbildung stark beeinflusst und einzelne Varianten, die häufiger sind, gegenüber mehreren Varianten mit geringem Abstand bevorzugt werden. Vernachlässigt man Wiederholungen und die Reihenfolge mit festgelegtem Abstand, stellt man außerdem fest, dass Spuren mit vielen ähnlichen Aktivitäten näher beieinander liegen als Spuren mit wenigen unterschiedlichen Aktivitäten, unabhängig von der Spurenlänge. Abbildung 28 zeigt die Anzahl der Artikel pro Cluster auf der linken Seite und die Anzahl der Auftragskennungen pro Cluster auf der rechten Seite. Insgesamt sind die Histogramme für die Anzahl der Spuren ähnlich wie die Histogramme für Auftragskennungen und Artikel. Alle drei weisen eine linksschiefe Verteilung auf, was die Granularität der Spuren und Artikel in Abhängigkeit von ihrem jeweiligen Cluster verdeutlicht.

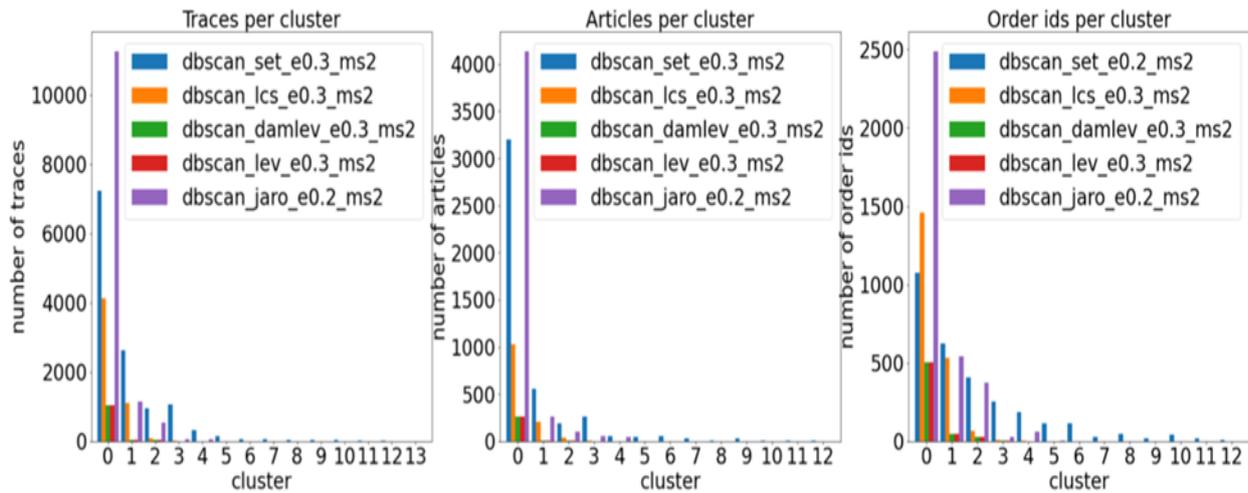


Abbildung 28: Granularitätsanalyse hinsichtlich der Anzahl der Ist-Prozessspuren (links), Anzahl an Artikeln (Mitte) und Auftragsnummern (rechts) per Cluster für DBSCAN

Process Discovery

Zur Bewertung der Qualität von Prozessmodellen werden in der Regel vier Aspekte herangezogen (Vgl. Van der Aalst & Carmona, 2022):

- Recall
- Präzision
- Generalisierung und
- Einfachheit

Der Recall (d. h. Replay Fitness) berücksichtigt die Wiederholbarkeit der Spuren im Ereignisprotokoll im Modell. Die Präzision behandelt ereignisprotokollunabhängiges Verhalten. Idealerweise sollte ein solches Verhalten im Modell nicht zugelassen werden. Es wurde bereits angeführt, dass ein Prozessmodell mit zu viel unerkanntem Verhalten nicht zum Ereignisprotokoll passt. Im Gegensatz dazu wird bei der Genrealisierung nicht gesehenes, aber gültiges Verhalten im Prozessmodell berücksichtigt. Andernfalls passt sich das Modell zu stark an. Schließlich verlangt die Einfachheit, dass das Entdeckungsmodell so einfach wie möglich sein sollte. Um die Ergebnisse in Bezug auf diese Metriken zu diskutieren, werden in Abbildung 30 die vier Cluster mit den meisten Varianten für DBSCAN mit Set-Distance, MinPts = 2 und $\epsilon = 0,3$ durch Prozessmodelle dargestellt. Jedes Cluster, d. h. jeder Artikelbereich, umfasst Spuren aus der Herstellung derselben Artikel sowie verschiedener Artikel mit ähnlichen Varianten, während 89 % der Artikel völlig unverbundene Artikel und Varianten ausschließen. Auch wenn es für lohnend erachtet wird, die 21 % der Artikel, die mehreren Clustern zugeordnet sind, weiter zu untersuchen, versprechen die vorläufigen Ergebnisse bereits einen fairen Kompromiss zwischen Präzision und Generalisierung. Auf den ersten Blick, unabhängig von der Anzahl der Spuren, Varianten, Bestellnummern und Artikel, scheinen sich alle entsprechenden Petri-Netze in Bezug auf die Einfachheit relativ ähnlich zu sein, verglichen mit dem in Abbildung 22 (c). Eine leicht quantifizierbare Methode zur Messung der Einfachheit eines Workflow-Modells ist das Zählen der benötigten Kanten, um die häufigste Spur wiederzugeben. Im Prozessmodell für Cluster 0, dargestellt in Abbildung 30 (a), durchläuft z. B. die erste Variante „Lasern“, ausgehend vom Startpunkt über einen unsichtbaren Übergang, insgesamt sechs Kanten. Zum Vergleich: Um die häufigste Variante „Lasern“ im Modell (c) aus

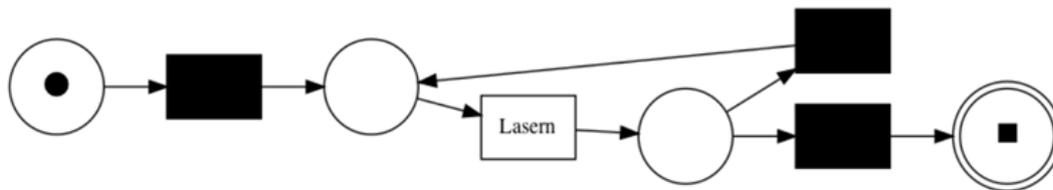
Abbildung 22 wiederzugeben, werden vom Startort aus fünf unsichtbare Übergänge und 11 Kanten durchlaufen, bis der Übergang zu „Lasern“ erreicht wird. Nachfolgend werden bis zum Endort dann weitere acht unsichtbare Übergänge und 15 Kanten bewältigt. Dies entspricht insgesamt 26 durchlaufene Kanten für die häufigste Variante mit einer einzigen Aktivität. Daraus ergibt sich eine Komplexitätsreduktion von 77 %. Die häufigsten Varianten eines jeden Clusters sind in nachfolgender Abbildung 29 dargestellt. Sie vergleicht die Anzahl der passierten Kanten in den entsprechenden Petri-Netzen in Abbildung 22c und 30, um die häufigsten Varianten wiederzugeben.² In jedem Cluster weisen die Petri-Netze in den späteren Prozessmodellen im Vergleich zu den früheren Prozessmodellen eine deutliche Komplexitätsreduzierung auf, d. h. sie sind einfacher. Es ist zu beachten, dass Cluster mit mehr Varianten tendenziell komplexer sind als Cluster mit weniger Varianten, daher können wir für die übrigen Cluster in Abbildung 28 eine bessere Einfachheit annehmen.

Cluster	Most common variant	Number of edges in fig. 2c	Number of edges in fig. 6	Reduction (%)
#0	['Lasern']	26	6	77
#1	['Lasern', 'Kanten-Biegen']	36	12	66
#3	['Lasern', 'Kanten-Biegen', 'Grundieren']	58	24	59
#10	['Lasern', 'Senken', 'Gewindeschneiden']	36	14	61

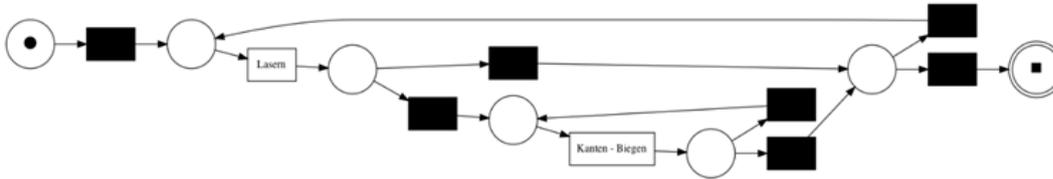
Abbildung 29: Häufigste Varianten eines jeden Clusters

Da die Verwendung von Trace Clustering zum Auffinden granularitätsbewusster Teilprozesse das Ereignisprotokoll auf die Teilmengen von Spuren in den resultierenden Clustern segmentiert, betrachten wir die von DBSCAN klassifizierte Ausreißer erneut, um den Recall zu bewerten. Wie im zweiten Liniendiagramm in Abbildung 26 zu sehen ist, sind nur weniger als 4 % der Spuren in keinem Cluster enthalten. Da die Konformitätsprüfung auf Prozessmodellen und Spuren als Input beruht, beeinflusst die Verwendung mehrerer einfacher Modelle anstelle eines einzigen, das auf alle Spuren passt, die Ergebnisse. Betrachtet man das unzureichend angepasste Petri-Netz (c) in Abbildung 22, könnten mehr Varianten mit dem Modell übereinstimmen, auch wenn sie unerwünschtes Verhalten zeigen. Im Gegensatz wird das Petri-Netz in Abbildung 30 nach dem Clustering der Spuren betrachtet. Stark abweichende Spuren für einen bestimmten Artikel könnten zu einem anderen Artikel gehören, anstatt ein starker Ausreißer desselben zu sein, gleiches gilt für Artikelserien. Je gleichmäßiger die Anzahl der Varianten pro Cluster verteilt ist und je höher der Prozentsatz der injektiven Cluster ist, desto größer wird dieser Effekt.

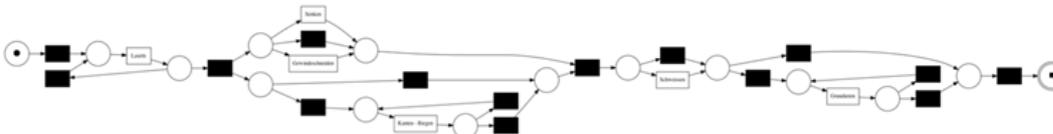
² Abbildung 29 wurde in Zusammenhang mit einer wissenschaftlichen Veröffentlichung erzeugt. Aus diesem Grund ist die Spalte „fig2c“ als „Abbildung 22c“ und die Spalte „fig6“ als „Abbildung 30“ zu lesen.



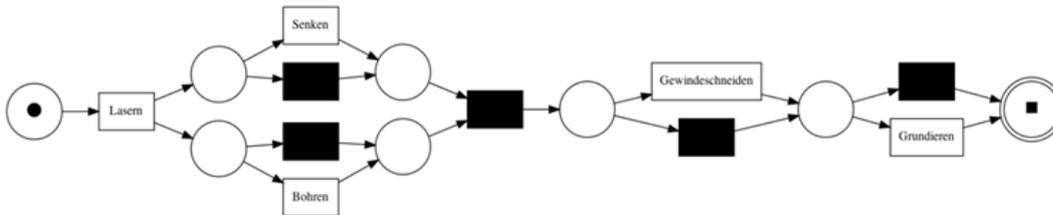
(a) Cluster 0 shows 7227 traces, 3 variants, 3201 articles and 1074 order ids



(b) Cluster 1 shows 12630 traces, 5 variants, 561 articles and 623 order ids



(c) Cluster 3 shows 1058 traces, 15 variants, 260 articles and 516 order ids



(d) Cluster 10 shows 32 traces, 5 variants, 5 articles and 31 order ids

Abbildung 30: Clustering-Ergebnisse mit DSCAN und mengenbasiertem Editierabstand. Nach dem Trace Clustering sind die Modelle weniger komplex als in Abbildung 22 (c)

Conformance Checking

Nach Ansicht der Experten vor Ort gibt es in dieser Fallstudie zwei Arten von konzeptionell nicht-konformen Spuren: Erstens können Wiederholungen derselben Tätigkeit, die wir in 192 Fällen gefunden haben, durch Ausschussteile verursacht werden, z. B. wenn die Qualität einer Tätigkeit nicht gewährleistet werden kann, so dass Artikel aussortiert werden und die Tätigkeit wiederholt wird, z. B. [„Lasern“, „Lasern“], d. h. zweimal lasern. Zweitens handelt es sich bei Ereignissen in falscher Reihenfolge, die wir in 2301 Fällen gefunden haben, wahrscheinlich um Fehler bei der Datenerfassung, da einige mechanische Schritte in falscher Reihenfolge unmöglich sind z. B. [„Montieren“, „Lasern“], d. h. erst montieren und dann lasern. Um diese Art von Anomalien zu finden, verringern wir bei der Erstellung der entsprechenden Prozessmodelle den Schwellenwert der akzeptierten Spuren – der derzeit auf 100 % gesetzt ist –, was z. B. das in Abbildung 30 (a) gezeigte Modell sein könnte. Dennoch ist es insbesondere bei der zweiten Art von Anomalie möglich, dass die anomale Spur mit der ähnlichsten Variante von DBSCAN als Ausreißer klassifiziert wird, was wir in zukünftigen Arbeiten untersuchen wollen.

Für diese Fallstudie wurden weitere Konformitätsprüfungs- und Anomalieerkennungsmethoden, die in den Basistechniken in 2.5.3 bereits vorgestellt wurden, in einem interaktiven Dashboard Madd4pm gesammelt.

2.5.6. Fazit

Obwohl die Anwendung von Process Mining in der Industrie weit verbreitet ist, bleiben in der Praxis einige Herausforderungen offen. Diese Herausforderungen hängen sowohl mit der Digitalisierung von physischen Materialflussprozessen als auch mit den lokalen Gegebenheiten von KMU zusammen. Eine davon ist die Suche nach der richtigen Granularität für eine angemessene Korrelation von Ereignisdaten hinter hierarchischen Bezeichnern und Klassenattributen in der Ereignissammlung. Gemischt granulare Ereignisdaten behindern die Anwendbarkeit von Process Mining, z. B. sind die entdeckten Modelle oft zu komplex und verbergen die wahre Prozesslogik des Fertigungsprozesses. Darüber hinaus variieren die Ergebnisse der Konformitätsprüfung in Bezug auf Auffindbarkeit und Verständlichkeit in Abhängigkeit von der Granularität der Case IDs. In ProCheck wurde ein Konzept für einen datenperspektivischen, unüberwachten Ansatz für die Korrelation von Prozessdaten entwickelt, der die Transparenz und Erklärbarkeit in realen Prozessen, insbesondere in der Fertigungsindustrie, erhöht.

Es konnte anhand von Realdaten gezeigt werden, dass Ansätze zum Clustering von Prozessspuren die Komplexität innerhalb der Prozesse reduzieren und gleichzeitig Familien ähnlicher Artikel und Produkte identifizieren, wobei dieser Ansatz durch die Verwendung echter Zeitstempel anstelle von fiktiven Zeitstempeln erweitert werden konnte. Derzeit beschränkt dies die Erkennung von nicht-konformen Prozessspuren auf die Anzahl der Vorkommnisse in Aktivitäten ohne Berücksichtigung von Positionsanomalien. Darüber hinaus eröffnet die Integration von Sensor- und Buchungsprozessdaten, in Zukunft für die Korrelation von Ereignisdaten, aber auch für weitere Process Mining-Techniken, viele Möglichkeiten. Dies schließt Verfahren wie Leistungsanalysen, prädiktive Überwachung in Offline- und Online-Szenarien mit ein. In ProCheck sind interaktive Dashboards, Start2pm und Madd4pm, sowie mehrere Notebooks, die für die Entwicklung und Evaluierung von Trace Clustering Ansätzen verwendet werden, entstanden.

2.6. Arbeitspaket 6 – Technische und prozessbezogene Evaluation

Die Ergebnisse in ProCheck wurden sowohl während der Projektlaufzeit iterativ als auch im Anschluss gemeinsam mit dem PA in einem Workshop evaluiert. In einer PA-Veranstaltung zur Mitte der Projektlaufzeit wurden die Ergebnisse des Kick-Offs erneut aufgegriffen, das Projektvorhaben gemeinsam mit den Fachexperten auf die Praxisrelevanz (v. a. Anforderungen des Mittelstandes) bewertet und der derzeitige Stand der Arbeiten beurteilt. Basierend auf der Diskussion mit den Teilnehmenden wurden die Forschungsarbeiten in der zweiten Projekthälfte weitergeführt. Abschließend wurden die erzielten Projektergebnisse in einem Abschlussmeeting präsentiert und erneut gemeinsam bewertet. Es zeigte sich wiederum eine hohe Relevanz des Themas für die Analyse von Materialflussprozessen bei KMU. Diese qualitative Evaluierung der Projektergebnisse erfolgte ebenfalls regelmäßig mit den in ProCheck involvierten Anwendungspartnern, insbesondere mit FP 1.

Da die finale Kombination der beiden Teilbereiche CPS und Process Mining aufgrund der in Kapitel 2.4 geschilderten Punkte nicht final vollzogen werden konnte, war die geplante technische und prozessbezogene Evaluation des Gesamtsystems nicht durchführbar. Stattdessen wurde sich auf die einzelnen Lösungskomponenten konzentriert und diese separat evaluiert. Die (quantitative) Evaluation der beiden Teilbereiche CPS und Process Mining wurde in Kapitel 2.4 bzw. 2.5 bereits detailliert ausgeführt, weshalb an dieser Stelle lediglich nochmals ein Überblick gegeben wird. Für die iterative Evaluation spielten wöchentliche Abstimmungstermine mit FP 1 in Form

von Videokonferenzen eine entscheidende Rolle. Für die Arbeiten in AP 5 betraf dies insbesondere die Bewertung von Daten im System auf ihre Richtigkeit. Dies war für die durchgeführte Event Data Korrelation entscheidend, da ohne diesen Schritt die Anwendung von Process Mining-Verfahren keine verlässlichen Ergebnisse geliefert hätte. Weiterhin galt es die durch Process Mining detektierten Cluster gemeinsam zu evaluieren. Auf diesem Weg konnten Abweichungen auf Prozessebene erkannt werden. Die durch die Anstrengungen im Forschungsprojekt erzeugten Cluster der Fertigungsprozesse wurden von den Ansprechpartnern des FP 1 als realistisch beurteilt. Die entwickelten Lösungen stiften im Produktionsalltag Nutzen und werden für die Fertigungssteuerung und -planung verwendet. Die eingesetzte CPS-Technologie und die damit erzeugten Daten konnten in engem Zusammenspiel der FE 1 und des FP 1 bewertet werden. Für das AP 3 und AP 4 lässt sich zusammenfassen, dass ein CPS aus rein stationären Sensorknoten nur unter günstigen Umständen zuverlässige auftragsbezogene Zeitinformationen für ein Event-Log erzeugt werden kann. In den übrigen Fällen muss zumindest der Auftragsbezug über mobile Sensorik mit Lokalisierung oder zusätzliches Scanning (z. B. RFID oder Barcode) hergestellt werden. Es lässt sich festhalten, dass eine vollständige Erfassung der Fertigungsprozesse bei KMU herausfordernd ist und stark von den lokalen Gegebenheiten vor Ort abhängt.

2.7. Arbeitspaket 7 – Dokumentation und Ableitung des Handlungsleitfadens für CPS-basiertes PM

Um die in ProCheck erzielten Ergebnisse interessierten KMU, Software-/Systemhäusern, Prozessberatungen und Herstellern von CPS-Technologien zu Verfügung zu stellen, wurden die Arbeiten entsprechend dokumentiert. Hierfür wurden die beiden Fachbereiche CPS-Installation sowie Process Mining für physische Materialflussprozesse bei KMU separat betrachtet. Für alle Arbeitspakete wurde im Sinne einer agilen Arbeitsweise regelmäßig Reporting betrieben und der Stand der Forschungsdokumentation an den Projektverlauf angepasst. Auf diesem Weg sind eine Vielzahl an Dokumenten entstanden, welche die Projektergebnisse darstellen und interessierten KMU zur Verfügung gestellt werden können.

Der im Antrag genannte Handlungsleitfaden für CPS-basiertes Process Mining konnte als händisches Artefakt nicht vollendet werden. Dies lässt sich damit begründen, dass die finale Kombination der beiden Bereiche CPS und Process Mining in ProCheck nicht vollzogen werden konnte. Durch die in Kapitel 2.4 geschilderten Schwierigkeiten bei der Verheiratung von CPS und Process Mining im Projektverlauf wurde deutlich, dass die Eignung einer Produktivitätsumgebung zur Kombination beider Themen ein sehr anspruchsvolles Thema ist und immer erst vor Ort entschieden werden kann. Daher wird die Identifikation der tatsächlichen Nutzungspotenziale durch Anwendung unserer Projektergebnisse weiterhin eine begleitende fachmännische Beratung erfordern. Anstelle des Handlungsleitfadens sind separate Dokumente, die eine wichtige Basis für eine derartige Beratungsleistung darstellen, entstanden:

- Zum einen zur Installation passender Sensorik für kleinteiligen Produktionsumgebungen,
- zum anderen zur Anwendung von Process Mining auf Produktionsdaten (insb. Clustering-Verfahren).

Diese Dokumente beinhalten detaillierte Informationen für interessierte KMU zum Aufbau von CPS zur Digitalisierung ihrer Fertigungsprozesse, mit deren Hilfe KMU bei der Digitalisierung der Fertigung unterstützend begleitet und für zukünftige Datenanalyseverfahren vorbereitet werden. Durch die interdisziplinäre Beschäftigung mit beiden Themenbereichen und deren Beziehungen

zueinander entstand in den Forschungseinrichtungen außerdem ein tieferes fachliches Verständnis zur Analyse von Produktionsdaten, welches sich sowohl dazu eignet weitergehende Forschung zu betreiben als auch den meist nicht technikaffinen KMUs einen echten Mehrwert bietet.

Außerdem wurden in einer öffentlichen Online-Veranstaltung die Projektergebnisse vorgestellt und mit den Teilnehmenden diskutiert. Aufgrund der Teilnahme sowohl des PA als auch anderer Interessierter konnten die im Laufe der Projektphase durchgeführten Arbeiten nochmals an der Praxis gespiegelt und validiert werden. Es zeigte sich, dass gerade für KMU in der Kombination beider Ansätze immer noch großes Forschungspotenzial steckt und zukünftige Projekte die Arbeiten aus ProCheck weiterführen können. Sobald es gelingt, den Auftragsbezug mithilfe von Sensorik herzustellen, können die Potenziale erst richtig gehoben werden. Darunter fällt u. a. ein ganzheitliches, digitales Abbild der Fertigungsprozesse von KMU und die damit verbundenen Chancen für fortschrittliche, datengetriebene Prozessanalysen.

3. Ergebniszusammenfassung und Ausblick

Abschließend werden die Projektergebnisse zusammenfassend dargestellt, zueinander in Relation gesetzt und ein Ausblick gegeben. Für einen schematischen Überblick des im Projekt umgesetzten Vorgehens und der erzielten Ergebnisse dient Abbildung 31.

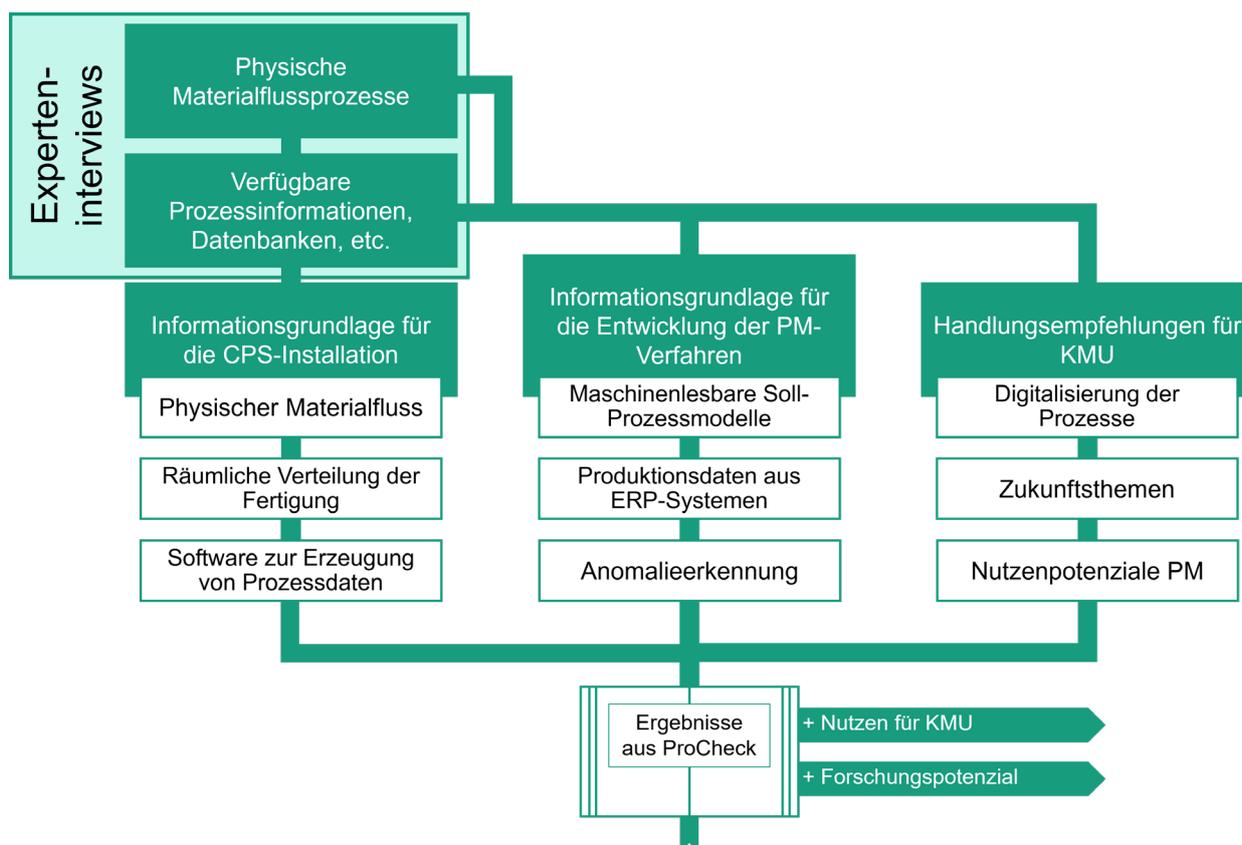


Abbildung 31: Projektübersicht und Inhalte

Grundlage der Projektarbeiten bildeten die in AP 1 durchgeführten Experteninterviews sowie die Vor-Ort-Aufnahme der Fertigungsprozesse. Mithilfe des Expertenwissens aus der Praxis gelang es, repräsentative Materialflussprozesse zu extrahieren. Durch die individuellen Antworten der

Praxispartner konnten die Soll-Prozesse innerhalb der Unternehmen ausgewählt und erfasst werden. Zudem wurden die wichtigsten Informationsflüsse, Datenbanken und deren Struktur ermittelt, was einen Überblick über den Grad der Digitalisierung der Prozesse ermöglichte. Die größte Herausforderung lag dabei in der Auswahl von für das Forschungsvorhaben geeigneten Prozessen. Kriterien wie die Verfügbarkeit von digitalen Prozessinformationen wurden dabei genauso betrachtet wie die Generalisierbarkeit der Materialflüsse und deren Durchführung. Die Arbeiten aus AP 1 schafften zum einen Verständnis über die physischen Bearbeitungsschritte innerhalb der Fertigung. Zum anderen konnte eine Wissensbasis über den Grad der Digitalisierung und vorhandene Prozessdaten, für die im weiteren Projektverlauf anschließenden Arbeitspakete bereitgestellt werden.

Diese Produktionsprozesse der Anwendungspartner galt es mithilfe der Arbeiten in ProCheck mit passender Sensorik auszustatten, Daten zu erheben und darauf aufbauend passende Process Mining-Verfahren anzuwenden.

Für die Bearbeitung von AP 2 waren die Vorarbeiten essenziell. Hier wurden in einem zweiten Schritt die ausgewählten Prozesse in maschinenlesbarer Form modelliert. Wichtig war an dieser Stelle vorab eine Anforderungsdefinition an die maschinenlesbare Modellierung von Soll-Prozessen. Nach der Festlegung der Anforderungen konnte mit Camunda ein passendes Modellierungswerkzeug ausgewählt, die Prozesse maschinenlesbar modelliert und abschließend auf ihre Funktionalität getestet werden.

Die Informationen über die Fertigungsprozesse und deren räumliche Verteilung waren maßgebend für den Aufbau und Betrieb eines CPS zur Erhebung von Prozessdaten. Da pandemiebedingt bei einem der Anwendungspartner (FP 2) erhebliche Ressourcenprobleme entstanden und eine Weiterarbeit im Projekt sehr unsicher war, wurde sich ab diesem Zeitpunkt vermehrt auf die Fertigungsprozesse von FP 1 fokussiert. Nach ausgiebiger Prüfung der Situation vor Ort konnte an zwei aussagekräftigen Teilbereichen eine CPS-Installation erfolgen und damit Produktionsprozessdaten erhoben werden.

Anschließend wurden die über die Sensorik erhobenen Daten analysiert und algorithmische Verfahren zur Event-Log Extraktion entwickelt. Aufgrund der erheblichen pandemiebedingten Verzögerungen im Projekt sowie der prozessualen Gegebenheiten vor Ort konnte die automatische Erzeugung von Event-Logs nicht zufriedenstellend erreicht werden. Um das Projektergebnis dennoch erreichen zu können, wurde im Projektkonsortium entschieden, tiefere Analysen auf den CPS-Daten durchzuführen.

Da die ursprünglich geplante Entwicklung der Process Mining-Methoden auf den reinen CPS-Daten nicht realisiert werden konnte, wurden vorhandene Prozessdaten aus dem ERP-System des Anwendungspartners genutzt. Dieser Schritt der weitgehend vom CPS-Aufbau entkoppelten Methodenentwicklung war notwendig, um den weiteren Projektverlauf sicherstellen zu können und das Projektergebnis nicht zu gefährden. Die im AP 5 entwickelten Verfahren zum Abgleich und zur Analyse von Produktionsprozessen sind daher in der Lage, sowohl Event-Logs sowohl aus ERP- als auch CPS-Rohdaten zu verarbeiten.

Die Vor-Ort-Besuche bei den Anwendungspartnern sowie der regelmäßige Kontakt zu den Mitgliedern des Projektbegleitenden Ausschusses gewährleisteten eine zielgerichtete Bearbeitung der Arbeitspakete.

Die hohen Teilnehmerzahlen bei den Projektveranstaltungen sowie der positive Zuspruch aus der Praxis zeigen, dass das Thema der datengetriebenen Prozessanalyse von Materialflüssen einen hohen Stellenwert bei den Unternehmen besitzt. Aus diesem Grund erscheint eine Fortsetzung und Weiterentwicklung der Forschungsarbeiten im Rahmen von ProCheck dringend notwendig. Ausgehend vom Forschungsstand nach Abschluss des Projekts ProCheck bieten sich neue Lösungsansätze, welche auf der CPS-Installation und der damit verbundenen erhöhten Datenverfügbarkeit aufbauen können. Mit dieser Verbesserung lässt sich der Digitalisierungsgrad der Prozesse steigern und weitere Anwendungspotenziale wie z. B. datengetriebene Analysen (Forecasts, Darstellung von Zusammenhängen, etc.) erschließen.

Im Nachgang zum Projekt sind diverse Maßnahmen zur Öffentlichkeitsarbeit geplant (siehe Kapitel 6).

4. Verwendung der Zuwendung

4.1. Forschungsstelle 1: Fraunhofer IIS

- Wissenschaftlich-technisches Personal (Einzelansatz A.1 des Finanzierungsplans): Es wurden 10 Mitarbeitende im Umfang von 28,54 Personenmonaten von den beantragen 30,06 eingesetzt. Die Differenz kommt v. a. aufgrund der zeitweisen Mitarbeit höher eingestufte Führungskräfte zustande, deren Mehrkosten nicht anerkannt werden konnten und daher von der Forschungsstelle getragen wurden.
- Geräte (Einzelansatz B des Finanzierungsplans): Es wurden keine Finanzmittel für Gerätebeschaffung beantragt.
- Leistungen Dritter (Einzelansatz C des Finanzierungsplans): Es wurden keine Leistungen für die Arbeit Dritter beantragt.

4.2. Forschungsstelle 2: Ludwig-Maximilians-Universität München

- Wissenschaftlich-technisches Personal (Einzelansatz A.1 des Finanzierungsplans): Es wurde 2 wissenschaftliche Mitarbeitende im Umfang von 16 Personenmonaten (einer davon mit 100 %, der andere mit 50 % einer Vollzeittätigkeit) eingesetzt.
- Geräte (Einzelansatz B des Finanzierungsplans): Es wurden keine Finanzmittel für Gerätebeschaffung beantragt.
- Leistungen Dritter (Einzelansatz C des Finanzierungsplans): Es wurden keine Leistungen für die Arbeit Dritter beantragt.

5. Notwendigkeit und Angemessenheit der geleisteten Arbeit

Die Arbeiten wurden entsprechend des Umfangs und ihrer Notwendigkeit von den eingesetzten wissenschaftlichen Mitarbeitern mit Unterstützung studentischer und wissenschaftlicher Hilfskräfte, den PA-Mitgliedern sowie von Vertretern weiterer interessierter Unternehmen durchgeführt. Als Orientierung diente dabei der genehmigte Arbeitsplan des Forschungsprojekts. Sofern von im Antrag genannten Methoden abgewichen wurde, beruhte das wie in den jeweiligen Kapiteln geschildert auf Erfordernissen, die sich im Projektverlauf ergaben. Die angestrebten Ergebnisse wurden erreicht.

Generell wurden aufgrund der Corona-Situation geplante ganz- und mehrtägige Vor-Ort-Workshops durch häufige kurze (Video-)Telefonate und informelle Feedback-Schleifen ersetzt, da pandemiebedingt in der Wirtschaft teilweise enge Zeitrestriktionen, Personal- und Ressourcenmangel auftraten. Die Unsicherheit bezüglich Verfügbarkeit der Anwendungspartner und dem weiteren Verlauf der Pandemie hat im Projekt u. a. zu erheblichen Verzögerungen geführt. Insbesondere der Aufbau der CPS-Umgebung konnte aus den genannten Gründen erst mit erheblicher Verzögerung beginnen.

Nach den in AP 3 dargestellten Analysen und in AP 4 aufgeführten Gründen wurde das Vorgehen im Projekt angepasst. Um die Projektergebnisse zu erreichen, wurde die Entwicklung der Process Mining-Methoden parallel zu AP 4 auf bereits bestehenden Prozessdaten der ERP-Systeme durchgeführt. Dies war dahingehend sinnvoll, da die Generierung von Event-Logs aus CPS-Rohdaten zu dieser Zeit im Projekt nicht umsetzbar war. Die Entwicklung der Methoden in AP 5 wurde so durchgeführt, dass bei einer erfolgreichen Generierung von Event-Logs aus den CPS diese für Process Mining genutzt werden können.

Ursprünglich geplante Vorstellungen des eingesetzten CPS zur Generierung von Event-Logs wurden ersetzt durch Beratungen von Unternehmen hinsichtlich des Aufbaus einer geeigneten CPS-Umgebung, Nutzenpotenziale und nächste Schritte zur Realisierung der Analyse mithilfe von Process Mining. Deshalb ist es nicht nur pandemiebedingt zu Abweichungen der Projektergebnisse gekommen, sondern auch aus inhaltlicher Notwendigkeit.

Die Ergebnisse entsprechen dennoch der Zielsetzung des Antrags, physische Materialflussprozesse mit CPS auszustatten und mithilfe von Process Mining zu analysieren.

6. Wissenschaftlich-technischer und wirtschaftlicher Nutzen insbesondere für KMU

Der wissenschaftlich-technischen Nutzen des Projekts liegt in der Synthese verschiedener wissenschaftlicher Ansätze und Methoden im Bereich der Produktion. In Zukunft sollen zahlreiche KMU von den Projektergebnissen profitieren können. Dazu zählen insbesondere Unternehmen des fertigenden Gewerbes mit vielen räumlich verteilten, physischen Materialflussprozessen. Durch die Anwendung können Unternehmen ohne Zutun Dritter ihre Fertigungsprozesse überwachen, steuern und Handlungsempfehlungen ableiten. Durch das strukturierte Vorgehen und den Mix zwischen Experteninterviews, Analysen und darauf aufbauenden Schritten wurde ein neuer Lösungsansatz geschaffen:

Beispielsweise konnten durch Process Mining gleichartige Aufträge geclustert und andersartige Aufträge als Anomalien identifiziert werden. Diese Informationen dienen den Entscheidungsträgern vor Ort Transparenz über die Fertigungsprozesse zu erhalten und einzelne Bearbeitungsschritte zu optimieren. Durch die Analyse der an den Stationen installierten Sensorik konnten detaillierte Informationen zur Auslastung von z. B. einzelnen Maschinen geliefert und dadurch eine gezielte Auslastungssteuerung angestoßen werden.

Im Anschluss an das Projekt sollen weitere Entwicklungen dafür sorgen, dass die Industrietauglichkeit des Systems sichergestellt werden kann. Gemeinsam mit Partnern aus dem Projekt treibt das Fraunhofer IIS diese Entwicklung. Dabei steht die Integration des Systems in den Fertigungs-

alltag im Mittelpunkt. Hierfür ist es entscheidend, mithilfe der eingesetzten Sensorik einen Auftragsbezug zu schaffen und diesen in den Daten abzubilden. Darauf aufbauend können die in ProCheck entwickelten PM-Verfahren eingesetzt werden, um tiefgreifende Analysen über physische Materialflussprozesse bei KMU durchführen zu können.

Die Ergebnisse von ProCheck münden darüber hinaus in zahlreichen Weiterentwicklungen in anderen Forschungsvorhaben. Das erarbeitete Wissen wird in Forschung und Lehre aktiv vermittelt.

Die im Langantrag vorgestellten Transfermaßnahmen für das Forschungsprojekt konnten im Rahmen der Projektlaufzeit größtenteils durchgeführt werden. Durch die anhaltende Corona-Pandemie konnten allerdings einige Live-Vorstellungen auf Messen etc. nicht wie geplant durchgeführt werden. Als Ausgleich dafür wurden zusätzliche Veröffentlichungen in Print- und Online-Medien durchgeführt. Die Maßnahmen haben zielführend dazu beigetragen das Projekt ProCheck interessierten Unternehmen nahezubringen und Teilnehmende für die PA-Veranstaltung zu gewinnen. Im Folgenden ist eine ausführliche Liste aller im Projekt geplanten und durchgeführten Transfermaßnahmen mit entsprechender zeitlicher Zuordnung aufgeführt.

Maßnahme A: Projektbegleitender Ausschuss (PA)	Der Projektbegleitende Ausschuss (PA) stellt ein tragendes Element in der Konzeption dieses Forschungsvorhabens dar. Er sichert den engen Praxisbezug und sorgt zugleich für eine frühzeitige Weitergabe von Untersuchungsergebnissen während der Projektlaufzeit.	A1 - Vorstellung des Projekts, Diskussion der geplanten Arbeiten und Einbringung von Anforderungen der KMU	Durchgeführt am 02.02.2021
		A2 - Vorstellung des eingesetzten CPS-Systems; Diskussion der Zwischenergebnisse und des weiteren Vorgehens in Workshops	Durchgeführt am 20.01.2022
		A3 - Abschlusspräsentation, Vorstellung der Fallstudien und des Leitfadens und Diskussion der erzielten Ergebnisse mit dem PA	Durchgeführt am 17.11.2022
Maßnahme B: Veranstaltungen	Ergebnistransfer in die Wirtschaft durch Vorstellung von (Teil-) Ergebnissen des Projekts auf Veranstaltungen	B1 - Vorstellung des eingesetzten CPS-Systems; Ergebnispräsentation auf Logistik-Veranstaltungen, z. B. Logistik-Forum oder Materialfluss-Kongress, sowie auf einem technischen Kongress, z. B. Wireless Congress	Corona-bedingt abgesagt
		B2 - Ergebnisvorstellung auf dem BVL-Kongress	Durchgeführt am 15.10.2021
		B3 - Ergebnispräsentation auf einer öffentlichen Projektabschlussveranstaltung	In Zusammenhang mit Maßnahme A3 durchgeführt am 17.11.2022

Maßnahme C: Internetdarstellung	Elektronische Verbreitung der Forschungsinhalte zur Information weiterer interessierter Unternehmen	C1 - Vorstellung des Projekts über Newsletter des Fraunhofer IIS	Februar 2020
		C2 - Internetauftritt des Forschungsprojekts auf den Webseiten des Fraunhofer IIS, der LMU und des BVL-Campus (plus ggf. der PA-Mitglieder)	Durchgeführt ³
Maßnahme D: Messepräsentation	Präsentation von (Teil-) Ergebnissen auf Messen	D1 - Präsentation des Projekts auf der „Logi-Mat“	Corona-bedingt abgesagt
		D2 - Präsentation des Projekts auf der „Hannover Messe“	
Maßnahme E: Veröffentlichungen	Zeitnahe Verbreitung der Ergebnisse durch Veröffentlichungen in Fachzeitschriften	E1 - Veröffentlichung von (Teil-) Ergebnissen in einschlägigen Fachzeitschriften wie „Logistik heute“, „DVZ – Deutsche Logistikzeitung“ und ähnlichen Zeitschriften der relevanten Industriezweige. ⁴	Zeiträume s. Tabelle 3
		E2 - Der Handlungsleitfaden zum Einsatz von CPS-basiertem Process Mining wird von den Forschungspartnern veröffentlicht.	Ersetzt durch andere Informationsmaterialien
Maßnahme F: Übernahme in die Lehre	Einbringung der (Teil-) Ergebnisse des Projekts in den Lehrbetrieb	F1 - Mitarbeit studentischer Hilfskräfte	Durchgehend während der Projektlaufzeit
		F2 - Betreuung von Abschlussarbeiten (Bachelor)	Titel: „Resource Usage Outlier Detection“ (Februar 2022)
		F3 - Betreuung von Abschlussarbeiten (Master)	Titel: „Temporal and Spatial Micro-Clustering for Trajectory Process Data“ (Dezember 2021)
Maßnahme G: Beratung	Ergebnistransfer an KMU ohne eigene Forschungsaktivitäten	G1 - Beratung von Unternehmen des aufgezeigten Ökosystems zum Einsatz von CPS-basiertem Process Mining	ab Dezember 2022

³ Für den Internetauftritt siehe <https://www.scs.fraunhofer.de/de/referenzen/procheck-process-mining-produktion.html> und <https://www.iis.fraunhofer.de/de/ff/lv/dataanalytics/anwproj/procheck.html>.

⁴ s. [Tabelle 3](#)

Maßnahme H: Forschungsbericht	Ergebnistransfer in die Wirtschaft	H1 - Der Schlussbericht wird auf den Webseiten des Fraunhofer IIS, der LMU und des BVL-Campus sowie in gedruckter Form verfügbar sein	Bis Q2/2023
		H2 - Der Handlungsleitfaden zum Einsatz von CPS-basiertem Process Mining wird an interessierte KMU verteilt	Ersetzt durch andere Informationsmaterialien
Maßnahme I: Lehre	Vorstellung der Ergebnisse auf Veranstaltungen und Kongressen	J1 - Vorstellung der Ergebnisse auf Logistikveranstaltungen, bspw. „Logistik-Forum“ oder „Technologietag des Fraunhofer IIS“	Vorstellung auf der BPMDS (Business Process Modeling, Development, and Support) – Konferenz im Juni 2023
Maßnahme K: Messepräsentation	Zeitnahe Verbreitung der Ergebnisse in die Wirtschaft durch Messeauftritte	K1 - Vorstellung der Ergebnisse auf der „LogiMat“ und „transport logistic“	Voraussichtlich Q2/2023
		K2 - Präsentation des Projekts auf der „Hannover Messe“	
Maßnahme L: Dissertationen	Verbindung des Themas mit einer möglichen Dissertation	L1 - Nach Abschluss des Forschungsprojekts kann das Thema im Rahmen einer Dissertation weiter verfolgt werden	In Arbeit

Beitrag	Typ
Fleischmann, L.: Combining Process Mining and Cyber-Physical Systems for material flow processes to ensure transparency (BVL-Kongress Pre-Event, Oktober 2021)	Vortrag inkl. Diskussion
Fleischmann, L.; Buck, M.: Process Mining von Materialflüssen. „Pro-Check“ – Forschungsprojekt zur Verifikation innerbetrieblicher Materialflussprozesse (Technische Logistik 7-8/2022)	Veröffentlichung in der Fachpresse
Fleischmann, L.; Vaupel, T.; Janda, P.: Fehlende Transparenz von auftragsbezogenen, physischen Produktionsabläufen. Cyber-Physische Systeme zur Erhebung von Kennzahlen (ident Jahrbuch 2023)	Veröffentlichung in der Fachpresse
Fleischmann, L.: Fachartikel zum Einsatz von Process Mining für physische Materialflussprozesse (ident Ausgabe 5/23)	Geplant Q3/2023: Veröffentlichung in der Fachpresse

Maldonado, A.; Fleischmann, L.; Seidl, T. (2023): Granularity aware Trace Clustering in Manufacturing: A Case Study.	Eingereicht Q1/2023: Wissenschaftliche Publikation für die BPMDS 2023 ⁵
Maldonado, A. (2023): Granularity aware Trace Clustering in Manufacturing: A Case Study.	Geplant bis Q4/2023: Vortrag auf einem internationalen Kongress wie z. B. ICPM, ISF, ICBPMN, BPMDS, etc.

Tabelle 3: Veröffentlichungen zu ProCheck

7. Durchführende Forschungsstellen

Forschungsstelle 1: Fraunhofer-Institut für Integrierte Schaltungen IIS Arbeitsgruppe für Supply Chain Services

Anschrift: Nordostpark 84,
90411 Nürnberg

Telefon/Fax: 0911 58061-9553 / 0911 58061-9599

Leiter der Forschungsstelle: Prof. Dr.-Ing. Albert Heuberger

Projektleiter: Lukas Fleischmann

Das Fraunhofer-Institut für Integrierte Schaltungen IIS in Erlangen ist eine weltweit führende anwendungsorientierte Forschungseinrichtung für mikroelektronische und informationstechnische Systemlösungen und Dienstleistungen. Es ist heute das größte Institut der Fraunhofer-Gesellschaft. Die Forschung am Fraunhofer IIS orientiert sich an zwei Leitthemen:

- Audio und Medientechnologien: Hier prägt das Institut seit mehr als 30 Jahren die Digitalisierung der Medien.
- Kognitive Sensorik: hier erforscht das Institut Technologien für Sensorik, Datenübertragungstechnik, Datenanalysemethoden sowie die Verwertung von Daten im Rahmen datengetriebener Dienstleistungen und entsprechender Geschäftsmodelle. Damit wird die Funktion des klassischen „intelligenten“ Sensors um eine kognitive Komponente erweitert. Die Forschungsergebnisse finden Anwendung in der vernetzten Mobilität, in Kommunikations- und Anwendungslösungen für das Internet der Dinge, in der Digitalisierung der

⁵ Für weitere Informationen siehe: <https://sites.google.com/view/bpmlds/>

menschlichen Wahrnehmung (Human Sensing), im Produkt- und Materialmonitoring sowie in Business Analytics in Versorgungsketten (Supply Chains).

Am Vorhaben ProCheck nehmen zwei Bereiche am Standort Nürnberg teil:

- Der Bereich „Lokalisierung und Vernetzung LV“ erforscht und entwickelt im Schwerpunkt drahtlose Ortungs- und Vernetzungstechnologien für Industrial Internet of Things (IIoT) Anwendungen. In der Abteilung Kommunikationsnetze wurden in den letzten 10 Jahren verschiedene Technologieplattformen für die Realisierung von CPS entwickelt.
- Der Bereich „Arbeitsgruppe für Supply Chain Services SCS“ beschäftigt sich seit 1995 aus wirtschaftswissenschaftlicher Sicht mit der Optimierung der Supply Chain: In unserem logistischen Verständnis, das explizit den gesamten Wertschöpfungsprozess umfasst, verbinden wir Technologie-Expertise und Service-Know-how und finden so branchenübergreifend Lösungen, die die Möglichkeiten der gesamten Supply Chain berücksichtigen. Insbesondere die Gruppe Process Intelligence analysiert und gestaltet Prozesse der Logistik und Produktion im Kontext der digitalen Wertschöpfung.

Forschungsstelle 2: Ludwig-Maximilians-Universität München (LMU)

Institut für Informatik, Lehrstuhl für Datenbanksysteme und Data Mining

Anschrift: Oettingenstr. 67, 80538 München

Telefon/Fax: 0911 58061-9553 / 0911 58061-9599

Leiter der Forschungsstelle: Prof. Dr. Thomas Seidl

Der Lehrstuhl für Datenbanken und Data Mining an der LMU München hat eine mehr als 20-jährige Erfahrung in der Grundlagenforschung zu Methoden in den Bereichen Data Science, Data Mining, maschinelles Lernen, künstliche Intelligenz und Datenbanktechnologien in Ingenieur-, Wirtschafts-, Lebens- und Geisteswissenschaften. Ein wesentlicher Schwerpunkt der aktuellen Forschung ist das Gebiet Process Mining, insbesondere Event Stream Mining, Online Drifterkennung in Prozessen und Wissenstransfersystemen. Über verschiedene etablierte Formate wie dem Data Science Lab der LMU bringt die Arbeitsgruppe ihre Forschungsergebnisse durch gemeinsame Projekte mit Firmen auch regelmäßig in die Praxis.

8. Anhang

8.1. Befragte Experten im Rahmen von AP 1

Im Rahmen der Interviewstudie aus AP 1 wurden folgende Unternehmen zur gegenwärtigen Lage hinsichtlich der physischen Materialflussprozesse und deren Grad der Digitalisierung befragt:

Unternehmen	Art
Gmöhling Transportgeräte GmbH	Fertigung von Transportgeräten und -behältern
Aufzugswerke Schmitt+Sohn GmbH & Co. KG	Fertigung von Aufzügen (Betrachtung der Maschinenträger am Standort Nürnberg-Kornburg)

Tabelle 4: Befragte Experten im Rahmen von AP 1

8.2. Im Projekt involvierte Unternehmen während der Projektlaufzeit

Unternehmen	Art
AST-X	CPS-Hersteller
Rauschert	Fertigendes Gewerbe
PASS Stanztechnik AG	Fertigendes Gewerbe
Jörg Battel	Freiberufler
Salt-Solutions	Technologien für Logistik & Produktion
Cosmino AG	Softwareentwicklung
Bayern Innovativ GmbH	Forschungsträger
FAU Erlangen-Nürnberg	Universität
KIMO Industrial Electronics GmbH	U. a. Hersteller elektronischer Geräte
Celonis	Softwarehersteller
Comarch	IT-Dienstleister und Softwarehersteller
Qualigon	Unternehmensberatung
Hiller GmbH	Maschinenbau
CNX Consulting	Unternehmensberatung
Aufzugswerke Schmitt+Sohn GmbH & Co. KG	Fertigendes Gewerbe
Bosch IO	Softwarehersteller
Gmöhling Transportgeräte GmbH	Fertigendes Gewerbe
Sykatech GmbH	Systempartner für Metalltechnik

Tabelle 5: Im Projekt involvierte Unternehmen während der Projektlaufzeit

Literaturverzeichnis

- Arnold, D. (2006): Intralogistik. Potentiale, Perspektiven, Prognosen. Berlin Heidelberg: Springer-Verlag GmbH (VDI-Buch)
- Bahl, P.; Padmanabhan, V. (2000). RADAR: an inbuilding RF-based user location and tracking system. In: INFOCOM 2000. Nineteenth Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings. IEEE Bd. 2 IEEE, S. 775–784
- Baier, T., Mendling, J., Weske, M.: Bridging abstraction layers in process mining. *Information Systems* 46, 123–139 (2014)
- Berti, A., van Zelst, S.J., van der Aalst, W.M.P.: Process Mining for Python (PM4Py) : Bridging the Gap Between Process- and Data Science. In: ICPMD 2019, ICPM Demo Track 2019. pp. 13–16. RWTH Aachen, <https://publications.rwth-aachen.de/record/774274>
- Burattin, A. (2017). Online conformance checking for petri nets and event streams. Online Proceedings of BPM Demo Track. CEUR-WS. org
- Burattin, A. (2018). Streaming process discovery and conformance checking. In *Encyclopedia of Big Data Technologies*. Springer
- Burattin, A., Sperduti, A., & van der Aalst, W. M. (2014, July). Control-flow discovery from event streams. In *IEEE Congress on Evolutionary Computation* (pp. 2420-2427).
- Burattin, A., van Zelst, S. J., Armas-Cervantes, A., van Dongen, B. F., & Carmona, J. (2018, September). Online conformance checking using behavioural patterns. In *International Conference on Business Process Management* (pp. 250-267). Springer, Cham
- Carrera, V. J. L., Zhao, Z., & Braun, T. (2018, September). Room Recognition Using Discriminative Ensemble Learning with Hidden Markov Models for Smartphones. In *2018 IEEE 29th Annual International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)* (pp. 1-7). IEEE
- Dakic, Dusanka; Stefanovic, Darko; Cosic, Ilija; Lolic, Teodora & Medojevic, Milovan (2018). Business process mining application: a literature review. In: *Proceedings of the 29th DAAAM International Symposium*. DAAAM International. Vienna
- De Medeiros, A. A., van der Aalst, W. M., & Weijters, A. J. M. M. (2008). Quantifying process equivalence based on observed behavior. *Data & knowledge engineering*, 64(1), 55-74
- De Weerd, J., Wynn, M.T.: Foundations of process event data. *Process Mining Handbook*. LNBIP 448, 193–211 (2022)
- Dumas, M., La Rosa, M., Mendling, J., & Reijers, H. A. (2018). *Fundamentals of business process management* (2. Auflage). Heidelberg: Springer

- Eißen J. (2019). Prozessoptimierung 4.0 – Process Mining als Erfolgsfaktor der digitalen Transformation. <https://www.capgemini.com/de-de/2019/06/process-mining/>
- Ester, M., Kriegel, H.P., Sander, J., Xu, X., et al.: A density-based algorithm for discovering clusters in large spatial databases with noise. In: kdd. vol. 96, pp. 226–231 (1996)
- Feldhorst S, ten Hompel M (2016). Bewegungsklassifikation mithilfe mobiler Sensoren. Logistics Journal : Proceedings, Vol. 2016
- Förderprojekt Inwipro „Intelligenter Warenbegleitträger zur industriellen Prozesssteuerung“, Laufzeit 2017-2019
- Ganschar, O. (2013): Produktionsarbeit der Zukunft - Industrie 4.0. Studie. Hg. v. D. Spath. Stuttgart: Fraunhofer Verlag
- Hanak A., Janda P., Meyer S. (2018). Optimizing Production Processes with Wireless Smart Sensors and Tracking. Wireless Congress 2018 Munich
- Hassani, M., Siccha, S., Richter, F., & Seidl, T. (2015). Efficient process discovery from event streams using sequential pattern mining. In Computational Intelligence, 2015 IEEE Symposium Series on Computational Intelligence, pp. 1366-1373
- Leemans, S.J., Fahland, D., Van Der Aalst, W.M.: Discovering block-structured process models from event logs containing infrequent behaviour. In: Business Process Management Workshops: BPM 2013 International Workshops, Beijing, China, August 26, 2013, Revised Papers 11. pp. 66–78. Springer (2014)
- Lerch C., Jäger A., Maloca S. (2017). Wie digital ist Deutschlands Industrie wirklich? Arbeit und Produktivität in der digitalen Produktion. In: Mitteilungen aus der ISI-Erhebung - Modernisierung der Produktion 71. Fraunhofer ISI
- Lieb, C.; Klenk, E.; Galka, S.; Keuntje, C. (2017). Einsatz von Routenzugsystemen zur Produktionsversorgung - Studie zu Planung, Steuerung und Betrieb. Lehrstuhl fml. Garching
- Marin-Castro, H.M., Tello-Leal, E.: Event log preprocessing for process mining: A review. Applied Sciences 11(22) (2021). <https://doi.org/10.3390/app112210556>, <https://www.mdpi.com/2076-3417/11/22/10556>
- Miclo, R., Fontanili, F., Marquès, G., Bomert, P., & Luras, M. (2015, August). RTLS-based Process Mining: Towards an automatic process diagnosis in healthcare. In: 2015 IEEE International Conference on Automation Science and Engineering (CASE) (pp. 1397-1402). IEEE
- Nielsen, F.: Hierarchical Clustering, pp. 195–211. Springer International Publishing, Cham (2016). https://doi.org/10.1007/978-3-319-21903-5_8, https://doi.org/10.1007/978-3-319-21903-5_8
- Pflaum, A., Hohmann, C., Hofmann, B., Boppert, J., & Klötzer, C. (2014). Industrie 4.0 und CPS – Bedarfe und Lösungen aus Sicht des Mittelstands. München: bayme vbm
- Pfohl HC (2018). Produktionslogistik. In: Logistiksysteme. Springer Vieweg, Berlin, Heidelberg

- Pourmirza, S., Peters, S., Dijkman, R., Grefen, P.: Bpms-ra: a novel reference architecture for business process management systems. *ACM Transactions on Internet Technology (TOIT)* 19(1), 1–23 (2019)
- Rajkumar, R.R., Lee, I., Sha, L. and Stankovic, J. (2010). Cyber-physical systems: the next computing revolution. *Proceedings of the 47th Design Automation Conference* (pp. 731-736). ACM
- Rozinat, A., Van der Aalst, W.M.: Conformance checking of processes based on monitoring real behavior. *Information Systems* 33(1), 64–95 (2008)
- Sanislav, Teodora & Miclea, Liviu. (2012). Cyber-physical systems - Concept, challenges and research areas. *Control Engineering and Applied Informatics*. 14. 28-33
- Statistisches Bundesamt (Destatis) (2018). Produzierendes Gewerbe. Kostenstruktur der Unternehmen des Verarbeitenden Gewerbes sowie des Bergbaus und der Gewinnung von Steinen und Erden, Fachserie 4 Reihe 4.3. Statistisches Bundesamt, Wiesbaden
- Telematik Markt. (2019). „R2D“: Forschungsprojekt zur Digitalisierung der industriellen Produktion. Abgerufen 27. September 2019, von <https://telematik-markt.de/telematik/r2d-forschungsprojekt-zur-digitalisierung-der-industriellen-produktion>
- Van der Aalst, W. (2016): *Process Mining. Data Science in Action*. Second Edition. Heidelberg: Springer
- Van Der Aalst, W.: Process discovery from event data: Relating models and logs through abstractions. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* 8, e1244 (02 2018). <https://doi.org/10.1002/widm.1244>
- Van Der Aalst, W.: *Process mining: data science in action*, vol. 2. Springer (2016)
- Van Der Aalst, W.M., Carmona, J.: *Process mining handbook*. Springer Nature (2022)
- Vieritz H., Meisen T., Vaegs T., Jeschke S., Hassani M., Beecks C., Seidl T., Matthias Priesters, Irene Mittelberg, Paula Niemietz, Tatiana Serbina, Stella Neumann: e-cosmos: Cluster-Analysen zeitelastischer, multimodaler Daten in der Linguistik. *LingUnite RWTH Aachen Tag der Sprachforschung*. (Poster) (2013)
- W.M.P. van der Aalst, A. Adriansyah, and B. van Dongen. *Replaying History on Process Models for Conformance Checking and Performance Analysis*. *WIREs Data Mining and Knowledge Discovery*, 2(2):182–192, 2012
- Weijters, A. J. M. M., van Der Aalst, W. M., & De Medeiros, A. A. (2006). *Process mining with the heuristics miner-algorithm*. Technische Universiteit Eindhoven, Tech. Rep. WP, 166, 1-34
- Weijters, A.J., Van der Aalst, W.M.: Rediscovering workflow models from event-based data using little thumb. *Integrated Computer-Aided Engineering* 10(2), 151–162 (2003)
- Zukunftsbild „Industrie 4.0“, Herausgeber und Veranstalter, Bundesministerium für Bildung und Forschung, Referat IT-Systeme, 53170 Bonn