GEFÖRDERT VOM
Bundesministerium
für Bildung
und Forschung

Abschlussbericht

des im Rahmen der Fördermaßnahme KMU-innovativ: IKT

geförderten Verbundprojekts

DaWiS

Entwicklung einer integrierten Lösung für

das Data Farming und die Wissensentdeckung in Simulationsdaten

(Projektlaufzeit: 01.01.2020 – 31.12.2021)

des

Fachgebiets Informationstechnik in Produktion und Logistik der TU Ilmenau

Förderkennzeichen:	01 IS 190 42B
Projektleiter:	Prof. Dr.-Ing. Steffen Straßburger
Telefon:	+49 3677 69-4051
E-Mail:	steffen.strassburger@tu-ilmenau.de

Das diesem Bericht zugrundeliegende Vorhaben wurde mit Mitteln des Bundesministeriums für Bildung, und Forschung unter dem Förderkennzeichen 01 IS 190 42B gefördert.
Die Verantwortung für den Inhalt dieser Veröffentlichung liegt beim Autor.

Zur Wahrung berechtigter Interessen des Zuwendungsempfängers oder Dritter oder aus anderen sachlichen Gesichtspunkten aus dem Bericht vertraulich zu behandelnde Einzelheiten sind (z.B. zur Wahrung der Priorität bei Schutzrechtsanmeldungen) mit ausdrücklich Hinweisen gekennzeichnet.

Inhaltsverzeichnis

1 Aufgabenstellung und Zielsetzung 5

2 Voraussetzungen für die Projektbearbeitung 7

3 Planung und Ablauf des Vorhabens 8

4 Angeknüpfter wissenschaftlicher und technischer Stand 10

5 Zusammenarbeit mit anderen Stellen 11

6 Verwendung der Zuwendung und des erzielten Ergebnisses mit Gegenüberstellung der vorgegebenen Ziele 12

6.1 GZ1 – Methoden des smarten Experimentdesigns evaluieren und implementieren 12

6.2 GZ2 – Methoden zur (cloudbasierten) Verteilung von Experimenten entwickeln..... 13

6.3 GZ3 – Data-Mining- und Visual-Analytics-Methoden vorsehen und evaluieren..... 15

6.4 GZ4 – Integration und Umsetzung einer integrierten Plattform 16

6.5 GZ5 – Erprobung der entwickelten Lösung in Pilotanwendungen 19

6.6 Gesamtziel – Transferieren des Data Farmings und der Wissensentdeckung in Simulationsdaten in die Praxis 23

6.7 Gesamteinschätzung der erreichten Ziele 23

7 Wichtigste Positionen des zahlenmäßigen Nachweises.....24

8 Notwendigkeit und Angemessenheit der geleisteten Arbeit 25

9 Voraussichtlicher Nutzen und Verwertbarkeit der Ergebnisse im Sinne des fortgeschriebenen Verwertungsplans 29

10 Während der Durchführung des Vorhabens dem ZE bekannt gewordener Fortschritt auf dem Gebiet des Vorhabens bei anderen Stellen 30

11 Erfolgte oder geplante Veröffentlichungen des Ergebnisses.....30

1 Aufgabenstellung und Zielsetzung

Simulation ist ein etabliertes Werkzeug zur Planung und Steuerung komplexer Produktions- und Logistiksysteme. Insbesondere ist Simulation eine Schlüsselkomponente auch hinsichtlich der Herausforderungen im Kontext von Industrie 4.0 (Krückhans und Meier 2013). Der Ablauf traditioneller Simulationsstudien ist hierbei üblicherweise darauf ausgelegt, einen vorher definierten Projektumfang bzw. ein konkretes (Projekt-) Ziel durch manuelles Experimentieren zu erfüllen, z. B. die Optimierung eines Fertigungslayouts (Law 2003; Sargent 2013).

Mit steigender Rechenleistung und der allgemeinen Verfügbarkeit von Big-Data-Infrastrukturen und cloudbasierten Lösungen sowie Fortschritten auf dem Gebiet des Data Mining ist eine weitere Anwendungsmöglichkeit für Simulationsmodelle erwachsen: das Durchführen einer sehr großen Bandbreite von Experimenten, um versteckte, vorher unbekannte und möglicherweise nützliche Wirkzusammenhänge aufzudecken. Gerade in komplexen Systemen können möglicherweise Beziehungen, Probleme oder sogar Lösungen vorhanden sein, die über das definierte Projektziel einer klassischen Simulationsstudie hinausgehen, aber dennoch zur Entscheidungsunterstützung beitragen können. Grundlage bildet die Methodik des Data Farming (Horne und Meyer 2010). Hierbei wird ein vorab validiertes Simulationsmodell als Datengenerator genutzt, um mit Hilfe von intelligentem Experimentdesign und High Performance Computing ein möglichst großes oder gar vollständiges Spektrum an Modell- bzw. Systemverhalten (Wirkungsraum) abdecken zu können. Ähnlich wie ein Farmer, der sein Land möglichst effektiv und effizient kultiviert und damit versucht, seinen Ernteertrag zu maximieren, drückt die „Farming“-Metapher aus, dass hierbei der Datenertrag durch das Simulationsmodell maximiert werden soll (Sanchez 2014). Auf der erzeugten Datenbasis können dann mit Hilfe von Data-Mining- bzw. Visual-Analytics-Methoden interessante Zusammenhänge in den erzeugten Daten aufgedeckt werden (Fayyad et al. 1996; Keim et al. 2008). Ein entsprechendes Vorgehensmodell für die Wissensentdeckung in Simulationsmodellen im Kontext Produktion und Logistik wurde von Feldkamp et al. 2015 entwickelt.

Ziel und Aufgabenstellung des DaWis-Vorhabens war es, die im Bereich Produktion und Logistik bisher nur in akademischen Fallstudien betrachtete Methode der Wissensentdeckung in Simulationsdaten auf die Praxis zu transferieren. Dazu gab es folgende fünf Ziele zu erreichen:

GZ1 Methoden des smarten Experimentdesigns zu evaluieren und zu implementieren, hierbei ist auf die Abstraktion von der bzw. die Anpassbarkeit auf die konkrete Anwendung/Modell zu achten

GZ2 Methoden zur (cloudbasierten) Verteilung von Experimenten zu entwickeln, um die benötigte Rechenleistung zu gewähren

GZ3 Data-Mining- und Visual-Analytics-Methoden vorzusehen und geeignete Frameworks für die Implementierung zu evaluieren

GZ4 Eine integrierte Plattform zu konzipieren und umzusetzen

GZ5 Die Erprobung der entwickelten Lösung in Pilotanwendungen von Anwendungspartnern

Auf die fünf Ziele und die Zielerreichung wird in Abschnitt 6 eingegangen. Konkreter Ausgangspunkt für die Arbeiten im Projekt waren die Vorarbeiten von Feldkamp et al. 2015 und insbesondere die in Abbildung 1 dargestellten Schritte.

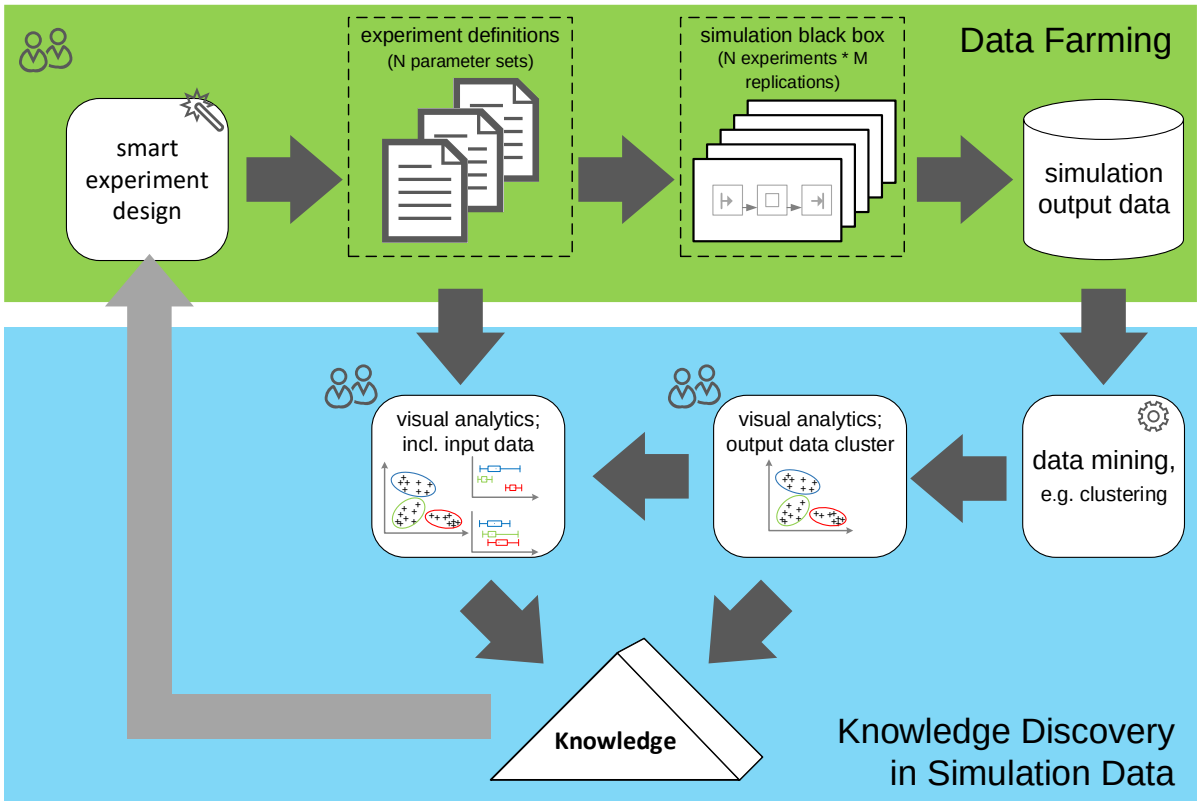

Abbildung 1: Wissensentdeckung in Simulationsdaten nach Feldkamp et al. 2015

2 Voraussetzungen für die Projektbearbeitung

Wesentliche wissenschaftliche Grundlagen des Vorhabens waren die Vorarbeiten und die Erfahrungen des Fachgebiets Informationstechnik in Produktion und Logistik der TU Ilmenau. Zu diesen Vorarbeiten gehört insbesondere auch eine Fallstudie in Kooperation mit Volkswagen und der Audi AG, bei der ein Simulationsmodell für die Intralogistik einer Fahrzeugendmontage untersucht wurde. Die entsprechenden Veröffentlichungen von Feldkamp et al. 2019 und Strassburger et al. 2018 waren die Impulse, die letztendlich zum DaWiS-Projekt geführt haben. Die Fallstudie hat das grundsätzliche Potenzial der Methoden rund um das Data Farming an einem konkreten Anwendungsfall in Produktion und Logistik gezeigt. Allerdings sind dabei und auch bei anderen Fallstudien des Fachgebiets die in Abbildung 1 **Fehler! Verweisquelle konnte nicht gefunden werden.**gezeigten Schritte im W esentlichen manuell und in unterschiedlichen Softwarewerkzeugen sowie auf den jeweiligen Einzelfall abgestimmt durchgeführt worden.

Eine wesentliche Motivation der Projektpartner war von Beginn an, die Schritte in eine (teil-)automatisierte, standardisierte und gleichzeitig hinreichend flexible Vorgehensweise einzubetten, die auch Nicht-Experten erlaubt, Data-Farming-Methoden zum Einsatz zu bringen. Genutzt werden sollte dazu (sofern möglich und sinnvoll) die beim Projektpartner SimPlan vorliegende Software SimAssist. SimAssist ist ein Werkzeug zur Unterstützung von Simulationsprojekten, für das die Grundlage in zwei vom Land Hessen geförderten Forschungsprojekten AssistSim (HA-Projekt-Nr. 185/09-15) und EDASim (HA-Projekt-Nr. 260/11-06) gelegt worden ist. Mit der Software können Eingangs- und Ergebnisdaten verwaltet sowie statistisch und graphisch aufbereitet worden. Eine umfassende Unterstützung von Simulationsexperimenten war zu Projektbeginn nicht Bestandteil der Anwendung. Die Idee der Projektpartner bei der Antragsstellung war daher, SimAssist zu erweitern und für das Ziel GZ4, die Konzeption und prototypische Umsetzung einer integrierten Plattform, in Betracht zu ziehen. Ein weiteres bei SimPlan vorliegendes Schnittstellentool, der SimController, erlaubt eine flexible Verknüpfung zwischen (verschiedenen) Simulationswerkzeugen und Optimierungs- oder Datenanalysesoftwarepaketen. Unter anderem stellt der SimController Dienste zum entfernten Starten, Kontrollieren und Beenden von Simulationsläufen zur Verfügung, so dass er sich potenziell für die Steuerung verteilter Simulationsläufe eignet. Mithin war der

SimController eine weitere Softwarekomponente, die (mit ggf. erforderlichen Anpassungen und Erweiterungen) für die Verwendung in DaWiS in Betracht gezogen wurde.

Neben diesen technischen und methodischen Vorarbeiten kam hinzu, dass bei den beiden Anwendungspartnern des Projekts, der Maschinenfabrik Reinhausen und der C-P-S Holding GmbH & Co. KG (im Folgenden kurz C-P-S), sehr konkrete Ideen für und großes Interesse an Fallstudien vorlagen, die für eine Erprobung der Wissensentdeckung mit Hilfe von Simulation gut geeignet erschienen.

3 Planung und Ablauf des Vorhabens

Die ursprüngliche Projektplanung umfasste sieben Arbeitspakete AP0 – AP6 und ist im Überblick in der folgenden Abbildung dargestellt.

Die fachlichen Arbeitspakete AP1 – AP4 waren an den Projektzielen GZ1 – GZ4 ausgerichtet: Im ersten Arbeitspaket AP1 lag der Fokus auf dem smarten Experimentdesign, im zweiten Arbeitspaket AP2 auf der Verteilung der Experimente und im dritten Arbeitspaket AP3 auf Data Mining- und Visual Analytics Methoden. Die Strukturierung dieser drei Arbeitspakete war ähnlich: auf eine Aufnahme von Anforderungen folgten eine Analyse des Stands der Technik und anschließend eine prototypische Umsetzung und ein Test. Die Arbeiten in Arbeitspaket AP4 an der Integration und den Benutzungsschnittstellen zielten auf die Konzeption und Umsetzung einer Softwareplattform gemäß Projektziel GZ4 ab. In Arbeitspaket AP5 wurde Dokumentationen als Grundlage für die Methodenanwendung durch Nicht-Experten erstellt. In dem Arbeitspaket AP6 wurden die Methoden und Dokumentationen an den Pilotanwendungen evaluiert (GZ5).

Begleitet werden sollte die Projektabarbeitung durch ein kontinuierliches Projektmanagement (AP0). Die geplanten Meilensteine haben sich am sowohl am Gesamtzeitraum des Projektes orientiert als auch daran, dass nach dem ersten Projektjahr die methodischen Arbeitspakete AP1-AP3 mit einer Validierung abgeschlossen sein würden.

Projekttitel: Entwicklung einer integrierten Lösung für das Data Farming und die Wissensentdeckung in Simulationsdaten

Projektstart: 01.01.2020

Nr.	Arbeitspakte / Meilenstein	Start / Datum	Dauer in Wochen	Aufwand in Personenmonaten
0	**Projektmanagement und Vermarktung**	**01.01.2020**		
0.1	Projektmanagement	01.01.2020	104	SIM: 2
0.2	Öffentlichkeitsarbeit/wissenschaftliche Publikation	01.03.2020	96	SIM: 0,5; ITPL 0,5
0.3	Verwertungsplan und geistige Eigentumsrechte	01.03.2020	96	SIM: 0,5; ITPL 0,5
M0.0	Meilenstein Kick Off des Projektes	07.01.2020	0	Alle
M0.1	Meilenstein Zwischenbericht inklusive Präsentation	17.02.2021	0	Alle
M0.2	Meilenstein Endbericht inklusive Präsentation Projektträger	15.12.2021	0	Alle
1	**Methoden für smartes Experimentdesign**	**08.01.2020**		
1.1	Analyse der Anforderungen	08.01.2020	8	SIM: 0,5; ITPL :1; C-P-S: 0,5; MR: 0,5
1.2	Auswahl bzw. Konzeption geeigneter Experimentdesignmethoden	02.03.2020	16	ITPL: 2
1.3	Implementierung der Methoden	30.03.2020	16	SIM: 3
1.4	Validierung und Tests	20.07.2020	6	SIM: 0,5; ITPL: 1
2	**Experimentverteilung**	**04.03.2020**		
2.1	Analyse der Anforderungen	04.03.2020	8	SIM: 0,5; ITPL :1; C-P-S: 0,5; MR: 0,5
2.2	Analyse bestehender Techniken der Verteilung & Virtualisieru	29.04.2020	12	ITPL: 2
2.3	Definition der Datenstrukturen	22.07.2020	8	ITPL: 1,5; SIM: 1
2.4	(Cloud) Infrastruktur, Einbindung	16.09.2020	14	SIM :1,5
2.5	Validierung und Test	23.12.2020	6	SIM: 0,5; ITPL: 1
3	**Data-Mining- und Visual-Analytics-Methoden**	**10.02.2020**		
3.1	Analyse der Anforderungen	10.02.2020	8	SIM: 0,5; ITPL :1; C-P-S: 0,5; MR: 0,5
3.2	Analyse der Methoden zum Data Mining	06.04.2020	12	ITPL: 2,5
3.3	Analyse der Visual-Analytics-Methoden	01.06.2020	12	ITPL: 2,5
3.4	Implementierung	24.08.2020	18	SIM: 3
3.5	Validierung und Test	28.12.2020	6	SIM: 0,5; ITPL: 1
4	**Integration und Nutzerinterfaces**	**01.10.2020**		
4.1	Integration der Methoden	01.10.2020	12	SIM: 1,5
4.2	Gestaltung der Nutzerinterfaces	24.12.2020	12	SIM: 1,5
4.3	Tests	18.03.2021	6	SIM: 0,5; ITPL: 0,5
5	**Dokumentation und Erstellung der Howtos**	**18.03.2021**		
5.1	Erstellen der Dokumentation bzw. des Handbuchs	18.03.2021	20	SIM: 1; ITPL: 1
5.2	Erstellen der Howtos insbesondere zum Anpassen der Modell	29.04.2021	10	SIM: 1; ITPL: 1
6	**Validierung in der Pilotanwendung**	**08.01.2020**		
6.1	Erstellung/Validierung Simulationsmodell des Pilotszenarios	22.03.2021	20	SIM: 2; C-P-S: 1,5; MR: 1,5
6.2	Pilotanwendung	09.08.2021	19	SIM: 2; ITPL: 4; C-P-S: 3; MR: 3

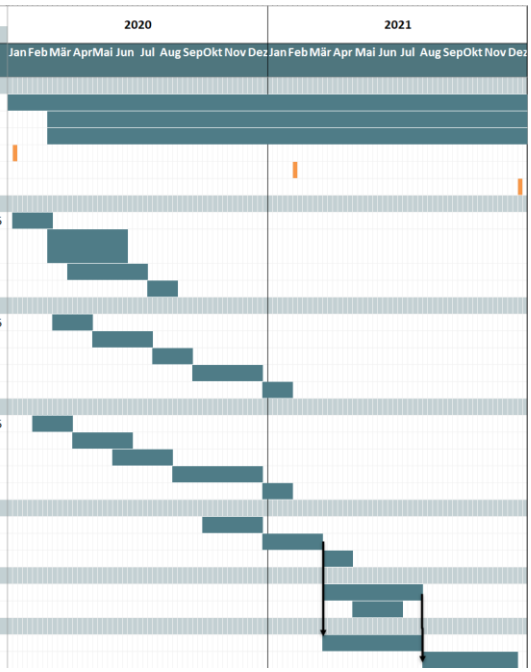

Abbildung 2: Projektplan und Meilensteine – Planung und erreichter Status

Tatsächlich hat es gegenüber diesem geplanten Ablauf lediglich eine nennenswerte Abweichung gegeben: Die Bearbeitung von Arbeitspaket 6 und hier insbesondere von Teilaufgabe 6.1 „Erstellung/Validierung Simulationsmodell des Pilotszenarios“ wurde in das erste Projektjahr vorgezogen. Der Grund war, dass sich bei der Aufnahme der Anforderungen in den Arbeitspaketen AP1-AP3 schnell gezeigt hat, dass es allen Projektbeteiligten deutlich leichter fällt, Anforderungen zu benennen und zu spezifizieren, wenn das entlang konkreter Anwendungsfälle geschieht. Im Gegenzug wurden die Arbeiten an Arbeitspaket AP5 in das letzte Projekthalbjahr verschoben.

Organisatorisch hat die Corona-Pandemie die Projektarbeit insofern beeinflusst, als das mit wenigen Ausnahmen ab Mitte März 2020 bis zum Projektende im Dezember 2021 so gut wie keine persönlichen Treffen mehr möglich waren. Das galt insbesondere für größere Runden wie etwa die Konsortial- und Meilensteintreffen. Erfreulicherweise verlief die Umstellung auf Online-Konferenzen unter Verwendung der einschlägigen Softwaretools völlig reibungslos. Auch kam es zu keinen in der inhaltlichen Arbeit des Projektes spürbaren wirtschaftlichen Beeinträchtigungen von Projektpartnern. Insgesamt ist unter den gegebenen Umständen die Zusammenarbeit im Konsortium über die gesamte Projektdauer hinweg reibungslos verlaufen.

4 Angeknüpfter wissenschaftlicher und technischer Stand

Methodisch wurde auf Grundlagen zu Experimentdesign, Experimentverteilung, Data Mining und Visual Analytics zurückgegriffen, die der TU Ilmenau aus Vorarbeiten bekannt waren (vgl. Abschnitt 2) bzw. die den im Folgendem angegebenen Quellen entstammen. Softwaretechnisch wurde eine Reihe vorliegender Komponenten, Bibliotheken und Applikationen miteinander verknüpft und in geeigneter Form erweitert und ergänzt.

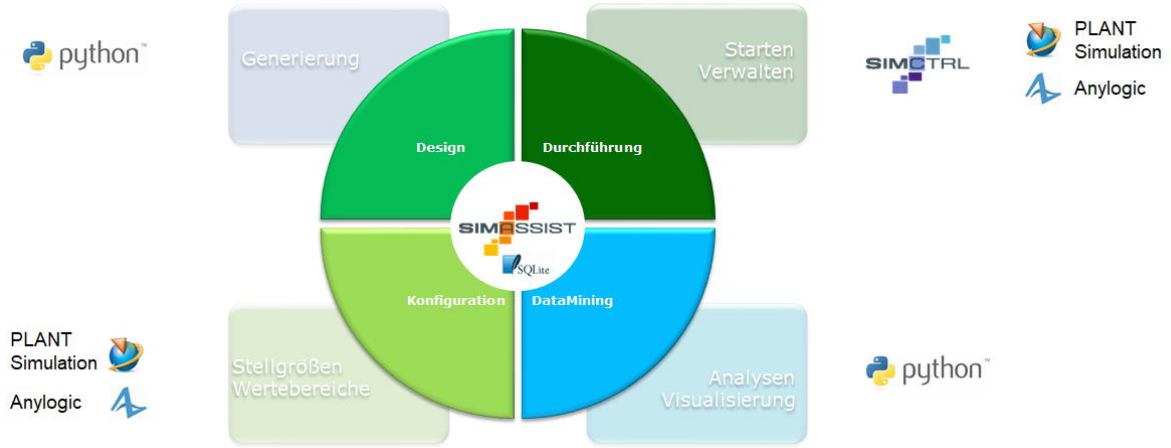

Abbildung 3: Überblick über die Komponenten des Softwareprototypen

Abbildung 3 zeigt im Überblick, dass für das Experimentdesign ebenso wie für Data Mining und Visual Analytics in Python implementierte Verfahren zum Einsatz gekommen sind. Dafür wurde auf geeignete Open-Source-Bibliotheken und eigene Entwicklungen zurückgegriffen. Die Experimentverteilung hat die Software SimController (mit einer Reihe geeigneter Erweiterungen) übernommen. Als Simulationswerkzeuge zur tatsächlichen Durchführung der Experimente nach Einstellung der entsprechenden Stellgrößen sind im Vorhaben Plant Simulation von Siemens und Anylogic von der gleichnamigen Firma verwendet worden. Allerdings kann hier mit überschaubarem Aufwand auch jedes andere Simulationswerkzeug eingebunden werden. Zur Haltung von Massendaten wurde eine SQLite-Datenbank verwendet, und als Rahmenapplikation für den gesamten Data Farming Ablauf von Design über Durchführung und Analyse (Data Mining) kam die Software SimAssist zum Einsatz. Details zu den Ergebnissen finden sich in Kapitel 6.

Die folgende Auflistung gibt einen Überblick über wesentliche Teile der verwendeten Fachliteratur, die sich sowohl auf Grundlagen des Data Farmings und Visual Analytics bezieht als auch auf Erweiterungen für den Bereich Simulation in Produktion und Logistik, die speziell

auf Vorarbeiten an der TU Ilmenau zurückzuführen sind. Daneben wurden Grundlagenquellen aus dem Bereich der Simulation und Digitalen Fabrik im Allgemeinen verwendet.

Fayyad, U.M.; Piatetsky-Shapiro, G.; und Smyth, P. (1996): From Data Mining to Knowledge Discovery in Databases. In: AI Magazine 17, S. 37–54.

Feldkamp, N.; Bergmann, S. und Strassburger, S. (2015): Knowledge Discovery in Manufacturing Simulations. In: Taylor, S.J.E; Mustafee, N. und Son, Y.-J. (Hrsg.): Proceedings of the 3rd ACM SIGSIM Conference on Principles of Advanced Discrete Simulation. ACM, New York, S. 3–12.

Feldkamp, N.; Bergmann, S.; Strassburger, S.; Sokoll, K. und Clausing, M. (2019): Knowledge Discovery in Simulation Data: An Industrial Case Study on Intralogistics in an Automotive Production. In: Haas, P.J.; Mustafee, N.; Rabe, M. und Bae, K.-H.G. (Hrsg.): Proceedings of the 2019 Winter Simulation Conference. IEEE, Piscataway.

Horne, G.; Åkesson, B.; Meyer, T. und Anderson, S. (2014): Data farming in support of NATO. Final Report of Task Group MSG-088. Neuilly-sur-Seine Cedex: North Atlantic Treaty Organisation (STO technical report, TR-MSG-088).

Keim, D.A.; Mansmann, F.; Schneidewind, J.; Thomas, J. und Ziegler, H. (2008): Visual Analytics: Scope and Challenges. In: Simoff, S.J.; Boehlen, M.H. und Mazeika, A. (Hrsg.): Visual Data Mining: Theory, Techniques and Tools for Visual Analytics. Springer, Berlin und Heidelberg, S. 76-90.

Krückhans, B.; Meier, H. (2013): Industrie 4.0 – Handlungsfelder der Digitalen Fabrik zur Optimierung der Ressourceneffizienz in der Produktion. In: Dangelmaier, W.; Laroque, C. und Klaas, A. (Hrsg.): Proceedings der 15. ASIM Fachtagung Simulation in Produktion und Logistik 2013. Entscheidungsunterstützung von der Planung bis zur Steuerung. HNI-Verlagsschriftenreihe, Paderborn, S. 31–40.

Law, A.M. (2003): How to conduct a successful simulation study. In: Chick, S.; Sanchez, P.J.; Ferrin, D. und Morrice, D.J. (Hrsg.): Proceedings of the 2003 Winter Simulation Conference. IEEE, Piscataway, S. 66–70.

Sanchez, S. M. (2014): Simulation Experiments: Better Data, Not Just Big Data. In: Tolk, A.; Diallo, S.D.; Ryzhov, I.O.; Yilmaz, L.; Buckley S. und Miller J.A. (Hrsg.): Proceedings of the 2014 Winter Simulation Conference. IEEE, Piscataway, S. 805–816.

Sargent, R.G. (2013): Verification and validation of simulation models. In: Journal of Simulation 7 (1), S. 12–24.

5 Zusammenarbeit mit anderen Stellen

Die Zusammensetzung des Konsortiums mit einem technischen Partner (SimPlan), einem Anwendungspartner (C-P-S), einem assoziierten Anwendungspartner (Maschinenfabrik Reinhausen) und der wissenschaftlichen Unterstützung durch das Fachgebiet Informationstechnik in Produktion und Logistik TU Ilmenau, hat sich als sehr produktiv herausgestellt. Einerseits waren die Verantwortlichkeiten und Kompetenzen klar voneinander abgegrenzt, andererseits war alles Wissen im Projekt vorhanden, um zielstrebig Ergebnisse

erarbeiten zu können. Dabei entsprach die gewählte Arbeitsteilung dem Kompetenzschwerpunkt der einzelnen Partner: Wissenschaftliche Fundierung mit entsprechenden Grundlagen aus der Forschung sowohl hinsichtlich Data Farming und Methoden des Data Mining als auch in Bezug auf Simulation durch die TU Ilmenau, Anwendungs- und Produktions-Know-How von C-P-S und der Maschinenfabrik Reinhauen sowie Softwareentwicklung und Simulation von SimPlan. Dabei hat die Zusammenarbeit ausgezeichnet funktioniert; jeder Partner hat sich stets intensiv eingebracht und für den Projektfortschritt engagiert.

6 Verwendung der Zuwendung und des erzielten Ergebnisses mit Gegenüberstellung der vorgegebenen Ziele

Dieser Abschnitt greift die für die TU Ilmenau relevanten und in Abschnitt 1 zusammengestellten Ziele gemäß Vorhabenbeschreibung auf, gibt sie jeweils kurz inhaltlich wieder und enthält eine Einschätzung des erreichten Stands.

6.1 GZ1 – Methoden des smarten Experimentdesigns evaluieren und implementieren

Gemäß Vorhabenbeschreibung waren Anforderungen an das Experimentdesign zusammenzutragen und mit dem aktuellen Stand der Forschung abzugleichen. Im Anschluss sollten ausgewählte Methoden des Experimentdesigns in einen Softwareprototypen überführt und getestet werden. Als eine der Hauptaufgaben dieses Teilziels hatte die TU Ilmenau das Identifizieren von geeigneten Experimentdesignmethoden. Dabei wurden die folgenden Methoden wie die vollständige Enumeration („n^k“), das Latin Hypercube Sampling („LHS“), das Nearly Orthogonal Latin Hypercube Sampling („NOLHS“) und das Central Composite Design zusammengetragen. Weiterhin war eine wichtige Aufgabe der TU Ilmenau im Zusammenhang mit diesem Ziel insbesondere die Entwicklung der Python-Skripte zur Generierung der ausgewählte Experimentdesignmethoden.

Als Unterstützung für den Anwender, der ein geeignetes Verfahren für das Experimentdesign im konkreten Anwendungsfall auswählen möchte, wurde eine Handreichung in Form eines Entscheidungsbaumes entwickelt, der in Abbildung 4 dargestellt ist. Ein gewisses Grundverständnis über die Ansätze für die Experimentplanung ist erforderlich, aber durch das zu Verfügung stellen von ausführlichen Hilfefunktionen (vgl. Abschnitt 6.6) kann der Anwender die Funktionsweise der Experimentdesignmethoden immer nachlesen.

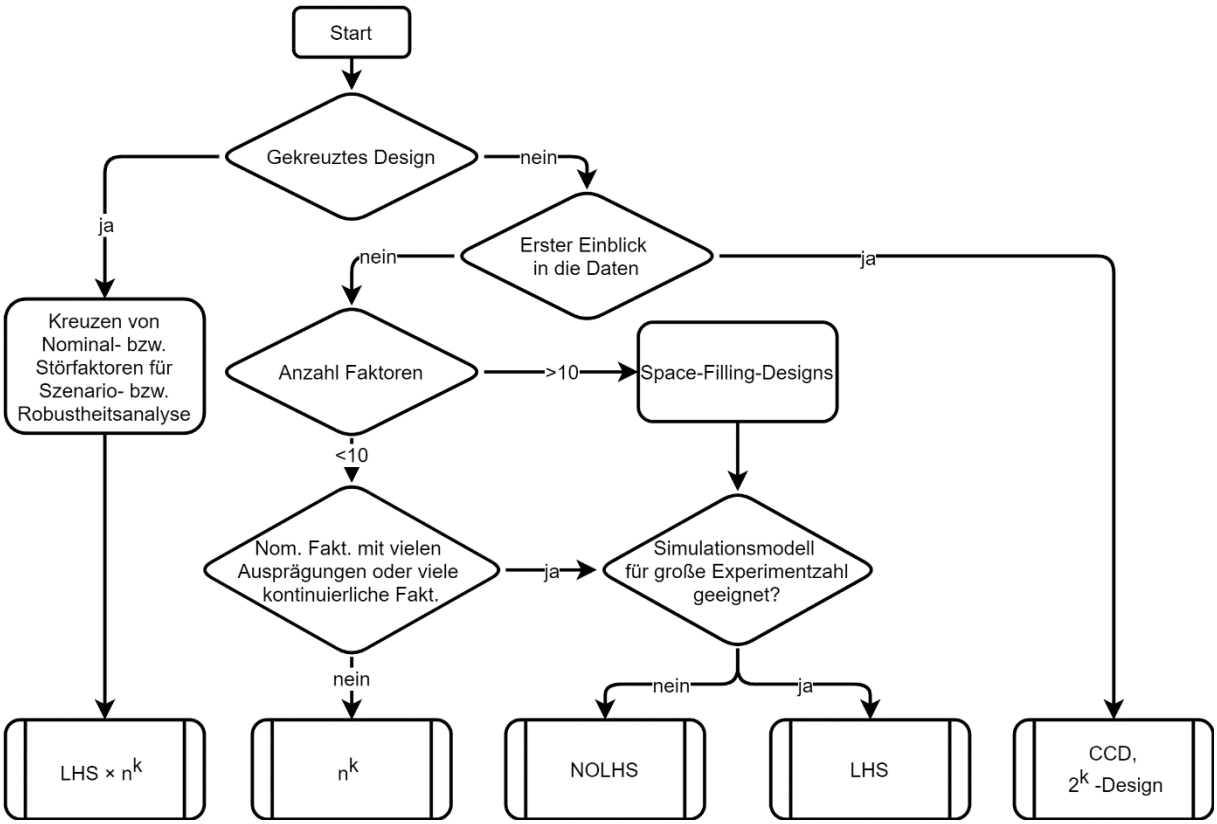

Abbildung 4: Entscheidungsbaum zur Auswahl geeigneter Experimentdesignmethoden.

6.2 GZ2 – Methoden zur (cloudbasierten) Verteilung von Experimenten entwickeln

Zur Erreichung dieses Ziels sollte gemäß Vorhabenbeschreibung sowohl eine Verteilung von Simulationsexperimenten in lokalen Rechnernetzen (Private Clouds) als auch in cloudbasierten Infrastrukturen ermöglicht werden. Ferner sollte der Verteilungsmechanismus unabhängig von Simulationswerkzeugen und natürlich auch von Simulationsmodellen sein. Im Einzelnen haben die TU Ilmenau und SimPlan unterstützt durch die Anwendungspartner die folgende Liste von Anforderungen für die Experimentverteilung zusammengestellt:

- Simulationsläufe sollen auf mehrere Rechner im Netzwerk verteilbar sein
- Mehrere Simulationsläufe auf einem Rechner sollen möglich sein
- Die komplette Kontrolle über den Simulator muss gegeben sein (Starten, Stoppen)
- Es muss einen kontinuierlichen Statusaustausch zwischen „Experimentverteiler“ und Simulation geben
- Es muss mindestens die Interaktionsmöglichkeiten Abbruch und Neustart geben
- Die verwendete Technologie sollte möglichst allgemeingültig sein (keine/wenig Abhängigkeiten)
- Kommunikation und Kommunikationsinhalte müssen getrennt sein
- Das Einrichten der Infrastruktur muss einfach sein

Im Rahmen der Realisierung der Verteilung und Virtualisierung hat die SimPlan AG die bereits vorhandene Schnittstellensoftware SimController so erweitert, dass eine große Anzahl von Experimenten verteilt und gehandhabt werden kann.

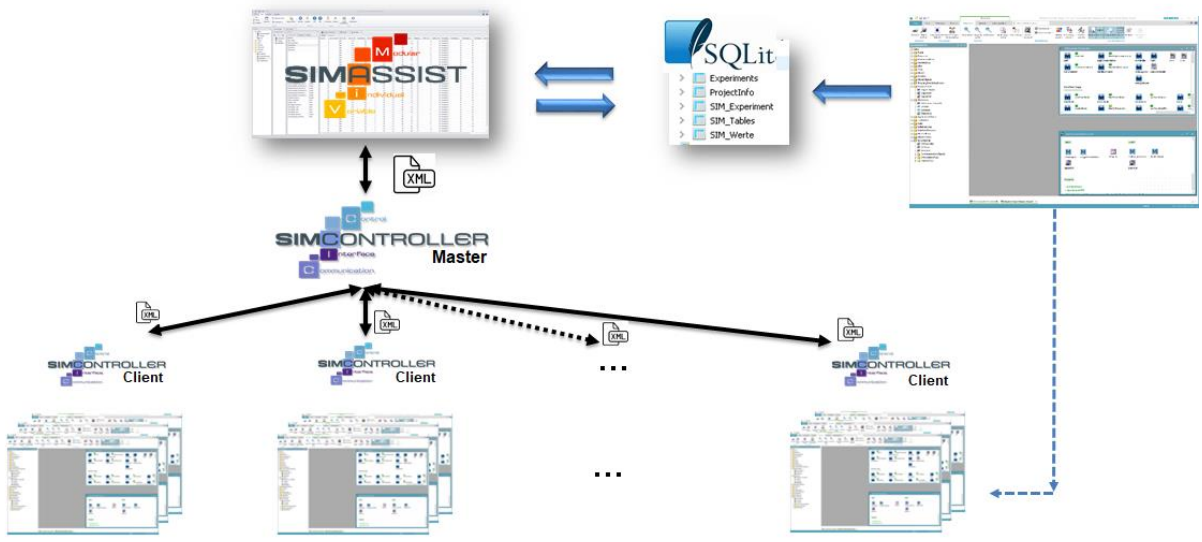

Abbildung 5: Schema der Experimentverteilung unter Verwendung des SimControllers.

Abbildung 5 zeigt den schematischen Aufbau des Zusammenspiels der Softwarekomponenten bei der Experimentverteilung. Die ausgetauschten XML-Dateien enthalten die Eingangsparameter für den jeweiligen Simulationslauf. Ergebnisdaten werden dezentral in SQLite Datenbanken geschrieben und am Ende eines Laufs wieder in die zentrale Instanz übermittelt. Dabei ist die in Abbildung 5 dargestellte Architektur unabhängig von der eingesetzten Simulationssoftware. Die Abbildung enthält exemplarisch die Darstellung für Plant Simulation aber für andere Werkzeuge wie etwa Anylogic sind der Ablauf und das Zusammenspiel der Komponenten identisch. Eine wesentliche Aufgabe der TU Ilmenau war hier das Testen der Architektur mit für das Data Farming typischen Experimentplänen, um Limitationen aufzudecken. Dadurch konnte im Zusammenspiel mit der SimPlan AG abschließend eine Architektur entwickelt werden, dessen Funktionsfähigkeit auch für sehr große Experimentpläne gegeben ist. Diese Funktionsfähigkeit konnte insbesondere am Anwendungsbeispiel der Maschinenfabrik Reinhausen (vgl. Abschnitt 6.5) eindrucksvoll nachgewiesen werden. Die TU Ilmenau hat unter Verwendung einer von SimPlan bereitgestellten Cloud-Infrastruktur mit vier Simulator-Instanzen an rund 4 Tagen mehr als 60.000 Experimente erfolgreich durchgeführt.

6.3 GZ3 – Data-Mining- und Visual-Analytics-Methoden vorsehen und evaluieren

Im Rahmen dieses Teilziels waren Data-Mining- und Visual-Analytics-Methoden zu evaluieren und prototypisch zu implementieren. Die Evaluierung haben die Partner unter Federführung der TU Ilmenau im Hinblick auf die Eignung der Methoden im Zusammenspiel mit Data Farming Ansätzen und im Hinblick auf die Eignung für den Einsatz im Bereich Produktion und Logistik vorgenommen.

	Untersuchung der Beziehung zw. Eingangs- und Ergebnisdaten																		
	Mustererkennung und Klassenbildung																		
	Charakterisierung von Ergebnisdaten									Visualisierungen									
	Datenvorverarbeitung																		
	Input	Parameter	Verfahren / Methode	Variablenskalierung					Priorität	Boxplot	Histogramm	Kuchen-diagramm	Mosaikplot	Balken-diagramm	Scatterplot	SPLOM	Parallel-koordinanten	Spinnennetz-diagramm	Heatmap	Graphen
			Vorverarbeitung																	
Dimensionsreduktion	Kategorische Eingangsparameter		One-Hot-Encoding	kategorisch zu binär	✓				Notwendig											
			Statistische Analyse für große Datenmengen																	
Statistische Analyse	Ergebnisdaten		Lagemaße und Verteilungen	alle		✓	(✓)		Notwendig	✓	✓									
	Ergebnisdaten		Streuungsmaße	alle		✓			Optional	✓	✓									
	Eingangs - und Ergebnisdaten		Korrelationsanalyse	metrisch		✓		✓	Notwendig										✓	
	Ergebnisdaten		Robustheitsmaße	metrisch		✓			Nicht Notwendig										✓	
			Charakterisierung und Konzeptualisierung																	
	Ergebnisdaten	Anzahl der Bins	Histogrammanalyse inklusive Binning	metrisch			✓		Notwendig		✓			✓						
	Eingangs-, Ergebnis- und Klassendaten	Klassenauswahl	Klassenvergleich und Diskriminierung	min. eine nominale Dimension				✓	Notwendig	✓	✓	✓	✓	✓			✓	✓		
			Assoziationsanalyse																	
Algorithmische Analyseverfahren	Eingangs - und Ergebnisdaten		Frequent Pattern Mining	nominal			✓	✓	Nicht Notwendig											✓
			Unüberwachtes Lernen																	
	Ergebnisdaten	Ergebnisdatenauswahl und Klassenanzahl	Kmeans	metrisch			✓		Notwendig		✓									
	Ergebnisdaten	Ergebnisdatenauswahl und Klassenanzahl	Gaussian Mixture Models	metrisch			✓		Notwendig		✓									
	Ergebnisdaten		Hauptkomponentenanalyse	metrisch			✓		Nicht Notwendig						✓					
			Überwachtes Lernen																	
Modelbildende Verfahren	Eingangs - und Ergebnisdaten	Filterfaktor	Regressionsanalyse	metrisch auf metrisch				✓	Notwendig						✓					
	Eingangs - und Ergebnisdaten		Logistische Regression	metrisch auf nominal				✓	Optional						✓					
	Eingangs - und Klassendaten	Pruning	Klassifikationsbäume	gemischt auf nominal				✓	Notwendig											✓
			Weitere Verfahren																	
Visual Analytics	Eingangs-, Ergebnis- und Klassendaten		Mehrdimensionale Mustererkennung	alle			✓	✓	Notwendig						✓	✓	✓		✓	✓
Beschreibung der Interaktionen									Interaktivität	Nicht Notwendig	Notwendig	Nicht Notwendig	Nicht Notwendig	Notwendig	Notwendig	Optional	Notwendig	Nicht Notwendig	Nicht Notwendig	Optional
Selektieren	Navigieren	Arrangieren	Wechseln	Filtern					Interaktionen		Filtern, Wechseln			Filtern	Filtern, Navigieren, Selektieren	Filtern, Selektieren	Arrangieren, Filtern, Navigieren, Selektieren			Selektieren
hervorheben, markieren, betonen	fokussieren, zoomen, durchblättern	sortieren, einstufen, indexieren	von Parametern, Darstellungen	Werte, Parameter					Priorität	Nicht Notwendig	Notwendig	Optional	Nicht Notwendig	Notwendig	Notwendig	Optional	Notwendig	Optional	Notwendig	Notwendig

Abbildung 6: Systematische Einordnung, Verknüpfung und Priorisierung von Data-Mining- und Visual-Analytics-Methoden.

Die Ergebnisse der Evaluation sind in verschiedenen Darstellungen zusammengefasst, wie eine in Abbildung 6 dargestellt ist. Auf dieser Basis erfolgte die Priorisierung durch die TU Ilmenau für die Entwicklung und Implementierung der Methoden.

Bei der Umsetzung der Verfahren wurde wieder viel Wert auf Flexibilität, Wartbarkeit und Erweiterbarkeit gelegt. Wie Abbildung 7 schematisch zeigt, werden die Simulationsergebnisse zusammen mit ggf. weiteren Parametern für das jeweilige Data-Mining-Verfahren aus der SimAssist-Rahmenapplikation an die eigentlichen Data-Mining-Algorithmen übergeben. Die Algorithmen selbst müssen dann wieder in Python-Code vorliegen, werden ausgeführt und geben die Ergebnisse dann wieder an die Rahmenapplikation zur Visualisierung zurück. Dieses Vorgehen ermöglich ähnlich wie bei den Methoden für das Experimentdesign eine einfache Erweiterbarkeit durch das Einbinden weiterer Data-Mining- oder Visualisierungsverfahren, die

als Python-Code mit entsprechenden Schnittstellen vorliegen müssen. Die TU Ilmenau hat sich vor allem auf die Entwicklung der Data-Mining- und Visualisierungsverfahren auf der Python-Seite konzentriert.

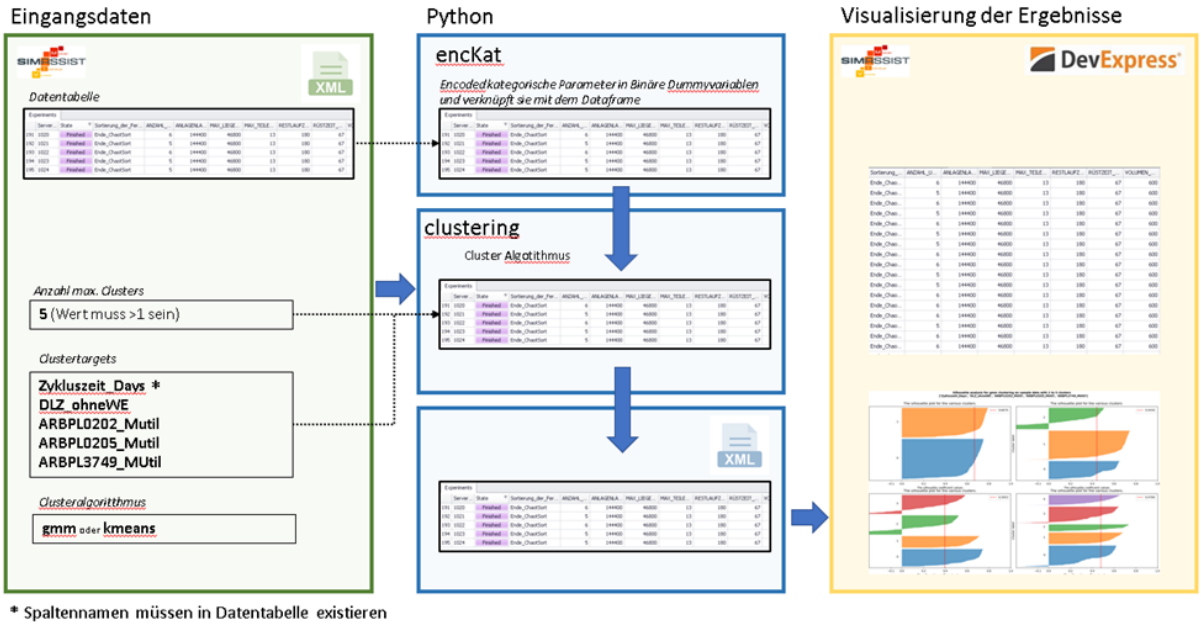

Abbildung 7: Schematischer Datenfluss bei der Einbindung von Data-Mining- und Visualisierungsmethoden.

6.4 GZ4 – Integration und Umsetzung einer integrierten Plattform

Aufgabe der integrierten Plattform ist es, den Ablauf von dem Experimentdesign über die Experimentverteilung bis zur Auswertung durch Data Mining Verfahren und zur Ergebnisvisualisierung durchgängig zu unterstützen. Das Projektteam hat sich entschieden, dazu neue Module (ein neue sogenanntes „Plug-In“ mit dem Namen „4farm“) in der offenen und erweiterbaren Plattform SimAssist vorzusehen. Dieses Plug-In bildet den gesamten in Abbildung 3 dargestellten Ablauf ab. In der Abbildung ist zu erkennen, dass neben Design, Verteilung und Auswertung auch noch eine Konfiguration erforderlich ist. Für die Konfiguration hat die SimPlan AG simulatorspezifische Bausteine entwickelt, welche das Spezifizieren der Stellgrößen und Ergebnisdaten sowie das Verknüpfen des Simulationsmodells mit der Projektdatenbank übernehmen.

Ebenfalls Bestandteil der Konfiguration und unmittelbare Vorbereitung des Experimentdesigns ist die Festlegung der beim Design zu berücksichtigenden Faktoren, Faktorgrenzen und ggf. der Schrittweiten (vgl. Abbildung 8). Dieser Schritt erfolgt in der Softwareplattform und nicht mehr in den jeweiligen Simulationssoftwarepaketen.

Berücksichtigung finden die Eingabegrößen (Stellgrößen), die zuvor im Simulationsmodell parametrisiert worden sind.

Design | Experiments

Kategorie

Name	Typ	Subkategorie	Default	Aktiv	Werte	...	...	I...
Einlastung								
PAR_Sortierung_der_Fertigungsaufträge	Enum	Sortierung	Ende_ChaotSort	☑	Ende_ChaotSort;AUFNR;Auftragsmenge [ST];Spätestes_E...	...	1	
Mitarbeiter								
PAR_ANZ_MA10_VERFÜGBAR	Integer	Verfügbarkeit	4	☑	Min: 1; Max:4; Step: 1	...	1	
PAR_MA11_2_VERFÜGBAR	Boolean	Verfügbarkeit	True	☑	True;False	...	1	
PAR_MA2_VERFÜGBAR	Boolean	Verfügbarkeit	True	☑	True;False	...	1	
PAR_MA9_VERFÜGBAR	Boolean	Verfügbarkeit	True	☑	True;False	...	1	
Puffer								
PAR_ANZAHL_UNVOLLSTÄNDIG_PUFFER_FÜGEN	Integer	Anzahl	5	☑	Min: 2; Max:10; Step: 1	...	1	
Schliessanlage								
PAR_RÜSTEN_SCHLIESSANLAGE_AKTIV	Boolean	Rüsten	True	☐	True	...	1	
Waschen								
PAR_MAX_TEILE_WASCHKORB	Real	Anzahl	13	☑	Min: 6; Max:24; Step: 1	...	1	
PAR_RESTRIKTIONEN_FÜGEPLATZ_AN	Boolean	Restriktionen	True	☐	True	...	1	
PAR_VOLUMEN_WASCHKORB	Real	Volumen	600	☑	Min: 200; Max:1000; Step: 100	...	1	
PAR_MAX_LIEGEZEIT_NACH_WASCHEN	Real	Zeit	46800	☑	Min: 14400; Max:57600; Step: 14400	...	1	
PAR_RESTLAUFZEIT_WASCHANLAGE	Real	Zeit	180	☑	Min: 120; Max:600; Step: 60	...	1	
PAR_RÜSTZEIT_WASCHKORB	Real	Zeit	67	☑	Min: 50; Max:120; Step: 5	...	1	

PAR_Sortierung_der_Fertigungsaufträge

Diskrete Werte:

- ☑ Wert
- ☑ Ende_ChaotSort
- ☑ AUFNR
- ☑ Auftragsmenge [ST]
- ☑ Spätestes_Ende
- ☑ Materialnummer_ISOLATOR
- ☑ ChaotSort

Abbildung 8: Festlegen der Faktoren und deren Wertebereiche.

Abbildung 9 zeigt die Einbindung der Experimentdesignmethoden in das 4farm-Modul. Der Anwender muss nur noch die Methode auswählen und die benötigten Parameter festlegen. Anschließend werden die ihm die generierten Parameterkombinationen in Tabellenform in der Softwarelösung angezeigt.

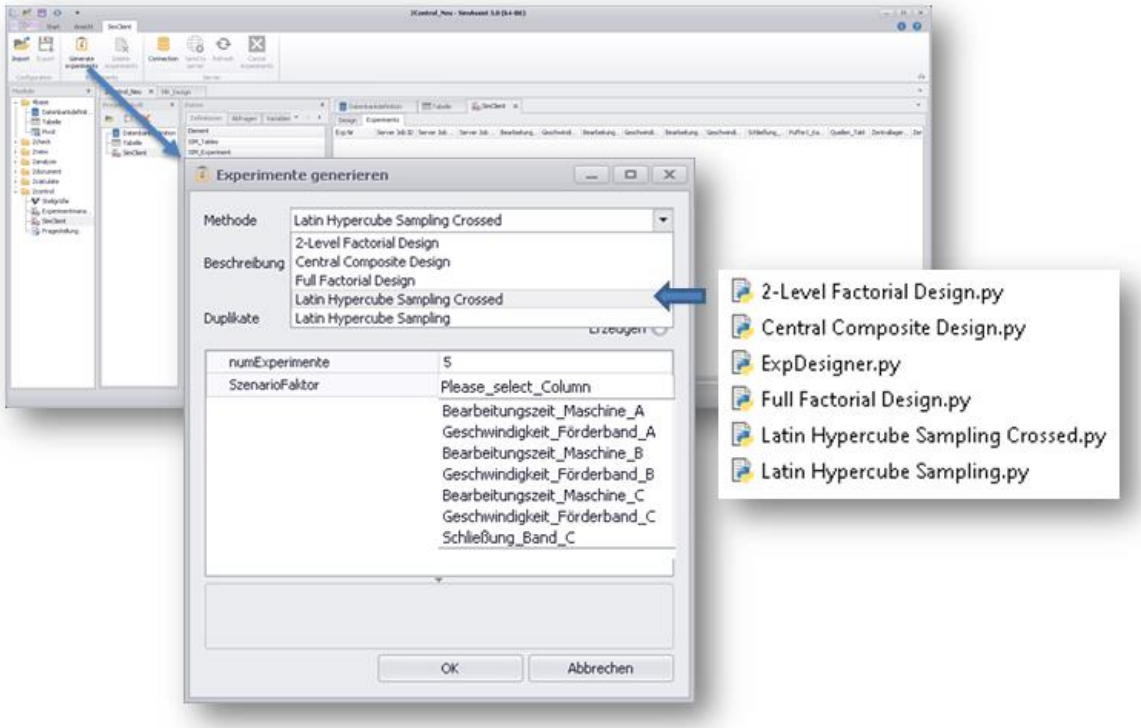

Abbildung 9: Auswahl der Experimentdesignmethode im 4farm-Modul.

Für die Experimentverteilung werden während jedes Simulationslaufs Statusinformationen ausgetauscht, die wie in Abbildung 10 beispielhaft dargestellt, zentral in der Softwarelösung visualisiert werden. Dadurch bekommt der Anwender noch während der Berechnung der Experimente erste Informationen über die Ergebnisdaten und mögliche Abbrüche in Läufen.

Design | Experiments

Exp Nr	Server J...	Server Job ...	Server Job Progress	KPI_DLZ_Truck A...	KPI_DLZ_Truck B...	KPI_Verbrauch_Pr...
51	1723	Finished	100 %	3721,18622324097	7862,89464324...	54264
50	1724	Started	71 %			
49	1725	Started	72 %			
48	1726	Finished	100 %	2931,70201137141	8109,91975358...	54264
47	1727	Started	0 %			

Abbildung 10: Verfolgung des Experimentfortschritts.

Die Einbindung der Data-Mining- und Visual-Analytics-Methoden erfolgt über einfache „Drag and Drop“- und „Dropdown“-Bausteine. Dadurch kann der Anwender einfach die gewünschten Daten und Parameter auswählen. In Abbildung 11 ist die Integration beispielhaft für das Clustering dargestellt. Bei den weiteren Methoden unterscheiden sich nur die methodenspezifischen Input- und Parameterfelder, welche mittels Metainformationen in den Skripten spezifiziert werden.

Name: Clustering
Category: KDS
Description: v.0.2
Python method: clustering.py

Input
default
- Stellgrößen [Numeric]
- Zielgrößen [Numeric]
- Kategorische Parameter [Any]
- Clustering-Zielgrößen [Any]
  - KPI_Auslastung_Produktion_Puffer_Leer [Double]
  - KPI_Auslastung_Stapler_5t_1 [Double]
  - KPI_Stillstand_Produktion_h [Double]
  - PAR_Szenario [Int64]
  - KPI_Auslastung_Produktion_Puffer_Voll [Double]
- Nicht zugewiesen

Parameter

Clusteranzahl	5
Clustering	kmeans
KategorischerParameter	

KategorischerParameter

OK | Cancel | Apply

Abbildung 11: Einbindung der Data-Mining- und Visual-Analytics-Methoden.

Im Rahmen des Vorhabens ist es gelungen, alle Bestandteile des in Abbildung 3 dargestellten Ablaufs prototypisch zu implementieren und in die Plattform SimAssist einzubinden.

6.5 GZ5 – Erprobung der entwickelten Lösung in Pilotanwendungen

Im DaWiS-Projekt wurden mit beiden Anwendungspartnern jeweils eine Pilotanwendung betrachtet. Zunächst wird hier der Anwendungsfall von C-P-S beschrieben bevor im Anschluss der Anwendungsfall der Maschinenfabrik Reinhausen diskutiert wird.

C-P-S – Versorgung einer Automobilproduktion mit Batterien

Der C-P-S Anwendungsfall umfasst die Anlieferung, die Zwischenpufferung und die sequenzgetreue Bereitstellung von zwei unterschiedlichen Batterietypen für den Einbau in der Produktion, sowie das Leergutrückführung.

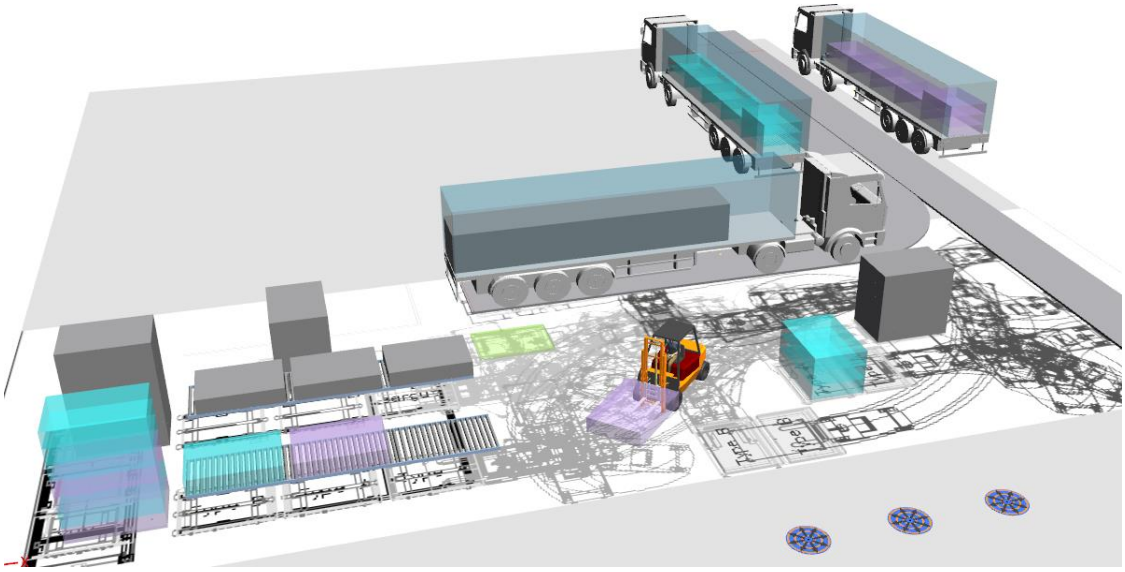

Abbildung 12: Visualisierung des C-P-S Anwendungsfalls

Im Screenshot der Simulatorvisualisierung (siehe Abbildung 12) sind die anliefernden LKW, ein Gabelstapler (je nach Szenario sind auch mehrere Stapler möglich) und (als Quader dargestellt) die Ladungsträger mit den Batterien zu erkennen. Folgende Leitfragen bildeten die Basis der Wissensentdeckung in Simulationsdaten:

- Unter welchen Rahmenbedingungen ist die gewählte Anzahl an eingesetzten Stapler ausreichend, um keinen Versorgungsabriss oder andere Rückstausituation zu verhindern?
- Welchen Einfluss hat der Staplertyp und die damit verbundenen Limitierungen auf die Systemleistung?
- Welche Fördertechnikkapazität ist ausreichend?
- Welchen Einfluss hat der Anlieferzyklus?
- Welchen Einfluss hat der prozentuelle Anteil der beiden Batterietypen?

Basierend auf den Fragestellungen wurde das Modell mit dem entsprechenden Detaillierungsgrad und den notwendigen Stellgrößen aus folgenden Gruppen versehen:

- Szenario: Insgesamt wurden drei Grundszenarien festgelegt, welche definieren wieviel Stapler von welchem Typ eingesetzt werden sollen.
- Produktionsvolumen: Die Stellgrößen dieser Gruppe legen das Produktionsvolumen und den Anteil des jeweiligen Batterietyps fest.
- Anlieferung: Neben dem Anlieferzyklus und der daraus resultierenden LKW-Beladung können auch Trailerwechselzeit und die Stapelgrößen verändert werden.
- Kapazitäten: Hierunter fallen die kapazitiven Beschränkungen der Fördertechnik für Voll- und Leergut

Auch die Kennzahlen lassen sich ähnlich wie die Stellgrößen in Gruppen unterteilen:

- Produktionsversorgung: Hierunter fallen sowohl die kumulierten Abrisse in der Produktionsversorgung, Füllstände der ver- und entsorgenden Fördertechnik als auch die Vorlaufzeiten.
- Stapler: Für jeden Stapler werden detaillierte Statistiken über Zustände, Mengengerüste und Auslastungen über den Simulationszeitraum mitgeführt.
- Anlieferung: Neben den Standzeiten und Beladezuständen werden auch die Verspätungen und die Rückstaus in der Anlieferung als Kennzahlen berechnet.
- Pufferflächen: Auch für die Batterietypen spezifischen Pufferplätze erfolgt eine Erfassung der wesentlichen Kennzahlen.

Mit dem Modell wurden durch das FG ITPL und vom Anwendungspartner C-P-S selbst zahlreiche Experimente durchgeführt und die generierten Daten entsprechend analysiert. Es wurde das gesamt verfügbare Portfolio an Data-Mining-Methoden und Visualisierungen erfolgreich zur Unterstützung der Fragestellungen eingesetzt. So wurden u.a. Cluster gebildet, welche es erlaubten, besonders gute Konfigurationen zu erkennen (siehe Abbildung 13) oder auch Entscheidungsbäumen genutzt welche das Ableiten von Regeln erlauben (siehe Abbildung 14). Schließlich konnte ein Satz von Managementempfehlungen erarbeitet werden.

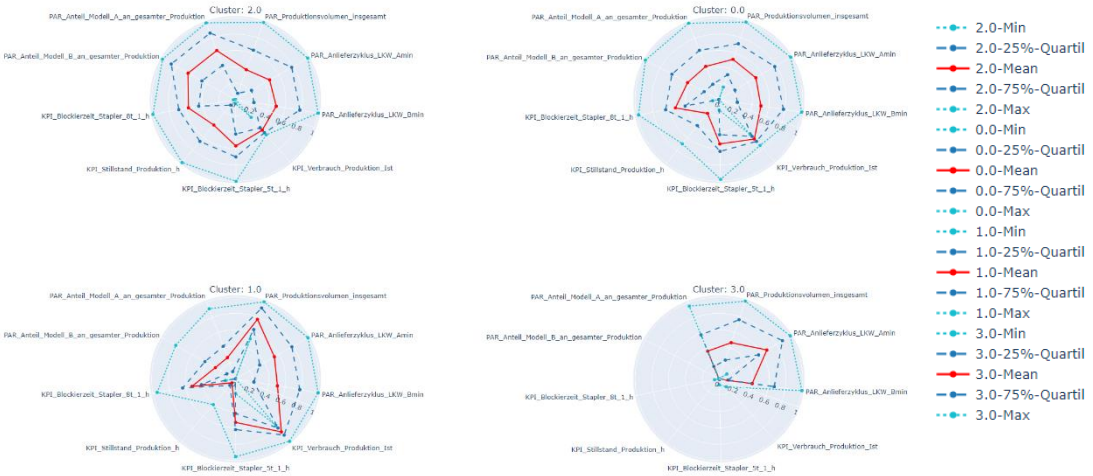

Abbildung 13: Visualisieren von KPIs und Faktoren mittels Radarplots für 4 erzeugte Cluster

- Beispiel für die Ableitung einer Regel

- Abgeleitete Regel:
  WENN
  "Prozentsatz des Batterietyps B" ≤ ***19,5***
  UND
  "Verbrauch der Produktion" > 37,5
  UND
  "Prozentsatz des Batterietyps A" ≤ 28,5
  DANN
  Cluster 1 (good performer)

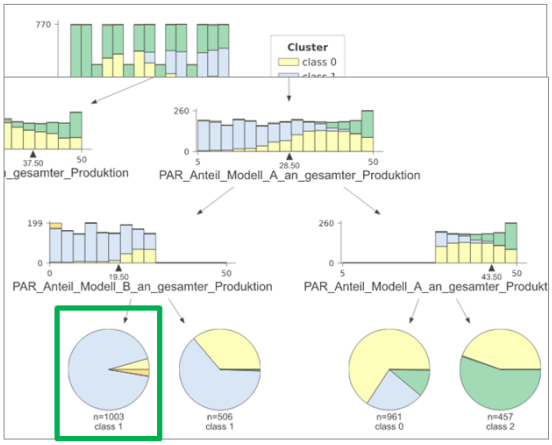

Abbildung 14: Beispiel für die Ableitung von Entscheidungsregeln im C-P-S Anwendungsfall

MR – Produktion RECOTEC

In Abbildung 15 ist einen Ausschnitt der Visualisierung des Modells der RECOTEC-Fertigung der Maschinenfabrik Reinhausen zu sehen. In dieser 2. Fallstudie wurde die Produktion unterschiedlicher Varianten von Isolatoren betrachtet. Hierbei durchlaufen teilweise unterschiedliche Varianten unterschiedliche Arbeitsabläufe bei denen zudem die Prozesszeiten und die Anlagenrüstzeiten sehr stark variieren.

Folgend Leitfragen sollten untersucht werden:

- Welchen Einfluss hat die Einlastungsreihenfolge der Produktionsaufträge?
- Wie wirken sich Limitierungen von Pufferkapazitäten zwischen den einzelnen Prozessschritten auf das System aus?
- Gibt es Abhängigkeiten, die die Leistungsfähigkeit des Systems beeinflussen?

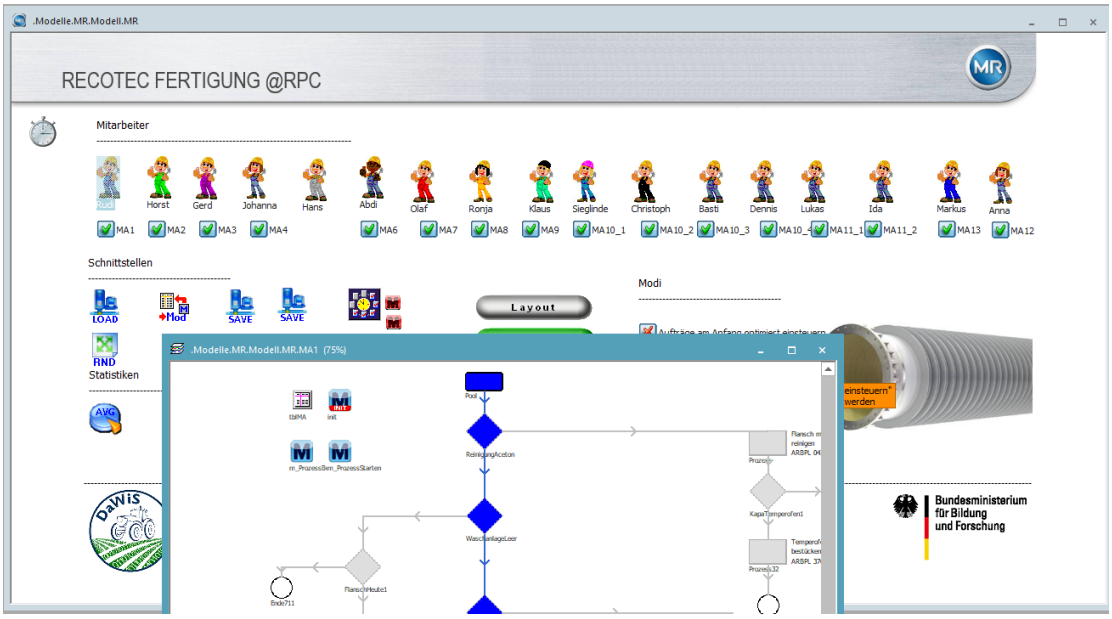

Abbildung 15: Ausschnitt aus dem Modell der RECOTEC-Fertigung – Anwendungsfall Maschinenfabrik Reinhausen

Im Rahmen der Erstellung des Modells wurden die einzelnen Prozessabläufe und Einflussfaktoren definiert und im Modell als Prozessablaufpläne gemäß vorgegebenem Regelwerk hinterlegt. Darin wurden auch die entsprechenden Stellgrößen und deren Einfluss auf die Prozessabläufe berücksichtigt. Auch hier lassen sich wieder entsprechende Stellgrößengruppen unterschieden:

- Produktionsaufträge: Hierunter fallen die Auswahl von Strategien zur Priorisierung und Gruppierung von Produktionsaufträgen bei der Einlastung der Komponenten in den ersten Prozessschritt.
- Kapazitäten: Es gibt eine Vielzahl von Anlagen und Prozessschritten, die aufgrund ihrer Kapazität den Prozessablauf beeinflussen.
- Flächen: Hierunter fällt die flächenmäßige Beschränkung von Puffern zwischen einzelnen Anlagen
- Prozesszeiten: Hierunter fallen Stellgrößen die vor allem die manuellen Tätigkeiten im Prozessabläufe beeinflussen.

Zur Bewertung der Ergebnisse wurden zusammen mit MR die entsprechenden Kennzahlen und deren Berechnungsmethodik festgelegt. Letzteres ist erfolgt, um auch die Vergleichbarkeit mit den Kennzahlen aus der Realanlage zu ermöglichen. Auch diese lassen sich wiederum in einige Gruppen zusammenfassen:

- Fertigungsaufträge: Neben der durchschnittlichen Durchlaufzeit eines Fertigungsauftrags wird auch die Gesamtabarbeitungsdauer aller Fertigungsaufträge über den Simulationszeitraum ermittelt.
- Auslastung: Die durchschnittliche Auslastung wird sowohl für Anlagen als auch für Mitarbeiter berechnet. Für die Anlagen werden zudem noch die kumulierten Rüstzeiten als Kennzahl aufbereitet. Hierzu werden die seitens MR bereitgestellten Berechnungslogiken verwendet.
- Sonstiges: Da es eine Vielzahl von Sonderabläufen gibt, z.B. ein Prozessschritt muss wiederholt werden da eine Liegezeit überschritten wurde, werden auch solche prozessspezifischen zeitlichen Verschwendungen statistisch erfasst.

Das Modell wurde zusammen mit MR basierend auf den bereitgestellten Detaildaten validiert. Mit dem validierten Modell hat die TU Ilmenau dann mehr als 60.000 Simulationsexperimente durchgeführt. Auch hier konnte Wissen generiert werden und eine Reihe von Management-Empfehlungen abgeleitet werden, von denen die Maschinenfabrik Reinhausen einige in ihre Abläufe integriert hat.

6.6 Gesamtziel – Transferieren des Data Farmings und der Wissensentdeckung in Simulationsdaten in die Praxis

Die prototypische Implementierung des gesamten Data Farming Ablaufs im Prototyp einer kommerzialisierbaren Software liegt als Ergebnis des Projektes vor. Die Erprobung durch die Praxispartner ist bereits im Projekt erfolgt, vor allem Anwender von C-P-S konnten unmittelbar mit dem Werkzeug arbeiten.

Sicher erfordert eine kommerzielle Verwertbarkeit noch die eine oder andere Weiterentwicklung gerade auch in Bezug auf die Unterstützung der Anwender. Aber auch dazu wurden in dem Projekt bereits viele Hilfen geschrieben und entwickelt, die es unerfahrenen Anwender ermöglicht sich an einfachen Beispielen mit grafischer Unterstützung in die Verfahren einzulesen. In Abbildung 16 ist beispielhaft ein Ausschnitt aus einer Hilfestellung zu dem Latin Hypercube Sampling dargestellt, welches an einer einfachen Grafik die Raumabdeckung durch die Experimente visualisiert und eine kurze Erläuterung zu der Anwendung gibt.

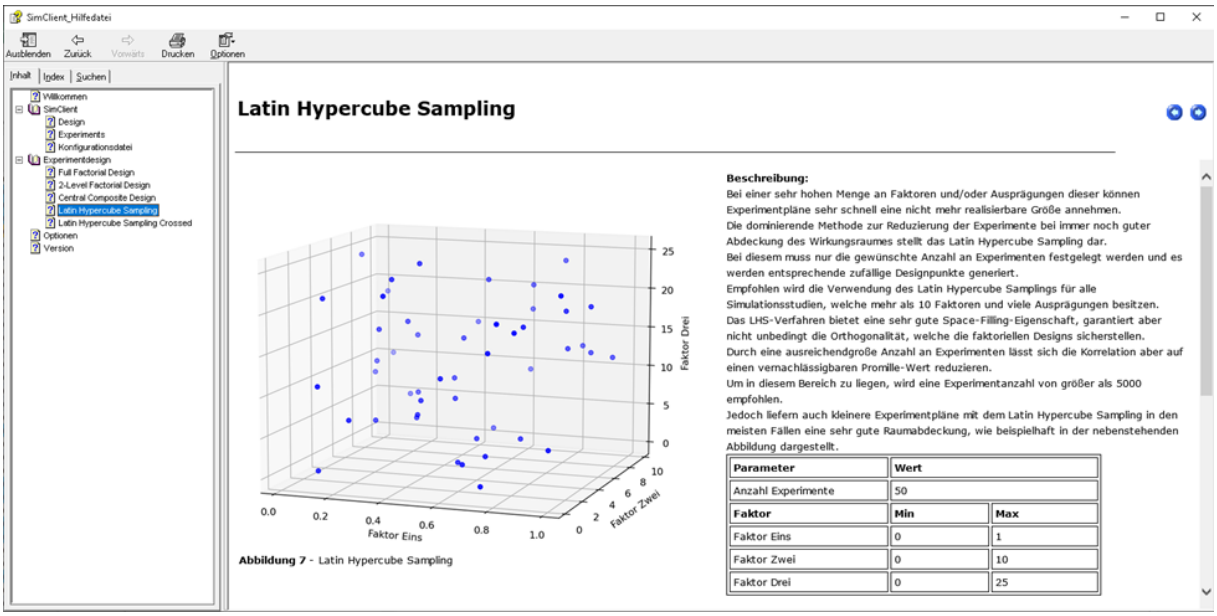

Abbildung 16: Ausschnitt aus der Hilfestellung zu dem Latin Hypercube Sampling.

6.7 Gesamteinschätzung der erreichten Ziele

Nach eigener Einschätzung hat die TU Ilmenau und das gesamte Vorhaben die selbstgesetzten Ziele aus der Vorhabenbeschreibung alle erreicht. Insbesondere ist das Hauptziel, der Transfer der Entwicklungen im Bereich Data Farming in eine integrierte Software, erreicht. Kleine Abstriche bezüglich der Zielerreichung sehen die Projektpartner bei Ziel GZ3, den Data-

Mining- und Visual-Analytics-Methoden. Hier wäre eine noch stärkere Assistenz des Benutzers angesichts der vielfältigen Anwendungs- und Interpretationsmöglichkeiten durch verschiedene Interaktionsmöglichkeiten und Visualisierungen wünschenswert gewesen. An dieser Stelle besteht weitergehender Forschungs- und Entwicklungsbedarf.

7 Wichtigste Positionen des zahlenmäßigen Nachweises

Die genehmigten Vorgaben des Antrags bzgl. der verfügbaren Budgets wurde eingehalten, aufgrund der pandemischen Lage über fast die gesamte Projektlaufzeit und den bereist beschriebenen kleinen Verschiebungen im Projektplan sowie auf Grund der konkreten Eingruppierung des Projektmitarbeiters kam es zu Abweichungen zwischen den einzelnen Posten, siehe Tabelle 1. Details sind den Mitabrufen / Abrechnungen zu entnehmen.

Tabelle 1: Aufwände des FG ITPL im Vergelich IST/SOLL .

Zeitraum	Aufwände ITPL	SOLL	IST	Delta
Gesamt	**Gesamtsumme (in €)**	**170.155,63**	**173.055,04**	**3.585,46**
	Personalkosten (in €)	134.856,36	135.880,19	1.023,83
	Personalmonate (in PM)	24	25	1
	Reisekosten Inland (in €)	4.940,00	1.481,41	-3.458,59
	Reisekosten Ausland (in €)	2.000,00	7.422,64	5.422,64
	Abschreibung auf Anlagen oder Material (in €)	-	-	-
	Projektpauschale (in €) *	28.359,27	28.270,24	597,58
	Kassenbestand		-3.430,24	
2020	Personalkosten (in €)	67.013,52	61.200,86	-5.812,66
	Personalmonate	12	12	0
	Reisekosten Inland (in €)	2.380,00	338,59	-2041,41
	Reisekosten Ausland (in €)	0	0	0
	Abschreibung auf Anlagen oder Material (in €)	-	-	-
2021	Personalkosten (in €)	67.842,84	74.679,33	6.836,49
	Personalmonate (in PM)	12	13	1
	Reisekosten Inland (in €)	2.560,00	1.142,82	-1.417,18
	Reisekosten Ausland (in €)	2.000,00	7.422,64	5.422,64
	Abschreibung auf Anlagen oder Material (in €)	-	-	-

* gilt nur für Hochschulen und Universitätskliniken

8 Notwendigkeit und Angemessenheit der geleisteten Arbeit

In diesem Abschnitt werden die einzelnen Arbeitspakete näher dargestellt einschließlich ihrer Laufzeit, der vorkalkulierten und der eingesetzten Personenmonate sowie einer Beschreibung der Umfänge. Das Format entspricht der Vorgabe seitens des Projektträgers; die Gliederung entspricht den sieben Arbeitspaketen aus der Tabelle in Abbildung 2 in Abschnitt 3 Alle Angaben beziehen sich ausschließlich auf das FG ITPL der TU Ilmenau.

AP 0 – Projektmanagement			
Geplante Laufzeit:	03/2020 – 12/2021	Insgesamt kalkulierte PM:	1 PM
Tatsächliche Laufzeit:	03/2020 – 12/2021	Insgesamt eingesetzte PM:	1 PM
Kurzbeschreibung: Inhalt des APs waren die Publikation der Projektfortschritte/Ergebnisse, die Öffentlichkeitsarbeit sowie die Berichtserstellung und des Verwertungskonzeptes im allgemeinen.			
Geleistete Arbeiten: ITPL hat beim Berichtswesen und rund um Termine und Workshops den Projektkoordinator SimPlan unterstützt sowie maßgebliche die wissenschaftlichen Publikationen vorangetrieben.			
Anmerkungen: Die ausgewiesenen PMs für AP0 umfassen die Arbeiten bis einschließlich Dezember 2021. Tatsächlich sind darüber hinaus noch Umfänge für die Berichtserstellung und Publikationen angefallen, die in der Nachkalkulation nicht enthalten sind.			

AP 1 – Methoden für smartes Experimentdesign			
Geplante Laufzeit:	01/2020 – 08/2020	Insgesamt kalkulierte PM:	4 PM
Tatsächliche Laufzeit:	01/2020 – 11/2020	Insgesamt eingesetzte PM:	4 PM

Kurzbeschreibung: Ziele des APs waren die Aufnahme der Anwenderanforderungen an das Experimentdesign, die Auswahl geeigneter Verfahren vor dem Hintergrund dieser Anforderungen und die testweise Implementierung. Das FG ITPL war hier als Know-How-Träger als auch der Scripterstellung an sich tätig.
Geleistete Arbeiten: Das FG ITPL hat gemeinsam mit SimPlan die Anforderungsworkshops durchgeführt und hat anschließend die Verfahren für das Experimentdesign evaluiert und prototypisch in Python-Scripten (nach Vorgaben des Rahmenwerkes durch SImPlan) implementiert. Abschließend erfolgten Tests sowohl der Scripte allein als auch im entstanden SimAssist Modul.
Anmerkungen: -

AP 2 – Experimentverteilung			
Geplante Laufzeit:	03/2020 – 01/2021	Insgesamt kalkulierte PM:	5,5 PM
Tatsächliche Laufzeit:	03/2020 – 04/2021	Insgesamt eingesetzte PM:	5 PM
Kurzbeschreibung: Ziele des APs waren die Konzeption und Entwicklung von Techniken zur flexiblen Experimentverteilung, die modell- und simulatorneutral und zur Verteilung in lokalen Rechnernetzen (Private Clouds) sowie in cloudbasierten Infrastrukturen sein sollten. ITPL war hier insbesondere bei der Anforderungserhebung, der Analyse der Möglichkeiten sowie den Test der durch SimPlan implementierten Lösung (SimController) involviert.			
Geleistete Arbeiten: Ähnlich wie in AP1 hat sich das FG ITPL gemeinsam mit der SimPlan AG um die Aufnahme der Anforderungen gekümmert alternativen geprüft. Zusätzlich hat das FG ITPL die entstanden Infrastruktur intensiv getestet und beim Debuggen unterstützt.			
Anmerkungen: Auf Grund der guten Vorarbeiten der SimPlan AG, konkret in Form des SimControllers, welcher adaptiert werden konnte, vielen die Aufwände zur Evaluation kleiner aus dagegen waren die Aufwände für das Testen großer als erwartet. In Summe konnte 0,5 PM Aufwand beim FG ITPL eingespart werden.			

AP 3 – Data Mining- und Visual Analytics-Methoden			
Geplante Laufzeit:	02/2020 – 01/2021	Insgesamt kalkulierte PM:	7 PM
Tatsächliche Laufzeit:	02/2020 – 05/2021	Insgesamt eingesetzte PM:	7,5 PM
Kurzbeschreibung: Ziele des APs waren die Auswahl und Bereitstellung geeigneter Data Mining- und Visual Analytics-Methoden für die Benutzenden. Wie schon in AP1 und AP2 war das FG ITPL hier bei der Aufnahme der Anforderungen beteiligt und hat weite Teile der Algorithmen/Scripte ausgewählt sowie prototypisch evaluiert und implementiert, so dass diese durch SimPlan in die Gesamtlösung integriert werden konnten.			
Geleistete Arbeiten: ITPL war wiederum in die Aufnahme der Anforderungen involviert und hatte den Schwerpunkt bei der Auswahl, Evaluation und Implementierung der Verfahren in entsprechend Python-Scripte.			
Anmerkungen: Der tatsächliche Aufwand bei der Evaluation und Implementierung viel leicht höher als geplant aus.			

AP 4 – Integration und Nutzerinterfaces			
Geplante Laufzeit:	10/2020 – 04/2021	Insgesamt kalkulierte PM:	0,5 PM
Tatsächliche Laufzeit:	10/2020 – 08/2021	Insgesamt eingesetzte PM:	1,5 PM
Kurzbeschreibung: Ziel dieses APs war in erster Linie die Integration der in den übrigen Arbeitspaketen entstandenen Prototypen in eine durchgängige Softwareapplikation. Schwerpunkt waren die Unterstützung der SimPlan AG bei Konzeption sowie die Test der entstanden Lösung.			
Geleistete Arbeiten: Test und Unterstützung beim Debuggen der im Rahmen der Arbeitspakete AP1 – AP3 entstandenen Softwareprototypen in der Rahmenapplikation SimAssist 4farm-PlugIn sowie der modellunabhängige Anbindungen an die Simulationswerkzeuge Plant Simulation und Anylogic.			

Anmerkungen: Auf Grund des sehr flexiblen Mechanismus war insbesondere der Testaufwand höher als geplant, die Flexibilität erleichterte aber das Arbeiten an anderen Stellen des Projektes, z. B. durch die gute Trennbarkeit der Entwicklung in den AP 1 und 3.

AP 5 – Dokumentation und Erstellung der Howtos			
Geplante Laufzeit:	03/2021 – 07/2021	Insgesamt kalkulierte PM:	2 PM
Tatsächliche Laufzeit:	06/2021 – 11/2021	Insgesamt eingesetzte PM:	2 PM
Kurzbeschreibung: Ziel des APs es, Handreichungen zusammenzustellen, die es Anwendern der Data Farming Lösung so einfach wie möglich machen sollen, die entstandene Software zu nutzen und die Anbindung an Simulationswerkzeuge (exemplarisch Plant Simulation und Anylogic) durchzuführen.			
Geleistete Arbeiten: ITPL hat sowohl an der internen Dokumentation als auch an entsprechenden Lehrmaterial insbesondere für die die verschiedenen Experimentdesign- und Data Mining-Verfahren mitgewirkt.			
Anmerkungen: -			

AP 6 – Validierung in der Pilotanwendung			
Geplante Laufzeit:	03/2021 – 12/2021	Insgesamt kalkulierte PM:	4 PM
Tatsächliche Laufzeit:	*03/2020* – 12/2021	Insgesamt eingesetzte PM:	4 PM
Kurzbeschreibung: Ziel dieses APs war der Test der im Vorhaben entwickelten Verfahren an Pilotanwendungen der beiden Anwendungspartner. Aufgabe des FG ITPL war die Unterstützung des Anwendungstests.			
Geleistete Arbeiten: Das FG ITPL hat insbesondere die Anwendungspartner MR und C-P-S unterstützt an den erstellten Modellen die Data Farming Anwendung zu erproben. Über die ursprünglich vorgesehenen Umfänge hinaus wurden die Pilotanwendungen intensiv			

zum Ableiten der Anforderungen an die Verfahren eingesetzt. Des Weiteren konnten in beiden Studien Managementempfehlungen abgeleitet werden.
Anmerkungen: Die Erstellung der Pilotanwendungen wurde gegenüber der ursprünglichen Projektplan weit in das erste Projektjahr hinein vorgezogen, weil sich herausgestellt hat, dass es allen Projektbeteiligten deutlich leichter fällt, Anwendungen entlang konkreter Piloten zu diskutieren und zu veranschaulichen (an Stelle einer ansonsten eher abstrakten Methodendiskussion).

9 Voraussichtlicher Nutzen und Verwertbarkeit der Ergebnisse im Sinne des fortgeschriebenen Verwertungsplans

Eine wirtschaftliche Verwertung durch die TU Ilmenau ist keine explizit eingeplante Maßnahme. Allerdings ist denkbar, dass Teile der entwickelten Lösung intellektuelles Eigentum der SimPlan AG und der TU Ilmenau werden, so dass sich aus der wirtschaftlichen Verwertung durch die SimPlan AG mittelbar eine wirtschaftliche Verwertung für Ilmenau ergibt.

Die wissenschaftlich-technische Verwertung durch das Fachgebiet ITPL erfolgt zum einen durch die Einbindung der Projektergebnisse in die Lehrveranstaltungen des Fachgebietes und somit direkt in die Ausbildung angehender Simulationsexperten, Fertigungsplaner, Data-Scientists etc. Hierbei kommen mehrere potenzielle Lehrveranstaltungen zu den Themen Digitale Fabrik, Simulation oder auch Industrie 4.0 in Frage. Konkret ergeben sich für die Einbindung der Projektergebnisse die folgenden Lehrveranstaltungen:

- Methoden und Werkzeuge der Digitalen Fabrik
- Fabrikbetriebe
- Simulation 1
- Steuerung von Produktionssystemen
- Industrie 4.0
- Data Science für industrielle Anwendungen

Auch die Verwendung der entwickelten Python-Skripte in zukünftigen Fallstudien oder Abschlussarbeiten ist denkbar. Drei der im Rahmen der wissenschaftlichen Verwertung bei der TU Ilmenau vorgesehenen Publikationen sind bereits erfolgt (vgl. Abschnitt 11). Eine

weitere befindet sich in der Einreichung. Insofern sind wesentliche Aspekte der wissenschaftlichen Anschlussfähigkeit aus heutiger Sicht bereits erreicht und bieten in Bezug auf die Einbindung in der Lehre viele Möglichkeiten. Aus der Kooperation heraus sind Folgeprojekte aktuell bereits in der Antragsentwurfsphase.

Insgesamt entsprechen Nutzen und Verwertungsmöglichkeiten aus heutiger Sicht den Planungen zu Beginn des Vorhabens in vollem Umfang.

10 Während der Durchführung des Vorhabens dem ZE bekannt gewordener Fortschritt auf dem Gebiet des Vorhabens bei anderen Stellen

Es sind im Projektverlauf keine vorhabensrelevanten Ergebnisse Dritter bekannt geworden, die nicht vorher bekannt waren. Es gab insoweit auch keine Notwendigkeit einer Anpassung der Ziele des Projektes. Als Informationsquelle dienten kleinere Recherchen während der Projektdauer und eine umfassende Literaturrecherche gegen Ende des Projektes zu Data Farming und Wissensentdeckung in Simulationsdaten. Dabei haben sich aber Potenziale für Anschlussforschung ergeben. Zum Beispiel im Zusammenspiel von Data Farming und Künstlicher Intelligenz oder Synergien mit anderen BMBF geförderten Projekten (AIDA-Vis). Es sind aber keine Arbeiten bekannt geworden, die zu einer veränderten Einschätzung des DaWiS-Vorhabens und der erzielten Ergebnisse geführt haben.

11 Erfolgte oder geplante Veröffentlichungen des Ergebnisses

In die folgenden Veröffentlichungen rund um das DaWiS-Vorhaben ist die TU Ilmenau involviert:

- Projektwebseite DaWiS – ITPL: https://www.tu-ilmenau.de/universitaet/fakultaeten/fakultaet-maschinenbau/profil/institute-und-fachgebiete/fachgebiet-informationstechnik-in-produktion-und-logistik/projekte/dawis
- Genath, Jonas; Bergmann, Sören; Spieckermann, Sven; Stauber, Stephan; Feldkamp, Niclas: Entwicklung einer integrierten Lösung für das Data Farming und die Wissensentdeckung in Simulationsdaten. In: Proceedings der ASIM FT - Simulation in Produktion und Logistik 2021, (2021), S. 377-386
- Genath, Jonas; Bergmann, Sören; Feldkamp, Niclas; Straßburger, Steffen: Automatisierung im Prozess der Wissensentdeckung in Simulationsdaten - Cha-

rakterisierung der Ergebnisdaten. In: Proceedings der ASIM FT - Simulation in Produktion und Logistik 2021, (2021), S. 367-376

- Genath, Jonas; Bergmann, Soeren; Straßburger, Steffen; Stauber, Stephan; Spieckermann, Sven: An Integrated Solution for Data Farming and Knowledge Discovery in Simulation Data: A Case Study of the Battery Supply of a Vehicle Manufacturer. In: Proceedings 2021 Winter Simulation Conference, Phoenix, AZ, USA, (2021)
- Genath, Jonas; Bergmann, Soeren; Straßburger, Steffen; Spieckermann, Sven; Stauber, Stephan: Data Farming und Wissensentdeckung in Simulationsdaten - Entwicklung einer integrierten Lösung. ZWF - Zeitschrift für wirtschaftlichen Fabrikbetrieb, 3/2022
- Genath, Jonas; Bergmann, Sören; Spieckermann, Sven; Stauber, Stephan; Feldkamp, Niclas: Development of an Integrated Solution for Data Farming and Knowledge Discovery in Simulation Data SNE - Simulation Notes Europe (eingereicht)
- Straßburger, Steffen: Data Farming using Plant Simulation and SimAssist. Plant Simulation User Day, 30th May Berlin (Vortrag)