



REPORT 46



ASIM SST 2024

Tagungsband Kurzbeiträge



ASIM SST 2024

27. ASIM Symposium Simulationstechnik

Universität der Bundeswehr München



4.9. – 6.9.2024

Herausgegeben von Oliver Rose und Tobias Uhlig

ASIM SST 2024

Tagungsband Kurzbeiträge

27. ASIM Symposium Simulationstechnik

4.9.-6.9.2024

Universität der Bundeswehr München

Herausgegeben von
Oliver Rose, Tobias Uhlig, Univ. der Bundeswehr München

ISBN ebook ISBN 978-3-903347-64-9

DOI 10.11128/arep.46

ARGESIM Report 46

ASIM Mitteilung 189

ARGESIM Verlag, Wien, 2024

www.argesim.org

Bibliographic Data:

Publisher: ARGESIM Publisher, Vienna

Title: Tagungsband Kurzbeiträge ASIM SST 2024

Subtitle 27. ASIM Symposium Simulationstechnik, 4.9.-6.9.2024, Universität der Bundeswehr München

Author(s):

Editor(s): Oliver Rose, Tobias Uhlig, Universität der Bundeswehr München

Series: ARGESIM Reports

Series Editors: Felix Breitenecker, Thorsten Pawletta, ASIM

Volume: ARGESIM Report no. 46

ISBN ebook: 978-3-903347-64-9, ARGESIM Verlag

DOI: 10.11128/arep.46

ASIM ID: ASIM Mitteilung 189

Publication Date: September 1, 2024

Number of Pages: vi + 56 pages

Cover: © Adobe Stock

Copyright © 2024 ASIM & ARGESIM Publisher

Copyright Information / Regulations ARGESIM

ARGESIM is a non-profit scientific society generally aiming for dissemination of information on system simulation - from research via development to applications of system simulation. ARGESIM's primary publication is the journal SNE – Simulation Notes Europe with open access to all contributions; generally, the authors retain the copyright of their

SNE contributions. This copyright regulation holds also for *ARGESIM Reports* and *ARGESIM Advances in Simulation* publishing conference volumes for ASIM, MATHMOD, and EUROSIM (in consideration of copyright regulations for related conference publications) and monographs on system simulation (mainly PhD theses).

About ARGESIM

ARGESIM is a non-profit society generally aiming for dissemination of information on system simulation from research via development to applications of system simulation. ARGESIM is closely co-operating with EUROSIM, the Federation of European Simulation Societies, and with ASIM, the German Simulation Society. ARGESIM is an 'outsourced' activity from the Mathematical Modelling and Simulation Group of TU Wien, there is also close co-operation with TU Wien (organisationally and personally). *ARGESIM Publisher* organizes publishing activities, with ISBN root 978-3-901608-xx-y and DOI root 10.11128/xx...x.

ARGESIM's activities are:

- Publication of the scientific journal SNE - Simulation Notes Europe (Membership Journal of EUROSIM, the Federation of European Simulation Societies) → www.sne-journal.org
- Organisation and Publication of the ARGESIM Benchmarks for Modelling Approaches and Simulation Implementations → www.argesim.org/benchmarks/
- Publication of the series ARGESIM Reports (for monographs in system simulation, and proceedings of simulation conferences and workshops) → www.argesim.org/publications/
- Publication of the special series FBS Simulation - Advances in Simulation / Fortschrittsberichte Simulation (monographs in co-operation with ASIM, the German Simulation Society)
- Organisation of the Conference Series MATHMOD Vienna (triennial, in co-operation with EUROSIM, ASIM, and TU Wien) → www.mathmod.at
- Administration and support of ASIM (German Simulation Society → www.asim-gi.org) and of EUROSIM (Federation of European Simulation Societies → www.eurosim.info)

ARGESIM – Arbeitsgemeinschaft Simulation News – Working Committee Simulation News – SNE Publication
Mommengasse 19/8, 1040 Vienna, Austria; Tel +43-1-58801-10111, -10115; Fax +43-1-58801-910111
Email: office@argesim.org, office@sne-journal.org; WWW: www.argesim.org, www.sne-journal.org
Incorporated Austrian Society ZVR No 213056164 – EU VAT ID No ATU 72054279
Bank Account: ARGESIM, IBAN AT07 2011 1828 9115 0800, BIC GIBAATWWXXX, ERSTE BANK VIENNA

VORWORT

ASIM lädt zum 27. ASIM Symposium Simulationstechnik, zu einer bereits klassischen ASIM-Tagung. Traditionell findet das **ASIM Symposium Simulationstechnik (SST)** alle zwei Jahre im Herbst statt, alternierend Tagung **ASIM SPL**, der internationale Fachtagung Simulation in Produktion und Logistik. Mit **ASIM SST 2024** ist die Tagung nach Verschiebungen und virtuellen Durchführungen in der Corona-Zeit wieder in den planmäßigen ASIM-Tagungskalender zurückgekehrt.

ASIM SST 2024 bietet traditionellerweise zwei Arten von Beiträgen, die alle einem Peer Review Prozess unterzogen wurden und sich in der Publikationsform unterscheiden:

- **ASIM SST 2024 Langbeiträge** (6-10 Seiten, deutsch oder englisch)
mit individueller DOI für jeden Beitrag e-Tagungsband:
Tagungsband Langbeiträge ASIM SST 2024, ARGESIM Report 47, ASIM Mitteilung 190, ISBN ebook 978-3-903347-65-6 ARGESIM Verlag
- **ASIM SST 2024 Kurzbeiträge** (2-5 Seiten, deutsch oder englisch), mit Tagungsband-DOI:
Tagungsband Kurzbeiträge ASIM SST 2024, ARGESIM Report 46, ASIM Mitteilung 189, ISBN ebook 978-3-903347-64-9 ARGESIM Verlag

ASIM SST 2024 wird veranstaltet an der Universität der Bundeswehr München am Campus Neubiberg, Veranstalter sind Oliver Rose und Tobias Uhlig vom Institut für Technische Informatik, unterstützt vom lokalen Organisationskomitee.

Das Programmkomitee bestand aus den ASIM-Mitgliedern Felix Breitenecker, Kurt Chudej, Walter Commerell, Olaf Hagendorf, Lukas Hollenstein, Andreas Körner, Christoph Laroque, Xiaobo Liu-Henke, Daniel Lücknerath, Thorsten Pawletta, Niki Popper, Oliver Rose, Michael Striebel, Tobias Uhlig, Oliver Ullrich, Sigrig Wenzel, und Jochen Wittmann. Gemeinsam mit dem SNE Review Board wird es Tagungsbeiträge zur Publikation in ASIM's Scientific Journal *Simulation Notes Europe* vorschlagen.

Zum **ASIM SST 2024** haben sich etwa 60 Fachleute zusammengefunden, zwei Drittel davon ASIM-Mitglieder. Dieser Tagungsband, *Kurzbeiträge ASIM SST 2024* publiziert die 13 angenommenen Kurzbeiträge – mit Volume DOI. Auch für Kurzbeiträge gibt es die Möglichkeit für Postconference Publications in ASIM's Scientific Journal *Simulation Notes Europe SNE*.

Für Hauptvorträge konnten wir zwei Simulationsfachleute gewinnen.

Hans Ehm, Head of Supply Chain Innovation der Infineon Technologies AG und
Felix Breitenecker, langjähriger ASIM-Sprecher.

ASIM SST 2024 bot auch interessantes Begleitprogramm: Mittwoch, 4.9.2024 gab es einen kleinen Empfang in der Münchener Innenstadt mit anschließenden Themen-Stadtführungen, und Donnerstag, 5.9.2024 eine Besichtigung des Flugsimulators mit anschließendem Abendbankett im Offizierscasino.

Wir Organisatoren möchten allen herzlich danken, die das **ASIM SST 2024** zu einem Erfolg machen: den Teilnehmerinnen und Teilnehmern, den Autoren der Beiträge, dem Programmkomitee für das Reviewing, dem ARGESIM Verlag für die Publikationen, und unseren Tagungsmitarbeitern.

Oliver Rose, Tobias Uhlig, September 2024

Inhaltsverzeichnis – List of Content

27. ASIM Symposium Simulationstechnik

Kurzbeiträge und Abstracts Hauptvorträge

Herausforderungen und (Simulations) Lösungen in Halbleiter Supply Chains Abstract Hauptvortrag; <i>H. Ehm</i>	1
Simulationstechnik und ASIM - Gestern, Heute, Morgen Abstract Hauptvortrag; <i>F. Breitenecker</i>	3
Mathematical model for spatial and temporal risk analysis of West Nile Virus Transmission in Bavaria, Germany; <i>O. Chinonso Mbaoma, S. M. Thomas, C. Beierkuhnlein</i>	5
Entwicklung einer Bewertungsmethode für Simulationsergebnisse im Kontext industrieller Symbiose: Ein Zielsystem zur Steigerung von Ressourceneffizienz und Resilienz <i>B. Vollack, M. Boelcke, T. Schmidt</i>	9
Entwicklung einer verteilten Simulationsinfrastruktur auf einem Raspberry Pi-Cluster mit Kubernetes und KNIME; <i>M. Großmann, M. Leißau, C. Laroque</i>	13
Wird Simulation durch KI überflüssig? - eine subjektive Bestandsaufnahme - <i>J. Wittmann, T. Clemen</i>	17
Beschleunigte Erstellung von Simulationsmodellen für Produktions- und Logistikprozesse mithilfe von GPT-basierten Large Language Models; <i>R. Krämer, J. Heger</i>	21
Concept development of coupling Simulation and Machine Learning in Supply Chains <i>Y. Mai Thi, M. Münnich, F. Zumpe, M. Callefi, P. Grzona</i>	25
The Filtering Effect on Simulated Signals under Consideration of Entropy Methods <i>A. Edthofer, A. Körner</i>	29
Konzeption eines Vorgehensmodells zur Auswahl von Entscheidungsunterstützungsmethoden in der Produktion und Logistik; <i>K. Langenbach, M. Rabe</i>	33
CMOS Technologies for OTAs: A High-Dimensional Comparison <i>B. Ohse, J. Kampe, C. Schneider</i>	37
Data-adaptive dynamic simulation via structured Dynamic Mode Decomposition <i>C. Bonenberger, S. Scholz, N. Scheiter</i>	41
Simulation and Control of 2-Dimensional Anisotropic Heat Conduction <i>S. Scholz, C. Bonenberger, N. Scheiter, L. Berger</i>	45
Ermittlung der Flottengröße für Systeme Autonomer Mobiler Roboter (AMR) <i>I. Meinhardt, F. Schulze, T. Schmidt</i>	49
Simulation-Based Analysis of Storage Strategies for an Inland Container Terminal <i>M. Mowe, J. Manemann, L. Jurgeleit, M. Kiefer, U. Clausen</i>	53

Langbeiträge (L) sind zu finden im

ARGESIM Report 47 Tagungsband Langbeiträge ASIM SST 2024

ISBN ebook 978-3-903347-65-6, DOI 10.11128/arep.47, ARGESIM Report 47, ASIM Mitteilung 190

Kurzbeiträge (K) und Abstracts der Hauptvorträge (H) sind zu finden im

ARGESIM Report 46 Tagungsband Kurzbeiträge ASIM SST 2024

ISBN ebook 978-3-903347-64-9, DOI 10.11128/arep.46, ARGESIM Report 46, ASIM Mitteilung 189

AUTORENINDEX – INDEX OF AUTHORS

27. ASIM Symposium Simulationstechnik

Kurzbeiträge - Abstracts Hauptvorträge

M. Apostolakis	(L 95)	S. Jacobitz	(L 129, 175)
A. Apostolatos	(L 95)	D. Jammer	(L 211, 219)
C. Appel	(L 35)	M. Jelibaghu	(L 203)
C. Beierkuhnlein	5	P. Junglas	(L 211, 219)
J. Benz	(L 19)	L. Jurgeleit	5
L. Berger	45 , (L 103)	J. Just	(L 19)
S. Bergmann	(L 81)	J. Kampe	37
K. Blank	(L 19)	V. Karumuri	(L 121)
D. Bodnár	(L 113)	S. Kerber	(L 95)
M. Boelcke	9	M. Kiefer	53
C. Bonenberger	41, 45	A. Konzept	(L 137)
F. Breitenecker	3	A. Körner	29
A. Buchele	(L 61)	R. Krämer	21
M. Callefi	25	C. Krull	(L 113, 121)
K. Chudej	(L 1, 9, 35)	M. Kühn	(L 167)
U. Clausen	53	K. Langenbach	33
T. Clemen	17	C. Laroque	13 , (L 143)
A. Dam	(L 43)	C. Lau	(L 19)
J. Dettelbacher	(L 61)	D. Lebiedz	(L 103)
A. Edthofer	K 29	M. Lechner	(L 73)
H. Ehm	1	M. Leißau	13 , (L 143)
M. Eley	(L 203)	T. Li	(L 129)
N. Feldkamp	(L 81)	X. Liu-Henke	(L 129, 175)
R. Ferrara	(L 175)	Y. Mai Thi	25
P. Flender	(L 129)	J. Manemann	53
J. Franke	(L 15)9	C. May	(L 159)
G. Gaßner	(L 143)	O. Mbaoma	5
M. Göllner	(L 129, 175)	I. Meinhardt	(L 49)
S. Groß	(L 95)	S. Miller	(L 95)
M. Großmann	13	M. Mowe	53
P. Grzona	25	M. Münnich	25
T. Hammer-Jakobsen	(L 43)	N. Nakhost-Lotfi	(L 43)
K. Handrich	(L 167)	A. Niotis	(L 1
J. Heger	21	B. Ohse	37
C. Hervás	(L 95)	F. Özkul	(L 185, 193)
A. Hitz	(L 137)	A. Palatnik	(L 203)
S. Hötzel	(L 19)	F. Pappert	(L 151)

T. Pawletta	(L 211, 219)	J. Stromberger	(L 61)
S. Pawletta	(L 211, 219)	P. Studenic	(L 4)3
N. Popper	(L 43)	L. Suchy	(L 159)
F. Raabe	(L 67)	R. Sutherland	(L 185, 193)
M. Rabe	33	S. Thomas	5
H. Radner	(L 43)	T. Thorwart	(L 203)
B. Reick	(L 137)	C. Tung	(L 95)
S. Reitelshöfer	(L 159)	T. Uhlig	(L 27)
J. Rohlfis	(L 9)	C. Urach	(L 43)
O. Rose	(L 27, 73, 151, 203)	J. van Rensburg	(L 95)
L. Roth	(L 203)	M. Vega Torres	(L 95)
J. Ruffer	(L 143)	T. Voigt	(L 143)
N. Scheiter	41, 45	B. Vollack	9, (L 27)
F. Schicks	(L 19)	N. von der Ahe	(L 19)
T. Schmidt	9, 49, (L 167)	S. Wenzel	(L 185, 193)
C. Schneider	37	T. Wiedemann	(L 51)
S. Scholz	41, 45, (L 103)	J. Wittmann	17, (L 19, 67)
F. Schulze	K 49	S. Xie	(L 27)
N. Schwab	(L 43)	X. Xu	(L 227)
D. Seufferth	(L 151)	G. Zauner	(L 43)
T. Stamm	(L 43)	T. Zhang	(L 27)
H. Stein	(L 151)	D. Zimmer	(L 89)
J. Steinbichl	(L 67)	R. Zöllner	(L 167)
		F. Zumpe	25

Langbeiträge (L) sind zu finden im

ARGESIM Report 47 Tagungsband Langbeiträge ASIM SST 2024

ISBN ebook 978-3-903347-65-6, DOI 10.11128/arep.47, ARGESIM Report 47, ASIM Mitteilung 190

Kurzbeiträge (K) und Abstracts der Hauptvorträge (H) sind zu finden im

ARGESIM Report 46 Tagungsband Kurzbeiträge ASIM SST 2024

ISBN ebook 978-3-903347-64-9, DOI 10.11128/arep.46, ARGESIM Report 46, ASIM Mitteilung 189

Herausforderungen und (Simulations) Lösungen in Halbleiter Supply Chains und Supply Chains die Halbleiter beinhalten

Hans Ehm

Infineon Technologies AG, Am Campeon 1-15, 85579 Neubiberg, Deutschland; hans.ehm@infineon.com

Halbleiterproduktion ist hochkomplex

Während Halbleiter eine intrinsisch lange Produktionszeit haben, haben die Produkte, die Halbleiter beinhalten eine verhältnismäßig kurze Lebenszeit (weil man mit dem nächsten, besseren Halbleiter schon wieder noch wettbewerbsfähigere Produkte entwickeln kann).

Die Halbleiterindustrie ist ein Befähiger zur Reduzierung von CO₂ in deren Anwendungen, insbesondere wenn Hochleistungshalbleiter mit geringer Verlustleistung zum Einsatz kommen. In diesem Umfeld haben sich globale Supply Chain Lösungen als differenzierender Wettbewerbsfaktor herauskristallisiert, welche die physikalischen Flexibilitäten täglich optimal nutzen.

Simulation und KI in der Halbleiterproduktion

Modellierungen und Simulationen finden dafür auf allen Ebenen (Maschine, Fabrik, interne- und End-to-End-Supply-Chain) statt, dabei kommen sowohl Diskrete Event Simulation (DES), Agenten-basierte Simulation, als auch System Dynamics Ansätze zur Anwendung.

Semantische Netze, die mit Simulationen Hand in Hand gehen, helfen Daten zu strukturieren und ermöglichen KI-basierte Verbesserungen, die durch die Semantik auf vorhandenes Wissen aufbauen.

CV Hans Ehm

Hans Ehm ist Senior Principal Supply Chain der Infineon Technologies AG und leitet die Abteilung supply chain Innovation. Er hat Physik an der Hochschule München (Dipl. Ing (FH)) und Mechanical Engineering an der Oregon State University (M.S./OSU) studiert. Während seiner 40-jährigen Tätigkeit in der Halbleiterindustrie hatte er Management Positionen und Beratungspositionen im Frontend (Wahrfertigung) im Backend (Assembly und Test) und in der globalen Supply Chain inne.



Hans Ehm war im Aufsichtsrat der camLine Holding AG und von APICS SCC (heute ASCM), der weltweit größten Organisation für supply chain und Besitzer des SCOR Models.

2015 erhielt er den LEO-Award (Logistics Excellence Optimisation) in der Kategorie "Manager" von der DVZ Media Group.

Er initiierte und leitete zahlreiche deutsche und europäische Förderprojekte wie SC³: semantically connected semiconductor supply chains. Er ist seit 2023 Sherpa des EU High Level Forums für Standardisierung

2023 hielt er die Keynote der INFORMS/Wintersimulation MASM (Modeling and Simulation for Semiconductor Manufacturing and supply Chain) Konferenz.

Seine zahlreichen Veröffentlichungen zu Modellierung und Simulationen sind auf Reserach Gate unter Hans Ehm Lab zu finden

Simulationstechnik und ASIM - Gestern, Heute, Morgen

Felix Breitenecker

Institute for Analysis and Scientific Computing, TU Wien, Wiedner Hauptstraße 8-10,
1040 Vienna, Austria; *felix.breitenecker@tuwien.ac.at

Die Systemsimulation hat sich seit ihrem Beginn in den siebziger Jahren des vorigen Jahrhunderts sehr in die Breite entwickelt, in Anwendungsgebieten, und im Bereich der Modellbildung und Methoden.

Shannons Definition für Simulation als “the process of designing a model of a real system or of evaluating various strategies for the operation of the system” erlaubt vielfältige methodische Entwicklungen und lädt zur Anwendung in verschiedensten Gebieten ein. Eine Übersicht über diese Entwicklungen sprengt sogar den Rahmen einer Monographie.

ASIM, die “Arbeitsgemeinschaft Simulation“, ein Fachausschuss der GI (Gesellschaft für Informatik), hat viele dieser Entwicklungen mehr oder weniger beachtet, couragiert oder notwendigerweise mitgemacht, teilweise beeinflusst, und sich mit der Entwicklung aber auch mit dem strukturellen Umfeld geändert.

Eine kurzweiligere Möglichkeit zur Darstellung dieser Entwicklungen ist die punktuelle Betrachtung von bestimmten meist *kontroversiellen Eigenschaften* der Entwicklung, die interessante Einsichten bietet.

Diese Eigenschaften sind z.B. allgemein - speziell, direkt – überraschend, erwartet - unerwartet, neu - wiederholt, klassisch – exotisch, notwendig – möglich, praktikabel – dogmatisch, linear – nichtlinear, zu früh – zu spät, Fluch oder Segen, erforscht – noch offen, numerisch – symbolisch, genau – nicht genau, einige nun kurz skizziert.

Die weißen Flecken: erforscht – noch offen

Ein oberflächlicher Blick auf den *Simulationskreislauf* erweckt den Eindruck, dass alle Schritte – von Modellbildung bis Validierung hinreichend methodisch erforscht sind. Ein genauerer Blick zeigt allerdings, dass die *Validierung* nur in einigen Anwendungsbereichen hinreichend abgedeckt ist. Ein nahezu *weißer Fleck* auf der Landkarte der Entwicklungen sind *Analysemethoden* – die derzeit kaum über die klassische (nichtlineare) Zeitbereichsanalyse hinausgehen.

An sich sprechen frühe methodische Ansätze vom Experiment als *Anwendung einer Analysemethode auf ein Modell*, heute verwenden wir fast ausschließlich die numerische Zeitbereichskeule.

Der Selbstbetrug: linear – nichtlinear

Ein Gral in der Systemsimulation: die Systemsimulation beherrscht *nichtlineare Systeme*. Die numerische Zeitbereichskeule meldet erfolgreich die tatsächliche Berechnung von nichtlinearen Zusammenhängen – und täuscht damit. Denn irgendwann landen die Algorithmen bei nichtlinearen Gleichungen, Newton hilft erfolgreich mit seinem Verfahren – durch *Linearisierung*. Spezialisten können dann überlegen, ob eine lokale a-priori - Linearisierung dasselbe Ergebnis liefert.

Faule Agenten: zu früh/spät, überraschend

Agentenbasierte Simulation hat sich zum anerkannten Werkzeug entwickelt. Heute wird ein *PCE Agent* als *Path Computation Element* identifiziert, doch frühe Arbeiten sprechen von *Perception-Cognition-Experience* – Agenten, mit eigener Intelligenz. Zu frühe Arbeiten, in Zeiten kleiner Rechnleistungen.

Bekannt ist der Zusammenhang zwischen agentenbasierten Modellen und PDE- bzw. ODE-Modellen im Bereich von Diffusionsprozessen. Die ODEs/PDEs beschreiben dann über direkte Ableitungen im Wesentlichen Dichten von Agenten. Es reizt die Frage nach der umgekehrten Richtung: gibt es zu einer ODE ein agentenbasiertes Modell mit ebendiesem Grenzwertverhalten?

Segen oder Fluch: praktikabel - dogmatisch

Modellbildungsmethoden wie *Modelica*, *System Dynamics*, *DEVs* sind anerkannt und erleichtern die Systemsimulation wesentlich. Sie sind ein praktikables Werkzeug zur Erstellung standardisierter Modelle.

Die mathematisch-dogmatische Sicht lehnt diese Modellbildungsebene ab, aus mehreren Gründen. Die Methoden sind aber in der Tat ein Segen für rasche Entwicklung, aber der Fluch folgt nach. *System Dynamics* beinhaltet als selbsterfüllende Prophezeiung ein lineares Verhalten, und *Modelica* bzw. ähnliche Systeme verführen die Entwickler zu mangelnder Innovationsfreude – sehr ähnliche Produkte.

Big Data und KI: neu – neu aufgelegt

Big Data und KI sind der neue Gral in der Simulation. Ein detaillierter Blick in den Simulationskreislauf zeigt die Forderung zur Datenaufbereitung – die mit Big Data nun zum Hauptzweck wird. Und Einsatz von KI – an sich komplexe nur teilweise neue Analyseverfahren, verbunden mit hoher Rechenpower. Und dahinter die Frage nach einer Kausalität.

Migration, Archäologie, Poesie: klassisch – exotisch, notwendig, möglich

Natur- und Ingenieurwissenschaften sind seit Beginn klassische Domänen der Systemsimulation, Biologie und Medizin folgten bald, ebenso Ökonomie, etc. Systemsimulation wurde zu einem notwendigen Werkzeug.

Aber Systemsimulation hat ihre Meriten auch in exotischen Anwendungen, wie *Archäologie* und *Poesie*, und *Migration* – aber wozu ?

Und ASIM

Wie eingangs erwähnt, hat ASIM diese Entwicklungen mehr oder weniger beachtet, couragiert oder notwendigerweise mitgemacht, teilweise beeinflusst, und sich mit der Entwicklung aber auch mit dem strukturellen Umfeld geändert.

CV Felix Breitenecker

Felix Breitenecker studied ‘Applied Mathematics’ at TU Wien. After a PhD Thesis on Optimal Control he achieved a habilitation on ‘Mathematics of Control and Simulation’ and got a position as assistant professor at TU Wien. After guest professor positions at University Glasgow, at University Budapest, at University Ljubljana and at other universities, he acts since 1992 as professor for Mathematical Modelling and Simulation at TU Wien (Technical University Vienna).



He covers a broad research area, from mathematical modelling to simulator development, from numerical mathematics to symbolic computation, from biomedical and mechanical simulation to process simulation, and he also deals with ‘unusual’ simulation areas, like archaeology and poetry. In teaching area, he is organizing e-learning development for basic mathematics, basic simulation, and simulation applications.

Felix Breitenecker is active in various simulation societies: president and past president of EUROSIM since 1992, and member of INFORMS, SCS and others.

Since 1984 he is board member of ASIM, the German simulation society, his main society activity – he served in many positions for ASIM.

He was engaged in many projects in the area of modelling and simulation, partners being industry and research institutions. Felix Breitenecker has published about 400 scientific publications, and he is author of three books and editor of 25 books (proceedings and monographs).

Since 1995 he is Editor in Chief of EUROSIM’s and ASIM’s scientific journal SNE Simulation Notes Europe.

Felix Breitenecker believes that good research requires good education, and so he encouraged many students for a thesis in modelling and simulation, resulting in up to now 38 PhD theses, and 175 master theses.

Mathematical model for spatial and temporal risk analysis of West Nile Virus Transmission in Bavaria, Germany.

Oliver Chinonso Mbaoma^{1*}, Stephanie Margarete Thomas^{1,2}, Carl Beierkuhnlein^{2,3}

¹Department of Biogeography, University of Bayreuth, Universitaetsstr. 30, 95447 Bayreuth, Germany;
**oliver.mbaoma@uni-bayreuth.de*

²Bayreuth Center of Ecology and Environmental Research, BayCEER, University of Bayreuth, Universitaetsstr. 30, 95447 Bayreuth, Germany

³Geographical Institute of the University of Bayreuth, GIB, Universitaetsstr. 30, 95447 Bayreuth, Germany

Abstract. The spatial and temporal trends of mosquito-borne diseases have been shifting. Recently, there has been an increase in cases of West Nile Fever across Germany. In this study, we investigated the spatial and temporal risk of West Nile virus transmission in Bavaria using a mathematical model. The model incorporated epidemiological, climatic, and bird distribution data. It generated spatially and temporally explicit graphs depicting the health states of mosquitoes and birds. Additionally, risk maps identify areas in Bavaria at heightened risk for West Nile virus transmission.

Introduction

West Nile fever is an illness triggered by West Nile virus (WNV), a single strand RNA arbovirus that from the Flaviviridae family and one of the most widely spread flaviviruses [1]. In 1937, the first case was isolated at West Nile district in Uganda and has since been occurred in all continents [2]. Although WNV normally circulates in a sylvatic cycle between mosquitoes and birds as hosts, mosquitoes can also spread the virus to people and other animals, including horses [2],[3]. Most of these animals and humans only experience mild viremia and are unable to serve as hosts for the virus [4]. In 2018, the virus was detected in 12 birds and two horses in eastern Germany [5]. The first human autochthonous case of the infection was detected in 2019 near Leipzig [6]. Favorable climatic conditions with high temperatures in summer promoted the spread of the virus. While mosquitoes tend to spread the virus locally, migratory birds are suspected to play a role in disseminating the virus over long distances to new areas [5]. With information on bird surveillance for

WNV, competence of vector mosquitoes and amplification suitability of hosts, we were able to develop a climate-driven process-based mechanistic model calibrated with functional traits of vector mosquito and host birds to predict the spatial and temporal risk of WNV infection in in Bavaria.

1 Materials and Method

For this research, we adopted a similar approach applied to replicate the spatial and temporal pattern of WNV infection in Germany between 2018 and 2022 by Mbaoma et al, [7].

1.1 Materials

Climate and bird data were used to parameterize and force the model adequately. Cases of WNV infection were collected from the Animal Diseases Information System database of the FLI from 2018 to 2022 [5]. Bird distribution data was collected from E-bird online database [8]. Temperature, rainfall and humidity data used to drive our model were collected from E-OBS database [9].

1.2 Model Description

With a clear understanding of the entomology, life characteristics, and functional ecology of both the WNV vector and its hosts, we used the mechanistic approach which was developed by Mbaoma et al., 2024 [7] to generate spatial and temporal predictions of potential WNV infection outbreak in Bavaria.

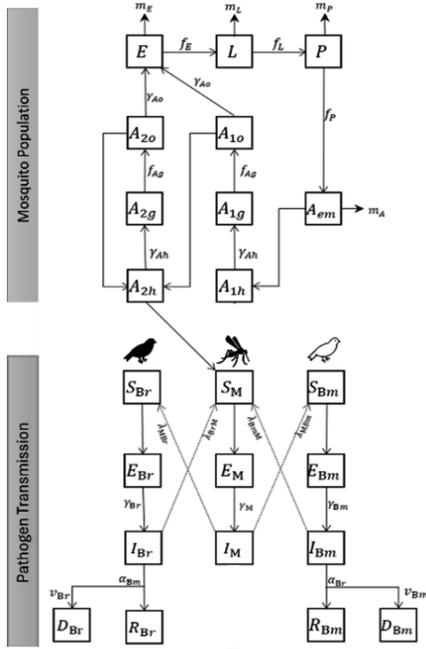


Figure 1: Flowchart of WNV epidemic model showing two sections depicting mosquito population and pathogen transmission between mosquito, resident birds and migratory birds.

Mathematical epidemiology and population biology were used to formulate the fundamentals of the compartmental model driven by several ordinary differential equations. The model consists of two sections: the first describing mosquito population while the second described pathogen transmission.

As described in the paper by Mbaoma et al. [7], disease-free mosquito population was explained by the following ODE.

$$\begin{cases}
 \dot{E} = d\gamma_{A_0}(\beta_1 A_{1_0} + \beta_2 A_{2_0}) - (\mu_E + df_E)E \\
 \dot{L} = df_E E - (m_L(1 + L/k_L) + f_L)L \\
 \dot{P} = f_L L - (m_P + f_P)P \\
 \dot{A}_{em} = f_P P \sigma e \left(-\mu_{em}(1 + P/k_P) \right) - (m_A + \gamma_{Aem})A_{em} \\
 \dot{A}_{1h} = \gamma_{Aem} A_{1h} - (m_A + \mu_r + \gamma_{Ah})A_{1h} \\
 \dot{A}_{1g} = \gamma_{Ah} A_{1h} - (m_A + \mu_{pr} + f_{Ag})A_{1g} \\
 \dot{A}_{1_0} = f_{Ag} A_{1g} - (m_A + \mu_r + f_{A_0})A_{1_0} \\
 \dot{A}_{2h} = f_{A_0}(A_{1_0} + A_{2_0}) - (m_A + \mu_r + \gamma_{Ah})A_{2h} \\
 \dot{A}_{2g} = \gamma_{Ah} A_{2h} - (m_A + \mu_{pr} + f_{Ag})A_{2g} \\
 \dot{A}_{2_0} = f_{Ag} A_{2g} - (m_A + \mu_r + f_{A_0})A_{2_0}
 \end{cases} \quad (1)$$

The process of infection between mosquito and bird population is explained by the following sets of Ordinary differential equations, with Equation 2,3 and 4 explaining health states of mosquitoes, migratory birds and

residential birds respectively.

$$\begin{aligned}
 \frac{dS_M}{dt} &= (-\lambda_{BrM}(T) + \lambda_{BmM}(T))S_M + A_{2h} - m_A(T)S_M \\
 \frac{dE_M}{dt} &= (\lambda_{BrM}(T) + \lambda_{BmM}(T))S_M - \gamma_M(T)E_M - m_A(T)E_M \\
 \frac{dI_M}{dt} &= \gamma_M(T)E_M - m_A(T)I_M
 \end{aligned} \quad (2)$$

$$\begin{aligned}
 \frac{dS_{Br}}{dt} &= \left(b_{Br} - (b_{Br} - m_{Br}) \frac{N_{Br}}{K_{Br}} \right) N_{Br} - \lambda_{MBr}(T)S_{Br} - m_{Br}S_{Br} \\
 \frac{dE_{Br}}{dt} &= \lambda_{MBr}(T)S_{Br} - \gamma_{Br}E_{Br} - m_{Br}E_{Br} \\
 \frac{dI_{Br}}{dt} &= \gamma_{Br}E_{Br} - \alpha_{Br}I_{Br} - m_{Br}I_{Br} \\
 \frac{dR_{Br}}{dt} &= (1 - v_{Br})\alpha_{Br}I_{Br} - m_{Br}R_{Br} \\
 \frac{dD_{Br}}{dt} &= v_{Br}\alpha_{Br}I_{Br}
 \end{aligned} \quad (3)$$

$$\begin{aligned}
 \frac{dS_{Bm}}{dt} &= \left(b_{Bm} - (b_{Bm} - m_{Bm}) \frac{N_{Bm}}{K_{Bm}} \right) N_{Bm} - \lambda_{MBm}(T)S_{Bm} - m_{Bm}S_{Bm} \\
 \frac{dE_{Bm}}{dt} &= \lambda_{MBm}(T)S_{Bm} - \gamma_{Bm}E_{Bm} - m_{Bm}E_{Bm} \\
 \frac{dI_{Bm}}{dt} &= \gamma_{Bm}E_{Bm} - \alpha_{Bm}I_{Bm} - m_{Bm}I_{Bm} \\
 \frac{dR_{Bm}}{dt} &= (1 - v_{Bm})\alpha_{Bm}I_{Bm} - m_{Bm}R_{Bm} \\
 \frac{dD_{Bm}}{dt} &= v_{Bm}\alpha_{Bm}I_{Bm}
 \end{aligned} \quad (4)$$

Cross infection between mosquitoes and amplifying hosts which are birds of interest was defined in Equations 5 to 8.

$$\lambda_{MBr}(T) = \delta_M F_b k(T) p_{MBr} \phi_{Br} \frac{I_M}{K_M} \quad (5)$$

$$\lambda_{MBm}(T) = \delta_M F_b k(T) p_{MBm} \phi_{Bm} \frac{I_M}{K_M} \quad (6)$$

$$\lambda_{BrM}(T) = \delta_M F_b k(T) p_{BrM} \frac{I_{Br}}{K_{Br}} \quad (7)$$

$$\lambda_{BmM}(T) = \delta_M F_b k(T) p_{BmM} \frac{I_{Bm}}{K_{Bm}} \quad (8)$$

Basic reproductive number was computed based on the next-generation matrix approach applied in the precursor model by Mbaoma et al [12] using the equation below:

$$R_0 = \sqrt{\frac{\left[\frac{\gamma_M(T)\beta_{Mbr}(T) + \beta_{MBm}(T)}{(\gamma_M(T) + m_M(T))m_M(T)} \frac{S_{Br} + S_{Bm}}{K_{Br} + K_{Bm}} \right] \left[\frac{\gamma_{Br}\beta_{Br}(T)}{(\gamma_{Br} + m_{Br})(\alpha_{Br} + m_{Br})K_{Br}} \frac{S_{Br}}{K_{Br}} \right] + \left[\frac{\gamma_{Bm}\beta_{Bm}(T)}{(\gamma_{Bm} + m_{Bm})(\alpha_{Bm} + m_{Bm})K_{Bm}} \frac{S_{Bm}}{K_{Bm}} \right]}{\left[\frac{\gamma_{Bm}\beta_{Bm}(T)}{(\gamma_{Bm} + m_{Bm})(\alpha_{Bm} + m_{Bm})K_{Bm}} \frac{S_{Bm}}{K_{Bm}} \right]}}$$

1.3 State Variables, Parameters and Functions.

All state variables, parameters and functions used to develop this model, including calibration and validation processes have been described in a precursor publication for WNV epidemic model by Mbaoma et al. [7].

2 Results and Discussion

2.1 WNV Transmission Risk Projection for Bavaria

The inverse calibrated WNV epidemic model generated spatial and temporal abundance of mosquito. It also generated different health states for hosts of interests which include residential birds and migratory birds. Due to limited occurrence recorded in Bavaria (just one case), we assumed our model was performing at optimum having been previously validated across Germany [7]. Functional traits of mosquito which includes fecundity, development rates, mortality rates, biting rates and extrinsic incubation rates of WNV pathogen all responded to deviations in temperature [10],[11]. The model was able to simulate various health states of mosquito and birds, with seasonal variation. Two compartments of interest from the mosquito population health state are shown (Fig. 2). Also, we were able to simulate risk maps of WNV transmission for Bavaria between 2017 and 2022 at NUTS3 levels, identifying potential hotspots for WNV infection outbreaks (Fig. 3).

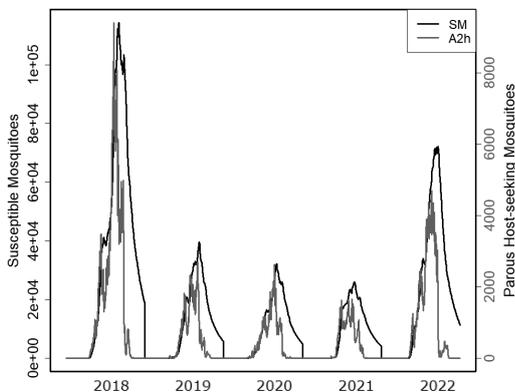


Figure 2: Simulated time series of health states for susceptible (S_M) and parous host-seeking mosquitoes (A_{2h}) between 2018 and 2022.

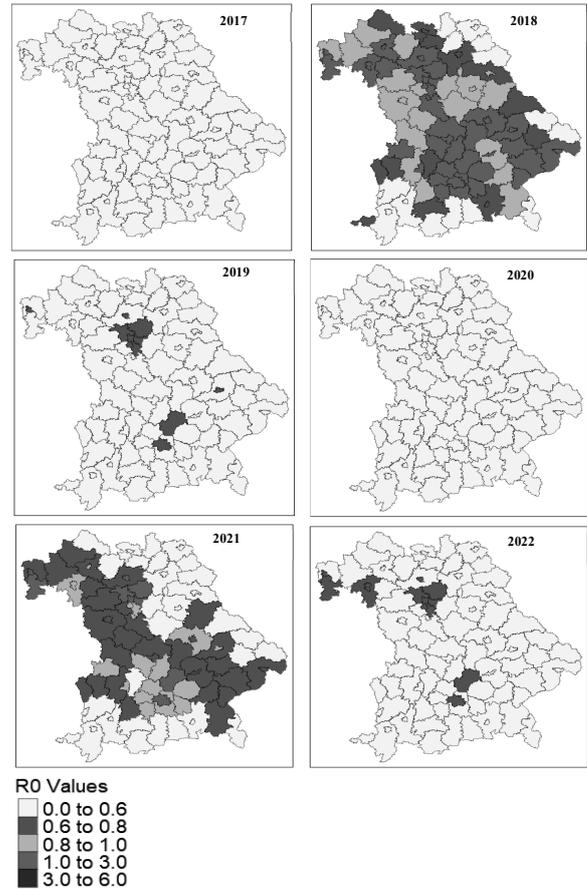


Figure 3: WNV transmission risk across Bavaria at NUTS3 level from 2017 to 2022 estimated from daily R0 values between August and October (week 30 to 42) when the peak of the WNV infections were reported by FLI.

2.2 Seasonal variation of WNV infection Risk

Several climate sensitive mosquito-borne diseases display seasonal variation [12]. In our model, several factors, including vector population, vectoral capacity, vector competence and amplifying host competence influenced R0 rates. Infection parameters, such as mosquito biting rate and extrinsic incubation rate were important drivers which was consistent with the role of environmental suitability in driving infectious disease transmission [10]. The model was able to simulate seasonal trend of R0 which represents WNV transmission risk (Fig. 4).

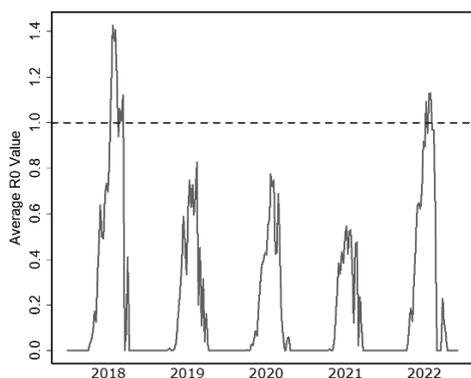


Figure 4: The risk of West Nile Virus (WNV) infection outbreaks, expressed as R_0 values, was assessed across Bavaria from 2018 to 2022. The red line represents the simulated daily R_0 values aggregated on a weekly basis, while the blue dashed horizontal line indicates the threshold at which infection transmission becomes active.

3 Conclusion

We developed a process based WNV epidemic model which was able to simulate areas prone to WNV transmission in Bavaria. From our results, it was obvious that WNV infection which is endemic in eastern Germany may likely spread to the South with several hotspots identified in Bavaria. Changing climate and migratory pattern of short distance migratory birds have been identified as key factors that could accelerate WNV infection spread. The model considered juvenile and adult stages of mosquito life cycle in several compartments, and two bird taxa with migratory and residential birds. Functional traits of mosquitoes were driven by climate. Host-feeding preference of vector, transmission probability and mosquito to host ratio were all accounted for in the model. In addition, spatial heterogeneity of climate forcings, vectors and host species were considered. We developed a model that could be deployed at the backend of a warning system for *Culex pipiens* mosquito population occurrence and present and future WNV transmission risk.

References

- [1] De Filette, M., Ulbert, S., Diamond, M.S. et al. Recent progress in West Nile virus diagnosis and vaccination. *Vet Res* 43, 16 (2012). [https://doi.org/10.1186/1297-](https://doi.org/10.1186/1297-9716-43-16)

- [9716-43-16](https://doi.org/10.1186/1297-9716-43-16).
- [2] Chancey, C., Grinev, A., Volkova, E., Rios, M. (2015). The global ecology and epidemiology of West Nile virus. *BioMed research international*. p. 376230. doi: 10.1155/2015/376230.
- [3] Holicki, C. M., Ziegler, U., Răileanu, C., Kampen, H., Werner, D., Schulz, J., Silaghi, C., Groschup, M. H., Vasić, A. (2020). West Nile Virus Lineage 2 Vector Competence of Indigenous *Culex* and *Aedes* Mosquitoes from Germany at Temperate Climate Conditions. *Viruses*. 2020 May 19;12(5):561. doi: 10.3390/v12050561
- [4] Bowen, R. A., & Nemeth, N. M. (2007). Experimental infections with West Nile virus. *Curr Opin Infect Dis*. Jun;20(3):293-7. doi: 10.1097/QCO.0b013e32816b5cad. PMID: 17471040
- [5] Animal Diseases Information System TSIS. Available at: <https://tsis.fli.de/Reports/Info.aspx> (Accessed: 16 February 2023).
- [6] Robert Koch Institut RKI (2019): https://www.rki.de/DE/Content/Infekt/EpidBull/Archiv/2019/Ausgaben/40_19.pdf?blob=publicationFile.
- [7] Mbaoma, O. C., Thomas, S. M., & Beierkuhnlein, C. (2024). Spatiotemporally Explicit Epidemic Model for West Nile Virus Outbreak in Germany: An Inversely Calibrated Approach. *Journal of epidemiology and global health* (accepted)
- [8] Sullivan, B. L., Wood, C. L., Iliff, M. J., Bonney, R. E., Fink, D., & Kelling, S. (2009). E-bird: A citizen-based bird observation network in the biological sciences. *Biological Conservation*, 142(10), 2282-2292. <https://doi.org/10.1016/j.biocon.2009.05.006>
- [9] Cornes, R. C., van der Schrier, G., van den Besselaar, E. J., & Jones, P. D. (2018). An ensemble version of the E-OBS temperature and precipitation data sets. *Journal of Geophysical Research: Atmospheres*, 123(17), 9391-9409.
- [10] Mordecai, E. A., Caldwell, J. M., Grossman, M. K., Lippi, C. A., Johnson, L. R., Neira, M., ... & Villena, O. (2019). Thermal biology of mosquito-borne disease. *Ecology letters*, 22(10), 1690-1708. doi: 10.1111/ele.13335
- [11] Shocket, M. S., Verwillow, A. B., Numazu, M. G., Slamani, H., Cohen, J. M., El Moustaid, F., Rohr, J., Johnson, L. R., & Mordecai, E. A. (2020). Transmission of West Nile and five other temperate mosquito-borne viruses peaks at temperatures between 23°C and 26°C. *eLife*, 9, e58511. <https://doi.org/10.7554/eLife.58511>
- [12] Tjaden, N. B., Caminade, C., Beierkuhnlein, C., & Thomas, S. M. (2018). Mosquito-borne diseases: advances in modelling climate-change impacts. *Trends in parasitology*, 34(3), 227-245. doi:10.1016/j.pt.2017.11.006.

Entwicklung einer Bewertungsmethode für Simulationsergebnisse im Kontext industrieller Symbiose: Ein Zielsystem zur Steigerung von Ressourceneffizienz und Resilienz

Björn Vollack¹, Manuel Boelcke¹, Thorsten Schmidt¹

¹Professur für Technische Logistik, TU Dresden, Münchner Platz 3, 01187 Dresden; technische.logistik@tu-dresden.de

Abstract. Die Bedeutung von Energie- und Ressourceneffizienz hat in den letzten Jahren stark zugenommen. Industrial Symbiosis (IS) steigert u. a. die Ressourceneffizienz mithilfe energetisch-stofflicher Austauschbeziehungen zwischen Unternehmen. Neben der Effizienz kommt auch der Resilienz eine immer stärkere Bedeutung zu. IS kann auch hier durch ein breites Netz an Alternativbeziehungen einen Beitrag zur Vermeidung von Produktionsunterbrechungen leisten. Die Austauschbeziehungen in einem Industriecluster können simuliert werden. Es ist aber eine standardisierte Methode zur Bewertung der Simulationsergebnisse notwendig, um unterschiedliche Varianten miteinander vergleichen zu können. In diesem Beitrag wird ein Zielsystem präsentiert, das ökonomische, ökologische, symbiotische und resiliente Aspekte betrachtet. Verschiedene Gewichtungen der Zielfunktionen ermöglichen es, individuelle Schwerpunkte zu setzen. Die Methode wird anhand eines Beispiels mit vier Unternehmen und verschiedenen Simulationsmodellvarianten demonstriert.

Einleitung

Die letzten Jahre haben die Bedeutung von Energie- und Ressourceneffizienz eindrucksvoll herausgestellt. Insbesondere in Industrieparks lassen sich durch die Anwendung von Industrial Symbiosis (IS) Potentiale heben. IS beschreibt dabei ein Konzept, welches durch den Austausch von Wasser, Energie, Abfällen und Nebenprodukten zwischen Unternehmen u. a. die Ressourceneffizienz steigert [1], [2].

Neben der Effizienz spielt für Unternehmen die Resilienz, d. h. die Widerstandsfähigkeit z. B. auf Ressourcenausfälle reagieren zu können, eine zunehmend

bedeutendere Rolle [3], [4]. Auch in dieser Hinsicht bietet IS Möglichkeiten, zum einen, weil nicht alle Material- und Energiebedarfe extern gedeckt werden müssen – Stichwort Autarkie –, zum anderen, weil ggf. durch die Vernetzung von Unternehmen Alternativen zur bisherigen Bezugsquelle bzw. dem bisherigen Abnehmer vorhanden sind.

Derzeit sind die Gestaltung und der Betrieb von Industrieparks häufig nicht auf diese internen Austauschprozesse ausgerichtet. Das Forschungsprojekt „Energieeffiziente Industrieclusteroptimierung“ (EnICO) hat das Ziel, die energetisch-stofflichen Austauschbeziehungen in einem Industriecluster simulationsbasiert zu optimieren. Durch Simulation werden gesamtoptimale Interaktionsbeziehungen unter Berücksichtigung verschiedener Zielfunktionen ermittelt.

Basis hierfür ist zunächst eine standardisierte Methode zur Bewertung der Simulationsergebnisse, die auch unterschiedliche Gewichtungspräferenzen im Hinblick auf die Zielfunktionen seitens der Anwendenden berücksichtigt. Dafür ist zunächst ein Zielsystem entwickelt worden, welches mehrere Zielfunktionen vereint und es ermöglicht, unterschiedliche Bewertungsschwerpunkte zu setzen.

1 Stand der Forschung

Ein Industriecluster ist die Ansammlung mehrerer benachbarter Unternehmen, die in Austauschbeziehungen stehen. Dies erfüllt die grundlegende Anforderung der räumlichen Nähe, die CHERTOW als wesentliche Voraussetzung für industrielle Symbiose (IS) benennt, auch wenn inzwischen zusätzlich unternehmensinterne und überregionale Symbiosen Gegenstand der Forschung

sind [1]. In der der Literatur werden solche Cluster auch als Eco Industrial Park (EIP) bezeichnet [1], [5]. Eines der bekanntesten Beispiele ist der EIP in Kalundborg, Dänemark [6].

MAIWALD ET AL. haben in der Simulationsumgebung SimulationX¹ eine Modellbibliothek entwickelt, die die dynamische Simulation eines EIP ermöglicht [7]. Eine automatisierte Bewertung bzw. Optimierung ist bisher nicht vorgesehen, wird von den Autoren aber als eine wichtige Weiterentwicklung ihrer Simulation angesehen.

In der Literatur existieren zahlreiche Kennzahlensysteme für EIPs. Diese werden häufig nach ökonomischen, ökologischen und sozialen Kennzahlenarten klassifiziert [8]. Eine wichtige Kennzahl zum Themengebiet Resilienz beschreiben VALENZUELA-VENEGAS ET AL. Dieser sogenannte Resilienzindikator besitzt den Wertebereich [0, 1], wobei der Wert 1 bestmöglicher und der Wert 0 keiner Resilienz entspricht. Er setzt sich aus dem Netzwerkkonnektivitätsindex NCI und der Flussersatzfähigkeit ϕ zusammen [9]:

$$\text{Resilienzindikator} = \frac{1}{2} * NCI + \frac{1}{2} * \phi \quad (1)$$

Der Netzwerkkonnektivitätsindex NCI des Industrieclusters betrachtet die Anzahl der Verbindungen. So kann ein Unternehmen bspw. nur über eine einzige Verbindung zu einem anderen Unternehmen verfügen oder aber zu mehreren anderen. Je mehr Verbindungen innerhalb des EIP bestehen, desto höher ist die Wahrscheinlichkeit, dass beim Ausfall eines Unternehmens zumindest eine andere Bezugs- bzw. Abnahmequelle existiert. Die Flussersatzfähigkeit ϕ bewertet für jedes Unternehmen, inwieweit bei dessen Ausfall die entfallenden Ressourcenflüsse kompensiert werden können. Denn ein solcher Ausfall sorgt dafür, dass das Unternehmen nicht mehr als Abnehmer seiner Eingangsprodukte und Emitent der Ausgangsprodukte zur Verfügung steht. Es wird für jedes Unternehmen der Anteil der nicht kompensierbaren Verluste bestimmt [9].

Da der Fokus des Forschungsprojektes weniger auf soziale Aspekte gerichtet ist, werden insbesondere ökonomische und ökologische Kennzahlen in die Bewertungsmethode überführt. Aufgrund seiner Bedeutung wird auch der Resilienzindikator explizit mit aufgenommen.

2 Beschreibung der Methode

Für die Einordnung von Simulationsergebnissen ist eine standardisierte Bewertung notwendig. Nachfolgend sind die diesem Beitrag zugrunde liegenden Schritte dargestellt:

1. Auswahl relevanter Kennzahlen und Überführung dieser in geeignete Zielfunktionen.
2. Gewichtung der Zielfunktionen unter verschiedenen Zielstellungen.
3. Berücksichtigung der Entscheiderpräferenz hinsichtlich der Gewichtung dieser Zielstellungen.

Die Auswahl der Kennzahlen erfolgt auf Basis einer Literaturrecherche und einem Abgleich mit den im Simulationsmodell zur Verfügung stehenden Informationen. Kennzahlen mit nur geringer Aussagekraft werden nicht berücksichtigt, um die Komplexität möglichst gering zu halten und in der Folge die Nachvollziehbarkeit zu erhöhen. Das Ergebnis ist ein Zielsystem mit sechs Zielfunktionen:

- Z_1 : Kosten der Verbindungen minimieren
- Z_2 : Externen Strombezug minimieren
- Z_3 : Externen Wärmebezug minimieren
- Z_4 : Externen Ressourcenbezug minimieren
- Z_5 : Anteil symbiotischer Beziehungen maximieren
- Z_6 : Resilienzindikator maximieren

Die Zielfunktion Z_1 berücksichtigt sowohl den Invest als auch die Betriebskosten einer symbiotischen Beziehung (z. B. Fernwärmeleitung, LKW-Transport), denn nicht jede theoretisch mögliche Verbindung ist ökonomisch auch sinnvoll. Die Zielfunktionen Z_2 bis Z_4 adressieren den Anspruch, dass möglichst alle innerhalb eines EIP anfallenden Strom-, Wärme-, Ressourcenmengen intern weitergenutzt werden sollen. Dem trägt auch Z_5 Rechnung, indem der Anteil symbiotischer Beziehungen zwischen einzelnen Unternehmen maximiert werden soll. Der Resilienzindikator in Z_6 orientiert sich an den Überlegungen von VALENZUELA-VENEGAS ET AL. und wurde an die SimulationX-Umgebung angepasst [9].

Im zweiten Schritt werden die Gewichtungen der Zielfunktionen für verschiedene Zielstellungen (ökonomisch, ökologisch, symbiotisch, resilient) mithilfe des Analytic Hierarchy Process (AHP) ermittelt [10]. Dabei werden die Zielfunktionen in paarweisen Vergleichen hinsichtlich ihrer Wichtigkeit für die einzelnen Zielstellungen gegenübergestellt. In Summe ergeben die Gewichtungen (Wertebereich [0, 1]) für jede Zielstellung 1.

¹ <https://www.esi-group.com/products/system-simulation>

Dieser Schritt wurde im Rahmen des Projekts beispielhaft mit Erkenntnissen aus der Literaturrecherche durchgeführt (vgl. Tabelle 1). In Zukunft ist die Einbeziehung von Unternehmen in die Bewertung geplant, um das Ergebnis zu validieren und ggf. anpassen zu können.

	ökonomisch	ökologisch	symbiotisch	resilient
Z ₁	0,091	0,042	0,043	0,039
Z ₂	0,187	0,172	0,098	0,093
Z ₃	0,187	0,196	0,098	0,093
Z ₄	0,442	0,456	0,098	0,093
Z ₅	0,058	0,076	0,374	0,230
Z ₆	0,035	0,058	0,289	0,452

Tabelle 1: Gewichtungen der Zielfunktionen unter den verschiedenen Zielstellungen

Schließlich wird mittels der Point-Allocation-Methode die Entscheiderpräferenz berücksichtigt, d. h. in der finalen Anwendung kann entschieden werden, welches Gewicht den einzelnen Zielstellungen zukommen soll [11]. Dazu werden 100 Punkte auf die vier Zielstellungen verteilt und somit ein Gesamtgewicht gemäß der Zielstellungspräferenz berechnet.

3 Anwendung der Methode

Das Vorgehen bei der Anwendung der Methode wird im Folgenden anhand eines einfachen und fiktiven Beispiels dargestellt. Dafür wird ein Industriecluster mit vier Unternehmen betrachtet: Ein Stahlproduzent (I1_STEE), ein Blockheizkraftwerk (I2_CHP), ein Papierhersteller (I3_PAMI) und ein Kunststoffhersteller (I4_CHEM). Jedes Unternehmen stellt ein Hauptprodukt her, bei dessen Produktion Nebenprodukte anfallen.

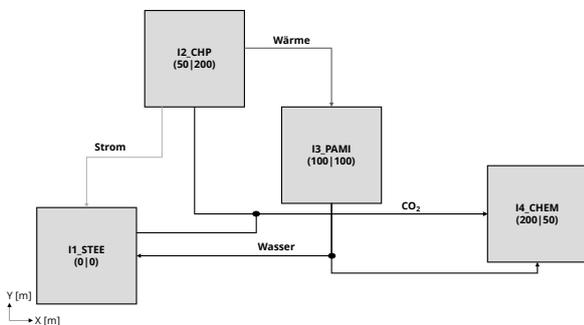


Abbildung 1: Schematische Darstellung der symbiotischen Austauschbeziehungen der Unternehmen (Variante 1 und 2)

Abbildung 1 zeigt eine schematische Darstellung der Austauschbeziehungen der Unternehmen. Die schwarzen Verbindungen stellen Materialverbindungen dar, rote Verbindungen Wärmeverbindungen und gelbe Verbindungen stehen für Stromverbindungen.

Es werden vier Simulationsmodellvarianten (SV) des Simulationsmodells erstellt, die sich durch die Art ihrer Austauschbeziehungen unterscheiden und mit der vorgestellten Methode bewertet werden:

- SV 1: Mehrere symbiotische Beziehungen für Material-, Wärme- und Stromflüsse.
- SV 2: Die symbiotischen Verbindungen sind analog zu Variante 1, jedoch ergänzt um die Betrachtung von Verlusten.
- SV 3: Alle benötigten Energie- und Materialmengen werden extern bezogen, es findet kein symbiotischer Austausch statt.
- SV 4: Grundsätzlicher Aufbau analog zu Variante 2, jedoch werden Wasser und Strom extern bezogen.

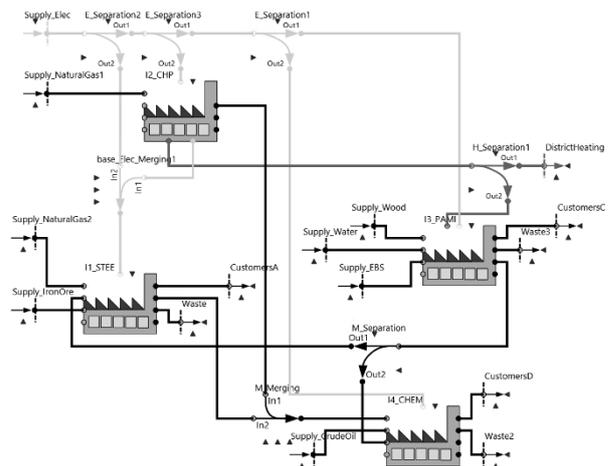


Abbildung 2: Vollständiges Simulationsmodell SV 1

In Abbildung 2 ist stellvertretend das vollständige Simulationsmodell SV 1 mit allen berücksichtigten Austauschbeziehungen dargestellt. Es wird die Simulationsbibliothek von MAIWALD ET AL. verwendet, die um einen Verlustbaustein, der u. a. die Berücksichtigung von Übertragungsverlusten bei Fernwärmeleitungen ermöglicht, erweitert wurde [7].

Jede Variante des Simulationsmodells wird mit der Methode bewertet. Genutzt werden die Gewichtung der Zielfunktionen gemäß Tabelle 1 und die folgende Entscheiderpräferenz: ökonomisch: 30; ökologisch: 30; symbiotisch: 25; resilient: 15. Diese fiktive Entscheiderpräferenz stellt eine ausgewogene Verteilung mit leichtem Fokus auf ökonomische und ökologische

Zielstellungen dar.

Die in Tabelle 2 dargestellten Gesamtnutzwerte ergeben folgende Rangfolge der Simulationsvarianten: $SV 1 > SV 2 > SV 3 > SV 4$. Dies bedeutet, dass der Gesamtnutzwert von Variante SV 1 den Gesamtnutzwert aller anderen Varianten übertrifft und mit der gewählten Gewichtung und Entscheiderpräferenz die gesamt optimale Lösung darstellt. Zu beachten ist, dass die geringen Unterschiede zwischen SV 1 und SV 2 lediglich auf die Berücksichtigung von Übertragungsverlusten zurückzuführen sind. Diese sind im Beispielmmodell sehr gering und deshalb erst in der sechsten Nachkommastelle ausweisbar. Hierbei handelt es sich eher um eine theoretische Unterscheidung mit wenig praktischer Relevanz, da in der Realität auf jeden Fall mit Verlusten zu rechnen ist. Insgesamt ist erkennbar, dass die Simulationsvarianten SV 3 und SV 4 deutlich schlechter bewertet sind. Dies liegt daran, dass hier das Symbiosepotential nicht vollständig ausgeschöpft wird.

SV 1	SV 2	SV 3	SV 4
0,211665	0,211662	-0,522877	-0,019082

Tabelle 2: Gesamtnutzwerte der Simulationsmodellvarianten (SV) der Szenarien

4 Zusammenfassung

Die hier vorgestellte Methode zur Bewertung von Simulationsergebnissen im Kontext industrieller Symbiose ermöglicht es, Simulationsergebnisse systematisch auszuwerten und zu vergleichen. Dabei kann der Anwender je nach Präferenz zwischen verschiedenen Zielstellungen gewichten. Eine Anwendung zur Bewertung ausgewählter manuell erstellter Simulationsmodellvarianten wie in Abschnitt 3 ist somit möglich. Eine besondere Stärke der Methode ist aber auch, dass sie zur Bewertung von vielzähligen automatisiert erstellten Simulationsmodellvarianten eingesetzt werden kann und eine Nutzung im Rahmen einer automatisierten Optimierung möglich ist.

In einem nächsten Schritt sind die Gewichtungen des AHP (vgl. Abschnitt 2) zu prüfen. Hier bietet es sich an, die Paarvergleiche z. B. durch Unternehmensvertreter durchführen zu lassen, um für einen möglichen Praxiseinsatz valide Gewichtungen zu erhalten. Anschließend kann die Methode im Rahmen einer praxisorientierten simulativen Optimierung eingesetzt werden.

5 Förderhinweis

Dieses Forschungsvorhaben wurde durch die Deutsche Forschungsgemeinschaft mit einer Laufzeit von November 2020 bis Oktober 2023 gefördert (GEPRIS-Nr.: 439187891; SCHM 2689/17-1; RO 2126/8-1).



Literatur

- [1] Chertow, M. R., Industrial symbiosis. Literature and taxonomy. *Annual Review of Energy and the Environment*, 2000, 25, 313–337.
- [2] VDI Zentrum Ressourceneffizienz. *Industrielle Symbiose*. <https://www.ressource-deutschland.de/themen/kreislaufwirtschaft/kreislauffuehrung-im-verarbeitenden-gewerbe/unternehmenskooperationen/>, abgerufen am 07.06.2024.
- [3] Gabler Wirtschaftslexikon. *Resilienz*. <https://wirtschaftslexikon.gabler.de/definition/resilienz-52429/version-275567>, abgerufen am 07.06.2024.
- [4] Kleemann, F. C., Frühbeis, R., Resiliente Lieferketten in der VUCA-Welt. *Supply Chain Management für Corona*, Brexit & Co., 2021, Springer Gabler, Wiesbaden.
- [5] Martin, M., Svensson, N., Eklund, M., Who gets the benefits? An approach for assessing the environmental performance of industrial symbiosis. *Journal of Cleaner Production*, 2015, Vol. 98, 263–271.
- [6] N. B. Jacobsen, N. B., Industrial Symbiosis in Kalundborg, Denmark. A Quantitative Assessment of Economic and Environmental Aspects. *Journal of Industrial Ecology*, 2006, Vol. 10, No. 1-2, S. 239-255.
- [7] Maiwald, M., Kosmol, L., Pieper, C., Schmidt, T., ES-ProNet: A Model Library for the Dynamic Simulation of Industrial Symbiosis. *International Journal of Modeling and Optimization*, 2020, Vol. 10, No. 1, S. 1-7.
- [8] Fraccascia, L., Giannoccaro, I., What, where, and how measuring industrial symbiosis: A reasoned taxonomy of relevant indicators. *Resources, Conservation & Recycling*, 2020, Vol. 157, 1-11.
- [9] Valenzuela-Venegas, G., Henríquez-Henríquez, F., Boix, M., Montastruc, L., Arenas-Araya, F., Miranda-Pérez, J., Díaz-Alvarado, F. A., A resilience indicator for Eco-Industrial Parks. *Journal of Cleaner Production*, 2018, 174, 807–820.
- [10] Saaty, T. L., Decision making - the Analytic Hierarchy and Network Processes (AHP/ANP). *Journal of Systems Science and Systems Engineering*, 2004, 13 (1), 1-35.
- [11] Khan, J. A., Ur Rehman, I., Hayat Khan, Y., Javed Khan, I., Rashid, S., Comparison of Requirement Prioritization Techniques to Find Best Prioritization Technique. *International Journal of Modern Education and Computer Science*, 2015, 7 (11), 53–59.

Entwicklung einer verteilten Simulationsinfrastruktur auf einem Raspberry Pi-Cluster mit Kubernetes und KNIME

Maurice Großmann^{1*}, Madlene Leißau¹, Christoph Laroque¹

¹Forschungsgruppe Industry Analytics, Westsächsische Hochschule Zwickau, Kornmarkt 1, 08056 Zwickau, Deutschland; *maurice.grossmann@fh-zwickau.de

Abstract. The development of simulation-based decision support systems requires large amounts of experimental data, which are generally not sufficiently available in companies. Adequate design of experiments with large-scale simulation experiments can take into account many factors in the subsequent decision making process. At the same time, the number of necessary simulation runs increases. In order to support the process of experiment planning and distributed simulation execution, the authors have developed a low-code-based infrastructure in KNIME that distributes and evaluates the simulation runs for computation in a Kubernetes cluster. This should help to reduce the time required for experiments and make the systems more applicable.

1 Motivation

Die Visionen von Industrie 4.0 und Industrie 5.0 stellen die Produktionsplanung und -steuerung vor neue Herausforderungen. Wirtschaftliche Trends führen dazu, dass Methoden und Werkzeuge der diskreten ereignisorientierten Simulation (DES) und simulationsgestützten Optimierung immer wichtiger werden.

Sie können daher nach Vieira et al. [1] als Schlüsseltechnologie im Zeitalter von Industrie 4.0 angesehen werden. DES-Modelle dienen bspw. nicht nur der Absicherung konkreter Produktionspläne und/oder Produktmixe, sondern ermöglichen auch umfangreiche Simulationsexperimente zur Vorhersage und Bewertung von Szenarien und der Generierung synthetischer Daten zur Entscheidungsunterstützung. Voraussetzung dafür sind DES-Modelle, die das reale System mit seinen dynamischen Prozessen möglichst gut abbilden, wobei der Detaillierungsgrad der Modellierung von der Zielsetzung abhängt.

Die Erstellung von DES-Modellen und die

nachfolgende Durchführung umfangreicher Simulationsexperimente stellen hohe Anforderungen in Bezug auf Zeit, (Experten-)Wissen und technische Ressourcen, wodurch Methoden und Vorgehensweisen wie das Data Farming zur automatisierten Verteilung und Durchführung von Simulationsexperimenten immer weiter in den Vordergrund rücken (bspw. [2][3]).

Im vorliegenden Beitrag dokumentieren die Autoren eine erstellte, automatisierte Simulationsinfrastruktur auf einem Raspberry Pi-Cluster als Low-Cost Alternative für kleinere Anwendungsszenarien gegenüber dedizierten Hochleistungs-Rechenclustern vor. Nach einer kurzen Darstellung themennaher Arbeiten soll der Fokus des vorliegenden Beitrags auf der Dokumentation der Implementierung dieses Clusters liegen und dessen Leistungsfähigkeit untersuchen. Abschließend fassen die Autoren wesentliche Erkenntnisse zusammen und geben einen Ausblick auf zukünftige Arbeiten.

2 Data Farming im Kontext der verteilten Simulation

In der Literatur finden sich zahlreiche themennahe Arbeiten, die sich mit Methoden und Vorgehensweisen wie Data Farming und der automatisierten Verteilung und Durchführung von Simulationsexperimenten beschäftigen. Data Farming zielt darauf ab, große Mengen an Simulationsdaten systematisch zu erzeugen und zu analysieren, um tiefere Einblicke in komplexe Systeme zu gewinnen [4]. Dabei wird angestrebt, auf einen möglichst umfassenden und qualitativ ausreichenden Datenbestand zurückzugreifen [5].

Król et al. [6] benennen genauso wie Sanchez et al. [4] und Pappert et al. [7] einen steigenden Bedarf an Rechenressourcen und Zeit und stellen dies in den

Zusammenhang mit einer steigenden Komplexität der zu betrachtenden Entscheidungs- und Planungsprobleme und somit der Größe und Detailtiefe der Simulationsmodelle. In Król et al. [6] können zudem die lange Zeit vorherrschenden Problemstellungen nachvollzogen werden, wie bspw. eine begrenzte Softwareunterstützung über alle Phasen des Data Farming hinweg und eine eingeschränkte Flexibilität und Skalierbarkeit in Bezug auf einmal gestarteter Faktorenkonfigurationen.

Voraussetzung für die effektive Durchführung von Data Farming ist eine leistungsfähige IT-Infrastruktur und die Verteilung von Simulationen auf mehrere Rechenknoten [2][6][7]. Anknüpfend daran beschreiben Pappert et al. [7] die Entwicklung einer stabilen und skalierbaren verteilten Infrastruktur zur Durchführung und Beschleunigung groß angelegter Simulationsexperimente basierend auf einem Kubernetes-Cluster.

Wenngleich im Ergebnis einer solchen skalierbaren Infrastruktur auf einem dedizierten Hochleistungs-Rechencluster eine beachtliche Anzahl an Faktorkonfigurationen von mehreren Tausend pro Tag erreicht werden kann [7], ist ein solches Rechencluster mit entsprechenden Kosten verbunden und steht nur in wenigen Fällen zur Verfügung. Das in diesem Beitrag beschriebene Raspberry Pi-Cluster soll und kann keine vergleichbaren Leistungszahlen erreichen, stellt aber eine günstige Alternative für kleinere Anwendungsszenarien dar.

3 Entwicklung einer verteilten Simulationsinfrastruktur

3.1 Aufbau des Raspberry Pi-Clusters

Für die Entwicklung der verteilten Simulationsinfrastruktur standen mehrere Kubernetes-Distributionen zur Auswahl. Aufgrund der ARMv8-Architektur der Raspberry Pi und der einfachen Installation entschieden sich die Autoren für Microk8s und gegen Alternativen wie K3s oder Minikube. Kubernetes ist eine Software zur Orchestrierung von Containern und erweitert die Funktionen der Containerisierungsplattform Docker. Die Microk8s-Distribution eignet sich besonders für Laptop- und Office-Workstations aufgrund ihrer geringen Größe und funktioniert problemlos auf einem Raspberry Pi.

Das Kubernetes-Cluster selbst ist ein Verbund einer undefinierten Menge an Rechensystemen. Meist besteht ein solches Cluster aus einer Master-Node, die alle Vorgänge zentral steuert, und einer unbegrenzten Anzahl an

Worker-Nodes, die die ihnen zugeteilten Arbeitsaufträge ausführen. Diese Arbeitsaufträge bestehen aus der kleinsten verteilbaren Einheit innerhalb eines Clusters, einem Pod. Ein Pod kann aus einem oder mehreren Docker-Containern bestehen. In Kubernetes werden immer komplette Pods und nicht einzelne Container angesteuert. Dabei können auch Replikationen von Pods und den darin befindlichen Containern skaliert werden. Replikationen sind Pods mit identischen Parameterkonfigurationen, deren Ergebnisse zur Erstellung von Statistiken genutzt werden. Ein Docker-Container stellt die beschreibbare Hülle für ein Docker-Image dar und bildet eine Abgrenzung zu anderen Containern.

Im Raspberry Pi-Cluster (siehe Abbildung 1) haben sich die Autoren für den Master-Node für ein ASUS ROG Flow Z13 mit einem Intel® Core™ i9-12900H Prozessor und 16 GB Arbeitsspeicher entschieden. Der Hauptgrund für die Entscheidung war die anvisierte Automatisierung des Verteilungs- und Auswertungsprozesses mit der Open-Source Entwicklungsplattform KNIME. KNIME als Entwicklungsumgebung bietet die Möglichkeit, wahlweise vollständig in einer Low- beziehungsweise No-Code-Umgebung und/oder einer bevorzugten Programmiersprache wie Python oder R zu arbeiten. Für die Implementierung des in diesem Beitrag beschriebenen Lösungsansatzes bedienen sich die Autoren an dem umfangreichen Spektrum zur Verfügung stehender Bausteine (Nodes) und ergänzen die bestehenden Funktionalitäten punktuell durch Python-Skripte.

Als Worker-Nodes dienen schließlich insgesamt 13 Raspberry Pi 4 mit einem 64-bit Quad Core ARM v8 Cortex-A72 Prozessor mit 1,5GHz, 4 GB LPDDR4 Arbeitsspeicher und einer 32 GB Micro SD als Festplattenspeicher. Des Weiteren ist ein zusätzlicher Raspberry Pi 4 für das Schreiben der benötigten Docker-Images reserviert. Alle Komponenten des Clusters sind über einen Switch im Intranet verbunden. Als Betriebssystem wird auf allen Geräten Ubuntu 22.04.3 verwendet. Während die Worker-Nodes ohne GUI auskommen, wurde auf dem Master-Node Ubuntu Desktop installiert, um die Vorteile der in KNIME aufgebauten Benutzeroberfläche zu nutzen und die gesamte Infrastruktur für die Anwender leichter zugänglich zu machen.

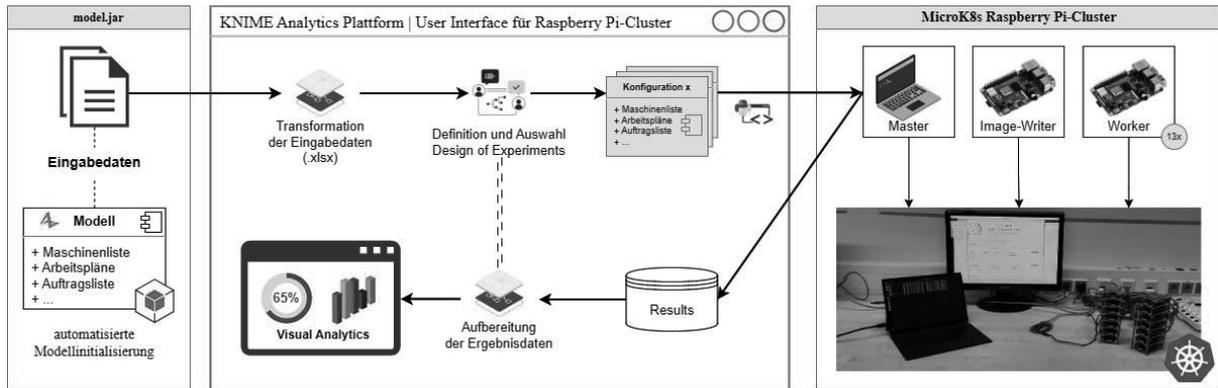


Abbildung 1: Konzeption der automatisierten Verteilungsstruktur

Nach der Installation des Clusters wurden die Raspberry Pi, die als Worker-Nodes deklariert wurden, dem Master-Node zugewiesen. Die Verwaltung des Clusters erfolgt über Terminalbefehle. Da die CPU-Architektur der Master-Node auf CISC (Complex Instruction Set Architecture) basiert, muss diese „getaint“ werden. CISC-Prozessoren haben einen umfangreichen Befehlssatz und werden hauptsächlich in Computern und Servern verwendet. Durch den „Taint“ können Regeln zur Verteilung auf dem Cluster festgelegt werden, um zu verhindern, dass Pods auf der Master-Node gestartet werden. Dies ist notwendig, da ein auf einem CISC-Prozessor erstelltes Image nicht auf einem RISC-Prozessor (Reduced Instruction Set Architecture) wie dem Raspberry Pi ausgeführt werden kann. RISC-Prozessoren haben einen einfacheren Aufbau, verwenden einen reduzierten Befehlssatz und werden vorwiegend in mobilen Endgeräten eingesetzt.

3.2 Automatisierung mit KNIME

Nachdem die Autoren bereits in früheren Arbeiten [8] die prinzipielle Machbarkeit von automatisierten Simulationsexperimenten im Zusammenhang mit KNIME gezeigt haben, soll im Folgenden auch die Verteilung auf einem Raspberry Pi Cluster anhand eines Workflows in KNIME realisiert werden.

Im Rahmen der angestrebten Automatisierung und systemübergreifenden Anwendbarkeit wurde der Workflow in KNIME so entwickelt, dass er ohne Anpassungen auf verschiedenen Betriebssystemen ausgeführt werden kann. Dies wurde durch eine intuitive grafische Benutzeroberfläche erreicht.

Anwender wählen zunächst das Simulationsmodell (als .jar) aus, wodurch alle notwendigen Dateien automatisch in den Workflow-Ordner importiert werden. Im nächsten Schritt können der Versuchsraum definiert und

das Experimentdesign ausgewählt werden, wobei derzeit ein vollfaktorieller Versuchsplan und ein Latin-Hypercube-Design (LHS) zur Verfügung stehen. Nach Vorschau und Bestätigung werden alle Faktorkonfigurationen automatisch im KNIME-Workflow als Datei geschrieben und bereitgestellt.

Das definierte Experiment kann dann gestartet werden, wobei zusätzliche Einstellungen wie die Anzahl der Replikationen vorgenommen werden können. Zunächst werden alle Konfigurationen aufgelistet, dann wird je Konfiguration ein Image erstellt und als Pod auf dem Cluster verteilt und repliziert. Über eine Terminalabfrage mittels Python liest der KNIME-Workflow das Protokoll der Pods aus; nach Abschluss einer Simulation werden die Daten in einer .xlsx-Datei innerhalb des Containers gesichert und auf die Master-Node kopiert. Der Workflow kopiert alle Ergebnisdateien aus den aktiven Pods und löscht anschließend alle verwendeten Images. Dieser Vorgang wird wiederholt, bis alle Konfigurationen durchlaufen sind.

Nach den Simulationen können die Ergebnisse visualisiert werden, um Rückschlüsse für die weitere Versuchsplanung zu ziehen.

4 Leistungsfähigkeit der Infrastruktur

Zum Vergleich und zur Bewertung der Leistungsfähigkeit des Raspberry Pi-Clusters wurde die automatisierte Simulationsinfrastruktur auch auf einem Server-Cluster getestet. Dieses bestand aus zwei virtuellen Maschinen mit identischen Betriebssystemen wie das Raspberry Pi-Cluster. Jede virtuelle Maschine war mit einem Intel Xeon Platinum 8362 Prozessor mit 2,80 GHz, 8 GB RAM und 128 GB Festplattenspeicher ausgestattet. Aufgrund der CISC-Architektur des Prozessors mussten die

Images neu erstellt werden und konnten nicht vom Raspberry Pi-Cluster verwendet werden. Eine VPN-Verbindung war erforderlich, um eine Verbindung zum Server herzustellen. In der Automatisierung wurden geringfügige Anpassungen am Python-Skript und am Workflow-Node zur Datenübertragung vorgenommen. Die Master-Node blieb gegenüber dem Raspberry Pi-Cluster unverändert. Dadurch konnte sichergestellt werden, dass der KNIME-Workflow weiterhin problemlos ausgeführt werden kann.

Auf beiden Clustern wurde die gleiche Experimentkonfiguration mit einem LHS-Design von 150 Versuchen und 13 Replikationen pro Versuch ausgeführt, was insgesamt 1950 Versuchsdurchläufe ergab. Die Raspberry Pis konnten nur eine Simulation gleichzeitig ausführen. Der Servercluster benötigte 12 Stunden und 36 Minuten, der Raspberry Pi-Cluster 12 Stunden und 43 Minuten, was einen zeitlichen Vorteil von 7 Minuten für den Servercluster ergibt. Hierbei wurden die Limitationen des Raspberry Pi-Clusters auf den Servercluster übertragen, was zu wenig Rechenleistung führte.

Für die jeweiligen Cluster wurden unterschiedliche Schwachstellen identifiziert. Während der Raspberry Pi-Cluster durch die geringere CPU-Leistung beim Starten und Stoppen der Pods sowie beim Schreiben der Ergebnisse gebremst wurde, war beim Server-Cluster die VPN- und SSH-Verbindung der limitierende Faktor. Die langsamere Datenübertragung verzögerte die Übertragung der Image-Informationen zum Start jedes neuen Durchlaufs im Vergleich zum Raspberry-Pi-Cluster. Dort wiederum verzögerte die langsame Schreibgeschwindigkeit der Ergebnisse den Start des nächsten Versuchsdurchlaufs.

5 Zusammenfassung und Ausblick

Die Autoren stellen in diesem Beitrag eine in KNIME entwickelte automatisierte Infrastruktur zur Verteilung von Simulationen vor, die komplexe Simulationsexperimente auf einer kostengünstiger Infrastruktur ermöglicht. Der entwickelte Workflow ist sowohl auf kleinen, leistungsschwachen Rechensystemen als auch auf leistungsstarken Servern anwendbar. Die jeweiligen Infrastrukturen zeigten individuelle Schwachstellen und bieten Potenzial zur Optimierung.

Zukünftig soll die Konfiguration von Simulationsexperimenten als simulationsgestützte Optimierungs-

schleife in den KNIME-Workflow integriert werden. Dies soll ermöglichen, die gewonnenen Daten zur Ableitung erfolgversprechender Parameterkonfigurationen zu nutzen. Zudem wird die Datenübertragung mittels VPN und SSH auf Raspberry Pi- und Server-Cluster untersucht, um Leerlaufzeiten zu reduzieren und die Gesamtperformance zu erhöhen.

References

- [1] Vieira, A. A. C., Dias, L. M. S., Santos, M. Y., Pereira, G. A. B., Oliveira, J. A. Setting an Industry 4.0 Research and Development Agenda for Simulation – A Literature Review. *International Journal of Simulation Modelling IJSIMM* 17(3); 2018 Sep, 377-390.
- [2] Lechler, T., Sjarov, M., Franke, J. Data Farming in Production Systems – A Review on Potentials, Challenges and Exemplary Applications. *Procedia CIRP* 96; 2021 Oct; 230–235.
- [3] Schlecht, M., de Guio, R., Köbler, J. Automated generation of simulation model in context of industry 4.0. *International Journal of Modelling and Simulation*; 2023 May; 1–13.
- [4] Sanchez, S. M., Sanchez, P. J., Wan, H. Work smarter, not harder: A tutorial on designing and conducting simulation experiments. In: Kim, S.; Feng, B.; Smith, K.; Masoud, S.; Zheng, Z.; Szabo, C. and Loper, M. (Eds.). *Proceedings of the 2021 Winter Simulation Conference (WSC)*, Phoenix (USA), 2021 Dec.
- [5] Leißau, M., Laroque C. Backward-Oriented Decision and Planning Approaches in Production Scenarios: A Systematic Literature Review and Potential Solution Approach. In: Mota, M.M.; Scala, P. (Eds.). *Proceedings of the EUROSIM Congress 2023*, Cham (Switzerland): Springer Nature Switzerland.
- [6] Król, D., Wrzeszcz, M., Kryza, B., Dutka, L., Kitowski, J. Massively Scalable Platform for Data Farming Supporting Heterogeneous Infrastructure In: Zimmermann, W. (Eds.). *The Fourth International Conference on Cloud Computing, Grids, and Virtualization*, Valencia (Spain), 2021 May; 144-149.
- [7] Pappert, F. S., Seufferth, D., Stein, H., Rose, O. Using Kubernetes to Improve Data Farming Capabilities. In: Corlu, C. G., Hunter, S. R., Lam, H., Onggo, B. S., Shortle, J., Biller, B. (Eds.). *Proceedings of the 2023 Winter Simulation Conference (WSC)*, San Antonio (USA), 2023 Dec; 2052-2063.
- [8] Leißau, M., Laroque, C. Reverse Engineering the Future – An Automated Backward Simulation Approach to on-Time Production in the Semiconductor Industry. In: Corlu, C. G., Hunter, S. R., Lam, H., Onggo, B. S., Shortle, J., Biller, B. (Eds.). *Proceedings of the 2023 Winter Simulation Conference (WSC)*, San Antonio (USA), 2023 Dec; 2040-2051.

Wird Simulation durch KI überflüssig? - eine subjektive Bestandsaufnahme -

Jochen Wittmann¹, Thomas Clemen²

¹Hochschule für Technik und Wirtschaft, Wilhelminenhofstraße 75A, 12459 Berlin

²Hochschule für Angewandte Wissenschaften Hamburg, Berliner Tor 7, 20099 Hamburg

Abstract. Ausgehend von den aktuellen Erfolgsberichten durch den Einsatz von KI-Methoden untersucht dieses Paper die Beziehung zwischen den Problemlösungsmethoden KI und Systemsimulation. Eine Analyse der Arbeitsschritte der beiden Ansätze stellt den grundlegenden Unterschied zwischen dem Black-Box-Ansatz der Lernmethoden und dem Glass-Box-Ansatz von strukturerklärenden Simulationsmodellen heraus. Anschließend kann der wechselseitige Nutzen der beiden Ansätze an vier Use-Cases erläutert werden. Als weiteres Ergebnis der Analyse stellt sich die Frage, inwieweit Simulation und KI-Verfahren zu gleichen, beziehungsweise zu jeweils anderen Ergebnissen führen können. Dazu wird das Konzept für eine strukturelle Analyse vorgestellt, das auf der Idee beruht, die Schnittmengenbeziehung zwischen den Ergebnissen der KI und der Methode Simulation zu untersuchen.

1 KI in Konkurrenz zu klassischen Simulationsmethoden?

Prognosen zur Entwicklung von Bedarf und inhaltlicher Ausgestaltung von Ingenieurberufen zeigen zwar einerseits, dass gerade für kreative Aufgaben wohl auch weiterhin die menschliche Intelligenz unersetzbar sein wird, andererseits könnten Anteile von um die 50 Prozent der Arbeitszeit durch Verfahren der künstlichen Intelligenz ersetzt oder zumindest erheblich unterstützt werden.

([1], [2], [3], [4]) Diese Prognosen sollen hier im Hinblick auf die Methode der dynamischen Systemsimulation näher untersucht werden. Auch in diesem Bereich gibt es erste Ansätze, Simulationsmodelle automatisch oder zumindest KI-unterstützt zu generieren, allerdings zeigen sich dabei derzeit wohl noch die Grenzen der KI. [5] Andererseits gibt es vielfältige Anwendungsbeispiele, bei denen KI-Einsatz und klassische Simulation erfolgreich kombiniert werden. Diese Beispiele werden wir im Abschnitt mit den möglichen Use-Cases noch detailliert aufgreifen.

Um die gegenseitigen Vorteile herausarbeiten zu können, soll aber zunächst die Arbeitsweise der Methode Simulation dem generellen Vorgehen beim datengetriebenen maschinellen Lernen (ML) (als Teilbereich der KI) gegenübergestellt werden. Danach werden dann die potenziellen Effizienzgewinne der Kombination der Methoden präziser dargestellt. Nach diesen, bereits in der Praxis etablierten Einsatzfeldern von Kopplung und/oder Kombination von KI und Simulation, wird zum Abschluss ein Ansatz vorgestellt, der durch Experimente auf struktureller, mengentheoretischer Basis Aussagen darüber erlaubt, ob die beiden Ansätze tatsächlich (nur) gleiche Ergebnisse hervorbringen oder eventuell qualitativ unterschiedliche Lösungen finden (können).

2 Analyse der Arbeitsschritte der isolierten Methoden

2.1 Methode Simulation

Ziel beim Einsatz der Methode Simulation ist stets die Entwicklung eines Modells in einem Formalismus, der die algorithmische Abarbeitung des Zeitfortschrittes im Rahmen eines Simulationslaufes möglich macht. Die so aus der Modellbeschreibung abgeleiteten Trajektorien der Modellgrößen über die Zeit erlauben Aussagen über das modellierte System im zuvor vereinbarten Validitätsbereich des Modells. Entscheidender Arbeitsschritt bei dieser Vorgehensweise ist die Aufstellung des Formalismus bzw. der Regeln, nach denen sich die zeitliche Dynamik des abgebildeten Systems entfaltet. (Abb. 1) Dabei gibt es zwei unterschiedliche Ansätze: Im allgemeinen Fall müssen Experimente am realen System ausgeführt werden und deren Ergebnisse dann allein mit menschlicher Intelligenz auf Abhängigkeiten und Regeln analysiert werden. Die Gesamtheit dieser Regeln stellt das for-

male Modell dar, das dann – in der Regel in einer Simulationsumgebung – algorithmisch weiterverarbeitet werden kann. Besteht bereits Vorwissen über das zu modellierende System, indem auf eine allgemeine Theorie (z.B. Energieerhaltung, Netzwerktheorie, ...) zurückgegriffen werden kann, so sind keine expliziten Systemexperimente notwendig. Vielmehr werden die Gesetze der Theorie angewandt, aus denen sich das formale Modell ableiten lässt. In beiden Fällen steht aber am Ende ein Regelsatz, mit dessen Hilfe sich die Eingangsdaten (bestehend aus Anfangszustand und Parameterbelegung) in einen Satz von Ausgangsdaten (Ergebnisse des Simulationslaufes) transformieren lassen. Betrachtet man diesen Ansatz in Form einer Input-Output-Box, so ist diese Box also transparent, das modellierte System wird durch ein erklärendes Strukturmodell in der Glass-Box repräsentiert.

2.2 Methode ML

Im Gegensatz zur Methode Simulation arbeiten viele Lernverfahren der KI mit einem Black-Box-Ansatz: Wesentlich sind hierfür als Ausgangspunkt wiederum bereits gesammelte Systemdaten. Diese müssen zunächst sorgfältig vorverarbeitet werden. Die Vorverarbeitung umfasst in der Regel die Homogenisierung sowie – im Fall von überwachten Lernverfahren - die Klassifikation (*labelling*) des Inputs gemäß der zu erwartenden, korrekten Systemantwort. Durch diese Klassifikation der Inputdatensätze mittels menschlicher Intelligenz wird Systemwissen in den Prozess eingebracht, damit sich danach in einer automatisierten Lernphase ohne weiteres Zutun ein Modell entwickeln kann. Dieses Modell antwortet in der Arbeitsphase (Modellanwendung) auf einen beliebigen Input mit einem Output, der – wiederum nur im vereinbarten Validitätsbereich - der gelernten Klassifikation entspricht.

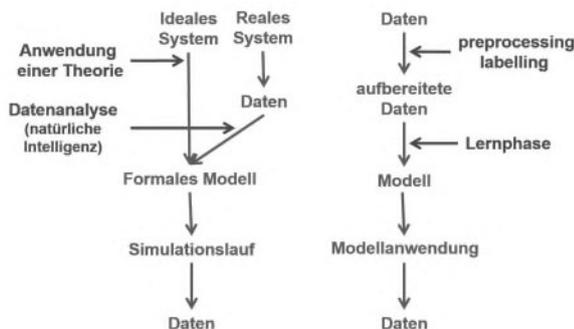


Abb. 1: Arbeitsschritte bei Modellbildung und datengetriebenen Lernverfahren

2.3 Vergleich der Ansätze

Ein wesentlicher Unterschied des entstandenen Modells liegt in der bereits erläuterten Black-Box-Sicht der Lernverfahren einerseits und der Glass-Box-Sicht der erklärenden Simulationsmodelle andererseits. Entsprechend erfordern die Lernverfahren für jeden einzelnen Input eine Aussage über den dazugehörigen korrekten Output aber keine Info über die Kausalzusammenhänge. Wie sich der Output aus dem jeweiligen Input ableitet, muss nicht als Regelwissen eingebracht werden. Entsprechend kann bei der Interpretation der Ergebnisse von Lernmodellen aber auch die Warum-Frage systematisch nicht geklärt werden, weil nicht notwendig isomorphe Beziehungen zwischen System und Modell ausgebildet sind, die die Kausalkette des Systems nachbilden.

Ein weiterer Unterschied liegt in der Menge der benötigten Systemdaten: Die Lerndaten müssen nicht nur sämtliche mögliche Klassen an Ergebnisdaten beinhalten, sie müssen auch in so großer Zahl vorliegen, dass die statistischen Ansätze beim automatischen Lernen von einer ausreichend großen Grundgesamtheit ausgehen können. Demgegenüber können die Regeln des Glass-Box-Modells auch durch wenige Daten, beim Ansatz über die Theorie sogar ganz ohne Systemdaten, gefunden bzw. hergeleitet werden.

Allerdings treten bei der Aufstellung des Regelsatzes für Simulationsmodelle häufig Unsicherheiten auf, die eine algorithmische Modellspezifikation wegen der Forderung der Eindeutigkeit nicht zulässt. In diesen Fällen hilft man sich durch die Einführung von randomisierten Parametern, die definierten Verteilungen gehorchen. Allerdings zieht dieser Ansatz praktisch immer und notwendigerweise eine aufwendige statistische Behandlung der Simulationsergebnisse nach sich. So relativieren sich die Einsparungen bezüglich der Menge an Input- bzw. Lerndaten am Ende durch einen höheren Aufwand auf der Outputseite, weil die Ergebnisse eines Simulationslaufes alleine keine Aussagekraft haben und im Sinne eines Zufallsexperimentes durch eine Vielzahl von Simulationsläufen statistisch abgesichert werden müssen.

Mit diesen allgemeinen Überlegungen im Hintergrund sollen nun aber konstruktive Synergien der beiden unterschiedlichen Ansätze in Form von vier Use-Cases vorgestellt werden.

3 Use-Cases zur kombinierten Nutzung von KI- und Simulationsmethoden

Die im folgenden beschriebenen Use-Cases beleuchten Ansätze zur kombinierten Nutzung von KI-Methoden und Simulation, die aktuell tatsächlich realisiert sind und nicht nur in Form von Forschungsergebnissen oder Pilotstudien vorliegen. Dazu wurde eine Google-Suche mit den Suchbegriffen „KI und Simulation“ im Web durchgeführt. Direkt und intensiv beschäftigen sich die hier aufgeführten Seiten mit diesem Thema:

- Simplan [6]
- Siemens “Simultelligenz” [7]
- Anylogic [8]
- Mathworks [9]
- Fraunhofer IISB Erlangen [10]
- Merkle CAE Solutions GmbH [11]

Jeder dieser Anbieter setzt zwar etwas andere Akzente, allerdings lassen sich für eine Übersicht auch starke Gemeinsamkeiten und Überschneidungen finden, die zu den hier vorgestellten typischen Use-Cases zusammengefasst werden können. Obwohl eine Fundstelle keinen direkten Beitrag zur Herausarbeitung der Use-Cases beigetragen hat, soll sie hier doch als Bemerkung am Rande Beachtung finden: Wenn eine Kopplung von KI und Simulation konzeptuell und softwaretechnisch gelingt, so weist das LRZ München darauf hin, dass dann auch die grundlegend unterschiedlichen Hardwarekonzepte, die für eine effiziente Berechnung notwendig werden, kombiniert und optimiert werden müssen. [12]

3.1 ChatBot nutzen für Knowledge Engineering und Systemanalyse

Der erste Use-Case unterstützt einen Arbeitsschritt bei der Entwicklung von Simulationsmodellen, der in der Beschreibung in Abschnitt 2 nicht explizit erwähnt wurde, aber dennoch einen wichtigen Teil der Arbeit auf dem Weg zu einem korrekten, vollständigen und konsistenten Modell ausmacht: die Systemanalyse und das Knowledge Engineering, das in der Analysephase mit Experten aus der Zieldomäne des zu entwickelnden Modells oft in Form von Interviews durchgeführt wird. Diese binden die Zeit der Experten mit Systemwissen. Viele Fragen bezüglich der Modellentwicklung könnten mit Hilfe eines Chatbots, der auf die vorhandene Datenbasis der Anwendungsdomäne zugreift geklärt werden. Sicherlich werden nicht alle Fragen beantwortet werden

können, allerdings dürfte ein solcher Einsatz die Zeit, die sonst für Experteninterviews benötigt wird, erheblich reduzieren. Forschungsseitig wurde ein ähnlicher Ansatz von Freydenlund et al. [5] untersucht.

3.2 Online-Sensordaten durch trainierte Lernmodelle analysieren

Ein zweiter Use-Case ist die Kopplung eines KI basierten, vortrainierten Modells mit online-Sensordaten unter Echtzeitbedingungen. Allerdings setzt das natürlich voraus, dass das KI-Modell zuvor mit allen potenziellen Zuständen des Systems trainiert wurde. Darüber hinaus stellt sich die Frage, ob das bloße Erkennen der Unregelmäßigkeit ausreicht, oder ob nicht zur Ergreifung von Gegenmaßnahmen eben doch auch das Regelwerk des Systems transparent sein muss.

3.3 Lerndaten für KI durch Simulationsmodelle erzeugen

Häufig ist es problematisch, konsistente Lerndaten für die Lernphase der KI zu erhalten. Solche Daten könnten künstlich durch klassische Simulation generiert werden. Vorteil: sie sind strukturiert, klassifiziert und vorbehaltlich Rechenzeit in beliebiger Menge zu erzeugen. Allerdings fällt der doppelte Modellierungsaufwand sofort ins Auge: es ist sowohl ein Simulationsmodell als auch ein KI-Modell im Einsatz. Glass-Box-Systemverständnis muss vorhanden sein, um Lerndaten sinnvoll generieren zu können. Die Echtzeit-Schnelligkeit des KI-Ansatzes wird also mit erheblichem Mehraufwand in der Konzeptionsphase erkauft. Darüber hinaus wird das Problem der „seltenen Ereignisse“ relevant: Selbst wenn diese durch vorherige Simulation als Lerndaten explizit generiert wurden, muss sichergestellt sein, dass sie von der KI als solche korrekt erkannt und gelernt werden und nicht als Ausreißer ignoriert werden.

3.4 Test des KI-Modells in virtueller Simulationsumgebung

Umgekehrt kann ein Simulationsmodell zum Test des KI-Modells verwendet werden: Das KI-Modell ist dabei mit einem digitalen Zwilling in Form eines Simulationsmodells der Zielumgebung gekoppelt und so kann sein späterer Einsatz im realen System gefahrlos in dieser virtuellen Umgebung getestet werden. Auch in diesem Szenario müssen Simulationsmodell und KI-Modell entwickelt werden.

Ein ähnliches Szenario ergibt sich, wenn KI-gestützte

Simulationsmodelle uni- oder bi-direktional mit IoT-Sensoren, Dronen oder anderen physikalischen Entitäten gekoppelt werden. [13] Durch diese Kopplung wird die kontinuierliche Anpassung des Simulationszustandes am realen Systemzustand ermöglicht.

3.5 „hybride“ Simulationsmodelle

Der letzte Use-Case beschäftigt sich mit der Situation, dass in einem komplexen, modular-hierarchisch aufgebauten Modell das Systemwissen nicht ausreicht, um das Verhalten einzelner Komponenten regelbasiert zu beschreiben. Bei bekannter Schnittstelle könnten diese Komponenten durch ein zuvor auf Black-Box-Basis trainiertes KI-Modell ersetzt werden. Dann sind in dieser Stelle keine Kausalbeziehungen zu spezifizieren, allerdings wird bei einem derartigen Modelldesign dann auch die Interpretation der Ergebnisse des Gesamtmodells komplizierter, weil keine durchgehende Kausalkette vorliegt.

4 Sind die Ansätze bezüglich ihrer Ergebnisse äquivalent?

Wie die jeweils zitierten Quellen zeigen, sind die in Abschnitt 3 vorgestellten Use-Cases Stand der Technik und können pragmatisch an die Aufgabenstellung angepasst und bezüglich Kosten und Nutzen bewertet werden. Allerdings stellt sich die Frage, ob es nicht darüber hinausgehende Potenziale gibt, die durch diese Use-Cases noch nicht abgedeckt sind. Ein Anhaltspunkt ist die oftmals geschilderte Erfahrung, dass durch eine automatische Datenanalyse Systembeziehungen aufgedeckt werden, die zuvor unbekannt waren.

Mit diesem Gedanken im Hintergrund kann man sich die Frage stellen, ob die Ergebnisse aus klassischer Simulation auf der einen, und die durch KI-Ansätze generierten Ergebnisse auf der anderen Seite, äquivalent sind bzw. sein können. Oder ist es vorstellbar, dass KI-basierte Verfahren Ergebnisse produzieren, die durch klassische Simulation nicht zu erzielen wären? Dasselbe gilt natürlich analog auch für die umgekehrte Richtung: gibt es Ergebnisse des Simulationsansatzes, die durch KI-Methoden nicht explorierbar sind?

Einen Ausblick auf zukünftige Forschungsaktivitäten soll die folgende Aufgabenstellung darstellen. Gegeben ist ein komplexes, ggf. adaptives Input-Output-System, das einmal durch ein regelbasiertes Simulationsmodell und

parallel durch ein KI/ML-Modell modelliert wird. Betrachtet man nun die Menge der Ergebnisse, die die beiden Ansätze liefern, so ergeben sich im allgemeinen Fall drei unterschiedliche Fälle: Die Mengen der Ergebnisse der Ansätze überdecken sich vollständig (a), sie sind disjunkt (b), oder aber es liegt der Fall einer echten Schnittmenge (c) vor. Es ist zu vermuten, dass die Fälle b) und c) realistischer sind, da man in der Praxis ja bereits beide Ansätze einsetzt und zumindest vergleichbare Ergebnisse erhält. Allerdings stellt sich dann die Frage, ob tatsächlich eine vollständige Äquivalenz (Fall c) vorliegen kann, oder ob es Systemzustände gibt, die nur unter Anwendung eines der Ansätze sichtbar werden. Die vorläufige Empfehlung der Autoren ist deshalb, Simulation und KI „gemeinsam zu denken“. Diese kombinierte Exploration des Zustandsraumes eines komplexen Systems verspricht interessante und neue Erkenntnisse zur Systemdynamik.

Literaturverzeichnis

- [1] H. K. Nordas und F. Klügl, „Drivers of Automation and Consequences for Jobs in Engineering Services: An Agent-Based Modelling Approach,“ *Front. Robot. AI*, Bde. %1 von %2Volume 8 - 2021, 10 05 2021.
- [2] P. Daugherty, B. Ghosh und K. Narain, „A new era of generative AI for everyone,“ 01 06 2024. [Online]. Available: <https://www.accenture.com/content/dam/accenture/final/accenture-com/document/Accenture-A-New-Era-of-Generative-AI-for-Everyone.pdf>.
- [3] R. Strack, M. Carrasco und P. Kolo, „The Future of Jobs in the Era of AI,“ 01 06 2024. [Online]. Available: <https://www.bcg.com/publications/2021/impact-of-new-technologies-on-jobs>.
- [4] T. Pultarova, „Will artificial intelligence replace engineers?,“ 01 06 2024. [Online]. Available: <https://www.imeche.org/news/news-article/feature-will-artificial-intelligence-replace-engineers>.
- [5] E. Frydenlund, J. Martinez, J. J. Padilla, K. Palacio und D. Shuttleworth, „Modeler in a box: how can large language models aid in the simulation modelling process?,“ *Simulation: Transactions of the Society for Modeling and Simulation International*, pp. 1-23, 2024.
- [6] Simplan, „Künstliche Intelligenz,“ [Online]. Available: <https://www.simplan.de/services/kuenstliche-intelligenz/>. [Zugriff am 01 06 2024].
- [7] SIEMENS AG, „Simultelligenz,“ [Online]. Available: <https://www.siemens.com/de/de/unternehmen/stories/forschung-technologien/kuenstliche-intelligenz/simultelligenz.html>. [Zugriff am 01 06 2024].
- [8] „Anylogic,“ [Online]. Available: https://www.anylogic.com/features/artificial-intelligence/?utm_source=youtube&utm_medium=social-organic. [Zugriff am 01 06 2024].
- [9] Mathworks, „4-proven-applications-for-your-ai-algorithms,“ [Online]. Available: <https://de.mathworks.com/company/technical-articles/4-proven-applications-for-your-ai-algorithms.html>. [Zugriff am 01 06 2024].
- [10] Fraunhofer IISB Erlangen, „Augmented Simulation,“ [Online]. Available: https://https://www.iisb.fraunhofer.de/en/research_areas/simulation/ai-augmented-simulation.html. [Zugriff am 01 06 2024].
- [11] Merkle CAE Solutions, „Künstliche Intelligenz,“ [Online]. Available: <https://www.merkle-partner.de/leistungen/kuenstliche-intelligenz>. [Zugriff am 01 06 2024].
- [12] „Leibniz Supercomputing Centre München,“ [Online]. Available: https://www.lrz.de/presse/ereignisse/2021-05-04-SuperMUC-NG-Phase-2_DE/. [Zugriff am 01 06 2024].
- [13] T. Clemen, N. Ahmady-Moghaddam und U. Lenfers, „Multi-Agent Systems and Digital Twins for Smarter Cities,“ *Proceedings of the 2021 ACM SIGSIM Conference on Principles of Advanced Discrete Simulation, 1(1)*, 45-55., pp. 45-55, 2021.

Beschleunigte Erstellung von Simulationsmodellen für Produktions- und Logistikprozesse mithilfe von GPT-basierten Large Language Models

Roman Krämer^{1*}, Jens Heger¹

¹Institut für Produktionstechnik und -systeme, Leuphana Universität Lüneburg, Universitätsallee 1, 21335 Lüneburg, Deutschland; *roman.kraemer@leuphana.de

Abstract. In dieser Arbeit wird die beschleunigte Erstellung von Python Code für Simulationsmodelle für Discrete Event Simulation (DES) mithilfe des Large Language Models (LLM) ChatGPT von OpenAI am Beispiel von Produktions- und Logistikprozessen demonstriert. Die Implementierung und Ausführung der Simulation erfolgt in einer Python-Umgebung unter Nutzung des quelloffenen Frameworks „SimPy“. Es werden notwendige Schritte aufgezeigt, um reproduzierbare Ergebnisse in einer nicht-datengenerierenden Produktionsumgebung zu erzeugen. Möglichkeiten und Aufwand mit quelloffenen und frei verfügbaren Mitteln, ad hoc Simulationsmodelle zu erstellen und Simulationen als Entscheidungshilfe zu nutzen, werden betrachtet. Die Potenziale und Limitierungen sowie mögliche Lösungen werden anhand einer exemplarischen sequenziellen Fertigungslinie aufgezeigt.

Einleitung

In Anbetracht kürzerer Produktlebenszyklen und der hiermit verbundenen Notwendigkeit häufiger Reorganisationen der Fertigung, haben sich ereignisdiskrete Simulationen als eine unverzichtbare Methode in der modernen Produktionsplanung etabliert. Die Erstellung und Durchführung von Simulationsstudien erfordert im Vorfeld jedoch oftmals einen erheblichen Ressourcenaufwand in Form von Fachpersonal und Fachwissen, Zeit sowie spezialisierter und kostspieliger Software.

Eine hier häufig anzutreffende Konstellation beteiligter Parteien besteht aus Prozessexperten, der Entscheidungsebene und den Simulationsexperten, welche in vielen Fällen externe Dienstleister sind. Allein die Abstimmung von Informationen und Zielformulierung über diese Parteien hinweg macht die Erstellung zu einem iterativen und daher langsamen Prozess vor allem in kleineren und mittleren Unternehmen, in denen keine

Erfahrungen und Routinen in Bezug auf die Methode Simulation etabliert sind. Dieser Ressourcen- und Organisationsaufwand schränkt die Einsätze von Simulationen daher häufig eben dort ein, wo sie am meisten benötigt werden. Optimierungspotenziale werden so nicht erkannt und genutzt.

Mit der stetig besseren Verfügbarkeit leistungsfähiger Large Language Models (LLMs) wie ChatGPT von OpenAI oder LLaMA3 von Meta eröffnen sich neue Möglichkeiten, die Erstellung von Simulationsmodellen erheblich zu beschleunigen und den Ressourcen- und Organisationsaufwand zu reduzieren.

Es wurde demonstriert, dass LLMs funktionalen Code erstellen können, der als Grundlage für Simulationen im Bereich Logistik dienen kann [1]. Auch die automatisierte Erstellung von Simulationsmodellen in großen, datengenerierenden Produktionsumgebungen hat nicht an Aktualität verloren [2, 3].

Dieser Beitrag positioniert sich zwischen diesen Anwendungsfällen: Es werden die beschleunigte Erstellung eines Simulationsmodells und die Durchführung der Simulationsstudie mithilfe eines frei verfügbaren LLMs von OpenAI und des Python Frameworks „SimPy“ am Beispiel einer exemplarischen Fertigungslinie in einer nicht-datengenerierenden Produktionsumgebung durchgeführt. Es werden keine Systeme zur Datenakquise aufgebaut.

1 Produktionsszenario als Basis der Modellbeschreibung

Als Beispiel soll ein simplifiziertes Produktionsszenario einer sequenziellen Fertigungslinie dienen, welche in ähnlicher Form in vielen realen Produktionen anzutreffen ist. Die Fertigungslinie ist gegliedert in:

1. Rohmateriallager
2. Produktionsmaschine 1
3. Pufferlager
4. Produktionsmaschine 2
5. Warenlager

Für den Transport zwischen den Lagern und Maschinen sowie deren Bestückung sind zwei AGVs (Automated Guided Vehicle) zuständig.

1.1 Produktionssequenz und -ablauf

AGV 1 entnimmt die für das Produkt benötigte Menge Rohmaterial aus dem Rohmateriallager und bestückt Produktionsmaschine 1. Diese beginnt mit der Bearbeitung. Nach abgeschlossener Bearbeitung rutscht das Zwischenprodukt in das Pufferlager. AGV 2 entnimmt das fertige Zwischenprodukt aus dem Pufferlager und bringt es zu Produktionsmaschine 2. Nach abgeschlossener Bearbeitung rutscht das Endprodukt weiter ins Warenlager.

Die Bearbeitungs-, Handhabungs- und Transportzeiten unterliegen keiner stochastischen Verteilung. Sie werden hier zur Veranschaulichung und zu Zwecken der Reproduzier- sowie Nachvollziehbarkeit als konstant angenommen.

Zeitschritte	t (in min)
AGV 1: Materiallager – Maschine 1	2
AGV 2: Pufferlager – Maschine 2	2
Bearbeitungszeit Maschine 1	10
Bearbeitungszeit Maschine 2	20
Simulationsdauer gesamt	480

Tabelle 1: Systemeigenschaften und -parameter

Der Anfangsbestand im Rohmateriallager beträgt 100 Einheiten und es wird eine Einheit für die Erstellung eines Endprodukts benötigt. Das Ergebnis der Simulation soll die Darstellung der Bestände der drei Lager in Stück, Auslastung der beiden Produktionsmaschinen sowie Nutzungsdauer der beiden AGV sein. Diese Werte sollen zur weiteren Bewertung visuell als Diagramme ausgegeben werden.

Das Szenario der Produktionssequenz ist so gewählt, da hier wegen der hohen Bearbeitungszeit in Produktionsmaschine 2 ein Flaschenhals vorliegt und das Ergebnis somit gut validiert und veranschaulicht werden kann.

1.2 Auswahl des LLM

Es kommt ChatGPT von OpenAI in der über den Browser abrufbaren Web-Version zum Einsatz. Das zugrundeliegende Modell ist ChatGPT 3.5 Turbo in der Version „gpt-3.5-turbo-0125“ mit einem Kontextfenster von 16385 Token (ca. 12300 englische Wörter) [4]. Das Modell kann nach Registrierung und Anlegen eines Nutzerkontos bei OpenAI frei genutzt werden.

Sämtliche Instruktionen werden in Englisch an ChatGPT übergeben. Dies geschieht vor dem Hintergrund, dass die Trainingsdatensätze für Sprache und Code in Englisch jeweils den größten Umfang haben und hier somit bessere Ergebnisse zu erwarten sind.

Für die Ausführung des von ChatGPT erzeugten Codes werden SimPy 4.1.1 und Python 3.10 genutzt.

2 Durchführung

Mit einem intuitiven Vorgehen, wie es von Konversationen mit LLMs gewohnt ist, können Ergebnisse schnell erzeugt werden. Diese sind in ihrer Ausprägung sowie Qualität und Reproduzierbarkeit allerdings nicht für den gedachten Anwendungsfall weiterverwendbar. Durch die Sequenz und feste Zuordnung von Parametern zu den jeweiligen Schritten entsteht ein Gesamtkontext, der in eben dieser Sequenz mit den schrittbezogenen Parametern detailliert beschrieben werden muss. ChatGPT ist so ausgelegt, immer Antworten zu generieren und trifft hierfür falls nötig Annahmen, wenn nicht alle Informationen aus dem Eingabe-Prompt hervorgehen. Dies geschieht ohne explizite Mitteilung an den Nutzer und wird allgemein als "Halluzination" bezeichnet.

In alltäglichen Anwendungsfällen und Problemstellungen werden solche Annahmen zwar meist passend getroffen, die Erstellung von Simulationsmodellen im Produktionskontext scheint zumindest zum jetzigen Zeitpunkt keinen solchen Anwendungsfall darzustellen. Aus diesem Grund müssen die Prompts hier präzise formuliert werden, sodass der Interpretationsspielraum stark minimiert und idealerweise eliminiert wird. Da der Simulationskontext durch den Anwendungsfall klar definiert ist, kann dies ebenso global in "custom instructions" erfolgen. Diese individuellen Hinweise werden im Hintergrund genutzt, um alle nutzergenerierten Prompts für die Eingabemaske unsichtbar zu ergänzen und um den Kontext in den Hinweisen zu erweitern.

3 Erstellung der Instruktionen

Die individuellen Hinweise wurden zunächst iterativ ausgehend vom erzeugten Output des LLM angepasst und bedarfsweise erweitert. Hierfür wurde der von ChatGPT jeweils erzeugte Code ausgeführt und der IST-Output der Simulation mit dem in Abschnitt 1.1 beschriebenen SOLL-Output abgeglichen. Die hier gesammelten Erkenntnisse wurden dann genutzt, um ChatGPT die individuellen Hinweise selbst verfassen zu lassen. Die individuellen Hinweise und der Prompt zu deren Erstellung können in dieser Repository gefunden werden: <https://github.com/romankraemer/ASIM2024>

Der Fokus lag darauf, den Spielraum für Interpretationen thematisch möglichst eng einzugrenzen, um zügig, vollständigen und mit allen Parametern ausgestatteten Code erzeugen zu lassen, der im Idealfall nach einer Überprüfung durch den Nutzer lediglich noch ausgeführt werden muss und keine großen Anpassungen und Debugging erfordert.

Neben der „harten“ fachbezogenen Vorgabe, SimPy zu nutzen, um Code für Simulationen in Python zu erzeugen und Definition sowie Anweisungen zur Erstellung von Plots der Ergebnisse, erweisen sich „weiche“ Vorgaben in ebenso hohem Maße der Qualität dienlich. Insbesondere seien an dieser Stelle genannt:

- ChatGPT explizit instruieren, den Nutzer nach allen Informationen und erforderlichen Parametern zu fragen, um keine Annahmen treffen zu müssen
- Dem Nutzer eine Zusammenfassung des Modells präsentieren, bevor der Python Code generiert wird

Für den ersten Punkt ist anzunehmen, dass die Kenntnis über SimPy und die in diesem Framework enthaltenen Klassen und Methoden mit ihren jeweils erforderlichen und möglichen Parametern zu einer maßgeblichen Eingrenzung des Kontexts beiträgt und die Modellerstellung hierdurch zügig in die erforderliche Richtung gesteuert wird. Der zweite Punkt erleichtert das Auffinden von unpräzisen Angaben und hilft so vor allem dem Nutzer bei der Validierung des Modellcodes.

Es sollte beachtet werden, die Instruktionen nicht zu fallspezifisch zu formulieren, um eine generelle Anwendbarkeit zur Erzeugung von Simulationsmodellen für Produktionsprozesse nicht einzuschränken.

4 Ergebnisse

Wie in Abschnitt 1.2 beschrieben, wird ChatGPT 3.5 Turbo in der Version „gpt-3.5-turbo-0125“ mit individuellen Hinweisen genutzt. Das Fertigungsszenario wird wie in Abschnitt 1.1 dargelegt mit den Fertigungs- und Simulationsparametern, einer Beschreibung des Simulationsziels und den gewünschten Ausgabewerten sowie deren Darstellung in Textform beschrieben und an ChatGPT überreicht. Da ChatGPT in den individuellen Hinweisen explizit instruiert wurde, Fragen zu stellen, wenn die Angaben nicht ausreichend sind, werden an dieser Stelle auch evtl. aufkommende Fragen beantwortet. Die vollständige Konversation kann hier abgerufen werden: <https://github.com/romankraemer/ASIM2024>

4.1 Simulationsergebnisse

Abbildung 1 zeigt, dass der Bestand des Warenlagers nach abgeschlossener Anlaufphase der Produktionslinie im 20-minütigen Takt, welcher durch Produktionsmaschine 2 vorgegeben wird, um jeweils ein neues Endprodukt ansteigt.

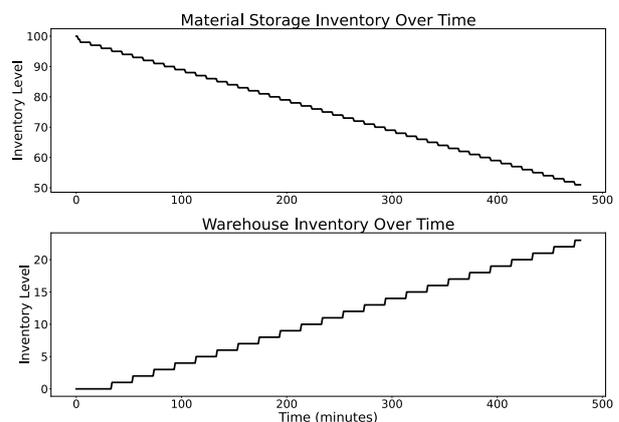


Abbildung 1: Bestand im Rohmaterial- und Warenlager

Der Bestand im Rohmateriallager sinkt, entsprechend des Rohmaterialverbrauchs in Produktionsmaschine 1, alle 10 Minuten. Abbildung 2 zeigt den Bestand im Pufferlager, welcher stetig ansteigt, da Produktionsmaschine 2 im Vergleich zu Maschine 1 eine doppelt so hohe Bearbeitungszeit hat.

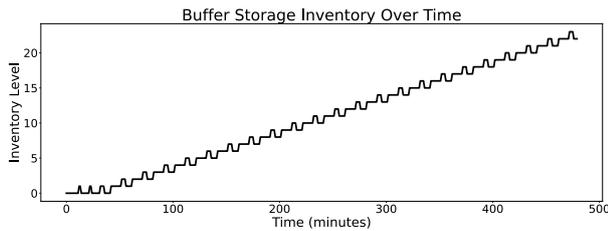


Abbildung 2: Bestand im Pufferlager

In Abbildung 3 ist die Belegung der beiden Produktionsmaschinen dargestellt.

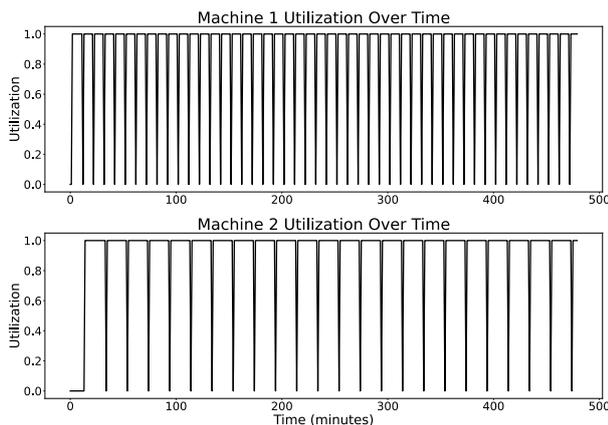


Abbildung 3: Belegung der Produktionsmaschinen

Abbildung 3 verdeutlicht hier nochmals, dass die Bearbeitungszeit in Produktionsmaschine 2 einen Flaschenhals für das Gesamtsystem darstellt.

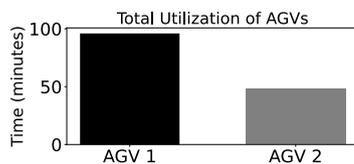


Abbildung 4: Nutzungsdauer der beiden AGV

Abbildung 4 zeigt die kumulierte Nutzung in Minuten für die beiden AGV. Die Simulationsergebnisse zeigen, dass durch den Einsatz einer zweiten Produktionsmaschine parallel zu Produktionsmaschine 2 die Linie optimiert und der Gesamtdurchsatz verdoppelt werden kann. Die Ergebnisse des von ChatGPT erstellten Simulationsmodells wurden also korrekt und vollständig umgesetzt.

5 Fazit und Ausblick

Der Beitrag zeigt den Einsatz eines GPT-basierten LLM zur Erstellung eines Simulationsmodells im Kontext der

Produktion. Für die Modellierung einfacher Abläufe kann ein ChatGPT Grundmodell, welches um individuelle Hinweise erweitert wird, zügig erfolgreich eingesetzt werden. Die strukturierte Abfrage von Systemeigenschaften und Erstellung von Simulationsmodellen wird mit einem solchen Ansatz vereinfacht und beschleunigt. Jedoch wird deutlich, dass die Vereinfachung an dieser Stelle teilweise einen hohen Verifizierungs- und Validierungsaufwand mit sich bringt und daher Expertise in Umgang und Erstellung von Simulationen erfordert. Bei allzu spezifischen Anforderungen an das Modell oder Ausgabewerte wird erkennbar, dass ChatGPT vermutlich nicht auf die Erstellung von Simulationen trainiert ist, da teilweise viele erklärende Schritte notwendig sind. Dies kann sich nach einiger Zeit negativ auf die Ausgabe auswirken, da Informationen anfangen, aus dem Kontextfenster zu fallen.

Trotz dieses Umstands sind keine LLMs des absoluten Top-Segments notwendig, um nützliche Ergebnisse zu erzeugen. Agentenbasierte Architekturen mit eng definiertem Kontext und Erweiterungen mit Retrieval Augmented Generation (RAG) Ansätzen oder eine Kombination aus beidem stellen daher vielversprechende Möglichkeiten dar, um das Potenzial solcher LLMs auch für die Übersetzung komplexer Produktionsabläufe in Simulationsmodelle nutzen zu können.

Für Simulationsexperten kann auch ein Vorgehen wie in diesem Beitrag dienlich sein, um z. B. bereits in ersten Besprechungsterminen schnell fundierte und anschauliche Gesprächsgrundlagen zu schaffen.

Literaturverzeichnis

- [1] I. Jackson, M. Jesus Saenz und D. Ivanov, "From natural language to simulations: applying AI to automate simulation modelling of logistics systems", *International Journal of Production Research*, Jg. 62, Nr. 4, S. 1434–1457, 2024.
- [2] M. Carl May, C. Nestroy, L. Overbeck und G. Lanza, "Automated model generation framework for material flow simulations of production systems", *International Journal of Production Research*, Jg. 62, 1-2, S. 141–156, 2024.
- [3] J. Garrido und J. Sáez, "Integration of automatic generated simulation models, machine control projects and management tools to support whole life cycle of industrial digital twins", *IFAC-PapersOnLine*, Jg. 52, Nr. 13, S. 1814–1819, 2019.
- [4] OpenAI, Models. [Online]. Verfügbar unter: <https://platform.openai.com/docs/models> (Zugriff am: 30. Mai 2024).

Concept development for coupling Simulation and Machine Learning in Supply Chains

Yen Mai Thi¹, Marc Münnich², Florian Zumppe², Mario Callefi³, Pierre Grzona^{3*}

¹Westfälische Hochschule Zwickau, Fakultät Wirtschaftswissenschaften, Zwickau, Germany, yen.mai.thi@fh-zwickau.de

²Fraunhofer-Institut für Werkzeugmaschinen und Umformtechnik IWU, Reichenhainer Straße 88, 09126 Chemnitz, Germany; florian.zumppe@iwu.fraunhofer.de, marc.muennich@iwu.fraunhofer.de

³Technische Universität Chemnitz, Professur Fabrikplanung und Intralogistik, Erfenschlager Str. 73, 09125 Chemnitz, Germany; mario.callefi@mb.tu-chemnitz.de, * pierre.grzona@mb.tu-chemnitz.de

Abstract

This study gives insights into the development of a simulation and machine learning combination concept in the field of additive manufacturing (AM). The method of systems engineering (SE) is used to determine the relationship between different technical goals in the field of semantic modelling of supply chains to validate the results of a matchmaking model for the digital AM supply platform to increase its network resilience. The concept serves as a base for further implementation.

Introduction

The aim of this study is to present the current state of research on the coupling of simulation and machine learning (ML) in the supply chain environment as part of a research project to create a platform-based simulation service. During the last year several prerequisites could be elaborated, whereas one is the possible coupling scenarios for simulation and ML [1]. From 4 cases based on VDI 3633 Blatt 12:2020 the connection of simulation followed by a ML approach was chosen as a basis for further investigation [2].

Additionally, the different approaches in the case for determining the resilience of a supply chain were investigated by *Grzona et al.* [3]. Based on the literature review, different use cases were identified and the aimed for the simulation studies. Time and Costs were the most named aiming for investigation in the literature and the usage of reinforcement learning and neural networks as ML's techniques of ML. By conducting survey also, the opportunity to increase the ML knowledge among the simulation experts was elaborated. [3]

1 Methodology

The research objective to achieve is to create a concept for the coupling of simulation and ML in an exemplaric value network to assess the previously named aims. This leads to 2 Research Questions:

RQ1: What can an exemplaric Value network look like?

RQ2: How can a coupling concept look like?

A systems engineering (SE) approach was chosen to develop the concept for implementing the coupling case further. The SE approach represents a well-established methodology for the design of systems, encompassing the principles of systems thinking. The SE process model and the methods and tools employed in the problem-solving process support a systematic transfer from an actual state to a desired state. The understanding of the elements and relations within the system, as well as their relations to the environment and environmental systems, constitutes a fundamental aspect of system thinking. [4]

These systematic perspectives can also be employed in conjunction with technical systems in sociotechnical systems, which incorporate the aspects of people and society [5]. Furthermore, the overall lifecycle of a system is considered from the development, through the realization, and finally the usage of the system [4].

2 Concept development

Ahrens [5] proposed a 6 Phases approach for a SE cycle (**Figure 1**). As the scope of the design is determined by the project a brief introduction will be given, followed by an explanation of the connection to the matchmaking process. From this the technical goals will be described

which lead to the methods used to create in the penultimate step the overall concept.

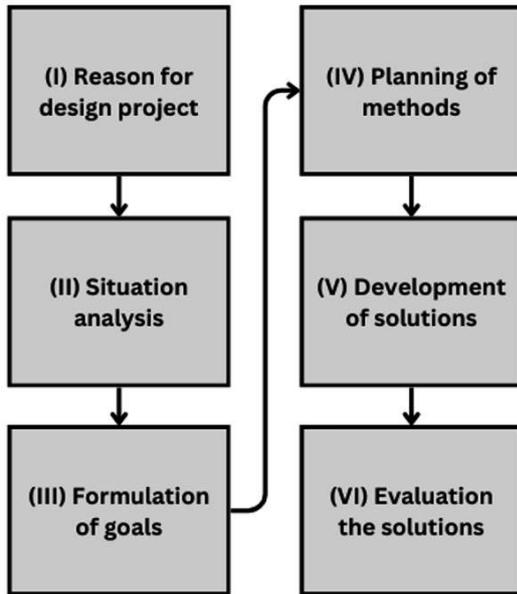


Figure 1: Six phases of SE process according to Ahrens [5]

2.1 Reasons for the design project

The aim of the project is to develop a digital platform for additive manufacturing (AM) to increase the resilience of value networks with the means of artificial intelligence and ontologies [6]. As the data of the system elements – in this case AM-resources – is based on the semantic data model it is needed to incorporate these concepts into the overall process [7].

The AM process chain has five phases, each with tools, models, and method requirements. Product design aims to create a digital object using computer-aided design (CAD) software. The pre-process is typically done in computer-aided planning (CAP) software to optimize the building room, slice the model, and set the machine parameters. During the printing process, a series of measures are employed to monitor the building process and, if necessary, to make in-situ adjustments to the machine parameters to reduce the rate of defects. This is supported by software for computer-aided manufacturing (CAM). Following the printing, several steps are typically carried out during the post-processing stage, including cleaning and separating support structures and verifying the physical product and process measures. The final stage is the finishing process, which employs a combination of techniques, including thermal finishing, to harden the products, mechanical methods, and quality

control, which may be destructive or non-destructive. Each of these steps requires a specific amount of time [8].

2.2 Situation analysis

The Simulation/ML core should be used to validate and support the general matchmaking process on the platform [9]. This is driven by the semantic data model for the capabilities of the different roles in the process chains. Out of the process chain, a graph is generated for the value network for which multiple combinations of suppliers and partners need to be found by matchmaking. As in this step, the behavior of this network is not evaluated and assessed in the dimensions of resilience, time, price, and quality. This leads to the models for the Simulation/ML (Figure 2).

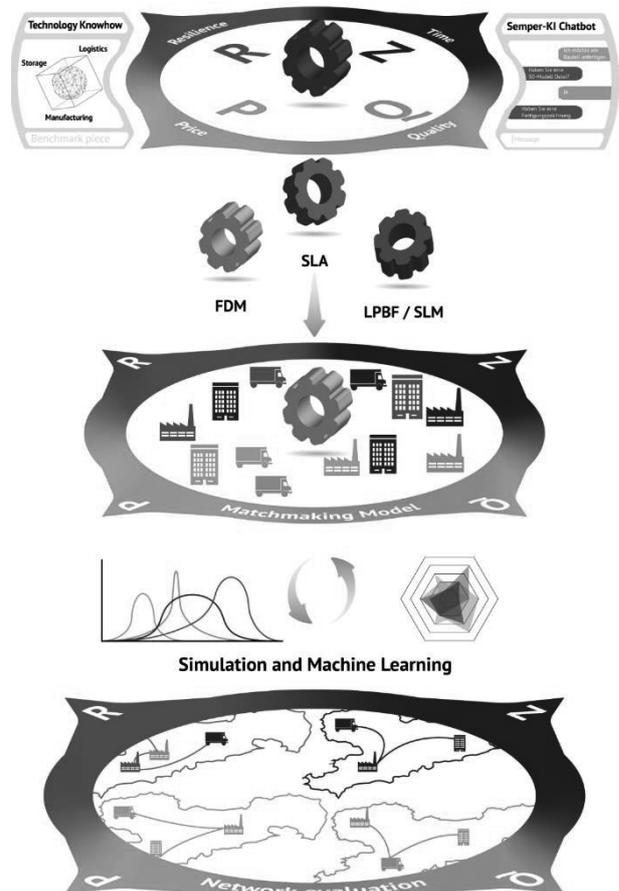


Figure 2: Meta model of matchmaking model and Simulation/ML core © InfAI e.V. Leipzig

2.3 Formulation of goals

The technical goals were elaborated in a continuous group work. As mentioned previously, a semantic database is needed. In this case, a specific AM logistics ontology is worked with, which is connected to the AM

ontology [7]. As the concept needed to be proven first, the focus was not on developing new simulation tools but on incorporating existing ones. Therefore, Siemens Tecnomatix Plant Simulation, AnyLogic, and Anylogistix were chosen for their ability to interact with external software through interfaces.

Besides the work with these tools, the solution should be able to analyze the data from multiple simulation experiments so that a data basis for whether a reinforcement learning or neural network approach can be used.

As the stakeholders connected to the work are from different domains and knowledge levels, a highly comprehensive model should be developed. At first, it should include the aspects of time and cost, but it should be easy to adapt to the other aspects.

2.4 Planning of methods

For the problem-solving cycle, the Case study method was chosen because of the explorative and unique character of the problem to solve [10].

A first test case was defined based on a product consisting of three different parts manufactured by the three partners and delivered to one customer. Additionally, the partners have different production capabilities, and the product needs several steps of post-processing. These steps consist of refining the parts, assembly, quality control, packaging, and delivery to the customer.

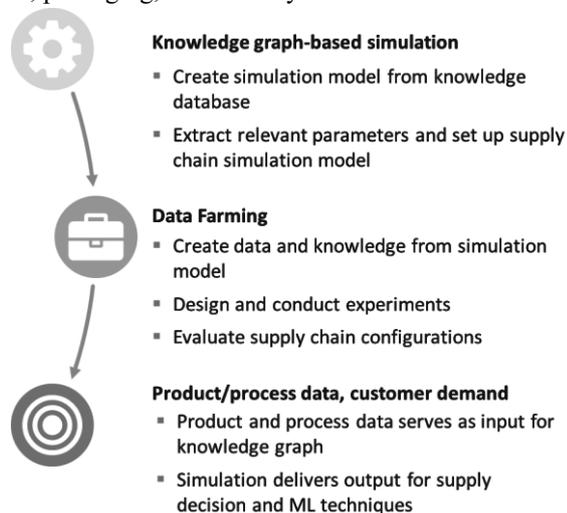


Figure 3: Pipeline for Simulation/ML ©Fraunhofer IWU, Chemnitz

2.5 Development of the solution

The pipeline for the concept is developed (Figure 3) to achieve the beforehand-described needs. To connect the different systems, KNIME, an open-source and versatile

data mining tool, is used. In parallel, it gives the opportunity to work with ML techniques and fulfills the node-oriented interface—the design goals.

Answering the questions concerning the time of an overall lead time is highly dependent on the specific capabilities of the partners, their geographic location, and processing times. This could be evaluated with Anylogistix in the experiments. The different AM technologies have a high influence, such as the powder bed fusion process, which has other optimizing criteria compared to a fused deposition modeling process concerning building room usage and the number of parts manufactured parallel [11].

For the cost assumption, it was mandatory to evaluate, besides the time needed to print the material, the different cost of each resource in the network and their connection to each other [12]. This resulted in a partner and activity-based cost formula for each subject in the value network. The solution was developed by one currently self-defined case, which follows the assumption that each partner could have multiple roles in the supply chain. This role model can be understood on the ontological level as domain-specific knowledge. Additionally, the knowledge for the specific value network is represented by the knowledge graph derived from the database and used by the simulation software. The results are then transferred back to the knowledge graph on the platform. This data will then be used to train the ML algorithms.

2.6 Evaluation

The results of the SE process could not be fully evaluated at the system level but at the level of the parts of the system. The simulation will be run using both simulation tools connected to the KNIME data flow model, and their results will be compared to a deterministic model of the value network. The results of the deterministic model will then be recalculated with classic spreadsheet software to validate the gained solutions.

3 Conclusion and limitations

The research questions could be answered, but as the work is still in process, the concept was proven in part with only one case study, which only includes manufactured parts. It needs to be further evaluated with different and more complex value networks. The study gives an overview of a data-driven platform's overall data flow process involving simulation techniques.

Additionally, Slicing software like Slic3r or hardware-specific tools should be used to better estimate the in-process time. As the data backend is not yet fully available, the data stream as input was manually added, so the amount of data is not yet sufficient to run ML techniques in a meaningful way. This leads to the next step of further including the knowledge graph stack on a proof-of-concept stage.

This also leads to further investigations to assess which type of ML technique is more suitable for the connection with the simulation tools used. Due to the project's characteristics, a difference or performance comparison to other concepts is not made, which would give more insight into possible performance or quality benefits on the decision level.

Acknowledgments

The project on which this report is based was funded by the German Federal Ministry of Economic Affairs and Climate Protection under the funding code MM - MMMKO01416521 - 0I1MK22001G. The responsibility for the content of this publication lies with the authors.

Disclosure of Interests.

The authors have no competing interests to declare that are relevant to the content of this article.

References

- [1] Münnich, M., Zumpe, F., Grzona, P., 2023. Simulation und Maschinelles Lernen in Supply Chains, in: ASIM WS 2023 Kurzbeiträge und Abstract-Beiträge. ASIM Workshop STS/GMMS/EDU 2023. 6.– 7. März 2023. AR-GESIM Publisher Vienna, pp. 15–18.
- [2] VDI 3633 Part 12, 2020. Simulation of systems in materials handling, logistics, and production: Simulation and optimisation. Beuth Verlag GmbH, Berlin, 24 pp.
- [3] Grzona, P., Mai Thi, Y., Zumpe, F., Münnich, M., 2023. Requirements for a ML- and platform-based simulation service. 20. ASIM Fachtagung Simulation in Produktion und Logistik 2023, p. 187.
- [4] Haberfellner, R., Weck, O. de, Fricke, E., Vössner, S., 2019. Systems Engineering. Springer International Publishing, Cham.
- [5] Ahrens, V., 2024. Systems Engineering and Management: Ein Generischer Ansatz Zur Gestaltung und Lenkung Sozio-Technischer Systeme. Springer Vieweg. in Springer Fachmedien Wiesbaden GmbH, Wiesbaden, 612 pp.
- [6] BMWK, 2023. KISS - KI-gestütztes Rapid Supply Network: KI-basierte semantische Plattform zur rapiden Etablierung von Wertschöpfungsnetzwerken in Krisenzeiten. Bundesministerium für Wirtschaft und Klimaschutz. https://www.digitale-technologien.de/DT/Redaktion/DE/Standardartikel/KuenstlicheIntelligenzProjekte/KuenstlicheIntelligenz_Alle_Projekte/ki-projekt_KISS.html. Accessed 9 February 2023.
- [7] Hartmann, H., 2024. Semper-KI Ontology V.09. Institut für Angewandte Informatik (InfAI) e.V. <https://schema.semper-ki.org/>. Accessed 2 June 2024.
- [8] Lachmayer, R., Ehlers, T., Lippert, R.B., 2024. Design for Additive Manufacturing. Springer Berlin Heidelberg, Berlin, Heidelberg.
- [9] Chen, X., Thi, Y.M., Sprigode, J., Riedel, R., Grzona, P., 2024. Partner Selection in Additive Manufacturing Networks. submitted, in: , APMS 2024 Conference (8.-12. September). Production Management Systems for Volatile, Uncertain, Complex and Ambiguous (VUCA) Environments.
- [10] Yin, R.K., 2018. Case study research and applications: Design and methods, Sixth edition ed. SAGE, Los Angeles, Calif., 319 pp.
- [11] Tagliaferri, V., Trovalusci, F., Guarino, S., Venettacci, S., 2019. Environmental and Economic Analysis of FDM, SLS and MJF Additive Manufacturing Technologies. Materials (Basel, Switzerland) 12 (24).
- [12] VDI 4405 Part 1, 2003. Activity-based cost analysis in internal logistics. Beuth Verlag GmbH, Berlin, 24 pp.

The Filtering Effect on Simulated Signals under Consideration of Entropy Methods

Alexander Edthofer^{1*}, Andreas Körner¹

¹Institute of Analysis and Scientific Computing, TU Wien, Wiedner Hauptstraße 8-10, 1040 Vienna, Austria;
*alexander.edthofer@tuwien.ac.at

Abstract. In this work, we investigate the impact of low-pass filters on two entropy methods, Permutation Entropy (PE) and Entropy of Difference (EoD), using simulated noise signals. Colored noise, specifically white, pink and brown noise, was generated and filtered with pass-band frequencies of 30 Hz, 40 Hz and 50 Hz, alongside unfiltered signals. The PE and EoD values were computed to analyze the effects of filtering. The results indicate that both entropy measures decrease with lower pass-band frequencies. PE can effectively distinguish between pink and brown noise with and without a lowpass filter, while EoD shows similar differentiation only with filtering. These findings highlight the sensitivity of entropy measures to lowpass filtering, with implications for their application in EEG analysis.

Introduction

Simulated signals such as noise are often used for modeling and comparison of situations or occurrences of real-world phenomena. Pink noise is very common in biosystems, as these are stochastic, self-organizing and their equilibrium is at the lowest energy level possible [1]. In the context of biomedicine, noise can be thought of as an idealized or abstracted signal. The brain activity of awake humans resembles pink noise when measured by an electroencephalogram (EEG) [2]. During unconsciousness, higher frequencies are not as present compared to the awake stage. Therefore, brain activity in this stage can be better compared to brown noise [3].

In the field of EEG analysis, usually there are band-pass filters applied on the measured signals [4]. The permutation entropy (PE) [5], first introduced in 2002, is a commonly used parameter in research in this field of application [6, 7], the entropy of difference (EoD) [8] is new and not yet established in EEG analysis, but seems promising. In this work, we investigate the impact of lowpass filters on these two entropy methods under the use of raw simulated signals using colored noise.

The effect of linear filters on white noise using the PE has already been shown in [9]. However, no other colored noise was considered. In the case of EoD, no such research has been conducted.

The computations were performed on a laptop with 16 GB RAM, an AMD Ryzen 5 5500U processor with operating system Microsoft Windows 10 Pro using MATLAB version R2023b.

1 Methods

1.1 Noise Signals

Stochastic processes can be used to generate noise signals. The most common ones for model analysis are white, pink and brown noise [10], which also coincide with the simulated signals for the application of EEG analysis. Their power spectral densities (PSD) in general are given by $S(f) = \frac{L(f)}{|f|^\alpha}$ with L being a positive, slowly varying or even constant function. White noise has a uniform distribution of the frequencies, i.e. $S_w(f) = L(f)$, which means $\alpha = 0$. A mathematical description is derived by the time-derivative of a Brownian motion process [11]. Pink noise is defined with $\alpha = 1$ as $S_p(f) = \frac{L(f)}{|f|}$, i.e. higher frequencies appear with lower amplitude. Brown noise has an even stronger decrease than pink as $\alpha = 2$, which results in $S_b(f) = \frac{L(f)}{|f|^2}$. The simulated signals in this work are generated using the MATLAB function `dsp.ColoredNoise`.

The noise signals are compared with two entropy methods. There are different lowpass-filters applied to the signals, which are compared as well. For this, the MATLAB function `lowpass` is used.

1.2 Permutation Entropy

The PE was first introduced in [5]. A given times series $(x_t) = (x_1, \dots, x_N)$ is divided in tuples of length m , which is called the order. For each tuple, an ordinal pat-

tern is determined, for length m there are $m!$ possible combinations. The PE is defined as

$$PE = -\frac{1}{\log(m!)} \sum_{i=1}^{m!} p_i \log p_i, \quad (1)$$

where p_i defines the probability of occurrence of pattern i and the base of the logarithm is two. The coefficient $-\frac{1}{\log(m!)}$ represents a normalization factor, such that $PE \in [0, 1]$. If two values in one tuple are equal, i.e. $x_i = x_j$, we choose the rule that $x_i < x_j$ for $i < j$. A detailed mathematical description is given in [7].

1.3 Entropy of Difference

An alteration of the PE is the EoD, defined by Pasquale Nardone in [8]. Again, a time series (x_t) is divided in the same amount of tuples of length m . For this method, one considers the neighboring values within a tuple and compares them. The encoding is given only by + and -, depending if there is an increase or decrease between the values. If two neighboring values would be equal, we again apply the rule that $x_i < x_j$ for $i < j$. For tuples of length m , there are 2^{m-1} possible patterns that can be achieved. The EoD is then defined as

$$EoD = -\frac{1}{m-1} \sum_{i=1}^{2^{m-1}} p_i \log p_i, \quad (2)$$

where p_i again defines the probability of pattern i . The base of the logarithm is two, such that the coefficient $-\frac{1}{m-1}$ is also a normalization factor, i.e. $EoD \in [0, 1]$.

2 Results

For our study, colored noise with $5 \cdot 10^6$ sample points was created. Firstly, the simulated signal is lowpass filtered by a passband frequency of either 30 Hz, 40 Hz or 50 Hz. Considering also unfiltered signals, this makes a total of four different scenarios. These passband frequencies were chosen as these are also used in the application of EEG analysis [4]. Secondly, the sample was rearranged to 1665 vectors each containing 3000 values. Comparing such a vector to a recorded signal with a sampling frequency of 200 Hz, this would correspond to 15 s. Next, the simulated signals were decoded in the respective patterns of the PE and EoD and afterwards the corresponding entropy value were calculated. The results of the two different entropies are given in the case of orders $m = 3$ and $m = 7$. For

the PE, the most common orders are between 3 and 7 [5]. A graphical representation of the results for the 1665 samples is given in Figure 1 for $m = 3$ using boxplots, which were created using the MATLAB function `boxplotgroup`. The y-axis refers to the entropy value, which are defined in equations (1) and (2), shown between 0.5 and 1 as these are the minimal and maximal value that appear in our study. The four different filter scenarios are indicated on the x-axis.

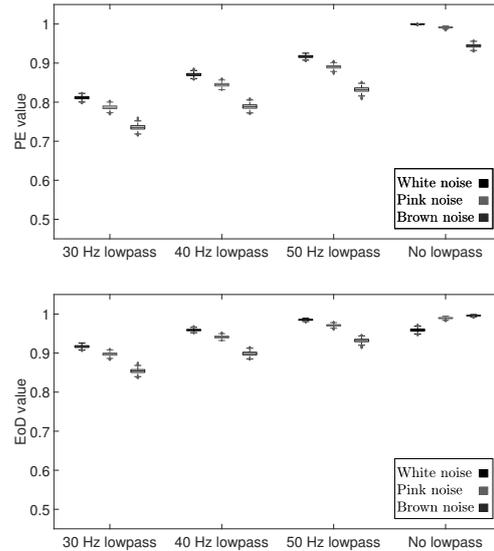


Figure 1: Entropy values of order $m = 3$ and different filters for white, pink and brown noise.

For order 3, a decrease of the PE value from white to pink to brown noise is observable, independent of the applied filter. EoD does not show the same behavior, as for no lowpass filter, there is an increase of the values from white to brown noise. In general, a decrease in the lower passband frequencies is observable for PE and EoD with a more prominent decrease in the case of PE. This is reasonable because, when lowpass filtering is done, fewer patterns occur. The PE can distinguish between pink and brown noise for all four scenarios as none of the respective boxplots overlap. The EoD manages this task only for the three scenarios in which lowpass filtering was applied. For the PE with $m = 3$ and no filtering, the results coincide with the ones of [10].

The results for order $m = 7$ are shown in Figure 2, created with the same function and settings as before. The PE for order 7 shows a similar course to $m = 3$ although the decrease for a lower passband frequency is even stronger. However, it still separates pink and

brown noise well, as the boxplots again are not overlapping. The EoD for order $m = 7$ cannot distinguish between pink and brown noise even for the filtering scenarios. The values for white noise are in all cases lower than those for pink noise.

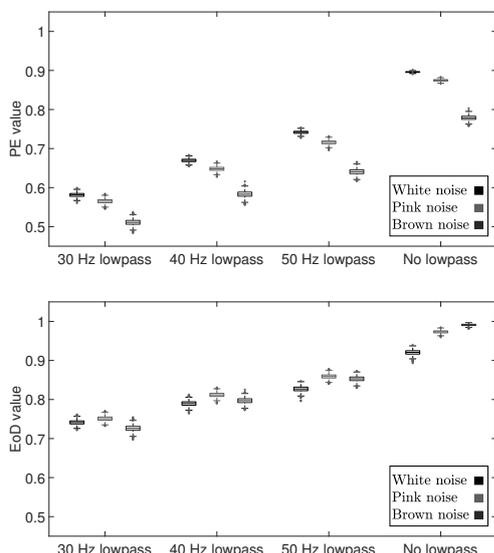


Figure 2: Entropy values of order $m = 7$ and different filters for white, pink and brown noise.

3 Discussion

In this work, we tested different lowpass filters on colored noise and investigated their impact on PE and EoD. Passband frequencies of 30 Hz, 40 Hz and 50 Hz were compared with no lowpass filtering, as these three are most commonly used in the field of EEG analysis [4]. The results show that the EoD behaves differently, if no filter is set. For any other of the three scenarios, EoD behaves similar to the PE for order $m = 3$. For the order $m = 7$, the results differ, as with noise does not achieve the highest EoD value, but pink noise.

One can see as well that the lower the passband frequency was set for the lowpass filter, the lower the entropy values get. For the PE and white noise, this was already indicated in [9]. We also showed the effect on other types of noise as well as a similar impact on the EoD. However, the decrease in value in the latter case is not as strong as for the PE.

We also considered a highpass filter, usually there is a passband frequency of 0.5 Hz, but this did not show any effect on the values of the two entropy methods.

The quality of EoD in comparison to PE, especially in the application of EEG analysis, will be examined in more detail in future studies.

Acknowledgement

The authors would like to thank their research partner the AG Neuromonitoring of Dr. Matthias Kreuzer from the Technical University of Munich for their continued guidance in the field of EEG analysis and, additionally, Clara Horvath for proofreading.

References

- [1] Szendro P, Vincze G, Szasz A. Pink-noise behaviour of biosystems. *European Biophysics Journal*. 2001; 30:227–231.
- [2] Buzsaki G. *Rhythms of the Brain*. Oxford university press. 2006.
- [3] Colombo MA, Napolitani M, Boly M, Gosseries O, Casarotto S, Rosanova M, Brichant JF, Boveroux P, Rex S, Laureys S, et al. The spectral exponent of the resting EEG indexes the presence of consciousness during unresponsiveness induced by propofol, xenon, and ketamine. *NeuroImage*. 2019;189:631–644.
- [4] Sanei S, Chambers JA. *EEG signal processing*. John Wiley & Sons. 2013.
- [5] Bandt C, Pompe B. Permutation entropy: a natural complexity measure for time series. *Physical review letters*. 2002;88(17):174102.
- [6] Berger S, Schneider G, Kochs EF, Jordan D. Permutation entropy: too complex a measure for EEG time series? *Entropy*. 2017;19(12):692.
- [7] Edthofer A, Feldhammer I, Fenzl T, Körner A, Kreuzer M. Permutation Entropy as a Conceptual Model to Analyse Brain Activity in Sleep. *arXiv preprint arXiv:231207129*. 2023;pp. arXiv–2312.
- [8] Nardone P, Sonnino G. Entropy of Difference: A New Tool for Measuring Complexity. *Axioms*. 2024; 13(2):130.
- [9] Dávalos A, Jabloun M, Ravier P, Buttelli O. The impact of linear filter preprocessing in the interpretation of permutation entropy. *Entropy*. 2021;23(7):787.
- [10] Chen Z, Ma X, Fu J, Li Y. Ensemble Improved Permutation Entropy: A New Approach for Time Series Analysis. *Entropy*. 2023;25(8):1175.
- [11] Kasdin NJ. Discrete simulation of colored noise and stochastic processes and 1/f/sup/spl alpha//power law noise generation. *Proceedings of the IEEE*. 1995; 83(5):802–827.

Konzeption eines Vorgehensmodells zur Auswahl von Entscheidungsunterstützungsmethoden in der Produktion und Logistik

Katharina Langenbach^{1*}, Markus Rabe¹

¹Fachgebiet IT in Produktion und Logistik, Technische Universität Dortmund, Leonhard-Euler-Str. 5, 44227 Dortmund, Deutschland; *katharina.langenbach@tu-dortmund.de

Abstract. Die Komplexität der heutigen Produktion und Logistik hat zur Folge, dass wiederkehrende Entscheidungen mit Hilfe von Entscheidungsunterstützungsmethoden (EUM) getroffen werden. Die Wahl einer geeigneten EUM ist jedoch komplex, insbesondere unter Berücksichtigung der zur Verfügung stehenden Daten. Ein Vorgehensmodell unterstützt Anwender aus der Industrie bei dieser Auswahl, indem es eine strukturierte, nachvollziehbare und wiederholbare Herangehensweise vorgibt. Dieser Beitrag beschreibt einen Vorschlag für ein Vorgehensmodell, das Fragestellungen der Produktion und Logistik einordnet und die Wahl einer geeigneten EUM sowie Verfahren zur Aggregation und Disaggregation von Daten unterstützt.

Einleitung

In der Planung, Umsetzung und Steuerung im Bereich der Produktion und Logistik sind vielfältige Entscheidungen zu treffen. Dabei unterscheiden sich die Entscheidungen beispielsweise im Zeithorizont oder in der zu berücksichtigenden Systemkomplexität [1]. Im Rahmen der digitalen Transformation ist insbesondere eine steigende Systemkomplexität zu beobachten [2]. Da Entscheidungen für komplexe Systeme ebenfalls komplex sind, ist es sinnvoll, sich bei den Entscheidungen technisch unterstützen zu lassen.

Für die Entscheidungsunterstützung können unterschiedliche Methoden angewendet werden. Diese Methoden weisen wiederum Anforderungen auf, damit sie sinnvolle Ergebnisse erzeugen. Dies hat zur Folge, dass je nach zu lösender Fragestellung der Entscheidungsunterstützung unterschiedliche Methoden zum Einsatz kommen sollten [1]. Die Auswahl geeigneter Methoden ist dabei herausfordernd und von großer Bedeutung [3].

Alle Entscheidungsunterstützungsmethoden (EUM) verwenden Daten. Diese liegen meist jedoch nicht in der geeigneten Aggregationsstufe vor. Sie sind entweder zu

feingranular oder zu stark verdichtet. Daher kann eine Datenaggregation oder -disaggregation (DAD) erforderlich sein.

In diesem Beitrag wird ein grundsätzliches Vorgehen zur Wahl und Durchführung von EUM vorgestellt, wobei auch die Bereitstellung von Daten in einer geeigneten Aggregationsstufe thematisiert wird.

1 Grundlagen

Die Lösung von Problemen beruht in den meisten Fällen auf der Beantwortung spezifischer Fragen. Durch die Formulierung von Fragen kann eine Problemstellung klar umrissen werden. Dabei kann sich die Art der Fragestellung je nach Domäne oder Managementebene stark unterscheiden. In der Produktion und Logistik treten etwa Fragen zu Auswirkungen von Änderungen der Produktionskapazitäten, der Produktionsplanung oder der hergestellten Produkte auf [4].

Insbesondere, wenn es Zielkonflikte bei Entscheidungen gibt oder die Auswirkungen von Einzelentscheidungen auf andere Entscheidungen betrachtet werden sollen, ist der Einsatz von entscheidungsunterstützenden Methoden sinnvoll [5]. Drei Methoden werden im Folgenden umrissen.

Tabellenkalkulationen oder (einfache) mathematische Modelle werden häufig in der Produktion und Logistik eingesetzt, wenn schnelle und einfache Berechnungen durchzuführen sind und Abhängigkeiten innerhalb eines Systems gut abzubilden sind [6]. Dabei werden häufig Vereinfachungen getroffen, wie statische Durchschnittswerte.

Bei komplexeren Systemen und zur Berücksichtigung von Zufälligkeiten und Wechselwirkungen, die in

der Wirklichkeit eine große Rolle spielen, kann *Simulation* eingesetzt werden [7]. Dabei ist stets die Simulationswürdigkeit der Fragestellung zu beachten [4].

Die Berücksichtigung aller Charakteristika einer Simulation, wie zeitliche oder zufällige Gegebenheiten, ist nicht für alle Fragestellungen aus dem Bereich der Produktion und Logistik erforderlich [8]. So können beispielsweise in der Transport- und Tourenplanung *heuristische* oder *exakte Verfahren* eingesetzt werden [9].

Um die Nachteile einzelner Methoden auszugleichen, kann es sinnvoll sein, hybride Methoden bzw. Kombinationen von Methoden zu verwenden. Dies zeigt sich zum Beispiel bei der Optimierung. Eine Optimierung darf strenggenommen nur im Falle vollständig vorliegender Informationen durchgeführt werden, also insbesondere für deterministische Planungsmodelle. Allerdings sind Unsicherheiten allgegenwärtig, was oft zu stark vereinfachten Modellen von realen Systemen führt [10]. Eine Kombination von entscheidungsunterstützenden Methoden, beispielsweise eine simulationsbasierte Optimierung, kann deshalb nützlich sein.

Jedoch sind nicht für alle Methoden dieselben Daten in der gleichen Form erforderlich. Zusätzlich kann es erforderlich sein, dass zur Beantwortung einer Fragestellung unterschiedliche oder hybride EUM angewendet werden, was wiederum Auswirkungen auf die Daten hat. Dabei ist zu beachten, dass die Qualität der Daten mit über die Vertrauenswürdigkeit der Ergebnisse der Methoden entscheidet [11]. Hierbei spielt insbesondere die Eignung der Daten neben anderen Qualitätsaspekten eine entscheidende Rolle. Unter Eignung ist in diesem Fall vornehmlich ein geeigneter Detaillierungsgrad der Daten zu verstehen. Um Daten anzupassen, können Techniken der Datentransformation angewendet werden [12].

Die Abhängigkeit von betrachteter Fragestellung, Wahl der EUM und vorliegender Daten, die transformiert werden können, verdeutlicht, dass für ein strukturiertes, reproduzierbares und vertrauenswürdigen Vorgehen ein Vorgehensmodell hilfreich sein wird.

2 Konzeptionelles Vorgehen

Zur Beantwortung von komplexen Fragestellungen ist ein strukturiertes, nachvollziehbares und wiederholbares Vorgehen erforderlich. Dabei sollten die Spezifika von Fragestellungen und vorhandenen Randbedingungen beachtet werden. Dennoch sollte möglichst ein allgemeines Vorgehen vorgegeben werden, sodass Anwender systematisch unterstützt werden. Im Folgenden wird hierfür

ein geeignetes Vorgehensmodell vorgeschlagen.

2.1 Entscheidungsprozess

Wenn eine Entscheidung im Bereich der Produktion und Logistik getroffen werden muss, ist dies mit einer oder mehreren Fragen verknüpft (siehe Abschnitt 1). Die Fragen werden stets unter Berücksichtigung von Randbedingungen wie dem betrachteten Zeithorizont beantwortet. Auf Grundlage der Fragen und Randbedingungen wird eine EUM ausgewählt, mit der eine fundierte und nachvollziehbare Entscheidung getroffen werden kann.

Zur Durchführung jeder EUM werden Daten benötigt (siehe Abschnitt 1). An diesem Punkt ergeben sich unterschiedliche Möglichkeiten. Oft liegen Daten in einer nicht geeigneten Aggregationsstufe vor. Würde die gewählte EUM unter Verwendung solcher Daten angewendet werden, hätte dies wenig vertrauenswürdige bis falsche Ergebnisse zur Folge. Daher müssen die Daten in Abhängigkeit von der vorliegenden Fragestellung angepasst werden. Liegen die Daten nach dieser Anpassung in geeigneter Form vor oder lagen sie von Anfang an in passender Form vor, kann die EUM angewendet und können nach Analyse der Ergebnisse Erkenntnisse abgeleitet werden. Basierend auf diesen Erkenntnissen können anschließend Entscheidungen getroffen werden.

2.2 Eingabegrößen

Wie bereits in Abschnitt 2.1 erläutert, spielen mehrere Eingabegrößen sowohl bei der Auswahl von EUM als auch bei der Wahl einer geeigneten DAD eine wichtige Rolle. Die Eingabegrößen beider Teilprozesse weisen Ähnlichkeiten auf und werden im Folgenden kurz beschrieben.

Die Grundlage für die Auswahl sowohl der EUM als auch des Ausmaßes der erforderlichen DAD bildet die zugrundeliegende Fragestellung. Diese legt die Schwerpunkte für die Beantwortung der Fragen fest, gibt Hinweise zur Priorisierung von Zielen bei Zielkonflikten und bestimmt den erforderlichen Detaillierungsgrad der Ergebnisse. Die Fragestellung kann jedoch nicht isoliert betrachtet werden. Sie steht immer im Zusammenhang mit den Randbedingungen wie dem verfügbaren Zeitrahmen, den zur Verfügung stehenden Ressourcen und dem vorhandenen Vorwissen. Entsprechend erfolgt die Auswahl der EUM stets basierend auf der Fragestellung und den Randbedingungen.

Für die Umsetzung der gewählten Methode werden Daten benötigt (siehe Abschnitt 1). Die Struktur und die

Aggregationsstufe der zu verwendenden Daten bestimmen, ob eine EUM angewendet werden kann. Welche Daten grundsätzlich benötigt werden, definiert die Fragestellung. Insbesondere entscheidet sie mit darüber, in welcher Aggregationsstufe die Daten vorliegen müssen. Darüber hinaus beeinflusst die EUM auch die erforderliche Datenstruktur. Zusätzlich müssen weitere Randbedingungen, wie vorhandene Ressourcen zur Umsetzung unterschiedlicher Transformationsmethoden, beachtet werden.

2.3 Gruppierung

Die Auswahl einer EUM sowie der geeigneten DAD erfordert eine gründliche Analyse der gegebenen Umstände. Mögliche auftretende Fragestellungen weisen unterschiedliche Komplexitäten auf. Um die darauf basierende Auswahl einer EUM zu vereinfachen, ist es sinnvoll, Fragestellungen unter Berücksichtigung von gegebenen Randbedingungen zu gruppieren. Anhand der Eigenschaften der Fragestellung sowie der durch die jeweilige Gruppe gegebenen Randbedingungen definierten Anforderungen können Empfehlungen abgeleitet werden, welche EUM genutzt werden sollte. Dies erfolgt durch den Vergleich von Charakteristika der Fragestellungen mit den Anforderungen und Stärken der einzelnen EUM. Die Einbeziehung von Randbedingungen und der durch diese gegebenen Gruppierung ermöglicht, Expertenwissen in den Prozess einzubringen.

Die frühzeitige Wahl einer EUM ist entscheidend, da eine Methode wiederum spezifische Anforderungen an die erforderlichen Daten stellt. Dies betrifft insbesondere die Struktur und Aggregationsstufe der Daten. Anhaltspunkte für die geeignete Aggregationsstufe der Daten liefern die Randbedingungen (siehe Abschnitt 2.2), insbesondere die zugrundeliegende Fragestellung. Die Wahl einer geeigneten DAD stellt dabei eine große Herausforderung für den Anwender dar. Durch die Erstellung von Gruppen ergibt sich auch hier die Möglichkeit, in einer vereinfachten und nachvollziehbaren Weise ein geeignetes Verfahren zu wählen. Dabei erfolgt die Erstellung von Gruppen durch die inhaltliche und von der Fragestellung und den Randbedingungen vorgegebene Datenaggregationsstufe, die Berücksichtigung der vorhandenen Daten und den Abgleich mit den Anforderungen und Potentialen unterschiedlicher DAD-Verfahren ebenfalls unter Berücksichtigung der Randbedingungen.

2.4 Ablauf

Aus den Überlegungen der Abschnitte 2.1 bis 2.3 ergibt

sich ein grundsätzliches Vorgehensmodell, das in Abbildung 1 dargestellt ist. Dabei wird deutlich, dass es sich bei der Wahl einer EUM sowie eines entsprechenden DAD-Verfahrens um ein komplexes und mehrschichtiges Vorgehen handelt. Dieses Vorgehen verfolgt das Ziel, die zugrunde liegende Fragestellung zu beantworten. Der Auswahl geht jeweils die Einordnung von Eingabegrößen (siehe Abschnitt 2.2) in Gruppen (siehe Abschnitt 2.3) voraus, um der Individualität der Fragestellung und der vorliegenden Randbedingungen gerecht zu werden. Erst danach kann die EUM umgesetzt werden. Nach der Umsetzung und Analyse der Ergebnisse der EUM können unterschiedliche Szenarien eintreten: Die Fragestellung kann vollständig beantwortet sein, eine Präzisierung kann erforderlich sein oder es kann sich eine neue Fragestellung ergeben. Dies wird durch den gestrichelten Pfeil in Abbildung 1 veranschaulicht.

Durch den in Abbildung 1 strukturierten Ablauf wird eine nachvollziehbare und reproduzierbare Entscheidung für die Wahl und Umsetzung einer EUM unterstützt. Dadurch ist es Anwendern aus der Industrie möglich, komplexen Fragestellungen aus dem Bereich der Produktion und Logistik zu begegnen.

3 Ausblick

Die in Abschnitt 2 beschriebene Methodik stellt einen ersten Ansatz zur Entwicklung eines Vorgehensmodells dar. Daraus ergeben sich Forschungsbedarfe, die eine detaillierte Ausarbeitung und Anwendung des Vorgehensmodells ermöglichen.

Zunächst sind die zu beantwortenden Fragestellungen zu analysieren, um ihre Komplexität unter Berücksichtigung von Randbedingungen einordnen zu können. Eine solche Einordnung bildet die Grundlage für die Auswahl einer geeigneten EUM. In diesem Zusammenhang ist es zudem erforderlich, die Anforderungen einzelner EUM zu sammeln und zu analysieren. Eine solche Analyse ermöglicht, die Anforderungen und Potentiale von EUM mit den Anforderungen der jeweils betrachteten Fragestellung abzugleichen und darauf basierend die Auswahl der geeigneten Methode zu treffen. Dabei sind insbesondere bestehende Ansätze zur Auswahl von EUM aus der Produktion und Logistik sowie verwandten Bereichen zu berücksichtigen.

Zusätzlich sind Verfahren zur DAD bereitzustellen, um den Einsatz der gewählten EUM zu ermöglichen. Dabei sind die Anforderungen, Stärken und Schwächen der einzelnen DAD-Verfahren zu berücksichtigen.

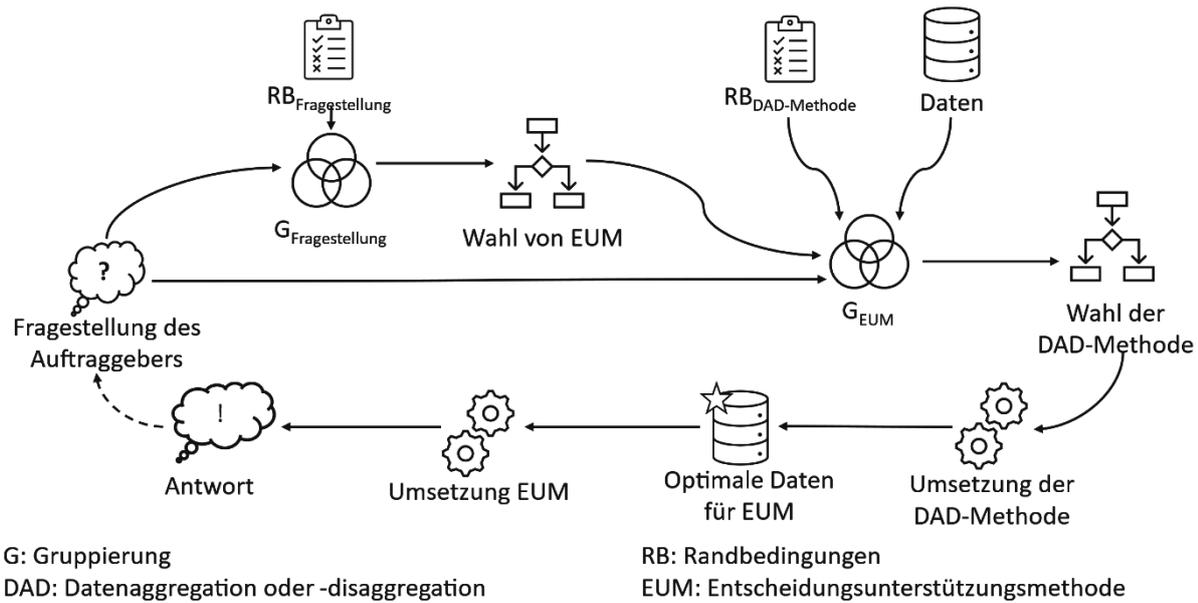


Abbildung 1: Schematische Darstellung des Vorgehensmodells

Um das vorgestellte Vorgehen in der Praxis anwenden zu können, sind die einzelnen Phasen und Phaseergebnisse sowie Rollen innerhalb des Vorgehensmodells zu definieren. Dabei ist eine möglichst automatisierte Verifikation und Validierung der einzelnen Phaseergebnisse zu berücksichtigen, um die Glaubwürdigkeit der generierten Antworten zu erhöhen und das Vertrauen in die Zuverlässigkeit des neu entwickelten Vorgehensmodells zu steigern.

Literatur

- [1] Arnold, D.; Isermann, H.; Kuhn, A.; Tempelmeier, H.; Furmans, K.: Handbuch Logistik. Berlin, Heidelberg: Springer 2008.
- [2] Lang, S.; Schenk, M.; Reggelin, T.: Towards Learning- and Knowledge-Based Methods of Artificial Intelligence for Short-Term Operative Planning Tasks in Production and Logistics: Research Idea and Framework. IFAC-PapersOnLine 52 (2019) 13, S. 2716–2721.
- [3] Renzi, C.; Leali, F.; Di Angelo, L.: A Review on Decision-Making Methods in Engineering Design for the Automotive Industry. Journal of Engineering Design 28 (2017) 2, S. 118–143.
- [4] Bicalho-Hoch, A.L.; Ozkul, F.; Wittine, N.; Wenzel, S.: A Tool-Based Approach to Assess Simulation Worthiness and Specify Sponsor Needs for SMEs. In: Feng, B.; Pedrielli, G.; Peng, Y., Shashaani, S.; Song, E.; Corlu, C.G.; Lee, L.H. et al. (Hrsg.): Proceedings of the 2022 Winter Simulation Conference (WSC), S. 1818–1829, Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.
- [5] Triantaphyllou, E.: Multi-Criteria Decision Making Methods. In: Pardalos, P.M.; Hearn, D.; Triantaphyllou, E. (Hrsg.): Multi-Criteria Decision Making Methods: A Comparative Study. Boston, Massachusetts: Springer US 2000, S. 5–21.
- [6] VDI 3633 Blatt 9: Simulation von Logistik-, Materialfluss- und Produktionssystemen: Tabellenkalkulation (TK) im Umfeld der Simulation. Berlin: Beuth, 2019.
- [7] VDI 3633 Blatt 1: Simulation von Logistik-, Materialfluss und Produktionssystemen: Grundlagen. Berlin: Beuth, 2014.
- [8] Gutenschwager, K.; Rabe, M.; Spieckermann, S.; Wenzel, S.: Simulation in Produktion und Logistik. Berlin, Heidelberg: Springer 2017.
- [9] Günther, H.-O.; Tempelmeier, H.: Supply Chain Analytics: Operations Management und Logistik. Norderstedt: Books on Demand 2020.
- [10] Juan, A.A.; Faulin, J.; Grasman, S.E.; Rabe, M.; Figueira, G.: A Review of Simheuristics: Extending Metaheuristics to Deal with Stochastic Combinatorial Optimization Problems. Operations Research Perspectives 2 (2015), S. 62–72.
- [11] Rabe, M.; Spieckermann, S.; Wenzel, S.: Verifikation und Validierung für die Simulation in Produktion und Logistik: Vorgehensmodelle und Techniken. Berlin, Heidelberg: Springer 2008.
- [12] García, S.; Luengo, J.; Herrera, F.: Data Preprocessing in Data Mining. Cham: Springer International Publishing 2015.

CMOS Technologies for OTAs: A High-Dimensional Comparison

Benedikt Ohse^{1*}, Jürgen Kampe¹, Christopher Schneider¹

¹Ernst-Abbe-Hochschule Jena – University of Applied Sciences, Jena, Germany;
{Benedikt.Ohse, Juergen.Kampe, Christopher.Schneider}@eah-jena.de

Abstract. We present a comparison of two CMOS (complementary metal-oxide-semiconductor) technologies as applied to operational transconductance amplifiers (OTAs). Using multi-objective optimization techniques, we approximate the so-called Pareto fronts of the performance spaces for different OTAs, which allows us to compare their best compromise performance settings. Due to an efficient, parallelized approximation method, we are able to study eight-dimensional performance spaces for practical examples.

1 Introduction

In the field of integrated circuits, new designs are often presented and compared to existing designs. The technology behind their fabrication and simulation is usually left aside. In this paper, we will present an approach to compare different technologies based on operational transconductance amplifiers (OTAs).

Our comparison applies concepts from *multi-objective optimization*, see e.g. [1, 2]. All attainable performances of a circuit (e.g., gain, bandwidth, active area) form its *performance space*. In general, these performances are competitive—increasing one performance usually results in decreasing another one. Therefore, we aim to approximate the so-called *Pareto front* of the performance space, which consists of all non-dominated combinations of performance parameters, see Section 2.1. In literature, this approximation is often obtained by the well-known Normal-Boundary Intersection method (NBI), see e.g. [3, 4, 5]. In contrast to that, we use a *box approximation method* in this paper, which is based on [6, 7], see Section 2.2. A visualization concept known as *Parallel Coordinates* [8] (see Section 2.3) is then applied to compare two different technologies for several OTAs, see Section 3. Therefore, eight-dimensional performance spaces are considered.

2 Multi-Objective Optimization

2.1 Definitions and Problem Statement

All scalars and vectors are written in lowercase (v), matrices in uppercase (M), and sets calligraphic style (\mathcal{S}). Relations for vectors (\leq, \geq) should be understood component-wise.

Design parameters $x \in \mathbb{R}^n$ for integrated circuits (e.g., channel widths and lengths, capacities) are subject to various constraints and form the *design space*

$$\mathcal{D} = \{x \in \mathbb{R}^n \mid c(x) \geq 0, x_\ell \leq x \leq x_u\}.$$

Here, $c: \mathbb{R}^n \rightarrow \mathbb{R}^p$ is a vector-valued function and $[x_\ell, x_u] \subseteq \mathbb{R}^n$ denotes lower and upper bounds. The choice of the design parameters for a circuit determines its performance parameters $f(x) \in \mathbb{R}^m$ (e.g., gain, bandwidth, and power dissipation). They form the *performance space* of a circuit, denoted by

$$\mathcal{P} = \{y \in \mathbb{R}^m \mid y = f(x), x \in \mathcal{D}\}.$$

For the evaluation of $c(x)$ and $f(x)$, computationally expensive SPICE-simulations (Simulation Program with Integrated Circuit Emphasis) are necessary [9]. The performances $f_i(x)$ typically compete with each other, so it is not possible to optimize all of them at the same time. Therefore, we aim at computing compromise solutions—so-called *Pareto solutions*. This can be done by solving the multi-objective optimization problem (MOP)

$$\min_{x \in \mathcal{D}} f(x) = [f_1(x), f_2(x), \dots, f_m(x)] \quad (\text{MOP})$$

For (MOP), a point $\bar{x} \in \mathcal{D}$ is called *Pareto solution*, if there exists no $x \in \mathcal{D}$ with $f_i(x) \leq f_i(\bar{x})$ for all $i \in \{1, \dots, m\}$ and with $f_j(x) < f_j(\bar{x})$ for at least one $j \in \{1, \dots, m\}$. The image point $f(\bar{x})$ is said to be *non-dominated* by all other image points $f(x)$, $x \in \mathcal{D}$.

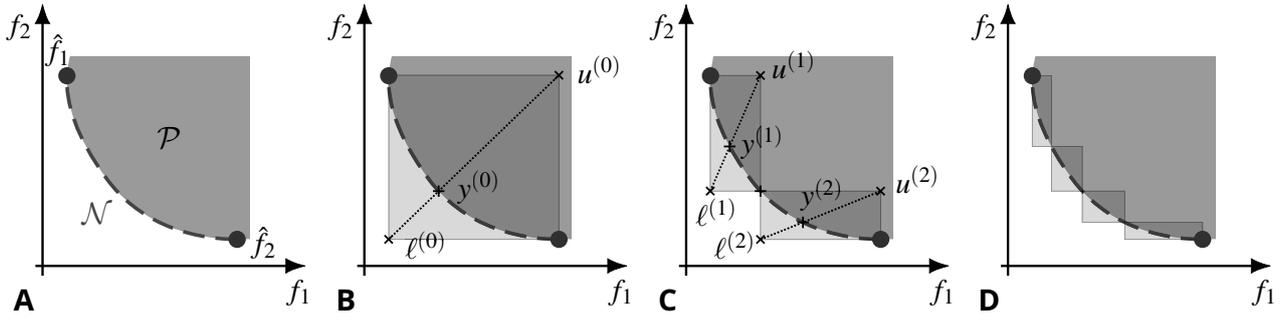


Figure 1: Iterative process of the *box approximation method* for the Pareto front \mathcal{N} of a two-dimensional performance space \mathcal{P} .

A: computation of individual minima **B:** initial box **C:** first box updates **D:** resulting box coverage of \mathcal{N}

The set of all non-dominated image points is called the *Pareto front* \mathcal{N} of (MOP). In the following, we present a method to approximate this set.

2.2 Box Approximation Method

The method to approximate the Pareto front \mathcal{N} is based on the algorithm `AdEnA` [7, Algo. 3]. Some modifications are made, the most important being the parallelization concept. The key idea is the computation of a *box-coverage* for the Pareto front. A box \mathcal{B} is defined by a lower bound $\ell \in \mathbb{R}^m$ and an upper bound $u \in \mathbb{R}^m$:

$$\mathcal{B} = [\ell, u] = \{y \in \mathbb{R}^m \mid \ell \leq y \leq u\}$$

All lower and upper bounds form the sets $\mathcal{L} \subseteq \mathbb{R}^m$ and $\mathcal{U} \subseteq \mathbb{R}^m$, resp. An enclosure of the Pareto front is then given by $(\mathcal{L} + \mathbb{R}_+^m) \cap (\mathcal{U} - \mathbb{R}_+^m)$. The iterative process of the algorithm is illustrated in Figure 1 and briefly described below.

Parallel Approximation

The algorithm starts with an individual minimization of all performances f_i , see Fig. 1-A. The components of the m resulting individual minima \hat{f}_i are then used to define the bounds $\ell^{(0)}, u^{(0)} \in \mathbb{R}^m$ of the initial box \mathcal{B}_0 , see Fig. 1-B.

A refinement of the box-coverage is then iteratively obtained by solving scalarized versions of (MOP), given by

$$\begin{aligned} \min_{x \in \mathcal{D}, t \in \mathbb{R}} \quad & t && (\text{SUP}(\ell, u)) \\ \text{s.t.} \quad & f(x) \leq t(u - \ell) + \ell \end{aligned}$$

Although, the solution \bar{x} of $(\text{SUP}(\ell, u))$ is only *weakly*

Pareto optimal, the presented method still works correctly, see [7, Def. 3.1, Def. 3.2, Lemma 4.7]. The two points $\bar{y} = f(\bar{x})$ and $\hat{y} = \ell + \bar{t}(u - \ell)$ are then used to update the bounds for the child boxes. The lower bound ℓ is updated with [7, Algo. 2] for \hat{y} , and the upper bound u is updated with [7, Algo. 1] for \bar{y} . Using these updated bounds, all child boxes can be defined. This process is shown in Fig. 1-B,C,D.

All of the boxes are independent of each other and can therefore be efficiently computed in parallel. Approximating Pareto fronts for performance spaces can take up to several days (using sequential computation), so parallelization is important for the application.

Stopping Criterion

The algorithm runs until all boxes of the approximation are small enough. To decide if a box is small enough, a stopping criterion is used. Here, we use a combination of two criteria. The first criterion is met if the smallest edge of a box is less than a given tolerance ϵ_1 , see [7]. The second criterion is met if the diagonal of a box is smaller than a given tolerance ϵ_2 .

2.3 High-Dimensional Visualization

Decision makers benefit from good visualizations of performance spaces and their Pareto fronts. In two and three dimensions, such visualizations are possible in Cartesian coordinate systems, see e.g. [5, 10]. For higher dimensions, radar plots or Parallel Coordinates plots [8] can be used. We use the latter, see Fig. 2-4. Here, all coordinate axes are placed in parallel. Each m -dimensional point is then represented as individual polygonal chain with m vertices (coordinate values) on

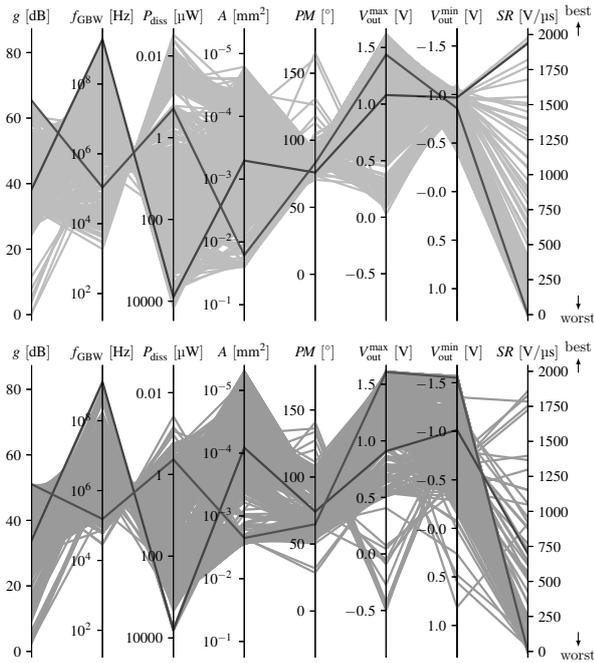


Figure 2: Points on the approximation of the Pareto front \mathcal{N} of the performance space \mathcal{P} of Circuit-1 in ■ 180 nm (top) and ■ 150 nm (bottom). The combinations for the highest ■ g and ■ f_{GBW} are highlighted.

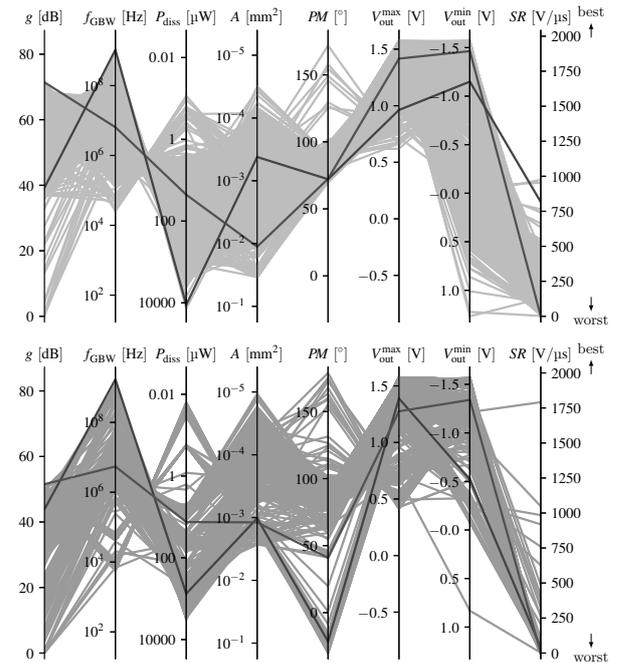


Figure 3: Points on the approximation of the Pareto front \mathcal{N} of the performance space \mathcal{P} of Circuit-2 in ■ 180 nm (top) and ■ 150 nm (bottom). The combinations for the highest ■ g and ■ f_{GBW} are highlighted.

the parallel axes. These plots allow a high level of interaction for decision makers.

3 Technology Comparison

For our comparison, we use three circuits with different complexities. Circuit-1 (lowest complexity) is equivalent to OTA-B-N-1 from [5, Tab. 2], Circuit-2 with a cascode is equivalent to OTA-C-1 from [5, Tab. 2], and Circuit-3 is equivalent to Circuit-1, but with a folded cascode. The eight performance parameters of interest are gain g , bandwidth f_{GBW} , power consumption P_{diss} , active area A , phase margin PM , maximum output voltage V_{out}^{max} , minimum output voltage V_{out}^{min} , and slew rate SR . As technologies, we use a 150 nm and a 180 nm complementary metal-oxide-semiconductor (CMOS) technology from different manufacturers.

The approximations of all Pareto fronts \mathcal{N} are computed by the box approximation method, where we used KNITRO [11] to solve the scalar problems $(SUP(\ell, u))$. Their Parallel Coordinates plots are shown in Fig. 2 (Circuit-1), Fig. 3 (Circuit-2), and Fig. 4 (Circuit-3). The results for the 180 nm technology are shown in

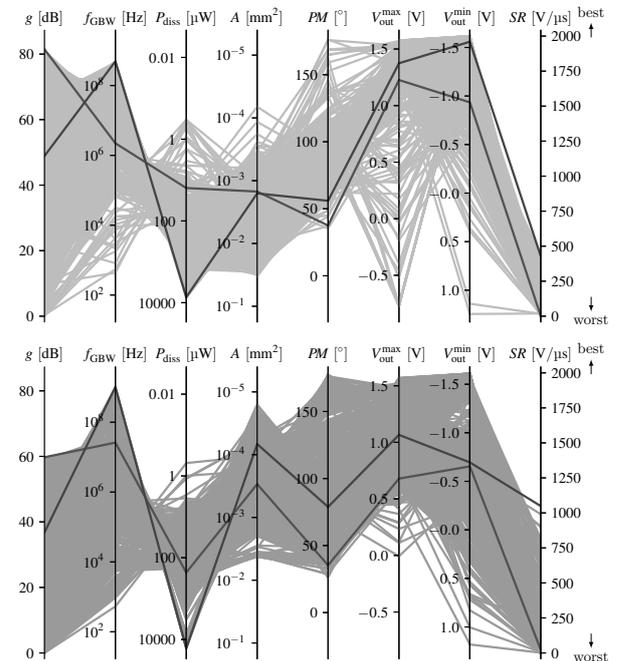


Figure 4: Points on the approximation of the Pareto front \mathcal{N} of the performance space \mathcal{P} of Circuit-3 in ■ 180 nm (top) and ■ 150 nm (bottom). The combinations for the highest ■ g and ■ f_{GBW} are highlighted.

the upper subfigures, the results for 150 nm are shown in the lower subfigures. In each plot, we highlighted the Pareto solutions with highest gain (red) and highest bandwidth (purple), resp. These visualizations allow for comparisons to be made between the designs as well as the technologies.

As expected, the results vary slightly by technology. For each circuit, with the larger technology of 180 nm, higher gains g can be achieved. While the maximum bandwidth f_{GBW} is similar for both technologies, the corresponding gain values, for example, are higher for the larger technology. Often, the smaller technology of 150 nm requires a smaller active area A . Power requirements P_{diss} vary by technology and circuit.

Comparing all the circuits in 180 nm, we see that Circuit-3 (highest complexity) has the highest gain g , but also the lowest bandwidth f_{GBW} . The power dissipation P_{diss} , the area requirement A , and the slew rate SR are best for Circuit-1 (lowest complexity). The phase margin PM is similar for all circuits. Maximum output voltages $V_{\text{out}}^{\text{max}}$ are similar for all circuits, while the minimum output voltage $V_{\text{out}}^{\text{min}}$ is worse for Circuit-1

4 Conclusion

The presented box approximation method and Parallel Coordinates visualization scheme allow a high-dimensional comparison of different technologies for integrated circuits, as shown in the case of OTAs. We compared two CMOS technologies for three OTAs of varying complexity, while simultaneously monitoring eight performance parameters. In the future, libraries of Pareto fronts of performance spaces for different designs and technologies would allow decision makers to choose favorable combinations in the circuit design process.

Acknowledgement

B.O. is supported by the graduate fellowship of Thuringia in Germany.

References

- [1] Ehrgott M. *Multicriteria Optimization*. Springer. 2005.
- [2] Jahn J. *Vector Optimization*. Springer. 2011.
- [3] Das I, Dennis JE. Normal-Boundary Intersection: A New Method for Generating the Pareto Surface in Nonlinear Multicriteria Optimization Problems. *SIAM Journal on Optimization*. 1998;8(3):631–657.
- [4] Mueller-Gritschneider D, Graeb H, Schlichtmann U. A Successive Approach to Compute the Bounded Pareto Front of Practical Multiobjective Optimization Problems. *SIAM Journal on Optimization*. 2009; 20(2):915–934.
- [5] Ohse B, Schreiber D, Kampe J, Schneider C. Efficient Approximation of Performance Spaces for Analog Circuits via Multi-Objective Optimization. In: *2023 Design, Automation & Test in Europe Conference & Exhibition (DATE)*. Antwerp, Belgium: IEEE. 2023; pp. 1–6.
- [6] Eichfelder G, Warnow L. An approximation algorithm for multi-objective optimization problems using a box-coverage. *Journal of Global Optimization*. 2022; 83(2):329–357.
- [7] Eichfelder G, Warnow L. Advancements in the computation of enclosures for multi-objective optimization problems. *European Journal of Operational Research*. 2023;310(1):315–327.
- [8] Inselberg A, Dimsdale B. Parallel coordinates: a tool for visualizing multi-dimensional geometry. In: *Proceedings of the First IEEE Conference on Visualization: Visualization '90*. San Francisco, CA, USA: IEEE Comput. Soc. Press. 1990; pp. 361–378.
- [9] Nagel LW. SPICE2: A Computer Program to Simulate Semiconductor Circuits. PhD Thesis, EECS Department, University of California, Berkeley. 1975. Issue: UCB/ERL M520.
- [10] Schreiber D, Ohse B, Kampe J, Schneider C. Performance Space Supported Design of Analog Electronic Circuits. *IFAC-PapersOnLine*. 2022; 55(20):13–18.
- [11] Byrd RH, Nocedal J, Waltz RA. Knitro: An Integrated Package for Nonlinear Optimization. In: *Large-Scale Nonlinear Optimization*, edited by Pardalos P, Di Pillo G, Roma M, vol. 83, pp. 35–59. Boston, MA: Springer US. 2006; Series Title: Nonconvex Optimization and Its Applications.

Data-adaptive Dynamic Simulation via Structured Dynamic Mode Decomposition

Christopher Bonenberger^{1,2*}, Stephan Scholz^{1,2}, Nico Scheiter¹

¹Ravensburg-Weingarten University of Applied Sciences (RWU), 88250 Weingarten, Germany

²Ulm University, 89081 Ulm, Germany; *bonenbch@rwu.de

Abstract. We consider the task of system identification and simulation via Dynamic Mode Decomposition (DMD) and physics-informed DMD for a simple heat conduction problem. In this regard, we consider the trade-off between data-driven and model-based simulation at the example of the one-dimensional heat equation. Thereby, we highlight the similarities between *dynamic simulation* and *data-driven modeling*. More precisely, we show how *physics-informed DMD* can be used to learn a data-adaptive finite-difference model and how this relates to the inherent limitations of finite-difference simulation.

Introduction

In recent years the scientific community has witnessed remarkable progress in data-driven modeling based on different machine learning (ML) models. However, the universality and effectiveness of these models—most often *Deep Learning* (DL) models—usually comes at the price of explainability and the need for large data sets. Scientific machine learning (SciML) tackles these issues by task-specific modifications of classical ML algorithms (e.g. incorporating prior knowledge leads to *domain-aware* or *physics-guided ML*). These techniques allow to utilize information provided through available data while leveraging prior knowledge about the system at hand. This ideally results in interpretable and robust models, which are able to incorporate physical laws and allow for effective use of data.

In the following we consider simulation (and model design in particular) at the intersection of data-driven and model-driven approaches. Based on the prominent example of heat conduction (e.g. [1], p. 475) we use several types of Dynamic Mode Decomposition (DMD, [2]) that implement different levels of *informedness* up to a data-adaptive method that basically implements a finite-difference model as known from dynamic simulation (e.g. [3]). The resulting scenario allows an almost seamless transition of decided system modeling to

data-adaptive system identification. In this setting, we show how well-known results from numerical analysis and computer simulation can help to understand and improve the use of ML techniques.

1 Methods

In the following we briefly review the required methodologies. For the sake of brevity a detailed consideration of basic DMD, physics-informed DMD and finite-difference methods is foregone. Instead reference is made to [2], [1], [4] and [3] respectively.

1.1 DMD, physics-informed & structured DMD

In short, *basic DMD* is concerned with data-driven estimation of a linear operator that approximates the dynamics of a system at hand. More precisely, assume we are given data $\mathbf{X} \in \mathbb{R}^{N \times T}$ consisting of T observations $\mathbf{x}_t \in \mathbb{R}^N$ that are equidistantly sampled in time. The system dynamics that map \mathbf{x}_t to \mathbf{x}_{t+1} may now be approximated by solving the optimization problem

$$\min_{\mathbf{A} \in \mathbb{R}^{N \times N}} \left\{ \sum_{t=1}^{T-1} \|\mathbf{x}_{t+1} - \mathbf{A}\mathbf{x}_t\|_2^2 \right\}. \quad (1)$$

The goal function may be rewritten as $\|\mathbf{X}_2 - \mathbf{A}\mathbf{X}_1\|_F$, where $\mathbf{X}_1, \mathbf{X}_2 \in \mathbb{R}^{N \times T-1}$ hold the first and last $T-1$ columns of \mathbf{X} respectively and $\|\cdot\|_F$ denotes the Frobenius norm. Now the direct least-squares solution may be written in terms of the pseudo-inverse $\mathbf{X}_1^\dagger = \mathbf{V}\Sigma^\dagger\mathbf{U}^\top$ using the Singular Value Decomposition (SVD) $\mathbf{X}_1 = \mathbf{U}\Sigma\mathbf{V}^\top$ (e.g. [2]). Typically, a low-rank approximation \mathbf{A}_r of \mathbf{A} is found by projection onto the leading left singular vectors, which amounts to $\mathbf{A}_r = \mathbf{U}_r\mathbf{A}\mathbf{U}_r$. In this regard the columns of $\Psi = \mathbf{U}\Phi$, where Φ are eigenvectors $\mathbf{A}_r\Phi = \Phi\Lambda_r$, are referred to as *dynamic modes*.¹

¹In general, approximating system dynamics by a linear operator is tied Koopman operator theory, which allows to trade nonlinear

With regard to the choice of r we follow [5].

Now, *physics-informed DMD* (piDMD, see [4]) is a regularized form of basic DMD. In short, we solve eq. (1) subordinate to certain constraints on \mathbf{A} , which implement prior knowledge about the system at hand. As an example, constraining \mathbf{A} to be unitary implements a conservative system. In the following, we will make use of piDMD for *spatially local* systems, i.e. we constrain \mathbf{A} to be a banded matrix in order to model local dependencies of diffusive systems (cf. [4]).

Finally, we make use of *structured DMD* (sDMD, see [6]), which regularizes basic DMD by writing \mathbf{A} in terms of Q elementary matrices \mathbf{S}_q and seeking an associated coefficient vector $\mathbf{a} \in \mathbb{R}^q$, i.e.

$$\min_{\mathbf{a} \in \mathbb{R}^q} \left\{ \sum_{t=1}^{T-1} \left\| \mathbf{x}_{t+1} - \left(\sum_{q=1}^Q a_q \mathbf{S}_q \right) \mathbf{x}_t \right\|_2^2 \right\}.$$

This amounts to solving the linear system $\mathbf{Z}\mathbf{a} = \mathbf{c}$ with

$$[\mathbf{Z}]_{i,j} = \sum_{t=1}^{T-1} \mathbf{x}_t^T \mathbf{S}_j^T \mathbf{S}_i \mathbf{x}_t \text{ and } [\mathbf{c}]_i = \sum_{t=1}^{T-1} \mathbf{x}_t^T \mathbf{S}_i^T \mathbf{x}_{t+1} \quad (2)$$

with $\mathbf{Z} \in \mathbb{R}^{Q \times Q}$ (see [6]). By choosing appropriate elementary matrices \mathbf{S}_q , sDMD can be used to incorporate prior knowledge, e.g. choosing Toeplitz matrices \mathbf{S}_q leads to a spatially shift-invariant system. We will use this formalism in order to realize data-adaptive system identification based on the finite-difference methods.

1.2 Diffusive Systems, Solutions & Simulation

We consider a simple one-dimensional insulated diffusive system, namely heat conduction on a bar of length L . This implies Neumann boundaries, i.e. $x'(t,0) = x'(t,1) = 0$, for the diffusion equation

$$\frac{\partial x(t,s)}{\partial t} = \alpha \frac{\partial^2 x(t,s)}{\partial s^2} \quad s \in [0,L], t > 0, \quad (3)$$

where $x(t,s)$ denotes the temperature at point s and time t and $\alpha \in \mathbb{R}^+$ is the *diffusivity constant*. The analytical solution is found from the convolution integral $g(t,s) * x(0,s) = x(t,s)$, where $x(0,s)$ denotes the initial distribution and $g(t,s)$ denotes a Gaussian distribution (e.g. [1], p. 475 et seqq.). The spatial convolution

finite-dimensional systems for infinite-dimensional linear systems (provided an appropriate *observable function*). However, in this work we are dealing with linear systems only.

amounts to pointwise multiplication in frequency domain and considering the boundary conditions yields

$$x(t,s) = \sum_{\gamma=0}^{\infty} \hat{c}_\gamma \exp\left(-4\alpha t \left(\frac{\pi\gamma}{2L}\right)^2\right) \cos\left(\frac{2\pi\gamma s}{2L}\right), \quad (4)$$

where \hat{c}_γ denotes the cosine Fourier series coefficients² of the initial distribution $x(0,s)$.

Beneath the analytical solution we consider numerical solutions to eq. (3) based on finite differences (e.g. [3], p. 179). A one dimensional centered space finite difference scheme can be written as

$$\mathbf{x}_{k+1} = \frac{\alpha\Delta t}{(\Delta s)^2} \mathbf{D}\mathbf{x}_k + \mathbf{x}_k, \quad k \in \mathbb{N}, \quad (5)$$

where $\mathbf{x}_k := \mathbf{x}((k+1)\Delta t)$, Δt and Δs denote temporal and spatial sampling intervals and \mathbf{D} is a tridiagonal matrix that is *almost* Toeplitz (except for the boundaries terms), i.e.

$$\mathbf{D} = \begin{bmatrix} -2 & 2 & 0 & \cdots & 0 \\ 1 & -2 & 1 & 0 & \cdots & 0 \\ & & & \ddots & & \\ 0 & \cdots & 0 & 1 & -2 & 1 \\ 0 & \cdots & & 0 & 2 & -2 \end{bmatrix} \in \mathbb{R}^{N \times N}. \quad (6)$$

Thereby the condition on stability is (cf. [3], p. 181)

$$C = \frac{2\alpha\Delta t}{(\Delta s)^2} \leq 1 \text{ or rather } \Delta t \leq \frac{(\Delta s)^2}{2\alpha}. \quad (7)$$

1.3 piDMD and sDMD for Diffusive Systems

In the context diffusive systems as described above, it is reasonable to use piDMD for spatially local systems (cf. [4]). In respect of eq. (5), particularly the matrix \mathbf{D} , we impose tri-diagonality constraints on \mathbf{A} . A solution to this kind of piDMD is provided in [4], p. 15.

A more specific modeling is possible via sDMD. Here, we focus on the following three different variants.

FD-sDMD. Let $\mathbf{T}_{\mathbf{c},\mathbf{r}} \in \mathbb{R}^{N \times N}$ denote a quadratic Toeplitz matrix defined by the vectors $\mathbf{c}, \mathbf{r} \in \mathbb{R}^N$, its first column and row respectively (necessarily their first components are equal, i.e. $[\mathbf{c}]_1 = [\mathbf{r}]_1$). Moreover, let $\mathbf{E}_{i,j} \in \mathbb{R}^{N \times N}$ denote a zero matrix with a single component in row i and column j being one, i.e. $[\mathbf{E}_{i,j}]_{i,j} = 1$.

² $x(0,s)$ can be expressed by such a series via symmetric extension, leading to $2L$ instead of L in the denominators of eq. (4).

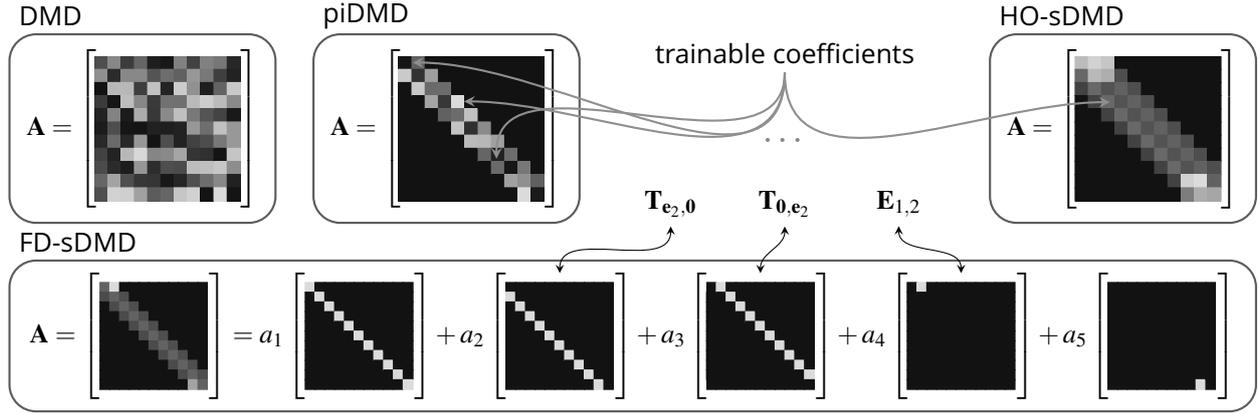


Figure 1: An overview about the different variants of DMD and coefficients that are trained (with $N = 11$). In the matrices \mathbf{A} each color represents a different trainable coefficient. For FD-sDMD the sum of elementary matrices is explicitly shown.

Finally, \mathbf{I} is the identity matrix and \mathbf{e}_k denotes its k -th column. Now, we may express the structure of eq. (5) via sDMD in terms of the elementary matrices

$$\mathbf{S}_1 = \mathbf{I}, \mathbf{S}_2 = \mathbf{T}_{\mathbf{e}_2, \mathbf{0}}, \mathbf{S}_3 = \mathbf{T}_{\mathbf{0}, \mathbf{e}_2}, \mathbf{S}_4 = \mathbf{E}_{1,2}, \mathbf{S}_5 = \mathbf{E}_{N, N-1}.$$

In this setting, sDMD yields an adaptive finite-difference model and that also learns boundary conditions from data (FD-sDMD, see fig. 1).

PE-sDMD. On top of sDMD one might incorporate more prior knowledge about the discretized system \mathbf{D} (this implies knowledge about the boundary conditions). In this case sDMD reduces to parameter estimation (PE-sDMD), since we solely consider a single structure matrix $\mathbf{S}_1 = \mathbf{D}$ (cf. eq. (6)). More precisely, in consideration of eq. (5) we solve

$$\min_{a_1 \in \mathbb{R}} \left\{ \sum_{t=1}^{T-1} \|\mathbf{x}_{t+1} - a_1 \mathbf{S}_1 \mathbf{x}_t - \mathbf{x}_t\|_2^2 \right\},$$

such that $a_1 = \sum_t \mathbf{x}_t^T \mathbf{S}_1^T (\mathbf{x}_{t+1} - \mathbf{x}_t) / \sum_t \mathbf{x}_t^T \mathbf{S}_1^T \mathbf{S}_1 \mathbf{x}_t$ (cf. eq. (2)) and $\mathbf{A} = a_1 \mathbf{S}_1 + \mathbf{I}$.

HO-sDMD. In contrast to the refinement of FD-sDMD to PE-sDMD, we might also be interested in more flexible structures. As there are countless possibilities, it is reasonable to follow familiar models. Here we consider a structure that loosely resembles a higher-order finite difference scheme (HO-sDMD) via

$$\mathbf{S}_1 = \mathbf{I}, \mathbf{S}_2 = \mathbf{T}_{\mathbf{e}_2, \mathbf{0}}, \mathbf{S}_3 = \mathbf{T}_{\mathbf{0}, \mathbf{e}_2}, \mathbf{S}_4 = \mathbf{T}_{\mathbf{e}_3, \mathbf{0}}, \mathbf{S}_5 = \mathbf{T}_{\mathbf{0}, \mathbf{e}_3}$$

and $\mathbf{S}_6, \dots, \mathbf{S}_{15}$ being matrices $\mathbf{E}_{i,j}$ that dissolve the Toeplitz structure at the boundaries. Visualizations of the structures resulting from DMD, piDMD and the different variants of sDMD are provided in fig. 1.

2 Experiments & Discussion

Now, we present some results of a detailed comparison of the aforementioned variants of DMD and finite-difference simulation at the example of flow prediction for a simple diffusive system. The special aspect of this setup is our ability to bring together these two different paradigms in a novel way, i.e. we are able to *interpolate* between data-driven and model-driven simulation.³

Experimental Setup. We consider the problem of heat conduction at the example of a unit length copper rod, i.e. we fix $L = 1\text{m}$ and set α based on the specific thermal capacity, thermal conductance and density of copper ($\alpha \approx 0.1\text{m}^2\text{s}^{-1}$). Moreover, we consider the time interval $t \in [0, 1]$. For all following numerical evaluations the initial (heat) distribution is described by a second order polynomial with random coefficients, i.e.

$$x(0, s) = p_0 + p_1 s + p_2 s^2, \quad p_0, p_1, p_2 \sim \mathcal{N}(0, 1),$$

where $\mathcal{N}(0, 1)$ denotes a zero mean and unit variance normal distribution. Exemplary spatiotemporal data is shown in fig. 2. The analytical solution is computed according to section 1.2 using the discrete Fourier trans-

³Code is available under <https://github.com/cmab92/ASIM2024-DD-MD-DMD>.

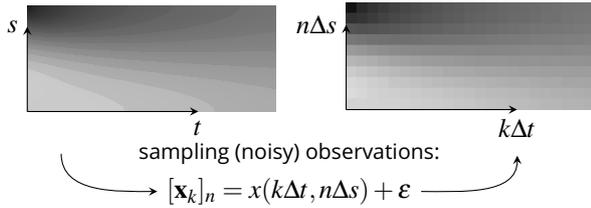


Figure 2: Analytical solution of a 1D heat conduction problem (left panel) and sampled data with $N = 11$ and $\Delta t \approx 0.043\text{s}$ such that $C = 1$. The measurement noise is modeled via additive Gaussian noise, i.e. $\varepsilon \sim \mathcal{N}(0, \sigma^2)$ (by default $\sigma^2 = 0.005$).

form to approximate the coefficients of the initial distribution. Discrete data is generated by equidistant sampling of the analytical solution at N points in space. Temporal sampling is bound to eq. (7) and we consider different stability settings by adjusting C (cf. fig. 3). As indicated in fig. 2 the sampling process is combined with additive noise. As usual for data-adaptive algorithms, we perform a train-test split, i.e. given T observations $\mathbf{x}_t \in \mathbb{R}^N$ a fraction of $\lceil RT \rceil$ observations is used for training and remaining data is used to evaluate performance. We measure the average relative prediction error over all test samples. Simulation-based predictions use the last available training sample.

Results. The different graphs in fig. 3 show relative prediction errors for varying train-test ratio, noise level and sampling respectively. The top graph shows that model-driven prediction is clearly superior to data-driven models although the latter improve with increasing train-test ratio R . The middle panel indicates that model-building also pays off in the presence of noise and again more structure increases performance. Finally, the results in the lower panel allow to draw the conclusion that especially the compromise between data-driven and model-driven solutions can be very effective. Notably FD-sDMD is quite robust against a violation of the stability condition.

Outlook. Our results are a first attempt to “simulation-informed DMD”. We have shown at a simple example, that DMD and piDMD can be improved by incorporating techniques that are typically used in dynamic simulation. The results are promising and we believe that a similar approach can be used for different ML algorithms. However, future work should focus on an evaluation of the proposed technique in

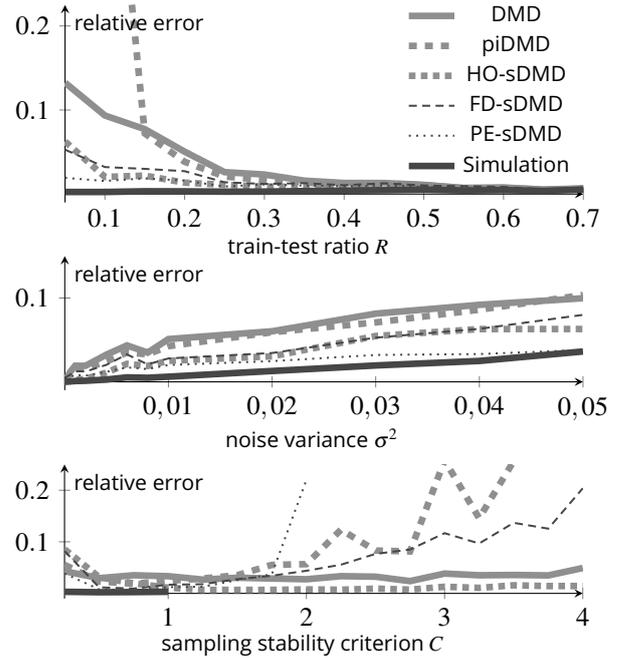


Figure 3: Relative prediction error—the median of 100 runs—for different parameter sweeps. Missing values indicate numerical instability. By default we use $C = 1, \sigma^2 = 0.005$ and $R = 0.25$ (with $N = 21$).

more complex real-world applications.

References

- [1] Bracewell R. *The Fourier transform and its applications*, vol. 3. McGraw-Hill. 2000.
- [2] Schmid PJ. Dynamic mode decomposition and its variants. *Annual Review of Fluid Mechanics*. 2022; 54:225–254.
- [3] Özişik MN, Orlande HR, Colaço MJ, Cotta RM. *Finite difference methods in heat transfer*. CRC press. 2017.
- [4] Baddoo PJ, Herrmann B, McKeon BJ, Nathan Kutz J, Brunton SL. Physics-informed dynamic mode decomposition. *Proceedings of the Royal Society A*. 2023;479(2271):20220576.
- [5] Gavish M, Donoho DL. The optimal hard threshold for singular values is $4/\sqrt{3}$. *IEEE Transactions on Information Theory*. 2014;60(8):5040–5053.
- [6] Bonenberger C, Scholz S, Schneider M. From Data-Driven to Model-Driven Learning via Structured Dynamic Mode Decomposition. In: *63rd IEEE Conference on Decision and Control (CDC)*. IEEE. 2024; .

Simulation and Control of 2-Dimensional Anisotropic Heat Conduction

Stephan Scholz^{1a*,2}, Christopher Bonenberger^{1b,2}, Nico Scheiter¹, Lothar Berger^{1a}

¹Hochschule Ravensburg-Weingarten (RWU), 88250 Weingarten, Germany

*stephan.scholz@rwu.de, ^a Control and Process Engineering, ^b Institut für Künstliche Intelligenz

² Universität Ulm, 89081 Ulm, Germany

Abstract. In this contribution, we provide a simulation and control approach for a two-dimensional heat conduction problem. In particular, we spatially approximate the two-dimensional heat conduction problem with anisotropic thermal conductivity and transfer it via eigenvalue computation to a sampled-time state space model. Furthermore, we design a static feed-forward filter to reach a reference temperature and a full state feedback with linear-quadratic regulation to guarantee a stable closed-loop behavior. Finally, we demonstrate the performance of the derived closed-loop system.

Introduction

Many industries, like semiconductor fabrication, require challenging thermal processing which need to be simulated and controlled very precisely, see e.g. [1]. The thermal dynamics modeling and its control design cover a wide range in the literature, depending on the geometry and research focus, see e.g. [1, 4] and [2, Ch. 2, 6, 8, 9]. The control design for small models can be simple and practical but they are limited for enhancements. In contrast, purely theoretical approaches might be too complex for realistic applications. In our contribution, we propose an extendable and easy-to-implement approach as a 2-dimensional geometry with multiple actuators along one boundary side and multiple sensors on the opposite. This is a simplified model of the realistic 3-dim. situation, see [3]. We approximate the spatial derivatives via finite differences and obtain a high-dim. state space, which is solvable through eigenvalue computation in Section 1. Based on these results, we derive the time discrete solution, the state feedback with linear-quadratic regulation and the reference tracking in Section 2. Finally, we visualize in Section 3 the proper operating closed-loop behavior.

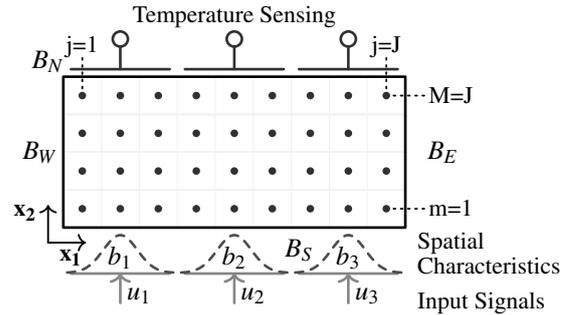


Figure 1: Geometry with finite difference nodes $x_{j,m}$, actuation on B_S and measurement on B_N .

1 Two-dimensional Heat Conduction

We assume a rectangle $\Omega = (0, L) \times (0, W)$ with length $L > 0$, width $W > 0$, see Fig. 1. We note the position $x = (x_1, x_2)^\top \in \bar{\Omega}$. The rectangle has the boundary $\partial\Omega = \bar{\Omega} \setminus \Omega = B_W \cup B_E \cup B_S \cup B_N$ with the sides $B_W = \{0\} \times [0, W]$ (west), $B_E = \{L\} \times [0, W]$ (east), $B_S = [0, L] \times \{0\}$ (south) and $B_N = [0, L] \times \{W\}$ (north). The object consists of a solid material with density $\rho > 0$, specific heat capacity $c > 0$ and anisotropic thermal conductivity $\lambda_1 > 0$ along x_1 -axis (or length) and $\lambda_2 > 0$ along x_2 -axis (or width). The anisotropy describes the physical situation to conduct heat faster along one axis compared to the other axis. We summarize these material properties as diffusivity constants $\alpha_1 := \frac{\lambda_1}{c\rho}$ and $\alpha_2 := \frac{\lambda_2}{c\rho}$. The evolution of temperature in the plate $\theta : [0, T] \times \Omega \rightarrow \mathbb{R}$ solves the two-dimensional heat equation

$$\frac{d}{dt}\theta(t, x) = \alpha_1 \frac{\partial^2}{\partial x_1^2} \theta(t, x) + \alpha_2 \frac{\partial^2}{\partial x_2^2} \theta(t, x) \quad (1)$$

the eigenvalues μ_i and eigenvectors ψ_i of A with $i \in \{1, \dots, JM\}$ then we find the matrix exponential as

$$\exp(A t) = V^{-1} \text{diag}(e^{\mu_1 t}, \dots, e^{\mu_{JM} t}) V \quad (9)$$

with $V = [\psi_1, \dots, \psi_{JM}]$. The article [5] states explicit formulas for the pair (μ_i, ψ_i) for a 1-dim. heat equation problem, which we extend for our 2-dim. problem. We assume $f(z, n) := \cos(z(2n-1)\pi)$ and state the eigenvalues of the 2-dim. problem as

$$\begin{aligned} \mu_{j,m} = & -2p_1 [1 - f([j-1]/J, 1)] \\ & -2p_2 [1 - f([m-1]/M, 1)] \end{aligned} \quad (10)$$

with $(j, m) \in \{1, \dots, J\} \times \{1, \dots, M\}$, $p_l = \alpha_l / \Delta x_l^2$, $l \in \{1, 2\}$, and the eigenvectors $\psi_i = (\psi_{i,1}, \dots, \psi_{i,JM})^\top$ via

$$\psi_{(j,m),(\tilde{n}_j, \tilde{n}_m)} = f\left(\frac{j-1}{2J}, \tilde{n}_j\right) f\left(\frac{m-1}{2M}, \tilde{n}_m\right) \quad (11)$$

using indices $(\tilde{n}_j, \tilde{n}_m) \in \{1, \dots, J\} \times \{1, \dots, M\}$. The proof of this assumption is omitted here, the correctness of (μ_i, ψ_i) can be verified by evaluating $A\psi = \mu\psi$ with cosine identities.

2 Controller Design

We sample the temperature in time as $\Theta(n\Delta T) =: \Theta(n)$ for $n \in \{0, \dots, \lfloor \frac{T}{\Delta T} \rfloor\}$ with sampling time $\Delta T > 0$ and derive from Eqs. (8-11) the time-sampled system

$$\Theta(n+1) = A_D \Theta(n) + B_D u(n) \quad (12)$$

with matrices

$$\begin{aligned} A_D &:= \exp(A\Delta T) = V^{-1} \text{diag}(e^{\mu_1 \Delta T}, \dots, e^{\mu_{JK} \Delta T}) V, \\ B_D &:= \int_0^{\Delta T} \exp(A[\Delta T - \tau]) B d\tau. \end{aligned}$$

The input signal $u(n) := -K \Theta(n) + W r(n)$ is designed such that a state feedback $K \in \mathbb{R}^{N_u \times J K}$ stabilizes the system and a static feed-forward filter $W \in \mathbb{R}^{N_u \times N_y}$ steers the measured temperatures to a static reference signal $r \in \mathbb{R}^{N_y}$. The feedback matrix K is found by solving the discrete infinite-horizon linear-quadratic regulator problem

$$\min_u J = \sum_{n=1}^{\infty} \Theta(n)^\top Q \Theta(n) + u(n)^\top R u(n)$$

with subject to Eq. (12). The feedback matrix K is computed numerically by solving the equation $K = [R + B_D^\top P B_D] B_D^\top P A_D$ in which matrix P is the solution of the discrete time algebraic Riccati equation

$$P = Q + A_D^\top P A_D - [A_D^\top P B_D] [R + B_D^\top P B_D]^{-1} [B_D^\top P A_D].$$

If the number of actuators equals the number of sensors as $N_u \equiv N_y$, then we find the filter matrix W via the assumptions $[A_D - B_D K] \Theta(n) + B_D W r \rightarrow \Theta(n)$ and $y(n) = C \Theta(n) \rightarrow r$ for $n \rightarrow \infty$ as

$$W = -[C(A_d - B_d K - I)^{-1} B_d]^{-1}.$$

For more details on the control design, we refer to introductory text books, e.g. [6, Ch. 7].

3 Simulation Case Study

We apply the proposed concepts on an example of a steel plate with model parameters as in Table 1 and controller configuration as in Table 2. The initial temperature distribution is $\theta_0(x_1, x_2) = 10 \sin(2\pi \frac{x_1}{L})$. We design the controller such that it is forced to act quickly $Q \gg R$ and the static filter shall steer the three measured temperatures towards the reference value $r = (5, 5, 5)^\top$. The computed input signal and the resulting measurement temperatures are visualized in Fig. (2, 3) and show a convergent behavior. In Fig. 4 we notice the transient temperature evolution along the boundary side B_N . We implemented the simulation with JULIA programming language [7] and solved the algebraic Riccati equation with the library `MATRIXEQUATIONS.JL` [8]. The full source code is available [9].

Discussion & Conclusion

In real world applications, we have to deal with thermal emissions like convection and radiation on each boundary side, which we did not consider in this contribution. Moreover, in 3-dim. objects we can only measure the temperatures on the boundary, not inside the object.

L	W	J	K	(λ_1, λ_2)	ρ	c
0.3	0.1	30	10	(40, 60)	8000	400

Table 1: System Parameters for Simulation

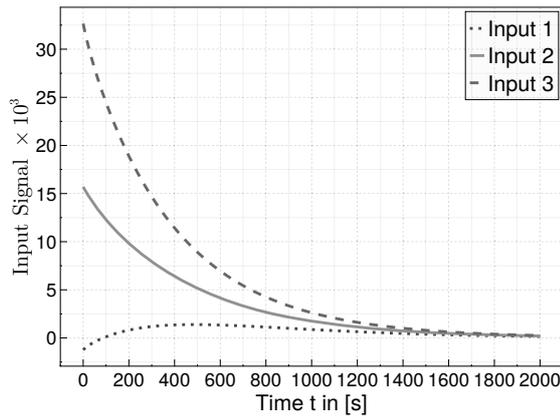


Figure 2: Input signals u_n with $n \in \{1,2,3\}$ applied on boundary side B_S .

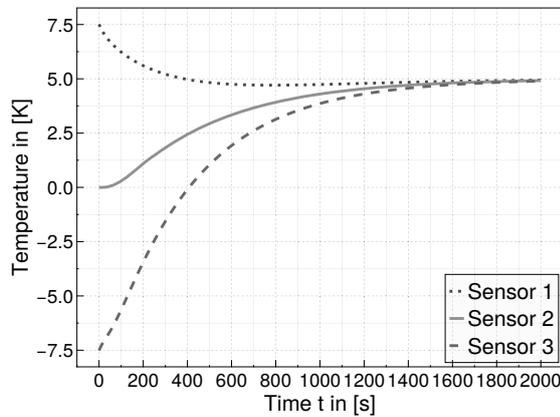


Figure 3: Measured temperatures y_n with $n \in \{1,2,3\}$ on boundary side B_N .

However, as we require access to all temperatures for the full state feedback, we have to compute them with a state observer. In a nutshell, we derived a state space control design for a 2-dim. heat equation and showed its performance. Further research will focus on the extension of our approach for systems with thermal emissions and 3-dim. geometries.

Actuators		Sensors		ΔT	Q	R
N_u	(m, M, v)	N_y	(m, M, v)			
3	(1,40,2)	3	(1,20,1)	2	$10^5 I$	I

Table 2: Controller Parameters for Simulation

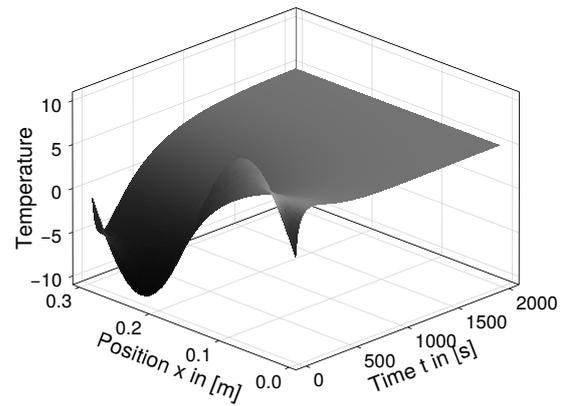


Figure 4: Temperature evolution on boundary side B_N .

References

- [1] Tay A, Tan KK, Zhao S, Lee TH. Predictive Ratio Control of Multizone Thermal Processing System in Lithography, *IFAC Proceedings Volumes*. 2008; 41(2). 10863–10868.
- [2] Meurer T. Control of higher-dimensional PDEs: Flatness and backstepping designs. Springer Science & Business Media, 2012.
- [3] Scholz S, Berger L. Hestia. jl: A Julia Library for Heat Conduction Modeling with Boundary Actuation. *Simul. Notes Eur.* 2023; 33(1): 27–30, doi: 10.11128/sne.33.sn.10634.
- [4] Scholz S, Berger L. Optimization-based Reference Tracking for Two-Dimensional Multiple Source Heating, *2024 IEEE 63rd Conference on Decision and Control*. 2024. (accepted, forthcoming).
- [5] Yueh WC, Eigenvalues of several tridiagonal matrices. *Applied Mathematics E-Notes [electronic only]*. 2005; 5: 66–74.
- [6] Åström KJ, Murray R. Feedback systems: an introduction for scientists and engineers. Princeton university press, 2021.
- [7] Bezanson J, Edelman A, Karpinski S, Shah VB. Julia: A fresh approach to numerical computing. *SIAM Review*. 2017; 59(1): 65-98.
- [8] Varga A, Karrasch D, Legat B, Aluthge D. andreavarga/MatrixEquations.jl: v2.4.2 (v2.4.2). Zenodo. 2024. Available: <https://doi.org/10.5281/zenodo.10546715>.
- [9] Scholz S. stephans3/StateSpace2DHeatConduction.jl: StateSpace2DHeatConduction.jl v0.1.0 (v0.1.0). Zenodo. 2024. Available: <https://doi.org/10.5281/zenodo.11350656>.

Ermittlung der Flottengröße für Systeme Autonomer Mobiler Roboter (AMR)

Ingolf Meinhardt^{1*}, Frank Schulze¹, Thorsten Schmidt¹

¹Institut für Technische Logistik, TU Dresden, Germany; *ingolf.meinhardt@tu-dresden.de

Abstract. Autonome Mobile Roboter sind in der Lage, sich selbstständig durch Lagerhallen, Produktionsstätten und andere Arbeitsumgebungen zu navigieren und Materialtransporte zu übernehmen. Für die Systemplanung stellt sich die Frage, wie man die richtige Flottengröße bestimmt. Analytische Ansätze werden oft als nicht ausreichend angesehen, weil AMR keiner vorgegebenen Route folgen müssen und sich selbst den kürzesten Weg suchen, Hindernisse erkennen und umfahren können. Zudem können Steuerungsaspekte und Lastszenarien nicht adäquat in einem Berechnungsmodell berücksichtigt werden. Es zeigt sich allerdings, dass analytische Verfahren für die Planung und Flottendimensionierung überraschend gut abschneiden.

Einführung

Autonome mobile Roboter (AMR) sind mehr als ein neuer Trend in der Intralogistik. Sie kommen bspw. immer häufiger als Alternative zu Fahrerlosen Transportfahrzeugen (FTF) zum Einsatz, übernehmen Transporte innerhalb eines Lagers oder versorgen Kommissionierplätze mit Material und entsorgen Abfall. Sie zeichnen sich durch ihre Flexibilität und Wendigkeit beim Einsatz aus, weil sie Hindernissen wie Personen oder entgegenkommenden Fahrzeugen ausweichen können.

Inzwischen werden AMR in vielen Unternehmen weltweit eingesetzt, mit steigender Tendenz. Dennoch gibt es kaum Publikationen darüber, wie die Flottengröße für ein AMR-System zu bestimmen ist. Die Empfehlung lautet zumeist: Simulation.

1 Literatur

Vielfach wird darauf verwiesen, dass (a) wegen der Autonomie der AMR und ihrer höheren Flexibilität nicht einmal Fahrwege und Fahrzeiten exakt bekannt sind und (b) verschiedene, auch nicht direkt quantifizierbare Para-

meter (wie beispielsweise separate Bahnhöfe/Haltebuchten oder variierende Transportanforderungen) Einfluss haben können und daher eine analytische Betrachtung nicht ausreichend sei. Berücksichtigt man jedoch, dass der Einsatz von AMR im industriellen Umfeld trotzdem überwiegend auf vorhandenen, meist orthogonal angeordneten Transportwegen erfolgt, reduziert sich Flexibilität auf das präzise Erkennen von anderen (sich bewegenden) Fahrzeugen oder Personen bzw. von (stationären) Hindernissen und dem Ausweichen oder Anpassen der Route. Und Autonomie bedeutet oft auch, dass Aufträge nicht zentral vergeben werden, sondern AMR selbstständig danach suchen. Für die Flottendimensionierung ist das unerheblich.

Damit lohnt sich ein Blick auf allgemeine Verfahren zur Berechnungen von Fahrzeuganzahlen. Grundsätzlich lassen sich zwei Vorgehensweisen unterscheiden:

Kimura [5], *Koo* [6], *Rall* [11] u.a. bilden das Fahrzeugsystem als Bediensystem ab mit vielen Bedienstationen und Warteschlangen. Ziel ist dabei, durch schrittweise Erhöhung der Fahrzeuganzahl die Warteschlangen an kritischen Punkten zu reduzieren und damit Blockierungen nach Möglichkeit zu vermeiden. Dazu sind aufwändige Berechnungen bis zur Erfüllung eines Konvergenzkriteriums wiederholt auszuführen.

Der andere, größere Anteil der Autoren orientiert dagegen auf eine Zeitbilanz: Die gesamte Fahrzeugzeit, die innerhalb einer Zeitperiode für eine gegebene Transportaufgabe benötigt wird, wird ins Verhältnis zur Länge dieser Zeitperiode gesetzt und ergibt die erforderliche Fahrzeuganzahl. Dabei wird von vier Schritten eines Arbeitsspiels ausgegangen: Lastaufnahme, Lastfahrt, Lastabgabe und Leerfahrt (*Großeschallau* [4], *Kuhn* [7], *Maxwell & Muckstadt* [10]).

In *Rajotia et al.* [12] werden 3 Phasen betrachtet, wobei die drei o.g. ersten Schritte zur Phase I zusammengefasst werden, weil die Zeiten dafür eindeutig aus Transportmatrix und Fahrzeitmatrix

ermittelt werden können. In Phase II verbleiben die Leerfahrten zum nächsten Beladeort, und Phase III steht für Zeiten für Warten und Blockieren (sowohl für Warten auf den nächsten Transportauftrag als auch Warten in einem Stau). Beide Phasen sind nicht eindeutig prognostizierbar und erfordern verschiedene Annahmen (z.B. Netztopologie, Steuerung). Bei AMR reduzieren sich diese Zeiten allerdings aufgrund des Ausweichverhaltens, es verbleiben minimale Zeitverluste beim Passieren von Kreuzungen bzw. vor Haltestellen zur Lastaufnahme bzw. -abgabe.

Großeschallau [4] berechnet einen minimalen Leerfahrtaufwand durch zwei Annahmen: An einer Station ankommende Fahrzeuge werden soweit als möglich für Transporte verwendet, die an dieser Station beginnen (lokaler Ausgleich). Das vermeidet unnötige Leerfahrten. Der verbleibende Fahrzeugüberschuss bzw. Fahrzeugbedarf ist zwischen den Stationen auszugleichen (Leerfahrtausgleich). Dabei werden Leerfahrten angenommen, die von Stationen mit Fahrzeugüberschuss zu den *nächstgelegenen* Stationen mit Fahrzeugbedarf führen. Das sorgt für möglichst kurze Leerfahrten. *Maxwell & Muckstadt* [10] gehen ähnlich vor, wobei hier die Nettoflüsse pro Station (aus gelieferten und abgeholtene Ladeeinheiten) genutzt werden, um ein klassisches Transportproblem zu formulieren. Die Lösung sind die benötigten Leerfahrten.

Beide Verfahren liefern eine untere Grenze für den Zeitbedarf bzw. die minimal notwendige Fahrzeuganzahl. *Rajotia et al.* [12] kritisieren dies, weil ein „beladenes Fahrzeug nach der Lieferung und Abgabe seiner Ladung möglicherweise keine andere Ladung vorfindet, die an derselben Station abgeholt werden kann.“ Aus diesem Grunde verzichten sie in ihrem Modell auf den lokalen Ausgleich und kommen so auf deutlich mehr Leerfahrten. Auch *Kuhn* [7] verzichtet auf den lokalen Ausgleich und verteilt die Leerfahrten von jeder Station proportional auf die Fahrzeugbedarfe.

Malmborg [8] bestimmt eine obere Grenze für die Leerfahrzeit, indem der Leerfahrzeugüberschuss auf die am weitesten entfernte Station verteilt wird. *Egbelu & Tanchoco* [3] und *Malmborg* [9] erkennen aber schließlich auch an, dass ein solches Vorgehen nicht ausreicht, weil Steuerungsregeln einen nicht zu vernachlässigenden Einfluss auf die Leistung des Systems haben.

Beisteiner [1] stellt zwei Ansätze vor: **(a)** Für jede Fahrt mit Last wird eine gleich lange Leerfahrt angenommen. **(b)** Es erfolgt ein lokaler Ausgleich je

Station und die verbleibenden Leerfahrten werden über die mittlere Distanz der Lastfahrten zurückgelegt. Beide Ansätze sind sehr grobe Schätzungen und reagieren empfindlich auf ungünstige Netztopologien.

2 Vergleich der Berechnungsansätze

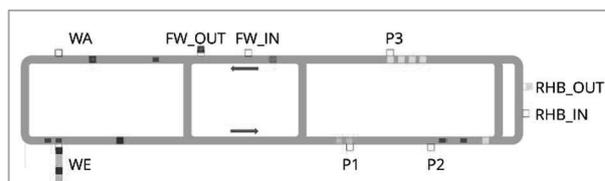


Abbildung 1: Layout des FTS (Beispiel)

Die im vorigen Abschnitt beschriebenen Berechnungsansätze wurden auf ein Beispiel (Abbildung 1) angewandt. Die Abstände zwischen den Stationen liegen in einer Entfernungsmatrix *E* vor, aus der unter Nutzung der Fahrzeugparameter eine Fahrzeitmatrix *F* abgeleitet wird. Die geforderten Durchsätze definiert die Transportmatrix *T*. Die Ergebnisse der Berechnungen fasst Tabelle 1 zusammen. Es wird deutlich, dass die unterschiedlichen Verfahren zu großen Abweichungen bei den Leerfahrten führen und dementsprechend auch zu abweichenden erforderlichen Fahrzeuganzahlen.

Tabelle 1: Vergleich der Berechnungsergebnisse

Autor	Fahrtweg [m]	Fahrzeit [s]	Gesamter Zeitbedarf [s]	Fahrzeuganzahl
Lastfahrten				
	14.793	16.100		
Leerfahrten				
Großeschallau	6.845	7.902	35.762	12,4
Rajotia	10.695	11.901	39.761	13,8
Beisteiner (a)	14.793	16.100	43.960	15,3
Kuhn	23.509	24.750	52.610	18,3
Malmborg	19.061	20.368	48.228	16,7

Zu ähnlichen Ergebnissen kommen auch *Dukic et al.* [2], wobei bei ihnen vor allem die Genauigkeit der Abschätzung von Leerfahrzeiten bei unterschiedlichen Betriebsszenarien gegenüber der Simulation im Fokus stand und nicht die dafür erforderliche Flottengröße.

3 Das Simulationsmodell

Zur Bewertung der analytischen Verfahren wurde ein Simulationsmodell erstellt, insb. um den Einfluss der Fahrzeuganzahl systematisch zu analysieren. Als Startwert wird die lt. Tabelle 1 bestimmte Minimalanzahl von 13 Fahrzeugen gewählt. Außerdem werden verschiedene Steuerungsstrategien genutzt. Da zur Demonstration bewusst ein einfaches Layout gewählt wurde, hat dies hier aber nur geringen Einfluss.

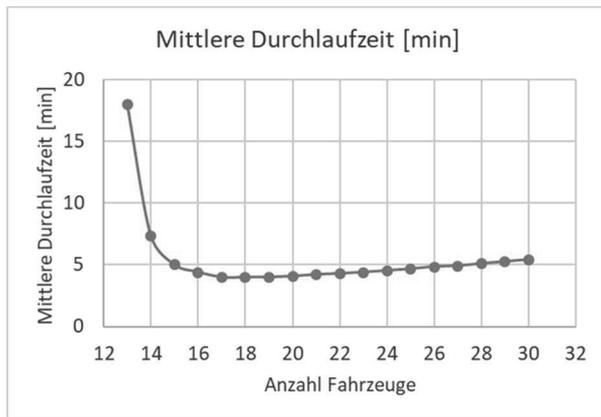


Abbildung 2: Mittler Durchlaufzeit in Abhängigkeit von der Fahrzeuganzahl

Abbildung 2 zeigt bei der minimalen Fahrzeuganzahl noch eine sehr große Durchlaufzeit. Obwohl der geforderte Durchsatz erreicht wird, bilden sich an einzelnen Stationen lange Warteschlangen unerledigter Transportaufträge. Durch Hinzunahme weiterer Fahrzeuge verkürzen sie sich, die Durchlaufzeit sinkt. Bei 17 Fahrzeugen wird das Minimum erreicht, um danach wieder leicht zu wachsen. Im Untersuchungsbereich lässt sich eine Zunahme der Behinderungen beobachten, wodurch die Durchlaufzeit wieder anwächst, was aber noch zu keinem Durchsatzverlust führt. Das entspricht dem in *Valmiki et al.* [13] beschriebenen Verhalten, wonach eine größere Fahrzeugflotte den Durchsatz bis zur Sättigungsgrenze erhöht, danach aber zur Verringerung des Durchsatzes führt. Umgekehrt bedeutet das: im Bereich der Sättigungsgrenze kann die Anzahl an Fahrzeugen ohne Leistungsverlust verringert werden.

Eine detaillierte Analyse zeigt aber noch mehr: Mit wachsender Fahrzeuganzahl nimmt die mittlere Länge der Leerfahrwege zu und entsprechend auch die Leerfahrzeit (Abbildung 3). Eine geringe Fahrzeuganzahl verknappt die Transportressourcen und sorgt

dafür, dass an mehr Stationen Ladeeinheiten auf Abholung warten. Dementsprechend führen weit häufiger kurze Leerfahrten zu benachbarten Stationen.

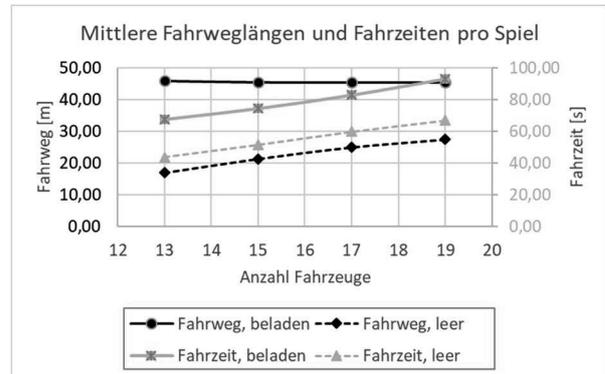


Abbildung 3: Fahrweglängen und Fahrzeiten abhängig von der Fahrzeuganzahl

Gleichzeitig steigt aber auch die Fahrzeit mit Last, obwohl hier die mittleren Weglängen gleich bleiben. Ursache ist die Zunahme an Behinderungen vor den Stationen. Das betrifft im Beispiellayout sowohl Fahrzeuge, die an der Station zur Lastaufnahme bzw. -abgabe halten, als auch Fahrzeuge mit einem anderen Ziel, welche diese Station nur passieren. Letzteres ließe sich durch Haltebuchten oder breitere Transportwege mit Überholmöglichkeit vermeiden.

Für das Beispiel wird eine extrem einfache Steuerungsstrategie gewählt: Fahrzeuge verbleiben nach Abgabe ihrer Ladung an Ort und Stelle, bis es einen neuen Transportauftrag gibt (Variante A). Zwei weitere Varianten sind implementiert, entfalten aber erst bei stärker vernetzten Topologien ihre Wirkung: Variante B schickt, falls es an der aktuellen Station keine Arbeit gibt, das Fahrzeug zur nächstgelegenen Station oder Kreuzung, so dass sich alle Fahrzeuge permanent bewegen. Variante C definiert einen zentralen „Parkplatz“ für nicht benötigte Fahrzeuge mit zusätzlichen Fahrzeiten von und zum Parkplatz.

Die nächsten Schritte bestehen darin, diese Untersuchungen auf andere Netztopologien (uni- vs. bidirektionale Pfade, unterschiedliche Vernetzungsgrade) und verschiedene Lastprofile (z.B. Schiefast) auszudehnen, um den Einfluss relevanter Faktoren zu quantifizieren.

4 Ergebnisse

AMR zeichnen sich durch ihre hohe Flexibilität im Einsatz, die Dynamik bei der Gestaltung des Fahrwegs

und Geschwindigkeitsvorteile gegenüber FTF aus. Aus Sicht der Systemplanung ändert das jedoch nichts daran, dass AMR sich zwar nicht pfadgebunden, aber dennoch genau wie FTF auf einem definierten Transportwegenetz bewegen und Zeiten zum Aufnehmen und Abgeben der Last sowie zum Fahren mit und ohne Last benötigen. In dieser Hinsicht unterscheiden sich AMR nicht von FTF.

Daher lassen sich analytische Verfahren, die ganz allgemein für Fahrzeugsysteme entstanden sind, auch auf AMR übertragen. Schwankungen der Fahrweglängen (infolge von Hindernisumfahrungen) führen zu deutlich weniger Abweichungen als unterschiedliche Verfahren zur Bestimmung von Leerfahrten – allein Leerfahrten in der Planungsphase überhaupt zu berücksichtigen, ist schon Fortschritt genug – oder zusätzliche Wartezeiten bei temporärer Häufung der Fahrzeuganzahl vor Haltestellen.

Forschungsbedarf besteht weiterhin hinsichtlich der genauen Abschätzung der Wirkung des Zusammenspiels von Last (zeitliche und örtliche Charakteristik der Transportaufträge), Layout (Topologie des Pfadnetzwerks) und Logik (Strategien zur Fahrzeugdisposition) auf die erforderliche Flottengröße.

References

- [1] Beisteiner, F.: Strategies for the employment of vehicles in an automated guided transportation system. *Proceedings of 2nd International Conference on Automated Guided Vehicle Systems, Stuttgart, Germany*. 1983.
- [2] Dukic, G., Opetuk T., Cajner, H., Jakovljević, M.: Simulation analysis of models for estimation of empty travel time of vehicles in non-automated material handling systems. *FME Transactions*. 2018, 46. 336-341. DOI: 10.5937/fmet1803336D.
- [3] Egbelu, P.J., Tanchoco, J.M.A. (1984): Characterization of automatic guided vehicle dispatching rules. In: *International Journal of Production Research.*, 22 (3) (1984), pp. 359-374, <https://doi.org/10.1080/00207548408942459>
- [4] Großeschallau, W. (1984). Materialflußrechnung: Modelle und Verfahren zur Analyse und Berechnung von Materialflußsystemen. Dissertation, Dortmund
- [5] Kimura, T.: Approximating the Mean Waiting Time in the GI/G/s Queue. *Journal of the Operations Reserch Society of Japan*. 1991, Vol. 34, No. 2, pp. 173–186.
- [6] Koo, P-H., Jang, J., & Suh, J.: Estimation of Part Waiting Time and Fleet Sizing in AGV Systems. In: *The International Journal of Flexible Manufacturing Systems*. 2005, 16, 211–228.
- [7] Kuhn, A.: Efficient planning of AGVS by analytical methods. *Proceedings of the 2nd International Conference on Automated Guided Vehicle Systems*. 1983, pp. 1-10.
- [8] Malmberg, C. J.: A model for the design of zone control automated guided vehicle systems. *International Journal of Production Research*. 1990, Vol 28 (10), 1741 - 1758.
- [9] Malmberg, C.J., & Shen, Y.: Heuristic dispatching models for multi-vehicle materials handling systems. *Applied Mathematical Modelling*. 1984, 18, 124-133.
- [10] Maxwell, W.L., & Muckstadt, J.A.: Design of Automatic Guided Vehicle Systems. In *IIE Transactions*, 1982, 14, 114-124. <https://doi.org/10.1080/05695558208975046>
- [11] Rall, B.: Analyse und Dimensionierung von Materialflusssystemen mittels geschlossener Warteschlangennetze. *Wissenschaftliche Berichte des Institutes für Fördertechnik und Logistiksysteme der Universität Karlsruhe*. 1998, Vol. 46, Dissertation.
- [12] Rajotia, S., Shanker, K., & Batra J.L.: Determination of optimal AGV fleet size for an FMS. *International Journal of Production Research*. 1998, Vol 36:5, pp 1177-1198, <https://doi.org/10.1080/002075498193273>
- [13] Valmiki, P., Reddy, A.V., Panchakarla, G., Kumar, K.R., Purohit, R., & Suhane, A.: A Study on Simulation Methods for AGV Fleet Size Estimation in a Flexible Manufacturing System. *Materials Today: Proceedings*. 2018, 5, 3994-3999.

Simulation-Based Analysis of Storage Strategies for an Inland Container Terminal

Maximilian Mowe¹, Jona Manemann¹, Lasse Jurgeleit¹, Maximilian Kiefer¹, Uwe Clausen¹

¹Institute of Transport Logistics, TU Dortmund University, Leonhard-Euler-Straße 2, 44227 Dortmund, Germany

Abstract. Due to the transformation of the logistics sector, the use of trimodal container terminals is becoming increasingly important. This results in major challenges, such as managing the constantly increasing volume of shipments while improving internal processes efficiency. Adapting existing storage strategies has proven to offer great potential for raising productivity. In this context, the following paper develops an alternative storage strategy for container terminals. A simulation model for an exemplary trimodal container terminal is developed to evaluate its functionality. In addition, the potential is analyzed based on a comparison with the currently used storing strategy.

Introduction

The logistics sector is important to the economies worldwide, e.g., the third-largest economic sector in Germany, with a total volume of € 293 billion in 2021. Logistics service providers, such as transportation and handling, account for 50 % of the volume [1]. Despite its economic importance and impressive figures, the logistics sector faces major challenges. Logistics has to fulfill customer requirements in an increasingly volatile environment and has failed to achieve sustainability targets so far. One lever in this respect is a modal shift from road to combined rail-road transport. A forecasted growth of rail by 73 % by 2030 compared to 2010 in Germany is particularly attributable to rail/road transshipment as a key growth driver [2]. With freight transport accounting for 19.4 % of Germany's total greenhouse gas emissions in 2021 [3], modal shift and other measures to reduce emissions are vital, and operators of trimodal container terminals connecting road, rail, and water modes of transportation in container freight transport need solutions to cope with growth efficiently.

One approach to overcoming these challenges for a trimodal container terminal is to improve internal processes. These range from improving the storage strategy to developing additional space and optimizing the use of

resources. However, new strategies must be tested before they are applied in reality. One method that makes this possible is simulation. Simulation allows to create a virtual image of reality to test changes based on this image, which can then be transferred to reality [4].

In order to support those who are facing these challenges, this paper aims to demonstrate the possibilities for optimizing operational strategies using simulation. Therefore, an exemplary container terminal is presented, which aims to compare two storage strategies by using simulation. The storage strategies to be compared differentiate as follows: the current situation with storage excluding the removal path, and the target situation with storage close to the delivery points.

For this purpose, our paper is divided into five sections. After a brief overview on current research on simulation for operational strategies of container terminals, the used simulation procedure is presented. Followed by developing the simulation model and conducting related experiments, an outlook on future work in this field of research is given.

1 Literature Review

A systematic literature review based on the guidelines of Durach et al. [5] is conducted to evaluate the common use of the method for planning operational strategies of container terminals. The following search terms and combinations of these are used to identify relevant sources: “simulation”, “optimization”, “container terminal”, “stack”, “strategies”, and “terminal operation”. Publications from the databases Scopus, Web of Science, Google Scholar, and Science Direct are examined for this purpose.

After full-text analysis, 18 results emerged. The evaluation of the results shows that the methodological approach of simulation is used for optimizing operational strategies. However, no publication was found that deals specifically with the storage strategies. In a related study,

Clausen et al. [6] constructed a simulation model to identify the optimal operational configuration for a terminal. The model determines the most effective combination of crane control and resource management strategies for each load the terminal can accommodate based on a specific set of operational conditions. Novaes et al. [7] examine in their study the potential of data science to enhance container terminal operations, improve efficiency, boost throughput, and strengthen competitiveness in the shipping sector. Decision-making within container terminals, particularly in determining optimal container stacking locations, is a significant challenge due to the multitude of factors at play. By analyzing the datasets, new strategies and policies can be simulated to minimize container rehandling operations.

Overall, the papers identified indicate that simulation has applications in a variety of use cases in the area of planning the operating strategies of container terminals. For this reason, the simulation method is described below before the model developed for this use case is presented.

2 Method

“Simulation is the representation of a system with its dynamic processes in an experimentable model with the aim of reaching findings which are transferable to reality” [8]. To examine a system with simulation in a structured manner, multiple procedure models can be used (e.g. [4], [9], [10]). The procedure models provide users with information for appropriately carrying out simulation studies. In German-speaking countries, VDI Guideline 3633 Sheet 1 is widely used and selected for this study because the system under investigation is located in Germany. The associated Verification and Validation (V&V) during the study follows Rabe et al. [11]. Their procedure model for V&V aligns with the phases of VDI 3633, which makes its application suitable for this study.

The phases defined by VDI 3633 guide users from task definition to the experiments and analysis. Within the procedure, there is a clear separation between data and model, and each phase produces a result [4]. The intermediate results are the subject of the V&V. According to Rabe et al. [11], each phase is examined in itself and additionally concerning or against previous phase results.

3 Simulation Modeling

According to the introduction, the objective is to analyze a given storage strategy for container handling and compare it with a zoning strategy. The investigations are conducted for two inland container terminals. CT1 is tri-modal (see Figure 1) and CT2 is bimodal. CT2 contains a road and a rail lane, which are located next to each other. Prior to the actual strategy comparison, the preliminary goal is to model the terminals in their respective system state with sufficient accuracy. The subsequent comparison is aimed at the container movements and, therefore, focuses on the internal handling processes. Accordingly, the system under consideration ends with the entry and exit of transport vehicles.

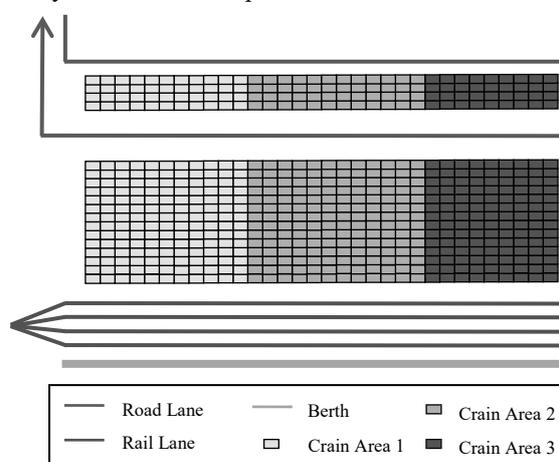


Figure 1: Schematic representation of container terminal 1

For this study, CT’s activities are roughly divided into three handling processes. Incoming containers are stored and retrieved to leave the terminal. Additionally, a relocation process may be necessary for stacked containers. Containers enter the system by a mode of transportation (train, truck, or ship). Within the storing process, a given strategy first determines the position. After that, a handling resource (e.g., portal crane) moves the container to its destination. The retrieval begins with the mode of transportation arriving at the terminal. The time of arrival may be known with a containers arrival or is defined while the container is already stored. Before moving the container, it is checked if the container is movable (i.e., the container has no containers stacked above itself). Movable containers get loaded onto the mode of transportation and leave the system. For non-movable containers, the relocation process for the container above is ini-

tiated before moving the desired container. At the beginning of the relocation process, it is also checked whether the container is movable. For movable containers, the position is determined by the given storage strategy. For non-movable containers, the relocation process is initiated for the container above, and the desired container is moved afterward. Accordingly, the relocation process is recursive. For this study, we only consider relocations because of retrieval processes.

The handling resources, containers, and modes of transport are being modeled as agents, which define their behavior within the terminal. The model, including the described processes and agents, is implemented with the simulation software AnyLogic 8.9 as a discrete event simulation using agents. To ensure the correct implementation and validation of the logistical process flows, the model was extensively examined. For example, individual transport processes were examined more closely using animations, trace analyses of process times, and a structured walk-through.

4 Experiments & Results

Within the experimental framework, the current terminal is analyzed in the first simulation scenario using five different degrees of capacity utilization with varying numbers of in- and outbound transport vehicles and containers. The second scenario implements the developed storing strategy with three different weighting factors for retrieval distance. In order to achieve valid results, the confidence interval method is used to determine the number of replications to be carried out per parameter variation. The study showed that 30 replications are sufficient to ensure stochastic reliability results. The duration of each simulation run lasts seven days with a two-day transient phase regarding to the method of Welch [12].

The developed storage strategy follows the existing requirements for container storage. Currently, the optimal storing location is identified based on the distance from its current position and under consideration of time restrictions such as the retrieval date. Additionally, the new storage strategy also considers the distances for retrieving containers. For this purpose, the storage area is divided into various sections running along the road and rail lanes. According to this, the optimal storing location can be determined in order to ensure the shortest possible retrieval distance to the lanes. Therefore, our calculation

is characterized by three elements. “A” indicates the distance of the containers current location to the available storage position. “B” represents the direct retrieval distance to the lanes for the trucks and trains. In addition, the weighting of the retrieval distance is managed by factor “p”.

$$\begin{aligned} &\text{Distance to optimal storing position} \\ &= \min \{A + p * B\} \end{aligned} \tag{1}$$

For validation of this equation, its functionality is verified in the simulation model. For this purpose, the straight distance between retrieval containers and the lane for the trucks and trains is analyzed, as this will significantly influence the overall efficiency of the internal handling processes. Since ships are mostly used to deliver the containers, this mode of transport is not considered in the evaluation. Figure 2 shows the results of the conducted simulation study.

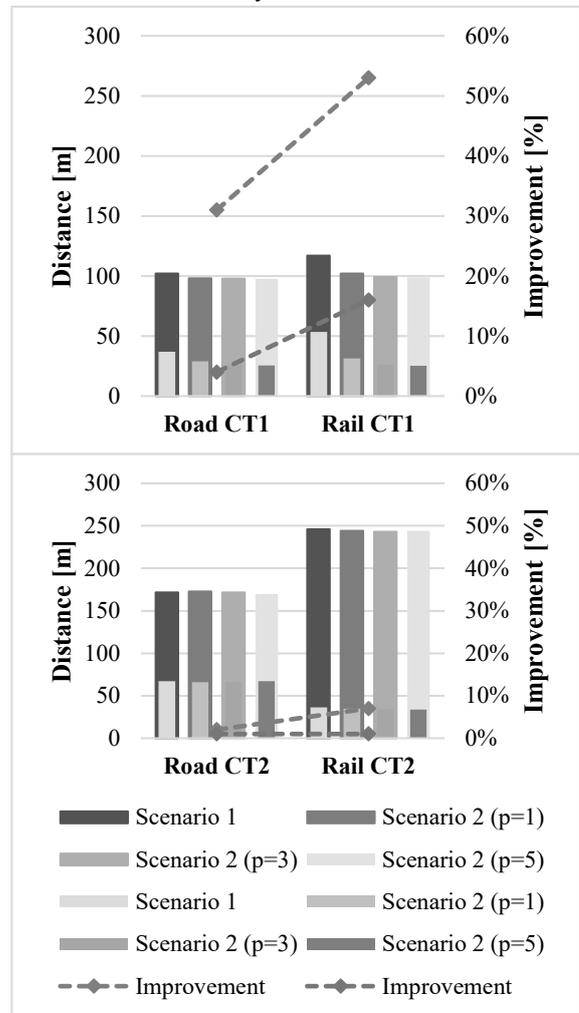


Figure 2: Comparison of covered distances for retrieval of containers

From the grey-colored bars in Figure 2, it can be observed that the covered distances of portal cranes for retrieval of containers at CT1 decrease for all weighting factors when applying the storing strategy compared to initial scenario 1. Depending on the mode of transportation, this improvement amounts to 31 % respectively 53 %, as shown by the grey dashed line. The improvement for CT2 is less significant. In general, containers are stored closer to the retrieval area, thus indicating that optimization of the storing position by the equation works. For a complete evaluation of its potential, the real distance from containers to the transport vehicles must be analyzed. This is shown by the green bars in Figure 2, which indicate that the real distance is significantly greater. Furthermore, the green dashed line demonstrates that the improvement achieved compared to scenario 1 is not as great as previously suggested. At CT2, this positive effect is almost neutralized. Furthermore, the total duration of dispatching processes for trucks and trains at both terminals will be analyzed to validate the potential of the new storing strategy. However, at either location there has been no significant reduction for both modes of transport. For some scenarios, dispatching times have even increased.

Overall, the results of the simulation scenarios indicate that the new storing strategy does not significantly improve the internal processes of the container terminal. Due to a high utilization of the lane, trucks can only stop in an assigned area but not at the level of containers storage location. Although containers will be placed closer to the lane, it will not affect its distance to the real loading position of the truck. This difficulty also occurs when loading the trains, as each container is linked to a fixed railway wagon, which may not be positioned at the same level as containers storing location.

5 Conclusion

This paper compares two different storage strategies while focussing on increased productivity in a tri- and bi-modal container terminal. To evaluate their potential in practice, simulation has been used. Based on VDI 3633, a simulation model was developed that combines agent-based and discrete-event simulation. Evaluation based on industry-specific key performance indicators revealed that our newly developed storing strategy works effectively and can improve internal processes. However, the

expected improvement cannot be achieved due to the specific infrastructural and operational conditions of both terminals considered in this example. In order to fully realize its potential, further work must be conducted to apply the storing strategy in combination with the scheduling of incoming modes of transportation. In addition, this strategy must be modified depending on the use case and its infrastructural conditions.

References

- [1] BMDV: Masterplan Binnenschifffahrt. 2018. Federal Ministry for Digital and Transport. <https://www.dslv.org/de/themen/masterplan-binnenschifffahrt>
- [2] Boysen J. et al.: Status quo des Güterverkehrssystems in Deutschland – eine Metastudie unter besonderer Betrachtung der Vernetzung des Verkehrs.
- [3] The Umweltbundesamt: Emissionen des Verkehrs. 2023.
- [4] VDI Association of German Engineers e.V.: Simulation of systems in materials handling, logistics and production: Part 1(VDI 3633). Berlin: Beuth, 2014.
- [5] Durach C. F. et al.: A New Paradigm for Systematic Literature Reviews in Supply Chain Management. *J Supply Chain Manag*, 53 (4) 2017, p. 67–85. doi:10.1111/jscm.12145.
- [6] Clausen U. et al.: CONTSIM–Container Terminal Management with Simulation. *Procedia - Social and Behavioral Sciences*, 54 2012, p. 332–340. doi:10.1016/j.sbspro.2012.09.752.
- [7] Novaes Mathias T. et al.: Operational performance evaluation of a container terminal using data mining and simulation. *Asian Transport Studies*, 10 2024, p. 100127. doi:10.1016/j.eastsj.2024.100127.
- [8] VDI Association of German Engineers e.V.: Simulation of systems in materials handling, logistics and production: Terms and definitions(VDI 3633). Berlin: Beuth Verlag, 2018.
- [9] Law A. M.: Simulation modeling and analysis. New York, NY: McGraw Hill Education, 2015.
- [10] Banks J. (ed): Handbook of simulation: Principles, methodology, advances, applications, and practice. New York: Wiley, 1998.
- [11] Rabe M. et al.: A new procedure model for verification and validation in production and logistics simulation. In: Winter Simulation Conference, 2008 (Editor: S. J. Mason). Piscataway, NJ: IEEE, 2008, pp 1717–1726. 10.1109/WSC.2008.4736258
- [12] Welch P. D.: The statistical analysis of simulation results. *The Computer Performance Modeling* 1983, 1983:268–328.

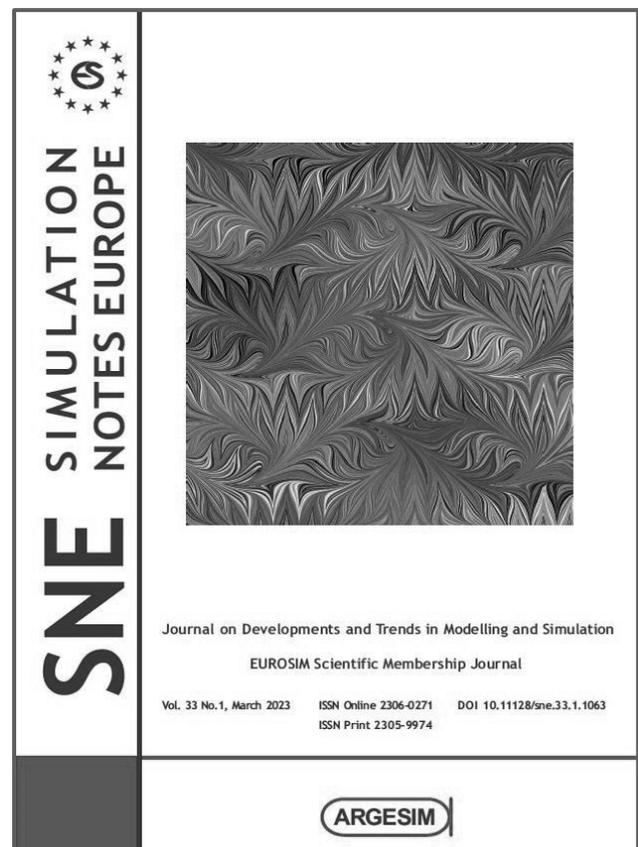
SNE SIMULATION NOTES EUROPE

Simulation Notes Europe (**SNE**) provides an international, high-quality forum for presentation of new ideas and approaches in simulation - from modelling to experiment analysis, from implementation to verification, from validation to identification, from numerics to visualisation - in context of the simulation process.

SNE seeks to serve scientists, researchers, developers and users of the simulation process across a variety of theoretical and applied fields in pursuit of novel ideas in simulation and to enable the exchange of experience and knowledge through descriptions of specific applications. **SNE** puts special emphasis on the overall view in simulation, and on comparative investigations, as benchmarks and comparisons in methodology and application. Additionally, **SNE** welcomes also contributions in education in / for / with simulation.

SNE is the official membership journal of **EUROSIM**, the Federation of European simulation societies and simulation groups, and the scientific membership journal of **ASIM**, the German simulation society. **SNE** is open for post-conference publications and for special issues organized by **EUROSIM** societies, e.g. **ASIM** thematic special issues or **ASIM** post-conference special issues.

SNE is primarily an electronic journal and follows an open access strategy, with free download in basic layout. Members of **EUROSIM** societies, as **ASIM**, **SIMS**, e.g. are entitled to download **SNE** in an elaborate and extended layout. Print **SNE** is available for specific groups of **EUROSIM** societies.



ISBN ebook 978-3-903347-64-9
ARGESIM Report 46
www.argesim.org

DOI 10.11128/arep.46
ASIM Mitteilung 189
www.asim-gi.org