Multi-stage production planning with special consideration of energy supply and demand

Vom Fachbereich Rechts- und Wirtschaftswissenschaften
der Technischen Universität Darmstadt

zur Erlangung des akademischen Grades
Doctor rerum politicarum
(Dr. rer. pol.)

genehmigte Dissertation
von Dipl.-Wi.-Ing. Konstantin Biel, MBA
geboren in Stuttgart

Referent: Prof. Dr. Christoph Glock
Korreferentin: Prof. Dr. Anne Lange

Tag der Einreichung: 04.08.2017
Tag der mündlichen Prüfung: 23.11.2017

Darmstadt, 2017
D 17

Acknowledgments

This dissertation marks the end of an exciting journey, which I embarked on four years ago. Various people accompanied me on this journey and helped me stay on track, which I am very grateful for. I am particularly thankful to my advisor, Prof. Dr. Christoph Glock, who gave me the opportunity to pursue my PhD at his Institute of Production and Supply Chain Management of Technische Universität Darmstadt. Christoph always provided me with a lot of freedom – be it with regard to the choice of the topic of my dissertation, be it with regard to my wish to go abroad on a research stay, or be it simply with regard to the way of handling day-to-day work. Along with this freedom, he offered steady advice on how to approach research-related questions, but also on how to make decisions in general in all kinds of situations. I am very sure that I will continue to greatly benefit from the highly structured and logical way of thinking that I learned from Christoph far beyond my time as a PhD student. I would also like to thank Prof. Dr. Anne Lange for her practical advice over the course of the preparation of this dissertation and for serving as my co-advisor.

Furthermore, I would like to express my gratitude to Prof. Dr. John W. Sutherland and Prof. Dr. Fu Zhao, who gave me the opportunity to spend four exciting months at the Division of Environmental and Ecological Engineering of Purdue University. While my research prior to this stay abroad had mostly adopted a management-driven perspective, Prof. Sutherland and Prof. Zhao urged me to approach energy-aware production planning from a rather engineering-based angle, which substantially deepened my understanding of this interdisciplinary research field. I would also like to especially thank Prof. Sutherland's and Prof. Zhao's research group for integrating me into their research and leisure activities from day one, which eventually gave me the chance to experience a real Chinese New Year celebration in the Midwest of the United States. In the context of my research stay at Purdue University, I would further like to acknowledge the generous financial and organizational support of the German Academic Exchange Service, which provided me with a doctoral scholarship.

In addition, I would like to thank my colleagues in Darmstadt, who I enjoyed working with very much. I am very thankful to Dr. Eric Grosse and Maximilian Schneider for their extensive support of my application for the scholarship of the German Academic Exchange Service, which enabled me to visit Purdue University in the first place. In general, Eric helped me a lot to never get lost in the jungle of forms that I needed to fill in either for receiving a travel grant to attend a conference or for finally submitting this dissertation. I am also very grateful to Fabian Beck and Sabrina Göttmann, who offered constant support when it came to formulating equations in the mathematically most proper and appealing way. Moreover, discussing last weekend's soccer results with Fabian often was a welcome distraction from everyday work. Sharing an office with Sabrina in my final year in Darmstadt was a lot of fun and led to in-depth conversations about the symmetry of the figures in this dissertation, goals in life, and ombré hairstyles. Overall, Sabrina's positive vibes and encouragement helped me a lot to drag myself across the finish line. In addition, I would like to thank Martin Wirth and Hamid Abedinnia, who I could always refer to with regard to operations research-related issues. The conversations with Hamid frequently turned into profound discussions about life in Iran and the highly complex structure of the Iranian political system, which considerably expanded my horizon. The same applies to the exchange of ideas regarding the future of energy-aware production planning and its most promising industrial applications with Timm Weitzel, which I enjoyed a lot. I am also very grateful for Timms advice on how to illustrate complex mathematical algorithms and results in a clear and concise way.

Furthermore, I would like to thank David Henninger and Valentin Ruser for their tireless help in recording wind speed forecasts for one of my research projects.

Among my colleagues, I would like to particularly express my gratitude to Jörg Ries. The two and a half years we shared an office gave rise to countless discussions on research ideas, startup concepts, political developments, and whiskey as an investment opportunity. I am especially thankful for Jörg's valuable advice and feedback on all of my research efforts, which was sometimes devastating, sometimes encouraging, but always blunt and honest. In a similar way, Yannick Zapf expressed his thoughts on my research activities, which I am very grateful for as well. Yannick made sure that I sufficiently took account of the practitioner's perspective at all times and that my model assumptions satisfied the requirements of real-life industrial applications.

Like most PhD students, I experienced both relaxing and stressful times. My family can most certainly tell a thing or two about the latter times in particular. I am very grateful to my parents, who always paved the way for me that ultimately led up to this dissertation. I would also like to thank my brother, Alexander, for his hands-on advice on how to make goal-oriented decisions and on how to maneuver out of one or another dilemma. Above all, I would like to express my gratitude to my beloved Galina for her endless support and particularly for bearing with my impatience at all times.

Without all this guidance and encouragement, I am sure that I would have lost my way even more often than I eventually did. However, I believe that losing one's way is an incremental part of pursuing a PhD as academic research is supposed to be about exploring unknown fields, about identifying dead ends, and about refuting long-established ways of thinking and models that were believed to capture the reality for far too long. With this dissertation, I hope to contribute to advancing the area of production planning, which has finally started to realize its vast potential to enhance energy efficiency and reduce greenhouse gas emissions in manufacturing processes. At the same time, this dissertation is meant to take account of human contributions in manufacturing processes and to show how gearing production planning towards diverging human characteristics can substantially enhance system performance.

Darmstadt, November 2017 Konstantin Biel

Zusammenfassung

Die vorliegende kumulative Dissertation umfasst fünf Artikel, die in verschiedenen wissenschaftlichen Zeitschriften veröffentlicht wurden. Alle fünf Artikel befassen sich mit der Produktionsplanung in mehrstufigen Produktionssystemen. Aufgrund unterschiedlicher Schwerpunkte der Artikel ist die vorliegende Dissertation in zwei Teile unterteilt. Teil A umfasst die Artikel 1 bis 4 und beschäftigt sich mit der Berücksichtigung verschiedener energetischer Aspekte in der mehrstufigen Produktionsplanung. Teil B besteht aus Artikel 5 und untersucht den Einfluss von Lern- und Vergessenseffekten auf den Produktionsfluss in mehrstufigen Produktionssystemen. Abgesehen von den unterschiedlichen inhaltlichen Ausrichtungen unterscheiden sich die Artikel auch im Hinblick auf die verwendeten Methoden. In Artikel 1 wird ein systematischer Überblick über die Literatur zur energieeffizienten Produktionsplanung erstellt. Die verbleibenden vier Artikel entwickeln dagegen mathematische Modelle zur Entscheidungsunterstützung unter Berücksichtigung verschiedener Energieeffizienzmaßnahmen (Artikel 2 bis 4) sowie unter Berücksichtigung menschlicher Faktoren (Artikel 5). Die Artikel 2 bis 4 lösen die aufgestellten mathematischen Modelle analytisch. Artikel 5 setzt dagegen diskrete, ereignisorientierte Simulation ein, um effektive Produktionssteuerungspolitiken herzuleiten. Die folgenden Abschnitte fassen die fünf Artikel kurz zusammen.

Artikel 1 durchsucht die Literatur systematisch nach wissenschaftlichen Artikeln, die quantitative Modellierungsansätze zur Berücksichtigung energetischer Aspekte in der Produktionsplanung entwickeln. Die gefundenen Modelle werden anschließend nach kurz- und mittelfristigen Planungsmodellen klassifiziert und auf Basis der mathematischen Modellierung der berücksichtigten energetischen Aspekte verglichen. Aufbauend auf dieser Synthese werden zukünftige Forschungsrichtungen aufgezeigt, welche die Artikel 2 bis 4 aufgreifen.

Artikel 2 untersucht, wie die Energieeffizienz eines zweistufigen Produktionssystems durch die Einbindung der Rückgewinnung von Abwärme aus den Produktionsprozessen in die Produktionsplanung gesteigert werden kann. Die Abwärme wird mithilfe eines Organic Rankine Cycles verstromt und zur Unterstützung der Deckung des Energiebedarfs der Produktionsstufen verwendet. Zur Entwicklung eines ganzheitlichen Produktionsplanungsmodells wird zunächst der thermodynamische Prozess der Umwandlung der Abwärme in elektrische Energie mathematisch modelliert. Diese mathematische Modellierung wird im nächsten Schritt in ein zweistufiges Losgrößenmodell integriert. Anschließend wird ein Verfahren zur Lösung dieses integrierten Modells entwickelt, welches diejenigen Werte für die Losgröße, die Transportlosanzahl sowie die Produktionsraten der beiden Produktionsstufen berechnet, welche die Summe aus Produktions- und Energiekosten minimiert. In einer numerischen Studie wird untersucht, unter welchen Rahmenbedingungen das integrierte Produktionsplanungsmodell die Umsetzung von Energieeffizienzzielen fördert und inwiefern die Integration der Abwärmerückgewinnung die Produktionsplanungsentscheidungen beeinflusst.

Artikel 3 erweitert das Modell aus Artikel 2, indem zusätzlich zum Abwärmerückgewinnungssystem ein Energiespeicher betrachtet wird. Dadurch kann die aus der Abwärme rückgewonnene Energie bei Bedarf gespeichert werden, wodurch Energierückgewinnung und Nutzung der rückgewonnen Energie zeitlich entkoppelt werden. Zur Einbindung des Energiespeichers in das Produktionsplanungsmodell eines seriellen, mehrstufigen Produktionssystems werden zunächst die Lade- und Entladevorgänge des Energiespeichers mathematisch modelliert und in ein gemischt-ganzzahliges lineares Optimierungsmodell integriert. Dieses Optimierungsmodell berechnet, in welchen Perioden im Planungshorizont die Produktionsstufen produzieren sollen und wie der Energiespeicher betrieben werden soll, um die

Summe aus Produktions- und Energiekosten zu minimieren. Abschließend wird in einer numerischen Studie die Leistungsfähigkeit des Optimierungsmodells unter Berücksichtigung zeitabhängiger Energiepreise untersucht.

Wie die Artikel 2 und 3 entwickelt auch Artikel 4 ein mathematisches Modell zur energieeffizienten Produktionsplanung. Im Unterschied zu den Artikeln 2 und 3 setzt Artikel 4 dabei jedoch nicht auf Unterstützung der Energieversorgung des Produktionssystems durch Abwärmerückgewinnung, sondern durch Windkraftanlagen. Eine große Herausforderung der Integration von Windenergie in die Produktionsplanung erwächst aus der Unsicherheit der Energiebereitstellung durch Windkraftanlagen infolge sich im Zeitablauf teilweise stark verändernder Windgeschwindigkeiten. Zur Abbildung der Variabilität der generierten Windenergie wird in Artikel 4 zunächst eine große Menge an Szenarien erzeugt, welche die in den jeweiligen Planungsperioden zur Verfügung stehende Windenergie beschreiben. Eine systematisch reduzierte Teilmenge dieser Szenarien dient im nächsten Schritt als Eingabeparameter für ein gemischt-ganzzahliges lineares Optimierungsmodell, welches vor Beginn des Planungszeitraums einen Maschinenbelegungs- und einen Energiebezugsplan für ein mehrstufiges Flow-Shop-System berechnet. Das Ziel des Optimierungsmodells besteht in der gleichzeitigen Minimierung der gewichteten Durchlaufzeit sowie der erwarteten Energiekosten. Der Energiebezugsplan, welcher festlegt, ob die vom Produktionssystem benötigte Energie von den Windkraftanlagen bezogen wird oder aus dem Stromnetz bezogen werden muss, wird anschließend sukzessive in jeder Planungsperiode angepasst, wenn die realen Windgeschwindigkeiten und damit die reale Energiebereitstellungskapazität der Windkraftanlagen feststehen. In einer numerischen Studie wird schließlich das integrierte Produktionsplanungsmodell unter unterschiedlichen Windbedingungen und zeitabhängigen Energiepreisen evaluiert.

In Teil B der Dissertation beschäftigt sich Artikel 5 ebenfalls mit der effizienten Steuerung mehrstufiger Produktionssysteme. Während sich die Artikel 1 bis 4 mit der effizienten Abstimmung von Produktionsentscheidungen auf die Bereitstellung von rückgewonnener Produktionsprozessabwärme und Windenergie fokussierten, stellt Artikel 5 die effiziente Abstimmung von Produktionsentscheidungen auf menschliche Faktoren wie Lernen und Vergessen ins Zentrum der Betrachtung. Zu diesem Zweck wird zunächst ein diskretes, ereignisorientiertes Simulationsmodel entwickelt, welches ein serielles, mehrstufiges Produktionssystem unter Berücksichtigung von Lernen und Vergessen der zuständigen Arbeiter beschreibt. In einer zweigeteilten Simulationsstudie werden im ersten Schritt Produktionsparameter identifiziert, welche die Systemleistung in signifikanter Weise beeinflussen. Darauf aufbauend werden flexible Puffermanagementregeln erarbeitet, mit deren Hilfe vorher identifizierte, negative Einflüsse bestimmter Produktionsparameter eingeschränkt werden sollen. Im zweiten Teil der Simulationsstudie wird die Leistungsfähigkeit der Puffermanagementregeln unter diversen Produktionsparameterszenarien systematisch überprüft.

Abstract

This cumulative dissertation consists of five papers published in different scientific journals. All five papers are concerned with multi-stage production planning. Due to differing foci of the papers, this dissertation is divided into two parts. Part A embraces Papers 1 to 4 and contributes to a research stream that investigates energy aspects in multi-stage production planning. Part B features Paper 5 and investigates the influence of worker learning and forgetting on multi-stage production systems. Aside from the differing foci, the five papers also vary in the methodologies employed. The first paper presents a systematic review of the state of the art of decision support models for energy-efficient production planning. The remaining four papers develop mathematical models for supporting production planning decisions considering different measures to foster energy efficiency (Papers 2 to 4) and considering human factors (Paper 5). Papers 2 to 4 analytically solve the developed mathematical models. In contrast, Paper 5 draws on discrete-event simulation to derive effective production control policies. The following paragraphs summarize the five papers.

Paper 1 systematically reviews the literature on quantitative decision support models which integrate energy considerations into mid-term and short-term production planning of manufacturing companies. The sampled articles are then classified and synthesized with regard to the characteristics of the modeling approaches representing different energy aspects. Based on the discussion of the sampled articles, Paper 1 identifies future research opportunities in the area of energy-aware production planning and thereby sets the stage for Papers 2 to 4 of this dissertation.

Paper 2 studies how waste heat rejected by manufacturing processes in a two-stage production system can be utilized to foster energy-efficient production planning. Among the different ways of recovering waste heat, Paper 2 focuses on the conversion of waste heat into electricity using an Organic Rankine Cycle (ORC). To this end, it first describes this thermodynamic conversion process mathematically and then integrates it into a lot sizing model such that the electricity from the recovered waste heat supports the energy supply of the production stages. In a next step, Paper 2 proposes a solution procedure which derives optimal values for the lot size, the production rates of the two production stages, and the number of shipments between the two production stages that minimize production- and energy-related costs. In a numerical analysis, Paper 2 investigates how considering waste heat recovery in production planning can effectively reduce energy consumption in manufacturing and how it impacts production planning decisions.

Paper 3 extends the model developed in Paper 2 and studies the use of an ORC-based waste heat recovery system (WHRS) combined with an electrical energy storage system (EESS). With the help of an EESS, generation and consumption of electricity from the WHRS can be decoupled. Using mixed integer linear programming (MILP), Paper 3 proposes a mathematical model that integrates time-varying energy prices alongside the technological processes of the WHRS and the EESS into the production planning problem of a serial multi-stage production system. This MILP model determines when production stages should process and how the WHRS and the EESS should be operated to optimize production- and energy-related costs. In a numerical analysis, Paper 3 examines how attaching an EESS to a WHRS can enhance its relevance for energy-aware production planning, particularly through providing the opportunity to store energy generated from waste heat in times of low energy prices and to then use it in times of high energy prices.

Similar to Papers 2 and 3, Paper 4 also contributes to the research stream on energy-aware production planning. However, in contrast to the preceding papers, it focuses on the integration of onsite wind power into production scheduling of a flow shop system. Coordinating production scheduling and the energy supply from an onsite wind turbine poses a major challenge to researchers and practitioners as the intermittent character of wind power due to the vagaries of wind speed adds a stochastic component to production scheduling. The approach suggested in Paper 4 overcomes this challenge by first generating a large number of wind power scenarios that characterize the variability and the time dependence of wind power over time. A systematically reduced subset of these wind power scenarios subsequently serves as an input to a two-stage stochastic optimization procedure. Based on the reduced wind power scenario set, this procedure first computes a production schedule and energy supply decisions that minimize the total weighted flow time and the expected energy cost. The energy supply decisions derive whether the electricity generated by the wind turbine during a given time slot should be used to support the energy supply of the machines or be fed into the grid and thus determine the amount of electricity that needs to be drawn from the grid to guarantee an uninterrupted energy supply of the machines. These energy supply decisions are adjusted in a second step in real time as the actual wind power data are gradually revealed. In a numerical example, the effectiveness of the procedure in incorporating energy supply from non-dispatchable renewable energy sources (RES) in production scheduling is shown under various conditions.

Part B of this dissertation consists of Paper 5. As Papers 1 to 4, Paper 5 is also concerned with efficiently managing multi-stage production systems. Papers 1 to 4 concentrated on how to effectively tailor the operation of production stages to energy supply from WHRSs or RES and time-varying energy prices. In contrast to these works, Paper 5 focuses on how to attune the operation of production stages to human characteristics such as individual worker learning and forgetting. To this end, Paper 5 first develops a generic simulation model of a serial multi-stage production system subject to learning and forgetting effects. Subsequently, it carries out an extensive simulation experiment to identify parameters of the production stages and their interactions which exercise a significant influence on system performance. Paper 5 then proposes flexible buffer management rules to counteract the impact of adverse production parameter combinations detected in the preceding simulation experiment. In a second simulation experiment, the performance of these buffer management rules is evaluated under various input parameter combinations.

Outline

Table of Contents

List of Abbreviations

AES	Applied energy sources
Btu	British thermal units
C	Celsius
cdf	Cumulative distribution function
CPP	Critical peak pricing
DER	Distributed energy resources
DR	Demand response
ecdf	Emprirical cumulative distribution fuction
EDF	Electricity demand function
EE	Energy-efficient
EEPP	Energy-efficient production planning
EESS	Electrical energy storage system
ESS	Energy storage system
EUR	Euro
FES	Final energy sources
FFS	Flexible flow shop
FIFO	First in - first out
FJS	Flexible job shop
FMS	Flexible manufacturing system
FS	Flow shop
GHG	Greenhouse gas
h	Hour
ILC	Interruptible load contract
J	Joule
JS	Job shop
K	Kelvin
kg	Kilogram
kJ	Kilojoule
KPI	Key performance indicator
kW	Kilowatt
kWh	Kilowatt hour
l	Liter
LCOE	Levelized cost of electricity
LHS	Latin Hypercube Sampling

LM	Load management
LT	Load tracking
m	Meter
MILP	Mixed integer linear programming
min	Minute
mt	Metric ton
ORC	Organic Rankine Cycle
PFS	Permutation flow shop
PP	Production planning
R&D	Research and development
RBP	Rolling blackout policy
RES	Renewable energy sources
s	Second
TOU	Time-of-use
TWh	Terrawatt hour
U.S.	United States of America
USD	United States dollar
VIF	Variance inflation factor
W	Watt
WHRS	Waste heat recovery system

List of Figures

Paper 1

Paper 2

Paper 3

Paper 4

Paper 5

List of Tables

Introduction

Paper 1

Paper 2

Paper 3

Paper 4

Paper 5

Introduction

This cumulative dissertation consists of five papers published in different scientific journals (see Table 1). All five papers are concerned with multi-stage production planning. Due to differing foci of the papers, this dissertation is divided into two parts. Part A embraces Papers 1 to 4 and contributes to a research stream that investigates energy aspects in multi-stage production planning. Part B features Paper 5 and investigates the influence of worker learning and forgetting on multi-stage production systems. Aside from the differing foci, the five papers also vary in the methodologies employed. The first paper presents a systematic review of the state of the art of decision support models for energy-efficient production planning. The remaining four papers develop mathematical models for supporting production planning decisions considering different measures to foster energy efficiency (Papers 2 to 4) and considering human factors (Paper 5). Papers 2 to 4 analytically solve the developed mathematical models. In contrast, Paper 5 draws on discrete-event simulation to derive effective production control policies. The following paragraphs provide an introduction to the research areas the papers contribute to and explain the research gaps the papers are looking to fill.

Paper 1 provides an introduction into the research stream of energy-efficient production planning. This research stream aims at expanding the field of traditional production planning by considering various energy-related aspects. Traditional production planning approaches typically solely aim at minimizing production-related goals such as inventory holding cost and machine setup cost (e.g., Goyal and Szendrovits, 1986; Silver, 1990; Glock, 2010), makespan (e.g., Lai et al., 1997; Janiak, A., 1998; Laub et al., 2007), or flow time (e.g., Azizoğlu et al., 2001; Chung et al., 2002; Xiong et al., 2015). However, the ongoing depletion of non-renewable resources, rising energy prices, and the advancing recognition of the impact of manufacturing processes on greenhouse gas (GHG) emissions have increasingly urged production planners to broaden their perspective. As a consequence, a new research stream has gained traction over the past decades that strives for also taking account of energy-related goals in production planning aside from the traditional production-related objectives. Prototypes of such energy-related goals are the minimization or restriction of energy consumption (e.g., Mouzon et al., 2007; May et al., 2015; Wang et al., 2015), energy cost (e.g., Moon and Park, 2014; Zanoni et al., 2014; Zhang et al., 2014), or energy-related GHG emissions (e.g., Fang et al., 2011; Liu, 2014; Sharma et al., 2015). The aim of Paper 1 is to review this research stream and to synthesize the plethora of works that integrate energy aspects into production planning in various ways. In contrast to related reviews of Giret et al. (2015) and Gahm et al. (2016), Paper 1 analyzes and classifies the relevant articles from a highly model-driven perspective and thus fills an important research gap. Furthermore, it not only considers machine scheduling models, but also master production scheduling and capacity planning as well as lot sizing models. In this way, Paper 1 provides the first comprehensive overview of modeling approaches to integrate energy aspects into production planning problems which Hax and Meal (1975) referred to as mid-term and short-term production planning problems. Based on the discussion of the sampled articles, Paper 1 identifies future research opportunities in the area of energy-aware production planning and thereby sets the stage for Papers 2 to 4 of this dissertation.

Table 1: Overview of papers included in this cumulative dissertation.

Focus	Authors	Title	Journal
Multi-stage production planning considering waste heat recovery, energy storage, and renewable energy sources	K. Biel, C.H. Glock	Systematic literature review of decision support models for energy-efficient production planning	Computers & Industrial Engineering
	K. Biel, C.H. Glock	On the use of waste heat in a two-stage production system with controllable production rates	International Journal of Production Economics
	K. Biel, C.H. Glock	Prerequisites of efficient decentralized waste heat recovery and energy storage in production planning	Journal of Business Economics
	K. Biel, F. Zhao, J.W. Sutherland, C.H. Glock	Flow shop scheduling with grid-integrated onsite wind power using stochastic MILP	International Journal of Production Research
Multi-stage production planning considering learning and forgetting effects	K. Biel, C.H. Glock	Governing the dynamics of multi-stage production systems subject to learning and forgetting effects: A simulation study	International Journal of Production Research

Paper 2 addresses one of these research opportunities and studies how waste heat rejected by manufacturing processes in a two-stage production system can be utilized to foster energy-efficient production planning. The potential of waste heat recovery to enhance the energy efficiency of manufacturing processes is evident, given that industrial energy consumption accounts for approximately one third of the world's total delivered energy (U.S. Energy Information Administration, 2015), and that about 40 to 50 percent of the energy consumed by the industrial sector is rejected as waste heat (Schaefer, 1995; Hung et al., 1997; López et al., 1998). Among the different ways of recovering waste heat, Paper 2 focuses on the conversion of waste heat into electricity using an Organic Rankine Cycle (ORC). To this end, it first describes this thermodynamic conversion process mathematically and then integrates it into a lot sizing model such that the electricity from the recovered waste heat supports the energy supply of the production stages. In a next step, Paper 2 proposes a solution procedure which derives optimal values for the lot size, the production rates of the two production stages, and the number of shipments between the two production stages that minimize production- and energy-related costs. Even though the fraction of the rejected waste heat that can technologically and economically be utilized is highly contingent on the characteristics of the waste heat, the waste heat recovery system (WHRS), and the ambient conditions (Hirzel et al., 2013), Paper 2 shows in a numerical analysis how considering waste heat recovery in production planning can effectively reduce energy consumption in manufacturing. As waste heat recovery has not been investigated in the context of operations management before, Paper 2 contributes to closing a major gap in the literature on energy-aware production planning.

One drawback of the way the WHRS is integrated into the production system in Paper 2 stems from the lack of flexibility regarding the use of the electricity generated from the waste heat. In fact, the electricity from the WHRS needs to be consumed immediately after its generation if no energy storage system is available. If an energy storage system was available, then generation and consumption of electricity from the WHRS could be decoupled. Paper 3 consequently addresses this link between energy generation and consumption and investigates the use of an ORC-based WHRS combined with an electrical energy storage system (EESS) in the context of a serial multi-stage production system. In contrast to Paper 2, which assumes a flat energy consumption charge, Paper 3 assumes time-varying energy prices, so-called time-of-use (TOU) prices. By means of TOU pricing profiles, utility companies attempt to smooth their customers' energy consumption throughout the day to avoid peak loads which entail substantial costs for them. To this end, utility companies attach high prices to energy consumption in times of generally higher total energy demand and low prices to energy consumption in times of generally lower total energy demand (Moon and Park, 2014). Using mixed integer linear programming (MILP), Paper 3 proposes a mathematical model that integrates a TOU pricing scheme alongside the technological processes of the WHRS and the EESS into the production planning problem of the serial multi-stage production system. This MILP model determines when production stages should process and how the WHRS and the EESS should be operated to optimize production- and energy-related costs. In a numerical analysis, Paper 3 examines how attaching an EESS to a WHRS can enhance its applicability, particularly through providing the opportunity to store energy generated from waste heat in times of low energy prices and to then use it in times of high energy prices. Overall, Paper 3 contributes to responding to the research question of how waste heat recovery can be integrated into production planning to effectively promote energy-efficient manufacturing.

Similar to Papers 2 and 3, Paper 4 also adds to the research stream on energy-aware production planning. However, in contrast to the preceding papers, it focuses on the integration of non-dispatchable renewable energy sources (RES), such as wind and solar power, into traditional production planning.

The need for powerful decision support models that connect production planning with the energy supply from non-dispatchable RES is directly tied to the substantial increase in global installed renewable energy capacity over the last decade (Whiteman et al., 2016). This expansion of the use of RES has been accompanied by a rising awareness of the benefits RES may provide manufacturing companies with. On the one hand, firms are increasingly installing onsite RES to reduce their dependence on the dictate of TOU pricing profiles (Lantz, 2016). On the other hand, manufacturing companies rely on RES to reduce their energy-related GHG emissions to cut expenses for emission certificates and to foster an environmentally conscious reputation (Tognetti et al., 2015). As wind power has gained particular attention among the different types of RES over the course of the recent expansion of RES (U.S. Department of Energy, 2015), Paper 4 focuses on the integration of onsite wind power into production scheduling of a flow shop system. Coordinating production scheduling and the energy supply from an onsite wind turbine poses a major challenge to researchers and practitioners as the intermittent character of wind power due to the vagaries of wind speed adds a stochastic component to production scheduling. Moon and Park (2014) and Liu (2016) were the first to propose production scheduling models which considered non-dispatchable RES. Moon and Park (2014), however, assumed that the model could choose in each time slot the amount of energy from RES to be used from a predetermined interval. Liu (2016) resorted to a similar approach, but generated the interval boundaries of the available energy from RES randomly from a uniform distribution. Yet, neither of the two approaches realistically portrays the intermittency of wind power in general and the time dependence of energy generated by a wind turbine in successive time slots in particular (Morales et al., 2010). Thus, the production scheduling models of both Moon and Park (2014) and Liu (2016) are limited in their capacity to effectively incorporate the uncertainty attached to the energy supply from non-dispatchable RES.

The MILP approach that can be used for coordinating production scheduling and the energy supply from an onsite wind turbine suggested in Paper 4 overcomes this challenge by first generating a large number of wind power scenarios that characterize the variability and the time dependence of wind power over time. A systematically reduced subset of these wind power scenarios subsequently serves as an input to a two-stage stochastic optimization procedure. Based on the reduced wind power scenario set, this procedure first computes a production schedule and energy supply decisions that minimize the total weighted flow time and the expected energy cost. The energy supply decisions derive whether the electricity generated by the wind turbine during a given time slot should be used to support the energy supply of the machines or be fed into the grid and thus determine the amount of electricity that needs to be drawn from the grid to guarantee an uninterrupted energy supply of the machines. These energy supply decisions are adjusted in a second step in real time as the actual wind power data are gradually revealed. In a numerical example, the effectiveness of the procedure in incorporating energy supply from non-dispatchable RES in production scheduling is shown under various conditions. Thus, Paper 4, which concludes Part A of this dissertation, successfully contributes to answering the research question of how the uncertainty induced into production scheduling by the intermittency of non-dispachable RES can efficaciously be handled while simultaneously considering both production- and energy-related goals.

Part B of this dissertation consists of Paper 5. As Papers 1 to 4, Paper 5 is also concerned with efficiently managing multi-stage production systems. Papers 1 to 4 concentrated on how to effectively tailor the operation of production stages to energy supply from WHRSs or RES and time-varying energy prices. In contrast to these works, Paper 5 focuses on how to attune the operation of production stages to human characteristics such as individual worker learning and forgetting. For decades, researchers and practitioners have realized the necessity to consider worker learning and forgetting to optimize production control of manufacturing processes relying on human contributions. Worker learning occurs in consequence of multiple repetitions of a task and reduces the time required to produce one unit of output by an individual or a team (Yelle, 1979). In contrast, forgetting stems from interruptions of the production routine and results in increased processing times of one unit of output due to lost experience (Globerson et al., 1989). Hence, production stages which require worker interventions are prone to changing their production rates over time. Failure to acknowledge these changes when setting up production control policies may lead to suboptimal system performance (Smunt, 1987; Bogaschewsky and Glock, 2009). For this reason, researchers in operations management have proposed various approaches to incorporate learning and forgetting effects into different production planning problems such as production scheduling (e.g., Dolgui et al., 2012; Gordon et al., 2012; Pan et al., 2014), assembly line balancing (e.g., Shafer, 2001; Bukchin and Cohen, 2013; Li and Boucher, 2017), or lot sizing (e.g., Jaber and Bonney, 1999; Jaber and Khan, 2010; Glock and Jaber, 2013b), which Paper 5 is also concerned with. However, these approaches largely rely on static decision support models which are unable to react to the dynamics introduced into production systems through constantly changing production rates. As the production flow through a coherent production system heavily depends on the interactions of the production stages, efficient production control of these dynamics needs to be capable of describing such interactions. Glock and Jaber (2013a) showed that it is possible to derive a closed-form expression of the interactions of a simple two-stage production system. However, the interactions of production stages within more complex production systems can only be modeled analytically by means of nonlinear differential equations. For this reason, the use of simulation techniques to optimize the control of production processes including dynamic components has in the wake of soaring computing capacities enjoyed an increasing popularity (Kleijnen, 2008).

The works of Finch and Luebbe (1995) and Mak et al. (2014) are advocates of this trend in the area of production planning under consideration of worker learning. Paper 5 builds on these articles and significantly extends their focus as it additionally takes account of forgetting effects, considerably deepens the investigation of interactions across multiple production stages, and proposes real-time production control policies to handle the dynamics resulting from constantly changing production rates. To this end, Paper 5 first develops a generic simulation model of a serial multi-stage production system subject to learning and forgetting effects. Subsequently, it carries out an extensive simulation experiment to identify parameters of the production stages and their interactions which exercise a significant influence on system performance. Paper 5 then proposes flexible buffer management rules to counteract the impact of adverse production parameter combinations detected in the preceding simulation experiment. In a second simulation experiment, the performance of these buffer management rules is evaluated under various input parameter combinations. In this way, Paper 5, which concludes this dissertation, contributes to responding to the research question of how production control can react in real time to dynamics in the production flow caused by worker learning and forgetting, which traditional, rigid production policies are incapable of.

Overall, this dissertation adds to the literature a systematic literature review of the state of the art of decision support models for energy-aware production planning (Paper 1), three decision support models to foster energy-aware manufacturing by integrating a WHRS, an EESS, and non-dispatchable RES into production planning (Papers 2 to 4), and one simulation study on how to control the dynamics in production systems subject to learning and forgetting effects (Paper 5). Despite the scientific character of the papers, the literature synthesis and the numerical analyses clearly underscore the practical applicability of the reviewed and developed decision support models and production control policies and hence their relevance for practitioners in the manufacturing industry. Detailed managerial implications can be found in the final section of each paper. In addition, these sections elaborate on the limitations of the chosen research approaches and highlight future research opportunities.

References

Azizoğlu, M., Çakmak, E., Kondakci, S., 2001. A flexible flowshop problem with total flow time minimization. *European Journal of Operational Research*, 132 (3), 528-538.

Bogaschewsky, R., Glock, C.H., 2009. Rational inefficiencies in lot-sizing models with learning effects. *Zeitschrift für Betriebswirtschaft*, 4, 37-57.

Bukchin, Y., Cohen, Y., 2013. Minimising throughput loss in assembly lines due to absenteeism and turnover via work-sharing. *International Journal of Production Research*, 51 (20), 6140-6151.

Chung, C.S., Flynn, J., Kirca, O., 2002. A branch and bound algorithm to minimize the total flow time for m-machine permutation flowshop problems. *International Journal of Production Economics*, 79 (3), 185-196.

Dolgui, A., Gordon, V., Strusevich, V., 2012. Single machine scheduling with precedence constraints and positionally dependent processing times. *Computers & Operations Research*, 39 (6), 1218-1224.

Fang, K., Uhan, N., Zhao, F., Sutherland, J.W., 2011. A new approach to scheduling in manufacturing for power consumption and carbon footprint reduction. *Journal of Manufacturing Systems*, 30 (4), 234-240.

Finch, B.J., Luebbe, R., 1995. The impact of learning rate and constraints on production line performance. *International Journal of Production Research*, 33 (3), 631-642.

Gahm, C., Denz, F., Dirr, M., Tuma, A., 2016. Energy-efficient scheduling in manufacturing companies: A review and research framework. *European Journal of Operational Research*, 248 (3), 744-757.

Giret, A., Trentesaux, D., Prabhu, V., 2015. Sustainability in manufacturing operations scheduling: A state of the art review. *Journal of Manufacturing Systems*, 37, 126-140.

Globerson, S., Levin, N., Shtub, A., 1989. The impact of breaks on forgetting when performing a repetitive task. *IIE Transactions*, 21 (4), 376-381.

Glock, C.H., 2010. Batch sizing with controllable production rates. *International Journal of Production Research*, 48 (20), 5925-5942.

Glock, C.H., Jaber, M.Y., 2013a. Learning effects and the phenomenon of moving bottlenecks in a two-stage production system. *Applied Mathematical Modelling*, 37 (18–19), 8617-8628.

Glock, C.H., Jaber, M.Y., 2013b. A multi-stage production-inventory model with learning and forgetting effects, rework and scrap. *Computers & Industrial Engineering*, 64 (2), 708-720.

Goyal, S., Szendrovits, A.Z., 1986. A constant lot size model with equal and unequal sized batch shipments between production stages. *Engineering Costs and Production Economics*, 10 (3), 203-210.

Hax, A.C., Meal, H.C., 1975. *Hierarchical integration of production planning and scheduling*. In: Geisler, M.A. (Ed.), Logistics: TIMS Studies in Management Science, vol. 1, North-Holland, Amsterdam, pp. 53-69.

Hung, T.C., Shai, T.Y., Wang, S.K., 1997. A review of Organic Rankine Cycles (ORCs) for the recovery of low-grade waste heat. *Energy*, 22 (7), 661-667.

Hirzel, S., Sontag, B., Rhode, C., 2013. *Industrielle Abwärmenutzung*. Fraunhofer-Institut für System- und Innovationsforschung (ISI), Karlsruhe.

Iversen, E.B., Morales, J.M., Møller, J.K., Madsen, H., 2016. Short-term probabilistic forecasting of wind speed using stochastic differential equations. *International Journal of Forecasting*, 32, 981-990.

Jaber, M.Y., Bonney, M., 1999. The economic manufacture/order quantity (EMQ/EOQ) and the learning curve: Past, present, and future. *International Journal of Production Economics*, 59 (1), 93-102.

Jaber, M.Y., Khan, M., 2010. Managing yield by lot splitting in a serial production line with learning, rework and scrap. *International Journal of Production Economics*, 124 (1), 32-39.

Janiak, A., 1998. Minimization of the makespan in a two-machine problem under given resource constraints. *European Journal of Operational Research*, 107 (2), 325-337.

Kleijnen, J.P.C., 2008. *Design and analysis of simulation experiments*. Springer, New York, NY.

Lai, T.-C., Sotskov, Y.N., Sotskova, N.Y., Werner, F., 1997. Optimal makespan scheduling with given bounds of processing times. *Mathematical and Computer Modelling*, 26 (3), 67-86.

Lantz, E., Sigrin, B., Gleason, M., Preus, R., Baring-Gould, I., 2016. *Assessing the future of distributed wind: Opportunities for behind-the-meter projects*. National Renewable Energy Laboratory, Golden, CO.

Laub, J.D., Fowler, J.W., Keha, A.B., 2007. Minimizing makespan with multiple-orders-per-job in a two-machine flowshop. *European Journal of Operational Research*, 182 (1), 63-79.

Li, Y., Boucher, T.O., 2017. Assembly line balancing problem with task learning and dynamic task reassignment. *International Journal of Advanced Manufacturing Technology*, 88 (9-12), 3089-3097.

Liu, C.-H., 2014. Approximate trade-off between minimisation of total weighted tardiness and minimisation of carbon dioxide (CO_2) emissions in bi-criteria batch scheduling problem. *International Journal of Computer Integrated Manufacturing*, 27 (8), 759-771.

Liu, C.-H., 2016. Mathematical programming formulations for single-machine scheduling problems while considering renewable energy uncertainty. *International Journal of Production Research*, 54 (4), 1122-1133.

López, L., Blanco, J.M., Bonilla, J.J., Bacza, S., Sala, J.M., 1998. Determination of energy and exergy of waste heat in the industry of the Basque country. *Applied Thermal Engineering*, 18 (3-4), 187-197.

Mak, L.C., Wong, W.K., Leung, Y.S., 2014. A simulation analysis of the impact of production lot size and its interaction with operator competence on manufacturing system performance. *Simulation Modelling Practice and Theory*, 49, 203-214.

May, G., Stahl, B., Taisch, M., Prabhu, V., 2015. Multi-objective genetic algorithm for energy-efficient job shop scheduling. *International Journal of Production Research*, 53 (23), 7071-7089.

Moon, J.-Y., Park, J., 2014. Smart production scheduling with time-dependent and machine-dependent electricity cost by considering distributed energy resources and energy storage. *International Journal of Production Research*, 52 (13), 3922-3939.

Morales, J.M., Mínguez, R., Conejo, A.J., 2010. A methodology to generate statistically dependent wind speed scenarios. *Applied Energy*, 87 (3), 843-855.

Mouzon, G., Yildirim, M.B., Twomey, J., 2007. Operational methods for minimization of energy consumption of manufacturing equipment. *International Journal of Production Research*, 45 (18-19), 4247-4271.

Pan, E., Wang, G., Xi, L., Chen, L., Han, X., 2014. Single-machine group scheduling problem considering learning, forgetting effects and preventive maintenance. *International Journal of Production Research*, 52 (19), 5690-5704.

Schaefer, H., 1995. *Energiewirtschaftliche Bedeutung der Nutzung von Abfallenergien*. In: Fratzscher, W., Stephan, K. (Eds.), Abfallenergienutzung - Technische, wirtschaftliche und soziale Aspekte, vol. 2, Akadmie Verlag, Berlin, pp. 42-60.

Shafer, S.M., Nembhard, D.A., Uzumeri, M.V., 2001. The effects of worker learning, forgetting, and heterogeneity on assembly line productivity. *Management Science*, 47 (12), 1639-1653.

Sharma, A., Zhao, F., Sutherland, J.W., 2015. Econological scheduling of a manufacturing enterprise operating under a time-of-use electricity tariff. *Journal of Cleaner Production*, 108, 256-270.

Silver, E.A., 1990. Deliberately slowing down output in a family production context. *International Journal of Production Research*, 28 (1), 17-27.

Smunt, T.L. 1987. The impact of worker forgetting on production scheduling. *International Journal of Production Research*, 25 (5), 689-701.

Tognetti, A., Grosse-Ruyken, P.T., Wagner, S.M., 2015. Green supply chain network optimization and the trade-off between environmental and economic objectives. *International Journal of Production Economics*, 170, 385-392.

U.S. Department of Energy, 2015. *Revolution ... now: The future arrives for five clean energy technologies – 2015 update*. Washington, DC.

U.S. Energy Information Administration, 2015. *Annual energy outlook with projections to 2040*. Washington, DC.

Wang, S., Lu, X., Li, X.X., Li, W.D., 2015. A systematic approach of process planning and scheduling optimization for sustainable machining. *Journal of Cleaner Production*, 87, 914-929.

Xiong, F., Xing, K., Wang, F., 2015. Scheduling a hybrid assembly-differentiation flowshop to minimize total flow time. *European Journal of Operational Research*, 240 (2), 338-354.

Yelle, L.E. 1979. The learning curve: Historical review and comprehensive survey. *Decision Sciences*, 10 (2), 302-328.

Zanoni, S., Bettoni, L., Glock, C.H., 2014. Energy implications in a two-stage production system with controllable production rates. *International Journal of Production Economics*, 149, 164-171.

Zhang, H., Zhao, F., Fang, K., Sutherland, J.W., 2014. Energy-conscious flow shop scheduling under time-of-use electricity tariffs. *CIRP Annals - Manufacturing Technology*, 63 (1), 37-40.

Whiteman, A., Rinke, T., Esparrago, J., Elsayed, S., 2016. *Renewable energy capacity statistics 2016*. International Renewable Energy Agency, Masdar.

Part A Multi-stage production planning considering waste heat recovery, energy storage, and renewable energy sources

Paper 1 Systematic literature review of decision support models for energy-efficient production planning[1]

Authors: Konstantin Biel, Christoph H. Glock

Type of publication: Journal article

Publication details: *Computers & Industrial Engineering*, 101, 243-259, 2016.

Abstract

Following the scarcity of resources, rising energy prices, and an increasing awareness of the role manufacturing plays in the generation of greenhouse gas emissions, the consumption of energy has more and more been the subject of research on production planning over the last decade. Even though recent years have witnessed a dramatic increase in the number of works published in this area, several related research questions have been opened up without sufficiently linking research approaches and research insights. The aim of this paper is to investigate the links between these questions and to highlight how the modeling approaches developed for different manufacturing systems, energy pricing policies, and energy efficiency criteria can benefit from each other and lead to more advanced energy-efficient production planning approaches. Therefore, this paper provides a review of the state-of-the-art of decision support models that integrate energy aspects into mid-term and short-term production planning of manufacturing companies. The paper first highlights the increasing importance of energy consumption in manufacturing and shows how considering energy consumption in production planning can contribute to more energy-efficient production processes. Subsequently, the paper outlines the review methodology used and descriptively analyzes the sampled papers. Afterwards, the selected papers are categorized according to the production planning tasks considered. From this classification, gaps in the existing literature are derived and potential areas for future research are suggested.

Keywords:
Energy efficiency; Sustainable manufacturing systems; Energy consumption; Production planning; Scheduling; Review

---

[1] Reprinted from *Computers & Industrial Engineering*, 101, Biel, K., Glock, C.H., Systematic literature review of decision support models for energy-efficient production planning, 243-259, 2016, with permission from Elsevier.

1 Introduction

Over the last decades, more and more researchers have directed their attention towards energy-efficient (EE) production planning (PP). The increasing popularity this research stream has enjoyed goes hand in hand with the growth in worldwide energy consumption, which – in light of the scarcity of resources and rising energy prices – is one of the main drivers of these research efforts. Global primary energy consumption has increased by more than 50 percent from 343 British thermal units (Btu) in 1993 to 524 Btu in 2012 (U.S. Energy Information Administration, 2015). Figure 1 shows that Asia and Oceania have contributed the greatest share to this development, with the dynamic Chinese economy being one of the main drivers of primary energy consumption in Asia. It is striking that only a single country among the top ten energy consumers – Germany – displayed a slightly negative compound annual growth rate of −0.2 percent from 1993 to 2012.

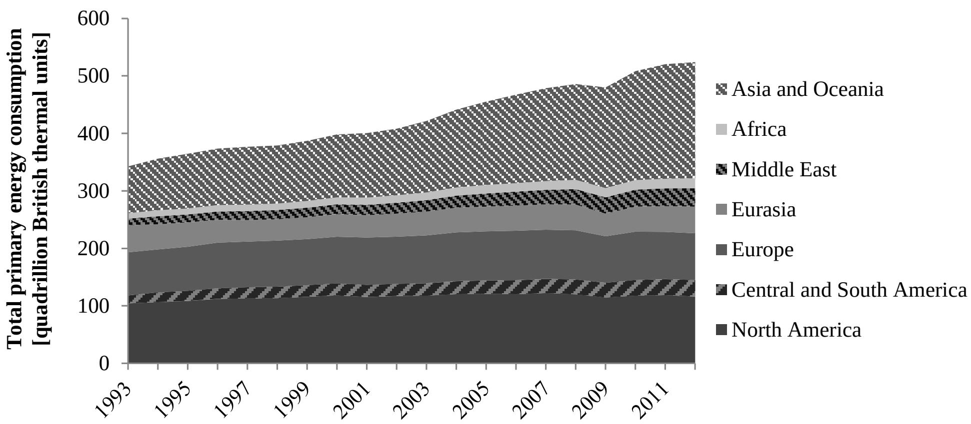


Figure 1: Total primary energy consumption by regions from 1993 to 2012 (U.S. Energy Information Administration, 2015).

In 2010, the industrial sector was responsible for 39.4 percent of the overall energy consumption. Despite the ongoing transition of many OECD countries from manufacturing economies to service economies, this share is expected to decrease only marginally to reach 37.4 percent in 2040 (U.S. Energy Information Administration, 2013). The energy consumption of the industrial sector largely originates from manufacturing industries. In 2010, the chemical industry alone was accountable for 19 percent, the iron and steel industry for 15 percent, and the nonmetallic minerals industry for 7 percent of the energy consumption within the industrial sector (U.S. Energy Information Administration, 2013). Hence, it is obvious that measures to enhance energy efficiency in manufacturing bear a great potential for reducing energy consumption and energy-related CO_2 emissions, and that these approaches will not lose their relevance in the decades to come.

In general, the literature on EE manufacturing can be divided into two streams of research: (I) studies aiming at reducing energy consumption by technological advancements of production processes (e.g., Hasanbeigi and Price, 2012; Neugebauer et al., 2011; Zhou et al., 2016) and (II) studies trying to reduce energy consumption by adjusting managerial parameters of the production process, which we term energy-efficient production planning (EEPP). The paper at hand solely reviews papers that belong to the second category of the EE manufacturing literature. In general, the aim of EEPP models is to compute production plans which do not only take account of traditional PP objectives, such as the minimization of inventory holding cost, setup cost, or total completion time, but also of energy-related objectives, such as the minimization of energy consumption, energy cost, or energy-related greenhouse gas (GHG) emissions, and of energy-related constraints, such as the compliance with a maximum contracted power demand or with a GHG emission constraint. In comparison to approaches seeking to technologically enhance the energy efficiency of manufacturing systems, EEPP models have become increasingly popular in practice as their adoption is usually not tied to large investments. Hence, research on EEPP has gained considerable attention in recent years, which resulted in numerous publications, especially in the last decade. However, as many of these papers were published just recently, most of the existing modeling approaches are not well connected and only a few integrated decision support models for EEPP have been suggested so far.

This review aims at providing both researchers and practitioners with a structured overview of decision support models that take account of energy efficiency issues in PP and at showing how energy efficiency criteria have been modeled in the respective works. Practitioners may use this review as an introduction into how decision support models may help to improve energy efficiency in manufacturing and as a guideline to identify an appropriate decision support model for a particular application. Researchers may use this review to get an overview of (traditional) PP problems that have already been modified to take account of energy efficiency issues, and they may use our synthesis to identify future research opportunities.

The explicit focus of our review on modeling details for EEPP differentiates it from existing review papers in this area. Duflou et al. (2012), for instance, conceptually categorized methods and techniques that help to improve energy efficiency in manufacturing according to the system scale level they belong to. Similarly, Haapala et al. (2013) reviewed management concepts, methods, and tools for sustainable manufacturing and distinguished between approaches striving for EE manufacturing processes and approaches striving for EE manufacturing systems. Garetti and Taisch (2012) provided a general overview of current sustainable research initiatives, research clusters, and standards. Brandenburg et al. (2014) structured the literature on sustainable supply chain management according to various modeling features. Even though the approach used in their paper for clustering the literature is related to the approach adopted in this paper, the focus of their review was on environmental factors and social aspects of supply chain planning models. Giret et al. (2015) and Gahm et al. (2016) are the reviews that are closest related to the paper at hand. Our work, however, differs from these works as we review and compare modeling options for EEPP in detail, while the reviews of Giret et al. (2015) and Gahm et al. (2016) discussed the sampled papers on a quite aggregated level without analyzing in detail how energy aspects were integrated into traditional scheduling approaches from a modeling point of view. In addition, we suggest a different and more model-driven classification for decision support models for EEPP. Furthermore, in contrast to Giret et al. (2015) and Gahm et al. (2016), we do not solely focus on machine scheduling models, but also review EE master production scheduling and capacity planning approaches as well as EE lot sizing models, which helps to give a more comprehensive overview of recent developments in the field.

The remainder of the paper is organized as follows: Section 2 introduces the review methodology used and descriptively analyzes the search results, while Section 3 proposes a classification scheme for the identified articles. Sections 4 to 6 review how energy efficiency criteria were modeled in the decision support models contained in our sample. Finally, Section 7 concludes the paper, presents managerial insights and suggestions for future research opportunities, and shows how the discussed EEPP models may benefit from each other.

2 Review methodology and descriptive analysis

To completely synthesize the literature on decision support models for EEPP, we conducted a systematic literature review. The methodology employed was adopted from Glock and Hochrein (2011) and Hochrein and Glock (2012). First, we defined two sets of keywords, one related to PP approaches and a second one related to energy efficiency, based on the keywords of a pre-sample of frequently cited articles on EEPP (see Table 1). Each keyword from the first set was combined with each keyword from the second set to generate the final list of keywords that contained 3·6=18 keyword combinations. Subsequently, the scholarly databases Business Source Premier and Scopus were searched for peer-reviewed articles featuring these keyword combinations either in their title, abstract, or list of keywords that were published up to the year of 2015.

Table 1: Overview of keyword sets used in the literature search.

Keywords related to production planning approaches	**Keywords related to energy efficiency**
Production planning	Energy
Inventory	Electricity
Scheduling	Load management
	Sustainable
	Smart grid
	Demand side management

The relevance of the articles identified during the literature search was verified by first reading the paper’s abstract and, in case it was not removed from the sample at this stage, by reading the entire paper. Subsequently, a snowball search was performed by examining all articles that were cited in the papers contained in our sample for relevance. Finally, an inverse search was performed in which we checked all articles that cited one of our sampled papers for relevance.

As energy efficiency, in general, has received a lot of attention in recent years, we defined clear limits to the literature search to keep this paper focused. The focus of the review was on decision support models for medium- and short-term PP problems that aim at minimizing or constraining the energy consumption, energy cost, peak power consumption, or energy-related GHG emissions of manufacturing industries. Hence, we excluded papers that

- focus on the minimization of energy consumption in a production plant without considering the connection between PP decisions and the consumption of energy obtained from external sources. Articles that fall into this area often study plants with in-house power generation units or that make use of cogeneration opportunities such as combined heat and power plants or paper and pulp production plants (e.g., Sarimveis et al., 2003; Waldemarsson et al., 2013). Hence, the decisions prepared by the corresponding models are rather of technological than of managerial nature;
- concentrate on reducing pollution instead of energy consumption in manufacturing (e.g., Yue and You, 2013);

- promote EE manufacturing from the perspective of utilities supplying energy to the manufacturing companies instead of focusing on the perspective of the manufacturing companies (e.g., Zugno et al., 2013);
- suggest (online) production/inventory control strategies (e.g., Cataldo et al., 2015; Fernandez et al., 2013; Sun and Li, 2013) as their model characteristics contrast strongly with the model characteristics of the (offline) a priori decision support models which we focus on in this review;
- do not present quantitative models, but rather general frameworks or software applications (e.g., Küster et al., 2013; Raileanu et al., 2015);
- do not focus on the machine-level, but rather on an (abstract) system-level (e.g., Herrmann and Thiede, 2009).

In total, we identified 89 relevant articles. Figure 2 shows how research on EEPP has evolved over time. The soaring number of articles published in recent years impressively emphasizes the relevance of the topic studied in this paper and the need for a review that structures existing planning approaches from a modeling point of view.

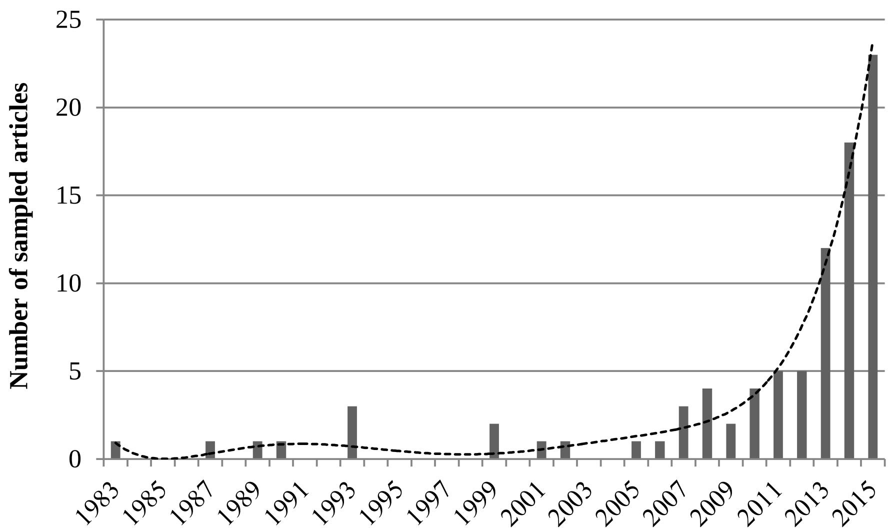


Figure 2: Number of sampled articles on energy-efficient production planning per year.

Figure 3 shows which journals published the highest number of papers contained in our sample. It is striking that 40 percent of the sampled articles were published in only six journals; in total, 43 journals published at least one of the sampled papers.

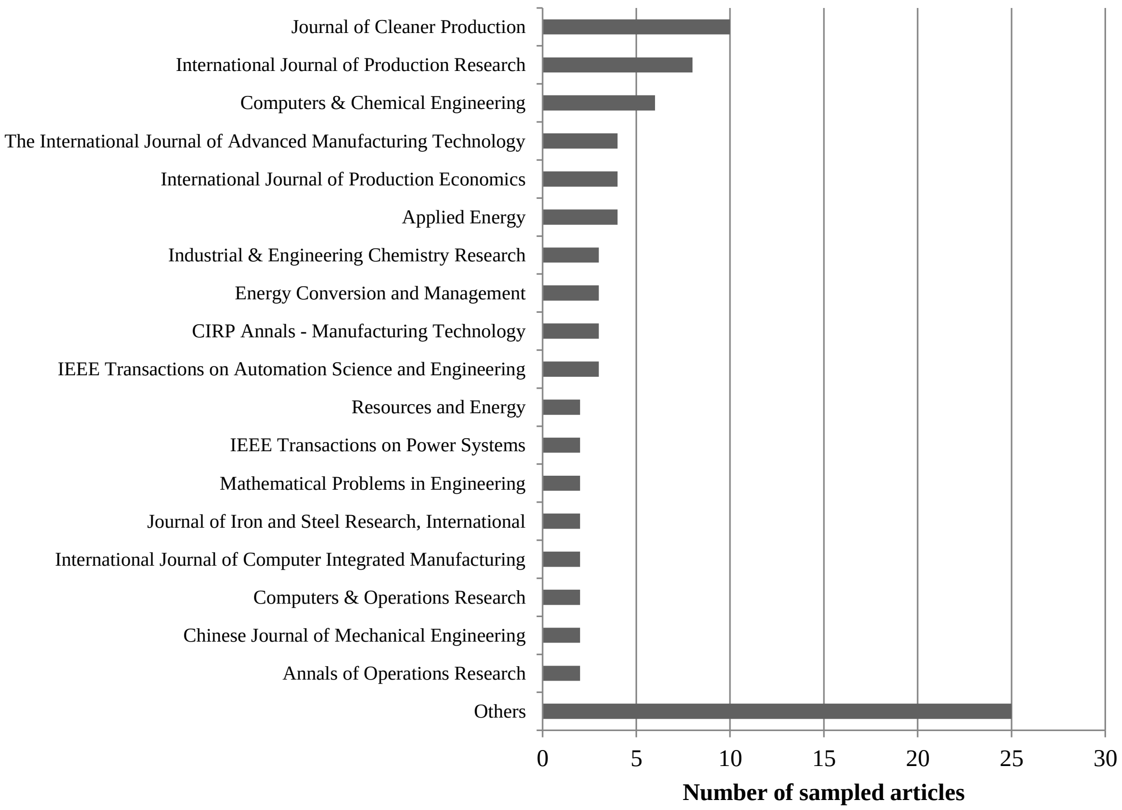


Figure 3: Number of sampled articles on energy-efficient production planning per journal.

To keep the length of the paper within reasonable limits, the intention of this review is not to discuss all 89 sampled papers in detail in the following. Instead, we first suggest an approach for classifying the sampled papers (see Section 3) and then discuss the most important and innovative modifications of traditional PP approaches with a special focus on opportunities for modeling energy efficiency criteria (see Sections 4 to 6).

3 Classification scheme

To align this review with the goals laid out in Section 1, we classify the sampled articles based on the traditional PP problems they belong to. This classification dates back to Hax and Meal (1975), who classified PP tasks according to a hierarchical planning structure. This hierarchical planning structure distinguishes PP problems with respect to their time horizon and aggregation from long-term, aggregate PP to short-term, detailed PP. For this review, we refined this classification scheme by further distinguishing between different EE machine scheduling approaches as explained in Section 6 (see Figure 4).

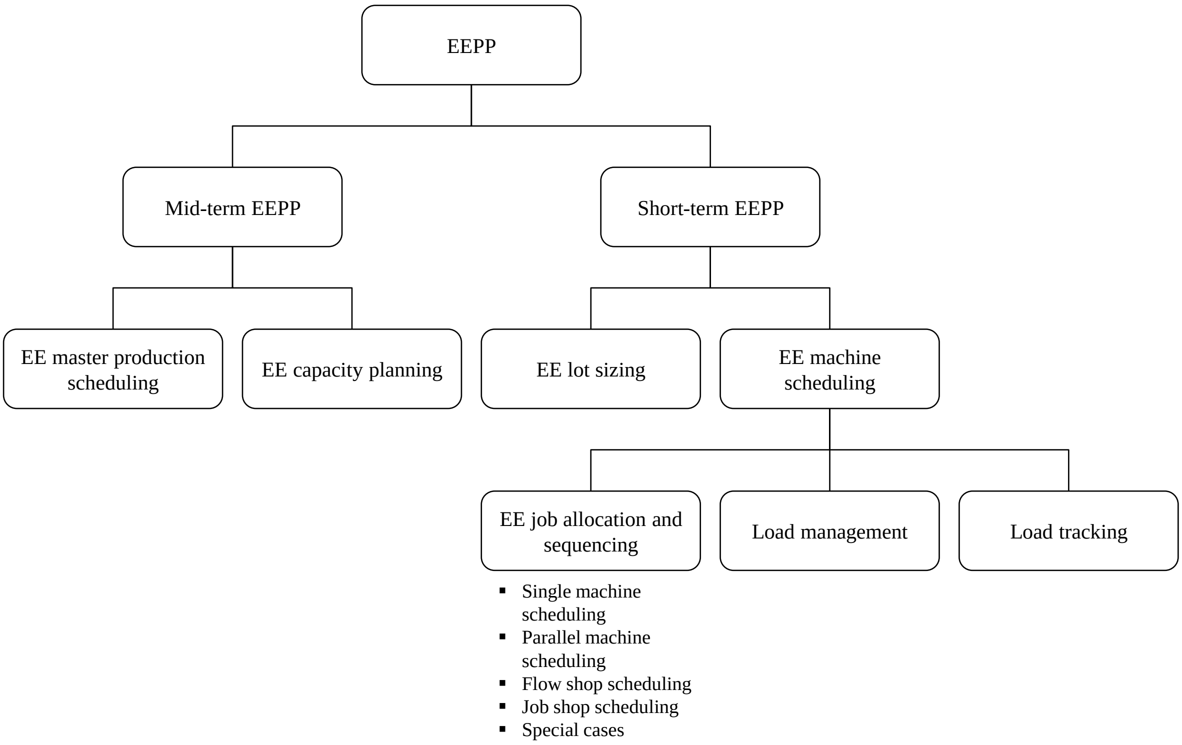


Figure 4: Classification scheme of energy-efficient production planning models.

As the focus of this review is on PP models that foster energy efficiency of existing manufacturing equipment while neglecting long-term investments into new production equipment, only the mid-term and the short-term planning levels are of direct relevance to this paper. Section 4 discusses papers that are concerned with EE master production scheduling and capacity planning. Section 5 then discusses EE lot sizing models, and Section 6 focuses on EE machine scheduling models. As one intention of this review is to improve the comparability of existing decision support models for EEPP, the notations used in the sampled papers are unified and simplified as far as possible in the following. Indices, sets, parameters, and variables will be introduced where first used in the paper. In addition, an exhaustive overview of the notation is provided in Appendix A. When reviewing the decision support models, the focus of our analysis will be on the approach used for modeling energy efficiency criteria. In most cases, these criteria were modeled by introducing an additional objective function, while in some cases constraints were added to the optimization problem. If energy efficiency aims are included in the objective function, we will discuss how energy efficiency can be increased in the given model. Additional objectives that could be relevant, such as the minimization of the total makespan, for example, will not be discussed in detail in this review as they have already been focused on in other literature reviews. For a comprehensive overview of the classification of the sampled articles, the reader is referred to Appendix B.

4 Energy-efficient master production scheduling and capacity planning

As illustrated in Figure 4, master production scheduling and capacity planning are considered as mid-term planning tasks. Master production scheduling and capacity planning determine how given plant locations and an established production system derived from the long-term production plan should best be utilized to achieve the company’s objectives. Hence, master production scheduling is responsible for calculating production quantities, inventory levels, and associated workforce levels to meet the demand in a given season, which is commonly divided into time intervals spanning a week to a month. The objective of master production scheduling usually is to develop an aggregated production plan for the given planning horizon, and not to formulate a detailed machine schedule (Fleischmann et al., 2005).

Denton et al. (1987) were the first to investigate how energy efficiency issues – in this particular case time-varying energy tariffs that feature both energy consumption and power demand charges – impact master production scheduling. The authors considered a scenario where a company produces a single product in an aggregated single production step. The company faces decision problems on three hierarchical levels: (I) each year, the company has to decide whether to produce uniformly over the year or at higher levels in some periods and lower ones in others; (II) each week, the company has to decide whether to operate only on weekdays, or on weekends as well; (III) each day, the company has to decide whether to operate one, two, or three shifts per day, and to which degree capacity should be utilized in each shift. On an annual basis, the developed model minimizes fixed cost, cost of holding capital stock, cost of capital services, cost of materials, cost of labor, cost of holding inventory, and cost of energy. The cost of energy is comprised of the cost of electricity, TC^{elct}, of natural gas, TC^{ntGs}, and of oil, TC^{oil}. Both TC^{elct} and TC^{ntGs} are composed of an energy consumption charge, $TC^{elct,EC}$ and $TC^{ntGs,EC}$, and a power demand charge, $TC^{elct,DC}$ and $TC^{ntGs,DC}$. $TC^{elct,EC}$ is calculated as follows:

$$\begin{aligned} TC^{elct,EC} &= \textstyle\sum_{ssn=1}^{2} 130 \cdot \sum_{shft=1}^{3} \left(C^{elct,base} + C^{elct}_{ssn,shft}\right) \cdot E^{elct}_{ssn,shft} \\ &\quad + \textstyle\sum_{ssn=1}^{2} 26 \cdot \sum_{shft=4}^{9} \left(C^{elct,base} + C^{elct}_{ssn,shft}\right) \cdot E^{elct}_{ssn,shft}. \end{aligned} \tag{1}$$

In Equation (1), one year is equally divided into two seasons, where one season contains 130 workdays during the week and 26 weekends. $C^{elct,base}$ corresponds to an electricity base charge which does not change over time. In contrast, the electricity consumption charge, $C^{elct}_{ssn,shft}$, varies per season ssn and per shift $shft$. $E^{elct}_{ssn,shft}$ corresponds to the consumption of electricity per season and shift. $TC^{elct,DC}$ is calculated as follows:

$$\begin{aligned} TC^{elct,DC} &= \textstyle\sum_{ssn=1}^{2} 6 \cdot C^{elct,PDC}_{ssn} \cdot MD^{elct,PDC}_{ssn} \\ &\quad + \textstyle\sum_{ssn=1}^{2} 6 \cdot C^{elct,ODC}_{ssn} \cdot \max\left\{MD^{elct,ODC}_{ssn} - MD^{elct,PDC}_{ssn}, 0\right\}. \end{aligned} \tag{2}$$

Hence, in each season, a monthly demand charge is due, which depends on the maximum power demand in a peak demand period, $MD^{elct,PDC}_{ssn}$, and the maximum power demand in an off-peak demand period, $MD^{elct,ODC}_{ssn}$. While $TC^{ntGs,EC}$ is calculated in exact conformance with Equation (1), $TC^{ntGs,DC}$ varies neither with the season nor with the demand pattern. In contrast to the price of electricity and natural gas, the price of oil is not assumed to vary over time.

Fethke and Tishler (1989) also investigated the PP problem of a company that considers time-varying energy prices in defining its master production schedule. In contrast to Denton et al. (1987), the authors directly related the energy consumption of a process to its production rate. Unlike Denton et al. (1987) and Fethke and Tishler (1989), Bakhrankova (2009) and Kondili et al. (1993) investigated EEPP in the context of multi-product continuous production processes. Bakhrankova (2009), on the one hand, focused on better capacity utilization by synchronizing production and distribution planning with respect to energy costs. Kondili et al. (1993), on the other hand, allocated products to be processed to machines and time slots and determined which machine should be left idle in a given time slot and when a change-over on a machine from one product quality to another one should be scheduled. Castro et al. (2009) also studied the master production scheduling problem of a multi-product continuous production facility. However, the key feature of their model is the introduction of a continuous-time formulation which aims at taking account of time-varying energy prices more effectively. The objective here is to produce in time intervals associated with low energy prices as much as possible. Castro et al. (2011) extended this model by adding a penalty cost to the energy-related cost, ERC, to account for violations of the maximum contracted power demand

In contrast to all articles discussed thus far in Section 4, Nilsson and Söderström (1993) studied how different electricity demand functions (EDFs) influence master production scheduling. Therefore, they investigated a production process with an approximately linear EDF, a process with an increased non-linearity in the EDF, and a production process with a non-linear, negative-sloped EDF. To minimize system cost, the authors varied the production flow within given limits with regard to two different energy tariffs that both consider an energy consumption rate and a power demand rate. This paper was extended by Nilsson (1993) by accounting for electricity tariffs with increased and decreased differentiation in addition. In this case, electricity prices during daytime are much higher or only slightly higher compared to nighttime prices. The aim of the model was to find the production schedule with optimal ERC and to study the influence of different EDFs and of the differentiation of the electricity tariff.

Karwan and Keblis (2007) were the first to consider real-time electricity pricing in master production scheduling of an air separation plant and to compare the resulting production schedule to production schedules under a time-of-use (TOU) pricing scheme[2] and another dynamic pricing scheme. Mitra et al. (2012) again investigated energy consumption in master production scheduling and took different machine operating modes (including start-up and shut-down) into account. By determining production levels, operating modes, inventory levels, and product sales, the proposed model minimizes the sum of inventory holding cost, cost of operating mode changes, $OMCH$, and ERC:

$$OMCH = \sum_{plnt=1}^{PLNT} \sum_{om \in OM} \sum_{om' \in OM} C_{om,om',plnt}^{trans} \cdot \sum_{t=1}^{T} ch_{om,om',plnt,t}, \tag{3}$$

$$ERC = \sum_{plnt=1}^{PLNT} \sum_{t=1}^{T} C_{plnt,t}^{E} \cdot \sum_{om \in OM} \sum_{i=1}^{n} pw_{i,om,plnt}^{prc} \cdot z_{i,om,plnt,t}. \tag{4}$$

In Equation (3), $ch_{om,om',plnt,t}$ indicates whether there is a transition from operating mode om to operating mode om' in plant $plnt$ between time interval $t-1$ and time interval t. $C_{om,om',plnt}^{trans}$ represents the corresponding transition cost. In Equation (4), $pw_{i,om,plnt}^{prc}$ represents the correlation between the power consumption and the production level when plant $plnt$ manufactures product i in operating mode om, which is determined using a multivariate regression. $z_{i,om,plnt,t}$ measures the production output of product i in plant $plnt$ in operating mode om in time interval t. Additionally, power consumption is not allowed to exceed the maximum contracted demand, MCD_t, in a given time interval t. The model of Mitra et al. (2012) was extended by Mitra et al. (2014), who developed a multi-scale

[2] TOU tariffs a priori divide a day into periods according to different power demand levels (e.g., off-peak, mid-peak, and on-peak) and attach high energy prices to on-peak periods and lower energy prices to off- and mid-peak periods. Utilities use TOU tariffs to induce their customers to smooth their power demand across the day (Zhang et al., 2014).

capacity planning model under demand uncertainty that integrates both strategic and operational planning.

Latifoğlu et al. (2013) proposed the first decision support model for companies using interruptible load contracts (ILCs). Under such contracts, utilities are allowed to schedule a maximum of K power interruptions in the planning horizon to avoid demand peaks. As the timing of the load interruptions is usually uncertain in advance, Latifoğlu et al. (2013) developed a robust planning approach that minimizes ERC under all possible interruption scenarios without compromising demand fulfillment.

Choi and Xirouchakis (2014) were the first to study EEPP for a multi-product flexible manufacturing system. The authors proposed a new methodology for estimating energy consumption and material flows for different production plans for a part type. By determining the production quantity of a part type in a given period and the corresponding process plan, the developed model minimizes the weighted sum of energy consumption, inventory holding cost, and backorder cost. This approach was extended by Choi and Xirouchakis (2015), who proposed a three-tier holistic PP approach. On the first planning level, this approach identifies technological instructions describing how to produce a product. On the second planning level, it evaluates the consequences of energy consumption and environmental impacts associated with various system reconfigurations. On the third planning level, the approach minimizes energy consumption and inventory holding cost and maximizes the production throughput by determining which production quantities of the part types should be processed according to which production plan.

Very specific EEPP models were presented by Schulte Beerbühl et al. (2015) and Kopanos et al. (2015). While Schulte Beerbühl et al. (2015) investigated the master production scheduling and capacity planning problem of a flexible electricity-to-hydrogen-to-ammonia plant considering energy cost, Kopanos et al. (2015) studied EEPP in an air separation plant with a special focus on a network of parallel compressors.

5 Energy-efficient lot sizing

As illustrated in Figure 4, lot sizing is a short-term PP task. The basic objective of this planning task is to determine a production lot size in a batch production process that balances setup or ordering cost and inventory carrying cost (Fleischmann et al., 2005). The paper of Collier and Ornek (1983) was probably the first work that integrated energy efficiency considerations into a PP problem. The authors developed a (Q, r)-inventory control model for a serial multi-stage production system. The model determines the lot size, Q, and the reorder point, r, such that the sum of inventory holding cost of finished and unfinished products and cost of energy used for setting up the machines is minimized.

In contrast, Özdamar and Birbil (1999) considered energy cost in the context of the multi-item capacitated lot sizing and loading problem. In the scenario studied in their paper, two decision problems arise: First, the family lot size of the different products needs to be determined. Secondly, to reduce energy cost, the number of active kilns used for processing the products has to be minimized. Similarly, Yildirim and Nezami (2014) studied an integrated production and preventive maintenance planning problem in a multi-product, multi-period, single machine manufacturing environment. Since product processing times are affected by machine degradation, preventive maintenance and repair of degraded machines can shorten job processing times and consequently lower energy consumption. The authors developed a decision support model for the simultaneous planning of preventive maintenance and PP that considers energy efficiency. The model minimizes the overall cost of production, inventory, maintenance, repair, and ERC by determining product outputs in a given period and scheduling preventive maintenances and setups:

$$ERC = \sum_{i=1}^{n} \sum_{t=1}^{T} z_{i,t} \cdot PW^{prc} \cdot l_{i,t}^{prc} \cdot C_t^E. \tag{5}$$

Here, $l_{i,t}^{prc}$ is the processing time of product i in time interval t. Without preventive maintenance at the end of a time interval, $l_{i,t}^{prc}$ increases by a constant percentage α^{NPM} in the following time interval. Accordingly, the total energy consumption and ERC increase. In contrast, if preventive maintenance measures are adopted, $l_{i,t}^{prc}$ will increase by a smaller percentage $\alpha^{PM} < \alpha^{NPM}$. Yet, the initial $l_{i,t}^{prc}$ cannot be restored.

Zanoni et al. (2014) were the first to integrate different operating modes of machines associated with different levels of energy consumption into lot sizing. They studied a two-stage production system with controllable production rates and intermediate buffer stocks. The authors investigated two production strategies: A continuous production strategy, where both machines produce all batches without interruption, and an interrupted production strategy, where the first machine is switched into an idle operating mode after processing a batch and restarted just in time to avoid stockouts at the second machine. In addition to the lot size, Q, the authors optimized the number of batches shipped between the two production stages, sh, as well as the production rates of the production stages, p_1 and p_2, that affect the overall energy consumption of the production process. The model minimizes inventory holding cost, setups cost, shipment cost, and ERC (idle, switching, and production costs).

In addition to production energy consumption, Bazan et al. (2015) took account of CO_2 emissions from the production and transportation operations in two single-vendor (manufacturer) single-buyer production-inventory systems with applicable multi-level emission taxes and energy usage required for production. These two models minimize the total supply chain cost consisting of the manufacturer's and retailer's inventory holding costs, the manufacturer's setup cost and the retailer's fixed ordering cost, transportation cost, the cost of CO_2 emissions from production, $EMPC$, and from transportation, $EMTC$, a penalty cost from emissions, $PENC$, and ERC:

$$EMPC = \left(\left(\epsilon_1^{CO_2} \right)^2 \cdot p_1 - \epsilon_2^{CO_2} \cdot p_1 - \epsilon_3^{CO_2} \right) \cdot d \cdot C^{CO_2}, \tag{6}$$

$$EMTC = trc \cdot \frac{d}{Q} \cdot FV \cdot \epsilon^{CO_2,F} \cdot C^{CO_2}, \tag{7}$$

$$PENC = \sum_{emLmt=1}^{EmLmt} \Xi_{emLmt} \cdot C_{emLmt}^{CO_2}, \tag{8}$$

$$ERC = \underbrace{\left(e_0^{\prime,prc} + \frac{E_1^{\prime,prc}}{p_1} \right)}_{\substack{Specific\ energy \\ consumed\ [kWh/unit]}} \cdot d \cdot C^E \ . \tag{9}$$

In Equation (6), $\epsilon_1^{CO_2}$, $\epsilon_2^{CO_2}$, and $\epsilon_3^{CO_2}$ are emission function parameters. In Equation (7), FV corresponds to the fuel volume required per truck per trip, and $\epsilon^{CO_2,F}$ denotes the amount of CO_2 emissions from fuel per liter consumed. In Equation (8), $emLmt$ represents an index of emission limits with $emLmt \in \{1,2, \dots, EmLmt\}$, where the penalties for exceeding emission limits increase with each emission limit, and Ξ_{emLmt} indicates whether emissions of a company exceed the emission limit

$emLmt$. In Equation (9), $e_0^{\prime,prc}$ corresponds to the specific energy usage coefficient depending on work-piece material, tool geometrics, and spindle drive characteristics, and $E_1^{\prime,prc}$ corresponds to the absolute energy usage coefficient depending on the machine tool. The developed models derive the optimal values for the manufacturer's production rate (within given limits), the manufacturer-retailer coordination multiplier (number of shipments of size Q in a manufacturer's cycle), the number of trucks of a given capacity per shipment, trc, and the number of items shipped per batch.

6 Energy-efficient machine scheduling

Machine scheduling is the second short-term PP activity considered in this review (see Figure 4). Works that fall into this stream of research are divided into two categories in the following: Section 6.1 reviews works that extended existing machine scheduling models to take account both of traditional scheduling objectives and energy efficiency considerations. Sections 6.2 and 6.3 then discuss decision support models that try to achieve energy-related goals after the machine schedule has been established.

6.1 Energy-efficient job allocation and sequencing

Machine scheduling is a short-term PP task that allocates jobs to machines and that determines the sequence of jobs allocated to the same machine. In most scheduling problems, a finite set of n jobs $J = \{J_i\}_{i=1}^{n}$ needs to be allocated to and sequenced on a finite set of m machines $M = \{M_k\}_{k=1}^{m}$. Each job consists of a finite set of ϕ_i ordered operations $O_i = \{O_{i,g,k}\}$, where $O_{i,g,k}$ is the gth operation of job J_i processed on machine M_k. While $l_{i,g,k}^{prc}$ corresponds to the processing time of $O_{i,g,k}$, $s_{i,g,k}$ represents the start time of $O_{i,g,k}$, and $c_{i,g,k}$ is the completion time of $O_{i,g,k}$. If a job only requires a single operation or a single machine production system is investigated, the presented sets and variables are simplified accordingly (Liu et al., 2014c).

6.1.1 Single machine scheduling

Mouzon et al. (2007) studied the single machine scheduling problem where n jobs have to be processed in the order of their arrival, and minimized both total completion time, TCT, and total energy consumption, TEC. First, they developed dispatching rules which determine whether to leave the machine idle, or to turn it off between jobs. These rules are based on the concept of break-even duration. The break-even duration, l^{BE}, represents the 'least amount of duration required for a turn off/turn on operation (i.e., time required for a setup, l^{stp}) and the amount of time for which a turn off/turn on operation is logical instead of running the machine at idle' (Mouzon and Yildirim, 2008):

$$l^{BE} = \max\left\{\frac{PW^{stp} \cdot l^{stp}}{PW^{idle}}, l^{stp}\right\}. \tag{10}$$

TEC is calculated as

$$TEC = \sum_{i=1}^{n-1}\left(c_{i+1} - l_{i+1}^{prc} - c_i\right) \cdot PW^{idle} - \sum_{i=1}^{n-1} \rho_i, \tag{11}$$

where $\rho_i = \begin{cases} \left(c_{i+1} - l_{i+1}^{prc} - c_i\right)PW^{idle} - PW^{stp}l^{stp}, if\ c_{i+1} - l_{i+1}^{prc} - c_i > l^{BE} \\ 0, otherwise \end{cases}, \forall i \in \{1,2,\dots,n-1\}.$

A very similar model was developed by Liu et al. (2014b). Yet, instead of minimizing TEC, the model minimizes total CO_2 emissions, $TCDE$, and assumes that CO_2 emissions can be related to the energy consumption by a constant factor ϵ^{CO_2}.

Mouzon and Yildirim (2008) generalized the model of Mouzon et al. (2007) by taking account of jobs with unequal release dates that do not necessarily need to be scheduled in the order of their arrival. Thus, the developed model not only determines the timing and length of turn off/turn on periods, but also sequences the jobs in such a way that both the total tardiness, $TTRD$, and TEC are minimized. The same problem was also investigated by Yildirim and Mouzon (2012), whose objective was to minimize TCT instead of $TTRD$ along with TEC. Liu (2014) and Liu and Huang (2014), in contrast, both integrated CO_2 emission considerations into production scheduling of a single batch machine in the presence of dynamic job arrivals and assumed that jobs have unequal release times, due dates, processing times, and priorities. The developed models minimize the total weighted tardiness, TWT, and $TCDE$ by determining whether to start processing a batch or to wait for more jobs to arrive in a given time interval. Thus, the models try to balance the negative effects of leaving a machine idle and the negative effects of operating an underutilized machine. Shrouf et al. (2014) investigated the single machine production scheduling problem where different operating modes, OM, (processing, idle, off) with different energy consumption levels exist. The proposed model minimizes TEC under the constraint that all jobs have to be processed in the given time horizon. Decision variables are the selection of the operating mode, $I_{om,t}$, and the transitions between operating modes, $ch_{om,om',t}$:

$$TEC = \sum_{t=1}^{T} C_t^E \cdot \left(\sum_{om\in OM} E_{om} \cdot I_{om,t} + \sum_{om\in OM} \sum_{om'\in OM} E_{om,om'}^{ch} \cdot ch_{om,om',t}\right). \tag{12}$$

Liu (2015) presented the first model that integrated renewable energy sources (RES) with an uncertain energy supply into a single machine scheduling problem. The energy supplied by the RES was assumed to be stored in a rechargeable battery. Supply uncertainty for the RES was modeled using interval numbers. To reduce CO_2 emissions, the machine does not use power from the electricity grid until the battery runs out of power. The author considered two cases: In the first case, the model simultaneously minimizes time-weighted flow time, $TWFT$, and $TCDE$ by assigning job J_i to time interval t with the help of the binary variable $b_{i,t}$:

$$TCDE = \epsilon^{CO_2} \cdot \sum_{t=1}^{T} \sum_{i=1}^{n} PW_i^{prc} \cdot l_i^{prc} \cdot b_{i,t} - (BatState_t + GQTY_t). \tag{13}$$

Here, $GQTY_t$ is an interval number that measures the amount of energy generated by RES in time interval t. $BatState_t$ is the state of the battery at the beginning of time interval t and also an interval number due to its dependence on $GQTY_t$. In the first case considered in the paper, the main objective is to explore how $TWFT$ and $TCDE$ are affected by the amount of power generated from RES in each time interval. In contrast, in the second case, the model solely minimizes $TWFT$ subject to a periodic as well as a rolling CO_2 emission constraint:

$$\sum_{\tau=t-RH+1}^{t} \epsilon^{CO_2} \cdot \sum_{i=1}^{n} PW_i^{prc} \cdot l_i^{prc} \cdot b_{i,\tau} - (BatState_\tau + GQTY_\tau) \le \mathrm{E}^{CO_2,\max} \cdot RH, \quad \forall t \in \{RH, RH+1 \ldots, T\}. \tag{14}$$

RH is the number of time intervals belonging to the rolling horizon and $E^{CO_2,\max}$ represents the total CO_2 emissions quota for a single time interval. In addition to evaluating the effect of the power generated from RES in each time interval on $TWFT$, this model gives some indication of how regulations on CO_2 emissions impact $TWFT$.

6.1.2 Parallel machine scheduling

Parallel machine scheduling may be considered as a generalization of the single machine scheduling problem and a special case of the flexible flow shop problem (see Section 6.1.3.2). Compared to the single machine scheduling problem, the allocation of jobs to machines is added as a new decision dimension – in addition to job sequencing – in the parallel machine scheduling problem (Pinedo, 2012). There are different types of production systems with machines in parallel. In the simplest case, the parallel machines are identical. In this case, the processing time of a job J_i is independent of the machines it is allocated to. However, in production systems with heterogeneous machines in parallel with different production rates, the processing time of a job J_i depends on the machine it is allocated to. In case of production systems with unrelated machines in parallel, the processing time of a job J_i does not only depend on the machine, but rather on the combination of machine and job. Thus, job J_i may be processed faster on machine M_k than on machine $M_{k'}$. Yet, job $J_{i'}$ may be processed faster on machine $M_{k'}$ than on machine M_k (Pinedo, 2012). Accordingly, the energy consumed to process jobs may differ from machine to machine.

Boukas et al. (1990) proposed the first parallel machine scheduling problem identified in this systematic literature review that takes energy efficiency considerations into account. The authors studied a production scheduling problem in a steel plant comprising four arc furnaces and three continuous casting machines. The developed model minimizes the total time needed to perform a given number of cycles for each of the four furnaces while complying with global power constraints and availability constraints of shared equipment. The model determines the starting times of each operation in each production cycle for each furnace as well as the power assigned to a furnace during a given cycle.

He et al. (2005) and He et al. (2008) both studied the parallel machine scheduling problem with energy consumption for the special case where all jobs are available at the beginning of the planning horizon. He and Liu (2010) extended these models by taking account of environmental concerns that prevent a job from being processed on a specific machine. Liang et al. (2015) extended the single machine scheduling model of Mouzon and Yildirim (2008) to the unrelated parallel machine scheduling problem. Unlike He et al. (2005), He et al. (2008), and He and Liu (2010), Liang et al. (2015) did not assume that all job release dates equal zero. Hait and Artigues (2011a), in contrast, investigated the production scheduling problem in a foundry as part of a pipe-manufacturing plant with respect to energy cost and operator unavailability. By determining the allocations of jobs to furnaces, the precedence relationships among different jobs, starting times of loading, melting, holding and unloading as well as break times for the operators, the model minimizes the sum of the energy cost incurred when holding a job in a furnace after production and the cost incurred when exceeding the maximum contracted power demand, MCD:

$$ERC = \sum_{t=1}^{T}\left(C^{E} \cdot PW^{hold} \cdot \sum_{i=1}^{n} l_{i,t}^{hold} + C^{PEN} \cdot \left(\frac{\sum_{i=1}^{n}\left(PW_i^{prc} \cdot l_{i,t}^{prc} + PW^{hold} \cdot l_{i,t}^{hold}\right)}{l_t} - MCD\right)\right). \quad (15)$$

Here, PW^{hold} is the power consumed when holding a job in a furnace after production, $l_{i,t}^{hold}$ is the time span job J_i is being held in a furnace after production in time interval t, and C^{PEN} corresponds to the penalty cost for violating MCD. Artigues et al. (2013) extended this model by taking penalties for due date violations into account. Similarly, Tang et al. (2015) investigated the reheat furnace scheduling problem, which is a special case of the parallel machine scheduling problem that assigns slabs to a reheat furnace, sequences the slabs for each furnace, and determines the feed-in time and the residence time for each slab in order to reduce unnecessary energy consumption. The model minimizes the weighted sum of the penalties for the deviation of the actual resident time from the desirable resident time, the changeover costs incurred by the differences of heating manners between adjacent slabs, and the penalties for inefficiently heating slabs.

In contrast to the models discussed so far, Moon et al. (2013) considered electricity prices that vary across peak-load, mid-load, and off-peak periods in the unrelated parallel machine scheduling problem subject to predefined processing times and due dates. In this model, machines may be put into an idle state that does not cause any electricity cost. By determining whether job J_i should be processed on machine M_k in time interval t, expressed by the binary variable $b_{i,k,t}$, the model minimizes the makespan, $MKSP$, multiplied by a penalty cost factor and ERC:

$$ERC = \sum_{k=1}^{m} \sum_{i=1}^{n} \sum_{t=1}^{T} b_{i,k,t} \cdot C_t^E \cdot PW_k^{prc} \cdot l_t, \tag{16}$$

where l_t corresponds to the length of time interval t. Similarly, Ding et al. (2015) studied the unrelated parallel machine scheduling problem under a TOU pricing scheme. In contrast to Moon et al. (2013), they only minimized ERC, while $MKSP$ needed to comply with a predetermined production deadline in a first step. Subsequently, Ding et al. (2015) modified their model to simultaneously minimize ERC and $MKSP$. Overall, the key feature of their work consists in the computational improvement of the introduction of time-varying energy prices into the machine scheduling problem which traditionally is time-independent. Liu (2014), which was already discussed in Section 6.1.1, also extended their batching problem to a production system with parallel machines.

Rager et al. (2015) based their research on the fact that final energy sources (FES) are not directly consumed by production resources. They are rather transformed by conversion units into applied energy sources (AES), such as steam or pressure. Thus, to enhance the energy efficiency of a production system, it is important to understand how FES and AES are related. The authors integrated this concept into an EE parallel machine scheduling problem. The amount of AES consumed by the parallel machines was assumed to depend both on the order of jobs on a machine and the processing time. The authors hypothesized that frequent load alternations result in inefficient conversion processes. Thus, they tried to minimize the demand for FES by minimizing the variance of demand for AES, $VarAESD$:

$$VarAESD = \sum_{t=1}^{T} \left(\sum_{i=1}^{n} \sum_{g=1}^{\phi_i} \sum_{k=1}^{m} b_{i,g,k,t} \cdot E_{i,g}^{AESD} - \bar{E}^{AESD} \right)^2. \tag{17}$$

Here, $b_{i,g,k,t}$ indicates whether operation $o_{i,g}$ is processed on machine k in time interval t, $E_{i,g}^{AESD}$ represents the demand for AES of operation $o_{i,g}$, and $\bar{E}^{AESD}$ represents the desired level of demand for AES.

Wang et al. (2015) went one step further than the approaches presented thus far in this section as they connected process planning and scheduling optimization to enhance sustainability in milling processes by taking account of energy consumption when machines are idle or need to be set up or when a tool needs to be changed. In the first step of a proposed two-step approach, the authors optimized the key milling parameters of individual machines to foster energy efficiency, surface quality, and productivity. Based on the optimized milling process parameters, the authors then developed a model for setting up the machines and scheduling multiple parts on the machines to minimize both $MKSP$ and TEC in a second step while accounting for different machine operating modes.

6.1.3 Flow shop scheduling

A flow shop (FS) system comprises two or more production stages in series, where each production stage consists of at least one machine. In a FS, each job J_i has to be processed on each machine M_k in the same machine sequence (Fang et al., 2011).

6.1.3.1 Basic flow shops

Fang et al. (2011) simultaneously minimized $MKSP$, $TCDE$, and the peak power consumption, PPC, of a machining FS with unlimited storage space between two successive machines using a multi-objective algorithm. The authors explicitly included PPC as the peak power consumption often significantly contributes to a company's cost of electricity:

$$PPC = \max\left\{ \begin{array}{c} \underbrace{\sum_{k'=1,k'\neq k}^{m}\sum_{i'=1}^{n}\sum_{h'=1}^{n}\sum_{v\in V} PW_{i',k',v}^{prc}\cdot \xi_{i',h',k',v,h,k}}_{Total\ power\ consumption\ of\ all\ machines\ besides\ machine\ k} \\ +\underbrace{\sum_{i=1}^{n}\sum_{v\in V} PW_{i,k,v}^{prc}\cdot b_{i,h,k,v}}_{\substack{Total\ power\ consumption \\ of\ the\ hth\ job\ on\ machine\ k}} \mid h\in\{1,2,\dots,n\}, k\in\{1,2,\dots,m\} \end{array} \right\}, \tag{18}$$

where $\xi_{i',h',k',v,h,k}$ indicates whether job $J_{i'}$ is the h'th job processed on machine $M_{k'}$ with speed v, which starts while the hth job is processed on machine M_k. $PW_{i',k',v}^{prc}$ denotes the power consumed when processing job $J_{i'}$ on machine $M_{k'}$ with speed v, and $b_{i,h,k,v}$ indicates whether job J_i is the hth job processed on machine M_k with speed v. $TCDE$ is calculated as

$$TCDE = \epsilon^{CO_2}\cdot\sum_{k=1}^{m}\sum_{v\in V}\sum_{i=1}^{n}\sum_{h=1}^{n} l_{i,k,v}^{prc}\cdot PW_{i,k,v}^{prc}\cdot b_{i,h,k,v}, \tag{19}$$

where $l_{i,k,v}^{prc}$ corresponds to the time required to process job J_i on machine M_k with speed v.

Zhang et al. (2014) neglected traditional scheduling objectives completely and only ensured a given production throughput using a constraint. In the objective function, they solely minimized ERC and $TCDE$ considering TOU tariffs. While ERC is calculated as in Moon et al. (2013), it is striking that the amount of CO_2 emissions per kWh, $\epsilon_t^{CO_2}$, is assumed to vary over time in the calculation of $TCDE$:

$$TCDE = \sum_{k=1}^{m}\sum_{i=1}^{n}\sum_{t=1}^{T}\epsilon_t^{CO_2}\cdot PW_{k,t}^{prc}\cdot b_{i,k,t}. \tag{20}$$

Thus, the authors took account of the fact that the objective of minimizing total electricity cost in the presence of a TOU tariff leads to a shift of energy consumption from on-peak periods to off-peak periods. In contrast, minimizing the carbon footprint leads to a shift in the opposite direction as the amount of CO_2 emissions per kWh is lower during on-peak hours than during the other periods. This is due to the fact that the base electricity load is often supplied by coal-based sources, and natural gas is used to cope with peak demands.

Sharma et al. (2015) investigated a two-stage assembly FS scheduling problem considering a TOU tariff with electricity and demand charges. Similar to Zhang et al. (2014), Sharma et al. (2015) accounted for the fact that minimizing energy cost and minimizing CO_2 emissions does not necessarily lead to the same production schedule. Depending on the energy procurement tariffs, the objectives may even be conflicting (Sharma et al., 2015). Consequently, the authors formulated (I) an economical, (II) an ecological, and (III) an econological objective. The economical objective (I) is to maximize the number of jobs in a shift and to minimize ERC, while the ecological objective (II) solely minimizes $TCDE$. The econological objective (III) combines the economical and the ecological objectives. To achieve objectives (I) to (III), the developed model determines the operating speed for job J_i on machine M_k as well as the maximum number of changes in operating speed per shift.

Similar to Wang et al.'s (2015) approach for the parallel machine scheduling problem (see Section 6.1.2), Lin et al. (2015) integrated processing parameter optimization and EE production scheduling for a discrete manufacturing flow shop. Zhang et al. (2015a) added a completely new aspect to the EE production scheduling literature by studying scheduling of multiple manufacturing factories under real-time electricity pricing. The authors assumed that there are FCT factories, with each factory having n_{fct}^{FS} identical FSs. Power loads can either be classified as non-shiftable loads ($PW_{fct,t}^{nShft}$), e.g., the indispensable lighting load, or as shiftable loads ($PW_{k,fct}^{prc}$), e.g., the loads of processing machines. The authors assumed that the real-time electricity price, $C^E(PW_t)$, exponentially depends on the current power demand, PW_t. Yet, from the perspective of the utilities, the current power demand comprises not only the power demand of the factories, PW_t^{mnfc}, but also the power demands of the residential buildings, PW^{rsdt}, and the commercial buildings, PW^{cmmc}. In their further analysis, Zhang et al. (2015a) considered a scenario in which the factories collaborate and a scenario in which they do not collaborate. In the collaboration scenario, the developed time-indexed integer programming model minimizes ERC of all FCT factories, given that all factories share their schedules with each other:

$$ERC = \sum_{t=1}^{T} C^E \left(\begin{array}{c} PW^{rsdt} + PW^{cmmc} \\ + \sum_{fct=1}^{FCT} n_{fct}^{FS} \cdot \left(\sum_{k=1}^{m_{fct}} PW_{k,fct}^{prc} \cdot b_{k,fct,t} + PW_{fct,t}^{nShft} \right) \end{array} \right) \cdot \sum_{fct=1}^{FCT} n_{fct}^{FS} \cdot \left(\sum_{k=1}^{m_{fct}} PW_{k,fct}^{prc} \cdot b_{k,fct,t} + PW_{fct,t}^{nShft} \right). \tag{21}$$

Here, $b_{k,fct,t}$ indicates whether machine M_k in factory fct is processing in time interval t. In the non-collaboration case, each factory ψ minimizes its individual ERC without power demand information from other factories and hence without information concerning the associated electricity price:

$$ERC = \sum_{t=1}^{T} C^E \left(\begin{array}{c} PW^{rsdt} + PW^{cmmc} \\ + \sum_{fct=1}^{FCT} n_{fct}^{FS} \cdot \left(\sum_{k=1}^{m_{fct}} PW_{k,fct}^{prc} \cdot b_{k,fct,t} + PW_{fct,t}^{nShft} \right) \end{array} \right) \cdot n_{\psi}^{FS} \cdot \left(\sum_{k=1}^{m_{\psi}} PW_{k,\psi}^{prc} \cdot b_{k,\psi,t} + PW_{\psi,t}^{nShft} \right). \tag{22}$$

6.1.3.2 Flexible flow shops

A flexible flow shop (FFS) is a generalization of the classical FS and of production systems with parallel machines as explained in Section 6.1.2 (Pinedo, 2012). In several publications, FFSs are also referred to as hybrid flow shops (e.g., Choi and Wang, 2012; Ribas et al., 2010; Voß and Witt, 2007). A FFS comprises two or more production stages in series, where each production stage consists of at least one machine, and where at least one production stage contains more than one machine. All jobs have to be processed in the same order on every production stage (Pinedo, 2012).

Bruzzone et al. (2012) investigated a FFS model of a machining process considering MCD. The authors assumed that a feasible production schedule has already been computed without accounting for the maximum power demand. The developed model adjusts this schedule by minimizing $MKSP$ and $TTRD$ while complying with MCD constraints and without altering the assignments of the jobs to the machines and the job sequences. Hence, only the starting times of the jobs are determined by the model subject to the following constraints:

$$\sum_{k=1}^{m} \sum_{i=1}^{n} \sum_{g=1}^{\phi_i} PW_{i,g,k}^{prc} \cdot b_{i,g,k,t} \leq MCD, \forall t \in \{1,2,\dots,T\}. \tag{23}$$

Here, $b_{i,g,k,t}$ indicates whether the gth operation of job J_i is processed on machine M_k in time interval t. A different approach was advocated by Dai et al. (2013), who integrated energy efficiency considerations not only as a constraint into a FFS scheduling problem. Instead – besides the minimization of $MKSP$ – they minimized TEC by assigning jobs to machines at the corresponding stages, by determining the sequence of operations on each machine, and by deciding on whether or not to turn off a machine when idle based on the concept of break-even duration suggested by Mouzon et al. (2007) and Mouzon and Yildirim (2008). In contrast, Luo et al. (2013) even accounted for time-varying energy prices and minimized ERC of a FFS scheduling problem along with $MKSP$.

Considering the same objectives, Tan et al. (2013) investigated the steelmaking process scheduling problem, which is a special case of the FFS scheduling problem that takes account of the specific requirements of steel production processes subject to variable electricity prices. Tan and Liu (2014) extended this article by developing a new solution methodology. Liu and Huang (2014), which was already discussed in Section 6.1.1, also addressed scheduling in the iron and steel industry and proposed a two-stage FFS consisting of a batch processing machine, followed by two parallel processing machines. All machines were assumed to have infinite buffers. Yet, in contrast to Tan et al. (2013), the authors considered PPC and $TCDE$ as energy efficiency criteria along with the traditional scheduling objective of minimizing TWT.

Castro et al. (2013) made use of the resource-task network discrete-time formulation developed by Castro et al. (2009) and Castro et al. (2011) to investigate the production scheduling problem in a melt shop of a steel plant considering different energy constraints. By assigning tasks to time intervals, the developed model minimizes both $MKSP$ and ERC under time-varying energy prices.

6.1.3.3 Permutation flow shops

A permutation flow shop (PFS) is a special case of a FS where jobs waiting to be processed on a machine are processed in the order of their arrival (FIFO) (Pinedo, 2012). Fang et al. (2013) integrated energy efficiency criteria in the form of a PPC constraint into a PFS scheduling problem that minimizes $MKSP$. PPC was formulated as in Fang et al. (2011) (see Equation (18)). Instead of minimizing $MKSP$, Liu et al. (2013) minimized TEC, which was formulated as a weighted sum of the idle times of each machine in their paper. It was assumed that machines are turned on as soon as the first job arrives and turned off when the last job has been processed. Hence, machines stay in an idle operating mode between the processing of two successive jobs. To minimize TEC, the developed model determines the idle time of machine M_k before starting operation $O_{i,k}$:

$$TEC = \sum_{k=1}^{m} WT_k \cdot \sum_{i=1}^{n} \max\{c_{i,k-1} - c_{i-1,k}, 0\}. \tag{24}$$

Here, WT_k is the weight assigned to machine M_k and reflects its impact on TEC.

6.1.4 Job shop scheduling

This section discusses decision support models for EE production scheduling in job shop (JS) environments. The difference between FSs and JSs is that in a JS jobs are not required to be processed on the same sequence of machines (Pinedo, 2012).

6.1.4.1 Basic job shops

Tang and Dai (2015) studied the classical JS scheduling problem considering energy consumption. Yet, they assumed that the assignment of jobs to machines and the processing order of all jobs have previously been determined. Hence, their model solely adjusts the production rates of the machines to minimize TEC. In contrast, the model of Liu et al. (2014c) does not assume a given production schedule, but rather determines starting times of machine operations and defines precedence relations for required job operations to minimize TWT and TEC. May et al. (2015) also considered energy consumption in the context of the classical JS scheduling problem. Yet, in contrast to Liu et al. (2014c), they took account of several machine operating modes (off, standby, idle, setup, processing) and minimized both $MKSP$ and TEC. Zhang et al. (2015b), who studied a JS scheduling problem in the process industry, also considered energy consumption for operating modes other than processing, setup, and idle. The authors used the term *Energy for Ready-Open-Close* to refer to energy consumed in the preparation and follow-up work before and after product processing and minimized TEC. Duerden et al. (2015) studied a JS scheduling problem as well. However, similar to the approach of Rager et al. (2015), the developed model minimizes the variance of energy consumption over time.

In contrast to the approaches discussed thus far in this section, Liu et al. (2015) investigated a classical JS scheduling problem of a manufacturing company subject to a so-called Rolling Blackout Policy (RBP). A RBP, initiated by the government, regularly interrupts the electricity supply of industrial customers for several days. As a consequence, the affected companies use private diesel generators to avoid breakdowns in production, which results in higher CO_2 emissions and electricity cost. The developed model minimizes TWT, TEC, and electricity cost for the case when a RBP is in effect, $RBPC$. The negative effects of the RBP are expressed as higher electricity cost during that period:

$$RBPC = \sum_{k=1}^{m} C^E \cdot \int_0^{\max\left\{c_{i,g,k} | i \in \{1,2,\ldots,n\}, g \in \{1,2,\ldots,\phi_i\}\right\}} PW_k^{prc} dt \;, \tag{25}$$

where the value of the electricity price, C^E, depends on whether the RBP is inactive or active. *TEC* is calculated similar to Liu et al. (2014c).

6.1.4.2 Flexible job shops

The flexible job shop (FJS) scheduling problem is a generalization of the classical JS scheduling problem. While in a classical JS, the sequence in which each job visits the machines is predetermined, it is part of the FJS scheduling problem to determine this sequence. Hence, each operation required by a job needs to be assigned to one machine out of a set of machines that are able to perform this operation. The result is a combined routing and scheduling problem (Brandimarte, 1993).

Moon and Park (2014) studied two FJS scenarios considering electricity cost. In the first scenario, the authors took account of time-dependent and machine-dependent electricity cost. By assigning operations to machines, determining the order of the assigned operations on each machine and by inserting idle periods, the model minimizes the sum of production overtime cost and ERC:

$$ERC = \sum_{k=1}^{m} \sum_{i=1}^{n} \sum_{\gamma=1}^{\mathrm{P}_k} \sum_{g=1}^{\phi_i} \sum_{t=s_{k,\gamma}}^{s_{k,\gamma}+l_{i,g,k}^{prc}-1} b_{i,g,k,\gamma} \cdot C_t^E \cdot PW_k^{prc}, \tag{26}$$

where $b_{i,g,k,\gamma}$ indicates whether the gth operation of job J_i is processed on machine M_k at priority γ out of the P_k priorities of machine M_k. In the second scenario, the authors additionally incorporated distributed energy resources (DER) and an energy storage system (ESS). Again, the model minimizes the sum of production costs related to $MKSP$ and ERC, which include the costs of DER and ESS. To solve this model, the authors developed a hybrid production and energy scheduling algorithm. This algorithm divides the model into two sub-models: (I) production scheduling (as in Scenario 1) with a given energy schedule and (II) energy scheduling with a given production schedule. Hence, by determining the amount of energy stored in a given interval and the amount of energy charged to and discharged from the ESS, the second sub-model minimizes the cost of distributed generators and battery operating cost, DGC:

$$\begin{aligned} DGC = {} & \sum_{t=0}^{T-1} (PBat_t \cdot (-C_t^E) + |PBat_t| \cdot CBat) \\ & + \sum_{t=0}^{T-1} \sum_{\zeta=1}^{NG} \big(GQTY_{\zeta,t} \cdot (-C_t^E) + GQTY_{\zeta,t} \cdot GCost_{\zeta,t}\big). \end{aligned} \tag{27}$$

Here, $PBat_t$ corresponds to the amount of energy charged to or discharged from the ESS in time interval t, $CBat$ is the operating cost of the EES, and $GQTY_{\zeta,t}$ represents the amount of energy generated by generator ζ out of the NG generators in time interval t.

In contrast to Moon and Park (2014), Jiang et al. (2014) minimized TEC to foster energy efficiency in a FJS scenario. Hence, they developed a model to simultaneously minimize a cost-weighted processing quality instability index of scheduling, processing cost, $MKSP$, and TEC. He et al. (2015) also minimized $MKSP$ and TEC of a FJS problem. Yet, in contrast to Jiang et al. (2014), they considered energy consumption which can differ from machine to machine during processing, tool switching, setup, and idle periods.

6.1.4.3 Flexible manufacturing systems

Flexible manufacturing systems (FMSs) basically are a further generalization of JSs. They feature high flexibility of resource allocation and part routing. On the one hand, a resource can be used for processing several operations of a job (shared resources), and on the other hand, several resources can process the same operation of a job (choice operations) (Le and Pang, 2013). Le and Pang (2013) studied such a FMS subject to power consumption uncertainties. To enhance the energy efficiency of the system, they developed a dynamic scheduling model to minimize the sum of tardiness penalties and TEC under power consumption uncertainties:

$$TEC = \sum_{k=1}^{m} \sum_{i=1}^{n} \sum_{g=1}^{\phi_i} \sum_{\theta=1}^{\Theta_i} b_{i,g,\theta,k} \cdot PW_{i,g,k}^{prc}(t) \cdot l_{i,g,k}^{prc}, \tag{28}$$

where each job J_i needs to be performed Θ_i times. $b_{i,g,\theta,k}$ indicates whether the gth operation of the θth repetition of job J_i is processed on machine M_k. It is important to note that the power input $PW_{i,g,k}^{prc}(t)$ is time-dependent and uncertain due to machine conditions, tool conditions, and workloads. Similarly, Garcia-Santiago et al. (2015) studied the FJS problem under shared resources. As in May et al. (2015), the authors took account of five machine operating modes (off, standby, idle, setup, processing). While complying with a predetermined production target, they minimized TEC by maximizing the time the machines spend in the standby operating mode. Similar to Le and Pang (2013), Zhang et al. (2013) also investigated dynamic scheduling in a FMS with the objective of minimizing TEC. Yet, the authors assumed that n jobs need to be scheduled a priori, and that n' new jobs arrive after production has been started. The arrivals of jobs as well as machine breakdowns were assumed to occur randomly. By determining whether the gth operation of job J_i is processed on machine M_k, the total time required to process all jobs as well as TEC are minimized. In contrast, Pang and Le (2014) used a weighted p-timed Petri Net to minimize both productive and idle energy consumption in a FMS, subjected to general production constraints.

6.1.5 Special cases

Energy efficiency considerations have also found their way into a few more specialized machine scheduling problems. Nilakantan et al. (2015), for instance, integrated energy efficiency considerations into a robotic assembly line balancing problem with the objective of simultaneously minimizing cycle time and TEC:

$$TEC = \max\left\{\sum_{i=1}^{n}\sum_{k=1}^{PS} E_{i,k}^{prc} \cdot b_{i,ps} \cdot a_{k,ps} \,|ps \in \{1,2,\dots,PS\}\right\}, \tag{29}$$

where $b_{i,ps}$ indicates whether job J_i is processed on stage ps, and $a_{k,ps}$ indicates whether robot k is allocated to stage ps.

Liu et al. (2014a) studied a so-called multi-stage multi-option seru production system considering the system's environmental and economic performance. A seru production system is a new production system in which the entire production process of a job is carried out within a single work cell, termed seru. By determining whether product i is manufactured in seru M_k, expressed by the binary variable $b_{i,k}$, and by deriving the corresponding production quantity, $z_{i,k}$, the developed model minimizes $MKSP$ and $TCDE$:

$$TCDE = \epsilon^{CO_2} \cdot \sum_{i=1}^{n}\sum_{k=1}^{m_i} b_{i,k} \cdot z_{i,k} \cdot e_{i,k}^{prc}, \tag{30}$$

subject to an overall CO_2 emission constraint per product i with $i \in \{1,2,\dots,n\}$:

$$\epsilon^{CO_2} \cdot \sum_{k=1}^{m_i} b_{i,k} \cdot z_{i,k} \cdot e_{i,k}^{prc} \leq \mathrm{E}_i^{CO_2} \cdot D_i. \tag{31}$$

Liu et al. (2012) investigated the electroplating process following three objectives: (I) The developed model minimizes the production cycle time of the hoist by determining the loaded move sequence and job lifting and releasing sequence, and starting and ending time of each loaded move. (II) By calculating the drag-in starting and ending times and drag-out starting and ending times in each rinse unit, the energy consumption is minimized. (III) To minimize water consumption, the water flow rates between rinse units are determined.

6.2 Load management

In contrast to job allocation and sequencing, load management (LM) and load tracking (LT) (see Section 6.3) are PP approaches that were developed solely to foster EE manufacturing. Both approaches originated from the fact that variances in power demand are a main cost driver of utilities. As a consequence, utilities have a strong interest in inducing their customers to flatten their power demand pattern such that cost-intensive power generation facilities covering peak demands need to be activated very rarely and investments in extending capacities can be avoided (Ashok, 2006). To influence the power demand pattern of their customers, utilities make use of energy supply contracts that, in case of an LM contract, penalize peak power consumption, or, in case of an LT contract, penalize any deviation from a predefined load curve (Nolde and Morari, 2010).

LM models are relatively homogeneous. In most cases, the models determine how a machine should be operated in a given time interval. This can either be an 'on-off' decision or a decision on the production rate. Santos and Dourado (1999), for example, studied the continuous production process of a mill in the kraft pulp and paper industry and minimized ERC by varying the production rates of the different mill departments, $p_{k,t}$:

$$ERC = \sum_{t=1}^{T} \sum_{k=1}^{m} p_{k,t} \cdot l_t \cdot PW_k^{prc} \cdot C_t^E . \tag{32}$$

In addition, Santos and Dourado (1999) minimized the number of production rate changes as these changes may reduce the efficiency of the production facility. However, most of the existing LM approaches solely focus on the first objective (Equation (32)). Ghobeity and Mitsos (2010) and Yechiel and Shevah (2012) developed models that determine the optimal production loads of seawater reverse osmosis desalination plants under time-varying energy prices. In contrast to Santos and Dourado (1999), Ierapetritou et al. (2002) considered time-varying, partly known and partly unknown energy prices in the context of LM of an air separation plant and also varied the production rates to flatten the load curve. As Ierapetritou et al. (2002), Zhu et al. (2011) also considered LM in an air separation process under time-varying energy prices, but with uncertain product demand. Instead of minimizing ERC, Zhu et al. (2011) maximized the revenue from the sales of products less the operating costs. The same approach was chosen by Yusta et al. (2010), who modeled a machining process and maximized profit while taking account of hourly variations in electricity prices on the spot market.

In contrast to the LM approaches discussed thus far, Ashok and Banerjee (2001), who studied LM of a flour mill, replaced the continuous decision variable $p_{k,t}$ by a binary decision variable $I_{k,t}$. This binary decision variable indicates whether machine M_k is operating in time interval t. Additionally, the authors introduced the binary variable $A_{k,t}$ that indicates whether a LM action takes place on machine M_k in time interval t. Each LM action causes a fixed cost. Hence, the following objective function can be derived:

$$ERC = \sum_{t=1}^{T} \sum_{k=1}^{m} \left(I_{k,t} \cdot l_t \cdot PW_k^{prc} \cdot C_t^E + A_{k,t} \cdot C^A \right) . \tag{33}$$

Ashok (2006) adjusted the model of Ashok and Banerjee (2001) for the use in batch-type LM, applied in mini steel-plants. Additionally, the author considered charges for MCD without including it into the optimization problem. This aspect was adjusted by Babu and Ashok (2008), who investigated LM in the electrolytic process industry. Besides charges for energy consumption and cost for LM actions, Ostadi et al. (2007) included a term in the objective function that penalizes the energy consumption in excess of MCD_t for the case of a factory with several production facilities. In contrast, Mikhaylidi et al. (2015) assumed that the electricity capacity in a given period is restricted. Yet, by the use of a rechargeable battery, energy can be charged from the grid to the battery in off-peak periods and discharged again in on-peak periods. By determining the processing times of the machines, the model minimizes operation postponement penalty cost, setup cost, cost for keeping machines in an idle operating mode, cost for storing electricity in the rechargeable battery, and cost for electricity consumed from the grid.

Wang and Li (2013) introduced LM into a multi-stage serial production system with intermediate buffers and machine reliability considerations, minimizing both ERC and power demand cost, PDC:

$$ERC = \sum_{k=1}^{m} \sum_{t=1}^{T} PW_k^{prc} \cdot l_t \cdot \pi_{k,t} \cdot I_{k,t} \cdot C_t^E, \tag{34}$$

$$PDC = \max \left\{ \frac{\sum_{\tau=t}^{t+L-1} \sum_{k=1}^{m} PW_k^{prc} \cdot I_{k,\tau} \cdot C_\tau^D}{L} | t \in \{1,2, \dots T\} \right\}, \tag{35}$$

where $\pi_{k,t}$ corresponds to the probability that machine M_k is processing during time interval t. L corresponds to the ceiling integer number of time intervals in any 15 min interval. This expression is necessary as the power demand cost is computed based on the highest average power demand measured in any on-peak 15 min interval during the billing period (Wang and Li, 2013). Bego et al. (2014) investigated LM in a serial multi-stage production system with intermediate buffers as well. Yet, instead of a TOU tariff, they considered a critical peak pricing (CPP) program[3] where participating customers have to identify a reservation capacity when signing the contract with the utilities. Electricity consumption exceeding this reservation capacity during the CPP intervals is penalized with an extremely high energy price. The decision variables of the developed model are the reservation capacity, RC, and the operating modes of the machines. The developed model minimizes ERC and the potential penalty cost due to the failure of timely fulfillment of target production:

$$ERC = C_1 + C_2 + C_3 + C_4 + C_5, \tag{36}$$

where C_1 corresponds to the electricity consumption cost during CPP intervals when the electricity consumed is higher than the level corresponding to the reservation capacity ($\Omega_t = 1$):

$$C_1 = \sum_{t \in CPP} \Omega_t \cdot \left(C_1^{RC} \cdot \left(\sum_{k=1}^{m} I_{k,t} \cdot PW_k^{prc} - RC \right) \cdot l_t + C_2^{RC} \cdot RC \cdot l_t \right). \tag{37}$$

C_2 is the electricity consumption cost during CPP intervals when the electricity consumed is not higher than the level corresponding to the reservation capacity ($\Omega_t = 0$):

$$C_2 = \sum_{t \in CPP} (1 - \Omega_t) \cdot C_2^{RC} \cdot \left(\sum_{k=1}^{m} I_{k,t} \cdot PW_k^{prc} \right) \cdot l_t. \tag{38}$$

C_3 and C_4 are the cost of the electricity consumed in the on- and off-peak periods during non-CPP intervals, respectively, calculated similar to Ashok and Banerjee (2001). C_5 is the cost for the reservation capacity.

Kong et al. (2014) developed a LM model for multiple furnaces in electric smelting plants. Apart from minimizing ERC, the model maximizes the total production output and the product quality (percentage of valuable metals in all products). A different model was proposed by Loganthurai et al. (2014), who studied the production process in the granite industry and who solely focused on minimizing PPC by

[3] CPP programs are an extension of TOU tariffs and consider CPP periods in addition to off-, mid-, and on-peak periods. The occurrence of CPP periods is usually announced by the utilities on short notice in consequence of weather forecasts, power demand on the previous day, etc. (Bego et al., 2014).

determining the optimal number of polishing machines to be operated during time interval t based on the operating time of cutting machines and compressor. Mohagheghi and Raji (2015) extended the LM approaches described so far by introducing a procedure which dynamically ranks loads and workstations of an industrial site as candidates for demand reduction. Based on this ranking, LM actions are scheduled to maximize power demand reduction without compromising the dynamic constraints of the overall plant to minimize TEC:

$$TEC = \sum_{k=1}^{m} \sum_{t=1}^{T} \xi_k \cdot PW_k^{prc} \cdot u_{k,t}. \tag{39}$$

Here, ξ_k corresponds to the demand response ranking of machine M_k, and $u_{k,t}$ is the utilization level of machine M_k in time interval t. In contrast to the previously discussed LM approaches, Yang et al. (2008) presented a specific LM model for a tandem rolling mill. In the considered scenario, a LM action is not determined by power reductions per se, but rather by the distribution of the intended thickness reductions of steel among three out of the five stands of the tandem cold mill.

6.3 Load tracking

Compared to the body of literature on LM, there are only a few works on LT, probably because LT contracts are not very common in practice. The LT problem originates from energy supply contracts in which utilities and customers agree on a specific load curve to be demanded by the customer. Deviations from this predefined load curve are penalized. Nolde and Morari (2010) studied such a case where a utility company provides a steel plant with a contracted load curve. To adjust production according to this contracted load curve, the authors developed an electrical load tracking scheduling model. The developed model minimizes the total load tracking error, $TrcErr$:

$$TrcErr = \sum_{t=1}^{T} \left| ld_t - \sum_{i=1}^{n} ldCntr_{i,t} \right|, \tag{40}$$

where $ldCntr_{i,t}$ corresponds to the power consumption that job J_i contributes to the load requirement of load interval t, and where ld_t is the contracted load in time interval t. To minimize the total load tracking error over the planning horizon, the model determines the starting and ending times of the tasks such that the contracted electrical load is tracked best. The model developed by Nolde and Morari (2010) was extended by Hait and Artigues (2011b), who suggested an alternative continuous time formulation of LT, which provides the basis for a considerably faster computation method than the one presented by Nolde and Morari (2010).

The model of Castro et al. (2013) also features an objective function driven by an incentive-based program, besides the objective driven by a price-based program already described in Section 6.1.3.2. This objective function minimizes ERC and costs incurred by deviations from a pre-contracted load curve, $DevC$:

$$DevC = \sum_{rs \in RS^{EN}} \sum_{hr \in Day} (C^{Pen,OC} \cdot OC_{rs,hr} + C^{Pen,UC} \cdot UC_{rs,hr}), \tag{41}$$

where $OC_{rs,hr}$ and $UC_{rs,hr}$ denote the over- and underconsumption of energy resource rs in hour hr, respectively, while $C^{Pen,OC}$ and $C^{Pen,UC}$ represent the corresponding penalties.

7 Conclusions, managerial insights, and implications for further research

Steadily increasing industrial energy consumption coupled with the progressing depletion of non-renewable resources has put energy efficiency on the agendas of both researchers and practitioners. Soaring energy prices have prompted the industrial sector to rethink its, at times, lax handling of energy consumption, while the growing concern of the society for environmental issues has induced politics to react with corresponding directives. On these grounds, research on EEPP has significantly increased in recent years, which led to numerous publications integrating energy efficiency considerations into existing PP models. Yet, existing modeling approaches in this emerging stream of research have thus far not been adequately linked and synthesized. The review at hand aimed at filling this gap by categorizing papers on EEPP into classical PP tasks from a technical and model-driven perspective.

Overall, the majority of the sampled articles on EEPP deal with job allocation and sequencing problems (51 articles; see Figure 5). However, it is striking that only 13 articles have been published in the area of JS scheduling, although JSs depict real-world manufacturing systems best. Yet, because of their flexibility, JS scheduling problems are also the hardest scheduling problems to solve.

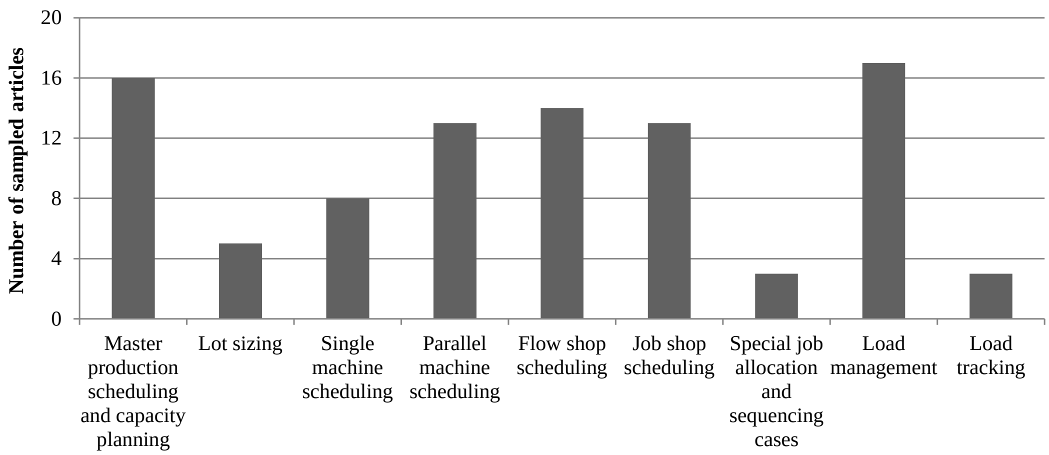
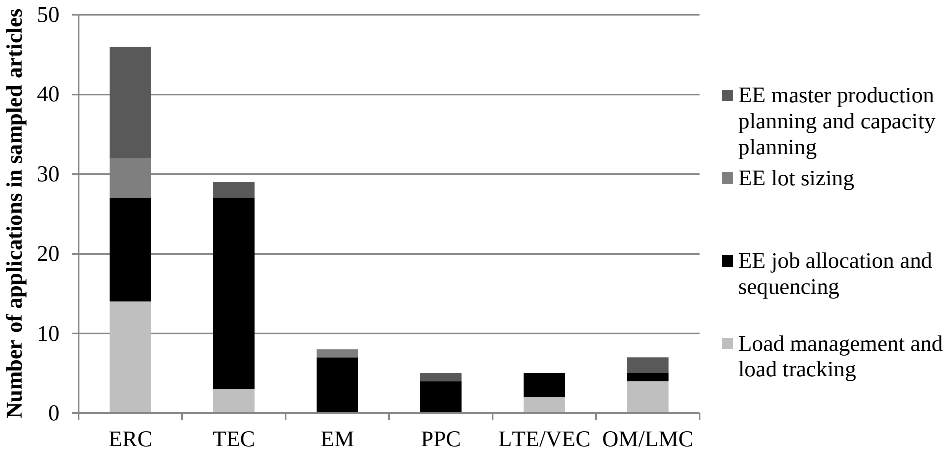


Figure 5: Number of sampled articles per production planning problem.

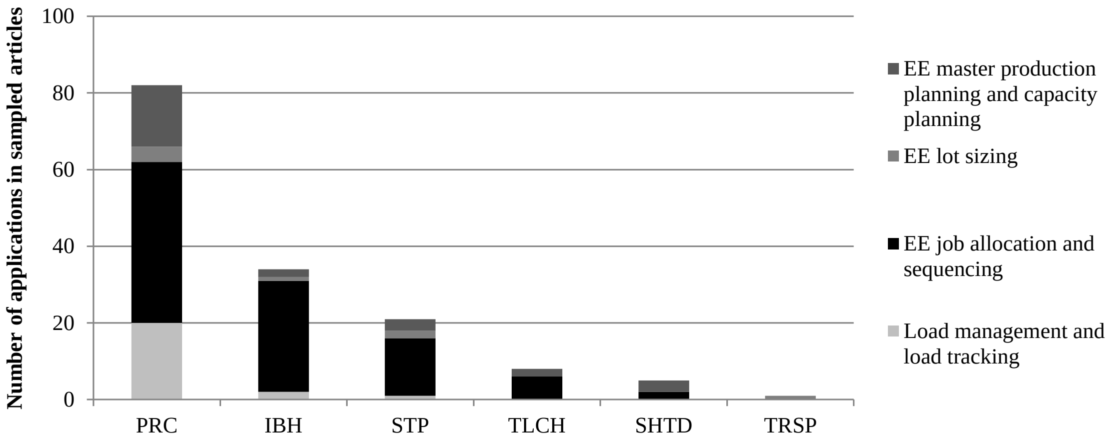
From Figure 6, we can infer that most works on EEPP reviewed in this paper integrated energy efficiency considerations into existing models by minimizing energy-related cost (46 articles). In most cases, the calculation of energy-related cost was based on price-driven demand response (DR) programs such as TOU tariffs, which are used by utilities to induce their customers to shift their energy consumption from high-demand to low-demand periods. Articles on job allocation and sequencing, in turn, predominately expressed EE considerations solely in the form of energy consumption (24 articles). At the same time, articles in this category were the only ones considering energy-related GHG emissions (7 articles), except for the EE lot sizing model of Bazan et al. (2015).



ERC = Energy-related cost, TEC = Total energy consumption, EM = Energy-related GHG emissions, PPC = Peak power consumption, LTE/VEC = Load tracking error/variance of energy consumption, OM/LMC = Cost of operating mode changes/load management actions

Figure 6: Number of applications of energy efficiency criteria in sampled articles.

The majority of the reviewed models took account of the energy consumed during processing (82 articles; see Figure 7). Yet, especially EE job allocation and sequencing approaches also gave weight to the energy consumed by idle machines (29 articles) and by machine setups (15 articles).



PRC = Processing, IBH = Idle, Basic, Holding, STP = Setup, TLCH = Toolchange, SHTD = Shutdown, TRSP = Transport

Figure 7: Number of applications of energy consuming machine operating modes in sampled articles.

From the analysis of the sampled articles, several managerial insights can be derived:

- EEPP models have proven to foster EE manufacturing in numerous ways. Manufacturing companies may use them to fulfil their usual production targets at lower energy consumption, lower energy cost, and less energy-related GHG emissions, respectively.

- The EEPP models reviewed in this article show that considering costs attached to energy consumption and CO_2 emissions is an important step in recognizing the true cost of manufacturing activities. PP models lacking energy aspects may induce practitioners to make PP decisions based on incomplete information, resulting in a suboptimal financial and environmental performance.
- EEPP models are very flexible with regard to both the specifications of the manufacturing processes they are intended to describe and the energy efficiency criteria considered crucial by a given company. Hence, practitioners may capitalize on the wealth of EEPP models that has been reviewed in this article when modeling their own manufacturing process along with the relevant energy efficiency criteria.
- The development of EEPP models requires a deep understanding of the production flow as well as of the energy consumption and CO_2 emissions of a given manufacturing process. Hence, practitioners should have recourse to experts from both the operations management field and the engineering field to arrive at high-quality PP solutions.
- In contrast to technological adjustments of manufacturing systems, the adoption of EEPP is usually not tied to large investments. Hence, EEPP models provide a comparably inexpensive opportunity, even for small to mid-size companies, to gear their manufacturing processes towards the rising importance of energy awareness.

Despite the past research efforts portrayed in this paper, there are several research gaps that provide opportunities for future research:

- Most of the discussed models integrated energy efficiency concerns only one-dimensionally. For instance, several papers simply aimed at minimizing energy consumption along with a traditional goal of PP. However, this approach neglects the economic reality of attached energy prices, which significantly vary over time. Furthermore, articles that took time-varying energy prices into account almost unanimously ignored the fact that the cheap energy consumed during off-peak periods is usually produced by coal-fired generation plants, while the expensive energy consumed during on-peak periods is produced by gas-fired generation plants. However, the amount of CO_2 emissions per kWh of power generated by coal-fired generation plants is significantly higher than the amount of CO_2 emissions resulting from gas-fired generation plants (Zhang et al., 2014). Hence, future research needs to account for the fact that objectives meant to foster energy efficiency, such as the simultaneous minimization of energy consumption, energy-related cost, and CO_2 emissions, may lead to conflicts in more advanced EEPP models (see Sharma et al., 2015; Zhang et al., 2014).
- The vast majority of the models aiming at the minimization of energy-related cost value the consumption of energy using prices that vary over time according to a predetermined schedule (TOU tariffs). Only Bego et al. (2014) incorporated the more advanced CPP scheme into their model. Karwan and Keblis (2007), Yusta et al. (2010), and Zhang et al. (2015a) considered real-time pricing. Power demand charges, which are usually responsible for a considerable share of the total energy-related cost, were not considered by the majority of articles, particularly not in the area of job allocation and sequencing. Overall, today's energy sourcing market offers a variety of energy procurement options, ranging from long-term forward contracts to spot market opportunities (Safdarian et al., 2014). Future research on EEPP should diversify its energy sourcing considerations and additionally take the price setting perspective of the utilities into account (Tsitsiklis and Xu, 2015). Furthermore, it is striking that overall EEPP models thus far predominately featured price-driven DR programs. Event-driven DR pro-

grams, such as ILCs, were only rarely taken account of in the reviewed articles (e.g., Latifoğlu et al., 2013). Hence, research on event-driven DR programs in EEPP may be intensified in the future as they become more popular in practice.

- A while ago, research striving to promote EE manufacturing from a technological point of view has already recognized the potential of integrating RES and ESSs into production systems. In contrast, the management-oriented PP community was much slower to incorporate these technological advancements into its models. Thus far, only Liu (2015) and Moon and Park (2014) took account of both RES and ESSs in a single-stage and a FJS scheduling problem, respectively, while Mikhaylidi et al. (2015) integrated an ESS into load management. However, overall PP research has by far not tapped the full potential RES and EESs provide to foster EE manufacturing. Future research needs to model the technological processes attached to generating energy from RES and storing it using different ESSs, and it has to extend existing PP approaches accordingly. A particular challenge will be the integration of the uncertainty attached to energy provided by RES into EEPP models.
- Lastly, almost all of the models reviewed in this paper assumed that CO_2 emissions are linearly dependent on energy consumption. Only Bazan et al. (2015) assumed a quadratic relationship based on the empirical study of Narita (2012). As mentioned before, Zhang et al. (2014) at least accounted for the fact that the amount of CO_2 emissions per kWh of energy depends on the source the energy is generated from. Yet, future research should intensify its efforts to realistically model the relationship between energy consumption and the corresponding CO_2 emissions as deviations from the often-assumed linear relationship may alter production plans considerably.

Looking at the dynamic evolution of research on EEPP, it is obvious that this review is limited in some respects. First, a different set of keywords may have led to another sample of articles. Secondly, the restriction to articles published in English and in academic, peer-reviewed journals may have excluded other relevant articles. Finally, the literature on the minimization of energy consumption within a production plant without considering the connection between PP decisions and energy consumption, which we intentionally excluded from this review, may provide additional insights into EEPP.

References

Artigues, C., Lopez, P., Hait, A., 2013. The energy scheduling problem: Industrial case-study and constraint propagation techniques. *International Journal of Production Economis*, 143 (1), 13-23.

Ashok, S., 2006. Peak-load management in steel plants. *Applied Energy*, 83 (5), 413-424.

Ashok, S., Banerjee, R., 2001. An optimization mode for industrial load management. *IEEE Transactions on Power Systems*, 16 (4), 879-884.

Babu, C.A., Ashok, S., 2008. Peak load management in electrolytic process industries. *IEEE Transactions on Power Systems*, 23 (2), 399-405.

Bakhrankova, K., 2009. Planning, productivity and quality in continuous non discrete production. *International Journal of Management & Enterprise Development*, 7 (1), 44-64.

Bazan, E., Jaber, M.Y., Zanoni, S., 2015. Supply chain models with greenhouse gases emissions, energy usage and different coordination decisions. *Applied Mathematical Modelling*, 39 (17), 5131-5151.

Bego, A., Li, L., Sun, Z., 2014. Identification of reservation capacity in critical peak pricing electricity demand response program for sustainable manufacturing systems. *International Journal of Energy Research*, 38 (6), 728-736.

Boukas, E.K., Haurie, A., Soumis, F., 1990. Hierarchical approach to steel production scheduling under a global energy constraint. *Annals of Operations Research*, 26, 289-311.

Brandenburg, M., Govindan, K., Sarkis, J., Seuring, S., 2014. Quantitative models for sustainable supply chain management: Developments and directions. *European Journal of Operational Research*, 233 (2), 299-312.

Brandimarte, P., 1993. Routing and scheduling in a flexible job shop by tabu search. *Annals of Operations Research*, 41 (3), 157-183.

Bruzzone, A.A.G., Anghinolfi, D., Paolucci, M., Tonelli, F., 2012. Energy-aware scheduling for improving manufacturing process sustainability: A mathematical model for flexible flow shops. *CIRP Annals - Manufacturing Technology*, 61 (1), 459-462.

Castro, P.M., Harjunkoski, I., Grossmann, I.E., 2009. New continuous-time scheduling formulation for continuous plants under variable electricity cost. *Industrial & Engineering Chemistry Research*, 48 (14), 6701-6714.

Castro, P.M., Harjunkoski, I., Grossmann, I.E., 2011. Optimal scheduling of continuous plants with energy constraints. *Computers & Chemical Engineering*, 35 (2), 372-387.

Castro, P.M., Sun, L., Harjunkoski, I., 2013. Resource–task network formulations for industrial demand side management of a steel plant. *Industrial & Engineering Chemistry Research*, 52 (36), 13046-13058.

Cataldo, A., Perizzato, A., Scattolini, R., 2015. Production scheduling of parallel machines with model predictive control. *Control Engineering Practice*, 42, 28-40.

Choi, S.H., Wang, K., 2012. Flexible flow shop scheduling with stochastic processing times: A decomposition-based approach. *Computers & Industrial Engineering*, 63 (2), 362-373.

Choi, Y.-C., Xirouchakis, P., 2014. A production planning in highly automated manufacturing system considering multiple process plans with different energy requirements. *The International Journal of Advanced Manufacturing Technology*, 70 (5-8), 853-867.

Choi, Y.-C., Xirouchakis, P., 2015. A holistic production planning approach in a reconfigurable manufacturing system with energy consumption and environmental effects. *International Journal of Computer Integrated Manufacturing*, 28 (4), 379-394.

Collier, P.I., Ornek, A., 1983. A mathematical model for energy and in-process inventory assessment in a manufacturing system. *Applied Energy*, 13 (4), 265-280.

Dai, M., Tang, D., Giret, A., Salido, M.A., Li, W.D., 2013. Energy-efficient scheduling for a flexible flow shop using an improved genetic-simulated annealing algorithm. *Robotics and Computer-Integrated Manufacturing*, 29 (5), 418-429.

Denton, F.T., Jefferies, K.L., Mountain, D.C., Robb, A.L., Spencer, B.G., 1987. The response of an industrial firm to alternative electricity rate structures: An optimization model for simulation applications. *Resources and Energy*, 9 (4), 327-346.

Ding, J., Song, S., Zhang, R., Chiong, R., Wu, C., 2015. Parallel machine scheduling under time-of-use electricity prices: New models and optimization approaches. *IEEE Transactions on Automation Science and Engineering*, 13 (2), 1138-1154.

Duerden, C., Shark, L.K., Hall, G., Howe, J., 2015. Minimisation of energy consumption variance for multi-process manufacturing lines through genetic algorithm manipulation of production schedule. *Engineering Letters*, 23 (1), 40-48.

Duflou, J.R., Sutherland, J.W., Dornfeld, D., Herrmann, C., Jeswiet, J., Kara, S., Hauschild, M., Kellens, K., 2012. Towards energy and resource efficient manufacturing: A processes and systems approach. *CIRP Annals - Manufacturing Technology*, 61 (2), 587-609.

Fang, K., Uhan, N., Zhao, F., Sutherland, J.W., 2011. A new approach to scheduling in manufacturing for power consumption and carbon footprint reduction. *Journal of Manufacturing Systems*, 30 (4), 234-240.

Fang, K., Uhan, N., Zhao, F., Sutherland, J., 2013. Flow shop scheduling with peak power consumption constraints. *Annals of Operations Research*, 206 (1), 115-145.

Fernandez, M., Li, L., Sun, Z., 2013. “Just-for-Peak” buffer inventory for peak electricity demand reduction of manufacturing systems. *International Journal of Production Economics*, 146 (1), 178-184.

Fethke, G., Tishler, A., 1989. The optimal choice of inputs under time-of-use pricing and fixed-proportions technology: An application to industrial firms∗. *Resources and Energy*, 11 (3), 241-259.

Fleischmann, B., Meyr, H., Wagner, M., 2005. *Advanced planning*. In: Stadler, H., Kilger C. (Eds.), Supply chain management and advanced planning: Concepts, models, software and case studies, 3rd ed., Springer, Berlin Heidelberg, pp. 81-106.

Gahm, C., Denz, F., Dirr, M., Tuma, A., 2016. Energy-efficient scheduling in manufacturing companies: A review and research framework. *European Journal of Operational Research*, 248 (3), 744-757.

Garcia-Santiago, C.A., Del Ser, J., Upton, C., Quilligan, F., Gil-Lopez, S., Salcedo-Sanz, S., 2015. A random-key encoded harmony search approach for energy-efficient production scheduling with shared resources. *Engineering Optimization*, 47 (11), 1481-1496.

Garetti, M., Taisch, M., 2012. Sustainable manufacturing: Trends and research challenges. *Production Planning & Control*, 23 (2-3), 83-104.

Ghobeity, A., Mitsos, A., 2010. Optimal time-dependent operation of seawater reverse osmosis. *Desalination*, 263 (1), 76-88.

Giret, A., Trentesaux, D., Prabhu, V., 2015. Sustainability in manufacturing operations scheduling: A state of the art review. *Journal of Manufacturing Systems*, 37, 126-140.

Glock, C.H., Hochrein, S., 2011. Purchasing Organization and Design: A literature review. *Business Research*, 4 (2), 149-191.

Haapala, K.R., Zhao, F., Camelio, J., Sutherland, J.W., Skerlos, S.J., Dornfeld, D.A., Jawahir, I.S., Clarens, A.F., Rickli, J.L., 2013. A review of engineering research in sustainable manufacturing. *Journal of Manufacturing Science and Engineering*, 135 (4), 041013-1-041013-16.

Hait, A., Artigues, C., 2011a. A hybrid CP/MILP method for scheduling with energy costs. *European Journal of Industrial Engineering*, 5 (4), 471-489.

Hait, A., Artigues, C., 2011b. On electrical load tracking scheduling for a steel plant. *Computers & Chemical Engineering*, 35 (12), 3044-3047.

Hasanbeigi, A., Price, L., 2012. A review of energy use and energy efficiency technologies for the textile industry. *Renewable & Sustainable Energy Reviews*, 16 (6), 3648-3665.

Hax, A.C., Meal, H.C., 1975. *Hierarchical integration of production planning and scheduling*. In: Geisler, M.A. (Ed.), Logistics: TIMS Studies in Management Science, vol. 1, North-Holland, Amsterdam, pp. 53-69.

He, Y., Li, Y., Wu, T., Sutherland, J.W., 2015. An energy-responsive optimization method for machine tool selection and operation sequence in flexible machining job shops. *Journal of Cleaner Production*, 87, 245-254.

He, Y., Liu, F., 2010. Methods for integrating energy consumption and environmental impact considerations into the production operation of machining processes. *Chinese Journal of Mechanical Enginering*, 23 (4), 428-435.

He, Y., Liu, F., Cao, H.-J., Li, C.-B., 2005. A bi-objective model for job-shop scheduling problem to minimize both energy consumption and makespan. *Journal of Central South University of Technology*, 12 (2), 167-171.

He, Y., Liu, F., Shi, J., 2008. A framework of scheduling models in machining workshop for green manufacturing. *Journal of Advanced Manufacturing Systems*, 7 (2), 319-322.

Herrmann, C., Thiede, S., 2009. Process chain simulation to foster energy efficiency in manufacturing. *CIRP Journal of Manufacturing Science and Technology*, 1 (4), 221-229.

Hochrein, S., Glock, C.H., 2012. Systematic literature reviews in purchasing and supply management research: A tertiary study. *International Journal of Integrated Supply Management*, 7 (4), 215-245.

Ierapetritou, M.G., Wu, D., Vin, J., Sweeney, P., Chigirinskiy, M., 2002. Cost minimization in an energy-intensive plant using mathematical programming approaches. *Industrial & Engineering Chemistry Research*, 41 (21), 5262-5277.

Jiang, Z., Zuo, L., Mingcheng, E., 2014. Study on multi-objective flexible job-shop scheduling problem considering energy consumption. *Journal of Industrial Engineering and Management*, 7 (3), 589-604.

Karwan, M.H., Keblis, M.F., 2007. Operations planning with real time pricing of a primary input. *Computers & Operations Research*, 34 (3), 848-867.

Kondili, E., Pantelides, C., Sargent, R., 1993. A general algorithm for short-term scheduling of batch operations—I. MILP formulation. *Computers & Chemical Engineering*, 17 (2), 211-227.

Kong, W., Chai, T., Ding, J., Yang, S., 2014. Multifurnace optimization in electric smelting plants by load scheduling and control. *IEEE Transactions on Automation Science and Engineering*, 11 (3), 850-862.

Kopanos, G.M., Xenos, D.P., Cicciotti, M., Pistikopoulos, E.N., Thornhill, N.F., 2015. Optimization of a network of compressors in parallel: Operational and maintenance planning - The air separation plant case. *Applied Energy*, 146, 453-470.

Küster, T., Lützenberger, M., Freund, D., Albayrak, S., 2013. Distributed evolutionary optimisation for electricity price responsive manufacturing using multi-agent system technology. *International Journal on Advances in Intelligent Systems*, 6, 27-40.

Latifoğlu, C., Belotti, P., Snyder, L.V., 2013. Models for production planning under power interruptions. *Naval Research Logistics*, 60 (5), 413-431.

Le, C.V., Pang, C.K., 2013. Fast reactive scheduling to minimize tardiness penalty and energy cost under power consumption uncertainties. *Computers & Industrial Engineering*, 66 (2), 406-417.

Liang, P., Yang, H.D., Liu, G.S., Guo, J.H., 2015. An ant optimization model for unrelated parallel machine scheduling with energy consumption and total tardiness. *Mathematical Problems in Engineering*, 2015, 1-8.

Lin, W., Yu, D.Y., Zhang, C., Liu, X., Zhang, S., Tian, Y., Liu, S., Xie, Z., 2015. A multi-objective teaching–learning-based optimization algorithm to scheduling in turning processes for minimizing makespan and carbon footprint. *Journal of Cleaner Production*, 101, 337-347.

Liu, C.-H., 2014. Approximate trade-off between minimisation of total weighted tardiness and minimisation of carbon dioxide (CO_2) emissions in bi-criteria batch scheduling problem. *International Journal of Computer Integrated Manufacturing*, 27 (8), 759-771.

Liu, C.-H., 2015. Mathematical programming formulations for single-machine scheduling problems while considering renewable energy uncertainty. *International Journal of Production Research*, 54 (4), 1122-1133.

Liu, C.-H., Huang, D.-H., 2014. Reduction of power consumption and carbon footprints by applying multi-objective optimisation via genetic algorithms. *International Journal of Production Research*, 52 (2), 337-352.

Liu, C., Dang, F., Li, W., Lian, J., Evans, S., Yin, Y., 2014a. Production planning of multi-stage multi-option seru production systems with sustainable measures. *Journal of Cleaner Production*, 105, 285-299.

Liu, C., Yang, J., Lian, J., Li, W., Evans, S., Yin, Y., 2014b. Sustainable performance oriented operational decision-making of single machine systems with deterministic product arrival time. *Journal of Cleaner Production*, 85, 318-330.

Liu, C., Zhao, C., Xu, Q., 2012. Integration of electroplating process design and operation for simultaneous productivity maximization, energy saving, and freshwater minimization. *Chemical Engineering Science*, 68 (1), 202-214.

Liu, G.-S., Zhang, B.-X., Yang, H.-D., Chen, X., Huang, G.Q., 2013. A branch-and-bound algorithm for minimizing the energy consumption in the PFS problem. *Mathematical Problems in Engineering*, 2013, 1-6.

Liu, Y., Dong, H., Lohse, N., Petrovic, S., Gindy, N., 2014c. An investigation into minimising total energy consumption and total weighted tardiness in job shops. *Journal of Cleaner Production*, 65, 87-96.

Liu, Y., Dong, H., Lohse, N., Petrovic, S., 2015. Reducing environmental impact of production during a Rolling Blackout policy – A multi-objective schedule optimisation approach. *Journal of Cleaner Production*, 102, 418-427.

Loganthurai, P., Rajasekaran, V., Gnanambal, K., 2014. Optimization of operating schedule of machines in granite industry using evolutionary algorithms. *Energy Conversion and Management*, 86, 809-817.

Luo, H., Du, B., Huang, G.Q., Chen, H., Li, X., 2013. Hybrid flow shop scheduling considering machine electricity consumption cost. *International Journal of Production Economics*, 146 (2), 423-439.

May, G., Stahl, B., Taisch, M., Prabhu, V., 2015. Multi-objective genetic algorithm for energy-efficient job shop scheduling. *International Journal of Production Research*, 53 (23), 7071-7089.

Mikhaylidi, Y., Naseraldin, H., Yedidsion, L., 2015. Operations scheduling under electricity time-varying prices. *International Journal of Production Research*, 53 (23), 7136-7157.

Mitra, S., Grossmann, I.E., Pinto, J.M., Arora, N., 2012. Optimal production planning under time-sensitive electricity prices for continuous power-intensive processes. *Computers & Chemical Engineering*, 38, 171-184.

Mitra, S., Pinto, J.M., Grossmann, I.E., 2014. Optimal multi-scale capacity planning for power-intensive continuous processes under time-sensitive electricity prices and demand uncertainty. Part II: Enhanced hybrid bi-level decomposition. *Computers & Chemical Engineering*, 65, 102-111.

Mohagheghi, S., Raji, N., 2015. Dynamic demand response: A solution for improved energy efficiency for industrial customers. *IEEE Industry Applications Magazine*, 21 (2), 54-62.

Moon, J.-Y., Park, J., 2014. Smart production scheduling with time-dependent and machine-dependent electricity cost by considering distributed energy resources and energy storage. *International Journal of Production Research*, 52 (13), 3922-3939.

Moon, J.-Y., Shin, K., Park, J., 2013. Optimization of production scheduling with time-dependent and machine-dependent electricity cost for industrial energy efficiency. *The International Journal of Advanced Manufacturing Technology*, 68 (1-4), 523-535.

Mouzon, G., Yildirim, M.B., 2008. A framework to minimise total energy consumption and total tardiness on a single machine. *International Journal of Sustainable Engineering*, 1 (2), 105-116.

Mouzon, G., Yildirim, M.B., Twomey, J., 2007. Operational methods for minimization of energy consumption of manufacturing equipment. *International Journal of Production Research*, 45 (18-19), 4247-4271.

Narita, H., 2012. Environmental burden analyzer for machine tool operations and its application. *INTECH Open Access Publisher*, 247-260.

Neugebauer, R., Wabner, M., Rentzsch, H., Ihlenfeldt, S., 2011. Structure principles of energy efficient machine tools. *CIRP Journal of Manufacturing Science and Technology*, 4 (2), 136-147.

Nilakantan, J.M., Huang, G.Q., Ponnambalam, S., 2015. An investigation on minimizing cycle time and total energy consumption in robotic assembly line systems. *Journal of Cleaner Production*, 90, 311-325.

Nilsson, K., 1993. Industrial production planning with optimal electricity cost. *Energy Conversion and Management*, 34 (3), 153-158.

Nilsson, K., Söderström, M., 1993. Industrial applications of production planning with optimal electricity demand. *Applied Energy*, 46 (2), 181-192.

Nolde, K., Morari, M., 2010. Electrical load tracking scheduling of a steel plant. *Computers & Chemical Engineering*, 34 (11), 1899-1903.

Ostadi, B., Moazzami, D., Rezaie, K., 2007. A non-linear programming model for optimization of the electrical energy consumption in typical factory. *Applied Mathematics and Computation*, 187 (2), 944-950.

Özdamar, L., Birbil, S.I., 1999. A hierarchical planning system for energy intensive production environments. *International Journal of Production Economics*, 58 (2), 115-129.

Pang, C.K., Le, C.V., 2014. Optimization of total energy consumption in flexible manufacturing systems using weighted p-timed Petri Nets and dynamic programming. *IEEE Transactions on Automation Science and Engineering*, 11 (4), 1083-1096.

Pinedo, M.L., 2012. *Scheduling: Theory, algorithms, and systems*. 4th ed., Springer, New York, NY.

Rager, M., Gahm, C., Denz, F., 2015. Energy-oriented scheduling based on evolutionary algorithms. *Computers & Operations Research*, 54, 218-231.

Raileanu, S., Anton, F., Iatan, A., Borangiu, T., Anton, S., 2015. *Optimizing power consumption in robotized job-shop manufacturing*. In: Borangiu, T., Thomas, A., Trentesaux, D. (Eds.), Service orientation in holonic and multi-agent manufacturing, Springer, Switzerland, pp. 13-27.

Ribas, I., Leisten, R., Framiñan, J.M., 2010. Review and classification of hybrid flow shop scheduling problems from a production system and a solutions procedure perspective. *Computers & Operations Research*, 37 (8), 1439-1454.

Safdarian, A., Fotuhi-Firuzabad, M., Lehtonen, M., 2014. Impacts of time-varying electricity rates on forward contract scheduling of DisCos. *IEEE Transactions on Power Delivery*, 29 (2), 733-741.

Santos, A., Dourado, A., 1999. Global optimization of energy and production in process industries: A genetic algorithm application. *Control Engineering Practice*, 7 (4), 549-554.

Sarimveis, H.K., Angelou, A.S., Retsina, T.R., Rutherford, S.R., Bafas, G.V., 2003. Optimal energy management in pulp and paper mills. *Energy Conversion and Management*, 44 (10), 1707-1718.

Schulte Beerbühl, S., Fröhling, M., Schultmann, F., 2015. Combined scheduling and capacity planning of electricity-based ammonia production to integrate renewable energies. *European Journal of Operational Research*, 241 (3), 851-862.

Sharma, A., Zhao, F., Sutherland, J.W., 2015. Econological scheduling of a manufacturing enterprise operating under a time-of-use electricity tariff. *Journal of Cleaner Production*, 108, 256-270.

Shrouf, F., Ordieres-Meré, J., García-Sánchez, A., Ortega-Mier, M., 2014. Optimizing the production scheduling of a single machine to minimize total energy consumption costs. *Journal of Cleaner Production*, 67, 197-207.

Sun, Z., Li, L., 2013. Opportunity estimation for real-time energy control of sustainable manufacturing systems. *IEEE Transactions on Automation Science and Engineering*, 10 (1), 38-44.

Tan, Y.-Y., Huang, Y.-L., Liu, S.-X., 2013. Two-stage mathematical programming approach for steelmaking process scheduling under variable electricity price. *Journal of Iron and Steel Research, International*, 20 (7), 1-8.

Tan, Y., Liu, S., 2014. Models and optimisation approaches for scheduling steelmaking–refining–continuous casting production under variable electricity price. *International Journal of Production Research*, 52 (4), 1032-1049.

Tang, D., Dai, M., 2015. Energy-efficient approach to minimizing the energy consumption in an extended job-shop scheduling problem. *Chinese Jounal of Mechanical Engineering*, 28 (5), 1048-1055.

Tang, L., Ren, H., Yang, Y., 2015. Reheat furnace scheduling with energy consideration. *International Journal of Production Research*, 53 (6), 1642-1660.

Tsitsiklis, J.N., Xu, Y., 2015. Pricing of fluctuations in electricity markets. *European Journal of Operational Research*, 246 (1), 199-208.

U.S. Energy Information Administration, 2013. *International energy outlook 2013*. Washington, DC.

U.S. Energy Information Administration, 2015. *Annual energy outlook with projections to 2040*. Washington, DC.

Voß, S., Witt, A., 2007. Hybrid flow shop scheduling as a multi-mode multi-project scheduling problem with batching requirements: A real-world application. *International Journal of Production Economics*, 105 (2), 445-458.

Waldemarsson, M., Lidestam, H., Rudberg, M., 2013. Including energy in supply chain planning at a pulp company. *Applied Energy*, 112, 1056-1065.

Wang, S., Lu, X., Li, X.X., Li, W.D., 2015. A systematic approach of process planning and scheduling optimization for sustainable machining. *Journal of Cleaner Production*, 87, 914-929.

Wang, Y., Li, L., 2013. Time-of-use based electricity demand response for sustainable manufacturing systems. *Energy*, 63, 233-244.

Yang, J.-M., Che, H.-J., Dou, F.-P., Zhou, T., 2008. Genetic algorithm-based optimization used in rolling schedule. *Journal of Iron and Steel Research, International*, 15 (2), 18-22.

Yechiel, A., Shevah, Y., 2012. Optimization of energy costs for SWRO desalination plants. *Desalination and Water Treatment*, 46 (1-3), 304-311.

Yildirim, M.B., Mouzon, G., 2012. Single-machine sustainable production planning to minimize total energy consumption and total completion time using a multiple objective genetic algorithm. *IEEE Transactions on Engineering Management*, 59 (4), 585-597.

Yildirim, M.B., Nezami, F.G., 2014. Integrated maintenance and production planning with energy consumption and minimal repair. *The International Journal of Advanced Manufacturing Technology*, 74 (9-12), 1419-1430.

Yue, D., You, F., 2013. Sustainable scheduling of batch processes under economic and environmental criteria with MINLP models and algorithms. *Computers & Chemical Engineering*, 54, 44-59.

Yusta, J.M., Torres, F., Khodr, H.M., 2010. Optimal methodology for a machining process scheduling in spot electricity markets. *Energy Conversion and Management*, 51 (12), 2647-2654.

Zanoni, S., Bettoni, L., Glock, C.H., 2014. Energy implications in a two-stage production system with controllable production rates. *International Journal of Production Economics*, 149, 164-171.

Zhang, H., Zhao, F., Fang, K., Sutherland, J.W., 2014. Energy-conscious flow shop scheduling under time-of-use electricity tariffs. *CIRP Annals - Manufacturing Technology*, 63 (1), 37-40.

Zhang, H., Zhao, F., Sutherland, J.W., 2015a. Energy-efficient scheduling of multiple manufacturing factories under real-time electricity pricing. *CIRP Annals - Manufacturing Technology*, 64 (1), 41-44.

Zhang, L., Li, X., Gao, L., Zhang, G., 2013. Dynamic rescheduling in FMS that is simultaneously considering energy consumption and schedule efficiency. *The International Journal of Advanced Manufacturing Technology*, 87 (5-8), 1387-1399.

Zhang, L., Luo, Y., Zhang, Y., Song, G., 2015b. Production scheduling oriented to energy consumption optimization for process industry based on self-adaptive DE algorithm. *International Journal of Control and Automation,* 8 (2), 31-42.

Zhou, L., Li, J., Li, F., Meng, Q., Li, J., Xu, X., 2016. Energy consumption model and energy efficiency of machine tools: A comprehensive literature review. *Journal of Cleaner Production*, 112, 3721-3734.

Zhu, Y., Legg, S., Laird, C.D., 2011. A multiperiod nonlinear programming approach for operation of air separation plants with variable power pricing. *AIChE Journal*, 57 (9), 2421-2430.

Zugno, M., Morales, J.M., Pinson, P., Madsen, H., 2013. A bilevel model for electricity retailers' participation in a demand response market environment. *Energy Economics*, 36, 182-197.

Appendix

Appendix A: Notation

Sets

CPP	Set of time intervals belonging to the critical peak pricing period
Day	Set of hours per day
J	Set of n jobs
M	Set of m machines
O_i	Set of ϕ_i ordered operations of job i
OM	Set of operating modes
RS^{EN}	Set of energy resources
V	Set of production speeds

Indices

$emLmt$	Emission limit with $emLmt \in \{1,2, \dots, EmLmt\}$
fct	Factory with $fct \in \{1,2, \dots, FCT\}$
g	gth operation of job J_i with $g \in \{1,2, \dots, \phi_i\}$
h	hth job processed on a machine with $h \in \{1,2, \dots n\}$
hr	Hour with $hr \in Day$
i	Job/product with $i \in \{1,2, \dots, n\}$
k	Machine (robot/seru) with $k \in \{1,2, \dots, m\}$
om	Operating mode with $om \in OM$
$plnt$	Plant with $plnt \in \{1,2, \dots, PLNT\}$
ps	Stage with $ps \in \{1,2, \dots, PS\}$
rs	Resource with $rs \in \{1,2, \dots, RS\}$
$shft$	Shift with $shft \in \{1,2, \dots, SHFT\}$
ssn	Season with $ssn \in \{1,2, \dots, SSN\}$
t	Time interval with $t \in \{1,2, \dots, T\}$
v	Production speed with $v \in V$
γ	Production priority with $\gamma \in \{1,2, \dots, \mathrm{P}_k\}$
ζ	Generator with $\zeta \in \{1,2, \dots, NG\}$
θ	Number of times job J_i is performed with $\theta \in \{1,2, \dots, \Theta_i\}$

Parameters

C^{A}	Cost of scheduling a load management action [EUR]
C^{CO_2}	Cost of CO_2 emissions [EUR/kg CO_2]
$C^{CO_2}_{emLmt}$	Emissions penalty for exceeding emissions limit $emLmt$ [EUR/year]
C^{D}_{t}	Power demand rate in the billing period time interval t belongs to [EUR/kW]
C^{E}	Cost of energy [EUR/kWh]
$C^{E}_{plnt,t}$	Cost of energy at plant $plnt$ in time interval t [EUR/kWh]
C^{E}_{t}	Cost of energy in time interval t [EUR/kWh]
$C^{elct}_{ssn,shft}$	Additional electricity consumption charge in shift $shft$ in season ssn [EUR/kWh]
$C^{elct,base}$	Base electricity consumption charge [EUR/kWh]
$C^{elctr,ODC}_{ssn}$	Additional monthly charge for peak power demand if peak occurs during off-peak period in season ssn [EUR/kW]
$C^{elct,PDC}_{ssn}$	Monthly charge for peak power demand during peak period in season ssn [EUR/kW]
C^{PEN}	Penalty cost factor for violating the maximum contracted power demand [EUR/kW]
$C^{Pen,OC}$	Penalty cost factor for load overconsumption [EUR/kW]
$C^{Pen,UC}$	Penalty cost factor for load underconsumption [EUR/kW]
C^{RC}_{1}	Consumption charge of energy for the energy consumed above the level corresponding to the reservation capacity RC during CPP intervals [EUR/kWh]
C^{RC}_{2}	Consumption charge of energy for the energy consumed that is not higher than the level corresponding to the reservation capacity RC during CPP intervals [EUR/kWh]
$C^{trans}_{om,om',plnt}$	Cost of transition from operating mode om to operating mode om' at plant $plnt$ [EUR]
$CBat$	Operating cost of the energy storage system [EUR/kWh]
D_i	Daily demand for product i [unit]
d	Product demand rate [unit/h]
E_{om}	Energy consumed in operating mode om [kWh]
$E^{AESD}_{i,g}$	Demand for applied energy sources of the gth operation of job J_i [kWh]
$E^{ch}_{om,om'}$	Energy consumed when transitioning from operating mode om to operating mode om' [kWh]
$E^{prc}_{i,k}$	Energy consumed when processing job J_i on machine M_k [kWh]
$E'^{,prc}_{1}$	Absolute energy usage coefficient depending on machine tool [kWh/year]
$e^{prc}_{i,k}$	Specific energy consumed when processing job J_i on machine M_k (product i in seru k) [kWh/unit]
$e'^{,prc}_{0}$	Specific energy usage coefficient depending on work-piece material, tool geometrics, and spindle drive characteristics [kWh/unit]

$\bar{E}^{AESD}$	Desired level of demand for applied energy sources [kWh]
FV	Fuel volume required per truck per trip [l]
$GCost_{\zeta,t}$	Cost of operating generator ζ in time interval t [EUR/kWh]
$GQTY_t$	Amount of energy generated by renewable energy sources in time interval t [kWh]
K	Maximum number of power interruptions utilities are allowed to schedule in the planning horizon [-]
L	Ceiling integer number of time intervals in any 15 min interval [-]
l_t	Length of time interval t [h]
l^{BE}	Break-even duration [h]
l_i^{prc}	Processing time of job J_i [h]
$l_{i,g,k}^{prc}$	Processing time of the gth operation of job J_i on machine M_k [h]
$l_{i,k,v}^{prc}$	Processing time of job J_i on machine M_k with speed v [h]
$l_{i,t}^{prc}$	Processing time of product i in time interval t [h]
l^{stp}	Time required for a machine setup [h]
ld_t	Contracted load in time interval t [kW]
MCD	Maximum contracted power demand [kW]
MCD_t	Maximum contracted power demand in time interval t [kW]
n_{fct}^{FS}	Number of identical flow shops at factory fct [-]
PW^{cmmc}	Power consumed by the commercial sector [kW]
PW^{hold}	Power consumed when holding a job in a furnace after production [kW]
PW^{idle}	Power consumed in the idle operating mode [kW]
$PW_{fct,t}^{nShft}$	Non-shiftable power demand of factory fct in time interval t [kW]
PW^{prc}	Power consumed when processing [kW]
PW_i^{prc}	Power consumed when processing job J_i [kW]
$PW_{i,g,k}^{prc}$	Power consumed when processing the gth operation of job J_i on machine M_k [kW]
$PW_{i,g,k}^{prc}(t)$	Power consumed when processing the gth operation of job J_i on machine M_k in time interval t [kW]
$PW_{i,k,}^{prc}$	Power consumed when processing job J_i on machine M_k with speed v [kW]
PW_k^{prc}	Power consumed by machine M_k when processing [kW]
$PW_{k,fct}^{prc}$	Power consumed by machine M_k at factory fct when processing [kW]
$PW_{k,t}^{prc}$	Power consumed by machine M_k when processing in time interval t [kW]
PW^{rsdt}	Power consumed by the residential sector [kW]
PW^{stp}	Power consumed when setting up a machine [kW]
$pw_{i,om,plnt}^{prc}$	Factor of correlation between the power consumption and the production level when plant $plnt$ manufactures product i in operating mode om [kW/unit]
RH	Number of time intervals belonging to the rolling horizon [-]

WT_k	Weight assigned to machine M_k [-]
α^{NPM}	Percentage increase of processing times if no preventive maintenance measures are adopted [%]
α^{PM}	Percentage increase of processing times if preventive maintenance measures are adopted [%]
$\mathrm{E}_i^{CO_2}$	Total CO_2 emissions quota for product i [kg CO_2]
$\mathrm{E}^{CO_2,\max}$	Total CO_2 emissions quota for a single time interval [kg CO_2]
ϵ^{CO_2}	Amount of CO_2 emissions per kWh [kg CO_2/kWh]
$\epsilon_t^{CO_2}$	Amount of CO_2 emissions per kWh in time interval t [kg CO_2/kWh]
$\epsilon_1^{CO_2}$	CO_2 emission function parameter [kg CO_2·year^2/unit^3] (see Bazan et al., 2015)
$\epsilon_2^{CO_2}$	CO_2 emission function parameter [kg CO_2·year/unit^2] (see Bazan et al., 2015)
$\epsilon_3^{CO_2}$	CO_2 emission function parameter [kg CO_2/unit] (see Bazan et al., 2015)
$\epsilon^{CO_2,F}$	Amount of CO_2 emissions from fuel per liter consumed [kg CO_2/l]
$\pi_{k,t}$	Probability that machine M_k is processing during time interval t [%]
ξ_k	Demand response ranking of machine M_k [-]
Ω_t	Binary variable which equals 1 if the power consumed during CPP interval t exceeds the reservation capacity RC, and 0 otherwise: $\Omega_t = \begin{cases} 1, if\ \sum_{k=1}^{m} I_{k,t} \cdot PW_k > RC \\ 0, otherwise \end{cases}, \forall t \in CPP$ [-]

Decision variables

$A_{k,t}$	Binary variable which equals 1 if a load management action is scheduled on machine M_k in time interval t, and 0 otherwise [-]
$a_{k,ps}$	Binary variable which equals 1 if robot k is allocated to stage ps, and 0 otherwise [-]
$b_{i,g,k,t}$	Binary variable which equals 1 if the gth operation of job J_i is processed on machine M_k in time interval t, and 0 otherwise [-]
$b_{i,g,k,\gamma}$	Binary variable which equals 1 if the gth operation of job J_i is processed on machine M_k at priority γ out of the P_k priorities of machine M_k [-]
$b_{i,g,\theta,k}$	Binary variable which equals 1 if the gth operation of the θth time job $J_i$is performed is processed on machine M_k, and 0 otherwise [-]
$b_{i,h,k,v}$	Binary variable which equals 1 if job J_i is the hth job processed on machine M_k with speed v, and 0 otherwise [-]
$b_{i,k}$	Binary variable which equals 1 if job J_i is processed on machine M_k, and 0 otherwise [-]
$b_{i,k,t}$	Binary variable which equals 1 if job J_i is processed on machine M_k in time interval t, and 0 otherwise [-]
$b_{i,ps}$	Binary variable which equals 1 if job J_i is processed on stage ps, and 0 otherwise [-]
$b_{i,t}$	Binary variable which equals 1 if job J_i is processed in time interval t, and 0 otherwise [-]

$b_{k,fct,t}$	Binary variable which equals 1 if machine M_k at factory fct is processing in time interval t, and 0 otherwise [-]
$BatState_t$	State of the battery at the beginning of time interval t [kWh]
c_i	Completion time of job J_i [h (continuous point in time)]
$c_{i,g,k}$	Completion time of the gth operation of job J_i on machine M_k [h (continuous point in time)]
$c_{i,k}$	Completion time of job J_i on machine M_k [h (continuous point in time)]
$ch_{om,om',plnt,t}$	Binary variable which equals 1 if there is a transition from operating mode om to operating mode om' at plant $plnt$ from time interval $t-1$ to time interval t, and 0 otherwise [-]
$ch_{om,om',t}$	Binary variable which equals 1 if there is a transition from operating mode om to operating mode om' in time interval t, and 0 otherwise [-]
$DevC$	Costs incurred by deviations from a pre-contracted load curve [EUR]
DGC	Cost of distributed generators and battery operating cost [EUR]
$E^{elct}_{ssn,shft}$	Electrical energy consumed in shift $shft$ in season ssn [kWh]
$EMPC$	Cost of CO_2 emissions from production [EUR]
$EMTC$	Cost of CO_2 emissions from transportation [EUR]
ERC	Energy-related cost [EUR]
$GQTY_{\zeta,t}$	Amount of energy generated by generator ζ out of the NG generators in time interval t [kWh]
$I_{k,t}$	Binary variable which equals 1 if machine M_k is switched on in time interval t, and 0 if it is switched off [-]
$I_{om,t}$	Binary variable which equals 1 if the machine is in operating mode om in time interval t, and 0 otherwise [-]
$l^{hold}_{i,t}$	Time job J_i is being hold in a furnace after production in time interval t [h]
$ldCntr_{i,t}$	Power consumption that job J_i contributes to the load requirement of load interval t [kW]
$MD^{elctr,ODC}_{ssn}$	Monthly peak power demand during off-peak periods in season ssn [kW]
$MD^{elctr,PDC}_{ssn}$	Monthly peak power demand during on-peak periods in season ssn [kW]
$MKSP$	Makespan [h]
$OC_{rs,hr}$	Overconsumption of energy resource rs in hour hr [kW]
$OMCH$	Cost of operating mode changes [EUR]
$p_{k,t}$	Production rate of machine M_k in time interval t [unit/h]
$PBat_t$	Amount of energy charged to or discharged from the energy storage system in time interval t [kWh]
PDC	Power demand cost [EUR]
$PENC$	Penalty cost from CO_2 emissions [EUR]
PPC	Peak total power consumption [kW]
PW_t	Power consumed in time interval t [kW]
PW^{mnfc}_t	Power consumed by factories in time interval t [kW]
Q	Production lot size [unit]
r	Reorder point [unit]
$RBPC$	Electricity cost for the case when a RBP is in effect [EUR]
RC	Reservation capacity [kW]

$s_{i,g,k}$	Start time of the gth operation of job J_i on machine M_k [h (continuous point in time)]
$s_{k,\gamma}$	Start time of the job processed on machine M_k at priority γ [h (continuous point in time)]
sh	Number of batch shipments [-]
TC^{elct}	Annual cost of electricity [EUR]
$TC^{elct,EC}$	Annual cost of electricity resulting from the energy consumption charge [EUR]
$TC^{elct,DC}$	Annual cost of electricity resulting from the power demand charge [EUR]
TC^{ntGs}	Annual cost of natural gas [EUR]
$TC^{ntGs,EC}$	Annual cost of natural gas resulting from the energy consumption charge [EUR]
$TC^{ntGs,DC}$	Annual cost of natural gas resulting from the power demand charge [EUR]
TC^{oil}	Annual cost of oil [EUR]
$TCDE$	Total CO_2 emissions [kg CO_2]
TCT	Total completion time [h]
TEC	Total energy consumption [kWh]
$TrcErr$	Total load tracking error [kW]
trc	Number of trucks of a given capacity per shipment [-]
$TTRD$	Total tardiness [h]
$TWFT$	Time-weighted flow time [h]
TWT	Total weighted tardiness [h]
$UC_{rs,hr}$	Underconsumption of energy resource rs in hour hr [kW]
$u_{k,t}$	Utilization level of machine M_k in time interval t [%]
$VarAESD$	Variance of demand for applied energy sources [$(kWh)^2$]
$z_{i,k}$	Amount of product i produced in seru k [unit]
$z_{i,om,plnt,t}$	Production output of product i in plant $plnt$ in operating mode om in time interval t [unit]
$z_{i,t}$	Production output of product i in time interval t [unit]
Ξ_{emLmt}	Binary variable which equals 1 if CO_2 emissions of a company exceed emission limit $emLmt$, and 0 otherwise [-]
$\xi_{i',h',k',v,h,k}$	Binary variable which equals 1 if job $J_{i'}$ is the h'th job processed on machine $M_{k'}$ with speed v, which starts while the hth job is processed on machine M_k, and 0 otherwise [-]

Appendix B: Classification of sampled articles

Production planning problem	Article	Energy efficiency objectives	Energy efficiency constraints	Energy supply contracts	Operating modes considered	Relation of emissions to energy consumption	Special features
Master production scheduling and capacity planning	Bakhrankova (2009)	ERC		TVCC	PRC		
	Castro et al. (2009)	ERC		TVCC	PRC		
	Castro et al. (2011)	ERC		TVCC, PCD	PRC		
	Choi and Xirouchakis (2014)	TEC			PRC, TLCH, IBH		
	Choi and Xirouchakis (2015)	TEC			PRC, TLCH, IBH		
	Denton et al. (1987)	ERC		TVCC, DC	PRC		
	Fethke and Tishler (1989)	ERC		TVCC, DC	PRC		
	Karwan and Keblis (2007)	ERC		RTP	PRC		
	Kondili et al. (1993)	ERC		TVCC	PRC		
	Kopanos et al. (2015)	ERC		TVCC	PRC, STP, SHTD		
	Latifoğlu et al. (2013)	ERC		TICC[4]	PRC		
	Mitra et al. (2012)	ERC, OM/LMC	MCD	TVCC	PRC, STP, SHTD		
	Mitra et al. (2014)	ERC, OM/LMC		TVCC	PRC, STP, SHTD		
	Nilsson (1993)	ERC		TVCC, DC	PRC		
	Nilsson and Söderström (1993)	ERC		TVCC, DC	PRC		
	Schulte Beerbühl et al. (2015)	ERC		TVCC	PRC		
Lot sizing	Bazan et al. (2015)	ERC, EM[5]		TICC	PRC, TRSP	QR	
	Collier and Ornek (1983)	ERC		TICC	STP		
	Özdamar and Birbil (1999)	ERC		TICC	PRC		
	Yildirim and Nezami (2014)	ERC		TVCC	PRC		
	Zanoni et al. (2014)	ERC		TICC	PRC, STP, IBH		

[4] Reduced flat energy consumption charge because of participation in interruptible load contract.
[5] Penalized with flat penalty per kg CO_2 and penalty which increases with amount of CO_2 emitted.

Production planning problem	Article	Energy efficiency objectives	Energy efficiency constraints	Energy supply contracts	Operating modes considered	Relation of emissions to energy consumption	Special features
Single machine scheduling	Liu (2014)	EM			PRC, IBH	LR	
	Liu (2015)	EM	MEM		PRC	LR	ESS, RES
	Liu and Huang (2014)	EM			PRC, IBH	LR	
	Liu et al. (2014b)	EM			STP, IBH	LR	
	Mouzon and Yildirim (2008)	TEC			STP, IBH		
	Mouzon et al. (2007)	TEC			STP, IBH		
	Shrouf et al. (2014)	TEC			PRC, STP, IBH, TLCH, SHTD		
	Yildirim and Mouzon (2012)	TEC			STP, IBH		
Parallel machine scheduling	Artigues et al. (2013)	ERC		TICC, PCD	PRC, IBH		
	Boukas et al. (1990)		MCD		PRC		
	Ding et al. (2015)	ERC		TVCC	PRC		
	Hait and Artigues (2011a)	ERC		TICC, PCD	PRC, IBH		
	He and Liu (2010)	TEC			PRC		
	He et al. (2005)	TEC			PRC		
	He et al. (2008)	TEC			PRC		
	Liang et al. (2015)	TEC			STP, IBH		
	Liu (2014)	EM			PRC, IBH	LR	
	Moon et al. (2013)	ERC		TVCC	PRC		
	Rager et al. (2015)	LTE/VEC			PRC		
	Tang et al. (2015)	OM/LMC, LTE/VEC			TLCH, IBH		
	Wang et al. (2015)	TEC			PRC, STP, TLCH, IBH		

Production planning problem		Article	Energy efficiency objectives	Energy efficiency constraints	Energy supply contracts	Operating modes considered	Relation of emissions to energy consumption	Special features
Flow shop scheduling	BFS	Fang et al. (2011)	PPC, EM			PRC	LR	
		Lin et al. (2015)	EM			PRC, IBH	LR	
		Sharma et al. (2015)	ERC, EM		TVCC, DC	PRC, STP, IBH	LR	
		Zhang et al. (2014)	ERC, EM		TVCC	PRC	TD	
		Zhang et al. (2015a)	ERC		RTP[6]	PRC		
	FFS	Bruzzone et al. (2012)		MCD		PRC		
		Castro et al. (2013)	ERC			PRC		
		Dai et al. (2013)	TEC			STP, IBH		
		Liu and Huang (2014)	PPC, EM			PRC, IBH	LR	
		Luo et al. (2013)	ERC		TVCC	PRC, IBH		
		Tan et al. (2013)	ERC		TVCC	PRC		
		Tan and Liu (2014)	ERC		TVCC	PRC		
	PFS	Fang et al. (2013)		MCD		PRC		
		Liu et al. (2013)	TEC			IBH		
Job shop scheduling	BJS	Duerden et al. (2015)	LTE/VEC			PRC, IBH		
		Liu et al. (2014c)	TEC			IBH		
		Liu et al. (2015)	ERC, TEC		TVCC (RBP)	PRC, STP, IBH		
		May et al. (2015)	TEC			PRC, STP, TLCH, IBH		
		Tang and Dai (2015)	TEC			PRC, STP, IBH		
		Zhang et al. (2015b)	TEC			PRC, STP, IBH, TLCH, SHTD		

[6] The electricity price is exponentially dependent on the current power demand.

Production planning problem		Article	Energy efficiency objectives	Energy efficiency constraints	Energy supply contracts	Operating modes considered	Relation of emissions to energy consumption	Special features
Job shop scheduling	FJS	He et al. (2015)	TEC			PRC, STP, TLCH, IBH		
		Jiang et al. (2014)	TEC			PRC		
		Moon and Park (2014)	ERC		TVCC	PRC		ESS, RES
	FMS	Garcia-Santiago (2015)	TEC			PRC, STP, IBH		
		Le and Pang (2013)	TEC			PRC		
		Pang and Le (2014)	TEC			PRC, IBH		
		Zhang et al. (2013)	TEC			PRC, IBH		
Special cases in job allocation and sequencing		Liu et al. (2012)	TEC			PRC		
		Liu et al. (2014a)	EM	MEM[7]		PRC	LR	
		Nilakantan et al. (2015)	TEC			PRC		
Load management		Ashok (2006)	ERC, OM/LMC		TVCC, DC	PRC		
		Ashok and Banerjee (2001)	ERC, OM/LMC		TVCC	PRC		
		Babu and Ashok (2008)	ERC		TVCC, DC	PRC		
		Bego et al. (2014)	ERC		TVCC, CPP	PRC		
		Ghobeity and Mitsos (2010)	ERC		TVCC	PRC		
		Ierapetritou et al. (2002)	ERC		TVCC	PRC, IBH		
		Kong et al. (2014)	ERC		TVCC, DC	PRC		
		Loganthurai et al. (2014)	PPC			PRC		
		Mikhaylidi et al. (2015)	ERC	TEC	TVCC	PRC, STP, IBH		ESS
		Mohagheghi and Raji (2015)	TEC			PRC		

[7] Per product.

Production planning problem	Article	Energy efficiency objectives	Energy efficiency constraints	Energy supply contracts	Operating modes considered	Relation of emissions to energy consumption	Special features
Load management	Ostadi et al. (2007)	ERC, OM/LMC		TVCC, PCD	PRC		
	Santos and Dourado (1999)	ERC		TVCC	PRC		
	Wang and Li (2013)	ERC		TVCC, DC	PRC		
	Yang et al. (2008)	TEC			PRC		
	Yechiel and Shevah (2012)	ERC		TVCC	PRC		
	Yusta et al. (2010)	ERC		RTP	PRC		
	Zhu et al. (2011)	OM/LMC		TVCC	PRC		
Load tracking	Castro et al. (2013)	ERC		PCD	PRC		
	Hait and Artigues (2011b)	LTE/VEC			PRC		
	Nolde and Morari (2010)	LTE/VEC			PRC		

Key:

BFS	Basic flow shop	LTE/VEC	Load tracking error/ variance of energy consumption	RBP	Rolling blackout policy
BJS	Basic job shop			RES	Renewable energy sources
CPP	Critical peak pricing	MCD	Maximum contracted power demand	RTP	Real-time pricing
DC	Power demand charge	MEM	Maximum emission	SHTD	Shutdown
ERC	Energy-related cost	OM/LMC	Cost of operating mode changes/ load management actions	STP	Setup
ESS	Energy storage system			TD	Time-dependent
FFS	Flexible flow shop	PCD	Penalty for power over-/underconsumption	TEC	Total energy consumed
FJS	Flexible job shop	PFS	Permutation flow shop	TICC	Time-invariant consumption charge
FMS	Flexible manufacturing system	PPC	Peak power consumption	TLCH	Toolchange
IBH	Idle, Basic, Holding	PRC	Processing	TRSP	Transport
LR	Linear relationship	QR	Quadratic relationship	TVCC	Time-varying consumption charge

Paper 2 On the use of waste heat in a two-stage production system with controllable production rates[1]

Authors: Konstantin Biel, Christoph H. Glock

Type of publication: Journal article

Publication details: *International Journal of Production Economics*, 181, 174-190, 2016.

Abstract

Industrial energy consumption accounts for approximately one third of the energy consumed by the four major end-uses of energy (i.e., residential, commercial, industrial, and transportation energy use). Manufacturing is thereby responsible for the majority of energy that is consumed in industry. The scarcity of resources, rising energy prices, and an increasing awareness that lowering energy usage is a prerequisite for sustainable production processes has induced researchers to consider energy consumption in the management of production systems. This paper contributes to this emerging stream of research by studying the role of waste heat in production planning and control. More specifically, it investigates the case where industrial waste heat can be converted into electricity, which can then be used to support operating the production stages. This paper introduces the generation and transformation of waste heat into a lot size model and investigates how lot sizing policies change if waste heat is used to operate the system. Special attention is paid to the scheduling of interruptions between production runs and the determination of optimal production rates. The results of the paper indicate that using waste heat resulting from production reduces the overall energy requirements of a production system. The inventory policies developed in this paper support an efficient use of waste heat.

Keywords:

Waste heat; Two-stage production system; Energy costs; Batch sizing; Variable production rate

[1] Reprinted from *International Journal of Production Economics*, 181, Biel, K., Glock, C.H., On the use of waste heat in a two-stage production system with controllable production rates, 174-190, 2016, with permission from Elsevier.

1 Introduction

According to the International Energy Outlook 2013, world energy consumption is expected to increase by 56 percent from 524 quadrillion British thermal units (Btu) in 2010 to 820 quadrillion Btu in 2040. Of the four major end-use sectors, the industrial sector, which includes companies from manufacturing industries (e.g., food, chemicals, and iron and steel) as well as from non-manufacturing industries (e.g., agriculture, mining, and construction; see U.S. Energy Information Administration, 2013), consumed 32 percent of the world's total delivered energy in 2012. The industrial energy consumption is expected to grow on average by 1.4 percent per year between 2010 and 2040 (U.S. Energy Information Administration, 2013). In 2012, 65 percent of the energy consumed by the industrial sector in the United States was used for manufacturing heat and power, while 17 percent was spent on non-manufacturing heat and power, and 18 percent on non-fuel uses. From 2012 to 2025, the U.S. manufacturing energy consumption for heat and power, which is of special importance for this paper, is expected to increase on average by 1.1 percent per year, where the growth is expected to slow down between 2025 and 2040, reaching on average 0.2 percent per year (U.S. Energy Information Administration, 2014). The question at the core of this paper is how production management can make an impact on the amount of energy consumed in manufacturing and how production management strategies can contribute to enhancing the energy efficiency of manufacturing systems.

In light of both an increasing energy use and rising energy prices, the reduction of energy consumption in manufacturing has received a top priority in theory as well as in practice. Existing approaches for the reduction of energy in industry can generally be divided into technological advancements and managerial measures. For insights into the literature of the first category, the reader is referred to Hasanbeigi and Price (2012), among others, who presented a comprehensive review of energy-efficient production technologies for the textile industry. However, their paper also provided numerous relevant references applicable to other industries. In the following, the focus will be on the second category, i.e., managerial measures that help to realize energy-efficient manufacturing processes. One popular measure in this context is energy-aware scheduling. This topic has been investigated by several authors, for example by Artigues et al. (2013), Luo et al. (2013), and Shrouf et al. (2014). Another popular research stream focuses on the reduction of energy consumption in reverse logistics. Papers that fall into this area are concerned with the recycling and reuse of materials and energy inputs required for production and recycling/rework (e.g., Govindan et al., 2015; Souza, 2013; Carrasco-Gallego et al., 2012). In a related line of thought, researchers try to use waste heat rejected during the production process to increase energy efficiency in production. Some authors tried to exploit cogeneration opportunities, i.e., to use the waste heat rejected by one process in the preheating step of another process (e.g., Havel and Šimovič, 2013; Salgado and Pedrero, 2008), while other authors studied the conversion of waste heat into electricity to improve energy efficiency. The system employed to execute this conversion is usually referred to as a Waste Heat Recovery System (WHRS) and is central to this paper. Its technological functionality will be explained in detail in Section 3. The applicability of WHRSs in industry is wide-ranging and includes, for instance, power plants (e.g., Wang et al., 2014; Uusitalo et al., 2014), cement factories (e.g., Karellas et al., 2013), or internal combustion engines (e.g., Zhu et al., 2014).

The authors deem the integration of a WHRS into a production system particularly promising to enhance energy efficiency in manufacturing. The reason for this becomes clear when considering the potential of industrial waste heat. According to Schaefer (1995), for example, on average 45 percent of the energy input of process heat in the German industry is rejected as waste heat during production. Hung et al. (1997) stated that the share of low-grade waste heat exceeds 50 percent of the total heat generated in industry. López et al. (1998) estimated that roughly 40 percent of the total energy consumed in industry is rejected as waste heat in a study of companies in the Basque country. Even though these numbers look promising at a first glance, it needs to be kept in mind that not the entire waste heat rejected during production can technologically and economically be utilized. Whether or not waste heat can technologically and economically be utilized depends on several factors, such as the energy content of the waste heat stream, the temperature difference between heat source and heat sink, or the geographic proximity of waste heat source and waste heat sink (Hirzel et al., 2013). Considering these factors, Pehnt et al. (2010) hypothesized that the amount of waste heat that can be utilized in a technologically and economically useful way is only 12 to 18 percent of the energy input. In a related study, McKenna and Norman (2010) concluded that 5 to 10 percent of the industrial energy consumption is rejected as reasonably useable waste heat. Thus, it appears that the overall potential of usable industrial waste heat is remarkable, but that there is room for further technological advancements to increase the amount of waste heat that can be utilized technologically and economically.

To foster the recovery of waste heat in manufacturing, this paper integrates a WHRS into a two-stage production system with controllable production rates, which is a problem that has, to the best of the authors' knowledge, not been studied in a holistic approach yet. Works from three different research streams are of relevance to this paper and need to be reviewed in the following: (I) the integration of energy aspects into production planning, (II) production management with controllable production rates, and (III) the use of an Organic Rankine Cycle (ORC) for recovering waste heat.

In the first research stream, the works of Artigues et al. (2013), Luo et al. (2013), and Shrouf et al. (2014) have already been referred to as advocates of energy-efficient production scheduling. However, one of the first models on energy-aware production planning was the one of Mouzon et al. (2007), which developed dispatching rules for the machine scheduling problem with the objective to minimize the energy consumed by the manufacturing equipment. The authors concentrated on analyzing non-bottleneck machines and found that the total energy consumption could considerably be decreased by turning off machines whose idle time exceeds a certain threshold. In a similar way, Fernandez et al. (2013) studied a serial manufacturing system with multiple machines and buffers in light of time-varying energy prices. To reduce power demand during peak periods, a so-called 'Just-for-Peak' buffer inventory is built up during off-peak periods without sacrificing system throughput. The model then decides whether or not to turn off a machine during on-peak periods, depending on whether the 'Just-for-Peak' buffer inventory is sufficiently large such that the downstream machines can continue to work in the on-peak periods without interruption. The objective of the model is to minimize the sum of the holding cost of the 'Just-for-Peak' inventory and energy consumption cost without compromising system throughput. Zanoni et al. (2014) studied a two-stage production system with controllable production rates and investigated the energy consumption of different production strategies. They showed that the energy-related cost can be cut significantly when taking energy aspects into account. Their work lays the foundations of the model presented in this paper.

As in Zanoni et al. (2014), we also consider the case where the production rates of the machines can be varied within given limits. Variable production rates have extensively been investigated in the context of the economic lot scheduling problem (e.g., Buzacott and Ozkarahan, 1983; Gallego, 1993; Elhafsi and Bai, 1997). Prior research has differentiated between a 'rigid case' and a 'flexible case', where the 'rigid case' describes a situation where production rates can only be varied before the start of a production run, while in the 'flexible case' changes in the production rate are also possible after the production run has been initiated. Glock (2010) and Glock (2011) studied the effect of variable production rates on the total costs of a two-stage and a multi-stage production system, respectively. In both works, the author showed that varying the production rate can reduce the inventory holding cost significantly. The work of Gutowski et al. (2006) linked the production rate of a machine to its energy consumption, which helps us to study the impact of varying production rates on energy efficiency enhancements associated with a WHRS.

Research on the technological background of WHRSs constitutes the third research stream that is of relevance to this paper. As our focus is on a WHRS that uses an ORC, we direct our attention to the literature that investigates this special type of WHRS. Hung et al. (1997), for example, presented a general overview of ORCs that can be used to recover waste heat. The authors investigated the thermodynamic properties of different ORCs and identified technological factors of the system and environmental conditions influencing the efficiency of ORCs. Dai et al. (2009) compared different ORCs and optimized the thermodynamic parameters of the ORCs to achieve a better performance using a genetic algorithm. Quoilin et al. (2011) extended the perspective from a merely thermodynamic point of view to a thermo-economic viewpoint. In contrast to Dai et al. (2009), Quoilin et al. (2011) considered labor cost and investment cost of the components of a WHRS in addition to the thermodynamic properties to optimize the system efficiency. They found that the optimal operating point (setting) when including the economic perspective differs from the optimal operating point (setting) when only the thermodynamic parameters are considered in the model. In addition to the works discussed in this section, several papers have been published that solely focus on the technological side of the ORC procedure. However, as the technological background of the ORC will be described in detail in Section 3, additional literature on the ORC will be introduced in this section.

The contribution of this paper is manifold. First, it integrates a WHRS into a two-stage production system with controllable production rates and thereby contributes to strengthening the link between the production management discipline and the engineering discipline. From a managerial point of view, this connection is necessary to establish a holistic production planning concept that takes account of scarce resources and constantly varying energy prices. From an engineering point of view, this link is highly valuable as it directly incorporates technological advancements into the industrial planning process and thereby highlights the proximity of research and development to real-life industry applications. Secondly, the model developed in this paper can be used to evaluate how operational decisions, such as production lot sizing or the setting of production rates, should be adopted when a WHRS is used in production. In addition, the model can be used as a decision support tool to assess whether or not a WHRS should be acquired and implemented. The model can also be employed to evaluate the robustness of a production system that is subject to changes in energy cost, changes in operating cost of the WHRS, or changes in the surrounding conditions of the WHRS.

The remainder of the paper is organized as follows: Section 2 describes the basic production system studied in this paper along with relevant assumptions and production strategies. Section 3 establishes the connection between the two-stage production system and the WHRS. Section 4 introduces an inventory model that considers a WHRS, and Section 5 develops a solution procedure for the model. Section 6 illustrates the behavior of the model with the help of a numerical example, and Section 7 concludes the paper.

2 Problem description

This paper investigates a production system that transforms raw materials into a finished product in two consecutive production steps. Figure 1 illustrates the structure of the production system, which resembles the one studied by Zanoni et al. (2014). We assume that the system faces a given demand, and that it has unlimited access to raw materials. After processing items at the first production stage, the processed items are stored in a buffer stock at the backend of the production stage (downstream). When a certain transport batch size has been reached in this buffer, the batch is shipped to the next buffer stock, which is located in front of the second production stage (upstream). The second production stage continuously removes items from its upstream buffer stock and transforms them into the finished products, which are then stored in a third (downstream) buffer until they are consumed by the final customer.

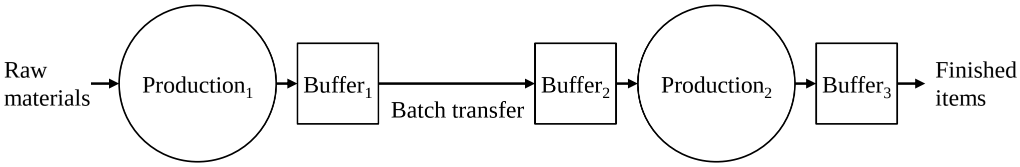


Figure 1: Product flow in two-stage production system.

Throughout the paper, the following terminology is used:

Definitions

CON	Continuous batch production policy
INT	Interrupted batch production policy
I	Idle operating mode
P	Production operating mode
S	Setup operating mode

Indices

i	Production stage with $i \in \{1,2\}$
j	Operating mode with $j \in \{P, I, S\}$
l	Production policy with $l \in \{CON, INT\}$
r	State point with $r \in \{1,2,3,4\}$

Parameters

A	Setup cost of the production system (i.e., cost for setting up both production stages) [EUR]
B_i	Binary variable which equals 1 if production stage i stays in the idle operating mode between two successive production cycles, and 0 if production stage i is switched off and switched on between two successive production cycles [-]
c	Cost of generating electricity from the ORC [EUR/kWh]
d	Demand rate [kg/h]
e	Cost of energy [EUR/kWh]
$\dot{EC}_{i,j}$	Power required by production stage i in operating mode j [kW]

f	Multiplication factor for the power required during the setup of the production system [-]
$hc_{1,2}$	Holding cost of buffer stocks 1 and 2 [EUR/(kg·h)]
hc_3	Holding cost of buffer stock 3 [EUR/(kg·h)]
h_r^{ORC}	Specific enthalpy of the organic fluid at state point r [kJ/kg]
$\dot{HR}_{i,j}$	Flow rate of waste heat recovered in operating mode j on production stage i [kW]
k_i	Energy required at production stage i to produce one unit [kWh/kg]
$\dot{m}_{i,j}^{ORC}$	Mass flow rate of the organic fluid during operating mode j on production stage i [kg/h]
$\dot{Q}_{i,j}^{WH}$	Flow rate of waste heat rejected in operating mode j by production stage i [kW]
SC	Shipment cost [EUR]
t_i^{on}	Setup time of production stage i [h]
W_i	Idle power of production stage i [kW]
W^{pump}	Power required by the pump [kW]
W^{turb}	Power generated by the turbine [kW]
α	Percentage of the initial energy input which is rejected as waste heat and led into the WHRS [%]
η^{ORC}	Efficiency of the ORC [%]
η^{pump}	Efficiency of the pump [%]
η^{turb}	Efficiency of the turbine [%]

Decision variables

$ERC_{i,j}^l$	Average energy-related cost in operating mode j on production stage i when employing production policy l [EUR/h]
$HC_{1,2}^l$	Average inventory holding cost at buffer stocks 1 and 2 when employing production policy l [EUR/h]
HC_3^l	Average inventory holding cost at buffer stock 3 when employing production policy l [EUR/h]
m	Number of shipments from buffer stock 1 to buffer stock 2 [-]
$OC_{i,j}^l$	Average cost of generating electricity from the ORC in operating mode j on production stage i when employing production policy l [EUR/h]
p_i	Production rate of production stage i [kg/h]
Q	Production lot size [kg]
SHC^l	Average shipment cost when employing production policy l [EUR/h]
SUC^l	Average setup cost when employing production policy l [EUR/h]
t_i^{cycle}	Time between two production cycles on production stage i [h]
TC^l	Average total cost of the system when employing production policy l [EUR/h]
$tPIC^l$	Average traditional production-inventory cost when employing production policy l [EUR/h]

Apart from what has already been stated, the paper makes the following assumptions:

(1) $p_1 \in [p_1^{min}, p_1^{max}] \geq p_2 \in [p_2^{min}, p_2^{max}] \geq d$, i.e., the production rates of both production stages can only be varied in predefined intervals. The production rate of the first production stage, p_1, is always equal to or higher than the production rate of the second production stage, p_2, which, in turn, is always equal to or higher than the demand rate, d;
(2) the production rates can only be varied before the start of production due to technological reasons. This scenario is usually referred to as the 'rigid case' in the literature. See, for example, Buzacott and Ozkarahan (1983), Silver (1990), and Glock (2011);
(3) processed items are shipped from the first production stage (buffer stock 1) to the second production stage (buffer stock 2) in batch shipments of equal sizes;
(4) one item of input material is required to produce one unit of the finished product on each stage;
(5) a given period of time, t_i^{on}, is required to set up production stage i;
(6) shortages are not allowed;
(7) costs of generating electricity from the ORC include the initial investment, operation and maintenance cost, cost of fuel, insurance, etc. (Tchanche et al., 2010).

Additional assumptions will be introduced where required.

As in Zanoni et al. (2014), two different types of production policies are studied in this paper. In the case of a continuous batch production policy (see Figure 2), a complete production lot of size Q is processed on the first production stage without interruption. Subsequently, the production stage remains in an idle operating mode or is switched off and switched on again when production is reinitiated (setup operating mode). Production at the second stage is initiated when buffer 3 runs out of stock.

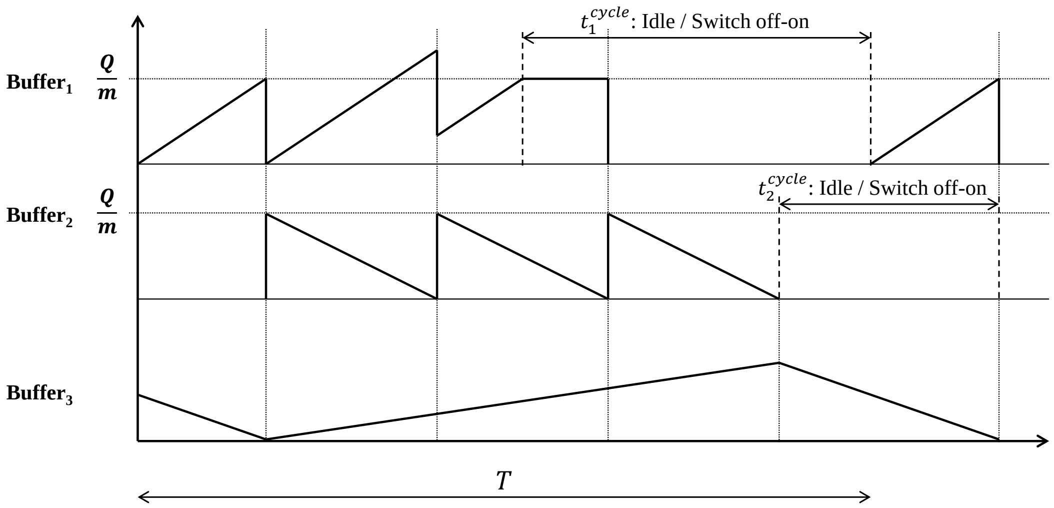


Figure 2: Inventory levels of the continuous batch production policy.

In the case of an interrupted batch production policy (see Figure 3), the first production stage splits up the lot size Q into m batches of size Q/m. After a batch has been completed, stage 1 is put into an idle operating mode, and production is resumed in time to avoid that buffer 2 runs out of stock. After a production lot has been completed, both production stages either remain in an idle operating mode or are switched off and on again as in the continuous production policy. Note that it is assumed that only the first production stage can be interrupted within a production run.

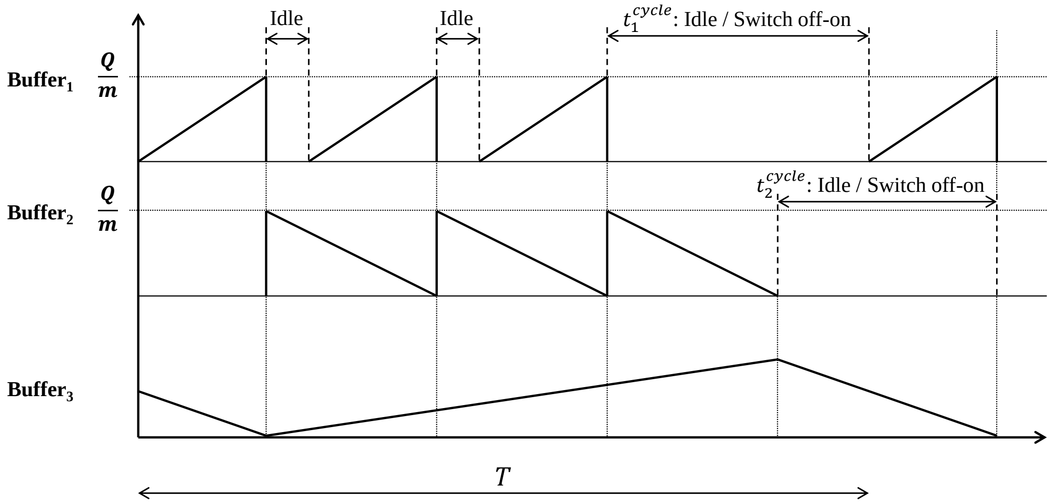


Figure 3: Inventory levels of the interrupted batch production policy.

Both production policies can be executed in four different ways, depending on whether a production stage remains in an idle operating mode between successive production cycles ($B_1 = 1$ for production stage 1, $B_2 = 1$ for production stage 2) or whether a production stage is switched off and on again between two successive production cycles ($B_1 = 0$ for production stage 1, $B_2 = 0$ for production stage 2). In total, this results in eight different production policies, which are summarized in Table 1.

Table 1: Overview of batch production policies.

Continuous batch production policies ($c/B_1/B_2$)	**Interrupted batch production policies ($i/B_1/B_2$)**
c/0/0	i/0/0
c/0/1	i/0/1
c/1/0	i/1/0
c/1/1	i/1/1

3 Integration of waste heat recovery system into the two-stage production system

This paper extends the two-stage production system introduced in Section 2 by considering energy requirements of the production stages and the conversion of recovered waste heat into electricity. The energy flow of the production system is illustrated in Figure 4. As can be seen, we assume that the production stages are operated using electricity from the grid. During all three different operating modes (production, idle, setup), the production stages reject waste heat, mainly in the form of exhaust gases. The idea at the core of this paper is to transform the waste heat generated during these operating modes into electricity, which can be used to operate the production stages. If the savings in energy consumption exceed the cost of operating the WHRS, then using waste heat increases the efficiency of the entire production system. There are several options to convert waste heat into electricity (Haddad et al., 2014). However, this paper solely focuses on the conversion of waste heat into electricity using an ORC. The rationale behind this choice is that we concentrate on industrial waste heat. As in most manufacturing processes the temperature of industrial waste heat is less than 370 °C, conventional measures to recover this waste heat, such as the application of a classical Rankine cycle, is economically not practicable. This is due to the fact that a classical Rankine cycle uses water instead of an organic working fluid. Over the course of the classical Rankine cycle, the water needs to be vaporized into steam and ultimately superheated to avoid steam condensation and erosion of the turbine blades, where the steam is expanded. Yet, the energy content of low-grade waste heat may not be high enough to superheat the steam (Johnson et al., 2008), which would reduce the lifetime of a steam turbine drastically and result in high maintenance and/or replacement cost (Quoilin et al., 2013). In contrast, the organic fluid used in an ORC does not require superheating and generally features a lower boiling point than water. Consequently, an ORC can recover low-grade waste heat significantly more efficiently than a classical Rankine cycle (Larjola, 1995). Furthermore, a classical Rankine cycle usually requires several turbines because of the extremely high pressure ratio[2] and the enthalpy drop[3] during the expansion of the steam. This, however, causes high maintenance and capital cost. In contrast, the pressure ratio and the enthalpy drop during the expansion of the ORC are much lower, which is why usually only a single-stage turbine is necessary, resulting in lower maintenance and capital cost (Andersen and Bruno, 2005). Moreover, this less complex turbine design along with the flexibility and high safety of the ORC further promotes the use of an ORC in the production system studied in this paper as it better complies with the requirements of a decentralized structure of the WHRS which we seek to apply as explained at the end of this section (Wei et al., 2007).

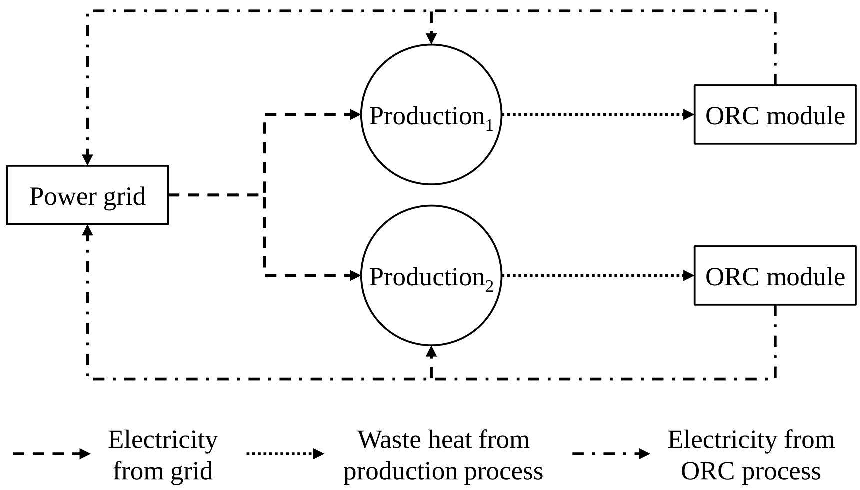


Figure 4: Energy flow in two-stage production system.

[2] The pressure ratio relates the turbine inlet pressure to the turbine outlet pressure (Yamamoto et al., 2001).
[3] Enthalpy refers to the thermodynamic potential of a state point, expressed in kJ. Hence, an enthalpy drop refers to the transition from a state point associated with a high thermodynamic potential to a state point associated with a low thermodynamic potential (Oliveira, 2013).

Figure 4 shows how the ORC mechanism is integrated into the energy flow of the production system: Waste heat rejected by the production stages is led into an ORC module, where it is transformed into electricity. This electricity can then either be used to run the production stages, which reduces the electricity requirement from the grid and which helps to make a more efficient use of the initial energy input, or it is fed into the grid in return for compensation. The model presented in this paper concentrates on the case where the generated electricity is used to run the production stages. Yet, the case where the generated electricity is fed into the grid can easily be developed from the model presented here.

Figure 5 gives an overview of the components of an ORC (Dai et al., 2009). First, the ORC makes use of a pump that provides the evaporator with an organic fluid. The evaporator heats and vaporizes the organic fluid using the waste heat rejected by the production processes. Subsequently, the high-pressure vapor is led into a single-stage turbine, where it is expanded and electric power is generated. In the last step, the condenser transforms the low-pressure vapor into a liquid again, which is then led to the pump. Subsequently, a new cycle starts (Wei et al., 2007). To describe the behavior of the system, four state points are defined (see Figure 5). The organic fluid is transferred from one state point to the next with the help of the ORC system components described above. Each of the four state points is characterized by its thermodynamic potential, which is referred to as enthalpy.

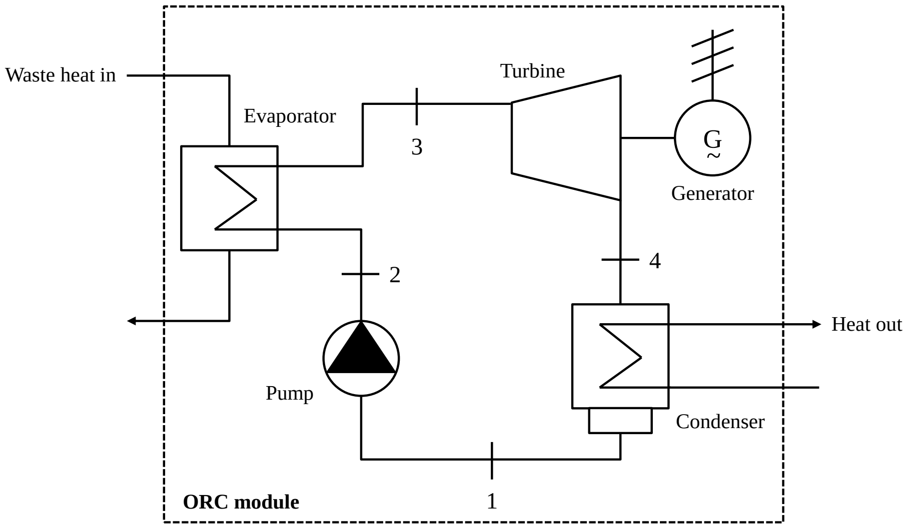


Figure 5: ORC system diagram.

To transfer the organic fluid from one state point to the next, energy is added to or withdrawn from the system. This can be done by either adding heat to or withdrawing heat from the system or by doing work on or doing work by the system. Thus, both work and heat represent the transfer of energy (Borgnakke and Sonntag, 2009). The SI unit of both work and heat is Joule (J), while the transfer rates of work and heat per unit time are measured in Watt (J/s = W). The energy (in the form of work or heat) required to transfer the organic fluid from one state point to the next is illustrated in Figure 6 with the help of a temperature-entropy process diagram.

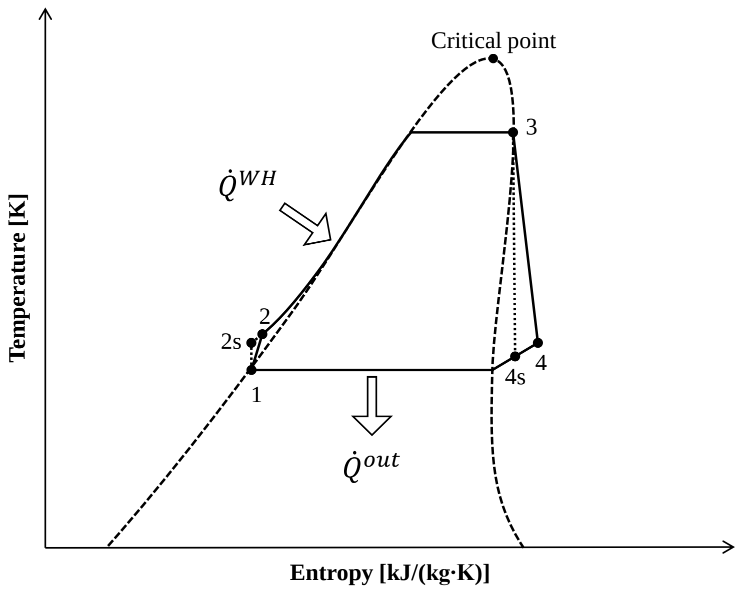


Figure 6: Typical temperature-entropy process diagram for the investigated ORC system.

The transitions between the different state points are described in the following:

- *State point 1 to state point 2:* The pump does work *on* the system by increasing the pressure of the working fluid in the transfer from state point 1 to state point 2. As stated above, state point r is characterized by its thermodynamic potential $\dot{m}^{ORC} \cdot h_r^{ORC}$ with $r \in \{1,2,3,4\}$. Thus, the work done by the pump per unit of time, W^{pump}, transfers the working fluid from the thermodynamic potential of state point 1 ($\dot{m}^{ORC} \cdot h_1^{ORC}$) to the thermodynamic potential of state point 2 ($\dot{m}^{ORC} \cdot h_2^{ORC}$), i.e., the work done by the pump per unit of time corresponds to the difference of the thermodynamic potentials of state points 1 and 2:

$$W^{pump} = \dot{m}^{ORC} \cdot \left(h_2^{ORC} - h_1^{ORC}\right) = \frac{\dot{m}^{ORC} \cdot \left(h_{2s}^{ORC} - h_1^{ORC}\right)}{\eta^{pump}}, \tag{1}$$

  whereas the change of state is not isentropic, i.e., entropy is generated. This means that losses are incurred by friction or heat dissipation within the pump during the system change from state point 1 to state point 2 (Borgnakke and Sonntag, 2009). Consequently, the efficiency of the pump $\eta^{pump} = \left(h_{2s}^{ORC} - h_1^{ORC}\right)/\left(h_2^{ORC} - h_1^{ORC}\right)$ is smaller than 1 in real-life applications as it cannot use its entire energy input for transferring the working fluid from state point 1 to state point 2.
- *State point 2 to state point 3:* The waste heat flow from the production processes, $\dot{Q}^{WH}$, is used to heat and subsequently vaporize the organic fluid. Thereby, the working fluid is transferred from state point 2 ($\dot{m}^{ORC} \cdot h_2^{ORC}$) to state point 3 ($\dot{m}^{ORC} \cdot h_3^{ORC}$). Thus, the waste heat flow from the production processes equals the difference of the thermodynamic potentials of state points 2 and 3:

$$\dot{Q}^{WH} = \dot{m}^{ORC} \cdot \left(h_3^{ORC} - h_2^{ORC}\right) = \dot{m}^{ORC} \cdot \left(h_3^{ORC} - \left(\frac{h_{2s}^{ORC} - h_1^{ORC}}{\eta^{pump}} + h_1^{ORC}\right)\right). \tag{2}$$

- *State point 3 to state point 4:* In the turbine, the high-pressure vapor is expanded and electric power is generated. The difference of the thermodynamic potentials between state point 3 ($\dot{m}^{ORC} \cdot h_3$) and state point 4 ($\dot{m}^{ORC} \cdot h_4$) is converted into mechanical power in the turbine, W^{turb}, which starts to rotate. By means of the shaft work, the turbine operates a generator, which produces electric power (Borgnakke and Sonntag, 2009). The turbine work done by the system per unit of time is equivalent to the difference of the thermodynamic potentials of state points 3 and 4:

$$W^{turb} = \dot{m}^{ORC} \cdot \left(h_3^{ORC} - h_4^{ORC}\right) = \dot{m}^{ORC} \cdot \eta^{turb} \cdot \left(h_3^{ORC} - h_{4s}^{ORC}\right). \tag{3}$$

  As the pump, the turbine does not work without losses. Turbine losses mainly result from the flow of the organic fluid through the turbine blades and passages. In addition, heat dissipation and turbine governing procedures reduce the efficiency of the turbine, η^{turb} (Borgnakke and Sonntag, 2009). Thus, the change of state is not isentropic, which is expressed by $\eta^{turb} = \left(h_3^{ORC} - h_4^{ORC}\right) / \left(h_3^{ORC} - h_{4s}^{ORC}\right) < 1$.
- *State point 4 to state point 1:* In the condenser, the low-pressure vapor is condensed into a liquid again by withdrawing heat from the system ($\dot{Q}^{out}$). Wei et al. (2007) modeled this process by considering the power required for fans which are used to cool the condenser. However, as the impact of this change of state is small compared to the impacts of the other three transitions, it is neglected in this paper.

The overall efficiency of the ORC can now be described as follows:

$$\eta^{ORC} = \frac{W^{turb} - W^{pump}}{\dot{Q}^{WH}} = \frac{\eta^{turb} \cdot \left(h_3^{ORC} - h_{4s}^{ORC}\right) - \frac{h_{2s}^{ORC} - h_1^{ORC}}{\eta^{pump}}}{h_3^{ORC} - \left(\frac{h_{2s}^{ORC} - h_1^{ORC}}{\eta^{pump}} + h_1^{ORC}\right)}. \tag{4}$$

We assume in this paper that waste heat is only rejected when the production system is in one of the three operating modes, i.e., when it produces, when it is idle, or when it is being set up. When the production system is down (switched off) and does not consume any energy, no waste heat is rejected. Clearly, the amount of waste heat recovered in the ORC depends on the amount of energy that was initially inserted into the system. Thus, it is necessary to analyze the power required during the different operating modes of the production stages.

According to Gutowski et al. (2006), the power required during production by production stage i depends on a constant and on the current production rate:

$$\dot{EC}_{i,P} = W_i + k_i \cdot p_i, \forall i \in \{1,2\}, \tag{5}$$

where W_i is the power required to keep the machine in an operating mode, and $k_i \cdot p_i$ is the power required to perform the operations. Thus, if the machine is idle and no operations are performed, the power requirement reduces to (see also Zanoni et al., 2014)

$$\dot{EC}_{i,I} = W_i, \forall i \in \{1,2\}. \tag{6}$$

The power that is required for setting up a production stage can also be modeled by referring to W_i, which has to be multiplied with the factor f to take account of additional energy requirements during the setup:

$$\dot{EC}_{i,S} = f \cdot W_i, \forall i \in \{1,2\}. \tag{7}$$

To increase the efficiency of energy usage in this system, the WHRS described above is installed. The flow rate of the waste heat recovered, $\dot{HR}_{i,j}$, is equivalent to the power output of the ORC. The power output results from the difference of the work done *by* the system per unit of time (W^{turb}), and the work done *on* the system per unit of time (W^{pump}) (Borgnakke and Sonntag, 2009). Thus, the flow rate of the waste heat recovered at production stage i in operating mode j corresponds to

$$\begin{aligned} \dot{HR}_{i,j} &= W^{turb} - W^{pump} \\ &= \dot{m}_{i,j}^{ORC} \cdot \left(\eta^{turb} \cdot \left(h_3^{ORC} - h_{4S}^{ORC}\right) - \frac{h_{2S}^{ORC} - h_1^{ORC}}{\eta^{pump}}\right), \forall i \in \{1,2\}, j \in \{P, I, S\}. \end{aligned} \tag{8}$$

Solving Equation (2) for $\dot{m}_{i,j}^{ORC}$ and inserting it into Equation (8) links the flow rate of the waste heat recovered by the ORC to the flow rate of the waste heat added to the ORC in the evaporator:

$$\dot{HR}_{i,j} = \frac{\dot{Q}_{i,j}^{WH} \cdot \left(\eta^{turb} \cdot \left(h_3^{ORC} - h_{4S}^{ORC}\right) - \frac{h_{2S}^{ORC} - h_1^{ORC}}{\eta^{pump}}\right)}{h_3^{ORC} - \left(\frac{h_{2S}^{ORC} - h_1^{ORC}}{\eta^{pump}} + h_1^{ORC}\right)}, \forall i \in \{1,2\}, j \in \{P, I, S\}. \tag{9}$$

We assume that the flow rate of the waste heat rejected in operating mode j by production stage i is proportional to the power required in operating mode j by production stage i: $\dot{Q}_{i,j}^{WH} = \alpha \cdot \dot{EC}_{i,j}$. Thus, the flow rate of the waste heat recovered by the ORC depends on the power initially required by the production system:

$$\dot{HR}_{i,j} = \frac{\alpha \cdot \dot{EC}_{i,j} \cdot \left(\eta^{turb} \cdot \left(h_3^{ORC} - h_{4S}^{ORC}\right) - \frac{h_{2S}^{ORC} - h_1^{ORC}}{\eta^{pump}}\right)}{h_3^{ORC} - \left(\frac{h_{2S}^{ORC} - h_1^{ORC}}{\eta^{pump}} + h_1^{ORC}\right)}, \forall i \in \{1,2\}, j \in \{P, I, S\}. \tag{10}$$

At first, WHRSs were primarily used in combination with continuous production processes as in this case the system can be centralized and easily be governed, especially because of the continuous waste heat flow. A typical example for this kind of production process is the cement clinker production (Karellas et al., 2013). As the paper at hand is concerned with lot size production processes – and consequently with different waste heat streams on different production stages in different operating modes – decentralized ORC modules immediately attached to the respective production stages seem more adequate for the scenario considered here than one (large) centralized WHRS to be used for recovering the waste heat of all production stages at once. In practice, there already are decentralized ORC modules that can be used for recovering waste heat during production. Large companies such as Siemens, General Electric, ABB, and Dürr Cyplan offer large systems for recovering waste heat, while some smaller players, such as Orcan Energy, recently started offering small and flexible modules, which can be installed immediately next to a production stage.

In order to work smoothly and to produce a constant output of energy, an ORC needs to reach a steady state (Dai et al., 2009), which means that the system operating conditions have to be constant with only minor changes for a sufficiently long period of time. For the case studied in this paper, this means that the production system has to remain in one operating mode for a sufficiently long period of time to ensure that the ORC reaches a steady state. If the operating modes were constantly changed during a short period of time, the turbine governing procedures would negatively impact the efficiency of the turbine and of the overall ORC, such that the efficiency of the WHRS would be significantly reduced. In light of this aspect, it is assumed hereafter that the periods in which the production stages remain in one of the three operating modes are always sufficiently long to guarantee that the ORC reaches a steady state. Under this assumption, it is technologically reasonable to make the transformation from Equation (8) to Equation (10), i.e., the direct link between running the production stages and immediately recovering waste heat from the energy consumed is technologically sound.

Another technological characteristic that needs to be addressed is the partial load behavior of the ORC, i.e., the way the performance of the ORC is affected by a change in the heat flow provided to the ORC between state points 2 and 3. In the application studied here, the question is how the efficiency of the WHRS is affected by differences in the flow of the waste heat, $\dot{Q}^{WH}$, associated with the different operating modes since we assumed that the waste heat rejected in operating mode j is proportional to the energy consumed in operating mode j. Obernberger et al. (2002) showed that for an ORC with an efficiency of 18 percent at nominal load, the efficiency dropped only by 1.5 percentage points at a partial load of 50 percent. This characteristic ensures that the ORC fits perfectly to the production system studied here as the energy consumption (and consequently the waste heat flows) varies across the operating modes. Because of the robustness of the ORC against changes in partial loads, it is technologically reasonable to use the same WHRS in the different operating modes.

4 The integrated model

The model developed in this paper integrates the WHRS presented in Section 3 into the two-stage production system illustrated in Section 2 and, thus, links the product flow to the energy flow. As a result, the total cost of the integrated system incorporates traditional production-inventory cost (inventory holding cost of the three buffer stocks, setup cost of both production stages, cost of shipping batches between the first and the second buffer), $tPIC$, energy-related cost for the different operating modes, ERC, and the cost of generating electricity from the ORC, OC. The integrated model supports decisions on the optimal usage of a WHRS and on the selection of operating modes during phases where the machines do not produce (idle operating mode vs. switching the machines off and on again). Depending on the selected production policy, the formulations of the cost components differ slightly.

4.1 Continuous batch production policy

As in traditional EPQ models, inventory holding cost represent the average cost of capital tied up in inventories. Inventory carrying cost of buffer stocks 1 and 2 as well as of the final buffer stock 3 are calculated as follows:

$$HC_{1,2}^{CON} = \left(\frac{1}{p_1} + \frac{1}{p_2} + (m-1) \cdot \left(\frac{1}{p_2} - \frac{1}{p_1}\right)\right) \cdot \frac{d \cdot Q}{2 \cdot m} \cdot hc_{1,2}, \tag{11}$$

$$HC_3^{CON} = (p_2 - d) \cdot \frac{Q}{2 \cdot p_2} \cdot hc_3 = \left(\frac{1}{d} - \frac{1}{p_2}\right) \cdot \frac{d \cdot Q}{2} \cdot hc_3. \tag{12}$$

At the beginning of each new production run, both production stages need to be set up. The average setup cost are calculated as

$$SUC^{CON} = \frac{A \cdot d}{Q}. \tag{13}$$

After one production batch of size Q/m has been completed at production stage 1, it needs to be shipped from the first buffer stock to the second buffer stock. Hence, the average shipment cost incurred amounts to

$$SHC^{CON} = \frac{SC \cdot m \cdot d}{Q}. \tag{14}$$

The energy cost varies across the three operating modes. During production, the average energy-related cost is derived from the difference between Equations (5) and (10):

$$\begin{aligned} ERC_{i,P}^{CON} &= \left(\dot{EC}_{i,P} - \dot{HR}_{i,P}\right) \cdot \frac{Q}{p_i} \cdot \frac{d}{Q} \cdot e \\ &= \left(W_i + k_i \cdot p_i - \frac{\alpha \cdot (W_i + k_i \cdot p_i) \cdot \left(\eta^{turb} \cdot \left(h_3^{ORC} - h_{4s}^{ORC}\right) - \frac{h_{2s}^{ORC} - h_1^{ORC}}{\eta^{pump}}\right)}{h_3^{ORC} - \left(\frac{h_{2s}^{ORC} - h_1^{ORC}}{\eta^{pump}} + h_1^{ORC}\right)} \right) \cdot \frac{d}{p_i} \cdot e \\ &= \left(\frac{W_i}{p_i} + k_i\right) \cdot \left(1 - \alpha \cdot \eta^{ORC}\right) \cdot d \cdot e, \forall\, i \in \{1,2\}. \end{aligned} \tag{15}$$

In the idle operating mode, the average energy-related cost corresponds to the difference between Equations (6) and (10):

$$\begin{aligned} ERC_{i,I}^{CON} &= \left(\dot{EC}_{i,I} - \dot{HR}_{i,I}\right) \cdot B_i \cdot \frac{d}{Q} \cdot \left(\frac{Q}{d} - \frac{Q}{p_i}\right) \cdot e \\ &= \left(W_i - \frac{\alpha \cdot W_i \cdot \left(\eta^{turb} \cdot \left(h_3^{ORC} - h_{4s}^{ORC}\right) - \frac{h_{2s}^{ORC} - h_1^{ORC}}{\eta^{pump}}\right)}{h_3^{ORC} - \left(\frac{h_{2s}^{ORC} - h_1^{ORC}}{\eta^{pump}} + h_1^{ORC}\right)} \right) \cdot B_i \cdot \left(\frac{1}{d} - \frac{1}{p_i}\right) \cdot d \cdot e \\ &= W_i \cdot \left(1 - \alpha \cdot \eta^{ORC}\right) \cdot B_i \cdot \left(\frac{1}{d} - \frac{1}{p_i}\right) \cdot d \cdot e, \forall\, i \in \{1,2\}. \end{aligned} \tag{16}$$

For setting up a production stage, the average energy-related cost equals the difference between Equations (7) and (10):

$$\begin{aligned} ERC_{i,S}^{CON} &= \left(\dot{EC}_{i,S} - \dot{HR}_{i,I}\right) \cdot (1 - B_i) \cdot \frac{d}{Q} \cdot t_i^{on} \cdot e \\ &= \left(f \cdot W_i - \frac{\alpha \cdot f \cdot W_i \cdot \left(\eta^{turb} \cdot \left(h_3^{ORC} - h_{4s}^{ORC}\right) - \frac{h_{2s}^{ORC} - h_1^{ORC}}{\eta^{pump}}\right)}{h_3^{ORC} - \left(\frac{h_{2s}^{ORC} - h_1^{ORC}}{\eta^{pump}} + h_1^{ORC}\right)} \right) \cdot (1 - B_i) \cdot \frac{d}{Q} \cdot t_i^{on} \cdot e \\ &= f \cdot W_i \cdot \left(1 - \alpha \cdot \eta^{ORC}\right) \cdot (1 - B_i) \cdot \frac{d}{Q} \cdot t_i^{on} \cdot e, \forall\, i \in \{1,2\}. \end{aligned} \tag{17}$$

This paper assumes that a certain period of time is required for setting up a production stage. Thus, switching off a production stage after producing one production lot and switching it on again when starting the production of the next production lot is only possible if its setup time, t_i^{on}, is equal to or shorter than the time between two successive production cycles, t_i^{cycle}. Hence, the following constraints need to be satisfied:

$$t_i^{cycle} = \frac{Q}{d} - \frac{Q}{p_i} \geq t_i^{on}, \forall\, i \in \{1,2\}. \tag{18}$$

Finally, the average cost of generating electricity from the ORC needs to be considered. This cost depends on the flow rate of the waste heat recovered:

$$\begin{aligned} OC_{i,P}^{CON} &= \dot{H}R_{i,P} \cdot \frac{Q}{p_i} \cdot \frac{d}{Q} \cdot c \\ &= \frac{\alpha \cdot (W_i + k_i \cdot p_i) \cdot \left(\eta^{turb} \cdot (h_3^{ORC} - h_{4s}^{ORC}) - \frac{h_{2s}^{ORC} - h_1^{ORC}}{\eta^{pump}}\right)}{h_3^{ORC} - \left(\frac{h_{2s}^{ORC} - h_1^{ORC}}{\eta^{pump}} + h_1^{ORC}\right)} \cdot \frac{d}{p_i} \cdot c \\ &= \left(\frac{W_i}{p_i} + k_i\right) \cdot \alpha \cdot \eta^{ORC} \cdot d \cdot c, \forall\, i \in \{1,2\}, \end{aligned} \tag{19}$$

$$\begin{aligned} OC_{i,I}^{CON} &= \dot{H}R_{i,I} \cdot B_i \cdot \frac{d}{Q} \cdot \left(\frac{Q}{d} - \frac{Q}{p_i}\right) \cdot c \\ &= \frac{\alpha \cdot W_i \cdot \left(\eta^{turb} \cdot (h_3^{ORC} - h_{4s}^{ORC}) - \frac{h_{2s}^{ORC} - h_1^{ORC}}{\eta^{pump}}\right)}{h_3^{ORC} - \left(\frac{h_{2s}^{ORC} - h_1^{ORC}}{\eta^{pump}} + h_1^{ORC}\right)} \cdot B_i \cdot \left(\frac{1}{d} - \frac{1}{p_i}\right) \cdot d \cdot c \\ &= W_i \cdot \alpha \cdot \eta^{ORC} \cdot B_i \cdot \left(\frac{1}{d} - \frac{1}{p_i}\right) \cdot d \cdot c, \forall\, i \in \{1,2\}, \end{aligned} \tag{20}$$

$$\begin{aligned} OC_{i,S}^{CON} &= \dot{H}R_{i,S} \cdot (1 - B_i) \cdot \frac{d}{Q} \cdot t_i^{on} \cdot c \\ &= \frac{\alpha \cdot f \cdot W_i \cdot \left(\eta^{turb} \cdot (h_3^{ORC} - h_{4s}^{ORC}) - \frac{h_{2s}^{ORC} - h_1^{ORC}}{\eta^{pump}}\right)}{h_3^{ORC} - \left(\frac{h_{2s}^{ORC} - h_1^{ORC}}{\eta^{pump}} + h_1^{ORC}\right)} \cdot (1 - B_i) \cdot \frac{d}{Q} \cdot t_i^{on} \cdot c \\ &= f \cdot W_i \cdot \alpha \cdot \eta^{ORC} \cdot (1 - B_i) \cdot \frac{d}{Q} \cdot t_i^{on} \cdot c, \forall\, i \in \{1,2\}. \end{aligned} \tag{21}$$

Summing up the cost components derived above – inventory holding cost, $HC_{1,2}$ and, HC_3, setup cost, SUC, shipping cost, SHC, energy-related cost, ERC, and ORC-related cost, OC – leads to the average total cost of the production system for the continuous batch production policy:

$$\begin{aligned} TC^{CON}(Q, p_1, p_2, m) &= HC_{1,2}^{CON} + HC_3^{CON} + SUC^{CON} + SHC^{CON} \\ &\quad + \sum_{i\in\{1,2\}} \sum_{j\in\{P,I,S\}} ERC_{i,j}^{CON} + \sum_{i\in\{1,2\}} \sum_{j\in\{P,I,S\}} OC_{i,j}^{CON} \\ &= \left(\frac{1}{p_1} + \frac{1}{p_2} + (m-1) \cdot \left(\frac{1}{p_2} - \frac{1}{p_1}\right)\right) \cdot \frac{d \cdot Q}{2 \cdot m} \cdot hc_{1,2} \\ &\quad + \left(\frac{1}{d} - \frac{1}{p_2}\right) \cdot \frac{d \cdot Q}{2} \cdot hc_3 + \frac{A \cdot d}{Q} + \frac{SC \cdot m \cdot d}{Q} \\ &\quad + \sum_{i\in\{1,2\}} \left(\frac{W_i}{p_i} + k_i + W_i \cdot B_i \cdot \left(\frac{1}{d} - \frac{1}{p_i}\right) + f \cdot W_i \cdot (1 - B_i) \cdot \frac{t_i^{on}}{Q}\right) \\ &\quad \cdot d \cdot \left((1 - \alpha \cdot \eta^{ORC}) \cdot e + \alpha \cdot \eta^{ORC} \cdot c\right). \end{aligned} \tag{22}$$

4.2 Interrupted batch production policy

If the production process of the first production stage is interrupted after each batch, the inventory holding cost for buffer stocks 1 and 2, $HC_{1,2}$, as well as the energy-related cost, $ERC_{1,I}$, and the ORC-related operating cost, $OC_{1,I}$, in the idle operating mode change compared to the case of a continuous batch production policy.

The average inventory holding cost for buffer stocks 1 and 2 can now be calculated as

$$HC_{1,2}^{INT} = \left(\frac{1}{p_1} + \frac{1}{p_2}\right) \cdot \frac{d \cdot Q}{2 \cdot m} \cdot hc_{1,2}. \tag{23}$$

The average energy-related cost in the idle operating mode of production stage 1 equals

$$\begin{aligned} ERC_{1,I}^{INT} &= \left(\dot{EC}_{1,I} - \dot{HR}_{1,I}\right) \cdot \frac{d}{Q} \cdot \left((m - 1 + B_1) \cdot \frac{Q}{m} \cdot \left(\frac{1}{p_2} - \frac{1}{p_1}\right) + B_1 \cdot Q \cdot \left(\frac{1}{d} - \frac{1}{p_2}\right) \right) \cdot e \\ &= \left(W_1 - \frac{\alpha \cdot W_1 \cdot \left(\eta^{turb} \cdot \left(h_3^{ORC} - h_{4s}^{ORC}\right) - \frac{h_{2s}^{ORC} - h_1^{ORC}}{\eta^{pump}}\right)}{h_3^{ORC} - \left(\frac{h_{2s}^{ORC} - h_1^{ORC}}{\eta^{pump}} + h_1^{ORC}\right)} \right) \\ &\quad \cdot \left(\frac{m-1+B_1}{m} \cdot \left(\frac{1}{p_2} - \frac{1}{p_1}\right) + B_1 \cdot \left(\frac{1}{d} - \frac{1}{p_2}\right) \right) \cdot d \cdot e \\ &= W_1 \cdot (1 - \alpha \cdot \eta^{ORC}) \cdot \left(\frac{m-1+B_1}{m} \cdot \left(\frac{1}{p_2} - \frac{1}{p_1}\right) + B_1 \cdot \left(\frac{1}{d} - \frac{1}{p_2}\right) \right) \cdot d \cdot e. \end{aligned} \tag{24}$$

The average cost of generating electricity from the ORC in the idle operating mode on production stage 1 amounts to

$$\begin{aligned} OC_{1,I}^{INT} &= \dot{HR}_{1,I} \cdot \frac{d}{Q} \cdot \left((m - 1 + B_1) \cdot \frac{Q}{m} \cdot \left(\frac{1}{p_2} - \frac{1}{p_1}\right) + B_1 \cdot \left(Q \cdot \left(\frac{1}{d} - \frac{1}{p_2}\right) \right) \right) \cdot c \\ &= \frac{\alpha \cdot W_1 \cdot \left(\eta^{turb} \cdot \left(h_3^{ORC} - h_{4s}^{ORC}\right) - \frac{h_{2s}^{ORC} - h_1^{ORC}}{\eta^{pump}}\right)}{h_3^{ORC} - \left(\frac{h_{2s}^{ORC} - h_1^{ORC}}{\eta^{pump}} + h_1^{ORC}\right)} \cdot \left(\frac{m-1+B_1}{m} \cdot \left(\frac{1}{p_2} - \frac{1}{p_1}\right) + B_1 \cdot \left(\frac{1}{d} - \frac{1}{p_2}\right) \right) \cdot d \cdot c \\ &= W_1 \cdot \alpha \cdot \eta^{ORC} \left(\frac{m-1+B_1}{m} \cdot \left(\frac{1}{p_2} - \frac{1}{p_1}\right) + B_1 \cdot \left(\frac{1}{d} - \frac{1}{p_2}\right) \right) \cdot d \cdot c. \end{aligned} \tag{25}$$

Thus, the average total cost of the system for the interrupted batch production policy equals

$$\begin{aligned} TC^{INT}(Q, p_1, p_2, m) &= HC_{1,2}^{INT} + HC_3^{INT} + SUC^{INT} + SHC^{INT} \\ &= + \sum_{i \in \{1,2\}} \sum_{j \in \{P,I,S\}} ERC_{i,j}^{INT} + \sum_{i \in \{1,2\}} \sum_{j \in \{P,I,S\}} OC_{i,j}^{INT} \\ &= \left(\frac{1}{p_1} + \frac{1}{p_2}\right) \cdot \frac{d \cdot Q}{2 \cdot m} \cdot hc_{1,2} + \left(\frac{1}{d} - \frac{1}{p_2}\right) \cdot \frac{d \cdot Q}{2} \cdot hc_3 + \frac{A \cdot d}{Q} + \frac{SC \cdot m \cdot d}{Q} \\ &\quad + \left(\begin{array}{c} \sum_{i \in \{1,2\}} \left(\frac{W_i}{p_i} + k_i + W_i \cdot B_i \left(\frac{1}{d} - \frac{1}{p_2}\right) \right) \\ + W_1 \cdot \frac{m-1+B_1}{m} \cdot \left(\frac{1}{p_2} - \frac{1}{p_1}\right) + \sum_{i \in \{1,2\}} f \cdot W_i \cdot (1 - B_i) \cdot \frac{t_i^{on}}{Q} \end{array} \right) \\ &\quad \cdot d \cdot \left((1 - \alpha \cdot \eta^{ORC}) \cdot e + \alpha \cdot \eta^{ORC} \cdot c \right). \end{aligned} \tag{26}$$

The constraints formulated in Inequation (18) remain valid for the interrupted batch production policy as well.

5 Solution of the models

This section derives optimal values for the lot size Q, the production rates p_1 and p_2, and the number of shipments m that minimize the total cost functions (22) and (26), respectively. The optimization problem for both production policies ($l \in \{CON, INT\}$) can be formulated as follows:

$$\text{Minimize} \qquad TC^l(Q, p_1, p_2, m) \tag{27}$$

$$\text{Subject to:} \qquad t_i^{cycle} \geq t_i^{on}, \forall i \in \{1,2\} \tag{28}$$

$$p_1 \geq p_2 \tag{29}$$

$$p_2 \geq d \tag{30}$$

$$Q > 0, m \in \mathbb{N}, p_i^{min} \leq p_i \leq p_i^{max}, \forall i \in \{1,2\} \tag{31}$$

Theorem 1. *If the constraints are neglected in a first step, it can easily be shown that Equations* (22) and (26) *are convex in Q for given values for p_1, p_2, and m. Thus, the optimal solutions for $Q^{CON,opt}$ and $Q^{INT,opt}$ equal*

$$Q^{CON,opt} = \left(\frac{d \cdot \left(A + SC \cdot m + \sum_{i \in \{1,2\}} f \cdot W_i \cdot (1 - B_i) \cdot t_i^{on} \cdot \left((1 - \alpha \cdot \eta^{ORC}) \cdot e + \alpha \cdot \eta^{ORC} \cdot c\right)\right)}{\left(\frac{1}{p_1} + \frac{1}{p_2} + (m-1) \cdot \left(\frac{1}{p_2} - \frac{1}{p_1}\right)\right) \cdot \frac{d \cdot hc_{1,2}}{2 \cdot m} + \left(\frac{1}{d} - \frac{1}{p_2}\right) \cdot \frac{d \cdot hc_3}{2}} \right)^{\frac{1}{2}}, \tag{32}$$

$$Q^{INT,opt} = \left(\frac{d \cdot \left(A + SC \cdot m + \sum_{i \in \{1,2\}} f \cdot W_i \cdot (1 - B_i) \cdot t_i^{on} \cdot \left((1 - \alpha \cdot \eta^{ORC}) \cdot e + \alpha \cdot \eta^{ORC} \cdot c\right)\right)}{\left(\frac{1}{p_1} + \frac{1}{p_2}\right) \cdot \frac{d \cdot hc_{1,2}}{2 \cdot m} + \left(\frac{1}{d} - \frac{1}{p_2}\right) \cdot \frac{d \cdot hc_3}{2}} \right)^{\frac{1}{2}}. \tag{33}$$

Proof. This follows from setting the first partial derivatives of Equations (22) and (26) with respect to Q equal to 0 and solving for $Q^{CON,opt}$ and $Q^{INT,opt}$, respectively. Convexity follows from the second partial derivatives of Equations (22) and (26), which are non-negative. ■

Substituting Equations (32) and (33) into Equations (22) and (26), respectively, yields the average total cost of the system for both production policies, which only depends on the decision variables p_1, p_2, and m:

$$\begin{aligned} TC^{CON,opt}(p_1, p_2, m) = d \cdot & \left(2 \cdot \begin{pmatrix} A + SC \cdot m + \sum_{i \in \{1,2\}} f \cdot W_i \cdot (1 - B_i) \cdot t_i^{on} \\ \cdot \left((1 - \alpha \cdot \eta^{ORC}) \cdot e + \alpha \cdot \eta^{ORC} \cdot c\right) \end{pmatrix} \right)^{\frac{1}{2}} \\ & \cdot \left(\left(\frac{1}{p_1} + \frac{1}{p_2} + (m-1) \cdot \left(\frac{1}{p_2} - \frac{1}{p_1}\right)\right) \cdot \frac{hc_{1,2}}{m} + \left(\frac{1}{d} - \frac{1}{p_2}\right) \cdot hc_3 \right)^{\frac{1}{2}} \\ & + \sum_{i \in \{1,2\}} \left(\frac{W_i}{p_i} + k_i + W_i \cdot B_i \cdot \left(\frac{1}{d} - \frac{1}{p_i}\right) \right) \\ & \cdot d \cdot \left((1 - \alpha \cdot \eta^{ORC}) \cdot e + \alpha \cdot \eta^{ORC} \cdot c\right), \end{aligned} \tag{34}$$

$$
\begin{aligned}
TC^{INT,opt}(p_1,p_2,m) = {} & d\cdot\left(2\cdot\begin{pmatrix} A+SC\cdot m+\sum_{i\in\{1,2\}} f\cdot W_i\cdot(1-B_i)\cdot t_i^{on} \\ \cdot\left((1-\alpha\cdot\eta^{ORC})\cdot e+\alpha\cdot\eta^{ORC}\cdot c\right)\end{pmatrix}\right)^{\frac{1}{2}} \\
& \cdot\left(\left(\frac{1}{p_1}+\frac{1}{p_2}\right)\cdot\frac{hc_{1,2}}{2\cdot m}+\left(\frac{1}{d}-\frac{1}{p_2}\right)\cdot hc_3\right)^{\frac{1}{2}} \\
& +\begin{pmatrix}\sum_{i\in\{1,2\}}\left(\frac{W_i}{p_i}+k_i+W_i\cdot B_i\cdot\left(\frac{1}{d}-\frac{1}{p_2}\right)\right) \\ +W_1\cdot\frac{m-1+B_1}{m}\cdot\left(\frac{1}{p_2}-\frac{1}{p_1}\right)\end{pmatrix} \\
& \cdot d\cdot\left((1-\alpha\cdot\eta^{ORC})\cdot e+\alpha\cdot\eta^{ORC}\cdot c\right).
\end{aligned}
\tag{35}
$$

Theorem 2. *If the integrality constraint on m is neglected, it can be shown that Equations* (22) and (26) *are quasi-convex in m for positive values of m and given values for p_1, p_2, and Q.*

Proof. Suppose M is a convex set in $\mathbb{R}^n$ and ξ_1 and ξ_2 are real-valued functions on M. Then a function $\Phi(m)=\xi_1(m)/\xi_2(m)$ is quasi-convex on M if ξ_1 is convex and ξ_2 is positive and linear on M (Avriel, 2003). To use this lemma, we reformulate Equation (22) as

$$
\begin{aligned}
TC^{CON}(m) = {} & \frac{\frac{SC\cdot d}{Q}\cdot m^2}{m}+\frac{d\cdot\left(\left(\frac{1}{p_2}-\frac{1}{p_1}\right)\cdot\frac{Q\cdot hc_{1,2}}{2}+\left(\frac{1}{d}-\frac{1}{p_2}\right)\cdot\frac{Q\cdot hc_3}{2}+\frac{A}{Q}\right)\cdot m}{m} \\
& +\frac{d\cdot\left(\sum_{i\in\{1,2\}}\left(\frac{W_i}{p_i}+k_i+W_i\cdot B_i\cdot\left(\frac{1}{d}-\frac{1}{p_i}\right)+f\cdot W_i\cdot(1-B_i)\cdot\frac{t_i^{on}}{Q}\right)\right)\cdot\left((1-\alpha\cdot\eta^{ORC})\cdot e+\alpha\cdot\eta^{ORC}\cdot c\right)\cdot m}{m} \\
& +\frac{\frac{d\cdot Q\cdot hc_{1,2}}{p_1}}{m},
\end{aligned}
\tag{36}
$$

and Equation (26) as

$$
\begin{aligned}
TC^{INT}(m) = {} & \frac{\frac{SC\cdot d}{Q}\cdot m^2}{m}+\frac{d\cdot\left(\left(\frac{1}{d}-\frac{1}{p_2}\right)\cdot\frac{Q\cdot hc_3}{2}+\frac{A}{Q}\right)\cdot m}{m} \\
& +\frac{d\cdot\begin{pmatrix}\sum_{i\in\{1,2\}}\left(\frac{W_i}{p_i}+k_i+W_i\cdot B_i\cdot\left(\frac{1}{d}-\frac{1}{p_2}\right)\right)+W_1\cdot\left(\frac{1}{p_2}-\frac{1}{p_1}\right) \\ +\sum_{i\in\{1,2\}} f\cdot W_i\cdot(1-B_i)\cdot\frac{t_i^{on}}{Q}\end{pmatrix}\cdot\left((1-\alpha\cdot\eta^{ORC})\cdot e+\alpha\cdot\eta^{ORC}\cdot c\right)\cdot m}{m} \\
& +\frac{d\cdot\left(\left(\frac{1}{p_1}+\frac{1}{p_2}\right)\cdot\frac{Q\cdot hc_{1,2}}{2}+W_1\cdot(B_1-1)\cdot\left(\frac{1}{p_2}-\frac{1}{p_1}\right)\cdot\left((1-\alpha\cdot\eta^{ORC})\cdot e+\alpha\cdot\eta^{ORC}\cdot c\right)\right)}{m}.
\end{aligned}
\tag{37}
$$

We set

$$
\begin{aligned}
\xi_1^{CON}(m) = {} & \frac{SC\cdot d}{Q}\cdot m^2+d\cdot\left(\left(\frac{1}{p_2}-\frac{1}{p_1}\right)\cdot\frac{Q\cdot hc_{1,2}}{2}+\left(\frac{1}{d}-\frac{1}{p_2}\right)\cdot\frac{Q\cdot hc_3}{2}+\frac{A}{Q}\right)\cdot m \\
& +d\cdot\left(\sum_{i\in\{1,2\}}\left(\frac{W_i}{p_i}+k_i+W_i\cdot B_i\cdot\left(\frac{1}{d}-\frac{1}{p_i}\right)+f\cdot W_i\cdot(1-B_i)\cdot\frac{t_i^{on}}{Q}\right)\right) \\
& \cdot\left((1-\alpha\cdot\eta^{ORC})\cdot e+\alpha\cdot\eta^{ORC}\cdot c\right)\cdot m+\frac{d\cdot Q\cdot hc_{1,2}}{p_1},
\end{aligned}
\tag{38}
$$

$$
\xi_2^{CON}(m) = m,
\tag{39}
$$

$$\xi_1^{INT}(m) = \frac{SC\cdot d}{Q}\cdot m^2 + d\cdot\left(\left(\frac{1}{d}-\frac{1}{p_2}\right)\cdot\frac{Q\cdot hc_3}{2}+\frac{A}{Q}\right)\cdot m$$
$$+d\cdot\left(\begin{array}{c}\sum_{i\in\{1,2\}}\left(\frac{W_i}{p_i}+k_i+W_i\cdot B_i\cdot\left(\frac{1}{d}-\frac{1}{p_2}\right)\right)+W_1\cdot\left(\frac{1}{p_2}-\frac{1}{p_1}\right)\\ +\sum_{i\in\{1,2\}}f\cdot W_i\cdot(1-B_i)\cdot\frac{t_i^{on}}{Q}\end{array}\right)$$
$$\cdot\left((1-\alpha\cdot\eta^{ORC})\cdot e+\alpha\cdot\eta^{ORC}\cdot c\right)\cdot m \tag{40}$$
$$+d\cdot\left(\begin{array}{c}\left(\frac{1}{p_1}+\frac{1}{p_2}\right)\cdot\frac{Q\cdot hc_{1,2}}{2}+\left(W_1\cdot(B_1-1)\cdot\left(\frac{1}{p_2}-\frac{1}{p_1}\right)\right)\\ \cdot\left((1-\alpha\cdot\eta^{ORC})\cdot e+\alpha\cdot\eta^{ORC}\cdot c\right)\end{array}\right),$$

$$\xi_2^{INT}(m) = m. \tag{41}$$

It can easily be shown that $\xi_1^{CON}(m)$ and $\xi_1^{INT}(m)$ are convex in m. On the other hand, $\xi_2^{CON}(m)$ and $\xi_2^{INT}(m)$ are positive and linear for $m > 0$. Thus, we conclude that Equations (22) and (26) are quasi-convex in $m > 0$ for given p_1, p_2, and Q. ∎

As a result of Theorem 2, and as every local minimum of a quasi-convex function $\Phi(m)$ is a global minimum unless $\Phi(m)$ is constant in a neighborhood of the local minimum (Greenberg and Pierskalla, 1971), an optimal solution for m may be calculated for given values of Q, p_1, and p_2 by increasing m stepwise from $m = 1$ until the following optimality condition holds:

$$TC^l(m^{opt}-1) \geq TC^l(m^{opt}) \leq TC^l(m^{opt}+1), \forall l \in \{CON, INT\}. \tag{42}$$

A comprehensive analytical investigation of Equations (34) and (35) with respect to p_1 and p_2 and a given value for m can be found in Appendices A and B, respectively. We show that if p_1^{opt} and p_2^{opt} influence the optimal solution, p_1^{opt} and p_2^{opt} will always be equal to p_1^{min} or p_1^{max} and p_2^{min} or p_2^{max}, respectively, depending on m as well as on the respective parameter values. Hence, only four combinations of the tuple (p_1, p_2) may be part of the optimal solution to the problem stated in Equations (27) to (31): (p_1^{max}, p_2^{max}), $(p_1^{min}, \min\{p_1^{min}, p_2^{max}\})$, (p_1^{max}, p_2^{min}), and (p_1^{min}, p_2^{min}). As a consequence, we can insert these four tuples one after another into Equations (34) and (35) and determine the number of shipments m that minimizes Equations (34) and (35), respectively, for each tuple. Subsequently, the optimal production lot size Q can be calculated with the help of Equations (32) and (33), respectively. If the resulting values for $Q^{CON,opt}$ and $Q^{INT,opt}$ violate the constraints defined in Inequation (18), the values for $Q^{CON,opt}$ and $Q^{INT,opt}$ need to be adjusted. Thus, $Q^{CON,opt}$ and $Q^{INT,opt}$ are increased in such a way that the number of times the production stages are switched off and on again between successive production cycles is reduced. Values for $Q^{CON,opt}$ and $Q^{INT,opt}$ are then calculated by treating Inequation (18) as an equation. Finally, by comparing the total cost associated with each of the four solutions obtained for the four possible combinations of the tuple (p_1, p_2), the optimal solution to the problem stated in Equations (27) to (31) can be identified. This procedure is summarized in the following algorithm:

Step 1: Set $P = \{(p_1^{max}, p_2^{max}), (p_1^{min}, \min\{p_1^{min}, p_2^{max}\}), (p_1^{max}, p_2^{min}), (p_1^{min}, p_2^{min})\}$ and $TC^{l,opt} = \infty$.

Step 2: Set (p_1, p_2) to the first tuple of P and delete this tuple from P.

Step 3: Set $m = 1$ and $TC^{help} = \infty$.

Step 4: Calculate TC^l from Equation (34) or (35), depending on whether $l = CON$ or $l = INT$.

Step 5: Calculate Q^l from Equation (32) or (33), depending on whether $l = CON$ or $l = INT$.

Step 6: If $Q^l \geq \max\left\{\frac{t_1^{on}}{\frac{1}{d}-\frac{1}{p_1}}, \frac{t_2^{on}}{\frac{1}{d}-\frac{1}{p_2}}\right\}$, go to *Step 7*. Else, set $Q^l = \max\left\{\frac{t_1^{on}}{\frac{1}{d}-\frac{1}{p_1}}, \frac{t_2^{on}}{\frac{1}{d}-\frac{1}{p_2}}\right\}$ and calculate TC^l from Equation (22) and (26), depending on whether $l = CON$ or $l = INT$.

Step 7: If $TC^l < TC^{help}$, set $m^{help} = m$, $p_1^{help} = p_1$, $p_2^{help} = p_2$, $Q^{help} = Q^l$, and $TC^{help} = TC^l$; set $m = m + 1$ and go to *Step 4*. Else, go to *Step 8*.

Step 8: If $TC^{help} < TC^{l,opt}$, set $m^{opt} = m^{help}$, $p_1^{opt} = p_1^{help}$, $p_2^{opt} = p_2^{help}$, $Q^{l,opt} = Q^{help}$, $TC^{l,opt} = TC^{help}$, and go to *Step 9*. Else, go to *Step 9*.

Step 9: If $P = \emptyset$, go to *Step 10*. Else, go to *Step 2*.

Step 10: End.

When the algorithm terminates, the optimal solution to the problem formulated in Equations (27) to (31) is given by $Q^{l,opt}$, m^{opt}, p_1^{opt}, and p_2^{opt}.

6 Numerical analysis

This section illustrates the behavior of the model with the help of a numerical example. The data characterizing the production and energy flow of the production system studied as well as the associated cost parameters (d, $hc_{1,2}$, hc_3, A, SC, p_i, f, W_i, k_i, e, t_i^{on}) are taken from a company from the manufacturing sector which uses two cutting processes in sequence (Zanoni et al., 2014). The cutting processes assumed here can, for example, be found in the automotive industry. As to the selection of working fluids to be used in ORCs, several studies have been published in the past that help to select an appropriate fluid for an application (e.g., Hung et al., 1997; Hung, 2001; Hung et al., 2010; Liu et al., 2004; Wang et al., 2011; Dai et al., 2009). In the numerical example presented in this study, Ammonia was used as working fluid, and the characteristics of the ORC (h_r, η^{pump}, η^{turb}) as well as of the waste heat were formulated as in Dai et al. (2009), who carried out a performance simulation study on the use of different organic fluids to be utilized in an ORC. The cost of generating electricity from the ORC, c, is taken from Tchanche et al. (2010), who conducted an economic feasibility study on the use of small-scale ORCs for recovering waste heat. The numerical value that we used for the percentage of the initial energy input rejected as waste heat and led into the WHRS, α, represents a compromise value of various assessments of the fraction of the waste heat that can be used in a technologically and economically useful way (e.g., Schaefer, 1995; López et al., 1998; McKenna and Norman, 2010). We point out that the WHRS used here is not necessarily tailored to the production system of the two cutting processes. Typical applications of WHRSs can usually be found in the cement, the steel working, or the glass industry (Invernizzi, 2013). Yet, as this paper is an extension of the work of Zanoni et al. (2014), we decided to investigate the same production system as Zanoni et al. did for reasons of comparability. Hence, the primary aims of this numerical analysis are (I) to study the behavior of the model, (II) to investigate how sensitive it is to changes in the input parameters, (III) to examine whether the introduction of a WHRS in a real-life industry application is economically reasonable at the moment, and if not (IV) to identify technological advancements that would be necessary to justify investments in WHRSs in the future. The data used for the numerical analysis are summarized in Table 2. As explained above, we did not study the combined production and waste heat recovery system as a whole in practice. Yet, all the data were collected from real-life applications.

Table 2: Data characterizing the production system and the WHRS of the numerical analysis.

d	100 kg/h	W_1	100 kW	h_1	314.16 kJ/kg
$hc_{1,2}$	0.04 EUR/(kg·h)	W_2	50 kW	h_{2s}	318.97 kJ/kg
hc_3	0.05 EUR/(kg·h)	k_1	0.2 kWh/kg	h_3	1692.15 kJ/kg
A	100 EUR/setup	k_2	0.05 kWh/kg	h_{4s}	1487.72 kJ/kg
SC	20 EUR/shipment	e	0.2 EUR/kWh	η^{pump}	0.6
p_1	[150,300] kg/h	t_1^{on}	0.1 h	η^{turb}	0.85
p_2	[120,280] kg/h	t_2^{on}	0.1 h	c	0.07 EUR/kWh
f	2	α	0.18		

Table 3 provides an overview of the results of the analysis for the scenario studied.

Table 3: Results of the numerical analysis for different production policies.

Production policy	m^{opt} **[-]**	p_1^{opt} **[kg/h]**	p_2^{opt} **[kg/h]**	Q^{opt} **[kg]**	$tPIC^{opt}$ **[EUR/h]**	ERC^{opt} **[EUR/h]**	OC^{opt} **[EUR/h]**	TC^{opt} **[EUR/h]**
c/0/0	2	300	280	792.82	36.06	15.65	0.12	51.83
c/0/1	2	300	120	831.22	34.16	21.67	0.17	55.99
c/1/0	4	300	150	1101.43	32.86	31.15	0.24	64.26
c/1/1	4	150	120	1127.20	31.94	34.24	0.27	66.44
i/0/0	2	300	280	796.91	35.88	15.88	0.12	51.88
i/0/1	4	300	120	1356.26	26.83	28.82	0.22	55.88
i/1/0	5	300	120	1512.11	26.58	32.74	0.26	59.58
i/1/1	5	300	120	1504.71	26.58	34.24	0.27	61.09

Table 4 compares the different production policies to the continuous production policy where both production stages remain in an idle operating mode between successive production cycles (c/1/1), which was chosen as the base case, bc. When comparing these results to the results obtained by Zanoni et al. (2014) without the integration of a WHRS, it can be seen that the performance of the production policies relative to the base case remains almost the same when a WHRS is added to the system.

Table 4: Comparison of batch production policies with base batch production policy (c/1/1).

Production policy z	$\frac{tPIC^{WHRS,z}-tPIC^{WHRS,bc}}{tPIC^{WHRS,bc}}$ **[%]**	$\frac{ERC^{WHRS,z}-ERC^{WHRS,bc}}{ERC^{WHRS,bc}}$ **[%]**	$\frac{OC^{WHRS,z}-OC^{WHRS,bc}}{OC^{WHRS,bc}}$ **[%]**	$\frac{TC^{WHRS,z}-TC^{WHRS,bc}}{TC^{WHRS,bc}}$ **[%]**
c/0/0	12.92	-54.30	-54.30	-21.99
c/0/1	6.96	-36.72	-36.72	-15.72
c/1/0	2.90	-9.01	-9.01	-3.28
c/1/1		Base case bc		
i/0/0	12.34	-53.63	-53.63	-21.92
i/0/1	-15.98	-15.82	-15.82	-15.90
i/1/0	-16.76	-4.38	-4.38	-10.33
i/1/1	-16.76	0.00	0.00	-8.06

The reason for this becomes obvious when the cost components associated with the standard production system without WHRS ($tPIC^{std}$, ERC^{std}, TC^{std}) are compared to the cost components of the production system that uses a WHRS ($tPIC^{WHRS}$, ERC^{WHRS}, TC^{WHRS}). Table 5 shows that the traditional production and inventory carrying costs remain almost constant when a WHRS is added to the system as Q^{opt}, m^{opt}, p_1^{opt}, and p_2^{opt} are hardly affected by the introduction of the WHRS in the numerical example studied. In contrast, the energy-related costs of the production system are reduced by almost identical percentages across all policies (differences can only be seen from the third decimal place onwards). If the cost of generating electricity from the ORC is taken into account in addition, the average total costs of the production systems with and without the WHRS differ by less than 1 percent across all policies for the scenario analyzed.

Table 5: Comparison of cost components with and without WHRS.

Production policy z	$\frac{tPIC^{WHRS,z}-tPIC^{std,z}}{tPIC^{std,z}}$ **[%]**	$\frac{ERC^{WHRS,z}-ERC^{std,z}}{ERC^{std,z}}$ **[%]**	$\frac{TC^{WHRS,z}-TC^{std,z}}{TC^{std,z}}$ **[%]**
c/0/0	0.00	-2.18	-0.43
c/0/1	0.00	-2.18	-0.56
c/1/0	0.00	-2.18	-0.70
c/1/1	0.00	-2.18	-0.74
i/0/0	0.00	-2.18	-0.44
i/0/1	0.00	-2.18	-0.74
i/1/0	0.00	-2.18	-0.79
i/1/1	0.00	-2.18	-0.80

The savings that result from the introduction of a WHRS obviously increase with the energy prices (see Figure 7). The jumps in the curves result from changes in the optimal production variables (Q^{opt}, m^{opt}, p_1^{opt}, and p_2^{opt}) as a result of an increase in the energy price, e.

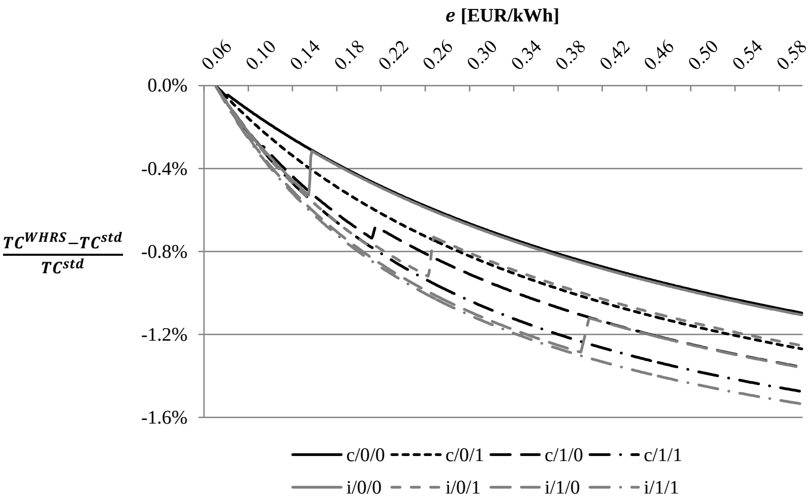


Figure 7: Effects of varying energy prices on total cost savings.

The saving potential of the WHRS, even in case of very high energy prices, is not promising for the real-life industry application considered here. To make an industrial-scale use of a WHRS in a batch production system economical, further technological advancements are necessary. Characteristics a WHRS would have to show to increase its effect on the production system are discussed in the following.

As was stated in the introduction, the overall amount of waste heat theoretically available exceeds the amount of waste heat which can technologically and economically be utilized today. From the figures cited in the introduction, we can infer that there is sufficient room for increasing the amount of waste heat that is led into the ORC, which would help to make more efficient use of the initial energy input. Secondly, the efficiency of the ORC could be further enhanced. Figure 8 indicates that, depending on the temperature of the waste heat stream and the temperature of the heat sink, the efficiency of the ORC can potentially be increased far beyond the 12.10 percent (see Equation (4)) that we assumed in our numerical example. For this reason, we consider a scenario that captures technological advancements to demonstrate the potential inherent in WHRSs in industrial applications next.

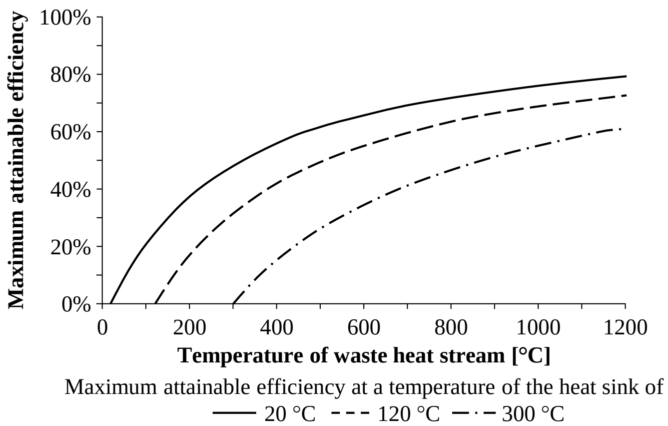


Figure 8: Maximum attainable efficiency of the conversion of waste heat into electricity (Hirzel et al., 2013).

Figure 9 shows the total cost saving potential of using a WHRS for our numerical example when either (I) the efficiency of the ORC, η^{ORC}, is set to 35 percent and the percentage of the initial energy input which is rejected as waste heat and led into the WHRS, α, is varied, or when (II) α is set to 35 percent and η^{ORC} is varied. Overall, the four interrupted production policies bear greater saving potentials than their respective non-interrupted counterparts. Within both the continuous (c/0/0, c/0/1, c/1/0, c/1/1) as well as the interrupted (i/0/0, i/0/1, i/1/0, i/1/1) production policies, the order of the four production policies regarding their saving potential is the same. The highest total cost savings that result from introducing a WHRS can be realized with the production policies where both production stages remain in an idle operating mode between successive production cycles (c/1/1 and i/1/1). Only slightly lower savings can be achieved when at least the first production stage remains in an idle operating mode between successive production cycles (c/1/0 and i/1/0). Production policy i/0/1 again leads to only slightly lower savings compared to production policy i/1/0. However, the total cost savings associated with the same policy for the case of a continuous production (c/0/1) lie well below the savings of production policies c/1/1 and c/1/0. Finally, the production policies c/0/0 and i/0/0 bear the lowest saving potentials. From this analysis, we conclude that the operating mode of the first production stage impacts the overall saving potential the most, which is not surprising as we only allowed the first production stage to be interrupted within a production run.

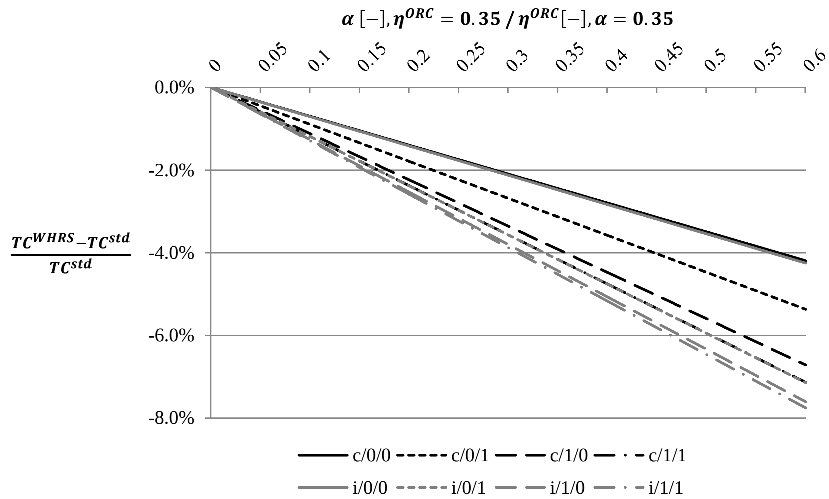


Figure 9: After technological advancements: Effect of varying waste heat stream or efficiency of ORC on total cost savings.

As a WHRS aims at making the production process more energy-efficient, it is of particular interest to study how the total costs are affected by the introduction of the WHRS when varying the energy consumption parameters of the production processes. Figure 10 illustrates the total cost savings resulting from the use of a WHRS as a function of the energy required at both production stages to produce one unit (k_1, k_2) as well as of the idle power of both production stages (W_1, W_2) in case of production policy i/1/1 after technological advancements have been achieved ($\alpha = 0.35$, $\eta^{ORC} = 0.35$). It is obvious that the benefits of a WHRS become even more apparent when it is used in the context of more energy-intensive production processes such as cement production, steel making, or glass production. Overall, the potential to enhance the energy efficiency of production processes by the use of a WHRS is significant, especially keeping in mind that only cost were considered in our study. If further performance factors are taken into consideration, such as greenhouse gas emissions, then making use of a WHRS would be even more appealing to users from industry.

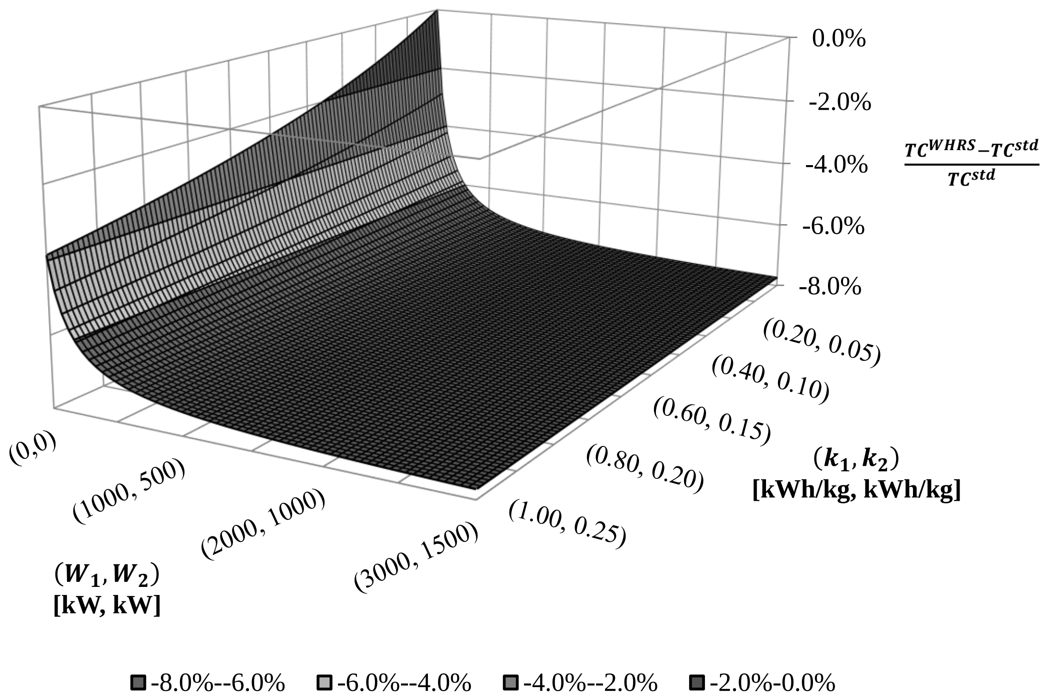


Figure 10: After technological advancements (α = 0.35, η^{ORC} = 0.35): Effect of varying energy consumption parameters of the production processes on total cost savings in case of production policy i/1/1.

7 Conclusions

This paper contributes to an emerging stream of research that concentrates on energy-aware production planning. The paper integrated a WHRS, which converts waste heat rejected by a production process into electricity to support operating the production stages, into a two-stage production system with controllable production rates. Thus, the paper established a link between managerial and technological approaches that both aim at enhancing the energy efficiency of production processes.

On the strategic level, the developed model can be applied to decide whether or not an investment in a WHRS is economically beneficial. On the tactical level, the model can be employed to determine how the traditional decision variables of inventory models are influenced by a WHRS, and to evaluate how the optimal solution is affected by changes in energy cost, operating cost of the WHRS, or the environment the WHRS is operated in. The numerical example studied implies that the impact of a WHRS on the total cost is larger for interrupted production policies than for continuous production policies. Besides, the WHRS was found to affect the total cost more if production stages remain in an idle operating mode instead of switching them off and on again. Yet, a more comprehensive simulation study would be necessary to identify scenarios in which the use of a WHRS is especially valuable.

The numerical example also showed that further technological advancements, which for instance increase the efficiency of the ORC or the fraction of the waste heat rejected by production processes that can technologically and economically be used, are necessary to make investments in WHRSs even more appealing to practitioners. However, even without technological advancements, the benefits of integrating a WHRS into a production system can further be exploited in different directions. Building on our research, it seems worthwhile to also consider emissions and to investigate how they can be reduced by using a WHRS. To this end, the emission of greenhouse gases needs to be related to the energy consumed in a first step. So far, this relationship has mostly been assumed to be linear (e.g., Fang et al. 2011; Liu, 2016; Sharma et al., 2015). Only Bazan et al. (2015) assumed a quadratic relationship, while Zhang et al. (2014) argued that the amount of emissions per kWh of energy consumed is associated with the source the energy is generated from, and hence varies over time. In a second step, it needs to be determined whether the amount of emissions should be minimized as part of the objective function (e.g., Fang et al. 2011; Zhang et al., 2014; Bazan et al., 2015; Sharma et al., 2015) or whether it is enough just to comply with constraints that limit the emissions per time period or the cumulative emissions over a specific (rolling) time horizon (Liu, 2016). From an environmental point of view, it may also be of interest to incorporate the selection and the sourcing process as well as the recycling or the disposal process of the organic fluid to be used in the ORC into the decision support model. This is due to the fact that various organic fluids differ greatly in terms of toxicity, flammability, ozone depleting potential, and greenhouse warming potential (Quoilin et al., 2013).

Beyond the consideration of environmental aspects, the advantages of a WHRS become even more explicit when it is combined with energy storage as investigated by Biel and Glock (2017). With the help of a WHRS, waste heat could be converted into electricity, which could subsequently be stored and be used when – depending on the energy supply contracts – power from the grid is particularly expensive. We leave these and other extensions for future research.

References

Andersen, W.C., Bruno, T.J., 2005. Rapid screening of fluids for chemical stability in Organic Rankine Cycle applications. *Industrial & Engineering Chemistry Research*, 44 (15), 5560-5566.

Artigues, C., Lopez, P., Hait, A., 2013. The energy scheduling problem: Industrial case-study and constraint propagation techniques. *International Journal of Production Economics*, 143 (1), 13-23.

Avriel, M., 2003. *Nonlinear programming: Analysis and methods*. Dover Publications, Mineola, NY.

Bazan, E., Jaber, M.Y., Zanoni, S., 2015. Supply chain models with greenhouse gases emissions, energy usage and different coordination decisions. *Applied Mathematical Modelling*, 39 (17), 5131-5151.

Biel, K., Glock, C.H., 2017. Prerequisites of efficient decentralized waste heat recovery and energy storage in production planning. *Journal of Business Economics*, 87 (1), 41-72.

Borgnakke, C., Sonntag, R.E., 2009. *Fundamentals of thermodynamics*. Wiley, New York, NY.

Buzacott, J.A., Ozkarahan, I.A., 1983. One- and two-stage scheduling of two products with distributed inserted idle time: The benefits of a controllable production rate. *Naval Research Logistics Quarterly*, 30 (4), 675-696.

Carrasco-Gallego, R., Ponce-Cueto, E., Dekker, R., 2012. Closed-loop supply chains of reusable articles: A typology grounded on case studies. *International Journal of Production Research*, 50 (19), 5582-5596.

Dai, Y., Wang, J., Gao, L., 2009. Parametric optimization and comparative study of organic Rankine cycle (ORC) for low grade waste heat recovery. *Energy Conversion and Management*, 50 (3), 576-582.

Elhafsi, M., Bai, S.X., 1997. The common cycle economic lot scheduling problem with backorders: Benefits of controllable production rates. *Journal of Global Optimization*, 10 (3), 283-303.

Fang, K., Uhan, N., Zhao, F., Sutherland, J.W., 2011. A new approach to scheduling in manufacturing for power consumption and carbon footprint reduction. *Journal of Manufacturing Systems*, 30 (4), 234-240.

Fernandez, M., Li, L., Sun, Z., 2013. “Just-for-Peak” buffer inventory for peak electricity demand reduction of manufacturing systems. *International Journal of Production Economics*, 146 (1), 178-184.

Gallego, G., 1993. Reduced production rates in the economic lot scheduling problem. *International Journal of Production Research*, 31 (5), 1035-1046.

Glock, C.H., 2010. Batch sizing with controllable production rates. *International Journal of Production Research*, 48 (20), 5925-5942.

Glock, C.H., 2011. Batch sizing with controllable production rates in a multi-stage production system. *International Journal of Production Research*, 49 (20), 6017-6039.

Govindan, K., Soleimani, H., Kannan, D., 2015. Reverse logistics and closed-loop supply chain: A comprehensive review to explore the future. *European Journal of Operational Research*, 240 (3), 603-626.

Greenberg, H.J., Pierskalla, W.P., 1971. A review of quasi-convex functions. *Operations Research*, 19 (7), 1553-1570.

Gutowski, T., Dahmus, J., Thiriez, A., 2006. *Electrical energy requirements for manufacturing processes*. In: Proceedings of the 13th CIRP International Conference on Life Cycle Engineering, Leuven, pp. 623-638.

Haddad, C., Périlhon, C., Danlos, A., François, M.-X., Descombes, G., 2014. Some efficient solutions to recover low and medium waste heat: Competitiveness of the thermoacoustic technology. *Energy Procedia*, 50, 1056-1069.

Hasanbeigi, A., Price, L., 2012. A review of energy use and energy efficiency technologies for the textile industry. *Renewable & Sustainable Energy Reviews*, 16 (6), 3648-3665.

Havel, P., Šimovič, T., 2013. Optimal planning of cogeneration production with provision of ancillary services. *Electric Power Systems Research*, 95, 47-55.

Hirzel, S., Sontag, B., Rhode, C., 2013. *Industrielle Abwärmenutzung*. Fraunhofer-Institut für System- und Innovationsforschung (ISI), Karlsruhe.

Hung, T.C., 2001. Waste heat recovery of organic Rankine cycle using dry fluids. *Energy Conversion and Management*, 42 (5), 539-553.

Hung, T.C., Shai, T.Y., Wang, S.K., 1997. A review of Organic Rankine Cycles (ORCs) for the recovery of low-grade waste heat. *Energy*, 22 (7), 661-667.

Hung, T.C., Wang, S.K., Kuo, C.H., Pei, B.S., Tsai, K.F., 2010. A study of organic working fluids on system efficiency of an ORC using low-grade energy sources. *Energy*, 35 (3), 1403-1411.

Invernizzi, C.M., 2013. *Closed power cycles: Thermodynamic fundamentals and applications*. Springer, London.

Johnson I, Choate W.T., Davidson, A., 2008. *Waste heat recovery: Technologies and opportunities in U.S. industry*. BCS, Inc. supported by the U.S. Department of Energy's Office of Energy Efficiency and Renewable Energy, Industrial Technologies Program, Laurel, MD.

Karellas, S., Leontaritis, A.D., Panousis, G., Bellos, E., Kakaras, E., 2013. Energetic and exergetic analysis of waste heat recovery systems in the cement industry. *Energy*, 58, 147-156.

Larjola, J., 1995. Electricity from industrial waste heat using high-speed Organic Rankine Cycle (ORC). *International Journal of Production Economics*, 41 (1-3), 227-235.

Liu, B.-T., Chien, K.-H., Wang, C.-C., 2004. Effect of working fluids on Organic Rankine Cycle for waste heat recovery. *Energy*, 29 (8), 1207-1217.

Liu, C.-H., 2015. Mathematical programming formulations for single-machine scheduling problems while considering renewable energy uncertainty. *International Journal of Production Research*, 54 (4), 1122-1133.

López, L., Blanco, J.M., Bonilla, J.J., Bacza, S., Sala, J.M., 1998. Determination of energy and exergy of waste heat in the industry of the Basque country. *Applied Thermal Engineering*, 18 (3-4), 187-197.

Luo, H., Du, B., Huang, G.Q., Chen, H., Li, X., 2013. Hybrid flow shop scheduling considering machine electricity consumption cost. *International Journal of Production Economics*, 146 (2), 423-439.

McKenna, R.C., Norman, J.B., 2010. Spatial modelling of industrial heat loads and recovery potentials in the UK. *Energy Policy*, 38 (10), 5878-5891.

Mouzon, G., Yildirim, M.B., Twomey, J., 2007. Operational methods for minimization of energy consumption of manufacturing equipment. *International Journal of Production Research*, 45 (18-19), 4247-4271.

Obernberger, I., Thonhofer, P., Reisenhofer, E., 2002. Description and evaluation of the new 1000 kWel Organic Rankine Cycle process integrated in the biomass CHP plant in Lienz, Austria. *Euroheat & Power*, 10, 1-17.

Oliveira, M.J., 2013. *Equilibrium thermodynamics*. Springer, Berlin Heidelberg.

Pehnt, M., Bödeker, J., Arens, M., Jochem, E., Idrissova, F., 2010. *Die Nutzung industrieller Abwärme – technisch-wirtschaftliche Potenziale und energiepolitische Umsetzung*. Institut für Energie- und Umweltforschung (ifeu), Fraunhofer-Institut für System- und Innovationsforschung (ISI), Institut für Ressourceneffizienz und Energiestrategien GmbH (IREES), Heidelberg Karlsruhe.

Quoilin, S., Declaye, S., Tchanche, B.F., Lemort, V., 2011. Thermo-economic optimization of waste heat recovery Organic Rankine Cycles. *Applied Thermal Engineering*, 31 (14), 2885-2893.

Quoilin, S., Van Den Broek, M., Declaye, S., Dewallef, P., Lemort, V., 2013. Techno-economic survey of Organic Rankine Cycle (ORC) systems. *Renewable & Sustainable Energy Reviews*, 22, 168-186.

Salgado, F., Pedrero, P., 2008. Short-term operation planning on cogeneration systems: A survey. *Electric Power Systems Research*, 78(5), 835-848.

Schaefer, H., 1995. *Energiewirtschaftliche Bedeutung der Nutzung von Abfallenergien*. In: Fratzscher, W., Stephan, K. (Eds.), Abfallenergienutzung - Technische, wirtschaftliche und soziale Aspekte, vol. 2, Akadmie Verlag, Berlin, pp. 42-60

Sharma, A., Zhao, F., Sutherland, J.W., 2015. Econological scheduling of a manufacturing enterprise operating under a time-of-use electricity tariff. *Journal of Cleaner Production*, 108, 256-270.

Shrouf, F., Ordieres-Meré, J., García-Sánchez, A., Ortega-Mier, M., 2014. Optimizing the production scheduling of a single machine to minimize total energy consumption costs. *Journal of Cleaner Production*, 67, 197-207.

Silver, E.A., 1990. Deliberately slowing down output in a family production context. *International Journal of Production Research*, 28 (1), 17-27.

Souza, G.C., 2013. Closed-loop supply chains: A critical review, and future research*. *Decision Sciences*, 44 (1), 7-38.

Tchanche, B.F., Quoilin, S., Declaye, S., Papadakis, G., Lemort, V., 2010. *Economic feasibility study of a small scale Organic Rankine Cycle system in waste heat recovery application*. In: Proceedings of the ASME 2010 10th Biennial Conference on Engineering Systems Design and Analysis, Istanbul, vol. 1, pp. 249-256.

U.S. Energy Information Administration, 2013. *International energy outlook 2013*. Washington, DC.

U.S. Energy Information Administration, 2014. *Annual energy outlook with projections to 2040*. Washington, DC.

Uusitalo, A., Honkatukia, J., Turunen-Saaresti, T., Larjola, J., 2014. A thermodynamic analysis of waste heat recovery from reciprocating engine power plants by means of Organic Rankine Cycles. *Applied Thermal Engineering*, 70 (1), 33-41.

Wang, C., He, B., Yan, L., Pei, X., Chen, S., 2014. Thermodynamic analysis of a low-pressure economizer based waste heat recovery system for a coal-fired power plant. *Energy*, 65, 80-90.

Wang, E.H., Zhang, H.G., Fan, B.Y., Ouyang, M.G., Zhao, Y., Mu, Q.H., 2011. Study of working fluid selection of Organic Rankine Cycle (ORC) for engine waste heat recovery. *Energy*, 36 (5), 3406-3418.

Wei, D., Lu, X., Lu, Z., Gu, J., 2007. Performance analysis and optimization of organic Rankine cycle (ORC) for waste heat recovery. *Energy Conversion and Management*, 48 (4), 1113-1119.

Yamamoto, T., Furuhata, T., Arai, N., Mori, K., 2001. Design and testing of the Organic Rankine Cycle. *Energy*, 26 (3), 239-251.

Zanoni, S., Bettoni, L., Glock, C.H., 2014. Energy implications in a two-stage production system with controllable production rates. *International Journal of Production Economics*, 149, 164-171.

Zhang, H., Zhao, F., Fang, K., Sutherland, J.W., 2014. Energy-conscious flow shop scheduling under time-of-use electricity tariffs. *CIRP Annals - Manufacturing Technology*, 63 (1), 37-40.

Zhu, S., Deng, K., Qu, S., 2014. Thermodynamic analysis of an in-cylinder waste heat recovery system for internal combustion engines. *Energy*, 67, 548-556.

Appendix

Appendix A: Properties of Equation (34) with respect to p_1 and p_2

$$TC^{CON,opt}(p_1,p_2,m) = d \cdot \left(\gamma \cdot \left(\frac{hc_{1,2}\cdot\left(\frac{2}{m}-1\right)}{p_1} + \frac{hc_{1,2}-hc_3}{p_2} + \frac{hc_3}{d}\right)\right)^{\frac{1}{2}} + \left(\frac{W_1\cdot(1-B_1)}{p_1} + \frac{W_2\cdot(1-B_2)}{p_2}\right)\cdot d\cdot\zeta + \beta \tag{A.1}$$

with

$$\gamma = 2\cdot\left(A + SC\cdot m + \left(\sum_{i\in\{1,2\}} f\cdot W_i\cdot(1-B_i)\cdot t_i^{on}\right)\cdot\left((1-\alpha\cdot\eta^{ORC})\cdot e + \alpha\cdot\eta^{ORC}\cdot c\right)\right)$$

$$\zeta = \left((1-\alpha\cdot\eta^{ORC})\cdot e + \alpha\cdot\eta^{ORC}\cdot c\right)$$

$$\beta = \sum_{i\in\{1,2\}}\left(k_i + \frac{W_i\cdot B_i}{d}\right)\cdot d\cdot\left((1-\alpha\cdot\eta^{ORC})\cdot e + \alpha\cdot\eta^{ORC}\cdot c\right)$$

Theorem A.1. *If* $B_1 = 0$ *and* $m \in \{1,2\}$, $p_1^{opt} = p_1^{max}$ *minimizes Equation (34) for given* p_2.

Proof. For $B_1 = 0$ and $m \in \{1,2\}$ Equation (A.1) reduces to

$$TC^{CON,opt}(p_1,p_2,m) = d \cdot \left(\gamma \cdot \left(\frac{hc_{1,2}\cdot\left(\frac{2}{m}-1\right)}{p_1} + \frac{hc_{1,2}-hc_3}{p_2} + \frac{hc_3}{d}\right)\right)^{\frac{1}{2}} + \left(\frac{W_1}{p_1} + \frac{W_2\cdot(1-B_2)}{p_2}\right)\cdot d\cdot\zeta + \beta \tag{A.2}$$

where $hc_{1,2}\cdot\left(\frac{2}{m}-1\right) \geq 0$ holds. Thus, it is obvious that $p_1^{opt} = p_1^{max}$ minimizes Equation (34) for given p_2 if $B_1 = 0$ and $m \in \{1,2\}$. ■

Theorem A.2. *If* $B_1 = 0$ *and* $m > 2$, *either* $p_1^{opt} = p_1^{min}$ *or* $p_1^{opt} = p_1^{max}$ *minimizes Equation (34) for given* p_2.

Proof. We show that Equation (A.2) does not have a local minimum in $p_1 \in \left[p_1^{min}, p_1^{max}\right]$ for given p_2 and $m > 2$. Therefore, we set the first partial derivative of Equation (A.2) with respect to p_1 to 0:

$$\frac{\delta TC^{CON,opt}(p_1,p_2,m)}{\delta p_1} = \frac{d\cdot\left(\frac{\gamma\cdot hc_{1,2}\cdot\left(1-\frac{2}{m}\right)}{p_1^2}\right)}{2\cdot\left(\gamma\cdot\left(\frac{hc_{1,2}\cdot\left(\frac{2}{m}-1\right)}{p_1}+\frac{hc_{1,2}-hc_3}{p_2}+\frac{hc_3}{d}\right)\right)^{\frac{1}{2}}} - \frac{W_1\cdot d\cdot\zeta}{p_1^2} = 0. \tag{A.3}$$

The only extremum of Equation (A.2) in p_1 can be found at

$$p_1^{opt} = \frac{hc_{1,2}\cdot\left(\frac{2}{m}-1\right)}{\gamma\cdot\left(\frac{hc_{1,2}\cdot\left(1-\frac{2}{m}\right)}{2\cdot W_1\cdot\zeta}\right)^2-\frac{hc_{1,2}-hc_3}{p_2}-\frac{hc_3}{d}}. \tag{A.4}$$

However, this extremum would only be a local minimum if the second partial derivative of Equation (A.2) with respect to p_1 was greater than 0 at p_1^{opt} for given p_2 and $m > 2$:

$$\frac{\delta^2 TC^{CON,opt}(p_1,p_2,m)}{\delta p_1^2} = -\frac{d\cdot\left(\frac{\gamma\cdot hc_{1,2}\cdot\left(1-\frac{2}{m}\right)}{p_1^2}\right)^2}{4\cdot\left(\gamma\cdot\left(\frac{hc_{1,2}\cdot\left(\frac{2}{m}-1\right)}{p_1}+\frac{hc_{1,2}-hc_3}{p_2}+\frac{hc_3}{d}\right)\right)^{\frac{3}{2}}} + \frac{d\cdot\left(\frac{\gamma\cdot hc_{1,2}\cdot\left(\frac{2}{m}-1\right)}{p_1^3}\right)}{\left(\gamma\cdot\left(\frac{hc_{1,2}\cdot\left(\frac{2}{m}-1\right)}{p_1}+\frac{hc_{1,2}-hc_3}{p_2}+\frac{hc_3}{d}\right)\right)^{\frac{1}{2}}} + \frac{2\cdot W_1\cdot d\cdot\zeta}{p_1^3} > 0. \tag{A.5}$$

$$-\frac{2\cdot d\cdot(W_1\cdot\zeta)^3}{\left(p_1^{opt}\right)^4\cdot\gamma\cdot hc_{1,2}\cdot\left(1-\frac{2}{m}\right)} > 0,\ \forall p_1^{opt} > 0, m > 2 \ \lightning \tag{A.6}$$

The contradiction we arrive at in Equation (A.6) proofs that Equation (A.2) does not have a local minimum $\forall p_1^{opt} > 0$ for given p_2 and $m > 2$. Consequently, either $p_1^{opt} = p_1^{min}$ or $p_1^{opt} = p_2^{max}$ minimizes Equation (34) for given p_2 if $B_1 = 0$ and $m > 2$. ■

Theorem A.3. *If $B_1 = 1$ and $m = 1$, $p_1^{opt} = p_1^{max}$ minimizes Equation (34) for given p_2.*

Proof. For $B_1 = 1$ and $m = 1$, Equation (A.1) reduces to

$$TC^{CON,opt}(p_1,p_2,m) = d\cdot\left(\gamma\cdot\left(\frac{hc_{1,2}}{p_1}+\frac{hc_{1,2}-hc_3}{p_2}+\frac{hc_3}{d}\right)\right)^{\frac{1}{2}} + \left(\frac{W_2\cdot(1-B_2)}{p_2}\right)\cdot d\cdot\zeta+\beta. \tag{A.7}$$

Thus, it is obvious that $p_1^{opt} = p_1^{max}$ minimizes Equation (34) for given p_2 if $B_1 = 1$ and $m = 1$. ■

Theorem A.4. *If $B_1 = 1$ and $m = 2$, any $p_1^{opt} \in \left[p_1^{min}, p_1^{max}\right]$ minimizes Equation (34) for given p_2.*

Proof. For $B_1 = 1$ and $m = 2$, Equation (A.1) reduces to

$$TC^{CON,opt}(p_1,p_2,m) = d\cdot\left(\gamma\cdot\left(\frac{hc_{1,2}-hc_3}{p_2}+\frac{hc_3}{d}\right)\right)^{\frac{1}{2}} + \left(\frac{W_2\cdot(1-B_2)}{p_2}\right)\cdot d\cdot\zeta+\beta. \tag{A.8}$$

Thus, it is obvious that any $p_1^{opt} \in \left[p_1^{min}, p_1^{max}\right]$ minimizes Equation (34) for given p_2 if $B_1 = 1$ and $m = 2$ as Equation (34) is independent of p_1 for $B_1 = 1$ and $m = 2$. ■

Theorem A.5. *If $B_1 = 1$ and $m > 2$, $p_1^{opt} = p_1^{min}$ minimizes Equation (34) for given p_2.*

Proof. For $B_1 = 1$ and $m > 2$, Equation (A.1) reduces to

$$TC^{CON,opt}(p_1, p_2, m) = d \cdot \left(\gamma \cdot \left(\frac{hc_{1,2} \cdot \left(\frac{2}{m} - 1\right)}{p_1} + \frac{hc_{1,2} - hc_3}{p_2} + \frac{hc_3}{d}\right)\right)^{\frac{1}{2}} + \left(\frac{W_2 \cdot (1 - B_2)}{p_2}\right) \cdot d \cdot \zeta + \beta, \tag{A.9}$$

where $hc_{1,2} \cdot \left(\frac{2}{m} - 1\right) < 0$ holds. Thus, it is obvious that $p_1^{opt} = p_1^{min}$ minimizes Equation (34) for given p_2 if $B_1 = 1$ and $m > 2$. ■

Theorem A.6. *If $B_2 = 0$ and $hc_{1,2} \geq hc_3$, $p_2^{opt} = p_2^{max}$ minimizes Equation (34) for given p_1 and m.*

Proof. For $B_2 = 0$, Equation (A.1) reduces to

$$TC^{CON,opt}(p_1, p_2, m) = d \cdot \left(\gamma \cdot \left(\frac{hc_{1,2} \cdot \left(\frac{2}{m} - 1\right)}{p_1} + \frac{hc_{1,2} - hc_3}{p_2} + \frac{hc_3}{d}\right)\right)^{\frac{1}{2}} + \left(\frac{W_1 \cdot (1 - B_1)}{p_1} + \frac{W_2}{p_2}\right) \cdot d \cdot \zeta + \beta. \tag{A.10}$$

Thus, it is obvious that $p_2^{opt} = p_2^{max}$ minimizes Equation (34) for given p_1 and m if $B_2 = 0$ and $hc_{1,2} \geq hc_3$. ■

Theorem A.7. *If $B_2 = 0$ and $hc_{1,2} < hc_3$, either $p_2^{opt} = p_2^{min}$ or $p_2^{opt} = p_2^{max}$ minimizes Equation (34) for given p_1 and m.*

Proof. We show that Equation (A.10) does not have a local minimum in $p_2 \in \left[p_2^{min}, p_2^{max}\right]$ for given p_1 and m. Therefore, we set the first partial derivative of Equation (A.10) with respect to p_2 to 0:

$$\frac{\delta TC^{CON,opt}(p_1, p_2, m)}{\delta p_2} = -\frac{d \cdot \gamma \cdot \left(\frac{hc_{1,2} - hc_3}{p_2^2}\right)}{2 \cdot \left(\gamma \cdot \left(\frac{hc_{1,2} \cdot \left(\frac{2}{m} - 1\right)}{p_1} + \frac{hc_{1,2} - hc_3}{p_2} + \frac{hc_3}{d}\right)\right)^{\frac{1}{2}}} - \frac{W_2 \cdot d \cdot \zeta}{p_2^2} = 0. \tag{A.11}$$

The only extremum of Equation (A.10) in p_2 can be found at

$$p_2^{opt} = \frac{hc_{1,2} - hc_3}{\gamma \cdot \left(\frac{-hc_{1,2} + hc_3}{2 \cdot W_2 \cdot \zeta}\right)^2 - \frac{hc_{1,2} \cdot \left(\frac{2}{m} - 1\right)}{p_1} - \frac{hc_3}{d}}. \tag{A.12}$$

However, this extremum would only be a local minimum if the second partial derivative of Equation (A.10) with respect to p_2 was greater than 0 at p_2^{opt} for given p_1 and m:

$$\frac{\delta^2 TC^{CON,opt}(p_1,p_2,m)}{\delta p_2^2} = -\frac{d\cdot\left(\gamma\cdot\left(\frac{hc_{1,2}-hc_3}{p_2^2}\right)\right)^2}{4\cdot\left(\gamma\cdot\left(\frac{hc_{1,2}\cdot\left(\frac{2}{m}-1\right)}{p_1}+\frac{hc_{1,2}-hc_3}{p_2}+\frac{hc_3}{d}\right)\right)^{\frac{3}{2}}} + \frac{d\cdot\gamma\cdot\left(\frac{hc_{1,2}-hc_3}{p_2^3}\right)}{\left(\gamma\cdot\left(\frac{hc_{1,2}\cdot\left(\frac{2}{m}-1\right)}{p_1}+\frac{hc_{1,2}-hc_3}{p_2}+\frac{hc_3}{d}\right)\right)^{\frac{1}{2}}} + \frac{2\cdot W_2\cdot d\cdot\zeta}{p_2^3} > 0. \tag{A.13}$$

$$\frac{2\cdot(W_2\cdot\zeta)^3}{\left(p_2^{opt}\right)^4\cdot\gamma\cdot(hc_{1,2}-hc_3)} > 0,\ \forall p_2^{opt} > 0, hc_{1,2} < hc_3 \ ↯ \tag{A.14}$$

The contradiction we arrive at in Equation (A.14) proofs that Equation (A.10) does not have a local minimum $\forall p_2^{opt} > 0$ and $hc_{1,2} < hc_3$ for given p_1 and m. Consequently, either $p_2^{opt} = p_2^{min}$ or $p_2^{opt} = p_2^{max}$ minimizes Equation (34) for given p_1 and m if $B_2 = 0$ and $hc_{1,2} < hc_3$. ■

Theorem A.8. *If $B_2 = 1$ and $hc_{1,2} \geq hc_3$, $p_2^{opt} = p_2^{max}$ minimizes Equation (34) for given p_1 and m. If $B_2 = 1$ and $hc_{1,2} < hc_3$, $p_2^{opt} = p_2^{min}$ minimizes Equation (34) for given p_1 and m.*

Proof. For $B_2 = 1$, Equation (A.1) reduces to

$$TC^{CON,opt}(p_1,p_2,m) = d\cdot\left(\gamma\cdot\left(\frac{hc_{1,2}\cdot\left(\frac{2}{m}-1\right)}{p_1}+\frac{hc_{1,2}-hc_3}{p_2}+\frac{hc_3}{d}\right)\right)^{\frac{1}{2}} + \left(\frac{W_1\cdot(1-B_1)}{p_1}\right)\cdot d\cdot\zeta + \beta. \tag{A.15}$$

Thus, it is obvious that $p_2^{opt} = p_2^{max}$ minimizes Equation (34) for given p_1 and m if $B_2 = 1$ and $hc_{1,2} \geq hc_3$, while $p_2^{opt} = p_2^{min}$ minimizes Equation (34) for given p_1 and m if $B_2 = 1$ and $hc_{1,2} < hc_3$ ■

The properties of the optimal solution of p_1 and p_2 with respect to Equation (34) are summarized in Table A.1.

Table A.1: Properties of the optimal solution of p_1 and p_2 with respect to Equation (34).

Batch production policies	$\boldsymbol{p_1^{opt}}$	$\boldsymbol{p_2^{opt}}$	$\boldsymbol{m}$
c/0/0	p_1^{max}	$p_2^{max}, if\ hc_{1,2} \geq hc_3$ $p_2^{min} \vee p_2^{max}, else$	$m \in \{1,2\}$
	$p_1^{min} \vee p_1^{max}$	$\min\{p_2^{max}, p_1^{opt}\}, if\ hc_{1,2} \geq hc_3$ $p_2^{min} \vee \min\{p_2^{max}, p_1^{opt}\}, else$	$m > 2$
c/0/1	p_1^{max}	$p_2^{max}, if\ hc_{1,2} \geq hc_3$ $p_2^{min}, else$	$m \in \{1,2\}$
	$p_1^{min} \vee p_1^{max}$	$\min\{p_2^{max}, p_1^{opt}\}, if\ hc_{1,2} \geq hc_3$ $p_2^{min}, else$	$m > 2$
c/1/0	p_1^{max}	$p_2^{max}, if\ hc_{1,2} \geq hc_3$ $p_2^{min} \vee p_2^{max}, else$	$m = 1$
	$[\max\{p_1^{min}, p_2^{opt}\}, p_1^{max}]$	$p_2^{max}, if\ hc_{1,2} \geq hc_3$ $p_2^{min} \vee p_2^{max}, else$	$m = 2$
	p_1^{min}	$\min\{p_2^{max}, p_1^{opt}\}, if\ hc_{1,2} \geq hc_3$ $p_2^{min} \vee \min\{p_2^{max}, p_1^{opt}\}, else$	$m > 2$
c/1/1	p_1^{max}	$p_2^{max}, if\ hc_{1,2} \geq hc_3$ $p_2^{min}, else$	$m = 1$
	$[\max\{p_1^{min}, p_2^{opt}\}, p_1^{max}]$	$p_2^{max}, if\ hc_{1,2} \geq hc_3$ $p_2^{min}, else$	$m = 2$
	p_1^{min}	$\min\{p_2^{max}, p_1^{opt}\}, if\ hc_{1,2} \geq hc_3$ $p_2^{min}, else$	$m > 2$

Appendix B: Properties of Equation (35) with respect to p_1 and p_2

$$TC^{INT,opt}(p_1, p_2, m) = d \cdot \left(\gamma \cdot \left(\frac{\frac{1}{m} \cdot hc_{1,2}}{p_1} + \frac{\frac{1}{m} \cdot hc_{1,2} - hc_3}{p_2} + \frac{hc_3}{d}\right)\right)^{\frac{1}{2}} + \left(\frac{\frac{1}{m} W_1 \cdot (1 - B_1)}{p_1} + \frac{\left(1 - \frac{1}{m}\right) \cdot W_1 \cdot (1 - B_1) + W_2 \cdot (1 - B_2)}{p_2}\right) \cdot d \cdot \zeta + \beta \tag{B.1}$$

with

$$\gamma = 2 \cdot \left(A + SC \cdot m + \sum_{i \in \{1,2\}} f \cdot W_i \cdot (1 - B_i) \cdot t_i^{on} \cdot \left((1 - \alpha \cdot \eta^{ORC}) \cdot e + \alpha \cdot \eta^{ORC} \cdot c\right)\right)$$

$$\zeta = \left((1 - \alpha \cdot \eta^{ORC}) \cdot e + \alpha \cdot \eta^{ORC} \cdot c\right)$$

$$\beta = \sum_{i \in \{1,2\}} \left(k_i + \frac{W_i \cdot B_i}{d}\right) \cdot d \cdot \left((1 - \alpha \cdot \eta^{ORC}) \cdot e + \alpha \cdot \eta^{ORC} \cdot c\right)$$

Theorem B.1. *If* $B_1 \in \{0,1\}$ *and* $m > 0$, $p_1^{opt} = p_1^{max}$ *minimizes Equation (35) for given* p_2.

Proof. This follows directly from evaluating the terms of Equation (B.1) which depend on p_1. ■

Theorem B.2. *If* $B_i \in \{0,1\}\ \forall i \in \{1,2\}$ *and* $\frac{hc_{1,2}}{m} \geq hc_3$, $p_2^{opt} = p_2^{max}$ *minimizes Equation (35) for given* p_1 *and* m.

Proof. This follows directly from evaluating the terms of Equation (B.1) which depend on p_2. ■

Theorem B.3. *If* $B_i = 1\ \forall i \in \{1,2\}$ *and* $\frac{hc_{1,2}}{m} < hc_3$, $p_2^{opt} = p_2^{min}$ *minimizes Equation (35) for given* p_1 *and* m.

Proof. For $B_1 = 1$ and $B_2 = 1$, Equation (B.1) reduces to

$$TC^{INT,opt}(p_1, p_2, m) = d \cdot \left(\gamma \cdot \left(\frac{\frac{1}{m} \cdot hc_{1,2}}{p_1} + \frac{\frac{1}{m} \cdot hc_{1,2} - hc_3}{p_2} + \frac{hc_3}{d}\right)\right)^{\frac{1}{2}} + \beta. \tag{B.2}$$

Thus, it is obvious that $p_2^{opt} = p_2^{min}$ minimizes Equation (35) for given p_1 and m if $B_i = 1\ \forall i \in \{1,2\}$ and $\frac{hc_{1,2}}{m} < hc_3$ ■

Theorem B.4. *If* $B_1 = 0$ *and* $B_2 = 0$ *and* $\frac{hc_{1,2}}{m} < hc_3$, *or* $B_1 = 0$ *and* $B_2 = 1$ *and* $\frac{hc_{1,2}}{m} < hc_3$, *or* $B_1 = 1$ *and* $B_2 = 0$ *and* $\frac{hc_{1,2}}{m} < hc_3$, *either* $p_2^{opt} = p_2^{min}$ *or* $p_2^{opt} = p_2^{max}$ *minimizes Equation (35) for given* p_1 *and* m.

Proof. We show that Equation (B.1) does not have a local minimum in $p_2 \in \left[p_2^{min}, p_2^{max}\right]$ for given p_1 and m. Therefore, we set the first partial derivative of Equation (B.1) to 0:

$$\frac{\delta TC^{INT,opt}(p_1, p_2, m)}{\delta p_2} = -\frac{d \cdot \gamma \cdot \left(\frac{\frac{1}{m} \cdot hc_{1,2} - hc_3}{p_2^2}\right)}{2 \cdot \left(\gamma \cdot \left(\frac{\frac{1}{m} \cdot hc_{1,2}}{p_1} + \frac{\frac{1}{m} \cdot hc_{1,2} - hc_3}{p_2} + \frac{hc_3}{d}\right)\right)^{\frac{1}{2}}} - \left(\frac{\left(1 - \frac{1}{m}\right) \cdot W_1 \cdot (1 - B_1) + W_2 \cdot (1 - B_2)}{p_2^2}\right) \cdot d \cdot \zeta = 0. \tag{B.3}$$

The only extremum of Equation (B.1) in p_2 can be found at

$$p_2^{opt} = \frac{\frac{1}{m} \cdot hc_{1,2} - hc_3}{\gamma \cdot \left(\frac{\left(hc_3 - \frac{1}{m} \cdot hc_{1,2}\right)}{2 \cdot \left(\left(1 - \frac{1}{m}\right) \cdot W_1 \cdot (1 - B_1) + W_2 \cdot (1 - B_2)\right) \cdot \zeta}\right)^2 - \frac{\frac{1}{m} hc_{1,2}}{p_1} - \frac{hc_3}{d}}. \tag{B.4}$$

However, this extremum would only be a local minimum if the second partial derivative of Equation (B.1) was greater than 0 at p_2^{opt} for given p_1 and m:

$$\frac{\delta^2 TC^{INT,opt}(p_1,p_2,m)}{\delta p_2^2} = -\frac{d\cdot\left(\gamma\left(\frac{\frac{1}{m}\cdot hc_{1,2}-hc_3}{p_2^2}\right)\right)^2}{4\cdot\left(\gamma\cdot\left(\frac{\frac{1}{m}\cdot hc_{1,2}}{p_1}+\frac{\frac{1}{m}\cdot hc_{1,2}-hc_3}{p_2}+\frac{hc_3}{d}\right)\right)^{\frac{3}{2}}} + \frac{d\cdot\gamma\left(\frac{\frac{1}{m}\cdot hc_{1,2}-hc_3}{p_2^3}\right)}{\left(\gamma\cdot\left(\frac{\frac{1}{m}\cdot hc_{1,2}}{p_1}+\frac{\frac{1}{m}\cdot hc_{1,2}-hc_3}{p_2}+\frac{hc_3}{d}\right)\right)^{\frac{1}{2}}} + 2\cdot\left(\frac{\left(1-\frac{1}{m}\right)\cdot W_1\cdot(1-B_1)+W_2\cdot(1-B_2)}{p_2^3}\right)\cdot d\cdot\zeta > 0. \tag{B.5}$$

$$\frac{2\cdot d\cdot\left(\left(\left(1-\frac{1}{m}\right)\cdot W_1\cdot(1-B_1)+W_2\cdot(1-B_2)\right)\cdot\zeta\right)^3}{\left(p_2^{opt}\right)^4\cdot\gamma\cdot\left(\frac{1}{m}\cdot hc_{1,2}-hc_3\right)} > 0, \forall p_2^{opt} > 0 \wedge \begin{pmatrix} \left(B_1 = 0 \wedge B_2 = 0 \wedge \frac{hc_{1,2}}{m} < hc_3\right) \\ \vee\left(B_1 = 0 \wedge B_2 = 1 \wedge \frac{hc_{1,2}}{m} < hc_3\right) \\ \vee\left(B_1 = 1 \wedge B_2 = 0 \wedge \frac{hc_{1,2}}{m} < hc_3\right) \end{pmatrix} ↯ \tag{B.6}$$

The contradiction we arrive at in Equation (B.6) proofs that Equation (B.1) does not have a local minimum $\forall p_2^{opt} > 0$ for given p_1 and m if $B_1 = 0$ and $B_2 = 0$ and $\frac{hc_{1,2}}{m} < hc_3$, or $B_1 = 0$ and $B_2 = 1$ and $\frac{hc_{1,2}}{m} < hc_3$, or $B_1 = 1$ and $B_2 = 0$ and $\frac{hc_{1,2}}{m} < hc_3$. Consequently, either $p_2^{opt} = p_2^{min}$ or $p_2^{opt} = p_2^{max}$ minimizes Equation (35) for given p_1 and m if $B_1 = 0$ and $B_2 = 0$ and $\frac{hc_{1,2}}{m} < hc_3$, or $B_1 = 0$ and $B_2 = 1$ and $\frac{hc_{1,2}}{m} < hc_3$, or $B_1 = 1$ and $B_2 = 0$ and $\frac{hc_{1,2}}{m} < hc_3$. ■

The properties of the optimal solution of p_1 and p_2 with respect to Equation (35) are summarized in Table B.1.

Table B.1: Properties of the optimal solution of p_1 and p_2 with respect to Equation (35).

Batch production policy	$\boldsymbol{p_1^{opt}}$	$\boldsymbol{p_2^{opt}}$	$\boldsymbol{m}$
i/0/0	p_1^{max}	$p_2^{max}, if\ \frac{hc_{1,2}}{m} \geq hc_3$ $p_2^{min} \vee p_2^{max}, else$	$\forall m$
i/0/1	p_1^{max}	$p_2^{max}, if\ \frac{hc_{1,2}}{m} \geq hc_3$ $p_2^{min} \vee p_2^{max}, else$	$\forall m$
i/1/0	p_1^{max}	$p_2^{max}, if\ \frac{hc_{1,2}}{m} \geq hc_3$ $p_2^{min} \vee p_2^{max}, else$	$\forall m$
i/1/1	p_1^{max}	$p_2^{max}, if\ \frac{hc_{1,2}}{m} \geq hc_3$ $p_2^{min}, else$	$\forall m$

Paper 3 Prerequisites of efficient decentralized waste heat recovery and energy storage in production planning[1]

Authors: Konstantin Biel, Christoph H. Glock

Type of publication: Journal article

Publication details: *Journal of Business Economics*, 87 (1), 41-72, 2017.

Abstract

Following the scarcity of resources, rising energy prices, and an increasing awareness of the role manufacturing plays in the generation of greenhouse gas emissions, the usage of energy has more and more been considered in research on production planning and scheduling in recent years. Time-varying energy prices, which have been introduced to penalize energy usage during peak-demand periods and which are supposed to smooth energy demand, have added a new aspect to this stream of research. This article studies how the integration of a waste heat recovery system, which can convert industrial waste heat into electrical energy, along with an electrical energy storage system can balance the positive and negative effects of energy peak prices on the production plan in a serial multi-stage production system. After developing an appropriate model, we investigate how the use of the waste heat recovery system and the electrical energy storage system impact production planning. In a numerical analysis, we investigate under which conditions the recovery of waste heat combined with the opportunity to store energy provides practitioners with an efficient tool to lower total energy usage and to better react to time-varying energy prices, and thus to reduce total energy cost.

Keywords:

Energy efficiency; Sustainable manufacturing system; Energy usage; Production planning; Waste heat recovery; Electrical energy storage system

[1] Reprinted from *Journal of Business Economics*, 87 (1), Biel, K., Glock, C.H., Prerequisites of efficient decentralized waste heat recovery and energy storage in production planning, 41-72, 2017, with permission from Springer-Verlag GmbH.

1 Introduction

Manufacturing companies increasingly recognize the importance of considering energy utilization in modern production planning. Various recent developments have contributed to this trend. For example, rising and (and sometimes also time-varying) electricity prices have brought the cost impact of energy usage to the attention of many companies. In addition, society has started to realize the negative impacts growing energy-related CO_2 emissions have on the environment. As a result, politics have begun to provide companies with strong incentives to reduce their energy usage and consequently their energy-related CO_2 emissions. Yet, various up-to-date reports regarding perspective industrial energy usage show that the topic of energy-aware manufacturing will not lose its relevance in the near future.

According to the Annual Energy Outlook 2015 of the U.S. Energy Information Administration, for example, the total energy usage in the United States is expected to rise from 27,665 TWh in 2012 to 30,986 TWh in 2040 (see Figure 1). In 2012, the industrial sector was accountable for 33 percent of the overall energy usage in the United States. Until 2040, this share is expected to reach 36 percent. Of the total amount of energy used in the US industrial sector, 11 percent were used in the form of electricity in 2012, and this share is expected to remain constant until 2040 (U.S. Energy Information Administration, 2015).

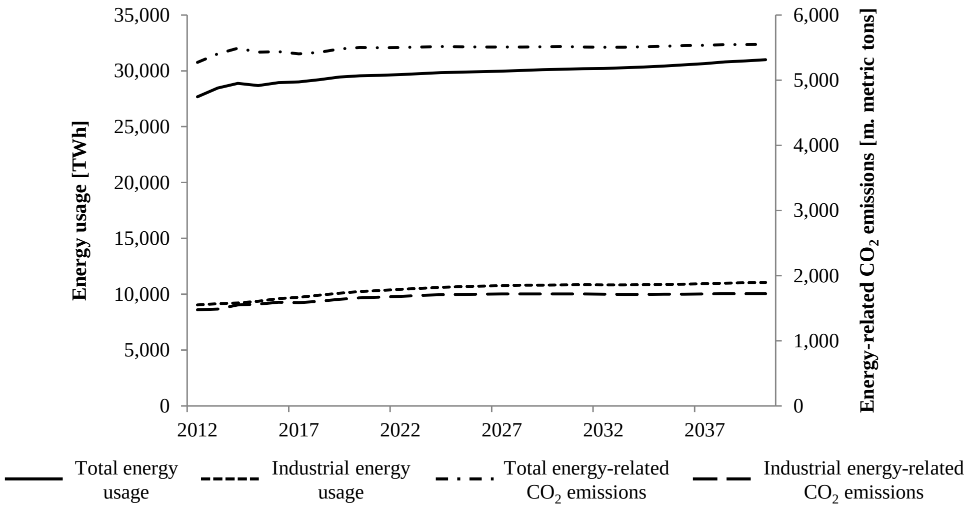


Figure 1: Energy usage and energy-related CO_2 emissions by end use in the United States (U.S. Energy Information Administration, 2015).

Figure 2 illustrates that 65 percent of the energy delivered to the industrial sector in the United States was used for producing heat and power for manufacturing processes in 2012. Another 17 percent were used for the production of heat and power for non-manufacturing processes, while the remaining 18 percent were spent on non-fuel uses (U.S. Energy Information Administration, 2014). These figures show that the industrial sector already has a major impact on the overall energy usage today and that this impact will increase in the future. Additionally, Figure 2 reveals that energy efficiency measures aiming at the reduction of heat and power demand in production provide a promising approach to lower energy usage in manufacturing.

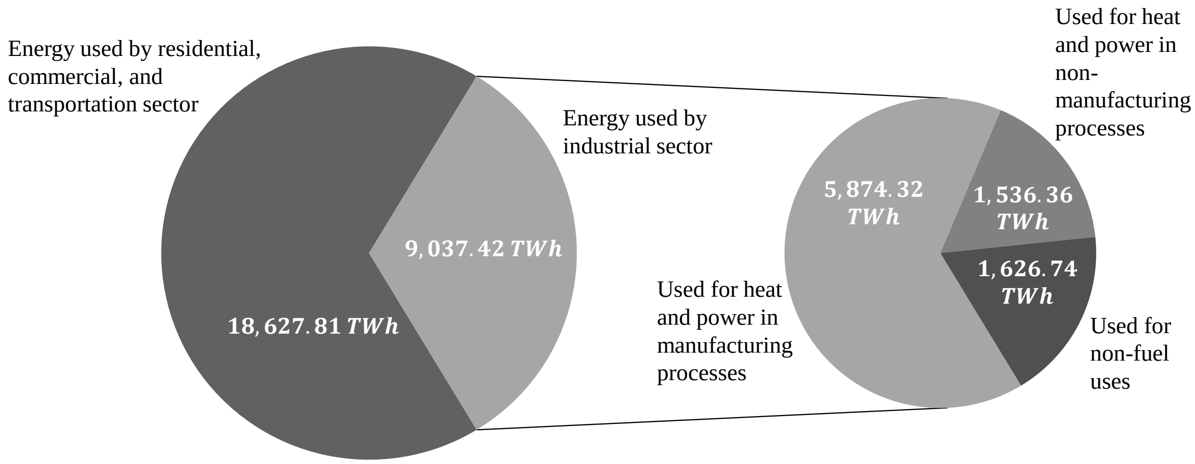


Figure 2: Distribution of energy usage by end use in the United States in 2012 with special focus on the industrial sector (U.S. Energy Information Administration, 2014).

Just as in the case of energy usage, the industrial sector is responsible for the largest share of energy-related CO_2 emissions. Figure 1 shows that while the industrial sector was accountable for 28 percent of the overall energy-related CO_2 emissions in 2012, it is expected to be responsible for 31 percent of the overall energy-related CO_2 emissions in 2040. In total, industrial energy-related CO_2 emissions are expected to increase by 0.6 percent each year from 2012 until 2040. It is evident that manufacturing activities are responsible for the vast majority of the energy-related CO_2 emissions in the industrial sector (1270 out of 1486 million metric tons (mt) in 2013 and 1383 out of 1723 million mt in 2040). At the same time, the use of electricity to run industrial processes causes the largest amount of emissions of all energy sources utilized in the industrial sector (531 million mt in 2013 and 592 million mt in 2040) (U.S. Energy Information Administration, 2015).

As a consequence of increasing energy usage and energy-related CO_2 emissions, the costs of electric power generation and transmission and distribution are expected to grow, leading to an expected increase of 18 percent in the average retail price of electricity in the United States until 2040 (U.S. Energy Information Administration, 2015). To smooth energy usage throughout the day, utility companies have established time-of-use (TOU) pricing profiles that penalize energy usage in periods of high total energy demand. This helps to avoid peak loads, which are a main cost driver of utility companies. According to Moon and Park (2014), electricity in high-demand – so-called on-peak – periods may even reach a price level that is eight times higher than the price level in low-demand – so-called off-peak – periods.

In light of these figures, it is not surprising that both researchers and practitioners have directed their attention to the development of approaches that help to increase energy efficiency in manufacturing in recent years. In 2012, between USD 310 and 360 billion were invested worldwide to improve energy efficiency of buildings, domestic appliances, transport and industry. These investments, which have partly been initiated by governmental directives (European Parliament and Council, 2012), exceeded supply-side investments in renewable electricity or in coal, oil and gas electricity generation (International Energy Agency, 2014). In 2011, continued improvements in energy efficiency led to energy savings of 16 TWh in eleven OECD countries; these energy savings exceeded the total final usage from all energy sources across the entire European Union. Additionally, the amount of energy saved through energy efficiency measures even surpassed the total final usage from any single fuel source in these countries in 2011 (International Energy Agency, 2014).

The largest share of improvements in energy efficiency has been achieved in the residential sector. Researchers have recently increased their efforts to close this gap for the industrial sector. Topics that have been studied in the literature include, for example, the energy usage of machines in different operating modes (e.g., Gutowski et al., 2006; Balogun et al., 2014) or changes in optimal lot sizing policies in case energy usage of the production equipment is considered (e.g., Zanoni et al., 2014; Bazan et al., 2015). Interestingly, the technical equipment that is available for increasing energy efficiency in production has thus far not received much attention in the literature on production planning. Systems that convert waste heat into usable electricity have been available on the market for a couple of years by now, and they are used successfully in many companies to reduce energy usage. In addition, more and more companies engage in the development of energy storage systems that can be used to store energy in times of oversupply (or low energy prices), such that the stored energy can be used in times of high energy demand. Even though there are a few works that study the behavior of such systems, it has remained largely unexplored under which conditions the use of such systems is economically beneficial, and how they impact the way production processes should be implemented. The paper at hand contributes to closing this research gap by developing the – to the best of the authors' knowledge – first decision support model for establishing a production plan for a serial multi-stage production system that uses a waste heat recovery system (WHRS) and an electrical energy storage system (EESS) subject to time-varying energy prices. Hence, the contribution of this paper is twofold. First, it integrates a WHRS along with an EESS into a serial multi-stage production system and thereby contributes to the interdisciplinary stream of research that connects production management from a managerial point of view to recent technical innovations fostering energy efficiency in manufacturing. Secondly, the model developed in this paper reveals for which daily production targets, energy usage patterns, and TOU tariffs, the WHRS combined with the EESS is most effective in reducing energy usage and consequently energy-related cost.

The remainder of the paper is organized as follows: Section 2 reviews the literature on energy efficient production management. Section 3 describes the problem studied, while Section 4 develops a mixed integer linear programming (MILP) model that integrates a WHRS and an EESS into a serial multi-stage production system. Section 5 presents a numerical analysis and investigates under which conditions it is economical to use a WHRS coupled with an EESS to enhance the energy efficiency of the production system. Section 6 derives managerial implications and concludes the paper.

2 Literature review

The literature relevant to the paper at hand can be divided into two domains of research. The first domain studies opportunities for improving energy efficiency in manufacturing from the perspective of production engineering. Works that fall into this area modify machine tools to lower friction losses during the production process or to reduce damping losses by optimizing the mechanical components of the machine tools (Neugebauer et al., 2011). A recent overview of energy-efficient (EE) production technologies is contained in the work of Hasanbeigi and Price (2012). Even though this comprehensive review concentrates on the textile industry, it refers the reader to numerous studies focusing on other industries. Quader et al. (2015) provided a state-of-the-art review of technological advancements fostering energy efficiency in the iron and steel industry. Madan et al. (2015) developed a step-by-step guideline of how technological improvements striving for more EE manufacturing may be introduced in a production process using the example of injection molding. Pons et al. (2013) finally measured how the implementation of such EE manufacturing technologies impacts a firm's overall performance.

The second domain of research relevant to this paper adopts an economic perspective and investigates how managerial actions may affect the energy usage of production systems. Gahm et al. (2016), for example, developed a framework which categorizes the literature on EE scheduling with regard to three major categories: (I) the location within the energy conversion chain where the managerial measures actually improve energy efficiency, (II) the way energy is supplied to the production system, and (III) the processes energy is required for. This research domain can be traced back to Denton et al. (1987), Nilsson (1993), and Nilsson and Söderström (1993), who were the first to investigate the impact of time-varying energy prices on production planning. Starting with these works, a rich literature on energy-aware production management has evolved and developed into various directions, with the most prominent ones being (I) machine scheduling, (II) load management, (III) lot sizing, and (IV) aggregate production planning.

The first research stream within the domain of energy-aware production management, machine scheduling, aims at computing short-term production schedules while taking account not only of traditional scheduling objectives, such as the minimization of the total weighted tardiness or the total completion time, but also of energy efficiency considerations, such as the minimization of energy usage or peak power usage. He et al. (2005) were among the first to consider energy usage of machines in a job shop scheduling problem. However, as they solely focused on energy usage instead of energy cost, they neglected the impact of time-varying energy prices. Mouzon et al. (2007) only concentrated on energy usage as well and used dispatching rules to schedule non-bottleneck machines in an energy-efficient way. Castro et al. (2009) introduced time-varying electricity prices into the scheduling of continuous plants and showed that considering energy-related cost in scheduling can result in cost savings up to 20 percent. Ghobeity and Mitsos (2010) showed that energy-aware scheduling is not only relevant in manufacturing, but that it is also of value in areas such as the seawater reverse osmosis. They found that the cost-saving potential is particularly high in the case of strongly varying energy prices. They referred to modern desalination plants, which produce both electricity and water, and suggested using these cogeneration plants to provide electricity to the grid in periods of high energy cost and using them for desalination in periods of low energy prices. Weinert et al. (2011) took account of energy usage in production processes by considering the energy used in a production operating mode as a single energy block. Based on this concept, they developed a methodology to rearrange these energy blocks to improve the energy efficiency in the considered production system. Bruzzone et al. (2012) studied a flexible flow shop model of a machining process subject to a maximum power demand constraint. Another work in this area is the one of Shrouf et al. (2014), who minimized the total energy usage in the single machine production scheduling problem considering different operating modes associated with different levels of energy usage. Zhang et al. (2014) investigated the flow shop scheduling problem in light of real-time electricity pricing.

The second research stream within the domain of energy-aware production management that is of relevance to this work, load management, aims at adjusting the load curve in such a way that production in on-peak periods is reduced and energy demand peaks are avoided. Ashok and Banerjee (2001) investigated how manufacturing companies may respond to TOU tariffs through industrial load management. Chao and Chen (2005) investigated whether it pays off for manufacturing companies to schedule their production in accordance with load management programs. To this end, the authors defined production and shutdown policies for randomly distributed on-peak periods. Ashok (2006) studied load management in steel plants and found that both manufacturing and utility companies can significantly reduce their energy-related costs through the participation in load management programs. Fernandez et al. (2013) and Sun et al. (2014) developed innovative approaches to reduce peak loads by accumulating a so-called 'Just-for-Peak' buffer prior to on-peak periods. During on-peak periods, production can be maintained by feeding production stages from the 'Just-for-Peak' buffer, while some production stages can be turned off to reduce energy usage in this period.

The third research stream within the domain of energy-aware production management, lot sizing, has just recently started to consider energy aspects. Zanoni et al. (2014) studied a two-stage production system and considered the energy usage of both stages during different operating modes. They compared alternative batch sizing policies, where multiple batches are either produced consecutively without or with interruption, and studied how the different batch sizing policies impact both the energy usage and the inventory and setup costs of the system. This paper was extended by Bazan et al. (2015), who added the generation of greenhouse gas emissions to the model and studied how a multi-level emission-taxing scheme influences the coordination of the system. Biel and Glock (2016) introduced a WHRS into the production system studied in Zanoni et al. (2014) and thus connected the lot sizing model to the technical process of recovering waste heat rejected during manufacturing to enhance the efficiency of production processes.

The last research stream within the domain of energy-aware production management relevant to this paper, aggregate production planning, studies the coordination of multi-stage production systems by deciding during which time intervals a production stage should be set up, idle, producing, or switched off. Mitra et al. (2012) developed a model to determine the optimal operating policies for continuous power-intensive processes subject to hourly-changing electricity prices. They considered different operating modes of the production system. Yet, they neglected the idle operating mode in which production stages remain in a 'ready-to-produce' state. Li and Sun (2013) studied a serial multi-stage production system with intermediate buffers. Production energy usage was dynamically controlled in their paper using Markov decision processes. TOU tariffs, in contrast, were not considered. This paper was extended by Sun and Li (2014), who considered four dimensions of the production system state: machine operation state, machine energy usage state, machine production activity state, and buffer state. Wang and Li (2013) also studied a serial multi-stage production system with intermediate buffers, but paid particular attention to the impact of time-dependent electricity prices and varying machine reliabilities on the production plan. Bego et al. (2014) studied a similar production system and developed a methodology that shows how companies should react to a different pricing profile termed Critical Peak Pricing (CPP)[2].

Our review of the literature shows that a rich stream of research exists in the area of energy-aware production management, and that several papers have taken account of time-varying electricity prices. Yet, the paper most relevant to our work is the one of Moon and Park (2014). They investigated the flexible job shop scheduling problem under time-varying and machine-dependent electricity cost. In addition, they integrated distributed energy resources and an energy storage system into their model. However, they assumed that idle machines do not consume energy. This assumption contrasts strongly with the findings of Gutowski et al. (2006), who showed that a significant amount of energy is required to keep machines in a 'ready-to-produce' state. Secondly, the energy storage system Moon and Park (2014) investigated in their model was charged by renewable energy resources, whose charging potential is hard to predict and to control. Furthermore, production activities and the energy potential to charge the energy storage system are unrelated. In contrast to Moon and Park (2014), the EESS in

[2] CPP programs are an extension of TOU tariffs and consider CPP periods in addition to off-, mid-, and on-peak periods. The occurrence of CPP periods is usually announced by the utilities on short notice in consequence of weather forecasts, power demand on the previous day, etc. (Bego et al., 2014).

the production system considered in this paper is charged with energy recovered from the production process using a WHRS. Thus, we establish a direct link between production activities and the amount of energy available to charge the EESS. Additionally, we model the processes of charging and discharging the EESS in more technical detail compared to Moon and Park (2014). Thus, we pay special attention to capacity and performance restrictions of the EESS and the way they impact the amount of energy that can be charged to or discharged from the EESS.

Research on the technological background of WHRSs and EESSs are also of relevance to this paper. Yet, as we introduce the functioning of those systems in Section 3, we refer the reader to the literature cited in this section.

3 Problem description

This study aims to develop a decision support model for establishing a production plan for a serial multi-stage production system that uses a WHRS and an EESS subject to time-varying energy prices. To this end, we first describe the production system (see Section 3.1) and discuss the operating modes of the production stages (see Section 3.2) and the energy usage attached to these operating modes (see Section 3.3). Subsequently, we integrate the WHRS (see Section 3.4) and the EESS (see Section 3.5) into the multi-stage production system.

3.1 Description of the production system and terminology

This paper investigates a serial multi-stage production system that transforms raw materials into finished products in m successive production steps. We assume that the demand for the finished product is known for the considered planning horizon, and that the first production stage has unlimited access to raw materials. Each unit is processed consecutively on every production stage. Units that have been processed on production stage i are stored in buffer stock i at the downstream side of this stage until they are processed on stage $i + 1$. Units that have been processed on the final production stage m are consumed immediately by the final customers. Figure 3 illustrates the structure of this production system.

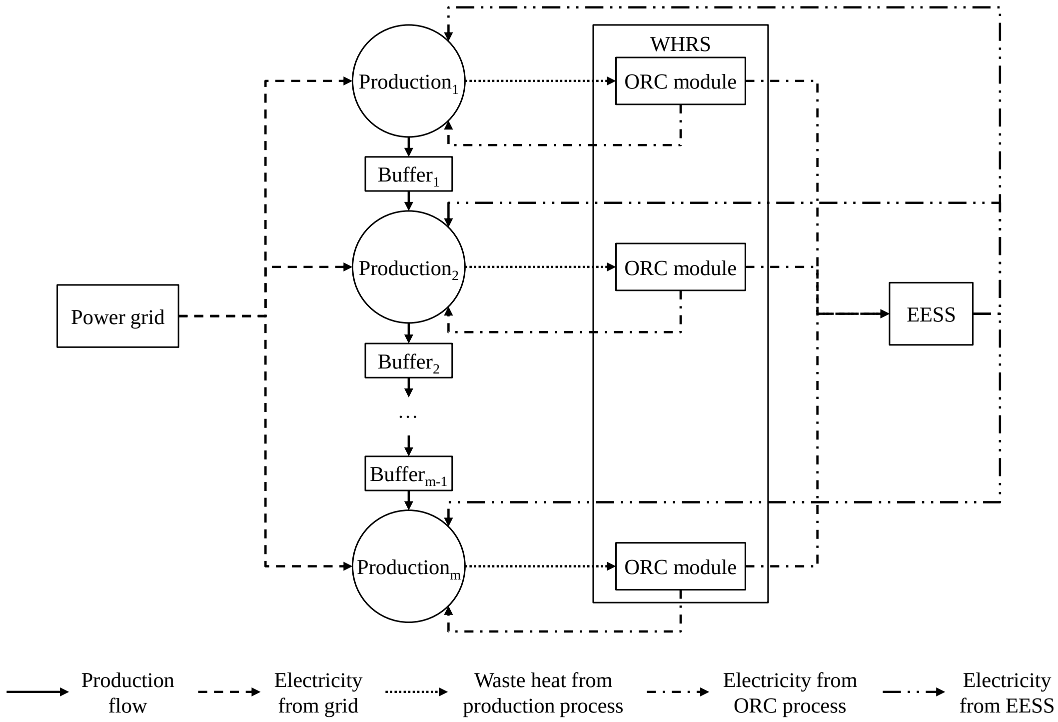


Figure 3: Production and energy flow in a serial multi-stage production system with WHRS and EESS.

In addition to the production flow, we include the energy flow into the coordination of the production system as our intention is to enhance the energy efficiency of the manufacturing process and to compute a production plan which optimally reacts to time-varying energy prices. Figure 3 shows that the energy required to run the production stages in different operating modes is not only provided by the grid (left part of the figure). Instead, the energy supply is supported by decentralized Organic Rankine Cycle (ORC) modules (center of the figure), which (partly) convert waste heat rejected by the production stages into electricity. To improve the use of waste heat over time, electrical energy can be stored in an EESS (right part of the figure). This way, the utilization of the electricity initially drawn from the grid is enhanced and the overall energy usage as well as the energy-related CO_2 emission are reduced. Details on how the waste heat recovery and the energy storage processes are incorporated into the MILP model used to derive a cost-optimal production plan for the production system considered are presented in Sections 3.4 and 3.5.

Throughout the paper, the following terminology is used:

Definitions

$Charge$	Scenario where the EESS is charged with energy
$Discharge$	Scenario where energy is discharged from the EESS
$Idle$	Idle operating mode
Off	Operating mode where the production stage is turned off
$Prod$	Production operating mode
$Setup$	Setup operating mode
$Store$	Scenario where the EESS stores the energy it already contains

Sets

ES	States of the EESS with $ES = \{Charge, Discharge, Store\}$
$OFFP$	Off-peak time slots
OM	Operating modes of the production stages with $OM = \{Idle, Off, Prod, Setup\}$
ONP	On-peak time slots

Indices

g	Shift with $g \in \{1,2,\dots,G\}$
i	Production stage with $i \in \{1,2,\dots,m\}$
r, s	States of the production stages with $r, s \in OM$
t	Time slot with $t \in \{1,2,\dots,T\}$
u	State of the EESS with $u \in ES$

Parameters

B_i^{ini}	Inventory level of the buffer stock at the downstream side of production stage i at the beginning of the planning horizon [ton]
B_i^{max}	Capacity (maximum inventory) of the buffer stock at the downstream side of production stage i [ton]
c^{EESS}	Cost of discharging energy from the EESS [EUR/kWh]
c^{ORC}	Cost of generating electricity by the ORC [EUR/kWh]
c_t	Energy usage charge in time slot t [EUR/kWh]
DOD	Maximum depth of discharge of the EESS [%]
d^{offp}	Power demand charge for maximum average power demand during off-peak time slots [EUR/kW]
d^{onp}	Power demand charge for maximum average power demand during on-peak time slots [EUR/kW]
E^{ini}	Amount of energy stored in the EESS at the beginning of the planning horizon [kWh]
E^{max}	Capacity of the EESS [kWh]
$EC_{i,r}$	Amount of energy consumed by production stage i in operating mode r [kWh]
f_i	Multiplication factor for the power required for setting up production stage i [-]
G	Total number of shifts [-]
h_1^{WH}	Specific enthalpy of the waste heat stream before transferring heat to the organic fluid [kJ/kg]
h_2^{WH}	Specific enthalpy of the waste heat stream after transferring heat to the organic fluid [kJ/kg]
$HR_{i,r}$	Amount of waste heat recovered on production stage i in operating mode r [kWh]
hc_i	Holding cost of the buffer stock at the downstream side of production stage i [EUR/(ton·h)]
k_i	Energy required at production stage i to produce one unit [kW/ton]
l	Length of a time slot [h]
m	Total number of production stages [-]

$\dot{m}^{WH}$	Mass flow rate of the waste heat stream [kg/h]
P^{max}	Maximum power (performance) energy can be charged to or discharged from the EESS [kW]
$p_{i,r}$	Production rate of production stage i in operating mode r [ton/h]
PR^{min}	Minimum number of units that need to be manufactured in the planning horizon [ton]
$\dot{Q}^{WH}$	Flow rate of rejected waste heat [kW]
sc_i	Cost for setting up production stage i [EUR/setup]
T	Total number of time slots [-]
T^C	Temperature of the cold reservoir (here: temperature at which the ORC rejects heat to the environment in the condenser) [K]
T^H	Temperature of the hot reservoir (here: temperature of the waste heat) [K]
T_1^{WH}	Temperature of the waste heat stream before transferring heat to the organic fluid [K]
T_2^{WH}	Temperature of the waste heat stream after transferring heat to the organic fluid [K]
W_i	Idle power of production stage i [kW]
w_t	Cost of labor in time slot t [EUR/h]
$wr_{i,r}$	Number of workers required at production stage i in operating mode r [-]
α	Percentage of the initial energy input rejected as waste heat by the production stages and led into the WHRS [%]
$\xi_{g,t}$	Binary variable that equals 1 if time slot t belongs to shift g, and 0 otherwise [-]
η^{Carnot}	Carnot efficiency [%]
η^{EESS}	Conversion efficiency of the EESS [%]
η^{ORC}	Efficiency of the ORC [%]

Decision variables

$B_{i,t}$	Inventory level of the buffer stock at the downstream side of production stage i at the end of time slot t [ton]
C_t	Amount of energy charged to the EESS in time slot t [kWh]
D_t	Amount of energy discharged from the EESS in time slot t [kWh]
DRC	Power demand-related cost [EUR]
E_t	State of charge of the EESS at the end of time slot t [kWh]
ERC	Total energy-related cost [EUR]
ERC^{EESS}	Energy-related cost for running the EESS [EUR]
ERC^{OM}	Energy-related cost for operating the production stages [EUR]
ERC^{WHRS}	Energy-related cost for running the WHRS [EUR]
ERS^{EESS}	Energy-related cost savings generated by using the EESS [EUR]
ERS^{WHRS}	Energy-related cost savings generated by using the WHRS [EUR]
H_t	Amount of energy recovered by the WHRS and immediately used for supporting the energy supply of the production stages in time slot t [kWh]
IHC	Total inventory holding cost [EUR]
LC	Total labor cost [EUR]

PD^{offp}	Maximum average power demand within all off-peak time slots of the planning horizon [kW]
PD^{onp}	Maximum average power demand within all on-peak time slots of the planning horizon [kW]
$S_{u,t}^{EESS}$	Binary variable that equals 1 if the EESS is in state u in time slot t, and 0 otherwise [-]
$S_{i,r,t}^{PS}$	Binary variable that equals 1 if production stage i is in state r in time slot t, and 0 otherwise [-]
SC	Total setup cost [EUR]
TC	Total cost of the production system [EUR]
WF_g	Number of workers employed in shift g [-]
$\chi_{i,r,u,t}$	Auxiliary variable to linearize $S_{i,r,t}^{PS} \cdot S_{u,t}^{EESS}$ [-]

3.2 Operating mode transitions

After a description of the general structure of the production system in Section 3.1, we next elaborate on the energy-related details of the integrated production system. Each production stage can be in one of four operating modes. It can (I) be switched off, (II) being set up, (III) be idle (while being switched on), or (IV) be producing. Possible transitions between the operating modes are illustrated in Figure 4. We note that only the transitions depicted as solid lines are used in normal operation. The dashed lines represent transitions which are technologically feasible, but irrational from an economic point of view in normal operation. For example, it would be possible to switch on and set up a production stage and then to directly switch if off again; however, as this would only lead to setup cost without production of output, we assume in the following that such transitions are not realized. When a production stage is switched off, it can either remain switched off or move into the setup operating mode that prepares the production stage for production. In the setup operating mode, a production stage is retooled depending on the production requirements. From the setup operating mode, a production stage can either move into the idle operating mode or directly into the production operating mode. The first transition is rather rare in practice, but may make sense in order to avoid energy-intensive setups in time periods of high energy costs. From the idle operating mode, a production stage can either move into the production operating mode or remain idle. Finally, a production stage that produces can either be switched off, be moved into the idle operating mode, or be kept in the production mode.

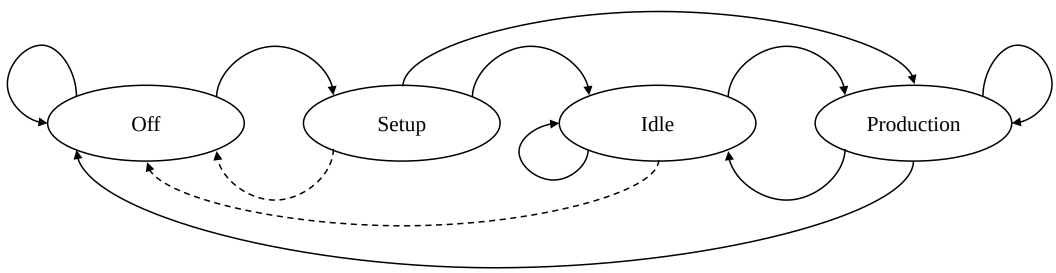


Figure 4: Possible transitions between the operating modes of a production stage.

3.3 Energy usage in different operating modes

We assume that all production stages are run using electric power. Depending on the operating mode, a production stage consumes a certain amount of electrical energy. If production stage i is switched off, it does not consume energy at all. If production stage i is idle, it can move into the production operating mode immediately. To facilitate this quick transition, production stage i is kept in a 'ready-to-produce' mode that consumes a constant amount of electrical energy (Gutowski et al., 2006):

$$EC_{i,Idle} = W_i \cdot l, \forall i \in \{1,2,\dots,m\}. \tag{1}$$

If production stage i produces, it does not only consume power in the amount of W_i kW to keep the production stage in a 'ready-to-produce' mode, but instead also requires a specific amount of electrical energy to perform the operations, where the energy usage is proportional to the production rate $p_{i,Prod}$ of the production stage (Gutowski et al., 2006):

$$EC_{i,Prod} = \left(W_i + k_i \cdot p_{i,Prod}\right) \cdot l, \forall i \in \{1,2,\dots,m\}. \tag{2}$$

The amount of energy required to set up production stage i can be assumed constant and generally larger than the amount of energy consumed in the idle operating mode. According to Zanoni et al. (2014), it can be modeled as

$$EC_{i,Setup} = f_i \cdot W_i \cdot l, \forall i \in \{1,2,\dots,m\}, \tag{3}$$

where $f_i > 1$ represents the factor by which $EC_{i,Setup}$ exceeds $EC_{i,Idle}$.

3.4 Integration of WHRS into the serial multi-stage production system

This paper analyzes under which conditions it is economically beneficial to reuse waste heat each production stage rejects when being set up, idle, or producing. In practice, there are three prominent ways of utilizing waste heat to enhance the energy efficiency of production processes. First, waste heat may be used for space heating or process heat in other production processes. Secondly, it may be used for heat pumps or absorption refrigerators where it lowers the heat initially required to take a working fluid from a lower to a higher temperature level, or vice versa. Lastly, it may be converted into electricity (Pehnt et al., 2010). In this paper, we exclusively focus on the latter use. For information on the other two uses, the reader is referred to Fang et al. (2013) and Liu et al. (2014).

3.4.1 Waste heat recovery using Organic Ranking Cycles

Several technical processes have been developed to convert waste heat into electricity (Haddad et al., 2014). The performance of these procedures varies with the temperature of the waste heat inserted into the system. In this paper, we limit our analysis to industrial production processes where the temperature of the waste heat typically does not exceed 370 °C. In this situation, research suggests to employ an ORC to convert the waste heat rejected by the production stages into electricity as this procedure performs especially well in the case of low-temperature waste heat due to its flexibility as well as its high safety and low maintenance requirements (Wei et al., 2007). The main components of an ORC are a pump, an evaporator, a turbine with a generator, and a condenser (see Figure 5). The pump provides the evaporator with the organic fluid within the cycle. The evaporator heats and vaporizes the organic fluid utilizing the waste heat from the production processes. Then, the high-pressure vapor is expanded in the turbine, which drives the generator and thus converts mechanical energy into electrical energy. After the expansion, the low-pressure vapor is liquefied in the condenser and led to the pump to start a new cycle (Wei et al., 2007).

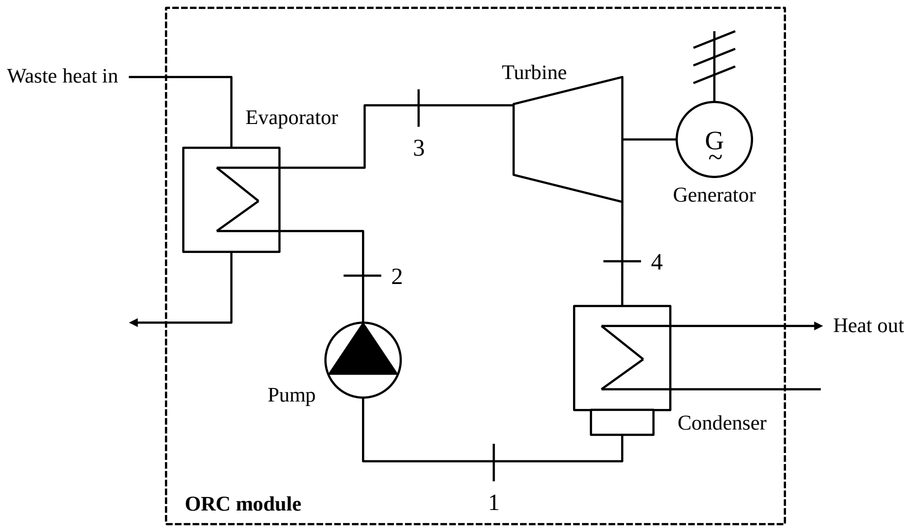


Figure 5: ORC system diagram.

The choice of the organic fluid to be used in the ORC is very important as it impacts the performance of the ORC from both a thermodynamic and an economic point of view. According to Invernizzi (2013), the selection process should take 'thermo-physical and thermodynamic properties, compatibility with the materials and the limits of thermal stability of the fluid, the health and safety characteristics, the fluid's availability and its cost' into account. Quoilin et al. (2011b) developed a model for optimally sizing ORCs with regard to these thermo-economic criteria and investigated their cycle performance for different organic fluids. In the paper at hand, we assume that the most suitable organic fluid for the application considered has been identified in advance, and thus exclude the selection of an appropriate fluid from the optimization problem. Further studies on the fluid selection for ORCs are those of Hung et al. (2010), Chen et al. (2010), and Wang et al. (2011).

The heat the waste heat stream transfers to the organic fluid, $\dot{Q}^{WH}$, can be calculated from the mass flow rate of the waste heat stream, $\dot{m}^{WH}$, and the thermodynamic potential, which is referred to as enthalpy, of the waste heat stream before (h_1^{WH}) and after (h_2^{WH}) transferring heat to the organic fluid:

$$\dot{Q}^{WH} = \dot{m}^{WH} \cdot (h_1^{WH} - h_2^{WH}). \tag{4}$$

The ORC can only convert a fraction of this waste heat added to the cycle into electrical energy. This fraction is referred to as the efficiency of the ORC, η^{ORC}, and it is limited by the Carnot efficiency, η^{Carnot}. The Carnot efficiency depends on the temperature of the hot reservoir, T^H (here: the temperature of the waste heat), and the temperature of the cold reservoir, T^C (here: the temperature at which the ORC rejects heat to the environment in the condenser):

$$\eta^{Carnot} = 1 - \frac{T^C}{T^H}. \tag{5}$$

In the present case, T^H ranges between the temperature of the waste heat stream before (T_1^{WH}) and after (T_2^{WH}) transferring heat to the organic fluid, and T^C is larger than or equal to the ambient temperature (Invernizzi, 2013). Hence, the mass flow and the composition of the waste heat stream as well as the difference between the temperature of the waste heat stream and the ambient temperature are of great importance to the efficiency of a WHRS. For a more detailed derivation of these relations, the reader is referred to Invernizzi (2013).

In general, the design of an ORC is optimized for a specific operating point (nominal load). Thus, the highest efficiency of converting waste heat into electrical energy in an ORC is achieved for a specific waste heat stream, characterized by its mass flow, its composition, and its temperature. Deviations from this waste heat stream in the form of partial loads result in a lower energy conversion efficiency of the ORC (Ibarra et al., 2014). Yet, several studies have shown that – in the case of ORCs – this decrease in conversion efficiency is rather small, such that the performance of ORCs is very robust against changes in partial loads. For instance, Obernberger et al. (2002) illustrated that the efficiency of the investigated ORC decreased only by 1.5 percentage points at a partial load of 50 percent from an efficiency of 18 percent at nominal load. Althaus et al. (2013) investigated the performance of six ORCs under varying waste heat streams and found that the efficiencies of the ORCs studied were neither at nominal load subject to large variations, nor at partial loads. Their final report stated that the efficiencies of the ORCs were approximately constant across partial loads of 40 to 50 percent of the nominal loads (Althaus et al., 2013). Thus, we can proceed on the assumption that the ORC can handle the variation of the waste heat streams resulting from the different modes production stages may operate in very well. For this reason, we neglect the slight performance drops of the ORC at partial loads in the developed model and assume a constant efficiency of the ORC, η^{ORC}.

Hereafter, the ORC is treated as a black box which merely converts the waste heat from the production stages into electricity. The amount of net energy the waste heat is converted into, $HR_{i,r}$, solely depends on (I) the amount of electrical energy consumed by production stage i, $EC_{i,r}$, (II) the fraction of this amount of energy that is rejected as waste heat over the course of the operation of the production stage and led into the WHRS, α, and (III) the efficiency of the ORC, η^{ORC}:

$$HR_{i,r} = \alpha \cdot \eta^{ORC} \cdot EC_{i,r}, \forall i \in \{1,2,\dots,m\}, r \in \{Idle, Prod, Setup\}. \tag{6}$$

For more information on the technological aspects of this conversion process, the reader is referred to Biel and Glock (2016).

3.4.2 Layout of the decentralized WHRS

As described in Section 3.1, we consider a decentralized WHRS, where one ORC module is installed at each production stage. We focus on this layout option as a single centralized WHRS would be associated with far higher investment cost and would not make it possible to equip production stages with ORC modules stepwise, a solution that is especially attractive to small and medium-sized companies. Additionally, the influence of heat losses would considerable grow in a centralized WHRS as the distances the waste heat would have to be transported significantly increase when the production stages and the ORCs are located far apart from each other. Furthermore, a decentralized WHRS seems more appropriate in the production system studied here as the production stages operate intermittently at different production rates and in varying operating modes and consequently reject different waste heat streams at different points in time. Regarding the periods of time the production stages remain in an operating mode, we assume that these periods are always sufficiently long to ensure that the ORC reaches a steady state. Otherwise, the efficiency of the ORC would be reduced (Dai et al., 2009).

3.5 Integration of EESS into the serial multi-stage production system

Along with the WHRS, an EESS is used in the production system considered here. Figure 3 illustrates how the ORC modules and the EESS are integrated into the production system. In each time slot, the EESS can either be charged with energy, store the energy it already contains, or discharge energy. As a consequence, the current state of the EESS determines for which purpose the energy recovered by the WHRS is used and thus how the energy requirements of the production stages are satisfied. The option where the electricity generated from the ORC modules is fed into the grid in return for compensation is not considered in this paper. Possible scenarios that may arise are summarized in Table 1. For more information on EESSs, the reader is referred to Zakeri and Syri (2015) and Chen et al. (2009).

Table 1: States of the EESS and their impacts on the WHRS and the energy supply of the production system.

EESS …	Energy recovered by the WHRS …	Production stages are supplied with energy from …
... is being charged with energy.	*... is charged to the EESS.*	*... the grid.*
... is storing energy.	*... supports the energy supply of the production system.*	*... the grid and the WHRS.*
... is discharging energy.	*... supports the energy supply of the production system.*	*... the grid, the WHRS, and the EESS.*

4 Model description

This section combines the serial multi-stage production system with the WHRS and the EESS in a MILP model. Apart from what has already been stated, the developed model is based on the following assumptions:

(1) one unit of input material is required to produce one unit of the finished product at each production stage;
(2) product shortages are not allowed;
(3) the planning horizon is divided into T time slots of equal length;
(4) the T time slots are distributed among G shifts;
(5) each production stage can only be in one operating mode in each time slot. As a consequence, energy usage is constant over each time slot. To model workloads that vary strongly within a time slot in an application of the model, time slots could be split up further to arrive at relatively homogeneous periods again;
(6) a setup of each production stage requires one time slot;
(7) the TOU pricing profile considers only two types of periods: off-peak periods, where energy usage and demand charges are low, and on-peak periods, where energy usage and demand charges are high. Thus, each time slot belongs either to the off-peak period or to the on-peak period of the TOU pricing profile. Furthermore, off- and on-peak periods consist of an integer number of time intervals;

(8) cost of generating electricity from the ORC include the initial investment, operation and maintenance costs, cost of fuel, insurance, etc. (Tchanche et al., 2010);
(9) cost of discharging energy from the EESS reflect the investment cost of the EESS. An EESS is assumed to provide a fixed amount of complete cycles of charging and discharging the EESS. Hence, a fixed maximum amount of energy can be discharged from the EESS across its entire lifecycle. The cost of discharging energy from the EESS can be derived by dividing the investment cost of the EESS by this fixed maximum amount of dischargeable energy (Chen et al., 2009; Zakeri and Syri, 2015);
(10) the power that is applied to charging energy to or discharging energy from the EESS is constant within one time slot, such that energy is charged to or discharged from the EESS evenly across one time slot;
(11) self-discharge losses of the EESS can be neglected as we use a flow battery for the EESS, which features a small self-discharge ratio per day (Chen et al., 2009);
(12) the time slots are sufficiently long to ensure that the ORC modules reach a steady state;
(13) the efficiency of the ORC modules, η^{ORC}, is constant across varying waste heat streams. This may lead to a slight over- or underestimation of energy-related cost at partial loads. Yet, these over- or underestimations are expected to be negligible as several experiments have shown that the performance of the ORC is very robust against changes in partial loads (see Sect. 3.4.1 and Obernberger et al., 2002; Althaus et al., 2013; Quoilin et al., 2011a).

Additional assumptions will be introduced where required.

4.1 Objective function

The intention of the model is to find a strategy for operating the production stages and the EESS at optimal cost while meeting a given demand. Thus, the objective is to minimize the total cost of the production system, TC, consisting of the total inventory holding cost, IHC, the total setup cost, SC, the total labor cost, LC, as well as the total energy-related cost, ERC:

$$\min TC = IHC + SC + LC + ERC. \tag{7}$$

The total inventory holding cost, IHC, results from multiplying the time-weighted inventories in all buffer stocks $i \in \{1,2,\dots,m-1\}$ by the corresponding inventory holding cost hc_i across the entire planning horizon:

$$IHC = \frac{l}{2} \cdot \sum_{t\in\{1,2,\dots,T\}} \sum_{i\in\{1,2,\dots,m-1\}} \left(B_{i,t-1} + B_{i,t}\right) \cdot hc_i. \tag{8}$$

Every time a production stage i needs to be set up in the planning horizon, setup cost sc_i is incurred:

$$SC = \sum_{t\in\{1,2,\dots,T\}} \sum_{i\in\{1,2,\dots,m\}} S^{PS}_{i,Setup,t} \cdot sc_i. \tag{9}$$

Depending on the time of the day, the cost of labor, w_t, may vary as a consequence of surcharges for night work or work on Sundays or on holidays. Thus, the total labor cost, LC, results from adding up the workforce employed in a shift multiplied by the cost of labor in the time slots belonging to a shift and the length of the time slots, l, across all shifts:

$$LC = \sum_{g\in\{1,2,\dots,G\}} \sum_{t\in\{1,2,\dots,T\}} WF_g \cdot \xi_{g,t} \cdot w_t \cdot l. \tag{10}$$

The total energy-related cost, ERC, equals the sum of the energy-related cost for operating the production stages, ERC^{OM}, the power demand-related cost, DRC, the energy-related cost for running the WHRS, ERC^{WHRS}, and the energy-related cost for running the EESS, ERC^{EESS}, reduced by the energy-related cost savings generated by using the WHRS, ERS^{WHRS}, and the energy-related cost savings generated by using the EESS, ERS^{EESS}:

$$ERC = ERC^{OM} + DRC + ERC^{WHRS} + ERC^{EESS} - ERS^{WHRS} - ERS^{EESS}. \tag{11}$$

The energy-related cost for operating the production stages equals the energy usage of the production stages multiplied by the energy usage charge in time slot t across the entire planning horizon:

$$ERC^{OM} = \sum_{t\in\{1,2,\dots,T\}} \sum_{i\in\{1,2,\dots,m\}} \sum_{r\in\{Idle,Prod,Setup\}} EC_{i,r} \cdot S_{i,r,t}^{PS} \cdot c_t. \tag{12}$$

The power demand-related cost equals the sum of the maximum average power demand within all off-peak time slots, PD^{offp}, and all on-peak time slots, PD^{onp}, multiplied by the corresponding power demand charges:

$$DRC = d^{offp} \cdot PD^{offp} + d^{onp} \cdot PD^{onp}. \tag{13}$$

The energy-related cost for running the WHRS depends on the amount of energy generated using the WHRS in each time slot multiplied by the cost of generating electricity from the ORC across the entire planning horizon:

$$ERC^{WHRS} = c^{ORC} \cdot \sum_{t\in\{1,2,\dots,T\}} \sum_{i\in\{1,2,\dots,m\}} \sum_{r\in\{Idle,Prod,Setup\}} HR_{i,r} \cdot S_{i,r,t}^{PS}. \tag{14}$$

The energy-related cost for running the EESS equals the amount of energy discharged from the EESS in time slot t weighted by the cost of discharging energy from the EESS across the entire planning horizon:

$$ERC^{EESS} = c^{EESS} \cdot \sum_{t\in\{1,2,\dots,T\}} D_t. \tag{15}$$

The energy-related cost savings generated by using the WHRS equals the amount of energy recovered by the WHRS and immediately used for supporting the energy supply of the production stages in time slot t multiplied by the energy usage charge in the respective time slot across the entire planning horizon:

$$ERS^{WHRS} = \sum_{t\in\{1,2,\dots,T\}} H_t \cdot c_t. \tag{16}$$

The energy-related cost savings generated by using the EESS equals the amount of energy discharged from the EESS in time slot t multiplied by the energy usage charge in the respective time slot across the entire planning horizon:

$$ERS^{EESS} = \sum_{t\in\{1,2,\dots,T\}} D_t \cdot c_t. \tag{17}$$

4.2 Constraints

Minimizing the objective function is subject to several constraints. We categorize these constraints into (I) production control and workforce constraints, (II) inventory constraints, and (III) energy-related constraints.

4.2.1 Production control and workforce constraints

Equations (18) to (23) control the operating modes of the production stages and the workforce assignments:

$$\sum_{r \in OM} S_{i,r,t}^{PS} = 1, \forall i \in \{1,2,\dots,m\}, t \in \{1,2,\dots,T\} \tag{18}$$

$$\sum_{r \in \{Off,Setup\}} S_{i,r,1}^{PS} = 1, \forall i \in \{1,2,\dots,m\} \tag{19}$$

$$\sum_{r \in \{Idle,Prod\}} S_{i,Prod,t}^{PS} \leq \sum_{r \in \{Idle,Prod,Setup\}} S_{i,r,t-1}^{PS}, \forall i \in \{1,2,\dots,m\}, t \in \{2,3,\dots,T\} \tag{20}$$

$$\sum_{t \in \{1,2,\dots,T\}} S_{i,Prod,t}^{PS} \cdot p_{M,Prod} \geq PR^{min} \tag{21}$$

$$\sum_{i \in \{1,2,\dots,m\}} \sum_{r \in OM} S_{i,r,t}^{PS} \cdot wr_{i,r} \cdot \xi_{g,t} \leq WF_g, \forall g \in \{1,2,\dots,G\}, t \in \{1,2,\dots,T\} \tag{22}$$

$$S_{i,r,t}^{PS} \in \{0,1\}, \forall i \in \{1,2,\dots,m\}, r \in OM, t \in \{1,2,\dots,T\} \tag{23}$$

Equation (18) assures that each production stage is only in one operating mode in each time slot. Equation (19) determines that all production stages are switched off at the beginning of the planning horizon. Equation (20) guarantees that a production stage is set up before it can be in the operating modes $Idle$ or $Prod$. Equation (21) ensures that the minimum number of units demanded in the planning horizon is manufactured. Equation (22) assures that the workforce remains constant across a shift. Furthermore, Equation (22) guarantees that the workforce employed in a shift is sufficiently large such that all production stages, which require different numbers of workers depending on their operating mode r, $wr_{i,r}$, can operate in a given time slot of this shift. The number of workers required in an operating mode may thereby differ from production stage to production stage. Equation (23) assures the binary character of $S_{i,r,t}^{PS}$.

4.2.2 Inventory constraints

Equations (24) to (27) control the buffer stock levels in the production system:

$$B_{i,t} = B_{i,t-1} + S_{i,Prod,t}^{PS} \cdot p_{i,Prod} - S_{i+1,Prod,t}^{PS} \cdot p_{i,Prod}, \forall i \in \{1,2,\dots,m-1\}, \\ t \in \{1,2,\dots,T\} \tag{24}$$

$$B_{i,t} \geq 0, \forall i \in \{1,2,\dots,m-1\}, t \in \{1,2,\dots,T\} \tag{25}$$

$$B_{i,t} \leq B_i^{max}, \forall i \in \{1,2,\dots,m-1\}, t \in \{1,2,\dots,T\} \tag{26}$$

$$B_{i,0} = B_i^{ini}, \forall i \in \{1,2,\dots,m-1\} \tag{27}$$

Equation (24) is the flow balance equation that links the buffer stock levels in consecutive time slots. This is done by ensuring that the buffer stock level at the downstream side of production stage i at the end of time slot t equals the sum of the corresponding buffer stock level at the end of the preceding time slot $t-1$ and the number of units produced by production stage i in time slot t, reduced by the number of units produced (consumed) by the succeeding production stage $i+1$ in time slot t. Equation (25) prohibits shortages, and Equation (26) ensures that the capacity constraints of the buffer stocks are not violated. Equation (27) defines the buffer stock levels at the beginning of the planning horizon.

4.2.3 Energy-related constraints

Equations (28) to (50) specify how the energy requirements of the production system are satisfied:

$$PD^{offp} \geq \sum_{i\in\{1,2,\dots,m\}} \sum_{r\in\{Idle,Prod,Setup\}} \frac{EC_{i,r}}{l} \cdot \begin{pmatrix} S_{i,r,t}^{PS} - \alpha \cdot \eta^{ORC} \\ \cdot \sum_{u\in\{Discharge,Store\}} S_{i,r,t}^{PS} \cdot S_{u,t}^{EESS} \end{pmatrix} - \frac{D_t}{l}, \quad \forall t \in OFFP \tag{28}$$

$$PD^{offp} \geq \sum_{i\in\{1,2,\dots,m\}} \sum_{r\in\{Idle,Prod,Setup\}} \frac{EC_{i,r}}{l} \cdot \begin{pmatrix} S_{i,r,t}^{PS} - \alpha \cdot \eta^{ORC} \\ \cdot \sum_{u\in\{Discharge,Store\}} \chi_{i,r,u,t} \end{pmatrix} - \frac{D_t}{l}, \quad \forall t \in OFFP \tag{29}$$

$$PD^{onp} \geq \sum_{i\in\{1,2,\dots,m\}} \sum_{r\in\{Idle,Prod,Setup\}} \frac{EC_{i,r}}{l} \cdot \begin{pmatrix} S_{i,r,t}^{PS} - \alpha \cdot \eta^{ORC} \\ \cdot \sum_{u\in\{Discharge,Store\}} S_{i,r,t}^{PS} \cdot S_{u,t}^{EESS} \end{pmatrix} - \frac{D_t}{l}, \quad \forall t \in ONP \tag{30}$$

$$PD^{onp} \geq \sum_{i\in\{1,2,\dots,m\}} \sum_{r\in\{Idle,Prod,Setup\}} \frac{EC_{i,r}}{l} \cdot \begin{pmatrix} S_{i,r,t}^{PS} - \alpha \cdot \eta^{ORC} \\ \cdot \sum_{u\in\{Discharge,Store\}} \chi_{i,r,u,t} \end{pmatrix} - \frac{D_t}{l}, \quad \forall t \in ONP \tag{31}$$

$$\sum_{u\in ES} S_{u,t}^{EESS} = 1, \forall t \in \{1,2,\dots,T\} \tag{32}$$

$$\frac{1}{\eta^{EESS}} \cdot C_t + H_t = \alpha \cdot \eta^{ORC} \cdot \sum_{i\in\{1,2,\dots,m\}} \sum_{r\in\{Idle,Prod,Setup\}} S_{i,r,t}^{PS} \cdot EC_{i,r}, \forall t \in \{1,2,\dots,T\} \tag{33}$$

$$E_t = E_{t-1} + C_t - \frac{1}{\eta^{EESS}} \cdot D_t, \forall t \in \{1,2,\dots,T\} \tag{34}$$

$$E_t \geq (1-DOD) \cdot E^{max}, \forall t \in \{1,2,\dots,T\} \tag{35}$$

$$E_t \leq E^{max}, \forall t \in \{1,2,\dots,T\} \tag{36}$$

$$E_0 = E^{ini} \tag{37}$$

$$\frac{1}{\eta^{EESS}} \cdot C_t \leq S_{Charge,t}^{EESS} \cdot E^{max}, \forall t \in \{1,2,\dots,T\} \tag{38}$$

$$\frac{1}{\eta^{EESS}} \cdot D_t \leq S_{Discharge,t}^{EESS} \cdot E^{max}, \forall t \in \{1,2,\dots,T\} \tag{39}$$

$$\frac{1}{\eta^{EESS}} \cdot D_t \leq E_{t-1} - (1-DOD) \cdot E^{max}, \forall t \in \{1,2,\dots,T\} \tag{40}$$

$$D_t \leq (1-\alpha \cdot \eta^{ORC}) \cdot \sum_{i\in\{1,2,\dots,m\}} \sum_{r\in\{Idle,Prod,Setup\}} S_{i,r,t}^{PS} \cdot EC_{i,r}, \forall t \in \{1,2,\dots,T\} \tag{41}$$

$$\frac{1}{\eta^{EESS}} \cdot C_t \leq \sum_{i\in\{1,2,\dots,m\}} \sum_{r\in\{Idle,Prod,Setup\}} S_{i,r,t}^{PS} \cdot S_{Charge,t}^{EESS} \cdot HR_{i,r}, \forall t \in \{1,2,\dots,T\} \tag{42}$$

$$\frac{1}{\eta^{EESS}} \cdot C_t \leq \sum_{i \in \{1,2,\dots,m\}} \sum_{r \in \{Idle,Prod,Setup\}} \chi_{i,r,Charge,t} \cdot HR_{i,r}, \forall t \in \{1,2,\dots,T\} \tag{43}$$

$$\frac{1}{\eta^{EESS \cdot l}} \cdot C_t \leq P^{Max}, \forall t \in \{1,2,\dots,T\} \tag{44}$$

$$\frac{1}{\eta^{EESS \cdot l}} \cdot D_t \leq P^{Max}, \forall t \in \{1,2,\dots,T\} \tag{45}$$

$$C_t \geq 0, \forall t \in \{1,2,\dots,T\} \tag{46}$$

$$D_t \geq 0, \forall t \in \{1,2,\dots,T\} \tag{47}$$

$$H_t \geq 0, \forall t \in \{1,2,\dots,T\} \tag{48}$$

$$S_{u,t}^{EESS} \in \{0,1\}, \forall u \in ES, t \in \{1,2,\dots,T\} \tag{49}$$

$$\chi_{i,r,u,t} \in \{0,1\}, \forall i \in \{1,2,\dots,m\}, r \in OM, u \in ES, t \in \{1,2,\dots,T\} \tag{50}$$

Equations (28) and (30) compute the maximum average power demand within all off-peak and on-peak time slots, respectively. As Equations (28) and (30) are not linear, they need to be linearized to generate a MILP model. Linearization, which is described in the Appendix, leads to Equations (29) and (31), respectively. Equation (32) assures that the EESS is only in one state in each time slot. Equation (33) determines that the energy that is recovered by the WHRS equals the sum of the gross energy that is charged to the EESS and the energy that is recovered by the WHRS and immediately used for supporting the energy supply of the production stages in time slot t. Equation (34) is the flow balance equation that links the states of charge of the EESS in consecutive time slots. The flow balance equation determines that the state of charge of the EESS at the end of time slot t equals the sum of the state of charge of the EESS at the end of the preceding time slot $t-1$ and the amount of energy fed into the EESS in time slot t, reduced by the gross energy discharged from the EESS in time slot t. Equation (35) guarantees that the state of charge of the EESS never falls below the maximum depth of discharge, while Equation (36) ensures that the state of charge of the EESS does not exceed its maximum capacity. Equation (37) defines the state of charge of the EESS at the beginning of the planning horizon. Equation (38) guarantees that energy is only charged to the EESS when the EESS is in state $Charge$, while Equation (39) assures that energy is only discharged from the EESS when the EESS is in state $Discharge$. Equation (40) ensures that the gross energy discharged in time slot t is smaller than or equal to the state of charge at the end of the preceding time slot $t-1$ less the maximum depth of discharge. Equation (41) asserts that the amount of energy discharged from the EESS in time slot t is smaller than or equal to the amount of energy required by the production system in time slot t. Equation (42) determines that the gross energy charged to the EESS is smaller than or equal to the energy recovered by the WHRS in time slot t. Since Equation (42) is not linear, it needs to be linearized to generate a MILP model. Linearization, which is described in the Appendix, leads to Equation (43). Equations (44) and (45) ensure that the power that is applied to charging energy to or discharging energy from the EESS lies below or equals the maximum charging/discharging power the EESS can provide. Finally, Equations (46) to (48) determine the non-negativity of the amount of energy that is charged to the EESS, the amount of energy that is discharged from the EESS, and the amount of energy that is recovered by the WHRS and immediately used for supporting the energy supply of the production stages in time slot t, respectively. Equations (49) and (50) assure the binary character of $S_{u,t}^{EESS}$ and $\chi_{i,r,u,t}$, respectively.

5 Numerical analysis

To explore the properties of the developed model under realistic conditions, we conduct a numerical analysis. The primary aims of this numerical analysis are (I) to illustrate the effectiveness of the EESS in supplementing the impact of the WHRS on energy-related cost, (II) to examine how the system performs for production processes of different energy intensity and under varying total product demand, and (III) to investigate how sensitive production plans are to changes in the energy usage charge and the power demand charge.

5.1 Production scenario settings and energy tariff

Industrial waste heat recovery is of interest for all manufacturing processes that reject enough waste heat such that a WHRS can be operated cost-efficiently. In this respect, the cement industry, where the temperature of waste heat reaches between 250 and 400 °C, the steel-working industry, and the glass industry with waste heat stream temperatures of 400 to 600 °C are prime examples (Invernizzi, 2013). For this reason, we derive the parameter values of the production system and the parameters describing the energy usage in this numerical analysis from Ashok (2006), who studied peak-load management in steel plants. We confine ourselves to a segment of the steel plant consisting of a bloom caster (production stage 1), a bloom mill (production stage 2), and a wire mill (production stage 3). These three production stages result in a serial production system with two intermediate buffers. The production and energy usage data are summarized in Table 2.

Table 2: Production and energy usage data used in the numerical analysis.

PR^{min}	160 ton	$p_{3,Prod}$	18 ton/h	f_2	2
hc_1	0.04 EUR/(ton·h)	W_1	500 kW	f_3	2
hc_2	0.05 EUR/(ton·h)	W_2	2900 kW	B_1^{ini}	0 ton
sc_1	1000 EUR/setup	W_3	3900 kW	B_2^{ini}	0 ton
sc_2	1000 EUR/setup	k_1	7.5 kWh/ton	B_1^{max}	150 ton
sc_3	1000 EUR/setup	k_2	31.4 kWh/ton	B_2^{max}	150 ton
$p_{1,Prod}$	20 ton/h	k_3	66.7 kWh/ton		
$p_{2,Prod}$	35 ton/h	f_1	2		

The data describing the WHRS and the EESS are presented in Table 3. The cost of generating electricity from the ORC, c^{ORC}, is described in Tchanche et al. (2010), among others, who studied small-scale ORCs for recovering waste heat. The efficiency of the ORC modules, η^{ORC}, is set to a value of 30 percent, which is reasonable if the waste heat stream has a temperature around 350 °C and the heat sink has a temperature around 70 °C (Hirzel et al., 2013). To store the energy recovered by the WHRS, we use a flow battery. The cost of discharging energy from this EESS, c^{EESS}, is taken from Chen et al. (2009).

Table 3: WHRS and EESS data used in the numerical analysis.

α	0.3	η^{EESS}	0.85	DOD	0.9
η^{ORC}	0.3	c^{EESS}	0.15 EUR/kWh	E^{max}	1500 kWh
c^{ORC}	0.07 EUR/kWh	E^{ini}	50 kWh	P^{max}	1500 kW

In this numerical analysis, we use the same representative TOU pricing profile Wang and Li (2013) used for the summer season from June till September (Orange and Rockland Utilities, 2012). This pricing profile is outlined in Table 4.

Table 4: Representative TOU pricing profile for the summer season.

Type of period	Time of day	c_t [EUR/kWh]	d^{offp} [EUR/kWh]	d^{onp} [EUR/kWh]
Off-peak	7 p.m. – 1 p.m.	0.08274	0	-
On-peak	1 p.m. – 7 p.m.	0.1679	-	18.8

In addition to inventory holding cost, setup cost, and energy-related cost, we consider cost of labor to highlight the trade-off between producing during high-price energy periods and producing at nighttime when extra pay for night work is due. The number of workers required by the three production stages depending on the operating modes is summarized in Table 5.

Table 5: Number of workers required by the three production stages in different operating modes.

Operating mode	Production stage 1	Production stage 2	Production stage 3
$Idle$	0	0	0
$Prod$	2	6	4
Off	0	0	0
$Setup$	3	3	3

Apart from what has already been stated, the numerical analysis is based on the following assumptions:

(1) the length of a time slot and the length of a setup of each production stage are normalized to 1 h;
(2) the planning horizon corresponds to one workday comprising two 8 h shifts from 6 a.m. to 2 p.m. and from 2 p.m. to 10 p.m.;
(3) for each day, a production quantity that needs to be processed is given;
(4) assuming one month to consist of 21 working days, the power demand-related cost needs to be divided by 21 as we only consider one day in this numerical analysis and the power demand charge is only paid once per billing period with reference to the maximum on-peak power demand within the monthly billing period;

(5) in the time from 6 a.m. to 8 p.m., the cost of labor equals 26 EUR/h. From 8 p.m. to 6 a.m., a surcharge of 25 percent is paid, resulting in a cost of labor of 32.5 EUR/h (U.S. Bureau of Labor Statistics, 2013).

The model was solved using the software IBM ILOG CPLEX Optimization Studio 12.6.1 on an Intel® Core™ i7-4770 CPU @ 3.40 GHz with 16 GB RAM.

5.2 Results

As the setup cost, the inventory holding cost, and the labor cost were not affected by the use of the WHRS and the EESS, and to illustrate the effectiveness of the EESS in supplementing the impact of the WHRS, we focus on the energy-related cost in the first part of the analysis. Compared to the base scenario in which neither the WHRS nor the EESS are utilized, the sole implementation of the WHRS leads to energy-related cost savings of 5.5 percent. Adding the EESS almost doubles these cost savings, reaching 10.6 percent. Figure 6 illustrates how the WHRS and the WHRS combined with the EESS influence the different components of the energy-related cost. Costs arising from the energy usage charge ($ERC^{OM} - ERS^{WHRS} - ERS^{EESS}$) are reduced by a similar percentage in the case of the sole use of the WHRS (−9.0 percent) and in the case of the use of WHRS and EESS (−9.6 percent) compared to the base scenario. The major advantage of the combined use of WHRS and EESS is revealed when investigating the costs arising from the power demand charge, DRC. While the sole use of the WHRS reduces the power demand-related cost by 9.0 percent compared to the base scenario, the combined use of WHRS and EESS reduces the power demand-related cost by 23.7 percent compared to the base scenario.

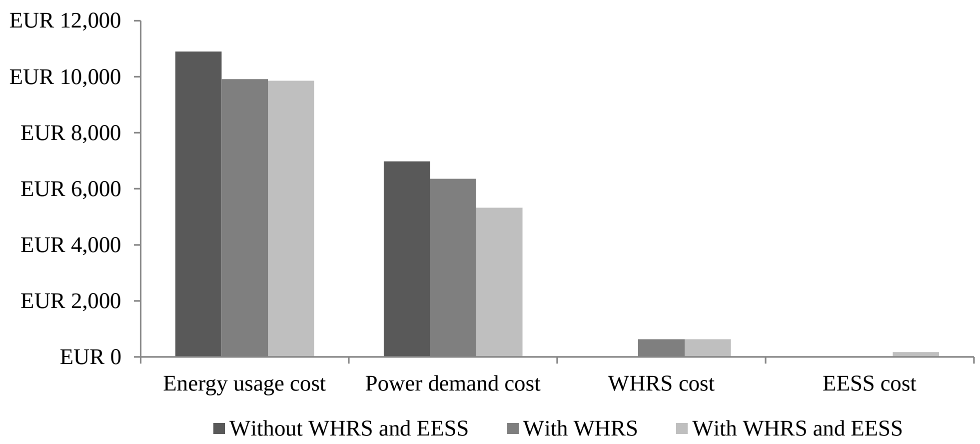


Figure 6: Energy-related cost dependent on the use of WHRS and EESS.

To examine how the system performs for production processes of different energy intensity, we varied the idle power of the production stages, W_i, as well as the energy required to produce one unit at the respective production stages, k_i, all by the same factor, starting from the base scenario EC^{base} = (W_1 = 500 kW, k_1 = 7.5 kWh/ton; W_2 = 2900 kW, k_2 = 31.4 kWh/ton; W_3 = 3900 kW, k_3 = 66.7 kWh/ton). Thereby, we hold the ratios of idle powers as well as of the amounts of energy required to produce one unit across the production stages constant. At the same time, we hold the ratio of idle power to the energy required to produce one unit constant for every production stage. As a consequence, we vary W_1 with a step width of 10 kW, W_2 with a step width of 58 kW, W_3 with a step width of 78 kW, k_1 with a step width of 0.15 kWh/ton, k_2 with a step width of 0.648 kWh/ton, and k_3 with a step width of 1.334 kWh/ton.

Figure 7 reveals that there seems to be a threshold energy usage of the production stages at which the total cost savings reach a maximum. This maximum in total cost savings is associated with the energy usage scenario EC^{opt} = (W_1 = 385 kW, k_1 = 5.775 kWh/ton; W_2 = 2233 kW, k_2 = 24.948 kWh/ton; W_3 = 3003 kW, k_3 = 51.359 kWh/ton) and amounts to −7.6 percent. In case the energy usage increases beyond EC^{opt}, the total cost savings decrease and reach a level of −5.3 percent in case of extremely energy-intensive production processes, where the savings potential of the WHRS combined with the EESS is completely exploited.

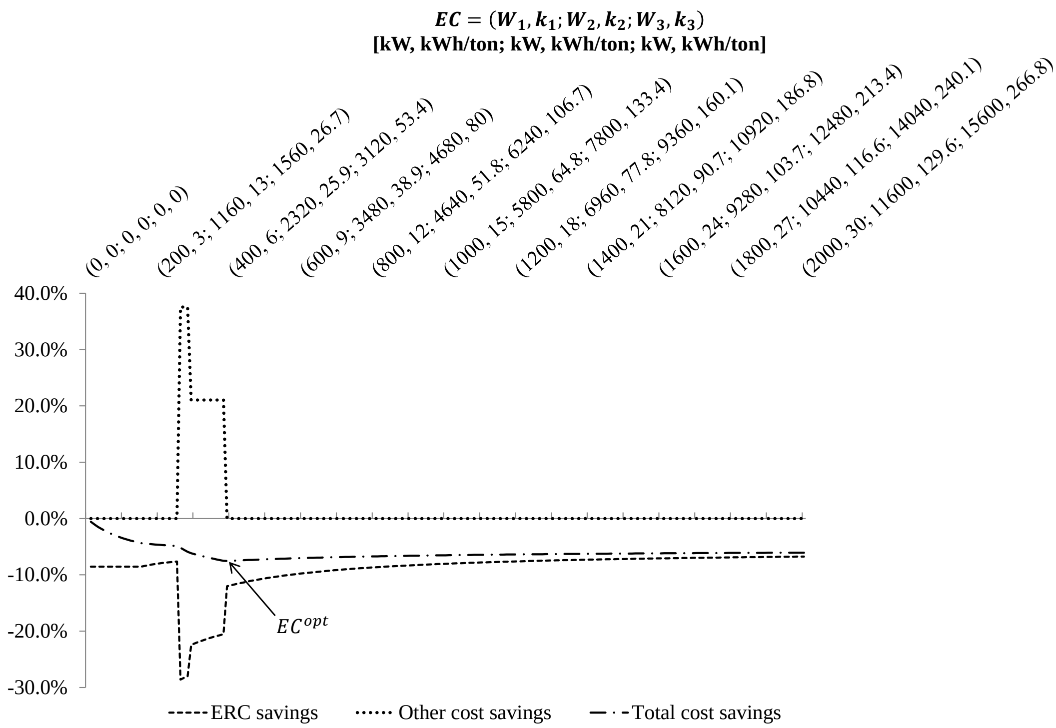


Figure 7: Cost savings dependent on energy usage of production stages.

The curve of the total cost savings in Figure 7 implies that the characteristics of the EESS (capacity and performance) need to be tailored to the energy used by the production stages to best utilize the savings potential of the considered WHRS along with the EESS. When the energy usage of the production stages lies below EC^{opt}, the EESS is oversized. Thus, its energy savings potential is only partly exploited. Yet, when the energy usage of the production stages exceeds EC^{opt}, the EESS is undersized. Hence, its energy savings potential is limited.

A similar phenomenon can be observed in Figure 8 when investigating the cost savings for different total demand levels, varying with a step width of 5 ton. The steep increase in total cost savings when shifting from a total demand of 105 ton to a total demand of 110 ton is caused by the fact that this additional demand requires more production during on-peak periods. As a consequence, without

WHRS and EESS, the total demand-related cost increases from 581.90 EUR to 6982.86 EUR. In contrast, when the WHRS and the EESS are used, the total demand-related cost merely increases from 0 EUR to 5327.11 EUR. Thus, employing both a WHRS and an EESS enables the production system to better absorb the ramifications of additional production periods during on-peak periods than in the case where such systems are not used. Yet, when the total demand reaches a certain level (here: shifting from 125 ton to 130 ton), more production stages need to concurrently produce during on-peak periods. As a result, the energy accumulated in the EESS during the off-peak period is not sufficient anymore to absorb the additional energy requirements of these additional production periods required within the on-peak period. Thus, it becomes clear that the WHRS and the EESS can only be used effectively if a fit between the characteristics (capacity and performance) of the EESS, the total demand (or production lots), and the TOU pricing profile is achieved.

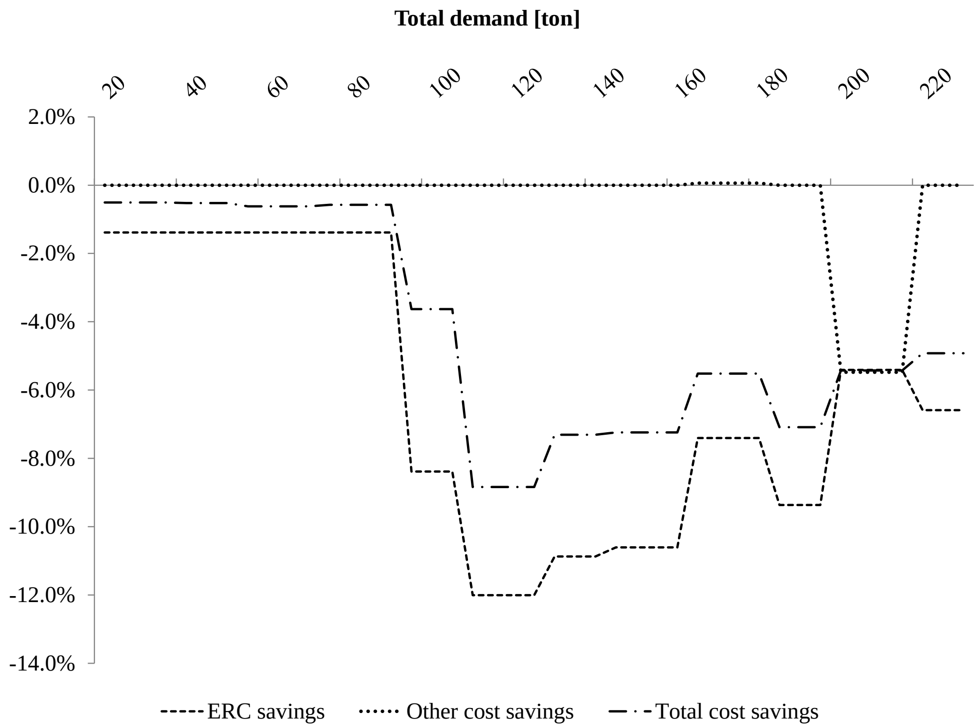


Figure 8: Cost savings dependent on total demand.

According to Moon and Park (2014), prospective electricity pricing profiles are expected to make the use of electricity in on-peak periods even more expensive, such that on-peak energy usage charges and on-peak power demand charges may be eight times as high as the corresponding off-peak charges. For this reason, we also investigate how soaring on-peak electricity prices impact total cost savings that result from using a WHRS and an EESS by varying the on-peak energy usage charge with a step width of 0.05 EUR/kWh and the on-peak power demand charge with a step width of 5 EUR/kW. Figure 9 reveals that both the increasing energy usage charge and the increasing power demand charge in on-peak periods lead to higher savings in total cost in almost all cases. This trend is only interrupted when energy usage during on-peak periods becomes so expensive that production phases are moved from on-peak periods to off-peak periods at nighttime where workers receive a surcharge. Subsequently, total cost savings again increase with the increasing on-peak energy usage charge and the increasing on-peak power demand charge. This phenomenon can be observed in Figure 9 when increasing the on-peak energy usage charge from 0.3 to 0.35 to 0.4 EUR/kWh while holding the on-peak power demand charge constant at 40 EUR/kW. Overall, it is clear that the profitability of the integrated system presented in this paper is expected to grow even further in the near future as energy usage charges and power demand charges are expected to rise.

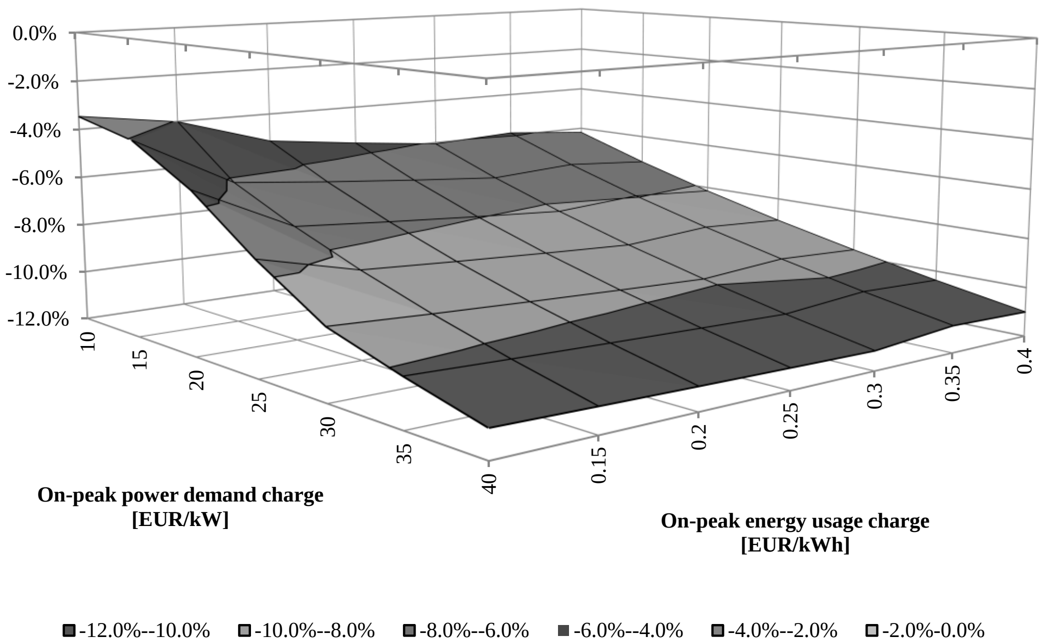


Figure 9: Total cost savings dependent on on-peak energy usage charge and on-peak power demand charge.

6 Conclusions and managerial implications

In consequence of the scarcity of resources, soaring energy prices, and a rising awareness of the impact energy-related CO_2 emissions have on the environment, energy aspects have more and more frequently been considered in production planning in recent years. The paper at hand studied a serial multi-stage production system subject to time-varying electricity prices. We integrated a WHRS along with an EESS into this production system and investigated under which conditions, specified by daily production targets, energy usage of the production stages, and TOU tariffs, the integrated system can best contribute to making manufacturing more energy-efficient. Prior decision support models for production planning were often criticized by researchers from the engineering field and practitioners for neglecting crucial technological restrictions. By modeling the technical processes of recovering waste heat and charging energy to or discharging energy from the EESS, we responded to these critics and developed a decision support model that represents real-life applications much more precisely. Thereby, we strengthened the interdisciplinary stream of research that connects classical production management to recent technological innovations promoting energy efficiency in manufacturing.

From the numerical analysis, the following managerial insights with respect to the use of a WHRS combined with an EESS can be derived:

- Attaching an EESS to a WHRS can substantially increase the effectiveness of the WHRS in reducing energy-related cost. Energy recovered by the WHRS in off-peak periods can be transferred to on-peak periods via the EESS and thus support the energy supply of the production system in periods of high energy prices. Consequently, the amount of energy drawn from the grid in on-peak periods decreases.
- Biel and Glock (2016) found that further technological advancements of WHRSs are necessary to stimulate the interest of practitioners. Yet, from the study at hand, we may conclude that supplementing a WHRS with an EESS may already justify the investment in a WHRS in its current technological state.
- Investments in an EESS need to be made in accordance with the specific energy requirements of the production system at hand. We illustrated that there are optimal characteristics (capacity and performance) of an EESS that exploit the energy saving potential of a WHRS best, given the energy requirements of a production system. If the EESS is too small in relation to the amount of energy recovered by the WHRS, the potential of the WHRS cannot be fully utilized. If the EESS is too large in relation to the amount of energy recovered by the WHRS, the energy saving potential of the EESS will only partly be realized and an investment in a smaller EESS would have been sufficient.
- For an effective use of a WHRS and an EESS, the dimensions of the EESS need to be aligned with both the total demand (or production lots) and the TOU pricing profile. If total demand is small enough to permit the company to avoid production in on-peak periods, the EESS will not significantly reduce energy-related cost. If total demand is high, in contrast, then an insufficient capacity of the EESS could restrict the effectiveness of the WHRS combined with the EESS.
- In light of soaring energy prices, the profitability of the use of a WHRS and an EESS will improve in the future. At the same time, companies can use EESSs to hedge risks arising from time-varying energy prices, especially when electricity tariffs move from predetermined TOU pricing profiles to hard-to-predict real-time pricing.

In a next step, it would be interesting to develop a model that combines the production planning decision with the sizing decision of an EESS, which has, for instance, already been studied by Schneider et al. (2016). Additionally, it seems desirable to also take energy-related CO_2 emissions into account and to investigate how they can be lowered by using a WHRS along with an EESS. Besides, other layout options of both the WHRS and the EESS may be worth investigating. Lastly, the model may be extended by modeling the relationship between the waste heat stream and the efficiency of the ORC in more detail. We leave these and other extensions for future research.

References

Althaus, W., Paucker, R., Grob, J., Bülten, B., Hunstock, B., 2013. *Entwicklungs-/Demonstrationsprojekt mit Feldversuch "ORC-Prozesse zur Abwärmenutzung an BHKW-Motoren"*. Fraunhofer-Institut für Umwelt-, Sicherheits und Energietechnik (UMSICHT), Oberhausen.

Ashok, S., 2006. Peak-load management in steel plants. *Applied Energy*, 83 (5), 413-424.

Ashok, S., Banerjee, R., 2001. An optimization mode for industrial load management. *IEEE Transactions on Power Systems*, 16 (4), 879-884.

Balogun, V.A., Kirkwood, N.D., Mativenga, P.T., 2014. Direct electrical energy demand in fused deposition modelling. *Procedia CIRP*, 15, 38-43.

Bazan, E., Jaber, M.Y., Zanoni, S., 2015. Supply chain models with greenhouse gases emissions, energy usage and different coordination decisions. *Applied Mathematical Modelling*, 39 (17), 5131-5151.

Bego, A., Li, L., Sun, Z., 2014. Identification of reservation capacity in critical peak pricing electricity demand response program for sustainable manufacturing systems. *International Journal of Energy Research*, 38 (6), 728-736.

Biel, K., Glock, C.H., 2016. On the use of waste heat in a two-stage production system with controllable production rates. *International Journal of Production Economics*, 181, 174-190.

Bruzzone, A.A.G., Anghinolfi, D., Paolucci, M., Tonelli, F., 2012. Energy-aware scheduling for improving manufacturing process sustainability: A mathematical model for flexible flow shops. *CIRP Annals - Manufacturing Technology*, 61 (1), 459-462.

Castro, P.M., Harjunkoski, I., Grossmann, I.E., 2009. New continuous-time scheduling formulation for continuous plants under variable electricity cost. *Industrial & Engineering Chemistry Research*, 48 (14), 6701-6714.

Chao, X., Chen, F.Y., 2005. An optimal production and shutdown strategy when a supplier offers an incentive program. *Manufacturing & Service Operations Management*, 7 (2), 130-143.

Chen, H., Cong, T.N., Yang, W., Tan, C., Li, Y., Ding, Y., 2009. Progress in electrical energy storage system: A critical review. *Progress in Natural Science*, 19 (3), 291-312.

Chen, H., Goswami, D.Y., Stefanakos, E.K., 2010. A review of thermodynamic cycles and working fluids for the conversion of low-grade heat. *Renewable & Sustainable Energy Reviews*, 14 (9), 3059-3067.

Dai, Y., Wang, J., Gao, L., 2009. Parametric optimization and comparative study of Organic Rankine Cycle (ORC) for low grade waste heat recovery. *Energy Conversion and Management*, 50 (3), 576-582.

Denton, F.T., Jefferies, K.L., Mountain, D.C., Robb, A.L., Spencer, B.G., 1987. The response of an industrial firm to alternative electricity rate structures: An optimization model for simulation applications. *Resources and Energy*, 9 (4), 327-346.

European Parliament and Council, 2012. Energy efficiency directive (2012/27/EU). *Official Journal of the European Union*, 55, L 315/1-L 315/56.

Fang, H., Xia, J., Zhu, K., Su, Y., Jiang, Y., 2013. Industrial waste heat utilization for low temperature district heating. *Energy Policy*, 62, 236-246.

Fernandez, M., Li, L., Sun, Z., 2013. "Just-for-Peak" buffer inventory for peak electricity demand reduction of manufacturing systems. *International Journal of Production Economics*, 146 (1), 178-184.

Gahm, C., Denz, F., Dirr, M., Tuma, A., 2016. Energy-efficient scheduling in manufacturing companies: A review and research framework. *European Journal of Operational Research*, 248 (3), 744-757.

Ghobeity, A., Mitsos, A., 2010. Optimal time-dependent operation of seawater reverse osmosis. *Desalination*, 263 (1), 76-88.

Gutowski, T., Dahmus, J., Thiriez, A., 2006. *Electrical energy requirements for manufacturing processes*. In: Proceedings of the 13th CIRP International Conference on Life Cycle Engineering, Leuven, pp. 623-638.

Haddad, C., Périlhon, C., Danlos, A., François, M.-X., Descombes, G., 2014. Some efficient solutions to recover low and medium waste heat: Competitiveness of the thermoacoustic technology. *Energy Procedia*, 50, 1056-1069.

Hasanbeigi, A., Price, L., 2012. A review of energy use and energy efficiency technologies for the textile industry. *Renewable &Sustainable Energy Reviews*, 16 (6), 3648-3665.

He, Y., Liu, F., Cao, H.-J., Li, C.-B., 2005. A bi-objective model for job-shop scheduling problem to minimize both energy consumption and makespan. *Journal of Central South University of Technology*, 12 (2), 167-171.

Hirzel, S., Sontag, B., Rhode, C., 2013. *Industrielle Abwärmenutzung*. Fraunhofer-Institut für System- und Innovationsforschung (ISI), Karlsruhe.

Hung, T., Wang, S., Kuo, C., Pei, B., Tsai, K., 2010. A study of organic working fluids on system efficiency of an ORC using low-grade energy sources. *Energy*, 35 (3), 1403-1411.

Ibarra, M., Rovira, A., Alarcón-Padilla, D.-C., Blanco, J., 2014. Performance of a 5 kWe Organic Rankine Cycle at part-load operation. *Applied Energy*, 120, 147-158.

International Energy Agency, 2014. *Energy efficiency market report 2014*. Paris.

Invernizzi, C.M., 2013. *Closed power cycles: Thermodynamic fundamentals and applications*. Springer, London.

Li, L., Sun, Z., 2013. Dynamic energy control for energy efficiency improvement of sustainable manufacturing systems using Markov Decision Process. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 43 (5), 1195-1205.

Liu, Y., Yang, J., Wang, J., Cheng, Z.-L., Wang, Q.-W., 2014. Energy and exergy analysis for waste heat cascade utilization in sinter cooling bed. *Energy*, 67, 370-380.

Madan, J., Mani, M., Lee, J.H., Lyons, K.W., 2015. Energy performance evaluation and improvement of unit-manufacturing processes: Injection molding case study. *Journal of Cleaner Production*, 105, 157-170.

Mitra, S., Grossmann, I.E., Pinto, J.M., Arora, N., 2012. Optimal production planning under time-sensitive electricity prices for continuous power-intensive processes. *Computers & Chemical Engineering*, 38, 171-184.

Moon, J.-Y., Park, J., 2014. Smart production scheduling with time-dependent and machine-dependent electricity cost by considering distributed energy resources and energy storage. *International Journal of Production Research*, 52 (13), 3922-3939.

Mouzon, G., Yildirim, M.B., Twomey, J., 2007. Operational methods for minimization of energy consumption of manufacturing equipment. *International Journal of Production Research*, 45 (18-19), 4247-4271.

Neugebauer, R., Wabner, M., Rentzsch, H., Ihlenfeldt, S., 2011. Structure principles of energy efficient machine tools. *CIRP Journal of Manufacturing Science and Technology*, 4 (2), 136-147.

Nilsson, K., 1993. Industrial production planning with optimal electricity cost. *Energy Conversion and Management*, 34 (3), 153-158.

Nilsson, K., Söderström, M., 1993. Industrial applications of production planning with optimal electricity demand. *Applied Energy*, 46 (2), 181-192.

Obernberger, I., Thonhofer, P., Reisenhofer, E., 2002. Description and evaluation of the new 1000 kWel Organic Rankine Cycle process integrated in the biomass CHP plant in Lienz, Austria. *Euroheat & Power*, 10, 1-17.

Orange and Rockland Utilities, 2012. *Service classification No. 20*. https://www.oru.com/_external/orurates/tariffsandregulatorydocuments/newyork/scheduleforelectricservice.html (retrieved April 27, 2015).

Pehnt, M., Bödeker, J., Arens, M., Jochem, E., Idrissova, F., 2010. *Die Nutzung industrieller Abwärme – technisch-wirtschaftliche Potenziale und energiepolitische Umsetzung*. Institut für Energie- und Umweltforschung (ifeu), Fraunhofer-Institut für System- und Innovationsforschung (ISI), Institut für Ressourceneffizienz und Energiestrategien GmbH (IREES), Heidelberg Karlsruhe.

Pons, M., Bikfalvi, A., Llach, J., Palcic, I., 2013. Exploring the impact of energy efficiency technologies on manufacturing firm performance. *Journal of Cleaner Production*, 52, 134-144.

Quader, M.A., Ahmed, S., Ghazilla, R.A.R., Ahmed, S., Dahari, M., 2015. A comprehensive review on energy efficient CO_2 breakthrough technologies for sustainable green iron and steel manufacturing. *Renewable & Sustainable Energy Reviews*, 50, 594-614.

Quoilin, S., Aumann, R., Grill, A., Schuster, A., Lemort, V., Spliethoff, H., 2011a. Dynamic modeling and optimal control strategy of waste heat recovery Organic Rankine Cycles. *Applied Energy*, 88 (6), 2183-2190.

Quoilin, S., Declaye, S., Tchanche, B.F., Lemort, V., 2011b. Thermo-economic optimization of waste heat recovery Organic Rankine Cycles. *Applied Thermal Engineering*, 31 (14), 2885-2893.

Schneider, M., Biel, K., Pfaller, S., Schaede, H., Rinderknecht, S., Glock, C.H., 2016. Using inventory models for sizing energy storage systems: An interdisciplinary approach. *Journal of Energy Storage*, 8, 339-348.

Shrouf, F., Ordieres-Meré, J., García-Sánchez, A., Ortega-Mier, M., 2014. Optimizing the production scheduling of a single machine to minimize total energy consumption costs. *Journal of Cleaner Production*, 67, 197-207.

Sun, Z., Li, L., 2014. Potential capability estimation for real time electricity demand response of sustainable manufacturing systems using Markov Decision Process. *Journal of Cleaner Production*, 65, 184-193.

Tchanche, B.F., Quoilin, S., Declaye, S., Papadakis, G., Lemort, V., 2010. *Economic feasibility study of a small scale Organic Rankine Cycle system in waste heat recovery application*. In: Proceedings of the ASME 2010 10th Biennial Conference on Engineering Systems Design and Analysis, Istanbul, vol. 1, pp. 249-256.

U.S. Bureau of Labor Statistics, 2013. *International comparisons of hourly compensation costs in manufacturing, 2012*. Washington, DC.

U.S. Energy Information Administration, 2014. *Annual energy outlook with projections to 2040*. Washington, DC.

U.S. Energy Information Administration, 2015. *Annual energy outlook with projections to 2040*. Washington, DC.

Wang, E., Zhang, H., Fan, B., Ouyang, M., Zhao, Y., Mu, Q., 2011. Study of working fluid selection of Organic Rankine Cycle (ORC) for engine waste heat recovery. *Energy*, 36 (5), 3406-3418.

Wang, Y., Li, L., 2013. Time-of-use based electricity demand response for sustainable manufacturing systems. *Energy*, 63, 233-244.

Wei, D., Lu, X., Lu, Z., Gu, J., 2007. Performance analysis and optimization of Organic Rankine Cycle (ORC) for waste heat recovery. *Energy Conversion and Management*, 48 (4), 1113-1119.

Weinert, N., Chiotellis, S., Seliger, G., 2011. Methodology for planning and operating energy-efficient production systems. *CIRP Annals - Manufacturing Technology*, 60 (1), 41-44.

Zakeri, B., Syri, S., 2015. Electrical energy storage systems: A comparative life cycle cost analysis. *Renewable & Sustainable Energy Reviews*, 42, 569-596.

Zanoni, S., Bettoni, L., Glock, C.H., 2014. Energy implications in a two-stage production system with controllable production rates. *International Journal of Production Economics*, 149, 164-171.

Zhang, H., Zhao, F., Sutherland, J.W., 2014. *Manufacturing scheduling for energy cost reduction in a smart grid scenario*. In: Proceedings of the ASME 2014 International Manufacturing Science and Engineering Conference collocated with the JSME 2014 International Conference on Materials and

Processing and the 42nd North American Manufacturing Research Conference, Detroit, MI, vol. 1, pp. V001T005A001-1-V001T005A001-10.

Appendix: Linearization of $S_{i,r,t}^{PS} \cdot S_{u,t}^{EESS}$

$$\chi_{i,r,u,t} \leq S_{i,r,t}^{PS}, \forall i \in \{1,2,\dots,m\}, t \in \{1,2,\dots,T\}, r \in OM, u \in ES \tag{A.1}$$

$$\chi_{i,r,u,t} \geq 0, \forall i \in \{1,2,\dots,m\}, t \in \{1,2,\dots,T\}, r \in OM, u \in ES \tag{A.2}$$

$$\chi_{i,r,u,t} \leq S_{u,t}^{EESS}, \forall i \in \{1,2,\dots,m\}, t \in \{1,2,\dots,T\}, r \in OM, u \in ES \tag{A.3}$$

$$\chi_{i,r,u,t} \geq S_{u,t}^{EESS} - \left(1 - S_{i,r,t}^{PS}\right), \forall i \in \{1,2,\dots,m\}, t \in \{1,2,\dots,T\}, r \in OM, u \in ES \tag{A.4}$$

Paper 4 Flow shop scheduling with grid-integrated onsite wind power using stochastic MILP[1]

Authors: Konstantin Biel, Fu Zhao, John W. Sutherland, Christoph H. Glock

Type of publication: Journal article

Publication details: *International Journal of Production Research*, In Press.

Abstract

Over the last decade, manufacturing companies have identified renewable energy as a promising means to cope with time-varying energy prices and to reduce energy-related greenhouse gas emissions. Among the different types of renewable energy sources, wind power has emerged as a major contributor to renewable energy generation capacities. To make efficient use of onsite wind power generation facilities in manufacturing, production scheduling tools need to consider the uncertainty attached to wind power generation along with changes in the energy procurement cost and in the products' environmental footprints. To this end, we propose a solution procedure that first generates a large number of wind power scenarios that characterize the variability in wind power over time. Subsequently, a two-stage stochastic optimization procedure computes a production schedule and energy supply decisions for a flow shop system. In the first stage, a bi-objective mixed integer linear program simultaneously minimizes the total weighted flow time and the expected energy cost, based on the generated wind power scenarios. In the second stage, energy supply decisions are adjusted based on real-time wind power data. A numerical example is used to illustrate the ability of the developed decision support tool to handle the uncertainty attached to wind power generation and its effectiveness in realizing energy-related objectives in manufacturing.

Keywords:

Production scheduling; Flow shop; Renewable energy; Energy consumption; Wind power

[1] Reprinted from *International Journal of Production Research*, Biel, K., Zhao, F., Sutherland, J.W., Glock, C.H., Flow shop scheduling with grid-integrated onsite wind power using stochastic MILP, In Press, with permission from Taylor & Francis.

1 Introduction

Governments, society, and industry are becoming ever more cognizant of the environmental issues surrounding the consumption of fossil fuels and the concomitant greenhouse gas (GHG) emissions. To mitigate the effects of the use of fossil fuels, there is growing use of a variety of renewable energy sources (RES), e.g., solar energy, wind energy, biomass, hydro energy, and geothermal energy. Part of this growth is attributable to government initiatives. For example, Germany introduced renewable energy in the late 1980s and accelerated this development in the aftermath of the Fukushima nuclear disaster in 2011 (Hake et al., 2015). Furthermore, the member states of the European Union agreed on target shares of energy from renewable sources in gross final energy consumption to be reached by 2020 (European Parliament and Council, 2009). For private companies, use of renewable energy is also being driven by economic factors. Companies have identified the minimization of their energy consumption and GHG emissions as well as the optimization of their energy supply strategy as a means to strengthen both their financial performance and their reputation as sustainable and environmentally responsible enterprises (Tognetti et al., 2015). As a result, global R&D expenditures on renewable energy technologies have more than doubled in the recent decade, reaching 11.7 billion USD in 2014 (McCrone et al., 2015).

Onsite renewable energy generation can serve as a valuable resource for manufacturing companies that employ energy-intensive production processes to reduce their dependence on utility companies and to react to the increasing use of time-varying electricity prices. Moreover, the use of RES can make a significant contribution towards decreasing the environmental impact of manufacturing processes. Among the different types of RES, wind power has emerged as a major contributor to renewable energy generation capacities in recent years. In the United States, for instance, 31 percent of the electricity generation capacity added between 2008 and 2014 was installed in the form of wind power generation facilities (U.S. Department of Energy, 2015). Among these facilities, the share of onsite wind power serving manufacturing facilities, farms, and rural or suburban homes has been rather small in the United States so far. Nevertheless, in a recent technical report of the National Renewable Energy Laboratory, Lantz et al. (2016) estimated that onsite wind development would be feasible for about 44 percent of the continental U.S. building stock, and that the maximum potential of onsite wind power generation exceeds the current total U.S. electricity demand. Based on this estimation and despite uncertainties regarding retail electricity rates and wind technology cost trends as well as regional variations, Lantz et al. (2016) expect the installed onsite wind capacity in the United States to double two times until 2030 (growth by 300 percent) and another time until 2050 (growth by 700 percent). However, to make efficient use of onsite wind power, companies need effective strategies to handle the intermittent character of wind power. This intermittency in the energy supply from wind turbines classifies wind power as a non-dispatchable energy source (Foley et al., 2012) since the power changes over time due to the vagaries of wind speed (this is illustrated in Figure 1; see also Iversen et al., 2016). The intermittency in the energy supply induces a stochastic component into production scheduling, which seeks to minimize energy-related objectives such as energy cost or GHG emissions along with traditional production scheduling goals such as makespan or flow time.

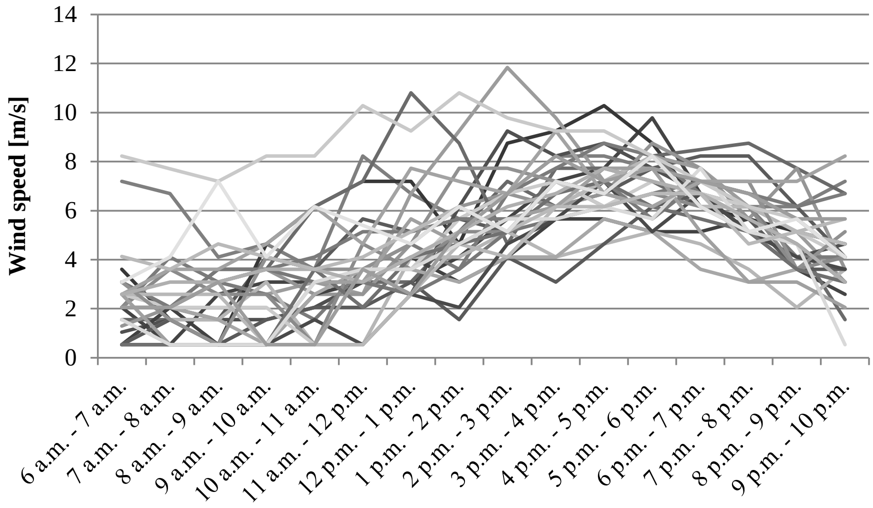


Figure 1: 30 days of wind speed observations between 6 a.m. and 10 p.m. at weather station KHWD in Hayward, California (measured at an anemometer height of 10 m).

This paper aims at strengthening the emerging stream of research that focuses on the integration of RES into production scheduling problems. Its contribution is twofold. First, it proposes a holistic mixed integer linear programming (MILP) problem that aims at minimizing both the total weighted flow time and the energy cost of a flow shop system. It is to be noted that for this problem, both the energy prices and a portion of the energy supply, provided by a wind turbine, vary over time. Secondly, the paper develops a solution procedure for the production scheduling problem in the form of a two-stage stochastic production scheduling and energy supply decision procedure that explicitly considers the stochastic and time-varying nature of wind power. To integrate the corresponding uncertainty into the model, the proposed solution procedure rests upon a large number of wind power scenarios generated from hourly wind speed forecasts. Overall, the proposed model enables decision-makers to efficiently coordinate production scheduling decisions with energy supply from onsite wind power generation facilities.

The remainder of the paper is organized as follows: Section 2 reviews the literature on energy-aware production scheduling with a special focus on the integration of RES. Section 3 describes the problem studied in this paper and formally translates this problem description into an MILP problem. Section 4 focuses on the treatment of the uncertainty resulting from the intermittent character of wind power generation and introduces the employed solution procedure. Section 5 presents a numerical study and highlights the merits of the developed model. Section 6 discusses managerial implications and concludes the paper.

2 Literature review

As the consideration of energy aspects in manufacturing has attracted much attention among researchers and practitioners in recent years, a rich body of literature on energy-aware production scheduling has evolved. For a comprehensive review of this research stream, the reader is referred to Biel and Glock (2016). One of the seminal works was contributed by Mouzon et al. (2007), which minimized both total completion time and energy consumption in a single-machine scheduling problem. Fang et al. (2011) were among the first to propose a model to minimize total makespan, peak power consumption, and GHG emissions in the context of a flow shop system. Similarly, May et al. (2015) developed a scheduling model to minimize makespan and energy consumption in a job shop environment. Two other works related to Fang et al. (2011) are Liu et al. (2017) and Lei et al. (2017), who optimized production scheduling in the context of a flow shop system with respect to energy efficiency goals. While Liu et al. (2017) developed a model to minimize energy consumption considering product quality, Lei et al. (2017) optimized workload balance and total energy consumption. Moon et al. (2013), who studied the unrelated parallel machine scheduling problem, and Luo et al. (2013), who investigated the hybrid flow shop problem, both considered so-called time-of-use (TOU) electricity pricing schemes. Such schemes are a typical means for utilities to attach higher prices to energy consumption during periods of generally higher demand for electricity (on-peak periods) and lower prices to periods of medium (partial-peak periods) or low demand for electricity (off-peak periods). Based on these time-varying energy prices, both Moon et al. (2013) and Luo et al. (2013) minimized energy-related cost along with total makespan.

Similarly, Zhang et al. (2014) considered a TOU pricing scheme in a flow shop scheduling problem. However, the authors exclusively focused on minimizing energy cost and GHG emissions while guaranteeing a predetermined production throughput. The novelty of the developed model principally lies in the recognition that energy consumption in off- and partial-peak periods is generally associated with higher GHG emission rates than in on-peak periods. This is due to the fact that electricity during off-peak and partial-peak periods is largely provided by coal-fired power plants, while gas-fired power plants, which usually emit less CO_2 per generated kWh of electricity than coal-fired power plants, contribute a far greater portion of the energy supply during on-peak periods (Zhang et al., 2014). Hence, minimizing energy-related cost and GHG emissions may actually be conflicting goals, and the model proposed by Zhang et al. (2014) can help to find a good trade-off.

In this line of thought, Sharma et al. (2015) studied a two-stage assembly flow shop scheduling problem while taking account of a TOU pricing scheme featuring both energy consumption and power demand charges. While the energy consumption charge was simply multiplied by the total energy consumption within a billing period, the power demand charge was multiplied by the peak power demand within this billing period (Sharma et al., 2015). Even though the power demand charge usually accounts for a major share of the overall electricity bill, the vast majority of the publications focusing on energy-aware production scheduling neglect it. Based on this pricing scheme, Sharma et al. (2015) concurrently minimized energy-related cost as well as GHG emissions and maximized the number of jobs processed in a shift.

Moon and Park (2014) were the first to integrate RES and fuel cells along with an energy storage system (ESS) into a production scheduling model. In the investigated flexible job shop scenario, the proposed model minimized the sum of the costs related to the total makespan, energy consumption from the public grid, and new technologies (e.g., RES and ESS) in the context of a TOU pricing scheme. For solving the model, the authors employed a hybrid production and energy scheduling algorithm. To arrive at a near-optimal solution, this algorithm alternately optimizes the production schedule and the energy supply decisions while treating the energy supply decisions and the production schedule, respectively, as inputs to the optimization. To incorporate the RES into the planning procedure, Moon and Park (2014) assumed that for each time period within the planning horizon, the minimum and maximum amount of energy that could be generated by the RES would be known in advance. Then, the model could choose the amount of energy to be generated by the RES from this interval for a given time period.

Liu (2016), who studied a single-machine scheduling problem, also considered RES and an ESS. However, in contrast to Moon and Park (2014), Liu (2016) explicitly incorporated the uncertainty attached to the RES energy supply. To this end, the author modeled the energy generated by the RES during each time slot of the planning horizon as an interval, with the interval boundaries randomly generated from a uniform distribution (instead of being predetermined). Based on these boundaries, the production schedule was determined in the optimization process. Using this representation of energy supply uncertainty, Liu (2016) proposed a first model which simultaneously minimized the total weighted flow time and GHG emissions with the help of the lexicographic-weighted Tchebycheff method. A second model minimized a single objective, the total weighted flow time, while considering both a periodic and a rolling GHG emission constraint.

The work at hand also integrates RES into a production scheduling problem. However, it differs from the works of Moon and Park (2014) and Liu (2016) in the way the uncertainty associated with the energy supply from a wind turbine is modeled. A very important characteristic of this uncertainty is the time dependence of the energy generated by a wind turbine for successive time periods (Morales et al., 2010). This time dependence has been examined in several studies that investigated the autocorrelation of wind speed observations at lags of different lengths of time. Corotis et al. (1977), for instance, analyzed the wind speeds at several sites in the Midwest of the United States and found that the hourly wind speed autocorrelation coefficients at a lag of one hour ranged from 0.6 to 0.8. Brett and Tuller (1991) reported autocorrelation coefficient values between 0.77 and 0.91 at a lag of one hour for various sites on the west coast of Canada. Furthermore, Corotis et al. (1977), Brett and Tuller (1991), and Brown et al. (1984) stressed the existence of a diurnal (daily) component in the investigated hourly wind speed time series. However, neither Moon and Park (2014) nor Liu (2016) addressed these characteristics that are associated with both wind speed and solar radiation time series (Hocaoğlu, 2010), even though Moon and Park (2014) considered solar power and Liu (2016) incorporated wind and solar power in their models. Additionally, they did not incorporate remedial actions in their respective models, i.e., they did not consider how to modify a predetermined schedule once the actual value of the renewable energy supply at a given time is known. For these two earlier works, the calculated production schedule and energy supply decisions solely rest upon the predetermined RES energy for each time period. That is, the production schedule and energy supply decisions do not take advantage of the actual wind or solar power data that are gradually revealed in real time. In contrast, the model proposed in this paper addresses the abovementioned time dependence of successive RES output values. To this end, our model considers a large number of hourly wind power scenarios that are fed into the production scheduling model. These wind power scenarios are based on hourly wind power point forecasts, and each scenario describes a time series of potential wind power that is generated by the wind turbine in each time slot across the entire planning horizon. Furthermore, we include real-time adjustment actions into our model to modify the energy supply decision when the actual RES energy supply at a given time is revealed. Consequently, the energy-related costs calculated with our model are expected to more accurately describe the energy-related costs that practitioners face when integrating RES into production scheduling. Lastly, we also consider power demand charges in the TOU pricing scheme, which is highly important for the practical relevance of the model as they are a major contributor to the overall energy-related costs (Wang and Li, 2015).

3 Model for flow shop scheduling and energy supply decision-making

This study aims to develop a mathematical programming model to support production scheduling and energy supply decisions for a flow shop system that has access to onsite wind power. In the following, Section 3.1 first describes the flow shop and energy supply system, and Section 3.2 then formalizes this description using MILP.

3.1 Description of the flow shop and energy supply system

The flow shop system comprises m machines (see Figure 2). Each job j, out of the set of jobs $J = \{1,2,\ldots,n\}$, needs to be processed on each machine i, out of the set of machines $M = \{1,2,\ldots,m\}$, in the order of the machine index. The energy required to run the machines is either taken from the grid or, if the wind is sufficiently strong, generated by the wind turbine. Energy generated by the wind turbine can either be used to run the machines, or it can be fed into the grid in return for compensation (feed-in tariff). The production schedule, on the one hand, defines in which time slot a job is to be processed on a given machine. The energy supply decisions, on the other hand, dictate whether the electricity generated by the wind turbine during a given time slot should be used to support the machines or be fed into the grid. Based on these decisions, the amount of electricity that needs to be drawn from the grid to guarantee an uninterrupted energy supply for the machines across the entire planning horizon is derived.

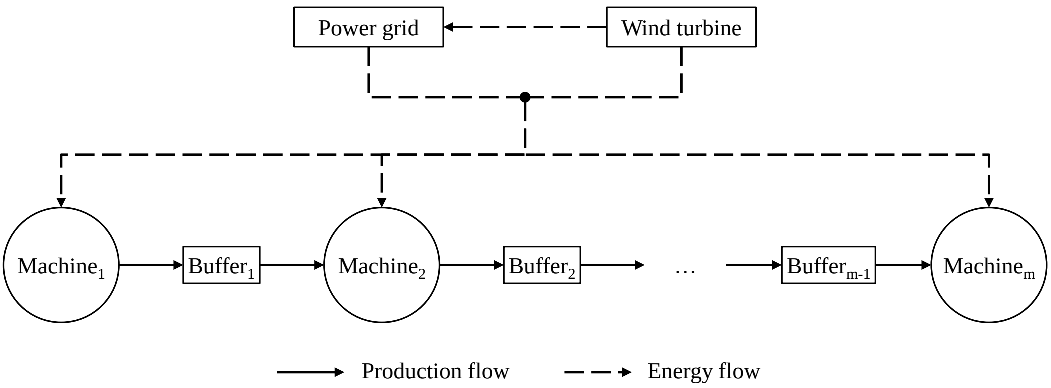


Figure 2: Production and energy flow in a flow shop system with a wind turbine.

Throughout the paper, the following terminology is used:

Sets

J	Set of jobs with $J = \{1,2,\ldots,n\}$
M	Set of machines with $M = \{1,2,\ldots,m\}$
OP	Set of off-peak time slots
PP	Set of partial-peak time slots
MP	Set of maximum-peak time slots
WP^o	Original set of wind power scenarios
WP^r	Reduced set of wind power scenarios

Indices

i	Machine with $i \in M$
j	Job with $j \in J$
k	Wind power forecast interval with $k \in \{1,2,\dots,K\}$
s	Wind power scenario with $s \in \{1,2,\dots,S\}$
t	Time slot with $t \in \{1,2,\dots,T\}$

Parameters

A	Swept area of the wind turbine rotor [m^2]
$C^p(V_{s,t})$	Power coefficient (overall efficiency) of the wind turbine at wind speed $V_{s,t}$ [-]
c_t	Energy consumption charge in time slot t [USD/kWh]
c^w	Levelized cost of electricity generated by the wind turbine [USD/kWh]
d^{mp}	Power demand charge for maximum power consumption in maximum-peak power time slots [USD/kW]
d^{pp}	Power demand charge for maximum power consumption in partial-peak power time slots [USD/kW]
d^p	Power demand charge for maximum power consumption in entire planning horizon [USD/kW]
H^f	Vector containing historical hourly wind power forecasts for the calibration period of length Γ with $H^f = \left(H_1^f, H_2^f, \dots, H_\Gamma^f\right)^T$ [kW]
H^o	Vector containing historical hourly wind power observations for the calibration period of length Γ with $H^o = (H_1^o, H_2^o, \dots, H_\Gamma^o)^T$ [kW]
$H^{f,o}$	Vector containing corresponding historical wind power forecasts and observations in the form of tuples for the calibration period of length Γ with $H^{f,o} = \left(\left(H_1^f, H_1^o\right), \left(H_2^f, H_2^o\right), \dots, \left(H_\Gamma^f, H_\Gamma^o\right)\right)^T$ [kW]
H^v	Vector containing historical wind power fluctuations for the calibration period of length Γ with $H^v = (H_1^o - H_2^o, H_2^o - H_3^o, \dots, H_{\Gamma-1}^o - H_\Gamma^o)^T$ [kW]
l	Length of a time slot [h]
P	Vector containing the hourly wind power point forecasts across the entire planning horizon with $P = (P_1, P_2, \dots, P_T)^T$ [kW]
$p_{i,j}$	Processing time of job j on machine i [h]
$q_{i,j}$	Power consumed by job j on machine i [kW]
r^{FiT}	Feed-in tariff [USD/kWh]
V^{cut-in}	Cut-in speed of the wind turbine [m/s]
$V^{cut-out}$	Cut-out speed of the wind turbine [m/s]
V^{rated}	Rated speed of the wind turbine [m/s]
$V_{s,t}$	Wind speed in time slot t of scenario s [m/s]
w_j	Weight representing the importance of job j [-]
Γ	Length of the calibration period of the scenario generation process [-]
ρ	Air density at the location of the wind turbine [kg/m^3]

$\tilde{\Psi}_s$	Wind power scenario containing the power generated by the wind turbine in each time slot t of scenario s in the entire planning horizon (uncertain) with $\tilde{\Psi}_s = (\tilde{\psi}_{s,1}, \tilde{\psi}_{s,2}, \dots, \tilde{\psi}_{s,T})^T$; Notice that the energy provided by the wind turbine becomes $\tilde{\psi}_{s,t} \cdot l$ and that $\tilde{\psi}^o_{s,t}$ ($\tilde{\psi}^r_{s,t}$) belongs to the original (reduced) set of wind power scenarios [kW]
ψ^{max}	Maximum power output of the wind turbine [kW]
ψ_t	Power actually generated by the wind turbine during time slot t [kW]
π_s	Probability of occurrence of scenario s; Notice that $\pi^o_{s,t}$ ($\pi^r_{s,t}$) belongs to the original (reduced) set of wind power scenarios [%]
ω_1	Weight of production-related objective 1 [-]
ω_2	Weight of energy-related objective 2 [-]

Decision variables

D^{mp}_s	Maximum power consumption in maximum-peak time slots in scenario s [kW]
D^{pp}_s	Maximum power consumption in partial-peak time slots in scenario s [kW]
D^p_s	Maximum power consumption in entire planning horizon in scenario s [kW]
DRC_s	Expected power demand-related cost of scenario s [USD]
EC	Expected total energy cost [USD]
ERC_s	Expected energy consumption-related cost of scenario s [USD]
$pd_{s,t}$	Power demanded from the grid across time slot t of scenario s [kW]
$pf_{s,t}$	Power fed into the grid across time slot t of scenario s [kW]
$TWFT$	Total weighted flow time [h]
$x_{i,j,t}$	Binary variable which equals 1 if job j is started on machine i during time slot t, and 0 otherwise [-]
$y_{i,j,t}$	Binary variable which equals 1 if machine i processes job j during time slot t, and 0 otherwise [-]
ϕ_t	Binary variable which equals 1 if the power generated by the wind turbine is used to run the machines in time slot t, and 0 if the power generated by the wind turbine is fed into the grid [-]
$\xi_{s,t}$	Binary variable which equals 1 if the power needed to run the machines in time slot t of scenario s exceeds the power generated by the wind turbine that is used to run the machines ($\sum_{i \in M} \sum_{j \in J} y_{i,j,t} \cdot q_{i,j} > \tilde{\psi}_{s,t} \cdot \phi_t$), and 0 otherwise [-]
$\theta_{s,t}$	Auxiliary variable to linearize $\xi_{s,t} \cdot \phi_t$ [-]
$\zeta_{i,j,s,t}$	Auxiliary variable to linearize $\xi_{s,t} \cdot y_{i,j,t}$ [-]

Apart from what has already been stated, we further assume that

(1) the processing order of the jobs can vary from machine to machine;
(2) each machine needs to finish a job j before processing another job $j' \in J \setminus \{j\}$;
(3) in each time slot, a machine can only process a single job;
(4) the capacity of buffer stocks between the machines is unlimited;
(5) the processing time of job j on machine i, $p_{i,j}$, is a multiple of the length of a time slot, i.e., all processing times are multiples of l;

(6) all jobs to be scheduled are known at the start of the planning horizon and need to be finished before the end of the planning horizon;
(7) the sizing decision of the wind turbine has already been made. Hence, the maximum power that the wind turbine can supply is known;
(8) electricity generated by the wind turbine is priced based on the levelized cost of electricity (LCOE) from the wind turbine. LCOE reflects" 'the per-kilowatt hour cost (in real dollars) of building and operating a generating plant over an assumed financial life and duty cycle' (U.S. Energy Information Administration, 2015). In the case of wind energy, LCOE takes account of capital costs, operations and maintenance costs as well as of financing costs, and it is based upon a given utilization rate of the wind turbine;
(9) the TOU pricing scheme is divided into an energy consumption charge (USD/kWh) and a power demand charge (USD/kW) that may vary across off-peak, partial-peak, and maximum-peak periods;
(10) the amount of electricity generated by the wind turbine during a future time slot is uncertain;
(11) the wind speed and consequently the wind power generated by the wind turbine stays constant during a time slot.

With these assumptions, the description of the model is complete and provides a sound basis for the modeling process in Section 3.2.

3.2 Modeling the flow shop and energy supply system via a bi-objective MILP problem

This section translates the flow shop scheduling and energy supply decision problem defined in Section 3.1 into an MILP problem. Unlike traditional flow shop scheduling models, our formulation needs to be time-indexed as the energy consumption charge and the power demand charge as well as the generated wind power vary over time.

3.2.1 Definition of production-related and energy-related goals

The proposed decision support model seeks to compute a production schedule and energy supply decisions that consider both a production-related goal and an energy-related goal. The production-related goal of the proposed model is to minimize the total weighted flow time, $TWFT$. This is a typical goal in production scheduling and takes the differentiation of jobs through assigned weights corresponding to holding cost or value previously added into consideration (Pinedo, 2012):

$$\min TWFT = \sum_{j \in J} w_j \cdot \sum_{t \in \{1,2,\ldots,T\}} \left(t \cdot x_{m,j,t} + \frac{p_{m,j}}{l} - 1 \right). \tag{1}$$

The energy-related goal of the MILP problem is to minimize the expected total energy cost, EC. In times when the LCOE of RES was far from being competitive to the LCOE of non-renewable energy sources, the primary benefit of RES was their potential to reduce energy-related GHG emissions. Today, due to significant technological progress over the last decades, wind turbines have already reached or are expected to reach grid parity soon in many regions of Europe and the United States. That is, the LCOE of RES is often equal to or lower than the LCOE of non-renewable energy sources (U.S. Energy Information Administration, 2015). Hence, wind power nowadays has the potential to reduce both carbon footprints and energy costs. In our application, this is particularly true as we consider a TOU pricing scheme where in times of generally high power demand, the energy price the company faces lies well above the LCOE of the onsite wind turbine. Thus, wind power cannot only be employed to reduce energy-related GHG emissions, but also to reduce energy cost, depending on the coordination of production planning and energy supply decisions.

The uncertainty attached to the energy supply from the wind turbine is incorporated into the MILP problem by means of a large number of wind power scenarios that describe the potential wind power time behavior across the entire planning horizon. The proposed model minimizes the expected value of EC, consisting of the sum of the expected energy consumption-related costs, ERC_s, and the expected power demand-related costs, DRC_s, across all wind power scenarios:

$$\min EC = \sum_{s\in\{1,2,\dots,S\}} \pi_s \cdot (ERC_s + DRC_s). \tag{2}$$

ERC_s embraces the cost of the energy purchased from the grid and the cost of the energy generated by the wind turbine less the revenue from the energy sold to the grid in scenario s:

$$ERC_s = l \cdot \sum_{t\in\{1,2,\dots,T\}} \left(pd_{s,t} \cdot c_t + \tilde{\psi}_{s,t} \cdot c^w - pf_{s,t} \cdot r^{FiT}\right). \tag{3}$$

In addition to the cost arising from energy consumption, the proposed model also considers the cost associated with the maximum power demand in the different peak periods, which is reflected in the considered TOU pricing scheme. Thus, DRC_s consists of the sum of the maximum power consumptions during the partial-peak period, D_s^{pp}, the maximum-peak period, D_s^{mp}, and the entire planning horizon, D_s^p, multiplied by the corresponding power demand charges:

$$DRC_s = D_s^{pp} \cdot d^{pp} + D_s^{mp} \cdot d^{mp} + D_s^p \cdot d^p. \tag{4}$$

3.2.2 Constraints to govern the production and energy flow in the flow shop system

During the optimization process, several constraints related both to the production process as well as to energy supply and consumption need to be considered. Equations (5) to (11) specify the production flow in the flow shop system:

$$y_{i-1,j,t} + y_{i,j,t} \leq 1, \forall i \in \{2,3,\dots,m\}, j \in J, t \in \{1,2,\dots,T\} \tag{5}$$

$$\sum_{\tau=1}^{t} x_{i-1,j,\tau} \geq \sum_{\tau=1}^{t} x_{i,j,\tau}, \forall i \in \{2,3,\dots,m\}, j \in J, t \in \{1,2,\dots,T\} \tag{6}$$

$$\sum_{t\in\left\{1,2,\dots,T-\frac{p_{i,j}}{l}+1\right\}} x_{i,j,t} = 1, \forall i \in M, j \in J \tag{7}$$

$$\sum_{j\in J} x_{i,j,t} \leq 1, \forall i \in M, t \in \{1,2,\dots,T\} \tag{8}$$

$$l \cdot \sum_{\tau\in\left\{t,t+1,\dots,t+\frac{p_{i,j}}{l}-1\right\}} y_{i,j,\tau} \geq x_{i,j,t} \cdot p_{i,j}, \forall i \in M, j \in J, t \in \left\{1,2,\dots,T-\frac{p_{i,j}}{l}+1\right\} \tag{9}$$

$$y_{i,j,t} + \sum_{j'\in J\setminus j} x_{i,j',t} \leq 1, \forall i \in M, j \in J, t \in \{1,2,\dots,T\} \tag{10}$$

$$x_{i,j,t}, y_{i,j,t} \in \{0,1\}, \forall i \in M, j \in J, t \in \{1,2,\dots,T\} \tag{11}$$

Equation (5) assures that each job j can only be processed on one machine in time slot t. Equation (6) guarantees that job j is processed on machine $i-1$ before it is processed on machine i, while Equation (7) ensures that each job j is processed on each machine i exactly once. Equation (8) assures that only one job is started on each machine i in each time slot t. Equations (9) and (10) guarantee that jobs are processed non-preemptively and that no job is started on a machine until the machine is finished with the previous job. Equation (11) assures the binary character of $x_{i,j,t}$ and $y_{i,j,t}$.

Equations (12) to (22) specify how the energy requirements of the flow shop system are satisfied:

$$pd_{s,t} = \left(\sum_{i \in M} \sum_{j \in J} y_{i,j,t} \cdot q_{i,j} - \tilde{\psi}_{s,t} \cdot \phi_t\right) \cdot \xi_{s,t}, \forall s \in \{1,2,\dots,S\}, t \in \{1,2,\dots,T\} \tag{12}$$

$$pd_{s,t} = \sum_{i \in M} \sum_{j \in J} \zeta_{i,j,s,t} \cdot q_{i,j} - \tilde{\psi}_{s,t} \cdot \theta_{s,t}, \forall s \in \{1,2,\dots,S\}, t \in \{1,2,\dots,T\} \tag{13}$$

$$pf_{s,t} = \tilde{\psi}_{s,t} + pd_{s,t} - \sum_{i \in M} \sum_{j \in J} y_{i,j,t} \cdot q_{i,j}, \forall s \in \{1,2,\dots,S\}, t \in \{1,2,\dots,T\} \tag{14}$$

$$\left(\sum_{i \in M} \sum_{j \in J} y_{i,j,t} \cdot q_{i,j} - \tilde{\psi}_{s,t} \cdot \phi_t\right) \cdot \xi_{s,t} \geq 0, \forall s \in \{1,2,\dots,S\}, t \in \{1,2,\dots,T\} \tag{15}$$

$$\left(\tilde{\psi}_{s,t} \cdot \phi_t - \sum_{i \in M} \sum_{j \in J} y_{i,j,t} \cdot q_{i,j}\right) \cdot (1 - \xi_{s,t}) \geq 0, \forall s \in \{1,2,\dots,S\}, t \in \{1,2,\dots,T\} \tag{16}$$

$$\sum_{i \in M} \sum_{j \in J} \zeta_{i,j,s,t} \cdot q_{i,j} - \tilde{\psi}_{s,t} \cdot \theta_{s,t} \geq 0, \forall s \in \{1,2,\dots,S\}, t \in \{1,2,\dots,T\} \tag{17}$$

$$\tilde{\psi}_{s,t} \cdot \left(\phi_t - \theta_{s,t}\right) + \sum_{i \in M} \sum_{j \in J} q_{i,j} \cdot \left(\zeta_{i,j,s,t} - y_{i,j,t}\right) \geq 0, \forall s \in \{1,2,\dots,S\}, t \in \{1,2,\dots,T\} \tag{18}$$

$$D_s^{pp} \geq pd_{s,t}, \forall s \in \{1,2,\dots,S\}, t \in PP \tag{19}$$

$$D_s^{mp} \geq pd_{s,t}, \forall s \in \{1,2,\dots,S\}, t \in MP \tag{20}$$

$$D_s^{p} \geq pd_{s,t}, \forall s \in \{1,2,\dots,S\}, t \in \{1,2,\dots,T\} \tag{21}$$

$$\phi_t, \xi_{s,t}, \theta_{s,t}, \zeta_{i,j,s,t} \in \{0,1\}, \forall i \in M, j \in J, s \in \{1,2,\dots,S\}, t \in \{1,2,\dots,T\} \tag{22}$$

Equation (12) establishes that the power demanded from the grid in each time slot t of each scenario s equals the difference between the machine power demand and the power generated by the wind turbine that is used for running the machines. Thus, if the production planner decides to use wind power for running the machines, the company demands less power from the grid. As Equation (12) is not linear, it needs to be linearized to arrive at an MILP formulation. Linearization of $y_{i,j,t} \cdot \xi_{s,t}$ and $\phi_t \cdot \xi_{s,t}$, which are described in Appendix A, yields Equation (13). Equation (14) determines the power fed into the grid in each time slot t of each scenario s for which the company will receive compensation. If the production planner decides to use wind power for running the machines, only excess wind power not required in production is fed into the grid. If wind power is not used in production, the entire wind power is fed into the grid in return for compensation. Equations (15) and (16) ensure that in case the power generated by the wind turbine that is used to run the machines is sufficient for the machines, the binary variables $\xi_{s,t}$ equal 0, and 1 otherwise. As Equations (15) and (16) are not linear, they need to be linearized to yield Equations (17) and (18) (see Appendix A). Equations (19) to (21) compute the maximum power demand in the partial peak period, the maximum peak period as well as in the entire planning horizon. Equation (22) assures the binary character of ϕ_t, $\xi_{s,t}$, $\theta_{s,t}$, and $\zeta_{i,j,s,t}$.

4 Solving the proposed MILP problem

The procedure to compute a production schedule and energy supply decisions based upon the MILP problem described in Section 3.2 is divided into four steps:

(1) Wind power scenario generation
(2) Wind power scenario reduction
(3) Solution of the bi-objective MILP problem to make production scheduling and energy supply decisions
(4) Real-time adjustments of energy supply decisions in response to actual wind power

The first two steps integrate the uncertainty attached to the energy supply from the wind turbine into the MILP problem (see Section 4.1). The last two steps solve the resulting two-stage stochastic production scheduling and energy supply decision problem (see Section 4.2).

4.1 Integration of uncertainty of wind power supply into the MILP problem

To integrate the uncertainty resulting from the intermittent character of the energy supply from the wind turbine into the production scheduling and energy supply decision model, a large number of wind power scenarios are generated. Each scenario consists of a wind power time series, where the data in the series are derived from wind speeds. Such wind speeds can easily be converted into wind power using Equation (23) (Sharma et al., 2013):

$$\tilde{\psi}_{s,t} = \begin{cases} 0, V_{s,t} < V^{cut-in} \\ \frac{1}{2} \cdot C^p\left(V_{s,t}\right) \cdot \rho \cdot A \cdot V_{s,t}^3, V^{cut-in} \leq V_{s,t} < V^{rated} \\ \psi^{max}, V^{rated} \leq V_{s,t} < V^{cut-out} \\ 0, V_{s,t} \geq V^{cut-out} \end{cases}, \forall s \in \{1,2,\ldots,S\}, t \in \{1,2,\ldots,T\}, \quad (23)$$

where $C^p\left(V_{s,t}\right)$ corresponds to the power coefficient (the overall efficiency) of the wind turbine at wind speed $V_{s,t}$, ρ is the air density at the location of the wind turbine, A represents the swept area of the wind turbine rotor, V^{cut-in} (V^{rated}, $V^{cut-out}$) corresponds to the cut-in (rated, cut-out) speed of the wind turbine[2], and ψ^{max} represents the maximum power output of the wind turbine (Sharma et al., 2013).

In recent years, two main approaches to generate short-term hourly wind power scenarios have evolved. While the first approach relies on historical wind power data, the second approach is based on wind power point forecasts combined with historical forecast errors (Foley et al., 2012). In this paper, we employ a scenario generation technique based on the second approach. As will become evident, our approach not only uses statistical methods to generate wind power scenarios from past wind power observations (Sharma et al., 2013), but also utilizes point forecasts calculated with the help of Numerical Weather Prediction models. Models of this type consider atmospheric processes such as advection, pressure gradients, and adiabatic heating and cooling as well as physical processes such as cloud and precipitation micro-physics (Al-Yahyai et al., 2010).

[2] The cut-in speed corresponds to the wind speed at which the turbine starts to generate power. The rated speed corresponds to the wind speed at which the turbine reaches its maximum level of power generation. The cut-out speed corresponds to the wind speed at which the turbine is shut down to protect it from excessive loads (Burton et al., 2011).

More specifically, we generate short-term hourly wind power scenarios using the procedure developed by Ma et al. (2013), who extended the work of Pinson et al. (2009). This procedure is capable of not only incorporating forecasting uncertainty, but also wind power variability into the scenario generation process. Figure 3 provides an overview of the entire process. First of all, this scenario generation process requires historical hourly wind power forecasts, $H^f = \left(H_1^f, H_2^f, \dots, H_\Gamma^f\right)^T$, along with the respective historical wind power observations, $H^o = (H_1^o, H_2^o, \dots, H_\Gamma^o)^T$, for the calibration period of length Γ. Hence, the vector $H^{f,o}$ summarises the corresponding historical forecasts and observations in the form of tuples: $H^{f,o} = \left(\left(H_1^f, H_1^o\right), \left(H_2^f, H_2^o\right), \dots, \left(H_\Gamma^f, H_\Gamma^o\right)\right)^T$. Besides H^f and H^o, the hourly wind power point forecasts across the entire planning horizon, $P = (P_1, P_2, \dots, P_T)^T$, need to be available. After normalizing the data, the elements of P are sorted into K predetermined equidistant forecast intervals as proposed by Bludszuweit et al. (2008) to assess the forecast error distribution of a given point forecast. The number of forecast intervals, K, that usually correlates with the length of $H^{f,o}$, determines the width of each interval. For instance, in the numerical example presented in Section 5, the calibration period embraced 1026 tuples of historical wind power forecasts and observations. Thus, we used ten forecast intervals, resulting in a width of each forecast interval of 0.1. Consequently, a normalized hourly point forecast of $P_t = 0.12$ would be sorted into the second forecast interval ($k = 2$). In the same way, each tuple of $H^{f,o}$ is subsequently sorted into the forecast interval according to the value of H^f of a tuple. For each forecast interval k, the empirical cumulative distribution function (ecdf), $F_k(\cdot)$, is then calculated based on the values of H^o of the tuples sorted into forecast interval k. This enables us to derive the forecast error distribution of a point forecast P_t falling into this interval (Ma et al., 2013).

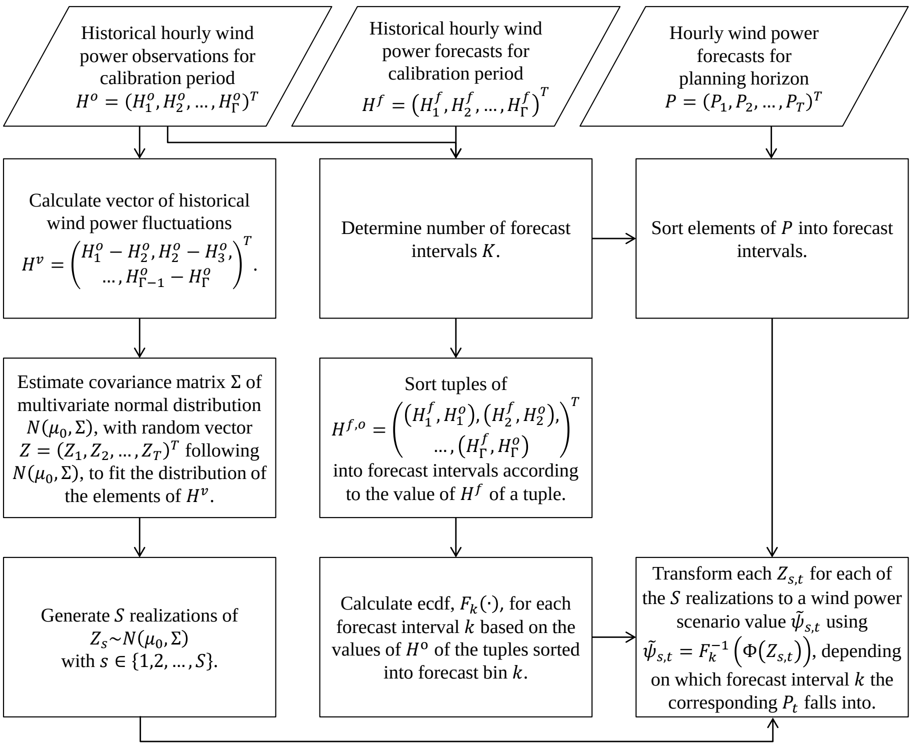


Figure 3: Flow chart describing the wind power scenario generation process based on Ma et al. (2013).

Pinson et al. (2009) and Ma et al. (2013) showed that a common random vector $Z = (Z_1, Z_2, \dots, Z_T)^T$ which follows a multivariate normal distribution, $Z \sim N(\mu_0, \Sigma)$, can be used to incorporate the interdependence of successive wind power forecast errors into the scenario generation process, independent of the different forecast error distributions of different forecast intervals. While μ_0 corresponds to a vector of zeros, the covariance matrix Σ equals

$$\Sigma = \begin{bmatrix} \sigma_{1,1} & \sigma_{1,2} & \cdots & \sigma_{1,T} \\ \sigma_{2,1} & \sigma_{2,2} & \cdots & \sigma_{2,T} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{T,1} & \sigma_{T,2} & \cdots & \sigma_{T,T} \end{bmatrix} \tag{24}$$

where $\sigma_{t,t'} = cov(Z_t, Z_{t'})$ with $t, t' \in \{1,2,\dots,T\}$. This covariance matrix summarizes the correlations of the random variables Z_t of different lead times and hence expresses the variability inherent in wind power time series. The structure of Σ is estimated to fit the distribution of the set of historical wind power fluctuations, $H^v = (H_1^o - H_2^o, H_2^o - H_3^o, \dots, H_{\Gamma-1}^o - H_\Gamma^o)^T$ (Ma et al., 2013). Using the multivariate normal random number generator in Matlab, the desired number of realizations of $Z_s \sim N(\mu_0, \Sigma)$ with $s \in \{1,2,\dots,S\}$ can be generated. Finally, each element $Z_{s,t}$ of each of these realizations needs to be transformed to the forecast error distribution of the forecast interval the corresponding point forecast P_t belongs to. The rationale behind this transformation is to identify the wind power scenario value $\tilde{\psi}_{s,t}$ whose functional value on the ecdf of forecast interval k, $F_k(\tilde{\psi}_{s,t})$, equals the functional value of $Z_{s,t}$ on the cumulative distribution function (cdf) of the normal distribution, $\Phi(Z_{s,t})$:

$$\tilde{\psi}_{s,t} = F_k^{-1}\left(\Phi(Z_{s,t})\right), \tag{25}$$

where $F_k^{-1}(\cdot)$ corresponds to the inverse of the ecdf of forecast interval k (Ma et al., 2013). This inverse transformation method was described by Liu and Der Kiureghian (1986) and is illustrated in Figure 4.

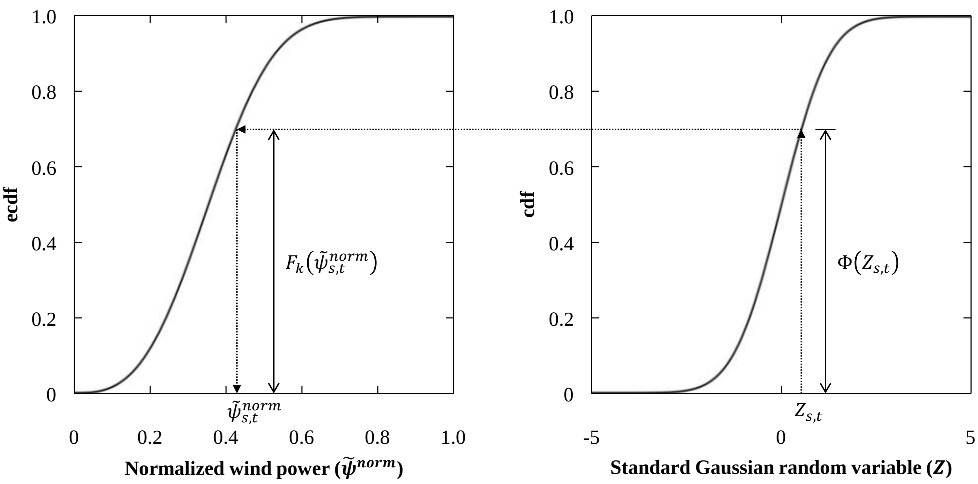


Figure 4: Graphical representation of the inverse transformation expressed in Equation (25).

To sufficiently account for the uncertainty attached to the energy supply from the wind turbines, Ma et al. (2013) recommend generating a set of at least 500 scenarios, WP^o, using the process described above, where the same probability of occurrence π_s^o is assigned to each scenario $s \in WP^o$. However, looking at the model proposed in Section 3.2, it is obvious that the computational burden of the solution process is directly linked to the number of scenarios considered. For this reason, we relied on the concept of scenario reduction. Scenario reduction aims at identifying a reduced subset, WP^r, of the original scenario set, WP^o, where the stochastic process described by WP^r is supposed to resemble the stochastic process described by WP^o as accurately as required by given probability distance measures (Gröwe-Kuska et al., 2003). Transferred to our application, this means that we are looking to single out as few wind power scenarios from WP^o as possible that jointly reflect the wind power supply uncertainty described by WP^o as sufficiently as demanded in terms of the closeness of the stochastic processes of WP^o and WP^r. In this way, the computational burden of the solution process can be reduced significantly at the expense of a presumably small loss of accuracy in describing the uncertainty of the wind power supply. In the original scenario set, each scenario $s \in WP^o$ is assigned the same probability of occurrence, π_s^o. In the reduced scenario set, each scenario $s' \in WP^r$ is assigned the combined probabilities of occurrence of all scenarios contained in WP^o the scenario in question is closest to, according to the given probability distance measures. In this paper, we adapted the scenario reduction algorithm developed by Li et al. (2016) to identify a reduced scenario set, WP^r, that concurrently minimizes the probability distances measures of the space distance, f^{sp}, and the moment distance, f^{mom}, of WP^r to WP^o (Li et al., 2016):

$$f^{sp} = \frac{1}{T} \cdot \sum_{s \in WP^o \setminus WP^r} \pi_s^o \cdot \min\left\{ \left| \widetilde{\Psi}_s - \widetilde{\Psi}_{s'} \right| \middle| s' \in WP^r \right\}, \tag{26}$$

$$f^{mom} = \max\left\{ \left(\begin{array}{c} \frac{1}{T} \cdot \sum_{t=1}^{T} \sum_{s \in WP^o} \pi_s^o \cdot \left(\tilde{\psi}_{s,t}^o - \sum_{s \in WP^o} \pi_s^o \cdot \tilde{\psi}_{s,t}^o \right)^\eta \\ - \frac{1}{T} \cdot \sum_{t=1}^{T} \sum_{s' \in WP^r} \pi_{s'}^r \cdot \left(\tilde{\psi}_{s',t}^r - \sum_{s' \in WP^r} \pi_{s'}^r \cdot \tilde{\psi}_{s',t}^r \right)^\eta \end{array} \right)^2 \middle| \eta \in \{1,2,3,4\} \right\}. \tag{27}$$

4.2 Two-stage stochastic production scheduling and energy supply decision method

In the next step, the reduced wind power scenario set, WP^r, serves as an input to a two-stage stochastic production scheduling and energy supply decision method (see Figure 5). In the first stage (before the start of the planning horizon, i.e., at $t = 0$), the proposed method allocates jobs to machines and time slots ($x_{i,j,t}$, $y_{i,j,t}$) and determines whether or not the energy generated by the wind turbine should be used to run the machines, or be fed into the grid (ϕ_t). This is achieved by solving the MILP problem defined in Section 3.2. As the MILP problem concurrently minimizes both $TWFT$ and EC, a bi-objective optimization problem results (see Section 4.2.1). The second stage of the two-stage stochastic scheduling problem then gradually adjusts the energy supply decisions, computed in the first stage, when the actual wind power data are revealed in each time interval $t \in \{1,2,\dots,T\}$ (see Section 4.2.2).

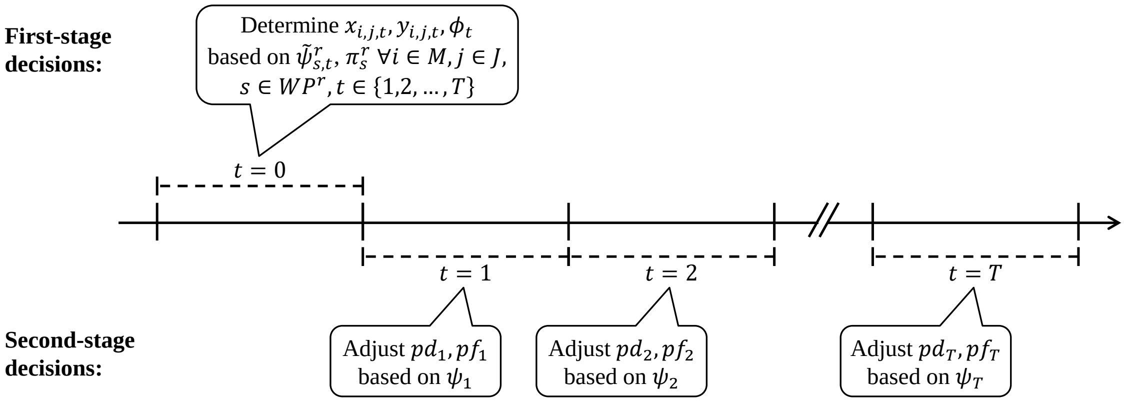


Figure 5: Overview of the decisions made during the application of the two-stage stochastic scheduling method.

4.2.1 Bi-objective optimization

As we seek to concurrently minimize $TWFT$ as well as EC, we face a bi-objective optimization problem. In practice, it is usually not possible to identify a solution that minimizes both objectives simultaneously (Antunes et al., 2004). For this reason, various techniques have been developed to arrive at a good trade-off between competing objectives (Ehrgott, 2005). Among these techniques, the weighted sum approach has become increasingly popular in recent years, particularly due to its simplicity in use and interpretation (Kim and Weck, 2005). The weighted sum approach first assigns a weight to each objective function. These weights represent the preferences of the decision-maker among the objectives considered. In the next step, the sum of the products of the objective functions and the assigned weights is minimized. As in our case $TWFT$ is expressed in hours, while EC is expressed in USD, the objective functions need to be normalized by means of the ideal point, Ω^i, and the nadir point, Ω^n. While $\Omega^i = \left(TWFT^{min}, EC^{min}\right)$ can be found by minimizing both objective functions independently, $\Omega^n = (TWFT^{max,PF}, EC^{max,PF})$ corresponds to the worst objective values on the Pareto optimal front. Thus, the normalized objective function of the weighted sum approach equals

$$\min\left(\omega_1 \cdot \left(\frac{TWFT - TWFT^{min}}{TWFT^{max,PF} - TWFT^{min}}\right) + \omega_2 \cdot \left(\frac{EC - EC^{min}}{EC^{max,PF} - EC^{min}}\right)\right), \tag{28}$$

where $\omega_1, \omega_2 \geq 0$ and $\omega_1 + \omega_2 = 1$ need to hold. By minimizing Equation (28) for various combinations of ω_1 and ω_2, a set of non-dominated solutions is generated. However, it is important to recognize that the weighted sum approach is only capable of computing non-dominated *vertex* solutions for bi-objective MILP problems. This is due to the fact that the non-dominated vertex solutions determine the boundary of the convex hull of the set of non-dominated solutions in the solution space of a bi-objective MILP problem. In contrast, the weighted sum approach is not capable of finding non-dominated solutions inside the convex hull as they are dominated by a convex combination of vertex solutions (see Appendix B and Antunes et al., 2004). However, it is not the main goal of this article to guarantee that the proposed solution procedure identifies all non-dominated solutions. In fact, this article rather aims at efficiently computing the boundary of the convex hull of the set of non-dominated solutions to assess the fundamental economic and environmental value of integrating RES into production scheduling.

4.2.2 Real-time adjustments of energy supply decisions

After computing a production schedule and energy supply decisions based on the reduced set of wind power scenarios in the first stage of the two-stage scheduling problem, the second stage modifies the energy supply decisions when the actual values of the wind power data are revealed over time. To this end, a real-time adjustment action, defined in Algorithm 1, modifies pd_t and pf_t in each time slot t, depending on the amount of energy actually supplied by the wind turbine in this time slot (ψ_t). Through these adjustments, the energy supply decisions are adapted to the actual wind power and the true energy cost associated with the production schedule and energy supply decisions can be calculated based on Equations (3) and (4), whereas the index of the wind power scenarios, s, can be neglected.

Algorithm 1: Real-time adjustment of energy supply decisions.

input: ψ_t, ϕ_t, $y_{i,j,t}$, $q_{i,j}$ $\forall i \in M, j \in J$
if $\phi_t = 1$ **then** // wind power is used to run machines
    if $\psi_t \leq \sum_{i \in M} \sum_{j \in J} y_{i,j,t} \cdot q_{i,j}$ **then** // wind power is smaller than or equal to power required by machines
        $pd_t = \sum_{i \in M} \sum_{j \in J} y_{i,j,t} \cdot q_{i,j} - \psi_t$
        $pf_t = 0$
    else // wind power exceeds power required by machines
        $pd_t = 0$
        $pf_t = \psi_t - \sum_{i \in M} \sum_{j \in J} y_{i,j,t} \cdot q_{i,j}$
    endif
else // wind power is fed into the grid
    $pd_t = \sum_{i \in M} \sum_{j \in J} y_{i,j,t} \cdot q_{i,j}$
    $pf_t = \psi_t$
endif
return pd_t, pf_t

5 Demonstration of the method in a numerical example

To evaluate the model and its intended effectiveness towards the integration of energy supply from an onsite wind turbine into production scheduling, we investigate a numerical example. By means of this numerical example, we pursue three primary goals:

(1) assess the value for manufacturing companies of investing in a wind turbine by comparing the set of non-dominated solutions when considering the wind turbine with the set of non-dominated solutions without wind turbine;

(2) evaluate the treatment of the uncertainty by comparing the set of non-dominated solutions associated with the first-stage decisions of the two-stage stochastic scheduling method with the corresponding set of adjusted solutions associated with the second-stage decisions;

(3) assess by how much the results of the proposed solution procedure deviate from the 'theoretical' optimum, in case the actual wind power data were known in advance.

5.1 Background for the numerical example

The entire numerical example is built around a hypothetical manufacturing firm located in Hayward, California. We consider a flow shop system consisting of three machines. This flow shop system needs to process three jobs within the planning horizon that spans two 8 h work shifts from 6 a.m. to 10 p.m. The planning horizon is split up into time slots of one hour (l). To accomplish the goals of this numerical example, ten random problem instances describing both the machines and the jobs to be processed were generated. The weights, representing the importance of the jobs, w_j, were randomly generated from a uniform distribution [1, 5]. The job processing times, $p_{i,j}$, were randomly generated from a uniform distribution [1 h, 4 h]. The processing power requirements of the machines, $q_{i,j}$, were randomly generated from a uniform distribution [50 kW, 200 kW]. Appendix C contains a comprehensive overview of the ten problem instances considered.

Electricity prices follow the predetermined TOU summer pricing scheme specified in Table 1, which is derived from a rate schedule for industrial customers of the Pacific Gas and Electric Company (Pacific Gas and Electric Company, 2016a). We chose this pricing scheme as the Pacific Gas and Electric Company serves the San Francisco Bay Area including Hayward, California, where our hypothetical manufacturing company is located. We assume that the working day we investigate and the associated workload represent a standard working day of the company considered. Hence, as in Wang and Li (2013), we suppose that the power peak load of this working day equals or almost equals the power peak load of the working days in a month. Consequently, assuming one month to consist of 21 working days, the power demand-related cost needs to be divided by 21. This is due to the fact that we only consider one day in this numerical example and the power demand charge is only paid once per billing period with reference to the maximum power consumption in the maximum-peak time slots, the partial-peak time slots, and the entire planning horizon within the monthly billing period. The feed-in tariff, r^{FiT}, employed in this study resembles the Electric-Renewable Market Adjusting Tariff of the Pacific Gas and Electric Company (2016b) (see Table 1).

Table 1: TOU summer pricing profile.

Type of period	**Time of day**	$\boldsymbol{c_t}$ **[USD/kWh]**	$\boldsymbol{d^p}$ **[USD/kW]**	$\boldsymbol{d^{pp}}$ **[USD/kW]**	$\boldsymbol{d^{mp}}$ **[USD/kW]**	$\boldsymbol{r^{FiT}}$ **[USD/kWh]**
Off-peak	9 p.m. – 8 a.m.	0.07634	16.89	-	-	0.08923
Partial-peak	8 a.m. – 11 a.m. and 6 p.m. – 9 p.m.	0.10141	16.89	5.05	-	0.08923
Maximum-peak	11 a.m. – 6 p.m.	0.13793	16.89	-	18.14	0.08923

Based on data provided by the U.S. Energy Information Administration (2015), the levelized cost of electricity generated by the wind turbine, c^W, is assumed to equal 0.0736 USD/kWh. For the wind power generation, we consider a wind turbine of type E-48 with a hub height of 50 m, produced by Enercon. This type of wind turbine features a cut-in wind speed, V^{cut-in}, of 3 m/s, a rated wind speed, V^{rated}, of 14 m/s, and a cut-out wind speed, $V^{cut-out}$, of 25 m/s. The swept area of the wind turbine rotor, A, amounts to 1810 m^2. For further information, particularly regarding the power coefficient, $C^p(V_{s,t})$, the reader is referred to Enercon GmbH (2015). The air density, ρ, is assumed to equal 1.22 kg/m^3 (Carta and Mentado, 2007).

The MILP problem defined in Section 3.2 was solved using the software IBM ILOG CPLEX Optimization Studio 12.5.1 with default settings on an Intel® CoreTM i7-4770 CPU @ 3.40 GHz with 16 GB RAM and the Windows 10 operating system. The CPU time for solving the MILP problem for one combination of ω_1 and ω_2 was, on average, 26.31 s considering the reduced wind power scenario sets, and, on average, 1.30 s when no wind power or solely the actual wind power scenario was considered.

5.2 Wind power scenarios for numerical example

In a first step, we investigated the ten random problem instances for a day featuring average wind conditions (July 1, 2016). In a second step, we singled out one problem instance and studied the developed production scheduling and energy supply decision procedure in case of very weak (October 3, 2016), very strong (May 20, 2016), and highly volatile wind conditions (July 8, 2016). For each test day, we generated 1000 hourly wind power scenarios, which span the entire planning horizon from 6 a.m. to 10 p.m., for the hypothetical manufacturing company in Hayward, California, based on the procedure described in Section 4.1. The input data required to generate the wind power scenarios (the historical wind power observations, H^o, and the historical wind power forecasts, H^f, for the calibration period (here: two months), and the wind power forecasts for the planning horizon, P) were based on wind speed data taken from the weather station KHWD at Hayward Executive Airport published by The Weather Company.[3] In the next step, we reduced the generated scenarios as explained in Section 4.1. Ultimately, we ended up with 53 (47, 58, 52) wind power scenarios for July 1, 2016 (October 3, 2016, May 20, 2016, July 8, 2016). Figure 6 presents the reduced wind power scenario set for July 1, 2016.

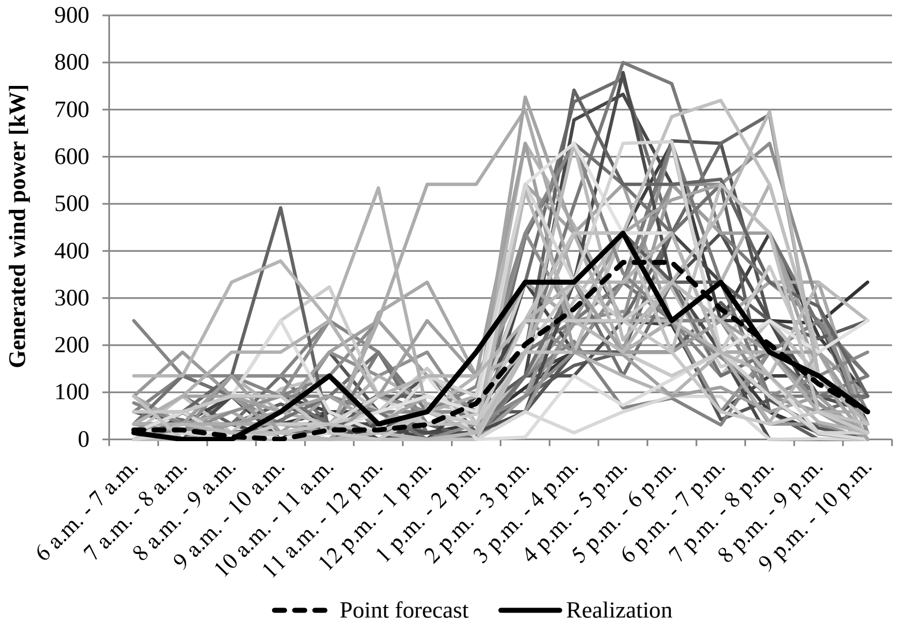


Figure 6: Reduced set of generated wind power scenarios compared with point forecast and realization for Hayward, California, for July 1, 2016, between 6 a.m. and 10 p.m.

5.3 Presentation of results

To compute the sets of non-dominated solutions of the ten problem instances studied, we systematically varied ω_1 and ω_2 at a step width of 0.1 and subsequently minimized Equation (28), taking account of Equations (5) to (22). As becomes evident from Figure 7, some combinations of ω_1 and ω_2 led to the same results. For this reason, the curves do not necessarily feature one point for each combination

[3] Historical wind speed forecasts and observations for the calibration period and wind speed forecasts for the planning horizon are available at https://www.wunderground.com/q/zmw:94580.4.99999.

of ω_1 and ω_2. Figure 7 illustrates the sets of non-dominated solutions for average wind conditions in case (I) no wind turbine was considered, (II) a wind turbine was considered and the first-stage decisions had already been made, and (III) a wind turbine was considered and the actual wind power data were known in advance. Besides, Figure 7 shows the sets of solutions of case (II) after they have been adjusted in the second stage of the proposed two-stage stochastic scheduling program. However, it needs to be recognized that the second-stage decisions are not based on an optimization algorithm. Instead, they feature actions that adjust the energy supply decisions of the first stage when the actual wind power data are gradually revealed while leaving the production scheduling decisions unchanged. Thus, the resulting sets of adjusted solutions may also contain solutions that are dominated by other solutions as illustrated in Figure 8.

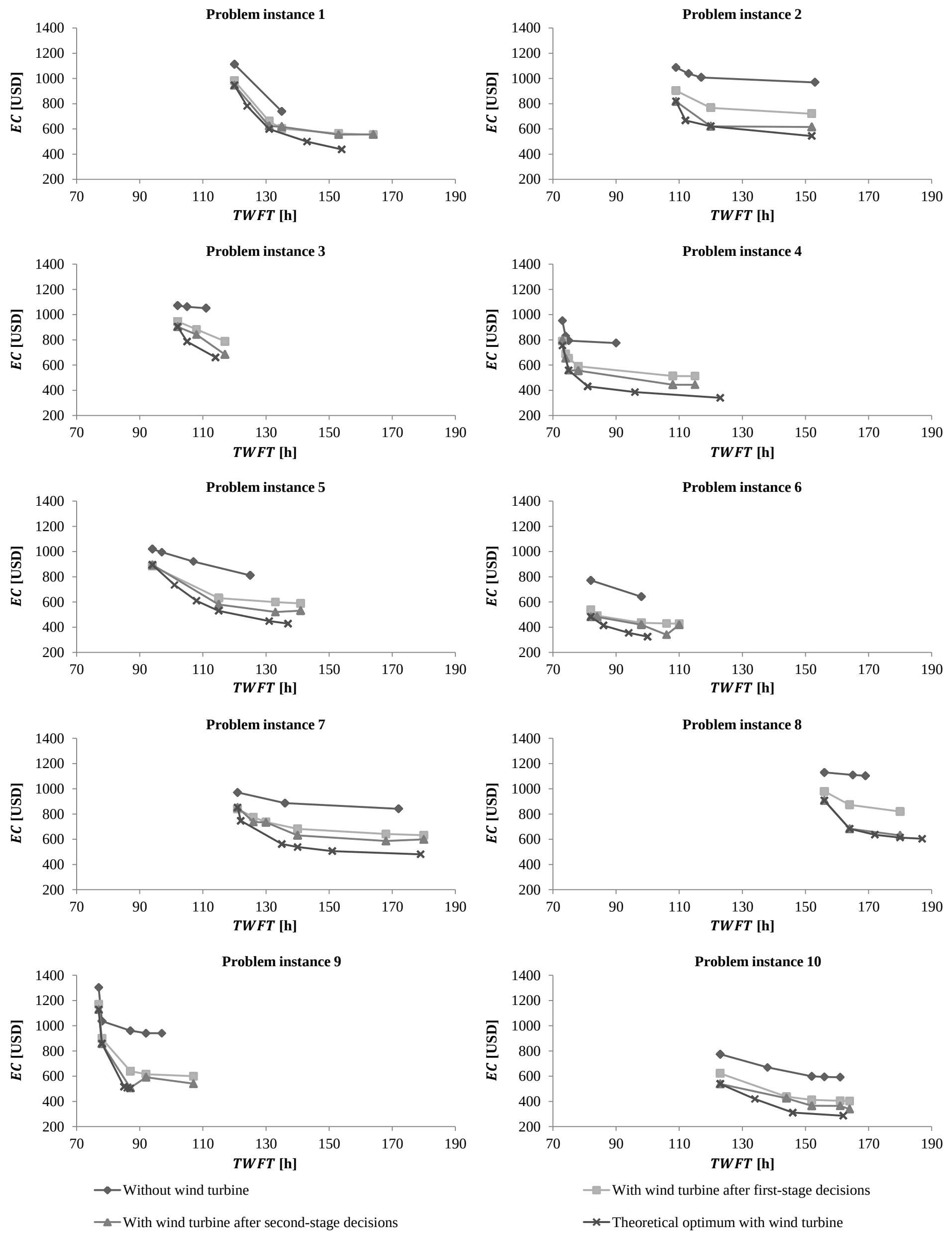


Figure 7: Sets of non-dominated and adjusted solutions of the ten random problem instances under average wind conditions.

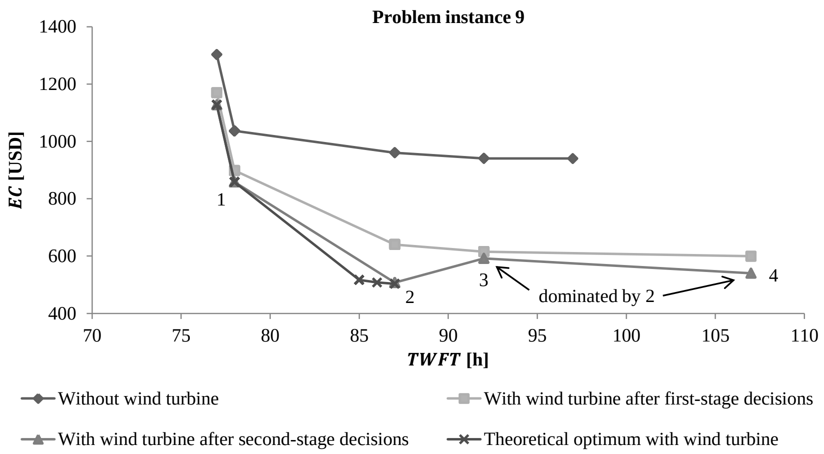


Figure 8: Sets of non-dominated and adjusted solutions of problem instance 9 under average wind conditions.

In general, Figure 7 indicates that the curves associated with the flow shop system that utilizes a wind turbine dominate the flow shop system without a wind turbine for all ten problem instances. Thus, the investment in a wind turbine seems to be worth it for the investigated application, at least when the production scheduling and energy supply decisions are coordinated. Furthermore, the proposed solution procedure seems to handle the uncertainty stemming from the intermittent character of the energy supply from the wind turbine very well in case of average wind conditions. This hypothesis rests upon the fact that the set of non-dominated solutions after the first-stage and the set of solutions adjusted in the second stage under consideration of a wind turbine lie very close together for most of the problem instances studied. Thus, the impact of the second-stage decisions is relatively small. To illustrate this point, a comparison of the energy cost resulting from the first-stage and second-stage solutions associated with the same preferences, expressed in ω_1 and ω_2, is sufficient. This is due to the fact that the production scheduling decisions and consequently $TWFT$ are not altered in the second stage of the two-stage stochastic scheduling program (see Figure 5). On average, the energy cost after the first-stage decisions deviated from the energy cost after the second-stage decisions by 8.9 percent. The minimum absolute deviation of the first-stage and second-stage solutions was 0.3 percent in the case of problem instance 1, while the maximum absolute deviation was 23.0 percent for problem instance 8. Overall, it is remarkable that in almost 90 percent of the combinations of ω_1 and ω_2 the second-stage decisions improved the solutions arrived at after the first-stage decisions. Yet, it needs to be kept in mind that no general conclusions may be drawn regarding the ability of the second stage to improve on solutions from the first stage as only one test day was investigated.

The value of the developed solution procedure is further highlighted when comparing the sets of solutions adjusted in the second stage with the sets of non-dominated solutions in case the actual wind power data were known in advance. Figure 7 shows that the corresponding solutions sets are quite close to each other in the majority of the problem instances and partly even feature identical solutions (see solution 1 in Figure 8). When solely looking at the solutions associated with the second-stage decisions, it is noteworthy that, starting from a pure minimization of the total weighted flow time, giving even only a little weight to energy cost may impact the outcome significantly. In case of problem instance 4, for example, increasing ω_2 from 0 to 0.15 decreased EC by 30.1 percent at the expense of an increase in $TWFT$ of 2.7 percent. In case of problem instance 9, increasing ω_2 from 0 to 0.05 cut

EC by 23.0 percent, while $TWFT$ grew by 1.3 percent. A further increase of ω_2 from 0.05 to 0.35 even decreased EC by 55.0 percent at the expense of an increase in $TWFT$ of 13.0 percent, compared to the corresponding values at $\omega_2 = 0$. Hence, giving only a little weight to the energy cost when computing the production plan can have a major impact on the overall production scheduling result, depending on the job processing times, the machine power requirements, and the importance of the jobs.

To identify potential limitations of the applicability of the developed method, we evaluated its performance under extreme wind conditions using problem instance 9. Figure 9 illustrates that the manufacturing firm can only benefit from the use of onsite wind power in case wind is sufficiently strong. In the case of strong wind conditions, it is particularly noteworthy that the curve of the theoretical optimum reduced to a single point as the actual amount of wind power generated on May 20, 2016, was higher than the power required by the machines during every time slot in the planning horizon. Hence, the scheduling problem becomes independent of the time-varying energy prices and its solution is solely driven by $TWFT$. Finally, Figure 9 shows that highly volatile wind conditions may pose a challenge to the coordination of production scheduling and energy supply decisions as the scenario generation procedure can hardly predict such rare conditions based on historic wind power fluctuations. As a consequence, the gaps between the non-dominated solutions after the first stage and the solutions adjusted in the second stage are larger than in case of less volatile wind conditions. Yet, the set of adjusted solutions is still fairly close to the theoretical optimum.

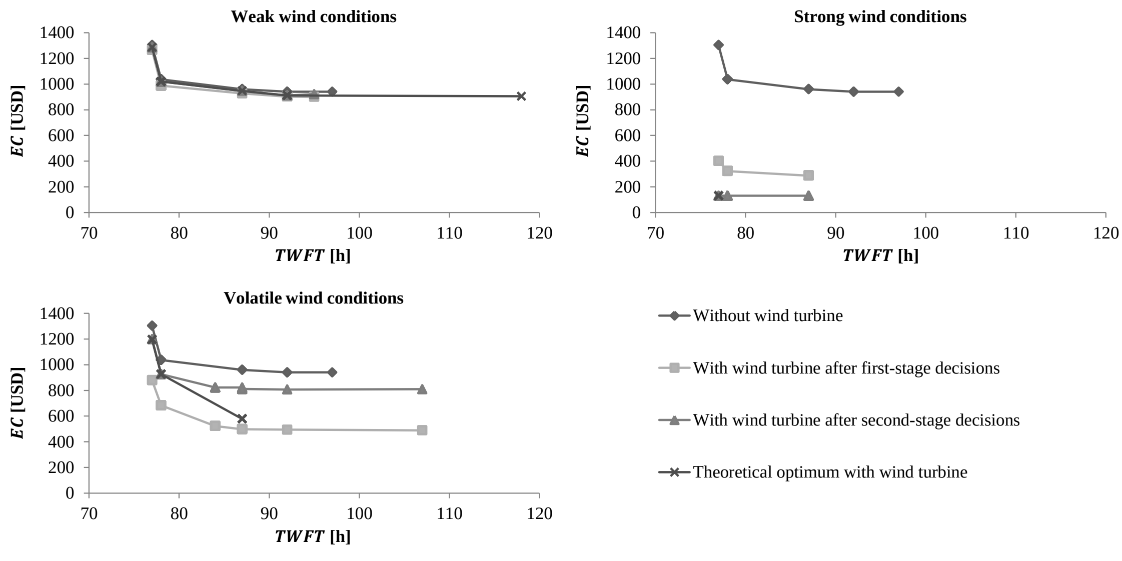


Figure 9: Sets of non-dominated and adjusted solutions of problem instance 9 under extreme wind conditions[4].

[4] Please note that it is possible that the sets of non-dominated solutions after first-stage decisions dominate the sets of non-dominated solutions associated with the theoretical optimum. This event may occur in case the expected wind power the first-stage decisions are based on exceeds the actual wind power the theoretical optimum is based on.

Overall, Figures 7 to 9 revealed partly remarkable differences in $TWFT$ and EC depending on whether or not an onsite wind turbine was considered and depending on the weights of the objectives. These differences can be traced back to diverging production schedules for the same problem instances. Based on problem instance 5, Figure 10 provides an indication of the extent to which the use of wind power as well as varying priorities of the objectives may impact production schedules. In case no wind turbine is considered, the curves reflecting the machine power demand and the power demanded from the grid are equal. In case a wind turbine is considered, these curves deviate in consequence of the use of wind power for running the machines. If minimizing $TWFT$ is the sole interest of the decision-maker ($\omega_1 = 1$, $\omega_2 = 0$), it is obvious that considering wind power will not impact the production schedule. When assigning more and more weight to the energy-related goal (i.e., decreasing ω_1 and increasing ω_2), the production schedules change. As a result, the machine power demand curve features local maxima in time slots associated with lower energy consumption charges and lower power demand charges (i.e., off- and partial-peak time slots) and a local minimum in time slots associated with high energy consumption charges and high power demand charges (i.e., maximum-peak time slots) in case no wind turbine is considered. Wind power clearly mitigates this trend: In time slots where the machine power demand curve without wind turbine reached local minima, the machine power demand curves with wind turbine feature rather high values or even local maxima (see Figure 10, $\omega_1 = 0.5$, $\omega_2 = 0.5$, and $\omega_1 = 0$, $\omega_2 = 1$). This is due to the large amounts of wind power available in these time slots which cause the curve of the power demanded from the grid to drop significantly. Increasingly prioritizing the energy-related goal leads to production schedules whose machine power demand curve gradually aligns with the curve of generated wind power. Overall, wind power can grant production planners considerable independence of the dictate of TOU pricing schemes, even if energy-related goals are at the center of their attention. Additionally, Figure 10 illustrates that the production schedules calculated based on the presented wind power scenario generation and reduction method are very close to the production schedules that are optimal for the wind power actually available. This again highlights the practical applicability of the developed method.

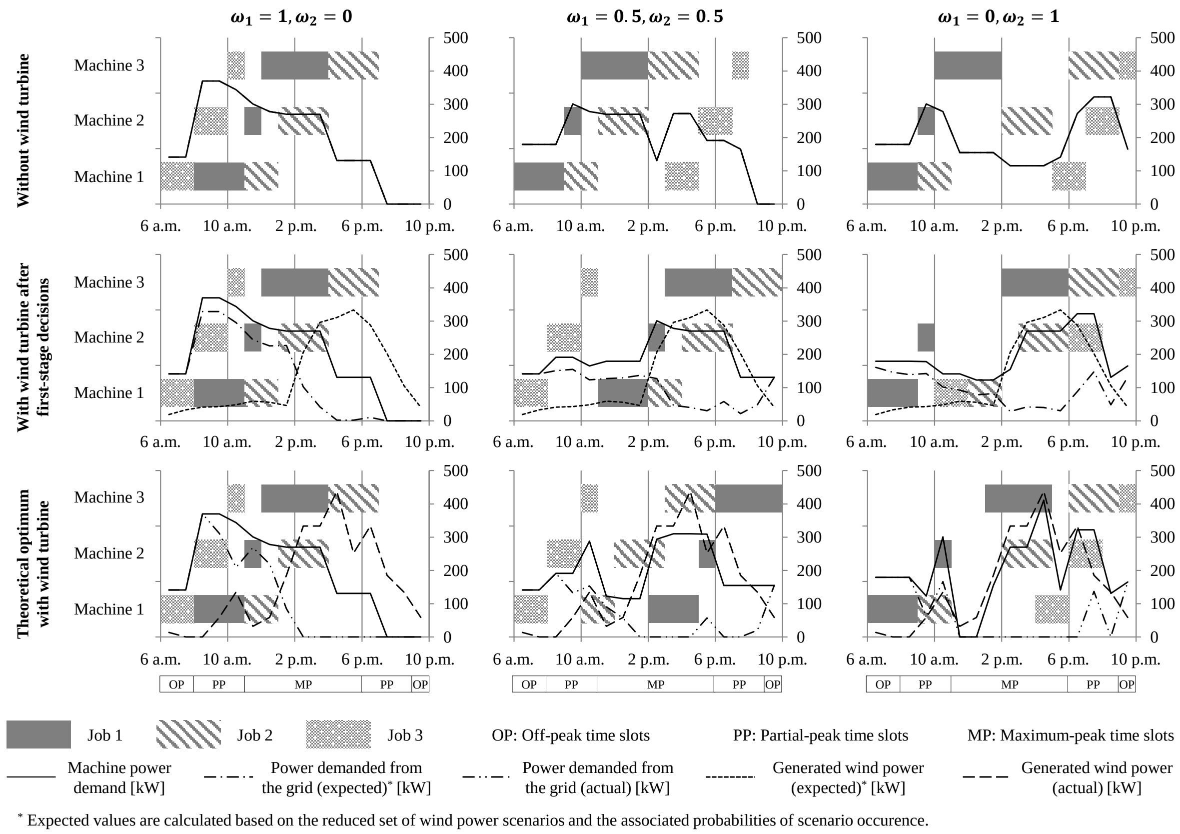


Figure 10: Comparison of production schedules of problem instance 5 under average wind conditions with exclusive focus on $TWFT$ (ω_1 = 1, ω_2 = 0), equally weighted focus on $TWFT$ and EC (ω_1 = 0.5, ω_2 = 0.5), and exclusive focus on EC (ω_1 = 0, ω_2 = 1).

6 Conclusions and implications

Global installed renewable energy capacity has increased significantly over the last decade, and wind power has been recognized as a main driver of this movement. Manufacturing companies, which have already been identified as major energy consumers, may capitalize on the use of wind power by reducing both their cost and energy-related GHG emissions. However, to make efficient use of the energy generated by wind turbines, practitioners are in need of effective decision support tools that incorporate the energy supply from wind turbines into the production scheduling process. The effectiveness of such tools is particularly contingent on their capacity to take account of the intermittent character associated with wind power. This paper proposed a decision support model that integrated this intermittency into a two-stage stochastic scheduling method by generating a large number of wind power scenarios. In this way, the model not only considered the variability inherent in wind power over time, but also the time-dependence of successive wind power outputs. After reducing the number of generated scenarios, they were fed into an MILP problem, which computed a production schedule as well as energy supply decisions in the first stage of the two-stage stochastic scheduling program. To this end, the proposed MILP problem concurrently minimized the total weighted flow time and the energy cost under consideration of a TOU tariff featuring both an energy consumption charge and a power demand charge. In the second stage, the energy supply decisions were gradually revised when the actual values of the wind power data were revealed. Thus, the holistic solution procedure presented in this paper may assist practitioners in computing production plans which strive for both production-related and energy-related objectives while considering power generated by non-dispatchable RES such as wind or solar power.

From the hypothetical case study that was based on real-life machine power requirements, real wind speed observations, and a current TOU tariff, several managerial implications can be derived:

- Depending on geographical and atmospheric circumstances as well as production process-related characteristics such as machine power requirements and job allocation and sequencing flexibility, RES can play a vital role in reducing energy cost and fostering environmental goals in manufacturing.
- The wind power scenario generation and reduction techniques employed in this paper are capable of sufficiently handling the uncertainty attached to the energy supply from wind turbines. As a result, the proposed solution procedure allows practitioners to compute production schedules and energy supply decisions that reach almost the same results as if the wind power scenarios were known in advance, even in case of extreme wind conditions.
- Shifting only a little weight from production-related goals to energy-related goals can have a major impact on the overall result. Hence, decision-makers may significantly decrease the energy cost at the expense of a minor increase in the total weighted flow time by including the energy cost in the objective function.
- Companies may use wind power as a valuable tool to considerably reduce the influence of time-varying energy prices on production scheduling, especially in case energy-related goals are crucial to the management.

In terms of future research, energy-related GHG emissions may be considered in the objective function or via another constraint. It may also be of interest to investigate the extent to which results change if not only the energy supply decisions are adjusted when the actual wind power is gradually revealed, but also the production schedule itself.

References

Al-Yahyai, S., Charabi, Y., Gastli, A., 2010. Review of the use of Numerical Weather Prediction (NWP) models for wind energy assessment. *Renewable and Sustainable Energy Reviews*, 14 (9), 3192-3198.

Antunes, C.H., Martins, A.G., Brito, I.S., 2004. A multiple objective mixed integer linear programming model for power generation expansion planning. *Energy*, 29 (4), 613-627.

Biel, K., Glock, C.H., 2016. Systematic literature review of decision support models for energy-efficient production planning. *Computers & Industrial Engineering*, 101, 243-259.

Bludszuweit, H., Dominguez-Navarro, J.A., Llombart, A., 2008. Statistical analysis of wind power forecast error. *IEEE Transactions on Power Systems*, 23 (3), 983-991.

Brett, A.C., Tuller, S.E., 1991. The autocorrelation of hourly wind speed observations. *Journal of Applied Meteorology*, 30 (6), 823-833.

Brown, B.G., Katz, R.W., Murphy, A.H., 1984. Time series models to simulate and forecast wind speed and wind power. *Journal of Climate and Applied Meteorology*, 23 (8), 1184-1195.

Burton, T., Sharpe, D., Jenkins, N., Bossanyi, E., 2011. *Wind energy handbook*. 2nd ed., Wiley, Chicester.

Carta, J. A., Mentado, D., 2007. A continuous bivariate model for wind power density and wind turbine energy output estimations. *Energy Conversion and Management*, 48 (2), 420-432.

Corotis, R.B., Sigl, A.B., Cohen, M.P., 1977. Variance analysis of wind characteristics for energy conversion. *Journal of Applied Meteorology*, 16 (11), 1149-1157.

Ehrgott, M., 2005. *Multicriteria optimization*. 2nd ed., Springer, Berlin Heidelberg.

Enercon GmbH, 2015. *Enercon product overview*. http://www.enercon.de/en/products/ep-1/e-48/ (retrieved July 6, 2016).

European Parliament and Council, 2009. Energy efficiency directive (2009/28/EC). *Official Journal of the European Union*, 52, L 140/16-L 140/62.

Fang, K., Uhan, N., Zhao, F., Sutherland, J.W., 2011. A new approach to scheduling in manufacturing for power consumption and carbon footprint reduction. *Journal of Manufacturing Systems*, 30 (4), 234-240.

Foley, A.M., Leahy, P.G., Marvuglia, A., McKeogh, E.J., 2012. Current methods and advances in forecasting of wind power generation. *Renewable Energy*, 37 (1), 1-8.

Gröwe-Kuska, N., H. Heitsch, and W. Römisch. 2003. *Scenario reduction and scenario tree construction for power management problems*. In: Proceedings of the 2003 IEEE Bologna Power Tech Conference, Bologna, vol. 3, pp. 1-7.

Hake, J.-F., Fischer, W., Venghaus, S., Weckenbrock, C., 2015. The German Energiewende – History and status quo. *Energy*, 92, 532-546.

Hocaoğlu, F.O., 2010. Novel analytical hourly solar radiation models for photovoltaic based system sizing algorithms. *Energy Conversion and Management*, 51 (12), 2921-2929.

Iversen, E.B., Morales, J.M., Møller, J.K., Madsen, H., 2016. Short-term probabilistic forecasting of wind speed using stochastic differential equations. *International Journal of Forecasting*, 32, 981-990.

Kim, I.Y., Weck, O.L., 2005. Adaptive weighted sum method for multiobjective optimization: A new method for Pareto front generation. *Structural and Multidisciplinary Optimization*, 31, 105-116.

Lantz, E., Sigrin, B., Gleason, M., Preus, R., Baring-Gould, I., 2016. *Assessing the future of distributed wind: Opportunities for behind-the-meter projects*. National Renewable Energy Laboratory, Golden, CO.

Lei, D., Zheng, Y., Guo, X., 2017. A shuffled frog-leaping algorithm for flexible job shop scheduling with the consideration of energy consumption. *International Journal of Production Research*, 55 (11), 3126-3140.

Li, J., Lan, F., Wei, H., 2016. A scenario optimal reduction method for wind power time series. *IEEE Transactions on Power Systems*, 31 (2), 1657-1658.

Liu, C.-H., 2016. Mathematical programming formulations for single-machine scheduling problems while considering renewable energy uncertainty. *International Journal of Production Research*, 54 (4), 1122-1133.

Liu, G.-S., Yang, H.-D., Cheng, M.-B., 2017. A three-stage decomposition approach for energy-aware scheduling with processing-time-dependent product quality. *International Journal of Production Research*, 55 (11), 3073-3091.

Liu, P.-L., Der Kiureghian, A., 1986. Multivariate distribution models with prescribed marginals and covariances. *Probabilistic Engineering Mechanics*, 1 (2), 105-112.

Luo, H., Du, B., Huang, G.Q., Chen, H., Li, X., 2013. Hybrid flow shop scheduling considering machine electricity consumption cost. *International Journal of Production Economics*, 146 (2), 423-439.

Ma, X.-Y., Sun, Y.-Z., Fang, H.-L., 2013. Scenario generation of wind power based on statistical uncertainty and variability. *IEEE Transactions on Sustainable Energy*, 4 (4), 894-904.

May, G., Stahl, B., Taisch, M., Prabhu, V., 2015. Multi-objective genetic algorithm for energy-efficient job shop scheduling. *International Journal of Production Research*, 53 (23), 7071-7089.

McCrone, A., Moslener, U., Usher, E., Grüning, C., Sonntag-O'Brian, V., 2015. *Global trends in renewable energy investment 2015*. Frankfurt School UNEP Collaborating Centre for Climate and Sustainable Energy Finance, Frankfurt (Main).

Moon, J.-Y., Park, J., 2014. Smart production scheduling with time-dependent and machine-dependent electricity cost by considering distributed energy resources and energy storage. *International Journal of Production Research*, 52 (13), 3922-3939.

Moon, J.-Y., Shin, K., Park, J., 2013. Optimization of production scheduling with time-dependent and machine-dependent electricity cost for industrial energy efficiency. *The International Journal of Advanced Manufacturing Technology*, 68 (1-4), 523-535.

Morales, J.M., Mínguez, R., Conejo, A.J., 2010. A methodology to generate statistically dependent wind speed scenarios. *Applied Energy*, 87 (3), 843-855.

Mouzon, G., Yildirim, M.B., Twomey, J., 2007. Operational methods for minimization of energy consumption of manufacturing equipment. *International Journal of Production Research*, 45 (18-19), 4247-4271.

Pacific Gas and Electric Company, 2016a. *Current electric rates for industrial customers (E-20 secondary firm)*. https://www.pge.com/tariffs/electric.shtml#INDUSTRIAL (retrieved July 6, 2016).

Pacific Gas and Electric Company, 2016b. *ReMAT feed-in tariff (Senate Bill 32)*. https://www.pge.com/en_US/for-our-business-partners/floating-pages/remat-feed-in-tariff/remat-feed-in-tariff.page (retrieved July 6, 2016).

Pinedo, M.L., 2012. *Scheduling: Theory, algorithms, and systems*. 4th ed., Springer, New York, NY.

Pinson, P., Madsen, H., Nielsen, H.A., Papaefthymiou, G., Klöckl, B., 2009. From probabilistic forecasts to statistical scenarios of short-term wind power production. *Wind Energy*, 12 (1), 51-62.

Sharma, A., Zhao, F., Sutherland, J.W., 2015. Econological scheduling of a manufacturing enterprise operating under a time-of-use electricity tariff. *Journal of Cleaner Production*, 108, 256-270.

Sharma, K.C., Jain, P., Bhakar, R., 2013. Wind power scenario generation and reduction in stochastic programming framework. *Electric Power Components and Systems*, 41 (3), 271-285.

Tognetti, A., Grosse-Ruyken, P.T., Wagner, S.M., 2015. Green supply chain network optimization and the trade-off between environmental and economic objectives. *International Journal of Production Economics*, 170, 385-392.

U.S. Department of Energy, 2015. *Revolution ... now: The future arrives for five clean energy technologies – 2015 update*. Washington, DC.

U.S. Energy Information Administration, 2015. *Levelized cost and levelized avoided cost of new generation resources in the annual energy outlook 2015*. Washington, DC.

Wang, Y., Li, L., 2013. Time-of-use based electricity demand response for sustainable manufacturing systems. *Energy*, 63, 233-244.

Wang, Y., Li, L., 2015. Time-of-use electricity pricing for industrial customers: A survey of U.S. utilities. *Applied Energy*, 149, 89-103.

Zhang, H., Zhao, F., Fang, K., Sutherland, J.W., 2014. Energy-conscious flow shop scheduling under time-of-use electricity tariffs. *CIRP Annals - Manufacturing Technology*, 63 (1), 37-40.

Appendix

Appendix A: Linearizations

Linearization of $\phi_t \cdot \xi_{s,t}$:

$$\theta_{s,t} \leq \xi_{s,t}, \forall s \in S, t \in \{1,2,\dots,T\} \tag{A.1}$$

$$\theta_{s,t} \geq 0, \forall s \in S, t \in \{1,2,\dots,T\} \tag{A.2}$$

$$\theta_{s,t} \leq \phi_t, \forall s \in S, t \in \{1,2,\dots,T\} \tag{A.3}$$

$$\theta_{s,t} \geq \phi_t - \left(1 - \xi_{s,t}\right), \forall s \in S, t \in \{1,2,\dots,T\} \tag{A.4}$$

Linearization of $\xi_{s,t} \cdot y_{i,j,t}$:

$$\zeta_{i,j,s,t} \leq \xi_{s,t}, \forall i \in M, j \in J, s \in S, t \in \{1,2,\dots,T\} \tag{A.5}$$

$$\zeta_{i,j,s,t} \geq 0, \forall i \in M, j \in J, s \in S, t \in \{1,2,\dots,T\} \tag{A.6}$$

$$\zeta_{i,j,s,t} \leq y_{i,j,t}, \forall i \in M, j \in J, s \in S, t \in \{1,2,\dots,T\} \tag{A.7}$$

$$\zeta_{i,j,s,t} \geq y_{i,j,t} - \left(1 - \xi_{s,t}\right), \forall i \in M, j \in J, s \in S, t \in \{1,2,\dots,T\} \tag{A.8}$$

Appendix B: Graphical representation of the relations between non-dominated, convex dominated, and dominated solutions

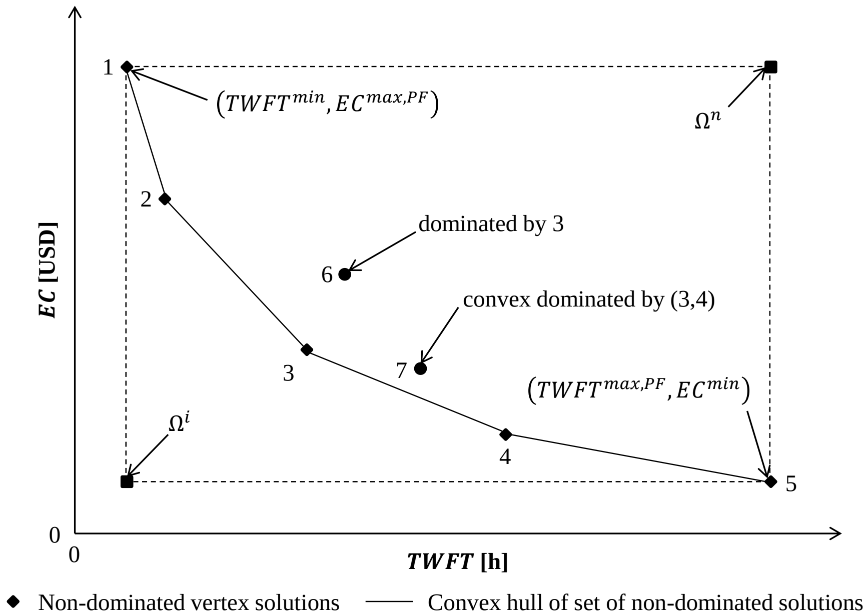


Figure B.1: Graphical representation of the relations between non-dominated, convex dominated, and dominated solutions

Appendix C: Overview of problem instances used in numerical example

Table C.1: Weights, representing the importance of the jobs, w_j with $j \in \{1,2,3\}$.

Problem instance	w_1	w_2	w_3
1	4	5	2
2	4	4	3
3	3	3	3
4	1	3	5
5	4	3	3
6	2	4	2
7	5	4	4
8	4	4	5
9	2	5	1
10	4	5	3

Table C.2: Job processing times $p_{i,j}$ with $i \in \{1,2,3\}$ and $j \in \{1,2,3\}$.

Problem instance	$p_{1,1}$	$p_{1,2}$	$p_{1,3}$	$p_{2,1}$	$p_{2,2}$	$p_{2,3}$	$p_{3,1}$	$p_{3,2}$	$p_{3,3}$
1	3	4	2	1	2	4	3	1	4
2	2	4	4	1	3	4	2	2	1
3	3	3	2	2	3	3	4	3	3
4	2	3	2	4	1	1	3	1	4
5	3	2	2	1	3	2	4	3	1
6	4	2	2	3	4	2	3	3	1
7	4	2	1	2	4	2	3	3	1
8	4	4	4	3	1	1	1	4	3
9	4	3	2	3	1	3	1	4	3
10	4	3	1	1	2	4	2	2	4

Table C.3: Processing power requirements $q_{i,j}$ with $i \in \{1,2,3\}$ and $j \in \{1,2,3\}$.

Problem instance	$q_{1,1}$	$q_{1,2}$	$q_{1,3}$	$q_{2,1}$	$q_{2,2}$	$q_{2,3}$	$q_{3,1}$	$q_{3,2}$	$q_{3,3}$
1	114	94	199	96	50	122	189	100	121
2	170	135	183	74	148	148	159	173	159
3	167	174	62	133	50	159	157	197	116
4	168	121	100	193	69	109	195	78	92
5	179	123	141	178	115	191	155	131	165
6	50	81	93	140	92	117	101	120	99
7	93	180	182	83	195	169	182	113	52
8	173	158	162	61	119	195	156	171	180
9	190	128	67	55	69	156	78	158	178
10	124	67	123	104	136	85	174	72	74

Part B Multi-stage production planning considering learning and forgetting effects

Paper 5 Governing the dynamics of multi-stage production systems subject to learning and forgetting effects: A simulation study[1]

Authors: Konstantin Biel, Christoph H. Glock

Type of publication: Journal article

Publication details: *International Journal of Production Research*, In Press.

Abstract

Managing production systems where production rates change over time due to learning and forgetting effects poses a major challenge to researchers and practitioners alike. This task becomes especially difficult if learning and forgetting effects interact across different stages in multi-stage production systems as rigid production management rules are unable to capture the dynamic character of constantly changing production rates. In a comprehensive simulation study, this paper first investigates to which extent typical key performance indicators (KPIs), such as the number of setups, in-process inventory, or cycle time, are affected by learning and forgetting effects in serial multi-stage production systems. The paper then analyzes which parameters of such production systems are the main drivers of these KPIs when learning and forgetting occur. Lastly, it evaluates how flexible production control based on Goldratt's Optimized Production Technology can maximize the benefits learning offers in such systems. The results of the paper indicate that learning and forgetting only have a minor influence on the number of setups in serial multi-stage production systems. The influence of learning and forgetting on in-process inventory and cycle time, in contrast, is significant, but ambiguous in case of in-process inventory. The proposed buffer management rules are shown to effectively counteract this ambiguity.

Keywords:
Learning; Forgetting; Production management; Multi-stage production system; Simulation

[1] Reprinted from *International Journal of Production Research*, Biel, K., Glock, C.H., Governing the dynamics of multi-stage production systems subject to learning and forgetting effects: A simulation study, In Press, with permission from Taylor & Francis.

1 Introduction

For decades, researchers and practitioners have recognized the impact of learning and forgetting effects on production rates in manufacturing processes requiring human contributions. Worker learning has been observed in numerous production environments, and it describes a scenario where the time required for producing one unit of output by an individual or a team decreases as a function of the output quantity and consequently experience (Yelle, 1979). The opposite effect, forgetting, incrementally lengthens the time a worker needs to process one unit of output, and it is assumed to be a function of production breaks and consequently lost experience (Globerson et al., 1989). As learning and forgetting change the production rates of production stages over time, researchers and practitioners face the challenge to effectively incorporate these dynamics into production planning. Neglecting worker learning and forgetting in planning production processes may lead to increased in-process inventory and higher inventory holding cost. Higher inventories may result from worker learning if learning leads to a faster production rate that ‘pushes’ inventory into the system and if the demand process is not able to keep up with the increased production rate (Bogaschewsky and Glock, 2009). At the same time, production stages with high learning coefficients may run out of input material if work stations feeding material to these production stages are unable to increase their production rates accordingly. The latter may lead to unplanned interruptions in the production process and unnecessary underutilization. Additionally, failure to recognize the effects of worker forgetting may induce production managers to consistently favor smaller production lot sizes over larger lot sizes to reduce inventory holding cost. Smunt (1987), however, showed that even minor forgetting effects can substantially increase optimal production lot sizes.

Over the last decades, the importance of considering learning and forgetting effects in the management of operations has been acknowledged by many researchers. Dolgui et al. (2012) and Pan et al. (2014), for example, took account of learning and forgetting in single machine (group) scheduling environments, while Gordon et al. (2012) presented a survey of the literature on scheduling with due date assignment considering the impact of learning on job processing. Grosse et al. (2013) investigated the effect of learning and forgetting on storage reassignment decisions in a warehousing environment, whereas Grosse and Glock (2015) studied the effect of worker learning on manual order picking processes.

Keachie and Fontana (1966), Adler and Nanda (1974), and Sule (1978) were among the first to emphasize the need to consider learning and forgetting effects in production lot sizing, which is also the topic of the paper at hand. Jaber and Bonney (1999) synthesized these and various other related approaches and provided an overview of possible problem extensions. Jaber and Khan (2010) investigated learning in a serial multi-stage production system and considered the reworking and disposal of defective items in addition. Taking multiple production stages into account enabled the authors to analytically examine how workers with different learning rates should be allocated to the different stages of the system, and how the lot size impacts the performance of the system under worker learning. Jaber and Khan (2010), however, did not take account of the transfer of learning between batches. The authors assumed instead that, after completing a batch, the experience accumulated during the production of the batch is completely lost, such that the production rates of the stages are reduced to their initial values each time a batch is finished. Glock and Jaber (2013b) extended the model of Jaber and Khan (2010) to allow for the transfer of learning between batches and additionally took account of forgetting during production breaks. Their numerical analysis indicated that serial multi-stage production systems perform best when workers along the production line learn equally fast. When learning rates differ among workers, the authors recommended assigning workers to production stages in ascending order of their learning rates. This contrasts with the findings of Jaber and Khan (2010), who advocated that workers with higher learning rates should be assigned to production stages located at the beginning of the serial production system.

The approaches of Jaber and Khan (2010) and Glock and Jaber (2013b) both statically determine batch size and number of batches. Yet, they make no recommendations on how to react to the dynamics introduced into the production system by learning and forgetting effects. That is, production planning decisions are not adjusted in response to changes in production rates, or, in other words, their production system is inflexible. To be able to dynamically react to changes in the production routine, some kind of flexibility needs to be inherent in the production system (Mandelbaum and Buzacott, 1990). Based on Beach et al. (2000), who provided a comprehensive review of flexibility in manufacturing, we categorize this sort of flexibility as internal, short-term (operational) flexibility. This type of flexibility refers to the ability of a production system to readjust between known production tasks in response to 'foreseeable events' (Carlsson, 1989; Hyun and Ahn, 1992; Upton, 1994). Such foreseeable events may be the unnecessary accumulation of inventory in the production system, which the system could counteract by an intentional adjustment of production rates. The latter aspect has been analyzed by Glock (2010, 2011), for example. In contrast to the works of Glock (2010, 2011), research that is closely related to the paper at hand assumes that such foreseeable events may materialize in varying production rates as a consequence of learning and forgetting.

The rate at which humans learn and forget depends on the specific characteristics of the individual and the task during which learning and forgetting occur. For companies, this means that workers are usually heterogeneous with respect to their learning potential, which results in different learning and forgetting rates for workers employed on the shop floor. Shafer et al. (2001) and Nembhard and Shafer (2008) took account of this fact by studying how differences in learning and forgetting rates of workers employed in production impact productivity. In extensive simulation studies, both works highlighted the importance of considering worker-individual learning and forgetting rates instead of assuming average rates for the entire workforce. In a related work, Nembhard (2001) developed a heuristic for assigning workers to tasks based on their individual learning rates, while Glock et al. (2012) proposed synchronizing production and demand by assigning additional workers to specific production stages or removing them if learning in production or demand makes changes to the workforce necessary. To ensure that inexperienced workers do not affect system throughput negatively, Bukchin and Cohen (2013) recommended appointing experienced workers to support their inexperienced colleagues during their learning period. It is evident that such production control approaches are particularly promising if workers are trained to perform different tasks. In this line of thought, Kim and Nembhard (2013) developed cross-training policies to decide which worker of a heterogeneous workforce to cross-train on which machine to make production control easier. Nembhard and Bentefouet (2014) extended this work and proposed policies for determining optimal levels of cross-training and optimal compositions of teams of multi-skilled (generalist) and specialized workers for a given production environment. Liu et al. (2016) also suggested to capitalize on the flexibility offered by multi-skilled workers and to dynamically reassign these workers to production stages over the course of the production process to balance varying production rates. Glock and Jaber (2013a) studied the challenge of bottlenecks that may shift their positions within the production system in consequence of a heterogeneous workforce where an initially slower production stage overtakes a previously faster one. To counter shifting bottlenecks, the authors proposed interrupting production from time to time and developed a heuristic method to determine not only the number of batch shipments and setups in the planning horizon, but also the position and length of interruption periods over time for the production stages to smooth the flow of material through the system. Deriving a closed-form expression to determine when and for how long to interrupt production is, however, very difficult, even for a simple two-stage production system (Glock and Jaber, 2013a). Hence, the complexity of this task increases significantly when transferred to a multi-stage production system. The complexity mainly results from numerous interactions between production stages as, for instance, the interruption of the first production stage may not only affect the production readiness of the second production stage, but ultimately of all downstream production stages to some extent. As the influence of such interactions can only be modeled analytically with the help of nonlinear differential equations, researchers and practitioners increasingly turn to simulation techniques to gain insights into the control of production processes featuring dynamic components (Kleijnen, 2008).

Finch and Luebbe (1995), for instance, studied the impact of worker learning on the performance of a serial five-stage production system in a simulation experiment. In this article, the authors resorted to the Theory of Constraints which postulates that the performance of a production system is determined by its constraints such as insufficient production capacities of production stages, for example (Goldratt and Cox, 1984). In light of this theory, Finch and Luebbe (1995) showed that increasing the production rates of production stages may have no or even a negative impact on system performance, in case these production stages do not constrain the system. Mak et al. (2014) also used simulation to investigate the impact of the production lot size and its interaction with operator competence on the performance of a U-shaped one-piece flow production system. In a numerical study, they showed that these interactions significantly influenced system performance, but also that the performance deteriorated at a certain lot size.

The paper at hand is related to the works of Finch and Luebbe (1995) and Mak et al. (2014), who simulated worker learning in a serial multi-stage production system as well. However, in contrast to both works, we also take account of forgetting effects and assess the influence of interactions among batch shipment size, initial production rates, learning as well as forgetting effects on system performance. Additionally, we distinguish between interactions at different positions in the production system and propose flexible, easy-to-use decision rules which help to react to the dynamics in the production flow caused by these interactions. Hence, the contribution of this paper is threefold. First, it proposes a generic simulation model of a serial multi-stage production system featuring learning and forgetting effects. Secondly, the paper extends prior research on the impact of interactions among production parameters including learning and forgetting on the performance of a serial multi-stage production system. Thirdly, it investigates how real-time production control can mitigate adverse effects associated with worker learning and forgetting (such as excessive inventory build-up or unplanned interruptions) based on the relevant interactions identified before. To the best of the authors' knowledge, these aspects have not been studied in the literature so far.

The remainder of the paper is organized as follows: Section 2 presents model assumptions and develops a simulation model. Section 3 carries out an extensive simulation study to analyze the impact of learning and forgetting on system performance for situations with and without flexible production control. Section 4 derives managerial implications and concludes the paper.

2 Model description

This paper focuses on a serial multi-stage production system in which a given number of items corresponding to a single production lot has to be manufactured. All items belong to the same type of product, and each item has to pass the entire production system in a given sequence. At each production stage, processed items are transported to the next production stage in equal-sized batch shipments. The structure of the production system, which is similar to the one studied by Glock and Jaber (2013b), is illustrated in Figure 1. We assume that demand is externally given, for example because an order has arrived at the system that needs to be manufactured. The first production stage is assumed to have ready and unlimited access to raw materials, and it is assumed that finished products are immediately consumed by the end customer once they leave the final production stage. After processing items at a production stage, the processed items are stored in a buffer at the backend (downstream) of the production stage, where the buffer stocks at the backend of stage i are referred to as $PostBuf_i$. When items are transported from stage i to stage $i+1$, they are stored in a buffer in front of the production stage (upstream) until they are processed, where the buffer stock in front of stage i is referred to as $PreBuf_i$. The transportation time between subsequent buffers is very small compared to the production time since the production stages are typically located in close proximity (Glock and Jaber, 2013b). Thus, the transportation time is neglected hereafter. Nevertheless, considering two separate buffer stocks between successive production stages is critical as such a setup requires batch shipments between these buffer stocks, and the size of these batch shipments heavily impacts the performance of the production system as will be shown in Section 3.

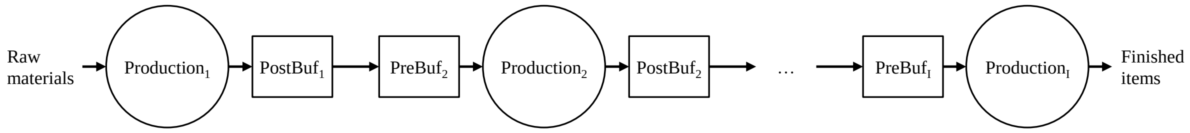


Figure 1: Example of a serial multi-stage production system.

2.1 Assumptions and notation

Apart from what has already been stated, the paper makes the following assumptions:

(1) all parameters are deterministic and constant over time;
(2) the planning horizon consists of J production runs of potentially unequal lengths, which are separated by $J-1$ interruption phases of potentially unequal lengths. A production run is thus constrained by the interruption before and after the production run (see Figure 2);
(3) before the start and upon completion of the production lot, the buffer stocks are empty. This enables the system to produce other types of products before and after manufacturing the production lot investigated here. Producing other types of products in the system may involve the same problems studied in this paper;
(4) each transition from an interruption phase to a production run requires a setup to be performed at the respective stage;
(5) one unit of input material is required to produce one unit of the finished product on each stage;
(6) forgetting cannot cause production rates to fall below initial production rates, which represent production rates for the case without experience;
(7) shortages are not allowed.

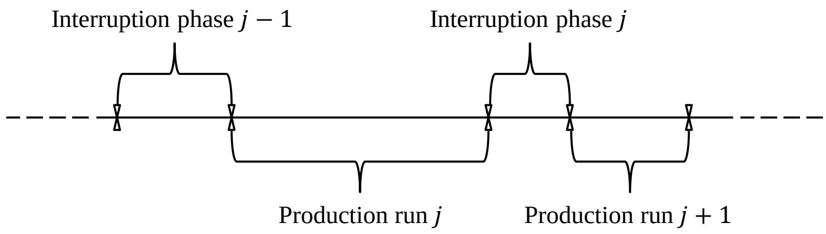


Figure 2: Separation of production runs and interruption phases.

Throughout the paper, the following terminology is used:

$\alpha_{j,i}$	Accumulated experience available on stage i at the beginning of production run j, measured in units of time
D	Total demand in the planning period
f_i	Exponent of the forgetting curve (forgetting coefficient) for stage i with $0 \leq f_i < 1$
I	Total number of stages of the production system
i	Index of production stage with $i \in \{1,2,\dots,I\}$
J	Total number of production runs
j	Index of production run with $j \in \{1,2,\dots,J\}$
k_i	Production count on stage i
l_i	Exponent of the learning curve (learning coefficient) for stage i with $0 \leq l_i < 1$
p_i	Production rate on stage i in items per unit of time
$PostBuf_i$	Buffer stock at the backend of stage i
$PreBuf_i$	Buffer stock in front of stage i
q	Size of a batch shipment that divides the production lot into partial lots, so-called batches
$T_{1,i}$	Time required to produce the first item on stage i
$\hat{T}_{1,i}$	Time for the first unit of the forgetting curve on stage i
$T_{F,j,i}$	Transition point from forgetting to learning after the jth interruption phase on stage i, measured in units of time, i.e., point in time when production is restarted after an interruption
$T_{k,i}$	Time required to produce the kth item on stage i
$\hat{T}_{f,i}$	Time required to produce one item after a production break of length $t_{f,i}$ on stage i
$T_{p,i}$	Time required to produce one item after a production period of length $t_{p,i}$ on stage i
$T_{T,j,i}$	Transition point from learning to forgetting after the jth production run on stage i, measured in units of time
$\hat{T}_{x,i}$	Time to produce one item after an interruption of x_i units
$t_{f,i}$	Time count of production interruption on stage i
$t_{f,j,i}$	Length of the interruption period on stage i after the jth production run
$t_{p,i}$	Time count of uninterrupted production on stage i
$t_{p,j,i}$	Length of the production period on stage i in the jth production run
x_i	Unit count of the forgetting curve on stage i

2.2 The learn-forget curve model

This paper assumes that learning and forgetting occur on all production stages of the system. According to Wright (1936), the time to produce the k_ith item on stage i can be calculated as

$$T_{k,i} = T_{1,i} \cdot k_i^{-l_i}. \tag{1}$$

The time to produce k_i items on stage i for the case where no prior experience is available can be calculated by integrating Equation (1) as follows:

$$t_{p,i} = \int T_{1,i} \cdot k_i^{-l_i} dk_i = \frac{T_{1,i}}{1-l_i} \cdot k_i^{1-l_i}. \tag{2}$$

Solving Equation (2) for k_i results in

$$k_i = \left(\frac{1-l_i}{T_{1,i}} \cdot t_{p,i}\right)^{\frac{1}{1-l_i}}. \tag{3}$$

Inserting Equation (3) into Equation (1) yields

$$T_{p,i} = T_{1,i} \cdot \left(\frac{1-l_i}{T_{1,i}} \cdot t_{p,i}\right)^{\frac{l_i}{l_i-1}}, \tag{4}$$

which represents the time that is required to produce one item on stage i after $t_{p,i}$ units of time of uninterrupted production. In other words, after $t_{p,i}$ units of time of uninterrupted production, stage i produces with a production rate of

$$p_i = \frac{1}{T_{p,i}}. \tag{5}$$

During production breaks, forgetting occurs. The forgetting curve of Carlson and Rowe (1976) suggests that the time for the x_ith unit of lost experience on stage i equals

$$\hat{T}_{x,i} = \hat{T}_{1,i} \cdot x_i^{f_i}, \tag{6}$$

where $\hat{T}_{1,i}$ represents the starting point of the forgetting curve. Integrating Equation (6) gives the time on the forgetting curve for x_i units of production interruption:

$$t_{f,i} = \int \hat{T}_{1,i} \cdot x_i^{f_i} dx_i = \frac{\hat{T}_{1,i}}{1+f_i} \cdot x_i^{1+f_i}. \tag{7}$$

Solving Equation (7) for x_i gives the abscissa of the move on the forgetting curve, measured in units that have not been produced:

$$x_i = \left(\frac{1+f_i}{\hat{T}_{1,i}} \cdot t_{f,i}\right)^{\frac{1}{1+f_i}}. \tag{8}$$

Inserting Equation (8) into Equation (6) yields

$$\hat{T}_{f,i} = \hat{T}_{1,i} \cdot \left(\frac{1+f_i}{\hat{T}_{1,i}} \cdot t_{f,i}\right)^{\frac{f_i}{1+f_i}}. \tag{9}$$

Equation (9) represents the time that is required to produce one item after a production break of $t_{f,i}$ units of time on stage i.

The learning and forgetting curves presented above are linked in such a way that the forgetting curve starts at the point in time when learning stops, and vice versa (see Figure 3). Consequently, the time that is required to produce the last item before the production process is interrupted corresponds to the starting point of the forgetting curve. The transition point, $T_{T,j,i}$, for $T_{p,i} = \hat{T}_{f,i}$ is consequently $T_{T,j,i} = t_{p,i} = t_{f,i}$. In other words, we can insert Equation (9) into Equation (4), which results in

$$\hat{T}_{1,i} = \left(T_{1,i} \cdot \left(\frac{1-l_i}{T_{1,i}} \cdot T_{T,j,i}\right)^{\frac{l_i}{l_i-1}} \cdot \left((1+f_i) \cdot T_{T,j,i}\right)^{\frac{-f_i}{1+f_i}}\right)^{1+f_i}. \tag{10}$$

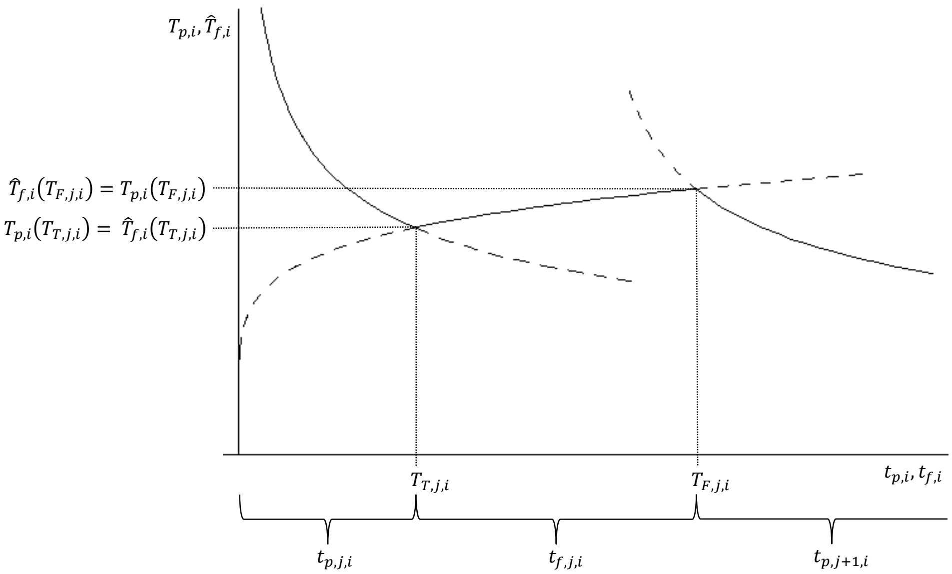


Figure 3: Learning and forgetting in production run j.

Inserting Equation (10) into Equation (9) gives the time required to produce one item after a production break of length $t_{f,i}$. When an interrupted production process is restarted at point $T_{F,j,i}$, some previously accumulated experience may be available in the system (see Figure 3). As at the transition point from forgetting to learning $\hat{T}_{f,i} = T_{p,i}$ holds, the amount of experience, $\alpha_{j,i}$, measured in production time that the production stage remembers, can be calculated by inserting Equation (9) into Equation (4) with $t_{p,i} = \alpha_{j,i}$ and $t_{f,i} = T_{F,j-1,i}$:

$$T_{1,i} \cdot \left(\frac{1-l_i}{T_{1,i}} \cdot \alpha_{j,i}\right)^{\frac{l_i}{l_i-1}} = \hat{T}_{1,i} \cdot \left(\frac{1+f_i}{\hat{T}_{1,i}} \cdot T_{F,j-1,i}\right)^{\frac{f_i}{1+f_i}}. \tag{11}$$

Solving Equation (11) for $\alpha_{j,i}$ results in

$$\alpha_{j,i} = \left(\frac{\hat{T}_{1,i}}{T_{1,i}} \cdot \left(\frac{1+f_i}{\hat{T}_{1,i}} \cdot T_{F,j-1,i}\right)^{\frac{f_i}{1+f_i}}\right)^{\frac{l_i-1}{l_i}} \cdot \frac{T_{1,i}}{1-l_i}. \tag{12}$$

2.3 The simulation model

The production system studied in this paper was modeled using the Java-based simulation software AnyLogic 7.3.6 Professional. The system (see example in Figure 1) consists of the first production stage (with no buffer stock in front of the stage), $I-2$ intermediate production stages (with buffer stocks in front and at the back of the stages), and a final production stage (with no buffer stock at the back of the stage). As long as the total demand, D, has not completely been satisfied by the final stage, an action chart is run through in every period. However, a production stage can only process items in a given period if enough input material is available in the upstream buffer, $PreBuf_i$, such that production can be executed with the current production rate. Thus, if the current production rate equals p_i items per unit of time, $PreBuf_i$ has to contain a minimum quantity of p_i items to enable the production stage to produce. If $PreBuf_i$ contains less than p_i items, production is not possible.

3 Simulation study

After integrating learning and forgetting effects into the serial multi-stage production system, we conducted a comprehensive simulation study. The primary goals of this study are (I) to assess the influence of learning and forgetting on system performance (Simulation Experiment 1) and (II) to investigate how internal, operational flexibility inherent in the production system may be used to control this influence (Simulation Experiment 2).

3.1 Experimental design

According to Kleijnen et al. (2005), simulation experiments mainly aim at understanding the behavior of real-life (production) systems over time and to identify and compare control policies that perform well for various input parameter scenarios. As we intend to evaluate the performance of a production system that produces a single lot consisting of a fixed number of items, D, which is split into equal-sized batches, we face a terminating system. Before and after processing this lot, the production system may process other products. Hence, the simulation model starts with empty buffers and no warm-up period is required (Kleijnen et al., 2005). As in Longo (2010), the debugging technique described in Dunn (1987) was used to verify the simulation model. Various combinations of different input parameters were fed into the model. Errors that occurred during the simulation run were analyzed and corrected.

To test whether learning and forgetting have an impact on the performance of the system and to subsequently identify key drivers of system performance (Simulation Experiment 1), we generated 500 input parameter scenarios using Latin Hypercube Sampling (LHS) (Loeppky et al., 2012) (see Section 3.2). For the test of the production control policies (Simulation Experiment 2), we used a Latin Hypercube Design featuring only 450 input parameter scenarios and considered only three fixed values of the batch shipment size (large, medium, small; see Section 3.3). LHS is associated with good space-filling properties, i.e., it is particularly suitable for 'exploring unknown, but potentially complicated response surfaces[2] with many quantitative factors' (Kleijnen et al., 2005). The corresponding input parameter values were sampled from the following intervals that were derived from the numerical study in Glock and Jaber (2013a):

- Batch shipment size q: [250, 2500]
- Initial production rates $T_{1,i}^{-1}$: [0.26, 0.53]
- Learning coefficients l_i: [0.03, 0.25]
- Forgetting coefficients f_i: [0.01, 0.24]

In all input parameter scenarios, a lot of $D = 10000$ items had to be processed by the production system. The performance of the production system was evaluated by tracking different key performance indicators (KPIs) for all stages, namely the total number of setups, the cumulative in-process inventory across all stages, and the cycle time, which all should be minimized when considered separately. These KPIs have often been used in simulation experiments of production systems (e.g., Mak et al., 2014). In contrast to Mak et al. (2014), we recorded cumulative values of the KPIs instead of average values. This is due to the fact that average values are only meaningful in case the production system reaches a steady state. However, as terminating systems frequently do not reach a steady state at all, average values have little meaning in such systems (Pidd, 2004). Hence, we also refrained from recording utilization rates of the production stages. Capitalizing on the generic structure of the simulation model, we ran the simulation experiments for a three-, a five-, and a seven-stage production system to also test whether the number of stages has an impact on the behavior of the KPIs.

The simulation study was carried out using an Intel® CoreTM i7-4770 CPU @ 3.40 GHz with 16 GB RAM and the Windows 10 operating system. In Simulation Experiment 1, the average CPU time for one simulation run featuring 500 parameter input scenarios was 7.1 (17.6, 40.6) seconds for the three-stage (five-stage, seven-stage) production system without considering learning and forgetting effects and 6.6 (83.3, 140.9) seconds when learning and forgetting effects were considered. In Simulation Experiment 2, the average CPU time for one simulation run featuring 450 parameter input scenarios was 5.5 (66.9, 104.5) seconds for the three-stage (five-stage, seven-stage) production system without production control and 9.2 (88.7, 148.8) seconds when production was controlled using different policies

3.2 Impact of learning and forgetting on system performance and identification of its drivers

Before analyzing the key drivers of the KPIs in detail, we investigated how worker learning and forgetting affect the KPIs to identify those KPIs that required further inspection. To this end, we compared the means of the total number of setups, the cumulative in-process inventory, and the cycle time under learning and forgetting, $\overline{KPI}^{w/\_lf}$, to the corresponding means for a situation where workers neither learned nor forgot, $\overline{KPI}^{w/o\_lf}$.[3] The Shapiro-Wilk normality test of the differences between the values with and without learning and forgetting of all three relevant KPIs indicated nonnormality of errors in the models. For this reason, we used the Wilcoxon signed-rank test, a non-parametric test, to find out whether learning and forgetting impacted the KPIs (Field et al., 2012). From Table 1, we can infer that, on average, the number of setups is not significantly affected by worker learning and forgetting for any of the considered production systems. Furthermore, the effect sizes of Pearson's r imply a

[2] A response surface is a model that approximates the relationship between input parameters and performance indicators, e.g., a regression equation (Kleijnen et al., 2005).
[3] Means were calculated across the 500 input parameter scenarios.

small effect (Cohen, 1992). In contrast, the influence of learning and forgetting on cumulative inventory and cycle time was found to significantly influence, on average, both the cumulative inventory and the cycle time (at the 1 percent level) for all investigated production systems. Additionally, for both KPIs, the effect sizes indicate a large influence of learning and forgetting, whereas the effect on cumulative inventory appears to slightly increase as the stages of the production system increase.

Table 1: Impact of learning and forgetting on the KPIs of the three-, five-, and seven-stage production system.

KPI	**Number of stages**	**Without learning and forgetting**		**With learning and forgetting**		**H$_0$: $\overline{KPI^{w/\_lf}} - \overline{KPI^{w/o\_lf}} = 0$**	
		Mean	Std. dev.	Mean	Std. dev.	p-value	Effect size r
Number of setups	3	12.80	10.86	12.99	11.85	0.6961	0.01
	5	27.41	21.23	27.06	20.77	0.8676	-0.01
	7	43.62	34.79	44.09	35.13	0.5964	0.02
Cumulative inventory	3	$9.63 \cdot 10^{7}$	$4.13 \cdot 10^{7}$	$5.22 \cdot 10^{7}$	$2.82 \cdot 10^{7}$	<0.0001	-0.56
	5	$1.90 \cdot 10^{8}$	$7.82 \cdot 10^{7}$	$1.01 \cdot 10^{8}$	$4.51 \cdot 10^{7}$	<0.0001	-0.60
	7	$3.79 \cdot 10^{8}$	$1.54 \cdot 10^{8}$	$1.50 \cdot 10^{8}$	$6.42 \cdot 10^{7}$	<0.0001	-0.61
Cycle time	3	37748.66	5925.64	16444.74	5478.09	<0.0001	-0.61
	5	47610.43	8381.28	21724.83	6232.53	<0.0001	-0.61
	7	74082.22	16262.90	26754.36	7529.52	<0.0001	-0.61

The strong increase of all three KPIs, both for the cases with and without learning and forgetting, that becomes evident from Table 1 as the system moves from three to seven production stages is simply due to the fact that the items to be processed have to pass through more production stages as the length of the production system increases. Thus, it is obvious that processing a lot in a five-stage production system, on average, requires more setups, that it leads to higher cumulative inventories, and that it takes longer cycle times than processing the same lot in a three-stage production system. However, comparing the rates at which the KPIs increase when transitioning from a three- to a five- to a seven-stage production system reveals interesting insights both for the cases with and without learning and forgetting: The average number of setups feature similar growth rates when transitioning from a three- to a five-stage production system without learning and forgetting (114 percent) and with learning and forgetting (108 percent) and when transitioning from a five- to a seven-stage production system (without learning and forgetting: 59 percent; with learning and forgetting: 63 percent). With respect to the average cumulative inventory and the average cycle times, this behavior was only found for the transition from a three- to a five-stage production system without learning and forgetting (cumulative inventory: 97 percent; cycle time: 26 percent) and with learning and forgetting effects (cumulative inventory: 93 percent; cycle time: 32 percent). In contrast, the transition from a five- to a seven-stage production system resulted in a 99 percent increase in the average cumulative inventory and a 56 percent increase in the average cycle time without learning and forgetting, but only in a 49 percent increase in the average cumulative inventory and a 23 percent increase in the average cycle time with learning and forgetting. This indicates that adding production stages to a longer production system appears to impact the cumulative inventory and the cycle time considerably less with learning and forgetting than without learning and forgetting as compared to adding production stages to a shorter production system.

In the subsequent investigation of the drivers of the KPIs, we exclusively focused on the cumulative inventory. We took no account of the number of setups as this KPI was not significantly affected by learning and forgetting. In addition, we also excluded drivers of the cycle time from a detailed analysis as learning and forgetting reduced the cycle time in all input parameter scenarios studied, and hence no ambiguous influence on system performance that required further investigation could be identified. In contrast, the impact of learning and forgetting on cumulative inventory was ambiguous. To identify the drivers of the cumulative inventory (dependent variable), we ran a multiple linear regression for each of the three production systems considered with the size of the batch shipments, the initial production rates, and the learning and forgetting coefficients as independent variables. Furthermore, we included all two-way interactions of these independent variables into the models. After centering the input data, the variance inflation factors (VIFs) indicated that no multicollinearity between the independent variables was present (see Table 2). The Durbin-Watson statistics and the corresponding p-values revealed that the residual terms of any two observations were almost certainly independent (see Table 2). From the plots of the residuals, we could infer that the assumptions of homoscedasticity and normally distributed errors were also met. Hence, the assumptions underlying a multiple linear regression are met and its results can be generalized beyond the input parameter scenarios considered (Field et al., 2012).

Table 2: Statistics of the multiple linear regressions of the three-, five-, and seven-stage production systems.

Number of stages	**Variance Inflation Factors**		**Durbine-Watson**		**Adjusted R^2**
	Mean	Maximum	statistic	p-value	
3	1.15	1.20	1.98	0.912	0.8605
5	1.42	1.65	2.09	0.316	0.8659
7	2.10	2.51	1.91	0.274	0.8732

As the regression analysis found that a large number of independent variables has a significant impact on the cumulative inventory, we solely discuss those independent variables whose impact was significant at the 1 percent level (see Table 3). An overview of the results of the regression analysis with a focus on independent variables with a significant influence on cumulative inventory at the 5 percent level can be found in Appendix A for the three-stage production system, in Appendix B for the five-stage production system, and in Appendix C for the seven-stage production system. From the values of the adjusted R^2 in Table 2, we can infer that 86 to 87 percent of the variance in the cumulative inventory can be explained by the independent variables considered for the investigated production systems. The standardized coefficients in Table 3 indicate that the batch shipment size and the learning coefficients of the production stages impact the cumulative inventory the most. The influence of the batch shipment size is positive, meaning that a larger batch shipment size leads to more cumulative inventory, given the other parameter values are at their mean. With regard to the influence of the learning coefficients, it is striking that the learning coefficient of the first production stage positively affects the cumulative inventory, while the learning coefficients of all other production stages negatively impact the cumulative inventory. Furthermore, it is of interest that the initial production rates $T_{1,i}^{-1}$ with $i \in \{2, \dots, I\}$ also have a negative influence on the cumulative inventory, while the corresponding interactions $T_{1,i}^{-1}: l_i$ with $i \in \{2, \dots, I\}$ are positive. Focusing on the five-stage production system, Figure 4a) indicates that the advantage of a high $T_{1,2}^{-1}$ over a low $T_{1,2}^{-1}$ resulting in less cumulative inventory diminishes when l_2 is fairly high. The interactions of the size of the batch shipment and the learning effects are all negative. This is especially noteworthy in terms of $q{:}\,l_1$ as q and l_1 individually both impact the cumulative inventory positively. As a consequence, the increase of q mitigates the positive impact of l_1 on the cumulative inventory (see Figure 4b)).

Table 3: Standardized regression coefficients of the independent variables with a significant influence on cumulative inventory (at the 1 percent level) of the three-, five-, and seven-stage production system.

Three-stage production system		**Five-stage production system**		**Seven-stage production system**	
Independent variable	Standardized coefficient	Independent variable	Standardized coefficient	Independent variable	Standardized coefficient
l_3	-0.46	q	0.67	q	0.74
l_2	-0.44	l_2	-0.30	l_2	-0.18
q	0.43	l_5	-0.26	l_7	-0.14
l_1	0.25	l_3	-0.25	l_4	-0.14
$T_{1,3}^{-1}$	-0.23	l_4	-0.20	l_6	-0.13
$T_{1,2}^{-1}$	-0.21	l_1	0.17	l_3	-0.13
$T_{1,2}^{-1}{:}\,l_2$	0.20	$T_{1,2}^{-1}{:}\,l_2$	0.16	l_5	-0.13
$T_{1,3}^{-1}{:}\,l_3$	0.17	$T_{1,3}^{-1}$	-0.16	$T_{1,2}^{-1}{:}\,l_2$	0.12
$l_1{:}\,l_2$	-0.16	$T_{1,2}^{-1}$	-0.14	$T_{1,4}^{-1}$	-0.12
$l_2{:}\,l_3$	-0.14	$T_{1,4}^{-1}{:}\,l_4$	0.12	$q{:}\,l_1$	-0.11
$l_1{:}\,l_3$	-0.11	$T_{1,3}^{-1}{:}\,l_3$	0.12	l_1	0.11
$T_{1,1}^{-1}{:}\,l_1$	0.10	$T_{1,5}^{-1}$	-0.11	$q{:}\,l_2$	-0.11
$T_{1,2}^{-1}{:}\,l_1$	-0.09	$q{:}\,l_1$	-0.11	$T_{1,6}^{-1}{:}\,l_6$	0.10
$q{:}\,l_1$	-0.08	$T_{1,5}^{-1}{:}\,l_5$	0.10	$T_{1,3}^{-1}$	-0.10
$l_3{:}\,f_1$	-0.07	$l_2{:}\,l_3$	-0.08	$T_{1,2}^{-1}$	-0.10
$q{:}\,l_2$	-0.07	$l_3{:}\,l_4$	-0.07	$q{:}\,l_4$	-0.10
$T_{1,3}^{-1}{:}\,l_2$	-0.06	$T_{1,3}^{-1}{:}\,l_1$	-0.07	$T_{1,3}^{-1}{:}\,l_3$	0.10
		$l_1{:}\,l_2$	-0.07	$T_{1,5}^{-1}{:}\,l_5$	0.09
		$q{:}\,l_2$	-0.07	$T_{1,7}^{-1}{:}\,l_7$	0.08
		$T_{1,4}^{-1}$	-0.07	$T_{1,5}^{-1}$	-0.08
		$T_{1,1}^{-1}{:}\,l_3$	-0.06	$q{:}\,l_6$	-0.08
		$T_{1,4}^{-1}{:}\,T_{1,5}^{-1}$	-0.06	$l_1{:}\,l_2$	-0.07
		$q{:}\,l_3$	-0.06	$q{:}\,l_3$	-0.07
		$q{:}\,l_5$	-0.06		
		$l_3{:}\,l_5$	-0.06		
		$l_1{:}\,l_4$	-0.05		
		$T_{1,4}^{-1}{:}\,l_1$	-0.05		

Another important interaction is inherent in the learning coefficients of successive production stages, $l_i{:}\,l_{i+1}$, which are all negative with respect to the cumulative inventory. Particularly, $l_1{:}\,l_2$ is of interest as it indicates that l_2 can mitigate the positive impact of l_1 on the cumulative inventory (see Figure 4c)). The same holds true for $l_1\colon l_3$ in case of the three-stage production system and $l_1\colon l_4$ in case of the five-stage production system. Aside from these two interactions, the regression analysis revealed a few important interaction effects that span several stages such as $T_{1,3}^{-1}\colon l_1$, $T_{1,1}^{-1}\colon l_3$, $T_{1,4}^{-1}\colon l_1$, as well as $l_3\colon l_5$ in case of the five-stage production system. Among these, the first three are of particular interest as again the positive influence of the production parameters associated with the first production stage is mitigated by $T_{1,3}^{-1}$, l_3, and $T_{1,4}^{-1}$, respectively. Figure 4d) shows that, in case of a low l_3, a low $T_{1,1}^{-1}$ leads to a lower cumulative inventory than a high $T_{1,1}^{-1}$. However, this relation is reversed in case of a high l_3. In case of a medium l_3, $T_{1,1}^{-1}$ basically has no effect at all on the cumulative inventory, provided that all other input parameters are at their average values. Lastly, it is noteworthy that, except for the interaction effect $l_3\colon f_1$ in case of the three-stage production system, forgetting does not exercise a significant effect on the cumulative inventory (at the 1 percent level), neither individually nor as part of an interaction.

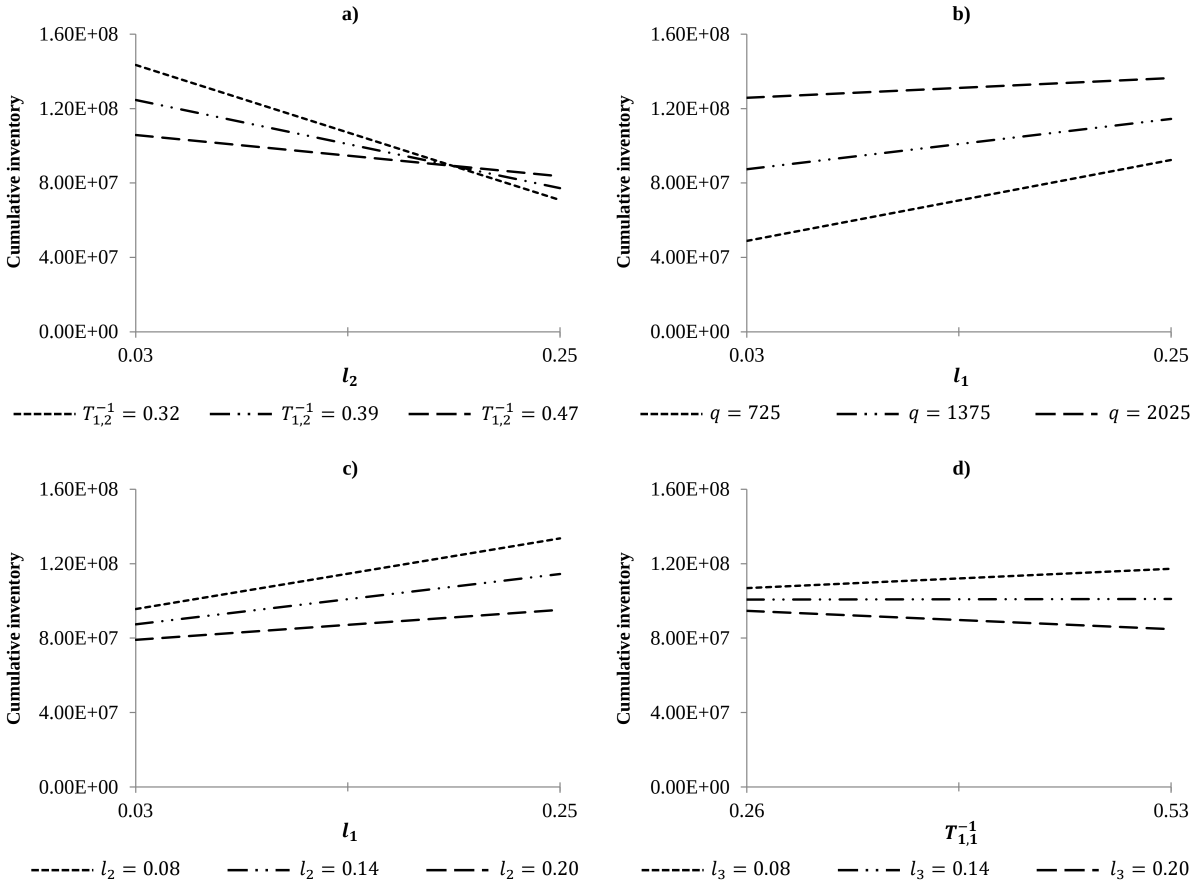


Figure 4: Visualization of the impact of selected two-way interactions on the cumulative inventory of the five-stage production system[4].

[4] Figures 4a)–d) each features three lines. Each of these lines results from setting one of the two independent variables of a two-way interaction to a fixed value and varying the other one on the x-axis. The first independent variable was fixed at three values, namely its mean value less one standard deviation, its mean value, and its mean value plus one standard deviation (with reference to the 450 input parameter values).

Overall, the regression analysis illustrated that several input parameters drive the cumulative inventory, and that it is affected by interactions of these parameters in a considerable and sometimes unexpected way. Furthermore, these interactions are not necessarily confined to parameters of the same stage or successive stages. Instead, there clearly are interdependencies spanning multiple stages and thus interdependencies that can barely be modeled in an analytical model. Across the three production systems investigated, the results are fairly homogeneous, indicating that the obtained results may be applicable to larger serial multi-stage production systems as well.

3.3 Production control through buffer management

From Section 3.2, we can infer that learning and forgetting do not significantly affect the number of setups, but always reduce the cycle time, while their influence on the cumulative inventory heavily depends on the relations of the input parameters. For this reason, we propose production control policies that particularly aim to control the cumulative inventory in the following.

3.3.1 Definition of buffer management rules

As the regression analysis indicated that certain input parameter combinations lead to unnecessary inventory build-up in the system, it may make sense to interrupt the production process from time to time. To manage the stages of the production system, a set of buffer management rules was developed that determine the points in time when a production stage is interrupted and restarted. More specifically, these rules define how to manage buffer stocks in front of the production stages, and when to initiate production at upstream stages to refill this buffer stock. This way, the buffer management rules flexibly adjust the production routine to altering properties of the production system (here: varying production rates in consequence of learning and forgetting). Hence, the buffer management rules require internal and short-term (operational) flexibility, introduced in Section 1, to be inherent in the production system such that production can be interrupted and restarted at virtually every point in time.

Typically, excessive inventory build-up in serial production systems occurs in front of the bottleneck stage as upstream stages produce at increasingly higher production rates than the bottleneck stages (Goldratt and Cox, 1984). There are different definitions of bottleneck stages in the literature. However, in the context of the current production system, the inventory definition seems appropriate. This definition identifies production stage i that features the highest $PreBuf_i$ in the production system as the bottleneck stage (Lawrence and Buss, 1995; Kuo et al., 1996; Zhai et al. 2011). The buffer management rules were formulated with reference to the drum-buffer-rope approach of Goldratt and Cox (1984), whose objectives are to ensure that (I) the bottleneck stage never runs out of material and that it is consequently always occupied, and that (II) upstream stages of the bottleneck are managed in such a way that no excessive inventory build-up occurs in front of the bottleneck stage. These objectives require that the bottleneck stage and its corresponding buffer stocks are closely monitored and that deviations from the planned production schedule are quickly responded to (Schragenheim and Ronen, 1991). In this paper, as suggested by Ronen and Starr (1990), active buffer management is only applied to the buffer stock directly in front of the bottleneck stage, $PreBuf_{BN}$ with $BN \coloneqq \{i \mid PreBuf_i \geq PreBuf_h, \forall\, h \neq i,\ i, h \in \{2,3,\dots,I\}\}$. If buffer management requires $PreBuf_{BN}$ to contain x items, for example, upstream production stages will produce as long as the stock level in $PreBuf_{BN}$ is lower than x items. However, as soon as the inventory level of $PreBuf_{BN}$ reaches x, upstream production stages are interrupted to avoid that excessive in-process inventory is built up. The buffer management rules monitor the inventory level in $PreBuf_{BN}$ and restart production once the inventory level falls below a specified limit. Through this constant monitoring, the buffer management rules enable the production system to flexibly react to the dynamics introduced into the system by learning and forgetting. In this paper, four different buffer sizes are investigated, where the size of a buffer was varied by an increment of $0.5 \cdot q$, resulting in four buffer management rules:

- Rule r#1: Interrupt all production stages i with $i \in \{1, \ldots, BN-1\}$ if $PreBuf_{BN} \geq 0.5 \cdot q$.
- Rule r#2: Interrupt all production stages i with $i \in \{1, \ldots, BN-1\}$ if $PreBuf_{BN} \geq 1.0 \cdot q$.
- Rule r#3: Interrupt all production stages i with $i \in \{1, \ldots, BN-1\}$ if $PreBuf_{BN} \geq 1.5 \cdot q$.
- Rule r#4: Interrupt all production stages i with $i \in \{1, \ldots, BN-1\}$ if $PreBuf_{BN} \geq 2.0 \cdot q$.

These buffer management rules were further fine-tuned to prevent constant bottleneck changes that may paralyze the production process through permanent production interruptions at alternating production stages. As explained in Section 3.1, we tested the buffer management rules only for three fixed batch shipment sizes (large, $q = 2000$, medium, $q = 1250$, and small, $q = 500$). This way, the focus of the analysis of the buffer management rules is on those input parameters that primarily influence the production rates and whose interactions the buffer management rules are intended to control. The 450 input parameter scenarios of $T_{1,i}^{-1}$, l_i, and f_i were constructed with LHS (see Section 3.1). To evaluate the performance of the buffer management rules, we compared them to the scenarios where large, medium, and small batch shipments are used and production is not controlled (*baselines b#1*, *b#2*, *b#3*).

3.3.2 Impact of the buffer management rules on the KPIs

Table 4 summarizes means and standard deviations of the KPIs associated with the baseline scenarios as well as the buffer management rules for the five-stage production system. The corresponding results for the three- and the seven-stage production system can be found in Appendices A and C, respectively. On average, the introduction of buffer management leads to an increase in the mean number of setups for all batch shipments sizes, with the increase being larger for smaller batch shipments. The difference to the baseline scenarios decreases with an increase in the buffer size as larger buffer sizes are less restrictive and lead to fewer interruptions of the production process. The means of the cumulative inventories are also clearly affected by the buffer management rules, and their impact increases as batch sizes are reduced. Except for two simulation setups, the average cumulative inventory levels are the lower, the stricter the buffer management rules are. In case of large batch shipments in the five- and the seven-stage production system, there seems to be an optimal buffer size close to $PreBuf_{BN} = 1.0 \cdot q$ (r#2) with regard to the cumulative inventory. The cycle times, in contrast, seem to be only marginally impacted by buffer management. For all three KPIs, the results indicate that the impact of the buffer management rules is slightly larger for shorter production systems. The standard deviations of the KPIs are partially quite substantial. However, multiple linear regressions on every KPI for the baselines and each buffer management rule revealed that the variances of the KPIs can largely be explained by the variance contained in the input scenarios.

Table 4: Descriptive analysis of the impact of buffer management rules on the KPIs of the five-stage production system.

		Number of setups		**Cumulative inventory**		**Cycle time**	
		Mean	Std. dev.	Mean	Std. dev.	Mean	Std. dev.
Large	b#1	14.15	3.71	1.31·108	3.86·107	24340.39	6242.59
batch	r#1	22.69	3.57	1.23·108	3.31·107	26284.56	6938.87
shipments	r#2	17.98	3.41	1.22·108	3.30·107	24347.14	6243.09
	r#3	16.73	3.23	1.26·108	3.46·107	24340.39	6242.59
	r#4	15.45	3.46	1.29·108	3.62·107	24340.39	6242.59
Medium	b#2	21.90	6.61	9.74·107	3.32·107	21512.91	5605.08
batch	r#1	36.66	6.55	8.40·107	2.43·107	23487.63	6353.78
shipments	r#2	29.00	6.42	8.43·107	2.60·107	21523.35	5603.07
	r#3	27.96	5.78	8.71·107	2.62·107	21512.91	5605.08
	r#4	26.22	5.67	9.02·107	2.75·107	21512.91	5605.08
Small	b#3	53.49	18.48	6.11·107	3.00·107	18585.13	5029.47
batch	r#1	94.01	20.23	3.95·107	1.56·107	20717.14	5891.86
shipments	r#2	74.46	19.33	4.27·107	2.07·107	18595.31	5024.04
	r#3	73.99	19.18	4.39·107	2.05·107	18585.50	5028.99
	r#4	71.97	18.07	4.55·107	2.04·107	18585.13	5029.47

These first descriptive results are generally supported by Figure 5, which shows the boxplots of the differences in the KPIs resulting from the introduction of the buffer management rules in the five-stage production system. The corresponding boxplots for the three- and the seven-stage production system can be found in Appendices A and C, respectively. Overall, it is remarkable that the range of the middle 50 percent of the difference values increases and diverges from zero with smaller batch shipment sizes. This indicates a growing, but less predictable impact of the buffer management rules at smaller batch shipments sizes. The only exception to this observation is the pairwise comparison of the buffer management rules for the cumulative inventory. With regard to the comparison of buffer management rules with the baselines, it is of interest that in almost all cases the range of the middle 50 percent of the difference values decreases and approaches zero, the less restrictive the buffer management rules are, i.e., the higher the allowed inventory levels in $PreBuf_{BN}$ are. From this observation, we can infer that the influence of the buffer management rules decreases, but becomes more predictable in case the rules are less strict. This becomes also evident from the pairwise comparisons of the buffer management rules, where the range of the middle 50 percent of the difference values increases and diverges from zero with increasing difference in $PreBuf_{BN}$ at which production is interrupted. Overall, r#2 seems to offer a good tradeoff between the size of the impact and the predictability as it leads to considerably fewer setups and a shorter cycle time than r#1 at almost the same level of cumulative inventory. r#3 and r#4 offer further (marginal) reductions in the number of setups and the cycle time at the expense of an increase in the cumulative inventory. Hence, depending on the KPIs most relevant in a given application, the buffer management rules should be determined.

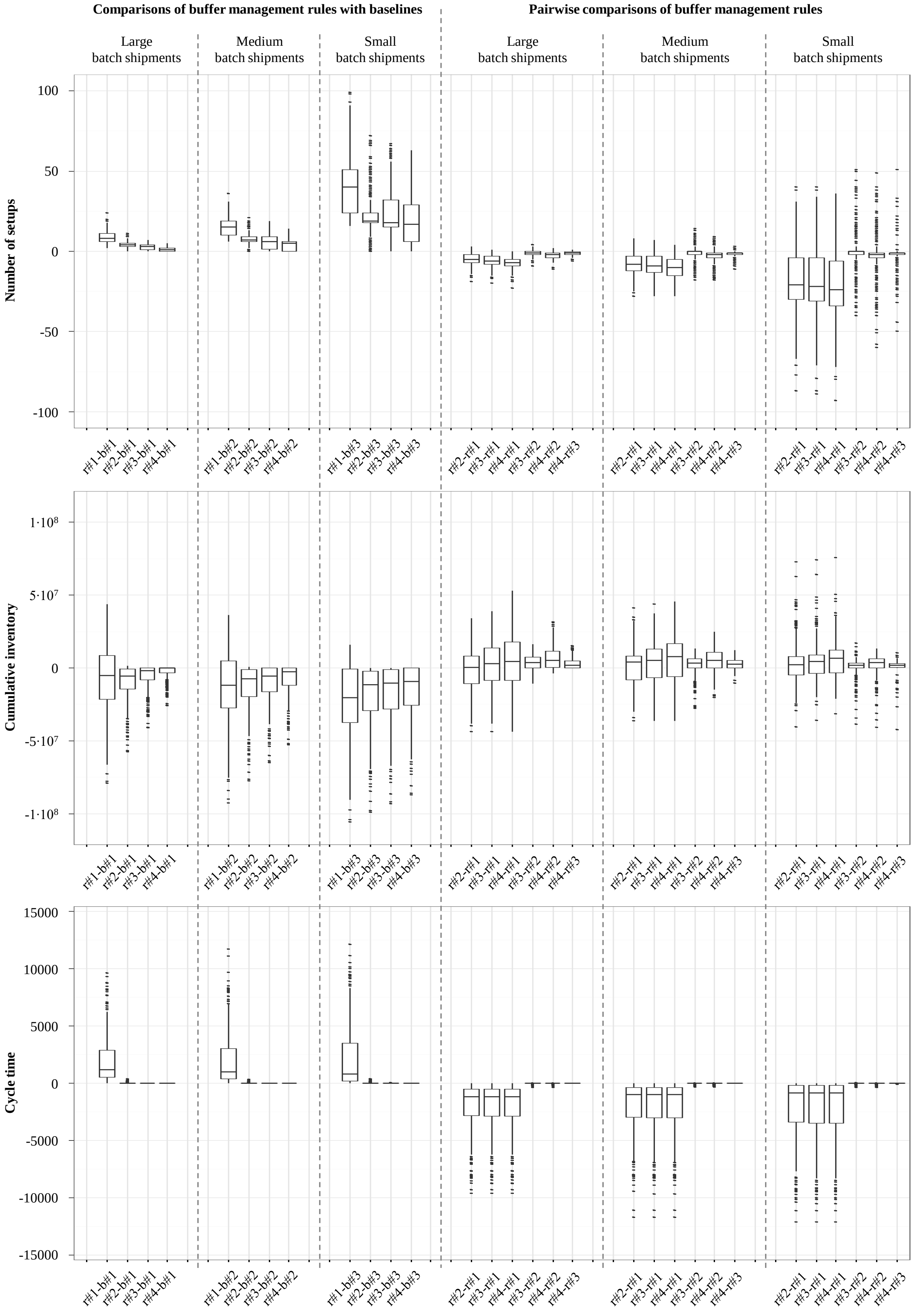


Figure 5: Boxplots of the differences in the KPIs resulting from the buffer management rules for the five-stage production system.

To assess whether the differences of the means associated with the baseline and the buffer management rules identified in the descriptive analysis are significant, we ran nine Friedman tests for each of the three production systems considered, i.e., one for each KPI for large, medium, and small batch shipments (Friedman, 1937). We chose the non-parametric Friedman test as several of the assumptions of the one-way repeated measures analysis of variance were violated. The results indicate that all KPIs are significantly impacted (at the 5 percent level) by applying different buffer management rules. To find out which means associated with the different buffer management rules significantly differ from each other, we ran pairwise Wilcoxon signed-rank tests (Field et al., 2012). Based on the Wilcoxon signed-rank test statistics, we estimated the standardized effect sizes of the pairwise comparisons according to Rosenthal (1991) to assess the extent to which the different buffer management rules influence the KPIs.

Figure 6 illustrates the obtained significance and effect size statistics. The comparison of the buffer management rules with the baselines (left part of Figure 6) revealed that all buffer management rules lead to a significant increase in the number of setups for all batch shipment sizes (at the 5 percent level) independent of the length of the production system. However, as already indicated by Figure 5, the effect sizes decrease with less restrictive buffer management rules, i.e., with rules that allow higher inventory levels in $PreBuf_{BN}$. Moreover, the effect sizes converge across the buffer management rules with decreasing q. However, bearing the boxplots of Figure 5 in mind, it is surprising that the effect sizes are almost equal, independent of the size of the batch shipment, instead of being larger for smaller batch shipments. The cause of this dissent may be the Wilcoxon signed-rank test itself that neglects difference values if they are equal to zero. In case of the use of large batch shipments in the five-stage production system for instance, while the differences in the number of setups of r#1 and b#1 were always different from zero, the corresponding differences between r#2 (r#3, r#4) and b#1 equalled zero in 38 (101, 176) out of 450 input parameter scenarios. Consequently, these difference values were not included in the calculation of the significance and effect size statistics. Hence, although the Wilcoxon signed-rank test seems to be the most appropriate test given the distribution of the dependent variables, the results may be partially inflated (Field et al., 2012). For this reason, it is crucial to evaluate the significance and effect size statistics featured in Figure 6 under consideration of the corresponding boxplots in Figure 5.

With respect to the cumulative inventory, all buffer management rules lead to a significant reduction in cumulative inventory (at the 5 percent level). While the descriptive analysis suggested that r#1 leads to the lowest cumulative inventory levels in almost all simulation setups, the standardized effect sizes indicate that r#2 to r#4 have a larger impact in terms of reducing cumulative inventory than r#1. On the one hand, this deviation may stem from the large variability of the influence of r#1, also identified in Figure 5. On the other hand, it may be due to the inflation of the test statistics explained above. With regard to the cycle times, only r#1 and r#2 have a significant positive impact (at the 5 percent level), with the effect of r#1 being substantially larger than the effect of r#2. In contrast, r#3 and r#4 do not significantly affect the cycle times (at the 5 percent level).

The results of the pairwise comparisons of the buffer management rules (right part of Figure 6) are mixed. For the number of setups, all buffer management rules lead to significantly different results (at the 5 percent level) for all batch shipment sizes, except for r#2 and r#3 in case of small batch shipments in the three-stage production system. For the cumulative inventory, all buffer management rules lead to significantly different results (at the 5 percent level) for all batch shipment sizes, except for r#1 and r#2 in case of large and medium batch shipments in the five-stage production system and in case of large batch shipments in the seven-stage production system. Again, a large number of equal values in the pairwise comparisons of r#2, r#3, and r#4 may have caused the corresponding, unexpectedly large effect sizes. For the cycle times, all buffer management rules lead to significantly different results (at the 5 percent level) for all batch shipment sizes, except for r#3 and r#4. However, only the effect sizes associated with the comparisons of r#1 with r#2, r#3, and r#4 are substantial and support the findings of the related boxplots in Figure 5. Overall, the results are fairly homogeneous across the different production systems, indicating that the results obtained may be applicable to larger serial multi-stage production systems as well.

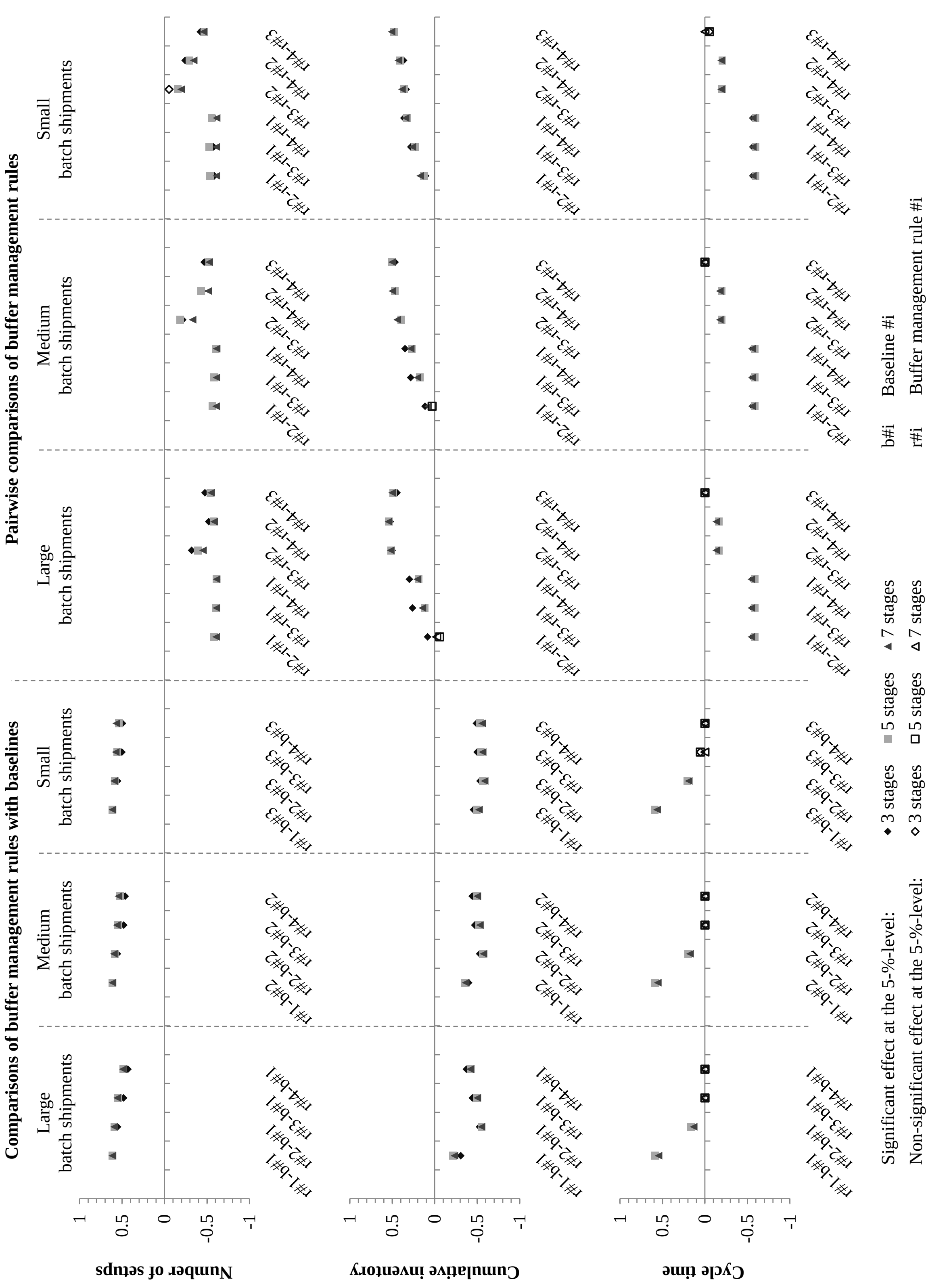


Figure 6: Significance and effect size statistics for the buffer management rules of the three-, five-, and seven-stage production systems.

4 Conclusions and managerial implications

This paper addressed the flexible control of serial multi-stage production systems where production rates change over time due to worker learning and forgetting. Prior solution approaches often suffered from their inability to take account of interactions across production stages, which may considerably impact system performance. Furthermore, they often promoted rigid production control policies that are not able to capture the dynamic character of constantly changing production rates. This paper first identified performance indicators typically used in production management that were significantly affected by worker learning and forgetting. In a second step, the input parameters and input parameter interactions driving these performance indicators were analyzed. Finally, the paper proposed flexible, real-time buffer management rules that aim at lowering excessive inventory build-up which may result from adverse input parameter combinations. Several managerial implications follow from the simulation study:

- Effective production control needs to take account of interactions of different stages in a production system, potentially even across multiple stages. In fact, the influence of production stage parameters that seem to worsen system performance when evaluated separately may be mitigated through production parameters of the same or other stages.
- The effectiveness of buffer management increases with decreasing batch shipment sizes. In case of all batch shipment sizes, the application of buffer management (rules r#2 to r#4) considerably reduces the cumulative inventory, while not (rules r#3 and r#4) or only marginally (rule r#2) increasing the cycle time compared with the case without buffer management (baselines b#1 to b#3).
- Different buffer sizes lead to significantly different results. Especially, very small buffer sizes can have a significant (adverse) impact on the KPIs, which differ from the effects slightly larger buffer sizes have. The results obtained in this paper give rise to the hypothesis that an optimal buffer size exists at which the effects of too strict buffer management (i.e., too many interruptions of the production process) and too lax buffer management (i.e., no intervention with or no control of the production process at all) are offset and buffer management is most effective. Yet, further research is necessary to gain more insights into these relationships.

One limitation of the paper at hand is that the results obtained are restricted to production systems similar to the one studied in this paper. Future research could therefore extend the model presented in this paper to a network setting, where a production stage may supply more than a single downstream stage, or receive products from more than a single upstream stage. Furthermore, it may be worth investigating alternative buffer stock configurations as the positioning of the buffer stocks clearly impact system performance. Another interesting research topic would be to consider random machine failures that could lead to the production of imperfect items as in Jaber and Khan (2010) and Glock and Jaber (2013b). In such scenarios, the advantages of simulation models over analytic models may come into play as simulation models can easily handle the interference of various sources of randomness.

References

Adler, G.L., Nanda, R., 1974. The effects of learning on optimal lot size determination—single product case. *AIIE Transactions*, 6 (1), 14-20.

Beach, R., Muhlemann, A.P., Price, D.H.R., Paterson, A., Sharp, J. A., 2000. A review of manufacturing flexibility. *European Journal of Operational Research*, 122 (1), 41-57.

Bogaschewsky, R., Glock, C.H., 2009. Rational inefficiencies in lot-sizing models with learning effects. *Zeitschrift für Betriebswirtschaft*, 4, 37-57.

Bukchin, Y., Cohen, Y., 2013. Minimising throughput loss in assembly lines due to absenteeism and turnover via work-sharing. *International Journal of Production Research*, 51 (20), 6140-6151.

Carlson, J.G., Rowe, A.J., 1976. How much does forgetting cost. *Industrial Engineering*, 8 (9), 40-47.

Carlsson, B., 1989. Flexibility and the theory of the firm. *International Journal of Industrial Organization*, 7 (2), 179-203.

Cohen, J., 1992. A power primer. *Psychological Bulletin*, 112 (1), 155-159.

Dolgui, A., Gordon, V., Strusevich, V., 2012. Single machine scheduling with precedence constraints and positionally dependent processing times. *Computers & Operations Research*, 39 (6), 1218-24.

Dunn, R.H., 1987. *The quest for software reliability*. In: Schulmeyer, G.G., McManus, J.I. (Eds.), Handbook of Software Quality Assurance, Van Nostrand Reinhold, New York, NY, pp. 342-384.

Field, A., Miles, J., Field, Z., 2012. *Discovering Statistics Using R*. Sage, Los Angeles, CA.

Finch, B.J., Luebbe, R., 1995. The impact of learning rate and constraints on production line performance. *International Journal of Production Research*, 33 (3), 631-42.

Friedman, M., 1937. The use of ranks to avoid the assumption of normality implicit in the analysis of variance. *Journal of the American Statistical Association*, 32 (200), 675-701.

Globerson, S., Levin, N., Shtub, A., 1989. The impact of breaks on forgetting when performing a repetitive task. *IIE Transactions*, 21 (4), 376-381.

Glock, C.H., 2010. Batch sizing with controllable production rates. *International Journal of Production Research*, 48 (20), 5925-5942.

Glock, C.H., 2011. Batch sizing with controllable production rates in a multi-stage production system. *International Journal of Production Research*, 49 (20), 6017-6039.

Glock, C.H., Jaber, M.Y., 2013a. Learning effects and the phenomenon of moving bottlenecks in a two-stage production system. *Applied Mathematical Modelling*, 37 (18–19), 8617-8628.

Glock, C.H., Jaber, M.Y., 2013b. A multi-stage production-inventory model with learning and forgetting effects, rework and scrap. *Computers & Industrial Engineering*, 64 (2), 708-720.

Glock, C.H., Jaber, M.Y., Zolfaghari, S., 2012. Production planning for a ramp-up process with learning in production and growth in demand. *International Journal of Production Research*, 50 (20), 5707-5718.

Goldratt, E.M., Cox, J., 1984. *The goal: An ongoing improvement process*. Gower, Aldershot.

Gordon, V., Strusevich, V., Dolgui, A., 2012. Scheduling with due date assignment under special conditions on job processing. *Journal of Scheduling*, 15 (4), 447-56.

Grosse, E.H., Glock, C.H., 2015. The effect of worker learning on manual order picking processes. *International Journal of Production Economics*, 170, Part C, 882-890.

Grosse, E.H., Glock, C.H., Jaber, M.Y., 2013. The effect of worker learning and forgetting on storage reassignment decisions in order picking systems. *Computers & Industrial Engineering*, 66 (4), 653-662.

Hyun, J.H., Ahn, B.H., 1992. A unifying framework for manufacturing flexibility. *Manufacturing Review*, 5 (4), 251-260.

Jaber, M.Y., Bonney, M., 1999. The economic manufacture/order quantity (EMQ/EOQ) and the learning curve: Past, present, and future. *International Journal of Production Economics*, 59 (1), 93-102.

Jaber, M.Y., Khan, M., 2010. Managing yield by lot splitting in a serial production line with learning, rework and scrap. *International Journal of Production Economics*, 124 (1), 32-39.

Keachie, E.C., Fontana, R.J., 1966. Effects of learning on optimal lot size. *Management Science*, 13 (2), B-102-B-108.

Kim, S., Nembhard, D.A., 2013. Rule mining for scheduling cross training with a heterogeneous workforce. *International Journal of Production Research*, 51 (8), 2281-2300.

Kleijnen, J.P.C., 2008. *Design and analysis of simulation experiments*. Springer, New York, NY.

Kleijnen, J.P.C., Sanchez, S.M., Lucas, T.W., Cioppa, T.M., 2005. State-of-the-art review: A user’s guide to the brave new world of designing simulation experiments. *INFORMS Journal on Computing*, 17 (3), 263-289.

Kuo, C.-T., Lim, J.-T., Meerkov, S.M., 1996. Bottlenecks in serial production lines: A system-theoretic approach. *Mathematical Problems in Engineering*, 2 (3), 233-276.

Lawrence, S.R., Buss, A.H., 1995. Economic analysis of production bottlenecks. *Mathematical Problems in Engineering*, 1 (4), 341-363.

Liu, C., Wang, J., Leung, J.Y.T., 2016. Worker assignment and production planning with learning and forgetting in manufacturing cells by hybrid bacteria foraging algorithm. *Computers & Industrial Engineering*, 96, 162-179.

Loeppky, J.L., Sacks, J., Welch, W.J., 2012. Choosing the sample size of a computer experiment: A practical guide. *Technometrics*, 51 (4), 366-376.

Longo, F., 2010. Design and integration of the containers inspection activities in the container terminal operations. *International Journal of Production Economics*, 125 (2), 272-283.

Mak, L.C., Wong, W.K., Leung, Y.S., 2014. A simulation analysis of the impact of production lot size and its interaction with operator competence on manufacturing system performance. *Simulation Modelling Practice and Theory*, 49, 203-214.

Mandelbaum, M., Buzacott, J., 1990. Flexibility and decision making. *European Journal of Operational Research*, 44 (1), 17-27.

Nembhard, D.A., 2001. Heuristic approach for assigning workers to tasks based on individual learning rates. *International Journal of Production Research*, 39 (9), 1955-1968.

Nembhard, D.A., Bentefouet, F., 2014. Selection policies for a multifunctional workforce. *International Journal of Production Research*, 52 (16), 4785-4802.

Nembhard, D.A., Shafer, S.M., 2008. The effects of workforce heterogeneity on productivity in an experiential learning environment. *International Journal of Production Research*, 46 (14), 3909-3929.

Pan, E., Wang, G., Xi, L., Chen, L., Han, X., 2014. Single-machine group scheduling problem considering learning, forgetting effects and preventive maintenance. *International Journal of Production Research*, 52 (19), 5690-704.

Pidd, M., 2004. *Computer simulation in management science*. 5th ed., Wiley, New York, NY.

Ronen, B., Starr, M.K., 1990. Synchronized manufacturing as in OPT: From practice to theory. *Computers & Industrial Engineering*, 18 (4), 585-600.

Rosenthal, R., 1991. *Meta-analytic procedures for social research*. Sage, Newbury Park, CA.

Schragenheim, E., Ronen, B., 1991. Buffer-management: A diagnostic tool for production control. *Production and Inventory Management Journal*, 32 (2), 74-79.

Shafer, S.M., Nembhard, D.A., Uzumeri, M.V., 2001. The effects of worker learning, forgetting, and heterogeneity on assembly line productivity. *Management Science*, 47 (12), 1639-1653.

Smunt, T.L. 1987. The impact of worker forgetting on production scheduling. *International Journal of Production Research*, 25 (5), 689-701.

Sule, D.R., 1978. The effect of alternate periods of learning and forgetting on economic manufacturing quantity. *AIIE Transactions*, 10 (3), 338-343.

Upton, D.M., 1994. The management of manufacturing flexibility. *California Management Review*, 36 (2), 72-89.

Wright, T.P., 1936. Factors affecting the cost of airplanes. *Journal of the Aeronautical Sciences*, 3 (4), 122-128.

Yelle, L.E., 1979. The learning curve: Historical review and comprehensive survey. *Decision Sciences*, 10 (2), 302-328.

Zhai, Y., Sun, S., Wang, J., Niu, G., 2011. Job shop bottleneck detection based on orthogonal experiment. *Computers & Industrial Engineering*, 61 (3), 872-880.

Appendix

Appendix A: Additional results of simulation study of three-stage production system

Table A.1: Results of the regression analysis of the three-stage production system with a focus on independent variables with a significant influence on cumulative inventory (at the 5 percent level).

Independent variable	Coefficient	Standardized coefficient	Standard error	t-value	Pr(>|t|)
(Intercept)	$5.30 \cdot 10^{7}$	0.03	$4.92 \cdot 10^{5}$	107.68	<0.0001
q	$1.86 \cdot 10^{4}$	0.43	$7.68 \cdot 10^{2}$	24.27	<0.0001
$T_{1,2}^{-1}$	$-7.62 \cdot 10^{7}$	-0.21	$6.47 \cdot 10^{6}$	-11.78	<0.0001
$T_{1,3}^{-1}$	$-8.23 \cdot 10^{7}$	-0.23	$6.37 \cdot 10^{6}$	-12.93	<0.0001
l_1	$1.13 \cdot 10^{8}$	0.25	$7.95 \cdot 10^{6}$	14.22	<0.0001
l_2	$-1.97 \cdot 10^{8}$	-0.44	$7.84 \cdot 10^{6}$	-25.14	<0.0001
l_3	$-2.04 \cdot 10^{8}$	-0.46	$7.85 \cdot 10^{6}$	-26.00	<0.0001
$T_{1,1}^{-1}: l_1$	$5.67 \cdot 10^{8}$	0.10	$1.03 \cdot 10^{8}$	5.50	<0.0001
$T_{1,2}^{-1}: l_1$	$-4.93 \cdot 10^{8}$	-0.09	$1.06 \cdot 10^{8}$	-4.66	<0.0001
$T_{1,2}^{-1}: l_2$	$1.15 \cdot 10^{9}$	0.20	$1.04 \cdot 10^{8}$	11.08	<0.0001
$T_{1,3}^{-1}: l_3$	$9.90 \cdot 10^{8}$	0.17	$1.04 \cdot 10^{8}$	9.50	<0.0001
$l_1: l_2$	$-1.15 \cdot 10^{9}$	-0.16	$1.27 \cdot 10^{8}$	-9.04	<0.0001
$l_2: l_3$	$-9.49 \cdot 10^{8}$	-0.14	$1.27 \cdot 10^{8}$	-7.47	<0.0001
$l_1: l_3$	$-7.78 \cdot 10^{8}$	-0.11	$1.26 \cdot 10^{8}$	-6.20	<0.0001
$q: l_1$	$-5.65 \cdot 10^{4}$	-0.08	$1.24 \cdot 10^{4}$	-4.55	<0.0001
$l_3: f_1$	$-4.88 \cdot 10^{8}$	-0.07	$1.25 \cdot 10^{8}$	-3.91	0.0001
$q: l_2$	$-4.67 \cdot 10^{4}$	-0.07	$1.25 \cdot 10^{4}$	-3.75	0.0002
$T_{1,3}^{-1}: l_2$	$-3.48 \cdot 10^{8}$	-0.06	$1.01 \cdot 10^{8}$	-3.46	0.0006
$q: T_{1,1}^{-1}$	$-2.68 \cdot 10^{4}$	-0.05	$1.05 \cdot 10^{4}$	-2.55	0.0110
$T_{1,1}^{-1}: l_2$	$-2.54 \cdot 10^{8}$	-0.04	$1.06 \cdot 10^{8}$	-2.40	0.0167
$T_{1,2}^{-1}: l_3$	$-2.33 \cdot 10^{8}$	-0.04	$1.00 \cdot 10^{8}$	-2.33	0.0203
$T_{1,2}^{-1}: f_1$	$2.16 \cdot 10^{8}$	0.04	$9.66 \cdot 10^{7}$	2.24	0.0255
$q: l_3$	$-2.52 \cdot 10^{4}$	-0.04	$1.17 \cdot 10^{4}$	-2.16	0.0314
f_2	$1.66 \cdot 10^{7}$	0.04	$7.78 \cdot 10^{6}$	2.14	0.0332
$T_{1,1}^{-1}: f_3$	$-1.98 \cdot 10^{8}$	-0.04	$9.71 \cdot 10^{7}$	-2.04	0.0420

Table A.2: Descriptive analysis of the impact of buffer management rules on the KPIs of the three-stage production system.

		Number of setups		**Cumulative inventory**		**Cycle time**	
		Mean	Std. dev.	Mean	Std. dev.	Mean	Std. dev.
Large	b#1	7.12	2.84	6.52·107	2.80·107	17576.80	5759.09
batch	r#1	13.56	2.50	5.47·107	1.94·107	19554.91	6411.53
shipments	r#2	9.68	2.12	5.59·107	2.09·107	17585.74	5753.55
	r#3	8.91	2.19	6.03·107	2.31·107	17576.80	5759.09
	r#4	8.05	2.46	6.28·107	2.51·107	17576.80	5759.09
Medium	b#2	10.37	4.90	5.05·107	2.67·107	16390.74	5468.37
batch	r#1	21.89	4.35	3.63·107	1.43·107	18588.20	6195.99
shipments	r#2	15.47	3.77	3.79·107	1.72·107	16401.48	5462.46
	r#3	14.56	3.64	4.15·107	1.80·107	16390.74	5468.37
	r#4	13.34	3.94	4.41·107	1.97·107	16390.74	5468.37
Small	b#3	23.71	13.30	3.48·107	2.64·107	15142.62	5198.55
batch	r#1	55.57	11.99	1.63·107	8.66·106	17672.00	6134.25
shipments	r#2	38.58	10.20	1.86·107	1.46·107	15153.01	5192.64
	r#3	38.35	9.55	1.99·107	1.30·107	15142.84	5198.34
	r#4	36.68	9.60	2.15·107	1.32·107	15142.62	5198.55

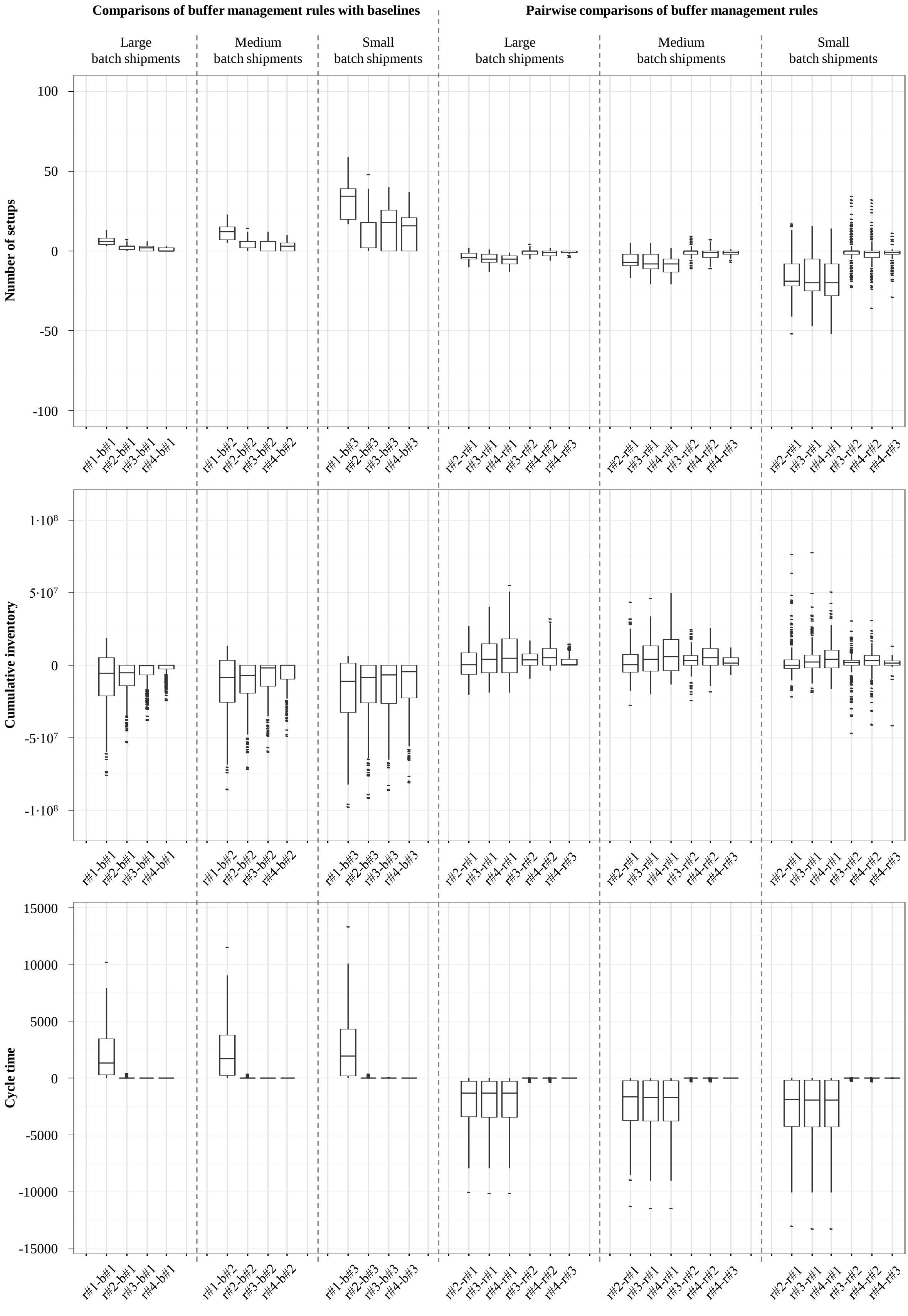


Figure A.1: Boxplots of the differences in the KPIs resulting from the buffer management rules of the three-stage production system.

Appendix B: Additional results of simulation study of five-stage production system

Table B.1: Results of the regression analysis of the five-stage production system with a focus on independent variables with a significant influence on cumulative inventory (at the 5 percent level).

Independent variable	Coefficient	Standardized coefficient	Standard error	t-value	Pr(>|t|)
(Intercept)	$1.01 \cdot 10^{8}$	-0.01	$8.43 \cdot 10^{5}$	119.73	<0.0001
q	$4.66 \cdot 10^{4}$	0.67	$1.31 \cdot 10^{3}$	35.43	<0.0001
l_1	$1.23 \cdot 10^{8}$	0.17	$1.34 \cdot 10^{7}$	9.20	<0.0001
l_2	$-2.16 \cdot 10^{8}$	-0.30	$1.36 \cdot 10^{7}$	-15.92	<0.0001
l_3	$-1.77 \cdot 10^{8}$	-0.25	$1.38 \cdot 10^{7}$	-12.77	<0.0001
l_4	$-1.41 \cdot 10^{8}$	-0.20	$1.39 \cdot 10^{7}$	-10.18	<0.0001
l_5	$-1.87 \cdot 10^{8}$	-0.26	$1.36 \cdot 10^{7}$	-13.76	<0.0001
$T_{1,3}^{-1}$	$-9.18 \cdot 10^{7}$	-0.16	$1.10 \cdot 10^{7}$	-8.31	<0.0001
$T_{1,2}^{-1}: l_2$	$1.47 \cdot 10^{9}$	0.16	$1.78 \cdot 10^{8}$	8.29	<0.0001
$T_{1,2}^{-1}$	$-7.96 \cdot 10^{7}$	-0.14	$1.07 \cdot 10^{7}$	-7.44	<0.0001
$T_{1,4}^{-1}: l_4$	$1.09 \cdot 10^{9}$	0.12	$1.76 \cdot 10^{8}$	6.16	<0.0001
$T_{1,3}^{-1}: l_3$	$1.08 \cdot 10^{9}$	0.12	$1.82 \cdot 10^{8}$	5.93	<0.0001
$T_{1,5}^{-1}$	$-6.55 \cdot 10^{7}$	-0.11	$1.12 \cdot 10^{7}$	-5.83	<0.0001
$q: l_1$	$-1.15 \cdot 10^{5}$	-0.11	$2.11 \cdot 10^{4}$	-5.47	<0.0001
$T_{1,5}^{-1}: l_5$	$8.98 \cdot 10^{8}$	0.10	$1.73 \cdot 10^{8}$	5.20	<0.0001
$l_2: l_3$	$-9.16 \cdot 10^{8}$	-0.08	$2.23 \cdot 10^{8}$	-4.10	0.0001
$l_3: l_4$	$-8.20 \cdot 10^{8}$	-0.07	$2.18 \cdot 10^{8}$	-3.76	0.0002
$T_{1,3}^{-1}: l_1$	$-6.44 \cdot 10^{8}$	-0.07	$1.78 \cdot 10^{8}$	-3.63	0.0003
$l_1: l_2$	$-7.82 \cdot 10^{8}$	-0.07	$2.18 \cdot 10^{8}$	-3.58	0.0004
$q: l_2$	$-7.55 \cdot 10^{4}$	-0.07	$2.25 \cdot 10^{4}$	-3.36	0.0009
$T_{1,4}^{-1}$	$-3.76 \cdot 10^{7}$	-0.07	$1.14 \cdot 10^{7}$	-3.29	0.0011
$T_{1,1}^{-1}: l_3$	$-5.92 \cdot 10^{8}$	-0.06	$1.87 \cdot 10^{8}$	-3.17	0.0017
$q: l_3$	$-6.57 \cdot 10^{4}$	-0.06	$2.14 \cdot 10^{4}$	-3.07	0.0023
$q: l_5$	$-6.34 \cdot 10^{4}$	-0.06	$2.11 \cdot 10^{4}$	-3.01	0.0028
$T_{1,4}^{-1}: l_5$	$-4.59 \cdot 10^{8}$	-0.06	$1.56 \cdot 10^{8}$	-2.95	0.0034
$l_3: l_5$	$-6.16 \cdot 10^{8}$	-0.06	$2.20 \cdot 10^{8}$	-2.80	0.0054
$l_1: l_4$	$-6.05 \cdot 10^{8}$	-0.05	$2.16 \cdot 10^{8}$	-2.80	0.0054
$T_{1,4}^{-1}: l_1$	$-4.67 \cdot 10^{8}$	-0.05	$1.78 \cdot 10^{8}$	-2.62	0.0092
$T_{1,1}^{-1}: f_2$	$4.37 \cdot 10^{8}$	0.05	$1.71 \cdot 10^{8}$	2.55	0.0113
$l_3: f_1$	$-5.41 \cdot 10^{8}$	-0.05	$2.16 \cdot 10^{8}$	-2.50	0.0128
$T_{1,5}^{-1}: l_1$	$-4.43 \cdot 10^{8}$	-0.05	$1.79 \cdot 10^{8}$	-2.48	0.0137
$l_2: l_5$	$-5.09 \cdot 10^{8}$	-0.05	$2.09 \cdot 10^{8}$	-2.44	0.0152
$T_{1,2}^{-1}: l_1$	$-4.11 \cdot 10^{8}$	-0.05	$1.76 \cdot 10^{8}$	-2.34	0.0197
$q: T_{1,5}^{-1}$	$-3.93 \cdot 10^{4}$	-0.04	$1.72 \cdot 10^{4}$	-2.28	0.0233
$q: T_{1,2}^{-1}$	$-3.98 \cdot 10^{4}$	-0.04	$1.75 \cdot 10^{4}$	-2.28	0.0235
f_2	$2.95 \cdot 10^{7}$	0.04	$1.32 \cdot 10^{7}$	2.23	0.0264

Independent variable	**Coefficient**	**Standardized coefficient**	**Standard error**	**t-value**	**Pr(>|t|)**
f_5	-2.86·10^{7}	-0.04	1.31·10^{7}	-2.18	0.0303
$f_1: f_3$	4.28·10^{8}	0.04	1.99·10^{8}	2.15	0.0322
$q: T_{1,3}^{-1}$	-3.60·10^{4}	-0.04	1.69·10^{4}	-2.13	0.0338
$q: T_{1,1}^{-1}$	-3.74·10^{4}	-0.04	1.84·10^{4}	-2.03	0.0430
$l_4: l_5$	-4.29·10^{8}	-0.04	2.18·10^{8}	-1.97	0.0495

Appendix C: Additional results of simulation study of seven-stage production system

Table C.1: Results of the regression analysis of the seven-stage production system with a focus on independent variables with a significant influence on cumulative inventory (at the 5 percent level).

Independent variable	**Coefficient**	**Standardized coefficient**	**Standard error**	**t-value**	**Pr(>|t|)**
(Intercept)	1.50·10^{8}	0.00	1.34·106	112.49	<0.0001
q	7.31·10^{4}	0.74	2.16·103	33.83	<0.0001
$T_{1,5}^{-1}: l_5$	1.22·10^{9}	0.09	2.93·108	4.15	<0.0001
$T_{1,6}^{-1}: l_6$	1.35·10^{9}	0.10	2.92·108	4.63	<0.0001
$q: l_1$	-1.77·10^{5}	-0.11	3.74·104	-4.73	<0.0001
$q: l_2$	-1.68·10^{5}	-0.11	3.68·104	-4.55	<0.0001
$q: l_4$	-1.55·10^{5}	-0.10	3.44·104	-4.52	<0.0001
l_2	-1.78·10^{8}	-0.18	2.29·107	-7.75	<0.0001
l_4	-1.42·10^{8}	-0.14	2.21·107	-6.42	<0.0001
l_7	-1.45·10^{8}	-0.14	2.31·107	-6.29	<0.0001
l_6	-1.35·10^{8}	-0.13	2.23·107	-6.05	<0.0001
l_3	-1.30·10^{8}	-0.13	2.28·107	-5.70	<0.0001
l_5	-1.29·10^{8}	-0.13	2.27·107	-5.66	<0.0001
$T_{1,4}^{-1}$	-9.96·10^{7}	-0.12	1.81·107	-5.50	<0.0001
$T_{1,2}^{-1}: l_2$	1.60·10^{9}	0.12	2.98·108	5.38	<0.0001
l_1	1.10·10^{8}	0.11	2.26·107	4.86	<0.0001
$T_{1,3}^{-1}$	-8.54·10^{7}	-0.10	1.85·107	-4.61	<0.0001
$T_{1,2}^{-1}$	-8.27·10^{7}	-0.10	1.80·107	-4.59	<0.0001
$T_{1,3}^{-1}: l_3$	1.25·10^{9}	0.10	3.09·108	4.04	0.0001
$T_{1,5}^{-1}$	-6.68·10^{7}	-0.08	1.88·107	-3.56	0.0004
$T_{1,7}^{-1}: l_7$	1.07·10^{9}	0.08	3.14·108	3.40	0.0008
$q: l_6$	-1.19·10^{5}	-0.08	3.69·104	-3.24	0.0014
$q: l_3$	-1.08·10^{5}	-0.07	3.42·104	-3.16	0.0018
$l_1: l_2$	-1.14·10^{9}	-0.07	3.63·108	-3.14	0.0019
$T_{1,1}^{-1}: l_1$	7.42·10^{8}	0.06	2.93·108	2.53	0.0121
$l_3: l_6$	-9.60·10^{8}	-0.06	3.80·108	-2.52	0.0123
$T_{1,6}^{-1}$	-4.52·10^{7}	-0.05	1.80·107	-2.51	0.0129
$T_{1,1}^{-1}: T_{1,6}^{-1}$	-5.26·10^{8}	-0.05	2.26·108	-2.32	0.0211

Independent variable	**Coefficient**	**Standardized coefficient**	**Standard error**	***t*-value**	**Pr(>|*t*|)**
$T_{1,3}^{-1}: l_1$	$-6.44 \cdot 10^8$	-0.05	2.90·108	-2.22	0.0273
$T_{1,5}^{-1}: l_3$	$-6.18 \cdot 10^8$	-0.05	2.86·108	-2.17	0.0313
$T_{1,4}^{-1}: l_4$	$6.41 \cdot 10^8$	0.05	2.96·108	2.17	0.0313
$l_2: l_4$	$-7.62 \cdot 10^8$	-0.05	3.56·108	-2.14	0.0330
$q: T_{1,7}^{-1}$	$-6.11 \cdot 10^4$	-0.05	2.88·104	-2.12	0.0350
$l_2: l_3$	$-8.10 \cdot 10^8$	-0.05	3.87·108	-2.09	0.0372
$l_1: l_3$	$-7.58 \cdot 10^8$	-0.05	3.71·108	-2.05	0.0418
$q: T_{1,6}^{-1}$	$6.15 \cdot 10^4$	0.05	3.08·104	2.00	0.0470
$T_{1,4}^{-1}: l_3$	$-5.94 \cdot 10^8$	-0.05	2.98·108	-1.99	0.0477
$f_4: f_6$	$6.79 \cdot 10^8$	0.05	3.44·108	1.97	0.0495

Table C.2: Descriptive analysis of the impact of buffer management rules on the KPIs of the seven-stage production system.

		Number of setups		**Cumulative inventory**		**Cycle time**	
		Mean	Std. dev.	Mean	Std. dev.	Mean	Std. dev.
Large batch shipments	b#1	21.55	3.68	1.95·108	4.60·107	30752.33	6593.88
	r#1	30.17	3.46	1.88·108	4.28·107	32422.54	7229.27
	r#2	25.14	3.24	1.86·108	4.20·107	30756.71	6592.30
	r#3	23.96	3.45	1.90·108	4.34·107	30752.33	6593.88
	r#4	22.74	3.63	1.93·108	4.45·107	30752.33	6593.88
Medium batch shipments	b#2	34.40	6.72	1.42·108	3.73·107	26015.14	5636.45
	r#1	48.91	5.76	1.28·108	3.03·107	27612.36	6250.01
	r#2	41.01	5.64	1.29·108	3.17·107	26022.78	5634.44
	r#3	39.80	5.78	1.32·108	3.23·107	26015.14	5636.45
	r#4	38.37	6.03	1.35·108	3.31·107	26015.14	5636.45
Small batch shipments	b#3	86.26	18.48	8.38·107	3.11·107	21140.23	4788.33
	r#1	124.12	16.00	6.00·107	1.83·107	22891.43	5490.86
	r#2	103.23	15.56	6.51·107	2.41·107	21147.94	4785.71
	r#3	102.25	14.92	6.68·107	2.42·107	21140.23	4788.33
	r#4	100.81	15.31	6.85·107	2.37·107	21140.23	4788.33

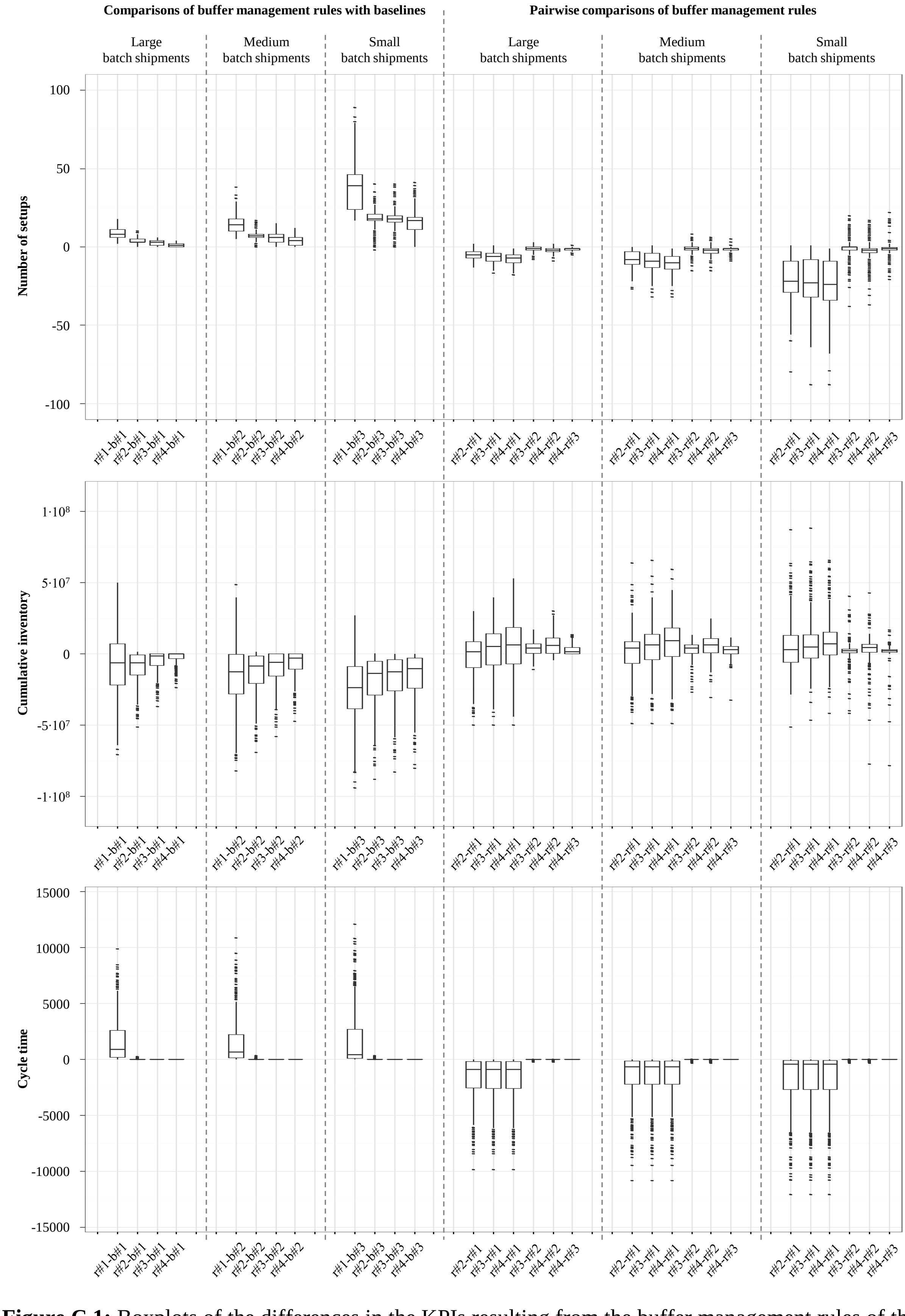


Figure C.1: Boxplots of the differences in the KPIs resulting from the buffer management rules of the seven-stage production system.

Academic CV of the author

05/2015 – 11/2017	Doctoral Candidate and Research Assistant, Institute of Production and Supply Chain Management, Technische Universität Darmstadt, Darmstadt, Germany (Advisor: Prof. Dr. Christoph H. Glock)
11/2015 – 02/2016	Visiting Scholar, Division of Environmental and Ecological Engineering, Purdue University, West Lafayette, IN, USA (Advisor: Prof. Dr. John W. Sutherland)
04/2013 – 04/2015	Doctoral Candidate and Research Assistant, Carlo and Karin Giersch Endowed Chair "Business Management: Industrial Management", Technische Universität Darmstadt, Darmstadt, Germany (Advisor: Prof. Dr. Christoph H. Glock)
10/2006 – 10/2012	Diploma in Industrial Engineering (Dipl.-Wi.-Ing.), Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany
09/2009 – 07/2010	Master of Business Administration (MBA), Union Graduate College (merged with Clarkson University in 01/2016), Schenectady, NY, USA