

Titre: Status-Centered Versus Self-Centered Demand Driven Execution
Title: System

Auteur: Lucas Alves Da Silva Santos
Author:

Date: 2026

Type: Mémoire ou thèse / Dissertation or Thesis

Référence: Alves Da Silva Santos, L. (2026). Status-Centered Versus Self-Centered Demand
Citation: Driven Execution System [Ph.D. thesis, Polytechnique Montréal]. PolyPublie.
<https://publications.polymtl.ca/76269/>

 **Document en libre accès dans PolyPublie**
Open Access document in PolyPublie

URL de PolyPublie: <https://publications.polymtl.ca/76269/>
PolyPublie URL:

**Directeurs de
recherche:** Maha Ben Ali, & Pierre Baptiste
Advisors:

Programme: Doctorat en génie industriel
Program:

POLYTECHNIQUE MONTRÉAL

affiliée à l'Université de Montréal

Status-Centered Versus Self-Centered Demand Driven Execution System

LUCAS ALVES DA SILVA SANTOS

Département de mathématiques et de génie industriel

Thèse présentée en vue de l'obtention du diplôme de *Philosophiæ Doctor*

Génie Industriel

Avril 2026

POLYTECHNIQUE MONTRÉAL

affiliée à l'Université de Montréal

Cette thèse intitulée :

Status-Centered Versus Self-Centered Demand Driven Execution System

présentée par **Lucas ALVES DA SILVA SANTOS**

en vue de l'obtention du diplôme de *Philosophiæ Doctor*

a été dûment acceptée par le jury d'examen constitué de :

Robert PELLERIN, président

Maha BEN ALI, membre et directrice de recherche

Pierre BAPTISTE, membre et codirecteur de recherche

Jean-Marc FRAYRET, membre

Mustapha NOUR EL FATH, membre externe

DEDICATION

To my father.

ACKNOWLEDGEMENTS

First, I would like to express my deepest gratitude to my research director, Maha Ben Ali. Thank you for your support during this period. Your meticulous approach to research has taught me invaluable lessons that I will carry forward. I am especially grateful for the balance you maintained; while you were always present when needed, you taught me a strong sense of autonomy, giving me the freedom to explore my own ideas and pave my own academic path.

My sincere thanks to my co-director, Pierre Baptiste. I am grateful for your availability and guidance throughout this process. You taught me to always look beyond the surface and seek the deeper meaning. Our discussions certainly pushed this research to new levels of novelty.

They say a journey of a thousand miles starts with a single step, so I would like to express my gratitude to my first research supervisor, Kleber Esposto. It has been such a long journey from the moment I was accepted as your master's student. Thank you for believing in me as a researcher.

I would also like to thank Jean-Marc Frayret. As my first teacher at Polytechnique Montréal, the lessons from your classes still resonate in my mind. Many of the ideas and thoughts that we discussed almost five years ago are now present in the foundations of this work.

I extend my gratitude to Robert Pellerin and Mustapha Nour El Fath, professors at Polytechnique Montréal and Université Laval respectively, for agreeing to serve on this jury. Thank you for the time dedicated to reading this research and for providing your valuable insights and opinions.

I am also grateful to everyone I encountered during this PhD journey — whether in conferences, classrooms, the office, or the corridors of the university. Each interaction contributed to my growth; if you are reading this, you are likely one of these people, and I thank you.

To my parents, my brother and my parents-in-law: Thank you for your endless encouragement. This chapter ends with a new one to begin, and I am sure that, no matter where I go from here, you will be there supporting me.

Finally, I reserve my most sincere and profound thanks for my wife. Thank you for embracing this journey with me. Your unwavering support and your belief in me, even when I thought I could not make it, were the cornerstone of my path. You are the most important part of my life. We started this journey as a duo and ended it as a trio. Thank you also for the greatest gift of all, our baby Matheus, and thank you for always being by my side.

RÉSUMÉ

La Planification des Besoins en Matériaux Pilotée par la Demande (DDMRP) a émergé comme une méthodologie permettant aux organisations de faire face à l'environnement de volatilité, incertitude, complexité et ambiguïté (VUCA) de la fabrication contemporaine. En découplant l'offre de la variabilité de la demande grâce au positionnement stratégique des buffers, le DDMRP atténue l'effet coup de fouet via une visibilité et une réactivité accrue. Cependant, alors que la phase de planification du DDMRP est dynamiquement réactive aux signaux du marché, la phase d'exécution reste régie par des heuristiques autocentrées qui ne parviennent pas à prendre en compte la réalité stochastique de l'atelier.

Cette thèse identifie une dichotomie structurelle critique : le moteur de la Planification Pilotée par la Demande (DDP) fonctionne comme une boucle de contrôle fermée, réactive aux signaux du marché, tandis que le moteur standard de l'Exécution Pilotée par la Demande (DDE) fonctionne comme un système en boucle ouverte, aveugle aux contraintes de capacité et aux perturbations de flux. L'objectif principal de cette thèse est d'améliorer la robustesse et la performance opérationnelle des *flowshops* gérés par le DDMRP en faisant passer la phase d'exécution d'heuristiques autocentrées à un système d'exécution dynamique centré sur l'état du système.

Ancrée sur la décomposition fonctionnelle des Systèmes de Pilotage de la Production, cette thèse atteint cet objectif à travers trois contributions distinctes, utilisant la Simulation à Événements Discrets (DES) pour reproduire la nature stochastique de l'environnement de production comme banc d'essai pour toutes les contributions. Premièrement, concernant le séquençement, la recherche propose un mécanisme d'affectation dynamique qui priorise les ordres en fonction de la disponibilité réelle des produits plutôt que des priorités de planification statiques. Deuxièmement, concernant le lancement des ordres, l'étude utilise les commandes urgentes comme indicateur d'hétérogénéité pour tester un mécanisme de lancement, retenant les ordres de réapprovisionnement standard lorsqu'une demande prioritaire est présente afin de protéger la robustesse de l'atelier. Troisièmement, concernant la génération et le lancement des ordres, la thèse couple le DDMRP avec l'apprentissage *automatique* pour proposer un processus intelligent de génération et de lancement des ordres de fabrication. Ce mécanisme utilise le temps d'écoulement prédictif afin de synchroniser le réapprovisionnement avec les conditions de l'atelier en temps réel.

Les résultats démontrent que les priorités statiques de planification ne parviennent pas à refléter l'urgence des ordres de fabrication en temps réel. Par conséquent, il est démontré que l'affectation dynamique surpasse le séquençement DDMRP standard pour prévenir les ruptures de stock au niveau local. De plus, le mécanisme de filtrage à l'étape du lancement des ordres prouve que capter l'hétérogénéité lors du lancement est essentiel pour maintenir la stabilité du flux. Enfin, l'intégration de l'apprentissage automatique permet une transition d'une génération et d'un lancement d'ordres de fabrication basé sur un facteur unique (soit le niveau de stock) vers une approche à deux facteurs (soit le niveau de stock et l'état du système), offrant une performance supérieure par rapport au DDMRP traditionnel.

Collectivement, ces contributions définissent un nouveau paradigme d'exécution pour le DDMRP. Les résultats confirment que fermer la boucle de rétroaction au niveau de l'exécution, en remplaçant les heuristiques autocentrées par des mécanismes de contrôle dynamiques basés sur l'état du système, améliore significativement la robustesse du système, réduisant les ruptures de stock et améliorant les taux de service dans des environnements à forte variabilité.

ABSTRACT

Demand Driven Material Requirements Planning (DDMRP) has emerged as a methodology for organizations to cope with the VUCA (Volatile, Uncertain, Complex and Ambiguous) environment of contemporary manufacturing. By decoupling supply from demand variability through strategic buffer positioning, DDMRP mitigates the bullwhip effect via enhanced visibility and responsiveness. However, while the planning phase of DDMRP is dynamically responsive to market signals, the execution phase remains governed by self-centered heuristics that fail to account for the stochastic reality of the shop floor.

This thesis identifies a critical structural dichotomy: the Demand Driven Planning (DDP) engine functions as a closed control loop, responsive to market signals, while the standard Demand Driven Execution (DDE) engine operates as an open-loop system, blind to capacity constraints and flow disruptions. The primary objective of this thesis is to enhance the robustness and operational performance of DDMRP managed flowshops by shifting the execution phase from self-centered heuristics to a dynamic, status-centered execution system.

Anchored on the functional decomposition of production control systems, this thesis achieves this objective through three distinct contributions, using Discrete Event Simulation (DES) to replicate the stochastic nature of the production environment as a testbed for all contributions. First, addressing Sequencing, the research proposes a dynamic dispatching mechanism that prioritizes orders based on real-time product availability rather than static planning priorities. Second, addressing Order Release, the study utilizes rush orders as a proxy for heterogeneity to test a release mechanism, withholding standard replenishment orders when high-priority demand is present to protect shop floor robustness. Third, addressing Order Generation and Order Release, the thesis couples DDMRP with Machine Learning to propose a ‘Smart’ Order Generation process. This mechanism uses predicted flow time to create a dynamic order generation and release gate to synchronize replenishment to real-time shop floor conditions.

The results demonstrate that static planning priorities fail to reflect real-time urgency. Therefore, dynamically dispatching is shown to outperform standard DDMRP sequencing in preventing stockouts at the local level. Furthermore, the gating mechanism at the order release stage proves that filtering heterogeneity at the release stage is essential for maintaining flow stability. Finally, the integration of Machine Learning enables a transition from single factor (inventory) to dual-

factor (inventory and system state) order generation and order release, yielding superior performance when compared to traditional DDMRP.

Collectively, these contributions define a new execution paradigm for DDMRP. The findings confirm that closing the feedback loop at the execution level, replacing self-centered heuristics with dynamic, status-based control mechanisms, significantly enhances system robustness, reducing stockouts and improving service levels in high-variability environments.

TABLE OF CONTENTS

DEDICATION	III
ACKNOWLEDGEMENTS	IV
RÉSUMÉ.....	V
ABSTRACT	VII
LIST OF TABLES	XIV
LIST OF FIGURES.....	XV
LISTE OF SYMBOLS AND ABBREVIATIONS	XVII
LIST OF APPENDICES	XX
CHAPTER 1 INTRODUCTION.....	1
1.1 Context	1
1.2 Research Problem.....	1
1.3 Thesis Structure.....	4
CHAPTER 2 LITERATURE REVIEW	6
2.1 Theoretical Foundations of Production Control Systems	6
2.1.1 Throughput-Oriented vs. WIP-Oriented Control	7
2.1.2 The Functional Decomposition of Production Control Systems.....	8
2.1.3 The Emergence of Hybrid Control Architectures	10
2.2 DDMRP Principles and Mechanics (Ptak & Smith, 2019)	11
2.2.1 Evolution of DDMRP Research: A Bias Toward Planning	15
2.3 DDMRP Through the Lens of PCSs Theory.....	16
2.3.1 Task 1: Sequencing and Dynamic Dispatching.....	17
2.3.2 Task 2: Order Release and Order Heterogeneity	19
2.3.3 Task 3: Order Generation and System State Prediction.....	22

2.4	Research Gaps	24
2.5	Literature Review Conclusion.....	25
CHAPTER 3 OBJECTIVES AND METHODOLOGY		26
3.1	Objectives.....	26
3.2	General Methodology.....	28
3.3	Case study	29
3.3.1	Main Elements of the DES Model	30
3.3.2	The Production Environment	31
3.3.3	Demand and Inventory Logic.....	32
3.3.4	Control Logic Mechanisms	32
3.4	Main Performance Metrics.....	32
CHAPTER 4 ARTICLE 1: DYNAMIC DISPATCHING OF DDMRP REPLENISHMENT ORDERS		34
ABSTRACT		34
RESUME.....		34
4.1	Introduction	35
4.2	DDMRP Concepts.....	36
4.2.1	Dispatching Rules in DDMRP Research	38
4.3	Dispatching Rule Models	41
4.3.1	Static Dispatching Rule Model (Static).....	41
4.3.2	Dynamic Dispatching Rule Model.....	42
4.4	Experiments.....	44
4.5	Results and Discussion.....	44
4.6	Managerial Insights.....	47
4.7	Conclusion.....	48

CHAPTER 5	ARTICLE 2: ROBUSTNESS ANALYSIS OF DEMAND DRIVEN MATERIAL REQUIREMENTS PLANNING FLOWSHOPS IN THE PRESENCE OF RUSH ORDERS	49
ABSTRACT		49
RESUME		49
5.1	Introduction	50
5.2	Literature Review	53
5.2.1	DDMRP Overview	53
5.2.2	Lead Time Calibration	55
5.2.3	Order Prioritization in DDMRP Flowshops	55
5.2.4	Management of Rush Orders	56
5.3	Methods	57
5.3.1	Case Study	58
5.3.2	Integration of Rush Orders	60
5.3.3	Experiments	62
5.3.4	Model and Lead Time Calibration	63
5.4	Results	65
5.4.1	Fulfillment of Non-Rush Orders	65
5.4.2	Fulfillment of Rush Orders	68
5.4.3	Loading Rate Effects	70
5.4.4	Lead Time Effects	70
5.4.5	Ratio of Rush Orders Effect	70
5.4.6	Impact of Order Prioritization Policies on DDMRP Robustness	71
5.5	Discussion	73
5.5.1	Theoretical Contributions	73

5.5.2	Practical Contributions.....	74
5.6	Concluding Remarks.....	75
CHAPTER 6 ARTICLE 3: SMART ORDER GENERATION IN DDMRP USING PREDICTED FLOW TIME.....		77
ABSTRACT.....		77
RESUME.....		78
6.1	Introduction.....	78
6.2	Related Literature.....	80
6.2.1	Context.....	80
6.2.2	Order Generation in DDMRP.....	82
6.2.3	Flow Time Prediction.....	83
6.2.4	Research Gap and Paper Contributions.....	84
6.3	Methodology.....	84
6.3.1	The Proposed Smart Order Generation Process.....	84
6.3.2	Traditional DDMRP Order Generation Processes.....	87
6.4	Application to a non-preemptive DDMRP flowshop.....	88
6.4.1	Case Study.....	88
6.4.2	Order Generation Policies.....	90
6.4.3	Flow Time Prediction.....	93
6.4.4	Experimentation.....	93
6.5	Results and Discussion.....	95
6.5.1	Delaying order generation if the system is overloaded.....	95
6.5.2	Delaying order generation if the system is underloaded.....	96
6.5.3	Delaying order generation if the system is overloaded or underloaded.....	98
6.5.4	Delaying order generation using CONWIP logic.....	99

6.5.5	Comparative Analysis of Order Generation Policies	101
6.6	Conclusion.....	103
CHAPTER 7 GENERAL DISCUSSION.....		104
7.1	Reminder of Research Objectives	104
7.2	Synthesis of Results	104
7.3	Theoretical Implications.....	105
7.4	Managerial Implications.....	107
7.5	Unexpected Findings.....	109
CHAPTER 8 CONCLUSION AND RESEARCH PERSPECTIVES		111
8.1	Synthesis of Contributions	112
8.2	Limitations of the Contributions	113
8.3	Research Perspectives	114
8.4	Personal Reflection	115
REFERENCES.....		116
APPENDICES.....		125

LIST OF TABLES

Table 4.1 Average Daily Usage (ADU) in units	41
Table 4.2 Processing Times in Seconds	42
Table 5.1 Average Daily Usage (ADU) in units	59
Table 5.2 Processing Times (seconds)	60
Table 5.3 DDMRP system and rush orders' characteristics.....	62
Table 5.4 LT levels according to the loading rate.....	64
Table 6.1 Average Daily Usage (ADU) in units	89
Table 6.2 Processing Times (seconds)	90
Table 6.3 Smart Order Generation policies description.....	91
Table 6.4 List of tested scenarios (SC)	94
Table A.1 Simulation Results for STATIC dispatching rule (95% CI)	125
Table A.2 Simulation Results for NFOHP dispatching rule (95% CI)	126
Table A.3 Simulation Results for DAOHP dispatching rule (95% CI)	127
Table B.1 Acronyms and their definitions	128
Table D.1 Detailed Feature Description.....	131
Table D.2 Machine Learning Algorithms Performance Comparison	133

LIST OF FIGURES

Figure 1.1 Self-Centered Demand Driven Execution	3
Figure 1.2 Proposed Status-Centered Demand Driven Execution	4
Figure 1.3 Thesis Structure	5
Figure 2.1 Decomposition of Manufacturing Control Tasks (Adapted from Lodding 2013).....	9
Figure 2.2 The five components of DDMRP (Ptak & Smith, 2019).....	12
Figure 2.3 DDMRP buffer zones	12
Figure 2.4 DDMRP Replenishment Order Generation and Sequencing steps.....	14
Figure 3.1 Positioning of contributions regarding each PCS function and Specific Objectives...	27
Figure 3.2 General Methodology	28
Figure 3.3 Representative Six-Machine DDMRP Flowshop	30
Figure 3.4 Generic Flowshop Model.....	31
Figure 4.1 DDMRP buffer zones and Net Flow Position (NFP)	37
Figure 4.2 Product X's NFP evolution over time (a) and relative priority evolution (b).....	39
Figure 4.3 The DDMRP flowshop managed by the Static dispatching rule	42
Figure 4.4 The DDMRP flowshop managed by the Net Flow On-Hand Positioning (NFOHP) dispatching rule	43
Figure 4.5 The DDMRP flowshop managed by the Dynamic Available On-Hand Positioning (DAOHP) dispatching rule.....	43
Figure 4.6 Impact of the Loading Rate on the Service Level and Average Inventory.....	45
Figure 4.7 Impact of the Loading Rate on Stockouts and Backorders.....	46
Figure 4.8 Impact of the Loading Rate on On-Hand Alerts Percent and MBOT	46
Figure 5.1 DDMRP buffer zones	54
Figure 5.2 Methodology Steps	57
Figure 5.3 The studied flowshop.....	58

Figure 5.4 Process Priority (PP) and Order Release (OR) Rush Order Prioritization policies	61
Figure 5.5 Evolution of the LT values according to the loading rate.....	64
Figure 5.6 Average Inventory Positioning (I) vs. Production Service Levels for non-rush orders (PSL _{nr}) and the whole system (PSL _a)	66
Figure 5.7 Average Inventory Positioning (I) vs. Customer Service Levels for non-rush orders (CSL _{nr}) and the whole system (CSL _a)	67
Figure 5.8 Rush Orders Production Service Level (PSL _r) Variation.....	69
Figure 5.9 Rush Orders Customer Service Level (CSL _r) Variation	69
Figure 5.10 Comparison of PP and OR rush order prioritization policies	72
Figure 6.1 DDMRP Buffers	81
Figure 6.2 Flow of the Smart Order Generation Process	85
Figure 6.3 Six Machine DDMRP Flowshop	89
Figure 6.4 Performance Comparison of WIP-Limiting Smart Order Generation Policies	96
Figure 6.5 Performance Comparison of Inventory-Pacing Smart Order Generation Policies	97
Figure 6.6 Performance Comparison of Zonal Smart Order Generation Policy	99
Figure 6.7 Performance Comparison of CONWIP Order Generation Policies.....	100
Figure 6.8 Comparison of order generation policies across different LT configurations	102
Figure C.1 Distribution of FT Observations with an 80% LR	130
Figure D.1 Distribution and Boxplot of the target variable (left) and Correlation Matrix for features and target (right).....	132
Figure D.2 Models R ² Comparison on train vs. test data (left) and MAE (right)	133
Figure D.3 Residual analysis plots for the XGBoost model	134
Figure D.4 SHAP feature importance chart (left) and summary plot (right)	135

LIST OF SYMBOLS AND ABBREVIATIONS

ADU	Average Daily Usage
ANN	Artificial Neural Network
BOM	Bill of Materials
BO	Backorders
BPR	Buffer Penetration Ratio
CONWIP	Constant Work In Process
CSL	Customer Service Level
CSLa	Average Customer Service Level
CSLnr	Customer Service Level (Non-Rush Orders)
CSLr	Customer Service Level (Rush Orders)
DDAE	Demand Driven Adaptive Enterprise
DDE	Demand Driven Execution
DDP	Demand Driven Planning
DDMRP	Demand Driven Material Requirements Planning
DDOM	Demand Driven Operating Model
DES	Discrete Event Simulation
DLT	Decoupled Lead Time
DRL	Deep Reinforcement Learning
EODD	Earliest Operation Due Date
FIFO	First Come First Served
FT	Flow Time
I	Average Inventory
IMR	Immediate Release

LR	Loading Rate
LT	Lead Time
LTF	Lead Time Factor
MAE	Mean Absolute Error
MBOT	Mean Backorder Fulfillment Time
MES	Manufacturing Execution System
ML	Machine Learning
MPS	Master Production Scheduling
MRP	Material Requirements Planning
NFE	Net Flow Equation
NFP	Net Flow Position
OHA	On-Hand Alert
OR	Order Release (Prioritization Policy)
PCS	Production Control System
pFT	Predicted Flow Time
PP	Process Priority (Prioritization Policy)
PSL	Production Service Level
PSLa	Average Production Service Level
PSLnr	Production Service Level (Non-Rush Orders)
PSLr	Production Service Level (Rush Orders)
RL	Reinforcement Learning
RMSE	Root Mean Square Error
SHAP	SHapley Additive exPlanations
SO	Stockouts

TOC	Theory of Constraints
TOG	Top of Green
TOR	Top of Red
TOY	Top of Yellow
VF	Variability Factor
VUCA	Volatility, Uncertainty, Complexity, and Ambiguity
WIP	Work-In-Process
WLC	Workload Control
α	Lower Threshold Factor (Smart Order Generation)
β	Upper Threshold Factor (Smart Order Generation)

LIST OF APPENDICES

APPENDIX A Simulation Results	125
APPENDIX B Acronyms and their definitions.....	128
APPENDIX C Lead Time Calibration using surrogate model	129
APPENDIX D Machine Learning Algorithm Selection for Predictive Analytics	131

CHAPTER 1 INTRODUCTION

1.1 Context

In the contemporary manufacturing landscape, characterized by volatility, uncertainty, complexity and ambiguity (VUCA), traditional Production Control Systems (PCSs) often fail to align inventory levels with actual market demand. Material Requirements Planning (MRP) systems, the dominant paradigm for decades, struggle in modern supply chains due to their reliance on forecast accuracy. This dependency frequently leads to a misalignment between inventory levels and actual consumption (Ptak & Smith, 2019).

Since the introduction of MRP, new technologies and theories on operations management have been developed, shaping the configuration of supply chains. However, the fundamental logic of MRP has remained largely unchanged: linking the planning and scheduling of dependent demand items directly to the demand of parent items (Miclo et al., 2018). In VUCA environments, these rigid dependencies frequently result in a “bi-modal distribution” of inventory, characterized by simultaneous stockouts of critical parts and overstocks of non-critical items (Ptak & Smith, 2019).

To counter these limitations, the Demand Driven Operating Model (DDOM) was proposed as a methodology to decouple supply from demand variability through strategic positioning of inventory buffers. The operational engine of this model, Demand Driven Material Requirements Planning (DDMRP), integrates aspects of MRP, Lean, and the Theory of Constraints (TOC) to shift production from a push-based forecast system to a pull-based consumption system (Ptak & Smith, 2018). Literature shows that DDMRP offers significant improvements when compared to traditional MRP (El Marzougui et al., 2024; Miclo et al., 2018), especially when managing variability (Butturi et al., 2021), optimizing inventory (Math et al., 2024), compressing lead times (Shofa et al., 2018; Velasco Acosta et al., 2019), and improving service levels (Kortabarria et al., 2018).

1.2 Research Problem

Despite the documented benefits of DDMRP, existing research has predominantly focused on the planning phase of the methodology (Dessevre et al., 2021; Dessevre et al., 2023; Martin et al., 2023). Conversely, the execution phase remains controlled by self-centered heuristics that do not account for real-time shop floor status. This creates a dichotomy where planning adapts to real-

time market signals, while execution relies on self-centered planning parameters ignoring shop floor state. From an operational perspective, the reliance on planning metrics for execution introduces a latency error. Orders and priorities derived from the Net Flow Position (NFP) represent a static snapshot taken at the moment of order generation, which validity decays as the time passes and orders progress through the shop floor. To analyze this structural limitation, this thesis adopts the functional decomposition of PCSs. According to Lödding (2013), a robust PCS must address four control tasks: Order Generation, Order Release, Sequencing, and Capacity Control. Lödding (2013) also distinguishes between open-loop and closed-loop approaches in PCSs. In open loop control the decisions are made based on planned values (e.g., production schedules, material requirements) without feedback from the actual shop floor. In closed loop control the decisions are continuously adjusted based on feedback from actual production performance (e.g. WIP levels, backlog, throughput). Modern manufacturing increasingly requires closed-loop mechanisms to cope with the VUCA environment.

When considering the functional tasks of PCSs, a distinction must be drawn between shop floor load and output potential: the first three tasks (i.e., sequencing, order release and order generation) are responsible for managing the demand side (the load) on the shop floor, while Capacity Management module controls the supply side (the output potential). When standard DDMRP proposed by Ptak and Smith (2019) is mapped against this framework, a critical research gap emerges specifically within the demand-side execution tasks.

- Sequencing: Order prioritization is determined by the planning metric Net Flow Position (NFP) and the Buffer Penetration Ratio (BPR), a metric that reflects the urgency of an order based on the depletion of the safety zone (i.e., the red zone). These metrics do not reflect real-time material availability (i.e., On-Hand Availability) or the status of the shop floor.
- Order Release: Once generated, orders are subject to an ‘Immediate Release’ (IMR) policy. This assumes infinite shop floor capacity and fails to account for disruptions caused by heterogeneous orders streams.
- Order Generation: Standard DDMRP relies on the NFP falling below a static Top of Yellow (TOY) threshold. This decision is binary and inventory-centric, ignoring the current workload or congestion of the production system.

The reliance on these self-centered rules limits DDMRP's capacity to adapt to dynamic environments. Figure 1.1 illustrates the functional mapping of the standard DDMRP methodology. The system begins with the Demand Driven Planning (DDP), which functions as a closed-loop system (represented by green dashed arrows). The DDP receives demand information directly from the client to continuously adjust buffer dimensioning. This planning phase generates the NFP signal, which encapsulates the information of available stock (On-Hand) and pending demand (On-Order and Qualified Demand).

The information carried by the NFP signal is what governs the Demand Driven Execution (DDE) phase, dictating order generation, order release, and sequencing. However, as shown in Figure 1.1, the execution phase operates as an open-loop system (represented by red dashed arrows). Once the NFP signal is received, the standard execution heuristics rely solely on these internal planning calculations. In this thesis, we define these standard DDMRP heuristics as 'Self-Centered' because they focus exclusively on the status of their own inventory buffers, completely ignoring the real-time physical status, capacity constraints, or congestion of the shop floor.

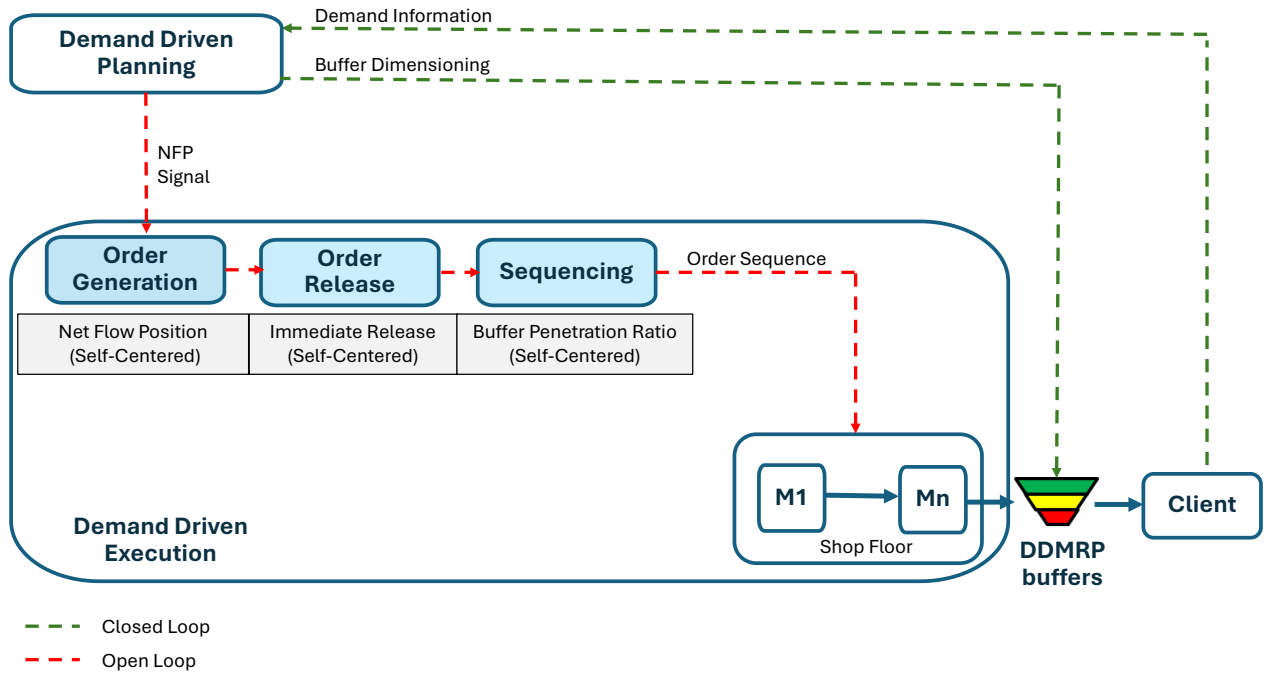


Figure 1.1 Self-Centered Demand Driven Execution

To overcome this structural limitation, this thesis proposes a Status-Centered Demand Driven Execution system illustrated in Figure 1.2. The fundamental difference lies in the establishment of

three distinct feedback loops that transmit real-time shop floor status back to the execution control tasks. To enrich the Demand Driven Execution (DDE) phase, this thesis proposes to explore three new streams of information:

1. Considering the real-time On-Hand Availability to ensure the sequencing reflects actual physical inventory risks, and thus, moves from a static heuristic in favor of dynamic dispatching.
2. Considering order heterogeneity in the order release task to implement a gating mechanism to manage heterogeneous flow of orders to increase flowshop stability and performance.
3. Consider predicted flow time when generating a replenishment order. Thus, the order generation task moves beyond the static NFP trigger through integration of the predicted flow time, creating a dynamic order approval gate that controls flowshop occupation to reduce inventory while maintaining performance.

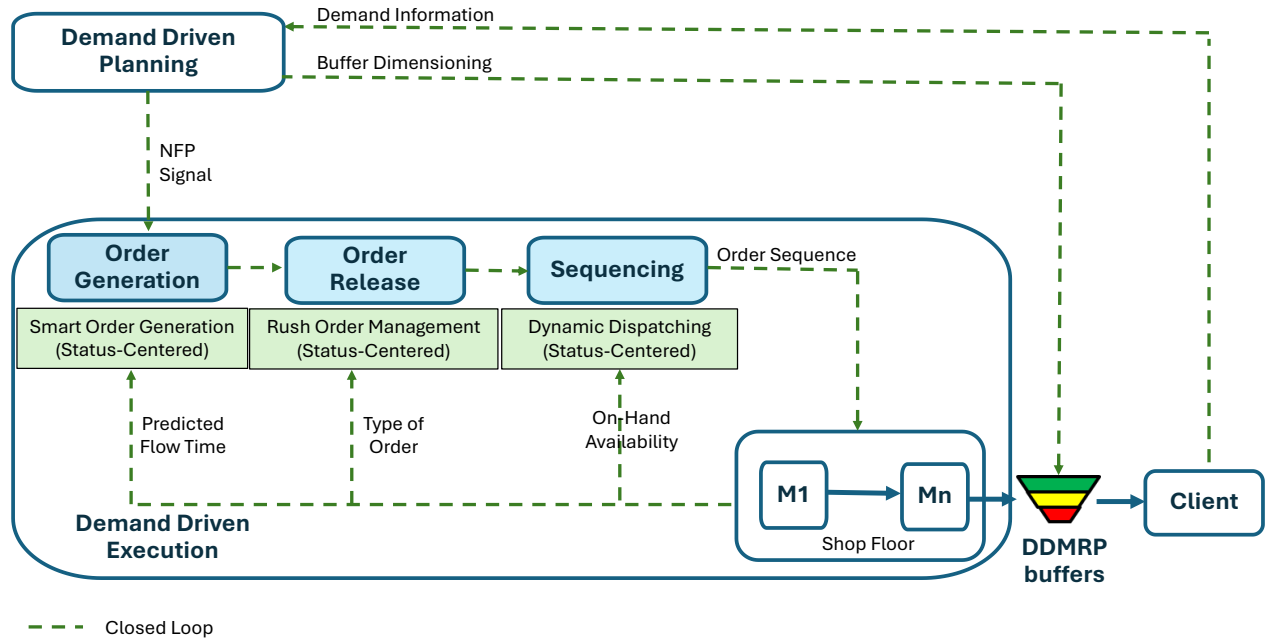


Figure 1.2 Proposed Status-Centered Demand Driven Execution

1.3 Thesis Structure

This thesis is organized into eight chapters, structured according to the continuous research cycle illustrated in Figure 1.3. Chapter 1 sets the industrial context and defines the core research problem

surrounding the static nature of DDMRP execution. Chapter 2 presents a comprehensive literature review establishing the theoretical foundation using the PCSs framework, ultimately isolating the specific research gaps addressed.

Based on this theoretical foundation, Chapter 3 formally defines the research objectives. It then justifies the methodological choices, specifically the use of Discrete Event Simulation (DES), and outlines the generic case study used as the experimental testbed for the subsequent contributions.

The core contributions are presented in Chapters 4, 5 and 6. These chapters consist of three independent research articles, encompassing modeling, experimentation, and detailed result analysis for the proposed mechanisms for sequencing, order release and order generation.

Following the contributions, Chapter 7 provides a general discussion that synthesizes the outcomes across all execution tasks to evaluate the shifting to the ‘Status-Centered Demand Driven Execution’ system. Finally, Chapter 8 delivers the conclusion and recommendations, closing the research loop by demonstrating how the insights result in feedback reshaping the initial context.

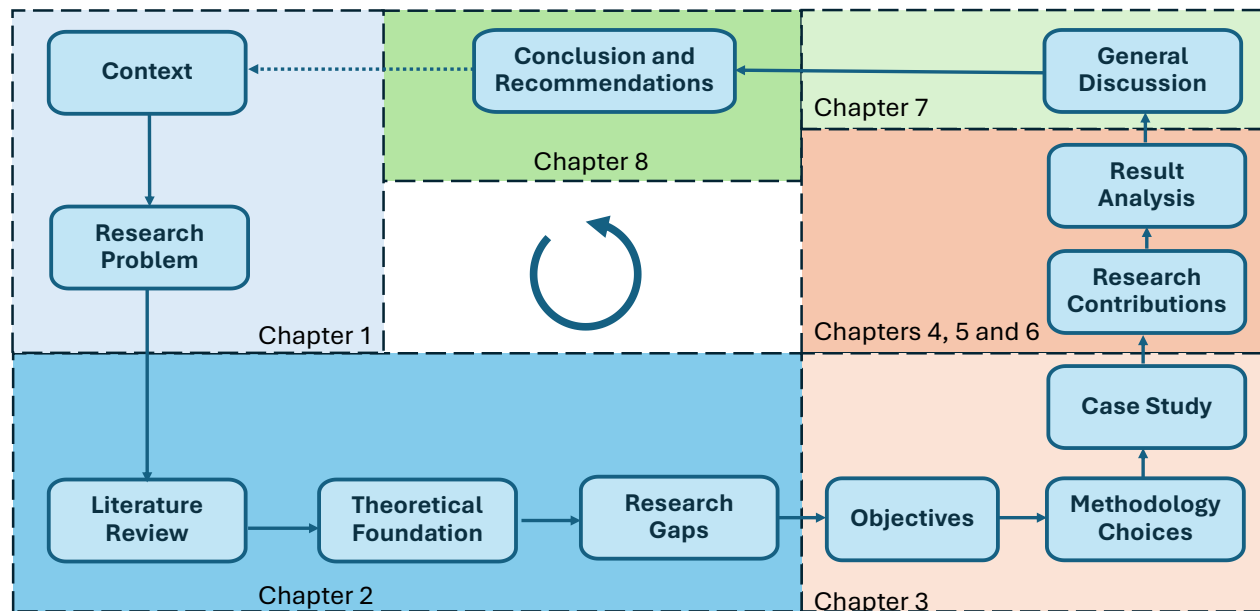


Figure 1.3 Thesis Structure

CHAPTER 2 LITERATURE REVIEW

This chapter establishes the theoretical and empirical foundations of the research. It is structured to systematically map widely accepted principles of Production Control Systems (PCSs) against the specific operational mechanics of Demand Driven Material Requirements Planning (DDMRP). The review is organized in seven main sections. Section 2.1 delineates the theoretical framework of PCSs, explicitly defining the four primary control tasks as proposed by Lodding (2013). Section 2.2 summarizes DDMRP principles and mechanics and synthesizes the current state of literature, contrasting the extensive research on the planning phase with the limited attention rendered to the execution phase. Section 2.3 analyzes the functioning of DDMRP through the theoretical lens of PCSs, identifying the structural dichotomy between the closed-loop planning and the open-loop execution. This section also deconstructs the execution phase across the control tasks that manage load: Sequencing, Order Release and Order Generation. For each task, the analysis identifies the specific limitations of standard DDMRP heuristics and presents production research domains aiming to address these limitations. Section 2.4 summarizes the research gaps that justify the objectives of the contributions presented in the subsequent chapters. Finally, Section 2.5 concludes the literature review establishing the main findings that will be used throughout the contributions.

2.1 Theoretical Foundations of Production Control Systems

PCSs constitute the information flow and decision logic that govern physical material flow within a factory (Karrer et al., 2012). Beyond simple execution, the functional scope of a PCS joins material requirements planning, demand management, capacity planning, and the detailed scheduling and sequencing of jobs. The primary objective of these functions are the reduction of Work-In-Process (WIP), the minimization of shop floor throughput times, the optimization of stockholding costs, and the improvement of delivery date adherence (Stevenson et al., 2005).

Given its status as a critical success factor for industrial organizations, the domain of production control has evolved significantly, resulting in the emergence of diverse PCSs methodologies over the last decades (Bagni et al., 2020). However, the effectiveness of any specific PCS is intrinsically linked to the environment in which it operates. As noted by Hopp and Spearman (2011), a simple flow line may function effectively with rudimentary planning tools, whereas a complex job shop can present significant operational challenges even when managed with sophisticated control logic

To categorize these systems, literature often distinguishes them based on their primary control variable. Fundamentally, a PCS can strictly control either WIP or Throughput (Bagni et al., 2020).

2.1.1 Throughput-Oriented vs. WIP-Oriented Control

As established by Hopp and Spearman (2011), a PCS can exert direct control over either Throughput (push) or WIP (pull), but cannot independently dictate both simultaneously. This research classifies PCSs architectures based on this primary control orientation.

In Throughput-Oriented Systems, production is driven by a schedule that specifies what must be produced and when. The defining characteristic of this approach is that throughput is the independent control variable, while WIP levels are observed as a dependent outcome (Hopp & Spearman, 2011). The most prominent examples of this logic are Material Requirements Planning (MRP), MRP II and Basestock systems (Lödding et al., 2003).

The operational core of these systems relies on the explosion of the Bill of Materials (BOM) against a Master Production Schedule (MPS) to calculate required lot sizes and release dates (Thevenin et al., 2020). However, this logic typically assumes infinite capacity and deterministic lead times (Stevenson et al., 2005). Because the system releases work based on a schedule rather than the actual status of the shop floor, it lacks a mechanism to restrict WIP accumulation during capacity constraints. Consequently, mismatches between the forecast-driven plan and execution reality often necessitate frequent rescheduling. This responsiveness comes at the cost of “system nervousness”, where minor changes in the MPS amplify into significant disruptions in planned order releases (Chu & Hayya, 1988). While safety stocks and safety lead times are often employed to buffer this uncertainty (Guide & Srivastava, 2000), they do not address the lack of WIP control.

In contrast, WIP-Oriented Systems invert this logic: the WIP level is the controlled independent variable, while throughput is the observed outcome. Instead of releasing orders based on a scheduled date, these systems authorize production only when the system status permits, effectively coupling the start to the completion of previous work (Hopp & Spearman, 2011).

The primary mechanism for this control is Workload Control (WLC). Although specific implementations vary, the core of WLC is the use of a pre-shop pool to hold unreleased orders, thereby decoupling order generation from order release (Hutter et al., 2017). The theoretical justification for this approach is grounded in Little’s Law (Equation 2.1), which demonstrates that

for a given throughput rate, the cycle time of a system is directly proportional to the WIP level (Hopp & Spearman, 2011). By strictly limiting the workload in the production units, these systems aim to reduce queue times, stabilize flow, and prevent the operational constraints associated with flooded manufacturing lines (Graves et al., 1995). Prominent examples of this category include Kanban, Constant Work in Process (CONWIP), and POLCA (Stevenson et al., 2005).

$$Cycle\ Time = \frac{WIP}{Throughput} \quad (2.1)$$

Recent reviews of the PCSs landscape indicate a significant shift toward this paradigm. Bagni et al. (2020) identified that among PCSs emerging in the last two decades, the vast majority (11 out of 13) utilize WIP as the primary control variable. However, while decentralized WIP-oriented systems like Kanban perform exceptionally well in repetitive flow environments, Lödding (2013) note that few are adequately designed for environments characterized by high variety and complex routing. This limitation suggests a need for hybrid or advanced control mechanisms in environments that require both the planning visibility of throughput-oriented systems and the flow stability of WIP-oriented control.

2.1.2 The Functional Decomposition of Production Control Systems

While the distinction between Throughput-Oriented and WIP-Oriented systems describes the trigger mechanism of material flow, there are specific operational decisions that govern shop floor behavior. To provide this structural lens, this thesis adopts the functional decomposition of PCSs. According to Lödding (2013) the performance of a manufacturing system is not the result of a single decision, but the interaction of four distinct, yet interconnected, manufacturing control tasks: Order Generation, Order Release, Sequencing and Capacity Control. Each task functions as a specific control loop with distinct inputs, decision horizons, and impacts on the performance. Figure 2.1 is a representation of the manufacturing control model, connecting the manufacturing control tasks, to its actuating variables, control variables and objectives.

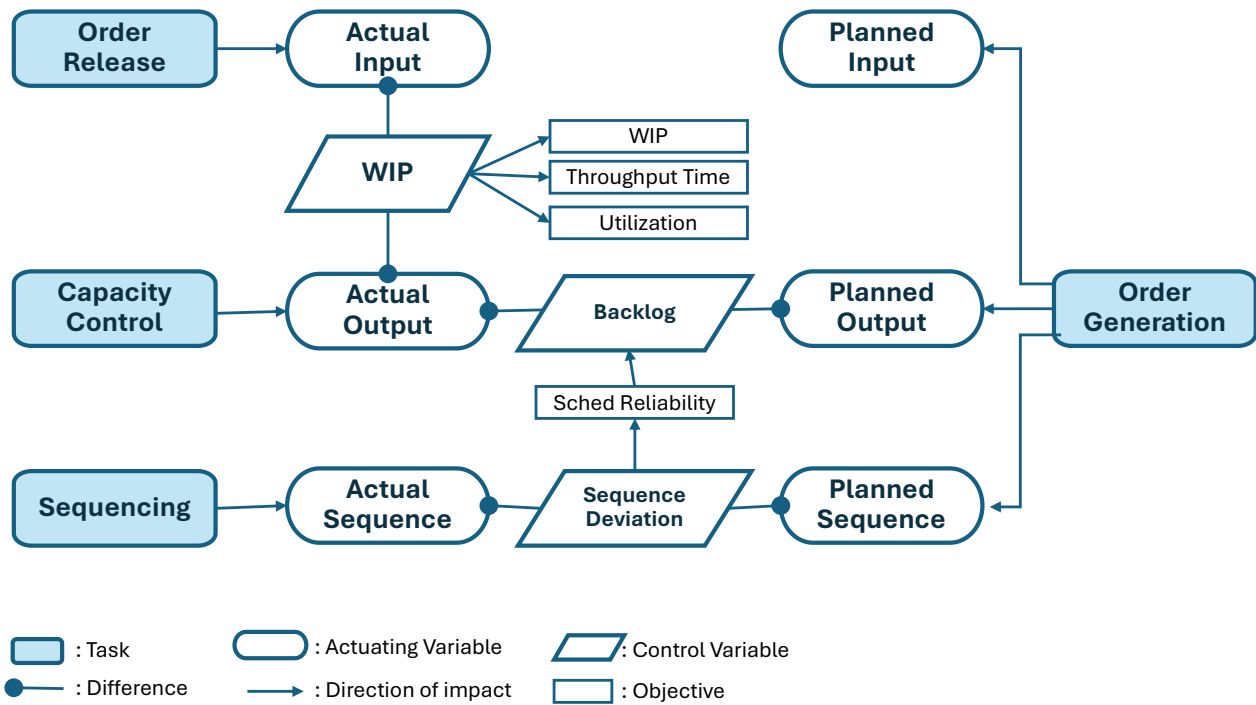


Figure 2.1 Decomposition of Manufacturing Control Tasks (Adapted from Lodding 2013)

Order Generation (Load Creation): Order generation constitutes the interface between external market (or forecasting modules) and the internal production planning system. Its primary function is to transform actual or forecasted demand into manufacturing orders. Lödding (2013) identifies this task as the determinant of the Planned Load. If the rate of order generation exceeds the system's theoretical capacity over time, the inevitable result is a buildup of backlog, regardless of the efficiency. In this sense, order generation sets the upper bound of the system performance but does not control the actual flow dynamics in the shop floor. Although generating orders is a production planning task. In make-to-stock productions the order generation is to some extent conducted using very simple procedures (e.g. DDMRP). These methods, especially because of how quickly decisions must be made, are frequently classified as a manufacturing control method, although they carry out planning tasks. Despite not explicitly determining planned values, based on the logic of the procedure, it is possible to derive useful planning values (Lödding, 2013).

Order Release (Load Limiting): Distinct from generation, Order Release acts as the 'gatekeeper' between the planning domain and the physical execution domain. While generation creates the need for a product, release authorizes the beginning of work and the allocation of resources. In WLC literature, this function is identified as the primary mechanism for stabilizing WIP and

throughput times (Land & Gaalman, 1996). By decoupling the generation from the release, the system can utilize a ‘pre-shop pool’ to dampen demand variability with an effective order release being superior to complex dispatching rules for controlling lead times, as it prevents the flooding of the shop that degrades system performance (Thürer et al., 2012). The Order Release controls the WIP by interacting with Capacity Control. Whenever there is a mismatch between the actual output defined by order release and the planned output defined by capacity control, the WIP increases or decreases.

Sequencing (Flow Optimization): Once an order is released and enters the physical system, it joins a queue at a specific resource. Sequencing (often referred to as dispatching) determines the processing rank of these waiting jobs. While Order Release controls the global load and average throughput times, Sequencing manages the local flow efficiency and the schedule reliability of individual orders (Pinedo, 2022). This task involves a constant trade-off: minimizing changeovers by grouping similar orders (efficiency) versus prioritizing orders with approaching due dates (reliability). In dynamic environments, the efficacy of sequencing relies on its ability to utilize real-time shop floor status rather than static planning priorities (Haupt, 1989).

Capacity Control (Potential Adjustment): Unlike the previous three tasks, which manage the demand side (the load), Capacity Controls manage the supply (the output potential). This task involves the medium-term adjustment of capacity to match the released workload. Distinguishing from long-term Capacity Planning (i.e. acquiring new machines), Capacity Control refers to the operational flexibility measures, such as authorizing overtime, reallocating cross-trained workforce, or outsourcing specific operations. Ideally, capacity control acts as a feedback loop where the current WIP level triggers a temporary adjustment in processing power to maintain flow stability (Lödding, 2013).

2.1.3 The Emergence of Hybrid Control Architectures

The binary classification of PCSs architectures reveals inherent limitations when applied to VUCA environments. While throughput-oriented systems suffer from ‘system nervousness’ due to forecast errors, pure WIP-oriented systems often lack the planning visibility required for high variability scenarios. To reconcile these constraints, hybrid architectures have emerged, aiming to combine the global visibility and flow stability. Lödding (2013) expands on how the PCSs must evolve from open-loop to closed-loop approaches to deal with uncertainty. The open loop approaches are

essentially planning tools, while closed loop approaches are true control systems. Considering hybrid systems where the PCSs tasks are governed by the execution phase (e.g., DDMRP), relying on open loop approaches for execution often leads to mismatches, causing lateness or excess WIP.

DDMRP positions itself as an evolution of traditional methods through the use of this hybrid approach. By strategically positioning inventory buffers to decouple supply from demand variability, DDMRP theoretically enables a pull-based generation of orders within a push-based planning framework, offering a structural solution to the operational constraints of traditional PCSs. However, the validation of DDMRP as a robust hybrid system requires analysis through the PCSs control functions established in Section 2.1.2. Specifically, it requires determining if DDMRP applies the feedback loops of WIP-oriented control to its execution tasks, or if it merely relies on self-centered parameters derived from its throughput-oriented planning structure.

2.2 DDMRP Principles and Mechanics (Ptak & Smith, 2019)

To address the limitations of traditional push and pull systems identified in the previous sections, the Demand Driven Operating Model (DDOM) was proposed as a comprehensive methodology to align production with actual market requirements. The operational engine of this model, DDMRP, integrates aspects of MRP, Lean Manufacturing, and the Theory of Constraints (TOC) to effectively decouple supply from demand variability. Fundamentally, the objective of DDMRP is to construct a robust supply chain design capable of mitigating the bullwhip effect through enhanced visibility and responsiveness, thereby improving performance.

DDMRP methodology is structured around five sequential components that define the systems configuration and operation, as shown in Figure 2.2. The first three components establish the model's static and dynamic parameters. Strategic inventory positioning will determine where the decoupling points are placed. Buffer profiles and levels will determine the amount of protection at those decoupling points. Dynamic adjustments define how that level of protection flexes up or down based on operating parameters, market changes, and planned or known future events.

The fourth and fifth elements define the actual operational aspects of a DDMRP system: Demand Driven Planning (DDP) which generates supply orders, and Demand Driven Execution (DDE), which manages open orders to ensure synchronization.

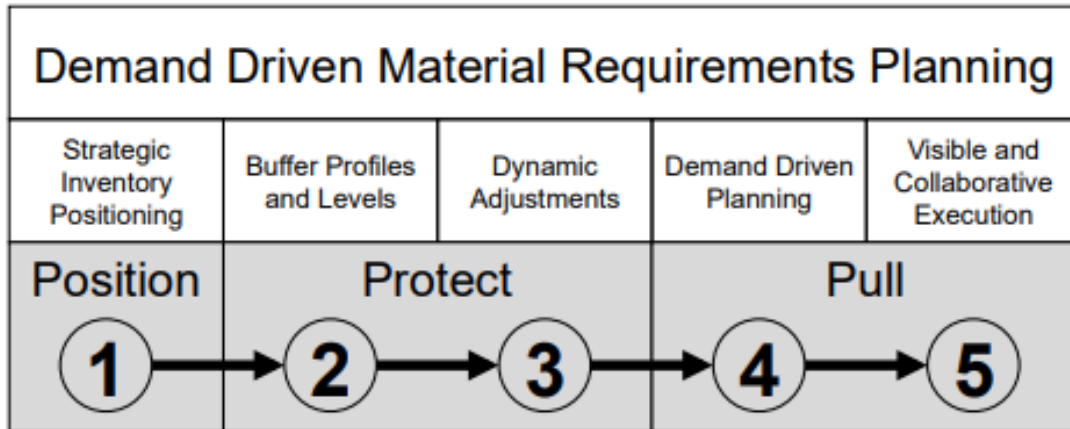


Figure 2.2 The five components of DDMRP (Ptak & Smith, 2019)

In DDMRP, the buffers serve a dual theoretical function: acting as a decoupling point to break the propagation of variability, damping the bullwhip effect, and as a control mechanism for order generation. Unlike traditional safety stocks, which are static buffers against forecast error, DDMRP buffers are dynamic stocks composed of three functional zones: (1) The Green Zone which determines the frequency and size of order generation; (2) The Yellow Zone which covers the expected demand during the lead time; and (3) The Red Zone which provides safety against variability in demand and supply (Ptak & Smith, 2019). This structure attempts to operationalize the positioning logic of the TOC within the calculation logic of MRP. Figure 2.3 is a representation of a DDMRP buffer.

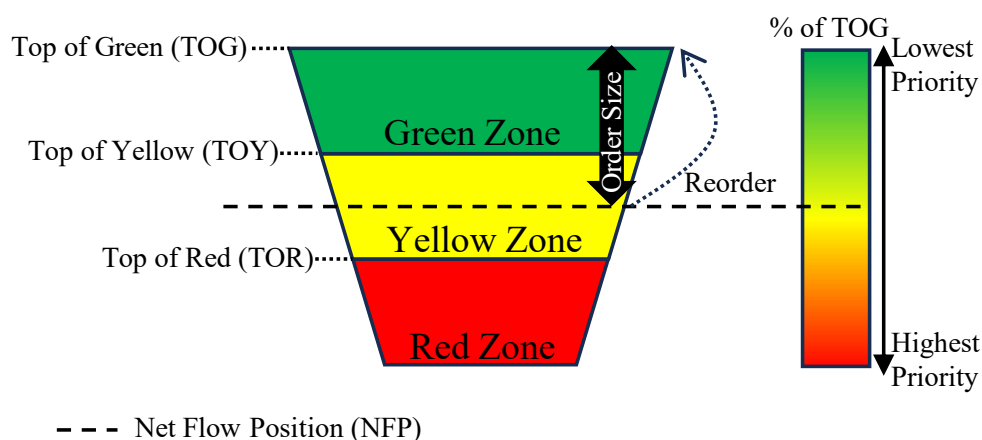


Figure 2.3 DDMRP buffer zones

The buffer zones are proportional to the Average Daily Usage (ADU) and the Decoupled Lead Time (DLT) – or the Lead Time (LT) for non-decoupled product structures. The variability of both the ADU and the LT are considered in terms of the Lead Time Factor (LTF) and Variability Factor (VF). The standard calculation logic is expressed in Equations 2.1 through 2.5:

$$\text{Red Base} = \text{ADU} \times \text{LT} \times \text{LTF} \quad (2.1)$$

$$\text{Red Safety} = \text{ADU} \times \text{VF} \quad (2.2)$$

$$\text{Red Zone} = \text{Red Base} + \text{Red Safety} \quad (2.3)$$

$$\text{Yellow Zone} = \text{ADU} \times \text{LT} \quad (2.4)$$

$$\text{Green Zone} = \text{ADU} \times \text{LT} \times \text{LTF} \quad (2.5)$$

Consequently, the upper boundaries (Top) of each zone, which trigger specific operational behaviors (Eqs. 2.6 to 2.8), are defined as:

$$\text{Top of Red (TOR)} = \text{Red Zone} \quad (2.6)$$

$$\text{Top of Yellow (TOY)} = \text{TOR} + \text{Yellow Zone} \quad (2.7)$$

$$\text{Top of Green (TOG)} = \text{TOY} + \text{Green Zone} \quad (2.8)$$

The order generation and sequencing in DDMRP is governed by the Net Flow Position (NFP) which generates a supply order signal (time and quantity) for buffer replenishment (Ptak & Smith, 2019). The DDP process continuously calculates the NFP for each decoupled part using Equation 2.9, incorporating shop floor information regarding qualified spikes, orders due today, and backlogs. A replenishment signal is generated whenever the NFP drops below the Top of Yellow (TOY). The quantity requested is calculated to restore the NFP to the Top of Green (TOG). Figure 2.4 represents the replenishment order generation and sequencing steps for DDMRP.

$$\text{Net Flow Position (NFP)} = \text{On - Hand} + \text{On - Order} - \text{Qualified Demand} \quad (2.9)$$

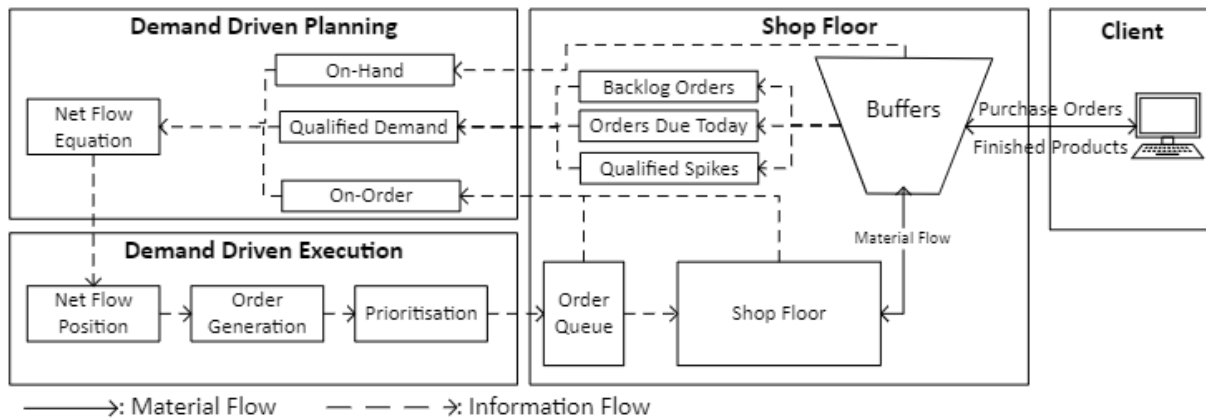


Figure 2.4 DDMRP Replenishment Order Generation and Sequencing steps

It is critical to distinguish the nature of the NFP from the physical On-Hand inventory. The NFP is a flow capability metric, which includes stock that does not yet exist physically (On-Order) to ensure the system keeps moving. Conversely, the On-Hand is an availability metric, representing what is available to process. Once an order is generated, the prioritization of these orders on the shop floor (sequencing) is derived from the buffer status. Priority is assigned based on the Buffer Penetration Ratio (BPR) or the percentage of the NFP relative to the TOG, where orders with a lower BPR have a higher priority. However, this creates a theoretical conflict: prioritizing based on NFP optimizes for future flow, potentially at the expense of immediate availability and shop floor efficiency.

Despite the methodological definition of distinct phases for Planning and Execution, research has not been evenly conducted across both areas. A review of DDMRP literature reveals a significant asymmetry: while the planning aspects of DDMRP such as buffer sizing (Azzamouri et al., 2022; Martin et al., 2023), positioning (Abdelhalim et al., 2021; Bayard, Stephanie et al., 2021; Habbadi et al., 2025) and parameter optimization (Duhem et al., 2023; Lahrichi et al., 2022; Werth et al., 2023) are subject of numerous research, the operational rules for managing orders at the execution level remain largely heuristic and self-centered.

This creates a dichotomy where the planning is dynamic and highly researched, while the execution is governed by static rules such as Immediate Release (IMR) and BPR sequencing. As highlighted in the previous analysis of PCSs, these static mechanisms may be insufficient in VUCA environments. The next section brings the contrast between the predominance of planning-focused

DDMRP research with the lack of execution-focused DDMRP research, effectively mapping the research gap that this thesis aims to address.

2.2.1 Evolution of DDMRP Research: A Bias Toward Planning

Since the formalization of the DDMRP concept by Ptak and Smith (2011), the academic exploration of DDMRP has evolved significantly. As the literature expanded, research efforts shifted from proving the concept's viability to challenging its internal mechanics. This progression can be divided into two streams of research: a first stream dedicated to comparative validation against traditional systems, followed by a second stream focused on improving the method itself.

The initial research efforts were primarily concerned with demonstrating the superiority of DDMRP over traditional MRP in industrial applications. Miclo et al. (2015) used Discrete Event Simulation (DES) to quantitatively compare the two approaches, demonstrating that DDMRP yielded higher service levels and lower inventory positions. Subsequent studies confirmed these findings across various contexts: Kortabarria et al. (2018) and Shofa et al. (2018) validated the methodology in standard flow environments, while Velasco Acosta et al. (2019) demonstrated its applicability in complex manufacturing environments with complex BOMs. Thürer et al. (2020) further assessed performance in multi-stage assembly systems, identifying that while DDMRP mitigates bottlenecks, the positioning of buffers requires careful trade-off analysis of WIP levels.

As the efficacy of the model was established, the second stream of research aimed to challenge the subjectivity of the original rules and propose robust mathematical alternatives. This stream focused on the 'Equation' side of DDMRP. Lee and Rim (2019) proposed mathematical models to dimension safety stocks, replacing the standard heuristic with an optimized formula. Similarly, Damand et al. (2022a) applied genetic algorithms to determine the optimal values for LTF and VF, arguing that standard guidelines failed to provide precise parameters. Dessevre, Martin, et al. (2019) introduced models to parametrize buffers dynamically using the workshop state, demonstrating that adjusting the DLT could reduce stock levels without compromising service. Abdelhalim et al. (2021) addresses the complexity of decoupling point placement, describing optimization models for buffer allocation in hybrid MTO/MTS systems.

While these two streams represent significant theoretical advancement, critical analysis reveals a persistent structural bias. Both the streams operate almost exclusively within the DDP domain. They seek to optimize the size of the order generated and how large the buffer should be.

Conversely, the DDE domain, specifically order release and sequencing, remains largely governed by self-centered heuristics. The operational reality of how these optimized orders are managed once they enter a constrained shop floor has received minimal attention.

Critiques of this limitation have begun to emerge. Baptiste (2018) noted that under current DDMRP rules, it is impossible to anticipate or delay an order based on the observed loading level of the workshop. Thürer et al. (2020) observed that while DDMRP mitigates bottlenecks, it can increase WIP levels due to the lack of workload control mechanisms. Furthermore, Dessevre et al. (2021) demonstrated that manipulating planning parameters (e.g. DLT) is an inefficient way to control flow times, suggesting that capacity management and execution control are more effective levers.

Despite these observations, standard DDMRP execution still relies on self-centered heuristics, such as Immediate Release (IMR) for order release and the BPR for order sequencing. There is a scarcity of research investigating how these static execution rules perform under high-variability conditions. Few study considering execution aspects often consider homogeneous customer orders and preemptive flowshops (Fernandes et al., 2023; Fernandes et al., 2024).

This thesis posits that the academic exploration of DMRP has reached a saturation point regarding parameter optimization. As noted by Dessevre et al. (2021), manipulating planning parameters is an inefficient method for controlling shop floor flow times. Therefore, a ‘third stream’ of DDMRP research is required, shifting the focus from the generation of the plan to the dynamics of its execution. This stream must move beyond static validation and mathematical optimization to develop a status-centered execution system that allow the DDMRP execution engine to adapt to the stochastic reality of the shop floor.

2.3 DDMRP Through the Lens of PCSs Theory

The task of evaluating the structural robustness of DDMRP requires the mapping of its operational mechanics against the fundamental PCSs tasks. A robust PCS functions as a closed control loop, where feedback constantly adjusts the input variables to stabilize performance against stochastic variability (Lödding, 2013).

When analyzing DDMRP through this framework, a structural paradox appears. The DDP phase theoretically constitutes a closed control loop: the buffer profiles (input) are dynamically adjusted based on the ADU and LT characteristics (feedback). However, despite being responsive to market

changes, the planning engine remains blind to the stochastic variability of the shop floor. The signal to order generation is based solely on inventory status relative to demand, without receiving feedback on the actual status of production, capacity or congestion. Consequently, the NFP signal to order generation ignores capacity constraints, assuming infinite capacity, relying on the execution phase to manage the load, potentially overloading the system during periods of high demand variability.

Furthermore, the DDE phase lacks a dynamic feedback mechanism, operating as an open loop system, based on self-centered heuristics, without receiving feedback to adjust execution. For the order generation, it receives the signal from the planning phase, ignoring the status of the shop floor. Considering the order release stage, once an order is generated, current DDMRP research relies on an immediate release for all orders. A system that releases work without an explicit limit on WIP is mathematically a push system (Hopp & Spearman, 2011), exposing the shop floor to a degradation of cycle times as utilization increase, as demonstrated by Dessevre et al. (2021).

Finally, some of the execution heuristics (specifically BPR and NFP for sequencing) introduce a critical latency error. The execution priority is derived from the NFP calculated at the moment of order generation. In a stochastic environment, the validity of this information decays as the orders traverses the routing. Relying on a self-centered, planning-derived metric to make real-time dispatching constitutes an open-loop control failure. The system fails to update the priority based on real-time state changes, leading to a misalignment between the planned urgency and the actual needs of the shop floor.

The following sections deconstruct this open-loop paradigm across the control tasks that manage the shop load. The analysis proceeds from the local machine level (sequencing) to the flow control level (order release) and finally to the system-loading level (order generation).

2.3.1 Task 1: Sequencing and Dynamic Dispatching

Sequencing, often referred to as dispatching, is the control task responsible for selecting the next job to be processed from a queue at a specific resource (Blackstone et al., 1981). The objective of sequencing is to ensure schedule reliability by prioritizing orders based on their urgency relative to the current system state (Lödding, 2013). Unlike higher-level planning functions that manage aggregate volumes, dispatching operates at the local resource level, determining the specific sequence of execution based on the state of the shop floor (Hopp & Spearman, 2011).

While dispatching and scheduling are often discussed interchangeably, they represent fundamentally different approaches to shop floor control. Scheduling establishes the timelines, delivery, and availability of manufacturing processes, whether they are personnel, inventory, or capital, offering a guide as to how the manufacturing will progress prior to execution (Lam & Ip, 2011). In contrast, dispatching is a dynamic decision-making process that occurs during execution (Hopp & Spearman, 2011). The difficulty of the dispatching problem arises from the fact that, first, given n orders queued at a workstation, there are $n!$ ways to sequence those jobs and, second, some shop conditions at another workstation might influence the optimal sequence of orders at the present workstation (Blackstone et al., 1981).

Dispatching by nature can be comprised of very simple or extremely complex priority rules. Some simple rules are ‘Select a job at random’, or ‘Select the job that has been waiting for the longest’ (Blackstone et al., 1981). These rules may be used at order release and order sequencing. Priority rules at order sequencing may interact with priority rules at order release (also known as the pool sequencing rules). While dispatching rules defines the sequence by which orders are processed at the machines, pool sequencing rules defines the sequences by which orders are considered for release to the shop floor each time order release is triggered (Gómez Paredes et al., 2020).

In the standard DDMRP methodology, order prioritization is derived directly from the buffer status calculated during the planning phase, in the moment of order generation. Ptak and Smith (2019) define priority based on the BPR, which is a function of the NFP relative to the TOG. The lower the percentage of the NFP within the buffer, the higher the execution priority. This approach assumes that the urgency of replenishing inventory (planning) is identical to the urgency of processing the order on the shop floor (execution).

Literature traditionally classifies dispatching heuristics into two categories: simple (static) rules and complex (dynamic) rules. Simple rules, such as First-In-First-Out (FIFO) or Earliest Operation Due Date (EODD), prioritize jobs based on a single attribute that remains constant or changes linearly with time. Studies indicate that these single-attribute rules are often ineffective in optimizing global shop performance because they fail to account for the wider state of the system (Barman & Lisboa, 2010). To address these limitations, the dynamic dispatching approach uses state-dependent variables, such as queue lengths, real-time WIP levels, or actual material availability to determine sequencing (Haupt, 1989).

From a production control perspective, relying on the NFP for shop floor sequencing introduces a critical latency error. As shown in Equation 2.9, the NFP includes the On-Order quantity (i.e. stock that has been ordered but not yet received). Consequently, a buffer may have a high NFP (low priority) due to a large volume of open supply orders, even if the physical On-Hand inventory is critically low. In a stochastic flowshop, this disconnection can lead to inefficiencies in the sequencing decision: an order needed immediately to prevent a stockout may be deprioritized in favor of an order for a buffer that is physically stable but has a lower NFP (Alves et al., 2024).

Recent literature in DDMRP has begun to question the efficacy of static, planning-based sequencing in dynamic environments. While some studies suggest that NFP-based sequencing outperforms Earliest Operation Due Date (EODD) or First-In-First-Out (FIFO) rules (Fernandes et al., 2023), these comparisons typically assume ideal execution conditions, such as preemptive flowshops. Furthermore, standard DDMRP sequencing does not dynamically adjust to shop floor disturbances. This limitation suggests the need for Dynamic Order Dispatching rules that utilize real-time execution variables rather than planning variables projected from the past. In the context of this research, dynamic dispatching offers a theoretical pathway to resolve the limitations of standard DDMRP sequencing. By transitioning from the static BPR to dynamic rules based on real-time On-Hand availability, the system can align execution priorities with the physical reality of the shop floor, rather than the theoretical projections of the planning engine.

2.3.2 Task 2: Order Release and Order Heterogeneity

In the context of production control systems, a fundamental distinction must be made between demand variability and order heterogeneity. While demand variability refers to fluctuations in the volume of demand over time, order heterogeneity refers to the co-existence of distinct classes of manufacturing orders within the same production stream, distinguished by a variety of factors such as, but not limited to, their urgency, service level requirements, origin of the order signal, processing complexity or commercial priority.

Standard DDMRP methodology is explicitly designed to dampen demand variability through the strategic resizing of inventory buffers (Ptak & Smith, 2019). However, these mechanisms focus on the magnitude of the signal, not its classification. Once an order is generated, standard DDMRP execution rules treat the stream as homogeneous, implicitly assuming all orders share the same

priority progression. This creates a control gap in environments especially when the production system is subject to heterogeneous orders streams in the same flow.

In theory, DDMRP acknowledges the existence of heterogeneous order streams. Ptak and Smith (2019) explains that DDP receives three different order signals: standard buffer replenishment orders, qualified demand spikes, and non-buffered items. However, to simplify execution, the DDP phase mathematically unifies these distinct signals into a single metric (i.e., the NFP). Consequently, the majority of DDMRP research (Azzamouri et al., 2022; Dessevre et al., 2023; Fernandes et al., 2024; Thürer et al., 2020) assumes that the resulting order stream is entirely homogeneous, treating all orders as routine buffer depletion without differentiation regarding their origin or client importance. Disregarding this heterogeneity masks critical execution challenges, as the standard DDMRP engine treats urgent orders, exceptional demand spikes or an order from an important client identically to routine stock replenishment.

In broader manufacturing literature, these urgent, exceptional demands, which DDMRP blends into the standard flow, are formally classified as rush orders. In manufacturing environments, production orders assigned with the highest priority to fulfill urgent customer requirements are defined as rush orders (Mundt et al., 2022). Unlike standard orders, their demand typically occurs before the standard delivery time, resulting in the preemption of existing schedules (Chang et al., 2024; Yao & Liu, 2009). The literature predominantly treats these orders as special cases or disturbances within the production process (Trzyna et al., 2012). From a system dynamics perspective, Jäger and Roser (2018) note that prioritizing rush orders does not alter the average Flow Time (FT) of the system; rather, it redistributes the waiting times, reducing the FT for the rush orders at the expense of increasing the FT variance for standard orders.

The management of rush orders is generally approached from two different theoretical perspectives: rush order acceptance and rush order scheduling. Research on acceptance focuses on the managerial decision of whether to admit a rush order into the system based on capacity and profitability. Methodologies in this domain often employ utility values to evaluate the potential impact on system performance (Aqlan et al., 2017) or heuristic models that weigh the economic reward against the cost of prioritization (Chen, 2010). Regarding the capacity planning, Hendry et al. (2008) suggest reserving 10-20% of capacity specifically for rush order acceptance. Thürer et al. (2009) further demonstrate that maintaining a rush order share of up to 30% of the capacity

ensures the stability of Flow Times. More recently, Mundt et al. (2022) developed a method for reserving production capacities to allow rush order acceptance without overloading the system.

Once accepted, the challenge shifts into determining how to process these urgent orders amidst regular orders (i.e. dispatching). Complex algorithm approaches have been proposed for job shops and hybrid flowshops, including genetic algorithms (Ren et al., 2021) and match-up rescheduling strategies (Moratori et al., 2010). However, in flowshop environments, rush order scheduling is primarily executed through dispatching rules. The literature identifies two primary strategies for this prioritization. The first strategy is to prioritize rush orders at every machine in the flowshop, as demonstrated by Neuner et al. (2022) and Trzyna et al. (2012). The second strategy is to prioritize rush orders only on the pool of orders at the order release stage, as demonstrated by Thürer et al. (2009).

From a DDMRP perspective, the presence of rush orders in the shop floor introduces levels of complexity to the order release task. Order Release serve as the ‘gatekeeper’ between the planning phase and the shop floor. According to Lödding (2013), the primary objective of this control function is to decouple the rate of order generation from the rate of production, thereby limiting WIP. By restraining the release of orders, the system regulates WIP, thereby stabilizing throughput times in accordance with Little’s Law (Hopp & Spearman, 2011).

In contrast to the load-limiting logic advocated by Hopp and Spearman (2011), current DDMRP research typically employs an Immediate Release (IMR) policy. Once an order is generated by the planning phase, it is immediately authorized for execution (Duhem et al., 2023; Fernandes et al., 2024; Velasco Acosta et al., 2019). This policy implicitly assumes that the regulation of the NFP in the planning phase is sufficient to prevent shop floor congestion. However, this assumption holds only in homogeneous environments with infinite capacity. In stochastic flowshops facing heterogeneous demand, IMR fails to account for the distinction between standard replenishment orders and rush orders for example.

The lack of a formal order release mechanism adapted to heterogeneous orders streams exposes the shop floor to increased WIP when variability increases. Recent studies have indicated that while DDMRP stabilizes the supply chain, it can inadvertently increase shop floor WIP when compared to systems with strict order release controls like CONWIP (Thürer et al., 2020). Specifically, Immediate Release alone is not sufficient to manage heterogeneity in the order stream. This

heuristic hinders DDMRP applications, since in real-world applications, manufacturers often contend with unforeseen, high priority demands that require expedited processing.

To maintain the robustness of the system when faced with heterogeneous orders, the execution phase requires a dynamic prioritization mechanism at the release stage. The current literature does not address this execution issue. In the context of this thesis, the introduction of rush orders serves as a proxy for order heterogeneity, challenging the standard DDMRP assumption of priority being defined by the buffer status only. By evolving from the IMR, which does not account for order heterogeneity beyond the NFP, this thesis proposes using rush orders to test dynamic release mechanisms. This effectively transitions the Order Release task in DDMRP from a static authorization based on inventory availability to a dynamic control based on the heterogeneous status of orders.

2.3.3 Task 3: Order Generation and System State Prediction

Order Generation constitutes the interface between external demand and the internal production system. The primary function of this task is to transform demand data into manufacturing orders, thereby determining the Planned Load of the system (Lodding, 2013). While Order Release regulates the flow of work, Order Generation sets the volume of work to maintain service levels.

In DDMRP, order generation is governed by the NFP. A replenishment order is recommended whenever the NFP falls below the TOY threshold (Ptak & Smith, 2019). In DDMRP theory, there is a distinction between the recommendation of an order (triggered by the NFP) and the approval of an order (execution by a planner). However, the methodology does not formalize a specific logic for the approval step. Consequently, both in academic literature and practical implementation, it is typically assumed that recommended orders are automatically approved (Duhem et al., 2023; Fernandes et al., 2024; Velasco Acosta et al., 2019). This renders the order generation process a single-factor decision: it is driven exclusively by the buffer status, ignoring the current state of the production system.

This reliance on static inventory threshold for order generation creates an open-loop interaction between Planning and Execution. Since the standard order generation logic is ‘blind’ to the shop floor status, it cannot account for stochastic variability in capacity or lead times. If the shop floor is underutilized (low congestion), the standard logic generates orders based solely on the buffer consumption. This may result in orders being released and completed earlier than necessary,

inflating the average inventory and holding costs. On the other hand, if the system is overloaded, the standard logic continues to generate orders to restore the NFP. This behavior potentially exacerbates queues and violates Little's Law by dumping more load into a constrained system (Hopp & Spearman, 2011).

The lack of a formalized, state-aware approval mechanism represents a significant gap in both planning and execution phases. While recent studies have explored coupling DDMRP with CONWIP to limit WIP (Dessevre et al., 2023), there is a lack of research integrating predictive analytics to create status-centered mechanisms directly into the order generation decision.

Predictive analytics encompasses a variety of statistical techniques such as data mining, predictive modelling, and Machine Learning (ML) that analyze current and historical facts to make predictions about future or unknown events (Waller & Fawcett, 2013). Within this domain, ML is defined as a subset of algorithms that provides systems the ability to learn and improve from experience without being explicitly programmed (Murphy et al., 2019). In manufacturing control, ML algorithms are particularly valued for their ability to model complex relationships between shop floor variables which traditional analytical models fail to achieve.

Beyond general process optimization, predictive analytics and ML have been successfully applied to resolve stochastic problems within manufacturing and supply chain management. Recent literature demonstrates the versatility of the techniques. In the supply chain management domain, ML algorithms have been used to prediction of supply chain risks (Kumar & Sharma, 2023), selection of replenishment policies (Priore et al., 2018) and backorders anticipation (Ntakolia et al., 2021).

At the shop floor level, ML algorithms have been used to transition from static to dynamic control relying on the ability to accurately predict system states in stochastic environments. Murphy et al. (2019) demonstrate that ML techniques significantly outperform conventional static heuristic for Flow Time Estimation by incorporating real-time variables such as queue lengths and utilization. Building on this predictive capability, researchers have successfully introduced dynamic reasoning into traditionally static tasks. Schneckenreither et al. (2020) used Artificial Neural Networks (ANNs) to replace static planning parameters with dynamic lead times calculated through flow time estimation to improve the order release process. Concurrently, Xanthopoulos et al. (2019) applied Reinforcement Learning (RL) to create adaptive control policies for CONWIP, adjusting

the release of orders based on system dynamics rather than fixed rules. In the domain of inventory management, Hammeler et al. (2023) used Deep Reinforcement Learning (DRL) to substitute static reorder point policies with dynamically adjusted thresholds based on the state of supply chain.

As demonstrated by Murphy et al. (2019), employing advanced ML algorithms can significantly improve prediction accuracy. These models can use state vectors (e.g. queue lengths, machine availability, occupation, current demand) to output precise information for a specific order in real time. This capability highlights a critical limitation in standard DDMRP, which relies on a static Decouple Lead Time (DLT) to dimension buffers and trigger replenishment. Order Generation sets the upper bound of performance (Lödding, 2013), if this task is misaligned with the shop floor state, all the other functions are rendered useless. Relying on a static parameter implicitly assumes constant flow capability, creating a blind spot where the system generates orders regardless of actual shop floor state. As argued by Schneckenreither et al. (2020), using static lead times for decision in dynamic environments can lead to the “Lead Time Syndrome”, where congestion triggers longer planned lead times, resulting in earlier releases that further saturate the system.

This limitation suggests that the order generation process must evolve from a single-factor decision based on buffer status to a dual-factor decision incorporating the real-time state of the system. In this thesis, the integration of ML serves as the mechanism to bridge the gap between the static planning and dynamic execution capability. Evolving from the single-factor NFP, which ignores shop floor congestion, this thesis proposed using the predicted Flow Time (pFT) to create a smart approval gate for replenishment orders where the system can strategically withhold orders during periods of idleness or congestion, thereby optimizing the trade-off between the inventory efficiency and service levels. This effectively transitions the Order Generation task in DDMRP from a decision based on self-centered inputs to a dual-factor status-centered execution system based on both inventory needs and the real-time state of the production system.

2.4 Research Gaps

This chapter has critically examined the theoretical foundations of DDMRP through the functional lens of PCSs. While the literature demonstrates that DDMRP provides a robust framework for decoupling supply from demand variability through buffer positioning, significant research gaps remain unaddressed:

- The DDP phase functions as a closed control loop, responsive to market signals, yet the DDE phase operates largely as an open loop, governed by self-centered heuristics that ignore the stochastic reality of the shop floor.
- Regarding the sequencing task, the reliance on the static sequencing rules introduces a critical latency error. Another important consideration is that the priority at the moment of order generation is derived from the NFP, which fails to account for disturbances that occur after order release. This leads to a misalignment between the planned priority and the physical feasibility of execution.
- At the order release level, the standard Immediate Release (IMR) policy, adopted by the majority of DDMRP research, operates under the assumption of infinity capacity and order homogeneity. This assumption lacks a gating mechanism, leaving the shop floor vulnerable to disruptions, particularly when processing heterogeneous orders streams. Without a mechanism to distinguish between standard replenishment and high-priority requirements, the system risks saturating critical resources with non-essential work.
- Finally, at the order generation level, the Net Flow Equation triggers replenishment based solely on inventory thresholds, lacking the capability to assess the current state of the production system. This single-factor self-centered decision creates a blind spot where orders are generated to restore buffers regardless of whether the shop floor is idle or overloaded.

2.5 Literature Review Conclusion

In this chapter, we established the literature review to comprehensively understand the theoretical foundations of PCSs, the emergence of DDMRP as a hybrid approach and its mechanisms. The literature review identified a bias toward planning in DDMRP research and proposed to expand research into DDMRP execution aspects. The analysis of the method through the lens of PCSs theory identified gaps in the DDE regarding the reliance on self-centered heuristics and the lack of dynamic feedback of shop floor information to the functional tasks of sequencing, order release and order generation in DDE. The next chapter presents the thesis objectives and the general methodology.

CHAPTER 3 OBJECTIVES AND METHODOLOGY

This chapter is structured to present the objectives, general methodology, the case study and the main performance metrics used to evaluate the results. Section 3.1 formally defines the general and specific objectives pursued by the contributions. Section 3.2 details the general methodology, while section 3.3 focuses on the case study used throughout the contributions. Finally, Section 3.4 justifies the main performance metrics used.

3.1 Objectives

The primary objective of this thesis is to enhance the robustness and operational performance of DDMRP managed flowshops. To achieve this, we propose shifting the Demand Driven Execution (DDE) phase from static, self-centered heuristics to a dynamic, status-centered execution system.

To achieve this primary objective, three specific objectives (SO) are pursued: (SO1) establish and determine the extent which dynamic dispatching rules improve performance compared to static planning priorities; (SO2) validate an order release gating mechanism, in terms of prioritization policies, capable of maintaining flow stability when the system is subjected to heterogeneous order streams; and (SO3) propose and evaluate the performance implications of a status-centered order generation process that use predicted flow time as central decision metric.

This thesis aims to achieve each of the specific objectives through three distinct research contributions considering a flowshop of n machines as presented in Figure 3.1. Contribution 1 achieves SO1 by investigating the sequencing task presented in Chapter 4 of this thesis. Contribution 2 addresses SO2 by focusing on the order release task, presented in Chapter 5. Contribution 3 achieves SO3 by examining the order generation task and is presented in Chapter 6 of this thesis.

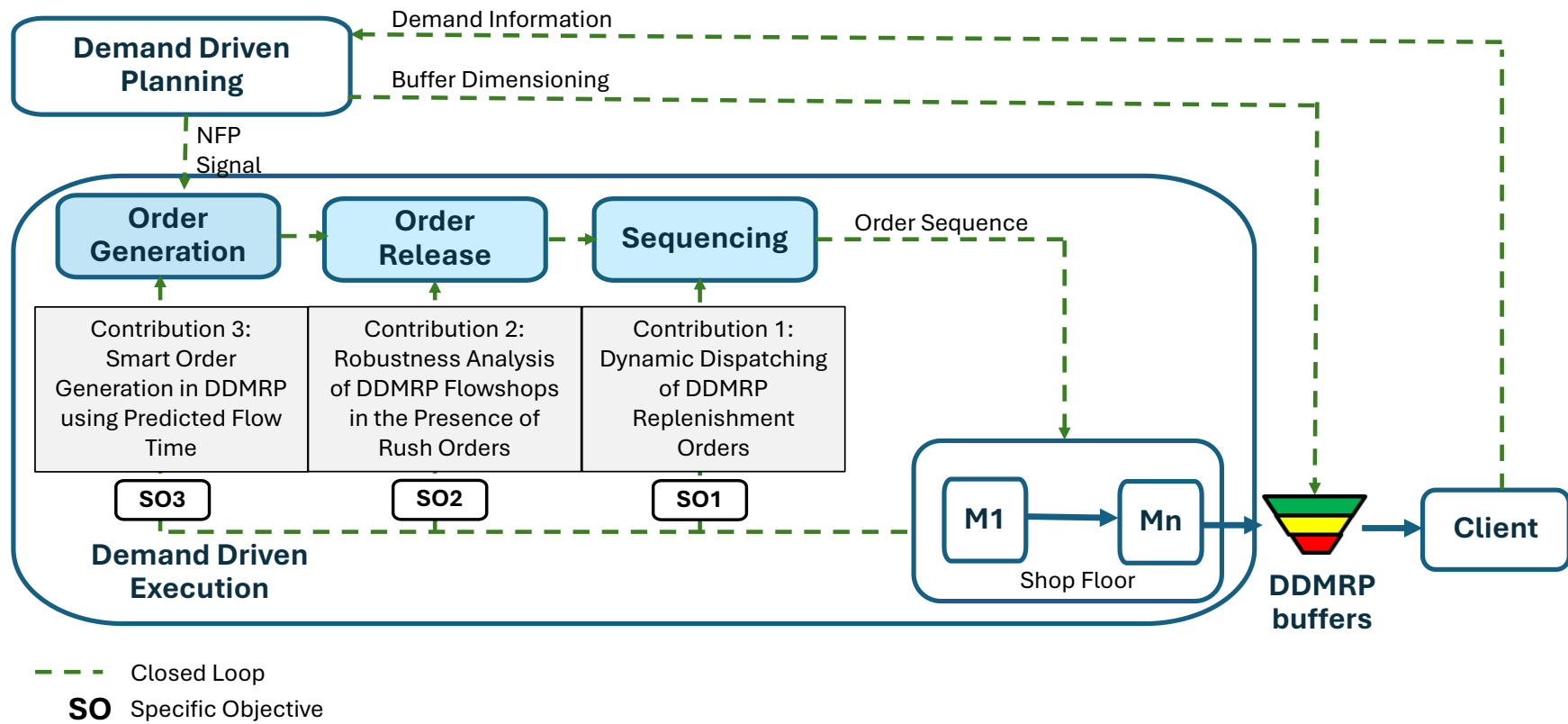


Figure 3.1 Positioning of contributions regarding each PCS function and Specific Objectives

3.2 General Methodology

The global methodology is structured as a chronological, five-step process illustrated in Figure 3.2.

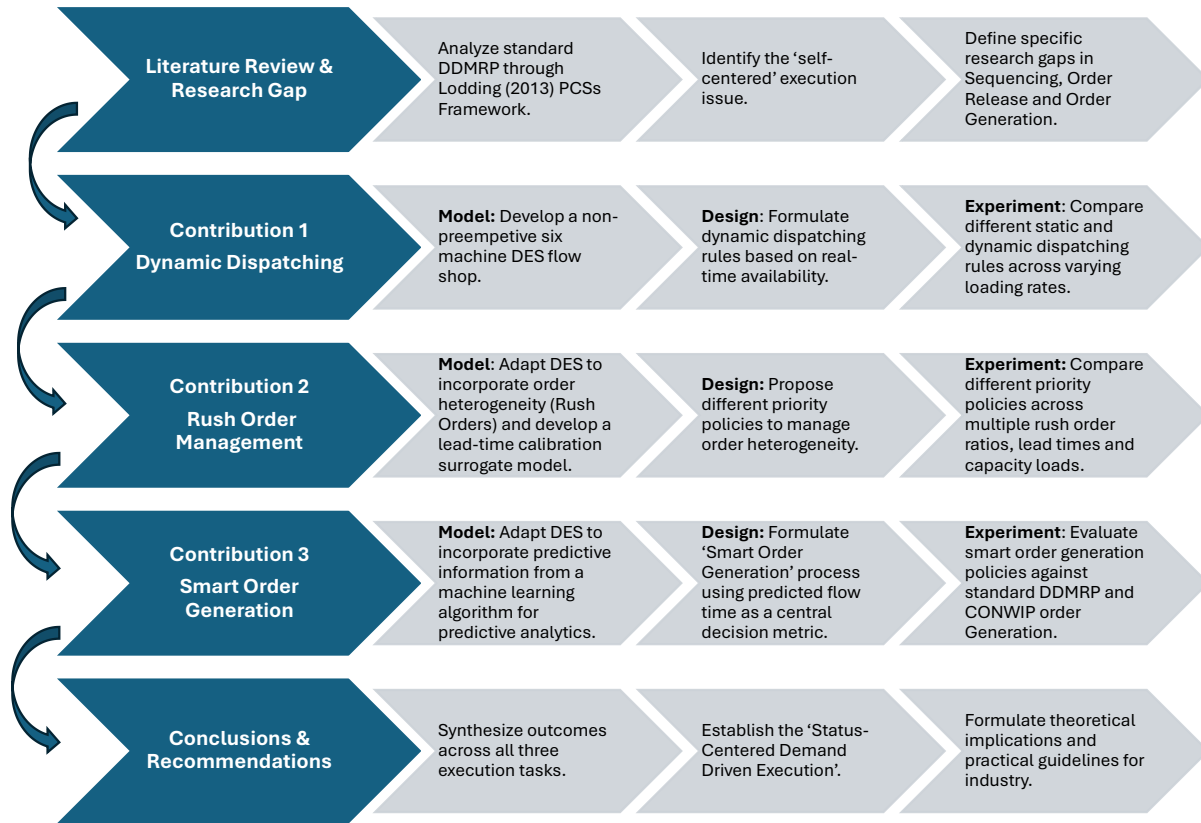


Figure 3.2 General Methodology

- **Step 1:** The methodology begins with a critical analysis of the standard DDMRP framework through the theoretical lens of PCSs. This step identifies the structural dichotomy of DDMRP: a closed-loop planning phase coupled with an open-loop execution phase. By mapping this flaw to Lodding (2013)'s functional tasks, we isolate three specific research gaps concerning the sequencing, order release and order generation tasks, setting the foundation for the three contributions of this thesis.
- **Step 2:** To address the sequencing gap, the methodology follows a three-part approach. First, the model is established by developing a non-preemptive, six-machine Discrete Event Simulation (DES) flowshop. This model is the base for all the three contributions. Next, the design phase formulates dynamic dispatching rules rather than static planning metrics. Finally, the experiment compares different static and dynamic dispatching rules across varying loading rates to measure the impact on the performance of DDMRP.

- **Step 3:** To address the order release gap, the base DES model is adapted to incorporate rush orders as a proxy for order heterogeneity and a lead-time calibration surrogate model is developed. The design phase proposes different priority policies to manage this order heterogeneity in the shop floor. The experiment phase compares these different priority policies across multiple rush order ratios, lead times, and capacity loads to measure the impact in system performance and determine the operational limits of DDMRP robustness.
- **Step 4:** Concerning the order generation gap, the DES model is adapted to incorporate predictive information from a machine learning algorithm. The design phase formulates a smart order generation process using the predicted flow time as a central decision metric and proposed smart order generation policies. The experiment evaluates these smart order generation policies against standard DDMRP and CONWIP order generation rules to measure the impact on the performance of DDMRP.
- **Step 5:** The final step serves to synthesize the outcomes across all three execution tasks. This synthesis aims to establish the ‘Status-Centered Demand Driven Execution’ system. Ultimately, this leads to the formulation of theoretical implications and practical guidelines for the industry, concluding the research.

3.3 Case study

This thesis utilizes a representative flowshop as a consistent test subject across all contributions (see Figure 3.3). The case study is modeled on a high-volume manufacturing system subject to process and demand stochastic variability characteristic of the environments where DDMRP is typically applied. The experimental model consists of a six-machine flowshop configured to produce 30 product references managed by DDMRP buffers. This setup is designed to capture the complexity of execution dynamics including stochastic processing times, capacity constraints and heterogeneous demand.

To validate the proposed status-centered execution system, it is necessary to capture the non-linear interactions between planning parameters and shop floor dynamics. Analytical models (such as Queueing Theory) are often insufficient for this purpose as they usually struggle to represent the stochastic behavior of high-variability flowshops (Mourtzis, 2019). Therefore, this research employs DES as the primary methodological approach.

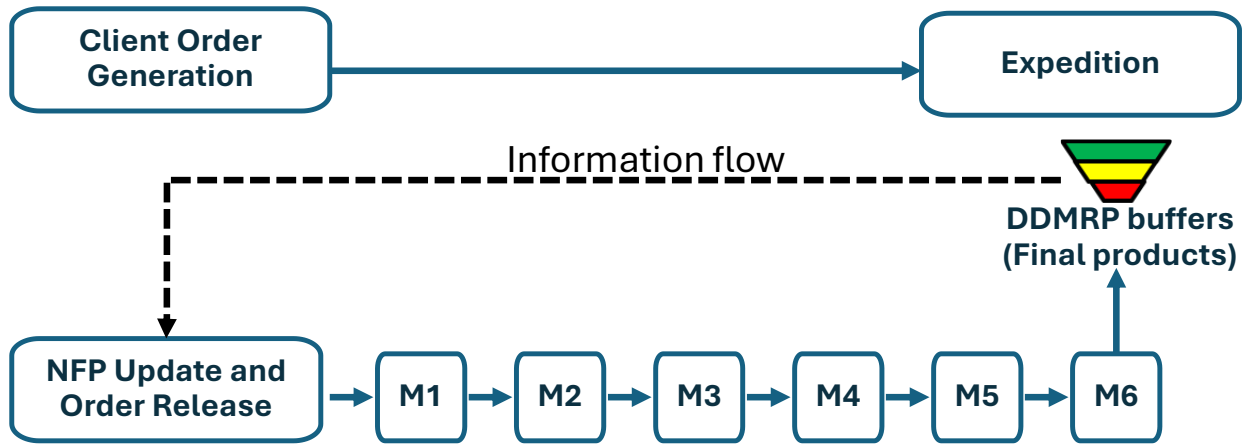


Figure 3.3 Representative Six-Machine DDMRP Flowshop

This choice was motivated by several reasons. First, simulation is a standard tool in Operations Management research, widely used to study complex production systems (Law, 2014). Unlike analytical models, simulation allows for the modeling of stochastic environments with high variability (Mourtzis, 2019), such as the proposed non-preemptive flowshop with demand and process variability used in this thesis. Furthermore, DDMRP is a methodology that relies on the interaction between multiple buffers and control points, with current DDMRP research relying on DES to test and validate different hypothesis (Dessevre et al., 2023; Duhem et al., 2023; Fernandes et al., 2024).

3.3.1 Main Elements of the DES Model

The experiments use a consistent generic simulation mode (Figure 3.4) across all contributions, inspired on the work of Dessevre et al. (2021). This ensures that the results are comparable and that the improvements observed can be connected to the proposed mechanisms rather than variations in the base model structure. The model was implemented using Arena 16.10 software.

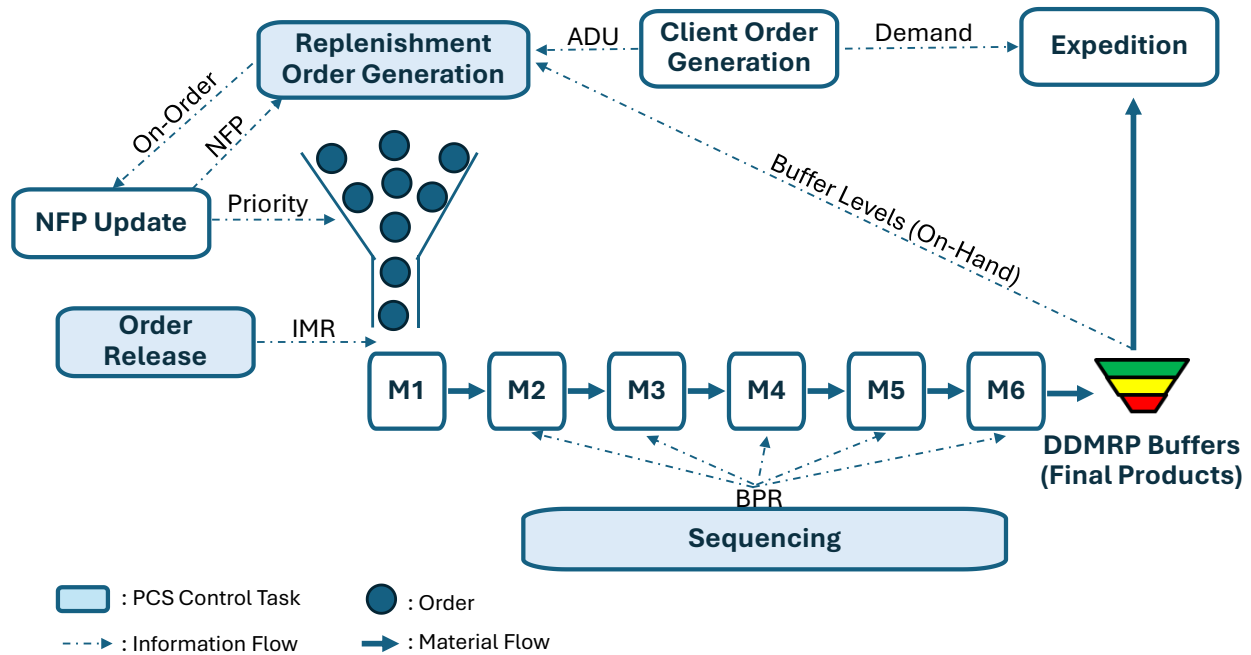


Figure 3.4 Generic Flowshop Model

The choice of the specific flowshop model (six machines, 30 products) is justified by its representativeness of the environment where DDMRP is typically applied: high-mix, constrained capacity, and significant variability. This flowshop was inspired by previous research on DDMRP which heavily experimented with the same configuration (Azzamouri et al., 2022; Dessevre et al., 2023; Dessevre et al., 2020). While we used a generic model rather than a specific industrial case for all contributions, this approach allows for the isolation of the control logic variables, ensuring generalization of the findings. By keeping the physical system constant across all three contributions, we ensure that the observed performance improvements could be attributed to the proposed status-centered execution system rather than variations in the flowshop structure.

3.3.2 The Production Environment

The simulation model consists of a non-preemptive flowshop with six sequential workstations (M1 to M6) selected to represent a constrained flowshop where downstream starvation and upstream blocking can occur, typical of DDMRP managed flowshops. The system produces 30 distinct finished product references (P1 to P30). All products follow an identical routing sequence from M1 to M6 but possess distinct processing time distributions. To introduce production variability, processing times (PT) are stochastic. They are modeled using a Triangular Distribution TRI(min, mode, max) applied to a base time, adjusted by a variability coefficient. The system operates at a

finite capacity. Workstations process one job at a time. The simulation does not allow order preemption (i.e. once an order begins processing, it must be completed before another order starts). There are small variations between the proposed models. Refer to Chapters 4, 5 and 6 for a full description of the production environment for each contribution.

3.3.3 Demand and Inventory Logic

The demand generation modules simulate a Make-to-Stock (MTS) environment characterized by high volatility. Customer demand is generated stochastically. The inter-arrival time of customer orders follows an exponential distribution. Each product (P1 to P30) is assigned a specific ADU. The daily quantity requested by each customer follows a normal distribution with the average equal to the ADU, ensuring that the system faces both volume and frequency variability. Each of the 30 products is managed by a standard DDMRP buffer profile. The buffer zones (Green, Yellow and Red) are calculated using the standard equations defined in Section 2.2, based on the specific ADU, Lead Time (LT), Variability Factors (VF) and Lead Time Factor (LTF) defined for the products.

3.3.4 Control Logic Mechanisms

The simulation model integrates the Planning and Execution logics of DDMRP, as illustrated in Figure 3.2. After each order is received, the system updates the NFP for every product using the standard equation defined in Section 2.2. The system evaluates the NFP: if it falls below the TOY threshold, a replenishment order is generated with a quantity calculated to restore the inventory to the TOG. Once generated, orders enter the production system logic. In the baseline scenarios used as a control group, orders are subjected to immediate release (IMR) and sequenced based on the combination of NFP and FIFO, or the BPR. These control logic modules serve as the baseline for comparison. Modifications to the order sequencing, release, and generation logic are introduced and detailed in Chapters 4, 5 and 6 respectively.

3.4 Main Performance Metrics

The evaluation of the outcomes of the simulation models relies on performance metrics aligned with the standard objectives of PCSs defined by Lödding (2013). The primary indicators common to all experiments are Service Level (SL) and Average Inventory (I). SL is quantified as the percentage of customers' orders fulfilled on time, serving as the primary measure of system responsiveness. Average Inventory is the average of finished inventory in the DDMRP buffers over

the simulation horizon. Secondary metrics, such as flow time and the frequency of stockouts, are also recorded to provide insights into the shop floor dynamics. Together, these metrics capture the fundamental trade-off between maintaining high market responsiveness and minimizing average inventory. The main objective of using these two primary indicators (i.e., SL and I) is to demonstrate that the transition to status-centered execution system across sequencing, order release and order generation consistently yields a superior operational trade-off compared to the baseline.

.

CHAPTER 4 ARTICLE 1: DYNAMIC DISPATCHING OF DDMRP REPLENISHMENT ORDERS

This article was submitted in April 4, 2024 and published in September 8, 2024 in the IFIP International Conference on Advances in Production Management Systems (APMS) in 2024 in Chemnitz. The version presented in this thesis is identical to the published version.

The authors of this article are Lucas Alves^{*1}, Maha Ben Ali¹ and Pierre Baptiste¹.

* Main author. Responsible for conceptualization, methodology, software, validation, formal analysis, investigation, data curation, writing and visualization.

¹ Department of Mathematics and Industrial Engineering, Polytechnique Montréal, Montréal, Canada.

ABSTRACT

The Demand Driven Material Requirements Planning (DDMRP) is a production control system originally tailored for the VUCA world. Despite research increasingly focusing on the dynamic aspects of this production control system, it has failed to address important execution aspects. This study bridges the literature gap by focusing on Demand Driven Planning and Execution, particularly the dispatching rules. The use of static dispatching rules, such as the one proposed in DDMRP, does not respond to the volatility and uncertainty of the VUCA world. Through the simulation of a six-machine flowshop, we investigated several dispatching rules to assess their impact on the system stability and service level. Our findings reveal that, when compared to a static dispatching rule, the use of a dynamic dispatching rule in the DDMRP execution significantly improves the service level and reduces the number of stockouts and the time to fulfill backorders.

RESUME

Le Demand Driven Material Requirements Planning (DDMRP) est un système de pilotage de la production conçu à l'origine pour le monde VUCA. Bien que la recherche se concentre de plus en plus sur les aspects dynamiques de ce système, elle a négligé des aspects importants de l'exécution. Cette étude comble cette lacune dans la littérature en se concentrant sur la Planification et l'Exécution pilotées par la demande, et particulièrement sur les règles d'affectation. L'utilisation de règles d'affectation statiques, telles que celle proposée dans le DDMRP, ne répond pas à la

volatilité et à l'incertitude du monde VUCA. À travers la simulation d'un flowshop à six machines, nous avons étudié plusieurs règles d'affectation afin d'évaluer leur impact sur la stabilité du système et le taux de service. Nos résultats révèlent que, comparée à une règle d'affectation statique, l'utilisation d'une règle d'affectation dynamique dans l'exécution du DDMRP améliore significativement le taux de service et réduit le nombre de ruptures de stock ainsi que le temps nécessaire pour combler les retards.

Keywords: Dynamic Order Dispatching, Demand Driven Planning and Execution, DDMRP, VUCA environments.

4.1 Introduction

Nowadays, organizations face accelerated change with increased levels of unpredictability and complexity, resulting in unforeseen problems and vulnerability which require the development of new skills and the adoption of new strategies (Taskan et al., 2022). Traditional push or pull Production Control Systems (PCSs) fail to respond to organizations' needs in this new environment. Thus, managers increasingly adopt hybrid PCSs (Miguel et al., 2021), such as the Demand Driven Material Requirements Planning (DDMRP) which is proposed as a responsive PCS adapted to VUCA environments and as an alternative to traditional PCSs designed for more stable and predictable environments.

Considering its operationalization, DDMRP has five sequential phases: Strategic Inventory Positioning; Buffer Profiles and Levels; Dynamic Adjustment; Demand Driven Planning (DDP), and Demand Driven Execution (DDE). The first three phases define the configuration of DDMRP, such as where the decoupling points are placed, the amount of protection at those decoupling points, and how that level of protection increases or decreases based on operating parameters. The fourth and fifth phases define the actual operational aspects of a DDMRP system (Ptak & Smith, 2019).

There is a clear distinction between the Planning and Execution phases of DDMRP operationalization. The planning of replenishment orders based on the Net Flow Positions (NFPs) of the buffered items is part of the DDP, while the management of open replenishment orders based

on the On-Hand quantities is part of the DDE (Ptak & Smith, 2019). While some studies related to DDMRP (e.g., (Dessevre et al., 2021; Dessevre, Martin, et al., 2019; Kortabarria et al., 2018; Shofa et al., 2018)) considered mainly the planning phase parameters (i.e. the NFP) to build the dispatching rules, others (e.g., (Miclo et al., 2018; Pekarcikova et al., 2019; Thürer et al., 2020; Velasco Acosta et al., 2019)) consider mainly the execution phase parameters (i.e. the On-Hand). The variety of approaches results in research not acknowledging how order dispatching can affect DDMRP performance facing dynamic changes characterized by VUCA environments.

According to (Fernandes et al., 2023), dispatching rules centered around the buffer status reflected by the NFPs surpass due-date based rules in DDMRP because orders of the same product in the same queue can have different priorities according to their NFP. However, static dispatching rules relying only on NFPs are not adapted to a dynamic production system. In the modern manufacturing industry, methods to support real-time decision-making, such as dynamic evaluation of dispatching rules, are an urgent requirement to respond to the uncertainty and complexity (Zhang et al., 2021). From a practical point of view, managing a queue with different orders of the same product with different priorities incurs excessive execution complexity and issues with lot management and changeover times. This paper contributes to the DDMRP literature by investigating the utility of dispatching replenishment orders using both DDMRP planning and execution parameters (i.e., the NFPs and the On-Hand quantities). Considering a six-machine flowshop managed by DDMRP, three different dispatching rules are evaluated by simulation at various loading rates and compared in terms of their impact on the system stability and service level.

This article is structured as follows. Section 2 brings the current literature state regarding the use of dispatching rules in DDMRP applications. In Section 3, we introduce the case study and the proposed dispatching rules. Sections 4 and 5 present respectively the experiments and the result analysis. Finally, in Section 6, a conclusion is made and some perspectives for future research work are depicted.

4.2 DDMRP Concepts

Conceived as a rethinking of Material Requirements Planning (MRP), the DDMRP draws elements of Lean Systems and the Theory of Constraints to improve the overall performance and

effectiveness of the MRP logic. It challenged some MRP practices, such as defining safety stocks based on demand forecasts (Miclo et al., 2018).

For DDMRP practice, a buffer is a level of stock that is sized and maintained to protect a decoupling point. It serves to absorb demand variability, lead time compression, and replenishment order generation in terms of the NFP (see Figure 4.1). A buffer is divided into three zones: (1) Green Zone – determining the frequency of order generation and minimum size of each order; (2) Yellow Zone – the demand coverage in the buffer while replenishment orders are fulfilled; (3) Red Zone – the safety embedded in the buffer position (Ptak & Smith, 2019).

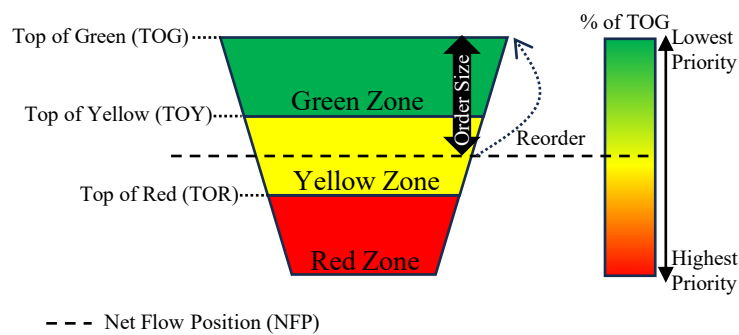


Figure 4.1 DDMRP buffer zones and Net Flow Position (NFP)

The buffer zones are proportional to the Average Daily Usage (ADU) and the Decoupled Lead Time (DLT) – or the Lead Time (LT) for non-decoupled product structures. The variability of both the ADU and the LT are considered in terms of the Lead Time Factor (LTF) and Variability Factor (VF). Refer to (Ptak & Smith, 2019) for suggested ranges. The sizing of each zone is demonstrated in the following equations:

$$\text{Red Base} = \text{ADU} \times \text{LT} \times \text{LTF} \quad (4.1)$$

$$\text{Red Safety} = \text{ADU} \times \text{VF} \quad (4.2)$$

$$\text{Red Zone} = \text{Red Base} + \text{Red Safety} \quad (4.3)$$

$$\text{Yellow Zone} = \text{ADU} \times \text{LT} \quad (4.4)$$

$$\text{Green Zone} = \text{ADU} \times \text{LT} \times \text{LTF} \quad (4.5)$$

After the three zones are sized, the Top Of each zone can be calculated. The following equations demonstrate the complete calculations of the buffers.

$$Top\ of\ Red\ (TOR) = Red\ Zone \quad (4.6)$$

$$Top\ of\ Yellow\ (TOY) = TOR + Yellow\ Zone \quad (4.7)$$

$$Top\ of\ Green\ (TOG) = TOY + Green\ Zone \quad (4.8)$$

As stated by (Ptak & Smith, 2019), replenishment order generation is a function of the buffers. A replenishment order is created if the NFP is below the TOY and the requested quantity corresponds to a replenishment to the TOG. Expressed by eq. 4.9, the NFP corresponds to the sum of the On-Hand (i.e. the current stock level of a particular item within a buffer) and the On-Order (i.e. the quantity of stock that has been ordered but not received) minus the Qualified Demand (QD). The QD is the past-due demand, plus today's demand, plus qualified order spikes in the future horizon (Ptak & Smith, 2019). The NFP is used to create orders which then need to be scheduled or prioritized at the production control level.

$$NFP = On - Hand + On - Order - Qualified\ Demand \quad (4.9)$$

4.2.1 Dispatching Rules in DDMRP Research

Dispatching consists of a priority rule that is used to select the next job to be processed from a set of jobs awaiting service (Blackstone et al., 1981). Efficient dispatching is critical for on-time deliveries, minimizing overdue and inventory costs, and adapting to dynamic and stochastic environments (Zheng et al., 2019). When considering the impact of dispatching on the flowshop performance, it is known that the change in the dispatching rules does not have an impact on the average Flow Time (FT) and the average Waiting Time (WT), but on the FT and the WT variability (Hubl et al., 2013). In fact, some replenishment orders can be accelerated at the expense of others, increasing the distribution width of the FTs and the distribution width of the WTs, even if the expected means remain unchanged (Jäger & Roser, 2018).

When analyzing the dispatching rule in DDMRP systems, it is crucial to take into account that one of the main objectives of DDMRP is to protect and promote the flow by using NFPs (Ptak & Smith, 2019). Initial DDMRP authors (Ptak & Smith, 2019) propose to use the NFP as a priority rule at the control production level. They note that the NFP characterizes the urgency for the order's creation, and therefore its priority. This priority (eq. 4.10) depends on the inverse of the value of the NFP over the TOG (i.e. smaller the value, the greater the priority) [3]. In this paper, we propose

using different parameters to calculate the priority and consequently dispatch replenishment orders in both planning and execution.

$$\text{Priority} = \text{NFP} / \text{TOG} \quad (4.10)$$

Initial DDMRP authors propose that the buffered item with the smallest value of the NFP over TOG at the moment of the order creation (i.e., the most ‘endangered’ buffer) has priority over the other buffered items in the planning stage (Ptak & Smith, 2019). In a static approach, priority is given on an order basis, while in a dynamic approach, priority is given on a product basis and in both cases they are dependent on the NFP (see Figure 4.2).

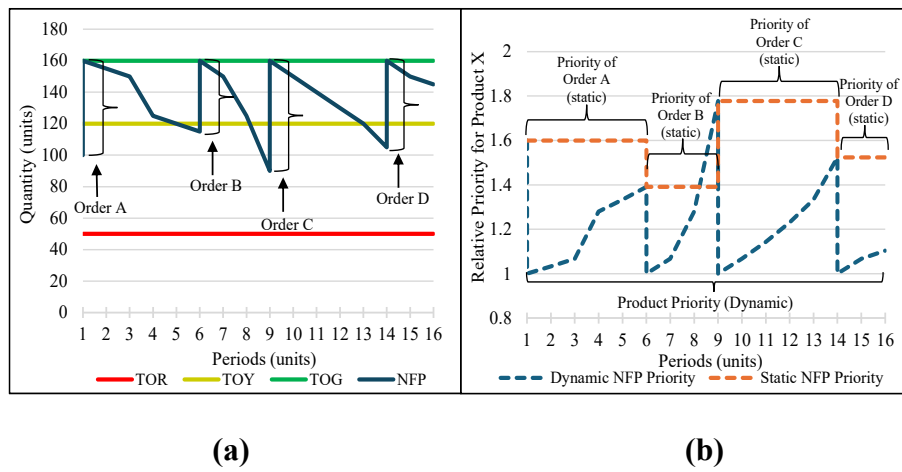


Figure 4.2 Product X's NFP evolution over time (a) and relative priority evolution (b)

When using the static NFP as proposed by [3], different orders of the same product on the same flow may have different priorities (i.e., the priority of order A is greater than that of order B and less than that of order C), while in a dynamic approach, different orders of the same product always have the same priority (i.e., all orders have the same priority that changes over time). For orders of the same product, the NFP at the moment of order generation determines the priority, with a smaller NFP resulting in a higher priority. According to (Ptak & Smith, 2019), the main function of the NFP is to dictate whether a replenishment order should be created for a buffer position while maintaining its integrity. The NFP is required for the planning phase, while the buffer integrity has to be considered in the execution to ensure that the buffer effectively performs its decoupling role.

While most research on DDMRP (Dessevre et al., 2021; Dessevre, Martin, et al., 2019; Kortabarria et al., 2018; Shofa et al., 2018) used the combination of NFP and First In First Out (FIFO) to dispatch replenishment orders, based only on DDP parameters, a few studies proposed different

approaches. In (Miclo et al., 2018; Pekarcikova et al., 2019), single dispatching rules are proposed for both DDP and DDE phases based only on DDE parameters. In (Miclo et al., 2018), the dispatching rule is a result of the percentage of the On-Hand over the Red Zone plus the Green Zone. For (Pekarcikova et al., 2019), the dispatching rule depends on the percentage of the On-Hand over the TOR. In (Velasco Acosta et al., 2019), the proposed dispatching rule is based on both DDP and DDE parameters. The authors used a single dispatching rule for DDP and DDE phases giving priority to replenishment orders with an NFP in the Red Safety Zone with a stockout alert, if none of the conditions apply, priority followed the FIFO rule. (Thürer et al., 2020) proceeds differently and uses the expected production start date as the dispatching rule (i.e. the earliest expected production start date has the priority) at both DDP and DDE phases.

Considering the buffer status to dispatch replenishment orders, using NFPs as recommended by [3] seems to outperform due-date oriented dispatching rules and FIFO rules (Fernandes et al., 2023). However, the performance of NFP is highly dependent on two factors: the reorder quantity associated with the replenishment orders and the fact that orders of the same product may have different priorities (Fernandes et al., 2023). This suggests that the static NFP rule proposed by [3] is not adapted to all situations, as using only NFPs fails to cover some issues in the DDE phase. To manage open replenishment orders in the execution phase, (Ptak & Smith, 2019) proposes to prioritize replenishment orders with issued On-Hand Alerts. An On-Hand Alert is generated whenever the On-Hand Position is below half of the TOR [3]. However, it is unclear in theory and practice how to use both On-Hand alerts and NFPs for dispatching considering dynamic production systems.

Despite the increase in research to evaluate dynamic aspects of DDMRP such as the decouple lead time (DLT) (Dessevre, Martin, et al., 2019), buffer sizing (Azzamouri et al., 2022; Martin et al., 2023), and parametrization (Damand et al., 2022a; Duhem et al., 2023), none of the previously mentioned applications considered a dynamic dispatching rule for DDMRP systems. Since (Fernandes et al., 2023) findings demonstrate that dynamically adjusting the NFP may not be the best option when dispatching replenishment orders, it is interesting to further investigate DDMRP dynamic dispatching rules. In contrast to the previous literature, this paper proposes to use real-time buffer data for order dispatching in a way that better connects the DDP and the DDE phases of the DDMRP methodology.

4.3 Dispatching Rule Models

This paper aims to investigate the utility of dynamic dispatching in DDMRP applications in terms of system stability and service level. To this end, first, we consider a Discrete Event Simulation (DES) model of a DDMRP flowshop, using ARENA 16.10. The model is inspired by (Azzamouri et al., 2022) and uses the static NFP rule as the most commonly used dispatching rule in DDMRP applications (Dessevre et al., 2021; Dessevre, Martin, et al., 2019; Kortabarria et al., 2018; Shofa et al., 2018). Then, we add two dynamic dispatching rules that will be compared to the static NFP rule (which serves as a benchmark for evaluation). Finally, we experiment with the model with the three dispatching rules in scenarios considering different loading rates and compare the two proposed dynamic dispatching rules to the static rule.

4.3.1 Static Dispatching Rule Model (Static)

The DES model represents a six-machine flowshop where final products managed with DDMRP buffers must pass through all the machines. We consider 5 operating days a week with 1 shift of 8 hours per day. Replenishment orders are generated every time the NFP is below TOY to restore the TOG quantity (section 2.1). We consider 30 products manufactured in the flowshop. The customer demands have a similar interarrival rate. The quantities requested by customer demands are calculated to meet the ADU presented in Table 4.1 with a 20% variability. To simplify, the visibility of customer demands is restricted to a 1-day simulation horizon. Thus, the model does not consider qualified spikes in the QD. If a customer demand cannot be met from the On-Hand inventory, it would be back-ordered to be fulfilled as soon as possible.

Table 4.1 Average Daily Usage (ADU) in units

Product	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10
ADU	65	23	55	63	19	38	29	33	61	34
Product	P11	P12	P13	P14	P15	P16	P17	P18	P19	P20
ADU	61	62	57	32	52	21	61	42	48	44
Product	P21	P22	P23	P24	P25	P26	P27	P28	P29	P30
ADU	35	52	23	56	55	37	32	51	55	52

The Processing Times (PT) (see Table 4.2) are randomly adjusted with a coefficient following a triangular law TRI (0.8; 1; 1.2). Transfer times between different processes are not considered neither changeover times between products. The variability factor is 0.50 for every product and the lead time factor is 0.75, 0.5, and 0.25 respectively for small, medium, and long PTs, following the factor ranges suggested by Ptak and Smith (Ptak & Smith, 2019). The Lead Time (LT) is 9.75 hours for all experiments. Buffers are randomly initialized between 50% and 100% of the TOG.

Table 4.2 Processing Times in Seconds

Products / Machines	M1	M2	M3	M4	M5	M6
P1 – P10 (small PT)	5	5	5	10	15	5
P11 – P20 (med. PT)	5	10	10	10	10	10
P21 – P30 (high PT)	5	15	15	10	5	15

The Static dispatching rule for replenishment orders considers the static value of the NFP over the TOG. The replenishment order with the smallest value has the highest priority. In the case of two replenishment orders having the same priority, the FIFO rule is used. This dispatching rule is used only for the first machine, all the subsequent machines use the FIFO dispatching rule (see Figure 4.3). For a description of DDMRP functioning refer to Ptak and Smith (Ptak & Smith, 2019).

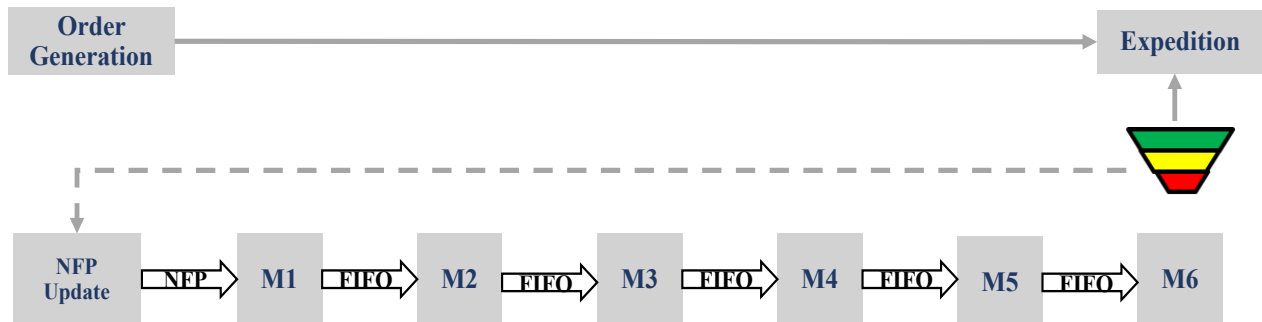


Figure 4.3 The DDMRP flowshop managed by the Static dispatching rule

4.3.2 Dynamic Dispatching Rule Model

To assess the performance of the dynamic dispatching rules, we adapted the static dispatching model to function with the two new dispatching rules. The first dynamic rule uses the **Net Flow On-Hand Positioning (NFOHP)**. We propose to use the static NFP to define the priority in the first machine queue, in such a way that the lowest value of the NFP over the TOG has the highest

priority. For the subsequent machines, the On-Hand Position (expressed by eq. 4.11 as the On-Hand stock divided by the TOR), is dynamically adjusted each time a decision is made about the next replenishment order to process. The replenishment order with the smallest On-Hand Position is placed first in the queue for processing (see Figure 4.4). In contrast to Ptak and Smith's (Ptak & Smith, 2019) which uses the On-Hand Position only for managing orders with an active On-Hand Alert, we propose to consider the On-Hand Position for all replenishment orders.

$$\text{On - Hand Position (OHP)} = \text{On - Hand} / \text{TOR} \quad (4.11)$$

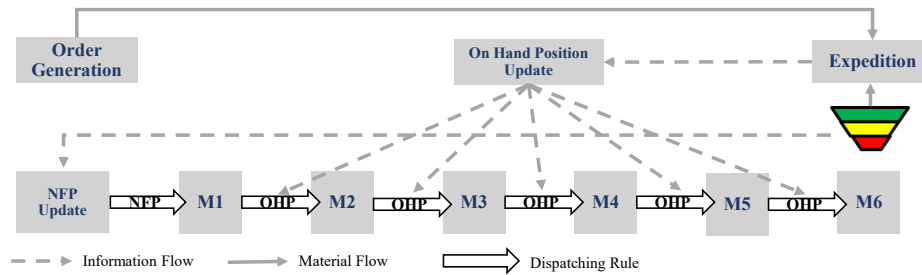


Figure 4.4 The DDMRP flowshop managed by the Net Flow On-Hand Positioning (NFOHP) dispatching rule

The second dynamic rule that we propose is the **Dynamic Available On-Hand Positioning (DAOHP)**. With this rule, we dynamically adjust what we call the Available On-Hand Position (AOHP) for the replenishment orders in all machines' queues each time a decision is made about the next replenishment order to process. As presented in eq. 4.12, we define the AOHP as the units in inventory minus the backorders and the demand due today. The replenishment order with the smallest AOHP is placed first in the queue for processing (see Figure 4.5).

$$\text{AOHP} = (\text{On - Hand} - \text{Backorders} - \text{Demand}) / \text{TOR} \quad (4.12)$$

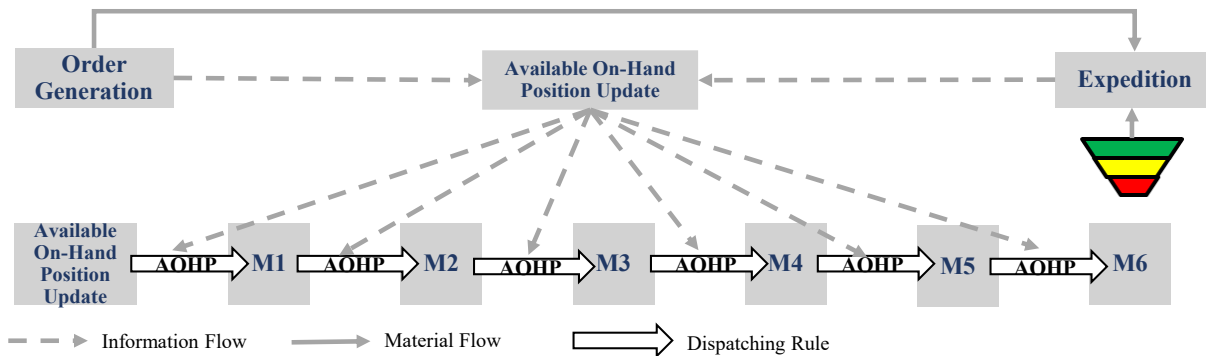


Figure 4.5 The DDMRP flowshop managed by the Dynamic Available On-Hand Positioning (DAOHP) dispatching rule

4.4 Experiments

To achieve the objective of assessing the dispatching rules' impact on the system stability and service level of DDMRP applications, we propose to experiment with a full factorial design with two factors characterizing the DDMRP system: the Loading Rate (LR) and the dispatching rule (Static, NFOHP, DAOHP). For each dispatching rule, 20 experiments (20 levels of LR from 80% to 99% in 1% increments) are simulated. A total of 60 experiments are conducted and the performance is measured using six performance indicators:

- Service Level (SL): the number of customer demands filled in time over the total number of customer demands.
- Average Inventory (I): the average inventory (Work In Process + Finished Products) during the experiment.
- Stockouts (SO): the number of times a buffer position reaches zero.
- Backorders (BO): the number of units delivered after the expected time.
- On-Hand Alerts (OHA) percentage: the percentage of On-Hand Alerts over the number of deliveries. An On-Hand Alert is generated whenever the On-Hand Position is below half of the TOR.
- Mean Backorder Fulfillment Time (MBOT): the mean time to fulfill a backorder.

Considering 50 replications for each experiment and 95% confidence intervals (CI) on estimates, the results were collected during 550 simulation days, after 50 warmup days.

4.5 Results and Discussion

The simulation results obtained for the three proposed dispatching rules are summarized in Appendix A. We can see that the three different rules had the same performance in highly loaded systems (i.e., LR greater than 95%). Figure 4.6 focuses on the results up to 95% of LR and illustrates the Service Level SL and the average Inventory I with the three dispatching rules. First, as expected, as the LR increases, the system fails to maintain its stability, as demonstrated by I increasing exponentially when the LR approaches 95%. Second, we notice that both static and dynamic rules are equally affected by increasing the LR. We can conclude then that the effectiveness of the dispatching rules is not connected to the LR.

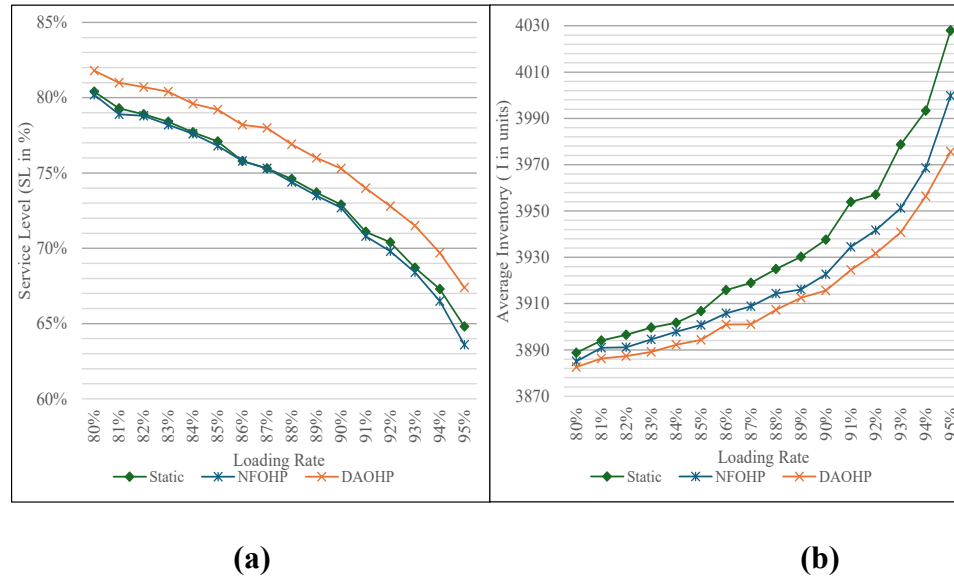


Figure 4.6 Impact of the Loading Rate on the Service Level and Average Inventory

While the Static rule and the NFOHP have mostly the same results in terms of SL, the DAOHP rule achieves a higher SL up to 95% of LR. This demonstrates that even if the NFOHP's parameters are dynamically adjusted, it fails to improve the DDMRP performance. This can be explained by the On-Hand not considering the backorders and the current demand and then failing to prioritize the most endangered buffers. In contrast, the DAOHP rule increases the SL compared to the static rule and the NFOHP. This is because the DAOHP rule prioritizes the most urgent replenishment orders for processing. In contrast to Ptak and Smith's (Ptak & Smith, 2019) On-Hand Position proposition which does not consider backorders and current demand, the DAOHP rule proposed in this paper prioritizes replenishment orders with the smallest availability. We can say that the DAOHP rule gives the DDMRP system the visibility of the most endangered buffer in real time and reorganizes it accordingly to define the next replenishment order.

Figure 4.7 shows the results of Stockouts and the Products Backordered. As expected, when the LR approaches 95% the number of Stockouts and Products Backordered starts to grow exponentially. It is interesting to note that up to 95% of LR, the DAOHP rule increases the buffer protection. When compared to the Static rule, the DAOHP rule consistently reduced the percentage of Stockouts (e.g., at 90% of LR, 9% reduction) and the number of products backordered (e.g., at 90% of LR, 12.6% of reduction).

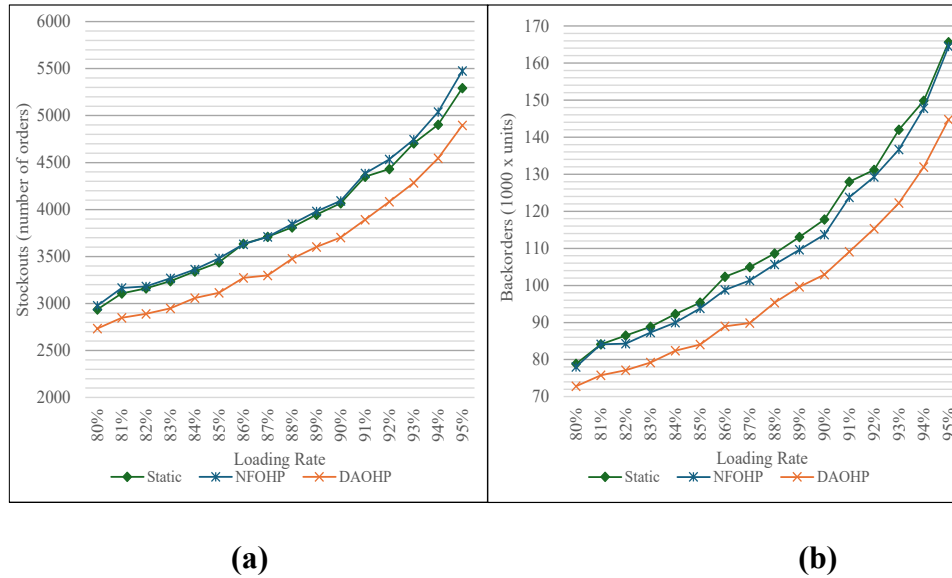


Figure 4.7 Impact of the Loading Rate on Stockouts and Backorders

Last but not least, with regards to the On-Hand Alerts percentage and the MBOT (see Figure 4.8), we can see that the effectiveness of a dispatching rule is connected to its ability to rapidly fulfill a backorder while reducing the number of On-Hand Alerts. Compared to the Static rule, the NFOHP rule has a smaller MBOT. Thus, this dynamic rule can reduce the time to fulfill backorders, avoiding stockouts. However, the NFOHP resulted in a higher number of On-Hand Alerts due to its lack of visibility of the most endangered buffers which resulted in the NFOHP rule and the Static rule having the same performance. In contrast, the DAOHP rule reduced both the On-Hand Alerts percentage and the MBOT, resulting in a higher performance compared to the other rules.

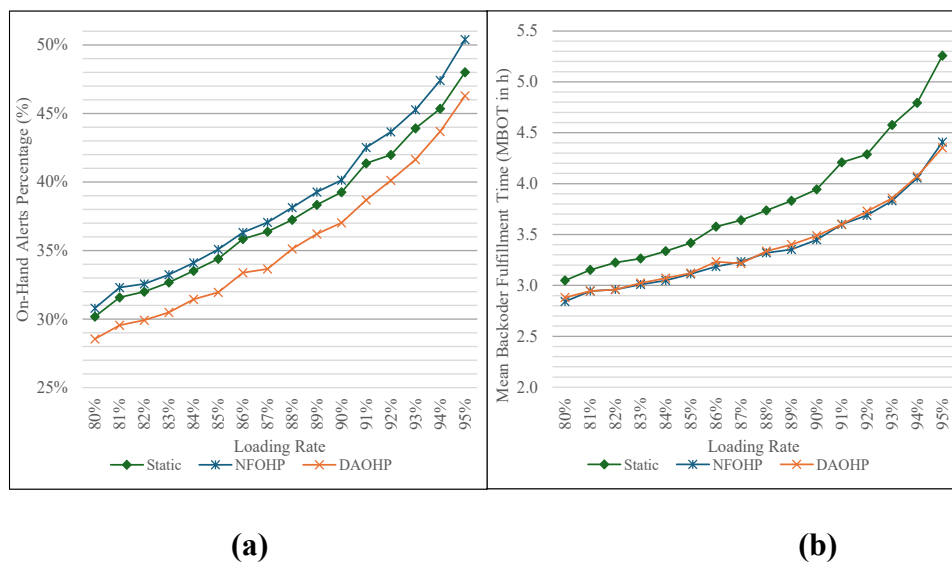


Figure 4.8 Impact of the Loading Rate on On-Hand Alerts Percent and MBOT

Finally, a comparison of the confidence intervals confirms the above results which indicate that, up to an LR of 95%, the DAOHP rule outperforms the Static and NFOHP rules, with the two latter performing similarly except for the MBOT. Both the Static and the NFOHP dispatching rules produced similar results mainly because of the routing characteristics of the flowshop (i.e., both rules treat the first machine queue similarly). The results between the Static and NFOHP rules may differ when examining other shop arrangements. For the DAOHP rule, since the same rule is applied to every queue, there is no reason for the results to differ in another production arrangement such as a job shop.

4.6 Managerial Insights

Dynamic dispatching rules that consider the DDMRP execution parameters can significantly improve service levels when compared to static rules. The DAOHP rule proposed in the study achieved higher service levels by prioritizing the most urgent replenishment orders for processing. The implementation of dynamic dispatching rules can reduce stockouts, backorders, and the time required to fulfill backorders. This improved responsiveness to demand fluctuations and disruptions is critical in today's volatile and uncertain business environment.

While traditional static dispatching rules in DDMRP cannot quickly adjust the production to changes in buffers, dynamic rules such as DAOHP can recognize the varying urgency levels of different buffers and prioritize accordingly. This tailored prioritization is consistent with the DDMRP philosophy of promoting flow and protecting decoupling points. The results of this study challenge the assumption of using the NFP to dispatch replenishment orders in the planning phase of DDMRP. The use of a dispatching rule based on real-time buffer data that incorporates backorder and future demand information can lead to better performance in DDMRP applications.

The implementation of dynamic rules such as the DAOHP requires the integration of real-time execution parameters, such as On-Hand inventory, backorders, and current demand with planning parameters such as buffer levels. This level of integration may require efforts to enable seamless data exchange between DDP and DDE modules. Another important consideration is the need to ensure high data quality and real-time visibility across the supply chain, which can be challenging, especially in multi-site, multi-echelon environments with disparate systems.

4.7 Conclusion

This paper investigates the problem of dynamically adjusting the dispatching rule in DDMRP applications. We contribute to the theory by extending the research on dynamic aspects of DDMRP that have not been addressed in the literature so far. Two new dynamic dispatching rules are proposed and compared to the conventional static dispatching rule under different loading rates. The results suggest that the dynamic rule DAOHP based on the Available On-Hand Positioning can significantly improve the service level, reduce the number of stockouts, and shorten the time to fulfill backorders.

Protecting the buffer position is one of the main objectives of DDMRP. However, buffer protection is different from inventory protection. At the execution level, considering the actual position of the inventory is key to improving the performance of the system. In this paper, using the Available On-Hand Position as the dispatching rule increased the SL due to its accuracy to represent real inventory status. The NFP does not represent the real inventory position because it considers orders that are already in the system. At the planning level, the NFP is an important parameter to determine the quantity to be replenished, but it should not be used at the execution level.

Despite the consistency of the results, there is a limitation related to the simulated flowshop. Further research should focus on uncovering the effects of using the static NFP as a prioritization parameter in DDMRP applications. Given the improvements resulting from using execution parameters for dispatching, it seems practical to consider other execution aspects to dispatch replenishment orders in DDMRP applications.

Acknowledgments. This work was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC).

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

CHAPTER 5 ARTICLE 2: ROBUSTNESS ANALYSIS OF DEMAND DRIVEN MATERIAL REQUIREMENTS PLANNING FLOWSHOPS IN THE PRESENCE OF RUSH ORDERS

This article was submitted in December 5, 2024 and published in January 13, 2026 in the Journal of Industrial and Production Engineering (JIPE). The version presented in this thesis is identical to the published version.

The authors of this article are Lucas Alves^{*1}, Maha Ben Ali¹ and Pierre Baptiste¹.

* Main author. Responsible for conceptualization, methodology, software, validation, formal analysis, investigation, data curation, writing and visualization.

¹ Department of Mathematics and Industrial Engineering, Polytechnique Montréal, Montréal, Canada.

ABSTRACT

The execution phase has received limited attention in the Demand Driven Material Requirements Planning (DDMRP) literature. In particular, the presence of rush orders in DDMRP systems is an unaddressed issue that is a common barrier to the application of the method. This paper aims to establish flowshops able to handle rush orders in a DDMRP system while maintaining high service levels for non-rush orders. Using Discrete Event Simulation, we designed a non-preemptive DDMRP flowshop to evaluate how rush orders affect service levels under different scenarios of rush order ratios, lead time settings, and prioritization policies. The results show that the DDMRP flowshop remains robust when rush orders account for 20% of capacity and that limiting prioritization in the production line ensures effective rush order management without compromising performance. These findings provide practitioners with insights into how to effectively maintain the robustness of DDMRP systems when reconciling rush and non-rush orders.

RESUME

La phase d'exécution a reçu une attention limitée dans la littérature sur le Demand Driven Material Requirements Planning (DDMRP). En particulier, la présence de commandes urgentes dans les systèmes DDMRP est une question non abordée qui constitue un obstacle fréquent à l'application de la méthode. Cet article vise à établir des flowshops capables de gérer les commandes urgentes

dans un système DDMRP tout en maintenant des taux de service élevés pour les commandes non urgentes. En utilisant la Simulation à Événements Discrets, nous avons conçu un flowshop DDMRP non préemptif afin d'évaluer comment les commandes urgentes affectent les taux de service selon différents scénarios de ratios de commandes urgentes, de paramétrages de délais et de politiques de priorisation. Les résultats montrent que le flowshop DDMRP reste robuste lorsque les commandes urgentes représentent 20 % de la capacité, et que limiter la priorisation sur la ligne de production assure une gestion efficace des commandes urgentes sans compromettre la performance. Ces conclusions fournissent aux praticiens des pistes sur la manière de maintenir efficacement la robustesse des systèmes DDMRP lors de la conciliation des commandes urgentes et non urgentes.

Keywords: Rush Orders; Order Prioritization; DDMRP; Discrete Event Simulation.

5.1 Introduction

Material Requirements Planning (MRP) systems often struggle in modern supply chains due to their reliance on inaccurate forecasts, which can lead to a mismatch between inventory and actual demand. To counter these limitations, Ptak and Smith proposed the Demand Driven Operating Model (DDOM), which shifts the focus from a forecast-driven push to a pull-based system guided by actual consumption (Ptak & Smith, 2019). This model integrates two key components: Demand Driven Execution, an agile framework for prioritizing shop floor orders based on real-time inventory buffer status, and Demand Driven Material Requirements Planning (DDMRP), the formal plan engine that manages those buffers. Literature shows that DDMRP offers a significant improvement over traditional MRP applications (El Marzougui et al., 2024; Miclo et al., 2018), especially when managing variability (Butturi et al., 2021), optimizing inventory (Math et al., 2024), compressing lead times (Shofa & Widyarto, 2016; Velasco Acosta et al., 2019), and improving service levels (Kortabarria et al., 2018). DDMRP outperforms traditional planning methods in high-volume manufacturing systems, particularly in environments characterized by both internal and external uncertainty (Corsini, 2023). However, the method is criticized for subjective parametrization (Krajčovič et al., 2024; Uzun Araz et al., 2024), the lack of capacity

management (Dessevre & Ben Ali, 2024), and the increased complexity for implementation when compared to traditional MRP (Bayard, S. et al., 2024).

While execution is critical to the effectiveness of any production control system, much of the literature on DDMRP tends to focus mainly on Demand Driven Planning aspects, such as Work-In-Process control (Dessevre et al., 2021; Dessevre et al., 2023), dynamic parametrization (Duhem et al., 2023; Martin et al., 2023), and buffer profiling (Azzamouri et al., 2022; Favaretto et al., 2023). Furthermore, DDMRP literature specifically addressing execution issues often focuses on homogeneous customer orders and preemptive flowshops (Fernandes et al., 2023; Fernandes et al., 2024). This narrow scope limits the applicability of DDMRP in many manufacturing environments, ranging from the precast concrete industry (Wang & Hu, 2018; Wang et al., 2021) to the semiconductor industries (Baykasoğlu & Ozsoydan, 2018; Psarommatis et al., 2021), which often face the challenge of handling rush orders.

Rush orders represent a critical execution issue for any production control system, including DDMRP, since placing rush orders is a common practice for companies to mitigate inventory risks (Chintapalli et al., 2017). Defined as the class of orders with the highest priority (Mundt et al., 2022), rush orders are a common response to quality issues, backorders, or technical requirements of the system, such as the need to consider different lead times for different parts. Handling rush orders poses an execution issue due to the operation priorities at the expense of non-rush orders, causing several negative consequences for overall operations (Arisha & Mahfouz, 2011). First, prioritizing rush orders delays the fulfillment of non-rush orders, which results in dissatisfied customers and missed service level agreements. In addition, the need to reallocate resources or adjust schedules for rush orders increases inventory levels and flow times, leading to instability. The volatility and uncertainty that generate rush orders are precisely the conditions that DDMRP is designed to mitigate. Despite this connection, the management of rush orders within DDMRP environments remains unexplored.

To address this gap, the objective of this study is to establish DDMRP flowshops able to handle rush orders by assessing their robustness when subjected to different rush order prioritization policies. This research employs a robustness analysis to assess the system's ability to absorb disruptions caused by rush orders without collapsing. This approach is employed because identifying resilient operational policies is of greater practical importance than finding a theoretical

optimum. We aim to guide practitioners on how to handle rush orders in a DDMRP flowshop while maintaining high service levels for non-rush orders. To achieve this objective, we propose to assess the impact of the ratio of rush orders on the performance of non-rush orders in DDMRP flowshops and to evaluate the impact of the extent of prioritization of rush orders in the production line on the robustness of the system. We employ a Discrete Event Simulation approach to capture the stochastic dynamics of rush orders disruptions, which static analytical models cannot adequately represent (Scrivano et al., 2023). Unlike existing DDMRP simulations models that typically assume preemptive flowshops (Fernandes et al., 2023) or homogeneous order streams (Fernandes et al., 2024), the proposed model incorporates specific execution-level prioritization logic within a non-preemptive environment. This allows the assessment of the model robustness under realistic, calibrated operational conditions. Using this simulation framework, we first analyze the impact of accepting rush orders on the service level, and secondly, explore the effects of this acceptance with different levels of rush order ratios, multiple lead time settings, and two prioritization policies. We conduct this analysis by comparing the average Inventory (I) for the different scenarios and two indicators for service level: Customer Service Level (CSL) and Production Service Level (PSL). In support of this analysis, to guide researchers and practitioners, we also provide a systematic procedure for Lead Time (LT) calibration within DDMRP applications.

The results of this study offer important implications for both theory and practice. Theoretically, this work addresses a critical gap by extending DDMRP research into the execution phase, providing a new perspective on order prioritization and system robustness under disruptive conditions. For practitioners, the study provides actionable guidance, demonstrating that system stability is best maintained by limiting the scope of rush order prioritization. Furthermore, this research offers a practical benchmark, showing that a DDMRP flowshop can remain robust while absorbing rush orders up to 20% of the capacity.

This paper is organized as follows. Section 2 presents the related literature and highlights the research contribution to the DDMRP literature. Section 3 outlines the methodology used and the experimental design. While Section 4 presents the simulation results, Section 5 brings a discussion of the practical and theoretical contributions. Finally, Section 6 concludes and discusses possible future research directions.

5.2 Literature Review

In this section, we first explain the MRP evolution, the key concepts of DDMRP, and review the importance of lead time calibration. Finally, we discuss order prioritization in DDMRP flowshops and the management of rush orders in manufacturing.

5.2.1 DDMRP Overview

Traditional MRP operates as a forecast-driven ‘push’ system. While effective in stable environments, its reliance on long-range forecasts makes it vulnerable to uncertainty, often propagating variability throughout the supply chain in what is known as the bullwhip effect (Saoud et al., 2025). To address these shortcomings, DDMRP was developed as a hybrid ‘push-pull’ methodology guided by the actual consumption. The core innovation is the use of strategically placed inventory buffers at decoupling points (Ptak & Smith, 2019). Inspired by methods based on the theory of constraints (e.g., the drum buffer rope and dynamic buffer management (Takami Narita et al., 2021)), DDMRP buffers protect critical points from disruptions while improving production responsiveness.

Several studies have pointed out the performance benefits of using DDMRP. The method's ability to effectively manage both internal and external variability is often cited as its primary advantage, particularly in complex and uncertain manufacturing environments (Alves et al., 2024; Azzamouri et al., 2021; Dessevire et al., 2023). Despite these benefits, there are several critiques of the method. A primary concern is the increased complexity, as DDMRP requires more parameters than a classic MRP, and the role of each parameter is not yet clearly defined (Bayard, S. et al., 2024). Another critique is the subjectivity in parametrization, which can lead to improperly sized buffers that fail to perform their intended function (Krajčovič et al., 2024).

DDMRP buffers are divided into three zones as shown in Figure 5.1: (1) the Green Zone which drives the frequency of order generation and the order size; (2) the Yellow Zone which represents the demand coverage in the buffer while replenishment orders are being fulfilled; and (3) the Red Zone which represents the safety embedded in the buffer (Ptak & Smith, 2019). When the Net Flow Position (NFP), represented by the dashed line, is below the Top of Yellow (TOY), an order is created to replenish the buffer to the Top of Green (TOG). The priority of the order is determined by the distance of the NFP to the TOG.

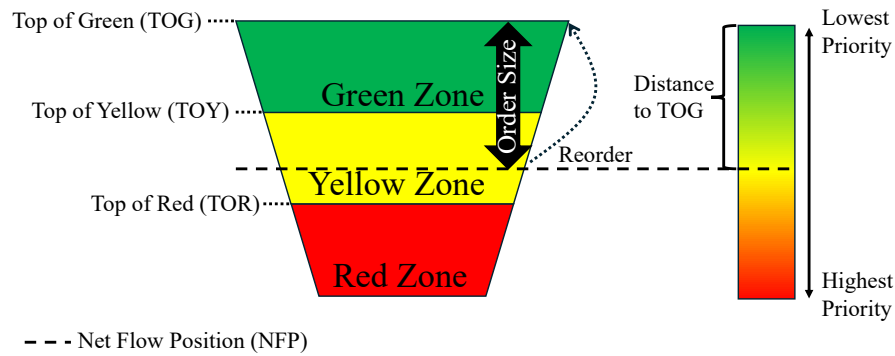


Figure 5.1 DDMRP buffer zones

The buffer zones are proportional to the Average Daily Usage (ADU) and the Decoupled Lead Time (DLT). While the Lead Time (LT) is the time allotted for the production of a part (Hopp & Spearman, 2011), the DLT is the longest cumulative coupled lead time chain in a manufactured item product structure, limited and defined by the placement of decoupling points within a product structure (Ptak & Smith, 2019). It should be noted that for non-decoupled product structures (i.e., a product without buffers between the processing steps), the DLT is equal to the LT. It is also important to distinguish between the LT, which is a planned production time, and the Flow Time (FT), which is an observed production time. Since it is common in the literature (e.g., Corsini (2023)) to confuse LT and FT, we propose clear definitions for LT and FT in Appendix B for a better performance comprehension.

DDMRP research measures performance in three categories: customer service, system performance, and inventory levels. Customer service is commonly measured through the Fill Rate (Corsini, 2023; Fernandes et al., 2024), On-Time Delivery (Damand et al., 2022b), or Customer Wait Time (Carmo-Silva & Fernandes, 2024; Fernandes et al., 2024) to evaluate the timely fulfillment of customer orders. System performance is often evaluated through metrics such as average Flow Time (Dessevre et al., 2023) and Throughput Time (Fernandes et al., 2023), aiming to measure the timely completion of the order on the shop floor. Inventory performance is usually evaluated by the average inventory (Alves et al., 2024; Fernandes et al., 2024).

Aligning with the literature, the present study utilizes the Customer Service Level (CSL) (i.e., the percentage of customers' orders filled on time directly from the buffer) to measure customer performance; the Production Service Level (PSL) (i.e., the probability of an order FT being smaller than the LT, inspired by Hopp and Spearman (Hopp & Spearman, 2011) definition of Service

Level) to measure system performance; and the Average Inventory (I) (i.e., the average of all finished products in stock during a period) for inventory performance.

5.2.2 Lead Time Calibration

The DLTs, and hence the LTs, are critical parameters in determining the buffer sizes of the parts managed by DDMRP. Improper calibration of these parameters results in poor performance (Lahrichi et al., 2022). However, DLT and LT calibration are not well understood in either theory or practice. For example, Younespour et al. (2024) dynamically adjust a part's DLT based on the upstreaming LT each time a buffer is depleted, but do not adjust the buffer sizes during the simulation. Corsini (2023) calibrates the Lead Time Factor (LTF) and the Variability Factor (VF) but does not explain how to calibrate the DLT. Fernandes et al. (2024) test different buffer levels but do not explain how the DLT parameters are adjusted. The same applies to Uzun Araz et al. (2024), who use a stochastic DLT in their application without describing its effect on the buffer levels or the performance of the system.

In most cases, when calculating DDMRP thresholds, the LT is treated as a management constant that is known before the release of production orders (Hopp & Spearman, 2011). In contrast to the existing DDMRP literature, where the LTs are considered as given and already known parameters (Damand et al., 2022a), this paper treats the LT as a management constant while proposing a systematic procedure to calibrate it.

5.2.3 Order Prioritization in DDMRP Flowshops

Regarding production order prioritization in DDMRP environments, Ptak and Smith (2019), the co-creators of the method, recommend the following: the order with the smallest percentage of the Net Flow Position (NFP) over the Top of Green (TOG) has the highest priority to be released to the first workstation in the flowshop (with a First In First Out – FIFO – basis in the case of equality). Then, a rule based on the On-Hand buffer penetration ratio (i.e., the On-Hand divided by the Top of Red (TOR)) is applied in the subsequent workstation queues. However, this rule does not account for the presence of rush orders in the system.

Several studies in the DDMRP literature prioritize replenishment orders using a combination of dispatching rules, such as the NFP for the first workstation and FIFO for the subsequent workstations (Azzamouri et al., 2022; Dessevre et al., 2021; Dessevre, Martin, et al., 2019;

Kortabarria et al., 2018; Shofa et al., 2018). While this approach is straightforward, its reliance on FIFO for most of the production line is a limitation, as it does not account for the urgency or inventory status of different orders; hence, its applicability for rush orders is not well established. This results in sequencing problems, especially with rush orders.

Other studies explore various order prioritization approaches in DDMRP (Miclo et al., 2015; Thürer et al., 2020; Velasco Acosta et al., 2019), but there is no consensus among them. For instance, Fernandes et al. (2024) compared several dispatching rules in a DDMRP flowshop, including the buffer penetration ratio rule, the NFP rule, and the Earliest Operation Due Date rule. The results show that this rule outperforms in high demand mix scenarios. However, a key limitation is the dependence on externally set due dates. In case of rush orders, aiming to have the smallest possible completion time, externally setting the due dates is not feasible, as it conflicts with the core concept of rush orders of being delivered as soon as possible.

The current DDMRP literature shows that the dispatching rule proposed by Ptak and Smith (2019) is not suitable for complex applications (Dessevre et al., 2023), complex product structures (Thürer et al., 2020), or highly volatile environments (Alves et al., 2024). This limitation raises questions about the robustness of DDMRP in diverse operational contexts. Despite the use of several different dispatching rules in DDMRP, the apparent consensus is to use NFP for the order release (i.e., the first workstation) and FIFO for the subsequent workstations.

5.2.4 Management of Rush Orders

In manufacturing, rush orders are production orders that are given the highest priority to meet urgent customer needs and delivery time requirements (Mundt et al., 2022). While most of the literature treats rush orders as special cases or disturbances in the production process (Trzyna et al., 2012), it has been shown that prioritizing certain orders and reducing their FTs does not affect the average FT or the waiting time of the whole system (Jäger & Roser, 2018).

Rush orders in manufacturing systems have been studied from two different perspectives: rush order acceptance and rush order scheduling. Considering the acceptance aspect of rush orders, most research has focused on developing methodologies to support the acceptance decision. Aqlan et al. (2017) propose a methodology for accepting rush orders using the utility value of each order. Chen (2010) develops a heuristic model to justify the acceptance of rush orders, considering the rewards and additional costs of acceptance. While Hendry et al. (2008) suggest a reserve of 10-20% of

capacity for rush order acceptance, Thüerer et al. (2009) provide evidence that maintaining a rush order share of up to 30% ensures the stability of FTs. More recently, Mundt et al. (2022) developed a method that reserves production capacity for rush orders, allowing firms to accept rush orders without overloading production. Unlike studies considering the acceptance of rush orders as a managerial decision, this paper treats it as a requirement of the system.

Research dealing with the rush order scheduling aspect has mainly focused on developing algorithms to improve production rescheduling in complex production arrangements. For example, Ren et al. (2021) use genetic algorithms in a job shop setting. He et al. (2019) employ a sorting genetic algorithm in a multistage hybrid flowshop. Milne and Wang (2014) propose a method to improve algorithms considering dynamic changes in demand priority to better allocate resources in such situations. When studying flowshops, the rush order scheduling is performed through prioritization using dispatching rules (i.e., a predefined rule that determines in which sequence released production orders should be processed at each workstation) with varying extents of prioritization in the production line (Neuner et al., 2022; Thüerer et al., 2009; Trzyna et al., 2012).

5.3 Methods

In this section, as demonstrated by Figure 5.2, we first present the case study and the hypothesis of the discrete event simulation model of a DDMRP flowshop inspired by Dessevre et al. [13] and designed using ARENA 16.20. Then, we describe the experiments to be carried out. Finally, we explain the procedure proposed to calibrate a DDMRP model and to calculate the Lead Time (LT) depending on the target service level.

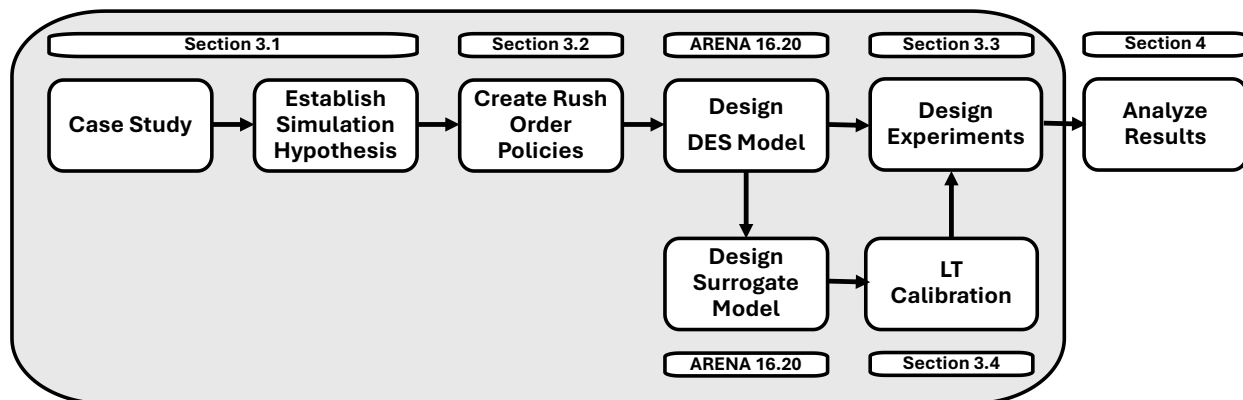


Figure 5.2 Methodology Steps

5.3.1 Case Study

In this study, we consider a six-machine (M1 to M6) non-preemptive flowshop where 30 finished products are managed with DDMRP buffers (see Figure 5.3), with Table 5.1 presenting the ADU for each Product (P1 to P30). The buffers are dimensioned according to Eqs. 1 to 6. The Variability Factor (VF) value is 0.50 for all products, and the Lead Time Factor (LTF) is 0.75, 0.5, and 0.25 for small, medium, and long processing times, using the ranges proposed by Ptak and Smith (2019).

$$\text{Red Zone} = \text{ADU} \times \text{DLT} \times \text{LTF} + \text{ADU} \times \text{VF} \quad (5.1)$$

$$\text{Yellow Zone} = \text{ADU} \times \text{LT} \quad (5.2)$$

$$\text{Green Zone} = \text{ADU} \times \text{LT} \times \text{LTF} \quad (5.3)$$

$$\text{Top of Red (TOR)} = \text{Red Zone} \quad (5.4)$$

$$\text{Top of Yellow (TOY)} = \text{TOR} + \text{Yellow Zone} \quad (5.5)$$

$$\text{Top of Green (TOG)} = \text{TOY} + \text{Green Zone} \quad (5.6)$$

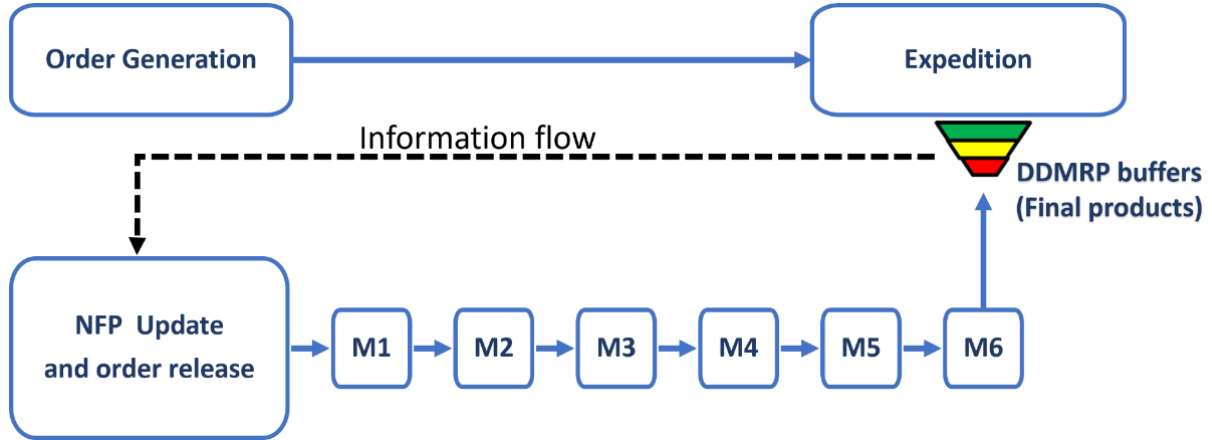


Figure 5.3 The studied flowshop

Production orders are generated according to DDMRP policy: a new production order is created if the Net Flow Position (NFP) is below the Top of Yellow (TOY) with a variable production lot size corresponding to a replenishment to the Top of Green (TOG). The NFP is calculated according to Eq. 5.7.

$$\text{On - Hand} + \text{On - Order} - \text{Qualified Demand} = \text{NFP} \quad (5.7)$$

where the On-Hand is the quantity of stock physically available (i.e., the current inventory level), the On-Order is the quantity ordered but not yet produced or received, and the Qualified Demand

is the sum of sales orders past due, sales orders due today, and qualified spikes. The order release priority depends on the percentage of the NFP over the TOG. If two orders have the same priority, the First In First Out (FIFO) rule is used.

Table 5.1 Average Daily Usage (ADU) in units

Product	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10
ADU	65	23	55	63	19	38	29	33	61	34
Product	P11	P12	P13	P14	P15	P16	P17	P18	P19	P20
ADU	61	62	57	32	52	21	61	42	48	44
Product	P21	P22	P23	P24	P25	P26	P27	P28	P29	P30
ADU	35	52	23	56	55	37	32	51	55	52

Following the discrete event simulation model developed by Dessevre et al. (2021), we assume that the arrival of customer orders follows an exponential distribution. The arrival rate depends on the product: 10 products have an arrival rate of 1/8 hours, 10 products have an arrival rate of 1/4 hours, and the last 10 have an arrival rate of 1/2 hours. The quantities requested by customer orders are calculated to meet the ADU. The visibility of customer orders is restricted to the next simulation day; the order spike threshold and the order spike horizon are not considered to calculate the qualified demand. If an order cannot be fulfilled from the on-hand inventory, it is back-ordered to be fulfilled as soon as possible.

In a flowshop with identical routing for all products, it is better to consider a single LT for all products than different LTs because the Flow Time (FT) is related to the queue size and processing time of all products (Dessevre, Martin, et al., 2019). Therefore, in this study, we consider the same LT for all products. The procedure used to calibrate the LT is described in Section 3.4. Initial processing times are given in Table 5.2. In each simulation, processing times are randomly adjusted with a coefficient following a triangular law TRI (0.8; 1; 1.2).

Table 5.2 Processing Times (seconds)

Products / Machines	M1	M2	M3	M4	M5	M6
P1–P10 (SMALL)	5	5	5	10	15	5
P11–P20 (MEDIUM)	5	10	10	10	10	10
P21–P30 (LONG)	5	15	15	10	5	15

Transfer times between different processes are not considered, nor are changeover times between products. Buffers are initialized randomly between 50% and 100% of the TOG. The DDMRP parameters for buffer sizing and qualified demand calculation are set at the beginning of each replication and are not updated during the experiment (i.e., static parameterization).

5.3.2 Integration of Rush Orders

To assess the impact of rush orders on DDMRP performance, we adapted the basic model of Dessevre et al. (2021) in a way to add rush orders and distinct rush order prioritization policies different from the dispatching rules proposed by Ptak and Smith (2019).

Rush orders

At the start of each replication, we randomly define a subset of products as rush products. Whenever a replenishment order for a rush product is created, it is considered a rush order. This avoids scenarios where two orders of the same product in the same queue have different statuses (rush and non-rush). We define the ‘*ratio of Rush Orders*’ as the percentage of rush product demand that occupies the capacity of the bottleneck machine (in our case, M6 in Figure 5.3). This ratio ranges from 0% to 50% of the demand in increments of 5%. Before each replication, considering the ADU for each product, the list of rush products is determined by randomly adding a product until the expected bottleneck occupancy percentage is reached, with a tolerance of 0.05%. By randomly assigning the list of rush products before each replication, we removed any potential impact that characteristics such as frequency, demand size, or processing times might have on the effects of prioritization for the experiments evaluated.

Rush Order Prioritization policies

In this study, we propose to maintain the apparent consensus to dispatch non-rush orders (i.e., combining NFP and FIFO) and to dispatch rush orders by giving priority to them either at the order

release or at all machines. To this end, we use the two rush order prioritization policies from the literature: first, a *Process Priority (PP) Prioritization Policy* that prioritizes rush orders in all machines on the shop floor (as in Trzyna et al. (2012) and Neuner et al. (2022)), and second an *Order Release (OR) Prioritization Policy* that prioritizes rush orders in the pool of orders to be released first (as in Thürier et al. (2009)).

We consider distinct queues to operationalize the policies, as illustrated by Figure 5.4. In the first machine, there are two queues, one for each group (rush orders and non-rush orders). The rush order queue (in red) always has priority over the non-rush order queue (in blue), meaning that products on the non-rush order queue are processed only if the rush order queue is empty. At the subsequent machines, the queues vary according to the policy, with specific queues for rush and non-rush orders or with a single mixed queue (in green).

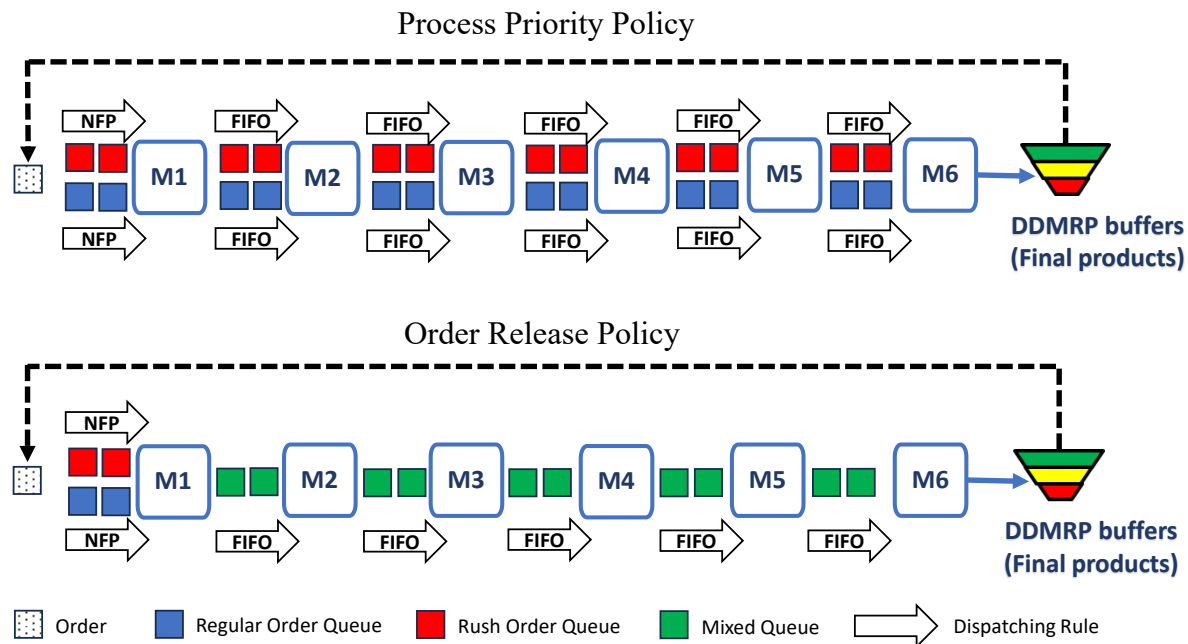


Figure 5.4 Process Priority (PP) and Order Release (OR) Rush Order Prioritization policies

We can see in Figure 5.4 that the PP policy has a larger extent of prioritization in the production line, with rush orders having their specific queue with the highest priority for all machines (M1 to M6) in the flowshop. The NFP dispatching rule is only used on the first machine. For the subsequent machines, both the rush orders queue and the non-rush orders queue are dispatched separately on a FIFO basis. In contrast, the OR policy has a lower extent of prioritization, with rush

orders having their prioritized queue (dispatched using the NFP rule) for the first machine only. For the subsequent machines, rush and non-rush orders follow a single mixed queue, dispatched on a FIFO basis.

5.3.3 Experiments

We consider a full factorial design including two factors characterizing the DDMRP system (the loading rate and the LT level) and two factors characterizing the rush orders (the rush order prioritization policy and the ratio of rush orders). Consequently, we have nine different DDMRP systems, with each DDMRP system being associated with a couple (loading rate, LT level). Table 5.3 summarizes the values taken by the varied factors. The procedure to determine LT values is given in Section 3.4.

Table 5.3 DDMRP system and rush orders' characteristics

DDMRP system		Rush Orders	
<i>Loading rate</i>	<i>LT level</i>	<i>Rush Order Prioritization Policy</i>	<i>Ratio of rush orders</i>
80%, 90%, 95%	Low, Medium, High	No prioritization (baseline), Process Priority (PP), Order Release (OR)	0% (baseline), 5%, 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, 50%

For each DDMRP system associated with a couple (loading rate, LT level), 20 experiments (2 rush order prioritization policies with 10 different percentages of Rush Orders) are compared to a baseline (an experiment with 0% rush orders and no rush order prioritization policy). Thus, a total of 189 (21 x 9) experiments are conducted. To compare DDMRP performance over these 189 experiments, we evaluate three performance indicators:

Production Service Level (PSL): Inspired by Hopp and Spearman (2011), this indicator measures the probability of the FT being smaller than the LT, which is defined as ‘the time allotted for producing a part on a line’ (see Appendix B). In the results analysis, we differentiate PSL for rush orders (PSL_r), for non-rush orders (PSL_{nr}), and the average PSL (PSL_a).

Customer Service Level (CSL): This indicator measures the percentage of customers' orders filled on time directly from the buffer. In our analysis, we differentiate CSL for rush orders (CSL_r), non-rush orders (CSL_{nr}), and the average CSL (CSL_a).

Average Inventory (I): average of finished products in the buffer for the period.

We consider 5 operating days a week with 1 shift of 8 hours per day. Each set of experiments has a loading rate target (80%, 90%, or 95%) for the bottleneck process, which is achieved by modulating the opening hours, only affecting the working hours of the machines, as the shop continues to receive orders during the 8-hour horizon. Considering 30 replications for each experiment and 95% confidence intervals on estimates, the results were collected during 550 simulation days, after 50 days of warmup.

5.3.4 Model and Lead Time Calibration

To calibrate the simulation model and to avoid potential influences of insufficient LT in the results, we need to set the LT levels depending on the loading rate setting for each experiment. Since the service level is influenced by manipulating the LT, the defined low LT for experiments with an 80% loading rate should be smaller than the low LT for experiments with a 90% loading rate.

In real applications, the LT is manipulated at different loading rates to achieve service level targets. In the same spirit, we propose to use a surrogate model to calibrate different levels of LT for each loading rate. We consider a low LT as a value that is greater than or equal to 60% of all the observed FTs in each scenario, resulting in at least a 60% SL. Similarly, we assume a medium LT as a value that is greater than or equal to 80% of the observed FTs and a high LT as a value that is greater than or equal to 90% of the observed FTs.

In the literature, the use of a surrogate model is considered a viable approach as a decision-making technique (Chua et al., 2022) and for parameterizing production control systems (Werth et al., 2023). Particularly, the FT can be analyzed through simulation to estimate LT (Pfeiffer et al., 2016). Since the PSL is the probability of the FT being less than or equal to the LT (Hopp & Spearman, 2011), for each loading rate, an appropriate LT value for the desired PSL is a value that is greater than a certain percentage of the total number of FT observations (i.e., to achieve an 80% SL, the LT must be greater than 80% of the FT observations). Therefore, we developed a surrogate model (i.e., a simplified representation of the simulation model) to observe several instances of the FT and to identify the values associated with the three LT levels (Low LT, Medium LT, and High LT). We then validated that the surrogate model, using these LT values, allowed us to reach the expected PSL levels for each loading rate. Details of how we used the surrogate model approach to calibrate the LT levels are in Appendix C.

Table 5.4 and Figure 5.5 present, for each loading rate, which LT values should be considered as Low, Medium, and High based on our assumption about the LT. For example, when experimenting with an 80% loading rate, the values 8.12h, 9.52h, and 10.64h are used, respectively, for Low LT, Medium LT, and High LT. Figure 5 also validates that as the loading rate approaches 100%, LTs approach infinity. This confirms Dessevre et al. (2021) observations that at higher loading rates, managers should refrain from adjusting the LT, as this results in elevated stock levels.

Table 5.4 LT levels according to the loading rate

	Loading Rate				
LT levels	80%	85%	90%	95%	99%
Low (PSL.60%)	8.12h	8.72h	9.72h	11.36h	25.20h
Medium (PSL.80%)	9.52h	10.24h	11.64h	13.76h	33.60h
High (PSL.90%)	10.64h	11.52h	13.20h	16.08h	44.80h

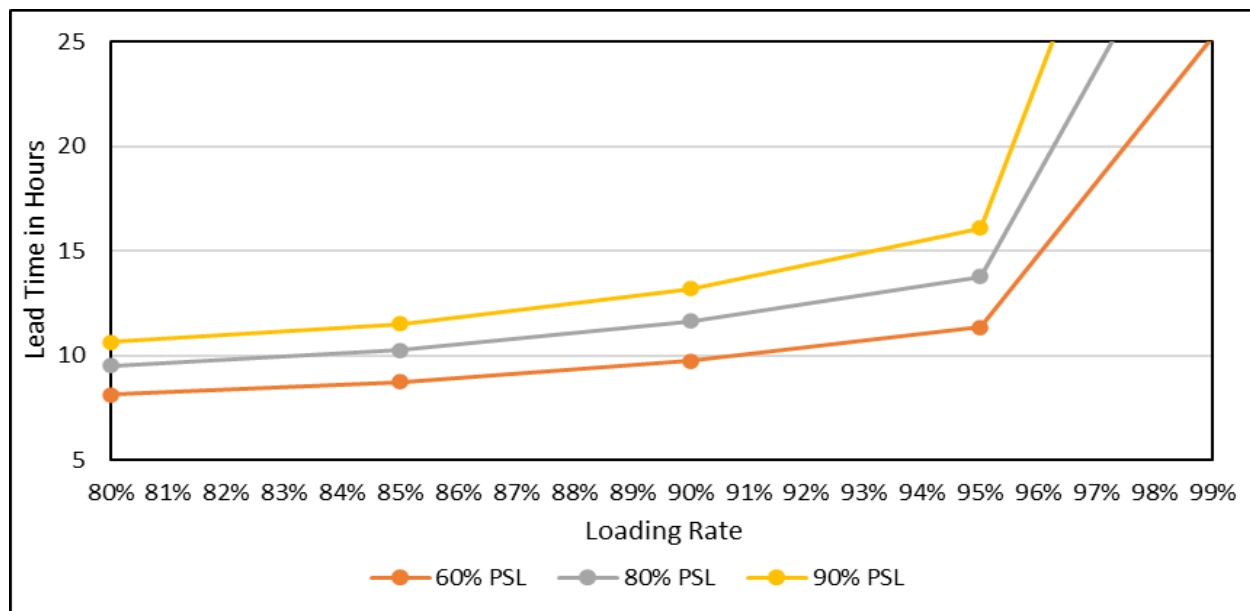


Figure 5.5 Evolution of the LT values according to the loading rate

5.4 Results

In this section, we present the results and analyze the robustness of the DDMRP flowshop when faced with rush orders. In our study, robustness refers to the ability of the system to maintain certain stability in the service level as the ratio of rush orders increases.

5.4.1 Fulfillment of Non-Rush Orders

Figure 5.6 presents the PSL for non-rush orders (PSL_{nr}) and the average PSL (PSL_a), respectively presented by red and orange dots. The PSL_a for the baseline (0% rush orders) is presented by the first black dot in each graph. Figure 5.7 presents the CSL for non-rush orders (CSL_{nr}) and the average CSL (CSL_a), respectively presented by red and orange dots. The CSL_a for the baseline (0% rush orders) is presented with black dots. Each of the nine main figures corresponds to a DDMRP system characterized by a couple (loading rate, LT level) along with three DDMRP buffer zones. These zones, illustrated as the sum of all the zones, serve as a reference for the average inventory (I) measured with the secondary axis on the right. In both Figures 5.6 and 5.7, the average inventory (I) is presented as a yellow line. The PSL and CSL of each DDMRP system are evaluated with the PP policy (left side) and OR policy (right side). Indicators are shown with 95% confidence intervals, marked by black dashes above and below.

We can observe in both Figures 5.6 and 5.7 that the average inventory (I) stays in the yellow zone, which is consistent with the DDMRP theory. The average inventory (I) depends on both the LT and the loading rate. As either the LT or the loading rate increases, the inventory also rises. However, the average inventory (I) remains completely independent of the rush order ratio and the extent of prioritization in the production line. As shown in the figures, neither changing the rush order ratio nor increasing the extent of prioritization affects the average inventory (I).

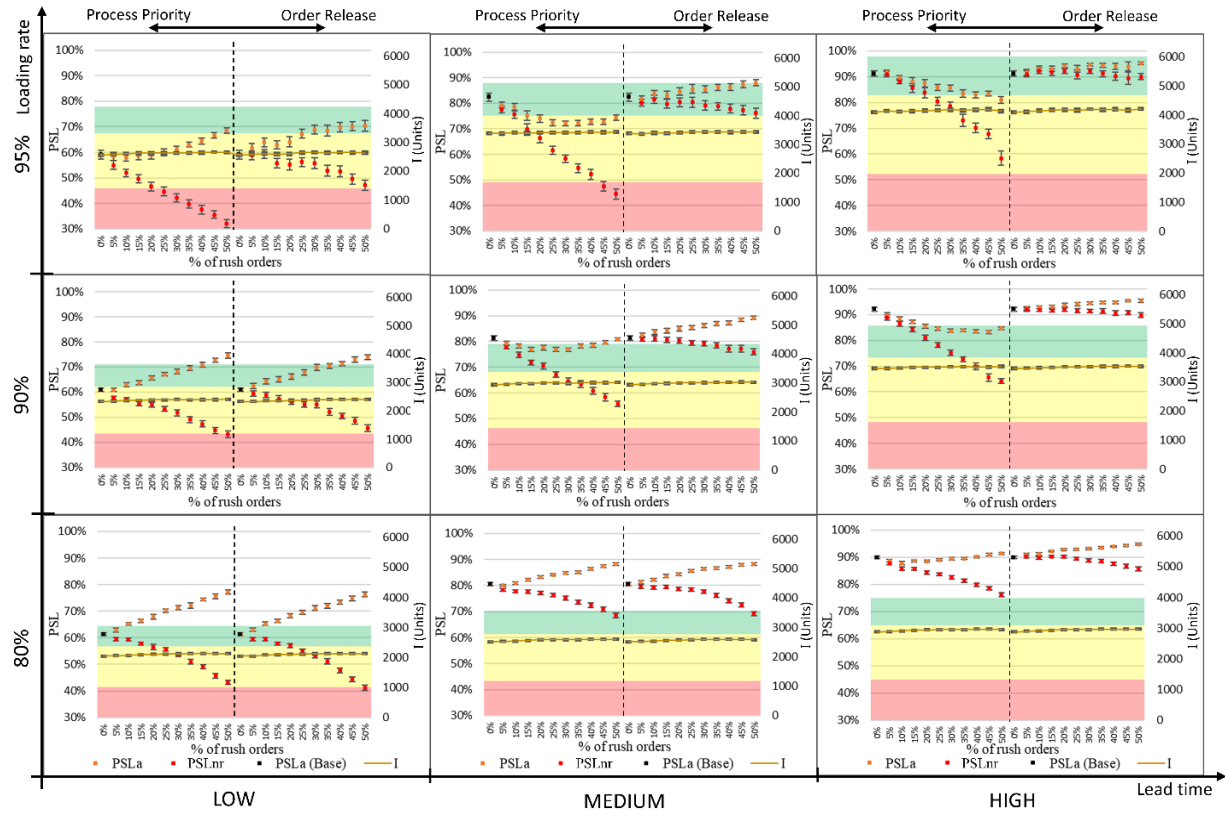


Figure 5.6 Average Inventory Positioning (I) vs. Production Service Levels for non-rush orders (PSLnr) and the whole system (PSLa)

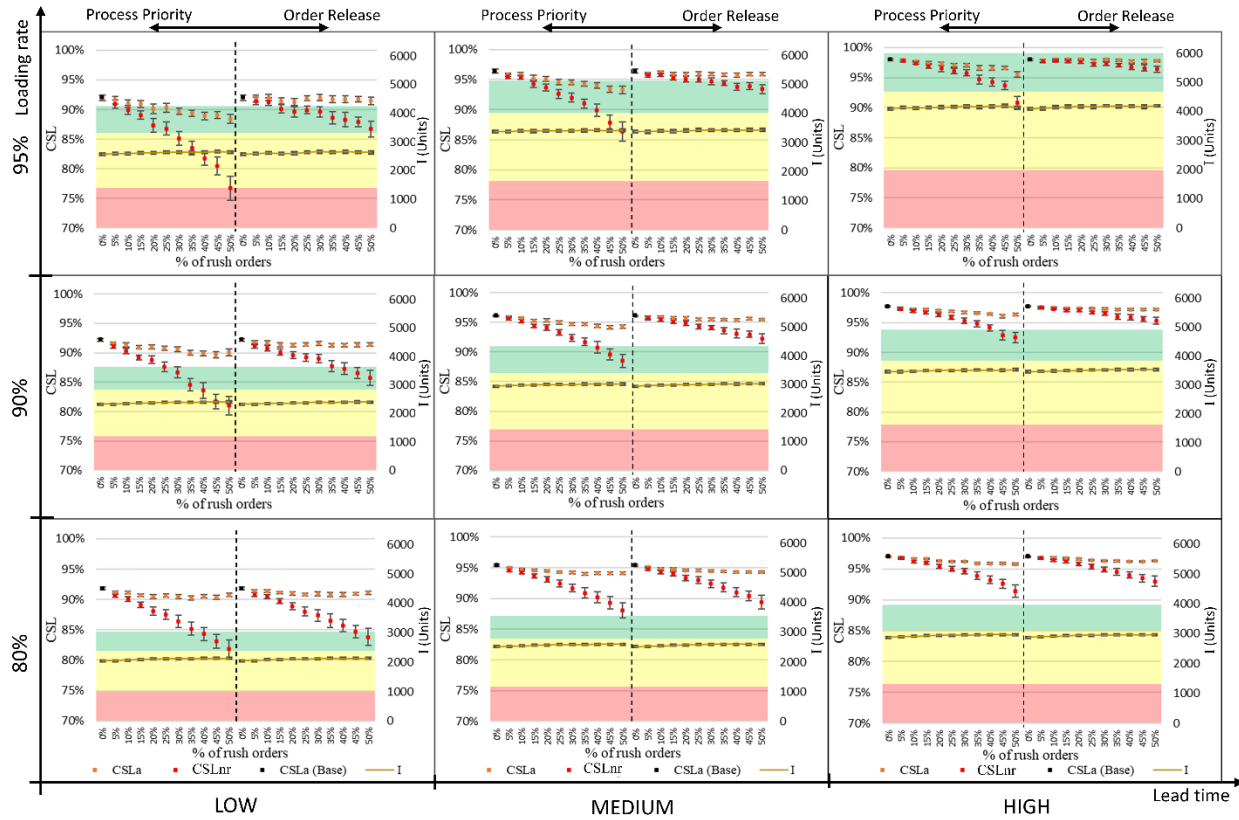


Figure 5.7 Average Inventory Positioning (I) vs. Customer Service Levels for non-rush orders (CSLnr) and the whole system (CSLa)

Figures 5.6 and 5.7 confirm the effectiveness of the DDMRP in maintaining higher levels of CSL despite the decline in the PSL. Under the baseline condition (0% rush orders), the nine DDMRP systems maintain the CSL above 90% even with PSL performance at 60%. When the PSLnr goes down to 30% (e.g., a 95% loading rate and a low LT case presented at the top left of Fig. 6 and Fig. 7), the CSLnr remains over 75%, regardless of the extent of prioritization in the production line. With small ratios of rush orders, the DDMRP systems perform exceptionally well without compromising the service levels of non-rush orders. These observations highlight the benefits of integrating DDMRP and rush order prioritization policies in handling rush orders within a DDMRP flowshop.

As expected, in every experiment, increasing the ratio of rush orders reduces the ability of the system to fulfill non-rush orders. However, PSLnr and CSLnr decrease very slowly up to 20% of the rush order ratio. For example, in low LT settings – the worst-performing experiments (see left of Fig. 7) – when using the OR policy up to a 20% rush orders ratio, the PSLnr decreases only 5% (from 60% to 55%).

The decline in the PSL seems more pronounced when using larger extents of prioritization (i.e., the PP policy), especially in 90% and 95% loading rate experiments (see top-left of Fig. 6). For the DDMRP system characterized by a loading rate of 90% and medium LT (see center Fig. 6), when the rush order percentage increases from 0% to 50%, the PSL_{nr} decreases from 80% to 55% with the PP policy and only from 80% to 76% with the OR policy.

In scenarios with low LTs, we note an increase in the PSL_a whenever we increase the ratio of rush orders. In systems where the LT is not sufficient, any type of priority in the system increases the PSL_a. When non-rush orders are always late, introducing a certain set of orders that are rarely late (i.e., rush orders) increases the PSL_a.

5.4.2 Fulfillment of Rush Orders

The indicators for the performance of rush orders are presented by blue dots in Figure 5.8 and Figure 5.9 for PSL_r and CSL_r, respectively. As in Figures 5.6 and 5.7, each of the nine main figures corresponds to a DDMRP system characterized by a couple (loading rate, LT level). The PSL_r and CSL_r are evaluated with the PP policy (left part) and OR policy (right part). The indicators are presented with 95% confidence intervals marked by black dashes above and below each point. Compared to the baseline (0% rush orders), the rush order prioritization policies achieve significant results in fulfilling rush orders without compromising the fulfillment of non-rush orders, thus maintaining robustness.

As expected, the performance of rush orders comes at the expense of the performance of non-rush orders. Unlike with non-rush orders, the extent of prioritization in the production line did not affect PSL_r and CSL_r up to a rush order ratio of 25%. Even with higher ratios of rush orders, the nine DDMRP systems maintain rush order service levels close to 100%. Both policies enable a PSL_r of over 90% and a CSL_r exceeding 95% for all experiments. The only larger decline observed happened to the PSL_r in the scenario with a 95% loading rate and low LT. To summarize, the extent of prioritization (i.e., the use of the PP policy or the OR policy) does not influence the fulfillment of rush orders, except for the DDMRP system with low LT and 95% loading rate. In this particular case, the PP policy outperforms the OR policy mainly because more priority points reduce the waiting time of rush orders in systems with longer queues.

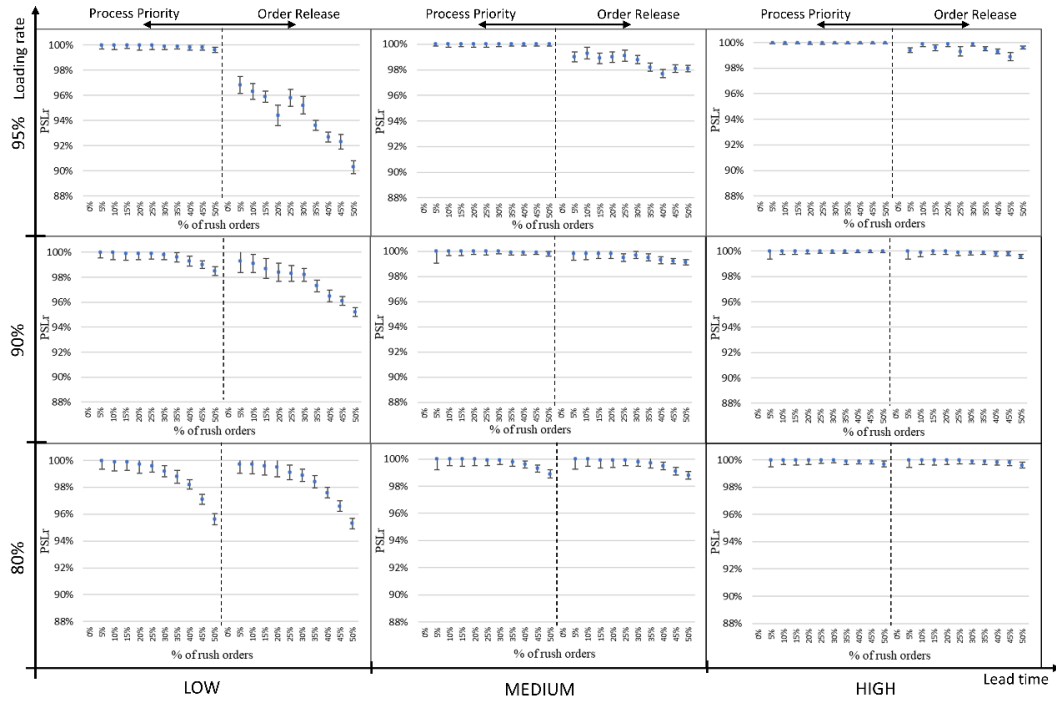


Figure 5.8 Rush Orders Production Service Level (PSLr) Variation

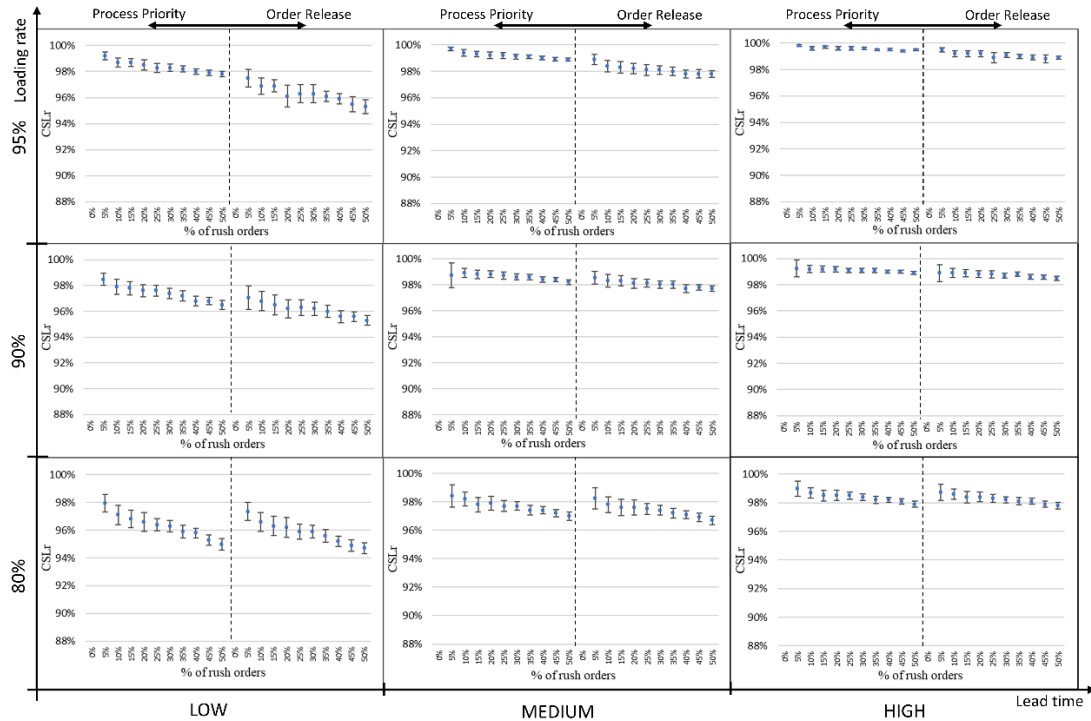


Figure 5.9 Rush Orders Customer Service Level (CSLr) Variation

5.4.3 Loading Rate Effects

For the baseline (0% of rush orders), the performance measured by the PSLa or CSLa is independent of the loading rate, regardless of the LT value. As the ratio of rush orders increases, the loading rate affects the extent to which the performance of non-rush orders decreases. When the loading rate increases, Figure 5.6 (and similarly in Figure 5.7) shows that the deviation of the PSLnr from the PSLa (and similarly the CSLnr from the CSLa) increases with the PP policy, whereas this deviation is reduced with the OR policy.

Regarding the fulfillment of rush orders (see Figures 5.8 and 5.9), we observe that the PP policy fulfills rush orders better as the loading rate increases (CSLr is close to 100% for loading rates of 90% and 95%), which is expected since this policy has a larger extent of prioritization. At low LT settings, increasing the loading rate reduces the service level of rush orders (evaluated by PSLr and CSLr) for the OR policy, but increases it for the PP policy. This is due to the longer queues resulting from the higher loading rates benefiting the policy with a larger extent of prioritization (i.e., the PP policy), especially when the processes are already occupied by a rush order.

5.4.4 Lead Time Effects

For the baseline (0% of rush orders), PSLa and CSLa increase as LT levels increase, regardless of the loading rate. In systems with rush orders, higher inventories, resulting from increasing the LT, lead to a reduction in the gap between the average performance (PSLa, CSLa) and the performance of the non-rush orders (PSLnr, CSLnr). Similarly, increasing LT improves the service levels of rush orders, regardless of the loading rate.

There is an important relationship between the LT and the robustness of the system. At low LTs, there are no experiments where the PSLnr equals the PSLa. However, as the LT increases, the PSLnr converges to the PSLa at low rush order ratios, indicating that the presence of rush orders in the system does not affect its robustness. These results highlight the connection between LT (and hence the inventory level) and the impact of the extent of prioritization in the production line on the performance of non-rush orders (and hence the system robustness).

5.4.5 Ratio of Rush Orders Effect

Increasing the ratio of rush orders obviously reduces the ability of the system to fulfill non-rush orders, impacting its robustness. Regarding PSL (see Fig. 6), when evaluating the OR policy, the

PSLa increases as the ratio of rush orders increases. In contrast, the PP policy does not show the same trend, as the PSLa decreases in medium and high LT experiments at 90% and 95% loading rates. This occurs because several late non-rush orders are replaced with on-time rush orders, ultimately increasing the flowshop average performance by raising the lateness of the already late non-rush orders.

Considering the CSL (see Fig. 7), CSLa remains stable, except for the PP policy at a 95% loading rate, due to DDMRP replenishment. For a higher ratio of rush orders, first, the results show that CSLnr is close to CSLa with medium and high LTs. Second, the OR policy resulted in a smaller decrease in CSLnr when compared to the PP policy. We conclude that the PP policy is more ‘costly’ to the rest of the system than the OR policy. Although the OR policy had a slightly lower CSLr than those obtained with the PP policy, the OR policy achieved a better trade-off between service levels for rush and non-rush orders.

5.4.6 Impact of Order Prioritization Policies on DDMRP Robustness

Figures 5.6 and 5.7 illustrate how the performance of non-rush orders is directly related to the loading rate in the DDMRP systems with rush orders. The two policies achieve equivalent performance at an 80% loading rate. At higher loading rates, the use of the OR policy (i.e., small prioritization extent) increases the PSLa as the rush order ratio increases, regardless of the LT. This behavior is not observed when the PP policy is employed (i.e., large prioritization extent).

The extent of rush order prioritization is critical to the robustness of the DDMRP system. Unlike the PP policy, the OR policy allows the system to accept more rush orders while maintaining system robustness. For the worst performing scenarios in terms of average inventory (I) (95% loading rate and high LT – see Figures 5.6 and 5.7, top right), PSLnr is close to PSLa for up to 20% of rush orders with the PP policy and for up to 25% of rush orders with the OR policy. Similarly, CSLnr is close to CSLa for up to 25% of rush orders with the PP policy and for up to 45% of rush orders with the OR policy.

To further evaluate the impact of the extent of prioritization in the production line on the system robustness, Figure 5.10 presents boxplots of the performance of both policies considering the different loading rates. The two graphs at the top of the figure evaluate the rush orders performance indicators CSLr and the PSLr, whereas the two graphs at the bottom assess the system robustness through the PSLa and the PSLnr. Each box presents the distribution of the policy performance. The

black line in the boxes illustrates the median, while the 'X' corresponds to the average value (the exact number is written in red and positioned just below the lower whisker). The lower section of the box corresponds to the first quartile of the sample, and the upper section corresponds to the third quartile. The whiskers range from the minimum to the maximum values, excluding the outliers. A point corresponds to the average result with an LT level and a ratio of rush orders (we have a total of 30 points corresponding to 3 LT levels and 10 different percentages, but few of them may be stacked on top of each other if they have the same value).

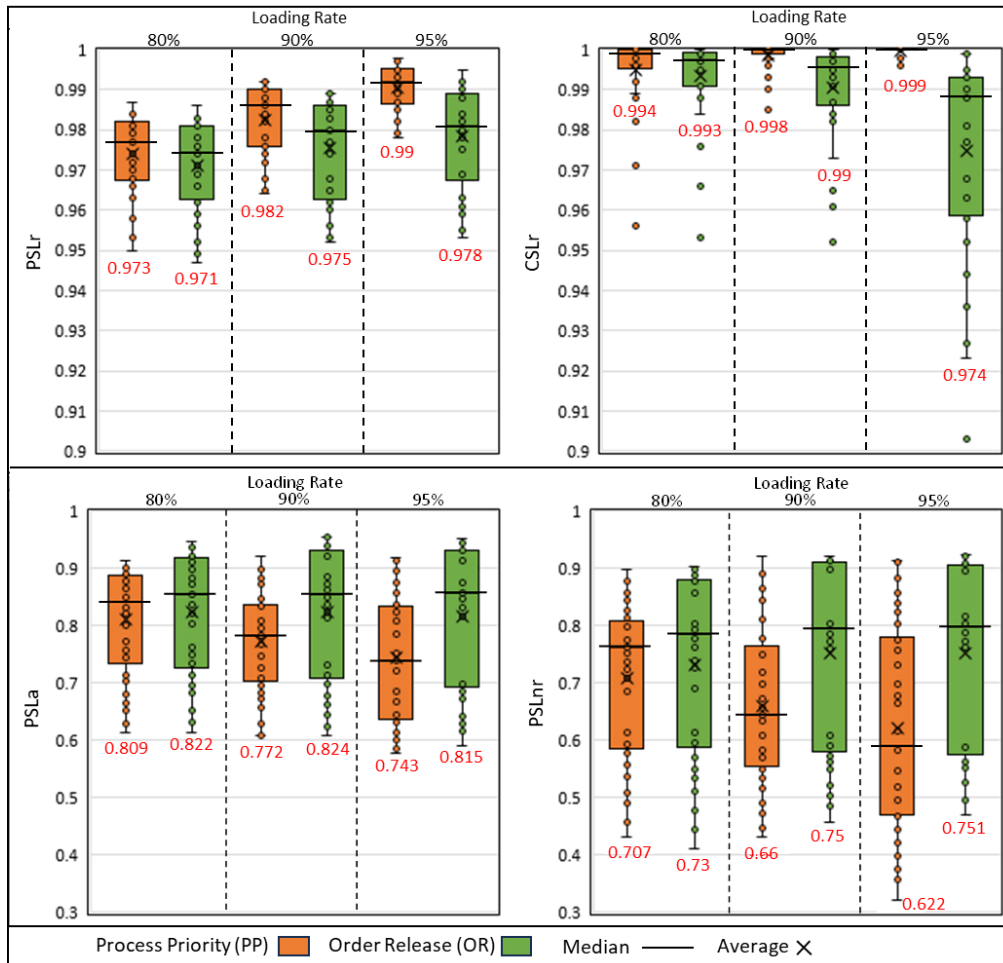


Figure 5.10 Comparison of PP and OR rush order prioritization policies

First, we can see that for the 80% loading rate case, the medians of each indicator obtained with the PP policy lie inside the respective OR policy box and vice versa. It can therefore be concluded that at an 80% loading rate, the extent of rush order prioritization has no impact on the robustness of the DDMRP system. However, at 90% and 95% loading rates, the medians of the PSLr and CSLr with the PP policy are above the corresponding OR policy box. It can therefore be concluded

that in a DDMRP flowshop with higher workloads, a larger prioritization extent (PP policy) outperforms the smaller prioritization extent (OR policy) in terms of service levels for rush orders due to the presence of more prioritization points. However, the system is more robust with a smaller prioritization extent (OR policy). This can be seen in Figure 5.10, as the medians of PSLa and PSLnr with the OR policy are above the PP policy box for loading rates of 90% and 95%.

5.5 Discussion

This study investigated the robustness of a DDMRP flowshop in the presence of rush orders, a known operational challenge not yet addressed in the DDMRP literature. Our findings demonstrate that the extent of prioritization is a more critical factor for maintaining robustness than the ratio of rush orders in the system. The results show that a policy limiting the prioritization of rush orders improves robustness, a finding that adds nuance to the broader literature on rush order management, which often focuses on reserving capacity (Hendry et al., 2008; Thürer et al., 2009) or complex rescheduling algorithms (Aqlan et al., 2017; Chen, 2010) rather than the scope of the prioritization itself. The following sections detail the theoretical and practical contributions of the work.

5.5.1 Theoretical Contributions

Current DDMRP literature has a focus on planning aspects like dynamic parametrization (Duhem et al., 2023; Martin et al., 2023) and buffer profiling (Azzamouri et al., 2022; Favaretto et al., 2023), leaving significant gaps regarding the execution phase (Alves et al., 2024). This study contributes by addressing a critical execution issue: the integration and management of rush orders. While previous research on DDMRP execution often assumed homogeneous orders and preemptive flowshops (Fernandes et al., 2023; Fernandes et al., 2024), our work provides an analysis in a more complex environment, broadening the applicability of DDMRP.

Our findings contribute to the ongoing discussion on order prioritization in DDMRP systems, where the efficacy of standard dispatching rules has been questioned in complex settings (Alves et al., 2024; Fernandes et al., 2024). This paper evaluated two distinct prioritization policies: Process Priority (PP), which considers only product characteristics for prioritization, and Order Release (OR), which builds upon using the NFP and product characteristics to manage prioritization. The results demonstrate that the extent of prioritization is a critical factor for system robustness. We found that a smaller extent of prioritization (i.e., OR policy) makes the system more robust,

particularly at higher loading rates. This finding adds a new dimension to DDMRP literature, suggesting that extensive prioritization of rush orders throughout the production line is not always optimal and detrimental to the performance of non-rush orders.

Furthermore, this study provides empirical evidence on the capacity share allocation to rush orders within a DDMRP framework. Our results show that a DDMRP flowshop remains robust when rush orders constitute up to 20% of the capacity of the bottleneck. This aligns with broader production control literature and provides a DDMRP-specific context, where rush orders occupying up to 10-20% of the capacity have been suggested as a viable practice (Hendry et al., 2008).

Finally, this paper addresses a significant methodological gap in DDMRP research: the Lead Time (LT) calibration. While most studies treat LT as a predetermined constant (Damand et al., 2022a; Dessevre, Martin, et al., 2019), this approach is a known source of subjectivity in parametrization (Krajčovič et al., 2024). We proposed and tested a systematic procedure to calibrate LT based on target service levels and system load. This replicable methodology allows for standardizing the setting of an important parameter, facilitating more consistent future research in DDMRP.

5.5.2 Practical Contributions

For practitioners, this research provides direct, actionable guidance on how to manage rush orders in a DDMRP environment while maintaining high service levels for all customers. The primary practical insight is that limiting the scope of prioritization is key to success. We recommend that managers prioritize rush orders only at the initial order release stage (the OR policy) rather than at every machine in the flowshop. This approach provides the best trade-off, ensuring rush orders are expedited while protecting the fulfillment of non-rush orders and maintaining the overall stability of the system, especially under high occupations.

This study demonstrates that managers can accept rush orders up to a certain threshold without destabilizing their operations. The finding that the system remains robust with a rush order ratio of up to 20% gives practitioners a concrete benchmark for accepting rush orders without jeopardizing performance. This addresses the previous understanding that introducing rush orders invariably leads to system-wide instability and low service level (Arisha & Mahfouz, 2011). Results show that DDMRP buffers are effective in protecting the Customer Service Level (CSL) even with declining Production Service Level (PSL), which is a critical benefit to maintain or increase on-time delivery.

Moreover, the subjective nature of parameter setting is a known barrier to DDMRP implementation (Krajčovič et al., 2024). This paper offers a practical solution by presenting a systematic procedure for LT calibration in DDMRP environments. Practitioners can use this method to set appropriate LTs for their specific operational context, considering their unique system loads and target service levels. This data-driven approach reduces implementation complexity and improves the performance and reliability of the DDMRP system, enhancing its applicability in real-world manufacturing environments.

5.6 Concluding Remarks

This paper contributes to DDMRP literature by focusing on one execution issue: the prioritization of rush orders in a DDMRP system. By simulating a non-preemptive DDMRP flowshop, we first analyze the impact of the evolution of the ratio of rush orders on the fulfillment of non-rush orders, and second, we evaluate how different prioritization policies affect the robustness of the DDMRP system.

The simulation results highlight the importance of limiting prioritization in the production line to maintain high service levels for non-rush orders. Therefore, whenever the goal is to provide the best possible service to rush orders at all costs, the use of larger extents of prioritization (i.e., PP policy) is justified. Otherwise, smaller extents of prioritization (i.e., OR policy) maintain the robustness of the system and guarantee the best trade-off between overall performance and rush order fulfillment. This study also demonstrates that it is crucial to calibrate the LT to avoid degrading non-rush order fulfillment. For this purpose, we provide practitioners with a method for calibrating the LT, useful for DDMRP and other types of production control systems.

The findings offer valuable insights into the practical issues faced by DDMRP applications and provide a viable strategy for DDMRP flowshops to work with different LTs without compromising the overall system performance. While this study is based on a specific simulated flowshop, the findings and principles apply to other similar manufacturing contexts. The two main insights of this study are tied to the fundamental dynamics of flowshops and DDMRP concepts rather than specific parameters of the studied case. Therefore, other manufacturing flowshops managed by DDMRP are expected to benefit from these principles. Although precise performance metrics may vary, the strategic approach of limiting rush order prioritization to the order release stage and systematically calibrating lead times offers guiding principles for improving robustness. The

proposed solution for handling rush orders in DDMRP flowshops is simple enough to be applied in practice, leading to wider applicability of DDMRP.

Finally, this study presents limitations related to both the simulation model and the DDMRP scope. First, the model assumes a linear flowshop with identical routing and does not account for operation constraints often present in real settings, such as changeover times or machine breakdowns. Second, the DDMRP buffer parameters were static, whereas dynamic adjustment mechanisms could be employed to respond to demand trends. Future research should extend this robustness analysis to more complex production environments, such as hybrid flowshops, and evaluate the effect of dynamic buffers adjustments on the impact of rush orders.

CHAPTER 6 ARTICLE 3: SMART ORDER GENERATION IN DDMRP USING PREDICTED FLOW TIME

This article was submitted in March 22, 2026 to the journal Computers & Industrial Engineering (CAIE). The version presented in this thesis is identical to the submitted version.

The authors of this article are Lucas Alves^{*1}, Maha Ben Ali¹ and Pierre Baptiste¹.

* Main author. Responsible for conceptualization, methodology, software, validation, formal analysis, investigation, data curation, writing and visualization.

¹ Department of Mathematics and Industrial Engineering, Polytechnique Montréal, Montréal, Canada.

ABSTRACT

Demand Driven Material Requirements Planning (DDMRP) methodology has a clear process for replenishment order recommendation but lacks a formalized process for order approval. This gap often leads to automatic order approval, which can be inefficient due to the mismatch between the order ready-date and actual consumption. This paper aims to develop a smart DDMRP order generation process that leverages real-time production system information to achieve measurable improvements in service performance and inventory efficiency. Using a discrete event simulation of a DDMRP-managed flowshop integrated with a machine learning algorithm we estimate a predicted flow time for each recommended order. This metric is used as the main argument for a dynamic approval gate to conditionally approve or withhold orders based on the anticipated state of the shop floor. We test several order generation policies with varying lead time thresholds and compare their performance against traditional DDMRP and Constant Work In Process (CONWIP) order generation. The results demonstrate that policies which delay order generation when the system is underloaded are more effective than both traditional DDMRP and CONWIP order generation, achieving an average inventory reduction of 8.8% with virtually no cost for the service level. Conversely, policies that delay order generation when the system is overloaded are proved redundant. Our findings suggest that the DDMRP order generation process should evolve from a single-factor decision based on buffer status to a dual-factor decision incorporating the predicted state of the system.

RESUME

La méthodologie Demand Driven Material Requirements Planning (DDMRP) dispose d'un processus clair pour la recommandation des ordres de réapprovisionnement, mais manque d'un processus formalisé pour leur approbation. Cette lacune conduit souvent à une approbation automatique des ordres, ce qui peut s'avérer inefficace en raison du décalage entre la date de disponibilité de l'ordre et sa consommation réelle. Cet article vise à développer un processus intelligent de génération d'ordres DDMRP qui tire parti des informations du système de production en temps réel afin d'obtenir des améliorations mesurables de la performance du service et de l'efficacité des stocks. À l'aide d'une simulation à événements discrets d'un flow shop géré en DDMRP, intégrée à un algorithme d'apprentissage automatique, nous estimons un temps d'écoulement prédictif pour chaque ordre recommandé. Cette métrique est utilisée comme argument principal pour une porte d'approbation dynamique permettant d'approuver ou de retenir sous condition les ordres en fonction de l'état anticipé de l'atelier. Nous testons plusieurs politiques de génération d'ordres avec des *lead times* variables et comparons leurs performances à celles de la génération d'ordres DDMRP traditionnelle et du système Constant Work In Process (CONWIP). Les résultats démontrent que les politiques qui retardent la génération d'ordres lorsque le système est sous-chargé sont plus efficaces que la génération d'ordres DDMRP traditionnelle et CONWIP, permettant une réduction moyenne des stocks de 8,8 % sans quasiment aucun impact sur le taux de service. Inversement, les politiques qui retardent la génération d'ordres lorsque le système est surchargé se sont révélées redondantes. Nos conclusions suggèrent que le processus de génération d'ordres DDMRP devrait évoluer d'une décision à facteur unique basée sur l'état du buffer vers une décision à deux facteurs intégrant l'état prédictif du système.

Keywords: DDMRP, Order Generation, Machine Learning, Flow Time Prediction, Discrete Event Simulation.

6.1 Introduction

Contemporary production environments are characterized by the volatile, uncertain, complex, and ambiguous (VUCA) world. This results in significant challenges, including rapidly

changing market demands and increased uncertainty in supply chains (Martin et al., 2023). In such environments, Demand Driven Material Requirements Planning (DDMRP) emerged as an alternative to traditional planning and control methods by combining elements from Lean, Theory of Constraints, and MRP (Miclo et al., 2018). DDMRP aims to protect and promote flow and agility in the face of volatile and uncertain environments by using strategically decoupled inventory buffers (Ptak & Smith, 2019).

Central to DDMRP execution is the Net Flow Equation (NFE), which triggers the replenishment order generation process once an inventory buffer falls below a predefined threshold (Ptak & Smith, 2019). However, a critical gap remains in literature. While the original methodology proposed by Ptak and Smith (2019) distinguishes between triggering the order recommendation and the subsequent order approval, it does not formalize the approval step. The majority of DDMRP research assumes that recommended replenishment orders are automatically approved and released for execution (Alves et al., 2024; Duhem et al., 2023; Fernandes et al., 2024). However, there is no clear consensus in the literature or in practice regarding how recommended DDMRP orders should be approved.

Recently, CONWIP (Constant Work in Process) has been proposed as an interesting approach for both order recommendation and approval (Dessevre et al., 2023; Fernandes et al., 2025). However, while CONWIP may be effective for Make-to-Order (MTO) configurations, automatic order approval in Make-to-Stock (MTS) scenarios may increase inventory, as replenishment orders' ready dates may often be much smaller than their actual due dates. Furthermore, automatic release with CONWIP overlooks valuable information about the state of the production system, which should be used to manage the flow of orders to the shop.

The absence of a formal order-approval step in DDMRP research raises the question of whether real-time production data can improve standard order generation. This paper aims to develop a smart DDMRP order generation process that leverages real-time production system information to achieve measurable improvements in service performance and inventory efficiency. We first propose a novel smart order generation process that integrates a machine learning model to predict flow time and leverages these predictions within alternative order-approval logics. Second, we evaluate how these order approval logics affect performance by simulating three classes of order generation policies in a non-preemptive DDMRP flowshop.

Our results reveal that, compared to traditional DDMRP and CONWIP order generation, using predicted flow time to strategically withhold order approvals in underloaded systems reduces inventory with minimal impact on service levels.

This paper is organized as follows. Section 2 brings the related literature and highlights the research contribution to DDMRP literature. Section 3 details the structure of the standard and the novel order generation process. Section 4 describes the case study, the methods, and the experimentation. Section 5 presents and analyzes the results from our experiments. Finally, conclusion and research perspectives are presented in Section 6.

6.2 Related Literature

6.2.1 Context

To address the shortcomings of traditional planning and control methods, DDMRP was developed as a hybrid push-pull methodology guided by actual consumption. The core innovation is the use of strategically positioned inventory buffers at decoupling points (Ptak & Smith, 2019). Inspired by methods based on the theory of constraints, DDMRP buffers protect critical points from disruptions while improving production responsiveness (Takami Narita et al., 2021).

DDMRP is divided into five sequential phases for system design and operation (Ptak & Smith, 2019). Most of the DDMRP research focuses on the first three phases: strategic inventory positioning (Achergui et al., 2020; Habbadi et al., 2025), buffer profiles and levels (Krajčovič et al., 2024; Martin et al., 2023), and dynamic adjustments (Damand et al., 2022a; Duhem et al., 2023; Uzun Araz et al., 2024). The first three phases essentially determine the initial configuration of the DDMRP system and its evolution to absorb and mitigate variability in both demand and supply (Fernandes et al., 2024). The other two phases are Demand Driven Planning (DDP) and Demand Driven Execution (DDE), which focus on operational aspects, such as order generation and release, and their execution. The core objective of DDP is the generation of replenishment orders designed to keep inventory and buffers at defined levels (Fernandes et al., 2023). The research for these two phases is scarce. In the DDP phase, Azzamouri et al. (2022) studied the impact of continuous order generation in a DDMRP managed flowshop and Dessevre et al. (2023) proposed the use of CONWIP as the order generation process in a divergent flowshop managed by DDMRP. In the DDE phase, the

research focuses on studying sequencing rules to govern replenishment orders execution (Alves et al., 2024; Carmo-Silva & Fernandes, 2024; Fernandes et al., 2023).

As shown in Figure 6.1, a DDMRP buffer is divided into three zones: (1) the Green Zone which drives the frequency of order generation and the order size; (2) the Yellow Zone which represents the demand coverage in the buffer while replenishment orders are being fulfilled; and (3) the Red Zone which represents the safety embedded in the buffer. The size of these three zones is dependent on the item's Average Daily Usage (ADU) and Decoupled Lead Time (DLT), which, for non-decoupled product structures is always equal to the LT, as demonstrated by Alves et al. (2026). The buffer replenishment to the Top of Green (TOG) occurs whenever the Net Flow Position (NFP) is below the Top of Yellow (TOY). The priority of the replenishment order is determined by the distance of the NFP to the TOG.

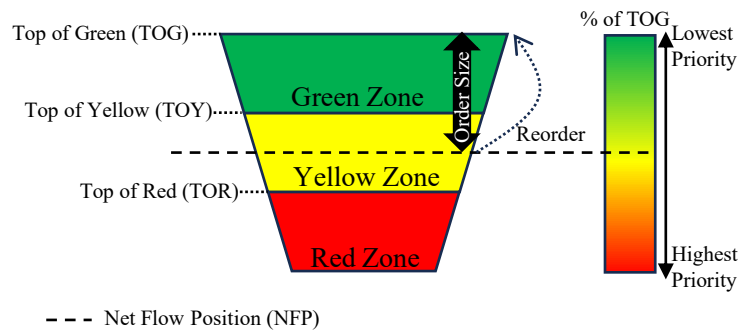


Figure 6.1 DDMRP Buffers

DDMRP buffers are central to the DDE, which is the operational engine of the framework, responsible for translating planning signals to the shop floor. The standard execution begins with a trigger from the NFP, which recommends a replenishment order. Once an order is released, its execution priority is based on the status of the inventory buffer (Ptak & Smith, 2019). Despite the growing body of research focusing on DDP and DDE aspects, there is no literature considering the order approval process, the missing link between DDP and DDE. The objective of this study is to fill this gap by formalizing a smart order generation process that considers both order recommendation and order approval.

6.2.2 Order Generation in DDMRP

The classic DDMRP order generation process assumes that an order should be recommended for a buffered position whenever the NFP falls below the TOY. Ptak and Smith (2019) mention that the DDMRP planning phase ends once the recommendation has been approved and becomes an open supply order, pointing out the need for an approval step in DDMRP applications. The authors make a brief reference to two situations, one in which orders are not approved at the time they are recommended (a timing issue) and another where orders are approved but with quantities that differ from the recommendation (a quantity issue). Since in both cases each recommended order requires an approval decision influenced by considerations beyond the NFP, the resulting orders are not always perfectly aligned with the basic NFP trigger, contradicting most of the current DDMRP research (e.g., (Dessevre et al., 2021; Miclo et al., 2018; Uzun Araz et al., 2024)) in which recommended orders are automatically approved.

The lack of formalization of the order generation process has led researchers to explore alternative approaches. Most notable is the integration of CONWIP and DDMRP. Developed by Spearman et al. (1990), CONWIP is a pull-oriented production control logic designed to regulate the total amount of Work-in-Process (WIP) within a manufacturing system. Operating under the principles of workload control, base CONWIP uses a fixed number of system-wide cards (or tickets) to authorize the release of new jobs onto the shop floor. A new production order is only allowed to enter the system when a card becomes available, which strictly occurs when a completed order exits the production. By enforcing a hard cap on the maximum WIP, this gating logic prevents shop floor flooding, maintains flow stability, and ensures predictable cycle time in accordance with Little's Law.

While one paper studied the use of CONWIP to sequence replenishment orders generated by the DDMRP system (Fernandes et al., 2025), another application evaluated CONWIP for order generation and order approval within the DDMRP framework (Dessevre et al., 2023). In both cases, the use of CONWIP rendered improvements over the traditional DDMRP framework; however, no further research was conducted aiming to evaluate the claimed improvements in alternative DDMRP settings. This research aims to propose a novel smart order generation

process and to investigate the timing of order approvals by comparing it with CONWIP and traditional DDMRP approaches.

6.2.3 Flow Time Prediction

Accurately estimating flow time is a significant challenge in dynamic manufacturing environments, yet it is critical for effective production control. Traditional planning systems often rely on static or historical average lead times, which fail to account for real-time shop floor conditions like congestion or resource availability (Hopp & Spearman, 2011). Flow time prediction has been a studied branch of operations management since 1960, with various proposed methods such as analytical approaches (e.g., Little's law and queueing theory), simulation (e.g., discrete event simulation), and data-driven estimation (e.g., predictive analytics and machine learning) (Mourtzis et al., 2014). With the popularization of artificial intelligence, the estimation methods using Machine Learning (ML) techniques, such as Artificial Neural Networks (ANNs), are becoming more widely used (Burggraf et al., 2020).

Several studies applied ML techniques to estimate both flow times and lead times in real and simulated environments. Gyulai et al. (2018) applied several machine learning algorithms to predict lead times in an optics flowshop. The authors compared Linear Regression (LR), Random Forests (RF), and Support-Vector Regression (SVR), demonstrating the increased performance of predictive analytics when compared to analytical methods to estimate the LT. Murphy et al. (2019) used a combination of machine learning algorithms, including Cubist datamining, ANN, H2o and neuralnet, RF, and XGBoost in a simulated environment to estimate real flow time values based on input variables obtained from the simulation model.

In both cases, the use of machine learning rendered significant improvements in prediction accuracy over traditional analytical methods. However, there is no research conducted aiming to integrate these predictive capabilities directly into the DDMRP execution phase. This research aims to use machine learning algorithms for predictive analytics to estimate the predicted flow time of recommended replenishment orders, integrating this metric to establish a novel smart order generation process.

6.2.4 Research Gap and Paper Contributions

This research addresses a recognized gap in DDMRP literature by clarifying the link between Demand Driven Planning (DDP) and Demand Driven Execution (DDE). The standard DDMRP methodology recommends a replenishment order whenever the NFP falls below the static TOY threshold without a formalized approach for order approval. This lack of formalization currently leads applications to consider automatic order release (e.g., Fernandes et al. (2024), Duhem et al. (2023)) implicitly assuming infinite capacity. While using the single-factor TOY threshold is effective to translate market consumption into supply signals, automatically approving orders without consideration for the shop floor state can lead to premature releases that inflate WIP without improving service performance.

Instead of relying on a single-factor decision based solely on buffer status, this paper introduces a dual-factor decision approach that considers both buffer levels and the real-time state of the production system in the order generation process. This approach extends the traditional DDMRP logic by integrating operational system conditions into the order release decision. Although recent literature explored coupling DDMRP order generation with alternate control logics such as CONWIP (e.g., Dessevre et al. (2023), Fernandes et al. (2025)), these approaches remain reactive. This study is the first to leverage real-time production system information aiming to improve service performance.

From a practical standpoint, this study is the first to develop a smart order generation process that integrates predictive analytics to estimate the predicted flow time of recommended orders to the order generation process. The proposed process introduces a dynamic approval gate that releases orders based on both inventory buffers and the current state of the production system. By incorporating machine learning techniques into the decision process, we contribute to the body of research aiming to enhance DDMRP performance through data-driven methods such as Cuartas and Aguilar (2022), Duhem et al. (2023) and, El Marzougui et al. (2024).

6.3 Methodology

6.3.1 The Proposed Smart Order Generation Process

This paper proposes a smart order generation process that transitions the order generation task from a reactive inventory trigger to a dual-factor predictive trigger. Building upon the

foundational NFP recommendation, the framework introduces an intelligent approval gateway that considers both the status of the inventory buffer and the predicted future state of the production system. The proposed smart process uses machine learning algorithms to estimate the predicted Flow Time (pFT) of a recommended order prior to its release. By treating the pFT as a dynamic representation of the current and anticipated shop floor load, the system can strategically authorize, delay, or withhold orders. This ensures that the generation of work is synchronized with the physical reality of the system, respecting the fundamental dynamics of shop floor queues by controlling the load on the shop floor level.

Figure 6.2 illustrates the workflow of the smart order generation process. First, the system retrieves the standard replenishment signal from the DDP together with the pFT, which is dynamically calculated by the machine learning algorithm using real-time shop floor state vectors. The subsequent dual factor decision logic is governed by external managerial policies (i.e., order generation policies) that establish the α and β threshold parameters, dictating the operational window for order generation. After this evaluation, the decision passes through a failsafe mechanism designed to prevent absolute system starvation. If order generation is withheld for too many consecutive cycles, the failsafe overrides the decision logic and forces order approval. Finally, the process outputs the decision to approve or deny an order.

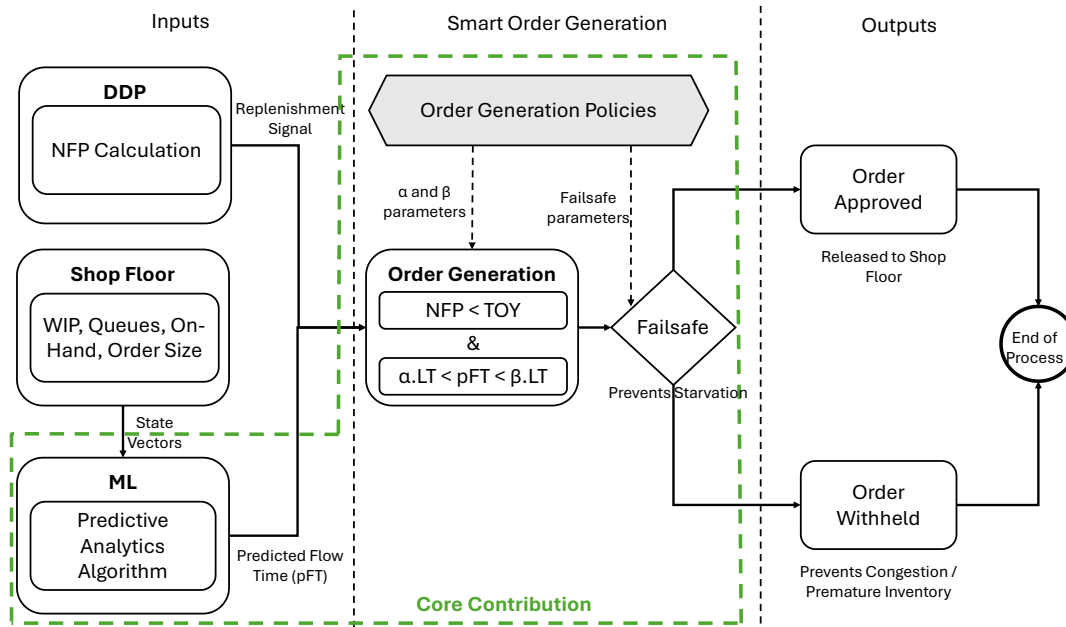


Figure 6.2 Flow of the Smart Order Generation Process

The proposed process alters the boundary between planning and execution in DDMRP. In standard applications, the signal generated by the DDP acts as an absolute command. In the proposed process, it acts as a recommendation (factor 1). Simultaneously, the predictive analytics engine continuously monitors the physical shop floor, translating state vectors into the pFT, a temporal metric related to the order recommendation (factor 2). The smart order generation process synthesizes both factors. By evaluating the inventory deficit against the anticipated system state, the process actively manages the load, transitioning order generation to a system aware process.

This approach effectively creates a dynamic and intelligent order generation process. For example, if the system is predicted to be very idle (very low pFT), the smart gateway might withhold an order to prevent it from being completed too early, reducing excess inventory. Similarly, if the system is predicted to be overly congested (very high pFT), the gateway might also withhold the order to avoid adding more work to an already stressed system. Crucially, unlike CONWIP, this framework does not maintain a queue of refused orders. If the smart gateway conditions are not met, the specific order recommendation is discarded, and the system awaits the next DDMRP generation trigger.

The proposed smart order generation process is mathematically defined as a conditional gating function, $G(t)$, acting upon the standard DDMRP recommendation. An order is generated and approved only if the NFP indicates the need for replenishment ($(NFP(t) \leq TOY)$ and the pFT falls within a window defined by the Lead Time (LT) and threshold factors α and β . Equation 6.1 is the representation of the proposed process.

$$G(t) = \begin{cases} 1 & \text{if } NFP(t) \leq TOY \wedge ((\alpha \cdot LT) < pFT(t) < (\beta \cdot LT)) \\ 0 & \text{otherwise} \end{cases} \quad (6.1)$$

Where:

- $G(t)$ is the decision to generate (1) or withhold (0) an order at time t .
- $NFP(t) \leq TOY$ is the standard DDMRP order recommendation trigger, based on buffer's status.
- $pFT(t)$ is the predicted flow time for a new order at time t , which reflects the current congestion and workload of the shop floor. This value depends on real-time system data.

- LT is the planned Lead Time for the item.
- α and β are threshold factors that define the smart gateway. They define the range for the pFT relative to the planned LT. An order is approved only if the pFT falls within this window.

To assess the implications of the proposed smart order generation process, we benchmark it against the conventional DDMRP order generation logic. In the next section, we define the two baseline processes that represent the current propositions from DDMRP literature (i.e., automatic approval and CONWIP).

6.3.2 Traditional DDMRP Order Generation Processes

The most common process for the approval of recommended orders in DDMRP is automatic approval. This approach is used in most DDMRP literature (Alves et al., 2024; Duhem et al., 2023; Fernandes et al., 2024) and serves as the comparison baseline. In this approach, every order recommendation is converted into an open supply order without any secondary checks. If the buffer position indicates a need for replenishment, the order is generated and released to the supply source (e.g., production or supplier), creating a pool of orders to be processed. Equation 6.2 is the mathematical representation of this process.

$$G(t) = \begin{cases} 1 & \text{if } NFP(t) \leq TOY \\ 0 & \text{otherwise} \end{cases} \quad (6.2)$$

Where:

- $G(t)$ is the decision to generate and approve an order at time t .
- $NFP(t)$ is the Net Flow Position of the buffer at time t .
- TOY is the reorder point for the buffer.

This approach ensures that the replenishment needs from the buffer are always addressed by the system. However, this immediacy can lead to inefficiencies, particularly in MTS environments. The automatic approval does not account for the current state of the production system, making it susceptible to concluding replenishment orders before their actual consumption, leading to unnecessary inventory.

A more controlled approach for replenishment order approval in DDMRP uses a CONWIP system (Dessevre et al., 2023) to generate and approve orders. In this approach, a

recommended order is created and converted into an open supply order only if the total number of orders in the system is below a predefined limit. As described by Hopp and Spearman (2011), a CONWIP system authorizes the release of a new order into the system only when an existing order has been completed. This maintains a constant, controlled level of WIP. Equation 6.3 is the mathematical representation of this process.

$$G(t) = \begin{cases} 1 & \text{if } NFP(t) \leq TOY \wedge (WIP(t) < L_C) \\ 0 & \text{otherwise} \end{cases} \quad (6.3)$$

Where:

- $G(t)$ is the decision to generate and approve an order at time t .
- $NFP(t)$ is the Net Flow Position of the buffer at time t .
- TOY is the reorder point for the buffer.
- $WIP(t)$ is the total amount of orders in production at time t .
- L_C is the predefined CONWIP Limit (i.e., the maximum number of orders allowed).

The values of L_C can be determined following the approach proposed by Dessevre et al. (2023). Different values of L_C can then be tested until an acceptable service level is achieved. The number of orders in the production system acts as a secondary check when CONWIP is used to order approval in DDMRP applications. An order may be recommended considering the buffers NFP; however, it is only approved if the number of orders in the system is below a predefined limit. This creates a two-part decision rule. This approach prevents the overloading of critical resources and regulates flow through the entire system by linking order release to the current state of the system.

6.4 Application to a non-preemptive DDMRP flowshop

To validate the proposed smart order generation process, we evaluate its performance against traditional DDMRP and CONWIP order generation processes in a non-preemptive DDMRP flowshop.

6.4.1 Case Study

We consider a six-machine non-preemptive flowshop (see Figure 6.3) where 30 finished products are managed with DDMRP buffers. Table 6.1 presents the average daily demand for the different products. Replenishment orders are generated according to DDMRP policy (i.e., a new

replenishment order is recommended if the NFP is below the ToY with a variable production lot size corresponding to a replenishment to the TOG). The order release priority depends on the percentage of the NFP over the TOG. If two replenishment orders have the same priority, First In First Out (FIFO) rule applies.

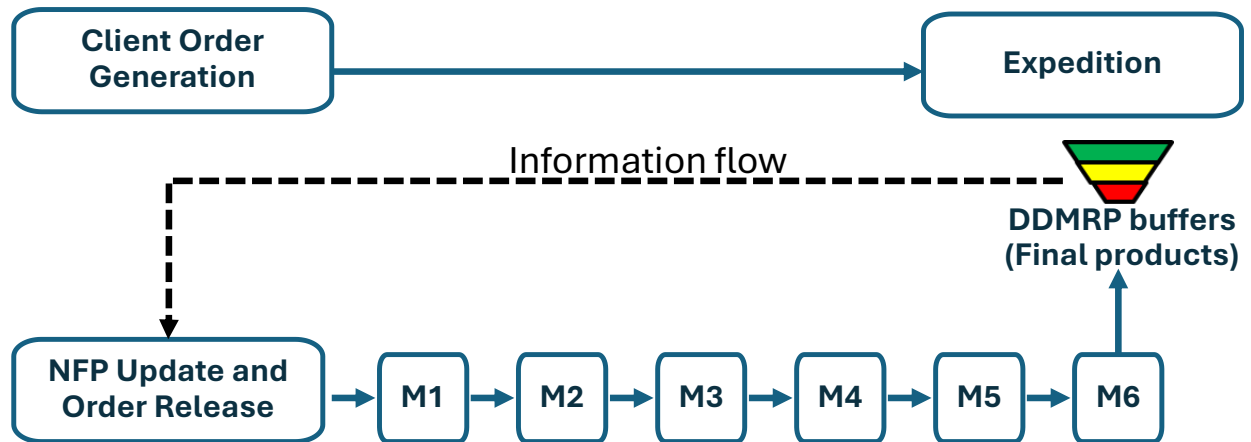


Figure 6.3 Six Machine DDMRP Flowshop

Table 6.1 Average Daily Usage (ADU) in units

Product	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10
ADU	111	39	94	107	32	65	49	56	104	58
Product	P11	P12	P13	P14	P15	P16	P17	P18	P19	P20
ADU	104	105	97	54	88	36	104	71	82	75
Product	P21	P22	P23	P24	P25	P26	P27	P28	P29	P30
ADU	60	88	39	95	94	63	54	87	94	88

Following the discrete event simulation model developed by Dessevre et al. (2021), we assume that the arrival of customer orders follows an exponential distribution with an arrival rate of 1/1.5 hours. The daily quantities for each product requested by customer orders are calculated to meet a normal distribution with the average equal to the ADU and a standard deviation of 25%. There is no order visibility for the next periods, and neither the order spike threshold nor the order spike horizon is considered. If an order cannot be fulfilled from the on-hand inventory, it is back-ordered to be fulfilled as soon as possible, following a FIFO rule.

In the considered flowshop, all products have the same routing. Due to the identical routing, we followed Dessevre, Baptiste, et al. (2019) recommendation to use a single Lead Time (LT) for all

the different products. The initial processing times are given in Table 6.2. For each simulation, the processing times are randomly adjusted with a coefficient following a triangular law TRI (0.8, 1, 1.2). Respecting the ranges of Ptak and Smith (2019), the Variability Factor (VF) is 0.50 for all products, and the Lead Time Factor (LTF) is 0.75, 0.5, and 0.25 for small, medium, and long processing times, respectively.

Table 6.2 Processing Times (seconds)

Products/Machines	M1	M2	M3	M4	M5	M6
P1–P10 (SMALL)	5	5	5	10	15	5
P11–P20 (MEDIUM)	5	10	10	10	10	10
P21–P30 (LONG)	5	15	15	10	5	15

Transfer time between processes is not considered, nor are changeover times between products. Buffers are initialized randomly between 50% and 100% of the TOG. The DDMRP parameters for buffer sizing and qualified demand calculation are set at the beginning of each replication and are not updated during the experiment (i.e., static parametrization).

6.4.2 Order Generation Policies

To evaluate the impact on the service level and inventory efficiency of the proposed smart order generation process, we implemented eight distinct order generations as presented in Table 6.3. Since all policies use the standard DDMRP equation for initial order recommendation ($NFP < TOY$), the core difference lies entirely in the approval condition.

As detailed in Table 6.3, the first five policies represent the proposed smart order generation policies. These policies evaluate a recommended order against its predicted flow time (pFT) to determine approval and use different thresholds.

Table 6.3 Smart Order Generation policies description

	Policy	α	β	Approval Condition	Type	Description
Smart Order Generation	Idle (IDL)	0	1	$0LT < pFT(t) < 1LT$	Smart	Approve if pFT is smaller than the defined LT.
	Margin Idle (IDL30+)	0	1.3	$0LT < pFT(t) < 1.3LT$	Smart	Approve if pFT is smaller than 1.3 times the defined LT.
	Occupied (OCC)	1	∞	$1LT < pFT(t) < \infty LT$	Smart	Approve if pFT is bigger than the defined LT.
	Margin Occupied (OCC30-)	0.7	∞	$0.7LT < pFT(t) < \infty LT$	Smart	Approve if pFT is bigger than 0.7 times the defined LT.
	Zone 30% (Z30 $\pm$)	0.7	1.3	$0.7LT < pFT(t) < 1.3LT$	Smart	Approve if pFT is between 0.7 times the LT and 1.3 times the LT.
Traditional DDMRP and CONWIP	Baseline (BSL)	-	-	<i>Automatic Approval</i>	Standard	Approve all recommended orders.
	CWIP 22 (CW22)	-	-	$WIP(t) < 22$	CONWIP	Approve if the sum of orders in production is smaller than 22.
	CWIP 25 (CW25)	-	-	$WIP(t) < 25$	CONWIP	Approve if the sum of orders in production is smaller than 25.

- **WIP-Limiting Policies (IDL and IDL30+):** Idle (IDL), these policies delay order generation if a recommendation occurs while the shop is overloaded (i.e., the pFT is greater than the LT). A new order is generated using the updated NFP when the next consumption occurs. These policies aim to prevent new orders into an already congested shop.
- **Inventory-Pacing Policies (OCC and OCC30-):** By approving orders only when the shop is Occupied (OCC), these policies delay order generation if a recommendation occurs while the shop is underloaded (i.e., the pFT is smaller than the LT). A new order is generated using the updated NFP when the next consumption occurs. These policies aim to prevent premature inventory accumulation.
- **Balanced Policy (Z30±):** By limiting order approval to periods when the shop is in a balanced Zone (Z30), this policy delays order generation if a recommendation occurs while the shop is either severely underloaded or overloaded. A new order is generated using the updated NFP when the next consumption occurs. This policy aims to prevent both premature inventory and new releases into a congested shop.

In this study, the failsafe mechanism threshold is set to six consecutive refusals due to the modeled arrival rate of 1.5 hours (i.e., on average every six orders are equal to the product's ADU). Throughout our experiments, this limit was never reached, peaking at a maximum of five consecutive refusals.

Finally, the last three policies in Table 3 represent the benchmark policies used to evaluate the performance of the smart framework.

- **Standard Baseline (BSL):** This represents the conventional DDMRP implementation. It uses a single-factor decision where every recommended order is automatically approved, assuming infinite shop floor capacity.
- **CONWIP policies (CW22 and CW25):** These policies represent an established execution model. They introduce a secondary capacity check, approving orders only if the total number of active orders currently in production remains below a predefined limit (L_c) of 22 or 25 orders respectively.

6.4.3 Flow Time Prediction

This study applies different machine learning algorithms to estimate the predicted flow time (pFT) of an order prior to its creation, enabling its use as the central decision metric in the smart order generation process. This approach is adopted due to superior performance of ML techniques over analytical methods (Gyulai et al., 2018) and their high prediction accuracy for flow time estimation, exceeding 87% in real production environments (Rosa et al., 2025).

A crucial step before the experimentation is to select the most suitable attributes (features) and the appropriate algorithm to generate an accurate pFT. In our research, we were inspired by the feature and algorithm selection methodology proposed by Djedidi (2025). A large dataset was generated by running a 25,000-day simulation of the case study model, capturing information on customer orders, the DDMRP system, and the flowshop state at the moment a replenishment order is generated. Ten features were collected, including Order Quantity, Net Flow Equation (NFE), On Hand inventory, Work in Process (WIP), and machine queues (M1–M6). Five supervised machine learning algorithms—Decision Tree, Random Forest, XGBoost, Support Vector Regressor (SVR), and Ordinary Least Squares (OLS)—were trained using 80% of the data, while 20% was reserved for testing. Model performance was evaluated using MAE, RMSE, and R^2 to assess prediction accuracy and detect potential overfitting. Among the tested models, XGBoost achieved the best predictive performance in terms of R^2 on the test dataset and was therefore selected for further analysis. Residual analysis confirmed good overall accuracy, although the model tends to underpredict higher values. Feature importance analysis using SHAP revealed that the last machine (M6) is the most influential predictor, followed by Order Quantity. Detailed data and a full explanation of this approach are provided in Appendix D.

6.4.4 Experimentation

For the DDMRP flowshop, we considered five different LT configurations (very low, low, medium, high, and very high). For each configuration, the 8 order generation policies presented in Table 3 are tested. Thus, we conducted a total of 40 SCenarios (see Table 6.4).

Table 6.4 List of tested scenarios (SC)

	Order Generation Policies							
LT Config.	BSL	IDL	IDL30+	OCC	OCC30-	Z30±	CW22	CW25
Very Low – 6.4h	SC1	SC6	SC11	SC16	SC21	SC26	SC31	SC36
Low – 8h	SC2	SC7	SC12	SC17	SC22	SC27	SC32	SC37
Medium – 9.6h	SC3	SC8	SC13	SC18	SC23	SC28	SC33	SC38
High – 11.2h	SC4	SC9	SC14	SC19	SC24	SC29	SC34	SC39
Very High – 12.8h	SC5	SC10	SC15	SC20	SC25	SC30	SC35	SC40

The LT configuration was defined to evaluate the process across a broad spectrum of system constraints. Starting from a standard low manufacturing time of 8 hours, we applied a constant positive and negative variation step of 20% to generate a range spanning from 6.4 hours to 12.8 hours. This extensive range, resulting in double the lowest setting for the highest setting, was designed to assess the smart order generation policies under contrasting operational conditions, allowing the simulation to cover scenarios with high tension and tight LTs to conservative scenarios and comfortable buffering. This ensures that the performance of the policies is validated with the system in both stressed and relaxed conditions.

To support our analysis, we use two key performance indicators:

- The average Service Level (SL), which is the ratio of orders fulfilled directly from the buffer without any waiting and the number of orders received.
- The average Inventory (I), which indicates the mean level of finished goods for the simulated period.

We consider 5 operating days a week with 1 shift of 8 hours per day. The loading rate for all the experiments was 90% for the bottleneck process (M6). Considering 50 replications for each experiment and 95% confidence intervals on estimates, the results were collected during 550 simulation days, after 50 days of warmup.

6.5 Results and Discussion

In this section, we present and discuss the results of the experimentation phase. In the results illustrations, different shapes represent the policy types, colors indicate different lead time (LT) settings and dashed horizontal and vertical whiskers denote the 95% confidence intervals for I and SL, respectively.

6.5.1 Delaying order generation if the system is overloaded

The objective of these experiments is to analyze the impact of WIP-Limiting policies which delay order generation when the system is overloaded (SC6 to SC15). Specifically, we tested policies that forbid order approval unless the pFT is predicted to be below a certain threshold defined by $[\alpha, \beta]$, where α is 0. The policies tested were IDL [$\alpha = 0, \beta = 1$] and IDL30+ [$\alpha = 0, \beta = 1.3$].

The results from our simulation, presented in Figure 6.4, showed that, across various LT settings, there was no statistically significant difference in performance between the policies and the traditional DDMRP order generation process. The explanation for this outcome lies in the fundamental definition and function of the LT within the DDMRP framework. In DDMRP applications, the LT is key to simultaneously combating demand signal distortion and supply continuity variability (Ptak and Smith, 2019). It is intentionally configured to be longer than the average manufacturing FT. This configuration aligns with the principles defined by Hopp and Spearman (2011) for achieving a target service level. To guarantee a certain service level (SL), which is a probability based on the distribution of the actual FT, the planned LT must be set at a higher percentile of the historical FT distribution, as demonstrated by Alves et al. (2026).

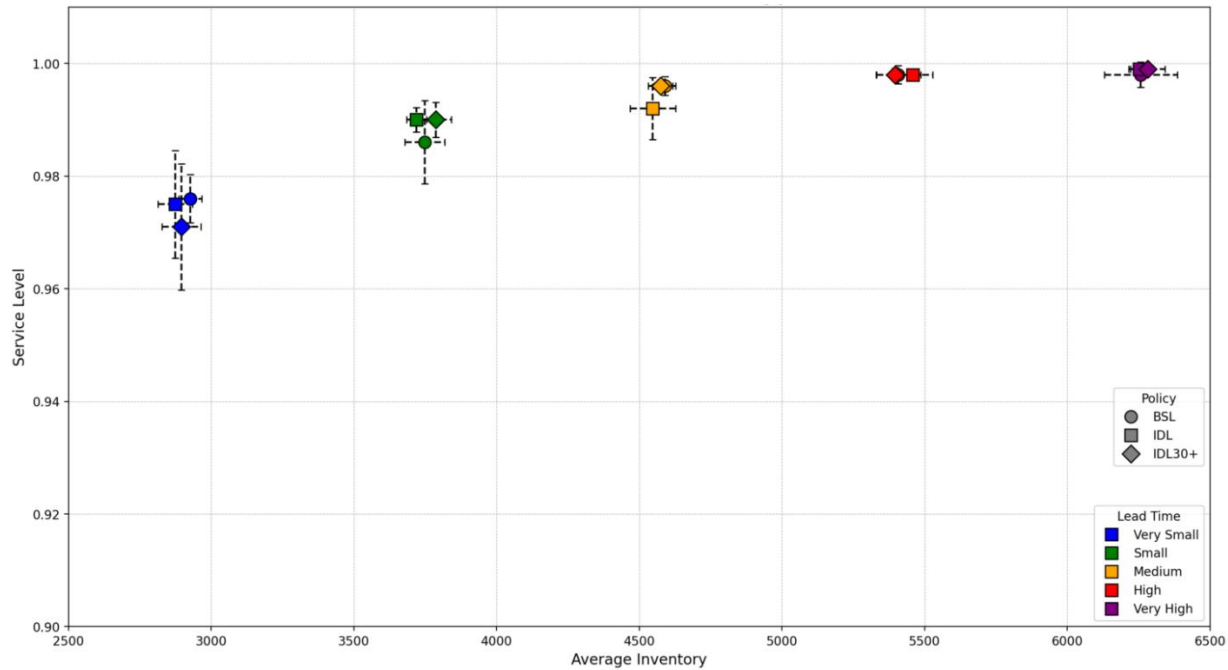


Figure 6.4 Performance Comparison of WIP-Limiting Smart Order Generation Policies

Consequently, in a properly configured DDMRP system, the condition $pFT < LT$ will be true most of the time. The additional ‘approval gate’ imposed to delay order generation if the system is overloaded is rarely triggered because the system is designed to have an LT that is consistently bigger than the average FT. Since orders are very rarely delayed (a condition that is almost always met), the operational behavior of the IDL and IDL30+ policies becomes nearly identical to the baseline DDMRP policy, which explains the lack of a statistically significant difference in the results. This finding suggests that the inherent buffering within DDMRP’s LT setting makes approving orders only when the system is underloaded is largely redundant.

6.5.2 Delaying order generation if the system is underloaded

These experiments analyzed the impact of inventory-pacing policies which delay order generation when the system is underloaded (SC16 to SC25). We tested policies that forbid order approval unless pFT is predicted to be above a certain threshold defined by the pair $[\alpha, \beta]$ where β is infinite. The specific policies were OCC policy $[\alpha = 1, \beta = \infty]$ and OCC30- policy $[\alpha = 0.7, \beta = \infty]$. Considering the average for all LT settings, the results from our simulation showed that these policies successfully reduce inventory by 8.8%, though this comes at the cost of a marginal reduction in the service level of 0.7 percentage points.

Analyzing the results by LT settings, presented in Figure 6.5, the policies resulted in mostly the same SL as traditional DDMRP, except at very low LT settings, in which the policies caused a reduction. For the average inventory, the tested policies always resulted in a reduction. Comparing policies, the proposed OCC resulted in smaller inventories at high and very high LT settings while maintaining the same SL as the OCC30- policy.

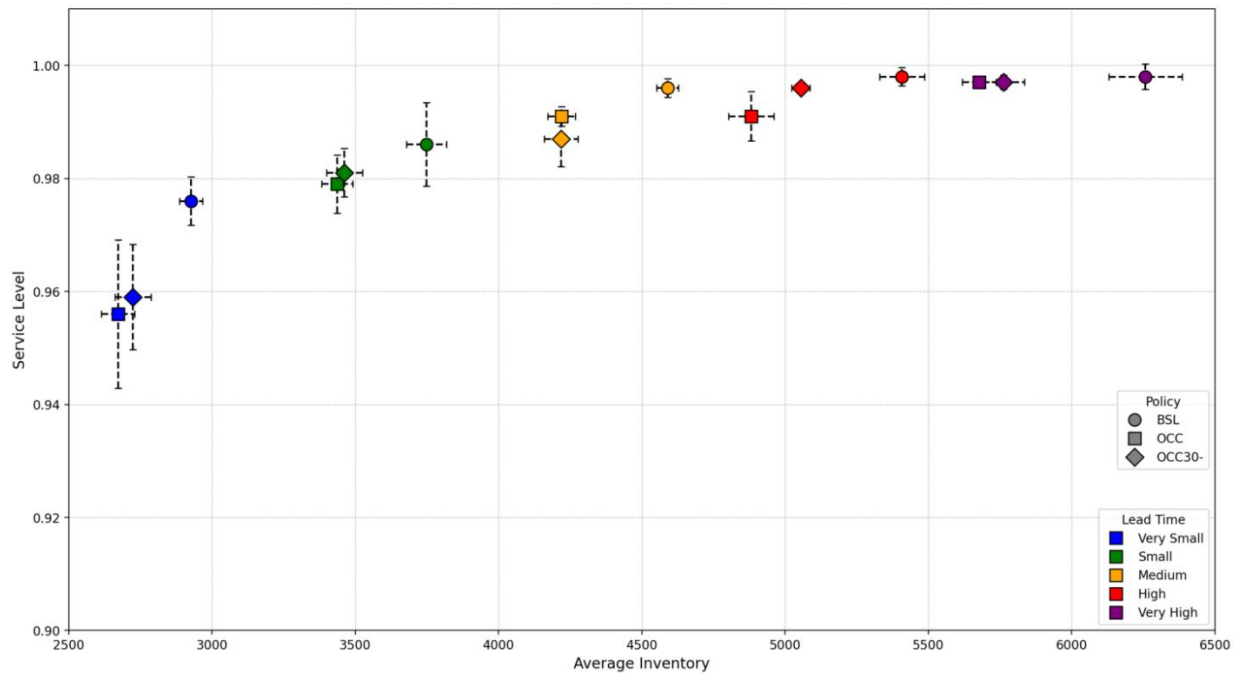


Figure 6.5 Performance Comparison of Inventory-Pacing Smart Order Generation Policies

To understand the proposed rationale behind the inventory-pacing policies, consider an analogy: approving a replenishment order is compared to leaving for an important meeting, where the pFT is the predicted travel time. Despite the average travel time (the LT) if your navigation app shows light traffic (a low pFT), you can afford to delay your departure to avoid arriving inefficiently early. Conversely, if the app shows heavy traffic (a high pFT), you must leave immediately to arrive on time. An inventory-pacing policy applies this exact logic by withholding orders from an underloaded system (the ‘low traffic’ state) to prevent the accumulation of early inventory. It permits the order to be released only when the system is sufficiently occupied (the ‘high traffic’ state), thereby synchronizing the start of work with the shop’s operational state to ensure a more fluid and efficient flow.

This logic directly explains the observed experimental results. The reduction in average inventory is a direct consequence of deliberately delaying orders when the system is

underloaded. This delay reduces the amount of Work-In-Process (WIP) in the system at any given moment, which, in accordance with Little's Law, leads to lower average inventory levels. However, this intended delay is also the source of the service level reduction in low LT settings. During this self-imposed delay, customer demand may continue to deplete the buffer. Therefore, delaying orders when the system is underloaded achieve a significant inventory reduction by introducing a calculated risk, trading a slight decrease in service level for a notable reduction in the average inventory.

6.5.3 Delaying order generation if the system is overloaded or underloaded

This set of experiments analyzed the impact of a balanced policy, which delays order generation if the system is either severely overloaded or (SC26 to SC30). Specifically, we tested a policy that forbids order approval unless the pFT falls within a balanced operational zone described by the pair $[\alpha = 0.7, \beta = 1.3]$. Considering the LT settings average, the simulation results showed that this approach resulted in an average inventory reduction of around 8% with a decrease in the service level of one percentage point.

When comparing the results with the different LT settings, presented in Figure 6.6, the balanced policy resulted in a significant decrease in SL up to medium LT. At high and very high settings, the policy had the same SL as traditional DDMRP while reducing average stock.

The explanation for this outcome lies in the dual nature of the restrictions imposed by the balanced policy. On one hand, the policy incorporates a lower threshold $[\alpha = 0.7]$, which prevents orders from entering an underloaded system. As explained in Section 5.2, this logic is the primary driver of inventory reduction, as it deliberately withheld orders to reduce the average inventory. On the other hand, the policy also imposes an upper threshold $[\beta = 1.3]$, forbidding order approval when the system is predicted to be overly congested. While the intent may be to avoid more work to an already stressed system, this restriction withheld orders precisely when the DDMRP order recommendation trigger is more critical, which is during periods of long anticipated FTs.

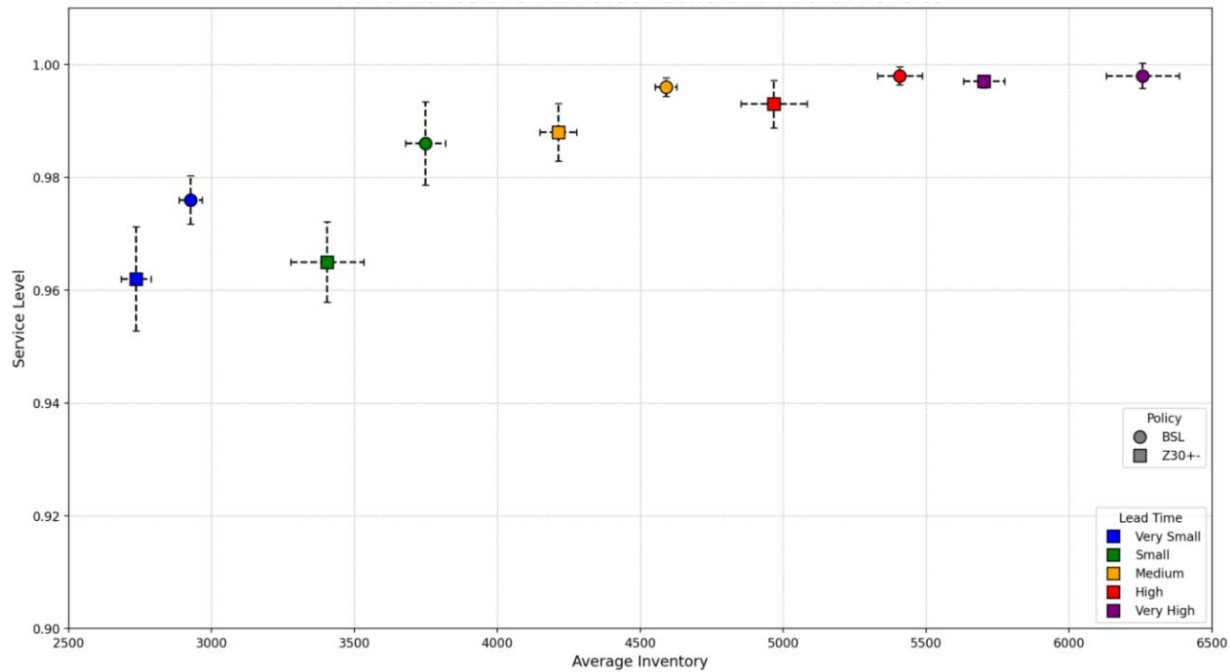


Figure 6.6 Performance Comparison of Zonal Smart Order Generation Policy

This inherent conflict within the policy design directly explains the observed results. Delaying orders from an underloaded system is effective, delivering a significant inventory reduction. However, the increased reduction in the service level when compared to the inventory-pacing policies is a consequence of delaying orders when the system is overloaded. The system is not frequently overloaded to the point where $pFT > 1.3LT$, but on the infrequent occasions that it does, the policy refusal in approving a replenishment order in a timely manner allows the buffer to be further depleted by demand. This effect is only mitigated at higher LT settings, in which the higher buffers absorbed the inventory disruption.

Consequently, the balanced policy represents a structural compromise that carries vulnerability. The results suggest that although these high-congestion events are infrequent, they are impactful enough to increase stockout risk and reduce the policy overall performance when compared to simpler one-side threshold policies.

6.5.4 Delaying order generation using CONWIP logic

In this final experiment, we evaluated the performance of delaying order generation using the CONWIP logic (SC31 to SC40), a concept previously explored in the literature by Dessevre et al. (2023). The policies were implemented by delaying order generation if a limit number

of orders in the system is attained, with two different levels tested: a more restrictive limit of 22 orders (CW22) and a less restrictive limit of 25 orders (CW25). As demonstrated in Figure 6.7, on average, the CW22 policy resulted in a 2.8% decrease in inventory with a 1.7 percentage point reduction in the service level, while the CW25 policy performance was statistically indistinguishable from the baseline.

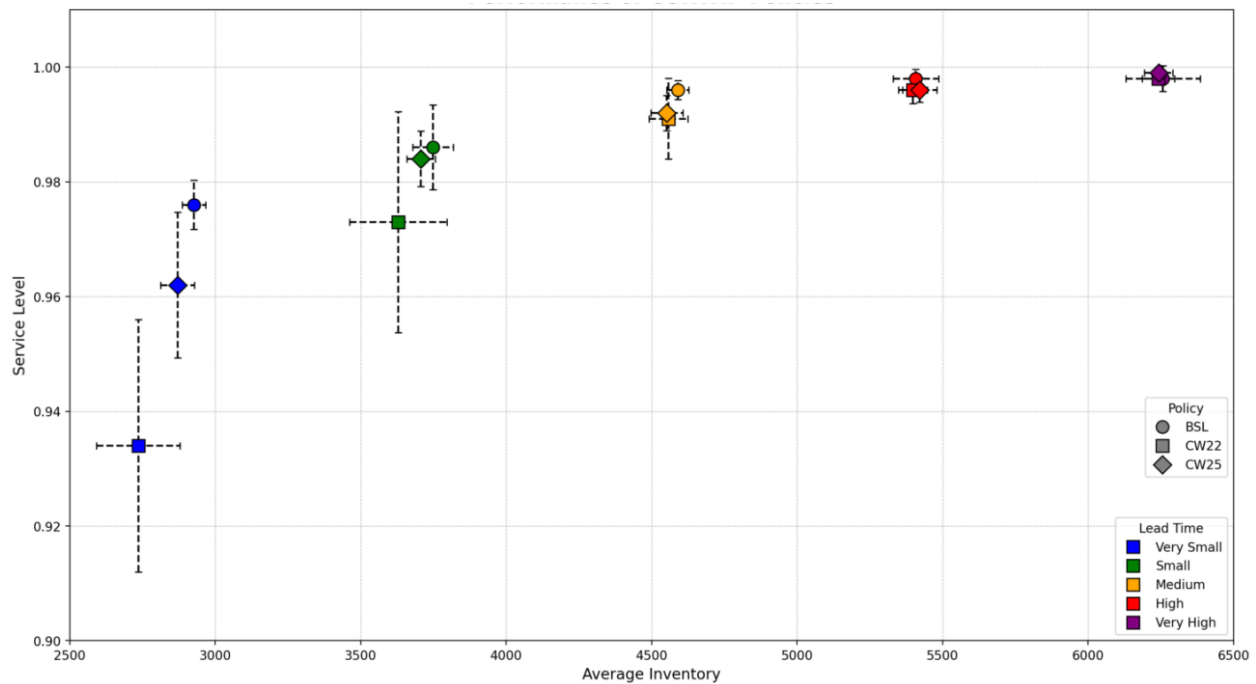


Figure 6.7 Performance Comparison of CONWIP Order Generation Policies

The explanation for this performance lies in a fundamental conceptual mismatch. CONWIP is, by design, a shop floor control logic focused on maintaining a constant level of WIP (Hopp & Spearman, 2011). The trigger for releasing a new order is the departure of a finished part, a signal that only indicates available capacity. This contrasts with the most effective smart order generation policies, which were successful due to strategically withholding orders from an idle system while considering the buffer's status.

The policies disconnection from the buffer status directly explains the outcomes. When faced with a more restrictive order limit (i.e., CW22 policy), the CONWIP policy was unable to approve the orders needed to maintain the buffer integrity, leading to a service level degradation higher than the smart order generation policies, particularly at very low LT settings. Ultimately, the CONWIP policy generated a steady flow of work but was 'blind' to the critical needs of the shop. As we increase the order limit (i.e., CW25 policy), the capacity

constraint becomes less binding, with the policy having the same performance as the unconstrained DDMRP. This behavior aligns with Dessevre et al. (2023) regarding CONWIP sizing in DDMRP, confirming that when CONWIP limit is set high enough to achieve performance comparable to DDMRP, its constraining effect becomes negligible.

6.5.5 Comparative Analysis of Order Generation Policies

A comparative analysis across all LT configurations, demonstrated in Figure 6.8, reveals a clear performance hierarchy among the tested order generation policies. The primary differentiating factor is the policies' effectiveness in managing the fundamental trade-off between inventory efficiency and service level. As expected, and consistent with DDMRP principles, increasing the planned LT augments both average inventory (I) and the service level (SL) across all policies. However, the logic used to approve orders determines the efficiency of this trade-off.

The analysis confirms that delaying order generation when the system is overloaded or using CONWIP logic is ineffective. The WIP limit proved redundant, offering no significant improvement over the baseline DDMRP. At the other extreme, the CONWIP policies were the most detrimental, degrading the service level without delivering significant inventory improvements due to their conceptual incompatibility with DDMRP buffer protection. The CW22 policy demonstrated the poorest performance trade-off.

In contrast, the policies that delay order generation when the system is underloaded (OCC and OCC30-) emerge as the most effective and robust strategy. These policies achieved significant inventory reduction of 8.8% at the cost of a marginal reduction of 0.7 percentage points in the service level. The consistent performance across all LT settings demonstrates their superior inventory efficiency without compromising service level. The results suggest that the key to improving the DDMRP order generation is not preventing releases into a congested system, but rather leveraging periods of system idleness. The inventory-pacing policies represent the most effective and broadly applicable enhancement, establishing a more efficient order generation.

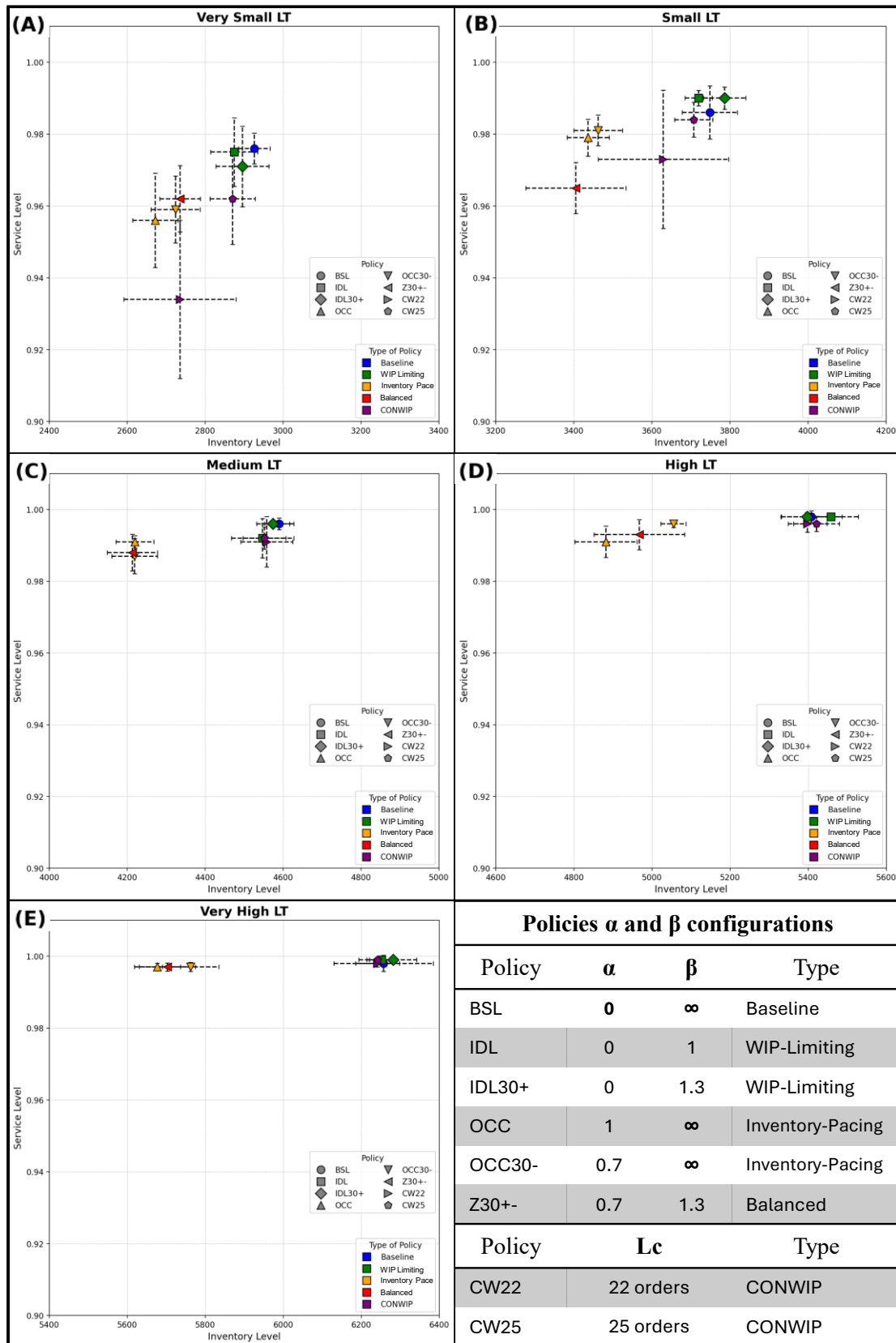


Figure 6.8 Comparison of order generation policies across different LT configurations

6.6 Conclusion

This research addressed a critical gap in the DDMRP methodology: the order approval step. We introduced and evaluated a novel smart order generation process that leverages real-time production system information through machine learning algorithms to achieve measurable improvements in service performance and inventory efficiency. Using pFT as the main decision metric we proposed order generation policies to enable state aware order generation.

Our findings demonstrate that the effectiveness of such policies is highly dependent on their logic. Policies that delay order generation when the system is underloaded proved to be the most effective. This approach resulted in an 8.8% reduction in the average inventory at the cost of a 0.7 percent reduction in service level. On the other hand, policies that delay order generation when the system is overloaded proved redundant. Furthermore, using CONWIP logic led to service level degradation without major improvements in inventory efficiency.

The primary implication of this study is that the DDMRP order generation process should evolve from a single-factor decision based only on buffer status to a dual-factor decision incorporating the predicted state of the production system. Critically, our results show that the most valuable state information is not congestion, but idleness. Withholding orders from an underloaded system has the potential to be used as an inventory reduction strategy, whereas restricting orders in a congested system is detrimental to performance. Consequently, the simpler one side threshold inventory-pacing policies represent the most robust and broadly applicable enhancement to traditional DDMRP framework.

While these results were validated in a specific flowshop environment, this paper provides a clear methodology for integrating predictive analysis into the DDMRP planning and execution framework. This opens several promising avenues for future research, including the application of different machine learning models, the integration of other uncertainty sources, such as supply variability in the modeling, and the validation of this smart order generation process in more complex environments such as job shops or divergent processes.

CHAPTER 7 GENERAL DISCUSSION

This chapter aims to discuss the thesis regarding methodological aspects and the results through the lens of the literature review. It synthesizes the findings, details the theoretical and managerial implications of transitioning Demand Driven Execution (DDE) to a status-centered execution system, and explains unexpected results encountered during the research. It does not repeat specific discussions and limitations of each contribution, as these were already discussed in the previous chapters.

7.1 Reminder of Research Objectives

This thesis sets out to resolve a structural dichotomy within the DDMRP methodology: while its planning capabilities are dynamic and responsive to the market, the standard DDE heuristics remain static and disconnected from the shop floor. Consequently, the primary objective of this research was to enhance the robustness and operational performance of DDMRP managed flowshops by shifting the execution phase from static, self-centered heuristics to a dynamic, status-centered execution system.

To achieve this goal, the research was decomposed into three specific objectives (SO) aligned with the functional tasks of Production Control Systems (PCSs). Specifically, this thesis aimed to: (SO1) establish and determine the extent to which dynamic, status-centered dispatching rules improve operational performance compared to static planning priorities; (SO2) validate an order release gating mechanism, in terms of prioritization policies, capable of maintaining flow stability when the system is subjected to heterogeneous order streams; and (SO3) propose and evaluate the performance implications of a status-centered order generation process that use predicted flow time as central decision metric.

By fulfilling these three specific objectives, this thesis provides a comprehensive status-centered demand driven execution system for highly variable environments.

7.2 Synthesis of Results

From a global perspective, the results obtained across the three contributions provide empirical evidence that the execution phase must transition from a self-centered logic (i.e., where orders are managed based solely on their individual buffer status) to a status-centered logic (i.e., where decisions are mediated by the real-time conditions of the production system).

Fundamentally, the collective findings demonstrate that DDMRP planning parameters cannot be repurposed as execution controls parameters in high-variability environments. The reliance on Net Flow Position (NFP) and the Buffer Penetration Ratio (BPR) introduces a critical latency issue. Because these metrics are calculated at the moment of order generation, they represent a static snapshot of the past. As the order progresses through the flowshop, this ‘self-centered’ priority often diverges from the physical reality of the inventory. By shifting the control variables to real-time execution metrics (i.e., On-Hand Availability) the DDMRP system significantly reduces stockouts and improves stability.

The improvements observed across the three tasks result from closing the execution feedback loop. In the sequencing task, the transition to dynamic dispatching proved that local, real-time availability results in a reduction of stockouts and consequently increased the service level. Considering the order release task, the validation of a gating mechanism demonstrated that standard DDMRP lacks the capacity to manage order heterogeneity. By using the order information to withhold standard orders when rush orders are present, the system effectively maintains its robustness, maintaining the service level for both rush and non-rush orders. For the order generation task, the integration of predictive analytics confirmed that the system can optimize the trade-off between inventory and service level by withholding orders during periods of idleness. This prevents the accumulation of premature inventory when order generation is subjected to a single-factor decision.

These findings redefine how DDMRP handles variability. Standard DDMRP relies on the sizing of physical inventory buffers to absorb demand and supply shocks. However, this thesis results demonstrate that inventory buffering alone is insufficient. By introducing status-centered execution system, the shop could better manage the control tasks, maintaining robustness, reducing stockouts and accumulation of early inventory, improving the trade-off between inventory and service level. This confirms that robust DDMRP performance requires not just the correct buffer sizing (planning), but intelligent active management of flow (execution).

7.3 Theoretical Implications

This thesis challenges the prevailing bias in DDMRP literature, which predominantly focuses on optimizing planning parameters, asserting instead that DDMRP research should continue towards analyzing and refining the DDE phase. By transitioning from self-centered heuristics to status-

centered execution system, this thesis extends PCSs reasoning across the three specific tasks studied.

The first contribution fundamentally challenges the use of planning metrics to guide execution in the order sequencing task. The prevailing assumption in DDMRP is that the NFP and the BPR are sufficient for order sequencing. Our findings prove that these static heuristics suffer from information latency and are ‘blind’ to local queueing dynamics. As demonstrated by Walraevens et al. (2022), queues with static priority dispatching suffer from starvation of low priority orders.

Another important implication challenges the DDMRP assumption that the planned urgency for sequencing remains valid throughout the production cycle. Our results demonstrate that in DDMRP environments, the order priority often diverges from the physical reality of the inventory it is meant to replenish. The traditional heuristics for DDMRP order sequencing are inventory-based and time-independent (i.e., the waiting time of an order does not influence its priority), resulting in scenarios in which frequent smaller orders tend to have a lower priority for execution than less frequent bigger orders. Theoretically, these findings establish that DDMRP execution prioritization must be decoupled from planning priority.

The second contribution expands the theoretical boundaries regarding heterogeneous order streams in the order release task. Despite mentioning an order approval step that may function as an order release gate, Ptak and Smith (2019) did not formalize a method. This led to the majority of DDMRP research considering that recommended orders should be immediately approved and released to the shop floor (Duhem et al., 2023; Fernandes et al., 2024; Velasco Acosta et al., 2019). The result is that standard DDMRP treats the release gate as an open valve, assuming the shop floor has infinite capacity. Our results demonstrate that this lack of control hinders the applicability of DDMRP when heterogeneous orders streams (such as rush orders) are introduced.

In the standard DDMRP theory, a shop floor is actually driven by three distinct order generation mechanisms: (1) normal buffer depletion via the NFE, (2) exceptional demand spikes for buffered products, and (3) direct customer orders for non-buffered products (Ptak & Smith, 2019). However most of the research simplifies this by assuming that all orders are generated via buffer depletion (Azzamouri et al., 2022; Dessevre et al., 2023; Fernandes et al., 2024; Thürer et al., 2020). In this context, the rush order ratio tested in this contribution serves a much broader theoretical purpose: it can be directly interpreted as the percentage of non-NFE orders (i.e., demand spikes and non-

buffered products) entering the system. Consequently, this study goes beyond the initial premise and provides the first empirical evidence of distinct order streams within a single DDMRP production system. Theoretically, it links DDMRP to Workload Control (WLC) theory, proving that the IMR assumption is flawed under order heterogeneity. It establishes that a formal order release mechanism is theoretically necessary to maintain flow stability in high variability DDMRP environments subjected to heterogeneous orders streams.

The third contribution alters the theoretical logic of replenishment in DDMRP. Standard theory treats order generation as a reactive, self-centered decision triggered by a single factor (i.e., $NFP < TOY$). This contribution demonstrates that this logic is prone to the ‘Lead Time Syndrome’ described by Schneckenreither et al. (2020). In dynamic environments, relying on static planning lead-times to dictate order generation forces earlier releases that inflate average inventory without improving the service level.

Furthermore, this research demonstrates the ‘early-release’ bias within this self-centered logic. In DDMRP practice, the Decoupled Lead Time (DLT) is not an average of observed flow times; rather, it is a planned time intentionally sized to be longer than the average manufacturing flow time to absorb demand and supply variability (Ptak & Smith, 2019). According to Hopp and Spearman (2011)’s definition of service level (i.e., the probability that the actual flow time is less than or equal to the planned lead time), guaranteeing high service levels mathematically requires setting the LT, and consequently the DLT, at a high percentile of the historical flow time distribution, as demonstrated by the second contribution. Consequently, the standard DDMRP order generation structurally forces orders into the shop floor ‘too early’ relative to the actual production speed. To resolve this, the proposed smart order generation introduces predictive analytics as a real-time capability check. By using the predicted Flow Time (pFT) as a second decision factor, it creates a status-centered correction. It dynamically delays order generation until the predicted shop floor occupation aligns with the DLT, effectively preventing premature inventory accumulation shifting order generation to a dual-factor status-centered control loop.

7.4 Managerial Implications

For practitioners, the implication of the results of this thesis is the necessity to decouple planning metrics from shop floor execution. While DDMRP provides a strong framework for inventory sizing and decoupling supply chains, relying on its planning parameters to manage execution

reduces system performance. Managers must focus on establishing high-quality, real-time data collection through modern Manufacturing Execution Systems (MES) to feed data to a status-centered execution phase.

The first practical implication drawn from this thesis is the need for a fundamental shift in how operators and supervisors prioritize work at the machine level. Practitioners must evolve from the DDMRP self-centered heuristics to status-centered metrics to sequencing. Planning parameters such as the NFP should be relegated to their intended purpose of buffer sizing and order generation. Because the NFP includes non-physical ‘On-Order’ quantities, using it for dispatching creates a false sense of urgency (or security) that misaligns with physical reality. To guide actual execution and prevent local stockouts, managers must implement dynamic dispatching rules based on real-time On-Hand availability, considering the backorders and the demand due. Operators should be directed to prioritize orders based on the immediate physical risk of a stockout, ensuring that the shop floor reacts to reality rather than a static planning snapshot.

The second implication requires production managers to refrain from the immediate release of replenishment orders. In practice, releasing orders to the shop floor the moment they are generated by the DDMRP logic reduces the performance, especially when heterogeneous streams of orders are present in the same flow. Practitioners should actively manage a pre-shop pool, establishing a formal order release gate. For rush orders, managers should control the level of prioritization in the various stages of the shop floor. Ensuring rush order priority only on order release stage is effective to filter heterogeneity, guaranteeing that the system remains robust, not compromising the service level for both rush and non-rush orders.

The third implication challenges the single-factor order generation in DDMRP. The research provides actionable guidance that order generation must evolve to a smart approval process. By adopting predictive analytics or incorporating real-time visibility of the shop floor actual workload, managers can synchronize the release of orders with the actual capability of the shop floor. Specifically, the results demonstrate the practitioners should strategically withhold the generation of standard replenishment orders during periods of system idleness. Delaying these releases until predicted flow time aligns with the planned lead time prevents premature inventory accumulation without compromising the service level, leading to a more efficient operation.

It is crucial for practitioners to recognize the conditions under which these recommendations apply. The status-centered execution system proposed in this thesis significantly outperforms traditional DDMRP execution under specific operation conditions of non-preemptive flowshops subjected to high-variability. The benefits of dynamic dispatching, release gating and smart order generation are most pronounced when the manufacturing system is operating near its capacity limits and when subjected to heterogeneous order streams. Under these high-load, high-variability conditions, transitioning to the proposed status-centered execution system becomes not just beneficial, but operationally essential to maintain flow stability and protect overall service levels.

7.5 Unexpected Findings

During the investigation of the execution control tasks, two significant unexpected findings that initially appeared contradictory to standard production control assumptions emerged. First, during the analysis of the sequencing task, an attempt was made to dynamically update the NFP continuously to govern dispatching priorities. Counterintuitively, this continuous recalculation reduced the performance of the dispatching rules compared to static heuristics. The data revealed that dynamically updating the NFP exacerbates the information latency issue. Because the NFP formula uses the On-Order quantities, the moment a new supply order is generated, the NFP immediately increases. This mathematically lowers the priority of that item in the queue, even though the physical On-Hand stock remains critically depleted. This anomaly conclusively proved that the NFP is fundamentally a planning metric and trying to dynamically update the NFP only masks the true physical needs of the system, confirming it is unfit for execution control.

Second, during the evaluation of the Smart Order Generation process, policies that restricted approval of orders in a highly occupied system were tested. This effectively placed a strict upper-bound cap on Work in Process (WIP), functioning similarly to a CONWIP system. However, in this specific scenario, the delaying order generation when the system is overloaded yielded the same results as traditional DDMRP, despite previous research applying CONWIP in DDMRP resulted in improvements (e.g., (Dessevre et al., 2023; Fernandes et al., 2025)). The anomaly suggests that the effectiveness of strict WIP capping is highly scenario dependent. In our DDMRP flowshop configuration, the true driver of efficiency was not capping orders during congestion, but rather withholding orders during periods of idleness to prevent premature inventory accumulation.

These unexpected results are critical in establishing the conditions for the proposed status-centered execution system. The experiments used a linear, non-preemptive flowshop with similar and constant routing. In such environments, the flow of materials is highly directional, meaning that preventing early releases sufficiently regulates the flow and prevents downstream saturation. Therefore, one of the boundary conditions is the application to linear flow environments.

Furthermore, the conclusion that a WIP cap on order generation is redundant cannot be generalized. In a more complex job shop environment characterized by highly variable routings, re-entrant flows, and shifting bottlenecks, a strict global WIP cap might still be a mandatory factor to prevent shop floor congestion. Consequently, while the principle of status-centered execution remains valid, translating the specific mechanisms used in this research (e.g., the removal of upper-bound limits) to highly non-linear manufacturing environments requires careful recalibration and represents a boundary to the current findings.

CHAPTER 8 CONCLUSION AND RESEARCH PERSPECTIVES

This final chapter synthesizes the research conducted in this thesis, demonstrating how the research objectives were achieved, outlining the core contributions, defining the limitations, and proposing directions for future research.

This thesis originated from a critical observation of the Demand Driven Material Requirements Planning (DDMRP) methodology addressing the problem of the structural dichotomy between DDMRP planning and execution phases. While the industry has embraced DDMRP for its dynamic ability to align planning with market demand, the execution phase remains governed by static, ‘self-centered’ heuristics that ignore the stochastic reality of the shop floor.

We relied on Discrete Event Simulation (DES) to capture the complex, non-linear interactions between control policies and shop floor dynamics. By modeling a non-preemptive flowshop subjected to high variability demand, stochastic processing times, and order heterogeneity, we were able to build an experimental testbed to evaluate the DDMRP transition from ‘self-centered’ heuristics to a dynamic, status-centered execution system.

By addressing the identified gaps, this thesis clearly achieves the main research objective of increasing the robustness and performance of DDMRP managed flowshops. The findings confirm that transitioning DDMRP to a status-centered execution system significantly improves robustness and performance. First, implementing dynamic dispatching based on real-time inventory availability overcomes information latency and prevents stockouts more effectively than standard DDMRP sequencing. Second, establishing an order release gating mechanism proves essential for managing heterogeneous order streams, allowing the system to safely absorb high-priority rush orders without degrading the performance of standard orders, thus increasing robustness. Finally, integrating predictive analytics effectively transitions order generation from a reactive to a predictive process that strategically withholds orders during idleness to reduce average inventory while preserving service levels.

Collectively, these status-centered control mechanisms effectively close the demand driven execution loop. By reconfiguring DDMRP execution from pushing planned orders to dynamically controlling the workload based on actual shop floor data, we established an execution system that evolved from ‘blind’ to a reactive and predictive system, addressing fundamental challenges of the Volatile, Uncertain, Complex and Ambiguous (VUCA) environments.

Specifically, the system mitigates volatility by introducing an order release gate, preventing temporary volume fluctuations and heterogeneous orders from causing disruptions on the shop floor, maintaining robustness of the system. It addresses uncertainty by integrating predictive analytics to replace static lead time assumptions, allowing the execution engine to proactively adapt to operational disruptions. Furthermore, the status-centered execution handles the complexity of heterogeneous order streams by filtering demands at the entry point and dynamically resolving resource utilization conflicts. Finally, it reduces the ambiguity created by the information latency of the planning metrics used to govern shop floor execution, ensuring that shop floor priorities accurately reflect the physical reality of the inventory rather than outdated planning calculations.

8.1 Synthesis of Contributions

This thesis challenges the prevailing bias in DDMRP literature, focused on optimizing planning parameters, by proving that research must be directed toward the Demand Driven Execution (DDE) phase. Furthermore, this research expands the theoretical boundaries of DDMRP by introducing Workload Control (WLC) theory and predictive control through Machine Learning (ML) to the DDMRP framework, moving DDMRP framework to status-centered predictive models. It establishes that robustness in execution requires mechanisms decouples from static planning assumptions.

Methodologically, this thesis introduces novel approaches for evaluating complex Production Control Systems (PCSs). By combining DES with ML algorithms (to predict flow times) and surrogate models (to calibrate lead times), it creates an advanced experimental testbed. This methodological approach allows for isolation and testing of specific control loops (i.e., order sequencing, release and generation) in highly stochastic environments, offering a reproducible framework for future research on PCSs.

For industrial practitioners, this thesis provides actionable alternatives to the standard DDMRP execution rules. It demonstrates that managers must refrain from using planning metrics to guide shop floor sequencing opting instead for the use of real-time data. It provides empirical evidence that practitioners must abandon the immediate release of orders and implement a pre-shop pool of orders to control standard orders when rush orders occur. Finally, it offers a smart order generation process that managers can adopt to prevent premature inventory accumulation, highlighting the necessity of using real-time Manufacturing Execution Systems (MES) data to drive operations.

8.2 Limitations of the Contributions

The first limitation, common to all contributions of this thesis, is the configuration of the production environment. We studied a linear flowshop with a constant routing structure. While this represents a common configuration in manufacturing, other environments such as job shops or complex assembly lines may present different dynamics. For instance, in a job shop with re-entrant flows, the impact of dynamic dispatching might be even more pronounced due to the routing complexity.

Secondly, our models assumed perfect information availability for the status-centered execution system. For all applications, we assumed the system had real-time visibility of the status of all queues and buffers. Information asymmetry is an important cause of the ambiguity faced by VUCA environments. In practice, data latency or inaccuracy in the execution system could affect the performance of these dynamic rules. While perfect information is an idealized assumption, modern Industry 4.0 Manufacturing Execution Systems (MES) increasingly provide near real-time visibility (Zhong et al., 2013) make these status-centered rules feasible in modern contexts.

Thirdly, a specific limitation applies to the first and third contributions, which operated under the assumption that all replenishment orders were uniformly generated through the consumption of DDMRP buffers. As previously highlighted, actual DDMRP environments must manage different order streams: standard buffer depletion, demand spikes, and non-buffered products (Ptak & Smith, 2019). While our second contribution addressed this heterogeneity, the first and third contributions were tested under the assumption of a purely buffer-driven order stream. The introduction of non-buffered orders could alter the flow time prediction or complicate the dynamic dispatching logic.

Notably, beyond the common limitations, each contribution has its own specific constraints, with the most pertinent listed below. First, regarding the sequencing task investigation, the study focused exclusively on non-preemptive scheduling without consideration for setup times, making them independent of the sequence. Furthermore, while the dynamic rule improved service levels, the study did not quantify the potential increase in ‘system nervousness’ (frequency of priority changes) compared to the static rules, which could cause an impact at the operator level.

Second, considering the order release task contribution, heterogeneity was modeled in a binary manner (rush orders vs. standard orders). In industrial reality, order heterogeneity is multi-dimensional, involving variations in routing complexity and customer strategic values for example. In these scenarios, the proposed rules may not be sufficient to deal with levels of heterogeneity.

Thirdly, regarding the order generation application, the predictive model was trained on data generated by the simulation itself. In a real-world application, the ML model would need training on historical industrial data, which may contain noise, missing values or patterns do not present in the simulation. The use of real data gathered over time would require a robust training strategy.

Finally, the scope of the thesis was limited to the execution phase. We did not investigate the feedback loop from execution back to planning (e.g., how status-centered control mechanisms results should influence the future sizing of buffers in the demand driven planning phase). This integration remains a promising avenue for future research.

8.3 Research Perspectives

While this thesis establishes a foundational framework for status-centered demand driven execution system, it opens several avenues for future exploration. The literature review demonstrated that shifting the direction of DDMRP research to execution is primordial. Parameter optimization does not respond to the actual needs of practitioners, as it does not reflect the execution issues observed when DDMRP is applied. The execution aspects of DDMRP in practice remains a ‘black box’ as organizations fill the lacks on the theory with heuristics of their own, dampening the evolution of the method to a more comprehensive and adapted system. The limitations identified in the previous chapters provide the roadmap to where the research concerning DDMRP execution should evolve.

Future studies should extend the dynamic dispatching and release rules to more complex environments, such as Job Shops with re-entrant flows or assembly lines with component synchronization constraints. The interaction between dynamic execution rules and sequence-dependent setup times remains a critical area to be explored. Another important avenue for future research is the development of training protocols that allow ML models to adapt dynamically as the DDMRP production system evolves through time (e.g., changes in the output rate, changes in the workforce, changes in the product structure).

This thesis focused on closing the demand driven execution loop. Another promising direction is to study the feedback from execution back to planning. For example, can the pFT generated by the ML model be used to dynamically update the Decouple Lead Time (DLT) in the buffer sizing equation? This would create a fully autonomous, self-correcting Demand Driven Adaptive Enterprise (DDAE) model. From a practical perspective, the dynamic mechanisms proposed here

rely on accurate, real-time data. Practitioners must prioritize the implementation of systems that improve data visibility and reliability ensuring accurate and readily available shop floor information. Investigating how the status-centered execution system reacts to information asymmetry is an important avenue to address the ambiguity of current VUCA environments.

While studying and discussing DDMRP execution, several other questions were raised but not explored. This reflection aims researchers looking for the next frontier in the Demand Driven Theory. Beyond all the possible reflections, the most important concerns the synchronization of different DDMRP flows. In assembly shops, how do you manage buffers of components that are used in more than one final product? How to calculate the DLT when different parts managed by DDMRP must be synchronized in the same flow? How do you manage the ready-date and due-date differences of parts of the same final product? Should we anticipate production? Should we postpone production? All these questions are still unanswered and DDMRP theory will greatly benefit from the answers.

8.4 Personal Reflection

Lastly, a final personal reflection for the readers. Through this thesis we offered a fundamental perspective on the nature of control in manufacturing. We often equate efficiency with the ability to release orders and initiate activity, operating under the assumption that being idle is a waste of resources. However, our results demonstrate that robustness requires discipline to withhold action.

This momentary pause to evaluate the shop floor state before releasing an order (the ‘step back’) is what distinguishes reactive system from an intelligent one. By assessing the context before committing to execution, we replace the chaos of saturation with the stability of synchronization. Ultimately, taking the time to evaluate the environment is not a delay; it is the prerequisite for speed and efficiency.

This principle extends far beyond the shop floor, offering a valuable lesson for our own lives. In an era of constant connectivity, we adhere to the Immediate Release mentality, accepting every task or request the moment it arrives, under the illusion that constant activity equates to productivity. However, just as the shop floor floods without a gating mechanism, our minds become saturated with unfinished work, leading to congestion and stress. Just as a status-centered approach, by learning to receive feedback from our own state, we move from the chaos of attempting everything to the satisfaction of finishing what matters.

REFERENCES

- Abdelhalim, A., Hamid, A., & Tiente, H. (2021, 7-9 Jun. 2021). *Optimisation of the Automated buffer positioning model under DDMRP logic* IFAC Symposium on Information Control Problems in Manufacturing (INCOM 2021), Budapest, Hungary. <https://doi.org/10.1016/j.ifacol.2021.08.067>
- Achergui, A., Allaoui, H., & Hsu, T. (2020, 24-27 Nov. 2020). *Strategic DDMRP's Buffer Positioning for hybrid MTO/MTS manufacturing* IEEE International Conference on Technology Management, Operations and Decisions (ICTMOD 2020). <https://doi.org/10.1109/ICTMOD49425.2020.9380588>
- Alves, L., Ben Ali, M., & Baptiste, P. (2024, 8-12 Sep. 2024). *Dynamic Dispatching of DDMRP Replenishment Orders* IFIP International Conference on Advances in Production Management Systems (APMS 2024), Chemnitz, Germany. https://doi.org/10.1007/978-3-031-71645-4_4
- Alves, L., Ben Ali, M., & Baptiste, P. (2026). Robustness analysis of Demand Driven Material Requirements Planning flowshops in the presence of rush orders. *Journal of Industrial and Production Engineering*, 1-21. <https://doi.org/10.1080/21681015.2026.2613022>
- Aqlan, F., Ahmed, A., Ashour, O., Shamsan, A., & Hamasha, M. M. (2017). An approach for rush order acceptance decisions using simulation and multi-attribute utility theory. *European Journal of Industrial Engineering*, 11(5), 613-630. <https://doi.org/10.1504/ejie.2017.087680>
- Arisha, A., & Mahfouz, A. (2011, 11-14 Oct. 2011). *The Analysis of Rush Orders Risk in Supply Chain: A Simulation Approach* MODSIM World Conference & Expo (MODSIM 2011), Hampton, USA.
- Azzamouri, A., Baptiste, P., Dessevre, G., & Pellerin, R. (2021). Demand Driven Material Requirements Planning (DDMRP): A systematic review and classification. *Journal of Industrial Engineering and Management*, 14(3), 439-456. <https://doi.org/10.3926/jiem.3331>
- Azzamouri, A., Baptiste, P., Pellerin, R., & Dessevre, G. (2022). Impact of the continuous and periodic assessment of a buffer replenishment on the DDMRP method. *International Journal of Production Research*, 61(16), 5637-5658. <https://doi.org/10.1080/00207543.2022.2109219>
- Bagni, G., Godinho Filho, M., Thürer, M., & Stevenson, M. (2020). Systematic review and discussion of production control systems that emerged between 1999 and 2018. *Production Planning & Control*, 32(7), 511-525. <https://doi.org/10.1080/09537287.2020.1742398>
- Baptiste, P. (2018, 8-11 July 2018). *DDMRP: Scheduling opportunities in case of complex BOMs* International Conference on Information Systems, Logistics and Supply Chain (ILS 2018), Lyon, France.
- Barman, S., & Lisboa, J. (2010). Cost performance of simple priority rule combinations. *Journal of Manufacturing Technology Management*, 21(5), 567-584. <https://doi.org/10.1108/17410381011046977>

- Bayard, S., Grimaud, F., & Delorme, X. (2021). Study of buffer placement impacts on Demand Driven MRP performance. *IFAC-PapersOnLine*, 54(1), 1005-1010. <https://doi.org/10.1016/j.ifacol.2021.08.119>
- Bayard, S., Grimaud, F., & Delorme, X. (2024). Demand Driven Material Requirement Planning: Core concepts and analysis of its behavior on a case study. *IFAC-PapersOnLine*, 58(19), 1054-1059. <https://doi.org/10.1016/j.ifacol.2024.09.134>
- Baykasoğlu, A., & Ozsoydan, F. B. (2018). Dynamic scheduling of parallel heat treatment furnaces: A case study at a manufacturing system. *Journal of Manufacturing Systems*, 46, 152-162. <https://doi.org/10.1016/j.jmsy.2017.12.005>
- Blackstone, J. H., Phillips, D. T., & Hogg, G. L. (1981). A state-of-the-art survey of dispatching rules for manufacturing job shop operations. *International Journal of Production Research*, 20(1), 27-45. <https://doi.org/10.1080/00207548208947745>
- Burggraf, P., Wagner, J., Koke, B., & Steinberg, F. (2020). Approaches for the Prediction of Lead Times in an Engineer to Order Environment—A Systematic Review. *IEEE Access*, 8, 142434-142445. <https://doi.org/10.1109/access.2020.3010050>
- Butturi, M. A., De Rosa, G., Balugani, E., & Gamberini, R. (2021, 28-29 Oct. 2021). *Understanding the Demand Driven Material Requirements Planning Scope of Application: A Critical Literature Review* Danube Adria Association for Automation & Manufacturing (DAAAM 2021), Vienna, Austria. <https://doi.org/10.2507/32nd.daaam.proceedings.067>
- Carmo-Silva, S., & Fernandes, N. O. (2024, 26-28 Jun. 2024). *Aligning Order Execution Priority with Customer Demand in DDMRP - an Assessment by Simulation* International Conference Innovation in Engineering (ICIE 2024), Povoação, Portugal. https://doi.org/10.1007/978-3-031-62684-5_24
- Chang, P., Yeng, L., Cheng, Y., & Lin, Y. (2024). Assessing network reliability in a hybrid flow shop with rush order insertion. *Reliability Engineering & System Safety*, 245. <https://doi.org/10.1016/j.ress.2024.109990>
- Chen, C. L. (2010, 13-16 Jul. 2010). *A heuristic model for justifying the acceptance of rush orders in multi-plants manufacturing supply chain with outsourcing* IEEE International Conference on Industrial Informatics (INDIN 2010), Osaka. <https://doi.org/10.1109/INDIN.2010.5549674>
- Chintapalli, P., Disney, S., & Tang, C. (2017). Coordinating Supply Chains via Advance-Order Discounts, Minimum Order Quantities, and Delegations. *Production and Operations Management*, 26(12), 2175-2186. <https://doi.org/10.1111/poms.12751>
- Chu, C. H., & Hayya, J. C. (1988). Buffering decisions under MRP environment: A review. *Omega*, 16(4), 325-331. [https://doi.org/10.1016/0305-0483\(88\)90069-2](https://doi.org/10.1016/0305-0483(88)90069-2)
- Chua, P., Moon, S., Ng, Y., & Ng, H. (2022). A Surrogate Model to Predict Production Performance in Digital Twin-Based Smart Manufacturing. *Journal of Computing and Information Science in Engineering*, 22(3), 1-17. <https://doi.org/10.1115/1.4053038>
- Corsini, R. (2023). Adaptive production control of two-product closed-loop supply chain dynamics under disruptions. *Journal of Industrial and Production Engineering*, 40(8), 638-660. <https://doi.org/10.1080/21681015.2023.2256962>

- Cuartas, C., & Aguilar, J. (2022). Hybrid algorithm based on reinforcement learning for smart inventory management. *Journal of Intelligent Manufacturing*, 34, 123-149. <https://doi.org/10.1007/s10845-022-01982-5>
- Damand, D., Lahrichi, Y., & Barth, M. (2022a). Parameterisation of demand-driven material requirements planning: a multi-objective genetic algorithm. *International Journal of Production Research*, 61(15), 5134-5155. <https://doi.org/10.1080/00207543.2022.2098074>
- Damand, D., Lahrichi, Y., & Barth, M. (2022b). A simulation-optimization approach to parameterize Demand-Driven Material Requirements Planning. *IFAC-PapersOnLine*, 55(10), 263-268. <https://doi.org/10.1016/j.ifacol.2022.09.626>
- Dessevre, G., Baptiste, P., Lamothe, J., & Pellerin, R. (2021). Visual charts produced by simulation to correlate service rate, resource utilization and DDMRP parameters. *International Journal of Production Research*, 61(3), 741-753. <https://doi.org/10.1080/00207543.2021.2015808>
- Dessevre, G., Baptiste, P., Martin, G., Lamothe, J., & Luras, M. (2019, 25-28 Jun. 2019). *Étude d'impact du paramétrage des temps de défilement sur la performance d'un déploiement de la méthode DDMRP* International Conference on Industrial Engineering (CIGI QUALITA 2019), Montreal, Canada.
- Dessevre, G., & Ben Ali, M. (2024). Analysis of a new dynamic capacity management approach in DDMRP: Application on a real industrial case. *Journal of Industrial Engineering and Management*, 17(2), 445-462. <https://doi.org/10.3926/jiem.6045>
- Dessevre, G., Lamothe, J., Pellerin, R., Ben Ali, M., Baptiste, P., & Pomponne, V. (2023). Comparison of pull management policies for a divergent process with DDMRP buffers: an industrial case study. *International Journal of Production Research*, 61(23), 8022-8042. <https://doi.org/10.1080/00207543.2022.2162997>
- Dessevre, G., Lamothe, J., Pomponne, V., Baptiste, P., Luras, M., & Pellerin, R. (2020, 22-24 Apr. 2020). *A DDMRP implementation user feedbacks and stakes analysis* International Conference on Information Systems, Logistics and Supply Chain (ILS 2020), Austin, United States. <https://hal-mines-albi.archives-ouvertes.fr/hal-02940566>
- Dessevre, G., Martin, G., Baptiste, P., Lamothe, J., Pellerin, R., & Luras, M. (2019, 25-27 Sep. 2019). *Decoupled Lead Time in finite capacity flowshop: a feedback loop approach* International Conference on Industrial Engineering and Systems Management (IESM 2019), Shanghai. <https://doi.org/10.1109/IESM45758.2019.8948198>
- Djedidi, K. (2025). *Prédiction des retards des ordres de réapprovisionnement dans un système de production géré par DDMRP* [Master's Thesis, Polytechnique Montréal]. PolyPublie. <https://publications.polymtl.ca/65779/>
- Duhem, L., Ben Ali, M., & Martin, G. (2023). Parametrization of a demand-driven operating model using reinforcement learning. *Computers in Industry*, 147, 1-12. <https://doi.org/10.1016/j.compind.2023.103874>
- El Marzougui, M., Messaoudi, N., Dachry, W., & Bensassi, B. (2024). A Model for Decision-making to Parameterizing Demand Driven Material Requirement Planning Using Deep Reinforcement Learning. *Production Engineering Archives*, 30(3), 377-393. <https://doi.org/10.30657/pea.2024.30.37>

- Favaretto, D., Marin, A., & Tolotti, M. (2023). A theoretical validation of the DDMRP reorder policy. *Computational Management Science*, 20(8), 1-28. <https://doi.org/10.1007/s10287-023-00443-5>
- Fernandes, N., Almeida, J., Ferreira, L., Ávila, P., & Carmo-Silva, S. (2025, 18-20 Jun. 2025). *Combining DDMRP and CONWIP: A Simulation Study of the Pool-Sequencing Rule* International Conference Innovation in Engineering (ICIE 2025), Prague, Czech Republic. https://doi.org/10.1007/978-3-031-94484-0_10
- Fernandes, N., Guedes, N., Thürer, M., Ferreira, L. P., & Ávila, P. (2023, 18-22 Jun. 2023). *How to Prioritize Replenishment Orders in DDMRP: A Simulation Study* Flexible Automation and Intelligent Manufacturing (FAIM 2023), Porto, Portugal. https://doi.org/10.1007/978-3-031-38165-2_12
- Fernandes, N., Thürer, M., & Silva, S. C. (2024). DDMRP relative priority for production execution: an assessment by simulation. *Operations Management Research*, 17, 1244-1256. <https://doi.org/10.1007/s12063-024-00503-2>
- Gómez Paredes, F., Godinho Filho, M., Thürer, M., Fernandes, N., & Jabbour, C. (2020). Factors for choosing production control systems in make-to-order shops: a systematic literature review. *Journal of Intelligent Manufacturing*, 33(3), 639-674. <https://doi.org/10.1007/s10845-020-01673-z>
- Graves, R. J., Konopka, J. M., & Milne, R. (1995). Literature review of material flow control mechanisms. *Production Planning & Control*, 6(5), 395-403. <https://doi.org/10.1080/09537289508930296>
- Guide, V. D. R., & Srivastava, R. (2000). A review of techniques for buffering against uncertainty with MRP systems. *Production Planning & Control*, 11(3), 223-233. <https://doi.org/10.1080/095372800232199>
- Gyulai, D., Pfeiffer, A., Nick, G., Gallina, V., Sihn, W., & Monostori, L. (2018). Lead time prediction in a flow-shop environment with analytical and machine learning approaches. *IFAC-PapersOnLine*, 51(11), 1029-1034. <https://doi.org/10.1016/j.ifacol.2018.08.472>
- Habbadi, S., El Mouayni, I., Herrou, B., & Sekkat, S. (2025). New Hybrid Method for Buffer Positioning and Production Control Using DDMRP Logic in Smart Manufacturing. *Journal of Manufacturing and Materials Processing*, 9(7). <https://doi.org/10.3390/jmmp9070219>
- Hammler, P., Riesterer, N., & Braun, T. (2023). Fully dynamic reorder policies with deep reinforcement learning for multi-echelon inventory management. *Informatik Spektrum*, 46(5-6), 240-251. <https://doi.org/10.1007/s00287-023-01556-6>
- Haupt, R. (1989). A survey of priority rule-based scheduling. *OR Spektrum*, 11(1), 3-16. <https://doi.org/10.1007/bf01721162>
- He, X., Dong, S., & Zhao, N. (2019). Research on rush order insertion rescheduling problem under hybrid flow shop based on NSGA-III. *International Journal of Production Research*, 58(4), 1161-1177. <https://doi.org/10.1080/00207543.2019.1613581>
- Hendry, L., Land, M., Stevenson, M., & Gaalman, G. (2008). Investigating implementation issues for workload control (WLC): A comparative case study analysis. *International Journal of Production Economics*, 112(1), 452-469. <https://doi.org/10.1016/j.ijpe.2007.05.012>

- Hopp, W. J., & Spearman, M. L. (2011). *Factory Physics* (3rd ed.). Waveland Press.
- Hubl, A., Jodlbauer, H., & Altendorfer, K. (2013). Influence of dispatching rules on average production lead time for multi-stage production systems. *International Journal of Production Economics*, 144(2), 479-484. <https://doi.org/10.1016/j.ijpe.2013.03.020>
- Hutter, T., Haeussler, S., & Missbauer, H. (2017). Successful implementation of an order release mechanism based on workload control: a case study of a make-to-stock manufacturer. *International Journal of Production Research*, 56(4), 1565-1580. <https://doi.org/10.1080/00207543.2017.1369598>
- Jäger, Y., & Roser, C. (2018, 26-30 Aug. 2018). *Effect of Prioritization on the Waiting Time* Advances in Production Management Systems (APMS 2018), Seoul, South Korea. https://doi.org/10.1007/978-3-319-99704-9_3
- Karrer, C., Alicke, K., & Günther, H. O. (2012). A framework to engineer production control strategies and its application in electronics manufacturing. *International Journal of Production Research*, 50(22), 6595-6611. <https://doi.org/10.1080/00207543.2012.658479>
- Kortabarria, A., Apaolaza, U., Lizarralde, A., & Amorortu, I. (2018). Material management without forecasting: From MRP to demand driven MRP. *Journal of Industrial Engineering and Management*, 11(4), 632-650. <https://doi.org/10.3926/jiem.2654>
- Krajčovič, M., Gabajová, G., Gašo, M., & Schickerle, M. (2024). Parameter Setting for Strategic Buffers in Demand-Driven Material Resource Planning through Statistical Analysis and Optimisation of Buffer Levels. *Applied Sciences*, 14(7), 3012-3036. <https://doi.org/10.3390/app14073012>
- Kumar, S., & Sharma, S. C. (2023). Integrated Model for Predicting Supply Chain Risk Through Machine Learning Algorithms. *International Journal of Mathematical, Engineering and Management Sciences*, 8(3), 353-373. <https://doi.org/10.33889/ijmems.2023.8.3.021>
- Lahrichi, Y., Damand, D., & Barth, M. (2022). A first MILP model for the parameterization of Demand-Driven MRP. *Computers & Industrial Engineering*, 174. <https://doi.org/10.1016/j.cie.2022.108769>
- Lam, C. Y., & Ip, W. H. (2011). Constraint priority scheduling using an agent-based approach. *Industrial Management & Data Systems*, 111(2), 246-263. <https://doi.org/10.1108/02635571111115164>
- Land, M., & Gaalman, G. (1996). Workload control concepts in job shops A critical assessment. *International Journal of Production Economics*, 46-47, 535-548. [https://doi.org/10.1016/s0925-5273\(96\)00088-6](https://doi.org/10.1016/s0925-5273(96)00088-6)
- Lee, C., & Rim, S. (2019). A Mathematical Safety Stock Model for DDMRP Inventory Replenishment. *Mathematical Problems in Engineering*, 2019, 1-10. <https://doi.org/10.1155/2019/6496309>
- Lödding, H. (2013). *Handbook of Manufacturing Control* (R. Rossi, Trans.; 1st ed.). Springer.
- Lödding, H., Yu, K. W., & Wiendahl, H. P. (2003). Decentralized WIP-oriented manufacturing control (DEWIP). *Production Planning & Control*, 14(1), 42-54. <https://doi.org/10.1080/0953728021000078701>

- Martin, G., Luras, M., & Baptiste, P. (2023). Dynamical multi-parameter sizing of DDMRP buffers in finite capacity flow-shops. *Computers & Industrial Engineering*, 175, 1-10. <https://doi.org/10.1016/j.cie.2022.108858>
- Math, M., Gopinath, D., & Biradar, B. S. (2024). Demand Driven Material Requirements Planning: An Inventory Optimization Model. *International Journal of Agricultural and Statistical Sciences*, 20(01), 105-111. <https://doi.org/10.59467/ijass.2024.20.105>
- Miclo, R., Fontanili, F., Luras, M., Lamothe, J., & Milian, B. (2015, 21-23 Oct. 2015). *MRP vs. demand-driven MRP: Towards an objective comparison* International Conference on Industrial Engineering and Systems Management (IESM 2015), Seville, Spain. <https://doi.org/10.1109/IESM.2015.7380288>
- Miclo, R., Luras, M., Fontanili, F., Lamothe, J., & Melnyk, S. A. (2018). Demand Driven MRP: assessment of a new approach to materials management. *International Journal of Production Research*, 57(1), 166-181. <https://doi.org/10.1080/00207543.2018.1464230>
- Miguel, H., González, G., Antonio, L., Moreno, G., Luisa, A., & Corralejo, G. (2021, 15-17 Dec. 2021). *Hybrid Push-Pull Systems: Classification, Optimal Methods and Uncertainty* International Conference on Computational Science and Computational Intelligence (CSCI 2021), Las Vegas, USA. <https://doi.org/10.1109/CSCI54926.2021.00167>
- Milne, R., & Wang, C. (2014). Enhancing mathematical programming models to account for demand priorities increasing as a function of delivery date. *Journal of Industrial and Production Engineering*, 31(1), 51-63. <https://doi.org/10.1080/21681015.2014.881426>
- Moratori, P., Petrovic, S., & Vázquez-Rodríguez, J. A. (2010). Integrating rush orders into existent schedules for a complex job shop problem. *Applied Intelligence*, 32(2), 205-215. <https://doi.org/10.1007/s10489-010-0215-6>
- Mourtzis, D. (2019). Simulation in the design and operation of manufacturing systems: state of the art and new trends. *International Journal of Production Research*, 58(7), 1927-1949. <https://doi.org/10.1080/00207543.2019.1636321>
- Mourtzis, D., Doukas, M., Fragou, K., Efthymiou, K., & Matzorou, V. (2014). Knowledge-based Estimation of Manufacturing Lead Time for Complex Engineered-to-order Products. *Procedia CIRP*, 17, 499-504. <https://doi.org/10.1016/j.procir.2014.01.087>
- Mundt, C., Beck, E. B., & Lödding, H. (2022, 25-29 Sep. 2022). *Determining Due Dates for Rush Orders* Advances in Production Management Systems (APMS 2022), Gyeongju, South Korea. https://doi.org/10.1007/978-3-031-16407-1_35
- Murphy, R., Newell, A., Hargaden, V., & Papakostas, N. (2019). Machine learning technologies for order flowtime estimation in manufacturing systems. *Procedia CIRP*, 81, 701-706. <https://doi.org/10.1016/j.procir.2019.03.179>
- Neuner, P., Haeussler, S., Fodor, J., & Blossey, G. (2022, 11-14 Dec. 2022). *Putting a Price Tag on Hot Lots and Expediting in Semiconductor Manufacturing* Winter Simulation Conference (WSC 2022), Singapore. <https://doi.org/10.1109/WSC57314.2022.10015256>
- Ntakolia, C., Kokkotis, C., Karlsson, P., & Moustakidis, S. (2021). An Explainable Machine Learning Model for Material Backorder Prediction in Inventory Management. *Sensors (Basel)*, 21(23). <https://doi.org/10.3390/s21237926>

- Pekarcıkova, M., Trebuna, P., Kliment, M., & Trojan, J. (2019). Demand driven material requirements planning. Some methodical and practical comments. *Management and Production Engineering Review*, 10(2), 50-59. <https://doi.org/https://doi.org/10.24425/mper.2019.129568>
- Pfeiffer, A., Gyulai, D., Kádár, B., & Monostori, L. (2016). Manufacturing Lead Time Estimation with the Combination of Simulation and Statistical Learning Methods. *Procedia CIRP*, 41, 75-80. <https://doi.org/10.1016/j.procir.2015.12.018>
- Pinedo, M. (2022). *Scheduling: Theory, Algorithms, and Systems* (6th ed.). Springer Nature.
- Priore, P., Ponte, B., Rosillo, R., & de la Fuente, D. (2018). Applying machine learning to the dynamic selection of replenishment policies in fast-changing supply chain environments. *International Journal of Production Research*, 57(11), 3663-3677. <https://doi.org/10.1080/00207543.2018.1552369>
- Psarommatis, F., Zheng, X., & Kiritsis, D. (2021). A two-layer criteria evaluation approach for re-scheduling efficiently semi-automated assembly lines with high number of rush orders. *Procedia CIRP*, 97, 172-177. <https://doi.org/10.1016/j.procir.2020.05.221>
- Ptak, C., & Smith, C. (2011). *Orlicky's material requirements planning* (3rd ed.). McGraw Hill.
- Ptak, C., & Smith, C. (2018). *The Demand Driven Adaptive Enterprise: Surviving, Adapting and Thriving in a VUCA World* (1st ed.). Industrial Press Inc.
- Ptak, C., & Smith, C. (2019). *DDMRP Demand Driven Material Requirements Planning: Version 3* (3rd ed.). Industrial Press Inc.
- Ren, X., Wang, X., Geng, N., & Jiang, Z. (2021, 23-27 Aug. 2021). *The Just-In-Time Job-Shop Rescheduling with Rush Orders by Using a Meta-Heuristic Algorithm* International Conference on Automation Science and Engineering (CASE 2021), Lyon. <https://doi.org/10.1109/CASE49439.2021.9551573>
- Saoud, P., Kourentzes, N., & Boylan, J. E. (2025). The importance of forecast uncertainty in understanding the Bullwhip effect. *International Journal of Production Research*, 1-22. <https://doi.org/10.1080/00207543.2025.2527957>
- Schneckenreither, M., Haeussler, S., & Gerhold, C. (2020). Order release planning with predictive lead times: a machine learning approach. *International Journal of Production Research*, 59(11), 3285-3303. <https://doi.org/10.1080/00207543.2020.1859634>
- Scrivano, S., Tan, B., & Tolio, T. (2023). Continuous-flow simulation of manufacturing systems with assembly/disassembly machines, multiple loops and general layout. *Journal of Manufacturing Systems*, 69, 103-118. <https://doi.org/10.1016/j.jmsy.2023.05.028>
- Shofa, M. J., Moeis, A. O., & Restiana, N. (2018). Effective production planning for purchased part under long lead time and uncertain demand: MRP Vs demand-driven MRP. *IOP Conference Series: Materials Science and Engineering*, 337, 1-7. <https://doi.org/10.1088/1757-899x/337/1/012055>
- Shofa, M. J., & Widyarto, W. O. (2016, 7-8 Dec. 2016). *Effective production control in an automotive industry: MRP vs. demand-driven MRP* 3rd International Conference on Engineering, Technology and Industrial Application (ICETIA 2016), Sukarta, Indonesia. <https://doi.org/10.1063/1.4985449>

- Spearman, M. L., Woodruff, D. L., & Hopp, W. J. (1990). CONWIP: a pull alternative to kanban. *International Journal of Production Research*, 28(5), 879-894. <https://doi.org/10.1080/00207549008942761>
- Stevenson, M., Hendry, L. C., & Kingsman †, B. G. (2005). A review of production planning and control: the applicability of key concepts to the make-to-order industry. *International Journal of Production Research*, 43(5), 869-898. <https://doi.org/10.1080/0020754042000298520>
- Takami Narita, V., Ikeziri, L., & Souza, F. (2021). Evaluation of dynamic buffer management for adjusting stock level: a simulation-based approach. *Journal of Industrial and Production Engineering*, 38(6), 452-465. <https://doi.org/10.1080/21681015.2021.1931493>
- Taskan, B., Junça-Silva, A., & Caetano, A. (2022). Clarifying the conceptual map of VUCA: a systematic review. *International Journal of Organizational Analysis*, 30(7), 196-217. <https://doi.org/10.1108/ijoa-02-2022-3136>
- Thevenin, S., Adulyasak, Y., & Cordeau, J. F. (2020). Material Requirements Planning Under Demand Uncertainty Using Stochastic Optimization. *Production and Operations Management*, 30(2), 475-493. <https://doi.org/10.1111/poms.13277>
- Thürer, M., Fernandes, N. O., & Stevenson, M. (2020). Production planning and control in multi-stage assembly systems: an assessment of Kanban, MRP, OPT (DBR) and DDMRP by simulation. *International Journal of Production Research*, 60(3), 1036-1050. <https://doi.org/10.1080/00207543.2020.1849847>
- Thürer, M., Silva, C., & Stevenson, M. (2009). Workload control release mechanisms: from practice back to theory building. *International Journal of Production Research*, 48(12), 3593-3617. <https://doi.org/10.1080/00207540902922810>
- Thürer, M., Stevenson, M., Silva, C., Land, M. J., & Fredendall, L. D. (2012). Workload Control and Order Release: A Lean Solution for Make-to-Order Companies. *Production and Operations Management*, 21(5), 939-953. <https://doi.org/10.1111/j.1937-5956.2011.01307.x>
- Trzyna, D., Kuyumcu, A., & Lödding, H. (2012). Throughput Time Characteristics of Rush Orders and their Impact on Standard Orders. *Procedia CIRP*, 3, 311-316. <https://doi.org/10.1016/j.procir.2012.07.054>
- Uzun Araz, O., Ilgin, M. A., Eski, O., & Araz, C. (2024). Fuzzy demand-driven material requirements planning: a comprehensive analysis of fuzzy logic implementation in DDMRP. *International Journal of Production Research*, 62(21), 7793-7811. <https://doi.org/10.1080/00207543.2024.2328770>
- Velasco Acosta, A. P., Mascle, C., & Baptiste, P. (2019). Applicability of Demand-Driven MRP in a complex manufacturing environment. *International Journal of Production Research*, 58(14), 4233-4245. <https://doi.org/10.1080/00207543.2019.1650978>
- Waller, M., & Fawcett, S. (2013). Data Science, Predictive Analytics, and Big Data: A Revolution That Will Transform Supply Chain Design and Management. *Journal of Business Logistics*, 34(2), 77-84. <https://doi.org/10.1111/jbl.12010>

- Walraevens, J., Van Giel, T., De Vuyst, S., & Wittevrongel, S. (2022). Asymptotics of waiting time distributions in the accumulating priority queue. *Queueing Systems*, 101(3-4), 221-244. <https://doi.org/10.1007/s11134-022-09839-7>
- Wang, Z., & Hu, H. (2018). Dynamic response to demand variability for precast production rescheduling with multiple lines. *International Journal of Production Research*, 56(16), 5386-5401. <https://doi.org/10.1080/00207543.2017.1414970>
- Wang, Z., Liu, Y., Hu, H., & Dai, L. (2021). Hybrid Rescheduling Optimization Model under Disruptions in Precast Production Considering Real-World Environment. *Journal of Construction Engineering and Management*, 147(4), 1-16. [https://doi.org/10.1061/\(asce\)co.1943-7862.0001976](https://doi.org/10.1061/(asce)co.1943-7862.0001976)
- Werth, B., Karder, J., Beham, A., & Altendorfer, K. (2023). Simulation-based Optimization of Material Requirements Planning Parameters. *Procedia Computer Science*, 217, 1117-1126. <https://doi.org/10.1016/j.procs.2022.12.310>
- Xanthopoulos, A., Chnitidis, G., & Koulouriotis, D. (2019). Reinforcement learning-based adaptive production control of pull manufacturing systems. *Journal of Industrial and Production Engineering*, 36(5), 313-323. <https://doi.org/10.1080/21681015.2019.1647301>
- Yao, J., & Liu, L. (2009). Optimization analysis of supply chain scheduling in mass customization. *International Journal of Production Economics*, 117(1), 197-211. <https://doi.org/10.1016/j.ijpe.2008.10.008>
- Younespour, M., Esmaelian, M., & Kianfar, K. (2024). Optimizing the strategic and operational levels of demand-driven MRP using a hybrid GA-PSO algorithm. *Computers & Industrial Engineering*, 193, 1-15. <https://doi.org/10.1016/j.cie.2024.110306>
- Zhang, L., Hu, Y., Tang, Q., Li, J., & Li, Z. (2021). Data-Driven Dispatching Rules Mining and Real-Time Decision-Making Methodology in Intelligent Manufacturing Shop Floor with Uncertainty. *Sensors (Basel)*, 21(14). <https://doi.org/10.3390/s21144836>
- Zheng, S., Gupta, C., & Serita, S. (2019, 16-20 Sep. 2019). *Manufacturing Dispatching Using Reinforcement and Transfer Learning* European Conference on Machine Learning and Principles and Practices of Knowledge Discovery in Databases (ECML-PKDD 2019), Wurzburg, Germany. https://doi.org/10.1007/978-3-030-46133-1_39
- Zhong, R., Dai, Q., Qu, T., Hu, G. J., & Huang, G. (2013). RFID-enabled real-time manufacturing execution system for mass-customization production. *Robotics and Computer-Integrated Manufacturing*, 29(2), 283-292. <https://doi.org/10.1016/j.rcim.2012.08.001>

APPENDIX A SIMULATION RESULTS

Table A.1 Simulation Results for STATIC dispatching rule (95% CI)

LR	SL (%)	I (units)	SO	BO (units)	OHA (units)	MBOT(hours)
80%	80.4 ± 0.18	3888.74 ± 12.92	2936 ± 31.46	78844 ± 1030	4526 ± 38.94	3.05 ± 0.03
81%	79.3 ± 0.2	3894.02 ± 15.32	3107 ± 35.02	84105 ± 1184	4744 ± 43.76	3.15 ± 0.03
82%	78.9 ± 0.23	3896.49 ± 14.38	3161 ± 38.99	86498 ± 1359	4800 ± 45.24	3.22 ± 0.03
83%	78.4 ± 0.23	3899.67 ± 15.68	3239 ± 39.94	88863 ± 1412	4902 ± 46.68	3.26 ± 0.03
84%	77.7 ± 0.27	3901.67 ± 18.89	3341 ± 46.59	92308 ± 1718	5027 ± 55.26	3.34 ± 0.03
85%	77.1 ± 0.2	3906.73 ± 14.54	3438 ± 33.78	95335 ± 1290	5153 ± 41.5	3.42 ± 0.03
86%	75.8 ± 0.33	3915.87 ± 24.31	3636 ± 56.86	102365 ± 2117	5382 ± 67.17	3.58 ± 0.04
87%	75.3 ± 0.33	3918.91 ± 23.61	3709 ± 55.81	104930 ± 2220	5467 ± 66.25	3.64 ± 0.05
88%	74.6 ± 0.29	3924.87 ± 22.97	3811 ± 49.24	108635 ± 1858	5591 ± 59.86	3.74 ± 0.05
89%	73.7 ± 0.36	3930.25 ± 26.65	3945 ± 62.13	113057 ± 2286	5752 ± 71.16	3.83 ± 0.05
90%	72.9 ± 0.3	3937.57 ± 21.88	4068 ± 50.59	117762 ± 1973	5898 ± 57.73	3.94 ± 0.04
91%	71.1 ± 0.45	3953.88 ± 33.36	4350 ± 74.42	128000 ± 2950	6221 ± 82.52	4.21 ± 0.08
92%	70.4 ± 0.47	3957.07 ± 34.42	4432 ± 78.9	131153 ± 3052	6294 ± 89.19	4.29 ± 0.07
93%	68.7 ± 0.59	3978.69 ± 44.11	4705 ± 97.69	142000 ± 3795	6597 ± 104.1	4.58 ± 0.11
94%	67.3 ± 0.55	3993.27 ± 46.22	4904 ± 90.23	149833 ± 3785	6809 ± 97.6	4.79 ± 0.11
95%	64.8 ± 0.72	4027.93 ± 58.26	5294 ± 116.8	165607 ± 4987	7209 ± 120.7	5.26 ± 0.16
96%	62.4 ± 0.86	4064.95 ± 71.39	5644 ± 139.7	180045 ± 6115	7562 ± 138.7	5.71 ± 0.19
97%	58.5 ± 1.35	4140.49 ± 116	6236 ± 214.8	205157 ± 9468	8141 ± 206.7	6.57 ± 0.35
98%	52.7 ± 2.13	4306.5 ± 215.52	7107 ± 333.2	245474 ± 15900	8942 ± 295.7	8.1 ± 0.71
99%	37.9 ± 3.71	5107.14 ± 569.4	9334 ± 581	355141 ± 30090	10838 ± 475.5	13.82 ± 2.2

Table A.2 Simulation Results for NFOHP dispatching rule (95% CI)

LR	SL (%)	I (units)	SO (units)	BO (units)	OHA (units)	MBOT(hours)
80%	80.2 $\pm$ 0.2	3885.08 $\pm$ 14.46	2977 $\pm$ 34.64	78044 $\pm$ 1187	4620 $\pm$ 43.35	2.84 $\pm$ 0.02
81%	78.9 $\pm$ 0.31	3890.96 $\pm$ 21	3165 $\pm$ 53.61	84108 $\pm$ 1899	4856 $\pm$ 66.29	2.94 $\pm$ 0.03
82%	78.8 $\pm$ 0.23	3891.08 $\pm$ 16.97	3181 $\pm$ 39.36	84269 $\pm$ 1360	4883 $\pm$ 48.18	2.96 $\pm$ 0.03
83%	78.2 $\pm$ 0.23	3894.47 $\pm$ 17.4	3268 $\pm$ 41.42	87297 $\pm$ 1379	4984 $\pm$ 50.75	3.01 $\pm$ 0.03
84%	77.6 $\pm$ 0.29	3897.84 $\pm$ 21.09	3361 $\pm$ 50.32	89990 $\pm$ 1706	5115 $\pm$ 61.21	3.05 $\pm$ 0.03
85%	76.8 $\pm$ 0.27	3900.74 $\pm$ 20.1	3481 $\pm$ 48.14	93848 $\pm$ 1666	5262 $\pm$ 61.51	3.11 $\pm$ 0.03
86%	75.8 $\pm$ 0.31	3905.75 $\pm$ 20.49	3631 $\pm$ 52.23	98775 $\pm$ 1733	5448 $\pm$ 60.98	3.18 $\pm$ 0.04
87%	75.3 $\pm$ 0.31	3908.78 $\pm$ 23.53	3709 $\pm$ 52.93	101320 $\pm$ 1861	5560 $\pm$ 67.65	3.23 $\pm$ 0.03
88%	74.4 $\pm$ 0.29	3914.32 $\pm$ 21.39	3846 $\pm$ 49.8	105636 $\pm$ 1735	5722 $\pm$ 58.58	3.32 $\pm$ 0.03
89%	73.5 $\pm$ 0.39	3916.14 $\pm$ 28.53	3981 $\pm$ 65.57	109625 $\pm$ 2360	5890 $\pm$ 77.57	3.35 $\pm$ 0.04
90%	72.7 $\pm$ 0.34	3922.62 $\pm$ 24.6	4092 $\pm$ 56.64	113720 $\pm$ 2005	6019 $\pm$ 65.49	3.45 $\pm$ 0.04
91%	70.8 $\pm$ 0.48	3934.49 $\pm$ 34.03	4385 $\pm$ 79.61	123749 $\pm$ 2993	6383 $\pm$ 88.08	3.6 $\pm$ 0.06
92%	69.8 $\pm$ 0.49	3941.69 $\pm$ 36.2	4535 $\pm$ 80.12	129247 $\pm$ 3001	6553 $\pm$ 91.57	3.69 $\pm$ 0.06
93%	68.4 $\pm$ 0.52	3951.21 $\pm$ 40.52	4744 $\pm$ 86.89	136678 $\pm$ 3255	6791 $\pm$ 97.19	3.83 $\pm$ 0.06
94%	66.5 $\pm$ 0.64	3968.55 $\pm$ 47.63	5038 $\pm$ 105.5	147799 $\pm$ 4024	7118 $\pm$ 114.9	4.06 $\pm$ 0.09
95%	63.6 $\pm$ 0.99	3999.74 $\pm$ 79.98	5474 $\pm$ 158.7	164444 $\pm$ 6712	7575 $\pm$ 158	4.41 $\pm$ 0.2
96%	61.1 $\pm$ 1.36	4029.7 $\pm$ 106.07	5842 $\pm$ 215.6	179340 $\pm$ 9147	7947 $\pm$ 211.1	4.78 $\pm$ 0.26
97%	55.5 $\pm$ 1.49	4113.73 $\pm$ 126	6694 $\pm$ 234.1	214139 $\pm$ 10400	8796 $\pm$ 217.7	5.62 $\pm$ 0.44
98%	51.7 $\pm$ 1.63	4182.68 $\pm$ 136.9	7241 $\pm$ 255.1	238022 $\pm$ 11490	9300 $\pm$ 233.6	6.36 $\pm$ 0.39
99%	35.6 $\pm$ 3.78	4974.96 $\pm$ 519.1	9679 $\pm$ 589.6	359376 $\pm$ 30800	11311 $\pm$ 465.7	12.14 $\pm$ 2.03

Table A.3 Simulation Results for DAOHP dispatching rule (95% CI)

LR	SL (%)	I (units)	SO (units)	BO (units)	OHA (units)	MBOT(hours)
80%	81.8 $\pm$ 0.22	3882.62 $\pm$ 16.38	2733 $\pm$ 37.66	72786 $\pm$ 1209	4285 $\pm$ 46.2	2.88 $\pm$ 0.02
81%	81 $\pm$ 0.23	3886.3 $\pm$ 19.52	2848 $\pm$ 40.66	75772 $\pm$ 1273	4439 $\pm$ 55.43	2.94 $\pm$ 0.03
82%	80.7 $\pm$ 0.	3887.32 $\pm$ 17.2	2891 $\pm$ 35.55	77123 $\pm$ 1218	4485 $\pm$ 48.25	2.96 $\pm$ 0.02
83%	80.4 $\pm$ 0.23	3889.14 $\pm$ 17.34	2948 $\pm$ 39.13	79200 $\pm$ 1363	4575 $\pm$ 49.73	3.02 $\pm$ 0.03
84%	79.6 $\pm$ 0.2	3892.24 $\pm$ 18.85	3058 $\pm$ 34.5	82368 $\pm$ 1158	4721 $\pm$ 45.5	3.07 $\pm$ 0.02
85%	79.2 $\pm$ 0.24	3894.35 $\pm$ 19.27	3114 $\pm$ 40	84063 $\pm$ 1382	4793 $\pm$ 50.04	3.12 $\pm$ 0.03
86%	78.2 $\pm$ 0.31	3900.94 $\pm$ 21.85	3274 $\pm$ 51.51	88951 $\pm$ 1795	5014 $\pm$ 61.94	3.23 $\pm$ 0.04
87%	78 $\pm$ 0.31	3901.03 $\pm$ 24.23	3299 $\pm$ 53.12	89840 $\pm$ 1821	5043 $\pm$ 66.62	3.22 $\pm$ 0.03
88%	76.9 $\pm$ 0.3	3907.32 $\pm$ 23.14	3476 $\pm$ 50.67	95364 $\pm$ 1768	5274 $\pm$ 61.56	3.34 $\pm$ 0.03
89%	76 $\pm$ 0.43	3912.54 $\pm$ 32.92	3603 $\pm$ 71.73	99617 $\pm$ 2437	5438 $\pm$ 86.92	3.4 $\pm$ 0.04
90%	75.3 $\pm$ 0.4	3915.62 $\pm$ 29.59	3703 $\pm$ 65.31	102945 $\pm$ 2365	5554 $\pm$ 76.3	3.49 $\pm$ 0.05
91%	74 $\pm$ 0.44	3924.55 $\pm$ 32.77	3893 $\pm$ 72.29	109112 $\pm$ 2557	5803 $\pm$ 82.51	3.6 $\pm$ 0.05
92%	72.8 $\pm$ 0.43	3931.62 $\pm$ 34.18	4084 $\pm$ 70.86	115302 $\pm$ 2668	6021 $\pm$ 82.35	3.73 $\pm$ 0.07
93%	71.5 $\pm$ 0.66	3940.86 $\pm$ 47.57	4284 $\pm$ 106.7	122236 $\pm$ 4027	6247 $\pm$ 117.7	3.86 $\pm$ 0.1
94%	69.7 $\pm$ 0.73	3956.32 $\pm$ 56.27	4546 $\pm$ 118.2	131928 $\pm$ 4522	6557 $\pm$ 128.9	4.07 $\pm$ 0.12
95%	67.4 $\pm$ 0.96	3975.58 $\pm$ 71.16	4897 $\pm$ 152.9	144702 $\pm$ 6049	6945 $\pm$ 158.5	4.35 $\pm$ 0.15
96%	63.6 $\pm$ 1.47	4016.6 $\pm$ 111.43	5473 $\pm$ 233.6	166345 $\pm$ 9533	7552 $\pm$ 231.3	4.81 $\pm$ 0.27
97%	57.9 $\pm$ 1.91	4103.17 $\pm$ 153.7	6319 $\pm$ 300.5	201786 $\pm$ 13530	8369 $\pm$ 273.5	5.75 $\pm$ 0.44
98%	53.7 $\pm$ 2.44	4219.27 $\pm$ 248.6	6956 $\pm$ 377.5	230211 $\pm$ 18240	8951 $\pm$ 322	6.76 $\pm$ 0.87
99%	36 $\pm$ 3.66	4763.71 $\pm$ 384.1	9616 $\pm$ 569.8	356796 $\pm$ 29170	11201 $\pm$ 453.5	10.74 $\pm$ 1.22

APPENDIX B ACRONYMS AND THEIR DEFINITIONS

Table B.1 Acronyms and their definitions

Acronym	Name	Definition
<i>LT</i>	Lead Time	Amount of time allotted for the production of a part (therefore a planned time).
<i>FT</i>	Flow Time	Average time from the release at the beginning of the routing until it reaches an inventory point at the end of the routing (therefore an observed time).
<i>DLT</i>	Decoupled Lead Time	Longest cumulative coupled lead time chain in the product structure of a manufactured item. Limited and defined by the placement of decoupling points within a product structure.
<i>NFP</i>	Net Flow Position	Value calculated using the Net Flow Equation. Determines the overall inventory status of a particular item managed using a DDMRP buffer and whether a supply order will be recommended against the buffered position.
<i>PSL</i>	Production Service Level	Probability of the FT being less than or equal to the LT.
<i>PSLr</i>	Production Service Level (Rush)	Probability of the rush orders FT being less than or equal to the rush orders LT.
<i>PSLnr</i>	Production Service Level (Non-Rush)	Probability of the non-rush orders FT being less than or equal to the non-rush orders LT.
<i>CSL</i>	Customer Service Level	Percentage of customer orders filled on time.
<i>CSLr</i>	Customer Service Level (Rush)	Percentage of customer rush orders filled on time.
<i>CSLnr</i>	Customer Service Level (Non-Rush)	Percentage of customer non-rush orders filled on time.
<i>I</i>	Inventory	Average inventory level for a scenario.

APPENDIX C LEAD TIME CALIBRATION USING SURROGATE MODEL

We developed a surrogate model as a representation of the simulated production system (see section 5.3.4). We aim to find range estimates for LT values at different loading rates. According to Hopp and Spearman (2011), the Production Service Level (PSL) is the probability of the Flow Time (FT) being smaller or equal to the Lead Time (LT), as expressed by Eq. C.1.

$$PSL = P(FT \leq LT) \quad (C.1)$$

For our experiments, we need to find the 3 LTs (Low, Medium, and High) corresponding to a PSL of 60%, 80%, and 90% at different loading rates. Since the main objective of the surrogate model is to calibrate the LT values at different loading rates, we created 20 scenarios varying from 80% to 99%, with 1% increases. For all scenarios, we use the initial LT equal to twice the sum of bottleneck processing times of all ADUs.

For each scenario, we considered 30 replications with a 550-day horizon. All observed FTs were collected after a 50-day warmup period. This leads for each scenario to a set of around 105.000 FT observations, depending on the scenario. We used the distribution function of the observed FT, considering $F^{-1}(0,6)$, $F^{-1}(0,8)$ and $F^{-1}(0,9)$ for Low, Medium, and High lead times respectively. Figure C.1 is an example of the output of one replication for the 80% loading rate scenario. The blue line represents the sum of the $FT < X$ at each point divided by the total number of observations. The orange line indicates around the 60% cumulative mark of FT observations. The yellow line indicates around 80% of FT observations, and the red line indicates around 90% of FT observations.

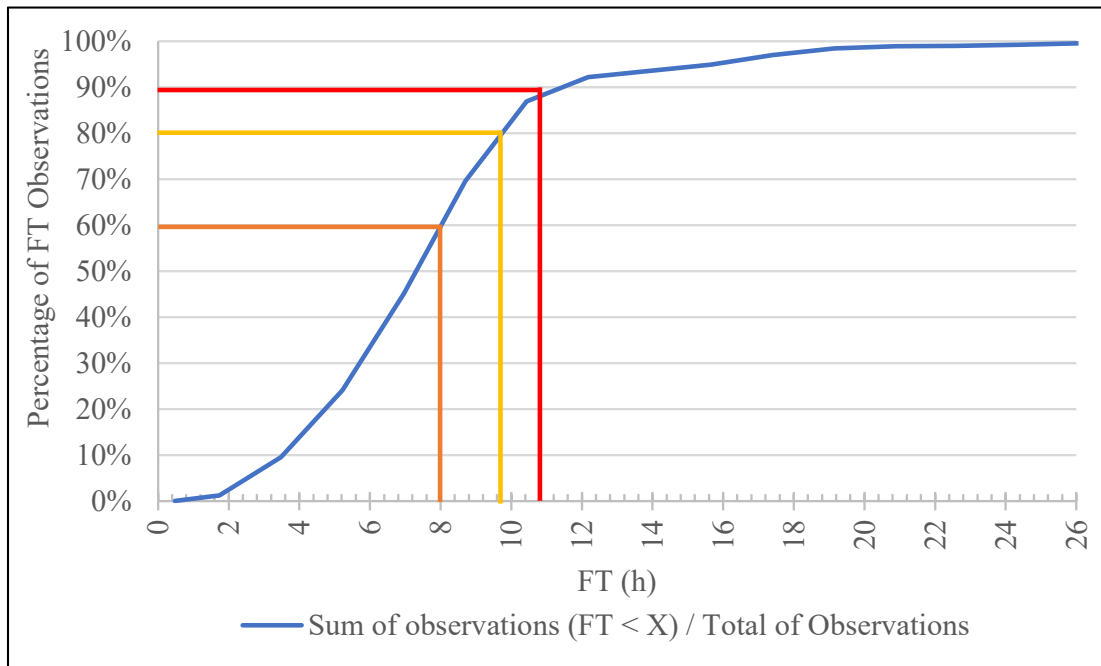


Figure C.1 Distribution of FT Observations with an 80% LR

Looking at the distribution, we calibrate the LT by finding estimates based on several observations of the FT. Analyzing the graph, for the 80% loading rate scenario, 60% of the FT observations are smaller than 8h, indicating that this value is a good estimate for the LT. Analyzing the entire distribution, we can verify that the value 8.12h is the closest value that is greater than 60% of the observed FTs in this scenario. This value was subsequently selected for use in the 80% loading rate and Low LT scenarios, as it provides a 60% service level with an 80% loading rate. The values were later validated in the experiments that produced results similar to our estimates.

APPENDIX D MACHINE LEARNING ALGORITHM SELECTION FOR PREDICTIVE ANALYTICS

This appendix provides supplementary materials and detailed methodology, inspired by Djedidi (2025) for the machine learning algorithm selection for the predictive analytics phase of the research.

D.1 Data generation and analysis

By running a 25000-day instance of the case study model, a large dataset was generated. The data contained information about the customer orders, the DDMRP system, and the state of the flowshop the moment a replenishment order is generated. Ten features were collected (see Table D.1) and used in the evaluation of the predictive algorithms: Order Quantity, Net Flow Equation (NFE), On Hand, Work in Process (WIP), and Machine Queues (M1 to M6).

Table D.1 Detailed Feature Description

Feature	Description
Order Quantity	Size of the recommended replenishment order
Net Flow Equation (NFE)	NFE calculation for the recommended order product
On Hand	The quantity On Hand for the recommended product
Work in Process (WIP)	The amount of work in process for the product of the recommended order
Queue in Machines (Mk)	Number of products in queue at Machine k, k=1..6

The target variable is heavily right-skewed, with most orders being completed in under 5 hours, but a long tail of orders takes longer, as shown in the distribution in Figure D.1 (left). A correlation analysis was performed to identify potential multicollinearity between features. The heatmap in Figure D.1 (right) shows the correlation coefficients between all variables. Notably, there are strong positive correlations between OnHand and NFE (0.79) and between Order Quantity and NFE (0.84).

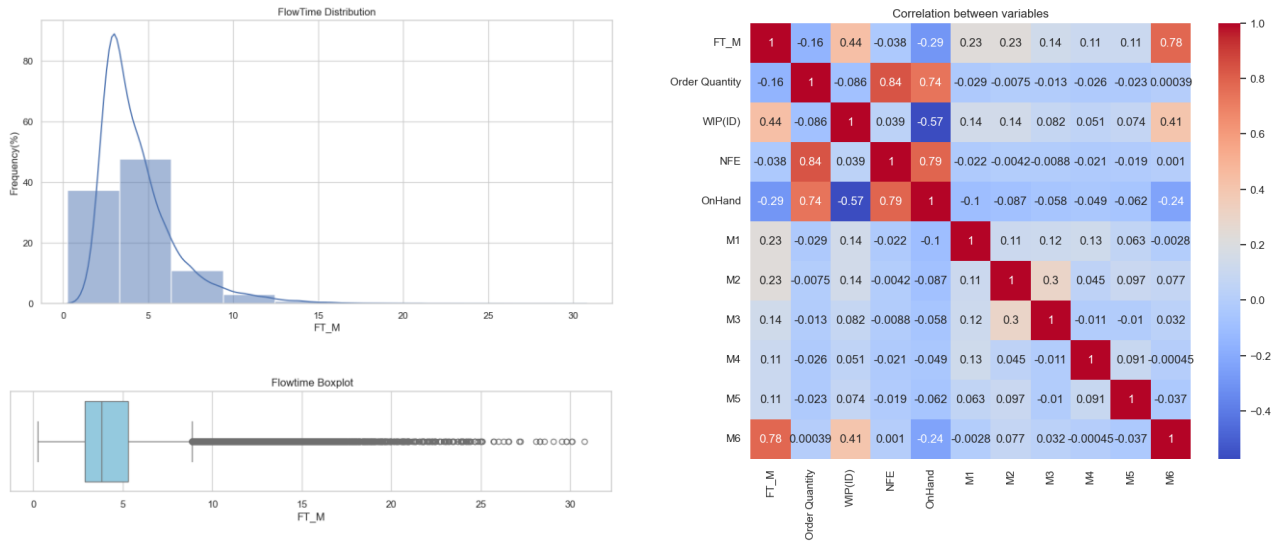


Figure D.1 Target variable distribution and boxplot (left) and features and target correlation matrix (right)

D.2 Predictive Models Training and Performance Comparison

Part of the data was used to train five supervised learning machine learning algorithms for predictive analytics: Decision Tree, Random Forest, XGBoost, Support Vector Regressor (SVR), and Ordinary Least Squares (OLS). We split the data, using 80% of the data for learning and 20% of the data for testing. For linear models, particular attention was paid to potential multicollinearity, which could affect the interpretability of feature importances. The performance of each model on the training and test datasets was measured using Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and R-squared (R^2), as shown in Table D.2. The XGBoost model demonstrated the best overall performance on the test set, achieving the highest R^2 value of 0.8162. Figure D.2 provides a visual comparison of the model's performance metrics on the train and test sets, illustrating the superior performance of the XGBoost algorithm.

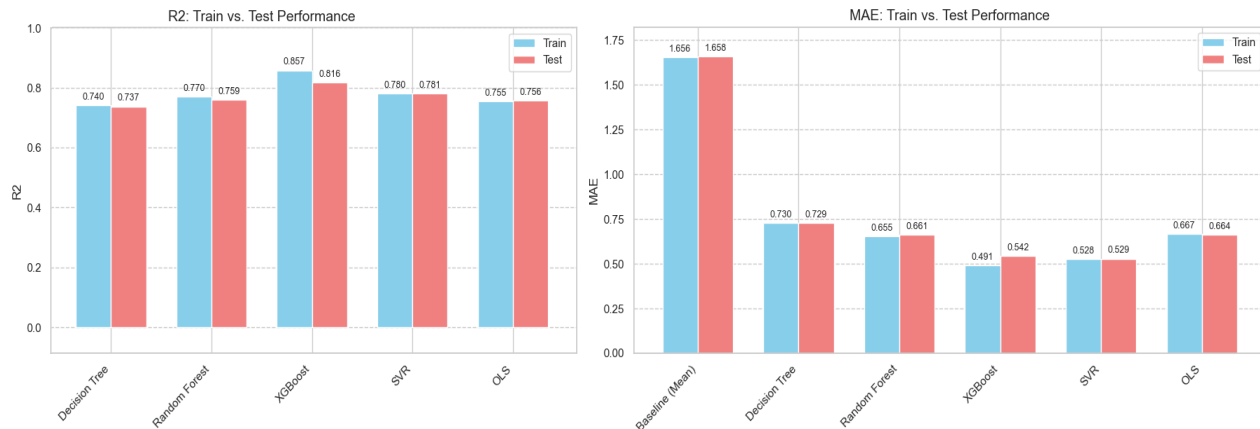


Figure D.2 Models R² Comparison on train vs. test data (left) and MAE (right)

Table D.2 Machine Learning Algorithms Performance Comparison

Model	RMSE(Train)	MAE(Train)	R ² (Train)	RMSE (Test)	MAE(Test)	R ² (Test)
Baseline (Mean)	2.2821	1.6564	NA	2.2801	1.6578	NA
Decision Tree	1.1634	0.7295	0.7401	1.1695	0.7286	0.7369
Random Forest	1.0954	0.6551	0.7696	1.1184	0.6614	0.7594
XGBoost	0.8619	0.491	0.8574	0.9776	0.5424	0.8162
SVR	1.0696	0.5279	0.7803	1.0677	0.5289	0.7807
OLS	1.1297	0.6666	0.755	1.1251	0.6642	0.7565

D.3 XGBoost Residual Analysis

As the best performing predictive algorithm, a residual analysis was conducted to validate the XGBoost model's predictive behavior (see Figure D.3). The results confirm that the model is unbiased, with the mean of residuals at 0.00. However, the analysis also reveals heteroscedasticity (i.e., the error variance increases for higher predicted values) and the tendency for underprediction, as evidenced by the right-skewed tail in the Q-Q plot.

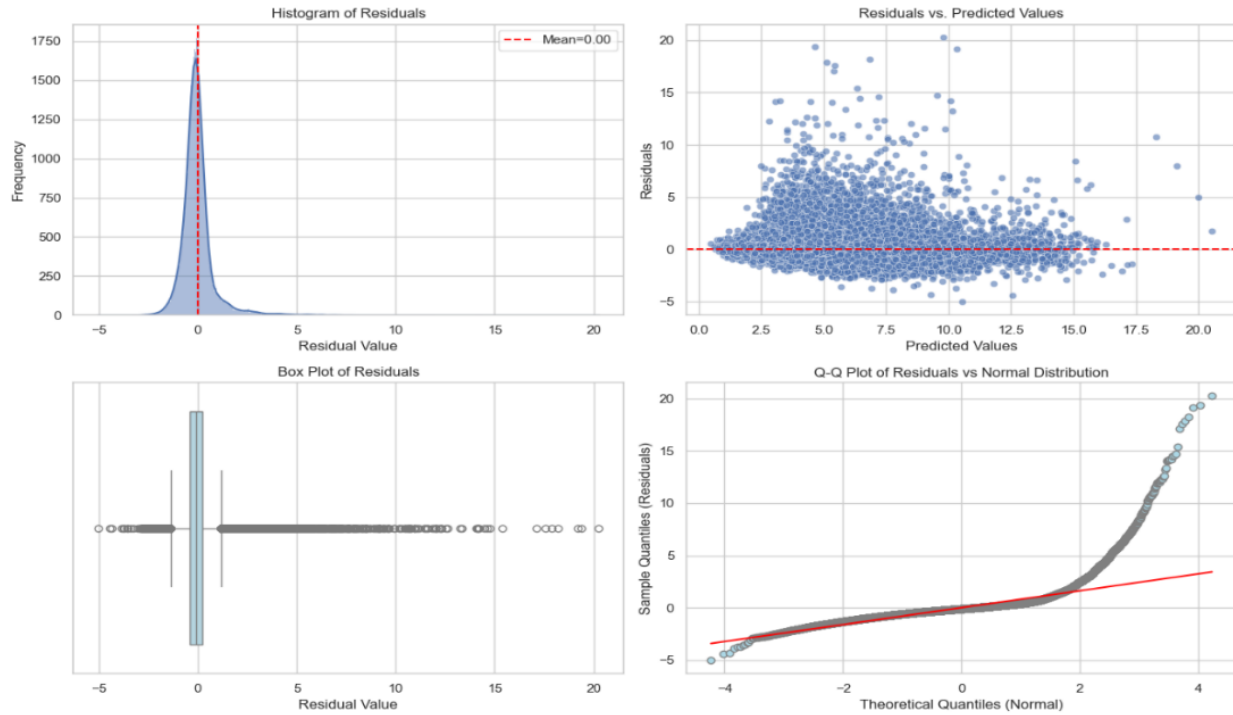


Figure D.3 Residual analysis plots for the XGBoost model

While these observations highlight limitations, particularly when predicting outliers' high flow time orders, the overall predictive power of the algorithm remains robust. The high R^2 value of 0.8162 and low overall error metric (RMSE = 0.9776, MAE = 0.5424) for the test set indicate that the model's performance is still suitable for the proposed application within the order generation system.

D.4 XGBoost Feature Importance (SHAP)

The feature importances for the final XGBoost model were calculated using the SHAP (SHapley Additive exPlanations) method. The analysis shows that M6 (queue at machine 6) is the most dominant predictor, with more than double the impact of the second most important feature, Order Quantity. The SHAP summary plot (see Figure D.4) illustrates how high (red) or low (blue) values of a feature impact the model's output. For M6, high values strongly push the prediction higher.

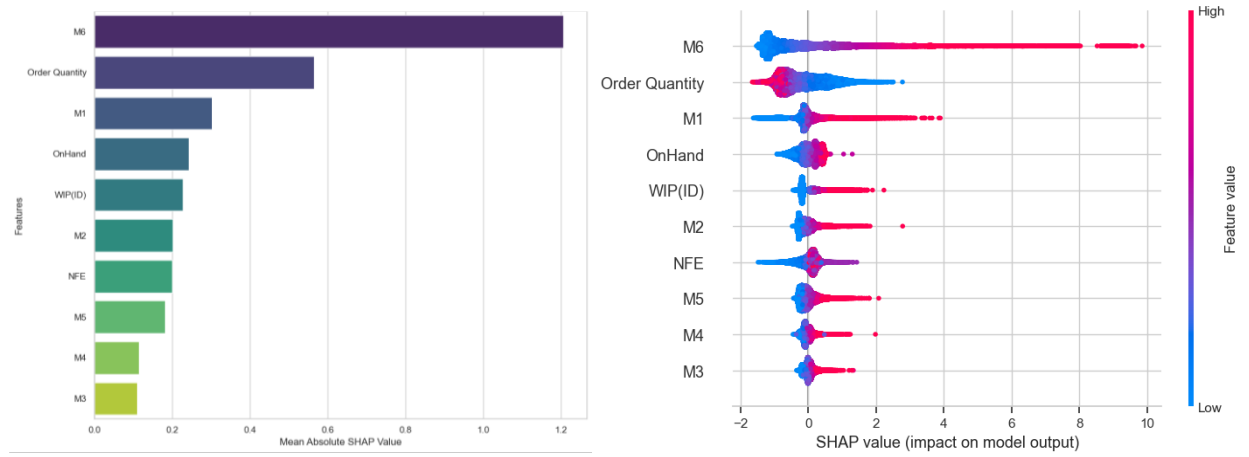


Figure D.4 SHAP feature importance chart (left) and summary plot (right)