

**Titre:** Une approche de décomposition pour le dimensionnement optimal des aires de stockage dans une ligne de production à fiabilité limitée  
**Title:**

**Auteur:** Thibault Gosset  
**Author:**

**Date:** 2011

**Type:** Mémoire ou thèse / Dissertation or Thesis

**Référence:** Gosset, T. (2011). Une approche de décomposition pour le dimensionnement optimal des aires de stockage dans une ligne de production à fiabilité limitée  
**Citation:** [Mémoire de maîtrise, École Polytechnique de Montréal]. PolyPublie.  
<https://publications.polymtl.ca/646/>

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

**URL de PolyPublie:**  
PolyPublie URL: <https://publications.polymtl.ca/646/>

**Directeurs de recherche:** Robert Pellerin, & Roland P. Malhamé  
**Advisors:**

**Programme:** génie électrique  
**Program:**

UNIVERSITÉ DE MONTRÉAL

UNE APPROCHE DE DÉCOMPOSITION POUR LE DIMENSIONNEMENT  
OPTIMAL DES AIRES DE STOCKAGE DANS UNE LIGNE DE PRODUCTION  
À FIABILITÉ LIMITÉE

THIBAUT GOSSET  
DÉPARTEMENT DE GÉNIE ÉLECTRIQUE  
ÉCOLE POLYTECHNIQUE DE MONTRÉAL

MÉMOIRE PRÉSENTÉ EN VUE DE L'OBTENTION DU DIPLÔME DE  
MAÎTRISE ÈS SCIENCES APPLIQUÉES  
(GÉNIE ÉLECTRIQUE)  
SEPTEMBRE 2011

UNIVERSITÉ DE MONTRÉAL

ÉCOLE POLYTECHNIQUE DE MONTRÉAL

Ce mémoire intitulé :

UNE APPROCHE DE DÉCOMPOSITION POUR LE DIMENSIONNEMENT  
OPTIMAL DES AIRES DE STOCKAGE DANS UNE LIGNE DE PRODUCTION  
À FIABILITÉ LIMITÉE

présenté par : M. GOSSET Thibault.

en vue de l'obtention du diplôme de : Maîtrise ès Sciences Appliquées

a été dûment accepté par le jury constitué de :

M. BAPTISTE Pierre, Ph.D., président.

M. MALHAMÉ Roland, Ph.D., membre et directeur de recherche.

M. PELLERIN Robert, Ph.D., membre et co-directeur de recherche.

M. OUALI Mohamed Salah, Doctorat, membre.

*À ma famille, mes proches et mes amis.*

# Remerciements

Je tiens tout d'abord à remercier Monsieur Roland Malhamé, mon directeur de recherche, qui s'est montré disponible tout au long de la réalisation de ce mémoire. Je le remercie pour l'inspiration, l'aide technique et financière et le temps qu'il a bien voulu me consacrer et sans qui ce mémoire n'aurait jamais vu le jour. Le sujet sur lequel nous avons travaillé m'a apporté une grande satisfaction intellectuelle.

Mes remerciements s'adressent également à Monsieur Robert Pellerin, mon co-directeur de recherche, pour le soutien qu'il a su m'apporter tout au long de ces deux années, ainsi que pour son aide financière.

Je souhaite présenter mes remerciements aux membres du jury, Monsieur Pierre Baptiste et Monsieur Mohamed-Salah Ouali, qui ont accepté de participer à ma soutenance.

J'adresse mes remerciements à toutes les personnes qui travaillent au Gerad pour leur soutien et leur aide technique dont j'ai pu bénéficier.

Je tiens également à remercier mes collègues de laboratoire, Fatima Mhada et Sadegh Bolouki, avec qui j'ai passé de très bons moments, j'ai eu des discussions très intéressantes et j'ai joué au soccer comme jamais avant. Je remercie aussi les colocataires du 5320 avec qui j'ai partagé une partie de ces deux superbes années à Montréal, et plus particulièrement Vincent Descotes avec qui je me suis lié d'amitié.

Enfin, je voudrais remercier mes parents, mes frères et soeurs, mes amis et mes proches qui m'ont été d'un grand soutien tout au long de la réalisation de ce mémoire.

Merci à tous et à toutes.

# Résumé

Le sujet de cette étude est un système de production composé de  $m$  machines en série, que l'on nomme ligne de production. La fiabilité des machines est représentée par un état de bon fonctionnement pour lequel la machine produit à un rythme donné et par un état de mauvais fonctionnement pour lequel la production de cette machine est arrêtée. Le temps passé dans chacun des états est représenté par une loi exponentielle. Ainsi, chaque machine peut être modélisée par une chaîne de Markov en temps continu à deux états. Finalement, des aires de stockage sont introduites dans toute la ligne en prévention des pannes aléatoires des machines de production.

Définir la taille optimale de ces aires de stockage est délicat. La fiabilité du système n'étant pas parfaite, une décision de redimensionnement des aires de stockage peut avoir des conséquences sur la production globale du système difficilement prévisibles. On est tenté de réduire les inventaires pour minimiser le coût de stockage mais le risque de rupture de la production est alors augmenté en conséquence. La production globale peut être fortement diminuée et ne plus satisfaire la demande du client ainsi que l'approvisionnement du fournisseur.

L'objectif de cette recherche est de développer un modèle mathématique de cette ligne afin de déterminer quelle politique de gestion des aires de stockage permet d'assurer une production moyenne déterminée, à un coût minimal de stockage des biens en cours de production.

Une première solution a déjà été proposée par (21) et aboutit sur une méthode de décomposition et agrégation de cette ligne de production. En considérant une demande constante et des taux de production décroissants dans la ligne de production, (21) mettent en évidence une propagation unidirectionnelle, de l'amont vers l'aval, des phénomènes d'arrêt de production qui se manifestent quand les aires de stockage se vident. Ils proposent une modélisation mathématique de la ligne et minimisent le coût de stockage des inventaires par programmation dynamique. La possibilité d'une méthode duale de leurs travaux a longtemps été soupçonnée. Elle considérerait un

approvisionnement constant et des taux de production croissants dans la ligne. La production serait alors considérée en flux poussé et les phénomènes de blocage de la production se réaliseraient quand les aires de stockage se remplissent. Dans cette étude, on examine la possibilité d'exploiter une telle méthode duale en suivant les étapes établies dans leurs travaux.

Dans le chapitre trois, on considère une ligne de production constituée d'une seule machine et d'une seule aire de stockage. De cette manière, il est possible d'obtenir des équations analytiques décrivant l'évolution du stock et son espérance en régime permanent. On se sert par la suite de ces équations pour obtenir une approximation de la moyenne théorique des inventaires de stockage dans une ligne de production constituée de plusieurs machines en série.

Dans le cas dual, les phénomènes de blocage de la production ne peuvent alors être provoqués que par le remplissage des aires de stockage en aval de la machine considérée. On modélise ces phénomènes par une réduction du temps de travail de la machine considérée qui dépend unidirectionnellement du blocage des autres machines en aval de la ligne. Cela est possible grâce à une approximation des processus de remplissage des aires de stockage.

On utilise aussi le principe de moyennage pour représenter l'approvisionnement de chacune des aires de stockage par une constante. Sous ces hypothèses, nous sommes en mesure de modéliser par des équations analytiques la ligne de production par des ensembles de base constitués d'une seule aire de stockage et d'une seule machine.

Selon le taux de service que l'on impose à l'approvisionnement en entrée du système, on considère un modèle de base différent qui prend en compte la possibilité d'un surplus d'aire de stockage ou non.

Finalement, une fois le modèle théorique construit, on vérifie sa validité par simulation Monte Carlo. On observe que la précision du modèle dépend du nombre de machines. On écrit enfin les équations de Bellman qui permettent d'obtenir le coût minimal de stockage des inventaires de la ligne par programmation dynamique. Cela est rendu possible grâce à la structure unidirectionnelle des phénomènes de blocage

dans la ligne et donc, grâce à la formulation du problème.

On observe pour conclure que les résultats obtenus permettent de valider cette méthode duale et on discute des travaux futurs envisageables.



# Abstract

We consider a production line composed of  $m$  unreliable machines in a tandem configuration. Each machine is represented by a two state Markov chain in continuous time in which states it can respectively process materials up to a specified maximum rate or its production is stopped. Furthermore, buffers are introduced between each machine of the line in order to limit the impact of machines' breakdown on the productivity of the line.

Disruption of material throughout the system may become an economical issue. For instance, the customer demand could be satisfied inconsistently, leading to supplementary inventory costs, unprofitable sales and customer dissatisfaction. Then, work-in-process produce by and available to machines play a major role in this system, making it possible for machines to continue to produce materials when one or several of them are being repaired thus maintaining up to a point a continuous flow of production through the line. The level of reliability of work stations must be taken into account in order to calculate a more efficient production plan and better schedule delivery of materials.

The objective of this study is to size storage areas of the production line so that disruptions of machines do not affect a continuous flow of parts throughout the system and at a minimum cost of holding inventories in the line.

A solution has been proposed by (21). They have developed a decomposition and aggregation methodology for the approximate performance analysis of the line when upstream to downstream machine capacities are monotone decreasing. In this case, a constant average production rate is imposed via a *pull* mechanism. The main advantage of this scheme is that causality propagates unidirectionally (upstream to downstream), thus making it possible to use dynamic programming as an optimization tool for storage areas sizing and cost optimization of holding inventories in the line. A dual method has long been thought possible. Increasing production rates would represent a *push* mechanism policy and causality would propagate in this for-

mulation from downstream to upstream. In this study, we examine the possibility of using this methodology for the dual case.

We first present a model of a line composed of one machine and one storage area only. In this case, analytical equations that describe the inventory evolution can be obtained by considering infinitesimal increments of possible variations of the stock and the boundary limits. We then develop assumptions and approximations to use those equations for analyzing the complete transfer line.

Under ergodicity of processes of the line and the so-called averaging principle, we can represent the rate of extraction of parts from each buffer by a constant. Machine states are more complex in the line environment and a work station works properly under specific constraints. We model the production states of the machines by an aggregation of the different states in which for each machine model the possibility of having the production blocked when its associated buffer becomes full.

Those approximations and assumptions allow us to use the previously developed analytical equations for a single machine line and to compose the line by those basic sets in series. A mathematical model of the line is then possible for calculating the expectation cost of the line and its average production. A specific model is also developed at the beginning of the system to have a supply service without disruptions.

Finally, we validate the model against Monte Carlo simulations and analyze where the imprecision could come from. We introduce the dynamic programming formulation that is possible because of the unidirectional propagation of phenomena. A recursive method is then possible and considers sub-problems that are expanded after optimization to find the minimum cost for the line with a defined average production rate.

# Table des matières

Dédicace . . . . .	iii
Remerciements . . . . .	iv
Résumé . . . . .	v
Abstract . . . . .	viii
Table des matières . . . . .	x
Liste des tableaux . . . . .	xiii
Liste des figures . . . . .	xiv
Liste des annexes . . . . .	xvi
Liste des sigles et abréviations . . . . .	xvii
Chapitre 1 Introduction . . . . .	1
Chapitre 2 Revue de littérature . . . . .	4
2.1 Introduction . . . . .	4
2.2 Approche d'analyse de lignes de production . . . . .	4
2.3 Revue critique . . . . .	8
Chapitre 3 Modélisation de la ligne de production . . . . .	9
3.1 Introduction . . . . .	9
3.2 Le modèle de base pour une machine . . . . .	10
3.3 Équations probabilistes d'évolution du stock . . . . .	14
3.4 Modèle pour la ligne de production . . . . .	25
3.4.1 Hypothèses pour exploiter le modèle de la ligne . . . . .	28
3.5 Modélisation mathématiques de la fonction coût . . . . .	37
3.6 Conclusion . . . . .	45

Chapitre 4	Simulation de la ligne de production . . . . .	47
4.1	Introduction . . . . .	47
4.2	Réalisation de la simulation . . . . .	47
4.2.1	Générer l’horizon temporel de chaque machine . . . . .	48
4.2.2	Tracer l’évolution des aires de stockage . . . . .	49
4.3	Résultats des simulations . . . . .	51
4.4	Conclusion . . . . .	58
Chapitre 5	Optimisation du coût de stockage des encours de la ligne de produc- tion . . . . .	60
5.1	Introduction . . . . .	60
5.2	Formulation du problème d’optimisation . . . . .	61
5.2.1	Programmation du problème d’optimisation . . . . .	62
5.3	Résultats de l’optimisation . . . . .	64
5.4	Conclusion . . . . .	70
Chapitre 6	CONCLUSION . . . . .	71
Références	. . . . .	74
Annexes	. . . . .	77
A.1	Main script . . . . .	77
A.2	Fonctions utilisées . . . . .	83
B.1	Main script . . . . .	91
B.2	Fonctions utilisées . . . . .	94
C.1	Introduction . . . . .	97
C.2	Building Blocks : Background Material . . . . .	98
C.3	Further modeling assumptions . . . . .	101
C.4	The Decomposition Approach . . . . .	103
C.4.1	Two approximations for converting the transfer line into iso- lated buffer-machine pairs . . . . .	104
C.4.2	Accounting for blocking effects downstream of machine $M_i$ . .	105
C.4.3	Two dynamic programs for buffering optimization . . . . .	106
C.5	Monte Carlo Simulation . . . . .	109
C.6	Dynamic Programming Results . . . . .	111

C.7 Conclusion . . . . .	115
C.8 appendix . . . . .	116
Références . . . . .	118

# Liste des tableaux

TABLEAU 3.1	Probabilité des différents états de fonctionnement de la machine	34
TABLEAU 4.1	Résultats de simulation pour une seule machine . . . . .	54
TABLEAU 4.2	Résultats de simulation pour une ligne de 2 machines . . . . .	54
TABLEAU 4.3	Résultats de simulation pour une ligne de 3 machines . . . . .	55
TABLEAU 4.4	Résultats de simulation pour une ligne de 4 machines . . . . .	55
TABLEAU 4.5	Résultats de simulation pour une ligne de 5 machines . . . . .	56
TABLEAU 4.6	Résultats de simulation pour une ligne de 10 machines . . . . .	56
TABLEAU C.1	Comparison of Theoretical and Simulation Results . . . . .	110
TABLEAU C.2	Results for a 5 Machine Line . . . . .	110

# Liste des figures

FIGURE 3.1	Représentation d'une station de travail . . . . .	10
FIGURE 3.2	Une trajectoire possible de l'évolution du stock . . . . .	12
FIGURE 3.3	Changement d'état possible : 2 <sup>e</sup> cas . . . . .	16
FIGURE 3.4	Changement d'état possible : 1 <sup>er</sup> cas . . . . .	16
FIGURE 3.5	Représentation de la ligne de production avec $m$ machines . .	25
FIGURE 3.6	Évolution possible du processus $I_{i+1}(t)$ . . . . .	27
FIGURE 3.7	Évolution possible de l'approvisionnement $u_{i-1}(t)$ . . . . .	28
FIGURE 3.8	Représentation de la station de travail $i$ au sein de la ligne de production . . . . .	28
FIGURE 3.9	Changements d'états de la machine $M_{m-1}$ . . . . .	33
FIGURE 3.10	Changement d'états de la machine $\tilde{M}_{m-1}$ . . . . .	35
FIGURE 3.11	Représentation de la station de travail affaiblie $i$ . . . . .	36
FIGURE 3.12	Solution graphique du point fixe . . . . .	40
FIGURE 4.1	Évolution de $Y_i(w)$ sur la longueur de l'horizon . . . . .	53
FIGURE 4.2	Évolution de $E[\sum X_{i,t}]$ sur la longueur de l'horizon . . . . .	59
FIGURE 5.1	Disponibilité optimale des aires de stockage . . . . .	64
FIGURE 5.2	Aire de sotckage optimale . . . . .	64
FIGURE 5.3	Disponibilité optimale des aires de stockage . . . . .	65
FIGURE 5.4	Aires de stockage optimales . . . . .	65
FIGURE 5.5	Coût global de la ligne en fonction du taux de service . . . . .	66
FIGURE 5.6	Disponibilité optimale des aires de stockage . . . . .	67
FIGURE 5.7	Aire de sotckage optimale . . . . .	67
FIGURE 5.8	Disponibilité optimale des aires de stockage . . . . .	67
FIGURE 5.9	Aires de stockage optimales . . . . .	67
FIGURE 5.10	Disponibilité optimale des aires de stockage . . . . .	68
FIGURE 5.11	Aire de sotckage optimale . . . . .	68
FIGURE 5.12	Coût global de la ligne en fonction du taux de service . . . . .	69
FIGURE 5.13	Disponibilité optimale des aires de stockage . . . . .	69
FIGURE 5.14	Disponibilité optimale des aires de stockage . . . . .	69
FIGURE C.1	Model of The Work Station With Upstream Buffer . . . . .	99

FIGURE C.2	$m$ Machines Production Line . . . . .	102
FIGURE C.3	Model of The $i$ th Work Station of The Line . . . . .	103
FIGURE C.4	4 State Markov Chain Aggregation . . . . .	107
FIGURE C.5	Optimal Availability of Buffers . . . . .	112
FIGURE C.6	Optimal Buffer Sizes . . . . .	112
FIGURE C.7	Optimal Availability of Buffers . . . . .	113
FIGURE C.8	Optimal Buffer Sizes . . . . .	113
FIGURE C.9	Optimal Availability of Buffers . . . . .	114
FIGURE C.10	Optimal Buffer Sizes . . . . .	114
FIGURE C.11	Global Cost of The Production Line Against Supply Efficiency	114



# Liste des annexes

ANNEXE A	Programme de simulation . . . . .	77
ANNEXE B	Programme d'optimisation . . . . .	91
ANNEXE C	Article pour soumission à IJPE . . . . .	96

# Liste des sigles et abréviations

$d$	Taux d'approvisionnement et de demande de la ligne
$m$	Nombre de machines de la ligne de production
$M_i$	$i^{eme}$ machine de la ligne de production
$k_i$	Cadence de production de la machine $i$
$T_{on}^i$	Temps aléatoire de bon fonctionnement de la machine $i$
$T_{off}^i$	Temps aléatoire de mauvais fonctionnement de la machine $i$
$p_i$	Paramètre de la loi exponentielle de la v.a. $T_{on}^i$
$r_i$	Paramètre de la loi exponentielle de la v.a. $T_{off}^i$
$\tilde{M}_i$	$i^{eme}$ machine fictive de la ligne de production
$\tilde{T}_{on}^i$	Temps aléatoire de bon fonctionnement de la machine fictive $i$
$\tilde{T}_{off}^i$	Temps aléatoire de mauvais fonctionnement de la machine fictive $i$
$\tilde{p}_i$	Paramètre de la loi exponentielle de la v.a. $\tilde{T}_{on}^i$
$\tilde{r}_i$	Paramètre de la loi exponentielle de la v.a. $\tilde{T}_{off}^i$
$X_i$	$i^{eme}$ Aire de stockage de la ligne de production
$z_i$	Valeur maximale de l'aire de stockage $i$
$\mathbf{P}_{z_i}$	Probabilité que l'aire de stockage $X_i$ soit en $z_i$
$\mathbf{P}_{o_i}$	Probabilité que l'aire de stockage $X_i$ soit en 0
$b_i$	Disponibilité de l'aire de stockage
$E[X_i]$	Espérance mathématiques de l'aire de stockage $X_i$
$T$	Horizon temporel de la simulation
$v.a.$	Variable aléatoire
$i.i.d.$	Indépendantes et identiquement distribuées
$\circ(h)$	Représente l'ensemble des fonctions négligeables devant $h$

# Chapitre 1

## Introduction

La modélisation des systèmes manufacturiers permet une meilleure gestion de la fabrication et ainsi d'optimiser l'utilisation des ressources de l'entreprise. Cependant, la fiabilité des stations de production pose un vrai problème de planification. La fiabilité étant par nature aléatoire, il est difficile de prévoir l'évolution exacte de la production. On s'intéresse alors à représenter le comportement d'une ligne de production composée de  $m$  machines en série et dont leur fiabilité est modélisée par des lois exponentielles. Des aires de stockage intermédiaires sont introduites à titre préventif vis-à-vis de l'interruption de production de ces machines.

L'introduction d'aires de stockage est, dans la pratique, une réponse courante apportée à ces possibles ruptures de production des stations de travail. Quand une ou plusieurs d'entre elles tombe(nt) en panne, les aires de stockage permettent aux autres machines de continuer à produire ou à s'alimenter en matière. De cette manière, un flux plus continu de production est réalisé par le système. On observe aussi des phénomènes d'ajustement des cadences de production et de blocage de la production des machines quand les aires de stockage se remplissent ou se vident. Ces phénomènes sont pour la plupart inter-dépendants et peuvent se propager dans toutes la ligne.

L'utilisation d'aires de stockage augmente cependant les inventaires de pièces qui n'ont pas terminé leur cycle de production et représente donc un coût supplémentaire de mobilisation des ressources de l'entreprise. On appelle ces inventaires des en-cours de production. Les réduire au minimum aurait pour inconvénient de diminuer radicalement l'efficacité de la ligne de production. On cherche donc un équilibre ou plus exactement une politique de gestion optimale de ces ressources pour laquelle l'efficacité du système est imposée et la minimisation du coût de stockage des en-cours de production est obtenue.

Un approvisionnement constant est imposé en entrée du système et la production est représentée par un flux continu. Ce genre de modèle peut être appliqué par exemple dans l'industrie pétrolière, pharmaceutique ou encore dans la production de masse de composants électriques.

(21) ont proposé une méthode d'optimisation approximative dans le cas où la production satisfait une demande constante en fin de ligne et non pas un approvisionnement constant. La ligne est alors opérée en flux tiré. Ils développent un modèle sous l'hypothèse que les machines ont des cadences de production décroissantes et, de cette manière, les phénomènes de blocage de la production des machines sont propagés unidirectionnellement, de l'amont vers l'aval. Grâce à des hypothèses additionnelles et certaines approximations, ils construisent la ligne de production à partir d'ensembles de base, composés d'une seule machine et d'une seule aire de stockage, dont les équations analytiques ont été développées par (15) et (2). Ils parviennent au final à minimiser par programmation dynamique le coût de stockage des en-cours de production.

Dans ce mémoire, on envisage l'extension de la méthode en vue de l'étude d'un cas dual pour lequel la ligne de production serait soumise à un approvisionnement constant, et sous l'hypothèse de machines ordonnées cette fois-ci selon des capacités de production croissantes. On utilise alors les idées de cette méthode pour résoudre le problème suivant : déterminer la taille optimale des aires de stockage qui assure une production moyenne stable dans le temps et à un coût minimal.

La structure de ce mémoire est la suivante. On présente dans le chapitre 2 la revue de littérature ainsi que l'environnement dans lequel se situe notre problème. Dans le chapitre 3, on développe le modèle de base dual. Celui-ci est composé d'une seule machine et d'une seule aire de stockage. Il considère un approvisionnement constant en entrée. Cela permet d'obtenir des équations analytiques. On détermine ensuite les hypothèses qui permettent de décomposer la ligne de production en ensembles de base pour lesquels les phénomènes de blocage et de pénurie de matière sont modélisés. Une fois le modèle mathématique approximatif établi. Dans le chapitre 4, on vérifie par simulation Monte Carlo la validité du modèle obtenu. Dans le chapitre 5, on détermine le coût minimal de stockage des en-cours de production par programmation

dynamique, en utilisant les équations de Bellman. Finalement dans le chapitre 6, on présente la conclusion et on discute des possibles recherches futures.

# Chapitre 2

## Revue de littérature

### 2.1 Introduction

Ce chapitre a pour objectif de faire état des travaux similaires dans ce domaine et de situer notre recherche en fonction de ce qui a déjà été publié. On présente dans un premier temps le sujet puis les méthodes d'analyse des différentes lignes de production.

L'étude de la fiabilité d'une ligne de production attire l'attention de nombreux chercheurs. Ce sujet a pour objectif de mieux comprendre les possibles interruptions de la production d'une ligne de fabrication afin de dimensionner les ateliers de manière optimale. Les aires de stockage ont une forte influence sur l'efficacité d'une ligne de production et c'est pourquoi de nombreuses publications ont été rédigées plus particulièrement sur ce sujet. Celles-ci portent aussi bien sur le dimensionnement d'une ligne de production que sur la politique à adopter pour piloter la ligne ou encore pour mesurer sa performance. On réalise dans ce chapitre une description des différentes machines qui ont été étudiées dans la littérature puis les différentes méthodes ayant été élaborées pour étudier des systèmes de production plus complexes.

### 2.2 Approche d'analyse de lignes de production

L'analyse exacte de lignes de production non fiables étant exclue, on utilise actuellement deux classes de méthodes pour modéliser approximativement ces lignes : la *méthode de décomposition* et la *méthode d'agrégation*. Celles-ci s'appuient sur l'utilisation d'un modèle de base pour reconstruire par étapes la ligne de production. Ces modèles de base qui représentent une unité de travail sont multiples et peuvent être regroupés selon leurs spécificités. On fait déjà la distinction entre les modèles pour lesquels les pannes sont *dépendantes d'opération*, c'est-à-dire dépendantes du

temps pendant lequel la machine produit correctement, et les modèles pour lesquels les pannes sont *dépendantes du temps*, c'est-à-dire qu'elles ne dépendent pas de l'état de fonctionnement de la machine mais simplement du temps. (5) répertoriaient déjà les différents modèles existants. On distinguera aussi les modèles selon la loi de distribution des temps de fonctionnement de la machine, selon la construction adoptée pour représenter une unité de travail ou encore, par le débit discret ou continu de matière imposé à la station de travail. Cette dernière caractéristique constitue de fait l'un des principaux critères de différenciation entre les modèles.

**En flux discret,** le temps est aussi discrétisé et la production se réalise unité par unité. Cette modélisation est particulièrement utilisée pour des ateliers de production pour lesquels chaque unité produite requiert un temps relativement important de transformation et un équipement spécialisé, comme dans l'industrie aéronautique. Les temps de fonctionnement de la machine sont généralement dépendants des opérations et déterminés par exemple par la probabilité que la machine produise un certain nombre de pièces durant son état de bon fonctionnement. Les auteurs (24), (25), (1) et (5) ont donné des solutions analytiques au problème et des outils d'analyse de performance des systèmes de production. On trouve aussi des applications industrielles telles que proposées par (3). Ces modèles en temps discret ont aussi été globalement étudiés par (4), (11), (14), (26), (10), ou encore (28).

**En flux continu,** les pannes sont principalement dépendantes du temps. On justifie cette modélisation en pratique lorsque la quantité de pièces produite par unité de temps est suffisamment importante, comme par exemple dans l'industrie chimique et la production de médicaments. En flux continu, pour des raisons de simplicité d'analyse, on note que les chercheurs ont une préférence pour une représentation des temps de fonctionnement par des lois exponentielles. Cela permet de représenter le comportement de la machine par une chaîne de Markov en temps continu. Les propriétés de non-vieillesse de la loi exponentielle sont toutefois discutables. (2) ont introduit un modèle composé d'une station et d'une aire de stockage pour laquelle l'inventaire peut être négatif et représente des retards de livraison (backlogs). Ils montrent que la politique optimale est de toujours produire au maximum lorsque la machine est en

état opérationnel. (15) montre que cette même politique est optimale dans le cas où les inventaires négatifs ne sont pas autorisés. Le modèle de deux machines en tandem a aussi été étudié dans ce cas par (17), (13), (11) et (30).

Les différents modèles de représentation d'une station de travail ont permis de résoudre bon nombre de problèmes industriels, mais dans le cas d'une ligne de production composée de plusieurs machines, de nouvelles problématiques se posent alors. La principale complication de ces modèles est la propagation de l'impact de pannes, que rapporte très clairement (30). Les différents phénomènes de blocage ou de pénurie sont difficilement modélisables et quantifiables dans une ligne de production car ils sont fortement interdépendants. Les méthodes d'agrégation et de décomposition, décrites ci-après, ont donc très vite été développées comme des outils pour répondre aux questions qui se posent sur la performance des lignes de production.

(30) proposait dans son article une méthode de décomposition des lignes de production par des stations de travail avec des machines en tandem que (11) a aussi utilisée pour développer une méthode qui permette de calculer la performance du système considéré. Chaque sous-système de la ligne est représenté par deux machines en tandem dans le cas d'une production en flux discret et caractérisées par leurs propres fiabilités. Un algorithme par itération successive permet d'aboutir aux résultats désirés de manière précise mais aussi de manière complexe et coûteuse en calcul. Par la suite, (14), (12) ont reformulé la technique de calcul associée à cette méthode de décomposition pour l'améliorer.

Le travail de (17) permet de calculer de manière approximative l'efficacité d'une ligne de transfert par une méthode d'agrégation. Cette méthode considère que l'ensemble des machines de la ligne sauf la dernière peuvent être considérées comme équivalentes à une machine et former un ensemble de deux machines en tandem séparées par une aire de stockage intermédiaire. Cette méthode est efficace pour le cas d'une ligne de trois machines mais n'offre pas de résultats probants pour une ligne avec plus de machines.

(28) proposent une méthode d'agrégation des stations de travail en remplaçant deux machines en tandem séparées par une aire de stockage par une seule machine.



Cette méthode permet ainsi de représenter toute la ligne de production par une seule machine pour ne plus former qu'un seul ensemble de deux machines en tandem. Elle permet aussi de n'imposer aucune restriction sur les paramètres des machines. Cependant, même si elle s'implante facilement, elle a le défaut d'avoir une erreur d'estimation trop importante sur les paramètres recherchés tels que l'efficacité de la ligne ou la production moyenne.

Plus récemment, (7) se sont intéressés au problème d'identification des machines qui réduisent la production globale d'une ligne de fabrication. Ces machines, que l'on appelle "goulot d'étranglement" (bottleneck), permettent de déterminer les points faibles du système et ainsi de limiter les ruptures de production par une meilleure allocation des ressources. Ils introduisent une définition de ce concept en fonction des paramètres du système de production. (8) proposent une méthode de détection de ces points faibles. Celle-ci permet le calcul des temps de pénurie de matière et de blocage de la production des machines du système par une méthode d'agrégation.

(21) ont proposé une méthode de décomposition et d'agrégation qui permet de dimensionner de manière optimale la taille des aires de stockage d'une ligne de production constituée de stations de travail en série et pour lesquels les taux de production sont décroissants. Cette méthode utilise un modèle de base comportant une machine et une aire de stockage et quantifie par des approximations rationnelles sur les processus de blocage de la production des machines les temps de séjour du stock dans les différents niveaux possibles. Cette méthode est utilisée pour une production en flux continu qui doit répondre à une demande constante. Cette méthode permet d'atteindre l'état optimal par un algorithme de programmation dynamique, ce qui assure sa robustesse. Elle permet d'ailleurs d'obtenir des résultats précis du coût de stockage des encours de production.

Toujours dans la même optique, (6) proposent, pour une ligne de production dont les machines partagent un même taux maximal de production, une méthode de dimensionnement des aires de stockage qui garantit un taux de production global de la ligne. À partir d'équations analytiques dans le cas d'une ligne de deux machines, ils formulent des hypothèses qui permettent de construire la ligne de fabrication à partir de l'interconnexion de tels ensembles. Ils déterminent des règles de construction de

la ligne pour optimiser le dimensionnement des aires de stockage. (9) proposent une analyse numérique d’une ligne de production dans le cas où les temps de fonctionnement des machines ne sont pas exponentiels mais dictés par une analyse empirique.

Les travaux qui suivent se placent dans cet environnement et étudient le cas d’une ligne de production dont les machines ont des taux de production croissants. On cherche à déterminer l’allocation optimale des aires de stockage dans la ligne en fonction de l’efficacité du système. La méthode reprend les étapes développées par (21) dans leur méthode alternative de décomposition-agrégation.

## 2.3 Revue critique

Les différentes méthodes développées précédemment donnent de nombreux résultats intéressants. Elles permettent de déterminer l’efficacité du système, où sont les “goulots d’étranglement” ou de décrire l’évolution du système. Cependant, dans le cas où l’on considère un flux continu de matière dans une ligne de production, aucune méthode ne permet aujourd’hui de modéliser avec précision l’évolution des inventaires dans le cadre d’une politique de production à flux poussée.

L’objectif de cette étude est de déterminer une méthode d’optimisation des ressources (le stock global des produits en cours de production), tout en assurant un taux moyen de production dans le système qui est opéré en flux poussé. On s’inspirera de la méthode mise en place par (21) dans le cas d’une production opérée en flux tiré.

# Chapitre 3

## Modélisation de la ligne de production

### 3.1 Introduction

L'objectif de ce chapitre est de déterminer le modèle mathématique utilisé pour décrire le coût global associé au stockage des inventaires de la ligne de production. On commence par définir et étudier le comportement du modèle de base, composé d'une seule aire de stockage et d'une seule machine, puis on développe l'analyse mathématique qui permet d'obtenir la probabilité que le stock se trouve dans n'importe quelle intervalle possible. On détermine ensuite les hypothèses qui permettent de considérer ce modèle de base dans la ligne de production et de prendre en compte les phénomènes de blocage et de pénurie qui apparaissent dans la ligne et qui se propagent. Enfin, on définit le modèle mathématiques qui représente le coût de stockage des en-cours de production de la ligne et que l'on souhaite minimiser.

La ligne de production est constituée de machines en série qui ne sont pas fiables. Si aucune aire de stockage n'était introduite dans la ligne de production, l'efficacité de celle-ci serait alors minimale et égale au produit de la fiabilité de chacune d'entre elles. En effet, la ligne serait alors productive uniquement quand toutes les machines (indépendantes entre elles) qui la composent seraient en état de bon fonctionnement. Ce résultat peut être très faible dès que l'on considère plusieurs machines ; c'est pourquoi on introduit des aires de stockage entre les machine. L'arrêt imprévisible d'une machine, au moins pour un temps, n'empêche plus les autres de continuer à fonctionner et un flux plus lissé de production est assuré dans la ligne. Alors que cette solution a pour avantage d'améliorer significativement l'efficacité de la ligne de production, elle a pour inconvénient d'augmenter les en-cours de production et donc le coût de mobilisation des ressources. De plus, des phénomènes de blocage

et de pénurie apparaissent dans la ligne quand les aires de stockage se vident ou se remplissent. Ces phénomènes sont difficilement modélisables car ils se propagent dans la ligne. Si le remplissage d'une aire de stockage bloque la production de la machine qui la précède, alors l'aire de stockage suivante se remplit plus rapidement et peut ainsi propager le phénomène dans la ligne. Au final, si l'on considère deux machines en tandem, la première voudrait que l'aire de stockage qui la suit soit plutôt vide pour pouvoir stocker sa production, alors que la seconde préférerait que cette même aire de stockage soit plutôt remplie pour ne pas craindre de pénurie.

Les étapes de la modélisation suivent le schéma suivant. On développe dans un premier temps un modèle pour une ligne composée d'une seule machine et d'une seule aire de stockage et pour laquelle on impose un taux d'approvisionnement constant. Ce modèle nous permet d'obtenir des équations analytiques d'évolution du stock. On formule alors des principes d'approximation pour décrire les phénomènes de pénurie et de blocage dans la ligne de production pour construire cette dernière à partir des ensembles de bases (une machine et une aire de stockage) pour lesquels on a des équations exactes.

## 3.2 Le modèle de base pour une machine

Une station de travail se compose d'une machine,  $M$ , qui alimente une aire de stockage infinie, et précédée d'une aire de stockage,  $X$ . On alimente l'aire de stockage finie à un rythme constant  $d$  et la machine s'alimente à son tour dans ce stock au rythme  $u(t)$ , tel que représenté dans la figure 3.1.

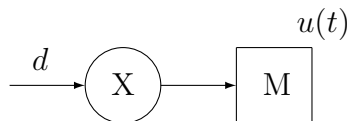


FIGURE 3.1 Représentation d'une station de travail

Le rythme de production  $u(t)$  dépend de l'état de fonctionnement de la machine ainsi que de l'état de l'aire de stockage. On considère dans un premier temps que cette

aire de stockage est finie, puis on verra pourquoi on peut aussi la considérer infinie positive.

La machine est non fiable et alterne entre deux états, marche, *ON*, et arrêt, *OFF*, respectivement de bon et mauvais fonctionnement. Dans l'état *ON*, la machine consomme de la matière au rythme  $k$  qui est supérieur au taux  $d$  contribuant ainsi à réduire le stock amont dans lequel elle s'alimente, et passe après un temps aléatoire,  $T_{on}$ , dans l'état *OFF*. Ce temps aléatoire est distribué suivant une loi exponentielle de paramètre  $p$ . Cette machine passe alors un autre temps aléatoire,  $T_{off}$ , dans l'état *OFF* durant lequel elle ne produit pas. Ce temps aléatoire est lui représenté par une loi exponentielle de paramètre  $r$ . On représente donc par le processus stochastique binaire 0/1  $\{M(t), t \in \mathbb{R}^+\}$  l'état de fonctionnement de la machine, qui est aussi un processus de Markov en temps continu. La cadence de production de la machine prend donc les valeurs suivantes en fonction de l'état de  $M(t)$  :

$$\begin{cases} u(t) = k > d & \text{si } M(t) = 1 \\ u(t) = 0 & \text{si } M(t) = 0 \end{cases} \quad (3.1)$$

L'aire de stockage,  $X$ , est finie de taille  $z$  et l'état du stock est représenté par le processus stochastique  $\{X(t), t \in \mathbb{R}^+\}$ , où  $x(t) \in [0, z]$ .  $x(t)$  représente une réalisation possible de la variable aléatoire  $X(t)$ , ou encore l'état actuel du stock. Avoir une aire de stockage finie impose une certaine dynamique au système. L'aire de stockage ne peut se vider que lorsque la machine est en état de fonctionnement (car sa cadence de production est plus importante que le taux d'approvisionnement). Dans ce cas, la machine ajuste alors sa cadence de production sur l'approvisionnement,  $d$ , car les excursions négatives (retard de livraison ou backlog) ne sont pas autorisées. On parle de *pénurie* car la machine ne produit pas à son taux maximal. Inversement, l'aire de stockage ne peut se remplir que lorsque la machine est en panne. Dans ce cas, c'est l'approvisionnement  $d$  qui n'est plus assuré et il y a rupture de l'approvisionnement dans le système. On parle de *blocage* de l'approvisionnement.

Finalement, la dynamique du stock répond à l'équation différentielle suivante :

$$\frac{\partial x(t)}{\partial t} = \begin{cases} d - u(t) & \text{si } x(t) \neq 0 \text{ ou } z \\ 0 & \text{si } x(t) = 0 \text{ ou } z \end{cases} \quad (3.2)$$

L'aire de stockage augmente et diminue selon l'état de fonctionnement de la machine et joue donc bien son rôle d'amortisseur. L'approvisionnement de l'aire de stockage peut se poursuivre lorsque la machine est en panne.

La figure 3.2 montre une trajectoire possible de l'aire de stockage et explique graphiquement les phénomènes de blocage et de pénurie.

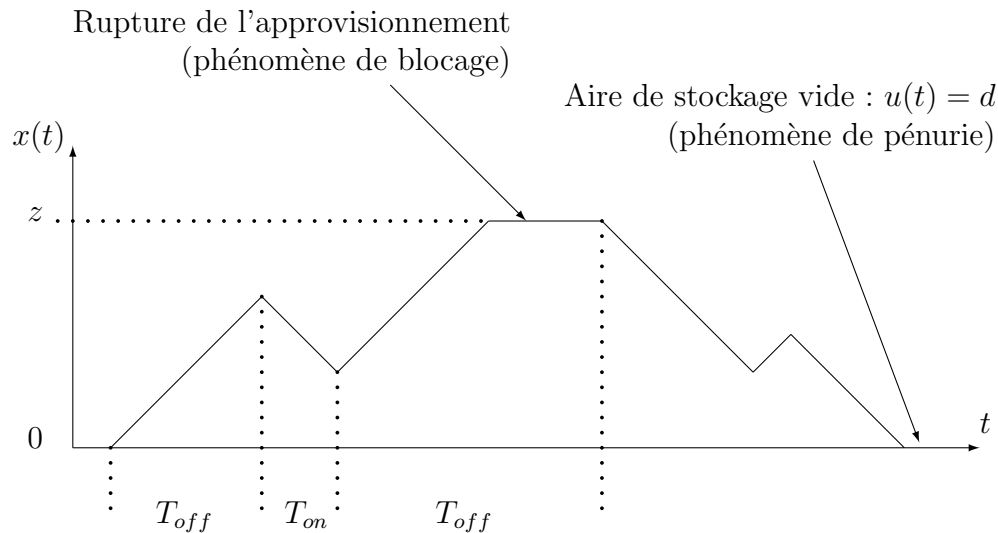


FIGURE 3.2 Une trajectoire possible de l'évolution du stock

Au départ, le stock est vide et la machine est en état de bon fonctionnement. Sa cadence de production,  $u(t)$ , est donc ajustée au taux d'approvisionnement  $d$ . Puis la machine tombe en panne, l'aire de stockage se remplit alors au taux  $d$ . La machine est à nouveau en état de fonctionnement, l'aire de stockage étant non vide, sa cadence est maximale et égale à  $k$ , l'aire de stockage se vide au taux  $k - d$ . La machine tombe à nouveau en panne, le stock augmente au taux  $d$ . L'aire de stockage est remplie, l'approvisionnement est arrêté. Lorsque la machine est à nouveau en état de marche,

l'approvisionnement est à nouveau possible et le stock se vide au taux  $k - d$ , jusqu'à se vider entièrement au bout d'un certain moment où l'on se retrouve au point de départ.

L'évolution du stock suit une dynamique très précise. On peut ainsi calculer la probabilité que celui-ci se trouve dans n'importe quel intervalle de l'aire de stockage,  $[0, z]$ , en considérant les variations infinitésimales possibles du stock et les conditions aux frontières. Les calculs sont détaillés dans la section qui suit.

$$\mathbf{P}_0 = \frac{r}{r+p} \frac{1-\mu}{1-\mu e^{-\alpha z}} \quad (3.3a)$$

$$\mathbf{P}_z = \frac{p}{r+p} \frac{(1-\mu)e^{-\alpha z}}{1-\mu e^{-\alpha z}} \quad (3.3b)$$

$$f_X(x) = \frac{p}{r+p} \frac{k}{k-d} \frac{\alpha e^{-\alpha x}}{1-\mu e^{-\alpha z}} \quad (3.3c)$$

avec,  $\mu = \frac{d}{k-d} \frac{p}{r}$  et  $\alpha = \frac{r}{d} - \frac{p}{k-d}$

La densité de probabilité,  $f_X(x)$ , ressemble à une loi exponentielle à un facteur multiplicatif près. Les probabilités d'être en 0 et  $z$  représentent les temps moyens passés en ces points.

Ces équations ne se trouvent pas telles quelles dans la littérature, mais des équations similaires ont cependant été développées par (15) dans le cas où la dynamique du stock (équation 3.2) est inversée. La machine alimente alors une aire de stockage qui est soumise à une demande constante. C'est le modèle utilisé par (21). On appellera cet ensemble *le modèle de Hu inversé*.

Ce modèle de base a été développé sous trois hypothèses principales :

1. L'approvisionnement en entrée est constant.
2. L'état de fonctionnement de la machine est décrit par 2 états de bon et mauvais fonctionnement dépendant seulement du temps.
3. La production de la machine n'est jamais bloquée autrement que par son propre mauvais fonctionnement.

Dans le cas d'une ligne de production avec plusieurs machines, ces trois hypothèses principales sont mises en défaut : l'approvisionnement de chaque modèle de base n'est plus constant, l'état de fonctionnement des machines n'est plus représenté par deux états seulement et la production peut être bloquée par l'état de remplissage des aires de stockage.

### 3.3 Équations probabilistes d'évolution du stock

La méthode de développement des équations de comportement du stock s'appuie sur le calcul des variations infinitésimales possibles de l'évolution de l'inventaire. Cette méthode est un cas particulier d'une méthode plus générale développée par (18).

Soit donc une station  $S$  composée d'une machine  $M$  et d'une aire de stockage  $X$ . Soit  $\{M(t), t \in \mathbb{R}^+\}$  le processus Markovien binaire *OFF/ON* ou 0/1 qui représente l'état de la machine. Soit  $\{X(t), t \in \mathbb{R}^+\}$ , le processus stochastique qui représente l'état du stock et  $x(t) \in [0, z]$  ses réalisations possibles. La dynamique d'évolution de l'état du stock répond à l'équation différentielle suivante :

$$\frac{\partial x(t)}{\partial t} = \begin{cases} d - u(t) & \text{si } x(t) \neq 0 \text{ ou } z \\ 0 & \text{si } x(t) = 0 \text{ ou } z \end{cases} \quad (3.4)$$

$$\text{où } u(t) = \begin{cases} k > d & \text{si } M(t) = 1 \\ 0 & \text{si } M(t) = 0 \end{cases} \quad (3.5)$$

Le processus  $\{M(t), t \in \mathbb{R}^+\}$  peut être représenté par un chaîne de Markov en temps continu et son générateur est :

$$\mathbf{Q} = \begin{pmatrix} -r & r \\ p & -p \end{pmatrix}$$

Si on pose  $\mathbf{P}_M(h) = (\mathbf{P}(M(t+h) = j | M(t) = i))_{(i,j) \in [0,1]^2}$ , on a l'égalité suivante



qui découle de la définition d'une chaîne de Markov en temps continu :

$$\mathbf{P}_M(h) = \mathbf{I} + \mathbf{Q}h + o(h) \quad (3.6)$$

$$\mathbf{P}(M(t+h) = 1 | M(t) = 0) = r.h + o(h) \quad (3.7)$$

$$\mathbf{P}(M(t+h) = 0 | M(t) = 1) = p.h + o(h) \quad (3.8)$$

où  $o(h)$  représente tous les termes négligeables devant  $h$ .

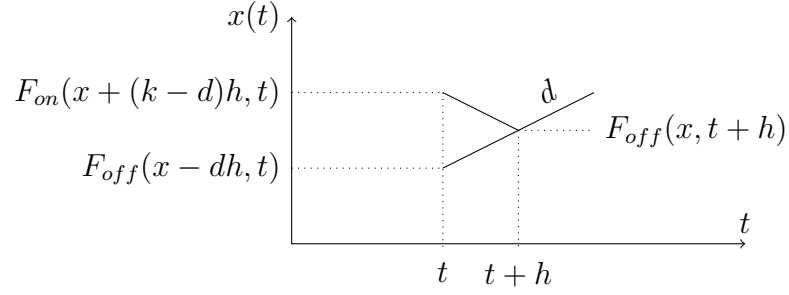
On définit les fonctions de répartition suivantes

$$\forall x \in [0, z] \quad \vec{F}(x, t) = \begin{pmatrix} F_{off}(x, t) \\ F_{on}(x, t) \end{pmatrix} = \begin{pmatrix} \mathbf{P}(X(t) \leq x, M(t) = off) \\ \mathbf{P}(X(t) \leq x, M(t) = on) \end{pmatrix} \quad (3.9)$$

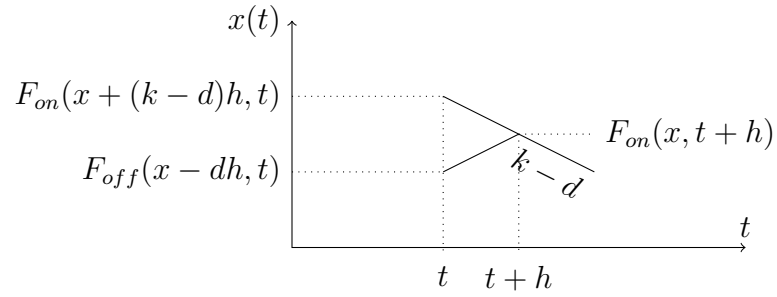
$$F_{X(t)}(x) = \mathbf{P}(X(t) \leq x) = F_{on}(x, t) + F_{off}(x, t) \quad (3.10)$$

On définit  $\mathbf{P}_0(t) = \mathbf{P}(X(t) = 0)$  et  $\mathbf{P}_z(t) = \mathbf{P}(X(t) \geq z)$  les probabilités d'être en  $x = 0$  et  $x = z$  respectivement.

**Évolution du stock entre 0 et z** En observant l'évolution du stock pendant un temps  $h$  suffisamment petit pour qu'un seul changement d'état de la machine soit possible, on peut prévoir les changements de dynamique du stock en appliquant le théorème de Bayes :

FIGURE 3.3 Changement d'état possible : 2<sup>e</sup> cas

$$\begin{aligned}
 F_{off}(x, t + h) &= \mathbf{P}(X(t + h) \leq x, M(t + h) = 0) \\
 &= \mathbf{P}(M(t + h) = 0 | M(t) = 1) \times F_{on}(x + (k - d)h, t) \\
 &\quad + \mathbf{P}(M(t + h) = 0 | M(t) = 0) \times F_{off}(x - dh, t)
 \end{aligned} \tag{3.11}$$

FIGURE 3.4 Changement d'état possible : 1<sup>er</sup> cas

$$\begin{aligned}
 F_{on}(x, t + h) &= \mathbf{P}(X(t + h) \leq x, M(t + h) = 1) \\
 &= \mathbf{P}(M(t + h) = 1 | M(t) = 1) \times F_{on}(x + (k - d)h, t) \\
 &\quad + \mathbf{P}(M(t + h) = 1 | M(t) = 0) \times F_{off}(x - dh, t)
 \end{aligned} \tag{3.12}$$

Ce qui revient à écrire après avoir utilisé l'équation (C.19) et avoir fait un développement de Taylor sur les fonctions densité :

$$\frac{F_{off}(x, t+h) - F_{off}(x, t)}{h} = -rF_{off}(x, t) + pF_{on}(x, t) - d\frac{\partial F_{off}(x, t)}{\partial x} + \frac{o(h)}{h} \quad (3.13)$$

$$\frac{F_{on}(x, t+h) - F_{on}(x, t)}{h} = rF_{off}(x, t) - pF_{on}(x, t) + (k-d)\frac{\partial F_{on}(x, t)}{\partial x} + \frac{o(h)}{h} \quad (3.14)$$

On obtient la dynamique instantanée de l'évolution du stock en faisant tendre le temps  $h$  vers 0 :

$$\frac{\partial \vec{F}(x, t)}{\partial t} = {}^t\mathbf{Q} \times \vec{F}(x, t) - \mathbf{J} \times \frac{\partial \vec{F}(x, t)}{\partial x} \quad (3.15)$$

avec,  $\mathbf{J} = \begin{pmatrix} d & 0 \\ 0 & d-k \end{pmatrix}$ .

**Évolution du stock en 0 et en  $z$**  On note que le niveau du stock est en  $z$  ssi la machine est en état de panne. Il y a donc une masse de probabilité  $\mathbf{P}_z(t)$  non nulle d'être en  $z$ . Durant un intervalle de temps infinitésimal  $h$ , cette probabilité vérifie :

$$\mathbf{P}_z(t+h) - \mathbf{P}_z(t) = -r.h\mathbf{P}_z(t) + d.h f_{off}(z^-, t)(1-r.h) + o(h) \quad (3.16)$$

L'équation ci-dessus exprime le fait que la conservation de masse de probabilité en  $z$  est le résultat net de perte de masse de probabilité suite à une réparation de la machine avec probabilité  $r.h + o(h)$ , et d'apport de masse de probabilité à partir de l'état *OFF* avec probabilité  $d.h f_{off}(z^-, t)$  pour un temps  $h$  petit. Par passage à la limite lorsque  $h$  tend vers zéro, on obtient l'équation différentielle suivante :

$$\frac{\partial \mathbf{P}_z(t)}{\partial t} = -r \mathbf{P}_z(t) + d f_{off}(z^-, t) \quad (3.17)$$

De plus, sachant que le transfert de masse de probabilité  $\mathbf{P}_z(t)$  lors de la réparation de la machine se fait vers  $z^-$  dans l'état  $ON$ , avec une vitesse  $k-d$ , on peut également écrire :

$$r \mathbf{P}_z(t) = (k-d) f_{on}(z^-, t) \quad (3.18)$$

En régime permanent, on obtient donc les équations d'équilibre :

$$\begin{aligned} \mathbf{P}_z \times r.h &= f_{off}(z^-) \times d.h \\ &= f_{on}(z^-) \times (k-d).h \end{aligned} \quad (3.19)$$

Le raisonnement est symétrique en 0.

$$\begin{aligned} \mathbf{P}_o \times p.h &= f_{off}(0^+) \times d.h \\ &= f_{on}(0^+) \times (k-d).h \end{aligned} \quad (3.20)$$

Au final, on s'intéresse au régime permanent de l'évolution du stock. L'opérateur dérivée partielle par rapport au temps  $\frac{\partial}{\partial t}$  s'annule alors et les fonction probabiliste perdent leur dépendance au temps. De plus, on dérive l'équation (3.15) pour exprimer les équations en fonction des densités de probabilité seulement. On obtient l'ensemble d'équations suivant :

$$\frac{\partial \vec{f}(x, t)}{\partial x} = \mathbf{J}^{-1} {}^t\mathbf{Q} \times \vec{f}(x, t) \quad (3.21a)$$

$$r \mathbf{P}_z = df_{off}(z^-) = (k-d) f_{on}(x = z^-) \quad (3.21b)$$

$$p \mathbf{P}_o = df_{off}(0^+) = (k-d) f_{on}(x = 0^+) \quad (3.21c)$$

Les fonctions recherchées vérifient aussi l'équation de normalisation :

$$\mathbf{P}_0 + \mathbf{P}_z + \int_0^z f_X(x)dx = 1 \quad (3.22)$$

**Résolution des équations** La solution de l'équation (3.21a) est la fonction  $\vec{f}(x) = \exp(x\mathbf{J}^{-1} {}^t\mathbf{Q})\vec{f}(0)$ , que l'on obtient après diagonalisation de la matrice :

$$\begin{aligned} \mathbf{J}^{-1} {}^t\mathbf{Q} &= \begin{pmatrix} -\frac{r}{d} & \frac{p}{d} \\ -\frac{r}{k-d} & \frac{p}{k-d} \end{pmatrix} = \begin{pmatrix} p & k-d \\ r & d \end{pmatrix} \begin{pmatrix} 0 & 0 \\ 0 & -\alpha \end{pmatrix} \begin{pmatrix} p & k-d \\ r & d \end{pmatrix}^{-1} \\ &= PDP^{-1} \end{aligned} \quad (3.23)$$

où,  $\alpha = \frac{r}{d} - \frac{p}{k-d}$ .

$$\begin{aligned} \vec{f}(x) &= \exp(x\mathbf{J}^{-1} {}^t\mathbf{Q})\vec{f}(0) = \exp(PDP^{-1}x)\vec{f}(0) \\ &= P \begin{pmatrix} 1 & 0 \\ 0 & e^{-\alpha x} \end{pmatrix} P^{-1}\vec{f}(0) \\ &= \left( \mathbf{I} + \frac{1 - e^{-\alpha x}}{\alpha} \mathbf{J}^{-1} {}^t\mathbf{Q} \right) \vec{f}(0) \end{aligned}$$

$$\Longleftrightarrow \begin{cases} f_{off}(x) &= f_{off}(0) + \left( \frac{p}{d}f_{on}(0) - \frac{r}{d}f_{on}(0) \right) \frac{1-e^{-\alpha x}}{\alpha} \\ f_{on}(x) &= f_{on}(0) + \left( \frac{p}{k-d}f_{on}(0) - \frac{r}{k-d}f_{off}(0) \right) \frac{1-e^{-\alpha x}}{\alpha} \end{cases}$$

$$\Longleftrightarrow \vec{f}(x) = \vec{f}(0)e^{-\alpha x} \quad \text{en utilisant l'équation (3.21c)} \quad (3.24)$$

En écrivant l'équation (C.5) en fonction de  $f_{off}(0)$  seulement, on peut finalement obtenir son expression :

$$(C.5) \iff f_{off}(0) \left( \frac{d}{p} + \frac{d}{r} e^{-\alpha z} + \frac{k}{k-d} \int_0^z e^{-\alpha x} dx \right) = 1 \quad (3.25)$$

$$\iff f_{off}(0) \left( \frac{d}{p} + \frac{d}{r} e^{-\alpha z} + \frac{k}{k-d} \frac{1 - e^{-\alpha z}}{\alpha} \right) = 1$$

$$\iff f_{off}(0) = \frac{p}{r+p} \frac{\alpha}{1 - \mu e^{-\alpha z}} \quad \text{où, } \mu = \frac{d}{k-d} \frac{p}{r} \quad (3.26)$$

ce qui permet d'obtenir les équations d'évolution du stock :

$$f_{on}(x) = \frac{r}{r+p} \frac{\alpha \mu e^{-\alpha x}}{1 - \mu e^{-\alpha z}} \quad (3.27a)$$

$$f_{off}(x) = \frac{p}{r+p} \frac{\alpha e^{-\alpha x}}{1 - \mu e^{-\alpha z}} \quad (3.27b)$$

$$\mathbf{P}_0 = \frac{r}{r+p} \frac{1 - \mu}{1 - \mu e^{-\alpha z}} \quad (3.27c)$$

$$\mathbf{P}_z = \frac{p}{r+p} \frac{(1 - \mu) e^{-\alpha z}}{1 - \mu e^{-\alpha z}} \quad (3.27d)$$

$$f_X(x) = f_{on}(x) + f_{off}(x) = \frac{p}{r+p} \frac{k}{k-d} \frac{\alpha e^{-\alpha x}}{1 - \mu e^{-\alpha z}} \quad (3.27e)$$

$$F(x) = \mathbf{P}_0 + \int_0^x f_X(u) du = \frac{1}{1 - \mu e^{-\alpha z}} - \frac{p}{r+p} \frac{k}{k-d} \frac{e^{-\alpha x}}{1 - \mu e^{-\alpha z}} \quad (3.27f)$$

avec,  $\mu = \frac{d}{k-d} \frac{p}{r}$  et  $\alpha = \frac{r}{d} - \frac{p}{k-d} = \frac{r}{d}(1 - \mu)$

Finalement, en comparant ces équations à celles obtenues par (15), on retrouve

de fortes similarités.

<u>Approvisionnement constant</u>	<u>Demande constante</u>
$\mathbf{P}_0 = \frac{r}{r+p} \frac{1-\mu}{1-\mu e^{-\alpha z}} \quad (3.28)$	$\mathbf{P}_0 = \frac{p}{r+p} \frac{1-\mu}{1-\mu e^{-\alpha z}} \quad (3.31)$
$\mathbf{P}_z = \frac{p}{r+p} \frac{(1-\mu)e^{-\alpha z}}{1-\mu e^{-\alpha z}} \quad (3.29)$	$\mathbf{P}_z = \frac{r}{r+p} \frac{(1-\mu)e^{-\alpha z}}{1-\mu e^{-\alpha z}} \quad (3.32)$
$f_X(x) = \frac{p}{r+p} \frac{k}{k-d} \frac{\alpha e^{-\alpha x}}{1-\mu e^{-\alpha z}} \quad (3.30)$	$f_X(x) = \frac{r}{r+p} \frac{k}{d} \frac{\alpha e^{-\alpha x}}{1-\mu e^{-\alpha z}} \quad (3.33)$
avec, $\mu = \frac{d}{k-d} \frac{p}{r}$ et $\alpha = \frac{r}{d} - \frac{p}{k-d}$	avec, $\mu = \frac{k-d}{d} \frac{r}{p}$ et $\alpha = \frac{p}{k-d} - \frac{r}{d}$

On remarque que la valeur propre  $\alpha$  d'un côté est l'exact opposé de l'autre. Cela peut s'expliquer par le fait que la dynamique du stock (équation 3.2) dans le cas que l'on étudie est inversée de celle où une demande constante est imposée. La valeur  $\mu$  d'un côté est alors l'exacte inverse de l'autre. Ainsi, il est possible d'écrire les équations de passage d'un cas à l'autre suivantes : (avec l'indice a/d pour approvisionnement/demande constante)

$\mathbf{P}_{0,d} = \mathbf{P}_{z,a} \quad \mathbf{P}_{z,d} = \mathbf{P}_{0,a} \quad (3.34)$
$f_{X_d}(x) = f_{X_a}(z-x) \quad F_{X_d}(x) = 1 - F_{X_a}(z-x) \quad (3.35)$
$E[X_d] + E[X_a] = z \quad (3.36)$

**Flux moyen dans le système** Les équations obtenues ont été déterminées en régime permanent. Elles représentent donc l'état d'équilibre du système. Il est intéressant de regarder la valeur de l'approvisionnement moyen et de la production moyenne à l'équilibre.

$$\begin{aligned}
d(1 - \mathbf{P}_z) &= d(1 - \frac{p}{r+p} \frac{(1-\mu)e^{-\alpha z}}{1-\mu e^{-\alpha z}}) \\
&= \frac{d}{1-\mu e^{-\alpha z}} (1 - \mu e^{-\alpha z} - \frac{p}{r+p} (1-\mu)e^{-\alpha z}) \\
&= \frac{d}{1-\mu e^{-\alpha z}} (1 - (\frac{p}{r+p} + \frac{r}{r+p} \mu) e^{-\alpha z}) \\
&= \frac{d}{1-\mu e^{-\alpha z}} (1 - \frac{p}{r+p} (1 + \frac{d}{k-d}) e^{-\alpha z}) \\
d(1 - \mathbf{P}_z) &= \frac{d}{1-\mu e^{-\alpha z}} (1 - \frac{p}{r+p} \frac{k}{k-d} e^{-\alpha z}) \tag{3.37}
\end{aligned}$$

$$\begin{aligned}
E[u] &= k \times (\mathbf{P}(M=1) - \mathbf{P}_o) + d \times \mathbf{P}_o \\
&= k(\frac{r}{r+p} - \frac{r}{r+p} \frac{1-\mu}{1-\mu e^{-\alpha z}}) + d \frac{r}{r+p} \frac{1-\mu}{1-\mu e^{-\alpha z}} \\
&= k \frac{r}{r+p} - (k-d) \frac{r}{r+p} \frac{1-\mu}{1-\mu e^{-\alpha z}} \\
&= \frac{1}{1-\mu e^{-\alpha z}} (k \frac{r}{r+p} (1-\mu e^{-\alpha z}) - (k-d) \frac{r}{r+p} (1-\mu)) \\
&= \frac{1}{1-\mu e^{-\alpha z}} (d \frac{r}{r+p} - k \frac{r}{r+p} \mu e^{-\alpha z} + (k-d) \frac{r}{r+p} \mu) \\
&= \frac{1}{1-\mu e^{-\alpha z}} (d - k \frac{r}{r+p} \mu e^{-\alpha z}) \\
E[u] &= \frac{1}{1-\mu e^{-\alpha z}} (d - d \frac{p}{r+p} \frac{k}{k-d} e^{-\alpha z}) \tag{3.38}
\end{aligned}$$

On remarque que l'approvisionnement moyen est bien égal à la production moyenne du système. Toute la matière qui entre dans le système en sort transformée. Il y a



donc bien conservation de la matière dans le système, lorsque celui-ci est à l'équilibre. On note ce flux moyen  $d_{moy}$  et on peut ainsi réécrire les fonctions plus simplement :

$$d \times (1 - \mathbf{P}_z) = d_{moy} \Leftrightarrow \mathbf{P}_z = \frac{d - d_{moy}}{d} \quad (3.39)$$

$$k \frac{r}{r+p} - (k-d) \mathbf{P}_o = d_{moy} \Leftrightarrow \mathbf{P}_o = \frac{\frac{r}{r+p}k - d_{moy}}{k-d} \quad (3.40)$$

**Dans le cas où la valeur propre  $\alpha$  s'annule,**  $f_{on}(x)$  et  $f_{off}(x)$  sont des constantes en fonction de  $x$ . Leur distribution est donc uniforme et le problème est plus simple à résoudre.

$$\mathbf{P}_0 = \frac{r}{r+p} \frac{1}{1 + \frac{r}{d}z} = \frac{d}{k} \frac{1}{1 + \frac{r}{d}z} \quad (3.41a)$$

$$\mathbf{P}_z = \frac{p}{r+p} \frac{1}{1 + \frac{r}{d}z} = \frac{k-d}{k} \frac{1}{1 + \frac{r}{d}z} \quad (3.41b)$$

$$f_X(x) = \frac{r}{d} \frac{1}{1 + \frac{r}{d}z} \quad (3.41c)$$

$$F(x) = \frac{\frac{r}{r+p} + \frac{r}{d}x}{1 + \frac{r}{d}z} = \frac{\frac{d}{k} + \frac{r}{d}x}{1 + \frac{r}{d}z} \quad (3.41d)$$

$$E(x) = \frac{z}{1 + \frac{r}{d}z} \left( \frac{p}{r+p} + \frac{1}{2} \frac{r}{d} z \right) \quad (3.41e)$$

avec,  $\alpha = 0 \Leftrightarrow \frac{r}{d} = \frac{p}{k-d} \Leftrightarrow \frac{r}{r+p}k = d$

On obtient le même résultat par un prolongement continu des équations (3.27) au point  $\alpha = 0$

**Dans le cas où l'aire de stockage n'a pas de limite,** on est assuré que l'approvisionnement ne sera jamais interrompu. On obtient alors les équations suivantes :

$$\alpha > 0 \quad z \rightarrow \infty$$

$$\mathbf{P}_o \rightarrow \frac{r}{r+p} (1 - \mu) \quad \mathbf{P}_z \rightarrow 0 \quad (3.42a)$$

$$f_X(x) \rightarrow \frac{p}{r+p} \frac{k}{k-d} \alpha e^{-\alpha x} \quad (3.42b)$$

$$\text{avec, } \mu = \frac{d}{k-d} \frac{p}{r} \text{ et } \alpha = \frac{r}{d} - \frac{p}{k-d}$$

La contrainte  $\alpha > 0$  est nécessaire pour que ces valeurs convergent et peut se réécrire de cette manière :

$$\alpha > 0 \quad \Leftrightarrow \quad \frac{k-d}{p} > \frac{d}{r} \quad \Leftrightarrow \quad \frac{r}{r+p} k > d \quad (3.43)$$

Cela indique que le temps moyen pendant lequel le stock est réduit  $(k-d)E[T_{on}] = \frac{k-d}{p}$  doit être supérieur au temps moyen pendant lequel le stock est augmenté  $dE[T_{off}] = \frac{d}{r}$ . Ou encore que la production moyenne de la machine  $kE[M(t) = 1] = k\frac{r}{r+p}$  doit être supérieure à l'approvisionnement  $d$ . Dans le cas inverse, le stock augmenterait à l'infini.

### 3.4 Modèle pour la ligne de production

La ligne de production est composée de  $m$  stations de travail. Chaque station de travail alimente avec sa production une aire de stockage en aval et s'alimente elle-même dans une aire de stockage en amont. On représente la ligne de production avec les flux de matière tel que dans la figure 3.5 :

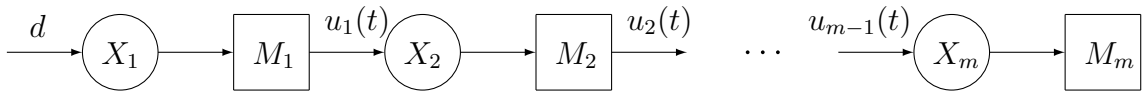


FIGURE 3.5 Représentation de la ligne de production avec  $m$  machines

Au sein de la ligne de production, des phénomènes nouveaux apparaissent et complexifient le comportement étudié précédemment pour le cas d'une seule machine. Les flux de production de chacune des stations de travail varient autrement que si ces stations étaient isolées de la ligne et pour lesquels le flux serait donné par l'équation (3.1). Les cadences de production s'ajustent à d'autres cadences ou sont bloquées selon les événements qui se produisent. Dans le cas dual de la méthode de (21), on fait l'hypothèse de capacités de production croissantes.

L'intérêt d'une telle configuration est que les phénomènes de pénurie ou de blocage de la production des machines se propagent alors dans un sens bien précis. Les phénomènes de blocage vont de l'aval vers l'amont et les phénomènes de pénurie de l'amont vers l'aval. Cette politique représente en quelque sorte un mécanisme de flux poussé. Les aires de stockage ont tendance à se remplir sous l'influence de l'approvisionnement imposé constant en début de ligne de façon à anticiper les ventes futures. On a alors les inégalités suivantes pour les cadences de production ainsi que leur comportement :

$$k_1 \leq k_2 \leq \dots \leq k_m \quad (3.44)$$

$$\begin{aligned}
& \text{Si } k_{i-1} \leq k_i \leq k_{i+1} \\
u_i(t) = & \left\{ \begin{array}{ll} u_{i-1}(t) & \text{si } X_i = 0 \\ 0 & \text{si } X_{i+1} = z_{i+1} \\ k_i & \text{sinon} \end{array} \right\} \quad \text{si } M(t) = 1 \\
& 0 \quad \text{si } M(t) = 0
\end{aligned} \tag{3.45}$$

On remarque bien que les phénomènes de blocage et d'ajustement de la production sont unidirectionnels. Cela permet de modéliser le blocage de la production de manière récursive. On se sert par la suite de cette propriété pour optimiser le coût de stockage des en-cours de production par programmation dynamique.

Dans le cas contraire, pour lequel on s'intéresse à un flux tiré qui réponde directement à la demande des clients, on adopterait une configuration des cadences de production décroissantes de sorte que les phénomènes seraient inversés et que les machines produiraient à des rythmes différents. On aurait alors le cas suivant :

$$\begin{aligned}
& \text{Si } k_{i-1} \geq k_i \geq k_{i+1} \\
u_i(t) = & \left\{ \begin{array}{ll} u_{i+1}(t) & \text{si } X_i = z_i \\ 0 & \text{si } X_{i+1} = 0 \\ k_i & \text{sinon} \end{array} \right\} \quad \text{si } M(t) = 1 \\
& 0 \quad \text{si } M(t) = 0
\end{aligned} \tag{3.46}$$

C'est le cas étudié par (21). Dans le cas que l'on étudie ici, on explique les phénomènes de blocage et de pénurie de la manière suivante :

**Le phénomène de blocage** se réalise lorsque l'aire de stockage directement en aval de la machine considérée est remplie. Cela a pour conséquence de bloquer l'approvisionnement de cette machine et donc d'*arrêter sa production*. Ce blocage peut avoir des conséquences en chaîne. Une fois bloquée, l'aire de stockage en amont se remplit plus rapidement et bloque à son tour la production de la machine en amont et ainsi de suite.

**Le phénomène de pénurie** se réalise lorsque l'aire de stockage en amont de la machine considérée se vide. Dans ce cas là, *la cadence de production de la machine considérée s'ajuste à celle de la machine juste en amont*. Cette dernière pourrait elle aussi être ajustée à d'autres machines en amont et ainsi de suite.

Dans l'environnement d'une ligne de production, la représentation d'un ensemble composé d'une machine et d'une aire de stockage doit prendre en compte ces phénomènes.

L'état de bon et mauvais fonctionnement de la machine est maintenant représenté par l'état de la machine  $\{M(t), t \in \mathbb{R}^+\}$  mais aussi par l'état de blocage de l'aire de stockage en aval de la machine considérée. La machine ne peut fonctionner correctement que lorsque celle-ci est en état de bon fonctionnement et que l'aire de stockage directement en aval est disponible pour recevoir la production de cette machine. On représente alors le phénomène de blocage de la machine (remplissage de l'aire de stockage en aval de la machine considérée) par le processus stochastique binaire 0/1  $\{I_{i+1}(t), t \in \mathbb{R}^+\}$ .

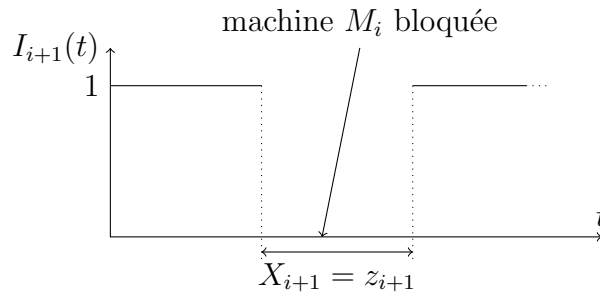


FIGURE 3.6 Évolution possible du processus  $I_{i+1}(t)$

L'approvisionnement de l'aire de stockage associée à la machine n'est plus constant en général. Il est égal à la production de la machine juste en amont ou à l'approvisionnement constant  $d$  en entrée pour le cas de la première aire de stockage et si cette demande est infinie.

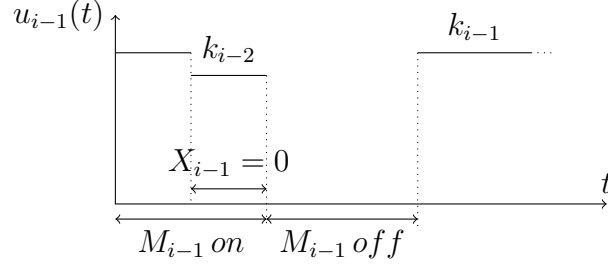


FIGURE 3.7 Évolution possible de l'approvisionnement  $u_{i-1}(t)$

On représente donc le modèle de la station de travail dans l'environnement de la ligne par la figure 3.8 suivante :

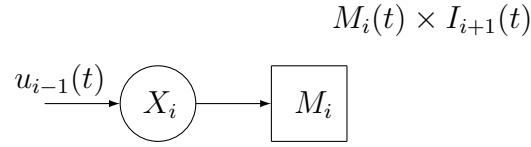


FIGURE 3.8 Représentation de la station de travail  $i$  au sein de la ligne de production

On notera par convention  $u_0(t) = d$  et  $I_{m+1}(t) = 1$ . Cette notation est ainsi valide pour tout  $i \in [1, m]$ . Dans le cas particulier de la première station de travail, on a bien un approvisionnement constant et dans le cas de la dernière station de travail, la machine ne sera jamais bloquée dans un premier temps. La ligne de production peut donc être interprétée comme un assemblage de ces modèles élémentaires.

### 3.4.1 Hypothèses pour exploiter le modèle de la ligne

Le modèle utilisé pour représenter les stations de travail dans la ligne de production ne permet pas d'utiliser les équations d'évolution du stock précédemment développées. L'objectif est donc de développer des hypothèses et des approximations pour pouvoir les utiliser. Celles-ci permettront de calculer l'espérance des en-cours de fabrication et de déterminer le coût lié à leur stockage.

On utilise dans un premier temps l'hypothèse d'ergodicité et le principe de moyennage des processus d'approvisionnement (21). De cette manière, il est possible de désolidariser les stations de travail entre elles. Par la suite, sous une hypothèse adéquate d'indépendance, on modélisera les phénomènes de blocage de la production par une réduction du temps pendant lequel les machines travaillent. On parle alors d'affaiblissement fictif de la machine. Ainsi, on construit la ligne de production par une séquence d'ensembles de base, pour lesquels ses caractéristiques ont été modifiées pour prendre en compte les phénomènes de blocage et de pénurie. On les appelle alors des ensembles fictifs.

### **Hypothèse d'ergodicité - Moyennage de l'approvisionnement**

La quantité de matière produite par chaque machine varie selon les événements. Les phénomènes de pénurie imposent aux machines d'ajuster leurs cadences de production et les phénomènes de blocage imposent l'arrêt de la production des machines de manière plus fréquente que normalement. L'aspect dynamique prend une part plus importante et il devient trop compliqué de décrire de manière exacte le comportement de la ligne comme on l'a fait pour le cas d'une seule machine. Cependant, on vérifie, qu'en régime permanent, la matière brute qui est consommée au début du système ressort de la ligne au bout d'un certain moment sous forme de produits finis. En fait, sur un temps suffisamment long, on vérifie que la quantité de matière qui entre dans le système est égale à ce qui en sort, de sorte qu'il y ait bien conservation de matière. En d'autres termes, il n'est pas possible d'accumuler de la matière indéfiniment dans la ligne.

Le taux de production moyen des machines en régime permanent est donc le même dans toute la ligne. Finalement, les phénomènes de blocage et de pénurie peuvent s'interpréter comme des phénomènes d'équilibrage de la production de la ligne de production pour assurer une production moyenne donnée au cours du temps. On a donc :

$$\forall i \in [1, m] \quad d_{moy} = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T u_i(t) dt \quad (3.47)$$

Utiliser l'*hypothèse d'ergodicité* permet de dire que la moyenne temporelle des réalisations d'un système ou processus converge vers l'espérance mathématique de ce système. Sous cette hypothèse, on a donc :

$$\forall t \in \mathbb{R}^+, \forall i \in [1, m] \quad E[u_i(t)] = d_{moy} \quad (3.48)$$

Au final, si l'on trouve les probabilités que chacune des réalisations possibles des cadences de production se produise, on peut donc se passer de la dynamique du système pour l'étudier.

Le *principe de moyennage* stipule que l'influence du processus  $u_i(t)$  sur le niveau moyen des encours de production  $E[X_{i+1}]$  est le même que si ce processus était remplacé par sa valeur moyenne,  $d_{moy}$ .

Sous l'application de ce principe, on peut maintenant considérer que chacune des aires de stockage est alimentée par un approvisionnement constant plutôt que par les valeurs variables des cadences de production des machines qui les alimentent en matière. On contourne ainsi une hypothèse du modèle de base qui avait été mise en défaut. Le taux en entrée de chaque ensemble de base peut être considéré constant. On le nomme  $\tilde{d}_i$ . Cette constante qui représente à la fois la production de la machine qui la précède et l'approvisionnement de l'aire de stockage qui la suit doit donc vérifier la contrainte que la valeur moyenne d'approvisionnement qui en résulte est égal à la production moyenne du système  $d_{moy}$ . Sachant que l'arrivée de l'approvisionnement est interrompue lorsque l'aire de stockage est remplie (c'est une conséquence de l'hypothèse de capacités de production croissantes dans la ligne), et que la probabilité que ceci arrive est  $\mathbf{P}_{z_i}$ , on a donc :

$$\tilde{d}_i(1 - \mathbf{P}_{z_i}) = d_{moy} \quad \Longleftrightarrow \quad \tilde{d}_i = \frac{d_{moy}}{(1 - \mathbf{P}_{z_i})} \quad (3.49)$$



La probabilité d'être en  $z_i$  joue un rôle important dans la ligne de production. On remarque que celle-ci permet de dimensionner l'approvisionnement constant de chaque station de travail,  $\tilde{d}_i$ , mais elle détermine aussi le blocage de la production d'une machine. Cette valeur est donc centrale dans cette étude. On la représente plus particulièrement comme le *coefficient de disponibilité* de l'aire de stockage et que l'on note  $b_i$  :

$$b_i = (1 - \mathbf{P}_{z_i}) \quad (3.50)$$

Cette valeur représente bien la proportion de temps pour laquelle l'aire de stockage n'est pas remplie. Tous les paramètres de la ligne sont fixés à priori sauf la taille des aires de stockage que l'on note  $z_i$ . Déterminer les  $z_i$  de la ligne détermine donc la probabilité que les aires de stockage soient remplies et donc, les  $b_i$ . Cette valeur permet donc en plus de dimensionner la ligne de production. C'est pourquoi on l'appelle *variable de décision*.

### Affaiblissement fictif de la machine

Le blocage d'une machine peut être une conséquence en chaîne du blocage de toutes les machines se situant en aval de celle que l'on considère. Pour bien comprendre ce mécanisme, on commence par étudier le comportement du dernier ensemble dans la ligne et les conséquences que ce blocage répercute sur la machine qui l'alimente  $M_{m-1}$ , juste en amont, et ensuite plus haut dans la ligne de production. On étudie ainsi comment la propagation de ce phénomène peut être pris en compte dans toute la ligne.

Le blocage de la machine  $M_{m-1}$ , représenté par le remplissage de l'aire de stockage  $X_m$  et par le processus  $\{I_m(t), t \in \mathbb{R}^+\}$ , a pour effet de réduire le temps moyen de fonctionnement réel de cette machine. On parle alors d'*affaiblissement de la machine*. Au final, cette machine affaiblie ne fonctionne correctement que lorsqu'elle est en état

de bon fonctionnement et que l'aire de stockage  $X_m$  est libre de recevoir la production de cette machine. Son état réel de fonctionnement peut donc être représenté par le processus stochastique suivant :

$$\{\tilde{M}_{m-1}(t), t \in \mathbb{R}^+\} \quad \text{où} \quad \tilde{M}_{m-1}(t) = M_{m-1}(t) \times I_m(t) \quad (3.51)$$

L'objectif est d'approcher ce processus aussi précisément que possible par une chaîne de Markov à 2 états afin, encore une fois, de contourner les hypothèses qui ont été mises en défaut dans le modèle de base avec une ligne composée d'une seule machine et d'une seule aire de stockage. Pour cela, on formule trois hypothèses :

- *La première approximation* consiste à considérer que les processus  $\{M_{m-1}(t), t \in \mathbb{R}^+\}$  et  $\{I_m(t), t \in \mathbb{R}^+\}$  sont indépendants,  $\forall i \in [1, m]$ .

En réalité, puisque  $k_{m-1} < k_m$ , la seule possibilité pour que la machine  $M_{m-1}$  se bloque est que celle-ci soit en état de fonctionnement et que la machine  $M_m$  (qui n'est jamais bloquée) ne soit pas en état de fonctionnement. Ainsi, le stock  $X_m$  se remplit. Les deux processus sont donc bien dépendants. Cependant, l'approximation s'explique qualitativement par le principe de moyennage. En effet, si remplacer le flux de production  $u_{m-1}(t)$  par sa moyenne n'a pas d'influence sur la valeur moyenne du stock  $X_m$ , alors, cette constante (le flux moyen) peut donc être considérée indépendante des deux processus  $\{M_{m-1}(t), t \in \mathbb{R}^+\}$  et  $\{I_m(t), t \in \mathbb{R}^+\}$ , de telle sorte que les deux processus peuvent être considérés comme indépendants entre eux.

- *La deuxième approximation* consiste à considérer que le processus  $\{I_m(t), t \in \mathbb{R}^+\}$  suit la loi d'une chaîne de Markov à 2 états.

Le processus  $\{I_m(t), t \in \mathbb{R}^+\}$  passe un certain temps dans l'état bloqué 0 puis un autre temps dans l'état de disponibilité avant de se bloquer à nouveau, puis il répète ce cycle à nouveau.

Le temps passé dans l'*état bloqué* est égal au temps nécessaire pour que la machine  $M_m$  se remette en marche, c'est-à-dire un temps exponentiel de paramètre  $r_m$ . En effet, cela est vérifié par la propriété de non-vieillessement de la loi exponentielle.

Le temps passé dans l'*état non bloqué*, qui est le temps nécessaire pour que l'aire de stockage se remplisse à nouveau, est distribué suivant une loi d'Erlang à phases. Cela a été démontrée par (18). L'approximation consiste donc à représenter ce temps par une loi exponentielle et non plus une loi générale d'Erlang à phases. On lui affecte le paramètre  $p'_m$ , et on représente dans la figure 3.9 les changements d'état de fonctionnement possibles de la machine  $M_{m-1}$ .

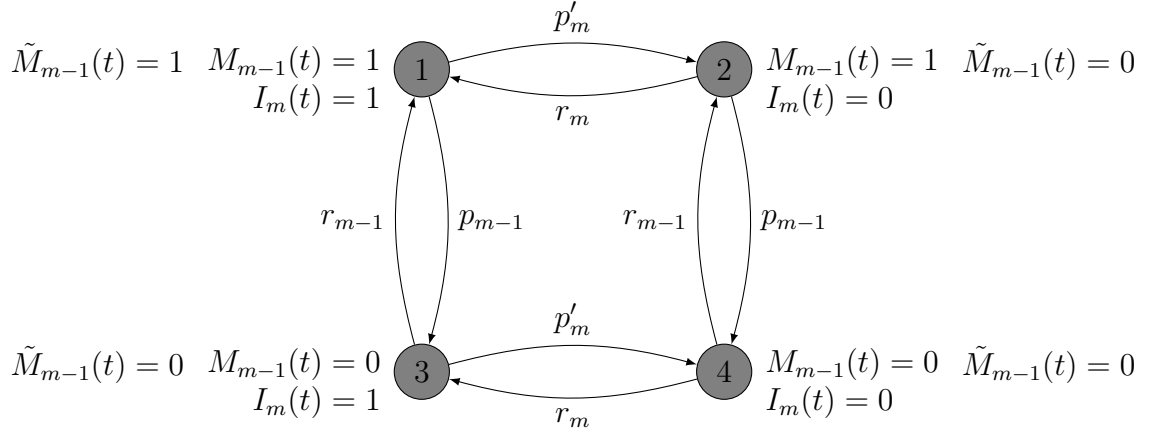


FIGURE 3.9 Changements d'états de la machine  $M_{m-1}$

En réalité, il n'est pas possible de passer de n'importe quel état à son voisin. Par exemple, pour que l'aire de stockage se remplisse, il est nécessaire que la machine soit en état de fonctionnement. A priori, il n'est donc pas possible de passer de l'état 3 à 4. Cependant, on rend cela possible par l'hypothèse faite précédemment d'indépendance des processus. L'intérêt de ces deux premières hypothèses est donc d'obtenir plus d'information sur le comportement approximatif du processus.

La probabilité que l'aire de stockage soit non remplie, en régime permanent, s'écrit  $\mathbf{P}(I_m = 1) = \frac{r_m}{r_m + p'_m}$ , elle est aussi égale à  $1 - \mathbf{P}_{z_m}$ , la probabilité de ne pas être en  $z_m$ . Il est donc possible de se passer de la valeur  $p'_m$  que l'on ne connaissait pas a priori, ainsi que de déterminer les probabilités de se trouver dans les différents états

possibles du processus :

$$b_m = 1 - \mathbf{P}_{z_m} = \mathbf{P}(I_m = 1) = \frac{r_m}{r_m + p'_m} \quad (3.52)$$

état 1	état 2	état 3	état 4
$\frac{r_{m-1}}{r_{m-1}+p_{m-1}}b_m$	$\frac{r_{m-1}}{r_{m-1}+p_{m-1}}(1-b_m)$	$\frac{p_{m-1}}{r_{m-1}+p_{m-1}}b_m$	$\frac{p_{m-1}}{r_{m-1}+p_{m-1}}(1-b_m)$
état de fonctionnement	état de non-fonctionnement		
$\frac{r_{m-1}}{r_{m-1}+p_{m-1}}b_m$	$1 - \frac{r_{m-1}}{r_{m-1}+p_{m-1}}b_m$		

TABLEAU 3.1 Probabilité des différents états de fonctionnement de la machine

L'état réel de fonctionnement de la machine, déterminé par la valeur de  $M_{m-1}(t) \times I_m(t)$ , est donc bien représenté par un état (unique) de fonctionnement de la machine et un état (multiple) de non fonctionnement.

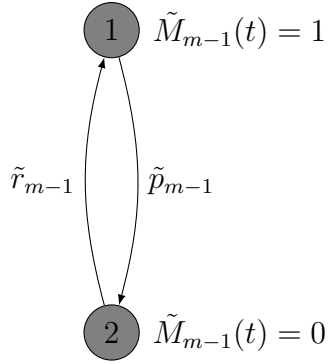
- *La troisième approximation* consiste à transformer la chaîne de Markov à 4 état en une nouvelle chaîne à 2 états.

En pratique, on considère des machines fonctionnant plus de 90 % du temps et des aires de stockage n'étant pas remplies plus de 10 % ou 20 % du temps. On néglige l'état 4 qui a donc une probabilité faible relativement aux deux autres.

L'état 1 étant le seul état de fonctionnement, on va rassembler les états 2 et 3 restants pour former une chaîne de Markov agrégée à deux états. La question se pose de savoir quel taux il est le plus judicieux d'affecter à la chaîne de Markov ? On quitte l'état 2 ou 3 pour aller dans l'état 1 avec un taux  $r_m$  ou  $r_{m+1}$  (voir figure 3.1).

Dans le cas d'une ligne *partiellement homogène*, ces taux sont égaux ( $r_{m-1} = r_m$ ). La question du choix de l'un ou l'autre ne se pose plus et on retient le taux  $r$  comme

celui de retour à l'état de fonctionnement. Le taux de sortie de l'état 1 se déduit par conservation des probabilités de fonctionnement. La figure suivante explique comment cette chaîne de Markov est construite :



Conservation des probabilités

$$\begin{aligned} \mathbf{P}(\tilde{M}_{m-1} = 1) &= \frac{\tilde{r}_{m-1}}{\tilde{r}_{m-1} + \tilde{p}_{m-1}} \\ &= \frac{r_{m-1}}{r_{m-1} + p_{m-1}} b_m \end{aligned} \quad (3.53)$$

Paramètres des lois exponentielles :

$$\tilde{r}_{m-1} = r \quad (3.54)$$

$$\tilde{p}_{m-1} = \frac{r + p_{m-1}}{b_m} - r \quad (3.55)$$

FIGURE 3.10 Changement d'états de la machine  $\tilde{M}_{m-1}$

Pour une ligne *non homogène*, ces taux ne sont pas identiques ( $r_{m-1} \neq r_m$ ). Ce cas est traité différemment et demande la prise en compte des états de fonctionnement de deux machines. Les états sont alors agrégés selon les événements possibles des différentes pannes de l'ensemble considéré. Il ne sera pas développé ici mais le lecteur intéressé pourra se référer à la thèse de (20) pour plus de détails.

En construisant ainsi le processus de fonctionnement de la machine  $M_{m-1}$ , on prend donc en compte la possibilité de blocage de celle-ci comme faisant partie entière de son modèle de représentation. Avec cette dernière approximation, on contourne finalement la dernière hypothèse mise en défaut dans le modèle d'une ligne composée d'une seule station de travail.

### Modèle approximatif de la machine $M_{m-1}$ et généralisation aux autres machines

En considérant le moyennage de l'approvisionnement et l'affaiblissement de la machine comme on vient de le faire pour la machine  $M_{m-1}$ , il est alors possible de représenter le fonctionnement de cette machine par le modèle utilisé pour une ligne

composée d'une seule station de travail. On est ainsi en mesure d'utiliser les équations d'évolution du stock que l'on connaît déjà. Celles-ci sont somme toute approximatives du cas réel.

L'intérêt double de cette méthode est qu'elle peut être étendue étape par étape à toute la ligne. Une fois que l'on a représenté l'état de fonctionnement de la machine  $M_{m-1}$  par le processus  $\{\tilde{M}_{m-1}(t), t \in \mathbb{R}^+\}$ , on peut appliquer ce même raisonnement pour la machine  $M_{m-2}$  et ainsi de suite dans toute la ligne. On peut ainsi représenter chaque station de travail d'une ligne de production composée de  $m$  machines par le modèle suivant :

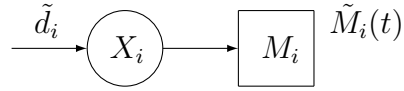


FIGURE 3.11 Représentation de la station de travail affaiblie  $i$

L'approvisionnement est considéré constant. Il est égal à  $\tilde{d}_i$ , et le comportement des machines est représenté par le modèle approximatif  $\tilde{M}_i(t)$ . On parle aussi d'une machine fictive, c'est-à-dire une machine pour laquelle le modèle représentatif de son comportement n'est pas le modèle original réel. On note les deux cas particuliers  $\tilde{d}_1 = d$ , l'approvisionnement du système en entrée, et  $\tilde{M}_m(t) = M_m(t)$  ; la dernière machine n'étant par hypothèse jamais bloquée, sa représentation est donc exacte.

Les équations de comportement se réécrivent ainsi :

$$\mathbf{P}_{0_i} = \frac{r}{r + \tilde{p}_i} \frac{1 - \mu_i}{1 - \mu_i e^{-\alpha_i z_i}} \quad (3.56)$$

$$\mathbf{P}_{z_i} = \frac{\tilde{p}_i}{r + \tilde{p}_i} \frac{(1 - \mu_i) e^{-\alpha_i z_i}}{1 - \mu_i e^{-\alpha_i z_i}} \quad (3.57)$$

$$f_{X_i}(x) = \frac{\tilde{p}_i}{r + \tilde{p}_i} \frac{k_i}{k_i - \tilde{d}_i} \frac{\alpha_i e^{-\alpha_i x}}{1 - \mu_i e^{-\alpha_i z_i}} \quad (3.58)$$

$$\text{avec, } \mu_i = \frac{\tilde{d}_i}{k_i - \tilde{d}_i} \frac{\tilde{p}_i}{r}, \alpha_i = \frac{r}{\tilde{d}_i} - \frac{\tilde{p}_i}{k_i - \tilde{d}_i}$$

$$\tilde{d}_i = \frac{d_{moy}}{b_i} \text{ et } \tilde{p}_i = \frac{r + p_i}{b_{i+1}} - r$$

La dépendance des phénomènes de blocage des machines est incluse dans la variable de décision  $b_i$ . Cette dernière est calculée en fonction des paramètres du modèle fictif de représentation du comportement de la machine  $M_i$  et en particulier en fonction de  $\tilde{p}_i$  qui dépend de  $b_{i+1}$ . Et ainsi de suite, la dépendance unidirectionnelle des phénomènes de blocage des machines est quantifiée par la variable de décision.

Finalement, la variable  $b_i$ ,  $\forall i \in [1, m]$ , comprend toute l'information nécessaire pour déterminer le comportement de la ligne de production. C'est pourquoi on lui a attribué le nom de variable de décision.

### 3.5 Modélisation mathématiques de la fonction coût

Il est dorénavant possible de modéliser les différentes stations de travail par des équations probabilistes qui décrivent l'état de remplissage des aires de stockage. En utilisant ces équations, il est donc possible de déterminer les inventaires moyens de la ligne et en leur affectant un coût unitaire  $c_i$  de stockage on obtient le coût global de la ligne de production.

$$E\left[\sum_{i=1}^m c_i \cdot X_i\right] = \sum_{i=1}^m c_i E[X_i] \quad (3.59)$$

Les  $(c_i)_{i \in [1, m]}$  représentent le coût que l'on affecte aux en-cours de fabrication à chacune des aires de stockage. Ils prennent en compte à la fois le coût de mobilisation des ressources ainsi que l'impact qu'ils peuvent avoir sur le temps moyen de nécessaire qu'il faut compter entre le placement d'une commande dans l'atelier et le moment où le produit devient disponible (lead time). La valeur ajoutée du bien produit augmentant suite aux opérations dans la ligne, on pourrait par exemple choisir ces coûts de stockage unitaires croissant dans la ligne. On peut aussi plus simplement prendre des coûts de stockage constants.

On calcule l'inventaire moyen de chacune des aires de stockage en prenant en compte les nouveaux paramètres des hypothèses formulées précédemment :

$$\begin{aligned}
 E[c_i \cdot X_i] &= c_i \left( \int_0^{z_i} x f_{X_i}(x) dx + z_i \mathbf{P}_{z_i} \right) \\
 &= c_i \frac{\tilde{p}}{r + \tilde{p}} \frac{1}{1 - \mu_i e^{-\alpha_i z_i}} \left( \frac{k_i}{k_i - \tilde{d}_i} \frac{1 - e^{-\alpha_i z_i}}{\alpha_i} - z_i \times \frac{\tilde{d}_i}{k_i - \tilde{d}_i} \frac{\tilde{p} + r}{r} e^{-\alpha_i z_i} \right)
 \end{aligned} \tag{3.60}$$

La taille de l'aire de stockage  $z_i$  peut être déterminée par la variable de décision  $b_i$  en développant l'équation (3.50) :



$$\begin{aligned}
(3.50) \Leftrightarrow \mathbf{P}_{z_i} = 1 - b_i &\Leftrightarrow \frac{\tilde{p}_i}{r + \tilde{p}_i} \frac{(1 - \mu_i)e^{-\alpha_i z_i}}{1 - \mu_i e^{-\alpha_i z_i}} = 1 - b_i \\
&\Leftrightarrow \left( \frac{\tilde{p}_i}{r + \tilde{p}_i} (1 - \mu_i) + \mu_i (1 - b_i) \right) e^{-\alpha_i z_i} = (1 - b_i) \\
&\Leftrightarrow \alpha_i z_i = \ln \left( \frac{\mu_i}{1 - b_i} \left( \frac{r}{r + \tilde{p}_i} \frac{k_i}{\tilde{d}_i} - b_i \right) \right) \quad \forall i \in [1, m]
\end{aligned} \tag{3.61}$$

avec,  $\mu_i = \frac{\tilde{d}_i}{k_i - \tilde{d}_i} \frac{\tilde{p}_i}{r}$  et  $\alpha_i = \frac{r}{\tilde{d}_i} - \frac{\tilde{p}_i}{k_i - \tilde{d}_i}$

Cependant, des conditions d'existence d'une solution s'imposent :

1. Pour que la fonction log soit bien définie, on vérifiera que :

$$\frac{\mu_i}{1 - b_i} \left( \frac{r}{r + \tilde{p}_i} \frac{k_i}{\tilde{d}_i} - b_i \right) > 0 \Leftrightarrow \frac{r}{r + \tilde{p}_i} b_{i+1} k_i > d_{moy} \tag{3.62}$$

Cette condition exprime le fait que la production moyenne de la machine considérée doit être plus faible que ce qu'elle réaliserait en moyenne si elle n'était pas limitée inférieurement. En d'autres termes, la production réelle de la machine est plus faible car cette dernière est ajustée lorsque l'aire de stockage est vide.

2. Pour avoir  $z_i > 0$ , on vérifie que :

$$\mathbf{P}_{z_i} = 1 - b_i < \frac{\tilde{p}_i}{\tilde{p}_i + r} \Leftrightarrow b_i > \frac{r}{r + \tilde{p}_i} b_{i+1} \tag{3.63}$$

Cette condition exprime le fait que la disponibilité de l'aire de stockage doit être plus importante que la probabilité de fonctionnement réel de la machine.

La proportion de temps passé en  $z_i$  est majorée par la proportion de temps de mauvais fonctionnement de la machine, et la probabilité de ne pas y être est donc minorée par la probabilité de bon fonctionnement de la machine.

On vérifie graphiquement que ces conditions sont bien nécessaires pour assurer l'existence d'une solution.

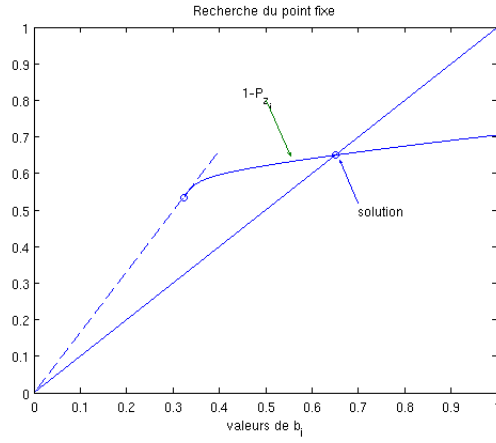


FIGURE 3.12 Solution graphique du point fixe

La courbe  $1 - \mathbf{P}_{z_i}$  commence à  $b_i = d_{moy}/k_i$  où elle prend la valeur  $\frac{r}{\tilde{p}_i+r}$  et a une tangente en ce point d'équation  $\frac{r}{\tilde{p}_i+r} \frac{k_i}{d_{moy}} \times b_i$  (la courbe en traits discontinus). On a donc bien la contrainte d'existence de la solution  $\frac{r}{\tilde{p}_i+r} \frac{k_i}{d_{moy}} \times b_i > b_i$ , qui est équivalent à  $\frac{r}{\tilde{p}_i+r} k_i \times b_{i+1} > d_{moy}$  comme pour le premier point.

Prendre en compte ces contraintes assure la condition  $\tilde{d}_i < k_i$  nécessaire pour développer le modèle à une machine. Cela implique donc la condition  $b_i \in ]\frac{d_{moy}}{k_i}, 1[$ . On représente ces contraintes dans l'ensemble  $\mathcal{A}_{i+1}(b_i)$  suivant qui détermine les valeurs possibles de  $b_{i+1}$  en fonction de  $b_i$  :

$$\mathcal{A}_{i+1}(b_i) = \left\{ b_{i+1} \in [0, 1] \text{ tels que } b_i > \frac{r}{r + p_i} b_{i+1} \text{ et } \frac{r}{r + p_i} b_{i+1} k_i > d_{moy} \right\} \quad (3.64)$$

Le cas de l'égalité pourrait être inclus dans cet ensemble mais il demande un

développement plus poussé. En pratique toutefois, il n'intervient pas réellement dans la ligne de production. Finalement, on obtient la fonction coût global des en-cours de production de chaque station de travail de la manière suivante :

$$E[c_i.X_i] = c_i \left( \int_0^{z_i} x f_{X_i}(x) dx + z_i \mathbf{P}_{z_i} \right) \quad (3.65)$$

$$\Longleftrightarrow$$

$$\forall i \in [1, m] \quad b_i \in ]\frac{d_{moy}}{k_i}, 1], \quad b_{i+1} \in \mathcal{A}_{i+1}(b_i), \quad b_{m+1} = 1$$

$$T_i(b_i, b_{i+1}) = \frac{c_i}{\alpha_i} \frac{\tilde{d}_i}{k_i - \tilde{d}_i} \left( \frac{k_i}{\tilde{d}_i} \left( b_i - \frac{r}{r + \tilde{p}_i} \right) - \frac{1 - b_i}{1 - \mu_i} \frac{\tilde{p}_i + r}{r} \ln \left( \frac{\mu_i}{1 - b_i} \left( \frac{r}{r + \tilde{p}_i} \frac{k_i}{\tilde{d}_i} - b_i \right) \right) \right) \quad (3.66)$$

avec,  $\mu_i = \frac{\tilde{d}_i}{k_i - \tilde{d}_i} \frac{\tilde{p}_i}{r}$ ,  $\alpha_i = \frac{r}{\tilde{d}_i} - \frac{\tilde{p}_i}{k_i - \tilde{d}_i}$ ,  
 $\tilde{d}_i = \frac{d_{moy}}{b_i}$  et  $\tilde{p}_i = \frac{r(1 - b_{i+1}) + p_i}{b_{i+1}}$

Dans le cas de la première aire de stockage, l'aire de stockage est limitée et il y a donc rupture de l'approvisionnement en entrée du système une fraction du temps non négligeable. En dimensionnant l'aire de stockage de manière plus ou moins grande, il est possible de déterminer quelle proportion du temps l'approvisionnement sera en rupture. Dans le cas où l'on veut s'assurer qu'il n'y ait pas de rupture de l'approvisionnement, il faut alors considérer une aire de stockage infinie, mais cette solution n'est pas réaliste d'un point de vue industriel.

La solution proposée pour garantir un approvisionnement sans rupture est de considérer une capacité de stockage supplémentaire extérieure à la ligne de production. On paie alors un coût fixe  $c_1.z_1$  pour réserver une aire de stockage de taille  $z_1$  pour l'arrivée de la matière brute au débit fixe  $d$ . Un coût unitaire additionnel  $c_2 > c_1$  est payé pour avoir accès à une aire de stockage supplémentaire. Ainsi, l'arrivée de matière brute n'est jamais interrompue. Cette capacité de stockage supplémentaire peut être considérée comme un coût d'utilisation de stockage supplémentaire non

planifié au sein même de l'usine.

La fonction coût de stockage pour la première machine se calcule en utilisant les équation (3.42). On rappelle que le coût d'utilisation de la première aire de stockage est proportionnelle à la taille de l'aire en question. On a donc :

$$\begin{aligned} T'(b_1 = 1, b_2) &= c_1 z_1 + c_2 \int_{z_1}^{\infty} (x - z_1) f_{X_1}(x) dx \\ &= c_1 z_1 + c_2 \left(1 - \frac{r}{p_1 + r} b_2\right) \frac{k_1}{k_1 - d} \frac{1}{\alpha_1} e^{-\alpha_1 z_1} \end{aligned} \quad (3.67)$$

$$\text{où } \alpha_1 = \frac{r}{d} - \frac{\tilde{p}_1}{k_1 - d} = \frac{r}{d} - \frac{r(1-b_2)+p_1}{b_2(k_1-d)} > 0$$

On détermine le coût optimal de l'aire de stockage  $X_1$  en dérivant par rapport à la variable  $z_1$  :

$$\begin{aligned} \frac{\partial T'(1, b_2)}{\partial z_1} = 0 &\iff c_1 - c_2 \frac{\tilde{p}_1}{\tilde{p}_1 + r} \frac{k_1}{k_1 - d} e^{-\alpha_1 z_1} = 0 \\ &\iff e^{-\alpha_1 z_1} = \frac{c_1}{c_2} \frac{\tilde{p}_1 + r}{\tilde{p}_1} \frac{k_1 - d}{k_1} \end{aligned} \quad (3.68)$$

L'existence d'une solution à cette équation impose la contrainte  $\frac{c_1}{c_2} \frac{k_1 - d}{k_1} \leq \frac{\tilde{p}_1}{\tilde{p}_1 + r}$ . On obtient donc la valeur du coût optimal de stockage interne et supplémentaire en intégrant cette dernière valeur dans la fonction (3.67) :

$$T'_{opt}(1, b_2) = \frac{c_1}{\alpha_1} \left( 1 + \ln \left( \frac{c_2}{c_1} \frac{\tilde{p}_1}{\tilde{p}_1 + r} \frac{k_1}{k_1 - d} \right) \right) \quad (3.69)$$

On remarque qu'il est possible que la valeur  $z_1$  soit nulle dans le cas particulier où  $\frac{c_1}{c_2} \frac{k_1 - d}{k_1} = \frac{\tilde{p}_1}{\tilde{p}_1 + r}$ . Dans ce cas là, la valeur du coût optimal de stockage est celui du coût d'utilisation de stockage supplémentaire. L'équation s'écrit :

$$T'_{opt}(1, b_2) = \frac{c_1}{\alpha_1} = \frac{c_2}{\alpha_1} \frac{\tilde{p}_1}{\tilde{p}_1 + r} \frac{k_1}{k_1 - d} \quad (3.70)$$

La fonction coût optimal de stockage interne et supplémentaire est vérifiée pour  $\alpha_1 > 0 \Leftrightarrow \frac{r}{r+p_1} b_2 k_1 > d$  et  $\frac{c_1}{c_2} \leq \frac{\tilde{p}_1}{\tilde{p}_1 + r} \frac{k_1}{k_1 - d}$ . De la même manière que précédemment, on intégrera ces contraintes dans l'ensemble  $\mathcal{A}'_2(b_1 = 1)$ .

Cette nouvelle formulation de la première aire de stockage amène à considérer deux modèles :

- Une ligne de production pour laquelle toutes les aires de stockage sont finies, y compris à l'arrivée des matériaux bruts, et pour laquelle on tolère un blocage de cette arrivée de matière avec une probabilité prédéterminée, considérée comme une contrainte de service. Ces périodes de blocage sont alors compensées par un débit de matière plus élevé en entrée pour maintenir en moyenne un débit  $d_{moy} = d$ .
- Une ligne de production pour laquelle toutes les aires de stockage sont finies et pour laquelle les ruptures d'approvisionnement ne sont pas tolérées et une aire de stockage supplémentaire est mise à disposition.

On cherche maintenant à minimiser le coût de stockage des en-cours de fabrication de la ligne. On formule les deux problèmes à résoudre de la manière suivante :

**Problème d'optimisation avec rupture d'approvisionnement tolérée :**  
**Pav**

$$\forall i \in [1, m] \quad b_i \in ]\frac{d_{moy}}{k_i}, 1], \quad b_{i+1} \in \mathcal{A}_{i+1}(b_i), \\ b_1 = t_s < 1, \quad b_{m+1} = 1$$

$$J_{tot} = \min_{(b_i)_{i \in [1, m+1]}} \sum_{i=1}^m T_i(b_i, b_{i+1}) \quad (3.71a)$$

$$T_i(b_i, b_{i+1}) = \frac{c_i}{\alpha_i} \frac{\tilde{d}_i}{k_i - \tilde{d}_i} \left( \frac{k_i}{\tilde{d}_i} (b_i - \frac{r}{r + \tilde{p}_i}) - \frac{1 - b_i}{1 - \mu_i} \frac{\tilde{p}_i + r}{r} \ln \left( \frac{\mu_i}{1 - b_i} \left( \frac{r}{r + \tilde{p}_i} \frac{k_i}{\tilde{d}_i} - b_i \right) \right) \right) \quad (3.71b)$$

avec,  $\mu_i = \frac{\tilde{d}_i}{k_i - \tilde{d}_i} \frac{\tilde{p}_i}{r}$ ,  $\alpha_i = \frac{r}{\tilde{d}_i} - \frac{\tilde{p}_i}{k_i - \tilde{d}_i}$   
 $\tilde{d}_i = \frac{d_{moy}}{b_i}$  et  $\tilde{p}_i = \frac{r(1 - b_{i+1}) + p_i}{b_{i+1}}$

**Problème d'optimisation avec rupture d'approvisionnement non tolérée :  $\mathbf{Pav}$**

$$\forall i \in [1, m] \quad b_i \in ]\frac{d_{moy}}{k_i}, 1], \quad b_{i+1} \in \mathcal{A}_{i+1}(b_i), \\ b_1 = t_s = 1, \quad b_{m+1} = 1$$

$$J_{tot} = \min_{(b_i)_{i \in [1, m+1]}} \left( T'_{opt}(b_1, b_2) + \sum_{i=2}^m T_i(b_i, b_{i+1}) \right) \quad (3.72a)$$

$$T_i(b_i, b_{i+1}) = \frac{c_i}{\alpha_i} \frac{\tilde{d}_i}{k_i - \tilde{d}_i} \left( \frac{k_i}{\tilde{d}_i} (b_i - \frac{r}{r + \tilde{p}_i}) - \frac{1 - b_i}{1 - \mu_i} \frac{\tilde{p}_i + r}{r} \ln \left( \frac{\mu_i}{1 - b_i} \left( \frac{r}{r + \tilde{p}_i} \frac{k_i}{\tilde{d}_i} - b_i \right) \right) \right) \quad (3.72b)$$

$$T'_{opt}(b_1, b_2) = \frac{c_1}{\alpha_1} \left( 1 + \ln \left( \frac{c_2}{c_1} \frac{\tilde{p}_1}{\tilde{p}_1 + r} \frac{k_1}{k_1 - d} \right) \right) \quad (3.72c)$$

$$\text{avec, } \mu_i = \frac{\tilde{d}_i}{k_i - \tilde{d}_i} \frac{\tilde{p}_i}{r}, \quad \alpha_i = \frac{r}{\tilde{d}_i} - \frac{\tilde{p}_i}{k_i - \tilde{d}_i}, \quad \alpha_1 = \frac{r}{d} - \frac{\tilde{p}_1}{k_1 - d} \\ \tilde{d}_i = \frac{d_{moy}}{b_i}, \quad \tilde{p}_i = \frac{r(1 - b_{i+1}) + p_i}{b_{i+1}} \text{ et } \tilde{p}_2 = \frac{r(1 - b_2) + p_1}{b_2}$$

Finalement, en écrivant  $T_1(b_1 = 1, b_2) = T'_{opt}(1, b_2)$  et  $\mathcal{A}_2(b_1 = 1) = \mathcal{A}'_2(1)$ , il est possible d'adopter la seule formulation **Pav** pour considérer les deux problèmes. La différence entre les deux problèmes se fait par le choix de la valeur  $b_1$ .

### 3.6 Conclusion

La ligne de production a été décomposée en ensembles de base d'une seule machine et d'une seule aire de stockage. Ces ensembles permettent d'obtenir des équations exactes décrivant l'évolution des inventaires sous certaines conditions qui n'étaient pas vérifiées à priori. Le principe de moyennage et l'aggrégation des états de bon et mauvais fonctionnement en une seule chaîne de Markov à deux états ont permis d'utiliser ces fonctions probabilistes pour décrire la valeur moyenne approximative

des inventaires de chaque aire de stockage.

La structure croissante des taux de production nominal des machines s'est avérée être l'hypothèse clé pour pouvoir intégrer les phénomènes d'ajustement et de blocage des taux de production des machines ainsi que leur propagation unidirectionnelle dans le modèle. Cela a ainsi mené à considérer deux modèles de lignes de production. Un pour lequel la rupture d'approvisionnement du système est permise et l'autre pour lequel cette rupture n'est pas permise. Les modèles déterminent ainsi le coût moyen d'utilisation des ressources du système.

Ce modèle de la ligne de production contient toute l'information nécessaire pour déterminer le dimensionnement des aires de stockage, la production moyenne du système, ainsi que le coût total de mobilisation des ressources de l'entreprise. L'intérêt est maintenant d'exploiter ce modèle ainsi que de l'optimiser pour permettre de répondre aux exigences économiques auxquelles font face les entreprises de production.



# Chapitre 4

## Simulation de la ligne de production

### 4.1 Introduction

En vue de valider les approximations proposées pour l'analyse et l'optimisation de la ligne de production, il est nécessaire de recourir à des simulations Monte Carlo, car le comportement des machines est aléatoire. Cette méthode consiste à faire un nombre suffisamment grand de tirages des variables aléatoires du système pour converger vers leurs valeurs moyennes (principe de la loi des grands nombres).

On va donc générer de manière aléatoire les temps de bon et mauvais fonctionnement,  $T_{on}^i$  et  $T_{off}^i$ , des  $m$  différentes machines de la ligne puis tracer l'évolution exacte des aires de stockage en fonction des tirages obtenus. Chaque tirage permet de construire une évolution possible des aires de stockage de la ligne qu'il faudra donc répéter un certain nombre de fois pour converger vers leurs valeurs moyennes. Les résultats obtenus par la simulation permettront ainsi de valider ou de réfuter les hypothèses qui ont été prises pour construire le modèle de la ligne.

Il faudra aussi vérifier que les tirages sont bien indépendants entre eux et que les propriétés statistiques des variables générées sont bien vérifiées.

### 4.2 Réalisation de la simulation

La partie non-déterministe de la simulation est le comportement des machines. Comme on l'a déjà vu, celles-ci passent des temps aléatoires,  $T_{on}^i$  et  $T_{off}^i$ , de bon et mauvais fonctionnement les uns à la suite des autres.

Une fois que ces temps de fonctionnement ont été générés, l'évolution des aires de stockage est entièrement connue : les cadences de production s'ajustent en fonction de l'état de la ligne de production et les aires de stockage se remplissent et se vident au rythme de ces cadences qui changent régulièrement.

On décompose donc la simulation de la ligne de production en 2 parties :

**Une partie aléatoire** qui consiste à faire des tirages aléatoires de l'état de fonctionnement des machines.

**Une partie déterministe** qui consiste à tracer l'évolution des aires de stockage en fonction des tirages que l'on a obtenu.

On réalise la programmation de la simulation à l'aide du logiciel Matlab. Le programme complet se trouve en annexe A.

#### 4.2.1 Générer l'horizon temporel de chaque machine

Pour créer l'horizon temporel des  $m$  machines, on génère des temps aléatoires de bon et mauvais fonctionnement les uns à la suite des autres jusqu'à atteindre l'horizon temporel voulu,  $T$ . Ces temps aléatoires sont *i.i.d.* et suivent une loi exponentielle de paramètre  $p_i$  quand la machine fonctionne correctement et de paramètre  $r_i$  quand la machine ne fonctionne pas correctement. On obtient donc l'horizon de la machine  $i$  lorsque l'on vérifie l'équation suivante :<sup>1</sup>

$$Horizon_i = T_{on}^{i,1} + T_{off}^{i,2} + \dots + T_{on}^{i,n-1} + T_{off}^{i,n} \geq T \quad (4.1)$$

Ainsi, pour  $m$  machines, on utilise l'algorithme suivant :

```
rand('state',100*clock); h=[]; H=zeros(n,m); l=1;
while min(h)<T
    H(l,:)=exprnd(1./p); H(l+1,:)=exprnd(1./r); h=sum(H); l=l+2;
end
```

La fonction `exprnd(mu)` génère des nombres pseudo-aléatoires qui suivent une loi ex-

---

1. il faudra  $n = \text{ent}(2 \times T(\frac{1}{p_i} + \frac{1}{r_i}) + 1)$  tirages en moyenne, car les temps de fonctionnement de la machine sont *i.i.d.*

ponentielle de moyenne  $\mu$ . C'est une transformation bijective de la fonction `rand` de Matlab.

Les propriétés statistiques qui nous intéressent sont principalement la moyenne des variables aléatoires et leurs indépendances entre elles. Le générateur de nombre pseudo-aléatoire interne de Matlab (`rand`) assure lui-même l'indépendance des tirages. On calcule ensuite la moyenne empirique des tirages pour chaque machine et on vérifie, sur l'horizon de la simulation, que la différence entre les moyennes théorique et empirique est inférieur à 1%, afin de s'assurer que le nombre de tirages de ces nombres pseudo-aléatoires ait été suffisante.

Ces critères sont nécessaires pour vérifier les propriétés statistiques des variables aléatoires générées, mais ils ne permettent pas de déterminer la longueur de l'horizon  $T$  choisi pour la simulation. Pour cela, on détermine un critère de convergence des grandeurs simulées qui nous intéresse (la moyenne des en-cours de production de chacune des aires de stockage et leur disponibilité moyenne). Ce critère est défini dans la section 4.3.

### 4.2.2 Tracer l'évolution des aires de stockage

Une fois que l'horizon de la simulation est généré, on peut alors tracer l'évolution des aires de stockage. Celles-ci évoluent dans le temps en fonction des cadences de production des machines,  $u_i(t)$ , qui changent régulièrement. Il existe 3 raisons pour lesquelles les cadences de production des machines sont amenées à changer : soit *une machine change d'état*, soit *une aire de stockage se remplit complètement* ou soit *une aire de stockage se vide complètement*. C'est-à-dire qu'on prend en compte les phénomènes de blocage et de pénurie en plus des changements d'état des machines.

Ainsi, entre deux événements consécutifs,  $E_j$  et  $E_{j+1}$ , les cadences de production des machines sont constantes et il est donc possible de faire évoluer les aires de stockages au rythme de ces premières avec la relation suivante :

$$X_i(t_{j+1}) = X_i(t_j) + u_i(t_j) \times (t_{j+1} - t_j) \quad (4.2)$$

où,  $t_j$  est le temps auquel arrive l'évènement  $E_j$ .

La méthode utilisée est intuitive et consiste à avancer évènement par évènement dans le temps en mettant à jour les cadences de production des machines et les aires de stockage jusqu'à atteindre l'horizon  $T$ . La principale difficulté de la simulation réside comme on peut le deviner dans la mise à jour des cadences de production des machines. En effet, sachant que chaque machine peut être amenée à adapter sa cadence de production à toutes les machines qui sont en amont d'elle, il existe alors  $(m + 1)!$  différents cas. Cependant, une remarque permettra de simplifier la programmation de la fonction, `rateOfProduction`, qui met à jour les cadences de production.

Pour  $m$  machines, l'algorithme proposé est le suivant :

```

M_bool=ones(1,m);      % toutes les machines sont actives
O_bool=zeros(1,m+1);   % aucune aire de stockage n'est vide
Z_bool=zeros(1,m+1);   % aucune aire de stockage n'est remplie
X=Z/2;                 % initialisation des stocks
U1=[d;k];              % taux auquel les aires de stockage sont augmentées
U2=[k;d];              % taux auquel les aires de stockage sont diminuées
t=0; j=1;
while t(j) < T
    t+=min(temps avant un prochain évènement);
    E=évènement produit; t(j+1)=t(j)+t+; X(j+1)=X(j)+(U1-U2)*t+;
    mise à jour de M_bool, O_bool et Z_bool en fonction de E;
    [U1,U2]=rateOfProduction(E,M_bool,O_bool,Z_bool,U1,U2);
    j=j+1;
end

```

La fonction `rateOfProduction` ajuste les taux de production en fonction de l'évènement qui s'est produit et des variables booléennes  $M_{bool}$ ,  $O_{bool}$  et  $Z_{bool}$  qui permettent de rendre compte de l'état actuel de la ligne de production.

Pour simplifier la programmation de cette fonction, on peut remarquer qu'*un évènement n'a pour effet que de modifier localement les cadences de production de la ligne de production*. Ainsi, connaissant l'état actuel de la ligne de production, il suffit de regarder comment l'évènement affecte les cadences de production localement pour

les mettre à jour.

Enfin, la précision attendue sera toujours relative à la précision obtenue pour la génération de l'horizon temporel.

### 4.3 Résultats des simulations

Il est maintenant question de définir un cadre expérimental pour les simulations et de comparer les résultats obtenus avec le modèle que l'on a réalisé.

Principalement, on s'assure que l'état 4 de non fonctionnement (voir figure 3.9 ) des stations de travail est bien de probabilité négligeable, que la longueur de l'horizon  $T$  choisi assure la convergence des grandeurs simulées qui nous intéressent, et que les propriétés statistiques des variables aléatoires générées sur cet horizon sont bien vérifiées. Pour chacun de ces trois points, on détermine les critères suivants :

1. Les paramètres des lois exponentielles sont choisis de telle sorte que les machines fonctionnent correctement 90% du temps. On vérifie a posteriori que ces valeurs multipliées par la probabilité de non disponibilité des aires de stockage permet bien de négliger la probabilité d'être dans l'état 4 de non fonctionnement (voir figure 3.9 ) des stations de travail. Ce critère permet de vérifier la troisième hypothèse qui a été formulée pour agréger les états de fonctionnement de la machine affaiblie, dans la section 3.4.1.
2. Les grandeurs simulées doivent se stabiliser autour d'une certaine valeur dans un intervalle égal au plus à 1% de cette valeur et ceci pendant un temps au moins égal à la moitié de l'horizon  $T$  choisi. Si ce critère de convergence n'est pas vérifié, l'horizon  $T$  de la simulation n'est pas validé et il doit alors être plus grand.
3. L'horizon de la simulation doit assurer une génération suffisamment importante des variables aléatoires du système pour que la différence entre la moyenne des temps de fonctionnement des machines et leurs valeurs théoriques soit de l'ordre

de 1%. Si ces deux derniers points sont vérifiés, alors l'horizon choisi  $T$  est validé.

Ces différents critères sont bien évidemment vérifiés pour les simulations que l'on a faites. On en discute après avoir montré les résultats.

**Distribution de l'état stationnaire** Le modèle que l'on a réalisé est valable en régime permanent. Or les conditions initiales choisies pour la simulation ne sont pas représentatives de ce régime. En effet, au départ de la simulation, toutes les machines sont en état de fonctionnement et les aires de stockage sont remplies de moitié. Les valeurs mesurées durant le lancement de la simulation présentent donc un biais dont il est nécessaire de s'affranchir pour garantir des résultats précis.

Pour détecter cette période de lancement, on utilise la méthode graphique développée par (29). Cette méthode consiste à réaliser  $n = 20$  simulations de  $T$  unités de temps ( $T$  est l'horizon choisi). On considère alors pour chaque unité de temps la moyenne empirique,  $Y_i$ , des  $n$  simulations. On lisse ensuite cette valeur,  $Y_i(w)$ , sur  $w$  unités de temps afin de détecter à partir de quand les valeurs du système convergent vers l'état stationnaire.  $Y_i$  et  $Y_i(w)$  sont obtenus de la manière suivante :

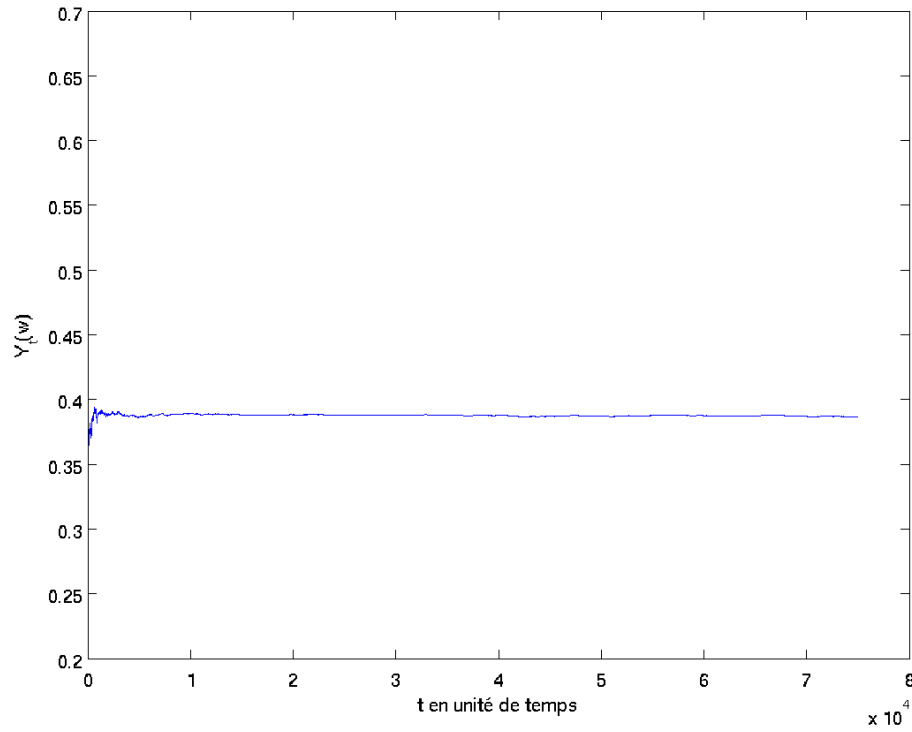
$$Y_i = \frac{1}{n} \sum_{j=1}^n Y_{i,j} \quad \forall i \in [1, T] \quad (4.3)$$

où  $Y_{i,j}$  est l'observation au temps  $i$  de la simulation  $j$ .

$$Y_i(w) = \frac{1}{2i+1} \sum_{s=1}^{2i+1} Y_s \quad \forall i \in [1, w] \quad (4.4)$$

$$Y_i(w) = \frac{1}{2w+1} \sum_{s=-w}^w Y_{i+s} \quad \forall i \in [w+1, T-w] \quad (4.5)$$

Pour obtenir les  $Y_{i,j}$ , il est nécessaire de découper la simulation en unité de temps. On obtient alors la courbe suivante dans le cas d'un système de trois machines, où  $n = 20$ ,  $T = 100000$ ,  $w = T/4$  :

FIGURE 4.1 Évolution de  $Y_i(w)$  sur la longueur de l'horizon

On remarque que le système converge vers le régime permanent en moins de 10.000 unités de temps, ce que l'on vérifiera aussi dans tous les autres cas. On déduit donc ce temps et les valeurs qui lui sont associées et on s'assure ainsi que la précision des grandeurs mesurées est meilleure.

On présente dans les tableaux qui suivent les résultats de simulation des lignes de production composées de 1, 2, 3, 4, 5 et 10 machines. Les paramètres choisis sont les suivants :

$$\begin{aligned}
 &\forall i \in [1, m] \\
 &r_i = 0.9 \quad p_i = 0.1 \quad k_1 = 3 \quad k_i = k_1 + (i - 1) \times 0.2 \\
 &c_i = 1 \quad d_{moy} = 1 \quad t_s = 0,95
 \end{aligned} \tag{4.6}$$

TABLEAU 4.1 Résultats de simulation pour une seule machine

variable	valeur simulée	valeur théorique	différence
$J_{tot}$	0,0699	0,0695	-0,57%
$\mathbf{P}_o$	0,8725		
$1 - \mathbf{P}_z$	0,9497	0,95	0,03%
$E[u_i]$	0,9997	1	-0,03%
$z_1$	0,8246		

Dans le cas d'une seule machine, on a des équations exactes, et on vérifie bien que les valeurs simulées sont très proches des valeurs théoriques, à moins de 0,6% de différence. La précision de ces valeurs permet de valider l'hypothèse d'ergodicité pour le modèle de base. Les valeurs simulées sont calculées par une moyenne temporelle des réalisations du système alors que les valeurs théoriques sont obtenues en considérant le régime permanent du système. Leur similarité prouve le bien fondé de cette hypothèse.

Pour les cas qui suivent, le débit en entrée du système a été diminué de 5%. Deux arguments motivent ce choix. D'abord, on observe que de cette manière les imprécisions sur les variables mesurées s'homogénéisent. Ensuite, les paramètres choisis sont tels que les contraintes sur le système sont importantes. Réduire le débit en entrée du système permet de stabiliser la production dans la ligne et d'éviter les conditions critiques d'utilisation du système.

TABLEAU 4.2 Résultats de simulation pour une ligne de 2 machines

variable	valeur simulée	valeur théorique	différence
$J_{tot}$	0,2179	0,2162	-0,78%
$\mathbf{P}_{o_1}$	0,7842		
$\mathbf{P}_{o_2}$	0,8267		
$1 - \mathbf{P}_{z_1}$	0,9538	0,95	-0,4%
$1 - \mathbf{P}_{z_2}$	0,9637	0,961	-0,3%
$E[u_i]$	0,9519	1	5,05%
$z$	[1,20 ; 1,11]		

Dans le cas d'une ligne de deux machines, la différence entre la valeur du coût total simulée et théorique est très faible. On note que la production moyenne simulée est la même pour les deux machines. On vérifie donc bien l'hypothèse de conservation



de matière dans le système. On remarque aussi que la diminution du débit de 5% se répercute sur leurs mesures et que les variables de décision sont très proches de leur valeur simulée. Ce modèle présente donc de très bons résultats et assure un dimensionnement exact de la ligne de production.

TABLEAU 4.3 Résultats de simulation pour une ligne de 3 machines

variable	valeur simulée	valeur théorique	différence
$J_{tot}$	0,3693	0,3607	-2,33%
$\mathbf{P}_{o_1}$	0,8032		
$\mathbf{P}_{o_2}$	0,8282		
$\mathbf{P}_{o_3}$	0,8871		
$1 - \mathbf{P}_{z_1}$	0,9465	0,95	0,37%
$1 - \mathbf{P}_{z_2}$	0,9452	0,945	-0,02%
$1 - \mathbf{P}_{z_3}$	0,9634	0,959	-0,46%
$E[u_i]$	0,9468	1	5,62%
$z$	[1,33;1,10;1,05]		

Les conclusions pour ce cas et les suivants sont similaires et sont expliquées à la suite des tableaux présentés ci-dessous.

TABLEAU 4.4 Résultats de simulation pour une ligne de 4 machines

variable	valeur simulée	valeur théorique	différence
$J_{tot}$	0,5207	0,5019	-3,61%
$\mathbf{P}_{o_1}$	0,8177		
$\mathbf{P}_{o_2}$	0,8237		
$\mathbf{P}_{o_3}$	0,8557		
$\mathbf{P}_{o_4}$	0,9161		
$1 - \mathbf{P}_{z_1}$	0,9447	0,95	0,56%
$1 - \mathbf{P}_{z_2}$	0,9323	0,937	0,5%
$1 - \mathbf{P}_{z_3}$	0,9366	0,938	0,15%
$1 - \mathbf{P}_{z_4}$	0,9598	0,957	-0,3%
$E[u_i]$	0,9463	1	5,67%
$z$	[1,39;1,11;0,98;0,99]		

TABLEAU 4.5 Résultats de simulation pour une ligne de 5 machines

variable		valeur simulée	valeur théorique			différence
$J_{tot}$		0,6798	0,6404			-5.796%
Machine n°		1	2	3	4	5
$z$		1,42	1,12	0,97	0,90	0,94
$\mathbf{P}_{o_i}$	Simulé	0,8171	0,8125	0,8334	0,8723	0,9331
$1 - \mathbf{P}_{z_i}$	Simulé	0,9415	0,9224	0,9213	0,9322	0,9576
$b_i$	Théorique	0,95	0,933	0,927	0,933	0,955
	Différence	0,9%	1,15%	0,62%	0,09%	-0,27%
$E[u_i]$	Simulé	0,9447				
	Théorique	1				
	Différence	5,85%				

TABLEAU 4.6 Résultats de simulation pour une ligne de 10 machines

variable		valeur simulée	valeur théorique			différence
$J_{tot}$		1,5135	1,3083			-13,56%
Machine n°		1	2	3	4	5
$z$		1,47	1,12	0,96	0,86	0,80
$\mathbf{P}_{o_i}$	Simulé	0,8052	0,7768	0,7742	0,7837	0,8018
$1 - \mathbf{P}_{z_i}$	Simulé	0,93	0,8912	0,8741	0,8665	0,8694
$b_i$	Théorique	0,95	0,927	0,912	0,902	0,897
	Différence	2,15%	4,02%	4,33%	4,09%	3,17%
$E[u_i]$	Simulé	0,9354				
	Théorique	1				
	Différence	6,91%				
Machine n°		6	7	8	9	10
$z$		0,77	0,75	0,75	0,77	0,91
$\mathbf{P}_{o_i}$	Simulé	0,8209	0,8472	0,8824	0,9298	0,9843
$1 - \mathbf{P}_{z_i}$	Simulé	0,8752	0,8858	0,9036	0,9287	0,9601
$b_i$	Théorique	0,896	0,899	0,908	0,925	0,954
	Différence	2,37%	1,49%	0,49%	-0,4%	-3,08%

On remarque que pour des lignes de production composées de 2 à 5 machines, les résultats sont encore très bons. Le dimensionnement des variables de décisions est très précis avec une erreur maximale de 1,5%. La conservation de matière dans le système est toujours vérifiée et on retrouve bien que les 5% de réduction du débit se répercutent sur les observations avec une différence maximale de 5,85%. Et finalement, le coût global de stockage des inventaires est lui aussi très bien dimensionné. La différence entre la grandeur simulée et la grandeur théorique est légèrement croissante avec le nombre de machines et atteint un maximum de 5,8%.

Pour le cas d'une ligne de 10 machines, on remarque encore que le dimensionnement des variables de décision et la prédiction du flux de matière dans le système est précis. On note tout de même que la différence entre le coût théorique et le coût simulé est cette fois nettement plus importante.

**Interprétation des résultats** La précision de la fonction coût totale, la somme des inventaires moyens de la ligne, est très bonne. On note une légère diminution de cette précision quand le nombre de machines considérées augmente. Dans le cas d'une ligne de 10 machines, cette imprécision est cette fois plus importante. Une explication de cette dernière imprécision pourrait être la suivante.

Aussi petit que peut être l'état 4 négligé, l'approximation de la chaîne de Markov ne prend pas en compte le fait qu'un temps de mauvais fonctionnement pourrait durer plus longtemps que le modèle le décrit. En passant de l'état 1 à 2, le processus pourrait passer par l'état 4 puis 3 pour enfin revenir à l'état 1 de fonctionnement (voir le schéma 3.9). Ainsi, le temps de bon fonctionnement moyen des machines dans la ligne serait sur-estimé, ce qui conduirait à trouver des inventaires théoriques plus faibles. Ce phénomène serait amplifié pour le cas d'une ligne de production de 10 machines.

Une part de l'imprécision vient aussi des hypothèses que l'on a prises. En pratique, les 4 états de la chaîne de Markov ne communiquent pas tous ensemble. Par exemple, si la machine est en panne, il n'est pas possible que l'aire de stockage en aval se remplisse et donc il n'est pas possible de passer de l'état 3 à 4. De cette manière, on prend en compte un phénomène qui génère une imprécision dans le modèle. Ceci dit, cette imprécision intervient pour une ligne de production composée de 10 machines seulement.

## 4.4 Conclusion

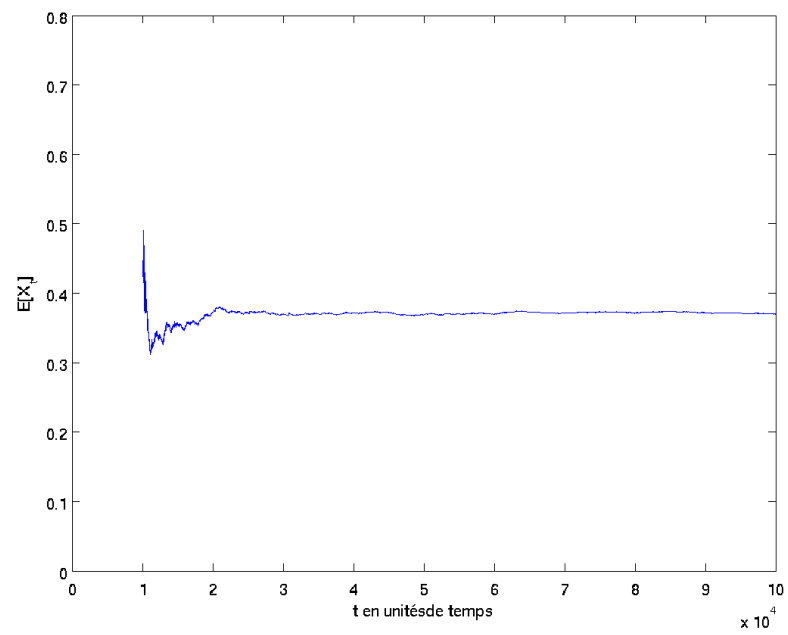
Au final, la précision globale du modèle est frappante pour des lignes de production allant jusqu'à 5 machines en série. Le dimensionnement des variables de décision est très précis et la prédiction du coût de stockage des inventaires dans la ligne de production très bon. Les hypothèses qui ont été réalisées pour construire le modèle sont donc validées et la méthode décrite représente une bonne analyse des phénomènes qui régissent l'évolution de la ligne de production.

**Validation des simulations** On montre ici que les trois points nécessaires pour valider les simulations du système sont bien vérifiés :

- L'état 4 pour lequel la machine est en panne et l'aire de stockage en aval est non disponible, est bornée par la valeur 0.0125 pour chacune des machines des différentes lignes de production. Cette probabilité peut donc être négligée à juste titre sans perdre beaucoup de précision sur le processus.
- Pour l'horizon choisi de  $T = 100.000$  unités de temps, dans le cas d'une ligne de 3 machines, la convergence de la somme des espérances des aires de stockage se comporte de la manière suivante ( voir figure 4.4).  
On vérifie que cette fonction se stabilise dans un intervalle inférieur à 1% de la valeur limite et ceci sur plus de la moitié de la longueur de l'horizon  $T$ . Cela a été vérifié pour toutes les simulations.
- Enfin, la différence entre les temps de fonctionnement empiriques moyens de chacune des machines et leur valeur théorique dans chacun des cas est bornée par 0,98%.

En vérifiant ces trois critères, les conditions de simulation du modèle et les résultats obtenus sont donc validés.

FIGURE 4.2 Évolution de  $E[\sum X_{i,t}]$  sur la longueur de l'horizon



# Chapitre 5

## Optimisation du coût de stockage des encours de la ligne de production

### 5.1 Introduction

Une fois notre modèle validé par simulation, on souhaite optimiser le coût de stockage des en-cours de production. Les taux de production des machines ont été préalablement choisis pour imposer une dépendance ordonnée entre les stations de travail; le coût de stockage attribué à une station de travail dépend de l'affaiblissement de sa machine, qui dépend lui-même de l'affaiblissement de toutes les machines en aval de sa position. On vérifie aussi que la fonction coût global est une somme de fonctions coûts,  $T_i(b_i, b_{i+1})$  dans la plupart des cas, qui sont monotones croissantes en fonction de  $b_i$  et monotones décroissantes en fonction de  $b_{i+1}$ . Ainsi, trouver le minimum de la fonction objectif est un problème qui se prête parfaitement à l'utilisation des équations de Bellman et de la programmation dynamique.

L'algorithme optimise un premier sous-problème, puis optimise un nouveau sous-problème qui inclut le premier sous-problème jusqu'à considérer la ligne de production en entière. La relation de dépendance à sens unique de la ligne de production permet de définir la manière dont il faut augmenter chaque sous-problème. Les variables de décision sont alors déterminées par la valeur qui minimise chaque sous-problème.

La structure de ce chapitre est la suivante. On met en place les équations de Bellman qui permettent de formuler notre problème à optimiser, puis on explique les étapes du programme qui a été réalisée. Enfin, on montre différents résultats intéressants obtenus puis on discute de leur signification.

## 5.2 Formulation du problème d'optimisation

Les problèmes **Pav** (3.71) et **Pav'** (3.72) sont une somme de fonction coût dont les variables de décision sont dépendantes entre elles de manière unidirectionnelle. La variable  $b_i$  dépend de la variable  $b_{i+1}$  et ainsi de suite jusqu'à la variable  $b_m$ . La direction naturelle de résolution du programme dynamique est donc dans le sens inverse, en commençant par la dernière machine, qui ne dépend pas de l'état d'affaiblissement des autres machines. Le remplissage de l'aire de stockage qui lui est associé détermine ainsi l'affaiblissement de la machine qui la précède. On commence donc par optimiser le coût de stockage des en-cours de production de cette machine, qui définit le premier sous-problème, puis on construit le sous-problème suivant en ajoutant une machine et ainsi de suite. Finalement, le problème d'optimisation dynamique se résout en posant les équations de Bellman (5.1) suivantes :

- La fonction coût instantanée :

$$g(X_i, b_i, b_{i+1}) = c_i \cdot X_i \quad (5.1a)$$

- Le coût à long terme du premier sous-problème :

$$J_m(b_m) = E[g(X_m, b_m, b_{m+1})] = T_m(b_m, b_{m+1}) \quad (5.1b)$$

- Le coût à long terme du ième sous-problème :

$$\begin{aligned} J_i(b_i) &= \min_{b_{i+1}} E[g(X_i, b_i, b_{i+1}) + J_{i+1}(b_{i+1})] \\ &= \min_{b_{i+1} \in \mathcal{A}_{i+1}(b_i)} (T_i(b_i, b_{i+1}) + J_{i+1}(b_{i+1})) \end{aligned} \quad (5.1c)$$

- Le coût à long terme du problème total :

$$J_1(b_1) = \min_{b_2 \in \mathcal{A}_2(b_1)} (T_1(b_1, b_2) + J_2(b_2)) \quad \text{si } b_1 = t_s < 1 \quad (5.1d)$$

$$J_1(b_1) = \min_{b_2 \in \mathcal{A}_2(b_1=1)} (T'_{opt}(b_1, b_2) + J_2(b_2)) \quad \text{si } b_1 = t_s = 1 \quad (5.1e)$$

Appliquer le principe d'optimalité revient donc à démontrer le résultat suivant :

$$J_{tot} = J_1(t_s) \quad \text{où on considère le problème } \mathbf{Pav} \text{ si } t_s < 1 \quad (5.2)$$

et le problème  $\mathbf{Pav}'$  si  $t_s = 1$

On note que la variable de dimensionnement  $z_i$  qui détermine la décision sur la taille de l'aire de stockage est calculée implicitement par le calcul des variables  $b_i$ .

La taille de l'espace d'état du  $i$ ème sous-problème est défini par la discrétisation de sa variable de décision  $b_i$ . Pour avoir une précision au millièm de cette variable, il faut donc la discrétiser en mille états à visiter. Plus on exige une précision importante et plus l'espace d'état est important, déterminant ainsi le nombre d'états à visiter par le programme.

### 5.2.1 Programmation du problème d'optimisation

L'algorithme pour résoudre le problème d'optimisation reprend les équations de Bellman (5.1) et construit les sous-problèmes à optimiser les uns à la suite des autres :



```

Calculer  $J_m$ ;  $l=m-1$ ;
while  $l>0$ 
  for  $i > d/k(l)$ 
    for  $j$  in  $A(i)$ 
      Calculer  $T_1(i,j)+J_{l+1}(j)$ ; % On calcule le coût du sous-problème
    end
     $J_l(i)=\min_j(T_1(i,j)+J_{l+1}(j))$ ; % On construit la fonction coût
                                         % du prochain sous-problème
  end
   $l=l-1$ 
end

```

On retrouve ensuite les valeurs optimales des variables de décision, que l'on appelle encore trajectoire d'état optimale (séquence des  $b_i$ ), et à posteriori, la séquence des décisions sur le dimensionnement des  $z_i$ , en repérant l'état associé au minimum des coûts de chaque sous-problème (en partant de la contrainte  $t_s$ ). Le principe d'optimalité de Bellman assure que le coût optimal du problème est déterminé par les solutions optimales des sous-problèmes emboîtés. On peut donc construire la ligne de production à partir des variables de décision obtenues  $b_i$  et déterminer ainsi la taille optimale des aires de stockage et le coût global de stockage des en-cours de production en fonction de la variable de contrainte  $t_s$ .

La construction du problème d'optimisation suit la méthode suivante :

Le premier sous-problème est représenté par la machine  $M_m$  et l'aire de stockage  $X_m$ . L'ensemble des politiques admissibles est de taille  $(1 - \frac{d_{moy}}{k_m})/e$ , où  $e$  est la valeur d'échantillonnage de la variable de décision  $b_m$ . De cette manière, on construit la fonction coût  $J_m(b_m)$ .

Le deuxième sous-problème est l'assemblage de la machine  $M_{m-1}$  et de l'aire de stockage  $X_{m-1}$  au premier sous-problème. L'ensemble des politiques admissibles est lui de taille  $(1 - \frac{d_{moy}}{k_{m-1}})/e$ . On construit la fonction coût  $J_{m-1}(b_{m-1})$  en choisissant pour chaque  $b_{m-1}$  le minimum de la fonction  $T_{m-1}(b_{m-1}, b_m) + J_m(b_m)$  sur les différents  $b_m$  admissibles dans l'ensemble des contraintes  $\mathcal{A}_m(b_{m-1})$ .

Et ainsi de suite, le  $i$ ème sous-problème est construit par l'assemblage de la machine  $M_i$  et de l'aire de stockage  $X_i$  au précédent sous-problème. L'ensemble des politiques admissibles est de taille  $(1 - \frac{d_{moy}}{k_i})/e$ . On construit la fonction coût  $J_i(b_i)$  en choisissant pour chaque  $b_i$  le minimum de la fonction  $T_i(b_i, b_{i+1}) + J_{i+1}(b_{i+1})$  sur les différents  $b_{i+1}$  admissibles dans l'ensemble des contraintes  $\mathcal{A}_{i+1}(b_i)$ .

On trouve ainsi la fonction optimal du système complet en fonction de  $b_1$ , ou encore de la contrainte  $t_s$  imposée aux industriels, une fois que l'on a considéré dans le sous-problème global la machine  $M_1$  et l'aire de stockage  $X_1$ .

### 5.3 Résultats de l'optimisation

Dans le cas où les valeurs choisies sont les mêmes que celles de la simulation (voir (4.6)), le temps moyen de bon fonctionnement des machines est de 90% du temps. On choisit, dans un premier temps, un taux de service égal à 95% en entrée de ligne.

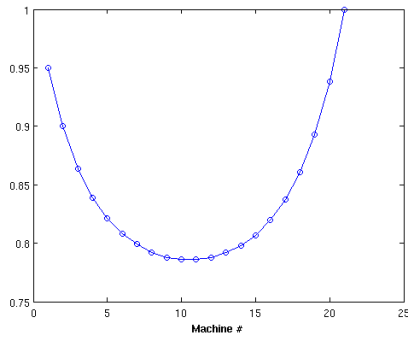


FIGURE 5.1 Disponibilité optimale des aires de stockage

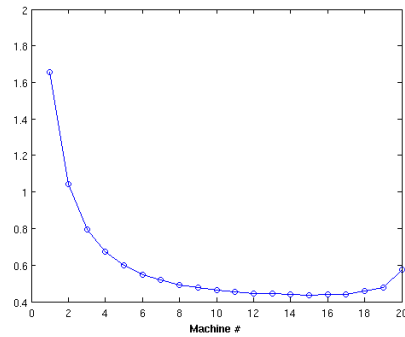


FIGURE 5.2 Aire de sotckage optimale

La répartition des valeurs optimales de la disponibilité des aires de stockage suit une courbe convexe et régulière, ainsi que la répartition des valeurs optimales des tailles des aires de stockage.

Dans un deuxième temps, on garde les mêmes paramètres mais cette fois-ci on choisit un taux de service en entrée égal à 1, c'est-à-dire qu'on résout le problème **Pav'** pour lequel une aire de stockage supplémentaire est mise à disposition et empêche l'interruption de l'approvisionnement. Le coût de cette aire de stockage supplémentaire est dix fois supérieur à celui des autres aires de stockage.

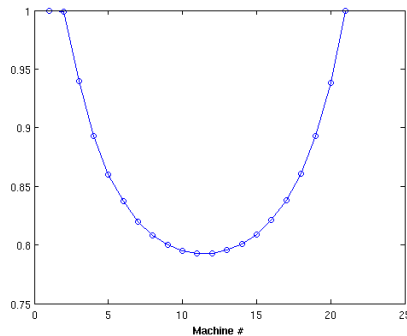


FIGURE 5.3 Disponibilité optimale des aires de stockage

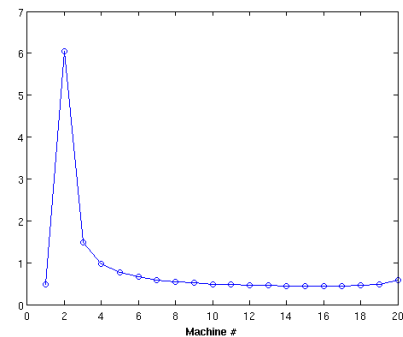


FIGURE 5.4 Aires de stockage optimales

On observe que la courbe reste toujours convexe et régulière d'un bout à l'autre de la ligne de production. Le coût de stockage supplémentaire étant dix fois supérieur au coût de stockage dans la ligne, on remarque que le coût minimal est atteint lorsque la disponibilité de la deuxième station de travail est aussi proche de 1 que possible, ceci en vue d'augmenter la capacité de stockage de la deuxième aire. En effet, il est plus intéressant de stocker dans la ligne un maximum d'inventaires à un coût moindre que de payer un coût de stockage en entrée de ligne dix fois plus important. Ce coût aurait pu être bien moins important mais il a été choisi de manière à amplifier le phénomène observable. Pour un coût moins important, la deuxième aire de stockage devient aussi moins importante comme on peut s'y attendre.

La courbe suivante montre le coût global de la ligne en fonction du taux de service choisi en entrée.

Le coût global de la ligne est croissant en fonction du taux de service choisi en début de ligne (figure C.11). Ce taux de service de la ligne de production peut aussi être interprété comme l'efficacité du système. En effet, on compare la production moyenne de ce système en fonction de l'approvisionnement imposé en début de ligne.

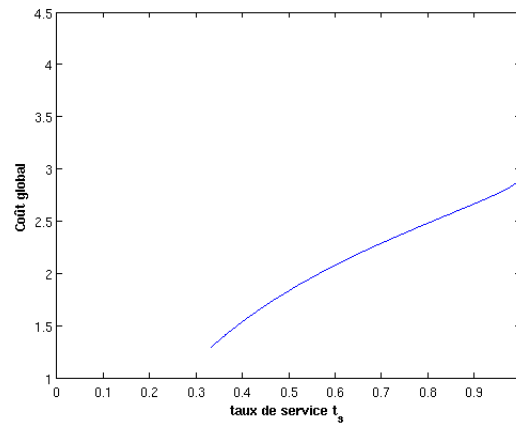


FIGURE 5.5 Coût global de la ligne en fonction du taux de service

On observe une discontinuité pour passer à un taux de service égal à 1. Cette différence est le coût à payer pour garantir à son fournisseur qu'il ne manquera jamais d'espace pour déposer les matériaux qu'il apporte. Ainsi, l'industriel est en mesure de prendre une décision chiffrée en fonction des contraintes de l'environnement dans lequel il évolue.

Dans le cas suivant, on choisit les mêmes valeurs que (21) pour observer le comportement dual, à savoir :

$$\begin{aligned} \forall i \in [1, m] \\ r_i = 0.5 \quad p_i = 0.2 \quad k_i = 2.5 \quad c_i = 1 \quad d_{moy} = 1 \end{aligned} \quad (5.3)$$

Pour ce cas, les machines travaillent correctement environ 71% du temps contre 90% du temps dans le cas précédent.

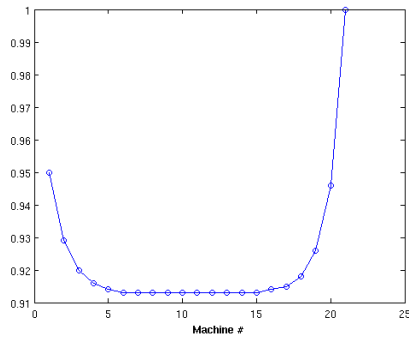


FIGURE 5.6 Disponibilité optimale des aires de stockage

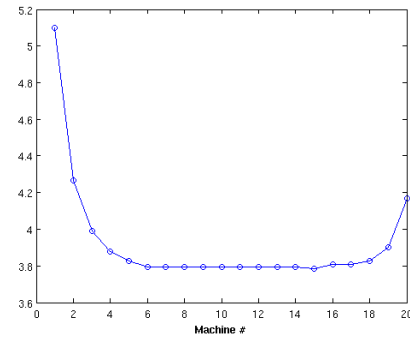


FIGURE 5.7 Aire de sotckage optimale

On observe, au centre de la ligne, que la diponibilité des aires de stockage atteint un seuil minimum où elle se stabilise. Ce seuil est la valeur qui minimise la fonction  $T_i(b_i, b_i)$  et représente le coût minimal atteignable en respectant les contraintes d'existence des solutions. En quelque sorte, la trajectoire d'état optimale décrite par les variables de décisions se rapprochent de cette valeur aussi rapidement que les contraintes d'existence d'une solution le permettent.

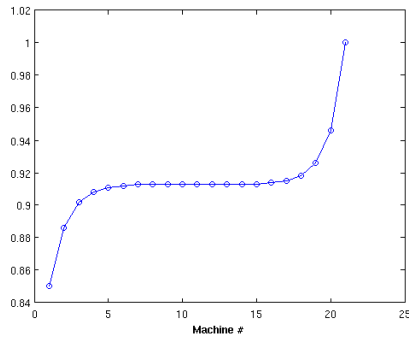


FIGURE 5.8 Disponibilité optimale des aires de stockage

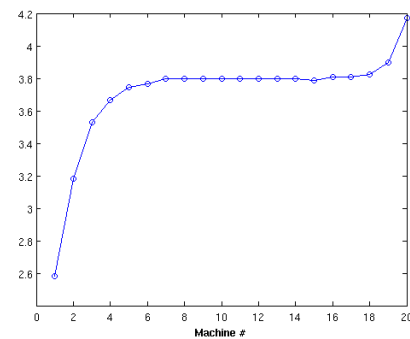


FIGURE 5.9 Aires de stockage optimales

En choisissant un taux de service plus faible que cette valeur, il en résulte le même phénomène mais cette fois inversée. Les valeurs de disponibilité des aires de stockage se rapprochent aussi rapidement du minimum atteignable que les contraintes lui permettent de le faire ; il s'agira cette fois d'une courbe croissante (figure 5.8). Finalement, plus une ligne de production est composée de machines en série, plus ces phénomènes

de début et de fin de ligne peuvent être considérés comme des phénomènes de bord négligeables.

Enfin, dans le cas où l'on exige qu'il n'y ait jamais de blocage de l'approvisionnement en début de ligne, les phénomènes observés s'accumulent.

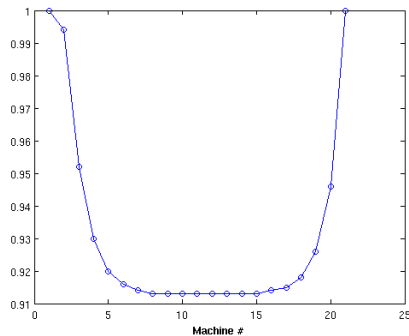


FIGURE 5.10 Disponibilité optimale des aires de stockage

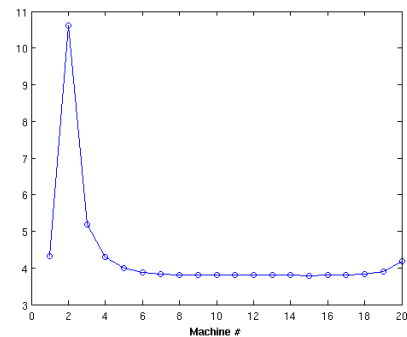


FIGURE 5.11 Aire de stockage optimale

Finalement, on remarque que les courbes suivent la même tendance que dans le cas dual de (21).

Le coût global de la ligne suit le même comportement. On observe cependant que le coût global de la ligne est cette fois beaucoup plus important. Cela s'explique par le fait que le temps de fonctionnement des machines est beaucoup plus faible, que les cadences de production sont moins importantes et qu'il est donc nécessaire d'augmenter les inventaires dans la ligne pour assurer un même niveau de service.

Les courbes que l'on vient de montrer ont été déterminées dans le cas où les  $k_i$  sont tous égaux. Cela ne tient pas compte des hypothèses faites sur la ligne de production mais cela permet d'expliquer les phénomènes observés après optimisation du coût de stockage des en-cours de production sans que les paramètres choisis n'influencent les observations. Dans les deux cas suivants, on fait varier les paramètres  $p_i$  et  $c_i$  pour observer quelles conséquences ils ont sur la répartition de la disponibilité optimale dans la ligne de production. La taille des aires de stockage optimale suit la même tendance que la disponibilité des aires de stockage.

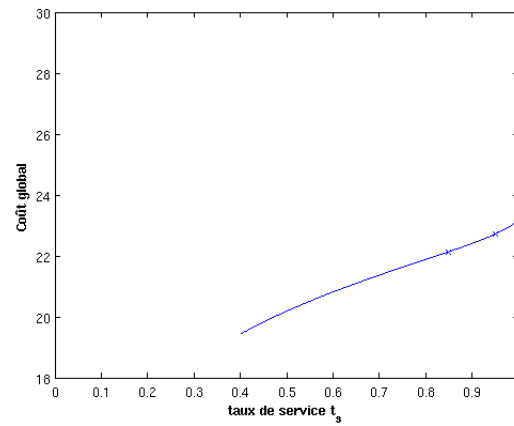


FIGURE 5.12 Coût global de la ligne en fonction du taux de service

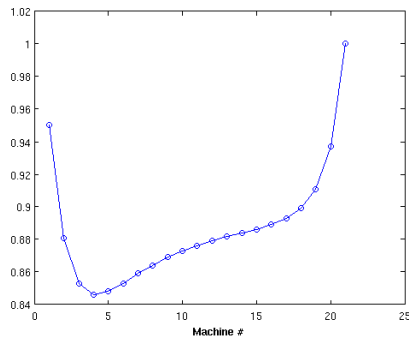


FIGURE 5.13 Disponibilité optimale des aires de stockage

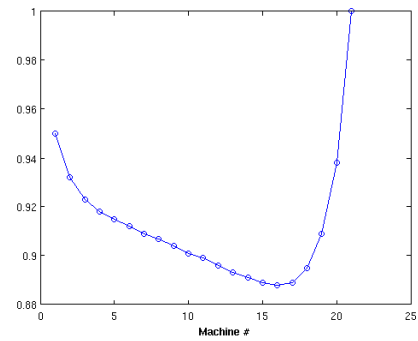


FIGURE 5.14 Disponibilité optimale des aires de stockage

Dans la figure 5.13, les paramètres  $p_i$  des lois exponentielles des temps de bon fonctionnement des machines vont en décroissant. Les machines en aval ont des temps moyens de bon fonctionnement qui sont plus importants que les machines qui sont en amont.

Dans la figure 5.14, le coût de stockage des en-cours de production est croissant dans la ligne. On observe alors que la disponibilité des aires de stockage suit elle une tendance décroissante, qui se répercute aussi sur la taille des aires de stockage. Le coût de stockage des aires de stockage étant plus important en fin de ligne, il est

économiquement plus intéressant de stocker des inventaires en début de ligne.

## 5.4 Conclusion

Au final, on observe que la répartition optimale des variables de la ligne de production suit une logique aussi bien économique que physique et que l'on observe des comportements en bout de ligne qui sont plus ou moins importants suivant les paramètres de la ligne étudiée et la longueur de la ligne. Les tendances des courbes obtenues sont identiques à celles obtenues précédemment dans le cas dual de (21). On obtient au final une méthode d'optimisation robuste qui permet de dimensionner une ligne de production à un coût minimal et pour lequel un taux de service en début de ligne peut être préalablement assuré.



# Chapitre 6

## CONCLUSION

La ligne de production que l'on vient de modéliser est caractérisée par des cadences de production croissantes dans la ligne de production et des paramètres identiques pour les temps de mauvais fonctionnement des machines. L'approvisionnement en entrée peut être assuré avec un taux de rupture déterminé à l'avance et peut aussi être garanti sans rupture, selon le modèle que l'on adopte. Ce cas représente en plusieurs points le cas dual du problème traité par (21). La structure des cadences de production est inversée et on considère un approvisionnement constant plutôt qu'une demande constante. De ce fait, les phénomènes de blocage et de pénurie sont eux aussi inversés et se propagent dans la direction opposée. On ne modélise plus le blocage des machines par la pénurie de matière mais par le remplissage des aires de stockage. Finalement, le modèle de base a été modifié et de nouvelles équations ont été développées.

La méthode de décomposition et agrégation de la ligne de production reste inchangée par rapport à (21). Les étapes sont les mêmes mais n'ont cependant pas été appliquées au même problème. Les résultats obtenus par la simulation permettent donc de conclure à nouveau que la méthode utilisée est précise et permet un dimensionnement correct des aires de stockage de la ligne de production. La méthode est donc validée dans un cadre expérimentale nouveau.

On note cependant que le modèle perd en précision pour une ligne de production composée de 10 machines. On explique cette imprécision par une sur-estimation du temps de bon fonctionnement des machines. Le modèle conduirait alors à prédire des inventaires moyens plus faibles et donc à augmenter l'écart avec les grandeurs simulées. Il est envisageable d'introduire un terme correctif pour pallier cet écart.

Une extension a été apportée au modèle en considérant une aire de stockage sup-

plémentaire en entrée du système. Celle-ci permet de garantir qu'il n'y ait pas de rupture d'approvisionnement. Le dual exact du modèle développé par (2) et utilisé par (21) n'a pas été utilisé car les excursions négatives n'auraient pas d'explications industrielles dans le cas que l'on a étudié. Anticiper la production par le maintien d'une quantité de stock toujours disponible n'est pas une solution retenue pour cette étude.

Au final, l'étude de la ligne de production permet d'obtenir une courbe qui décrit le coût global de stockge des en-cours de production en fonction de l'efficacité attendue du système. Cette courbe est aussi valide dans le cas où l'on demande une efficacité de 100% de la ligne de production. Optimiser le coût de stockage de la ligne a aussi permis de remarquer que les taux de diponibilité des aires de stockage sont déterminés de telle sorte qu'ils vérifient les contraintes d'existence des solutions tout en se rapprochant du minimum atteignable.

Les possibilités de développement du modèle sont nombreuses :

On peut s'intéresser à modéliser en bout de ligne l'extraction de la production. Celle-ci peut être représentée par différents modèles selon que les retards de livraison sont autorisés ou non. Cette modélisation pourrait alors être à nouveau utilisée dans le cas initial de (21) et pourrait permettre de représenter des lignes de production en flux poussé ou en flux tiré pour lesquelles l'approvisionnement et une demande constante sont imposés en même temps à la ligne de production.

Une autre possibilité serait de considérer une autre politique de gestion des ressources qui permettrait de considérer une architecture "CONWIP" (CONstant Work In Process) tel que cela a été développé par (19) avec le modèle initial en flux tiré. Il est aussi possible de considérer cette ligne de production dans le cas où différentes pièces peuvent passer dans la ligne.

Finalement, d'autres évolutions pourrait être envisagées pour des systèmes de production qui seraient composés de stations d'assemblage et de désassemblage. Le système ne serait alors plus une ligne de production mais un ensemble de plusieurs lignes de production qui pourraient se rejoindre ou se séparer à certains endroits de

la ligne. Tant que le phénomène de blocage reste unidirectionnel dans le système de production, il semble possible d'utiliser la méthode d'agrégation des états de fonctionnement des machines du système qui a été développée par (21) et permet d'optimiser le coût global du système considéré.

# Références

- [1] ARTAMONOV, G. T. (1976). Productivity of a two-instrument discrete processing line in the presence of failures. *Cybernetics and Systems Analysis*, 12, 464–468.
- [2] BIELECKI, T. et KUMAR, P. (1988). Optimality of zero-inventory policies for unreliable manufacturing systems. *Operations Research*, 36, 532–541.
- [3] BURMAN, M., GERSHWIN, S. B. et SUYEMATSU, C. (1998). Hewlett-packard uses operations research to improve the design of a printer production line. *Interfaces*, 28, 24 – 36.
- [4] BUZACOTT, J. A. (1967). Automatic transfer lines with buffer stocks. *International Journal of Production Research*, 5, 183 – 200.
- [5] BUZACOTT, J. A. et HANIFIN, L. E. (1978). Models of automatic transfer lines with inventory banks a review and comparison. *AIIE Transactions*, 10, 197 – 207.
- [6] CHIANG, S.-Y., HU, A. et MEERKOV, S. M. (2008). Lean buffering in serial production lines with nonidentical exponential machines. *IEEE Transactions on Automation Science and Engineering*, 5, 298–306. Exponential machines ;Lean buffering ;Serial production lines ;.
- [7] CHIANG, S.-Y., KUO, C.-T. et MEERKOV, S. (1998). Bottlenecks in markovian production lines : A systems approach. *IEEE Transactions on Robotics and Automation*, 14, 352–359. Markovian production lines ;.
- [8] CHIANG, S.-Y., KUO, C.-T. et MEERKOV, S. M. (2000). Dt-bottlenecks in serial production lines : theory and application. *IEEE Transactions on Robotics and Automation*, 16, 567–580. Aggregation technique ;Automotive component plant ;Down time bottlenecks ;Production line ;.
- [9] CHING, S., MEERKOV, S. M. et ZHANG, L. (2008). Assembly systems with non-exponential machines : Throughput and bottlenecks. *Nonlinear Analysis, Theory, Methods and Applications*, 69, 911–917. Assembly systems ;Bottlenecks ;Finite buffers ;Unreliable machines ;.

- [10] DUBOIS, D. et FORRESTIER, J. (1982). Productivité et encours moyens d'un ensemble de deux machines séparées par une zone de stockage. *RAIRO Automatique AP II*, 16, 105–132.
- [11] GERSHWIN, S. et SCHICK, I. (1978). Analytic methods for calculating performance measures of production lines with buffer storages. *Proceedings of the IEEE Conference on Decision and Control*, 618–624.
- [12] GERSHWIN, S. B. (1987). An efficient decomposition method for the approximate evaluation of tandem queues with finite storage space and blocking. *Operations Research*, 35, 291–305.
- [13] GERSHWIN, S. B. (1994). *Manufacturing Systems Engineering*. Prentice-Hall, INC., Englewood Cliffs New Jersey.
- [14] GERSHWIN, S. B. et SCHICK, I. C. (1983). Modeling and analysis of three-stage transfer lines with unreliable machines and finite buffers. *Operations Research*, 31, 354–380.
- [15] HU, J.-Q. (1995). Production rate control for failure-prone production systems with no backlog permitted. *Automatic Control, IEEE Transactions on*, 40, 291–295.
- [16] JINGSHAN, L. et MEERKOV, S. M. (2008). *Production Systems Engineering*. Springer.
- [17] KOSTER, M. D. (1987). Estimation of line efficiency by aggregation. *International Journal of Production Research*, 25, 615–626.
- [18] MALHAMÉ, R. et BOUKAS, E.-K. (1991). A renewal theoretic analysis of a class of manufacturing systems. *Automatic Control, IEEE Transactions on*, 36, 580 – 587.
- [19] MHADA, F. et MALHAMÉ, R. (2010). Approximate performance analysis of conwip disciplines in unreliable non homogeneous transfer line. *Annals of Operations Research*, 182, 213 – 233.
- [20] SADR, J. (2005). *Ligne de production non fiable ...* Thèse de doctorat, École Polytechnique de Montréal.
- [21] SADR, J. et MALHAMÉ, R. (2004). Decomposition/aggregation-based dynamic programming optimization of partially homogeneous unreliable transfer lines. *Automatic Control, IEEE Transactions on*, 49, 68 – 81.

- [22] SADR, J. et MALHAMÉ, R. (2004). Unreliable transfer lines : decomposition/aggregation and optimization. *Annals of Operations Research*, 125, 167 – 190.
- [23] SADR, J. et MALHAMÉ, R. P. (2004). A characterization of the maximum throughput of unreliable transfer lines. Nassau, Bahamas, vol. 3, 2566–2571. Decomposition techniques ; Iterated demand algorithm (IDA) ; Throughput maximization ; Unreliable transfer lines ;.
- [24] SHESKIN, T. J. (1976). Allocation of interstage storage along an automatic production line. 8, 146–152.
- [25] SOYSTER, A., SCHMIDT, J. et ROHRER, M. (1979). Allocation of buffer capacities for a class of fixed cycle production lines. *AIIE Transactions (American Institute of Industrial Engineers)*, 11, 140–146.
- [26] SOYSTER, A. et TOOF, D. (1976). Some comparative design aspects of fixed cycle production systems. *Naval Research Logistics*, 23, 437–454.
- [27] TERRACOL, C. et DAVID, R. (1987). An aggregation method for performance evaluation of transfer lines with unreliable machines and finite buffers. *Robotics and Automation. Proceedings. 1987 IEEE International Conference on*. vol. 4, 1333 – 1338.
- [28] TERRACOL, C. et DAVID, R. (1987). Performances d’une ligne composée de machines et de stocks intermédiaires. *Automatique, productique, informatique industrielle*, 21, 239–262.
- [29] WELCH, P. (1983). The statistical analysis of simulation results. *In The computer modeling handbook*, ed. S. Lavenberg, 268–328.
- [30] ZIMMERN, B. (1956). Etude de la propagation des arrêts aléatoires dans les chaînes de production. *Revue de Statistique Appl.*, 4, 85–104.

# Annexe A

## Programme de simulation

### A.1 Main script

```
%-----
%--- Simulation ---
%-----
%-----
% L'idée est de générer les temps exponentiels de manière aléatoire
% puis de faire évoluer notre système à chaque fois qu'un évènement arrive.
% (un évènement est soit le changement de fonctionnement d'une machine
% soit un stock qui se remplit ou se vide)
%-----

% Ce script permet de simuler l'évolution des aires de stockage d'une ligne de m machines

clear all; format short; clc;

%% Paramètres de la simulation

% Il est nécessaire de déclarer global toutes les variables.

global m; m=5;          % nombre de machines (pour une seule machine, il faut faire des
                        % modifications
global k; k=linspace(3,4,m);
                        % cadence des machines
global d; d=1;          % taux d'approvisionnement en entrée et en sortie
global e; e=.00001;     % Précision pour les variables
r=.5*ones(1,m); p=.25*ones(1,m);
                        % paramètres des lois exponentielles de bon et mauvais fonctionnement de
                        % chaque machine
c=ones(1,m+1);          % cout des encours de fabrication de chacune des aires de stockage

T=100000;               % horizon temporelle de la simulation
a=10000;                % temps pour lequel les mesures sont biaisées
```

### %% Initialisation l'état initial de la ligne de production

```

Z=4*ones(1,m);
X=zeros(m*T,m+1); % J'alloue de la mémoire pour chacune des aires m+1 aires de stockages
X(1,:)=Z/2; % Les stocks au départ de la simulation

M=ones(1,m); % État des machines : 0 en panne, 1 en bon fonctionnement
O_bool=zeros(1,m+1); % État des aires de stockage.
Z_bool=zeros(1,m+1);

K1=round([d,k]/e)*e; % Cadences auxquelles on augmente les aires de stockage.
K2=round([k,d]/e)*e; % Cadences auxquelles on diminue les aires de stockage.
K_history=zeros(m*T,m+2); % Pour sauvegarder les cadences au cours du temps
K_history(1,:)=K1,K2(m+1)];

```

### %% Construction de l'horizon aléatoire (toutes les machines sont ON au départ)

```

%-----
% Il est possible d'utiliser le générateur Mersene Twister
% Pour l'implanter suivre la démarche suivante :
% 1- Télécharger le fichier twister.zip à l'adresse :
%      http://www.mathworks.com/matlabcentral/fileexchange/6614
% 2- Extraire ce fichier dans le dossier de travail courant.
% 3- Taper 'mex -setup' dans le prompt de Matlab et choisir l'option 2
% 4- Taper 'mex twister.cpp', ce qui crée un fichier twister.mex64. dans le répertoire courant
%      (compile twister.cpp à l'aide de gcc)
% 5- La commande twister peut maintenant remplacer la fonction rand
% 6- On utilise alors la fonction -log(twister(1,m))./p à la place de exprnd(1./p)
%-----
rand('state',sum(100*clock));

l=1; u=0;
while min(u)<T
    U(l,:)=exprnd(1./p); U(l+1,:)=exprnd(1./r); u=sum(U); l=l+2;
end
Usave=U; size_=size(U); % On crée une sauvegarde de la génération de l'horizon

```

### %% Construction de l'évolution des stocks

```

u_=U(1,:); % Cette fonction représente le temps de fonctionnement des machines
            avant un prochain changement d'état.
E=1; t=zeros(1,m*T); % E est le compteur d'évènement et t le temps écoulé

```



% Boucle secondaire de construction de l'évolution des stocks

while t(E)<T

```
[t1,p1]=min(u_); t1=round(t1/e)*e;
% le temps avant le prochain changement machine et la position de la
% machine qui changera.
[t2,p2]=min_non_neg((Z-X(E,:))./pos_ou_nul(K1-K2)); t2=round(t2/e)*e;
% Le temps avant que la prochaine aire de stockage se remplisse et la
% position de cette aire de stockage
[t3,p3]=min_non_neg(-X(E,:)./neg_ou_nul(K1-K2)); t3=round(t3/e)*e;
% De même pour une aire de stockage qui se vide

X(E+1,:)=X(E,:)+(K1-K2)*min([t1,t2,t3]);
% On met à jour les aires de stockage en fonction du premier événement
% qui arrive
```

% En fonction de l'évènement qui s'est produit on réalise certaines mises à jour

```
if t1<min([t2,t3]) % Si l'évènement est un changement d'état d'une machine
    u_=u_-t1; % On diminue le temps qui repère le prochain changement machine
    % (après cette action u_(p1)=0 )
    M(p1)=1-M(p1); % On met à jour l'état de la machine
    U(:,p1)=[U(size_(1)*(p1-1)+2:size_(1)*p1),0];
    % On diminue l'horizon aléatoire de cette machine du temps de l'état
    % écoulé.
    % Dans le cas d'une machine, il faut remplacer «,0» par «:0»
    u_(p1)=U(1,p1); % On met à jour le prochain temps avant un nouveau changement
    % de cette machine
    t(E+1)=t(E)+t1; A=1; % On avance dans le temps et on repère l'évènement qui s'est produit

elseif t2<min([t1,t3]) % Si l'évènement est une aire de stockage qui se remplit
    u_=u_-t2; % On diminue le temps qui repère le prochain changement machine
    % (après cette action u_>0 )
    Z_bool(p2)=1; % On met à jour l'état de l'aire de stockage qui se remplit
    X(E+1,p2)=Z(p2); % On s'assure que le stock est bien à Z plutôt qu'à Z-epsilon
    t(E+1)=t(E)+t2; A=2; % On avance dans le temps et on repère l'évènement qui s'est produit

elseif t3<min([t1,t2]) % Si l'évènement est une aire de stockage qui se vide
    u_=u_-t3; % Même remarque que précédemment
    O_bool(p3)=1;
    X(E+1,p3)=0;
    t(E+1)=t(E)+t3; A=3;

elseif t2==t3 % Si l'évènement est une aire de stockage qui se vide est une aire de
    % stockage qui se remplit , les deux en même temps.
    u_=u_-t2;
    Z_bool(p2)=1; X(E+1,p2)=Z(p2);
    O_bool(p3)=1; X(E+1,p3)=0;
    t(E+1)=t(E)+t2; A=4;
```

```

elseif t1==t2          % Si l'évènement est une machine qui change d'état de fonctionnement et
                        une aire de stockage qui se remplit

    u_=u_-t1;
    M(p1)=1-M(p1);
    U(:,p1)=[U([size_(1)*(p1-1)+2:size_(1)*p1]),0];
    u_(p1)=U(1,p1);
    Z_bool(p2)=1; X(E+1,p2)=Z(p2);
    t(E+1)=t(E)+t1; A=5;

elseif t1==t3          % Si l'évènement est une machine qui change d'état de fonctionnement et
                        une aire de stockage se vide

    u_=u_-t1;
    M(p1)=1-M(p1);
    U(:,p1)=[U([size_(1)*(p1-1)+2:size_(1)*p1]),0];
    u_(p1)=U(1,p1);
    O_bool(p3)=1; X(E+1,p3)=0;
    t(E+1)=t(E)+t1; A=6;

end
% On met à jour les variables du système avant la prochaine itération

for i=1:m+1
    if round(X(E+1,i)/e)~=0
        O_bool(i)=0;
    end
    if round((Z(i)-X(E+1,i))/e)~=0
        Z_bool(i)=0;
    end
end

[K1,K2]=cadences_de_production(K1,K2,M,O_bool,Z_bool,A,p1,p2,p3);
                        % On met à jour les cadences de production
K1=round(K1/e)*e; K2=round(K2/e)*e;
E=E+1; K_history(E,:)=[K1,K2(m+1)];

end          % Fin de la boucle secondaire

t([E+1:end])=[]; X([E+1:end],:)=[]; K_history([E+1:end],:)=[];
                        % On supprime tout l'espace alloué qui est inutile

```

### %% Observation des sorties

```

To=zeros(1,m+1); Tz=zeros(1,m+1);
% Ces variables servent à calculer les temps passés en 0 et en Z
for l=1:m+1 % On commence une boucle pour chaque aire de stockage
    i=a+1;
    while i<=E % On avance évènement par évènement
        if X(i,l)~=0 && X(i,l)~=Z(l)
            i=i+1;

        elseif X(i,l)==0 % Si on est à 0 on regarde combien de temps on y reste et on incrémente
            % To de ce temps
            j=i;
            while i<=E && X(i,l)==0
                i=i+1;
            end
            if i-1-j~=0
                To(l)=To(l)+t(i-1)-t(j);
            end
        elseif X(i,l)==Z(l) % De même que précédemment mais en Z
            j=i;
            while i<=E && X(i,l)==Z(l)
                i=i+1;
            end
            if i-j-1~=0
                Tz(l)=Tz(l)+t(i-1)-t(j);
            end
        end
    end
end
end

```

### %% Evolution de l'aire de stockage par unité de temps

```

Xt=zeros(T,m+1); Aired=zeros(T,m+1); Airee=zeros(E-1,m+1);
i=1; j=2;

while i<=T && j<=E
    if round(t(j)*100)/100 < i;

        Airee(j,:)=Airee(j-1,:)+(t(j)-t(j-1)).5*(X(j,:)+X(j-1,:));
        Xt(i,:)=X(j,:)+(K_history(j,[1:m+1])-K_history(j,[2:m+2]))*(i-t(j));
        Aired(i,:)=Aired(j,:)+(i-t(j)).5*(X(j,:)+Xt(i,:));
        j=j+1;

    if round(t(j)*100)/100 > i;
        i=i+1;
    end
end

```

```

elseif round(t(j)*100)/100 > i;
    Xt(i,:)=Xt(i-1,:)+( K_history(j-1,[1:m+1])-K_history(j-1,[2:m+2]));
    Airet(i,:)=Airet(i-1,:)+.5*(Xt(i,:)+Xt(i-1,:));
    i=i+1;

else
    Xt(i,:)=X(j,:);
    if i-1 < round(t(j-1)*100)/100;
        Airee(j,:)=Airee(j-1,:)+.5*(X(j,:)+X(j-1,:))*(t(j)-t(j-1));
        Airet(i,:)=Airee(j,:);
    else
        Airet(i,:)=Airet(i-1,:)+.5*(Xt(i-1,:)+Xt(i,:));
        Airee(j,:)=Airet(i,:);
    end
    j=j+1; i=i+1;

end
end
Po=To/(a-t(E)); Pz=Tz/(a-t(E));
% On calcule la proportion de temps passé dans ces états

Prod_moy=([t,0]*[zeros(1,m+2);K_history]-[t([1:end1]),0,0]*[K_history;zeros(1,m+2)])./t(end);
% On calcule la production moyenne empirique

b=ones(T-a,m+1); for i=1:T-a b(i,:)=Airet(a,:); end
c=ones(T-a,m+1); for i=1:m+1 c(:,i)=[1:T-a]; end
EXs=(Airet([a+1:T],:)-b)./c;
% On calcule l'espérance empirique

outc=[EXs(end,:);Po;Pz;Prod_moy([2:end])];
% On met les variables en forme pour mettre tout ça dans un fichier texte
% On écrit enfin tout cela dans un fichier

outcomes=fopen('outcomes.txt','w');
fprintf(outcomes,'% -30s','horizon de chaque simulation'); fprintf(outcomes,'% -20d\n',T);
fprintf(outcomes,'%s\n','-----');
fprintf(outcomes,'% -29s','Coût total simulé'); fprintf(outcomes,'% -10.4f\n',sum(EXs(:,[1:m])));
fprintf(outcomes,'%s\n','-----');
fprintf(outcomes,'% -11s','EX simulé'); fprintf(outcomes,'% -11s','Po simulé');
fprintf(outcomes,'% -11s','Pz simulé'); fprintf(outcomes,'% -20s\n','Prod moy simulée');
fprintf(outcomes,'%s\n','-----');
fprintf(outcomes,'% -10.4f % -10.4f % -10.4f % -10.4f \n',outc);
fprintf(outcomes,'%s\n','-----');
fclose(outcomes); type outcomes.txt,

```

## A.2 Fonctions utilisées

### Prendre le minimum non négatif

```
function [ mini_non_neg, position ] = min_non_neg( K_ )
% Cette fonction permet de trouver le minimum non nul d'un vecteur donné.
% On l'utilise avec la fonction pos_ou_nul

for i=1:length(K_)
    if K_(i)< 0
        K_(i)=Inf;
    end
end
[mini_non_neg,position]=min(K_);
end
```

### Prendre la partie négative ou nulle

```
function [ K ] = neg_ou_nul( K_ )
% Cette fonction permet de considérer seulement les valeurs négatives d'un vecteur donné.

for i=1:length(K_)
    if K_(i)>0
        K_(i)=NaN;
    end
end
K=K_;
end
```

### Prendre la partie positive ou nulle

```
function [ K ] = pos_ou_nul( K_ )
% Cette fonction permet de considérer seulement les valeurs positives d'un vecteur donné.

for i=1:length(K_)
    if K_(i)<0
        K_(i)=NaN;
    end
end
K=K_;
end
```

### Ajuster les taux de production

```

function [ K1_, K2_ ] = cadences_de_production( K1, K2, M, O_bool, Z_bool, A, B, C, D )
%La fonction cadence de production met a jour les cadences de production en
%fonction des évènements et de l'état actuel de la ligne de production
% M est l'état de fonctionnement des machines (opérateur booléen)
% O_bool répertorie les stocks vides (opérateur booléen)
% Z_bool répertorie les stocks pleins (opérateur booléen)
% A définit l'évènement qui s'est produit (1 pour un changement machine, 2 si un stock s'est remplit,
% 3 si un stock s'est vidé et 4,5,6 si deux évènements se sont produits en même temps)
% B, C, D définissent respectivement la position de la machine ou du stock qui vient de changer

global m; global k; global d;

K1_=K1; K2_=K2;          % On va considérer les changements de cadences par rapport aux
                           originales

```

```

%-----
%   EVENEMENT : une machine change d'état
%-----
if A==1 || A==5 || A==6      % 3 cas possibles

    if M(B)==0                % Cas où la machine est passée de l'état on à off

        K1_(B+1)=0; K2_(B)=0;
                                % La cadence de production de cette machine est donc nulle

        % Prise en compte des répercussions dans la ligne

        if B~=m && O_bool(B+1)
                                % On vérifie si la machine juste en aval avait ajusté sa cadence de
                                % production à cette machine. Si la réponse est oui on propage cette
                                % nouvelle valeur à toutes les machines qui s'étaient ajustées à elle.

            i=B+1;

            while i<=m && O_bool(i)
                                % Propagation: tant que les stocks sont vides, ces machines ont donc
                                % ajustées leur cadence.

                K2_(i)=0; K1_(i+1)=0;
                i=i+1;
                if i==m+1 && O_bool(m+1)
                                % Si la propagation s'est produite dans toutes la ligne, la demande n'est
                                % donc plus assurée et devient nulle. (ce cas apparaît par exemple
                                % quand toutes les machines ont ajusté leur production à
                                % l'approvisionnement en début de ligne et que la demande en fin de
                                % ligne est inférieure ou égale à cet approvisionnement)

                    K2_(m+1)=0;
                end

            end

        elseif B==m && O_bool(m+1)
                                % De même, si la propagation s'est produite dans toutes la ligne, la
                                % demande n'est donc plus assurée et devient nulle

                K2_(m+1)=0;
            end
        end
    end
end

```

```

if B~=1 && Z_bool(B)
    % On vérifie si la machine juste en amont avait ajusté sa cadence de
    % production à cette machine. Si la réponse est oui on vérifie la
    % propagation de cette ajustement de cadences dans la ligne.

    i=B-1;

    while i>=2 && Z_bool(i+1)
        % Propagation: tant que les stocks sont vides, ces machines ajustent donc
        % leur cadence à celle qui est à l'origine du changement

        K2_(i)=0; K1_(i+1)=0;
        i=i-1;

        if i==1 && Z_bool(2) && Z_bool(1)==0
            % Si le stock 1 n'est pas rempli, on continue l'approvisionnement
            K2_(1)=0; K1_(2)=0;
        elseif i==1 && Z_bool(2) && Z_bool(1)
            % Si le stock 1 est rempli, on arrête l'approvisionnement
            K2_(1)=0; K1_(2)=0; K1_(1)=0;
        end
    end

    end

elseif B==1 && Z_bool(B)
    % Dans le cas où le premier stock s'est rempli, on arrête alors
    % l'approvisionnement en début de ligne

    K1_(1)=0;
    end

elseif M(B)==1
    % Cas où la machine est passée de l'état off à on

    K1_(B+1)=k(B); K2_(B)=k(B);
    % La cadence de production est donc de nouveau l'originale.

    if O_bool(B)
        % Si le stock en amont est vide, alors la machine doit ajuster son taux de
        % production
        K1_(B+1)=K1(B); K2_(B)=K1(B);
    end

    if Z_bool(B+1)
        % Si le stock aval est plein, alors la machine doit arrêter sa production
        K2_(B)=K2(B+1); K1_(B+1)=K2(B+1);
    end
end

```



```

if B~=m && O_bool(B+1)
    % Si le stock en aval est vide, alors il y avait eu ajustement par
    % propagation de la production de la machine en aval.
    i=B+1;

    while i<=m && O_bool(i) && M(i) && Z_bool(i+1)==0
        % Propagation: tant que les stocks sont vides et que la machine est en état
        % de fonctionnement, ces machines ajustent leur cadence à celle qui
        % est à l'origine du changement (Si cette machine avait elle-même
        % ajustée son taux à une autre machine, cela aurait déjà été pris en
        % compte dans les conditions précédentes).
        K2_(i)=K2_(B); K1_(i+1)=K1_(B+1);
        i=i+1;

        if i==m+1 && O_bool(m+1) & K2_(m)~=0
            % On vérifie que la demande est exploitable à nouveau
            K2_(m+1)=d;
        end
    end

elseif B==m && O_bool(B+1)
    % Si le dernier stock est vide on regarde comment doit évoluer la
    % demande finale
    if K1_(m+1)==0
        K2_(m+1)=0;
    elseif K1_(m+1)~=0
        K2_(m+1)=d;
    end
end
end

```

```

if B==1 && Z_bool(B)
    % Si le premier stock est plein on regarde comment doit évoluer
    % l'approvisionnement en début de ligne
    if K1_(B+1)~=0
        K1_(1)=d;
    elseif K1_(B+1)==0
        K1_(1)=0;
    end

elseif B~=1 && Z_bool(B)
    % Si le stock en amont est plein, alors il y avait eu blocage par
    % propagation de la production des machines.

    i=B-1;

    while i>=1 && Z_bool(i+1) && M(i)
        % Propagation: tant que les stocks sont pleins et que la machine est en
        % état de fonctionnement, ces machines retrouvent leur cadence de
        % production, à moins que la machine qui a changé d'état a eu sa
        % production bloquée par exemple par un stock plein en amont.

        if K2_(B)==0
            K2_(i)=0; K1_(i+1)=0;

        elseif K2_(B)~=0
            K2_(i)=k(i); K1_(i+1)=k(i);
        end

        if i==1 && M(1) && Z_bool(1) & K2_(B)==0
            % De même pour le cas de la première machine.
            K1_(1)=0;
        elseif i==1 && M(1) && Z_bool(1) & K2_(B)~=0
            K1_(1)=d;
        end

        i=i-1;

    end

end

end

end

end

```

```

%-----
%   EVENEMENT : un stock se remplit
%-----

if A==2 || A==4 || A==5
    % Dans ce cas, la machine en amont est bloquée.
    if C~=1
        K2_(C-1)=K2_(C); K1_(C)=K2_(C);
    elseif C==1
        K1_(C)=K1_(C+1);
    end
end

%-----
%   EVENEMENT : un stock se vide
%-----

if A==3 || A==4 || A==6
    % Dans ce cas, la machine ajuste son taux de production, et possiblement
    % plusieurs d'entre elles.
    if D~=m+1
        K1_(D+1)=K1_(D); K2_(D)=K1_(D);
        i=D;

        while i<=m && O_bool(i) && M(i)
            K1_(i+1)=K1_(D); K2_(i)=K1_(D);
            if i==m && O_bool(m+1)
                K2_(m+1)=K1_(D);
            end

            i=i+1;
        end

        elseif D==m+1;
            K2_(D)=K2_(D-1); K1_(D)=K2_(D-1);
        end
    end

end % Fin de la fonction

```

### Déterminer la période transitoire du système

```

%-----
%--- Warm Up ---
%-----

clear all;

global T; T=100000;
global m; m=2;

n=20; w=round(T/4);
Aire=zeros(T,m+1,n);

for i=1:n
    [Aire(:,i),t]=Simul(T,m); % La fonction Simul extrait l'évolution des aires de stockage calculé dans
                                la simulation ainsi que l'évolution du temps par événements
end

Y=zeros(T,m+1,n); Y_=zeros(T,m+1); Yw=Inf*ones(T,m+1);
Y(1,:)=Aire(1,:); Y([2:T],:)=Aire([2:T],:)-Aire([1:T-1],:);

for l=1:m+1
    for i=1:T
        Y_(i,l)=sum(Y(i,l,:))/n; % On calcule Yi tel que définit
    end
end

for l=1:m+1
    for i=1:w
        Yw(i,l)=sum(Y_([1:2*i-1],l))/(2*i-1);
                                % On lisse cette fonction
    end
    for i=w+1:T-w
        Yw(i,l)=sum(Y_([i-w:i+w],l))/(2*w-1);
    end
end

plot(Yw); % On crée un graphe pour déterminer la période de lancement

```

# Annexe B

## Programme d'optimisation

### B.1 Main script

```
%-----
%---- Optimisation ----
%-----

clear all; format compact; clc;

%% Paramètres de la ligne à optimiser

e=.001;                % echantillonnage (l'espace d'état de chaque variable est
                        % de dimension 1/e)
m=5;                   % nombre de machines
ts=.95;                % taux de service en début de ligne (si ts=1, on considère le cas de
                        % Bielecki et Kumar inversé )
r=.9*ones(1,m); p=.1*ones(1,m);
                        % paramètres des lois exponentielles
k=linspace(3,3.8,m);   % taux de production des machines
c=ones(1,m);           % coût de stockage des en-cours de production
d=1;                   % Production moyenne imposée de la ligne

H=waitbar(0, 'En cours d'optimisation');

%% Réalisabilité

for i=1:m
    if k(i)*r(i)/(r(i)+p(i))<=d
        print(' On ne satisfait pas les conditions de realisabilite de production ')
    end
end
```

### %% Optimisation

```

u=Inf*ones(1/e,1/e,m-1);          % fonctions convexes    $T_i(b_i, b_{i+1}) + J_{i+1}(b_{i+1})$ 
v=Inf*ones(1/e,m);                 % fonctions croissantes  $J_i(b_i)$ 

                                % calcul du coût de la machine m en fonction de  $b_m$ 
for i=round(r(m)/(r(m)+p(m))/e):1/e
    v(i,m)=c(m)*T_bi_bii(r(m),p(m),k(m),d,i*e,1);
end

l=m-1;
while l>1                          % calcul du coût de stockage de chacun des ensembles

    for i=round(d/k(l)/e+1):1/e      % borne inf de  $b_i$  jusqu'à 1

        for j=round(d/k(l)*(r(l)+p(l))/r(l)/e+1):min([round(i*(r(l)+p(l))/r(l)),1/e])
            % borne inf de  $b_{i+1}$  (en fonction de  $b_i$ ) jusqu'à sa borne sup

            u(i,j,l)=c(l)*T_bi_bii(r(l),p(l),k(l),d,i*e,j*e)+v(j,l+1);
        end

        v(i,l)=min(u(i,:,l));
    end

    waitbar((m-l)/(m-1),H);
    l=l-1;
end

J1=Inf*ones(1/e);                  % fonctions coût de la dernière machine  $J_1(b_1=1)$ 

for j=round(d/k(1)*(r(1)+p(1))/r(1)/e):min([round((r(1)+p(1))/r(1)*(1-c(1)/c2*(k(1)-d)/k(1))/e),1/e])

    alpha=r(1)/d-(r(1)*(1-j*e)+p(1))/(k(1)-d);

    J1(j)=c(1)/alpha*(1+log(c2/c(1)*k(1)/(k(1)-d)*(1-r(1)/(r(1)+p(1))*j*e)))+v(j,2);

end

v(1/e,1)=min(J1(:));              % fonction coût globale pour  $t_s=1$ 

```

### %% Solutions

```

B_sol(1)=ts; % taux de service en debut et en fin de ligne

if ts==1
    [mini,pos]=min(J1(:)); B_sol(2)=pos*e;
    alpha=r(1)/d-(r(1)*(1-B_sol(2))+p(1))/(k(1)-d);
    z(1)=1/alpha*log(c2/c(1)*k(1)/(k(1)-d)*(1-r(1)/(r(1)+p(1))*B_sol(2)));
else
    [mini,pos]=min(u(round(B_sol(1)/e),:,1)); B_sol(2)=pos*e;
    z(1)=z_(r(1),p(1),k(1),d,B_sol(1),B_sol(2));
end

for l=2:m-1

    [mini,pos]=min(u(round(B_sol(l)/e),:,l)); B_sol(l+1)=pos*e;
    z(l)=z_(r(l),p(l),k(l),d,B_sol(l),B_sol(l+1));
end

cout_tot=v(round(B_sol(1)/e),1);
z(m)=z_(r(m),p(m),k(m),d,B_sol(m),1);

%% Observation des sorties

figure(1); plot([B_sol,1],'-ob'); xlabel('Machine #','FontWeight','bold')
figure(2); plot(z,'o-b'); xlabel('Machine #','FontWeight','bold')
figure(3); plot([e:e:1-e],v([1:1/e-1],1),'-b',1,v(1/e,1),'ob',ts,v(round(ts/e),1),'xb');
xlabel('taux de service t_s','FontWeight','bold') , ylabel('Coût global','FontWeight','bold')

outc=[[1:m];B_sol;z];
outcomes=fopen('outcomes.txt','w');
fprintf(outcomes,'%s','coût total pour la ligne'); fprintf(outcomes,' %22.4f\n\n',cout_tot);
fprintf(outcomes,'%s','stock # i');fprintf(outcomes,'\t%s','b_i correspondant');
fprintf(outcomes,'\t%s\n','z_i correspondant');
fprintf(outcomes,'%s\n','-----');
fprintf(outcomes,'%2d %19.4f %23.4f\n',outc);
fclose(outcomes); type outcomes.txt,

```

## B.2 Fonctions utilisées

### Calculer l'espérance du stock

```
function EX=T_bi_bii(r,p,k,d,bi,bii)
% Calcul de l'esperance des en-cours de production en fonction de bi et bii.
% r,p,k,d sont les paramètre de la machine associée à cette aire de stockage.
% bi est le coefficient de disponibilité de Xi et bii de Xi+1

mu=d/(bi*k-d)*(r*(1-bii)+p)/r/bii;
alpha=r*bi/d-(r*(1-bii)+p)/(bi*k-d)*bi/bii;
EX= 1/alpha*d/(k*bi-d)*(bi*k/d*(bi-r/(r+p)*bii)- (1-bi)/(1-mu)*(r+p)/r/bii*...
    log(mu/(1-bi)*(r/(r+p)*bii*bi*k/d-bi)));

end
```

La fonction coût possède une discontinuité au point  $\alpha_i = 0$  comme on l'a vu avec les équations (3.41). On ne prend pas en compte ces équations dans le programme mais on s'assure d'éviter ce point en choisissant bien les paramètres de la ligne. Par exemple, en choisissant les paramètres de l'optimisation (5.3), on a :

$$\alpha_i = 0 \quad \Longleftrightarrow \quad b_i b_{i+1} = \frac{d_{moy}}{k_i} \frac{r + p_i}{r} = 0,56 \quad (\text{B.1})$$

Si on choisit un échantillonnage précis au centième, on passe forcément par cette valeur et alors le programme d'optimisation est faussé. Pour éviter cette valeur on pourra par exemple modifier très faiblement un des paramètres pour s'assurer que ce point n'est pas atteignable par le choix de l'échantillonnage. Par exemple :

$$\alpha_i = 0 \quad \Longleftrightarrow \quad b_i b_{i+1} = \frac{d_{moy}}{k_i} \frac{r + p_i + 0.001}{r} = 0,56008 \quad (\text{B.2})$$

Cette valeur ne change pas vraiment les solutions du problème d'optimisation et permet en plus d'éviter le point de discontinuité de la fonction.



### Calculer la taille du stock

```

function Z = z_( r,p,k,d,bi,bii )
% Calculer le seuil limite de l'aire de stockage Xi en fonction des paramètres de la machine
% r,p,k,d sont les paramètres de la machine associée à cette aire de stockage.
% bi est le coefficient de disponibilité de Xi et bii de Xi+1

mu=d/(bi*k-d)*(r*(1-bii)+p)/r/bii;
alpha=r*bi/d-(r*(1-bii)+p)/(bi*k-d)*bi/bii;
Z=log(bi/(1-bi)*mu*(r/(r+p)*k/d*bii-1))/alpha;

end

```

# Annexe C

## Article pour soumission à IJPE

**Title** Serial Production Line Under Push Mechanism With Exponential Reliability ; Inventory Size & Cost Optimization

**Authors** Thibault Gosset, École Polytechnique de Montréal & Gerad  
 Roland Malhamé, École Polytechnique de Montréal & Gerad  
 Robert Pellerin, École Polytechnique de Montréal & Cirrelt

**Abstract** We consider a transfer line subject to a constant rate of demand for finished parts. This constant rate of finished parts extraction can be enforced either through a pull mechanism or a push (supply rate of raw material) mechanism. Sadr and Malhamé have developed a decomposition/aggregation methodology (45; 46) for the approximate performance analysis of transfer lines with rate enforcement achieved through a pull mechanism. The main advantage of this approximation scheme is that causality propagates unidirectionally (upstream to downstream), thus making it possible to use dynamic programming as an optimization tool for buffer sizing under a Kanban architecture. In this paper, we develop the dual of this decomposition/aggregation methodology, one for which the constant rate of production is enforced through a push mechanism. Causality propagates in this formulation from downstream towards upstream. The resulting approximate performance estimates (mean storage costs under a service level constraint) are validated against Monte Carlo simulations, and optimal buffering levels are calculated.

**Keyword** Production Line Decomposition/Aggregation Markov Chain Dynamic Programming

## C.1 Introduction

Production systems inspire many researchers to develop mathematical models that best predict delivery of materials and that help calculate more efficient production plans. In this area, the unreliability of work stations presents a major issue and an important work of literature has been written on best policies for piloting such systems (36; 39; 40; 44; 47), their supply (31; 40) or their demand (42). Equations describing inventories evolution and evaluating the relation between quality and quantity become very complex for systems composed of many work stations. A globally preferred method for approaching this issue is to consider a decomposition of these systems into several tractable ones (34; 35; 37; 38; 45; 46). Production phenomena involved in the system may thus be analyzed on a local scale ; inventories in the system may interfere with machine production (blocking or starvation) when they are full and/or empty. Such phenomena tend to propagate like a shock wave in the transfer line, both upstream and downstream the faulty machine and make the system difficult to analyse. As a result, in (34; 35; 38), a system of sequentially coupled non linear equations is developed and is solved through both forward and backward iterations to determine all machine blocking and starvation effects. Sadr and Malhamé (45; 46) have developed a decomposition method whose main advantage when compared to previous work in this area, is that under a given demand enforced by a pull mechanism (KANBAN discipline) and an assumption of a transfer line with monotone decreasing production capacities, the calculation of starvation and blocking effects can be carried out sequentially from upstream towards downstream, with no iterations required. Note that the capacity related assumption produces the desirable result that the maximum production rate of a given machine is never limited by that of the machine preceding it in the transfer line. The main benefit of calculations flowing unidirectionally is the ability of minimizing the cost of holding inventories in the system through a dynamic programming scheme, thus yielding the optimal buffer sizing policy. The possibility of a dual approach, one where calculations would propagate from downstream to upstream has long been suspected. It is the goal of this paper to carry out such an approach. Here, instead of a pull mechanism to enforce constancy of mean production rate, a push mechanism at the raw material supply level plays that role. Furthermore, the machine capacities instead of decreasing, are assumed to be monotone increasing from upstream to downstream.

The remainder of the paper is organized as follows. In section C.2, we derive the probabilistic evolution of two simple single machine systems which are the push dual of the Hu (pull) machine (41) , and an appropriately modified push dual version of the Bielecki and Kumar (pull) machine (32), respectively. In section C.3, we present the dynamic model of our unreliable transfer line operated in a push mode, we state the associated parameters assumptions, and formulate two versions of an associated optimal buffer sizing problem. In section C.4, we state the key simplifying assumptions of machine decoupling approximation and demand averaging principle of Sadr and Malhamé (45; 46). We reformulate these assumptions in our push context, and use them to achieve a dual line decomposition approach which culminates into two dynamic programming optimization problems corresponding to the two versions of the buffer sizing problem of Section C.3. Section C.5 is dedicated to a simulation based validation of the quality of our approximate decomposition, while Section C.6 presents a sample of dynamic programming based optimal buffering results. Section C.7 concludes this paper by discussing the main limitations of our approach as well as presenting some possible extensions.

## C.2 Building Blocks : Background Material

In this section, two models that will next be used for the decomposition of the line are developed. We consider a system composed of one machine,  $M$ , and an upstream storage area,  $X$ , which is finite for the first model and unlimited for the second (Figure C.1 ).

The machine spends a random amount of time in either one of two states denoted  $\alpha(t)$  : state 1 is when the machine is fully operational, while state 2 is the failure mode. An exponential distribution is attributed to those random times with parameters  $p$  for state 1 and  $r$  for state 0. Thus, the machine can be represented by a two-state Markov chain in continuous time,  $\{M(t), t \in \mathbb{R}^+\}$ . A constant supply  $d$  is applied to the system as an input as long as the buffer is not full and  $u(t)$  is the machine production rate.

The machine is assumed to produce at its capacity  $k > d$  whenever operational. Furthermore, backlogs are not allowed in this system and the machine must adjust its production rate to  $d$  when the inventory is empty. The evolution of the inventory

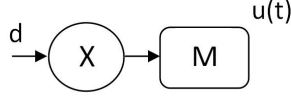


FIGURE C.1 Model of The Work Station With Upstream Buffer

is represented by the stochastic process in continuous time  $\{X(t), t \in \mathbb{R}^+\}$ . When the inventory is finite and equal to  $z$ , the process is ruled by the following differential equation :

$$\frac{dX(t)}{dt} = \begin{cases} d - u(t) & \text{if } \alpha(t) = 1 \text{ or } x(t) < z \\ 0 & \text{if } \alpha(t) = 0 \text{ and } x(t) = z \end{cases} \quad (\text{C.1})$$

Based on (43), it is possible to obtain as a special case the forward Kolmogorov equations associated with hybrid-state Markov process  $(\alpha(t), x(t))$ , together with boundary conditions, as follows :

$$\frac{\partial \vec{f}(x, t)}{\partial t} = -\mathbf{J} \frac{\partial \vec{f}(x, t)}{\partial x} + \mathbf{Q}^t \vec{f}(x, t) \quad (\text{C.2})$$

$$r\mathbf{P}_z(t) = df_{off}(z^-, t) = (k - d)f_{on}(x = z^-, t) \quad (\text{C.3})$$

$$p\mathbf{P}_o(t) = df_{off}(0^+, t) = (k - d)f_{on}(x = 0^+, t) \quad (\text{C.4})$$

where  $\mathbf{Q}$  is the matrix of transition rates of the two-state Markov chain,  $\mathbf{J} = \text{diag}(d, d - k)$ ,  $\vec{f}(\cdot, \cdot) = (f_{off}(\cdot, \cdot), f_{on}(\cdot, \cdot))^t$  and  $f_{off}(x, t)$  and  $f_{on}(x, t)$  are hybrid probability densities of states  $(0, x(t))$ , and  $(1, x(t))$  respectively, while  $f_X(\cdot, \cdot) = f_{on}(\cdot, \cdot) + f_{off}(\cdot, \cdot)$  is the density functions of  $X(t)$ .  $\mathbf{P}_o(t)$  and  $\mathbf{P}_z(t)$  are probability masses at time  $t$  that the inventory be empty or full, respectively. Those functions also verify the normalization condition :

$$\forall t \in \mathbb{R}^+ \quad \mathbf{P}_o(t) + \mathbf{P}_z(t) + \int_0^z f_X(x, t) dx = 1 \quad (\text{C.5})$$

The steady state distribution is extracted from this set of equations where partial derivatives with respect to time are set to zero in (C.2), and one solves a set of linear differential equations in the  $x$  variable with  $\mathbf{P}_o$  and  $\mathbf{P}_z$  constants, boundary conditions (C.3), (C.4), and normalization condition (C.5). It leads to the following solutions (see derivation in Appendix) :

$$\mathbf{P}_o = \frac{r}{r + p} \frac{1 - \mu}{1 - \mu e^{-\lambda z}} \quad (\text{C.6})$$

$$\mathbf{P}_z = \frac{p}{r+p} \frac{(1-\mu)e^{-\lambda z}}{1-\mu e^{-\lambda z}} \quad (\text{C.7})$$

$$f_X(x) = \frac{p}{r+p} \frac{k}{k-d} \frac{\lambda e^{-\lambda x}}{1-\mu e^{-\lambda x}} \quad (\text{C.8})$$

where  $\mu = \frac{d}{k-d} \frac{p}{r}$  and  $\lambda = \frac{r}{d} - \frac{p}{k-d}$

Different observations can be made in relation to the above equations.

- (i) The equations share great similarities with those in (41). Indeed, by temporarily associating an index  $s$  to quantities in the supply rate driven model, and  $d$  to quantities in the demand rate driven model, the following dual relationships can be established :

$$\mathbf{P}_{0,s} = \mathbf{P}_{z,d} \quad \mathbf{P}_{z,s} = \mathbf{P}_{0,d} \quad f_{X_s}(x) = f_{X_d}(z-x) \quad (\text{C.9})$$

$$E[X_s] + E[X_d] = z \quad F_{X_s}(x) = 1 - F_{X_d}(z-x) \quad (\text{C.10})$$

- (ii) If  $c$  is the storage cost per part, the inventory holding cost of the model,  $J_{tot}$ , can be obtained as follows :

$$J_{tot} = E[c.X] = c \left( \int_0^z x f_X(x) dx + z \mathbf{P}_z \right) \quad (\text{C.11})$$

- (iii) Since for any stable operation, the long term average supply must be equal to the long term production rate of the machine denoted  $d_{av}$ , it is possible to write :

$$d \times (1 - \mathbf{P}_z) = k \times \left( \frac{r}{r+p} - \mathbf{P}_o \right) + d \times \mathbf{P}_o = d_{av} \quad (\text{C.12})$$

Note that supply disruptions are allowed in this model configuration. This results into a net supply average,  $d_{av}$ , lower than the nominal supply rate,  $d$ , and a supply efficiency  $SE$  less than 1. More specifically :

$$SE = d_{av}/d = 1 - \mathbf{P}_z \quad (\text{C.13})$$

The choice of warehouse size  $z$  will have a direct influence on this efficiency. An infinite storage area would allow a 100% efficiency; however, all warehouse are finite. The second basic building block model considers the possibility of a supplementary storage area as an outsourced capacity whenever the need arises, i.e. when the dedicated storage area becomes full. In effect, the level of raw materials is no longer constrained and it is described by the following equations, for  $\lambda > 0$  and obtained from (C.6-C.8)

as  $z \rightarrow \infty$

$$\mathbf{P}_o \rightarrow \frac{r}{r+p} (1 - \mu) \quad \mathbf{P}_z \rightarrow 0 \quad (\text{C.14})$$

$$f_X(x) \rightarrow \frac{p}{r+p} \frac{k}{k-d} \lambda e^{-\lambda x} \quad (\text{C.15})$$

where,  $\mu = \frac{d}{k-d} \frac{p}{r}$  and  $\lambda = \frac{r}{d} - \frac{p}{k-d}$

In this model the holding cost of raw materials is defined as proportional to the fixed size  $z$  of the locally allocated storage space, and proportional to mean size of raw material overflows soliciting the outsourced capacity at respective per part costs  $c$  and  $c'$ . As a result, the total cost verifies :

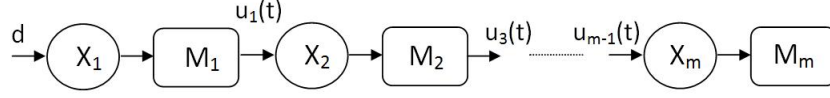
$$J_{tot} = cz + c' \left( \int_z^\infty (x - z) f_X(x) dx \right) \quad (\text{C.16})$$

These two models are designed as building blocks within the decomposition method of the production line described in the next section. Two models are considered ; one for which supply disruptions are tolerated, and one for which they are not.

### C.3 Further modeling assumptions

We consider a production line composed of  $m$  machines  $M_i$ ,  $i = 1..m$ . Machine  $M_i$  is assumed to evolve according to a two-state continuous time Markov chain labeled  $\alpha_i(t)$ . This process  $\alpha_i(t)$  has value 1 when the machine is operational and value 0 when the machine is failed. Given that the machine is operational, its failure rate is  $p_i$  whereas when failed, the associated repair rate is  $r_i$ . Storage areas,  $X_i$ , and maximum size  $z_i$  are introduced in the line and associated to each machine,  $M_i$  (Figure C.2). These areas make it possible for machines to continue working, at least for a while, whenever one or more neighboring machines fail, thus enhancing the overall productivity of the line. Then, storage areas are similarly denoted  $B_i$ , standing for "buffer", where  $X_i$  would referably be used for denoting the instantaneous level of the storage area. Each machine has a nominal production rate,  $k_i$ , and an instantaneous production rate,  $u_i(t)$ .

We consider that the finished products buffer is infinite and thus the last machine,  $M_m$ , is never blocked. Furthermore nominal production rates are monotone increasing

FIGURE C.2  $m$  Machines Production Line

throughout the line.

$$k_1 \leq k_2 \leq \dots \leq k_m \quad (\text{C.17})$$

Assumption (C.17) is key to building our decomposition method. Indeed, because  $k_m \geq k_{m-1}$ , starting downstream from buffer  $B_m$ , it can never become full unless machine  $M_m$  - itself assumed never blocked - is down, in which case machine  $M_{m-1}$  production is stopped. In turn,  $k_{m-1} \geq k_{m-2}$  implies that buffer  $B_{m-1}$  can never become full, unless machine  $M_{m-1}$  is down or effectively at zero production because of blocking. As a consequence, from the point of view of inventory  $X_m$ , if it is away from zero, the rate of parts extraction is entirely dictated by the state of machine  $M_m$ , i.e.  $k_m$  if  $M_m$  is operational and not blocked, or zero otherwise. The same argument can be applied to buffer  $B_{m-1}$  in the sense that if it is away from zero, the rate of parts extraction will be entirely dictated by the status of machine  $M_{m-1}$ , i.e. it is  $k_{m-1}$  if  $M_{m-1}$  is operational and not blocked, or zero otherwise. This argument can be propagated recursively proceeding from downstream towards upstream of the transfer line, and the definition of the production rate (C.18) follows, showing the unidirectional propagation of the phenomena (starvation and blocking) :

$$u_i(t) = \begin{cases} \begin{cases} u_{i-1}(t) & \text{if } X_i = 0 \\ 0 & \text{if } X_{i+1} = z_{i+1} \\ k_i & \text{otherwise} \end{cases} & \text{if } M_i(t) = 1 \\ 0 & \text{if } M_i(t) = 0 \end{cases} \quad (\text{C.18})$$

Together with the so-called demand averaging principle, an approximation to be discussed in Section C.4, these observations will allow to conclude that under Assumption (C.17), line performance evaluation for a given choice of buffer sizes can proceed unidirectionally from downstream towards upstream (unlike most other decomposition methods (34; 35; 37; 38) ). In effect, it is as though under Assumption (C.17) together with an externally imposed line production rate, there were in effect, a unidirectional



propagation of causality.

Recall that in this set up, the enforcement of overall line production rate is achieved through a push mechanism at supply. This is to be contrasted with the pull mechanism invoked instead by Malhamé and Sadr (45; 46) for enforcing the required production rate in which case, Assumption (C.17) had instead to correspond to monotone decreasing machine capacities, and calculations proceeded in the opposite direction. Buffer sizes play a fundamental role in affecting blocking probabilities of machines (and indirectly their related starvation probabilities). This in turn affects the minimal buffering requirements to secure a given production rate through the line.

In the next section we review approximation principles first invoked in (45; 46), which will permit the unfolding of our dual decomposition method.

## C.4 The Decomposition Approach

In general, a production line does not operate as efficiently as machines in isolation. Thus, the models described in Section C.2 cannot be directly used as building blocks in the transfer line without somehow accounting for the loss of productivity caused by their mutual interferences in the line. The supply input of the inventory work in process (wip)  $X_i$  is a stochastic flow of materials,  $u_{i-1}(t)$ , and in view of Assumption (C.17), the production of machine  $M_i$  must stop whenever the buffer immediately downstream,  $X_{i+1}$ , is full. Thus, the possibility of production for machine  $M_i$  can be represented by a binary stochastic process,  $\{I_{i+1}(t), t \in \mathbb{R}^+\}$ , equal to 1 whenever  $X_{i+1}$  is not full and 0 otherwise.  $I_{i+1}(t)$  will be called the buffer availability process. Each machine, together with its upstream inventory is then represented by the model in the Figure C.3, with the conventions  $u_0(t) = d$  (the system input supply), and  $I_{m+1}(t) = 1$  (machine  $M_m$  is never blocked).  $I_{i+1}(t)$  acts like a valve allowing or disallowing the flow of material out of  $M_i$ .

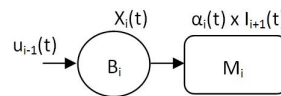


FIGURE C.3 Model of The  $i$ th Work Station of The Line

In the next subsections, we first revisit two approximations; the machine decoupling approximation and the demand averaging principle in (45; 46). The latter is being renamed in this context the supply averaging principle. These subsequently allow us, under Assumption (C.17) to represent a generic machine  $M_i$  within the transfer line by a two-state virtual isolated machine, with productivity adequately weakened so as to account for the interference effects of surrounding buffers. The parameters of the virtual  $M_i$  machines are recursively calculated starting from machine  $M_m$  and moving backwards upstream. Inventory steady-state distributions can then be sequentially characterized based on the isolated machine building block models of Section C.2. The analysis culminates in the formulation of two dynamic programs, one for an unbounded warehouse and a supply efficiency of 1, and the other one for a limited size warehouse.

#### C.4.1 Two approximations for converting the transfer line into isolated buffer-machine pairs

The first approximation, the so-called machine decoupling approximation, stipulates that the buffer availability process  $I_i(t)$  is a two state Markov process, independent of the state  $\alpha_{i-1}(t)$  of the machine feeding into it.

The second approximation, the supply averaging principle capitalizes on the fact that in the long run, if production rate  $d$  must be met by the transfer line, then every machine in the transfer line must in the long run produces parts at that rate, because of parts conservation in the system. However, it has been established in Section C.2 that a generic machine  $M_i$  has zero production whenever the buffer  $X_{i+1}$  it feeds into is full, i.e. whenever  $I_{i+1}(t) = 0$ . This means that, conditional on  $I_{i+1}(t) = 1$ , the long term production rate of machine  $M_i$ ,  $\tilde{d}_{i+1}$ , which supplies the next inventory  $X_{i+1}(t)$ , must be sufficiently higher than  $d$ , to make up for inability to produce when  $I_{i+1}(t) = 0$ . More specifically, in the steady-state, the following equation must hold :

$$E[u_i(t)] = \tilde{d}_{i+1} P(X_{i+1} < z_{i+1}) + 0 P(X_{i+1} = z_{i+1}) = d \quad (\text{C.19})$$

Note that in (C.19),  $\mathbf{P}(X_{i+1} < z_{i+1})$  is essentially  $E[I_{i+1}(t)]$  and will be defined as the availability of buffer  $X_{i+1}$ , and denoted  $b_{i+1}$  throughout the rest of the paper. Thus :

$$b_{i+1} = 1 - \mathbf{P}_{z_{i+1}} \quad (\text{C.20})$$

Now  $u_i(t)$  is a stochastic supply process which is constrained to be zero whenever  $I_{i+1}(t) = 0$ , and must also satisfy constraint (C.19). The supply averaging principle is the approximation whereby of all the stochastic supply processes consistent with the preceding constraints, one picks the simplest, i.e. a constant of value  $\tilde{d}_{i+1} = d/b_{i+1}$ . This provides a simple model of the influence of the part of the transfer line upstream of buffer  $B_{i+1}$  on the dynamics of  $X_{i+1}$ . In the next subsection, we use the machine decoupling approximation to model the effects of blocking on machine  $M_i$ , i.e. the part of the transfer line downstream of  $M_i$ .

#### C.4.2 Accounting for blocking effects downstream of machine $M_i$

Machine  $M_i$  in the transfer line has only one situation in which it is not blocked and is thus potentially able to produce if not starved. It corresponds to the product state  $\alpha_i(t) \times I_{i+1}(t)$  equal to 1, i.e. if  $M_i$  is operational and  $X_{i+1}$  not full. Otherwise, whenever either of  $\alpha_i(t)$ , or  $I_{i+1}(t)$  is zero, i.e. for the three other states of the  $(\alpha_i(t) \times I_{i+1}(t))$  process, the machine is unable to produce. Under the Machine decoupling approximation  $I_{i+1}(t)$  is a two state Markov chain independent of the Markov chain  $\alpha_i(t)$ , which we shall assume to have a blocking rate denoted  $p'_{i+1}$  and an unblocking rate denoted  $r'_{i+1}$ . Therefore, the product process  $(\alpha_i(t) \times I_{i+1}(t))$  describing machine  $M_i$  when accounting for blocking effects, is a four state Markov chain obtained as the cartesian product of the component chains, and shown in Figure C.4, where product states  $\{i, j\}$ ,  $i = 0, 1$ ,  $j = 0, 1$ , refer to  $\{\alpha_i(t) = i, I_{i+1}(t) = j\}$ . The ultimate approximation of machine  $M_i$  including blocking effects consists in aggregating these four states into a two-state Markov chain; that is one for which the machine produces if not starved, and one for which it cannot. Steps to achieve this aggregation are borrowed from (45; 46), and adapted to our case as follows.

The signal  $I_{i+1}(t)$  spends a random exponential time in state 0. As discussed in Section C.2 in relation to Assumption (C.17), buffer  $B_{i+1}$  can become full only if machine  $M_{i+1}$  is down or blocked, while the latter itself can become blocked only if some machine downstream of it is down. Of course, it is always possible that the root cause of blocking be multiple failures; the latter case is excluded from our analysis under the assumption that it is a very low probability event. Furthermore, if the transfer line is partially homogeneous (45), i.e. if all repair rates  $r_i$  of machines are

identical equal to  $r$ , then under the non multiple failures assumption, the rate  $r'_{i+1}$  at which  $I_{i+1}(t)$  moves from 0 back to 1 is precisely  $r$ ; that is the rate at which machine  $M_{i+1}$  go back to 1. As to the rate  $p'_{i+1}$  at which  $I_{i+1}(t)$  moves from 1 to 0, it is obtained indirectly via the calculation of the steady-state probability of buffer  $B_{i+1}$  not being blocked, i.e. using  $b_{i+1} = \frac{r}{p'_{i+1} + r}$ , assuming the  $b_{i+1}$  calculation has been carried out at a previous stage. This 4 state model is further aggregated into a two-state Markov model with rates  $\tilde{r}_i$  and  $\tilde{p}_i$  as shown in Figure C.4. State  $\{1, 1\}$  is redefined as state 1, the only operational state and all other non operational states are merged into 0. Following (45), in the partially homogeneous case,  $\tilde{r}_i$  is identified with  $r$ , while  $\tilde{p}_i$  is again calculated indirectly from the requirement that  $\mathbf{P}(M_i(t) \times I_{i+1}(t)) \stackrel{ind}{=} \frac{r}{r+p_i} \times b_{i+1} = \frac{r}{r+\tilde{p}_i}$ , leading to :

$$\tilde{p}_i = \frac{p_i + r(1 - b_{i+1})}{b_{i+1}} \quad (\text{C.21})$$

Note that the probability of machine  $M_i$  to be operational is effectively reduced by the factor  $b_{i+1}$ . In addition, note that following (46), one could develop an approach for computing the aggregated two-state representation of machine  $M_i$  including blocking effects in the general non homogeneous transfer line case, but we refrain from doing so here for simplicity and conciseness. Thus, the result is a virtual machine designated  $\tilde{M}_i$  which could be starved but is never blocked because any blocking effects are already accounted for and result in a less productive machine.

With the above assumptions and approximations we can now view the transfer line as a collection of single buffers feeding single two-state never blocked virtual Markovian machines. Such units are precisely the building blocks analyzed in Section C.2, and therefore steady-state expectations and probability densities of interest can be obtained from (C.6-C.8), while thanks to assumption (C.17), the parameters of the virtual machines are calculated sequentially from downstream towards upstream. Details follow in the next subsection which culminates into the formulation of two dynamic programs for buffering optimization.

### C.4.3 Two dynamic programs for buffering optimization

In the context of a limited warehouse size with tolerated supply interruptions, and delayed payback deliveries, the global cost of the line is equal to a linear combination

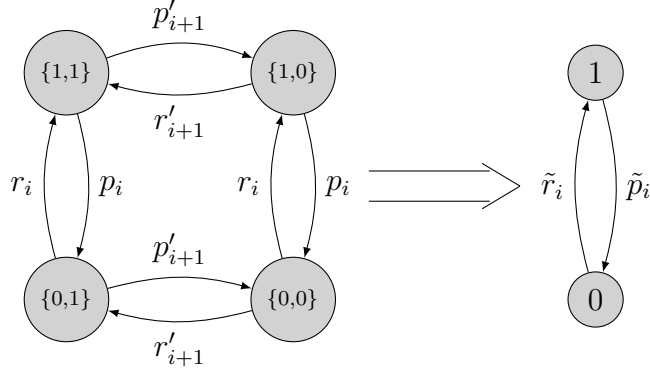


FIGURE C.4 4 State Markov Chain Aggregation

of the mean costs of all wips  $X_i$ ,  $i = 1..m$ , in the line and the objective is to minimize the expectation of this random cost to reach optimality over all possible buffer sizes,  $z_i$ ,  $i = 1..m$ , i.e. :

$$\min E[\sum_{i=1}^m c_i \cdot X_i] = \min_{(z_i \in [0, \infty])_{i \in [1, m]}} \sum_{i=1}^m c_i \left( \int_0^{z_i} x f_{X_i}(x) dx + z_i \mathbf{P}_{z_i} \right) \quad (\text{C.22})$$

Starting from two-state furthestmost downstream machine  $M_m$  (assumed never blocked), and considering that the wip  $X_m$  is fed from a constant stream at rate  $d/b_m$  as long as it is not full, (thanks to the supply averaging principle,) it is possible to use (C.6-C.8) to express  $b_m$  in terms of  $z_m$ , and to obtain the storage cost local to  $M_m$  as a function of  $z_m$ . This is the initialization of a backwards recursive calculation whereby, given  $b_{i+1}$ , the parameters of virtual machine  $\tilde{M}_i$  (possibly starved but never blocked) are calculated, with parameters  $r$  and  $\tilde{p}_i$  (Eq. (C.21)).  $\tilde{M}_i$  draws its parts from wip  $X_i$  itself considered fed by a constant stream  $d/b_i$ . Given  $z_i$ , it is related to  $b_i$  and  $b_{i+1}$  via equation (C.20) :

$$\lambda_i z_i = \ln \left( \frac{\mu_i b_i}{1 - b_i} \left( \frac{r}{p_i + r} \frac{k_i}{d} b_{i+1} - 1 \right) \right) \quad (\text{C.23})$$

where now  $\mu_i = \frac{\tilde{d}_i}{k_i - \tilde{d}_i} \frac{\tilde{p}_i}{r}$  and  $\lambda_i = \frac{r}{\tilde{d}_i} - \frac{\tilde{p}_i}{k_i - \tilde{d}_i}$

Based on (C.6-C.8), and using (C.23) to express  $z_i$  in terms of  $b_i$  and  $b_{i+1}$  viewed as surrogate decision variables preferable to  $z_i$  because bounded between 0 and 1, it

follows that :

$$\begin{aligned} \forall i \in [1, m] \\ T_i(b_i, b_{i+1}) = c_i E[X_i] &= \frac{c_i}{\lambda_i} \frac{\tilde{d}_i}{k_i - \tilde{d}_i} \left( \frac{k_i}{\tilde{d}_i} \left( b_i - \frac{r}{\tilde{p}_i + r} \right) + \frac{1 - b_i}{1 - \mu_i} \frac{\tilde{p}_i + r}{r} \times \right. \\ &\quad \left. \ln \left( \frac{\mu_i}{1 - b_i} \left( \frac{r}{\tilde{p}_i + r} \frac{k_i}{\tilde{d}_i} - b_i \right) \right) \right) \end{aligned} \quad (\text{C.24})$$

where two constraints relating  $b_i$  and  $b_{i+1}$  must be satisfied :

1.  $b_i > \frac{r}{r + p_i} b_{i+1}$  ; indeed, the right hand side is the minimum availability of buffer  $X_i$  attained for  $z_i = 0$ .
2.  $b_{i+1} > \frac{d}{k_i} \frac{r + p_i}{r}$  ; this is a production rate feasibility requirement, insuring in (C.23) that for  $b_i$  between 0 and 1, the expression inside the  $\ln(\cdot)$  remain positive, thus guaranteeing that there exists a solution.

Choosing to view  $b_{i+1}$  as dependent on  $b_i$ , and recalling that  $b_{i+1}$ , a probability, must lie on  $[0, 1]$ , these constraints are gathered in the set  $\mathcal{A}_{i+1}(b_i)$  :

$$\mathcal{A}_{i+1}(b_i) = \left\{ b_{i+1} \in [0, 1] \text{ s.t. } \frac{d}{k_i} \frac{r + p_i}{r} < b_{i+1} < b_i \frac{r + p_i}{r} \right\} \quad (\text{C.25})$$

In the context of a limited size warehouse and the possibility of outsourcing storage, for a given availability  $b_2$  of stock  $X_2$ , it is possible to optimize the size of the limited warehouse in expression (C.16) to achieve the best sharing of local and outsourced costs (see Appendix), yielding the following minimum cost function :

$$T'_{opt}(b_1 = 1, b_2) = \frac{c_1}{\lambda_1} \left( 1 + \ln \left( \frac{c_{1'}}{c_1} \frac{\tilde{p}_1}{\tilde{p}_1 + r} \frac{k_1}{k_1 - d} \right) \right) \quad (\text{C.26})$$

where recalling (C.14-C.15) for the definition of  $\lambda_1$ ,  $\lambda_1 > 0 \Leftrightarrow \frac{r}{r + p_1} b_2 k_1 > d$  and  $\frac{c_1}{c_{1'}} \leq \frac{\tilde{p}_1}{\tilde{p}_1 + r} \frac{k_1}{k_1 - d}$  (condition for having  $z > 0$ ).

In summary, two distinct dynamic programs can now be formulated, one for a limited warehouse, and the other one with raw material storage outsourcing :

The above two expressions represent an approximate evaluation of the global expectation of the costs of storage of raw materials and wip for transfer line operation. The supply efficiency of the system can be defined as the variable  $b_1$ . Note that the value  $b_1$  depends on the propagation of blocking phenomena in the line up from the

**Buffer sizing problem with tolerated supply disruptions under a constant average production rate  $d$  : Pav**

$$\forall i \in [1, m] \quad b_i \in ]\frac{d}{k_i}, 1[, \quad b_{i+1} \in \mathcal{A}_{i+1}(b_i), \quad b_{m+1} = 1$$

$$J_{tot} = \min_{(b_i)_{i \in [1, m+1]}} \sum_{i=1}^m T_i(b_i, b_{i+1}) \quad (\text{C.27})$$

with,  $T_i(., .)$  as in (C.24),  $\mu_i = \frac{\tilde{d}_i}{k_i - \tilde{d}_i} \frac{\tilde{p}_i}{r}$ ,  
 $\lambda_i = \frac{r}{\tilde{d}_i} - \frac{\tilde{p}_i}{k_i - \tilde{d}_i}$ ,  $\tilde{d}_i = \frac{d}{b_i}$  and  $\tilde{p}_i = \frac{r(1-b_{i+1})+p_i}{b_{i+1}}$

**Buffer sizing problem with zero tolerance for supply disruptions under a constant average rate of production  $d$  : Pav'**

$$\forall i \in [1, m] \quad b_i \in ]\frac{d}{k_i}, 1[, \quad b_{i+1} \in \mathcal{A}_{i+1}(b_i), \quad b_1 = b_{m+1} = 1$$

$$J_{tot} = \min_{(b_i)_{i \in [1, m+1]}} \left( T'_{opt}(b_1, b_2) + \sum_{i=2}^m T_i(b_i, b_{i+1}) \right) \quad (\text{C.28})$$

with,  $T_i(., .)$  and  $T'_{opt}(., .)$  as in (C.24) and (C.26),  
 $\mu_i = \frac{\tilde{d}_i}{k_i - \tilde{d}_i} \frac{\tilde{p}_i}{r}$ ,  $\lambda_i = \frac{r}{\tilde{d}_i} - \frac{\tilde{p}_i}{k_i - \tilde{d}_i}$ ,  $\tilde{d}_i = \frac{d}{b_i}$  and  $\tilde{p}_i = \frac{r(1-b_{i+1})+p_i}{b_{i+1}}$

never blocked machine  $M_m$  furthest downstream. Also, note a variant of problem Pav is one where a supply efficiency, say  $b_1^0$ , is imposed at the outset. In this case, it becomes a constraint in dynamic program Pav.

We next validate these models against Monte Carlo simulation.

## C.5 Monte Carlo Simulation

A Monte Carlo simulation is used to obtain the expected global cost of holding the wip stocks and finished parts inventory in the line, for a given choice of buffer sizes. The following two main steps have been followed in order to build the simulator :

- 1 Machines on/off random durations are first generated according to the given failure and repair rate parameters resulting after proper ordering into a marked point process.
- 2 Given the point process in 1, evolution of internal wip and inventory becomes deterministic on the given time horizon and can be propagated by considering production rates changes as the driving point process unfolds.

We make sure that the results converge over more than half the time of the chosen horizon, and the warmup portion of the simulated data is withdrawn from the results for better precision. The resulting values of the simulation are shown in the Table C.1 with  $\forall i \in [1, m]$   $r_i = 0.9$ ,  $p_i = 0.1$ ,  $k_1 = 3$ ,  $\Delta k = .2$ ,  $d = 1$  and  $c_i = 1$ .  $\Delta k$  is the steady increment from  $k_i$  to  $k_{i+1}$ , and the  $z_i^t$  vector is set to  $[0.8246]$ ,  $[1.2; 1.1]$ ,  $[1.33; 1.1; 1.05]$ ,  $[1.39; 1.11; 0.98; 0.99]$  and  $[1.42; 1.12; 0.97; 0.90; 0.94]$  for the considered 1, 2, 3, 4 and 5 machines lines respectively. These values are obtained from the dynamic programming formulation described in the next section.

TABLEAU C.1 Comparison of Theoretical and Simulation Results

# of machines	1	2	3	4	5
Simulated total cost	0.0699	0.2179	0.3693	0.5207	0.6690
Theoretical total cost	0.0695	0.2162	0.3607	0.5019	0.6404
Percent Relative Error	-0.57%	-0.78%	-2.33%	-3.61%	-4.27%

For a single machine line, analytical equations are exact and a weak difference shows that the system ergodic assumption is verified. Up to a 5 machines line the simulation results validate the theoretical model with a maximum difference of 5.796%. Details on internal mean wips and buffer availability relative estimation errors for the 5 machine line are given in Table C.2 : Inventory wip expectations display 5.7%

TABLEAU C.2 Results for a 5 Machine Line

Buffer #	1	2	3	4	5
$E[X_i]$					
Simulated	0.1684	0.1608	0.1356	0.1054	0.0702
Theoretical	0.1780	0.1529	0.1307	0.1053	0.0735
Relative percent error	5.7%	-4.9%	-3.6%	-0.09%	4.7%
$b_i$					
Simulated	0.9477	0.9281	0.9252	0.9362	0.9609
Theoretical	0.95	0.9326	0.9268	0.9332	0.9552
Relative percent error	0.23%	0.48%	0.17%	-0.32%	-0.59%

maximum relative errors, which are of the same order as global cost relative errors for the five machine line are given in Table C.2. It is important to note that buffer availability is estimated with a high degree of precision, uniformly within the line.



These simulation results justify the application of the theoretical model for buffer optimal sizing carried out in the next section.

## C.6 Dynamic Programming Results

The minimization problems in (C.27) and (C.28) are amenable to solution via dynamic programming. Indeed, the total cost is additive and each component  $T_i(b_i, b_{i+1})$  is dependant on downstream variables  $b_i, ..b_m$  only. It involves what can be viewed as a transition cost between state  $b_i$  at stage  $i$  and state  $b_{i+1}$  at stage  $i + 1$ . The implicit decision variable enforcing this transition is buffer size  $z_i$ . Therefore, the optimal path can be obtained recursively, with Bellman equations as follows for  $i = 1..m - 1$  :

- (i) Instantaneous cost of wip  $i$ ,  $X_i$  :

$$g(X_i) = c_i \cdot X_i \quad (\text{C.29})$$

- (ii) Minimum expected cost of  $X_m$  given  $b_m$  :

$$J_m(b_m) = E[g(X_m, b_m, b_{m+1} = 1)] = T_m(b_m, 1) \quad (\text{C.30})$$

- (iii) Minimum expected cost of wips from  $i$  to  $m$  in regards to  $b_i$  :

$$\begin{aligned} J_i(b_i) &= \min_{b_{i+1}} E[g(X_i) + J_{i+1}(b_{i+1})] \\ &= \min_{b_{i+1} \in \mathcal{A}_{i+1}(b_i)} (T_i(b_i, b_{i+1}) + J_{i+1}(b_{i+1})) \end{aligned} \quad (\text{C.31})$$

The function  $J_1(b_1)$ , which is the minimum transfer line cost as a function of the imposed supply efficiency  $b_1$  can be defined as follows to yield a unique set of equations for problems Pav and Pav' :

$$J_1(b_1) = \begin{cases} \min_{b_2} (T'_{opt}(1, b_2) + J_2(b_2)) & \text{if } b_1 = 1 \text{ \textbf{Pav}} \\ \min_{b_2} (T_1(b_1, b_2) + J_2(b_2)) & \text{if } b_1 < 1 \text{ \textbf{Pav}} \end{cases} \quad (\text{C.32})$$

In the following figures results are presented from the optimisation programming, with  $\forall i \in [1, 20]$   $r_i = 0.9$ ,  $p_i = 0.1$ ,  $k_1 = 3$ ,  $\Delta k = .2$ ,  $d = 1$  and  $c_i = 1$ .

In Figures C.5 and C.6, the supply efficiency is imposed at 0.95. The optimal availability of storage areas follows a convex and smooth curve starting from .95 the imposed supply efficiency up to 1 (the last machine is never blocked). For a lower

line efficiency, the optimal curve starts lower but has the same general shape with minimum availability somewhere towards the middle of the line. While the optimal availability curve appears to be almost symmetric with respect to the center of the line, this is not the case for the optimal buffer sizes curve. The largest buffer sizes are concentrated towards the upstream part of the line where there is a strong constraint on supply efficiency. They tend to taper down as one moves away from that point downstream since machine capacity is increasing (Assumption (C.17)), and production reduction by propagating blocking phenomena is less important as one goes further in the line. At the end of the line, the few last buffers increase again. Such an increase in buffering may be due to the fact that it permits a maximum capitalization on the infinite buffering assumed available for final inventory storage. Thus at the cost of a small increase in the size of buffer  $B_{20}$  which may not be that occupied on average, one can afford to significantly decrease buffer sizes when moving upstream towards the center of the line.

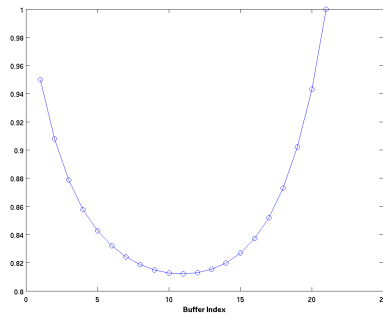


FIGURE C.5 Optimal Availability of Buffers

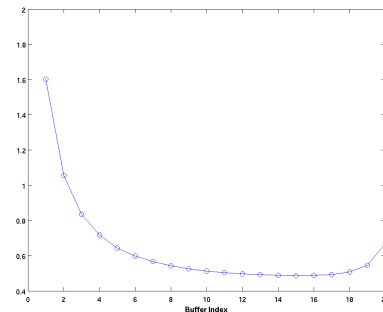


FIGURE C.6 Optimal Buffer Sizes

In Figures C.7 to C.10, the supply efficiency is set to 1 with the possibility of outsourcing storage of any overflow. The rate of arrival of raw materials is constant equal to  $d$ . The cost of storage of raw material overflow for the additional storage capacity is set at the significantly higher than local rate, of 5 in Figure C.7-C.8. In this case, the optimal distribution of storage areas availability starts at 1, and then appears to follow the same pattern as for the previous case with fluctuations of supply tolerated. Indeed, because of the structure of the dynamic program, for a given availability of supply at buffer 2, the problem of optimally sizing buffers 3, ...,  $m$  falls

in the class of problems Pav, and thus the same type of results should be expected. The optimal buffers have an irregular distribution. The warehouse size with fixed running cost is empty, while buffer 2 is associated with the largest storage space. A likely explanation of this phenomenon could be the following : the cost of having an outsourced storage capacity is five times higher than the regular one in the line ; thus it is cheaper to stock the maximum of materials in the line than to outsource storage. The question remains as to why storage goes to 2 instead of 1, and the answer is that buffer 2 is associated with a fluctuating instantaneous cost proportional to the current size of wip in it ; this is unlike buffer 1 where the running cost is fixed and proportional to its size  $z_1$ . When the per unit outsourcing cost increases from 5 to 8 in Figure C.9-C.10, some buffering is now associated with buffer 1, while the size of buffer 2 increases threefold. Note that for  $c' = 5$ , the mean cost of outsourcing storage is 1.1997, it slightly decreases to 1.1807 when  $c' = 6$ , and increases back to 1.4022 when  $c' = 8$ . In effect, the storage capacity of buffer 2 is increasingly utilized at the expense of the storage outsourcing component of the cost which decreases. Once buffer 2 is saturated, any further increase in outsourcing per unit costs will induce a filling of buffer 1 up to a point (because of its fixed instantaneous costs proportional to its size, which make it more expensive), past which, in the absence of better alternatives, the outsourcing cost component starts increasing again. In the mean time, the global cost of the line remains monotone increasing, from 3.7884 to 3.8299 up to 4.088, respectively. While the global cost evolves monotonically with  $c'$ , the outsourcing, buffer 1, and buffer 2 storage component costs occasionally evolve in a non monotonic fashion.

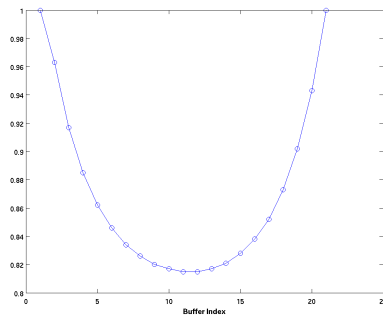


FIGURE C.7 Optimal Availability of Buffers

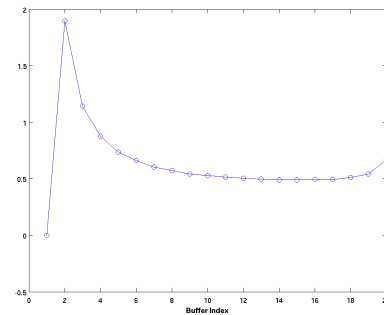


FIGURE C.8 Optimal Buffer Sizes

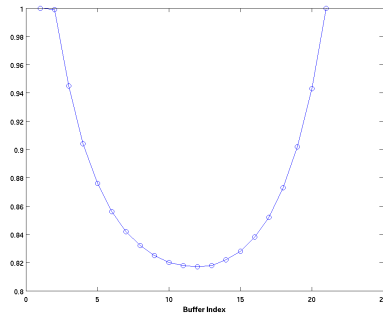


FIGURE C.9 Optimal Availability of Buffers

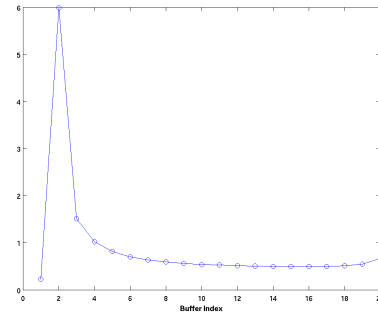


FIGURE C.10 Optimal Buffer Sizes

In Figure C.11, the minimal global running storage cost of the line is plotted as a function of the required supply efficiency, and is seen to be quasi linear with a positive slope. Furthermore, the gap between the two values  $J_{tot}(b_1 = 1)$  and  $\lim_{b_1 \rightarrow 1} J_{tot}(b_1)$  represents the cost of guaranteeing that raw materials can be delivered at a strictly constant pace.

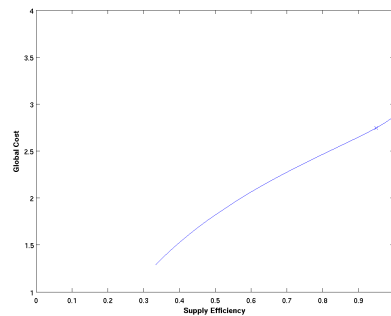


FIGURE C.11 Global Cost of The Production Line Against Supply Efficiency

The results of the optimization are of interest. They show that, using dynamic programming, it is possible to size the buffers of the line optimally for a specific supply efficiency, and for a given required parts production rate. In addition, the global cost of the line can be determined as a function of the required supply efficiency, which is of direct interest for the industrial manager. This is indeed a measure of the costs associated with making the internal random fluctuations of the transfer line, more or less transparent to its supplier.

## C.7 Conclusion

The decomposition approach developed by Malhamé and Sadr (45; 46) for the approximate performance analysis of unreliable transfer lines under Kanban policies with machines having monotone decreasing production capacities from upstream towards downstream, has been generalized to the dual case where these capacities are monotonically increasing instead. The main advantage of this decomposition methodology relative to other existing methods, is the possibility of propagating calculations unidirectionally (from downstream towards upstream in the current case). This allows one to identify sequential decision stages and to use dynamic programming both as an analytical as well as a numerical optimization tool for determining how lean a transfer line should be under specific constraints : a given long term production rate, and a fixed so-called supply efficiency measured by the probability of warehouse overflow when raw materials are brought at a steady pace consistent with the production rate requirement. It is observed that as the required supply efficiency increases, there is near linear increase in the corresponding minimal long term average storage costs in the transfer line, with an acceleration close to 100% efficiency. Thus the cost of supply efficiency, as measured by the degree to which the instantaneous rate of supply of raw materials can be shielded from the random fluctuations internal to the transfer line, can be assessed. This is of significance in the analysis of cascaded supply chains.

Concerning decomposition methods in transfer lines under Kanban policies, what one discovers is that decompositions are easiest in lines where machine production capacities are either equal, or evolve monotonically with the machine index. Indeed, if the capacities monotonically decrease from upstream towards downstream, the rate at which a given buffer gets emptied becomes a strictly local property, while if the capacities monotonically increase instead, it is the rate at which the buffer gets filled which is a local property. In the first case, the buffer fill rate is a complex non local (distant influences) stochastic process, and in the second case, the buffer emptying rate is the non local process. In both cases, under a fixed mean production rate requirement, an averaging principle is what allows one to circumvent the difficulty related to non locality, and thus carry out calculations entirely locally. It will be a challenge in the coming years to develop transfer line decompositions for lines possessing piecewise monotonic machine capacity sections, short of full forward and backwards propagat-

ing iterations such as in (34; 35; 37; 38). Finally, note that decomposition methods are also possible for non local policies such as CONWIP (44), and they could be helpful in developing a theory of production architecture in transfer lines (33).

## C.8 appendix

**Proof of equations (C.6-C.8)** In the steady state, (C.2) is written as  $\mathbf{J} \frac{d\vec{f}(x)}{dx} = \mathbf{Q}^t \vec{f}(x)$ , with solution  $\vec{f}(x) = \exp(x\mathbf{J}^{-1} {}^t\mathbf{Q}) \vec{f}(0)$ .

$$\mathbf{J}^{-1} {}^t\mathbf{Q} = \begin{pmatrix} p & k-d \\ r & d \end{pmatrix} \begin{pmatrix} 0 & 0 \\ 0 & -\lambda \end{pmatrix} \begin{pmatrix} p & k-d \\ r & d \end{pmatrix}^{-1} = PDP^{-1} \quad (\text{C.33})$$

where,  $\lambda = \frac{r}{d} - \frac{p}{k-d}$ . Then,

$$\begin{aligned} \vec{f}(x) &= \exp(x\mathbf{J}^{-1} {}^t\mathbf{Q}) \vec{f}(0) = P \exp(Dx) P^{-1} \vec{f}(0) \\ &= P \begin{pmatrix} 1 & 0 \\ 0 & e^{-\lambda x} \end{pmatrix} P^{-1} \vec{f}(0) \\ &= \left( \mathbf{I} + \frac{1-e^{-\lambda x}}{\lambda} \mathbf{J}^{-1} {}^t\mathbf{Q} \right) \vec{f}(0) \\ f_{off}(x) &= f_{off}(0) + \left( \frac{p}{d} f_{on}(0) - \frac{r}{d} f_{on}(0) \right) \frac{1-e^{-\lambda x}}{\lambda} \\ &= f_{off}(0) e^{-\lambda x} \quad \text{using equation (C.4)} \\ f_{on}(x) &= f_{on}(0) + \left( \frac{p}{k-d} f_{on}(0) - \frac{r}{k-d} f_{off}(0) \right) \frac{1-e^{-\lambda x}}{\lambda} \\ &= f_{on}(0) e^{-\lambda x} \quad \text{using equation (C.4)} \end{aligned} \quad (\text{C.34})$$

Equations (C.3) and (C.4) yield  $f_{on}(0) = \frac{d}{k-d} f_{off}(0)$ ,  $f_X(0) = \frac{k}{k-d} f_{off}(0)$ ,  $\mathbf{P}_o = \frac{d}{p} f_{off}(0)$ , and  $\mathbf{P}_z = \frac{d}{r} f_{off}(0)$ . Applying (C.5) to  $f_{off}(0)$ , one obtains :

$$\begin{aligned} (C.5) \quad &\Longleftrightarrow f_{off}(0) \left( \frac{d}{p} + \frac{d}{r} e^{-\lambda z} + \frac{k}{k-d} \int_0^z e^{-\lambda x} dx \right) = 1 \\ &\Longleftrightarrow f_{off}(0) = \frac{p}{r+p} \frac{\lambda}{1-\mu e^{-\lambda z}} \quad \text{where, } \mu = \frac{d}{k-d} \frac{p}{r} \end{aligned} \quad (\text{C.35})$$

which leads to equations (C.6), (C.7) and (C.8).

**Proof of equation (C.26)** Definition (C.16), together with equations (C.14) (C.15), leads to the following result.

$$J'_{tot} = cz + c' \frac{p}{r+p} \frac{k}{k-d} \frac{1+\lambda z}{\lambda} e^{-\lambda z} \quad (\text{C.36})$$

This function is convex with respect to  $z \in [0, \infty[$ . Indeed its derivate  $\frac{dJ'_{tot}(z)}{dz} = c - c' \frac{p}{r+p} \frac{k}{k-d} e^{-\lambda z}$  is first negative, zero at  $z^* = \frac{1}{\lambda} \ln(\frac{c'}{c} \frac{p}{r+p} \frac{k}{k-d})$  and positive for  $z > z^*$ . Therefore, the optimal cost (C.26),  $T'_{opt}$ , is  $J'_{tot}(z^*)$  and is reached at  $z = z^*$ .

# Références

- [31] J. Arnold, S. Minner, B. Eidam, Raw material procurement with fluctuating prices, *International Journal of Production Economics* 121 (2009) 353-364.
- [32] T. Bielecki, P. Kumar, Optimality of zero-inventory policies for unreliable manufacturing systems, *Operations Research* 36 (1988) 532-541.
- [33] A.M. Bonvik, S. Gerhwin, Beyond kanban : Creating and analyzing lean shop floor control policies, in : *Manufacturing and Service Operations Management*.
- [34] S.Y. Chiang, A. Hu, S.M. Meerkov, Lean buffering in serial production lines with nonidentical exponential machines, *IEEE Transactions on Automation Science and Engineering* 5 (2008) 298-306.
- [35] S.Y. Chiang, C.T. Kuo, S.M. Meerkov, Bottlenecks in Markovian production lines : A systems approach, *IEEE Transactions on Robotics and Automation* 14 (1998) 352-359.
- [36] J. Geraghty, C. Heavey, A comparison of hybrid push/hull and conwip/pull production inventory control policies, *International Journal of Production Economics* 91 (2004) 75-90.
- [37] S. Gershwin, I. Schick, Analytic methods for calculating performance measures of production lines with buffer storage, *Proceedings of the IEEE Conference on Decision and Control* (1978) 618-624.
- [38] S. Gershwin, Efficient decomposition method for the approximate evaluation of tandem queues with finite storage space and blocking, *Operations Research* 35 (1987) 291-305.
- [39] A. Gharbi, J.P. Kenné, M. Beit, Optimal safety stocks and preventive maintenance periods in unreliable manufacturing systems, *International Journal of Production Economics* 107 (2007) 422-434. *Operations Management in China*.
- [40] A. Hajji, F. Mhada, A. Gharbi, R. Pellerin, R. Malhamé, Integrated product specifications and productivity decision making in unreliable manufacturing systems, *International Journal of Production Economics* 129 (2011) 32-42.
- [41] J.Q. Hu, Production rate control for failure-prone production systems with no backlog permitted, *Automatic Control, IEEE Transactions on* 40 (1995) 291-295.



- [42] S. Ioannidis, An inventory and order admission control policy for production systems with two customer classes, *International Journal of Production Economics* 131 (2011) 663-673.
- [43] R. Malhamé, E.K. Boukas, A renewal theoretic analysis of a class of manufacturing systems, *Automatic Control, IEEE Transactions on* 36 (1991) 580-587.
- [44] F. Mhada, R. Malhamé, Approximate performance analysis of conwip disciplines in unreliable non homogeneous transfer line, *Annals of Operations Research* 182 (2010) 213-233.
- [45] J. Sadr, R. Malhamé, Decomposition/aggregation-based dynamic programming optimization of partially homogeneous unreliable transfer lines, *Automatic Control, IEEE Transactions on* 49 (2004) 68 - 81.
- [46] J. Sadr, R. Malhamé, Unreliable transfer lines : decomposition/aggregation and optimization, *Annals of Operations Research* 125 (2004) 167-190.
- [47] G.K. Saharidis, V.S. Kouikoglou, Y. Dallery, Centralized and decentralized control policies for a two-stage stochastic supply chain with subcontracting, *International Journal of Production Economics* 117 (2009) 117-126.