Optimisation Stochastique pour la gestion des lits d'hospitalisation sous incertitudes

Alexandre Mazier

▶ **To cite this version:**

Alexandre Mazier. Optimisation Stochastique pour la gestion des lits d'hospitalisation sous incertitudes. Autre. Ecole Nationale Supérieure des Mines de Saint-Etienne, 2010. Français. NNT : 2010EMSE0586 . tel-00667076

HAL Id: tel-00667076

https://theses.hal.science/tel-00667076

Submitted on 6 Feb 2012

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Ecole Nationale Supérieure des Mines
SAINT-ETIENNE

NNT : 2010 EMSE 0586

THESE

présentée par

Alexandre Mazier

Pour obtenir le grade de Docteur
de l'École Nationale Supérieure des Mines de Saint-Étienne

Spécialité : Génie Industriel

Optimisation stochastique pour la gestion des lits d'hospitalisation sous incertitudes

Soutenue à Saint-Étienne le 6 décembre 2010

Membres du jury

M. Jean-Charles Billaut	Professeur, Université François Rabelais, Tours	Président
M. Erik Demeulemeester	Professeur, Katholieke Universiteit Leuven, Belgique	Rapporteur
M. Alain Guinet	Professeur, Institut National des Sciences Appliquées de Lyon	Rapporteur
M. Michel Gourgand	Professeur, Université Blaise Pascal, Clermont-Ferrand	Examinateur
M. Xiaolan Xie	Professeur, École des Mines de Saint-Étienne	Directeur de thèse
M. Gille-André Cuchet	Directeur d'hôpital, CH de Firminy	Membre invité
Mme. Marianne Sarazin	Département d'information médicale, CH de Firminy	Membre invitée
M. Rémy Brunon	Médecin responsable du service des urgences, CH de Firminy	Membre invité

Spécialités doctorales :	**Responsables :**
SCIENCES ET GENIE DES MATERIAUX	J. DRIVER Directeur de recherche – Centre SMS
MECANIQUE ET INGENIERIE	A. VAUTRIN Professeur – Centre SMS
GENIE DES PROCEDES	G. THOMAS Professeur – Centre SPIN
SCIENCES DE LA TERRE	B. GUY Maître de recherche – Centre SPIN
SCIENCES ET GENIE DE L'ENVIRONNEMENT	J. BOURGOIS Professeur – Centre SITE
MATHEMATIQUES APPLIQUEES	E. TOUBOUL Ingénieur – Centre G2I
INFORMATIQUE	O. BOISSIER Professeur – Centre G2I
IMAGE, VISION, SIGNAL	JC. PINOLI Professeur – Centre CIS
GENIE INDUSTRIEL	P. BURLAT Professeur – Centre G2I
MICROELECTRONIQUE	Ph. COLLOT Professeur – Centre CMP

Enseignants-chercheurs et chercheurs autorisés à diriger des thèses de doctorat (titulaires d'un doctorat d'État ou d'une HDR)

AVRIL	Stéphane	MA	Mécanique & Ingénierie	CIS
BATTON-HUBERT	Mireille	MA	Sciences & Génie de l'Environnement	SITE
BENABEN	Patrick	PR 1	Sciences & Génie des Matériaux	CMP
BERNACHE-ASSOLANT	Didier	PR 0	Génie des Procédés	CIS
BIGOT	Jean-Pierre	MR	Génie des Procédés	SPIN
BILAL	Essaïd	DR	Sciences de la Terre	SPIN
BOISSIER	Olivier	PR 1	Informatique	G2I
BOUCHER	Xavier	MA	Génie Industriel	G2I
BOUDAREL	Marie-Reine	PR 2	Génie Industriel	DF
BOURGOIS	Jacques	PR 0	Sciences & Génie de l'Environnement	SITE
BRODHAG	Christian	DR	Sciences & Génie de l'Environnement	SITE
BURLAT	Patrick	PR 2	Génie industriel	G2I
COLLOT	Philippe	PR 1	Microélectronique	CMP
COURNIL	Michel	PR 0	Génie des Procédés	SPIN
DAUZERE-PERES	Stéphane	PR 1	Génie industriel	CMP
DARRIEULAT	Michel	IGM	Sciences & Génie des Matériaux	SMS
DECHOMETS	Roland	PR 1	Sciences & Génie de l'Environnement	SITE
DESRAYAUD	Christophe	MA	Mécanique & Ingénierie	SMS
DELAFOSSE	David	PR 1	Sciences & Génie des Matériaux	SMS
DOLGUI	Alexandre	PR 1	Génie Industriel	G2I
DRAPIER	Sylvain	PR 2	Mécanique & Ingénierie	SMS
DRIVER	Julian	DR 0	Sciences & Génie des Matériaux	SMS
FEILLET	Dominique	PR 2	Génie Industriel	CMP
FOREST	Bernard	PR 1	Sciences & Génie des Matériaux	CIS
FORMISYN	Pascal	PR 1	Sciences & Génie de l'Environnement	SITE
FORTUNIER	Roland	PR 1	Sciences & Génie des Matériaux	SMS
FRACZKIEWICZ	Anna	DR	Sciences & Génie des Matériaux	SMS
GARCIA	Daniel	MR	Génie des Procédés	SPIN
GIRARDOT	Jean-Jacques	MR	Informatique	G2I
GOEURIOT	Dominique	MR	Sciences & Génie des Matériaux	SMS
GRAILLOT	Didier	DR	Sciences & Génie de l'Environnement	SITE
GROSSEAU	Philippe	MR	Génie des Procédés	SPIN
GRUY	Frédéric	MR	Génie des Procédés	SPIN
GUY	Bernard	MR	Sciences de la Terre	SPIN
GUYONNET	René	DR	Génie des Procédés	SPIN
HERRI	Jean-Michel	PR 2	Génie des Procédés	SPIN
INAL	Karim	PR 2	Microélectronique	CMP
KLÖCKER	Helmut	DR	Sciences & Génie des Matériaux	SMS
LAFOREST	Valérie	CR	Sciences & Génie de l'Environnement	SITE
LERICHE	Rodolphe	CR CNRS	Mécanique et Ingénierie	SMS
LI	Jean-Michel	EC (CCI MP)	Microélectronique	CMP
LONDICHE	Henry	MR	Sciences & Génie de l'Environnement	SITE
MALLIARAS	George Grégory	PR 1	Microélectronique	CMP
MOLIMARD	Jérôme	MA	Mécanique et Ingénierie	SMS
MONTHEILLET	Frank	DR 1 CNRS	Sciences & Génie des Matériaux	SMS
PERIER-CAMBY	Laurent	PR 2	Génie des Procédés	SPIN
PIJOLAT	Christophe	PR 1	Génie des Procédés	SPIN
PIJOLAT	Michèle	PR 1	Génie des Procédés	SPIN
PINOLI	Jean-Charles	PR 0	Image, Vision, Signal	CIS
STOLARZ	Jacques	CR	Sciences & Génie des Matériaux	SMS
SZAFNICKI	Konrad	MR	Sciences & Génie de l'Environnement	SITE
THOMAS	Gérard	PR 0	Génie des Procédés	SPIN
TRIA	Assia		Microélectronique	CMP
VALDIVIESO	François	MA	Sciences & Génie des Matériaux	SMS
VAUTRIN	Alain	PR 0	Mécanique & Ingénierie	SMS
VIRICELLE	Jean-Paul	MR	Génie des procédés	SPIN
WOLSKI	Krzysztof	DR	Sciences & Génie des Matériaux	SMS
XIE	Xiaolan	PR 1	Génie industriel	CIS

Glossaire :

PR 0	Professeur classe exceptionnelle
PR 1	Professeur 1ère catégorie
PR 2	Professeur 2ème catégorie
MA(MDC)	Maître assistant
DR	Directeur de recherche
Ing.	Ingénieur
MR(DR2)	Maître de recherche
CR	Chargé de recherche
EC	Enseignant-chercheur
IGM	Ingénieur général des mines

Centres :

SMS	Sciences des Matériaux et des Structures
SPIN	Sciences des Processus Industriels et Naturels
SITE	Sciences Information et Technologies pour l'Environnement
G2I	Génie Industriel et Informatique
CMP	Centre de Microélectronique de Provence
CIS	Centre Ingénierie et Santé

Dernière mise à jour le : 9 mars 2010

A tous ceux qui me sont chers,

A la mémoire de mon grand-père.

Remerciements

Tout d'abord, je souhaite remercier M. Jean-Charles Billaut qui, en plus de m'avoir initié aux joies de la recherche opérationnelle avec l'ensemble de l'équipe Ordonnancement et Conduite du Laboratoire d'Informatique de l'université de Tours, m'a fait l'immense honneur de présider le jury de thèse.

Je tiens ensuite à remercier les deux rapporteurs de cette thèse : A M. Erik Demeulemeester pour l'effort et l'énergie dépensée afin d'apporter des remarques judicieuses sur un manuscrit écrit dans une langue étrangère. A M. Alain Guinet, dont j'ai eu plaisir de cotoyer dans différents projets tout au long de ces trois années, pour ces remarques et conseils privilégiants l'aspect « hôpital » et qui apportent un coté humain à la recherche. Je remercie également M. Michel Gourgand d'avoir accepté de participer à ce jury et pour les enseignements que j'ai pu tirer de nos quelques discussions.

J'adresse un merci tout particulier à M. Xiaolan Xie pour la grande qualité de son encadrement, son investissement dans ce travail, son coté très humain pourtant difficilement décelable au premier abord et pour m'avoir appris à surmonter l'effroi provoqué par le terme *stochastique*.

Merci au personnel de l'Ecole des Mines de Saint Etienne de m'avoir accueilli dans leurs locaux. Aux personnes du centre ingénierie et santé pour l'enrichissement personnel que j'ai pu tirer d'un centre de recherche pluri-disciplinaire. Un merci particulier a Françoise notre chère secrétaire qui s'accomplit toujours dans la bonne humeur de tout ce qu'on peut lui demander même en dernière minute. Enfin merci aux membres de l'équipe GIH pour leurs conseils avisés et pour les bons moments passés ensemble.

Tout ce travail n'aurait pas pu se faire sans le support du Centre Hospitalier de Firminy. J'en remercie son directeur M. Gille-André Cuchet pour avoir accepté de nous ouvrir les portes de son établissement, Mme Marianne Sarazin, médecin DIM, et Dr. Rémy Brunon, responsable des urgences pour leur investissement total dans ce projet et leur soutient sans faille. Je remercie également M. Sébastien Claude, responsable du service informatique, pour le support technique qu'il m'a apporté. Enfin, je remercie tout le personnel de l'établissement, médecins, cadres de santé, soignants, ambulanciers, secrétaires ... d'avoir répondu à toutes mes questions avec le sourire et pour tout le crédit qu'ils ont bien voulu porter à cette étude. J'espère que les résultats sont à la hauteur de vos attentes.

J'ai une pensée toute particulière à tout ceux avec qui j'ai eu plaisir à partager ces pauses cafés. Doctorants, post-doc, stagiaires, permanents, techniciens, secrétaires ... ces instants passés à parler de tout et n'importe quoi – « ouiii ils sont puérils, c'est parce que ca va mal dans leur couple » (c) – sont des moments importants que j'ai pu savourer en votre compagnie. D'ailleurs, qui fait le café ?

Ma dernière pensée va à toi pour qui je n'ai pas de qualificatifs à la hauteur des remerciements que je voudrais t'adresser. Toi qui m'a suivi dans les contrées foréziennes et alti-ligériennes. Toi qui m'a encouragé durant ces trois années et qui a toujours su me relancer dans mes périodes de doute. Toi qui a supporté mes humeurs quelque peu changeantes durant la rédaction de ce manuscrit. Tu mérites tout autant que moi ce diplome car sans toi tout ce travail de recherche n'aurait pas pu être ce qu'il est aujourd'hui. Je t'aime.

Table des matières

Introduction générale **17**
Le besoin d'outils d'aide à la décision dans le milieu de la santé . . . 17
Objectifs scientifiques . . . 18
Organisation de la thèse . . . 19

Chapitre 1 Gestion du flux de patients dans les systèmes hospitaliers **21**
1.1 Introduction . . . 21
1.2 Le flux de patients aux urgences . . . 22
1.2.1 Organisation des services d'urgences et Parcours patient . . . 23
1.3 Description des services recevant des patients urgents . . . 26
1.3.1 L'imagerie médicale . . . 26
1.3.2 Le plateau de biologie . . . 28
1.3.3 Les soins continus . . . 29
1.3.4 Le bloc opératoire . . . 29
1.3.5 Les services d'hospitalisation . . . 30
1.4 Le système d'information . . . 31
1.5 Les unités de soins . . . 32
1.5.1 Les unités de chirurgie . . . 32
1.5.2 Les unités de médecine . . . 33
1.6 Synthèse . . . 34
1.6.1 Les difficultés de gestion des unités de soins . . . 34
1.6.2 L'engorgement des services d'urgences . . . 35
1.6.3 Courte discussion sur les différences entre établissements privés et établissements publics . . . 36
1.7 Gestion du flux de patients et prise en charge des imprévus : Méthodologie . . . 37
1.7.1 Constat de départ . . . 37
1.7.2 Influer sur le flux des urgences par le management des unités de soins . . . 38
1.7.3 Objectifs de la thèse . . . 39

Chapitre 2 Management des services de soins et prise en charge des urgences **41**
2.1 Management et modélisation des systèmes hospitaliers . . . 41
2.1.1 La recherche de la performance . . . 42
2.1.2 Modélisation hospitalière . . . 44
2.2 Management du flux de patients aux urgences . . . 45
2.2.1 Gestion du flux de patients entrants . . . 45
2.2.2 Gestion du processus de soins . . . 48
2.2.3 Gestion du flux sortant . . . 49

2.3 Management des ressources . . . 51
2.3.1 Management des ressources de diagnostic . . . 51
2.3.2 Ordonnancement du bloc opératoire . . . 52
2.3.3 Ressources humaines et logistique hospitalière . . . 53
2.4 Management des lits des unités de soins . . . 55
2.4.1 Gestion à long terme . . . 55
2.4.2 Gestion à moyen terme . . . 58
2.4.3 Gestion à court terme . . . 63
2.5 Conclusion . . . 64

Chapitre 3 Choix des dates d'admission des patients 67
3.1 Introduction . . . 67
3.2 Le problème de planification des admissions de patients électifs . . . 68
3.2.1 Modélisation formelle . . . 70
3.2.2 Évaluation des politiques de décisions . . . 71
3.3 Le problème de planification journalier . . . 71
3.3.1 Formulation mathématique . . . 72
3.3.2 Évaluation de la fonction objectif . . . 73
3.4 Estimation des besoins pour les futurs patients . . . 74
3.5 Différentes approches de résolution . . . 76
3.5.1 Utilisation du nombre moyen approximé de patients inconnus . . . 76
3.5.2 Stratégie basée sur des taux de service . . . 77
3.5.3 Approche par optimisation de Monte Carlo . . . 77
3.6 Expérimentations numériques . . . 79
3.6.1 Cas d'étude et données utilisées . . . 79
3.6.2 Choix du taux de services . . . 81
3.6.3 Taille de l'échantillon pour l'optimisation de Monte Carlo . . . 82
3.6.4 Comparaison des 3 stratégies journalières . . . 83
3.6.5 Résolution du problème global par les stratégies proposées . . . 85
3.7 Synthèse . . . 87

Chapitre 4 Optimisation de la clé de répartition des patients 89
4.1 Présentation et enjeux . . . 89
4.2 Objectifs et évaluation d'une clé de répartition des patients . . . 91
4.2.1 Optimisation journalière de la clé de répartition . . . 91
4.2.2 Optimisation globale de la clé de répartition . . . 92
4.3 Introduction aux méthodes de recherches stochastiques . . . 93
4.3.1 Recherche aléatoire . . . 93
4.3.2 Approximation stochastique . . . 94
4.3.3 Approximation stochastique par perturbation simultanée . . . 95
4.3.4 Estimation du gradient par analyse de perturbation . . . 96
4.4 Optimisation de la clé de répartition des patients par la méthode de perturbation simultanée 97
4.4.1 Projection d'un point dans l'espace des solutions . . . 98
4.4.2 Différentes variantes pour le problème de recherche par perturbation simultanée . . . 99
4.5 Optimisation de la clé de répartition de patients par l'analyse de perturbation . . . 102
4.5.1 Recherche de la meilleure direction améliorante . . . 105
4.5.2 Algorithme pour la recherche de la clé de répartition . . . 108
4.6 Comparaison des approches et performances . . . 108

4.6.1 Étude expérimentale de la convergence des approches . . . 109
4.6.2 Performance des approches . . . 110
4.7 Conclusion . . . 113

Chapitre 5 Affectation des patients aux unités de soins 123
5.1 Affectation des patients aux unités de soins . . . 123
5.1.1 Introduction . . . 123
5.1.2 Description du problème . . . 124
5.2 Modélisation . . . 126
5.2.1 Modèle journalier . . . 126
5.2.2 Optimisation de Monte Carlo . . . 128
5.2.3 Estimations des besoins futurs par unité . . . 128
5.3 Affectation des patients aux chambres . . . 129
5.3.1 Description du problème . . . 129
5.3.2 Modélisation mathématique . . . 131
5.4 Insertion des urgences . . . 132
5.4.1 Considérations . . . 133
5.4.2 Adaptation du modèle général . . . 134
5.5 Expérimentations . . . 135
5.5.1 Performance de l'approche . . . 135
5.5.2 Couplage avec le modèle de planification des admissions . . . 137
5.5.3 Une simulation du processus de gestion du flux de patients . . . 141
5.6 Conclusion . . . 142

Conclusion générale 145
Gestion des patients dans les unités de soins : synthèse . . . 145
Perspectives de recherche . . . 145
Vers une gestion globale de l'hôpital . . . 146

Annexes 151

Annexe A Le flux de patient du CH de Firminy en 2008 151
A.1 Activité des urgences . . . 151
A.2 Dimensionnement de l'hôpital et taux d'occupation . . . 151
A.3 Flux de patients dans les services . . . 152
A.4 Durées de séjour . . . 152

Annexe B La tarification à l'activité 157
B.1 Coûts de production d'un soin . . . 157
B.2 Groupe Homogène de Malade . . . 158

Liste de publications 163

Bibliographie 165

Table des figures

1.1 Parcours patient simplifié 24
1.2 Classification CCMU d'un patient 26
1.3 Parcours patient en fonction du degré d'urgence 27
1.4 Description d'un système autonome 39

2.1 Parcours patients au sein de l'hôpital 42
2.2 Les 3 causes d'engorgement des urgences (extrait de Asplin *et al.* (2003)) 46

3.1 Variation de la proportion de patients connus et inconnus 80
3.2 Impact du choix du taux de service 82
3.3 Convergence de l'optimisation de Monte Carlo 83
3.4 Converge de l'optimisation de Monte Carlo pour les hôpitaux 2 et 3 84

4.1 Chemin nominal et chemin perturbé pour un système à file d'attente 97
4.2 Translation d'un point de $\mathcal{E}$ dans l'espace des solutions 99
4.3 Illustration de la projection dans l'espace des solutions ; cas en dimension 2 100
4.4 Un mauvais jeu de paramètres pour SPSA 115
4.5 Le problème des optima locaux 116
4.6 Un bon jeu de paramètres pour SPSA 117
4.7 10000 itérations de SPSA 118
4.8 Convergence de la méthode PSCRG1 pour l'hôpital de type 1 119
4.9 Convergences de la méthode PSCRG1 pour les hôpitaux de type 2 et 3 119
4.10 Convergence de la méthode PSCRG2 pour l'hôpital de type 1 120
4.11 Convergences de la méthode PSCRG2 pour les hôpitaux de type 2 et 3 120
4.12 Convergence de l'optimisation par analyse de perturbation APCR 121

5.1 Convergence de l'optimisation de Monte Carlo pour l'hôpital de type 1 136
5.2 Convergence de l'optimisation de Monte Carlo pour les hôpitaux de type 2 et 3 137

A.1 Distribution des durées de séjour des patients réguliers de médecine 154
A.2 Distribution des durées de séjour des patients urgents de médecine 155
A.3 Distribution des durées de séjour des patients réguliers de chirurgie 155
A.4 Distribution des durées de séjour des patients urgents de chirurgie 156

B.1 GHM : Eléments décisionnels 158
B.2 GHM : Eléments décisionnels (suite) 160
B.3 Exemple d'arbre décisionnel de groupage 161

Liste des tableaux

1.1 Echelle de Glasgow pour un adulte 24
1.2 La classification CCMU modifiée 25

2.1 Les 3 niveaux de gestion des lits 55
2.2 Comparaison entre DSAP et HRP 60

3.1 Distribution des durées de séjour en médecine à l'hôpital de Firminy 80
3.2 Convergence de l'optimisation de Monte Carlo 82
3.3 Classement des stratégies pour les types d'hôpitaux considérés. 85
3.4 Résultats de simulation 86

4.1 Performance des approches pour l'hôpital de type 1 112
4.2 Performance des approches pour l'hôpital de type 2 113
4.3 Performance des approches pour l'hôpital de type 3 114

5.1 Convergence de l'optimisation de Monte Carlo pour l'hôpital de type 1 136
5.2 Performances de l'approche d'affectation journalière pour l'hôpital de type 1 138
5.3 Performances de l'approche d'affectation journalière pour l'hôpital de type 2 138
5.4 Performances de l'approche d'affectation journalière pour l'hôpital de type 3 139
5.5 Simulation des affectations aux unités de soins dans un contexte réel 139
5.6 Simulation avec admission des urgences une fois par jour 140
5.7 Utilisation des lits par unité et par type d'hôpital 141
5.8 Nombre de mutations par service et par type d'hôpital 142

A.1 Activité du service d'urgences 152
A.2 Répartition de lits et des chambres 153
A.3 Taux d'occupation global et taux de remplissage par les urgences pour l'année 2008 153
A.4 Entrées/sorties en unités de courts séjours 154
A.5 Répartition des flux d'entrée de patients en médecine 154
A.6 Répartition des flux d'entrée de patients en chirurgie 154

Liste des Algorithmes

3.1 Evaluation d'une politique d'admission 71
3.2 Construction d'une solution pour le problème de bin-packing 74
3.3 Génération d'un échantillon 75
4.1 Evaluation d'une clé de répartition : approche journalière 92
4.2 Optimisation par analyse de perturbation 97
4.3 Algorithme SPSA 98
4.4 PSCR 100
4.5 PSCRJ 101
4.6 PSCRG2 102
4.7 Direction de plus forte descente 106
4.8 APCR 108
4.9 Mesure du gradient par analyse de perturbation 109
5.1 Génération de demandes en lits des patients inconnus pour les unités de soins 130

Introduction générale

L'INDUSTRIALISATION de la gestion des établissements hospitaliers entraîne pour les sciences de gestion et d'ingénierie des systèmes l'émergence de nouvelles problématiques de recherche. Rigueur budgétaire, statisfaction des patients et gestion du temps sont parmi l'ensemble des éléments nécessitant la mise en place de nouvelles organisations et une gestion plus rigoureuse des processus dans les systèmes hospitaliers. Ce phénomène est renforcé par les évolutions sociétales de cette décénie écoulée. L'augmentation du nombre de patients et leur dépendance aux soins rendent la prise de décisions sur la gouvernance des établissement inévitable et délicate. La recherche opérationnelle et l'aide à la décision sont des sciences qui peuvent aider les hôpitaux dans leurs choix. Cependant, il est nécessaire de bien comprendre ces systèmes particuliers afin de rendre les méthodes de résolution efficaces.

Le besoin d'outils d'aide à la décision dans le milieu de la santé

Depuis maintenant quelques années, la fréquentation des hôpitaux est à la hausse. Cette augmentation du nombre de malades à traiter n'est pas sans soulever différents problèmes de nature organisationnelle aux dirigeant de ces établissements. En premier lieu, la hausse de la fréquentation des établissements de soins entraîne de manière inéluctable une hausse des dépenses dans les systèmes de santé. Cette hausse se retrouve amplifiée par les progrès de la médecine et de la pharmacologie proposant des traitements toujours plus efficaces mais également toujours plus coûteux. Pour freiner cette inflation, il a été demandé aux établissements de diminuer leurs dépenses. Cela se traduit en pratique par un besoin impératif de gérer au plus juste les ressources disponibles.

Cette rationalisation des ressources disponibles couplée à l'augmentation du flux de patients n'est pas sans causer de fréquents problèmes de surcharge des hôpitaux. Les effets de telles situations sont très néfastes. Les affaires d'erreurs médicales éclatant au grand jour dans la presse sont de plus en plus nombreuses. Le personnel des hôpitaux, devant traiter de plus en plus de patients avec une quantité limitée de ressources, devient sujet au stress et aux erreurs. Clairement, les hospitaliers ont besoin d'aide pour faire face aux évolutions que leur impose la société.

Ces deux faits entraînent deux problèmes de gestion principaux. Tout d'abord les problèmes liés au ressources vont être de savoir comment gérer au mieux les ressources disponibles et comment les répartir équitablement entre les différents acteurs. La deuxième catégorie de problème concerne le flux de patients. La question se posant est comment maîtriser ce flux ? Les réponses à ces questions ne sont pas simples à formuler tant elles dépendent de nombreux paramètres. C'est précisement dans ces situations que les sciences de l'information peuvent aider à apporter des éléments de réponse.

Ces sciences sont au coeur de tous les domaines depuis la fin du siécle précédent. La démocratisation de l'informatique auprès de chacun représentera certainement dans les années à venir la révolution technologique phare du début du XXI^{e} siécle. L'industrie s'est appropriée les formidables opportunités offertes par cet outil dès son apparition. Les avancées de la recherche, l'augmentation de la puissance de calcul et le perfectionnement des méthodes issues de domaines tels les sciences de gestion de production offrent

des pistes de progrès non négligeables pour le management des sytèmes. La recherche opérationnelle, se définissant comme une science cherchant à répondre à la question de comment répartir au mieux des ressources afin de réaliser un objectif et en tenant compte de conditions multiples, fait partie des sciences de gestion dont les progrès sont conséquents. Les systèmes de santé, par leur gestion de plus en plus proche des systèmes industriels, aspirent de plus en plus à profiter de ces innovations dans le management des ressources.

Objectifs scientifiques

L'objectif que nous nous fixons tout au long de cette thèse va être de développer une approche d'aide à la décision à destination des responsables de la gestion des hôpitaux. Cettre approche va mettre en place des démarches d'optimisation afin de proposer à chaque instant les choix les plus pertinents. Un tel outil doit être construit en respectant certaines règles sous peine d'être inapplicable en pratique. A notre humble avis, ces règles sont au nombre de trois. Nous les considérons comme garantes de la qualité scientifique d'une recherche dédiée au monde de la santé et vont constituer, pour nous, trois objectifs à atteindre tout au long de notre travail. Nous décrivons ces trois règles dans les paragraphes qui suivent.

1. Une gestion humaine. L'hôpital possède bien sûr de nombreux points communs avec les systèmes industriels qui le rendent gérable par des outils de gestion similaires. Cependant nous ne pouvons pas le gérer comme une usine. La dimension humaine sous-jacente à la définition du terme « hospitalité » ne doit pas être oubliée lors de l'élaboration d'un nouvel outil de gestion. Nous ne parlons pas de produits et de machines mais de patients et de services de soins. Ainsi, développer une méthodologie aboutissant à une prise de décision ne considérant que des critères économiques est vouée à l'échec. Chaque décision doit être orientée patient et personnel soignant. A ce titre, nous nous refuserons à parler de minimisation de coûts financiers mais nous préferons parler de pénalités quand une décision entraîne un inconfort humain. Tout au long de ce mémoire, nous chercherons à minimiser ces pénalités et ainsi à favoriser le bien être de chacun face à la difficulté que peut présenter un processus de guérison.

2. Faire face aux aléas. Une des caractéristiques principales des hôpitaux est la présence d'événements aléatoires. Ces événements surviennent à tous les niveaux et sont sources de modifications intempestives de l'organisation des services afin d'absorber les perturbations induites. Le personnel hospitalier est une ressource très adaptative qui avec l'expérience a appris à se réorganiser en fonction des évolutions du système les entourant. En réalité, cette qualité, qui est leur, vient du fait qu'ils ne savent que très peu agir autrement qu'en subissant les divers aléas. Bien sûr, des tentatives ont été effectuées afin de changer cette attitude, mais la complexité liée à l'évolution dynamique d'un système aussi imposant qu'un établissement hospitalier envoie rapidement ces essais au rang d'infructueux. Il est nécessaire, nous en sommes convaincus, de devenir maître de ces évenements aléatoires et perturbateurs afin de mieux les anticiper. Une approche d'aide à la décision dans ce type de système , à n'importe quel niveau, sera forcément à un moment où un autre confrontée à des aléas. Il apparaît comme primordial, lors de chaque prise de décision, de toujours considérer ce qui peut se produire ultérieurement.

3. Une gestion centralisée. L'organisation des systèmes hospitaliers fait apparaître, quand on y regarde de plus près, une multitude de sous-systèmes, généralement cloisonnés, ayant chacun leur fonctionnement propre, parfois concurrentiels et ne communiquant quasiment pas entre eux. Ce type d'organisation est caractéristique des systèmes de grande taille. Pourtant, en coopérant, en mutualisant certaines de leurs ressources, ils auraient tout à y gagner. Pour ce faire, il est nécessaire d'insérer un « arbitre » ; une entité capable de garantir la bonne coopération de chacun avec tous. Nous pensons que l'amélioration des performances d'un système hospitalier passe par une grande coopération des services entre eux et ce afin d'obtenir une plus grande flexibilité à l'échelle de l'établissement. La décision, la gestion doit être centralisée dans un module ayant la connaissance globale du système afin de choisir

comment distribuer les tâches à accomplir à chacun en fonction de ses possibilités.

Organisation de la thèse

Cette thèse va mettre en place une démarche visant à gérer les lits des services d'hospitalisation de court séjour. Elle est organisée en 5 chapitres principaux pouvant se regrouper en deux parties. La première partie, comprenant les deux premiers chapitres, constitue le point de départ de nos travaux et peut être vue comme un panorama de l'existant. Le premier chapitre va se focaliser sur une présentation générale des systèmes hospitaliers en prenant comme point de vue celui du patient des urgences. Cette démarche exploratoire va nous permettre de nous intéresser à la gestion du flux de patients dans les différentes unités composant un hôpital. Nous nous attacherons tout particulièrement à pointer les dysfonctionnements et proposerons très brièvement quelques pistes d'améliorations possibles face à ces difficultés. Le chapitre suivant va dresser un état de l'art sur les recherches menées et ayant pour but d'améliorer la gestion du flux de patients dans un établissement hospitalier. Ces deux chapitres sont organisés en trois temps forts :

1. Management du flux de patients aux urgences
2. Gestion des ressources présentes à l'hôpital
3. Management des unités de soins

Le chapitre 1 va conclure en partant des lacunes relevées dans l'existant et proposer la base de la méthodologie que nous allons mettre en place afin de gérer le flux de patients dans les services de soins. Une hiérarchie de problèmes à résoudre sera notamment proposée et justifiée. Nous exposerons pour chaque choix que nous faisons dans quel but nous l'avons fait.

Cette conclusion peut être considérée comme l'introduction à la deuxième partie de cette thèse qui va détailler une approche de gestion des lits sous incertitude. Le premier problème que nous aborderons est celui du choix des dates d'admission des patients réguliers (Chap. 3). Différentes approches seront proposées. Chacune de ces propositions vont considérer une manière différente de prendre en compte les évolutions futures du système. Nous allons cependant rencontrer, durant cette recherche d'estimation des besoins futurs, un sous-problème de taille conséquente. Dans le but de répartir les patients électifs, et donc de représenter leur date d'admission, nous utiliserons une clé de répartition. Le choix, ou plutôt la recherche d'une telle clé est un problème d'optimisation en soi. Le chapitre 4, commençant par une introduction aux méthodes de recherche stochastique, y sera consacré et proposera des approches basées sur deux méthodes d'approximation stochastique de gradient que sont l'analyse de perturbation et la perturbation simultanée.

Le chapitre suivant (Chap. 5) traitera de deux problèmes de nature opérationnelle. En premier lieu, nous nous intéresserons à l'affectation des patients aux unités de soins ou plus généralement aux étages de l'hôpital. Cette affectation se doit de respecter les capacités d'accueil de chaque étage au mieux, et, si le besoin en est, de répartir les dépassements occasionnels de capacité dans les services. C'est à ce niveau que nous introduisons les patients en provenance des urgences. Le modèle général proposé pour les patients réguliers sera adapté pour répondre aux spécificités des besoins d'admission des patients urgents tout au long de la journée. Dans les deux problèmes, choix des admissions et choix des unités d'accueil que nous venons de mentionner, lors d'une prise de décision, l'hôpital est considéré dans sa globalité. Nous étions donc jusqu'à présent dans le cadre d'une gestion centralisée du flux de patients. Le deuxième problème opérationnel va lui se cantonner aux unités de soins indépendamment les unes des autres. Les patients étant affectés à l'unité considérée doivent à présent être affectés à une chambre. Compte tenu des contraintes pouvant exister entre les patients et les différents types de chambre, nous résolvons un problème de couplage patients/chambres. La gestion du flux de patients tel que nous la considérons dans ce mémoire s'achève avec cette dernière étape. Ce dernier chapitre se terminera par une simulation de

la gestion du flux de patients en mettant les unes après les autres les différentes méthodes d'aide à la décision que nous avons proposées.

Finalement la conclusion générale va, outre les perspectives offertes par nos travaux, ouvrir vers des besoins d'outils de gestion encore plus évolués en incluant toujours plus de ressources afin de gérer au mieux les séjours des patients.

Chapitre 1

Gestion du flux de patients dans les systèmes hospitaliers

Sommaire

1.1 Introduction
1.2 Le flux de patients aux urgences
1.3 Description des services recevant des patients urgents
1.4 Le système d'information
1.5 Les unités de soins
1.6 Synthèse
1.7 Gestion du flux de patients et prise en charge des imprévus : Méthodologie

Ce chapitre présente et formalise la gestion du flux des patients au sein d'un hôpital. L'exploration des services de l'hôpital s'est faite en partant du service traitant les urgences. En suivant les différents cheminements des patients à partir de ce service, nous nous sommes intéressés au fonctionnement des différentes unités rencontrées par la suite. Cette analyse se base sur une observation de terrain. Deux établissements hospitaliers ont accepté de participer à cette étude. Le premier est un établissement privé, le Centre Hospitalier Privé de la Loire (CHPL), où l'activité est essentiellement composée d'actes de chirurgie programmée. Le deuxième est un établissement public, le Centre Hospitalier de Firminy, hôpital général de taille moyenne où l'activité principale provient de l'activité urgente. Le présent chapitre s'est principalement focalisé sur le fonctionnement de ce deuxième établissement. Cependant, même si une généralisation peut aisément en être tirée, une brève discussion sur les spécificités des différents types d'établissements hospitaliers sera exposée en toute fin.

1.1 Introduction

Les services d'urgences (SU) des hôpitaux, communément désignés sous le vocable de "urgences", doivent s'accomplir de leur rôle tel qu'il est défini dans la loi. La mission des SU des hôpitaux a été définie par l'article R.6123-18 du Code de la Santé Publique qui stipule que :« *Tout établissement autorisé à exercer l'activité d'urgences mentionnée au §3 de l'article R. 6123-1 est tenu d'accueillir en permanence*

dans la structure des urgences toute personne qui s'y présente en situation d'urgence ou qui lui est adressée, notamment par le SAMU. » La circulaire DHOS (Direction de l'Hospitalisation et de l'Offre de Soins) du 16 avril 2003 précise que « *les services des urgences ont pour mission de prendre en charge, en priorité, les besoins de soins immédiats, susceptibles d'engager le pronostic vital et/ou fonctionnel, qui exigent, quels que soient l'endroit ou les circonstances, l'intervention d'un médecin formé à la prise en charge des urgences et les besoins de soins urgents, qui appellent la mobilisation immédiate d'un médecin ayant les compétences et les moyens d'intervenir* ».

Les SU se doivent donc de répondre en permanence à une obligation de soins. Quel que soit la pathologie ou le traumatisme, quelle que soit l'origine de l'individu, 24h/24 et 7j/7 les soins prodigués doivent être du plus haut niveau possible. Tout doit être mis en oeuvre afin de garantir cette obligation juridique. Cela va de l'organisation des équipes soignantes, des soins aux patients à la mise en place des aides nécessaires à sa sortie en passant par tous les examens médicaux nécessaires afin de déterminer la pathologie dont souffre le patient et enclencher le processus de soins adapté. En particulier, les services d'urgences se doivent de maintenir en permanence une capacité d'accueil et un accès aux soins optimal. Ce cadre de fonctionnement est commun à tous les SU de tous les établissements de santé de France. Il y a quelques années, nous pouvions distinguer dans le paysage français des Services d'Accueil et de traitement des Urgences (SAU) qui étaient le niveau d'équipement le plus élevé, les Unités de Proximité d'Accueil, de Traitement et d'Orientation des Urgences (UPATOU) qui prenaient en charge des cas plus simples, les Pôles Spécialisés d'urgences (POSU) qui accueillaient des urgences lourdes et spécifiques. Les décrets du 22 mai 2006 ont supprimé ces notions. La circulaire DHOS/01/2007/65 du 13 février 2007 relative à la prise en charge des urgences fait désormais référence aux « structures d'urgences » autorisées par l'ARH qui ont pour obligation de répondre aux mêmes exigences de qualité en termes d'effectifs et de compétences médicales et paramédicales.

Du point de vue de l'hôpital en tant que structure, les services d'urgences représentent une porte d'entrée par laquelle transite le flux de patients classiquement dénommé "flux urgent". Il s'oppose à un deuxième flux d'arrivée de patients dit "flux programmé". Dans le jargon et suivant les personnes, ce flux peut être également appelé "flux de patients électifs", "flux de patients réguliers" ou encore "flux de patients en entrées directes". Peu importe le vocable retenu, ce flux représente les patients pour qui leur admission a été choisie et annoncée au préalable. Le flux de patients urgents, à l'opposé, est classiquement défini par sa nature imprévisible. La notion de "prévisible" est à prendre avec certaines précautions. Comme nous le verrons dans cette thèse, le flux programmé n'est pas entièrement prévisible, du moins au sens premier du terme. Il est seulement connu à l'avance tandis que le flux des urgences est connu en temps réel, au fur et à mesure des arrivées de patients.

1.2 Le flux de patients aux urgences

En France, les patients se présentant aux services d'urgences peuvent être classés en trois catégories en fonction de leur provenance :

1. les patients amenés par le Service Départemental d'Incendie et de Secours(SDIS)
2. les patients amenés par le Centre 15 (SAMU)
3. les patients se présentant d'eux-mêmes

Dans les deux premiers cas, le besoin d'être soigné en urgences a été évalué par des professionnels. Dans la troisième catégorie, on peut distinguer deux sous-catégories suivant que le patient soit adressé par un médecin ou non. Le nombre de patients se présentant à l'accueil des services d'urgences n'est pas équi-réparti entre ces catégories. Les patients se présentant d'eux-mêmes sans avoir pris d'avis médical

auparavant sont largement prépondérants ce qui entraîne de nombreuses dérives dans l'utilisation des urgences. Les patients présentant des pathologies mineures dite de "bobologie" (pouvant être traitée par la médecine de ville) vont être consommateurs de ressources dédiées aux vraies urgences. Nous venons de soulever un premier problème de gestion du flux de patients urgents. Nous en verrons d'autres.

La demande en entrée de ces services ne cesse de croître. Les causes de cette croissance sont multiples. Tout d'abord, le fait avéré du vieillissement de la population française, l'entrée de ce que les démographes appèlent le "papi-boom", provoque inévitablement une forte augmentation des passages aux urgences pour des pathologies liées à l'âge. Ces pathologies ne sont pas les seules en hausse au cours des dernières années. Les activités à caractère non-urgentes, relevant de médecine générale, représentent aujourd'hui une grande partie des soins délivrés. Des raisons peuvent être identifiées comme la facilité et la gratuité d'accès à des soins de qualité avec du personnel de haute technicité ou bien encore l'évolution général des moeurs de la population. Une grande partie de la population française, souvent rurale, a de plus en plus de mal à accéder aux médecines généralistes de proximité. La disparition des médecins de campagne, l'abandon des tours de garde devenus impossibles à mettre en place ou encore l'éloignement domicile-travail sont les principales causes menant les patients à se présenter aux urgences pour des pathologies mineures.

Paradoxalement, ces activités, bien qu'embolisant les SU, sont une source de revenus rentables pour l'hôpital. Les SU ont appris à travailler avec ce dilemme. Leur organisation fonctionnelle et le mode de prise en charge des patients, tels qu'ils sont présentés dans la suite de ce chapitre, en découlent.

1.2.1 Organisation des services d'urgences et Parcours patient

Le processus de prise en charge d'un patient aux urgences dépend de l'organisation de ce service. Cependant, certaines configurations comme celle décrite ci-après sont fréquentes en France. Depuis la mise en place de la tarification à l'activité, la T2A (voir annexe B), les hôpitaux sont obligés de définir précisément le contexte de soins dans lequel entre chacun de ses patients. Ainsi pour les urgences, un patient recevant un simple soin avant de rentrer chez lui est différent d'un patient devant être gardé une demi-journée en observation, qui est lui même différent d'un patient devant être hospitalisé par la suite. D'un point de vue général, la tarification à l'activité requiert une gestion des ressources au plus juste ce qui est source de nombreux problèmes d'organisations et d'optimisations.

En règle générale, les services d'urgences sont scindés en deux parties : une partie externe et une partie d'hospitalisation de courte durée (UHCD). La partie externe traite les pathologies "mineures" et peut être vue comme une unité de consultation externe[1]. La partie UHCD reçoit, quand à elle, les patients dont l'état de santé est plus préoccupant et nécessite une mise en observation dans un lit. Le parcours patients pour cette organisation est représenté simplement par le schéma Fig. 1.1.

Classification des patients urgents

Dans la pratique, tous les patients ne suivent pas exactement ce schéma. Le parcours patient au sein des services d'urgences est fortement lié à la pathologie ainsi qu'au degré d'urgences du malade. Dans le domaine médical, et plus particulièrement dans le traitement de l'urgence, l'échelle de Glasgow, ou score de Glasgow (Teasdale et Jennett, 1974) est très utilisé afin d'évaluer le niveau de conscience d'un patient. Cette évaluation se base sur l'évaluation de trois critères : ouverture des yeux, réponse verbale et réponse motrice. En fonction de la qualité de chacun de ces critères, un score est obtenu allant de 3 (coma profond ou patient décédé) à 15 (patient en parfaite santé). Le détail de cette échelle est donnée Table 1.1(source Wikipédia).

1. La phase de « consultation médicale » également dénommée « examen clinique » comprend un interrogatoire, un examen physique et éventuellement la réalisation d'un ECG, la mesure de saturation en dioxygène ou l'utilisation d'une bandelette urinaire.

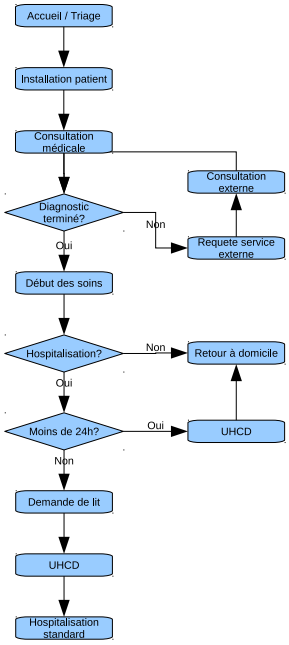

FIGURE 1.1 – Parcours patient simplifié

TABLE 1.1 – Echelle de Glasgow pour un adulte

Ouverture des yeux	Réponse verbale	Réponse motrice
1 - nulle	1 - nulle	1 - nulle
2 - à la douleur	2 - incompréhensible	2 - Extension stéréotypée (extension des extrémités des membres sup.)
3 - au bruit	3 - inappropriée	3 - flexion stéréotypée (flexion des extrémités des membres sup. et du tronc)
4 - spontanée	4 - confuse	4 - évitement (Retrait rapide avec éloignement face à l'agression)
	5 - normale	5 - orientée (mouvement tendant à faire disparaître la cause de la douleur)
		6 - aux ordres

Cependant cette évaluation présente certains biais. Par exemple une personne muette de naissance n'obtiendra au maximum qu'un score de 11. Il y a donc une nécessité d'éléments complémentaires qui ne sont pas forcément simples à obtenir dans une situation d'urgences. Une deuxième classification des patients urgents est proposée. Il s'agit de la Classification Clinique des Malades des Urgences (CCMU). Cette classification comprend dans sa version initiale 5 classes. C'est le médecin du SMUR ou l'urgentiste qui établit la classe d'appartenance d'un patient après examen clinique. Dans la version modifiée 2 classes ont été ajoutées (P et D) afin de pouvoir codifier tous les patients se présentant aux urgences (Table 1.2 et Fig. 1.2).

Parcours Patient

Cette classification nous permet d'établir plus en détail le parcours des patients au sein des services d'urgences en fonction de leur codification CCMU (Fig. 1.3). Le classement du patient dans cette codification est effectué par le médecin lors de la consultation. Cependant afin d'aiguiller les patients dès leur

TABLE 1.2 – La classification CCMU modifiée

CCMU	Définition	Exemples
P	Patient présentant un problème psychologique et/ou psychiatrique en l'absence de toute pathologie somatique	
1	L'état du patient est jugé stable sans besoin d'actes complémentaires pour l'établissement du diagnostic.	Maladies Bénignes : Angine, malaise vagal, ...
2	L'état du patient est jugé stable. Des actes complémentaires au diagnostic (radiographie, prise de sang, ...) ou thérapeutique (suture,réduction,...) sont nécessaires.	Sutures de plaie simple, fracture fermée, hypotension, brûlure 2^{eme} degré, ...
3	Patient dont l'état est susceptible de s'aggraver aux urgences	AVC (Glasgow >13), angor, fracture ouverte, appendicite, ...
4	Patient dont le pronostic vital est engagé sans besoin de réanimation immédiate	AVC (8<Glasgow<13), défaillance respiratoire, ventre de « bois », coma éthylique, ...
5	Patient dont le pronostic vital est engagé nécessitant des soins de réanimation immédiat	défaillance respiratoire aiguë (nécessitant intubation, trachéotomie, ...) défaillance circulatoire aiguë (nécessitant défibrillation, transfusion massive, ...), Glasgow < 8,
D	Patient décédé. Pas de réanimation entreprise	

arrivée, une estimation du degré d'urgence est faite lors de l'étape de triage par l'infirmière d'accueil et d'orientation (IAO). Les patients dont le pronostic vital est engagé sont placés en salle de déchoquage où se trouve le matériel nécessaire à la réanimation. Les autres patients doivent attendre avant d'être progressivement admis. Cette admission se fait en fonction de l'évaluation donnée par l'IAO et de la libération des places dans le service.

Dans l'étape suivante, la pratique diffère de la théorie. Administrativement, les patients sont d'abord vus en consultation externe dans un box. Ensuite, en fonction de leur pathologie, soit ils restent dans le box de consultation pour recevoir leurs soins soit ils sont admis en UHCD. Dans la réalité, les patients, dont on sait par expérience que leur passage aux urgences va durer un certain temps (attente examens complémentaires, soins longs, besoin d'hospitalisation évident), sont directement admis en UHCD où aura lieu la consultation.

A l'issu de cette consultation, dans la plupart des cas, des examens complémentaires se révèlent être nécessaires. Certains comme un bilan sanguin ou une radiographie osseuse sont extrêmement fréquents. La liste des examens complémentaires pouvant être pratiqués est longue. Notons que certains de ces examens ont lieu dans le service des urgences (ECG, bandelette urinaire, ...) et que d'autres nécessitent l'utilisation de ressources relevant d'une autre unité de l'hôpital (scanner, échographie, consultation spécialiste, ...). D'autres examens peuvent être demandés à mesure que les résultats sont connus et ce dans le but de diagnostiquer la pathologie [2] dont souffre le patient.

Le processus de soins peut alors débuter. Si le patient est stable, il est alors prêt à quitter le service. Dans le jargon des urgences, on dit que le patient est décisionné. Autrement dit, la pathologie diagnostiquée, la décision de garder le patient à l'hôpital ou de lui permettre de rentrer chez lui est prise par le

2. même si un diagnostic est évident, comme cela est le souvent le cas pour des traumatismes telle une fracture osseuse, l'obligation de moyen impose la réalisation de l'examen permettant de l'établir ; réalisation d'une radiographie dans le cas d'une fracture

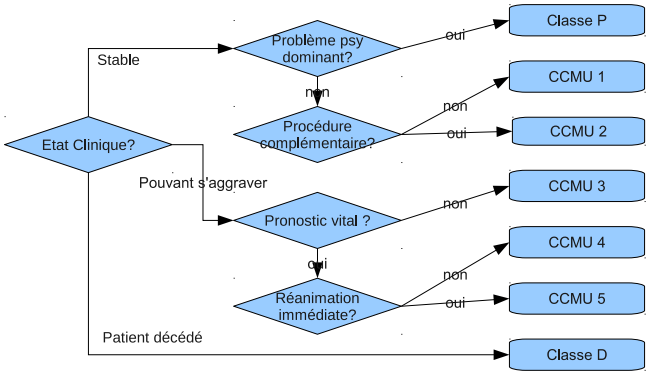

FIGURE 1.2 – Classification CCMU d'un patient

médecin. Dans le cas d'une décision d'hospitalisation, la dernière tâche consiste en la recherche d'un lit d'hospitalisation. Une période de négociation entre les responsables du SU (médecins et cadre de santé) et les cadres de santé des unités de soins va commencer afin de trouver une place pour le patient en attente aux urgences. Cette dernière phase est très problématique. Nous aurons l'occasion d'y revenir maintes fois dans ce mémoire.

1.3 Description des services recevant des patients urgents

Les services des urgences sont loin d'être des systèmes autonomes. Ils sont dépendants, tout au long de la prise en charge de leurs patients, de nombreuses ressources présentes soit sur place à l'hôpital soit sur des sites externes. D'une manière générale, chaque appel à un service externe est source d'allongement du temps de passage des patients et donc de perte de performance pour les urgences. Par ailleurs, tous les services devant traiter des patients des urgences doivent le faire en sus de leurs activités et fonctionnement qui leurs sont propres. Chaque patient ou tâche supplémentaire non-prévue les oblige à adapter leur planning de façon à le ou la prendre en charge. Nous allons détailler le fonctionnement des principales unités concernées par cette problématique et essayer de voir où se situe le problème réellement.

1.3.1 L'imagerie médicale

Le plateau d'imagerie médicale regroupe en plus de la radiologie conventionnelle des appareillages plus spécifiques tel un scanner. Il est au carrefour entre trois flux : interne, externe et urgences. Autrement dit, cette ressource est utilisée aussi bien par les services d'hospitalisation, qui souhaitesuivre l'état de guérison de leurs patients, par des rendez-vous de patients extérieurs adressés par un médecin de ville, que par les urgences afin d'établir le diagnostic d'un patient.

Le planning est fait pour permettre à ces trois flux de coexister sans se croiser. Pour des raisons d'hygiène, il est peu souhaitable qu'un patient externe croise un patient hospitalisé. Les deux risquent de se contaminer mutuellement ; le patient externe risquant de transmettre un virus (telle la grippe ou la gastro-entérite) qui viendra s'additionner à la pathologie du patient hospitalisé ; le patient hospitalisé risque, quand à lui, de transmettre une bactérie résistante contractée à l'hôpital au patient externe.

Les équipements d'imagerie, notamment le scanner, sont très onéreux. L'hôpital est souvent copropriétaire de ces équipements avec des radiologues de ville. Ceux-ci ont en contrepartie des plages d'utilisation qui leur sont réservées ce qui ajoute une contrainte pour l'établissement d'un planning efficace.

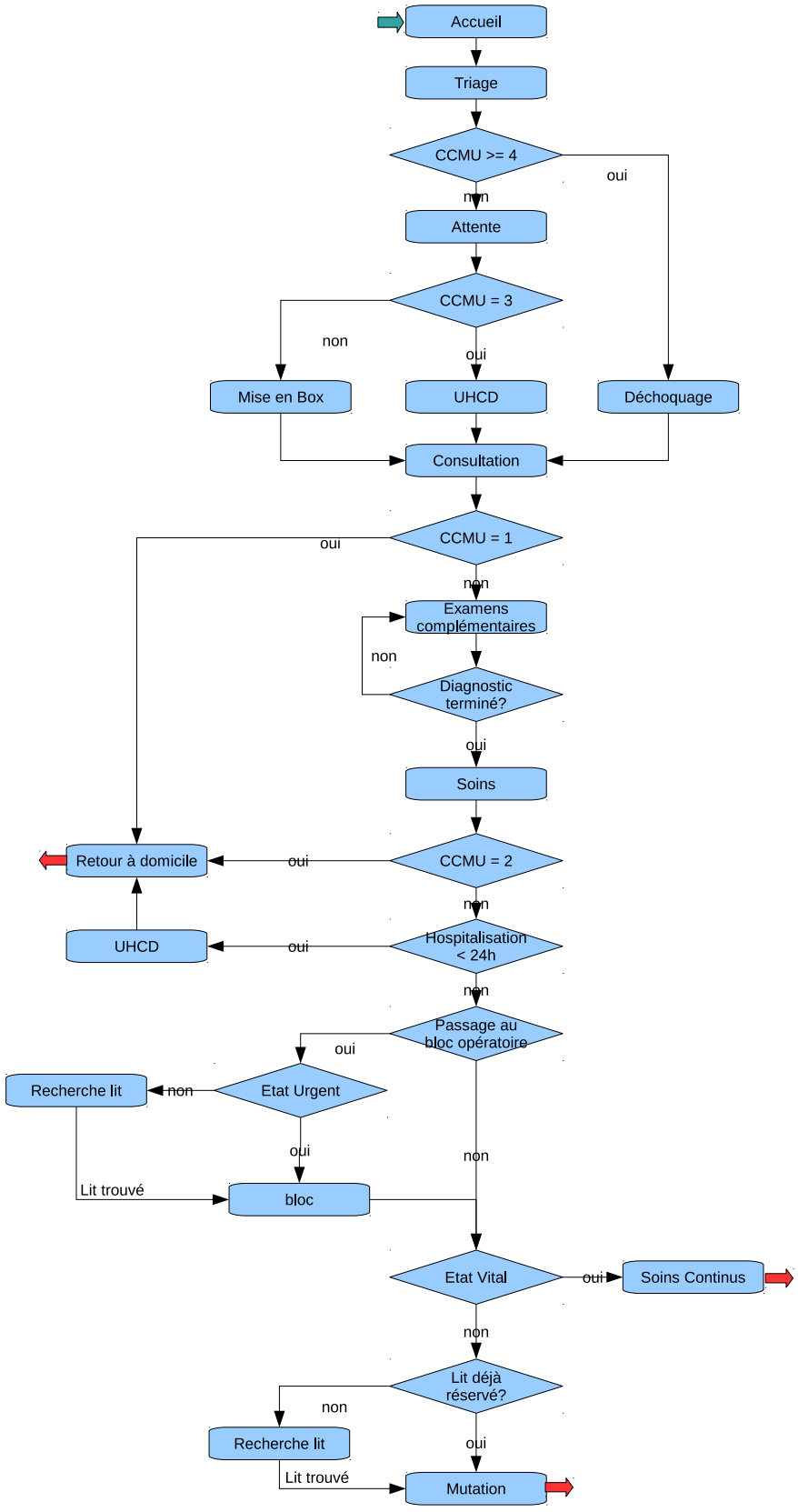

FIGURE 1.3 – Parcours patient en fonction du degré d'urgence

Une annexe de radiologie osseuse est présente au sein même du service des urgences. Elle est composée d'une équipe médicale de manipulateur en radiologie et de matériels radiologiques. Sa vocation est de faire passer les examens les plus courants aux patients urgents et ainsi d'éviter de perturber le planning du plateau d'imagerie. En période de garde, c'est l'équipe de cette annexe qui fait passer les examens plus complexes comme les scanners.

Pour la prise en charge des patients des urgences, mis à part pour les examens de radiographie osseuse, une étape de négociation entre les deux services est actuellement inévitable. Le différent oppose l'urgentiste qui veut faire passer son patient le plus rapidement possible et le service de radiologie qui se doit de respecter son planning, par respect des patients ayant rendez-vous . Il en résulte souvent que les examens demandés par les urgences ont lieu pendant les plages de fermeture théorique du service d'imagerie ; autrement dit pendant la pause de midi ou après 18h. Il existe probablement des solutions alternatives à cette attente. Cependant, elles ne sont pas forcément évidentes à élaborer d'un simple coup d'oeil sur un planning mural. Pendant les plages dédiées aux patients internes à l'hôpital, le problème de l'insertion et du décalage est moins contraignant. En effet, le patient étant déjà présent dans l'hôpital et dans sa chambre, il suffit de décaler l'horaire à laquelle les ambulanciers l'amèneront afin de faire passer en priorité un patient urgent. En généralisant le propos, il s'agit ici de trouver une méthode, une politique, d'insertion du patient urgent dans le planning en limitant les perturbations provoquées par celui-ci. Une certaine souplesse dans les planning doit être trouvée afin de permettre le retour à la normale le plus rapidement possible après la réalisation d'un examen non prévu initialement.

1.3.2 Le plateau de biologie

Le laboratoire d'analyses médicales, avant de pouvoir se mettre au service des unités de l'hôpital, doit répondre à de lourdes contraintes juridiques. Ces contraintes sont la cause, comme nous le verrons plus loin, de certains problèmes du point de vue des urgences.

L'activité du laboratoire dépend grandement de la période de la journée. Le matin, tous les services envoient les bilans sanguins de leurs patients ce qui est source d'une grosse charge de travail. Les analyses sont effectuées, soit par un technicien en analyses biomédicales soit par un automate d'analyse. Chaque échantillon est identifié par un code barre permettant de faire le lien avec le dossier médical du patient. Une fois l'analyse effectuée, les résultats sont disponibles via une interface Web consultable dans les services. Cet affichage doit être considéré seulement comme une indication du résultat final. Les données affichées doivent été vérifiées par le médecin biologiste qui s'assure que toutes les analyses ont été faites correctement et qu'aucune contamination externe (tube non stérile, ...) ne fausse le résultat. La validation intervient uniquement lors de la signature de la feuille de résultats ce qui en fonction de la disponibilité du biologiste peut avoir lieu plusieurs heures après la fin de l'analyse.

Pour les analyses effectuées sur place, le laboratoire s'engage à fournir les résultats dans la journée. Généralement, ils sont obtenus en 3h de temps. Pour les analyses devant être effectuées dans un laboratoire extérieur, il faut attendre le retour du résultat sur papier. La loi interdisant de recopier le résultat d'une analyse faite par un autre laboratoire, il n'est donc pas possible actuellement d'avoir accès aux résultats d'analyses faites à l'extérieur sur l'intranet.

Concernant les délais d'obtention des résultats, seul le début de matinée pause problème. L'activité à ce moment de la journée est très importante. La laboratoire de biologie doit en effet traiter l'afflux de prélèvements sanguins réalisés à jeun dans les services d'hospitalisations. Pour un bilan envoyé à 8h, les résultats ne sont fréquemment disponibles qu'en fin de matinée alors que 1h30 suffisent l'après midi. En théorie, la réceptionniste, en enregistrant un dossier patient pour les urgences, peut saisir le caractère urgent de l'analyse. Les tubes correspondants sont ensuite automatiquement traités prioritairement par l'automate. Dans la pratique, le passage prioritaire des analyses sur les automates n'est que rarement effectif pour divers raisons. La seule façon actuelle d'avoir des résultats plus rapidement pour les urgences

est de prévenir le laboratoire qu'il doit faire passer en priorité telle ou telle analyse. C'est ensuite le technicien en analyses biomédicales qui placera le tube en premier sur l'automate. L'action permettant de réduire le temps de présence d'un patient au service des urgences est donc entièrement manuelle.

1.3.3 Les soins continus

Le service de soins continus reçoit des patients nécessitant une surveillance accrue mais ne relevant pas de la réanimation. La limite entre les rôles et la définition de ces deux types de services est floue et mal définie par la loi. Elle reste donc assez subjective et dépend de l'appréciation du médecin. La règle d'admission d'un patient dans ce type d'unité peut être formulée, quelque peu ironiquement, comme suit : "*Un patient pour être accepté en soins continus doit être suffisamment malade mais pas trop*".

A Firminy, le service comporte 8 lits et la durée moyenne de soins dans ce service est d'environ 4 jours. Au bout de ce délai, le patient est transféré en réanimation, dans un autre établissement si nécessaire, ou bien il est transféré en unité d'hospitalisation. Il est rare dans ce type de médecine que l'état de santé du patient reste longtemps comme il était à son admission.

Dans le cas positif d'une amélioration, la sortie du patient se fait si possible dans le service le plus adapté à la pathologie. Dans tous les cas, il y a une coordination avec les urgences dans le processus de recherche de lit d'hospitalisation. Les sorties de ce service sont généralement prévues 24 ou 48h à l'avance. Dans ce cas, l'admission dans un autre service peut se faire comme dans le cadre du flux de patients programmés.

1.3.4 Le bloc opératoire

Le bloc opératoire représente une ressource centrale de l'hôpital. Il fait partie de l'image que l'on se fait d'un patient hospitalisé. Il est composé essentiellement de salles opératoires et d'une salle de réveil. A Firminy, il y a 5 salles opératoires auxquelles une spécialité est associée.

1. Orthopédie
2. Viscérale
3. ORL / Stomatologie
4. Gynécologie
5. Césarienne

Les chirurgiens se partagent les plages d'opération. Par exemple, en chirurgie viscérale, 3 chirurgiens ont chacun un jour de la semaine (de lundi à mercredi) qui leur est réservé. Les deux autres jours de la semaine, la salle présente des créneaux dit "partagés" où chacun selon ses besoins peut ajouter des patients. Ces plages sont également utilisées par des spécialités minoritaires n'ayant pas de salle dédiée. C'est le cas entre autres de l'ophtalmologie. Le cadre de santé veille à ce qu'aucun débordement n'ait lieu dans les flexibilités accordées. Les chirurgiens doivent déjà remplir leurs plages dédiées avant d'utiliser les créneaux partagés.

La salle de césarienne doit rester en permanence disponible dans l'éventualité d'un accouchement devant être effectué en urgence. Cette salle bien que spécifique est parfois utilisée pour une opération urgente d'une autre spécialité. Dans ce cas le cadre de santé du bloc opératoire avant d'autoriser cette intervention s'assure qu'une autre salle est, ou sera, libre rapidement pour recevoir une éventuelle césarienne. Cette salle restera bloquée jusqu'à ce que la salle où ont lieu normalement les césariennes soit à nouveau disponible (25% des accouchements de l'hôpital de Firminy ont lieu par césarienne ce qui est au dessus de la moyenne nationale qui est d'environ 20%). En considérant la moyenne nationale des accouchements

ayant lieu par césarienne, pour l'hôpital de Firminy, cela représente environ 300 accouchements par an soit près de 1 par jour. En ajoutant qu'un accouchement dans ces conditions utilise la salle opératoire pendant 2 à 3h, il est aisé de voir le besoin de garder cette salle disponible en permanence.

La salle de réveil avec 6 lits est de dimension assez petite. La recommandation est de 1,5 lits de réveil par salle d'opération, soit dans notre cas 8 lits. De plus, les patients sont parfois admis dans cette salle pour être préparés à leur intervention. C'est le cas par exemple de certaines anesthésies locales qui débutent 1h avant l'opération.

En temps normal, un lit d'hospitalisation dans un service de chirurgie est préalablement affecté au patient opéré dans le cadre de son hospitalisation. Dans le cas des urgences, c'est le service des urgences lui même qui doit trouver le lit d'aval avant d'envoyer son patient au bloc. Il arrive parfois qu'à la fin de l'intervention, ou plus précisément du réveil, le lit réservé ne soit pas encore disponible. Dans ce cas de figure le patient doit attendre en salle de réveil. Il peut être gardé au maximum pendant 3h. Avant de débuter l'intervention, le cadre s'assure du respect de cette règle. Si aucun lit n'est disponible dans ce délai, l'intervention est alors décalée.

Les urgences sont insérées dans le planning du bloc en fonction de la gravité de l'état du patient et en fonction de la disponibilité des lits d'aval. Les cas d'urgence vitale, en général des importants polytraumatismes, sont rares (quelques cas pas an). Ils sont traités prioritairement par le proche CHU qui possède des ressources plus adaptées. On est donc plus souvent en présence de semi-urgences ; des interventions non programmées mais qui peuvent être repoussées de quelques heures. Elles représentent tout de même 50% de l'activité du bloc et ne sont donc pas négligeables. Le planning du bloc opératoire est élaboré de façon à pouvoir prendre en charge ce flux quotidien.

Le passage par le bloc opératoire n'est pas fondamentalement, ici, un problème pour les urgences. Le problème, pouvant entraîner une présence non pertinente dans le service, se situe au niveau de la recherche du lit d'hospitalisation postopératoire. Nous sommes donc ramenés au problème de gestion des unités de soins.

1.3.5 Les services d'hospitalisation

Les services d'hospitalisation doivent réussir à gérer leur planning d'occupation des lits en fonction des patients déjà présents, des sortants, des arrivées programmées et des patients en attente aux urgences. Au problème d'affectation d'un patient à une chambre est associé quelques contraintes classiques comme la non mixité des chambres, éviter de faire se côtoyer des personnes présentant des pathologies incompatibles et essayer de respecter d'autres incompatibilités possibles (âge, religion, ...). Notons au passage que l'on ne prend pas en compte ici les souhaits des patients concernant l'hébergement en chambre simple ou double. Les chambres simples sont en général et avant tout réservées pour des patients dont l'état de santé nécessite une mise à l'écart.

Une certaine mutualisation des lits existe entre les différentes unités. Même si dans la mesure du possible un patient est affecté là ou le personnel sera le plus compétent pour les soins requis, il arrive que pour des raisons de manque de place, il soit affecté dans un autre service. Cette organisation bien que présentant de gros avantages dans la souplesse offerte pour la gestion des lits présente tout de même quelques inconvénients pour le personnel. Tout d'abord, le personnel infirmier ne connaît pas nécessairement les protocoles à exécuter. Il faut souvent contacter le service de la spécialité pour avoir confirmation du protocole à suivre. D'autre part, les médecins sont eux obligés de parcourir les différents étages de l'établissement pour voir leurs patients. Affecter un patient, dans un service qui n'est pas celui ou théoriquement il devrait être, est donc générateur d'un coût supplémentaire en terme de temps de travail, de risque d'erreurs, ... Cependant, actuellement, une fois affecté dans un service le patient ne sera pas retransféré pour rejoindre le service auquel il devrait théoriquement se trouver. Des exceptions sont parfois faites pour les patients relevant des services d'un autre bâtiment de l'hôpital.

Une fois les sorties puis les entrées programmées effectuées, les médecins urgentistes doivent faire avec les places restantes dans les services pour placer leurs patients. Compte tenu de l'organisation actuelle du travail dans les services, les mutations de patients depuis les urgences n'ont lieu qu'à partir de 14h. Il est pour l'instant quasi-impossible de placer un patient le matin. De plus, la gestion prévisionnelle est quasi-inexistante. Les planning d'occupation des lits sont très peu fait en tenant compte des entrées et des éventuels besoins en urgences du lendemain.

Le problème s'accentue le week-end. En effet, outre la fermeture de certains services (hospitalisation de semaine) diminuant le nombre de places disponibles, il est très compliqué de placer un patient. L'absence des médecins et des cadres fait que les sorties sont comme suspendues en fin de semaine. Pire, le lundi matin personne ne peut dire réellement combien de places vont être disponibles. Pendant ce temps les patients s'accumulent aux urgences. L'activation des lits supplémentaires est monnaie courante les dimanche et lundi. Même si la possibilité est offerte et permet indéniablement de libérer de précieuses places au sein de l'UHCD, sa pratique pose des problèmes d'organisation. Ajouter un lit dans une chambre double augmente considérablement les risques dus à la cohabitation de trois patients dans un lieu prévu pour les soins de seulement deux malades. De plus, il est souvent nécessaire de déplacer certains patients afin de pouvoir en accueillir un autre.

1.4 Le système d'information

Le Système d'Information Hospitalier (SIH) couvre l'ensemble des informations utilisées dans un établissement de santé. C'est une branche particulière des systèmes d'information. Leur rôle premier est de permettre à celui qui en a besoin d'obtenir l'information souhaitée. Dans le cadre des SIH, le système doit en particulier contenir les informations relatives aux patients, le dossier patient informatisé, qui sont soumises à l'approbation de la CNIL. Ces informations peuvent être présentées suivant plusieurs formes selon que l'utilisateur soit un cadre de santé qui souhaite gérer son service, le service facturation lors du calcul de la facturation d'un séjour, le département d'information médicale qui réalise des statistiques pour la T2A et ainsi de suite pour tous les acteurs de l'hospitalier. D'autres informations peuvent être gérées par le biais de modules de gestion de l'utilisation des ressources, de modules financiers, ...

Les SIH permettent donc de centraliser les informations nécessaires et de les rendre disponibles à chacun dans le but de remplir son rôle dans l'établissement. Le système d'information présent au CH de Firminy permet, entre autres, de connaître l'occupation des lits de l'hôpital et de faire de la gestion prévisionnelle. La possibilité est offerte de saisir des pré-entrées tout comme il est possible d'anticiper les sorties des patients en suivant l'état d'avancement de leurs soins. Malheureusement la prise en compte de toutes ces informations se révèle compliquée car devant être réalisée, dans la pratique, manuellement. La fonctionnalité n'est pas utilisée faute de pouvoir en tirer des bénéfices.

Pour la gestion du flux de patients, la seule utilisation généralement faite des SIH consiste à renseigner l'occupation des lits par les patients même si cette saisie est effectuée avec un différé de quelques heures. Afin de faire, ce qui peut être qualifié d'une "bonne" gestion des flux de patients, ils semblent important de s'accomplir de deux tâches : renseigner les informations dès qu'elles sont disponibles et faire de la gestion prévisionnelle.

Le renseignement de l'information est primordial. Sans la connaissance, il est impossible de pouvoir planifier quoi que ce soit à court, moyen ou long terme. L'information doit être connue dès que possible afin de pouvoir réagir au plus tôt tant sur le planning réel que sur le prévisionnel. Cela commence par le renseignement de l'occupation des lits en "temps-réel" ou, de manière plus réaliste, dans l'heure. Ainsi, la tâche de connaître la disponibilité des lits s'en trouve grandement simplifiée. Certains répondent à cette affirmation que, dans certains cas, il n'est pas judicieux de mettre un patient des urgences dans un lit libre l'après-midi. Ceci est justifié par le besoin de ce lit le lendemain matin pour une entrée programmée. La

saisie des pré-entrées dans le logiciel prend alors tout son sens car elle permet d'informer tout le monde sur les besoins internes au service.

Enfin, la gestion prévisionnelle, va permettre de se prémunir contre des situations indésirables conduisant à la paralysie des services. Chaque décision doit être prise en tenant compte de l'impact et des répercussions qu'elle va avoir. Humainement parlant, ce type de gestion est irréaliste. Ces informations, permettant la prise de décision, doivent être décidées en automatique par le SIH, à partir des informations connues et donc renseignées.

1.5 Les unités de soins

La Section 1.3.5, a présenté brièvement le fonctionnement des unités de soins ainsi que les problèmes que peuvent poser l'hospitalisation d'un patient en attente aux urgences. Voyons à présent plus en détail comment ces unités sont organisées. Intéressons nous notamment aux entrées directes : comment sont elles décidées ? comment sont-elles gérées au sein du service concerné ? De quelles souplesses pouvons nous bénéficier ?

Dans cette section, ne seront prises en considération que les unités recevant des patients urgents. En particulier, les services de longs et moyens séjours dont la gestion relève de problèmes très différents seront exclus de cette description. De même parmi les unités recevant des urgences, dites de court séjour,les unités de gynécologie, d'obstétrique et de pédiatrie, du fait de leurs spécificités médicales, ont des fonctionnements séparés des unités de soins conventionnelles et seront donc également exclues.

1.5.1 Les unités de chirurgie

Même si des services similaires peuvent exister dans certaines unités de médecine, quatre familles de services recevant des patients sont identifiables en chirurgie :

1. Consultation
2. Chirurgie Ambulatoire
3. Chirurgie de Semaine
4. Hospitalisation complète

La consultation est une étape impérative avant n'importe quel acte nécessitant une hospitalisation aussi courte soit elle. Le schéma de passage des patients aux urgences (Fig. 1.3) commence également par cette étape avant qu'une décision ne soit prise. Dans le cas particulier de la chirurgie, il est courant que ce soit le chirurgien lui-même qui décide de la date d'admission du patient. Cette date est forcément à mettre en corrélation avec l'organisation des blocs opératoires. Une organisation fréquemment rencontrée est celle dont nous avons discuté précédemment (section 1.3.4). Les chirurgiens possèdent des créneaux d'opérations qui leur sont alloués et durant lesquels ils programment leurs patients comme ils le désirent. Il est à noter que lors du choix de la date d'hospitalisation, les cadres de santé des services destinataires ne sont pas systématiquement consultés sur leur capacité à prendre en charge un patient supplémentaire ce qui peut entraîner des situations indésirables voir même problématiques. Si plusieurs chirurgies lourdes en terme de soins infirmiers sont programmées le même jour, il va s'en suivre des difficultés évidentes pour le personnel soignant de s'affranchir de leur charge de travail quotidienne.

En fonction de l'importance de l'acte chirurgical à réaliser, les patients sont affectés à l'un des 3 services restants. Sont entre autres affectés en chirurgie ambulatoire les patients dont l'intervention ne nécessite pas de garder le patient la nuit suivante. Ces services ne recevant pas de patients urgents, nous n'irons pas plus loin sur leur fonctionnement.

Les services d'hospitalisation de semaine et d'hospitalisation complète doivent eux s'organiser en fonction du flux de patients provenant des urgences. Le service de chirurgie de semaine est, comme son

nom l'indique, un service ouvert pendant la semaine et fermé le week-end. Toute l'organisation d'un service de ce type en découle. L'annexe A nous permet de voir que les durées de séjour en chirurgie sont généralement courtes. Une répartition dans les différents services de chirurgie peut être faite en considérant les types de chirurgie suivants :

- Chirurgie longue : durée de séjour supérieure à 5 jours
- Chirurgie courte : durée de séjour entre 3 et 5 jours
- Chirurgie très courte : durée de séjour entre 1 et 2 jours
- Chirurgie ambulatoire : durée de séjour inférieure à 1 jour

Les estimations des durées de séjours sont évaluées nationalement pour chaque type d'intervention et sont référencées dans la T2A. Les services de chirurgies possèdent la particularité de bien maîtriser ces durées. Les seules sources de dérogation à ces estimations sont dues aux difficultés à trouver une place dans un établissement de soins de suites quand la nécessité médicale l'impose et ne permet pas de passer par des alternatives comme les soins à domicile.

Les chirurgies longues ne peuvent être affectées qu'en service d'hospitalisation complète. Les chirurgies courtes et très courtes sont généralement affectées en service de semaine. Bien entendu, les chirurgies courtes sont préférentiellement admises en début de semaine ; les chirurgies très courtes en fin de semaine. De cette manière, on peut s'assurer que d'une part le service soit vide à la fin de la semaine et que d'autre part, le niveau de remplissage du service se maintienne tout au long de la semaine. Dans la théorie, avec une telle organisation, toutes les urgences devraient être admises en hospitalisation complète et ce même si elles relèvent de chirurgie courte. Le service de semaine ne recevrait alors que des patients programmés dont les durées de séjours sont maîtrisées.

Cependant, la proportion de patients relevant de chaque type de chirurgie varie d'une semaine à l'autre en fonction des planifications faites par les chirurgiens. Il est donc rarement possible de respecter l'organisation théorique. De plus lors d'une demande des urgences, si le service d'hospitalisation complète n'a aucune place disponible, le patient sera alors placé en service de semaine même s'il devra être déplacé à coup sûr en fin de semaine pour permettre la continuité de ses soins.

Pour conclure sur cette description des services de chirurgie, d'après l'annexe A, les services de chirurgie d'un hôpital comme celui de Firminy reçoivent en moyenne autant de patients programmés que de patients urgents.

1.5.2 Les unités de médecine

Les unités de médecines possèdent, a priori, plus de flexibilité sur le choix des dates d'admissions que celles de chirurgie. Elles ne sont pas directement dépendantes du choix du médecin d'hospitaliser un patient. En pratique, les médecins, de l'hôpital ou de ville, voient les patients en consultation puis éventuellement décident d'une hospitalisation suivant le cas de figure. Ils transmettent ensuite une liste de patients à admettre au cadre de santé concerné afin que ce dernier prévoie l'admission du patient. Les hospitalisations peuvent donc être davantage programmées puisque les cadres possèdent une liste de patients en attente d'admission.

Lors de l'admission d'un patient en chirurgie le choix est fait par le chirurgien en fonction de la disponibilité d'une ressource : le bloc opératoire. Dans le cas des patients de médecine, le choix est souvent fait en fonction de plusieurs ressources nécessaires tout au long du séjour du patient. Ainsi, l'admission d'un patient dans ces unités de soins doit coïncider avec une série de réservations de ressources annexes comme notamment des ressources d'exploration fonctionnelle utilisées pour le suivi de guérison du patient.

La gestion des entrées directes n'est pas chose aisée puisqu'il faut non seulement avoir un lit pour

accueillir le patient mais également que d'autres ressources aient des créneaux libres pour faire passer le-dit patient. Le problème de l'accueil des patients urgents s'en trouve d'autant plus compliqué. En effet, la première difficulté se posant est la même que pour tous les services d'un hôpital : avoir la capacité en terme de lits de recevoir des patients urgents. La deuxième difficulté repose sur les besoins d'examens complémentaires tout au long du séjour. Même si le patient a pu recevoir aux urgences les diagnostics établis normalement lors de l'accueil d'un nouveau patient dans un service de médecine, les éventuels examens ayant lieu durant le séjour doivent être planifiés rapidement. On n'a plus ici le délai avant l'admission d'un patient régulier qui permettait d'ajuster les réservations de ressources externes afin de les faire concorder ensemble.

Autant en chirurgie les admissions sont le plus souvent programmées longtemps à l'avance (plus d'un mois pour de l'orthopédie, environ 15 jours pour du viscéral), autant en médecine les admissions directes relèvent le plus souvent de semi-urgences[3]. Les problèmes médicaux ne se traitent pas comme les problèmes chirurgicaux. Prenons l'exemple d'un patient en déséquilibre diabétique. S'il n'est pas admis rapidement pour ajuster son traitement, son état risque de se dégrader très rapidement. Il en va de même pour des problèmes cardiaques ou des insuffisances respiratoires. Ces patients semi-urgents représentent une grande part des admissions directes effectuées dans les services de médecine et doivent être pris en charge très rapidement.

Dans un hôpital comme celui de Firminy, dans les services de médecines, près de 75% des patients admis en entrée directe relèvent de la semi-urgence et doivent être admis de préférence dans les 3 jours, au pire dans la semaine. Si on couple ce taux au faible taux d'occupation des services de médecine par des patients provenant du flux programmé (70% des services de médecine sont occupés par des patients provenant des urgences), la visibilité à une semaine est quasi nulle tant la connaissance des futurs patients intervient tardivement.

1.6 Synthèse

Avant de conclure ce chapitre, faisons un bilan des difficultés concernant la gestion du flux de patients d'un hôpital. Nous en retiendrons deux aspects :

1. La gestion des services de soins
2. La gestion des services d'urgences

Les deux sous-sections suivantes récapitulent les enjeux et buts de ces deux aspects.

1.6.1 Les difficultés de gestion des unités de soins

Les différentes unités des hôpitaux sont confrontées à des problèmes de gestion de leurs ressources tout en répondant le mieux possible à la demande. Elles doivent concilier un flux électif, lui même composé selon les cas d'un flux interne et d'un flux externe, et un flux urgent. Lors du choix de d'admission d'un patient électif, le problème se posant est de savoir si cette admission ne saturera pas le service, empêchant de ce fait la prise en charge de patients provenant des urgences dans les jours suivant. Réciproquement, lors de la prise en charge d'un patient urgent, comment être sûr de pouvoir bien admettre comme prévu le patient programmé dans 2 jours ?

En s'intéressant aux échelles de temps utilisées dans la mesure des durées moyennes de séjours (DMS), la difficulté de synchronisation entre les services de l'hôpital devient évidente. Au niveau des SU, tout va vite, et l'unité de raisonnement est l'heure. Quand l'on passe aux unités de soins, les séjours sont alors

3. Il y a également des entrées semi-urgentes en chirurgie. Une appendicectomie par exemple.

comptabilisés en jours. Mais ces problèmes d'échelle ne s'arrêtent pas là. Les services de soins de suite ou de moyens séjours raisonnent en semaine quand les maisons de repos et les maisons de retraite calculent les séjours en mois[4]. La synchronisation entre toutes ces unités de séjours de patients n'est pas chose aisée tant les modes de raisonnement sont différents.

D'autres considérations peuvent être développées comme les problèmes de planification d'actes complémentaires durant le séjour ou encore le problème d'appariement de patients dans les chambres composant le service.

Avec le temps, la connaissance et la pratique, les cadres de santé arrivent à gérer leurs services mais avec le souci permanent de n'avoir aucune garantie sur le fait que le lendemain la situation ne sera pas catastrophique.

1.6.2 L'engorgement des services d'urgences

L'engorgement est de loin le pire ennemi des services d'urgences et ce en tout point de vue. Les causes en sont multiples à commencer par l'augmentation du nombre de patients se présentant à l'accueil de ces services. La plupart des hôpitaux ont été construits il y a une trentaine d'année. A cette époque, le flux de patients de ce service n'était pas celui qu'il est aujourd'hui. Il n'est pas rare de voir des SU accueillant trois fois plus de patients que ce qui était initialement prévu. Même si l'organisation interne a évolué et progressé de manière à pouvoir absorber ce flux, les difficultés organisationnelles n'en demeurent pas moins présentes.

Chaque étape du processus de prise en charge du patient urgent est source de temps d'attente qui pourraient en général être réduits par une gestion plus fine des différentes ressources. Bien entendu, nul ne dit que cette gestion est facile à mettre en place. Les contraintes et obligations pesant sur chaque ressource font de leur gestion un problème d'optimisation à part entière. Reprenons l'exemple du scanner : les examens sont longs, de nombreuses sources différentes sont demandeuses de créneaux d'utilisation, ... De plus, la ressource étant onéreuse, la logique économique veut un taux d'utilisation élevé afin d'amortir l'investissement. Malheureusement en saturant le planning d'utilisation de la ressource, on limite la capacité à insérer des évènements imprévus.

Une fois que le patient a réussi à passer toutes les étapes de sa prise en charge au SU, lorsqu'il est décisionné, il devrait en toute logique libérer la place qu'il occupe dans le service. Pour les patients repartant à leur domicile, on ne constate en général aucun problème sur ce point. Pour ceux qui doivent être hospitalisés, on se heurte aux difficultés liées aux unités de soins que nous avons vue dans ce chapitre. De plus, en revenant sur la "vétusté" relative des hôpitaux, on constate que initialement, les capacités d'accueil étaient surdimensionnées par rapport au flux de patients de l'époque. Pendant les années 80-90 le nombre de lits d'hospitalisation a été diminué en réponse aux contraintes démographiques et économiques existantes alors. De nos jours, le nombre de séjours est en augmentation et les modifications apportées sur les hôpitaux deviennent incompatibles avec les nouvelles données médicales, économiques et démographiques. Cependant, grâce aux progrès de la médecine, les durées de séjours ont considérablement diminué permettant ainsi aux établissements de continuer à fonctionner avec leur ressources diminuées malgré l'augmentation de leur activité. Si cette augmentation devait se poursuivre, et surtout si elle devait continuer d'aller plus vite que les progrès de la médecine, peut être sera-t-il alors nécessaire d'inverser la tendance des dernières décennies et redimensionner à la hausse les services de soins.

4. La gestion de la sortie des unités de soins vers ces services de moyen et long séjours est un autre problème qui ne sera pas traité dans cette thèse.

1.6.3 Courte discussion sur les différences entre établissements privés et établissements publics

Un des objectifs de la T2A lors de son introduction était de supprimer les différences de rémunération entre les établissements privés et les établissement publics. Ne pouvant financer, par définition, une clinique privée sur le mode d'un hôpital public, la réforme a donc fait le choix d'appliquer le mode de financement des établissement privés, directement proportionnel aux actes pratiqués, aux établissements publics. Cependant selon le type d'établissement, des différences demeurent dans la gestion et l'admission du flux de patients [5].

Malgré la tentative d'harmonisation apportée par la T2A, les cliniques restent des établissements privés ayant donc une obligation de bénéfices. Des choix sont donc plus ou moins explicitement fait sur la nature des patients hospitalisés. On peut donc distinguer les patients de médecine, relevant souvent de semi-urgences, et avec des durées de séjours mal maîtrisées, des patients de chirurgie qui eux se planifient longtemps à l'avance et dont la durée d'hospitalisation est parfaitement connue. Il est plus qu'évident que ce deuxième type de patients est plus simple à gérer. Qui plus est, la chirurgie mettant en oeuvre des ressources de haute technicité, cette activité est source de revenus. Ce type d'activité est donc a priori plus lucratif. Chacun aura d'ailleurs déjà constaté que les cliniques privées traitent beaucoup de cas chirurgicaux. C'est la première différence que nous notons entre les deux types d'établissement. Pour des raisons économiques, les structures privées sont amenées à traiter un flux de patients bien établi. Les autres patients, tels ceux de médecine, sont donc souvent traités par les établissements publics faute de capacités d'accueil ailleurs.

Ce cas de figure peut cependant être nuancé. Premièrement, ce constat ne tient pas forcément dans des régions moins urbanisées. Il n'est pas rare que des petites villes aient vu les services de leur hôpital public fermer les uns après les autres et finalement ne garder que de la gériatrie ou des longs séjours. Pour ces villes, toute l'activité urgente et de courts séjours est assurée par la clinique locale. Deuxièmement, le fait que, dans les grandes métropoles, l'hôpital public est généralement représenté par un CHU. Ce type d'établissement, par définition, est dévoué à la recherche médicale et donc au traitement des pathologies moins bien maîtrisées. Les actes simples et bien maîtrisés sont alors effectués par les cliniques privées. Pour conclure sur cette différence dans les types d'activités, l'offre de soins en France est répartie entre les différents types d'établissements chacun ayant sa place, où, d'un point de vue plus économique, chacun occupe sa niche.

La deuxième et dernière différence dans la gestion du flux de patients que nous souhaitons mentionner concerne la prise en charge des urgences dans les unités de soins. Dans un hôpital privé, les patients électifs payent pour réserver leur place et être soigné. Il apparaît donc logique qu'ils soient prioritaires par rapport aux patients urgents. De plus, le remplissage de l'établissement étant pratiquement assuré à lui seul par le programmé, il est fréquent de voir les SU des établissements privés diriger leurs patients urgents en sortie vers d'autres hôpitaux. A l'inverse, les hôpitaux publics devant gérer avec l'engorgement de leur SU, doivent plus ou moins prioritiser ce flux. Remarquons au passage que dans les moeurs de la population, les gens se présentent plus fréquemment d'eux même dans un SU d'hôpital public.

Finalement, la T2A, même si elle est normalement construite pour prendre en compte les différents profils patients, tend à favoriser l'activité programmée. Les hôpitaux publics ne peuvent donc pas sacrifier leurs flux d'entrées directes au profit unique de leur SU. Même si la présence d'un forte demande des urgences cause parfois la déprogrammation de patients électifs, certaines pratiques actuelles tendent à faire revenir les patients vus aux urgences quelques jours après afin de les faire passer dans le programmé. Cela présente en plus l'avantage de libérer plus rapidement les places si précieuses des services d'urgences.

5. cette section se base sur les différences constatées lors de la visite d'un établissement privé et d'un établissement public auxquels viennent s'ajouter quelques propos tirés de différentes personnes ayant un lien avec le domaine médical.

1.7 Gestion du flux de patients et prise en charge des imprévus : Méthodologie

L'engorgement des urgences est un problème qui demeure de nos jours mal maîtrisé. Pour preuve, chacun sait que le rapport entre le temps de traitement effectif de la pathologie et le temps total passé dans le service est très petit. La majeure partie du temps passé par les patients dans un SAU est généralement due à l'attente de la suite du processus de prise en charge. Le fil directeur que nous nous fixons dans cette thèse va être de contribuer à la diminution du temps de passage des patients aux services des urgences. Autrement dit, nous cherchons à améliorer la prise en charge des patients de cette unité de l'hôpital. Le présent chapitre va exposer la démarche que nous nous proposons de mettre en place dans cet objectif.

1.7.1 Constat de départ

Le présent chapitre nous a permis de dresser un constat de départ pour notre étude. En particulier, il nous a permis, pour chaque composante de l'hôpital entrant en jeu lors du passage d'un patient, d'aborder entre autres :

1. Les différents problèmes de gestion et d'organisation auxquels doivent faire face les unités concernées.
2. Les causes de ces problèmes ; quel(s) point(s) précis de l'organisation actuelle fait(font) défaut(s).
3. Des solutions issues de règles de bonnes pratiques ou de travaux de recherche permettant de remédier aux difficultés rencontrées.

Concernant le management des SAU, dans le cadre de la réduction de leur engorgement, parmi les trois catégories de problèmes à résoudre exposées par Asplin *et al.* (2003), nous décidons de suivre les recommandations de la littérature, et notamment pour la France celles de la MeaH. Nous allons nous focaliser sur les problèmes d'admission des patients urgents dans les services d'hospitalisation. Nous ne remettons nullement en cause les améliorations pouvant être apportées par la gestion des deux autres catégories de problèmes. En effet, si l'on veut réduire le nombre de patients présents et en attente dans les SAU, il est bien évidemment nécessaire dans un premier temps de mieux réguler l'accès à ce service afin de ne garder que les patients ne pouvant être soignés par d'autres dispositifs du système de santé. La deuxième catégorie, le management du "throughput", consiste principalement en une chasse aux temps d'attente entre les différentes étapes du processus de prise en charge des patients urgents. Néanmoins, ces temps d'attentes, aussi réduit soit-il, existeront toujours car l'insertion de tâches non planifiées dans un planning ne peut se faire autrement. La dernière catégorie, le management des sorties, a un impact différent. Une fois diagnostiqué, un patient n'a en théorie plus de raison d'être présent dans le SAU. Le temps s'écoulant entre la fin du processus de prise en charge d'un patient et son départ effectif du SAU est autant de temps pendant lequel le patient occupe de la place, i.e. des ressources. Cela est nuisible pour la prise en charge des autres patients et entraîne une sorte de gaspillage de la capacité d'accueil du service.

Nous pouvons cependant distinguer les patients pouvant rentrer chez eux à la fin de leur passage au SAU de ceux devant être hospitalisés. Dans le premier cas de figure, le problème est mineur. En effet, tout au plus le patient attend soit une ambulance soit qu'un proche vienne le chercher. Dans le deuxième cas, comme nous l'avons déjà vu, le temps d'attente peut atteindre des durées intolérables et largement supérieures aux attentes pouvant survenir lors du processus de prise en charge. Il est fréquent de voir un patient attendre plus de 12 heures avant d'être admis dans un service d'hospitalisation. C'est pourquoi nous avons retenu ce dernier point comme étant susceptible d'apporter les plus grandes améliorations aux SAU. De plus, réduire ce temps d'attente avant hospitalisation est synonyme d'aider les unités

d'hospitalisation à intégrer des patients non planifiés plus facilement. La fluidification des SAU passe donc par une gestion plus précise et plus encadrée des patients dans les services de soins.

1.7.2 Influer sur le flux des urgences par le management des unités de soins

L'admission d'un patient urgent dans un service d'hospitalisation est la phase causant le plus d'attente non justifiée du patient au sein du SAU. Du point de vue des urgences, la seule action qui peut être menée pour réduire ce temps d'attente est d'anticiper dès que possible la phase de recherche d'un lit d'aval afin qu'une solution soit, si possible, trouvée quand le patient arrivera au bout du processus de prise en charge aux urgences. Cette anticipation, même si elle ne peut qu'être bénéfique, ne suffit pas. La principale cause vient des services de soins qui ne sont pas suffisamment bien organisés pour recevoir les patients urgents de manière optimale. Les responsables ne disposent généralement pas de tous les éléments nécessaires pour permettre une gestion efficace de la prise en charge des patients urgents.

Des solutions simples peuvent être apportées en se basant sur l'idée de faire de la place pour accueillir les patients urgents. A un niveau opérationnel, la MeaH (Cauterman et Engel, 2007) a pointé du doigt le manque de disponibilité des lits le matin repoussant du coup les admissions urgentes à l'après midi. La solution suggérée consiste à faire effectuer le plus de sorties possibles le matin afin de rendre les lits disponibles le plutôt possible. Sur un horizon de la semaine, la même étude a également pointé du doigt la surcharge des unités de soins en début de semaine due à un grand nombre d'admission de patients électifs. A l'opposé les fins de semaines présentent souvent des taux d'occupation plus faible. Cette surcharge est pénalisante car en saturant les services de soins, le SAU ne peut plus y placer ses patients. La recommandation de la MeaH porte sur la réalisation d'un lissage des taux d'occupation des services tout au long de la semaine.

Ces recommandations ne sont pas évidentes à mettre en place pour des hospitaliers. On pourrait en formuler beaucoup d'autres. Le domaine émergent de la télémédecine, qui permet de monitorer un patient à distance (Tian et Tianfield, 2003; Pour, 2006), pourrait être utilisé afin de décharger certains patients plutôt que prévu. Ces patients pourraient alors intégrer un processus d'hospitalisation à domicile grâce aux équipements de la télémédecine.

Quoi qu'il en soit, la principale préoccupation des cadres de santé des unités de soins est de pouvoir assurer leur planning de patients programmés avant de prendre en charge des patients urgents. Cela n'est pas sans raison. D'après Verstavel (2007) la T2A, telle qu'elle est définie, tend à favoriser les entrées directes. Ce n'est pas le seul paramètre entrant en jeu. La question qui se pose pour les cadres de santé est de savoir jusqu'à quand il est possible de recevoir des nouveaux patients en provenance des urgences avant de risquer de perturber le planning des admissions programmées. Pour améliorer la fluidité des urgences on pourrait formuler la question inverse : combien de patients programmés peuvent être planifiés avant de compromettre la prise en charge des patients urgents ? Sur un horizon très court, la construction du planning du lendemain, répondre à la question est à la portée d'une personne habituée. Sur un horizon plus lointain, quelques jours, il devient déjà compliqué de prévoir les évolutions du système. Sur un horizon d'une semaine, ce n'est plus possible. Pourtant la durée des séjours des patients dépasse fréquemment cet horizon au-delà duquel le simple être humain n'est plus capable d'anticiper les évolutions possibles d'un système aussi complexe que les systèmes hospitaliers.

Smith-Daniel *et al.* (1988) arguaient qu'il n'y a que deux manières de traiter les urgences :

1. Réserver un certains nombre de lits (ou ressources) pour les urgences
2. Prendre en compte les urgences potentielles lors de la planification des patients

Notons que dans le premier cas, les patients urgents ne sont explicitement pas pris en compte lors de la planification des patients. En fait ils sont considérés lors du dimensionnement du nombre de lits à réserver. Une telle approche suppose deux choses : (i) que le flux de patients urgent est assez stable

et proche de sa moyenne (ii) qu'en cas de sous-effectif, ou de sur-effectif, des patients urgents on sache redistribuer les ressources. Des approches semblables vont être vues dans l'état de l'art et notamment dans la gestion des ressources de diagnostic (section 2.3.1). Visiblement, en considérant la forte variabilité pouvant exister dans les demandes, cette approche semble difficilement généralisable à l'ensemble des services d'hospitalisation.

La deuxième proposition présente, quant à elle, la difficulté de réussir à évaluer le nombre de patients urgents potentiels. En allant même plus loin, il ne s'agit pas de prendre en compte uniquement les patients urgents potentiels mais également les patients électifs potentiels. Nous y reviendrons dans le prochain chapitre. Nous retiendrons cette approche dans cette thèse.

1.7.3 Objectifs de la thèse

Nous souhaitons, dans nos travaux présentés dans ce mémoire, fournir aux hospitaliers, les cadres de services, les médecins, les services d'urgences, un outil permettant d'aider à la gestion du flux de patients. Cet outil se veut être le garant d'une prise en charge efficace des patients électifs sans être au détriment de la prise en charge des patients provenant des urgences. Chaque entité de la chaîne de prise en charge du patient est considérée de manière neutre et, comme nous le verrons, est souvent ramenée à l'expression d'une capacité d'accueil. De ce fait, les choix effectués sont complètement impartiaux supprimant in fine d'éventuelles inégalités pouvant actuellement exister. Nous proposons pour chaque service des actions possibles en fonction d'une optimisation selon certains critères. Bien entendu, le personnel qualifié reste le seul décisionnaire final de par ses compétences et sa connaissance des cas particuliers pouvant survenir.

La gestion du flux de patients dans les unités de soins et plus particulièrement la gestion des lits peut être vue comme un système autonome : un système capable de s'autogérer. Ce type de système, dans sa description donnée par Kephart et Chess (2003), se base sur un ensemble de connaissance et la définition d'une politique de gestion. En suivant l'évolution de certaines caractéristiques du système ainsi géré, une analyse de l'état actuel du système est alors effectuée. Le résultat de cette analyse est alors utilisé par un module de décisions dont les sorties sont ensuite appliquées sur le système (voir figure 1.4). Un point remarquable de ce type de système est son caractère de boucle fermée. Les sorties d'une suite de décisions modifient l'état du système considéré et se retrouvent en entrée de la décision suivante. Une application d'un tel type de management appliquée au domaine de la santé se retrouve dans Carson *et al.* (1998).

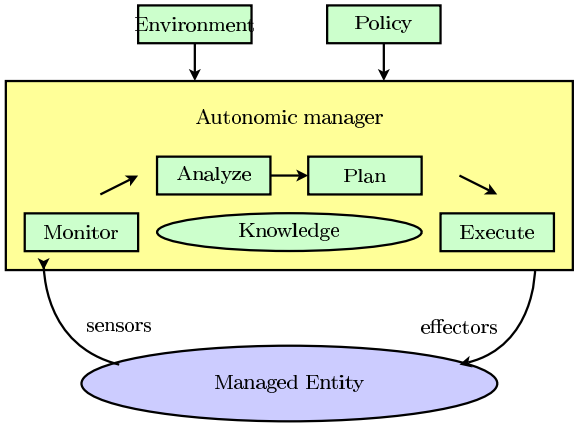

FIGURE 1.4 – Description d'un système autonome

En appliquant une telle démarche au management du flux de patients, l'étape de monitorage se concentrerait sur la surveillance de l'arrivée de nouveaux patients, le flux entrant. L'étape d'analyse, en utilisant la connaissance fournie par les données statistiques de l'hôpital, étudierait l'état d'occupation des

services de l'hôpital ainsi que les prévisions pour les périodes à venir. Ses prévisions seraient alors fournies à un module de planification qui définirait pour chaque patient son affectation. Cette planification serait faite en accord avec les préférences formulées par la direction de l'établissement. Le planning ainsi fournit deviendra alors le point d'entrée des décisions suivantes prises lors de l'arrivée de nouveaux patients. Cette modélisation est dans la réalité très proche du fonctionnement des systèmes hospitaliers.

Cette thèse va se consacrer à la définition des modules d'analyse et de planification qui pourrait intégrer un système autonome dans le cas des la gestion et de la prise en charge du flux de patients dans les unités de soins. Le caractère de boucle fermée, l'impact d'une décision à un instant donné sur la suite des décisions, doit clairement intégrer la méthode de planification. Aussi, un de nos objectifs va être de favoriser la prise de décision à chaque instant, à chaque évènement tout en gardant à l'esprit que les décisions prises auront un impact sur les décisions suivantes. Ainsi, une des clés de la réussite de la mise en place d'une telle démarche va résider dans la capacité à anticiper les possibles évolutions du système considéré et donc d'évaluer l'ensemble des conséquences que peut avoir un choix.

Plusieurs niveaux de prise de décision doivent être considérés afin de fournir un module de gestion du flux de patients. Les trois niveaux que nous proposons sont grandement inspirés de la littérature :

- Stratégique : dimensionner les services pour prendre en charge le flux de patients électifs et le flux de patients provenant des urgences.
- Tactique : planification des admissions en tenant compte des urgences et des évolutions du système
- Opérationnel : affectation des patients aux chambres, insertion des urgences

Il est important de noter que chaque niveau a besoin du précédent pour fonctionner de manière optimale. Ainsi, l'insertion des patients urgents du niveau opérationnel ne pourra se faire efficacement que si le planning des admissions a été construit de sorte à laisser libre suffisamment de lits. Pouvoir laisser libre suffisamment de lits signifie forcément que le dimensionnement du service correspond à la demande des patients qu'ils soient électifs ou urgents.

Nous ne traiterons pas du niveau stratégique dans cette thèse. Deux raisons nous poussent à cela. Tout d'abord la littérature sur le dimensionnement de service de soins est relativement conséquente et nous ne voyons pas de thématiques nouvelles qui n'auraient été déjà traitées. Deuxièmement, Les niveaux tactique et opérationnel ont un impact plus immédiat et plus significatif dans l'optique de prise en charge facilitée des patients urgents que nous souhaitons proposer. Nous allons donc nous intéresser par la suite à la gestion du patient depuis la décision de son hospitalisation jusqu'à sa sortie.

La seule ressource que nous prendrons explicitement en compte est le lit. En effet, même certains auteurs recommande de tenir compte d'un maximum de ressources lors de la planification des patients, nous pensons que la gestion de chaque ressource doit rester séparée. Pour cause, les ressources externes aux services de soins, comme nous l'avons vu dans le chapitre précédant, doivent concilier plusieurs flux de patients. Il est évident que, lors de la planification d'une admission, nous ne pouvons prendre en compte toutes ses problématiques de manière efficace. La modularité des services de soins dans leur gestion doit rester une force.

Dans la suite de ce mémoire nous allons tout d'abord présenter un état de l'art de la recherche pouvant avoir un impact sur la prise en charge des urgences. Nous nous attacherons ensuite à décrire les approches que nous proposons afin de gérer le niveau tactique et opérationnel de la gestion du flux de patients. Le chapitre 3 va considérer le problème de choix des dates d'admission des patients. Lors de la résolution de ce problème, nous allons mettre en avant un outil, une clé de répartition, que nous utilisons pour nous permettre d'anticiper sur les évolutions futures du système. La recherche de cette clé sera exposée dans le chapitre 4. Enfin le chapitre 5 présentera la prise en charge effective des patients, le niveau opérationnel de l'approche.

Chapitre 2

Management des services de soins et prise en charge des urgences

Sommaire

2.1 Management et modélisation des systèmes hospitaliers
2.2 Management du flux de patients aux urgences
2.3 Management des ressources
2.4 Management des lits des unités de soins
2.5 Conclusion

DEPUIS SES DÉBUTS la recherche opérationnelle s'est intéressée principalement à la résolution de problèmes de gestion avec application dans les domaines militaires et de l'industrie. Pourtant, on retrouve des travaux portant sur les milieux de la santé dès les années 60. Force était de constater que ces problèmes étaient d'un niveau de complexité supérieure aux problèmes industriels traités à l'époque. Avec l'évolution des techniques de résolutions de ces problèmes et le gain de performance significatif des ordinateurs, les vingt dernières années ont vu les études concernant la gestion des hôpitaux se multiplier. Cependant, les moeurs, les gouvernances, les orientations politiques qui changent continuellement la nature et les objectifs des problèmes de gestion en santé évoluent également. Un exemple parmi tant d'autre en France est la mise en place de la tarification à l'activité aux débuts des années 2000 qui a considérablement modifié les habitudes des établissements publics. La présente revue de littérature va synthétiser les travaux visant à optimiser le problème de management du flux de patients tel que nous le connaissons aujourd'hui ; tel qu'il a été présenté dans le chapitre précédent. Pour ce faire, nous aborderons le management des services d'accueil des urgences puis le management des services dont les urgences sont dépendantes. Plus précisément, nous aborderons les problématiques de gestion de ressources avant de nous intéresser en détail à une ressource particulière, la ressource "lit" des unités d'hospitalisation.

2.1 Management et modélisation des systèmes hospitaliers

Une des principales difficultés à la gestion d'un établissement hospitalier réside dans les nombreuses interconnexions entre les unités qui le composent. Chaque unité est plus ou moins dépendante des autres. La figure 2.1 extraite de la thèse de Sondes Chaabane (Chaabane, 2004) présente les différents parcours patients au sein de l'hôpital. Même si cette thèse portait sur la gestion des blocs opératoires (chemin en

gras dans le schéma), on voit ici la place centrale qu'occupent les services de courts-séjours dans le flux de patients. En particulier, il est évident à la vue de ce diagramme que ces services sont partagés entre le flux de patients provenant des urgences et le flux de patients admis en entrée directe. De même, il est aisé de se rendre compte de la difficulté que représente la gestion du flux de patients dans un hôpital. Le nombre de parcours possible est important, et ce, même sans prendre en compte les phénomènes de boucles ou de ré-entrances. La gestion des ressources, non présentée sur ce schéma, ajoute un niveau de complexité au problème du management de ce flux.

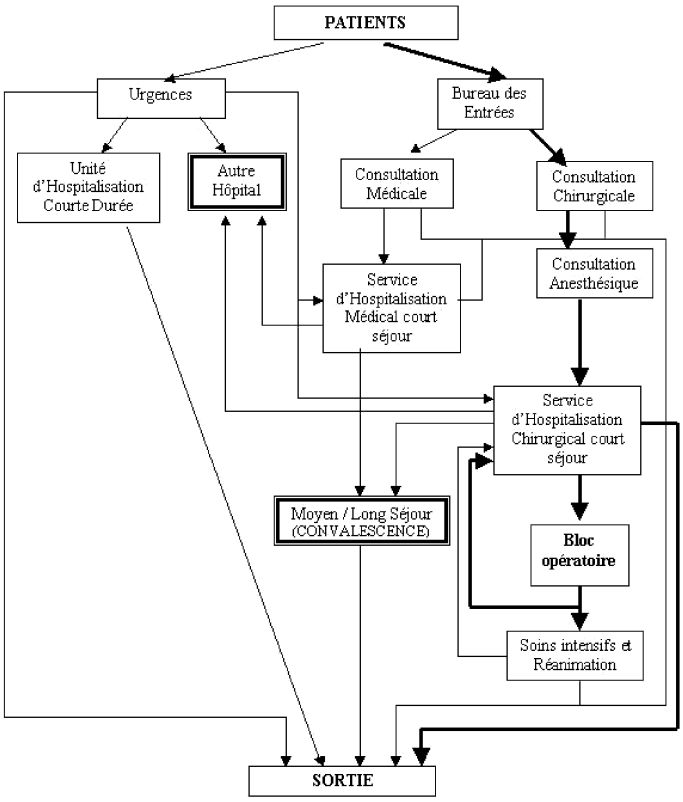

FIGURE 2.1 – Parcours patients au sein de l'hôpital

Enfin, et surtout, le principal obstacle à franchir, lors de la résolution de problèmes d'affectation de patients ou de ressources, est la prise en compte de la charge de travail nécessaire à l'absorption des flux aléatoires et perturbateurs tels les flux d'urgences. L'état de l'art de Smith-Daniel *et al.* (1988) traitaient déjà de la résolution par la recherche opérationnelle de ce problème de prise en charge lors de la planification des patients.

2.1.1 La recherche de la performance

L'amélioration des processus de soins est omniprésente dans les établissements hospitaliers. Les réformes, visant à une diminution des dépenses via une réduction des budgets ou une réorganisation des ressources, se succèdent, chacune apportant son lot de réponses aux problèmes existants mais faisant également émerger de nouvelles problématiques. Le point commun de ces réformes est une recherche d'amélioration des performances des hôpitaux. Cependant, les hôpitaux étant des systèmes complexes composés de nombreuses entités, la définition de performance va outre le simple équilibre financier ou le délai de guérison du patient. Lederman (2008), dans son mémoire, propose une définition de la performance en milieu hospitalier en identifiant quatre composantes :

- L'efficacité productive : La production est-elle optimale compte tenu des ressources allouées ?
- L'efficacité allocative : Un niveau de satisfaction plus élevé aurait-il été atteint si les ressources allouées avaient été employées différemment ?
- L'équité : L'égal d'accès aux soins est-il garanti ? Quel droit de regard pour les usagers ?
- L'efficacité médicale : Les soins dispensés donnent-ils les résultats escomptés ? Parvenir à réaliser dans la pratique une concertation plus grande des professionnels sur les cas qui nécessitent de la pluridisciplinarité.

Ces quatre composantes sont classiquement utilisées en économie de la santé. L'efficacité productive, appelée également efficience, concerne la réalisation d'un objectif à l'aide du minimum de moyens possibles. Une application concrète dans le domaine de la santé pourrait être l'amélioration de la productivité de certaines prestations entrant dans le processus de prise en charge. L'efficacité allocative représente la relation entre les coûts, la production et les besoins du client. Cela revient à optimiser l'allocation de ressources médicales en la corrélant à l'activité. L'équité vise à garantir que pour deux situations identiques, le résultat sera le même. En termes de performance hospitalière, cette garantie peut être obtenue en cherchant à réduire le temps de passage des patients dans les services et ainsi standardiser les traitements. Finalement, une analyse rapide sur l'efficacité médicale pourrait conclure que celle-ci est optimale lorsque les soins dispensés donnent les résultats escomptés. Mais alors que dire des protocoles de soins à faible taux de réussite comme les chimiothérapies ? La recherche de performance ne doit pas l'emporter sur l'éthique. L'efficacité médicale peut être améliorée en passant par des étapes de concertation entre les professionnels sur les cas posant problème.

Cette définition de la performance est très orientée vers des applications pratiques. Au niveau de la direction d'un établissement, d'autres mesures de la performance sont requises. Celles ci doivent permettre d'évaluer l'impact des politiques budgétaires, des politiques d'orientation vers des spécialités ou encore des politiques de recrutement menées dans l'établissement. La méthode DEA (Data Envelopment Analysis) présente un niveau d'abstraction suffisant pour être utilisée dans le but de mesurer de telles performances. Ozcan *et al.* (2010) le montrent dans le cas de l'évaluation des performances d'un hôpital universitaire Brésilien.

Des travaux de recherche sont menés sur les problématiques de gestion et d'amélioration des milieux hospitaliers depuis les années 70. Cependant la nature des problèmes de gestion a beaucoup évolué depuis. Les hôpitaux sont sans cesse obligés de s'adapter aux changements. La notion de processus de progrès permanent est omniprésente. Des approches très prisées dans le domaine de l'industrie comme le Lean Manufacturing ou les méthodes du Six Sigma peuvent apporter des améliorations significatives dans le cadre d'un processus d'amélioration continu. Leurs applications au domaine hospitalier permettent d'accroître l'efficacité des établissements (De Koning *et al.*, 2006; Paillassard et Castro, 2010). Cependant, la portée de cette philosophie ne peut prendre en compte tous les problèmes de gestion de l'hôpital. Certains problèmes comme par exemple les problèmes de planification doivent non seulement prendre en compte de nombreuses contraintes mais doivent aussi chercher à minimiser (ou maximiser) une fonction de coûts (ou de gains). Les méthodes issues de la recherche opérationnelle permettent de pallier à cette limite.

Les deux approches ne sont pas pour autant concurrentes. Le bloc opératoire est une ressource de l'hôpital où les deux approches peuvent conduire à des améliorations. La recherche opérationnelle a largement couvert les problématiques de planification de créneaux opératoires, de construction de planning opératoires ou encore de dimensionnement de nombre de salles d'intervention (voir section 2.3). Les approches issues des méthodes Six Sigma peuvent, elles, être utilisées, par exemple, dans le cadre d'une démarche de réduction des risques de contamination du patient par déchets infectieux tandis que le Lean peut être utilisé pour mieux gérer le processus de prise en charge du patient. D'une manière assez générale

on peut entrevoir une possible complémentarité entre ces philosophies. Par exemple, les approches de type Lean Manufacturing pourraient apporter leur lot d'améliorations d'un point de vue organisationnelle tandis que la recherche opérationnelle se focaliserait plus sur des thématiques d'aide à la décision en cherchant à proposer la meilleure solution.

2.1.2 Modélisation hospitalière

Toute analyse d'un système dans le but de l'améliorer passe par une étape de modélisation plus ou moins formelle selon la méthodologie employée. Les sciences issues du génie industriel n'échappent pas à cette règle et, qui plus est, nécessitent une description précise des processus étudiés et des ressources utilisées. Nous avons déjà vu que si nous devions résumer un système hospitalier par un adjectif, notre choix se porterait sur "complexe". De ce fait, cet adjectif convient également pour qualifier la modélisation d'un tel système. Les méthodes de modélisation existantes dans le domaine de l'ingénierie tel UML ou les réseaux de Pétri ne sont généralement pas suffisantes pour permettre de fournir un modèle complet d'un processus hospitalier si elles sont employées seules.

Différentes recherches ont été menées sur le thème de la modélisation hospitalière. Le but fondamental de ces modélisations est de réussir à représenter la dynamique des systèmes de santé. Parmi les outils du génie industriel les plus couramment utilisés, nous retrouvons la simulation. Lagergren (1998) présente l'apport de la modélisation et de la simulation dans la gestion des milieux hospitaliers. Pour un système donné, elle permettra, en effet, d'en évaluer les performances en fonction d'aléas auxquels il sera confronté. La simulation permet d'investiguer des organisations non encore testées de manière pratique et d'observer les conséquences de certaines décisions qui pourraient être prises au niveau stratégique. La théorie des réseaux de Pétri est un autre outil qui permet d'étudier la dynamique d'un système. Dans le cadre d'un contexte hospitalier, nous pouvons par exemple citer l'approche développée par Ramudhin *et al.* (2006) où encore celle de Darabi *et al.* (2009). Cependant, chacune de ces approches essaie de capturer un aspect particulier des systèmes hospitaliers. Un manque de généralisation à l'ensemble des aspects particuliers de ces systèmes dynamiques se fait ressentir. Alesky *et al.* (2008) ont proposé une approche générale visant à la modélisation et à la construction d'outils d'aide à la décision pour le milieu hospitalier. Cette approche se base sur l'utilisation de différentes techniques de modélisation en fonction du niveau de détails souhaité (de macroscopique à microscopique). L'utilisation de plusieurs méthodes générales de modélisation se retrouve dans le mémoire de thèse de Augusto (2008). Il a proposé une plateforme permettant la modélisation et la simulation des systèmes hospitaliers. Cette plateforme, med-PRO [1], associe la puissance et l'abstraction de UML pour l'aspect modélisation à une classe particulière des réseaux de Pétri nommée réseaux de Pétri de Santé pour l'aspect simulation.

Dans le cadre plus particulier de la gestion du flux de patients, il paraît pourtant simple de modéliser un processus de soins sous la forme d'un problème d'atelier où la suite des étapes du processus de soins serait assimilable à la gamme de fabrication. On pourrait alors s'inspirer de méthodes de résolution de flowshop sous incertitude existante dans la littérature. Par exemple, lors de la planification des admissions des patients, un parallèle pourrait être fait avec les travaux présentés par Luh *et al.* (1999). Dans cette étude, les arrivées des ordres de fabrication (les demandes d'admission de patients), sont probabilistes. Autrement dit, les dates d'arrivée, les durées de fabrication (durée de séjours), les dates échues (fenêtre d'admission) ou encore les priorités (urgences) ne sont connues qu'au dernier instant. L'objectif était de réduire les encours de production (les patients en attente de la suite de leur processus de soins) et la résolution proposée combine relaxation Lagrangienne et programmation dynamique.

Une autre approche basée sur des problèmes de type jobshop revient à étudier la faisabilité de l'insertion d'un patient dans un planning. A titre d'exemple citons Gröflin *et al.* (2008) qui s'intéressent

1. Medical Process-Resource-Organisation

à l'insertion de job dans un planning de multi-processor-task jobshop (chaque tâche d'un job requiert plusieurs ressources). Le lien qui peut être fait avec un patient nécessitant, un lit, une infirmière et un médecin lors de son admission est évident.

Cependant le caractère aléatoire de l'évolution d'un état de santé empêche généralement l'application, en général, de ce type de modélisation. En effet, la séquence des ressources qui vont être utilisées par le patient n'est que très rarement connue à l'avance. Il est donc nécessaire d'employer d'autres méthodes pour représenter un processus de soins. C'est ce que nous allons voir dans la suite de ce chapitre.

2.2 Management du flux de patients aux urgences

Le but recherché dans le management des urgences peut se résumer de manière quasi universelle comme la recherche d'une méthode permettant de décongestionner un service surchargé. D'un point de vue pratique, l'engorgement des services d'urgences se caractérise par le nombre de patients en état d'attente entre deux étapes de leur processus de soins. La situation d'attente peut paraître, même si moyennement désirable, normale dans une salle dite d'attente. Dans le service, cette attente pose beaucoup de problèmes car en plus de monopoliser des ressources, la présence de patients dans les couloirs des SAU va perturber les soins prodigués aux autres patients. Pines (2007) dans une lettre adressée aux éditeurs du journal *Academic Emergency Medicine* donne une définition de l'engorgement des urgences : « les urgences sont encombrées quand les ressources pour traiter la demande en soin sont inadéquates et mènent à une réduction de la qualité de soins ». L'auteur va même jusqu'à caricaturer la surcharge des services d'urgences en ironisant : « L'encombrement des urgences, c'est l'éléphant qui est dans la salle d'attente. Il est juste difficile d'évaluer son poids, à quel point il sent mauvais et quand est-ce qu'il va passer au travers du plancher. ». La gradation finale semble montrer l'impuissance des hospitaliers à réagir afin d'endiguer ce problème.

Il est pourtant crucial de parvenir à cerner, à définir, en premier les causes de l'engorgement des SAU afin de pouvoir y remédier. D'après Asplin *et al.* (2003), les causes sont à chercher tout au long du processus de soins et peuvent être regroupées en trois catégories (Fig. 2.2) :

1. les problèmes dus au flux entrant (input)
2. les problèmes rencontrés lors du traitement dans le SAU (throughput)
3. les problèmes dus au flux sortant (output)

Des méthodes permettant la mesure puis l'analyse dans le but de l'amélioration des services d'urgences ont été proposées par la suite par le même groupe d'auteur (Solberg *et al.*, 2003). En partant des résultats d'un conseil d'expert qui a proposé 113 mesures possibles, seul 38 ont été conservées dont 15 sur le flux entrant, 9 sur le flux interne et 14 sur le flux sortant des SAU. Deux critiques peuvent être formulées à l'égard de cette étude. D'une part, le nombre d'indicateurs à relever demeure important, même en se focalisant sur un flot particulier. Le personnel des services d'urgences est occupé par la gestion du flux de patients en attente de traitement qui, rappelons le, est généralement surdimensionné vis-à-vis des capacités d'accueil des SAU. Il n'est que très difficilement possible d'obtenir des observations qui ne peuvent être obtenues par le système d'information déjà présent. La deuxième critique que nous pouvons formuler porte sur l'utilisation qui sera faite des résultats. En effet, il semble peu aisé de tirer les bonnes conclusions sur autant d'indicateurs.

2.2.1 Gestion du flux de patients entrants

La première cause d'engorgement des SAU liée au flux de patients entrants est sans nul doute due à l'augmentation du nombre de patients se présentant à l'accueil de ces services. En effet, la capacité d'accueil des services d'urgences a stagné pendant que le flux entrant de patients n'a de cesse de croître.

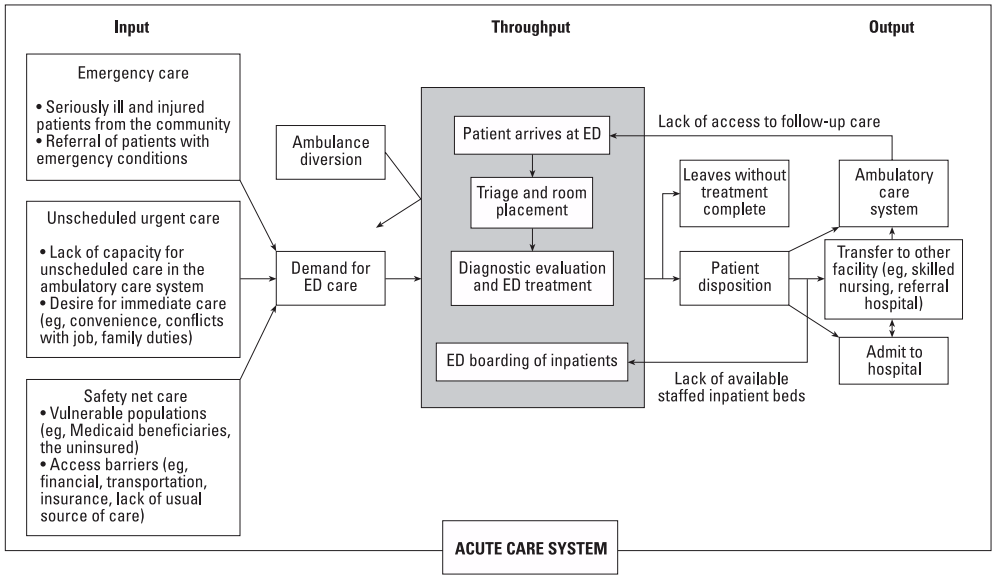

FIGURE 2.2 – Les 3 causes d'engorgement des urgences (extrait de Asplin *et al.* (2003))

L'étude de la DREES[2] de Baubeau *et al.* (2000) révèle une augmentation de +43%[3] de passages aux urgences entre 1990 et 1998. Deux raisons de cette augmentation peuvent être avancées. Tout d'abord le vieillissement de la population entraîne inévitablement une plus grande dépendance aux soins en général et, concernant les urgences, une plus grande fréquence de pathologie due à l'âge tel des douleurs abdominales ou bien encore des détresses respiratoires.

La deuxième cause de cette augmentation semble être avant tout être sociétale (Baubeau *et al.*, 2003a). Le nombre de patients se présentant d'eux-mêmes aux urgences, souvent pour des pathologies mineures, est en constante évolution. En effet, le rapport de la DREES mentionne que 15% des patients des urgences sont classés CCMU 1 (sans examen complémentaires). De plus, ces activités de bobologies sont majoritairement responsables de l'augmentation des passages aux urgences puisque le taux de patients hospitalisés après leur passage au SAU demeure stable à 20% (Baubeau *et al.*, 2003b; Carrasco, 2006).

Il semble qu'une des meilleures solutions pouvant être apportée à ce problème ne relève pas d'une approche managériale. Une revalorisation de la médecine générale accompagnée d'une sensibilisation de la population serait certainement une manière de procéder. L'inconvénient inhérent à ce type de solution réside dans les délais souvent longs avant d'obtenir des résultats. Bien entendu un premier pré-requis serait de mettre en place une telle approche.

Pourtant cette augmentation d'activité ne relevant pas de la médecine d'urgence est préjudiciable au bon fonctionnement des SAU puisque utilisant les mêmes ressources. Des solutions ont dû être mises en place pour pouvoir traiter ce flux. Afin de répondre aux besoins de permanence des soins qui ne peuvent plus être assurés par les médecins généralistes seuls, la mise en place des maisons médicales de gardes tend à être favorisée par le ministère de la santé (Grall, 2006). Malgré le bien fondé de ces établissements, leur utilisation ne semble pas à la hauteur des espérances en matière de délestage des filières d'urgences. Wang *et al.* (2009) ont proposé l'instauration d'une filière rapide dédiée à la prise en charge des patients relevant généralement de médecine générale. Cette filière se compose d'un box de consultation et d'un médecin qui est chargé de traiter les patients non-urgents. Cela revient en quelque sorte à intégrer la maison médicale de garde au sein du SAU. En s'appuyant sur un modèle de simulation, les auteurs ont montré l'intérêt

2. Direction de la Recherche, des Etudes, de l'Evaluation et des Statistiques

3. Ce problème n'est pas propre au cas français. Huang (1998) fait un constat similaire au Royaume Uni durant la fin des années 90.

que pouvait avoir ce type d'approche qui, en réservant quelques ressources pour les patients non-urgents, permet au final de laisser à disposition des vraies urgences les autres ressources du service.

La régulation des flux entrants de patients peut également se faire par l'intermédiaire des centres de régulation d'accès aux urgences (Centre 15, SAMU) ou par les professionnels de santé et notamment les pompiers. Le cas français a été étudié par Hermassi et Ladet (2008) qui, à l'aide de modélisations basées sur les réseaux de Pétri, fournissent une base d'étude des performances de l'ensemble de l'amont des urgences. Une thèse de doctorat tirant profit de cet outil est à suivre. Le pilotage des ambulances se retrouvent dans d'autres travaux. Par exemple, Stevenson *et al.* (2001) proposent un modèle de simulation permettant de déterminer l'hôpital de destination des ambulances en fonction de certains paramètres et notamment en fonction de la sévérité des blessures du patient urgent. Su et Shih (2003) ont, quant à eux, proposé un modèle d'évaluation du processus d'interventions urgentes par des ambulances afin d'éventuellement améliorer la prise en charge des patients.

Dans la littérature traitant de la gestion du flux de patients entrants dans les services d'urgences, on peut noter la présence de deux problèmes typiques du continent Nord Américain : la gestion de crise et la réduction des "diversion hours". La problématique de pilotage des réponses du système de santé en situation de crise n'est à proprement parler spécifique à cette zone géographique, mais c'est une thématique très présente dans les moeurs locaux de vouloir se savoir paré à toute éventualité. A titre d'exemple, nous citerons Kelen *et al.* (2006) qui, en classifiant les patients présents dans un établissement hospitalier, définit des listes de patients à faire sortir en priorité de l'hôpital afin de libérer des places pour l'afflux massif de patients aux urgences.

La seconde problématique, la réduction des "diversion hours" est un concept qui n'existe pas, ou très peu, en Europe. D'ailleurs, n'ayant pas trouvé de traduction adéquat pour désigner ce que sont les "diversion hours", nous avons choisi de conserver la forme anglo-saxonne. Dans ces pays, le pilotage des flux de patients urgents que l'on peut avoir en France par le centre 15 n'existe pas. Les patients souffrants sont systématiquement emmenés par ambulance au service d'urgences le plus proche. L'afflux de patients non-urgents dans les SAU est beaucoup plus important que celui que nous connaissons. Il arrive fréquemment que l'afflux de patients dans un SAU soit tel qu'il conduit à la paralysie totale du service. Dans ce cas là, l'hôpital avertit les autorités sanitaires locales qu'il ne peut plus recevoir de nouveaux patients. Les ambulances se présentant à cet hôpital sont donc déroutées vers un autre établissement. Cette période durant laquelle le service est fermé à toute nouvelles admissions est appelée "diversion hours". En France, un équivalent serait le plan blanc (un plan de gestion de crise) qui conduit également à dérouter les ambulances vers d'autres hôpitaux. La principale différence entre ces deux situations est que le plan blanc est très rarement déclenché alors que les "diversion hours" sont monnaies courantes dans les villes nord-américaines.

L'utilisation des "diversion hours" pose de nombreux problèmes pour la sécurité des patients. En premier lieu, le temps de transport vers un centre de traitement des urgences s'en trouve augmenté. Cependant, le passage en "diversion hours" n'améliore pas le problème d'engorgement des urgences. Bien souvent, le problème est seulement déplacé vers un autre hôpital qui le propage ensuite aux autres établissements de la région.

La littérature sur ce domaine est assez conséquente et est principalement à mettre à l'actif des hospitaliers. Nous retiendrons les travaux de Vilke *et al.* (2004) et de Asamoah *et al.* (2008). La première étude a pour but de mettre en évidence le phénomène oscillatoire et la dépendance existante entre les établissements quand l'un passe en statut de diversion. La démonstration est faite en prenant deux hôpitaux A et B. A se voit interdire de passer en diversion. L'impact de cette politique a eu pour conséquence de réduire le nombre de fois où B a dû passer en statut de diversion. Le deuxième impact est que, l'hôpital B passant moins en diversion, la décharge de son flux de patients sur A a moins eu lieu. Comme preuve, les auteurs ont mesuré la performance de A (nombre de patients traités), qui est restée sensiblement la

même. Il s'avère donc nécessaire de vouloir limiter l'utilisation des diversion hours.

C'est ce qu'ont cherché à faire Asamoah *et al.* (2008) qui se proposent de fixer un quota de temps maximum de diversion hours pour chaque hôpital. Ce temps doit être le strict minimum pour permettre à l'hôpital de réorganiser ses ressources. Le but recherché ici est de réduire les diversion hours afin de diminuer la probabilité que deux hôpitaux soit en même temps en statut de diversion. Les auteurs montrent que si tous les hôpitaux coopèrent, un temps maximum de 1h de diversion hours toutes les 8h présente un intérêt dans le traitement des urgences.

Pour conclure sur ce problème des diversion hours, les solutions proposées ne sont que palliatives à un problème qui trouve son origine dans l'organisation du système de santé. Cependant, comme dans le cas Français, une réforme du système de santé n'est pas quelque chose de simple à mettre en place.

2.2.2 Gestion du processus de soins

Les patients se présentant aux urgences doivent être traités au sein du service. Un des premiers pré-requis pour permettre un écoulement régulier du flux de patients urgents sans risquer les situations d'engorgement est d'avoir un service bien dimensionné par rapport au flux de patients reçus. Le dimensionnement d'une unité de soins est parmi les problèmes les plus traités dans la littérature du management hospitalier. Nous en verrons quelques exemples dans la section 2.4.1 dédiée au management des unités de soins. Cette problématique, appliquée aux services d'urgences, peut être traitée en s'inspirant des études qui seront exposées pour les unités de soins. Cependant, nous pouvons mentionner les travaux de Wiinamaki et Dronzek (2003) et de Miller *et al.* (2004) qui traitent spécifiquement du problème de dimensionnement d'un SAU. Dans la première étude, les auteurs partent d'une évaluation de l'existant et proposent un nombre de lits à ajouter au service dans le cadre d'une extension de celui-ci. Un modèle de simulation est utilisé à cette fin. Miller *et al.* (2004) sont plus généralistes puisqu'ils considèrent le dimensionnement des ressources matérielles, qu'il s'agisse de lits, de chambre ou d'appareillage, en tenant compte de la performance du processus de prise en charge.

Comme pour les systèmes hospitaliers en général, l'analyse de la prise en charge des patients aux urgences passe tout d'abord par une étape de modélisation. Chauvet et Gourgand (2008) ont proposé un état de l'art de ces méthodes et se sont particulièrement focalisés sur la collecte de données nécessaires lors de la définition d'un modèle. Ainsi, cette étude originale fait le tour des méthodes employées permettant l'acquisition de données (extraction de base de données, questionnaire, ...).

La définition de la performance du processus de soins n'est pas aisée à formuler. Li et Howard (2010) proposent, en se basant sur un processus de soins similaire à celui que nous avons exposé dans le chapitre précédent en figure 1.1, une méthode de mesure de la performance de ces systèmes en considérant quatre classes de patients (trauma, critique, aigu et express). En considérant un ensemble de paramètres pour chacune de ces classes de patients urgents, les auteurs donnent une formule analytique de l'évaluation de la performance d'un système de traitement des urgences. Cependant, ces mesures ne prennent pas en compte les phénomènes de boucles qui existent lors de l'étape de diagnostic. Pour ces auteurs, la notion de performance est liée au temps moyen de passage d'un patient au sein du SAU. Baesler *et al.* (2003), eux, considèrent que la performance d'un service d'urgence est liée au nombre de patients pouvant être traités avant de compromettre le traitement des autres patients. Par leur étude, il est alors possible de chiffrer la capacité maximum d'accueil de patients urgents dans le SAU.

Pour finir sur les méthodes de management du processus de soins aux urgences, notons que des travaux ont également été menés afin de gérer en pratique le flux de patients. Par exemple, Diefenbach et Kozan (2009) ont proposé une modélisation du processus de traitement des patients par un problème de type flowshop hybride, permettant ainsi dès l'arrivée du patient d'ordonnancer son passage dans les différentes ressources du service. La limite de cette approche semble pourtant assez rapidement atteinte puisque le nombre de patients urgents, leur parcours, la gamme, n'est pas définie dès leur arrivée. Ce genre d'approche

fonctionne uniquement sous une hypothèse forte qui est que pour une majeure partie des patients urgents, leur parcours de soins peut être établi dès leur arrivée. Dans le milieu médical, et notamment dans la médecine d'urgence, les décisions sont prises en fonction des examens médicaux réalisés. Cela rend presque impossible l'application d'un modèle classique d'ordonnancement d'atelier. La suite de décisions prises et donc le parcours patient au sein des urgences peut être vue statistiquement comme aléatoire et difficilement prévisible. Pour tenter de faire face à ces problèmes certains auteurs ont proposé de manager le processus de soins des services d'urgences par des approches multi-agents (Laskowski *et al.*, 2009; Daknou *et al.*, 2010). Chauvet *et al.* (2006) se sont intéressés à la gestion particulière des services d'urgences pédiatriques en proposant la mise en place d'un modèle de connaissance puis d'un modèle d'action et de résultats.

Il demeure cependant que la majorité du temps perdu lors du processus de soin est dû au problème de disponibilité et de coordination des ressources externes au service (Asplin *et al.*, 2003; Lederman, 2008). Le management des ressources de diagnostic dépasse le cadre des services d'urgences puisqu'elles sont partagées par l'ensemble de l'hôpital. La section 2.3.1 dressera un panorama des méthodes de gestion de ces ressources.

Sans entrer dans les détails du fonctionnement d'une ressource ou d'une autre, des améliorations peuvent être apportées. Lederman (2008) propose une méthode générale visant à mettre en place un « manager par les processus » qui en analysant les processus présents dans les coeurs de métiers entrant en jeu dans l'hôpital va émettre des propositions d'amélioration. Cela peut se voir dans la lignée des travaux de Samaha *et al.* (2003) qui, en utilisant un modèle de simulation, ont montré que les problèmes venaient des processus plutôt que des ressources. L'illustration de la tâche incombant au « manager par les processus » proposée par Lederman s'intéresse au temps de passage dans le SAU du centre hospitalier de la côte Basque. L'étude du flux de patients dans ce service a montré que l'utilisation d'actes de biologie et l'appel au service des brancardiers étaient les principaux facteurs d'allongement du temps de passage des patients. L'analyse a abouti à une réorganisation des missions des brancardiers qui doivent être d'une part plus disponibles pour le SAU et d'autre part seront chargés des acheminements entre le plateau de biologie et le SAU. Concernant les quatres composantes de la performance que l'auteur a définies les résultats sont les suivants :

- l'efficacité médicale est renforcée par le recentrage du SAU sur son coeur de métier (affranchissement des tâches de brancardage et d'acheminement des prélèvements biologiques), et la prise en charge plus rapide de certaines pathologies par les spécialistes d'« organes »
- l'efficacité productive et l'équité sont améliorées par la réduction du temps de passage des patients au SAU
- l'efficacité allocative est optimisée par la mise en cohérence de la ressource médicale avec l'activité, en générant au passage une économie annuelle de l'ordre d'un équivalent temps plein d'infirmier.

Nous pouvons tirer un enseignement important de cette étude. L'optimisation d'un service donné ne passe pas uniquement par une réorganisation de ce service mais éventuellement par la redéfinition des rôles des services annexes.

2.2.3 Gestion du flux sortant

Les patients arrivant en fin du processus de soins aux urgences se subdivisent en deux classes : ceux qui peuvent rentrer à leur domicile et ceux qui doivent être hospitalisés. Les patients de la première catégorie ne posent a priori pas de difficultés particulières. Ce n'est pas le cas de la deuxième catégorie de patients qui représente de loin la première source d'embolisation des urgences Moskop *et al.* (2009a). Divers facteurs peuvent être avancés comme étant la cause de l'incapacité récurrente des SAU à transférer

des patients vers les services de soins. Ainsi, Verstavel (2007) et Moskop *et al.* (2009b) indiquent que les patients électifs sont plus rentables que les patients urgents du fait, respectivement, de la T2A en France et du système de couverture sociale aux USA. Cette affirmation est contredite par Henneman *et al.* (2009) qui annoncent qu'entre 2003 et 2005, les patients provenant des urgences ont généré plus de bénéfices que les patients électifs en Grande Bretagne.

Verstavel (2007) constatent également que certains services jouent inégalement le jeu des urgences en ne favorisant pas, voir en ne prenant pas en compte, l'écoulement du flux de patients depuis ce département. Moskop *et al.* (2009b) généralisent ce constat en observant que les responsables des services annexes ne placent pas une haute priorité à l'optimisation du flux aux urgences. En fait, ils sont déjà bien occupés à faire tourner leur propre service avec leurs propres contraintes. On voit ici se dessiner un besoin d'aide à la gestion du flux de patients qui doit être transversal aux différentes unités constituant l'hôpital.

L'idée de mettre en place une centrale de management des lits présente dans ce sens des avantages évidents. Le premier est sans doute la mise en place d'un interlocuteur unique qui évite ainsi les appels à répétition depuis les urgences vers tous les autres services de l'hôpital. Le deuxième intérêt est bien sûr d'avoir une entité qui à chaque instant connaît les besoins et les disponibilités de chaque unité de soins. Cette centralisation de l'information permet de gérer l'occupation des lits en ayant une vision globale de l'hôpital. Un retour d'expérience de la mise en place d'une cellule de management des lits dans un hôpital de l'Oklahoma de taille moyenne et couvrant un territoire rural (hôpital similaire au CH de Firminy) nous est proposé par Hemphill et Nole (2005). Le mode de fonctionnement de cette cellule est assez simple puisqu'il consiste à affecter des codes de couleur (vert,orange,rouge) aux unités de l'hôpital afin de savoir dans quel état elles se trouvent. Cela permet de pouvoir repérer aisément où sont les problèmes et quelles sont les alternatives possibles. Les auteurs indiquent que suite à la mise en oeuvre de cette cellule les temps d'attente injustifiés dans leur SAU a été réduit de 63% et le nombre de refus d'accès au soin pour cause d'hôpital saturé (diversion hours) a été réduit de 54%.

D'autres études montrent qu'une gestion complète et centralisée du flux de patient peut être bénéfique pour l'écoulement des patients en attente aux urgences. Howell *et al.* (2008) prônent une gestion proactive des lits en passant par la mise en place d'un "directeur des soins". En plus de la gestion des affectations de patients, ce directeur se voit également octroyer la possibilité d'ouvrir de nouveaux lits dans les services afin de prévenir la congestion des urgences. Durant la période de tests de cette méthode, les auteurs ont pu constater une augmentation du nombre de patients traités aux urgences, une réduction du temps de passage pour les patients suivis d'hospitalisation et aucun changement notable pour les passages avec retour à domicile. A noter que comme conséquence de la fluidification des passages aux urgences, le nombre de diversion hours a également été réduit par cette approche. Les auteurs ont ensuite étendu leur approche pour le cas particulier des services de réanimation/soins intensifs avec des résultats similaires pour les patients les plus gravement atteints et pour qui la rapidité de la réponse médicale est très importante (Howell *et al.*, 2010).

L'intérêt de la mise en place d'un responsable des lits a également été abordé en France. Griveau (2007) définit une fiche métier d'un « cadre de l'unité de coordination des hospitalisations » et dresse un panorama des impacts positifs que cela peut avoir sur la prise en charge des urgences. D'après l'auteur, ce cadre de santé qui a pour rôle de coordonner les flux d'admission de patients de l'hôpital par une gestion centralisée tranche avec l'archaïsme actuel très cloisonné des hôpitaux. Le mémoire de Verstavel (2007) traite également de cette problématique. En analysant en détail le flux de patients aux urgences du centre hospitalier de Bretagne Sud, il dresse une liste des conséquences que peut avoir un engorgement de ce service. En particulier Verstavel souligne l'impact négatif sur la prise en charge globale des patients et l'altération des conditions de travail du personnel. Comme solution, l'auteur propose une liste d'indicateurs destinés à être utilisés par une commission de régulation des flux. Bien entendu, certains hôpitaux

ont déjà mis en place ce système de gestion avant que la littérature en fasse les louanges. Néanmoins, cette pratique est loin d'être généralisée à l'ensemble des établissements de soins.

Toutes les études que nous venons de citer s'accordent sur le fait qu'elles concernent la gestion des unités de soins pour permettre une prise en charge facilitée des urgences. Le lien de causalité directe entre services d'hospitalisation et performance des urgences avait auparavant été étudié et mis en avant dans les travaux de Bagust *et al.* (1999). En utilisant la simulation et en faisant varier onze paramètres dans des intervalles de valeurs fréquemment rencontrées, l'étude montre qu'à partir d'un taux de remplissage des lits de 85% la prise en charge des urgences devient de mauvaise qualité et qu'à partir de 90%, les situations de manques de lits deviennent fréquentes.

En guise de court résumé sur la gestion des urgences, au vu de la littérature sur le sujet, nous pouvons conclure que l'optimisation de ces services dépend principalement de la gestion des services de soins. Nous pouvons relever dans cette optique les travaux de Huang (1995, 1998) qui traitent du dimensionnement des unités de soins en tenant compte des délais d'attente aux urgences.

2.3 Management des ressources

Nous venons de le voir, les SAU sont largement dépendants des autres unités de l'hôpital. Nous nous proposons dans cette section de voir quelles sont les contraintes pesant sur ces services et quels sont les travaux de recherche qui ont été menés pour y répondre. Nous allons plus particulièrement nous intéresser aux ressources dites de diagnostics tel l'IRM ou le scanner mais aussi à une ressource sur laquelle l'activité urgente à un fort impact : le bloc opératoire. Pour être complet sur le management des ressources hospitalières, nous terminerons sur quelques mots concernant les activités de logistique hospitalière et les activités de gestion des ressources humaines.

2.3.1 Management des ressources de diagnostic

Comme nous l'avons déjà mentionné dans le précédent chapitre introduisant les systèmes hospitaliers, les ressources permettant le diagnostic de l'état clinique d'un patient sont confrontées à trois flux de patients :

- le flux de patients externes
- le flux de patients internes
- le flux de patients urgents

Chacun de ces trois flux se présente avec des caractéristiques différentes. Pour le flux de patients externes, il convient de trouver un rendez-vous le plus tôt possible afin de faire passer l'examen médical. Le patient se présentant à l'hôpital uniquement pour cet examen, il est important de respecter l'engagement pris. Pour les patients internes, l'horaire de passage peut être modulé en partie si besoin est. Cependant, cette modification doit demeurer mineure au risque d'allonger le séjour du patient dans l'établissement. Enfin pour les patients urgents, le problème est simple. Il faut les faire passer le plus rapidement possible. Une méthode basique d'affectation consisterait à affecter d'abord les patients urgents puis les moins urgents. Cependant, dans la pratique, les patients les moins prioritaires sont ceux planifiés le plus longtemps à l'avance et les plus urgents sont connus au dernier moment.

Une idée consiste à réserver un certain nombre de créneaux horaires pour faire passer les patients urgents. Cette idée se retrouve dans les travaux de Patrick et Puterman (2007) dans un cadre général et dans Geng *et al.* (2009) dans le cadre d'un service d'IRM[4]. Les questions se posant alors sont : que

4. Imagerie à Résonance Magnétique

faire si le nombre de créneaux réservés n'est pas suffisant ? que faire si l'ensemble des créneaux alloués ne sont pas utilisés ? Pour Patrick et Puterman si le nombre de créneaux est insuffisant alors le service fera des heures supplémentaires tandis que pour Geng *et al.*, les patients intégreront alors le processus de réservation standard même s'il est plus long. Les ressources de diagnostic étant généralement coûteuses, on ne peut se permettre de laisser vacant un créneau réservé non utilisé. Patrick et Puterman utilisent un ensemble de patients dont le rendez-vous est planifié ultérieurement, et qui se sont déclarés disponibles pour être appelés n'importe quand. Dans le cas des IRM, Geng (2010) y a consacré la majeure partie de sa thèse.

L'étude de la planification des rendez-vous dans les services d'IRM avait déjà été effectuée un peu plus tôt par Green *et al.* (2006). Les auteurs ont utilisé un modèle de Markov où chaque état i est caractérisé par le nombre de patients non programmés arrivant après le i^{eme} créneau de rendez-vous. L'aide à la prise de rendez-vous est un problème que l'on retrouve également dans l'étude de Sickinger et Kolisch (2009) qui considèrent la planification d'un service d'imagerie possédant deux scanners. Ces travaux vont donc un peu plus loin que ceux précédemment cités qui ne considéraient implicitement qu'une ressource agrégée.

Les patients externes, internes et urgents constituent trois classes de base des patients. On peut souvent, dans la pratique, identifier des sous-classes. Les services gérant des ressources de diagnostic allouent en fait des créneaux à chacune des classes. Patrick *et al.* (2008) présentent une approche permettant de planifier les patients en fonction de leur classe d'appartenance et des créneaux qui y sont alloués. L'activité dans le domaine de la santé étant souvant fluctuante, et en tout état de cause rarement constante, Vermeulen *et al.* (2009) ont proposé une méthode flexible d'allocation des créneaux à chaque classe. Le nombre de créneaux alloués va s'adapter en fonction de la demande actuelle et de ses prévisions d'évolution.

L'allocation dynamique est également au coeur des travaux de Day *et al.* (2010) qui considèrent par un modèle de décision markovien l'allocation des créneaux d'une unité de diagnostic cardio-vasculaires. D'autres études traitant de domaines particuliers de la médecine existent. Nous citerons Santibáñez *et al.* (2009) qui considère un hôpital de jour en cancérologie. L'objectif que se sont fixés les auteurs nous permet de conclure sur cette sous-section en rappelant que la politique de planification des rendez-vous doit permettre de réduire le temps d'attente des patients tout en maximisant l'utilisation des ressources.

2.3.2 Ordonnancement du bloc opératoire

Le bloc opératoire constitue une ressource dont la maximisation de son utilisation est essentielle. Cette ressource, située au coeur de l'activité des hôpitaux, de par la haute technicité du matériel et des équipes soignantes intervenant en son sein en font une des ressources les plus coûteuses. La littérature se rapportant à l'optimisation du fonctionnement de cette ressource est riche et abondante comme le montre le très bon état de l'art réalisé par Cardoen *et al.* (2010). Cette sous-section ne va qu'illustrer les problématiques de gestion de cette ressource.

Plusieurs sous-problèmes relatifs à l'optimisation du bloc opératoire peuvent être traités comme la création et l'ordonnancement du planning opératoire ou le choix des équipes d'intervention. Le problème de construction du planning du bloc opératoire peut se décomposer en trois niveaux (Beliën, 2006) :

1. Diviser les temps de salle opératoire entre chirurgien ou groupe de chirurgien
2. Construire le planning cyclique (master surgery schedule) affectation de salle et de temps de salle aux chirurgiens ou groupe de chirurgiens
3. Choix des dates d'interventions et création du planning des salles opératoires

La construction du planning du bloc opératoire, l'affectation de temps de salle aux chirurgiens, a été traitée, entre autres, dans Beliën et Demeulemeester (2007) et Beliën et Demeulemeester (2008). Dans

le premier cas, la construction du planning a été effectuée en tenant compte du taux d'occupation des lits en sortie du bloc opératoire afin de ne pas emboliser les services des soins. Le deuxième article est également intéressant puisqu'il prend en compte la planification du personnel infirmier en même temps que l'allocation de temps opératoire aux chirurgiens.

Le planning des chirurgiens défini, l'étape suivante dans la gestion du bloc opératoire va concerner la construction du planning opératoire : le choix des dates d'intervention. Nous pouvons citer les travaux de Carter et Ketabi (2009) qui cherchent à lisser la charge de travail du bloc opératoire sur une semaine. Nous en arrivons à l'ordonnancement des salles opératoires. En d'autres termes, cette phase va affecter les ressources nécessaires aux interventions prévues. Pour ce faire, Liu *et al.* (2010) ont converti le problème en un problème de partitionnement d'ensemble qu'ils couplent à de la programmation dynamique afin d'agréger l'ensemble des sous-états ainsi défini. Cardoen (2009) met l'accent sur les différents objectifs, parfois contradictoires, qui peuvent être avancés par les acteurs du bloc opératoire lors de la définition de ce que doit être un ordonnancement optimal. Finalement, six objectifs sont retenus parmi lesquels la minimisation des pics d'utilisation des lits de réveil ou bien la minimisation des heures supplémentaires dans les salles de réveil.

Finalement, nous pouvons mentionner quelques études qui dans la construction du planning des salles opératoires vont prendre en compte les arrivées aléatoires des patients urgents. Une manière de souligner l'impact que peuvent avoir ces arrivées sur le planning opératoire est exposé dans Tancrez *et al.* (2008). Parmi la littérature sur l'ordonnancement sous incertitude du bloc opératoire, nous retenons Denton *et al.* (2007), Pham et Klinkert (2008) qui se ramènent à un cas particulier du problème de jobshop nommé multi-mode blocking job shop (MMBJS) et Lamiri *et al.* (2008) qui utilisent la méthode optimisation de Monte Carlo.

La recherche en France sur l'optimisation des blocs opératoires n'est pas en reste. Sous l'impulsion des projets HRP et HRP2 financé par le cluster GOSPI[5] de la région Rhône Alpes, plusieurs thèses ont été réalisées. Kharraja (2003) s'est intéressé à la construction du planning opératoire. Chaabane (2004) a étudié l'ensemble des problématiques du bloc opératoire depuis la définition des plages d'ouverture jusqu'à l'ordonnancement des salles d'opération en passant par la programmation des interventions. Jebali (2004) élargit le périmètre du problème en l'étendant aux services de chirurgie et en définissant quatre niveaux décisionnels : la planification des admissions, la construction du programme opératoire sur l'horizon de T jours, la construction du programme opératoire journalier et le pilotage temps-réel du bloc opératoire. Le dernier de ces quatre niveaux, le pilotage temps-réel du bloc opératoire, va principalement consister à insérer les urgences dans le programme opératoire. La thèse de Lamiri (2007) se focalise sur ce point. En élargissant à la proche francophonie, nous pouvons ajouter la thèse de Fei (2006) ou les travaux de Hanset *et al.* (2010) qui vont déboucher sur une thèse de doctorat également.

2.3.3 Ressources humaines et logistique hospitalière

Afin d'être complet dans cette section consacrée à la gestion des ressources, nous allons terminer par un rapide aperçu des ressources que sont les ressources humaines et les activités de logistique hospitalière. Bien que ces ressources n'entrent pas directement en considération lors de la gestion du flux de patients, elles n'en demeurent pas moins importantes afin de garantir le bon fonctionnement de l'ensemble de l'hôpital. Le management du personnel englobe bien sûr l'ensemble des corps de métier présents à l'hôpital. Certaines activités, par exemple les équipes administratives, ne nécessitent pas de gestion particulière. Elles respectent les conditions de travail standard.

La planification des ressources humaines se réduit très souvent dans le domaine hospitalier à la conception du planning de roulement infirmier. La conception de ce planning est un domaine qui a suscité

5. Gestion et Organisation des Systèmes de Production.

énormément de travaux de recherche. Etant donné un nombre de vacation à assurer et un ensemble de contraintes relevant du droit du travail, le problème consiste à trouver un roulement permettant à chaque infirmière d'effectuer au bout du compte le même ensemble de vacations que les autres. Pour un exemple concret d'étude sur le domaine nous pouvons nous référer à Aickelin et Dowsland (2004) qui développent une approche par algorithme génétique pour résoudre le problème. Pour plus de références et un survol beaucoup plus complet, nous nous en remettons aux états de l'art de Cheang *et al.* (2003) et Burke *et al.* (2004). En restant dans le management des équipes d'infirmiers, d'autres problèmes plus opérationnels, ou à l'opposé stratégiques, peuvent être traités. Par exemple, lors de l'admission d'un patient dans un service, d'hospitalisation ou aux urgences, les infirmières doivent procéder à l'installation du patient dans un lit et éventuellement la pose de quelques dispositifs médicaux comme une perfusion ou un monitoring cardiaque. Ces tâches étant demandeuses en temps, une gestion de l'affectation du personnel infirmier à l'admission des patients est nécessaire et est traitée par Sundaramoorthi *et al.* (2010). L'aspect stratégique se retrouve dans les travaux de Mincsovics et Dellaert (2010) qui proposent de traiter le problème de gestion du budget lié aux équipes d'infirmiers en détaillant différentes catégories que sont les infirmiers permanents, les intérimaires et les heures supplémentaires.

La gestion des plannings des médecins est également un problème qui se pose. Nous pouvons citer les travaux de Pecora *et al.* (2008) qui, en cherchant à tenir compte des préférences des médecins, traitent du planning de présence des médecins urgentistes en salle de soins. Cette étude peut être mise en relation avec celle de Tan *et al.* (2002) qui eux traitaient de l'impact de l'augmentation du nombre de médecins présents en salle d'urgences. D'une manière plus générale sur la gestion des besoins de personnel en salle d'urgences, nous pouvons nous référer à Centeno *et al.* (2003) pour un aspect prévisionnel et à Glaa *et al.* (2006) pour la gestion du planning infirmier dédié à ce service. Une autre étude intéressante à mentionner ici est celle de Beliën et Demeulemeester (2006). Les auteurs traitent de la conception du planning des médecins stagiaires ; les externes ou internes en France. La contrainte est de respecter les objectifs de formation des stagiaires. Aussi, la construction de leur planning d'affectation doit prendre en compte l'ensemble des activités auxquelles ils doivent assister pendant leur stage.

Pour conclure sur la gestion du personnel, citons des travaux qui considèrent la gestion des ressources humaines en santé d'une manière générale et non pas focalisée sur telle ou telle profession. Brauner *et al.* (2003) présentent une application originale des méthodes de réécriture inspirées de la théorie des langages dans la planification du personnel médical. Enfin, les travaux de thèse de Trilling (2006) dévoué à la gestion des ressources humaines dans le cadre des plateaux médico-techniques tel le bloc opératoire commence par une étude de dimensionnement et de définition des vacations nécessaires pour faire fonctionner le service.

Intéressons nous à présent aux activités dites de logistique hospitalière. Nous pouvons considérer quelques grandes familles de ces problèmes comme la stérilisation des ustensiles médicaux, l'approvisionnement et la gestion des stocks de médicaments, les activités d'hôtellerie et le transport des patients. La littérature sur ces problématiques est une fois encore relativement volumineuse. Nous retiendrons deux thèses de doctorat : celles de Tlahig (2009) et de Kergosien (2010).

Tlahig considère la recherche d'une meilleure configuration des activités logistiques. En particulier le choix se pose entre internaliser, externaliser ou mutualiser certaines activités. Cela revient à choisir entre faire soit même ? faire faire ? ou faire ensemble ? Une étude de cas permettant de support à cette thèse est proposée et considère un service de stérilisation.

Kergosien traite des problèmes de gestion et de planification des flux logistiques, produits ou patients dans les hôpitaux. En particulier, il considère les cas des hôpitaux mono et multisites. En effet, dans les algorithmes proposés, la distinction est faite entre les flux intersites et les flux internes à un site particulier d'un complexe hospitalier.

Finalement, nous citerons deux études liées au projet NHE [6]. Chabrol *et al.* (2006) et Féniès et Rodier (2006) ont proposé des applications, parfois issues de problèmes généraux de logistiques, pour la gestion de la chaîne logistique hospitalière d'un établissement en phase de conception.

2.4 Management des lits des unités de soins

Dans le domaine hospitalier, les méthodes de gestion des lits sont nombreuses et variées. Ces méthodes mises en place généralement de manière empirique répondent aux problèmes de gestions rencontrés dans chaque établissement. Selon que l'on considère un grand hôpital de type CHU, une clinique privée de chirurgie ou un petit hôpital de province, il paraît logique de penser que les problématiques rencontrées ne sont pas les mêmes. Gossart *et al.* (2010) ont comparé les difficultés rencontrées dans seize hôpitaux, français et belges. L'étude a considéré huit hôpitaux de grande taille et huit hôpitaux de moyenne et petite taille. Il ressort de cette analyse que, malgré ce que chacun peut penser de prime abord, les problématiques rencontrées sont les mêmes quelle que soit la taille de l'hôpital. Ce résultat intéressant mériterait d'être exploité plus en détail et en premier lieu d'étendre l'échantillon de référence afin d'obtenir une statistique fiable.

Le management des unités de soins ne semble donc pas dépendre de la taille de l'établissement considéré. En 1995, Van Dierdonck et Roth avait proposé une approche en trois niveaux, stratégique, tactique et opérationnelle, permettant de regrouper les problèmes de gestion en fonction de leur but. Cette classification a été reprise par de nombreux auteurs dont Vissers *et al.* (2007). Le tableau 2.1 présente comment nous pouvons définir chacun de ces trois niveaux.

TABLE 2.1 – Les 3 niveaux de gestion des lits

Niveau	Horizon	Concept	Contenu
Stratégique	long terme (année)	définition d'orientations politiques	Fixer les objectifs de taux de service par groupe de patients et les objectifs d'utilisation des ressources
Tactique	moyen terme (mois/semaine)	définition des politiques d'admission	Quantité de ressources requises, mix des patients en fonction de leur groupe, préétablissement du planning opérationnel
Opérationnel	court terme (jour/temps réel)	règles d'ordonnancement	Détails des affectations des patients, insertion des urgences

Nous nous proposons dans les sous-sections suivantes de rapporter les travaux de recherche qui ont été menés sur chacun de ces niveaux de gestion.

2.4.1 Gestion à long terme

Evaluation de la performance des unités de soins

Nous le disions en introduction de cette section dédiée à la gestion des lits et des unités de soins, la gestion stratégique des unités de soins doit permettre de définir des objectifs politiques à long terme. La nature des objectifs à atteindre se traduit bien souvent par la recherche d'un taux de service cible. Par taux de service, nous entendons le rapport entre le nombre de patients traités, le temps d'attente avant la prise en charge et le taux de refus pour cause de service surchargé. Bien sûr, tout cela doit être effectué en tenant compte des coûts que représente l'allocation de ressources aux unités de soins. Un nombre insuffisant de lits provoque le départ des patients pour d'autres hôpitaux, l'augmentation

6. Nouvel Hôpital Estaing.

de la file d'attente et du stress. Par exemple, les patients des urgences vont empiéter sur les services de chirurgie ce qui va avoir pour double conséquence d'augmenter les temps d'attente pour ce service et de causer des reports d'admissions planifiées. A l'opposé, un nombre trop important de lits va gâcher, de par leur inutilisation, des ressources pourtant dures à obtenir (Gorunescu *et al.*, 2002).

Il apparaît donc comme primordial, pour la gestion à long terme, de définir des outils permettant l'évaluation de la performance d'un système. La simulation est un outil qui s'applique très bien dans ce cas. Elle permet, une fois les statistiques sur le flux de patients établies, de tester différentes organisations et d'en relever des indicateurs de la performance. Pour Kim *et al.* (1999), ces indicateurs sont la qualité de service et le taux d'occupation des unités de soins. Pour Williams *et al.* (2010) les mesures de la performance sont le temps d'attente et le taux d'utilisation des ressources. Pour Gorunescu *et al.* (2002), seul le taux d'utilisation de la ressource lit est important car de lui va découler les temps d'attentes constatés et la qualité de service. Enfin Creemers et Lambrecht (2010) considèrent quatre mesures de la performance dont le nombre d'heures supplémentaires.

Toutes ces études ont été faites pour des types de problèmes relevant de services particuliers et leur application à d'autres unités ne semble pas forcément possible. Kim *et al.* (1999) exposent clairement que leur modèle, défini pour un service de soins intensifs, ne convient pas pour l'étude d'unités présentant des durées de séjour longues. A l'inverse, Gorunescu *et al.* (2002) qui cherchaient à définir le taux de remplissage permettant un fonctionnement optimal d'un service de gériatrie ne considèrent dans leur modèle que des séjours longs (durée moyenne de 25 jours). Creemers et Lambrecht (2010) considèrent une approche très générale de la mesure de performance dans les systèmes de service. Par système de service, nous entendons bien entendu les systèmes de santé mais également les services au sens général de l'industrie où un client prend rendez-vous, passe par une file d'attente et reçoit son service en monopolisant plusieurs ressources. Le biais dans cette approche réside certainement dans sa trop grande généralité et donc le fait de ne pas prendre en compte des particularités propres aux systèmes de santé. Quoi qu'il en soit, cette approche semble pouvoir s'appliquer de préférence à des systèmes de type ambulatoire.

Finalement, nous retiendrons l'étude de Williams *et al.* (2010) comme étant la plus complète sur le domaine. Les auteurs s'intéressent en premier lieu au processus d'arrivée des patients avant de chercher à caractériser l'utilisation des ressources de l'hôpital. Il en résulte une étude de performance relativement globale ayant comme première application la mise en avant des points où se situent les ressources goulots.

Le problème de dimensionnement des services

Si l'on devait résumer à une seule problématique la gestion à long terme des unités de soins, ça serait sans aucun doute le problème du dimensionnement en nombre de lits. A la base de la résolution de ce problème se trouve des informations permettant de caractériser le flux de patients. Ces informations doivent être assez précises pour ne pas sous-estimer par la suite les besoins en termes de lits. Ainsi, des informations relatives aux profils de demandes mensuelles, hebdomadaires et même horaires pour les urgences doivent être utilisées afin d'obtenir une modélisation cohérente (Harper et Shahani, 2002). A l'aide de ces données, le processus d'arrivée des patients peut être caractérisé et ainsi permettre de définir des séries de scénario de demande. Pour certains services dont l'activité peut varier tout au long de la semaine, certains auteurs comme Lapierre *et al.* (1999) recommandent de s'intéresser à une journée en particulier, sans se préoccuper des autres, et ainsi générer les séries de demande par jour de la semaine. Cette approche permet de tirer pleinement profit de la variabilité inhérente aux jours de la semaine mais peut cependant présenter l'inconvénient de nécessiter des données trop précises. En effet, en cherchant des données pour chaque jour de la semaine pour chaque service, on réduit considérablement la taille de l'échantillon servant de base à la modélisation du flux de patients entrants. Le risque est alors d'avoir des estimations faussées par l'imprécision de l'échantillon considéré. D'autres auteurs proposent une approche plus modérée en considérant uniquement les jours ouvrés et les week-end (Bekker et de Bruin, 2010).

Le premier pré-requis ici est donc l'existence de données statistiques permettant la définition du modèle de flux de demande. Cependant, le problème du dimensionnement se pose principalement lors du choix de l'ouverture d'une nouvelle unité ; lorsqu'une unité est déjà existante, on se rapporte plus fréquemment à un problème de redistribution des ressources. La difficulté à surmonter en premier lieu est, dans le cas d'un nouveau service, l'absence de données historiques. Utley *et al.* (2003) ont proposé une technique se basant sur la définition d'une probabilité pour un patient d'être encore présent j jours après son admission. Ils ont appliqué leur approche pour le choix de la création d'une nouvelle unité de soins intermédiaires permettant de délester les soins intensifs. Ils ont ensuite généralisé leur approche dans Utley *et al.* (2005).

Le nombre de lits à affecter à une unité est ensuite conditionné par les objectifs fixés par la direction des établissements. Ces objectifs peuvent être assez variés. Par exemple, un objectif peut être une combinaison du taux d'occupation, d'un niveau de service cible et d'un taux d'admission de nouveaux patients pour Kokangul (2008), ou bien alors la combinaison du nombre de patients transférés vers un autre hôpital, le nombre de patients refusés et le taux d'occupation pour Ridge *et al.* (1998), ou encore le taux de refus seul pour Bekker et de Bruin (2010). Enfin Huang (1995) et Harper et Shahani (2002) prennent en considération le traitement des patients urgents lors de leurs choix de dimensionnement.

L'application de ces travaux se retrouve dans tout l'hôpital. Harper et Shahani (2002), Nguyen *et al.* (2005) et de Bruin *et al.* (2010) proposent une approche pour le dimensionnement d'un service de soins en général. Ramis *et al.* (2001) s'intéresse au dimensionnement d'un service ambulatoire de chirurgie, Asaduzzaman *et al.* (2010) à une unité néonatale. Un plus grand nombre d'études considère le dimensionnement de services de soins intensifs (Kokangul, 2008) et (Ridge *et al.*, 1998). Gallivan *et al.* (2002) se focalisent sur la variabilité des durées de séjour pour les services de soins intensifs en cardiologie. Enfin Litvak *et al.* (2008) apportent une autre vision pour le dimensionnement des services de soins intensifs : la coopération inter-hospitalière. Face à la difficulté que peut avoir un hôpital à prendre en charge des patients urgents dans leur service de soins intensifs, les auteurs proposent que des hôpitaux se regroupent et mettent à disposition un petit nombre de lits pour les urgences. L'étude présente une méthode qui, pour un taux d'acceptation cible donné, fournit le nombre de lits à réserver dans chaque établissement.

Finalement, nous noterons la dominance de deux familles de méthodes que sont la simulation (Ridge *et al.*, 1998; Lapierre *et al.*, 1999; Kim *et al.*, 1999; Ramis *et al.*, 2001; Harper et Shahani, 2002; Williams *et al.*, 2010) et la théorie des files d'attente (Huang, 1995; Gorunescu *et al.*, 2002; Bekker et de Bruin, 2010; de Bruin *et al.*, 2010; Asaduzzaman *et al.*, 2010; Creemers et Lambrecht, 2010).

Réallocation et redéploiement des ressources

La constante évolution des systèmes de santé que nous mentionnions en introduction de ce chapitre s'applique également pour les flux de patients. Ainsi, le flux de patients considéré lors du dimensionnement initiale d'un service peut s'avérer au bout de quelques années ne plus correspondre à la réalité. Il est alors nécessaire de redistribuer les ressources entre les différents services. L'idée de base est de redistribuer les lits des services présentant un faible taux d'occupation aux services dont la capacité est insuffisante (Akkerman et Knip, 2006; Cochran et Bharti, 2006). Cela permet d'une part une meilleure utilisation des ressources mais également une prise en charge plus rapide des patients en attente d'admission dans un service très demandé.

Vissers (1998) propose d'anticiper sur l'évolution démographique lors de la planification des ressources. Pour ce faire, il étudie, à partir des prévisions de croissance de la population et à partir des prévisions d'évolution des durées moyennes de séjour, les prévisions de demande de soins. A l'aide de cette prévision, l'auteur met en avant les ressources faisant défaut classiquement appelées ressources goulots. L'allocation de ces ressources est effectuée dans un premier temps afin de garantir le fonctionnement de chaque spécialité et de l'hôpital tout entier. Les autres ressources moins critiques sont allouées par la suite.

La réallocation de ressources peut aussi être vue comme la réservation de lits au sein d'un service pour certaines pathologies ou pour certain type de patients ciblés. C'est le cas de Huang (1998) qui étudie le redimensionnement des services d'hospitalisation en fonction du flux de patients aux urgences dans le but de fluidifier ce dernier. Kim *et al.* (2000) envisagent aussi d'adapter la politique de gestion des lits dans les services de soins intensifs en fonction du planning du bloc opératoire et de l'activité urgente.

Enfin, notons que chaque modification dans l'allocation des ressources représente un coût. C'est en cherchant à minimiser ce coût d'exploitation que Akcali *et al.* (2006) ont proposé une méthode s'inspirant des modèles de décision markovien et de la recherche du plus court chemin dans un graphe. Les auteurs construisent un graphe dont les sommets représentent des états de distribution des ressources et les arcs sont pondérés par le coût de passage d'un état à un autre tout en tenant compte des flux de patients probables. En réduisant le nombre d'états du système aux seuls changements réalistes de l'allocation de ressources, l'approche présente un rapport qualité/rapidité intéressant.

2.4.2 Gestion à moyen terme

La gestion à long terme que nous venons de voir considérait les patients uniquement sous forme de flux requérant des ressources. En d'autres termes, les patients étaient considérés comme une entité agrégée dont le besoin en ressources dépendait de la taille de cette entité. La gestion à moyen terme va commencer à considérer le patient en tant qu'entité propre ayant des caractéristiques et des besoins ou demandes à prendre en compte du mieux possible. Les ressources disponibles étant connues, la gestion à moyen terme va s'intéresser à la planification des arrivées des patients tout en tenant compte des futures demandes de soins si celles-ci peuvent interférer avec la prise en charge du patient.

En clair, la gestion à moyen terme concerne principalement la problématique de choix des dates d'admission des patients. Ce choix doit être fait en tenant compte des ressources critiques qui vont être nécessaires tout au long du séjour. Cette affirmation soulève deux questions. Tout d'abord quels sont ces ressources critiques, ces ressources dont la prise en compte est indispensable lors de la planification d'une admission ? Deuxièmement, comment doit-on procéder pour garantir que les ressources requises seront bien disponibles ?

Selon Gemmel et Van Dierdonck (1999), plus le nombre de ressources prises en considération est important plus la qualité du planning résultant s'en trouvera élevée. Selon ces auteurs, deux types de planification des admissions peuvent être mises en place. Pour la première qu'ils nomment « Decision Support system for Admission Planning », la réponse à la demande est donnée par affectation des ressources. Cette affectation de ressources est la résultante d'une succession de décisions hiérarchisées du niveau stratégique au niveau opérationnel. Ils y opposent une méthode qu'ils nomment « Hospital Ressource Planning » qui est empruntée au système ERP (Enterprise Ressource Planning), bien connu dans le milieu industriel, qui se base sur les besoins en différentes ressources. Les auteurs effectuent une comparaison de ces deux approches sur sept critères dont nous synthétisons les résultats Table 2.2.

La conclusion de cette étude indique que les systèmes de type HRP sont plus performants et surtout plus facilement implémentables en pratique. Cet avantage leur est donné de par leur flexibilité plus grande sur l'ensemble des ressources à prendre en compte. Cependant l'inconvénient majeur de ce type d'approche est de ne pas suffisamment considérer les demandes futures lors de l'étape de planification. Il s'ensuit un fort risque de temps mort dans l'utilisation de ces ressources. De surcroît, l'ensemble des ressources requises lors du séjour d'un patient ne sont pas obligatoirement connues dès la planification des admissions. D'autres ressources peuvent être gérées indépendamment et possèdent donc un système de planification qui leur sont propre. Ces ressources doivent donc être exclues des systèmes HRP. Nous en avons vu des exemples en section 2.3 dédiée au management des ressources.

La suite de cette section consacrée à la gestion à moyen terme, le niveau tactique, des unités de soins va maintenant s'intéresser à comment prendre en compte les besoins futurs puis aux choix des dates d'admission des patients.

Prévision de besoins futurs

La prédiction des besoins futurs peut se faire pour certaines ressources particulières comme par exemple la prédiction des besoins en poche de sang transfusable dans le bloc opératoire (Walczak et Scharf, 2000). Plus généralement, les modèles permettant la prévision de la charge de travail dans les jours à venir se sont intéressés à modéliser la probabilité de présence d'un patient dans un service. La ressource 'lit" est incontestablement une ressource critique à prendre en compte lors de la planification à moyen terme de l'activité des services de soins.

Une première condition à l'estimation du nombre de lits requis dans les jours à venir est de connaître la durée de séjour des patients. Certains services comme la chirurgie, dont les processus de soins sont bien maîtrisés, n'ont en général que peu de problèmes vis-à-vis de l'estimation de cette donnée. Pour les services de long séjour, cela peut être plus fluctuant. Marshall *et al.* (2004) proposent un modèle utilisant des distributions de probabilités conditionelles afin de déterminer la durée de séjour des patients en gériatrie. La pédiatrie est une autre discipline où les durées de séjour sont mal maîtrisées. Pour y faire face Walczak *et al.* (2002) utilisent un réseau de neurones afin de prédire les durées de séjour dès la consultation d'admission du patient. La durée de séjour d'un patient demeure une donnée très fluctuante et relativement imprévisible. En effet, chaque séjour est soumis à une multitude d'évolutions médicales possibles ce qui rend assez peu performantes des approches comme les deux que nous venons de citer.

Certains auteurs ont cherché, en se basant sur des statistiques existantes dans les services, telles les durées de séjour et les taux d'arrivée de nouveaux patients, à prédire l'évolution globale d'un service à partir de son état courant. Ce type d'approche présente l'avantage de partir d'une base déterministe avant d'y introduire des évènements stochastiques. C'est ce qu'ont fait Littig et Isken (2007) qui ont collecté les données de différents processus telle la programmation des admissions des patients ou la gestion de l'occupation des lits afin de proposer deux modèles de prévisions : l'un du flux d'entrée, l'autre du flux de sortie. Ces prévisions sont, ensuite, utilisées pour établir une estimation du taux d'occupation des lits dans les différents services sur les jours à venir. Dans le même registre, Pagel *et al.* (2010) ont étudié le cas d'un service de soins intensifs en pédiatrie où les auteurs, dans leur collecte de statistiques, ont séparé les cas provenant des urgences des patients électifs. Le modèle proposé permet ensuite de calculer la distribution de probabilité de la demande en lit sur les prochains jours. La recherche de la distribution de probabilité est un problème que nous avons déjà vu précédemment dans les travaux de Utley *et al.* (2005). Cette distribution globale est estimée à partir de celle qui considère qu'un patient est encore présent dans le service j jours après son admission.

Pour terminer sur l'estimation des futurs besoins, revenons sur un travail que nous avons également déjà cité précédemment et qui permet de traiter du cas des patients urgents : l'étude de Huang (1995). L'auteur a utilisé la théorie des files d'attente pour représenter un système où les patients arrivent aléatoirement, séjournent pour une période aléatoire de temps puis, soit ils libèrent le lit qu'ils occupaient soit ils sont transférés vers un autre service. Comme premier résultat de l'étude d'un tel système, l'auteur a montré que pour chaque jour de la semaine, l'utilisation de lits, dans un service, par des patients provenant des urgences suit une loi de Poisson.

TABLE 2.2 – Comparaison entre DSAP et HRP

	Decision Support system for Admission Planning (DSAP)	Hospital Ressource Planning system (HRP)
1. Esprit et but	Sélection des patients à admettre d'une certaine manière permettant la prise en charge des urgences. Prise en compte des capacités des ressources	Assurance que la bonne ressource est disponible au bon moment. Essaie d'éviter les blocages
2. gestion centralisée	C'est un des fondements des 2 approches. On peut cependant décentraliser le choix pour certaines unités ne partageant pas leur ressources.	
3. informations nécessaires	Capacité en lit, ressources infirmières et bloc opératoire. Biologie et radio (par ex) ne sont pas considérées comme critique pour la durée moyenne de séjour (à prouver ou bien assurer une charge stable et équilibrée de ces services)	Prend implicitement en compte toutes les ressources
4. classification des patients	classification par les ressources demandées (service)	Notion de groupe homogène de patients. Problème : cette notion n'est pas connue entièrement dès le départ.
5. projection sur besoins futurs	Utilisation de modèles statistiques pour prévision	Principe des demandes dépendantes
6. prise en charge des urgences	explicitement pris en compte par le modèle de prévision	Prédiction du nombre d'urgences pour réservation de capacité. Projection à partir du moment où les urgences sont admises
7. possibilité d'implémentation	Besoin de paramétrage pour chaque système. Nécessité de connaître l'ensemble des informations.	Testé par simulation. Besoin de calculer fréquemment le planning des admissions

Choisir la date d'admission des patients

Le choix de la date d'admission d'un patient électif est guidé par plusieurs paramètres. Tout d'abord le degré d'urgence médicale va permettre de définir une fenêtre de temps dans laquelle l'admission doit être faite. Ensuite, la date doit être choisie généralement de manière à limiter le temps d'attente du patient. Enfin, le taux d'occupation du(des) service(s) receveur(s) va largement influer sur le choix de cette date. En effet, comme déjà mentionné dans Gemmel et Van Dierdonck (1999), la disponibilité des ressources tout au long du séjour est cruciale pour garantir un planning efficace. La prise en compte des ressources lors de la planification des admissions joue également un rôle dans l'évaluation d'une politique d'admission. En effet, il est courant, et nous y reviendrons dans la partie suivante consacrée aux services ambulatoires, de chercher un planning cherchant à maximiser l'utilisation des ressources et notamment les ressources coûteuses (Garg *et al.*, 2010). Cependant, du point de vue patient, un planning optimal serait plutôt du type sans attente avant l'admission. La planification des admissions doit permettre de trouver un compromis entre ces objectifs parfois contradictoires. Dans ce but, Vissers *et al.* (2007) ont proposé une plateforme se basant sur la simulation afin de pouvoir évaluer une politique d'admission.

Nous retrouvons une étude portant sur le choix de la date d'admission des patients dès 1970 (Connors, 1970) où la prise en compte des ressources nécessaires, ici le lit, est faite en tenant compte de la probabilité qu'un patient soit toujours présent après un certain nombre de jours passés dans l'établissement. La disponibilité des ressources est prise en compte à plusieurs niveaux : hôpital, service et chambre. Un dépassement de capacité à l'un de ces niveaux entraîne une pénalité plus ou moins importante dans le calcul de la fonction objectif. Beaucoup d'études ont suivi comme le montre l'état de l'art de Smith-Daniel *et al.* (1988).

D'autres études ne considèrent pas uniquement le lit comme ressource à prendre en compte lors de la planification des admissions. Hutzschenreuter *et al.* (2008) se focalisent sur les services de chirurgie cardio-thoracique et prennent en compte comme ressource additionnelle les lits de réanimation. Dans le cas particulier traité dans cette étude, ce choix se justifie par le nombre élevé de patients admis dans un service de ce type avant de retourner dans le service de chirurgie.

Il n'est pas toujours possible de donner directement une date d'admission pour un patient. Pour certaines pathologies, la chirurgie cardio-thoracique par exemple, nécessitant des ressources pointues, très demandées et sujettes à recevoir des patients urgents, il est parfois préférable dans un premier temps de donner au patient une date prévisionnelle d'admission ou un intervalle de temps pendant lequel l'admission aura lieu. Ceci est particulièrement vrai pour les services de chirurgie avec l'utilisation du bloc opératoire. Il est commun de devoir décaler une intervention en fonction des variations de la demande de cette ressource (Jebali *et al.*, 2003).

Pour finir, notons que la planification des admissions peut aller au-delà du simple fait du choix d'une date. Il est possible de l'utiliser afin de prévoir, ou suivant le cas de se prémunir, d'un besoin d'ajustement de dernière minute des ressources. Par exemple, l'ouverture de lits supplémentaires afin de faire face à une augmentation temporaire de l'activité due aux urgences peut être anticipée. En tenant compte des dates prévues d'admissions, des fenêtres dans lesquelles les admissions peuvent être modifiées et du taux de patients provenant des urgences, Wang *et al.* (2007) proposent un outil permettant d'ajuster le nombre de patients à admettre un certain jour et, si besoin est, de décider de l'ouverture de lits supplémentaires.

Planification des rendez-vous ambulatoires

La planification des activités de soins ambulatoires est particulière au sens où les besoins de ressources, notamment les lits, ne sont pas perturbés par les besoins futurs. En effet, les services de soins ambulatoires fonctionnent de manière discrète puisque l'activité d'un jour donné est clairement démarquée de celle du lendemain par une période de fermeture. La gestion de la prise de rendez-vous pour ce type d'activité

est à rapprocher de la gestion des ressources de diagnostic (section 2.3.1). Certains articles déjà cités comme Sickinger et Kolisch (2009) ou Santibáñez *et al.* (2009) auraient leur place dans cette section. La ressource, qu'il s'agisse du médecin ou du bloc opératoire, est considérée comme coûteuse. L'objectif va donc être de garantir un taux d'utilisation élevé pour cette ressource. Ce postulat n'est pas sans poser de problème si la ressource considérée peut être amenée à traiter des patients urgents. En effet, des créneaux doivent être réservés pour permettre la prise en charge éventuelle de ce flux de patients.

D'une manière générale, on va donc chercher à optimiser l'utilisation d'une ressource tout en cherchant à garantir la qualité du service rendu au patient. Cette qualité de service, du point de vue du patient, se mesure en termes de temps d'attente. Gupta et Denton (2008), dans leur revue de littérature consacrée au sujet, considèrent deux types d'attente : directe et indirecte. Le premier type d'attente est tout simplement le temps passé par le patient dans la salle d'attente le jour de son rendez-vous. L'attente indirecte est, quant à elle, le temps qui va s'écouler entre la date de prise de rendez-vous et le jour de rendez-vous lui même. Même si l'attente directe est source d'inconfort pour le patient, elle reste globalement tolérable et a pour principal impact d'être la cause d'heures supplémentaires pour le personnel de l'hôpital. L'attente indirecte est beaucoup plus problématique. Une attente trop longue se révèle être synonyme d'impossibilité d'accès aux soins dans un délai raisonnable et donc potentiellement risquée pour la santé du patient (Murray et Berwick, 2003).

Un planning optimal de rendez-vous médicaux doit permettre de réduire les attentes des patients, de laisser des petites périodes de temps libre aux médecins et de finir à l'heure. Un paramètre important à prendre en compte est le taux de patients ayant rendez-vous qui ne se présentent pas. Ce taux est généralement compris entre 5 et 30% (Zeng *et al.*, 2010). Kaandorp et Koole (2007) ont montré qu'avec un taux constant et commun à tous les patients, le problème est multimodulaire[7] et peut donc être résolu par un algorithme de recherche locale. Zeng *et al.* (2010) ont poursuivi l'étude théorique de ce problème et ont démontré que, dans un cas ne respectant pas l'hypothèse émise par Kaandorp et Koole, le problème n'est alors plus multimodulaire. Afin de prendre en compte ces patients ne se présentant pas, Zeng *et al.* ont utilisé une pratique existante dite de « surbooking » consistant à affecter plus de patients que la capacité théorique ne le permet. Une telle pratique peut cependant se révéler dangereuse si elle est mal maîtrisée. Les auteurs nous proposent un algorithme de recherche locale permettant de construire le planning des rendez-vous avec surbooking.

Le problème se posant avec les patients ne se présentant pas à leur rendez-vous est double. En plus de leur non présence entraînant une période d'inutilisation des ressources qui leur étaient allouées, ces patients ont une probabilité assez forte de reprendre rendez-vous pour recevoir leur soin initialement prévu. Ce phénomène est pris en compte dans les travaux de Grenn et Savin (2008). L'approche proposée va même plus loin puisqu'elle considère que plus la file d'attente est longue plus la probabilité qu'un patient ne se présente pas est grande.

Pour finir avec la gestion des rendez-vous, nous allons nous intéresser aux travaux de Cayirli *et al.* (2006). L'étude recherche dans un premier temps à établir ce que les auteurs nomment le séquencement des patients ; autrement dit, le choix des dates de rendez-vous. La principale hypothèse formulée dans cette étude concerne une distinction sur les temps de service entre les nouveaux patients et les patients qui sont déjà venus et pour qui le temps de traitement est considéré comme moindre. Partant de cette distinction, six paramètres sont pris en compte pour définir un service d'un hôpital parmi lesquels la taille du service, le taux de patients ne se présentant pas, le taux de patient se présentant sans rendez-vous ou encore la ponctualité moyenne des patients. En utilisant la simulation, les auteurs testent six règles de séquencement qui effectuent un panachage des nouveaux patients et des patients déjà venus. Cette étude ne s'arrête pas là puisque les auteurs vont ensuite, sur le même principe, tester des règles permettant la

7. La notion de multimodularité est une extension de ce qu'est la convexité en deux dimensions pour une fonction multidimensionnelle.

résolution du problème à court terme qu'est l'ordonnancement des patients tout au long de la journée.

2.4.3 Gestion à court terme

Pour terminer sur la gestion des services de soins, nous allons à présent voir le versant très opérationnel de la gestion à court terme. En effet, ce niveau de gestion est organisé autour du patient et à sa gestion au jour le jour. Le problème est simple : des patients, qu'ils soient électifs et dont l'admission a lieu ce jour ou qu'ils proviennent des urgences, vont devoir être affectés à un lit. Le lit se situant dans une chambre, son choix va être contraint par certaines données comme la non mixité des chambres ou la présence de certains équipements médicaux. Cependant le premier choix à effectuer est celui du service qui va accueillir le patient. La logique et la raison veulent que le patient soit affecté là où le personnel est le plus compétent : dans l'unité de soins relevant du domaine de la pathologie. Cependant, de plus en plus d'établissements ont recours à une mutualisation des lits entre les services. Cela offre un gain en flexibilité pour la gestion opérationnelle des lits mais présente également des inconvénients. Le principal défaut de cette organisation est sans doute la polyvalence du personnel qui est requise pour permettre de prodiguer n'importe quel soin dans n'importe quel service. Les protocoles de soins peuvent être très différents d'une spécialité à l'autre et donc être méconnus par le personnel infirmier d'une spécialité donnée. Cela pose des problèmes de sécurité du patient. Pour pallier à cet inconvénient mais conserver l'intérêt de la mutualisation, certains hôpitaux ne mutualisent leurs lits qu'entre des services de spécialités proches. On parle alors de mutualisation partielle.

Globalement, l'ensemble des études menées sur la gestion à court terme des unités de soins s'accorde pour se fixer comme objectif de chercher à minimiser le nombre de changement de lits d'un patient tout au long de son séjour. Cette minimisation présente un double intérêt. Tout d'abord le confort du patient s'en trouve amélioré. Ensuite, du point de vue économique, effectuer une mutation représente un coût pour l'hôpital puisque ce changement va nécessiter des moyens humains pour être effectif. Wang *et al.* (2008) ont montré, sous certaines hypothèses et notamment en supposant connaître le planning des admissions sur un certain horizon, qu'il existe au moins un planning d'affectation avec aucun changement de lits durant le séjour des patients. Malheureusement, dans la réalité, le planning des admissions est rarement connu longtemps à l'avance et est, de plus, bien souvent perturbé par les arrivées non programmées des urgences.

Pour réduire le nombre global de mutations de patients entre services, Thomson *et al.* (2009) proposent une approche proactive en envisageant de muter des patients en prévision de besoins de place dans certaines unités. Les auteurs ont d'abord modélisé le problème sous forme d'un processus de décision markovien. Le nombre important d'états représentant l'ensemble des opérations possibles sur un horizon donné étant prohibitif, une heuristique a été proposée. Cette heuristique consiste en la définition de trois seuils pour chaque unité de soins. L'affectation et la mutation des patients a lieu suivant les trois étapes suivantes :

1. Transfert Proactif : transfert des patients d'un service j où le nombre de lits disponibles est inférieur à un certain seuil r_{j1} vers un service j' où le nombre de lits disponibles est supérieur à un autre seuil $r_{j'2}$.
2. Affectation primaire (patients dans le bon étage) : affectation des patients en attente à leur étage aussi longtemps qu'il y a de la place.
3. Affectation secondaire (patients dans un autre étage) : Affectation des patients dans des services avec au moins r_{j3} lits disponibles et en accord avec le service maximisant les revenus.

En comparant pour de petites instances cette heuristique avec la solution optimale fournie par l'approche markovienne, les auteurs montrent expérimentalement que leur approche fournit des résultats proches d'une solution optimale.

Demeester *et al.* (2010) ont, quant à eux, chercher à répondre au problème d'affectation des patients dans les chambres en plaçant une priorité importante sur le respect des équipements techniques requis par le patient. Pour ce faire, ils ont développé une méthode de recherche taboue avec voisinage variable visant, en premier objectif, à satisfaire les besoins techniques de chaque séjour patient.

Résumer la gestion à court terme des services de soins à l'affectation des patients à un lit reviendrait à en oublier un volet important. La prise en charge des patients urgents est un problème de tous les jours dans les services de soins. La MeaH [8] cible de manière très pratique les barrières à la prise en charge des patients urgents et indique des règles de bonnes pratiques à suivre dans les services (Cauterman et Engel, 2007). Parmi ces règles, nous en retiendrons une qui consiste à faire effectuer les sorties des patients le plus tôt possible dans la journée. Cette règle, pour être applicable, doit passer presque nécessairement par une anticipation la veille de la sortie d'un patient. La communication entre les médecins et le personnel soignant, et notamment les cadres de santé, doit être existante et efficace pour permettre une telle anticipation. Harrison et Escobar (2010) se sont intéressés à ce processus de prévision des départs de patients. Ils ont proposé un modèle permettant d'évaluer la probabilité de départ prochaine d'un patient en tenant compte de divers facteurs et notamment de la sévérité de la maladie.

Enfin, Bechar et Guinet (2006); Ben Bachouch *et al.* (2007) ont proposé des approches permettant d'insérer des nouveaux patients, patients urgents inclus, dans un planning d'affectation aux lits existant. Les deux études se sont basées sur un problème de programmation linéaire.

2.5 Conclusion

Cet état de l'art s'est intéressé au management de toutes les composantes qui peuvent avoir un lien avec les urgences et donc en favoriser le fonctionnement.

Nous constatons une forte prédominance d'études portant sur des ressources très pointues. Ordonnancement du bloc opératoire, règle de prise en charge des patients en soins intensifs ou allocation de créneaux horaires sur des ressources d'imagerie sont des problématiques qui reviennent souvent dans la littérature du management des systèmes hospitaliers. Cela peut se comprendre. Ces ressources sont, comparées à d'autres comme le lit, rares et très demandées. Il est important d'en maîtriser le fonctionnement et d'en optimiser l'amortissement économique de leur achat.

Comme l'a dit Gupta et Denton (2008), la recherche opérationnelle appliquée au contexte hospitalier peut permettre de réduire les coûts et d'améliorer l'accès aux soins pour le patient. L'impression que nous retenons des travaux récemment publiés et présentés dans ce chapitre est une forte prise en compte de l'aspect économique. La dimension patient n'est prise en compte que sous forme d'un taux d'admission cible ou d'une longueur de file d'attente. Le confort, les souhaits ne sont que très peu pris en considération. De ce fait, et par extension, les charges de travail pesant sur le personnel soignant ne sont pas prises en compte au niveau où elles devraient l'être. Chacun cherche à faire de son mieux face à sa mission. On assiste à l'apparition d'un cloisonnement dans un système qui a éminemment besoin d'être ouvert et communiquant entre ses entités.

Le problème de cloisonnement se ressent grandement au niveau des SAU. Ces unités ont un grand besoin d'interopérabilité afin de permettre un écoulement des patients qu'elles traitent. Le nombre d'études, relevant de la recherche opérationnelle ou du génie industriel, traitant ou prenant en compte les patients urgents est faible. Nous voyons deux causes. Soit le processus de prise en charge des urgences apparaît comme complexe à maîtriser et se retrouve délaissé au profit d'autres problématiques, soit l'importance des urgences est considérée comme mineure. Certains auteurs n'hésitent pas à annoncer que refuser une admission à un patient urgent n'a qu'un faible impact sur l'économie de l'hôpital.

8. Mission d'Expertise et d'Audit Hospitalier du ministère de la santé. Cette entité a pour but d'identifier et de tenter de remédier aux problèmes organisationnels des hôpitaux

Nous pensons que la prise en charge des patients urgents étant vitale pour l'être l'humain, elle doit l'être également pour l'hôpital. La recherche opérationnelle doit permettre d'atteindre les objectifs financiers déjà étudiés par le passé mais doit également incorporer la dimension hautement humaine de l'hôpital. Les barrières actuelles, le cloisonnement des services de soins, doivent être surmontées afin que chaque entité de l'hôpital travaille de concert avec les autres. Tels sont nos convictions, nos motivations pour la suite de ce travail.

Chapitre 3

Choix des dates d'admission des patients

Sommaire

3.1 Introduction
3.2 Le problème de planification des admissions de patients électifs
3.3 Le problème de planification journalier
3.4 Estimation des besoins pour les futurs patients
3.5 Différentes approches de résolution
3.6 Expérimentations numériques
3.7 Synthèse

AU NIVEAU TACTIQUE de l'approche décrite précédemment, on retrouve le problème de choix des dates d'admissions des patients réguliers. Gemmel et Van Dierdonck (1999) définissaient la planification des admissions comme une partie du processus d'admission qui commence lors de l'arrivée de la demande d'hospitalisation et se termine avec l'entrée effective du patient. Nous allons dans ce chapitre répondre à la question « *comment planifier les entrées directes tout en garantissant la possibilité de recevoir les futurs patients urgents de manière optimale ?* » en utilisant un modèle de programmation dynamique stochastique visant à réduire d'éventuelles situations de surcharge de l'hôpital. Ce problème, comme nous allons le voir, doit permettre de prendre en compte non seulement les besoins en lits des futurs patients urgents qui seront amenés à être hospitalisés mais également les besoins des futurs patients réguliers dont l'hospitalisation sera prononcée ultérieurement. Les deux flux doivent pouvoir coexister en minimisant leur impact mutuel. Une des principales difficultés que nous allons rencontrer porte sur la variabilité dans les flux d'arrivée de patients, la variabilité dans les durées de séjours et les évènements qui peuvent survenir aléatoirement pendant le séjour.

3.1 Introduction

Le problème de choix des dates d'admission des patients électifs a déjà été décrit dans ses grandes lignes dans les chapitres précédents. Cependant, nous allons nous attacher à synthétiser toutes les informations nécessaires permettant d'aboutir à une description concise du problème traité dans les sections suivantes.

Les capacités d'accueil d'un hôpital sont limitées par le nombre de lits théorique dont dispose l'établissement. En général, chaque établissement dispose d'une marge de lits supplémentaires permettant de faire face à des pics d'activité. Cependant lorsque le taux de remplissage moyen avoisine les 90%, il devient fréquent de se retrouver temporairement en situation de surcapacité. Ce mode de fonctionnement étant anormal, cela représente une situation indésirable qui a un coût financier ou organisationel. Il apparaît donc comme nécessaire de réussir à s'organiser, de trouver des flexibilités, de manière à limiter ces situations.

Clairement, les patients provenant des urgences ne permettent aucune flexibilité permettant d'agir sur le taux de remplissage de l'hôpital. Quand un patient est en attente au SU, il n'y a pas d'autres solutions que de l'admettre, de surcoît rapidement, dans un service[1]. Seuls les patients réguliers présentent une flexibilité sur laquelle nous pouvons jouer afin de tenter d'équilibrer la charge globale de l'hôpital : la date d'admission. Le choix de cette date doit être fait avec une attention particulière : l'admission d'un patient un jour j va avoir des répercussions sur l'ensemble de son séjour. Il devient primordial, donc, de pouvoir estimer les futures arrivées lors de la phase de planification des admissions. On s'autorise, de plus, à pouvoir replanifier des patients déjà programmés de manière à pouvoir mieux faire face aux évolutions du sytème entre la décision d'hospitalisation et la date d'entrée programmée.

Avant de formaliser mathématiquement le problème, revenons sur un type de patients présentés dans le chapitre 1 : les patients semi-urgents. Ces patients sont plus restrictifs car ils nécessitent une prise en charge rapide, idéalement dans les 3 jours, au pire dans la semaine. Si ces délais ne sont pas respectés on s'expose à une dégradation de l'état de santé du patient qui peut avoir deux conséquences :

1. Augmentation de la durée de séjour ou du moins des coûts de prise en charge du patient, celui-ci étant plus malade.
2. Entrée du patient par les urgences si la dégradation est trop importante.

Dans le deuxième cas, non seulement on a perdu la flexibilité du choix de la date d'admission mais en plus le patient devra vraisemblablement être hospitalisé plus longtemps[2].

La suite de ce chapitre va présenter le problème de décision dynamique sujet à une politique de planification journalière des admissions de patients électifs. Nous proposons 3 stratégies de planification, chacune considérant une estimation des besoins futurs en lit d'un manière différente, que nous comparerons aux données issue du Centre Hospitalier de Firminy.

3.2 Le problème de planification des admissions de patients électifs

A la fin de chaque journée, les médecins transmettent aux cadres de santé, ou au responsable des admissions, la liste des patients vus en consultation et dont l'hospitalisation est nécessaire. Le responsable a donc la connaissance des patients présents dans l'hôpital, des patients électifs déjà programmés dans les jours à venir et la liste des nouveaux patients à planifier. Nous faisons la supposition que l'annulation et la replanification d'une admission la veille de la date d'entrée prévue n'est pas possible. Les patients planifiés pour être admis lors de la prochaine période ne peuvent donc pas être replanifiés et seront admis. Tous les autres patients réguliers seront soumis à une replanification soumise aux évènements aléatoires suivants survenant à chaque période :

- nombre de nouveaux patients en provenance des urgences.

1. Certains établissement s'autorisent à envoyer des patients en attente aux urgences vers d'autres établissements. Devant la difficulté organisationnelle à admettre en urgence un patient déjà présent dans l'établissement, nous laissons, à la discrétion du lecteur, imaginer le cas de l'admission urgente d'un patient provenant d'un autre hôpital.

2. voir les statistiques des durées de séjour des patients urgents et des patients réguliers en annexe A.4

- nombre de nouveaux patients réguliers à planifier.
- nombre de départs des patients en fin de séjour.

Le premier objectif que nous nous fixons consiste à minimiser les dépassements de capacité. Cependant, nous devons prendre en compte d'autres objectifs comme les coûts, ou pénalités, liés à la planification d'un patient. Les patients ont souvent besoin de nombreux examens lors de leur admission. L'accès à certaines ressources, comme les équipements d'exploration fonctionnelle, doit donc être planifié. Il devient alors évident qu'admettre un patient le lendemain de sa consultation, où la décision d'hospitalisation est prise, va poser des problèmes d'organisation. Il risque d'en résulter une attente de ressource dès le début du séjour, ce qui est préjudiciable en terme de pertinence du séjour. Le début effectif du traitement va être retardé en attendant le résultat de l'examen ciblant précisément la pathologie dont souffre le patient. L'admission le lendemain de la date à laquelle le patient est connu doit donc être pénalisée de sorte à ne pas favoriser ce comportement.

Nous allons à présent modéliser le problème de planification des admissions en entrées directes sous la forme d'un modèle dynamique de programmation stochastique sur un horizon de H périodes où la capacité en lit est T à chaque période. A la fin de chaque période t, le nombre de patients urgents à admettre immédiatement [3] et le nombre de patients nouvellement connus à hospitaliser sont obtenus en fonction de différents processus aléatoires. Chaque patient peut être représenté par le triplet :

- r_i la date à laquelle l'hospitalisation a été décidée, i.e. date à laquelle le patient est connu des services.
- p_i la durée de séjour prévisionnelle.
- d_i le délai maximum pour l'admission ; un patient doit être admis au plus tard à la date $r_i + d_i$.

Plus spécifiquement, nous considérons les points suivants liés au processus aléatoire d'arrivée des patients :

- $\{n_\tau^r\}_{\tau=1}^{\infty}$, une série iid [4] de variables aléatoires, où n_τ^r est le nombre de patients réguliers nouvellement connus en période τ.
- $\{p_i^r\}_{i=1}^{\infty}$ la séquence iid de variables aléatoires des durées de séjour des patients réguliers.
- $\{d_i\}_{i=1}^{\infty}$ les délais maximaux pour l'admission des patients réguliers sont une séquence iid de variables aléatoires.
- $\{n_\tau^e\}_{\tau=1}^{\infty}$, une série iid de variables aléatoires, où n_τ^e est le nombre de patients urgents arrivant en période τ.
- $\{p_i^e\}_{i=1}^{\infty}$ la séquence iid de variables aléatoires des durées de séjour des patients urgents

Les distributions de probabilités de chacuns de ces éléments aléatoires peuvent être estimées à partir de données extraites du SIH de l'hôpital considéré. Certains de ces paramètres peuvent être très variables comme les durées de séjour qui varient de 1 à plus de 20 jours. D'autres peuvent être assez constantes. C'est le cas des délais d'admissions maximaux dans les unités de médecine où la majeure partie des patients relevant de semi-urgences, le délai est de l'ordre de quelques jours.

3. Dans la pratique les patients urgents ont été admis au cours de la période écoulée.
4. Indépendante et Identiquement Distribuée

3.2.1 Modélisation formelle

A la fin de chaque période τ, en fonction des patients nouvellement arrivés durant cette période, l'admission des patients réguliers est replanifiée et le nouveau planning des admissions est représenté par les variables de décision suivantes :

$$X_{it}^{\tau} = \begin{cases} 1 & \text{si le patient } i \text{ est planifié pour être admis en période } t \text{ dans le planning de la période } \tau \\ 0 & \text{sinon} \end{cases} \tag{3.1}$$

La suite de décisions prises pour le patient i est donc représentée par la matrice X_i et le vecteur X_i^{τ} représente la décision prise pour ce patient à la période τ. Il va de soit qu'à la fin d'une période, nous ne planifions, ou replanifions, que les patients qui ne sont pas encore admis. Nous pouvons donc définir la donnée A_i^{τ} qui indique si un patient i a été admis avant la fin de la période τ :

$$A_i^{\tau} = \sum_{\tau'=1}^{\tau-1} X_{i,\tau'+1}^{\tau'} \tag{3.2}$$

La décision d'admission des patients réguliers dépend du processus d'arrivée des patients et de leur durée de séjour. Elle doit donc être prise en tenant compte des deux variables stochastiques suivantes :

- $\overline{N}_{\tau}^{r}$ le nombre cumulatif de patients réguliers connus en périodes $1, \ldots, \tau$; i.e. $\overline{N}_{\tau}^{r} = n_1^r + \ldots + n_{\tau}^r$;
- N_{τ}^{e} le nombre de patients urgents présents en période τ. Cette valeur est une fonction des variables aléatoires iid $\{n_1^e, \ldots, n_{\tau}^e\}$ et des durées de séjour des patients concernés.

Le problème de choix des dates d'admission des patients réguliers peut être, comme annoncé, formulé comme un programme stochastique dynamique cherchant à minimiser le critère suivant :

$$F = \min_{\pi^1,\ldots,\pi^H} E\left[\sum_{\tau=1}^{H}\left(\sum_{i=1}^{\overline{N}_{\tau}^{r}} A_i^{\tau} f_{\tau}(X_i^{\tau-1}, X_i^{\tau}) + \beta\left(N_{\tau}^{e} - T + \sum_{i=1}^{\overline{N}_{\tau}^{r}} A_i^{\tau} \sum_{\tau'=\tau-p_i+1}^{\tau} X_{i\tau'}^{\tau}\right)^{+}\right)\right] \tag{3.3}$$

Où $(y)^+ = \max\{0, y\}$ et $\pi^1, ..., \pi^H$ sont les politiques de décisions d'admissions des différentes périodes. Elles dépendent du planning précédent, des variables stochastiques précédemment citées et des données de chaque patient connu. La politique de décision d'une période τ est définie comme suit :

$$X^{\tau} = \pi^{\tau}\left(X^{\tau-1}, \overline{N}_{\tau}^{r}, N_{\tau}^{e}, (r_i, p_i, d_i; i \in [1, ..., \overline{N}_{\tau}^{r} + N_{\tau}^{e}])\right) \tag{3.4}$$

Sous contraintes

$$X_i^{\tau} = X_i^{\tau-1} \quad \forall i / A_i^{\tau} = 1 \tag{3.5}$$

$$X_i^{\tau} = 0 \quad \forall t \leq r_i \vee \forall t / t \equiv week - end \vee \forall t > r_i + d_i \tag{3.6}$$

La fonction objectif (3.3) est composée de 2 termes. Le premier représente le coût total du planning. f_{τ} est la pénalité de planification d'un nouveau patient en période τ ou de replanification d'un patient déjà programmé. Cette fonction dépend des décisions qui ont été prises lors de la précédente période $X_i^{\tau-1}$ et de la décision courante X_i^{τ}. Plus précisément :

$$f_{\tau}(X_i^{\tau-1}, X_i^{\tau}) = \begin{cases} 0 & \text{si } X_i^{\tau-1} = X_i^{\tau} \\ \overline{f}(a_i^{\tau}, a_i^{\tau} - \tau) & \text{si } r_i = \tau \\ \overline{f}(a_i^{\tau}, a_i^{\tau} - r_i) - \overline{f}(a_i^{\tau-1}, a_i^{\tau-1} - r_i) + \delta & \text{si } r_i < \tau \end{cases} \tag{3.7}$$

où a_i^τ est la date d'admission prévue pour le patient i dans le planning de la période τ ; $a_i^\tau = \sum_t t * X_{it}^\tau, \forall i$. $\overline{f}(x, y)$ est la pénalité de planification d'un patient en période x, y période après qu'il soit connu. δ est une pénalité supplémentaire liée à la replanification d'un patient.

Le deuxième terme de la fonction objectif évalue la pénalité pour l'utilisation de surcapacité en lit pour chaque période. β est un paramètre de pondération entre les deux termes.

3.2.2 Évaluation des politiques de décisions

L'ensemble $\{\pi^1, ..., \pi^H\}$ définit la politique de contrôle du problème d'admission des patients électifs. Cette sous-section va décrire un algorithme permettant l'évaluation du planning résultant (Algorithme 3.1). En effectuant plusieurs réplications de cet algorithme, nous pouvons alors donner une mesure de la performance de n'importe quelle politique de contrôle du problème.

Algorithme 3.1 : Evaluation d'une politique d'admission

```
for τ ← 1 to H do
    foreach Patient i tel que X_{iτ}^{τ-1} = 1 do
        Admettre le patient i;
        A_i^{τ'} ← 1, ∀τ' ≥ τ;
    end
    N̄_τ^r ← N̄_{τ-1}^r + n_τ^r;
    N_τ^e ← card({i ∈ {1, ..., N_{τ-1}^e}/r_i + p_i > τ}) + n_τ^e;
    X^τ ← π^τ(X^{τ-1}, N̄_τ^r, N_τ^e, (r_i, p_i, d_i; i ∈ [1, ..., N̄_τ^r]));
end
```

L'optimisation globale de toutes les politiques de contrôle sur l'horizon entier est très compliquée du fait de la très grande taille de l'espace des solutions. Pour cette raison, nous n'allons pas résoudre directement le problème que nous venons d'exposer. A la place, nous allons nous intéresser aux choix des politiques de décisions de chaque période en définissant un problème de planification journalier.

3.3 Le problème de planification journalier

Le problème journalier de planification des admissions va, compte tenu des informations disponibles un jour donné, calculer un nouveau planning des admissions. On peut supposer, sans perte de généralité, que ce jour est le jour 0. Le but recherché ici est de planifier les patients électifs nouvellement connus durant la journée écoulée et éventuellement de replanifier les patients programmés non-encore admis. Pour ce problème, nous considérons un horizon de H' périodes.

Une fois encore, les données dont nous disposons sont la liste des patients déjà présents dans l'hôpital, la liste des patients planifiés pour être admis dans les prochaines périodes et la liste des nouveaux patients à programmer. L'évolution du système va être sujette à l'arrivée, tout du long de l'horizon de planification, de patients urgents et réguliers qui demeurent actuellement inconnus. Il est cependant primordial de tenir compte de ces futurs patients afin de garantir la qualité du planning de jour en jour. Pour le problème journalier, nous agrégeons ces paramètres stochastiques, pour les patients urgents et réguliers, dans un ensemble de variables aléatoires.

N_t va désigner le nombre de patients inconnus à l'instant de planification (l'instant 0) et qui seront présents, et donc nécessiteront un lit, à l'instant t. Cette variable représente donc le besoin en lit à l'instant t pour les futurs patients hospitalisés, qu'ils proviennent des urgences ou du flux d'entrées directes. Grâce

à elle, nous allons pouvoir anticiper sur le futur et ainsi éviter de planifier les admissions de manière myopique ou à l'aveugle. Rappelons que comme les durées de séjours peuvent être assez longues, il est important d'anticiper pour éviter les conséquences d'un mauvais choix initial. L'estimation de ce besoin futur sera détaillée un peu plus loin dans ce chapitre (Section 3.4).

C'est un truisme de dire que les patients déjà présents ne nécessitent pas de décisions sur leur date d'admission. Cependant leur présence pour un temps encore plus ou moins long va influer sur le choix concernant les patients en attente d'admission. Nous notons I l'ensemble de ces patients. La donnée T_t va représenter la capacité restante dans l'hôpital à l'instant t. Elle est calculée à partir de la capacité totale de l'établissement T diminuée du nombre de patients présents dès le départ et encore présent à l'instant t. Enfin, c_{it} va représenter le coût, ou plutôt la pénalité, associé à l'admission du patient électif i durant la période t. Ce coût est dérivé des fonctions f_τ et $\overline{f}$ (voir équation (3.7)).

Avant de passer à la modélisation mathématique du problème journalier de planification des admissions, notons que les durées de séjour des patients sont considérées dans cette étude comme déterministes. Parmi les travaux prenant en compte la nature stochastique de cette donnée, recitons (Utley *et al.*, 2005) et (Connors, 1970) qui utilisent la probabilité qu'un patient soit encore présent j jours après son admission. Ceci se justifie dans le contexte des études réalisées. Dans notre cas, plusieurs raisons nous font tendre vers le choix de durées déterministes. Tout d'abord depuis l'entrée en vigueur de la T2A, les durées de séjours tendent à se standardiser. Il donc possible d'avoir une bonne estimation de cette durée dès le début du séjour. Il demeure tout de même une incertitude relative que nous pouvons quantifier de plus ou moins quelques jours. Cependant plus l'on avance dans un séjour plus la date de sortie du patient peut être estimée précisément par le médecin. Ainsi, lors de chaque résolution du problème journalier, nous pouvons remettre à jour ces valeurs et avoir de cette manière des estimations précises des sorties des jours prochains. De plus, l'incertitude restante pour les jours plus lointains se retrouve mélangée avec les incertitudes, probabilistes, des patients inconnus. Il ne nous paraît donc pas indispensable d'ajouter une couche d'abstraction, par le biais des probabilités, pour l'estimation des durées de séjour.

3.3.1 Formulation mathématique

Dans la modélisation mathématique du problème journalier de planification des admissions, les variables de décisions considérées sont les suivantes :

- $X_{it} = \begin{cases} 1 & \text{si le patient } i \text{ est admis pendant la période } t \\ 0 & \text{sinon} \end{cases}$
- O_t : le dépassement de capacité moyen durant la période t

et la formulation mathématique correspondante :

$$(P) \qquad J^* = Minimiser\ J(X) = \sum_{i,t} c_{it} X_{it} + \beta \sum_t O_t \tag{3.8}$$

Sous les contraintes :

$$O_t = E_{N_t}\left[\left(N_t - T_t + \sum_i \sum_{t'=t-p_i+1}^{t} X_{it'}\right)^+\right] \qquad \forall t \in 1..H' \tag{3.9}$$

$$\sum_{t=1}^{d_i} X_{it} = 1 \qquad \forall i \in I \tag{3.10}$$

$$O_t \geq 0 \qquad \forall t \in 1..H' \tag{3.11}$$

$$X_{it} \in \{0,1\} \qquad \forall i \in I, \forall t \in 1..H' \tag{3.12}$$

La fonction objectif (3.8) minimise le dépassement de capacité moyen pondéré par un paramètre β et la somme des coûts liés aux affectations des patients. L'ensemble de contraintes (3.9) permet d'évaluer le dépassement de capacité moyen pour chaque période considérée. Finalement les contraintes (3.10) assurent que chaque patient est planifié une et une seule fois et que son admission intervient dans l'intervalle dans lequel il doit être admis.

Théorème 3.1 *Le problème journalier d'admissions des patients électifs* (P) *est NP-difficile au sens fort.*

Preuve: `Réduction polynomiale du problème de bin-packing`
Ce théorème est prouvé par réduction polynomiale du problème de bin-packing connu pour être NP-difficile au sens fort. Garey et Johnson (1979) définissent le problème de bin-packing comme suit : Etant donné un ensemble fini U d'items, une taille $s(u) \in \mathbb{Z}^+$ pour chaque $u \in U$, une capacité entière positive des bins (paniers) B et un entier positif K. Existe-t-il une partition de U en ensembles disjoints $U_1, U_2, ..., U_K$ tel que la somme des tailles des items dans chaque sous-ensemble U_i est B ou moins ?

Pour prouver la complexité du problème stochastique (P), nous définissions une transformation polynomiale du problème de bin-packing en un problème de décision : Existe-t-il un planning X^* tel que $J(X^*) = 0$?

A chaque instance du problème de bin-packing, nous définissions une instance I_1 du problème d'admission, où chaque patient est représenté par un item u, de la manière suivante :

- $H' = B$
- $T_t = K, \forall t$
- $N_t = 0, \forall t$
- $\beta = 1$
- $I = U$
- $p_u = s(u)$
- $c_{ut} = \begin{cases} 1 & \text{, si } t + p_u - 1 > B \\ 0 & \text{, sinon} \end{cases}$

Nous allons à présent montrer que pour le problème de planification I_1, une solution X^* tel que $J(X^*) = 0$ existe si et seulement si le problème de bin-packing a une solution.

Condition suffisante : Supposons que le problème de bin-packing a une solution $U_1, ..., U_K$. Considérons alors que chaque sous-ensemble U_j contient des patients planifiés dans le lit j. Le i^{ime} patient du sous-ensemble U_j, u_{ij}, est planifié pour être admis à la période $t = \sum_{i'=0}^{i-1} s(u_{i'j})$. Il est alors aisé de voir que le coût d'une telle solution X est égal à 0 puisqu'il n'y a aucun dépassement de capacité (comme on considère K bins alors au maximum K patients seront présents en même temps) et aucun coût de planification –ou d'heures supplémentaires – (pour chaque panier j, $\sum_i s(u_{ij}) \leq B$, l'horizon de planification).

Condition nécessaire : A présent, supposons qu'il existe un planning réalisable X^* tel que $J(X^*) = 0$. Nous construisons une solution au problème de bin-packing à l'aide de l'algorithme (3.2) polynomial qui suit en faisant un parallèle entre le remplissage des bins et un diagramme de Gantt représentant les séjours des patients dans les ressources lits :

Remarque : Comme $J(X^*) = 0$, il n'y a pas de dépassement de capacité dans le planning, donc il y a toujours assez de bins disponibles à chaque instant. De la même manière comme on ne dépasse pas l'horizon de planification, la capacité des bins est respectée.

□

3.3.2 Évaluation de la fonction objectif

Cette sous-section va nous fournir les moyens d'évaluer un planning X, solution du problème quotidien précédemment formulé. La valeur du critère (3.8) est conditionnée par une variable stochastique. Le

Algorithme 3.2 : Construction d'une solution pour le problème de bin-packing

```
1  Soit L une liste triée par ordre croissant de toutes les dates d'admission et de départ;
   // en cas d'égalité, la priorité est donnée aux dates de départ
2  Soit busy une liste des bins occupé à un instant;
3  Soit avail une liste des bins disponibles à un instant;
4  busy ← ∅;
5  avail ← tous les bins;
6  foreach d ∈ L do
7      if d est une date d'admission then
8          bin ← extraire le premier bin de avail;
9          Ajouter l'item correspondant à d dans bin;
10         Ajouter bin à busy;
11     else
12         bin ← extraire le bin contenant l'item correspondant à d de busy;
13         Mettre bin dans avail;
14     end
15 end
```

critère est sujet à l'évaluation de l'espérance de la variable aléatoire N_t. Comme nous ne connaissons pas l'expression analytique permettant ce calcul, nous allons l'estimer en utilisant la simulation de Monte Carlo.

Pour faciliter la lecture des modèles dans la suite de ce chapitre, nous noterons par PP_t le nombre de patients présents à l'instant t :

$$PP_t = \sum_i \sum_{t'=t-p_i+1}^{t} X_{it'}$$

Pour chaque période de l'horizon de planification, nous effectuons L observations des capacités requises pour les patients inconnus : $N_{t1}, N_{t2}, ..., N_{tL}$. Ensuite, pour chaque période, le dépassement est évalué comme étant une moyenne empirique des observations effectuées :

$$O_t \approx O_{tL} = \frac{1}{L} \sum_{l=1}^{L} \Big(PP_t + N_{tl} - T_t \Big)^+$$

Nous utilisons alors O_{tL} au lieu de O_t pour évaluer le critère $J(X)$. Notons au passage que par la loi des grands nombres, O_{tL} converge avec une probabilité de 1 vers O_t quand L augmente. Par conséquent, avec pour une grande valeur de L, nous sommes en mesure de donner une évaluation précise du coût du planning calculé.

3.4 Estimation des besoins pour les futurs patients

De la qualité de l'estimation du nombre N_t de patients inconnus présent à un instant t va dépendre la qualité du planning généré quotidiennement. Sous le vocable de « patients inconnus » nous incluons les patients urgents qui doivent être admis dans la période où ils sont connus et les patients électifs pour qui la date d'admission doit être décidée. Cette section va présenter une méthode permettant de caractériser la distribution des variables aléatoires N_t. Comme nous venons de le mentionner, N_t inclut la somme de deux composantes :

- N_t^e le nombre de patients urgents présents dans l'hôpital à l'instant t.

- N_t^r le nombre de patients réguliers présents en période t.

Dans les deux cas, la génération d'une série $\{N_t^\bullet\}_{t=1}^{H'}$, un échantillon iid, suit le même algorithme de base.

La manière dont nous estimons la composante aléatoire $N_t^\bullet$ est résumée dans l'algorithme (3.3). La supposition est faite que le processus stochastique d'arrivées et la distribution des durées de séjours pour les patients réguliers et urgents sont connus. Dans la pratique, ces données peuvent être déduites de l'historique de données de l'hôpital. L'exécution de cet algorithme permet d'obtenir un échantillon de valeurs sur l'horizon tout entier. En le répétant autant de fois que nécessaire, on obtient alors le nombre d'échantillons requis ; par exemple on répète l'algorithme L fois afin d'évaluer le planning journalier (voir section précédente).

Algorithme 3.3 : Génération d'un échantillon

```
1  for t ← 1 to H' do
2  |  A ← générer le nombre de nouveaux patients pour la période t;
3  |  foreach nouveau patient i do
4  |  |  r_i ← t;
5  |  |  a_i ← t;
6  |  |  p_i ← générer la durée de séjour;
7  |  end
8  |  Faire partir les patients sortant en période t;
9  |  N_t^• ← Nombre de patients présents;
10 end
```

Cet algorithme, dans sa forme énoncée, convient sans modification à la génération de la composante des patients urgents (i.e. $N_t^\bullet = N_t^e$). En effet les patients sont admis dans la période où ils sont connus ($r_i = a_i$). Cela ne fonctionne pas pour les patients réguliers qui sont admis après leur arrivée soit le lendemain, soit le surlendemain, soit ... . Le problème est ici de représenter la politique d'admission des patients électifs une fois qu'ils sont connus. A cette fin, nous définissons pour chaque jour une clé $\gamma = \{\gamma_1, \gamma_2, ..., \gamma_j\}$. Chaque nouveau patient connu en période t sera admis en période $t+1$ avec une probabilité γ_1, en période $t+2$ avec une probabilité γ_2, ... Remarquons que le choix de cette clé, la meilleure possible, est en lui même un problème d'optimisation relativement difficile. Le chapitre suivant y sera consacré.

Avant de conclure sur cet algorithme nous permettant d'estimer les besoins futurs en terme de lit, détaillons un peu la procédure de sortie des patients en fin de séjour. En fait, pour des raisons médicales, aucun départ n'a lieu durant le week-end[5]. Dans la pratique les sorties tombant le week-end sont soit avancées soit retardées en fonction de l'état de santé du patient. Plus précisément, pour un patient quittant "normalement" l'établissement un samedi, il y a une bonne chance pour que sa sortie soit avancée de 1 jour et qu'il parte vendredi. A contrario, la probabilité de devoir attendre encore 2 jours et donc sortir le lundi suivant est moindre. Un raisonnement similaire peut être mené pour les patients dont la sortie prévue tombe un dimanche.

Dans notre étude nous supposerons donc l'hypothèse qui suit. Si la date de départ $(a_i + p_i)$ d'un patient tombe un samedi (resp. un dimanche) alors sa sortie sera avancée (resp. retardée) de 1 jour avec une probabilité ρ. La sortie sera retardée (resp. avancée) de 2 jours avec une probabilité de $1 - \rho$. Afin d'être relativement proche de la réalité, nous estimons que ρ doit être de l'ordre de $0,8$. cela modélise

5. Rappelons qu'une sortie ne peut être prononcée que par un médecin. Les médecins titulaires ne travaillant pas le week-end, il s'en suit un blocage dans le processus de sortie. De même, pour des raisons similaires, aucune admission n'a lieu le week-end.

le fait qu'un patient guérissant un samedi à plus de chance de partir vendredi qu'un patient guérissant dimanche.

3.5 Différentes approches de résolution

Voyons à présent trois approches différentes pour résoudre le problème stochastique journalier de planification des admissions des patients électifs. Chacune de ces approches se base sur le modèle précédemment exposé (section 3.3) et tire différemment profit de l'estimation que nous faisons des variables aléatoires N_t. Le but recherché dans chaque approche est toujours d'anticiper les futures arrivées de manière à répartir la charge de travail au mieux. La première stratégie va utiliser une estimation de l'espérance du nombre de patients inconnus présent à une période t. La deuxième approche va garantir la présence de lits pour les futurs patients en accord avec un taux de service demandé. Enfin nous utiliserons une optimisation de Monte Carlo.

3.5.1 Utilisation du nombre moyen approximé de patients inconnus

Nous allons dans cette première approche transformer le modèle stochastique (P) en un modèle déterministe (P_1) en remplaçant chaque variable aléatoire N_t par une approximation du nombre moyen de patients présents $\widehat{N_t}$. Cette estimation est directement calculée à partir des processus d'arrivée des patients et de la distribution des durées de séjour.

Cette approche suppose donc que le cas moyen représente bien le nombre de patients à admettre dans le futur. Autrement dit, nous supposons que la variabilité est faible, que N_t a un petit écart-type ou bien encore qu'une sorte de régime stationnaire existe dans le processus d'admissions et de présences de patients.

Plus spécifiquement, nous définissons $\widehat{N_t}$ comme :

$$\widehat{N_t} = \begin{cases} \lambda t & \text{si } t \leq \overline{LoS} \\ \lambda * \overline{LoS} & \text{sinon} \end{cases}$$

où λ est le taux d'arrivée moyen des patients, urgents et électifs, soit $\lambda = E[n_1^r] + E[n_1^e]$. $\overline{LoS}$ est la durée moyenne de séjour avec $\overline{LoS} = (E[n_1^r]E[p_1^r] + E[n_1^e]E[p_1^e])/(E[n_1^r] + E[n_1^e])$.

Avec ces éléments, le modèle de planification devient le programme linéaire en nombre entier suivant :

$$(P_1) \quad Minimiser\ J_E(X) = \sum_{i,t} c_{it} X_{it} + \beta \sum_t O_t \tag{3.13}$$

Sous contraintes :

$$O_t \geq PP_t + \widehat{N_t} - T_t \qquad \forall t \in 1..H' \tag{3.14}$$

$$\sum_{t=r_i}^{d_i} X_{it} = 1 \qquad \forall i \in I \tag{3.15}$$

$$O_t \geq 0 \qquad \forall t \in 1..H' \tag{3.16}$$

$$X_{it} \in \{0,1\} \qquad \forall i \in I, \forall t \in 1..H' \tag{3.17}$$

Les contraintes (3.14) déterminent le besoin moyen de surcapacité en période t, en accord avec l'assomption que le nombre moyen de patients inconnus à l'instant 0 et présents en période t reste proche de sa moyenne.

3.5.2 Stratégie basée sur des taux de service

La seconde stratégie consiste à utiliser un taux de service permettant différent niveau de protection contre la variabilité des besoins en lit pour les patients futurs. Un paramètre α, le taux de service, est donc introduit afin de choisir ce niveau de protection.

L'approche de la section 3.4 est utilisée pour estimer la distribution de probabilité des variables aléatoires N_t. En générant suffisamment d'échantillons, nous sommes à même de calculer la probabilité $P(N_t = n)$ d'avoir n patients inconnus présents en période t. Cette distribution est ensuite utilisée pour déterminer le nombre $N_{\alpha t}$ de lits à réserver en période t afin de couvrir les besoins en lit pour les patients inconnus avec une probabilité α. Plus précisément, $N_{\alpha t}$ est calculé comme suit :

$$N_{\alpha t} = \min_{N} \left\{ \sum_{n=0}^{N} P(N_t = n) \geq \alpha \right\}$$

Le choix de la valeur du paramètre α n'est pas aisé et doit être fait en toute connaissance de cause. Un taux de service de 100% impliquerait que l'on souhaite prendre en compte toutes les possibilités. Cela n'a pas de sens car les cas les plus extrêmes sont également souvent les plus rares. A l'opposé, un taux de service de 50% reviendrait à se rapprocher de la stratégie précédente et considérer un cas médian.

La stratégie de planification journalière des admissions avec taux de service est déterminée en résolvant le programme linéaire (P_2) suivant :

$$(P_2) \quad Minimiser\ J_\alpha(X) = \sum_{i,t} c_{it} X_{it} + \beta \sum_{t} O_t \tag{3.18}$$

Sous les contraintes :

$$O_t \geq PP_t + N_{\alpha t} - T_t \qquad \forall t \in 1..H' \tag{3.19}$$

$$\sum_{t=r_i}^{d_i} X_{it} = 1 \qquad \forall i \in I \tag{3.20}$$

$$O_t \geq 0 \qquad \forall t \in 1..H' \tag{3.21}$$

$$X_{it} \in \{0, 1\} \qquad \forall i \in I, \forall t \in 1..H' \tag{3.22}$$

La seule différence concerne les contraintes (3.19) où l'estimation du dépassement de capacité est faite en fonction du taux de service que l'on se fixe.

3.5.3 Approche par optimisation de Monte Carlo

Finalement, la troisième stratégie va utiliser une approche de Monte Carlo pour résoudre le problème journalier d'admission des patients réguliers. L'idée de base de cette méthode est l'utilisation de différentes observations de la variable N_t pour pallier au fait nous ne connaissons pas son expression analytique. K différents scénarios de besoins en lit pour les patients inconnus sont générés à partir de l'algorithme présenté en section 3.4. Pour chaque période, nous obtenons donc une suite $N_{t1}, ..., N_{tK}$ où chaque terme N_{tk} est indépendant des autres. Le dépassement de capacité requis pour chaque période est ensuite supposé être la moyenne des besoins de chaque période.

L'esprit de cette méthode est donc le suivant : plus on fait d'observations plus on a de chance de couvrir les différents cas pouvant survenir en respectant le fait que certaines possibilités sont rares ; les cas les plus fréquents se retrouvent plus fréquemment dans les observations à l'inverse de cas limites.

Pour sa modélisation, cette stratégie nécessite une famille de variables de décisions supplémentaire. Soit o_{tk} le dépassement de capacité pour la période t observé dans l'échantillon k. Le problème de

planification se réécrit une fois de plus sous la forme d'un programme linéaire noté (P_3) :

$$(P_3)\ \ Minimiser\ J_K(X) = \sum_{i,t} c_{it}X_{it} + \frac{\beta}{K}\sum_{t,k} o_{tk} \tag{3.23}$$

Sous contraintes :

$$o_{tk} \geq PP_t + N_{tk} - T_t \qquad \forall t \in 1..H', \forall k \in K \tag{3.24}$$

$$\sum_{t=r_i}^{d_i} X_{it} = 1 \qquad \forall i \in I \tag{3.25}$$

$$o_{tk} \geq 0 \qquad \forall t \in 1..H' \tag{3.26}$$

$$X_{it} \in \{0,1\} \qquad \forall i \in I, \forall t \in 1..H' \tag{3.27}$$

L'ensemble de contraintes (3.24) permet de calculer le dépassement de capacité pour chaque période de chaque échantillon. Ensuite, dans le second terme de la fonction objectif, nous calculons le dépassement de capacité moyen pour chaque période, qui était auparavant désigné par O_t.

Théorème 3.2 *Soit X_K^* le planning solution optimale du problème d'optimisation de Monte Carlo (P_3). Avec une probabilité de 1, quand K tend vers l'infini, $J_K(X_K^*)$ converge vers le coût J^* de la solution optimale du problème de planification journalier (P) et X_K^* converge vers une solution optimale X^* du problème (P).*

Preuve: `Convergence de l'optimisation de Monte Carlo`

Nous allons donc prouver qu'une solution optimale du problème (P_3) converge vers une solution optimale du problème (P) quand K augmente. Dans cette optique nous allons commencer par prouver que le coût estimé d'une solution quelconque $J_K(X)$ converge vers le coût exact de la solution, $J(X)$. Tout d'abord considérons la fonction objectif réécrite sous la forme suivante :

$$J(X) = E\left[G(X,N)\right] \tag{3.28}$$

Où

$$G(X,N) = \sum_{it} c_{it}X_{it} + \beta\sum_{t}\left(N_t - T_t + \sum_{i}\sum_{\tau=t-p_i+1}^{t} X_{i\tau}\right)$$

Comme $N = \{N_1, ..., N_H\}$ est un vecteur de variables aléatoires, alors $G(X,N)$ est également une variable aléatoire. De plus, comme les scénarios $N^k = \{N_{1k}, ..., N_{Hk}\}, \forall k \in \{1, ..., K\}$ sont des observations i.i.d du vecteur N, les $G(X,N^k), \forall k$ sont également des observations aléatoires i.i.d de $G(X,N)$. Le critère du problème (P_3) s'écrit alors dans ce cas :

$$J_K(X) = \frac{1}{K}\sum_{k} G(X,N^k) \tag{3.29}$$

A présent, soit X une solution réalisable de (P). D'après la loi forte des grands nombres, $1/K\sum_k G(X,N^k)$ converge presque sûrement vers $E[G(X,N)]$ quand K augmente. Donc, pour n'importe quel $\epsilon \geq 0$ donné, il existe un entier $\overline{K_X} > 0$ tel que

$$|J_K(X) - J(X)| = |\beta|\left|\frac{1}{K}\sum_{k} G(X,N^k) - E[G(X,N)]\right| \quad < \epsilon \quad ,\forall K > \overline{K_X} \tag{3.30}$$

Soit $\overline{K} = \max_X \overline{K_X}$. Sans perte de généralité, nous pouvons réécrire, pour n'importe quelle solution X, l'équation précédente comme suit :

$$|J_K(X) - J(X)| < \epsilon, \forall K > \overline{K} \tag{3.31}$$

Cette formulation demeure vraie pour le cas particulier de X_k^* une solution optimale du problème (P_3),

$$|J_K(X_K^*) - J(X_K^*)| \leq \epsilon, \forall K > \overline{K}$$

De plus, par définition de X_K^* et de X^* il s'ensuit que pour tout $K > \overline{K}$

$$\begin{aligned} |J_K(X_K^*) - J_K(X^*)| &= J_K(X^*) - J_K(X_K^*) \\ &\leq J_K(X^*) - J_K(X_K^*) + J(X_K^*) - J(X^*) \\ &\leq |-J_K(X_K^*) + J(X_K^*)| + |J_K(X^*) - J(X^*)| \\ &\leq 2\epsilon \end{aligned} \tag{3.32}$$

A partir des équations (3.31) et (3.32), il vient que

$$\begin{aligned} |J_K(X_K^*) - J(X^*)| &= |J_K(X_K^*) - J_K(X^*) + J_K(X^*) - J(X^*)| \\ &\leq |J_K(X_K^*) - J_K(X^*)| + |J_K(X^*) - J(X^*)| \\ &\leq 3\epsilon \end{aligned}$$

comme en plus nous avons

$$\begin{aligned} |J(X_K^*) - J(X^*)| &= J(X_K^*) - J(X^*) \\ &\leq J(X_K^*) - J(X^*) - J_K(X_K^*) + J_K(X^*) \\ &\leq |J(X_K^*) - J_K(X_K^*)| + |-J(X^*) + J_K(X^*)| \\ &\leq 2\epsilon \end{aligned}$$

nous pouvons établir les deux résultats suivants : pour tout $\epsilon > 0$ il existe un entier fini $\overline{K}$ tel que

$$|J_K(X_K^*) - J(X^*)| \leq 3\epsilon \quad \forall K > \overline{K} \tag{3.33}$$

$$|J(X_K^*) - J(X^*)| \leq 2\epsilon \quad \forall K > \overline{K} \tag{3.34}$$

L'inégalité (3.33) implique que quand K augmente, le coût estimé $J_K(X_K^*)$ d'une solution optimale de (P_3) converge presque sûrement vers le coût exact $J(X_K^*)$. L'inégalité (3.34) montre que la solution optimale résultante de (P_3) converge presque sûrement vers une solution optimale X^* du problème (P) quand K augmente.

□

3.6 Expérimentations numériques

3.6.1 Cas d'étude et données utilisées

Pour nos expérimentations sur ce problème, nous nous basons sur le cas de l'hôpital de Firminy dont la présentation et quelques données sur son flux de patients sont données en section 1.5 et en annexe de ce mémoire. Nous choisissons de nous limiter au cas des services de médecine.

On suppose que les processus d'arrivée de patients réguliers et urgents suivent une distribution de Poisson de paramètres respectifs λ_r et λ_e. D'après les bases de données auxquelles nous avons eu accès, il apparaît que $\lambda_r = 4$ et $\lambda_e = 6$ représentent assez bien les flux d'arrivées dans le pôle de service concerné. Les distributions de durées de séjour sont estimées à l'aide du module « input analyzer » du logiciel de simulation SIMAN-ARENA. Il ressort que la distribution la plus représentative est la distribution Beta. Cependant les paramètres diffèrent quelque peu suivant que les patients soit urgents ou réguliers

					Paramètre loi Beta	
	Durée moyenne	Ecart type	Min	Max	Alpha 1	Alpha 2
Patients réguliers	9,51	7,3	1	58	1,55	8,36
Patients urgents	10,83	7,48	1	74	1,5	8,54

TABLE 3.1 – Distribution des durées de séjour en médecine à l'hôpital de Firminy

principalement à cause des durées de séjour sensiblement plus longues dans le premier cas (voir Table 3.1).

Afin de justifier numériquement l'intérêt de notre démarche, nous utilisons l'algorithme décrit en section 3.4 afin d'observer pour chaque jour la proportion de patients connus, la partie déterministe du problème, et la proportion de patients inconnus, la partie stochastique. Évidemment, plus le temps avance, plus la proportion de patients présents inconnus initialement va augmenter tandis que la proportion de patients connus va diminuer. La figure 3.1 permet d'observer ce phénomène.

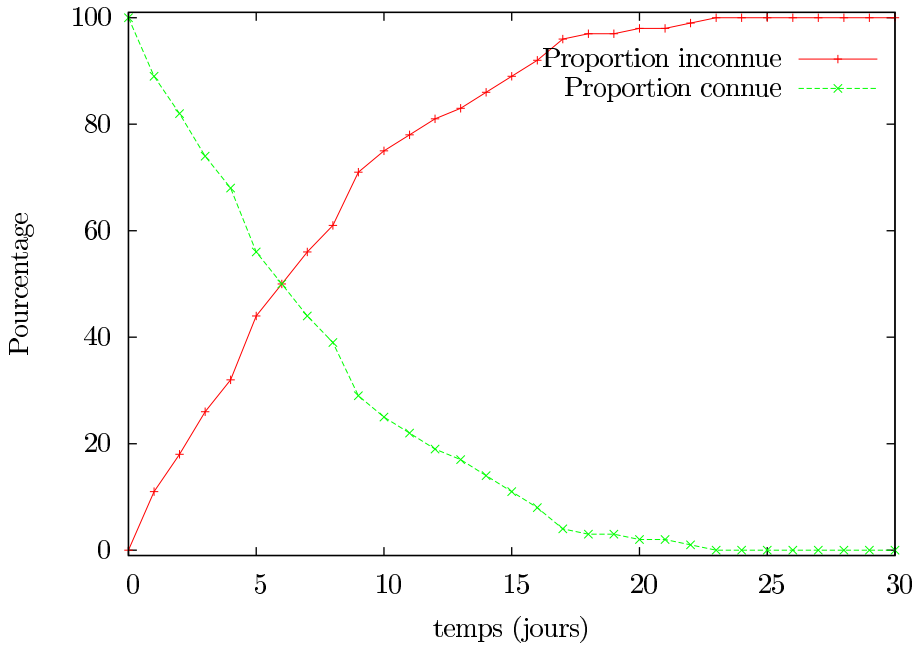

FIGURE 3.1 – Variation de la proportion de patients connus et inconnus

Il apparaît très clairement que dès le sixième jour, le nombre de patients inconnus devient plus important que le nombre de patients connus. Rappelons que la durée moyenne de séjour pour ces services est d'à peu près 10 jours. Autrement dit, à la fin d'un séjour moyen, plus de la moitié de l'effectif de patients a changé. Il est donc primordial lors de la planification d'un patient de tenir compte, comme nous le faisons, des futures arrivées.

Dans le cas considéré des services de médecine, le choix d'une période correspondant à un jour est assez naturel. L'horizon de planification H' doit être suffisamment long pour permettre à tous les patients connus d'être sortis à la fin. Sans quoi, nous ne pourrions évaluer qu'une partie de l'impact du choix d'admission d'un patient un jour t. Nous retenons $H' = 30$. Les patients réguliers se présentant à ce service étant en grande partie des patients semi-urgents, nous faisons l'hypothèse que tous les patients sont de ce type. En particulier, le délai maximum d'admission sera fixé à chaque patient à une semaine soit cinq jours ouvrés. Les pénalités d'admission des patients réguliers sont fixées de manière à favoriser

des admissions, calculées en nombre de jours ouvrés, en J+2 ou J+3. L'admission en J+1 est actuellement un délai trop court pour mettre en place la réservation des ressources, dont dépend le séjour, de manière convenable. Ce délai d'admission est donc plus pénalisé que les deux précédents. Enfin pour les autres délais d'admission, J+4 et J+5, la pénalisation est proportionnelle à la durée d'attente tout en étant supérieure à la pénalisation de J+1. La clé de répartition γ est choisie de manière à représenter ces pénalités. Nous prenons une clé où 15% des patients sont affectés en J+1, 50% en J+2 et 35% en J+3.

La capacité d'accueil nominale est fixée, comme à Firminy, à 100 lits. Le paramètre β de l'équation (3.8), représentant la pénalisation liée au dépassement de capacité, est fixé à 2,5. L'accent est ainsi mis sur la minimisation de cette partie de l'objectif. Ainsi une bonne solution dans nos expérimentations favorisera, par exemple, une admission en J+1 plutôt qu'une admission entraînant des dépassements de capacité par la suite.

En fonction de la stratégie considérée ou de si l'on souhaite évaluer la valeur du critère, le nombre d'échantillons devant être générés par l'algorithme 3.3 n'est pas du même ordre de grandeur. Si nous notons K le nombre d'échantillons nécessaires pour l'optimisation de Monte Carlo (Section 3.5.3), par W le nombre d'échantillons nécessaires à la définition de la loi de probabilité $P(N_t)$ (Section 3.5.2) et L le nombre d'échantillons pour évaluer correctement le critère (Section 3.3.2), alors on a l'ordre suivant : $K << W << L$. En effet, l'optimisation par l'approche de Monte Carlo tire profit du fait qu'un petit nombre d'échantillons doivent relativement bien représenter la réalité. La valeur de K doit donc être raisonnable de manière à ne pas générer trop de contraintes ce qui ralentirait considérablement la performance de l'optimisation. Pour l'estimation d'une loi de probabilité, le nombre d'échantillon généré doit être plus important de manière à être relativement précis. Notons que nous ne considérons qu'une estimation et donc qu'une grande précision n'est pas requise. Au contraire, l'évaluation de la fonction économique d'un planning donné par simulation de Monte Carlo se doit d'être la plus précise possible. Un très grand nombre d'échantillons doit être choisi. Compte tenu du temps nécessaire à la génération d'échantillons et des considérations précedemment mentionnée, nous choisissons empiriquement les valeurs $L = 10^6$ et $W = 10^4$. Le choix de K sera discuté un peu plus loin. Enfin, pour toutes les expérimentations qui vont suivre, les calculs seront effectués sur un PC tournant sous Linux 2.6 avec un processeur de 2.4GHz. Les modèles linéaires seront résolus à l'aide du solveur libre GLPK.

3.6.2 Choix du taux de services

La résolution du problème (P_2) décrit en section 3.5.2 utilise un paramètre α ; le taux de service. Nous allons nous intéresser à son choix de manière à ce que la méthode fournisse les meilleurs résultats possibles. Pour ce faire, 10 instances indépendantes sont générées puis résolues pour un taux de service variant de 55% à 95% par pas de 5%. Comme la valeur de l'objectif dépend de l'instance[6], nous appliquons une mise à l'échelle en se basant sur le coût obtenu pour un taux de service de 95%. La figure 3.2 présente la valeur moyenne de cet objectif mis à l'échelle pour chaque valeur de α.

Un premier résultat venant après la lecture de ce graphique concerne les valeurs élevées de α. Comme annoncé, ces taux de services fournissent un niveau de protection trop important menant à des solutions de mauvaise qualité. Deux valeurs peuvent être retenues pour le choix du paramètre α : 65% et 80%. Cependant une analyse plus détaillée des résultats montre que la dispersion des valeurs est moindre pour $\alpha = 80\%$ (écart-type= 2,53 pour 65% et 1,9 pour 80%). Nous retenons donc ce choix pour la suite de nos expérimentations.

6. Par nature certaines instances mènent obligatoirement à des coûts plus élevés que d'autres

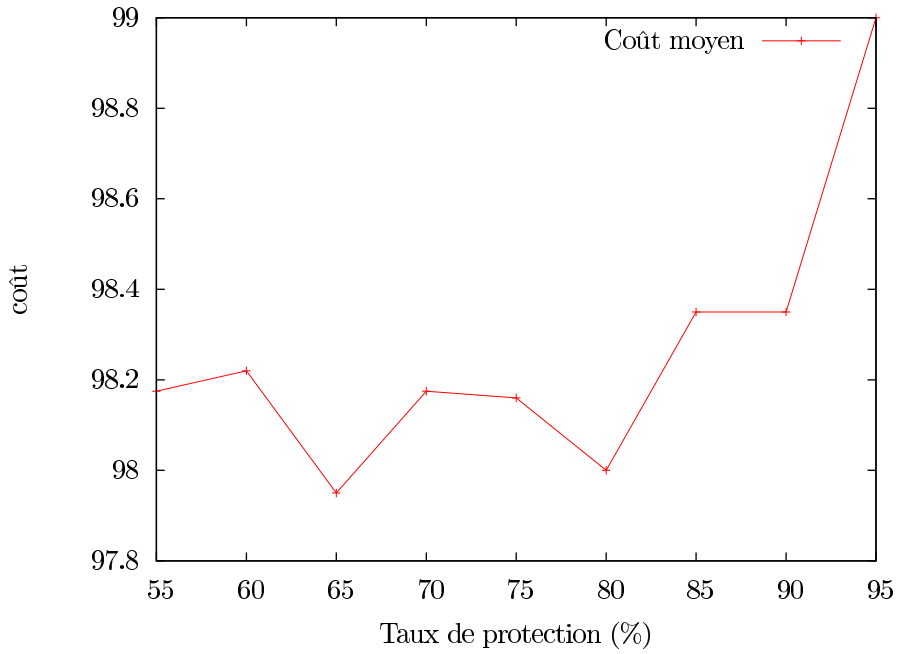

FIGURE 3.2 – Impact du choix du taux de service

3.6.3 Taille de l'échantillon pour l'optimisation de Monte Carlo

Afin de choisir la taille idéale de l'échantillon pour l'optimisation de Monte Carlo et la résolution du problème (P_3), revenons sur la convergence de la méthode. Nous avons démontré que cette méthode converge vers une solution optimale du problème journalier. Cependant, il va être intéressant de voir expérimentalement si cette convergence est atteinte rapidement. Pour ce faire, nous résolvons le problème (P_3) pour des valeurs de K variant dans $\{2; 5; 10; 20; 50; 100; 200; 500; 750; 1000\}$ et pour une instance donnée du problème journalier P. Pour chaque valeur de K nous effectuons 10 optimisations de Monte Carlo en regénérant à chaque fois l'échantillon d'observations. Les résultats de ces optimisations sont présentés table 3.2 et graphiquement figure 3.3.

TABLE 3.2 – Convergence de l'optimisation de Monte Carlo

K	Min.	Moy.	Max.	Ecart type.	CPU time(s)
2	152,35	155,01	164,03	3,58	0
5	152,35	153,67	157,12	1,63	0
10	151,46	153,15	155,13	1,07	0,1
20	151,46	152,02	153,00	0,70	0,1
50	151,46	151,96	152,06	0,44	0,1
100	151,46	151,46	151,46	0	0,3
200	151,46	151,85	152,44	0,51	2,5
500	151,46	151,56	152,44	0,31	20,1
750	151,46	151,46	151,46	0	54,9
1000	151,46	151,56	152,44	0,31	116,6

Il résulte que la méthode converge rapidement. Nous sommes donc capable d'atteindre facilement une solution optimale et pour un temps de résolution très raisonnable. Nous considérons qu'à partir de $K = 20$ les solutions sont de bonne qualité et qu'il n'est donc pas justifié de générer plus d'observations,

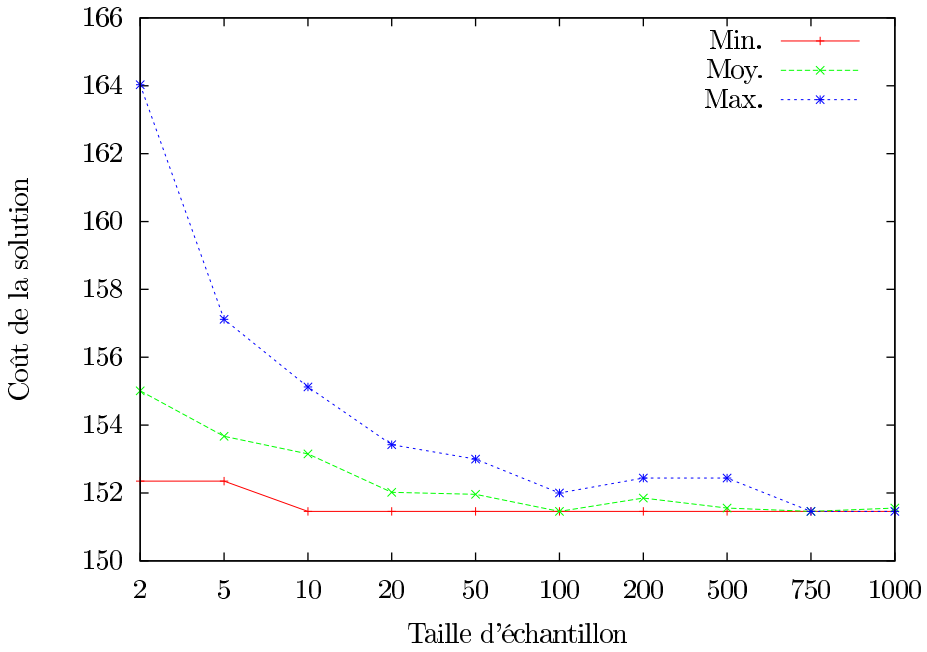

FIGURE 3.3 – Convergence de l'optimisation de Monte Carlo

ce qui de plus augmente la taille du problème et donc son temps de résolution. Pour la suite de cette étude nous conserverons donc la valeur de $K = 20$.

3.6.4 Comparaison des 3 stratégies journalières

Afin de tester et de comparer les performances des 3 stratégies journalières proposées sur différents cas de figures, nous définissons trois types d'établissement hospitalier comme suit :

- **Hôpital 1** un hôpital où le flux de patients urgents est plus important que le flux de patients réguliers ($\lambda_e > \lambda_r$)
- **Hôpital 2** un hôpital où les deux flux de patients sont équilibrés ($\lambda_e = \lambda_r$)
- **Hôpital 3** un hôpital où le flux de patients électif est le plus important ($\lambda_e < \lambda_r$)

Notons que d'après nos observations de terrains, un établissement de type 1 est un établissement similaire à celui de Firminy (dont l'activité provient principalement des urgences). Un établissement présentant des flux de type 2 se rapproche plus d'un grand établissement de type CHU. Finalement, les flux de patients d'un hôpital de type 3 représente la majorité des cliniques privées. Pour nos expérimentations, pour l'hôpital 1 représentant celui de Firminy, nous fixons les valeurs $\lambda_e = 6$ et $\lambda_r = 4$. Pour l'hôpital de type 2, nous fixons $\lambda_e = \lambda_r = 5$; pour l'hôpital de type 3 $\lambda_e = 4$ et $\lambda_r = 7$. Le nombre moyen de nouvelles arrivées par jour, $\lambda_e + \lambda_r$, est supérieur dans l'hôpital de type 3. En réalité, ce nombre moyen n'est valable que pour les jours de la semaine. λ_r étant nulle le week end, il a été nécessaire d'augmenter le flux des entrées programmées afin de conserver un hôpital dont le taux de remplissage reste élevé. Finalement, notons que pour les deux derniers types d'hôpitaux, la convergence rapide de l'optimisation de Monte Carlo est toujours vraie (Fig. 3.4).

Pour chaque classe d'établissement, 2 ensembles de 10 instances sont générés. Le premier ensemble est construit de manière à représenter des problèmes de planification du lundi tandis que le deuxième ensemble représente des problèmes du mercredi. Même si nous avons testé les méthodes sur tous les jours

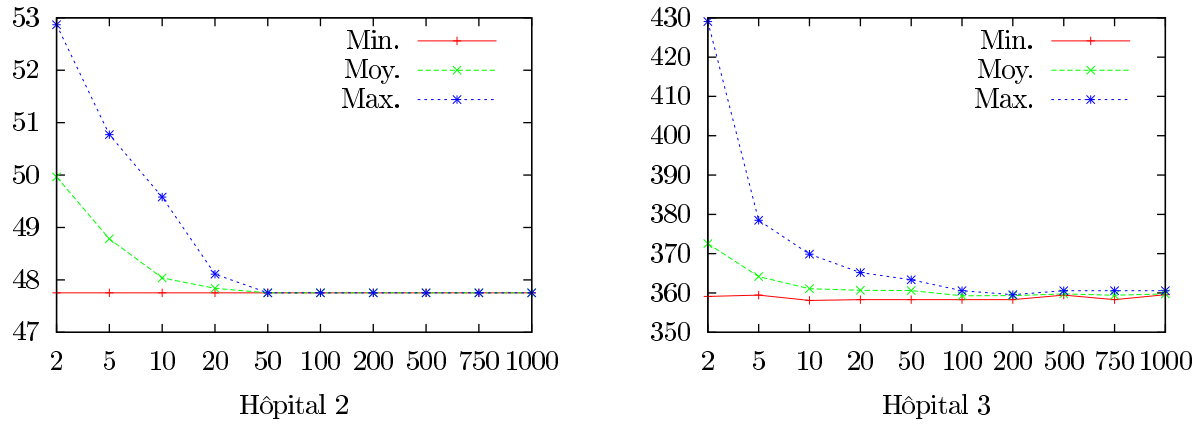

FIGURE 3.4 – Converge de l'optimisation de Monte Carlo pour les hôpitaux 2 et 3

de la semaine, nous ne présentons que les résultats de ces 2 jours. Le problème de planification du lundi est réputé dans la pratique pour être compliqué à la suite de la pause, ou plutôt de la perturbation, du week-end. Le problème de mercredi, le milieu de semaine, est là pour mettre en avant une éventuelle complexité sur la résolution du problème de planification qui serait liée au jour de la semaine. Dans chaque ensemble le nombre de patients à planifier est le même pour chaque instance. La seule différence réside dans l'état initial de l'hôpital. Ainsi, les comparaisons peuvent avoir lieu sur des problèmes de même taille.

Afin de soutenir l'intérêt des approches proposées, nous définissons pour les besoins de comparaison une quatrième méthode répondant à un algorithme simple et myopique. En particulier pour cette méthode nous supposons que :

- Les patients programmés ne peuvent être replanifiés
- Les patients nouvellement connus sont admis à des dates aléatoires choisies dans leur fenêtre d'admissibilité
- Les futures arrivées ne sont pas prises en compte

Nous faisons le choix d'une affectation aléatoire. Cela nous donne l'occasion d'avoir une stratégie proche de la réalité car en pratique, la répartition des admissions semble être uniformément répartie malgré les quelques règles existant dans le choix de cette date (voir section 1.5). Une autre stratégie simple affectant les patients le jour où la capacité disponible est maximale aurait pu être utilisée. Cependant dans notre cas, les futures arrivées n'étant pas prise en compte, cela reviendrait presque systématiquement à affecter les patients au plus tard de leur fenêtre d'admissibilité.

Chaque planning calculé par une des quatre stratégies proposées sera évalué à l'aide de la méthode présentée en section 3.3.2 et avec les besoins en lit N_{tl} estimés par l'algorithme de la section 3.2.2. A la suite de cette évaluation, pour chaque instance, nous réalisons un classement des approches : 1 pour la meilleure, 2 pour la deuxième, ... Le tableau 3.3 présente pour chaque type d'hôpital précédemment défini et pour chaque ensemble d'instances le classement moyen obtenu par chaque méthode. De plus nous mesurons l'écart relatif moyen entre le meilleur planning trouvé et le planning de chaque méthode dans la colonne Δ. Cet écart relatif est calculé en évaluant pour chaque instance la différence entre la solution trouvée par une méthode et la meilleure solution trouvée toutes approches confondues.

D'après ces résultats, il apparaît clairement en premier lieu que l'utilisation d'une moyenne estimée du nombre de patients futurs n'est pas un bon choix pour ce problème. La méthode est trop rigide et ne

TABLE 3.3 – Classement des stratégies pour les types d'hôpitaux considérés.

Hôpital 1				
	Ensemble 1		Ensemble 2	
	Rang	$\Delta(\%)$	Rang	$\Delta(\%)$
Moy. Approx. (P_1)	2,3	2,4	2,2	1,2
Taux de service (P_2)	2,8	1,6	2,5	1,7
Monte Carlo (P_3)	1,2	0,1	1,2	0,0
Myopique	3,4	2,6	4	6,4

Hôpital 2				
	Ensemble 1		Ensemble 2	
	Rang	$\Delta(\%)$	Rang	$\Delta(\%)$
Moy. Approx. (P_1)	3,3	8,7	3,5	11,6
Taux de service (P_2)	2,4	5,5	2,3	3,6
Monte Carlo (P_3)	1,5	0,4	1,4	1,2
Myopique	2,8	5,7	2,8	6,7

Hôpital 3				
	Ensemble 1		Ensemble 2	
	Rang	$\Delta(\%)$	Rang	$\Delta(\%)$
Moy. Approx. (P_1)	3	5,6	3,6	12,6
Taux de service (P_2)	3,6	3,6	2,4	3,8
Monte Carlo (P_3)	1,1	0,0	1	0,0
Myopique	3,3	5,8	3	6,2

permet pas de prendre en compte la variation existante dans l'occupation des services. Dans certains cas, où le flux de patients réguliers à admettre devient important (hôpital 2 et 3), l'approche fournit même de moins bons résultats que la méthode myopique. L'approche par taux de service donne des résultats peu probants dans certains cas. La difficulté du choix du taux de protection peut en être la cause. Peut être aurait-il été plus judicieux de définir une méthode dont le taux de protection varierait en fonction de cas de figures rencontrés. Cependant, il aurait fallu répondre à d'autres questions auparavant qui sont : comment doivent être choisis les taux de protection ? Comment prévoir dans quels cas de figures vont se situer les périodes à venir ? Peut être aurait-il été judicieux de définir un taux pouvant varier dans le temps, i.e. un paramètre α_t. Finalement, nous retiendrons de cette comparaison que l'approche par optimisation de monte Carlo fournit les meilleurs résultats.

3.6.5 Résolution du problème global par les stratégies proposées

Il nous reste finalement à étudier le problème global de choix des dates d'admission pour les patients réguliers. La sous-section précédente vient de comparer les différentes approches envisagées dans la résolution d'un problème journalier, mais comment se comportent-elles jour après jour dans le cadre d'une utilisation réelle ? Nous allons répondre à cette question en utilisant le schéma de simulation du problème global défini dans la section 3.2.2. Nous traitons le cas particulier où la politique de décision permettant de définir un nouveau planning d'une période à une autre, π_τ, est la même pour chaque période. Plus précisément, π_τ sera l'une des quatre approches utilisées (les 3 proposées + l'approche myopique). A chaque étape de la simulation, nous collectons les coûts de planification engendrés par le nouveau planning ainsi que les dépassements de capacité constatés. A chaque instant les patients à admettre sont hospitalisés, les

autres seront replanifiés lors de la génération du planning du jour en fonction des nouvelles informations acquises au fil des itérations.

Pour chaque réplication de la simulation, l'algorithme de génération de besoins en lits de la section 3.4 est utilisé pendant une période de warm-up pour générer un planning initial. La méthode myopique est conservée comme étant le seul point nous permettant de nous comparer avec le cas réel. Notons que le risque d'avoir un patient non planifié est inexistant. Dans l'approche myopique, tous les patients ont une date d'admission, qui n'est jamais remise en cause, définie dès qu'ils sont connus. Dans toutes les autres approches même si un patient voit son admission repoussée au fur et à mesure des planning établis, la date butoir de fin de sa fenêtre d'hospitalisation sera forcément choisie lors de la planification de la période précédente même si ce choix engendre des coûts d'admission élevés.

Les pré-expérimentations menées ont montré que pour faire face aux différences survenant d'une simulation à l'autre sur une période de 2 mois, 2000 réplications de l'algorithme de simulation sont nécessaires [7]. Les indicateurs collectés présentés dans le tableau 3.4 sont les suivants :

- Nombre de jours en dépassement de capacité
- Dépassement de capacité moyen sur les jours concernés
- Coûts de planning
- Coûts de la solution d'après la fonction objectif (Equation 3.3)

TABLE 3.4 – Résultats de simulation

		Nb dépassement		Dépassement moy.		Coûts		Obj.
		Moy.	Ec. Type	Moy.	Ec. Type	Moy.	Ec. Type	Moy.
Hôpital 1	Estim. Moy. (P_1)	3,06	3,8	5,11	3,89	2,86	2,11	124,97
	Taux(P_2)	4,43	4,9	6,1	4,6	1,59	1,25	115,55
	Monte Carlo (P_3)	3,37	5,1	5,0	3,4	0,74	0,25	64,47
	Myopique	6,51	5,7	6,7	4,8	1,78	1,57	161,97
Hôpital 2	Estim. Moy. (P_1)	0,94	2,1	4,4	4,0	3,84	2,84	125,53
	Taux(P_2)	1,09	2,3	4,3	3,5	2,44	1,96	84,84
	Monte Carlo (P_3)	0,96	1,7	3.9	1,9	0,82	0,7	33,98
	Myopique	2,83	3,5	5	4,1	2,27	2,0	103,54
Hôpital 3	Estim. Moy. (P_1)	0,98	2,0	4,0	4,0	5,71	4,26	183,58
	Taux(P_2)	3,83	4,6	5,2	4,0	2,98	2,52	139,03
	Monte Carlo (P_3)	1,45	3,6	4,5	3,7	1,61	1,17	64,67
	Myopique	5,21	5,1	6,8	5,4	3,09	2,65	180,71

En guise de remarque préalable, nous pouvons mentionner que toutes les méthodes proposées résolvent le problème journalier en moins d'une seconde. Sans surprise, les résultats observés pour la résolution d'un problème journalier sont amplifiés lors de planifications successives. La valeur de l'objectif des planning générés par l'optimisation de Monte Carlo est deux fois moindre que celui des autres stratégies. De plus, on peut constater que sur tous les critères observés, cette méthode reste la meilleure. Finalement notons que chaque jour, avec cette méthode, pour l'hôpital de type 1 en moyenne 3,33 patients ont été reprogrammés, pour l'hôpital de type 2 en moyenne 4.57 patients ont été reprogrammés et pour l'hôpital de type 3, le nombre de patients reprogrammés est en moyenne de 7,02. Dans les trois cas, nous constatons que ce nombre de patients reprogrammés est proche du nombre de nouveaux patients connus (paramètre λ_e).

7. Ceci est surtout vrai pour l'hôpital 1 dont le flux de patients urgents, non maîtrisé, est important. Pour les autres types d'hôpitaux, le nombre de réplications requises est moindre.

3.7 Synthèse

Dans ce chapitre, nous venons de voir et de résoudre le problème de choix des dates d'admission des patients en tenant compte d'événements aléatoires telle l'arrivée de patients urgents, qui peuvent se produire dans le futur. En partant de la modélisation du problème sous la forme d'un modèle de programmation dynamique stochastique, nous nous sommes intéressés en détails à comment prévoir les futurs besoins de lits puis à comment résoudre le problème de planification des admissions se posant chaque jour.

La prévision d'un scénario possible se fait en déroulant un algorithme itératif prenant en entrée des informations sur le flux de patients (processus d'arrivée et distribution des durées de séjour) et une clé de répartition des patients permettant de simuler la logique de délais avant admission d'un patient.

En se basant sur une utilisation différente des prévisions de possibles scénarios futurs, trois stratégies ont été proposées. L'une d'entre elles, l'optimisation de Monte Carlo, se démarque clairement des autres de par ses performances. Cette méthode peut de plus être considérée comme presque exacte pour le problème journalier de planification des admissions de par ses propriétés de convergences.

Enfin, notons que la méthode proposée est générique et peut s'appliquer aussi bien à des services de médecine, comme nous l'avons vu lors des expérimentations, qu'à des unités de chirurgie. Notons seulement que dans ce dernier type d'unité le passage au bloc opératoire peut contraindre le choix de replanification. Plusieurs alternatives peuvent être envisagées comme, pour commencer, la redéfinition d'une période comme un créneau de bloc et non plus une journée. Il s'ensuit généralement que certains créneaux ne peuvent pas accueillir, pour diverses raisons, certains patients. Deux adaptations sont envisageables : soit la pénalité d'affectation d'un patient à une de ces périodes est augmentée à un niveau plus élevé que les autres pénalités (ce qui revient à favoriser certaines périodes), soit on interdit formellement l'admission d'un patient dans le créneau correspondant en fixant la variable de décision associée. L'idée se généralise à chaque unité de soins requérant une ressource précise en même temps que l'admission d'un patient.

Chapitre 4

Optimisation de la clé de répartition des patients

Sommaire

4.1 Présentation et enjeux
4.2 Objectifs et évaluation d'une clé de répartition des patients
4.3 Introduction aux méthodes de recherches stochastiques
4.4 Optimisation de la clé de répartition des patients par la méthode de perturbation simultanée
4.5 Optimisation de la clé de répartition de patients par l'analyse de perturbation
4.6 Comparaison des approches et performances
4.7 Conclusion

LA RECHERCHE D'UN PLANNING des dates d'admission des patients réguliers et plus spécialement la définition des politiques d'admissions journalières permettant de mettre à jour quotidiennement le planning prévisionnel des admissions a été traitée dans le chapitre précédent. Il a notamment été établi comme résultat principal qu'une optimisation de Monte Carlo couplée à un algorithme de génération de scénarios fournissaient les meilleurs résultats. Cependant, l'algorithme de génération de scénarios utilise une donnée non triviale : une clé de répartition des patients électifs. Rappelons pour mémoire que pour les patients urgents, le choix de la date d'admission est sans alternatives possibles. Cela n'est pas le cas pour les patients électifs qui, une fois qu'ils sont connus, seront admis au fil des jours à venir en fonction de la politique d'admission des patients réguliers. La clé de répartition a pour but de simuler cette politique. Ce chapitre va s'intéresser à la recherche de la meilleure clé possible. Cette recherche se fera par deux méthodes dites de gradient : la perturbation simultanée et l'analyse de perturbation.

4.1 Présentation et enjeux

La clé de répartition des patients électifs joue un rôle important dans la qualité du planning des admissions généré quotidiennement puisque grâce à elle, les besoins futurs en lits sont estimés. Le problème de recherche de la meilleure clé peut se formuler comme étant la recherche de la clé permettant à la politique d'admission journalière de fournir les meilleurs planning possibles. D'un point de vue journalier

strict, il s'agirait donc de rechercher la clé qui pour un problème journalier donné permettrait d'obtenir le planning avec la plus petite valeur de la fonction objectif. Cependant, le rôle de la clé étant de représenter les futures admissions telles qu'elles seront décidées ultérieurement, une deuxième formulation du problème de recherche de la clé peut être faite : la recherche de la clé représentant le mieux la politique de décisions des admissions. Avec ces deux formulations, il apparaît que (i) la construction du planning des admissions dépend de la clé et que (ii) la clé dépend de la manière dont le planning est construit. Ce phénomène d'interdépendance peut sembler problématique au premier abord. Fort heureusement, ces deux formulations du problème vont dans le même sens et conduisent à rechercher une unique clé : celle qui permet d'obtenir les meilleurs planning au fil du temps et qui donc représente bien les futurs choix d'admissions.

On peut voir dans la recherche de la clé de répartition un autre intérêt que celui de permettre à la méthode d'optimisation journalière de fournir des performances optimales. En se basant sur le fait que, implicitement, la clé va représenter la politique d'admission induite par la méthode de planification des admissions retenues, une utilisation de cette clé comme une méthode de planification "optimisée" en elle même peut être faite. En effet, la clé fournit un précieux indicateur sur comment répartir les patients nouvellement connus dans les jours suivants en fonction du flux (urgent et électif) de l'hôpital considéré. Son utilisation par un cadre de santé cherchant à planifier les futures admissions se révèle donc judicieuse puisque la clé représente en substance quelle proportion de patients admettre pour chaque jour à venir. Bien entendu, la clé proposée se doit d'être suffisamment générique pour être la meilleure en moyenne parmi tous les cas pouvant se présenter. Il convient donc d'adapter ses décisions en fonction des cas particuliers du jour. De plus, aucune indication spécifique n'est donnée sur la répartition des séjours longs et séjours courts ; ce choix étant fait par la méthode d'optimisation des admissions. Celui qui souhaite utiliser directement la clé de répartition des patients réguliers doit s'équiper d'un ensemble de règles de gestion complémentaires afin de faire face aux ajustements particuliers devant être effectués.

D'après la définition de la section 3.4, la clé de répartition est un vecteur de probabilité définissant pour chaque jour de la semaine comment sont répartis les patients au cours des n jours suivants :

$$\gamma_j =^t \{\gamma_{j1}, \gamma_{j2}, ...\gamma_{jn}\}, \forall j \in \{lundi, mardi, ...\} \tag{4.1}$$

En particulier, chaque nouveau patient électif connu un jour j est admis i jours après où :

$$i = \arg\max_d \{\sum_{t=1}^{d} \gamma_{jt} \leq p\} \tag{4.2}$$

avec p un nombre aléatoire uniformément tiré entre 0 et 1.
Plus formellement, considérons les éléments suivants :

- $X(t)$ L'état du système à l'instant t (i.e. un planning d'admission des patients)
- γ la clé de répartition (paramètre du système)
- ω les variables aléatoires des évènements pouvant survenir (i.e. les nombres aléatoires dans une simulation)

Ce triplet définit ce que l'on peut appeler un scénario où la suite des états du système va évoluer en fonction des paramètres donnés en entrée et des évènements aléatoires survenant au fil du temps. Soit $L(X(t), \gamma, \omega)$ la mesure de performance d'un scénario. Cette mesure, comme dans le chapitre précédent, afin de représenter ce qui peut être considéré comme un bon planning, prend en compte les coûts de planning et les dépassements de capacité. Comme la série des $X(t)$ dépend des deux données γ et ω, utilisées dans la méthode d'optimisation des planning d'admissions, nous pouvons alors nous ramener à l'écriture suivante de la mesure de performance :

$$L(\gamma, \omega) \equiv L(X(t), \gamma, \omega) \tag{4.3}$$

Enfin, soit la fonction objectif mesurant l'espérance de la performance du système pour le couple (γ, ω) :

$$J(\gamma) = E\left[L(\gamma, \omega)\right] \tag{4.4}$$

A l'aide de ces éléments, nous pouvons définir le problème d'optimisation de la clé de répartition des patients électifs. Il s'agit de trouver la clé γ^* tel que :

$$\gamma^* = \arg\min_{\gamma}\{J(\gamma)\} \tag{4.5}$$

Sous les contraintes

$$\sum_{t=1}^{n} \gamma_{jt} = 1 \qquad \forall j \tag{4.6}$$

$$0 \leq \gamma_{jt} \leq 1 \qquad \forall j, \forall t \tag{4.7}$$

Ces deux ensembles de contraintes (4.6) et (4.7) définissent l'espace des clés de répartition, la clé d'un jour donné étant un vecteur de probabilité. Cette espace sera noté (Γ) dans la suite de ce chapitre.

4.2 Objectifs et évaluation d'une clé de répartition des patients

La mesure de performance d'une clé de répartition, la fonction $L(\gamma, \omega)$, peut être envisagée de deux manières suivant l'utilisation que l'on souhaite faire de cette clé. Ces deux utilisations de la clé peuvent se voir en fonction de la durée de validité que nous pouvons accorder à la clé de répartition. Ainsi, nous pouvons distinguer la recherche d'une meilleure clé journalière utilisée pour une instance particulière de la recherche d'une meilleure clé globale utilisée indifféremment de l'instance à optimiser.

4.2.1 Optimisation journalière de la clé de répartition

Dans la version journalière du problème de recherche de la meilleure clé de répartition des patients, nous allons considérer que, pour chaque instance rencontrée lors du problème journalier de planification des admissions, nous remettons en cause la clé utilisée et recherchons celle qui permettra à la méthode de planification de fournir les meilleurs résultats en terme de performance. L'évaluation d'une clé peut se faire de manière assez intuitive en résolvant le problème de choix des dates d'admissions puis en évaluant le planning proposé par simulation de Monte Carlo par exemple (voir section 3.3.2).

Cependant, cette évaluation ne prenant pas en compte les coûts d'affectation des patients affectés selon la politique représentée par la clé, il demeure un biais. En effet, admettre un patient i jours après qu'il soit connu représente un coût globalement croissant avec le temps. Si les coûts d'affectations ne sont pas pris en compte dans l'évaluation de la clé, alors affecter un patient à la fin de l'horizon de planification reviendrait au même, du point de vue de l'évaluation, qu'affecter un patient deux jours après qu'il soit connu. Peu importe la méthode d'optimisation retenue pour la recherche de la meilleure clé, elle convergerait vers une clé dénaturée où $\gamma_i = 0, \forall i \neq H$ et $\gamma_H = 1$ avec H l'horizon de planification. Avec une telle clé, il n'y aurait aucun nouveau patient avant la fin de l'horizon ce qui aurait comme conséquence de laisser libre la capacité aux patients à admettre. On aurait alors bien évidemment un planning de coût minimal mais faussé par la non prise en compte des futurs patients.

Trois éléments sont donc à prendre en compte lors de l'évaluation de la clé :

1. Les coûts du planning d'admission calculé avec cette clé
2. Dépassements de capacité sur l'horizon de planification
3. Coûts de planning moyens liés à l'admission des patients selon la clé

Par définition de la clé de répartition, nous connaissons le pourcentage de patients réguliers étant admis τ jours après être connus. En calculant le coût moyen correspondant à une telle admission, nous sommes alors en mesure de connaître les coûts de planning liés à la composante γ_τ de la clé. En sommant sur toutes les composantes, nous connaissons alors le coût moyen de planning induit par la planification des patients admis selon la clé de répartition.

Soit $\overline{C_j}$ le coût moyen qu'un patient soit admis j jours après qu'il soit connu (d'après la fonction f définie équation (3.7)) et n_t^r le nombre de patients réguliers nouvellement connus en t alors les coûts de planning moyens liés à la clé peuvent s'exprimer comme suit :

$$\begin{aligned} C = \sum_t \sum_{\tau=1}^{H} n_t^r \gamma_\tau \overline{C_\tau} \alpha^t &= \sum_{\tau=1}^{H} \sum_t n_t^r \gamma_\tau \overline{C_\tau} \alpha^t \\ &= \sum_{\tau=1}^{H} \overline{N^r} \gamma_\tau \overline{C_\tau} \end{aligned} \tag{4.8}$$

avec

$$\overline{N^r} = \sum_t \alpha^t n_t^r \tag{4.9}$$

un terme constant lors de l'évaluation de la clé et $0 << \alpha < 1$ un terme permettant d'équilibrer au fil du temps, dans l'évaluation du gradient, l'importance relative entre les coûts d'affectation et les pénalités de dépassement de capacité.

Finalement, nous pouvons formuler l'objectif de l'optimisation journalière de la clé de répartition comme suit :

$$\min \sum_{i,t} c_{it} X_{it} + \beta \sum_t O_t + C \tag{4.10}$$

Algorithme 4.1 : Evaluation d'une clé de répartition : approche journalière

Input : γ la clé à évaluer
Input : X un planning d'admission
1 Générer K échantillons o_t^kde besoin en lit selon γ;
2 $O \leftarrow \frac{\beta}{k} \sum_{t,k} o_t^k$;
3 $\overline{N^r} \leftarrow \sum_t \alpha^t n_t^r$;
4 $C_{key} \leftarrow \sum_{\tau=1}^{H} \overline{N^r} \gamma_\tau \overline{C_\tau}$;
5 $C_{plan} \leftarrow \sum_{i,t} c_{it} X_{it}$;
6 **return** $C_{plan} + O + C_{key}$;

L'avantage implicite de cette méthode est de permettre de trouver pour chaque instance le planning optimale sous la condition de disposer d'une méthode de recherche de clé fournissant la meilleure clé. En contrepartie, l'inconvénient qui se dessine est lié à la rapidité avec laquelle cette clé peut être trouvée étant donné qu'il va falloir effectuer une recherche pour chaque instance du problème.

4.2.2 Optimisation globale de la clé de répartition

La deuxième façon de considérer le problème d'optimisation de la clé de répartition consiste en la recherche d'une clé générale pouvant être utilisée quelque soit l'instance rencontrée. La clé doit être telle que lors des mises à jour successives du planning d'admission, le résultat global soit le meilleur possible.

L'évaluation de cette clé doit donc être faite sur un ensemble d'instances successives et non plus pour le cas d'une seule instance comme précédemment. Il paraît assez naturel d'utiliser une simulation pour l'évaluation d'une clé. Dans cette simulation, les mises à jour successives sont faites avec la méthode

d'optimisation des admissions retenue qui utilise la clé à évaluer. La méthode d'évaluation demeure la même que dans le problème général de planification des admissions en utilisant l'équation (3.3) et l'algorithme d'évaluation d'une politique d'admissions de la section 3.3.2 avec comme clé de répartition la clé à évaluer. Plusieurs réplications sont nécessaires.

Il n'est pas ici nécessaire de prendre en compte les coûts moyens de planning liés à la clé de répartition comme c'était le cas dans la recherche journalière. En effet, ces coûts représentant en réalité les coûts d'affectation des futurs patients seront implicitement pris en compte lors de leur admission effective dans la simulation.

Avec cette méthode le temps de calcul nécessaire à l'obtention de la meilleure clé n'est plus un problème. En effet, la clé étant calculée une et une seule fois, il importe peu si cette recherche dure plusieurs heures. Même si cette approche peut sembler séduisante du fait de sa généricité, il demeure un risque. Il n'est peut-être pas possible de trouver une clé suffisamment générique pour permettre globalement de minimiser les coûts des planning calculés au fil des jours.

4.3 Introduction aux méthodes de recherches stochastiques

La notion de stochasticité est fréquente dans les différentes méthodes offertes par la recherche opérationnelle. Souvent synonyme de laborieuses heures de révisions pour un étudiant de master, elle trouve tout son sens dans les applications réelles qui sont généralement soumises à des aléas et qui ne sont que rarement entièrement déterministes. Nous avons déjà employé le terme de "stochastique" dans le chapitre précédent sans pour autant le définir. On peut parler de processus stochastique dès lors qu'un phénomène de nature aléatoire va régir l'évolution du système considéré. Cette notion n'est pas propre au génie industriel. On en trouve des applications dans différentes sciences comme la physique, la biologie ou encore la météorologie où les aléas jouent un rôle très important sur l'évolution du système climatique.

Dans le domaine de l'optimisation, Gentle *et al.* (2004) définissent qu'il y a optimisation stochastique si :

- Il y a un bruit aléatoire dans la mesure de la fonction objectif

et/ou

- Lors de la recherche de solutions, il y a un choix aléatoire qui est fait sur la direction de recherche

Nous pouvons citer comme exemple pour le premier point le problème journalier de planification des admissions où l'objectif du modèle d'optimisation ne donnait qu'une estimation de la valeur réelle du critère. Concernant la deuxième propriété, notons qu'il est souvent utile d'introduire de l'aléa dans la recherche d'une solution et ce afin d'éviter des problèmes causés par des optimas locaux. Les algorithmes de voisinage appliquent en générale cette notion.

4.3.1 Recherche aléatoire

Les méthodes de recherche aléatoire sont nombreuses. La plus simple d'entre toute est probablement la méthode de recherche aléatoire aveugle (Blind Random Search). Elle consiste, à chaque itération, en la génération aléatoire d'une nouvelle solution sur l'espace des solutions. Si la solution est meilleure que toutes les solutions précédemment trouvées, alors on la conserve et on continue d'itérer jusqu'à trouver d'autres solutions améliorantes ou jusqu'à ce qu'un maximum d'itérations autorisé soit atteint.

Cette méthode, bien que poussant la simplicité à son maximum, présente tout de même l'intérêt de converger presque sûrement vers une solution optimale si l'espace de solutions est fini. Cependant la lenteur de convergence représente le principal défaut de cette méthode qui se rapproche d'une énumération

exhaustive. La "naïveté" principale de la méthode provient du fait qu'elle ne considère pas les résultats précédents lors de la recherche d'une nouvelle solution.

L'algorithme de recherche aléatoire localisée, définit par Matyas (1965) utilise une information de plus lors du calcul d'une nouvelle solution. En partant de la meilleure solution trouvée à une itération quelconque, un vecteur aléatoire est généré puis ajouté à cette solution afin d'en construire une nouvelle. Si la nouvelle solution est améliorante, alors elle est sauvegardée et les prochaines itérations partiront de cette solution.

On voit apparaître ici la notion de voisinage d'une solution, l'ensemble des voisins étant déterminé en fonction du mode génération du vecteur aléatoire. Cette idée de repartir d'une solution pour en construire une meilleure est la base des algorithmes que nous allons présenter à présent [1] et aboutissant à la méthode de perturbation simultanée.

4.3.2 Approximation stochastique

En étendant l'idée de la recherche locale, il apparaît comme presque évident que de pouvoir trouver du premier coup la meilleure solution voisine d'une solution courante serait un net avantage en terme de temps de calcul. En d'autres termes, comment choisir le vecteur modifiant la solution courante afin d'être sûr d'aller vers une solution améliorante. Le gradient d'une fonction

$$\nabla f = \frac{\partial f}{\partial X} = \left\{ \begin{array}{c} \frac{\partial f}{\partial X_1} \\ \vdots \\ \frac{\partial f}{\partial X_n} \end{array} \right\}$$

nous permet de connaître la direction de plus forte pente et donc la meilleure direction possible à partir d'un point donné. De plus, trouver les optimas d'une fonction f reviendrait à trouver les racines de l'équation $\nabla f = 0$. Malheureusement, il est courant de ne pas avoir à notre disposition la description analytique de la fonction objectif permettant le calcul du vecteur gradient.

La recherche du gradient d'une fonction en point, son estimation ou sa mesure même bruitée est un problème qui a suscité de nombreuses études dans le domaine de l'optimisation. La base des algorithmes d'approximation stochastique de gradient a été exposée par Robbins et Monro (1951). L'idée exposée par les auteurs concerne le cas où il est seulement possible d'avoir une mesure bruitée du gradient en un point X ; soit $g(X)$ cette mesure. L'algorithme décrit par Robbins et Monro se base sur l'équation récursive suivante dans le cas d'une minimisation :

$$X^{(k+1)} = X^{(k)} - a_k g(X^{(k)}) \tag{4.11}$$

où $X^{(k)}$ est la valeur de la solution à la k^{ieme} itération et a_k une séquence de pas positifs. Cette séquence doit respecter certaines propriétés dont nous discuterons un peu plus tard. Remarquons que comme $\nabla f = \frac{\partial f}{\partial X} \neq g$ alors cet algorithme n'est pas une méthode de descente de gradient classique.

De nombreux papiers ont découlé suite à cet article, chacun proposant généralement des améliorations dans l'estimation du gradient $g(X)$. Une des approches connaissant le plus grand succès est celle des différences finies (Finite Difference Stochastic Approximation, FDSA) proposée par Kieffer et Wolfowitz (1952). L'idée est ici d'évaluer le terme $f(X^{(k)} \pm \delta)$ pour chaque composante de $X^{(k)}$. Dans le cas linéaire à deux dimensions cela revient à calculer la pente. On a alors l'estimation du gradient suivante :

$$g(X^{(k)}) = \left[\begin{array}{c} \frac{f(X^{(k)}+c_k e_1)-f(X^{(k)}-c_k e_1)}{2c_k} \\ \vdots \\ \frac{f(X^{(k)}+c_k e_n)-f(X^{(k)}-c_k e_n)}{2c_k} \end{array} \right] \tag{4.12}$$

1. Même si certains algorithmes sont antérieurs à celui ci, le cheminement des idées reste le même

Le couple de valeur $\{a_k, c_k\}$ est appelé séquence de pas. Chacune des deux séries doit respecter certaines conditions afin de garantir la convergence de la méthode des différences finies :

- $a_k > 0$
- $c_k > 0$
- $a_k \to 0$ quand k croît
- $c_k \to 0$ quand k croît
- $\sum_{k=0}^{\infty} a_k = \infty$
- $\sum_{k=0}^{\infty} \frac{a_k^2}{c_k^2} < \infty$

Le choix de la séquence de pas est critique pour les performances de la méthode. Il est courant dans la littérature de choisir les deux définitions suivantes :

$$a_k = \frac{a}{(k+1+A)^{\alpha}} \text{ et } c_k = \frac{c}{(k+1)^{\beta}} \tag{4.13}$$

Il reste à l'utilisateur de choisir les paramètres a, c, α et β strictement positif et $A \geq 0$.

4.3.3 Approximation stochastique par perturbation simultanée

Un des inconvénients de la méthode d'approximation stochastique par les différences finies réside dans le nombre d'évaluations nécessaires de la fonction objectif. Pour chaque composante du vecteur solution X, la méthode nécessite 2 évaluations de la fonction f, soit au total $2n$ évaluations à chaque itérations si l'on considère que n est la dimension du vecteur X.

Si l'évaluation de la fonction f requiert la résolution d'un problème combinatoire ou le résultat d'une simulation, le temps de calcul multiplié par le nombre d'évaluations requises à chaque itération multiplié par le nombre d'itérations désirées se révèle être prohibitif.

Pour évaluer le gradient d'une fonction en un point, la méthode FDSA perturbe chacune des composantes de la solution courante. Il est nécessaire à chaque perturbation de calculer la valeur de la solution résultante, d'où le nombre de calculs prohibitifs. L'idée proposée par Spall (1992) est de perturber simultanément toutes les composantes de la solution. Cette méthode appelée approximation stochastique par perturbation simultanée (Simultaneous Perturbation Stochastic Approximation, SPSA) se base sur l'estimation suivante du gradient :

$$g(X^{(k)}) = \begin{bmatrix} \frac{f(X^{(k)}+c_k\Delta_k)-f(X^{(k)}-c_k\Delta_k)}{2c_k\Delta_{k1}} \\ \vdots \\ \frac{f(X^{(k)}+c_k\Delta_k)-f(X^{(k)}-c_k\Delta_k)}{2c_k\Delta_{kn}} \end{bmatrix} = \frac{f(X^{(k)}+c_k\Delta_k)-f(X^{(k)}-c_k\Delta_k)}{2c_k} \left[\Delta_{k1}^{-1}, \ldots, \Delta_{kn}^{-1}\right] \tag{4.14}$$

Où $\Delta_k = (\Delta_{k1}, \ldots, \Delta_{kn})$ est un vecteur de perturbation aléatoirement généré.

Les conditions de convergence demeurent les mêmes que dans FDSA pour la séquence de gain $\{a_k, c_k\}$. Des conditions un peu plus restrictives sont données pour la distribution des composantes Δ_{ki}. Cependant, l'auteur indique en illustration de la méthode exposée que l'utilisation d'une distribution de Bernoulli, à résultats ± 1, répond aux exigences théoriques du vecteur Δ afin de garantir la convergence de la méthode.

Une présentation claire de ce que sont les méthodes d'approximation stochastique, ce qu'est la méthode SPSA, des exemples de domaines d'applications ou encore des perspectives de recherche pour améliorer la méthode sont très bien exposés dans Spall (1998b).

La méthode de perturbation simultanée présente donc l'avantage de n'utiliser que 2 évaluations de la fonction objectif par itération. Dans les exemples donnés dans différents articles de Spall, il est clairement montré qu'en termes de nombre d'évaluation de la fonction économique, la méthode SPSA converge plus

rapidement. Des applications montrant la performance de la méthode par rapport à d'autres approximations stochastiques dans le cas de systèmes à évènements discrets sont montrées dans Fu et Hill (1997) mais aussi dans Gerencser *et al.* (2004) et Hill (2005). Enfin, des exemples de domaines d'applications ou encore des perspectives de recherche pour améliorer la méthode sont très bien exposés dans Spall (1998b).

La méthode tel que présentée dans sa version de base fonctionne dans le cas de problèmes non contraints. Cependant, on trouve dans la littérature des études permettant d'étendre les champs d'application de l'approximation par perturbation simultanée. Sadegh (1997) montre que lorsqu'une solution sort de l'espace des solutions admissibles, alors en la projetant vers le point le plus proche de l'espace des solutions, la convergence reste assurée. Dans le cas de problèmes fortement contraints, il n'est pas toujours évident de pouvoir trouver une telle projection qui, dans l'affirmative, doit être implémentable. D'autres techniques pour l'optimisation sous contraintes avec approximation stochastique définissent une fonction de pénalité pour chaque contrainte violée. Dans le cadre de SPSA, nous pouvons nous référer principalement à Wang et Spall (2003, 2008).

4.3.4 Estimation du gradient par analyse de perturbation

Parmi les algorithmes implémentant des méthodes d'optimisation par descente de gradient, nous pouvons distinguer deux catégories : les méthodes basées sur des approximations du gradient et les méthodes basées sur une mesure du gradient. Dans la première catégorie, les méthodes se basent sur des évaluations de la fonction objectif pour des points donnés. La méthode de perturbation simultanée, qui vient d'être abordée, s'inscrit dans cette catégorie. Dans le deuxième cas, la mesure du gradient, en accord avec les paramètres à optimiser, est faite sur l'ensemble des données, qui potentiellement peuvent être bruitées. Parmi les méthodes relevant de cette catégorie, nous allons nous intéresser à l'analyse de perturbation. Nous nous référons dans la littérature à Ho et Cao (1991); Glasserman (1991); Cao (1996); Fu et Hu (1997).

De manière informelle, l'analyse de perturbation se base sur l'étude d'une longue simulation appelée chemin nominal. La mesure du gradient est ensuite réalisée en évaluant l'impact de la modification d'un paramètre sur le reste de la simulation. Le chemin résultant de cette modification est appelé chemin perturbé. En d'autres termes, en observant un chemin, parfois appelé également la trajectoire, emprunté par un système, nous utilisons l'analyse pour répondre à la question : « Qu'arriverait-il si nous reproduisons la même trajectoire avec une petite perturbation d'un évènement à un instant t ? ».

La figure 4.1 illustre les notions de chemin nominal et chemin perturbé dans un cas simple de système à file d'attente. Le système considéré est composé d'un serveur traitant les demandes les unes après les autres et dans l'ordre de leurs arrivées (FIFO). Le temps de traitement d'une demande du serveur est soumis à des paramètres optimisables. Le nombre d'entités pouvant être présentes simultanément est limité à 4 dans l'exemple traité.

Dans cette illustration, avec la séquence d'arrivées données, la cinquième demande est rejetée faute de place disponible dans le système dans le chemin nominal. Supposons que la modification d'un paramètre du serveur conduise à traiter plus rapidement la première demande. Alors le chemin perturbé résultant permet de traiter la cinquième demande. Dans cet exemple, nous avons illustré un cas où la modification d'un paramètre conduit à une amélioration des performances du système (plus de demandes traitées). La réciproque, conduisant à une dégradation des performances, peut bien entendu être vraie.

Voyons à présent un peu plus formellement le principe de cette méthode. Soit θ un paramètre et Θ l'ensemble des paramètres du système. Soit ω le chemin nominal d'une simulation sur ce système et $L(\omega)$ une mesure de la performance du système. Le chemin ω peut être vu comme un ensemble d'élément τ_i, la date dans la simulation à laquelle survient l'évènement i. Déterminer l'effet d'un paramètre θ sur les performances du système revient à considérer trois sous-problèmes (Ho et Cao (1991)) :

1. Quel impact a le changement du paramètre θ sur le timing de τ_i ; i.e. $\frac{\partial \tau_i}{\partial \theta} = ?$

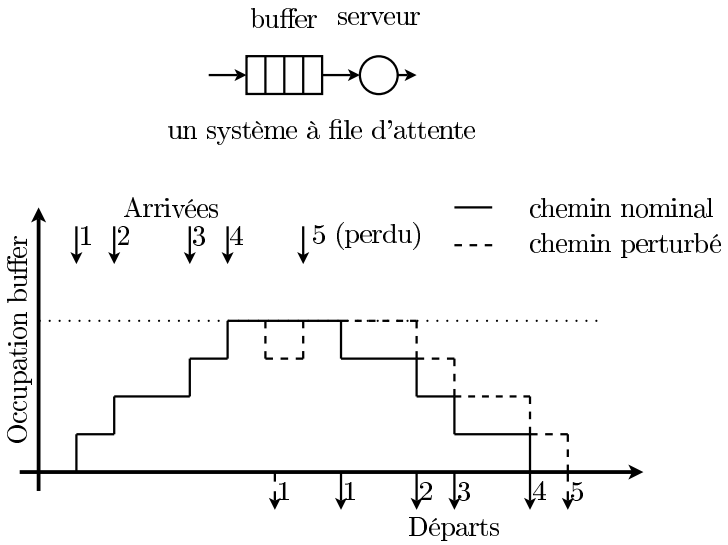

FIGURE 4.1 – Chemin nominal et chemin perturbé pour un système à file d'attente

2. Quel impact le changement de timing τ_i d'un évènement a sur les autres évènements, τ_k ; i.e. $\frac{\partial \tau_k}{\partial \tau_i} = ?$
3. Quel impact le changement de timing d'un ensemble d'évènement τ_k a sur la performance du système L ; i.e. $\frac{\partial L}{\partial \tau_k} = ?$

En mettant bout à bout ces sous-problèmes, on obtient la mesure de l'impact du changement du paramètre θ sur la performance du système :

$$\frac{\partial L}{\partial \theta} = \sum_i \frac{\partial L}{\partial \tau_k} \frac{\partial \tau_k}{\partial \tau_i} \frac{\partial \tau_i}{\partial \theta} \tag{4.15}$$

Cette méthode d'analyse est appelée analyse de perturbation infinitésimale puisque basée sur l'étude de petites perturbations. Pour finir, récapitulons les étapes d'un algorithme d'optimisation par analyse de perturbation (algo 4.2).

Algorithme 4.2 : Optimisation par analyse de perturbation

```
1 Initialiser les paramètres du système;
2 Définir par simulation un chemin nominal;
3 for k=1 to n do
4 |   g ← Mesure de ∂L/∂Θ;
5 |   Mise à jour des paramètres du système;
  |   /* par l'algorithme de Robbins-Monro par exemple                    */
6 |   Mise à jour du chemin nominal avec les nouveaux paramètres;
7 end
```

4.4 Optimisation de la clé de répartition des patients par la méthode de perturbation simultanée

Nous allons à présent nous intéresser à la résolution du problème de recherche de la clé de répartition des patients en utilisant la méthode d'approximation stochastique par perturbation simultanée.

De manière basique, l'algorithme que nous allons employer respecte la base décrite par James Spall dans ses différents ouvrages et est donné dans l'algorithme 4.3 pour un cas avec n itérations [2].

Algorithme 4.3 : Algorithme SPSA

$\gamma(0) \leftarrow$ trouver une clé solution initiale;
for k=*1* **to** n **do**
  $a_k \leftarrow a/(k+A+1)^{\alpha}$;
  $c_k \leftarrow c/(k+1)^{\beta}$;
  $\Delta \leftarrow$ générer un vecteur à valeur ± 1 selon une distribution de Bernoulli de paramètre 0.5;
  $\gamma^{+} \leftarrow \gamma^{(k-1)} + c_k * \Delta$;
  $\gamma^{-} \leftarrow \gamma^{(k-1)} - c_k * \Delta$;
  $g \leftarrow \frac{f(\gamma^{+}) - f(\gamma^{-})}{2 * c_k * \Delta}$;
  $\gamma^{(k)} \leftarrow \gamma^{(k-1)} - a_k * g$
end

L'hypothèse ici est faite de l'utilisation de la séquence de gain définie dans l'équation 4.13 correspondant aux méthodes standards de descente de gradient par l'algorithme FDSA. Les initialisations des paramètres $a, c, A, n, \alpha, \beta$ et p seront discutées plus en détail dans les expérimentations numériques en fin de chapitre (section 4.6).

Sous cet algorithme simple en apparence se cache deux difficultés d'implémentations que nous nous proposons de résoudre avant de pouvoir continuer avec la résolution en pratique de la recherche de clé de répartition :

1. Garantir que les clés perturbées γ^{+} et γ^{-} ainsi que la clé calculée à chaque itération $\gamma^{(k)}$ soit dans l'espace des solutions.
2. Le choix de la fonction d'évaluation f d'une clé

4.4.1 Projection d'un point dans l'espace des solutions

En partant du résultat précédemment cité et démontré dans Sadegh (1997) indiquant que lorsqu'un point est en dehors de l'espace des solutions alors en utilisant la solution la plus proche, la méthode de perturbation simultanée conserve ses propriétés de convergence, nous définissons une projection $\Pi : \mathbb{R} \rightarrow \Gamma$. Dans le cas de la recherche de la clé résultante de l'itération k, la projection va associer à chaque "clé" $\gamma^{(k)\prime}$, calculée à partir de l'approximation faite du gradient, la clé la plus proche dans l'espace des solutions. Nous procédons en deux étapes. Premièrement, nous projetons cette clé dans l'espace, $\mathcal{E}$ défini par $\sum_i X_i = 1$. La distance entre un point X et son projeté $Y \in \mathcal{E}$ doit être minimale. Comme l'espace contenant l'espace des solutions est régi par l'équation voulant que la somme des composantes du vecteur soit 1, nous posons comme paramètre la dernière composante du vecteur, la N^{ieme}. Soit la définition suivante de la distance entre deux points :

$$d(X,Y) = \sum_{i \neq N} (Y_i - X_i)^2 + (1 - \sum_{i \neq N} Y_i - X_N)^2$$

et sa dérivée en sa i^{eme} composante :

$$\begin{aligned} \frac{\partial d(X,Y)}{\partial Y_i} &= 2\,(Y_i - X_i) - 2\left(1 - \sum_{i \neq N} Y_i - X_N\right) \\ &= 2\,(Y_i - X_i) - 2\,(Y_N - X_N) \end{aligned}$$

2. il est aisé de modifier l'algorithme pour insérer tout autre condition d'arrêt jugée pertinente

Cette dérivée s'annule pour tout point Y tel que :

$$\begin{cases} Y_i = X_i + Y_N - X_N & ,\forall i \neq N \\ Y_N = \frac{1-\sum_{i\neq N} X_i + (N-1)X_N}{N} \end{cases}$$

Partant d'une solution $\gamma^{(k)\prime}$ en dehors de l'espace des solutions, nous sommes à présent capable de calculer son projeté Y dans un sous espace $\mathcal{E}$ contenant l'espace des solutions Γ. Partant de $\mathcal{E}$, il reste à garantir que chaque composante X_i soit bien dans l'intervalle $[0, 1]$. Partant d'une solution $X^{(k-1)} \in \Gamma$ et du projeté Y, nous cherchons le point $X^{(k)}$à la frontière de Γ et tel que les trois points suscités soient alignés. Nous définissons une translation permettant d'obtenir le point $X^{(k)} \in \Gamma$ recherché à partir du point Y comme illustré figure 4.2.

$$X^{(k)} = (1-\alpha)X^{(k-1)} + \alpha Y$$

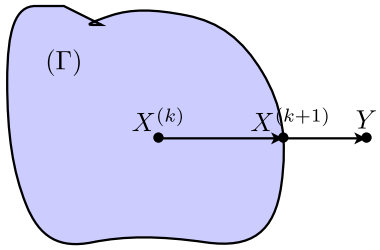

FIGURE 4.2 – Translation d'un point de $\mathcal{E}$ dans l'espace des solutions

La recherche du coefficient α est alors réalisée comme suit :

$$\alpha = \min_i \alpha_i$$

où α_i est calculé pour $i = 1$ à N en fonction des valeurs de Y_i :

$$\alpha_i = \begin{cases} 1 & , \text{ si } Y_i \in [0,1] \\ \frac{X_i^{(k-1)}}{X_i^{(k-1)} - Y_i} & , \text{ si } Y_i < 0 \\ \frac{X_i^{(k-1)} - 1}{X_i^{(k-1)} - Y_i} & , \text{ si } Y_i > 1 \end{cases}$$

et sous les contraintes $\alpha_i = \begin{cases} 0 & \text{si } \alpha_i < 0 \\ 1 & \text{si } \alpha_i > 1 \\ \alpha_i & \text{sinon} \end{cases}$

La projection Π d'un point de l'espace vers le point le plus proche dans l'espace des solutions que nous venons de définir est illustrée figure 4.3 dans le cas à 2 dimensions.

Etant donné la projection Π que nous venons de définir, nous pouvons donner la forme générale de l'algorithme nommé PSCR pour Perturbation Simultanée pour la recherche de Clé de Répartition (Algo. 4.4).

4.4.2 Différentes variantes pour le problème de recherche par perturbation simultanée

PSCRJ : Une clé de répartition par instance

La première approche que nous proposons traite du problème d'optimisation de la clé de répartition que nous avons qualifié de journalier. Considérant une instance, nous cherchons à trouver la clé offrant une

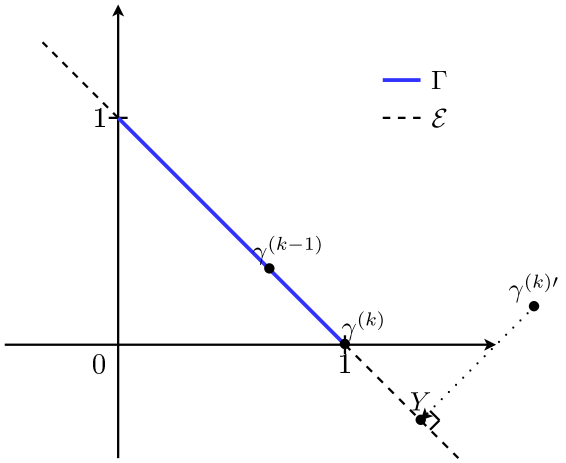

FIGURE 4.3 – Illustration de la projection dans l'espace des solutions ; cas en dimension 2

Algorithme 4.4 : PSCR

$\gamma^{(0)} \leftarrow$ Définir une clé initiale **for** *k=1* **to** n **do**
  $a_k \leftarrow a/(k+A+1)^{\alpha}$;
  $c_k \leftarrow c/(k+1)^{\beta}$;
  $\Delta \leftarrow$ générer un vecteur à valeur ± 1 selon une distribution de Bernoulli de paramètre 0.5;
  $\gamma^{+} \leftarrow \Pi(\gamma^{(k-1)} + c_k * \Delta)$;
  $\gamma^{-} \leftarrow \Pi(\gamma^{(k-1)} - c_k * \Delta)$;
  $g \leftarrow \frac{f(\gamma^{+}) - f(\gamma^{-})}{2 * c_k * \Delta}$;
  $\gamma^{(k)} \leftarrow \Pi(\gamma^{(k-1)} - a_k * g)$
end

solution de valeur minimal au problème de planification des admissions. L'algorithme PSCRJ (Algo 4.5) suit dans les grandes lignes l'algorithme général PSCR. La différence réside dans la nécessité de recalculer régulièrement le planning en fonction de la clé courante. Comme nous le disions en introduction, le planning dépend de la clé et la clé dépend du planning. Enfin, la fonction d'évaluation d'une clé, f suit l'algorithme correspondant à la section sur l'optimisation journalière de la clé de répartition (Algo. 4.1).

PSCRG1 : Une clé unique pour chaque instance

Cette deuxième approche se focalise sur la recherche d'une clé globale, une clé qui sera utilisée quelque soit l'instance considérée. L'algorithme, nommé PSCRG1, permettant cette recherche suit directement l'algorithme général (Algo. 4.4). La fonction d'évaluation f utilisée est telle que nous l'avons décrite en section 4.2.2. Elle s'appuie sur l'approche de simulation que nous avons proposée et qui vise à évaluer une politique d'admission. Cette politique d'admission va utiliser la clé de répartition à évaluer afin d'estimer les futurs besoins en termes de lits.

PSCRG2 : Une clé journalière globale

A mi-chemin entre la recherche d'une clé pour chaque instance rencontrée et la recherche d'une clé globale, nous pouvons formuler une approche considérant une clé globale pour chaque jour de la semaine. Ainsi, le problème revient à chercher, par exemple pour le lundi, une clé qui sera utilisée tous les lundi. La solution fournie par cette approche est donc un ensemble de clé de répartition ; une par jour de la

Algorithme 4.5 : PSCRJ

$\gamma^{(0)} \leftarrow$ Définir une clé initiale **for** *k=1* **to** n **do**
  if $k \; mod \; m == 0$ **then**
    Calculer un nouveau planning des admissions;
  end
  $a_k \leftarrow a/(k+A+1)^{\alpha}$;
  $c_k \leftarrow c/(k+1)^{\beta}$;
  $\Delta \leftarrow$ générer un vecteur à valeur ± 1 selon une distribution de Bernoulli de paramètre 0.5;
  $\gamma^{+} \leftarrow \Pi(\gamma^{(k-1)} + c_k * \Delta)$;
  $\gamma^{-} \leftarrow \Pi(\gamma^{(k-1)} - c_k * \Delta)$;
  $g \leftarrow \frac{f(\gamma^{+}) - f(\gamma^{-})}{2 * c_k * \Delta}$;
  $\gamma^{(k)} \leftarrow \Pi(\gamma^{(k-1)} - a_k * g)$
end

semaine.

Des adaptations doivent être apportées sur l'algorithme général employé jusqu'ici. En premier lieu, il convient de trouver une manière de procéder permettant de prendre en considérations plusieurs instances du jour considéré. En effet, pour un jour donné de la semaine, se contenter d'une seule instance reviendrait exactement au même que la recherche de clé journalière. Il va être nécessaire d'employer un ensemble d'instances afin de garantir le bon comportement global de la clé recherchée.

La solution que nous proposons de mettre en place consiste à utiliser une simulation tel celle permettant l'évaluation d'une politique d'admission où pour chaque jour j, le calcul du nouveau planning des admissions sera fait en tenant compte de la clé du jour donné. A chaque itération de l'algorithme de simulation, une optimisation de la clé de répartition correspondante au jour considéré sera effectuée par l'estimation de gradient selon la méthode de perturbation simultanée. Ainsi, la clé correspondant au lundi sera optimisée en fonction de la suite d'instances correspondantes rencontrées. La phase d'optimisation se faisant en fonction de l'instance rencontrée, il paraît logique d'utiliser l'évaluation journalière de la clé de répartition (Algo. 4.1)comme fonction f d'évaluation des clés perturbées.

Le changement d'instance est inévitable afin de permettre de trouver une clé généralement bonne quelle que soit l'instance rencontrée. Cependant, il existe un risque de dénaturer le problème et notamment l'estimation faite du gradient. En effet, le changement d'instance revient à changer la fonction que l'on cherche à optimiser. Le problème à résoudre n'étant plus le même, les ordres de grandeurs et les directions du gradient ne sont plus obligatoirement les mêmes. C'est un peu comme si l'on faisait une mesure du gradient d'une certaine fonction à une itération et une mesure du gradient d'une autre fonction à l'itération suivante. La méthode d'optimisation par perturbation simultanée est conçue pour l'optimisation d'un problème étant donnée une fonction objectif constante.

Cependant, l'idée avancée dans cet algorithme d'avoir des itérations de la méthode dite de perturbation simultanée tout au long d'une simulation semble appropriée pour la recherche d'une clé globale. La variante PSCRG2 va, en conservant le principe des itérations au fil d'une simulation, itérer sur chaque instance rencontrée un certain nombre de fois avant de passer à l'instance suivante dans la simulation. Cette répétition à pour but de permettre la convergence de la clé courante vers une meilleure clé pour l'instance considérée. Cette condition de répétition devient même impérative si l'on considère que la génération du vecteur de perturbation est aléatoire. Il est donc obligatoire de faire plusieurs itérations afin d'assurer la convergence de la méthode vers un optima. Le détail de PSRG2 est donné algorithme 4.6 où γ_j représente la clé du j^{eme} jour de la semaine.

De par sa construction, cette méthode de recherche d'une meilleure clé globale peut être vue comme

Algorithme 4.6 : PSCRG2

```
γ^(0) ← Définir une clé initiale Construire un état d'occupation initial de l'hôpital;
Construire une liste de patients programmés;
for k=1 to n do
    for j=1 to 7 do
        Effectuer les entrées/sorties du jour k * 7 + j;
        Ajouter les patients électifs nouvellement connu à la liste des programmés;
        for i=1 to m do
            g ← Approximer la composante j du gradient par SPSA avec l'évaluation journalière;
            γ_j^(k*m+i) ← Π(γ_j^(k*m+i-1) − a_{k*m+i} g);
        end
        Calculer le nouveau planning des admissions;
    end
end
```

étant une méthode cherchant la meilleure clé globale pour chaque jour de la semaine.

4.5 Optimisation de la clé de répartition de patients par l'analyse de perturbation

Nous allons à présent considérer la recherche d'une clé en utilisant la méthode d'analyse de perturbation décrite en section 4.3.4. Pour faire le lien avec la description théorique de cette méthode, dans le cas de la recherche de la meilleure clé de répartition des patients électifs, les évènements τ_i sont les dates d'arrivée des patients correspondant et les paramètres à optimiser sont les composantes de la clé de répartition et qui, donc, régissent bien les évènements du système considéré. Les perturbations que nous allons traiter, les modifications de la clé de répartition, vont donc avoir comme conséquences d'avancer ou de retarder les arrivées des patients.

Pour un planning d'admission donné, l'évaluation de la fonction objectif du problème de planification des admissions de patients réguliers, la performance, ne dépend que des dépassements de capacité au fil des périodes. En particulier, à un instant donné, le dépassement de capacité est donné par l'équation suivante :

$$O_t = E\left[\left(N_t^p - T_t + N_t^v(\gamma)\right)^+\right] \tag{4.16}$$

où N_t^p est le nombre de patients présents à l'instant t sur lesquels on ne peut agir. Ce nombre inclut les patients déjà présents à l'instant 0 et les patients en provenance des urgences qui ont été admis lors d'une période précédente et qui sont toujours présents. T_t désigne la capacité en lit de l'instant t et $N_t^v(\gamma)$ désigne le nombre de patients présents à l'instant t mais dont la valeur dépend de la clé de répartition γ.

Supposons que le nombre total de patients inconnus nous soit donné. Soit N le nombre total de patients inconnus dont l'admission est contrainte par la clé de répartition γ. Un patient inconnu i peut être représenté par le triplet $(r_i, a_i(\gamma), l_i)$ où, pour mémoire, r_i désigne la date à laquelle le patient est connu, l_i désigne la durée de séjour du patient et a_i est défini à partir de la clé de répartition comme

suit :

$$a_i = \begin{cases} r_i + 1 & , \text{ avec une probabilité } \gamma_1 \\ r_i + 2 & , \text{ avec une probabilité } \gamma_2 \\ \vdots & \\ r_i + n & , \text{ avec une probabilité } \gamma_n \end{cases}$$

Finalement soit $\omega = (r_i, a_i, l_i)_{i=1}^N$ le chemin nominal de la simulation et $L(\omega) = \sum_t O_t$.

Sans perte de généralité supposons qu'il existe une clé de répartition γ^i pour chaque patient i avec $\gamma^i = \gamma, \forall i$. Réécrivons la définition de O_t :

$$O_t = E\left[\left(N_t^p - T_t + \sum_{i=1}^{N} \mathbb{1}\{a_i \leq t < a_i + l_i\}\right)^+\right]$$

comme a_i dépend de γ^i alors $O_t = F(\gamma^1, \gamma^2, \ldots, \gamma^N)$ et donc

$$\begin{aligned} \frac{\partial O_t}{\partial \gamma} &= \frac{\partial F}{\partial \gamma^1}\frac{\partial \gamma^1}{\partial \gamma} + \ldots + \frac{\partial F}{\partial \gamma^N}\frac{\partial \gamma^N}{\partial \gamma} \\ &= \sum_{i=1}^{N} \frac{\partial F}{\partial \gamma^i} \end{aligned}$$

Afin de pouvoir évaluer chaque composante de cette dérivée, revenons sur l'équation définissant O_t et reformulons la :

$$\begin{aligned} O_t &= E\left[\left(N_t^p - T_t + \sum_{i=1}^{N} \mathbb{1}\{a_i \leq t < a_i + l_i\}\right)^+\right] \\ &= E\left[\left(N_t^p - T_t + N_{jt}^v + \mathbb{1}\{a_j \leq t < a_j + l_j\}\right)^+\right] \end{aligned}$$

avec

$$N_{jt}^v = \sum_{i \neq j}^{N} \mathbb{1}\{a_i \leq t < a_i + l_i\}$$

un terme indépendant de γ^j représentant le nombre de patients électifs inconnus, sauf j, présents à l'instant t dans le chemin nominal.

$$\begin{aligned} O_t &= E\left[E\left[\left(N_t^p - T_t + N_{jt}^v + \mathbb{1}\{a_j \leq t < a_j + l_j\}\right)^+ | \omega - (r_j, a_j, l_j)\right]\right] \\ &= E\left[\sum_{\tau} \left(N_t^p - T_t + N_{jt}^v + \mathbb{1}\{r_j + \tau \leq t < r_j + \tau + l_j\}\right)^+ \gamma_\tau^j\right] \end{aligned}$$

Il s'ensuit pour la dérivée partielle de O_t par rapport à la perturbation de la date d'arrivée du patient j que

$$\frac{\partial O_t}{\partial \gamma^j} = E\left[\sum_{\tau} \left(N_t^p - T_t + N_{jt}^v + \mathbb{1}\{r_j + \tau \leq t < r_j + \tau + l_j\}\right)^+ \dot{\gamma_\tau}\right]$$

donc pour la dérivée de O_t

$$\frac{\partial O_t}{\partial \gamma} = E\left[\sum_{j=1}^{N}\sum_{\tau} \left(N_t^p - T_t + N_{jt}^v + \mathbb{1}\{r_j + \tau \leq t < r_j + \tau + l_j\}\right)^+ \dot{\gamma_\tau}\right]$$

et pour le critère prenant en compte la somme des dépassements de capacité :

$$\frac{\partial \sum_t O_t}{\partial \gamma} = E\left[\sum_t \sum_{j=1}^{N}\sum_{\tau} \left(N_t^p - T_t + N_{jt}^v + \mathbb{1}\{r_j + \tau \leq t < r_j + \tau + l_j\}\right)^+ \dot{\gamma_\tau}\right] \tag{4.17}$$

Notons que dans le cas du calcul de composante i du gradient, $g_i = \frac{\partial \sum_t O_t}{\partial \gamma_i}$ alors, chaque composante de la clé de répartition étant indépendante, la dérivée $\dot{\gamma_\tau} = 1$ si $\tau = i$ et 0 sinon.

Dans le cas où le nombre de patients inconnus n'est pas connu, supposons $\overline{N}$ une borne de ce nombre. Il suffit alors de résoudre le problème avec $\overline{N} - N$ patients fictifs avec une date d'arrivée r_i supérieure à la fin de l'horizon.

Tout le problème qui se pose à présent réside dans l'évaluation de l'équation (4.17) qui est une mesure du gradient. Dans cette équation, le terme $\left(N_t^p - T_t + N_{jt}^v + \mathbb{1}\{r_j + \tau \leq t < r_j + \tau + l_j\}\right)^+$ représente le dépassement de capacité dans le chemin perturbé où le patient j a été admis en période $r_j + \tau$ au lieu de a_j.

Comme $\sum_\tau \dot{\gamma_\tau} = 0$, l'équation 4.17 est équivalente à

$$\frac{\partial \sum_t O_t}{\partial \gamma} = E\left[\sum_t \sum_{j=1}^{N} \sum_\tau \left(\left(N_t^p - T_t + N_{jt}^v + \mathbb{1}\{r_j + \tau \leq t < r_j + \tau + l_j\}\right)^+ - \left(N_t^p - T_t + N_t^v(\gamma)\right)^+\right) \dot{\gamma_\tau}\right] \tag{4.18}$$

Par conséquent, et afin de pouvoir mesurer le gradient lors de l'exécution de l'algorithme d'analyse de perturbation, nous définissons la variable $\Delta_{jt\tau}$, la variation de dépassement de capacité en période t par rapport au chemin nominal si le patient j est admis τ périodes après r_j ; i.e. en période $r_j + \tau$ au lieu de a_j, sa date d'admission nominale. A l'aide de cette variable, réécrivons la mesure du gradient faite par la méthode d'analyse de perturbation :

$$g = \frac{\partial \sum_t O_t}{\partial \gamma} = E\left[\sum_\tau \left(\sum_{j,t} \Delta_{jt\tau}\right) \dot{\gamma_\tau}\right]$$

En particulier pour chaque composante du gradient :

$$g_\tau = \sum_{j,t} \Delta_{jt\tau} = \sum_j \sum_{t=r_j+1}^{r_j+H+l_j} \Delta_{jt\tau} \tag{4.19}$$

Notons bien qu'ici nous sommons les différences qui peuvent survenir en la plus petite date d'admission possible, $r_j + 1$ et la plus grande date de départ possible due à une admission repoussée à son maximum, $r_j + H + l_j$. La variation de dépassement de capacité due à la perturbation de la date d'admission d'un patient se calcule comme suit

$$\sum_t \Delta_{jt\tau} = \begin{cases} \sum_{t=a'_j}^{a'_j+l_j-1} \Delta_t^+ + \sum_{t=a_j}^{a_j+l_j-1} \Delta_t^- & \text{, si } a'_j + l_j \leq a_j \text{ ou } a_j + l_j \leq a'_j \\ \sum_{t=a'_j}^{a_j-1} \Delta_t^+ + \sum_{t=a'_j+l_j}^{a_j+l_j-1} \Delta_t^- & \text{, si } a'_j \leq a_j \leq a'_j + l_j \\ \sum_{t=a_j}^{a'_j-1} \Delta_t^- + \sum_{t=a_j+l_j}^{a'_j+l_j-1} \Delta_t^+ & \text{, si } a_j < a'_j < a_j + l_j \end{cases} \tag{4.20}$$

où $a'_j = r_j + \tau$ est la date d'admission perturbée et Δ_t^+ (resp. Δ_t^-) est la variation de dépassement de capacité si la période t à un patient de plus (resp. de moins) :

$$\Delta_t^+ = \begin{cases} 1 & \text{, si } N_t - T_t \geq 0 \text{ et } t \leq H \\ 0 & \text{, sinon} \end{cases} \text{ et } \Delta_t^- = \begin{cases} -1 & \text{, si } N_t - T_t > 0 \text{ et } t \leq H \\ 0 & \text{, sinon} \end{cases}$$

avec N_t le nombre total de patients présents à la période t et T_t la capacité en lits de la période t.

Dans la définition de la variation de dépassement de capacité dans le chemin perturbé (équation (4.20)), le premier cas correspond à la situation où la période d'occupation nominale et perturbée sont disjointes. Le deuxième cas correspond à la situation où l'admission est avancée par rapport à l'admission nominal. Le troisième cas correspond à la situation où l'admission est repoussée.

Tout la démonstration qui vient d'être effectuée est faite sous l'hypothèse d'un planning fixé. Si la planning n'était pas fixe, comme c'est le cas dans la recherche globale de la meilleure clé de répartition

des patients, la condition de dérivabilité n'est pas respectée et le résultat pour la mesure du gradient n'est plus valable. Nous ne pouvons donc utiliser l'analyse de perturbation que pour l'optimisation journalière de la clé de répartition.

Il demeure encore que la mesure faite du gradient ne prend en compte pour l'instant que les dépassements de capacité. Les coûts moyens de planning liés à la clé doivent être ajoutés. L'expression de ces coûts est donnée équation (4.8). La dérivée de cette expression pour une composante τ de la clé est aisément calculable. Il vient finalement l'expression finale de la mesure du gradient :

$$g_\tau = \beta \sum \Delta_{jt\tau} + \overline{N^r}\ \overline{C_\tau}, \forall \tau \tag{4.21}$$

où $\overline{N^r}$ est la constante définie équation (4.9).

4.5.1 Recherche de la meilleure direction améliorante

Comme dans la méthode SPSA, lors de chaque itération, une fois la mesure du gradient effectuée, la mise à jour de la solution courante à l'aide de l'équation (4.11) peut nous mener à une solution ne respectant pas les contraintes établies pour la clé de répartition des patients.

Une alternative simple consiste, comme précédemment, à trouver le projeté le plus proche dans l'espace des solutions de la clé de l'itération k. Observons un peu plus précisément le comportement de cette construction lorsque nous sommes dans un cas limite où la solution courante se trouve à la frontière de l'espace des solutions ; i.e. au moins une des composantes du vecteur clé est nulle. Si le vecteur gradient $-g(X^{(k)})$ pointe en dehors de l'espace de solution, alors le projeté le plus proche dans l'espace des solutions du point $X^{(k+1)}$ est le point $X^{(k)}$.

Dans ce cas de figure, l'itération ne permet pas d'aller vers une meilleure solution et est donc en quelque sorte perdue. Cela peut se révéler être préjudiciable pour la performance de la méthode puisque le phénomène peut se répéter pendant une succession d'itérations ; il n'y a a priori que peu de chances que d'une itération à l'autre la direction du gradient varie fortement. Pourtant, une direction améliorante peut être déterminée en fonction du gradient mesuré permettant ainsi d'améliorer la solution courante.

En partant de la mesure du gradient faite lors de l'analyse de perturbation, nous pouvons définir la meilleure direction améliorante. Nous nous ramenons à l'algorithme suivant :

$$\gamma^{(n+1)} = \gamma^{(n)} + a_n d_n$$

où a_n est le pas à l'itération n et d_n est la direction de plus forte descente. d_n répond au problème suivant :

$$Minimiser \sum_\tau g_\tau d_\tau \tag{4.22}$$

sous les conditions

$$\sum_\tau d_\tau = 0 \tag{4.23}$$

$$\sum_\tau d_\tau^2 = 1 \tag{4.24}$$

$$d_\tau \geq 0, \text{ si } \gamma_\tau = 0 \tag{4.25}$$

$$d_\tau \leq 0, \text{ si } \gamma_\tau = 1 \tag{4.26}$$

Cette formulation du problème de recherche de la meilleure direction améliorante est quadratique. Cependant nous pouvons redéfinir cette recherche en considérant tout d'abord l'algorithme 4.7 lequel se base sur l'utilisation de la valeur moyenne des composantes considérées du gradient, $\overline{g}$.

Algorithme 4.7 : Direction de plus forte descente

Soit $Q = \{1, 2, \ldots, \tau, \ldots, H\}$;
$\overline{g} \leftarrow \frac{\sum_{\tau \in Q} g_\tau}{|Q|}$;
foreach τ *tel que* $g_\tau \geq \overline{g}$ *et* $\gamma_\tau = 0$ **do**
    $d_\tau \leftarrow 0$;
    Retirer τ de l'ensemble Q;
end
Actualiser $\overline{g}$ et re-itérer jusqu'à convergence;
$c \leftarrow \left(\sqrt{\sum_\tau (\overline{g} - g_\tau)^2}\right)^{-1}$;
foreach $\tau \in Q$ **do**
    $d_\tau \leftarrow c(\overline{g} - g_\tau)$;
end

Avec la solution fournie par cet algorithme il vient immédiatement que

$$\sum_\tau g_\tau d_\tau = c|Q| \left(\overline{g}^2 - \frac{1}{|Q|} \sum_{\tau \in Q} g_\tau^2 \right) \leq 0$$

ce qui montre que l'on a bien une direction améliorante. Intéressons nous à présent à montrer que cette direction est en fait la direction de plus forte descente.

Proposition 4.5.1 *Soit $\overline{g}$ la moyenne de toutes les composante g_τ. Il existe une solution optimale d tel que $d_\tau = 0$ si $g_\tau \geq \overline{g}$ et $\gamma_\tau = 0$*

Preuve:
Soit Q l'ensemble de toutes les périodes τ ; i.e. $H = |Q|$. Soit d une direction améliorante optimale telle qu'il existe une période τ^* où $g_{\tau^*} \geq \overline{g}$ et $\gamma_\tau = 0$. Partant de cette solution, nous construisons une autre solution réalisable d' où $d'_{\tau^*} = 0$ et $\sum_\tau g_\tau d_\tau \geq \sum_\tau g_\tau d'_\tau$. En retirant la valeur τ^* de Q, en mettant à jour $\overline{g}$, en remplaçant d par d' et en répétant ces opérations, on obtient un processus durant lequel $\overline{g}$ décroît. A la fin du processus, nous sommes en présence d'une solution améliorante d tel que $d_\tau = 0$ si $g_\tau \geq \overline{g}$ et $r_\tau = 0$.

Sans perte de généralité, supposons que $\tau^* = 1$. Considérons l'équation suivante :

$$d'_\tau = c\left(d_\tau + \frac{d_1}{H-1}\right), \forall \tau \neq 1 \text{ et } d'_\tau = 0$$

où c est un nombre non négatif tel que $\sum_\tau (d'_\tau)^2 = 1$. La faisabilité de d' est évidente si $\gamma_\tau < 1, \forall \tau$ et $d_1 \geq 0$. Dans le cas d'un $\gamma_{\tau^\circ} = 1$, comme $\gamma_\tau = 0 \forall \tau \neq \tau^\circ$, par construction $d_\tau \geq 0, d'_\tau \geq 0 \forall \tau \neq \tau^\circ$ et $d'_{\tau^\circ} = -\sum_{\tau \neq \tau^\circ} d'_\tau \leq 0$ ce qui prouve la faisabilité de d'. De plus

$$\begin{aligned} 1/c^2 &= \sum_{\tau \neq 1} \left(d_\tau + \frac{d_1}{H-1}\right)^2 \\ &= \sum_{\tau \neq 1} d_\tau^2 + \frac{2d_1 \sum_{\tau \neq 1} d_\tau}{H-1} + \frac{d_1^2}{H-1} \\ &= \sum_{\tau \neq 1} d_\tau^2 - \frac{d_1^2}{H-1} \\ &\leq 1 \end{aligned}$$

il suit donc que $c \geq 1$. De plus comme $g_1 \geq \overline{g}$:

$$\begin{aligned}
\sum_{\tau} g_\tau d'_\tau &= \sum_{\tau \neq 1} g_\tau c \left(d_\tau + \frac{d_1}{H-1} \right) \\
&= c \left(\sum_{\tau \neq 1} g_\tau d_\tau + \frac{\sum_{\tau \neq 1} g_\tau}{H-1} d_1 \right) \\
&\leq c \left(\sum_{\tau \neq 1} g_\tau d_\tau + \overline{g} d_1 \right) \\
&\leq c \left(\sum_{\tau \neq 1} g_\tau d_\tau + g_1 d_1 \right) \\
&\leq c \sum_{\tau} g_\tau d_\tau \\
&\leq \sum_{\tau} g_\tau d_\tau
\end{aligned}$$

où la relation due à l'optimalité de d, $\sum_\tau g_\tau d_\tau \leq 0$ est utilisée. Il s'ensuit que d' est meilleur que d donc d' est également une solution optimale.

□

Proposition 4.5.2 *Si $\gamma_\tau > 0$ pour tout τ tel que $g_\tau \geq \overline{g}$ alors la direction de plus forte descente est :*

$$d_\tau = c(\overline{g} - g_\tau)$$

où c est le facteur non négatif tel que $\sum_\tau d_\tau^2 = 1$.

Preuve:
Considérons la fonction de Langrange suivante :

$$L(\lambda, d) = \sum_{\tau} g_\tau d_\tau + \lambda \left(\sum_{\tau} d_\tau^2 - 1 \right)$$

avec $\lambda > 0$. Comme $\sum_\tau d_\tau = 0$,

$$L(\lambda, d) = \sum_{\tau \neq 1} (g_\tau - g_1) d_\tau + \lambda \left(\sum_{\tau \neq 1} d_\tau^2 + \left(\sum_{\tau \neq 1} \right)^2 - 1 \right)$$

La solution maximale de la fonction de Langrange est déterminée comme suit :

$$\begin{aligned}
\frac{\partial}{\partial d_\tau} L(\lambda, d) &= g_\tau - g_1 + 2\lambda \left(d_\tau + \sum_{\tau \neq 1} d_\tau \right) \\
&= g_\tau - g_1 + 2\lambda (d_\tau - d_1) \\
&= 0
\end{aligned}$$

et donc

$$d_\tau = \frac{1}{2\lambda} (g_1 - g_\tau) + d_1 \tag{4.27}$$

En sommant sur cette relation pour tout $\tau \neq 1$ et en utilisant le fait que $\sum_\tau d_\tau = 0$, il vient que

$$\begin{aligned}
-d_1 &= \frac{1}{2\lambda} \sum_{\tau \neq 1} (g_1 - g_\tau) + (H-1) d_1 \\
&= \frac{1}{2\lambda} H (g_1 - \overline{g}) + (H-1) d_1
\end{aligned}$$

et donc

$$d_1 = \frac{1}{2\lambda}(\overline{g} - g_1)$$

En substituant ce résultat dans l'équation (4.27) on a alors

$$d_\tau = \frac{1}{2\lambda}(\overline{g} - g_\tau)$$

La valeur de λ est déterminée tel que $\sum_\tau d_\tau^2 = 1$. Si la solution résultante respecte toutes les contraintes de direction améliorante admissible, alors c'est la plus grande direction améliorante. La faisabilité de d est évidente dans le cas où $\gamma_\tau < 1$ pour tout τ. Dans le cas où $\gamma_{\tau^\circ} = 1$ pour un τ° quelconque, comme $\gamma_\tau = 0, \forall \tau \neq \tau^\circ$, par construction, $d_\tau \geq 0, \forall \tau \neq \tau^\circ$ et $d_{\tau^\circ} = -\sum_{\tau \neq \tau^\circ} d_\tau \leq 0$ ce qui prouve l'admissibilité de d.

□

4.5.2 Algorithme pour la recherche de la clé de répartition

L'algorithme d'analyse de perturbation pour le problème journalier, nommé APCR (Analyse de Perturbation pour la Clé de Répartition), est présenté algorithme 4.8. La base de cet algorithme demeure la même que dans les méthodes de descente de gradient. Nous avons explicité au début de cette section sur l'optimisation de la clé de répartition par analyse de perturbation comment mesurer le gradient. Nous donnons en algorithme 4.9 comment cette mesure se calcule en pratique. La dernière instruction de l'algorithme APCR recalcule une nouveau planning des admissions une fois toutes les m itérations. Le changement de planning est contraire à l'hypothèse formulée dans cette section qui suppose un planning donné lors de la mesure du gradient. Cependant, comme nous l'avons plusieurs fois mentionné, le planning est lié à la clé de répartition. Il apparaît donc comme illogique de modifier la clé par descente de gradient sans mettre à jour le planning. Nous pouvons relativiser la violation de cette hypothèse en considérant le fait que malgré le calcul d'un nouveau planning, l'instance considérée, i.e. le problème, reste la même.

Algorithme 4.8 : APCR

Input : Etat d'occupation, Planning d'admission initial

```
γ^(0) ← Définir une clé initiale;
for k=1 to n do
    Définir un chemin nominal par simulation;
    g ← Calculer les composantes du gradient;
    d ← meilleure_direction_ameliorante(g, γ^(k));
    γ^(k+1) ← γ^(k) + a_k * d;
    if k mod m == 0 then
        Calcul du nouveau planning des admissions;
    end
end
```

4.6 Comparaison des approches et performances

Dans cette section consacrée aux expérimentations numériques sur les différents algorithmes proposés tout au long de ce chapitre, nous allons nous focaliser dans un premier temps sur le paramétrage de l'algorithme SPSA ainsi qu'à une étude expérimentale de sa convergence. Nous réaliserons ensuite un comparatif des différents algorithmes proposés (PSCRG1 et 2, PSCRJ, APCR).

Algorithme 4.9 : Mesure du gradient par analyse de perturbation

```
1  foreach t do
2  |  Calculer le nombre de place disponible à l'instant t du chemin nominal;
3  |  Calculer Δ_t^+ et Δ_t^- d'après l'équation 4.21;
4  end
5  foreach composante τ de la clé γ do
6  |  foreach patient régulier j do
7  |  |  Δ ← Somme des variations de dépassement de capacité tel que définit équation 4.20;
8  |  |  g_τ ← g_τ + Δ;
9  |  end
   |  /* Ajout de coûts moyen de planning                                    */
10 |  g_τ ← g_τ + N^r C_τ;
11 end
12 return g;
```

4.6.1 Étude expérimentale de la convergence des approches

Pour chacun des algorithmes proposés dans ce chapitre, une mesure de la performance de la clé, $\gamma^{(k)}$ de l'itération k sera effectuée dans le but de montrer qu'au cours des itérations, la méthode considérée converge bien vers un minimum. Notons qu'afin de conserver des temps de calculs raisonnables, la mesure de performance effectuée tout au long des itérations est volontairement imprécise[3]. Il en résultera dans les courbes de convergences présentées dans cette sous-section une oscillation autour de la valeur asymptotique.

Paramétrage de SPSA

La méthode SPSA présente de nombreux paramètres. Certains sont propres à la méthode comme le choix de la génération du vecteur de perturbation, les autres sont dus aux méthodes de descente de gradient et plus spécifiquement à la définition de la séquence de gain. Un guide d'implémentation de la méthode SPSA est donné dans Spall (1998a). Il est notamment précisé que l'utilisation d'une distribution de Bernoulli de paramètre *0,5* et à valeur dans ± 1 satisfait aux exigences théoriques pour la génération du vecteur de perturbation. En se basant sur différentes études menées sur les méthodes de gradient, il est ensuite détaillé comment choisir les paramètres a, c, A, α, β utilisés dans le calcul de la séquence de gains. En général, les indications fournies indiquent au mieux un intervalle dans lequel doit se trouver la valeur du paramètre. Le choix de la valeur exacte à donner à ce paramètre – plutôt vers la borne inférieure ? la borne supérieure ? entre les deux ? – reste donc à faire.

Afin de paramétrer la méthode SPSA, nous partons d'un cas simple à mettre en oeuvre sur lequel nous allons pouvoir expérimenter. Une instance du problème de planification des admissions (occupation de l'hôpital+liste de patient à admettre) est générée. En partant de cette instance, nous allons tester différents jeux de paramètres et observer l'évolution des solutions proposées au fil des itérations. L'évaluation pour les clé perturbées est faite à l'aide de la méthode d'évaluation du problème de recherche journalier (calcul d'un nouveau planning puis évaluation par simulation de Monte Carlo). Dans la suite de cette sous-section consacrée au paramétrage de la méthode SPSA, chaque jeux de paramètres sera testé cinq fois sur 1000 itérations de l'algorithme afin de tenir de la variabilité dans les directions prises lors des approximations stochastiques.

3. Par exemple, pour une évaluation par simulation, un minimum de 2000 réplications est nécessaire afin de faire face à la variabilité du problème. Nous ne faisons dans cette étude expérimentale que 100 réplications.

Malgré les précieuses informations données par Spall sur le choix des paramètres, on se retrouve avec une fourchette estimative dans laquelle doivent se situer chacun des termes a, c, A, α, β. Une mauvaise combinaison de ces paramètres peut facilement mener à des résultats tels que ceux présentés figure 4.4 où sur les cinq tests, aucun n'a convergé.

Le choix de bonnes valeurs pour ces paramètres est donc important afin d'obtenir une bonne convergence de la méthode vers un minimum local. En considérant la figure 4.5, on se rend compte que bien que partant du même point de départ (même instance et même clé initiale), le choix des paramètres a mené à une convergence vers deux minimums locaux. Finalement, un choix plus fin semble mener vers la même solution pour chaque réplication (Fig. 4.6). Nous garderons ces paramètres pour la suite des expérimentations comme étant des paramètres permettant a priori de converger vers une unique solution. Remarquons que nous n'avons pas d'information sur la qualité de cette solution vis-a-vis du minimum global. Cependant, nous espérons que la solution atteinte est suffisamment bonne pour permettre une amélioration de la clé de répartition. Nous nous intéresserons plus particulièrement à cette question de performance à la fin de cette section.

Pour terminer cette étude de paramétrage, nous avons chercher quel semblait être le nombre d'itérations nécessaires avant d'atteindre la valeur asymptotique. Nous avons donc observé la convergence de la méthode sur 10000 itérations (Fig. 4.7). Nous constatons que la convergence est obtenue au bout d'environ 3000 itérations. Cependant, il est intéressant de voir que dès 1000 itérations la solution est déjà proche de la valeur asymptotique.

Convergence des approches

Nous venons de voir lors du paramétrage de la méthode SPSA que cette dernière converge bien pour lors de la recherche de la meilleure clé journalière (algorithme PSCRJ).

Intéressons nous à présent à l'algorithme PSCRG1 dont l'étude expérimentale de convergence pour un hôpital de type 1 est présentée figure 4.8 et pour les hôpitaux de type 2 et 3 en figure 4.9.

De manière générale, pour les trois courbes, on constate que lors des 100 premières itérations, la décroissance a lieu par palier avant d'atteindre la valeur asymptotique aux alentours de la 150^{eme} itération.

L'étude de convergence de la méthode PSCRG2, figure 4.10 et 4.11, révèle qu'après quelques itérations où la méthode semble "stagner" la décroissance est brutale avant de se stabiliser comme pour l'algorithme PSCRG1 aux alentours de la 150^{eme} itération.

Finalement, la figure 4.12 présente la convergence de l'approche se basant sur l'analyse de perturbation. L'allure de la courbe est beaucoup plus conventionnelle que les précédentes avec une décroissance régulière de la valeur de l'objectif. Cette différence est cependant certainement due au fait que l'on considère ici le problème d'optimisation journalier de la clé de répartition et non le problème global. A titre de comparaison, avec l'approche de perturbation simultanée, il est préférable de comparer cette courbe avec celle présentée en figure 4.7.

4.6.2 Performance des approches

La convergence de nos approches vers un optimum local étant établie, nous allons à présent nous intéresser aux performances de l'optimisation de la clé de répartition sur la politique d'admission des patients. Pour les méthodes de recherche d'une clé globale (PSCRG1 et PSCRG2), la clé solution fournie par les deux algorithmes correspondants est ensuite utilisée dans la simulation d'évaluation des performances de la politique de choix des dates d'admission des patients électifs. Pour les méthodes d'optimisation journalières (PSCRJ et APCR), la recherche de la clé pour l'instance courante est effectuée préalablement au calcul du nouveau planning des admissions.

Nous menons nos investigations sur les trois types d'hôpitaux que nous avons définis dans le chapitre

précédent. Afin d'étudier également plus en détail le comportement de chaque approche, nous choisissons de partir de trois clés initiales :

1. Une clé équiprobable où chaque patient à une probabilité égale d'être affecté sur n'importe quel jour de l'horizon d'admission ; i.e. $\gamma_\tau = 1/H, \forall \tau$
2. Une clé dénaturée où tous les patients sont affectés le plus tard possible ; i.e. $\gamma_\tau = 0, \forall \tau \neq H$ et $\gamma_H = 1$
3. La clé empirique utilisée dans le chapitre précédent comme étant représentative, à notre sens, de la structure de coûts d'admission

Les tableaux 4.1, 4.2 et 4.3 représentent respectivement pour chacun des types d'hôpitaux l'évaluation de chaque approche d'optimisation de la clé de répartition en utilisant comme point de départ une des clés sus-citées. Plus précisément, pour chacune des clé initiales que nous testons, nous donnons les performances obtenues si cette clé était utilisée par la méthode d'optimisation de Monte Carlo lors de la phase de planification des admissions. Nous comparons ensuite cette performance avec celle obtenue avec une clé optimisée. En particulier, nous nous intéressons au pourcentage de gain obtenu, au nombre de jours en dépassement de capacité (colonne *Day over*), au nombre moyen de patients au dessus de la capacité théorique pour les journées en dépassement de capacité (colonne *Avg over*) et aux coûts moyens de planning des admissions effectuées (colonne *Costs*).

Pour l'analyse des résultats, nous allons détailler les performances constatées pour chaque type d'établissement avant de synthétiser et de conclure approche par approche.

Pour l'hôpital de type 1, qui est représentatif du flux du Centre Hospitalier de Firminy, dont les résultats sont exposés table 4.1, nous pouvons en premier lieu constater l'échec de la méthode PSCRG2 qui dans deux cas sur trois n'a pas amélioré la solution initiale. Le seul cas d'amélioration est constaté en partant d'une clé très mauvaise, la clé dénaturée. Cependant, le résultat atteint n'est pas meilleur que les résultats de base des deux autres clés initiales. Toutes les autres approches ont amélioré le résultat de départ. On peut remarquer dans notre cas d'étude que les améliorations ont porté principalement sur la diminution des coûts de planning. Les améliorations sur les dépassements de capacité sont plus mitigées. Parfois, le nombre de jour en dépassement de capacité augmente mais est compensé par la diminution du nombre moyen de patients au-delà de la capacité théorique. Parfois, le scénario inverse est observé. La méthode par analyse de perturbation, APCR, a fourni le meilleur résultat du point de vue de l'objectif. Cependant, cette approche se révèle pour cet hôpital être très sensible à la clé initiale. En effet, le résultat obtenu en partant de la clé dénaturée est très peu probant. Pour finir avec ce type d'établissement, nous notons la constance et la relative bonne performance de la clé globale fournie par la méthode PSCRG1.

Des résultats similaires sont observables pour l'hôpital de type 2 où les flux urgents et électifs sont équilibrés (Tab. 4.2). Nous pouvons cependant signaler les meilleurs résultats obtenus, par rapport aux précédents, par l'approche PSCRG2. Plus généralement, les deux algorithmes de recherche d'une clé globale, PSCRG1 et 2, ont tous deux fournis des résultats de bonne qualité et du même niveaux de performance indépendamment de la clé initiale. Signalons que comme précédemment, la méthode APCR a fourni le meilleur résultat même si cette performance n'est pas stable en fonction de la clé initiale.

Les résultats pour l'hôpital de type 3 (Tab. 4.3) suivent globalement la même tendance. La méthode PSCRG2 bien que constante dans ces résultats en termes d'objectif n'a amélioré que la solution liée au pire des cas initiaux. Nous noterons cette fois-ci la très bonne performance affichée par la recherche journalière de la clé par perturbation simultanée. Cependant, et comme pour APCR, les performances sont dépendantes de la clé initiale.

La première conclusion que l'on peut tirer de ces expérimentations est l'échec de l'approche PSCRG2. Rappelons que cette approche considérait une clé de répartition globale et commune pour chaque jour de la semaine. Dans le cas concret considéré dans nos expérimentations, il n'y a pas de phénomènes notables dans le flux de patients qui seraient liés aux jours de la semaine ; le taux d'arrivée des patients

TABLE 4.1 – Performance des approches pour l'hôpital de type 1

	Obj	Gain (%)	Day over	Avg over	Costs
Clé empirique	64,47	-	3,37	5,00	0,75
PSCRJ	64,00	0,74	3,73	4,72	0,68
PSCRG1	56,94	11,68	2,85	4,85	0,75
PSCRG2	68,49	-6,22	3,78	5,04	0,69
APCR	54,41	15,61	3,44	4,96	0,39
Clé équi-répartie	69,97	-	3,88	5,00	0,71
PSCRJ	63,32	9,50	4,15	4,94	0,40
PSCRG1	58,31	16,67	2,85	4,94	0,77
PSCRG2	70,26	-0,41	3,58	5,21	0,79
APCR	62,07	11,28	3,83	5,06	0,45
Clé dénaturée	109,31	-	5,60	5,89	0,89
PSCRJ	101,92	6,76	5,75	5,31	0,63
PSCRG1	57,10	47,77	2,69	5,01	0,78
PSCRG2	66,33	39,32	3,48	4,91	0,79
APCR	103,03	5,75	5,61	5,62	0,81

est constant. Cela peut permettre de formuler une explication sur les résultats obtenus par l'algorithme PSCRG2. Avoir une clé spécifique pour chaque jour de la semaine peut avoir un intérêt si les instances des différents jours sont fortement typées. Clairement, ce n'est pas le cas dans notre cas d'étude. Les instances de n'importe quel jour de la semaine sont similaires à celle des autres jours. Pour aller plus loin, il serait particulièrement intéressant de relever des données d'un service dont l'activité serait dépendante du jour de la semaine, et ce, afin de tester plus efficacement cette approche.

La deuxième approche de recherche d'un clé globale a donné de bon résultat. Le caractère important que l'on peut relever est sa stabilité quelque soit la clé initialement fournie. Cependant, la définition d'une clé représentant le mieux un cas moyen ne semble pas être suffisante pour faire face aux trop nombreuses variations pouvant exister dans l'arrivée des patients.

La force dans la stabilité des résultats proposés ne se retrouve pas dans les approches de résolution de la version journalière du problème de recherche de clé optimale. Néanmoins, il semble, que moyennant une bonne solution initiale, ces méthodes soient les plus efficaces pour tenir compte des aléas pouvant survenir. En comparaison, par rapport à la méthode APCR, la méthode PSCRJ de recherche d'une clé journalière par perturbation simultanée, n'a pas été aussi performante. Il semblerait que, pour deux méthodes résolvant le même problème, l'une soit plus robuste que l'autre. Nous pouvons ajouter les démarches théoriques que nous avons pu formuler pour la méthode APCR. En effet, l'algorithme de recherche de meilleure direction améliorante est sans aucun doute une force supplémentaire que nous avons pu inculquer à l'algorithme de base. La projection définie pour la méthode de perturbation simultanée ne garantit en effet pas d'aller toujours vers la meilleure solution. En fait en pratique, il se trouve même que dans les cas à la frontière de l'espace des solutions, l'algorithme doit mettre plusieurs itérations avant d'obtenir une direction de gradient permettant de sortir de ce point. Cela est dû au caractère aléatoire de la perturbation de la solution courante et par extension au caractère aléatoire du gradient.

Finalement, nous pouvons dresser un petit comparatif des méthodes en fonction de leur rapidité. Même si le temps passé à rechercher une clé globale peut raisonnablement être plus long que le temps passé à la recherche d'une clé journalière, le rapport entre les deux doit rester correct. Dans cette optique, l'optimisation par analyse de perturbation présente un avantage certain. En effet, toutes les méthodes par analyse de perturbation nécessitent à chaque itération deux évaluations de la fonction objectif. Cette éval-

TABLE 4.2 – Performance des approches pour l'hôpital de type 2

	Obj	Gain (%)	Day over	Avg over	Costs
Clé empirique	33,98	-	0,96	3,92	0,82
PSCRJ	31,60	7,00	0,80	3,65	0,81
PSCRG1	30,59	9,98	1,15	3,77	0,66
PSCRG2	33,54	1,28	0,55	3,59	0,95
APCR	22,02	35,18	0,97	3,80	0,43
Clé équi-répartie	32,70	-	1,20	3,75	0,72
PSCRJ	24,50	25,07	1,62	3,87	0,30
PSCRG1	30,31	7,30	1,12	4,03	0,63
PSCRG2	33,50	-2,47	0,50	3,79	0,96
APCR	28,84	11,80	1,28	3,98	0,54
Clé dénaturée	57,70	-	2,87	4,21	0,91
PSCRJ	62,23	-7,86	3,72	4,35	0,73
PSCRG1	31,04	46,21	0,81	3,75	0,78
PSCRG2	33,75	41,51	0,58	3,61	0,95
APCR	54,92	4,82	2,42	4,23	0,97

uation nécessite une ou plusieurs optimisations ce qui, multiplié par le nombre d'itérations de l'algorithme peut rapidement être déconcertant. L'approche APCR est basée sur un algorithme itératif nécessitant seulement de temps en temps une optimisation. En termes de temps de calculs, la méthode APCR fournit une solution en moins d'une minute. La méthode PSCRJ nécessite 3 optimisations par itération (2 pour l'évaluation du gradient + 1 pour le calcul du nouveau planning suite à l'obtention d'une nouvelle clé de répartition) ce qui à pour effet d'allonger le temps de calcul à 30 minutes avant d'obtenir une réponse pour le problème journalier. L'impact de la démultiplication du nombre d'optimisation requise est encore plus grand pour les approches globales qui se base sur des répetitions de simulation nécessition elle même maintes phase d'optimisation du planning d'admission des patients. L'approche PSCRG1 fournit une réponse au bout de quelques jours quand presque une semaine est nécessaire à la méthode PSCRG2.

4.7 Conclusion

Une clé de répartition des patients électifs était l'une des données utilisées pour permettre, lors de la planification des admissions de patients réguliers, d'estimer les demandes futures en besoin de lits d'hospitalisation par les patients encore inconnus. Cette clé est une donnée centrale dont vont dépendre certaines politiques de choix des admissions. C'est le cas de l'approche de planification se basant sur l'optimisation de Monte Carlo. Il était donc important de chercher à trouver la clé permettant d'obtenir les meilleurs résultats possibles pour le planning d'admission des patients.

Nous avons vu au travers de ce chapitre que la recherche de cette clé n'est pas un problème évident à résoudre. En premier lieu l'interdépendance entre l'estimation des besoins futurs et le mode d'admission inhérent à la politique choisie nous a forcé à proposer des alternatives à des méthodes de recherche locale, du type descente de gradient, issues de la littérature. En second lieu, le faible nombre d'alternatives existantes permettant de pallier aux difficultés liées à un espace des solutions sous contraintes nous a obligé à définir des transformations permettant de revenir à chaque itération dans l'espace des solutions. Parmi les algorithmes proposés, il est établi de loin, notamment en terme de temps de calcul, que la méthode se basant sur l'analyse de perturbation est actuellement la meilleure. Cependant cette approche demeure sensible à la solution initialement fournie. On peut alors se demander si l'on sera toujours capable

TABLE 4.3 – Performance des approches pour l'hôpital de type 3

	Obj	Gain (%)	Day over	Avg over	Costs
Clé empirique	64,67	-	1,45	4,45	1,62
PSCRJ	36,52	43,53	1,04	3,76	0,82
PSCRG1	51,51	20,36	1,47	4,58	1,16
PSCRG2	67,22	-3,94	1,55	4,46	1,66
APCR	49,20	23,92	1,56	4,55	1,05
Clé équi-répartie	57,91	-	1,95	4,35	1,22
PSCRJ	49,26	14,94	2,47	4,22	0,67
PSCRG1	53,43	7,73	1,61	4,46	1,17
PSCRG2	71,82	-24,02	1,45	4,52	1,85
APCR	58,28	-0,64	2,03	4,58	1,17
Clé dénaturée	104,82	-	5,16	5,04	1,33
PSCRJ	108,48	-3,49	6,55	4,68	1,04
PSCRG1	52,05	50,35	1,49	4,49	1,18
PSCRG2	68,30	34,85	1,38	4,63	1,75
APCR	109,45	-4,41	4,47	5,30	1,67

de fournir une clé initiale de bonne qualité en entrée de cette méthode. Dans la négative, une méthode de recherche de clé globale tel PSCRG1 serait une bonne alternative de par sa constance dans la qualité de la solution fournie.

Des améliorations sur la recherche de la clé de répartition peuvent très certainement être proposées. Pour commencer d'autres familles de méthode comme les algorithmes génétiques ou des méthodes de recherche locale guidée comme COMPASS (Hong et Nelson, 2006, 2007; Li *et al.*, 2009; Xu *et al.*, 2010), qui s'appuient sur la définition d'un voisinage de recherche le plus prometteur et une règle d'allocation de budget de simulation, auraient pu être mises en oeuvre. La recherche théorique n'est pas en reste pour ce problème. En effet, le problème d'optimisation d'une solution contrainte dans l'espace n'est pas assez traité dans la littérature et la recherche d'alternatives à l'utilisation d'une projection pour revenir dans l'espace des solutions admissibles peut s'avérer être pertinente pour l'avenir.

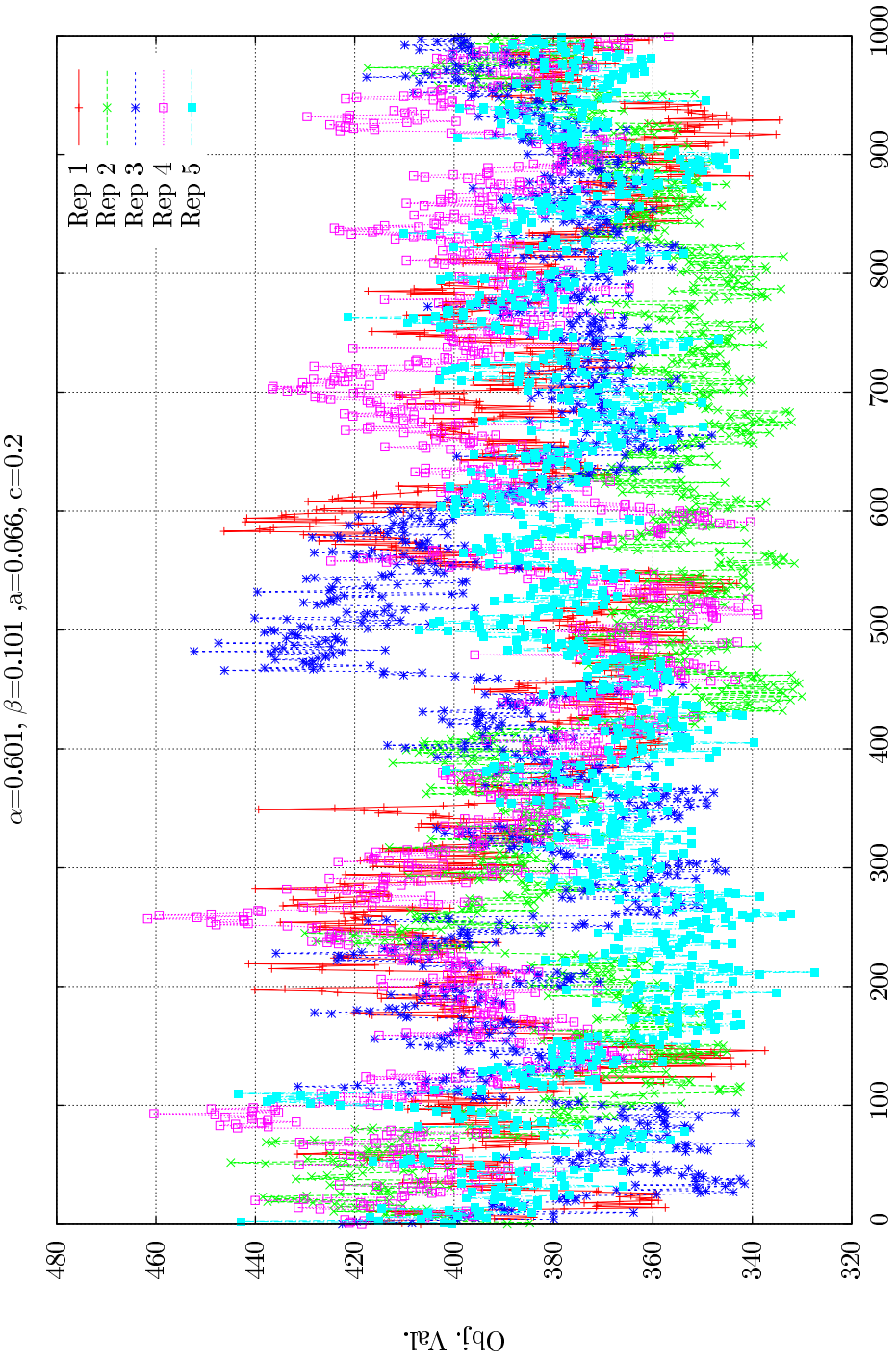

FIGURE 4.4 – Un mauvais jeu de paramètres pour SPSA

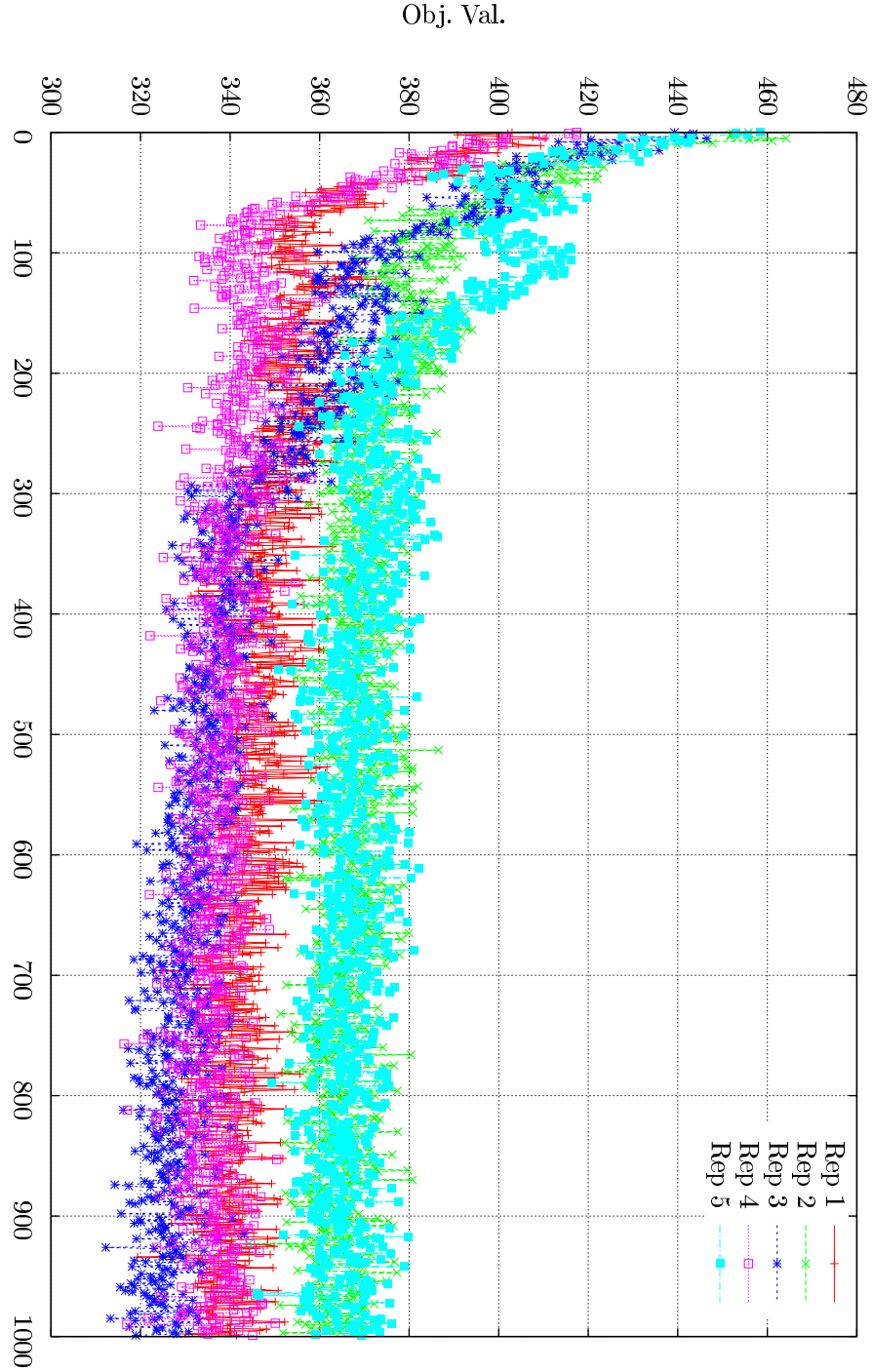

FIGURE 4.5 – Le problème des optima locaux

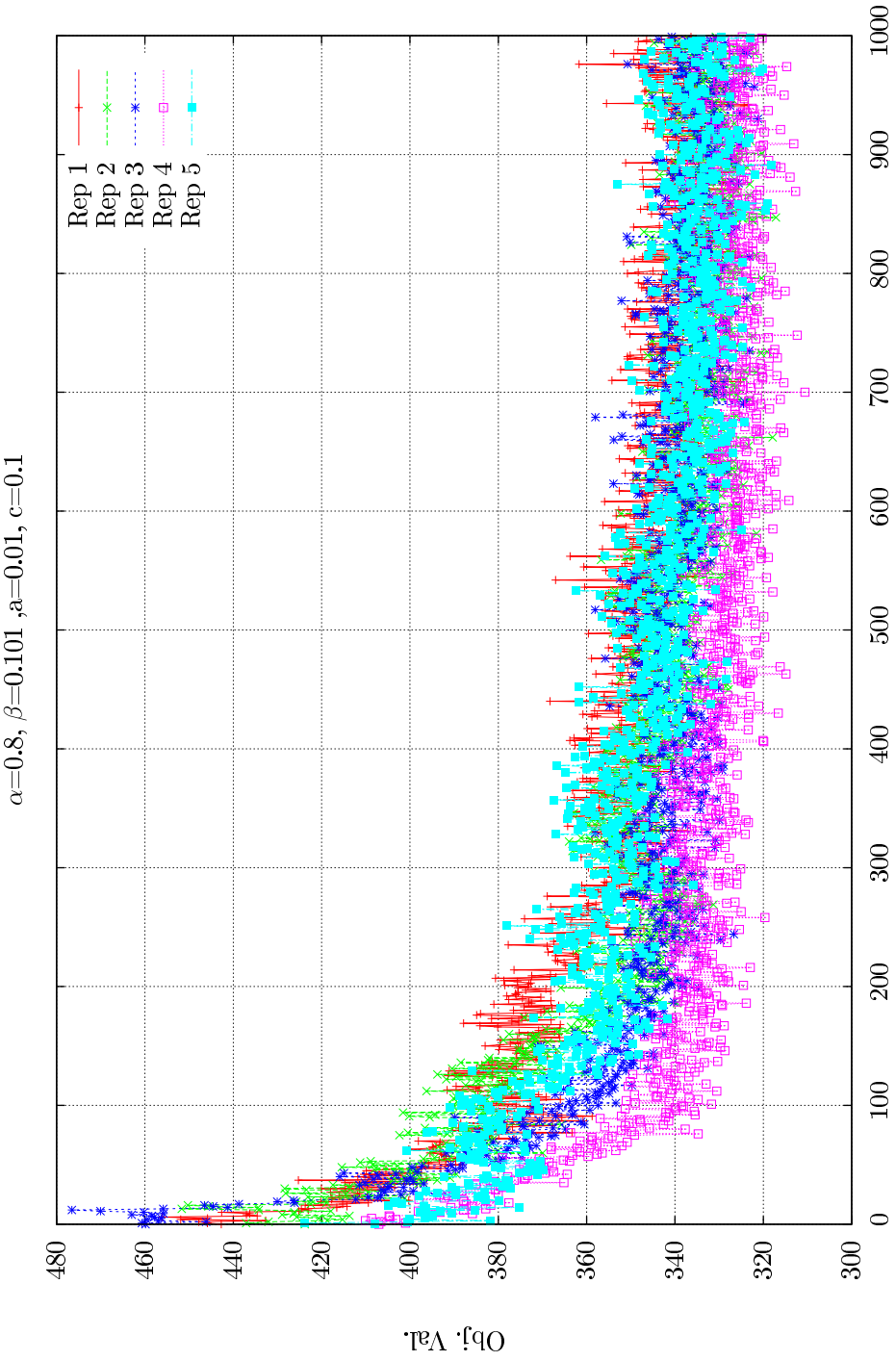

FIGURE 4.6 – Un bon jeu de paramètres pour SPSA

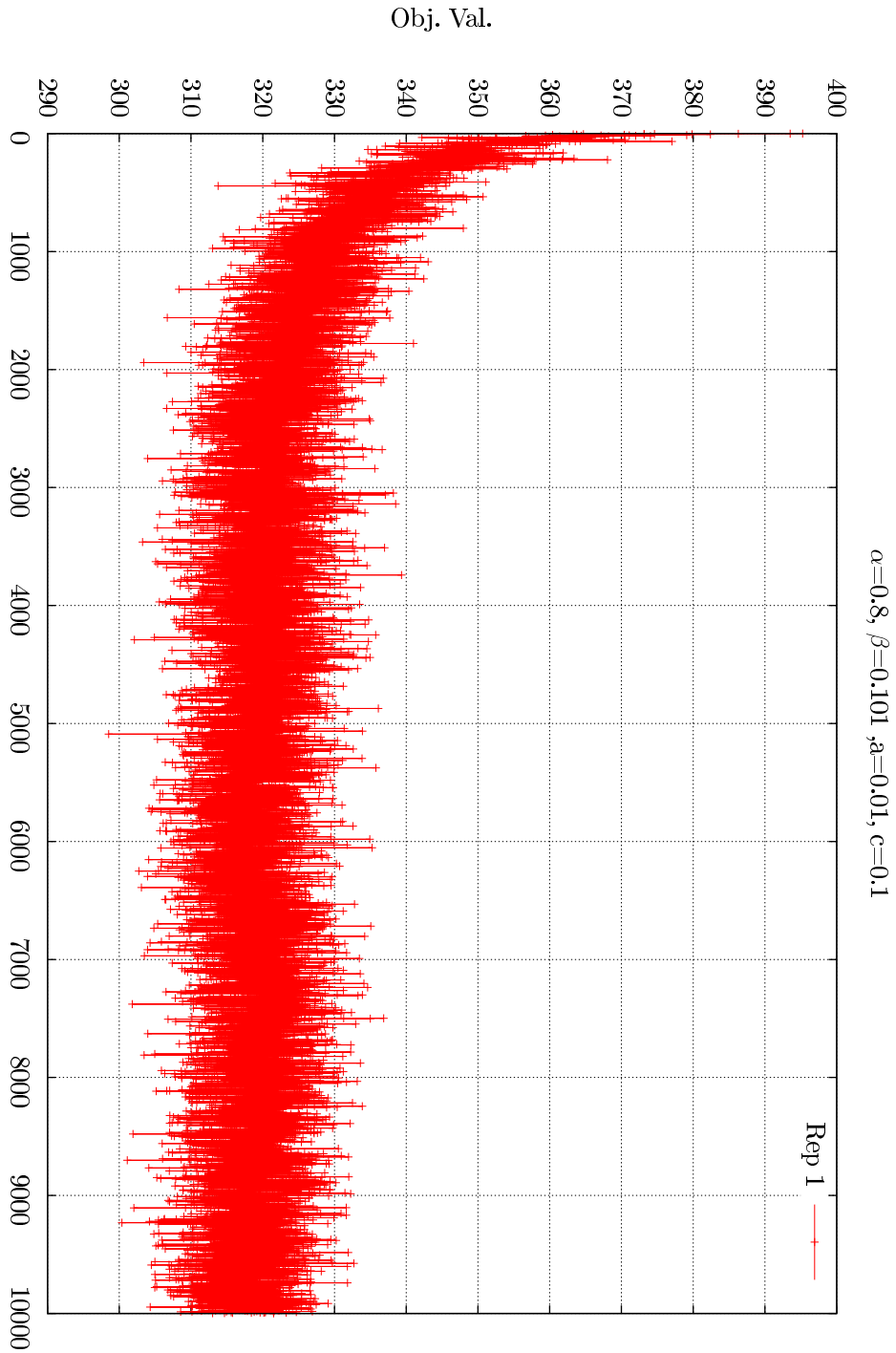

FIGURE 4.7 – 10000 itérations de SPSA

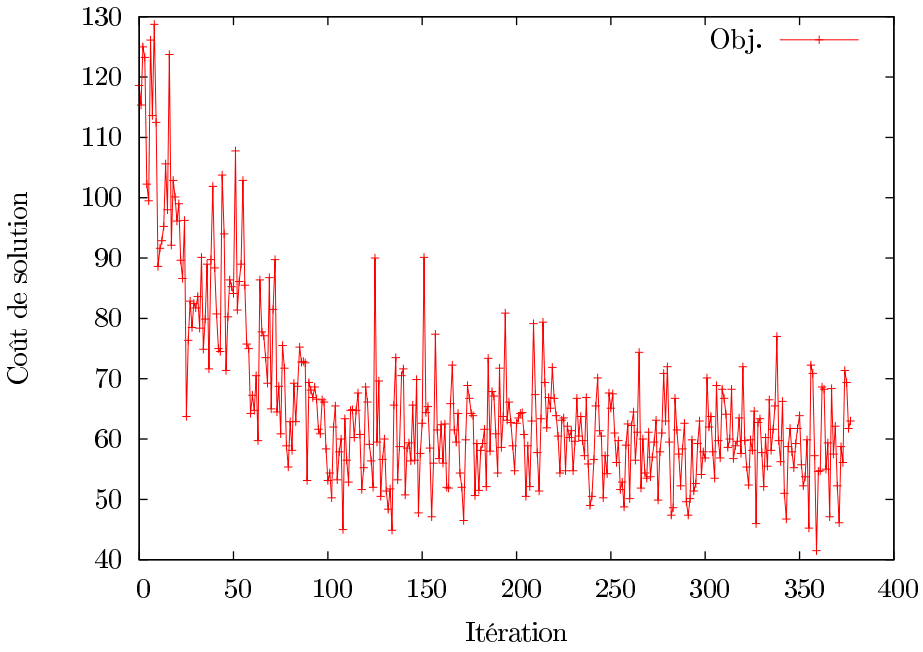

FIGURE 4.8 – Convergence de la méthode PSCRG1 pour l'hôpital de type 1

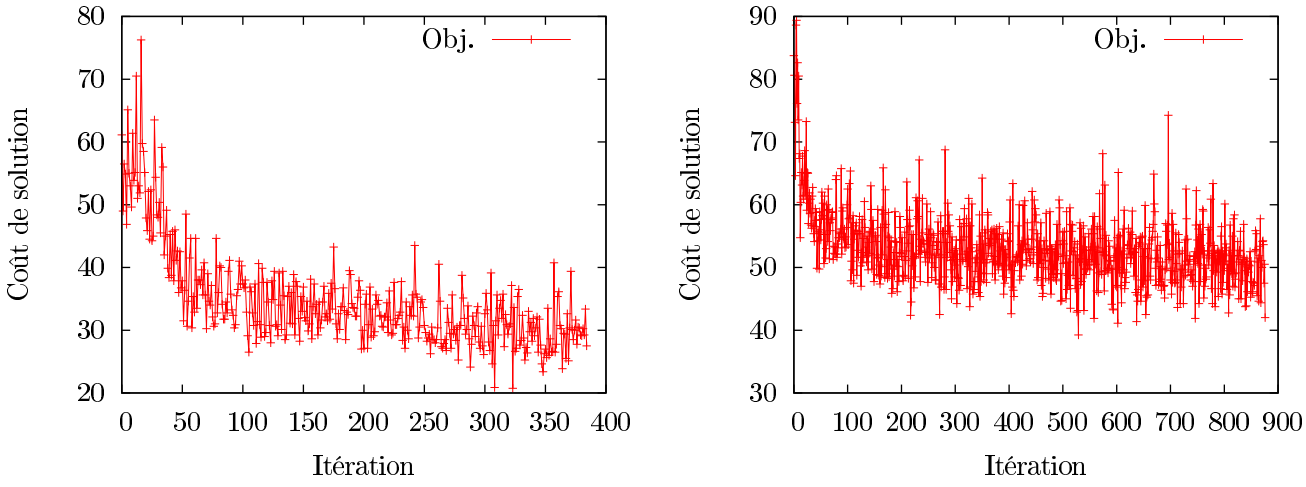

FIGURE 4.9 – Convergences de la méthode PSCRG1 pour les hôpitaux de type 2 et 3

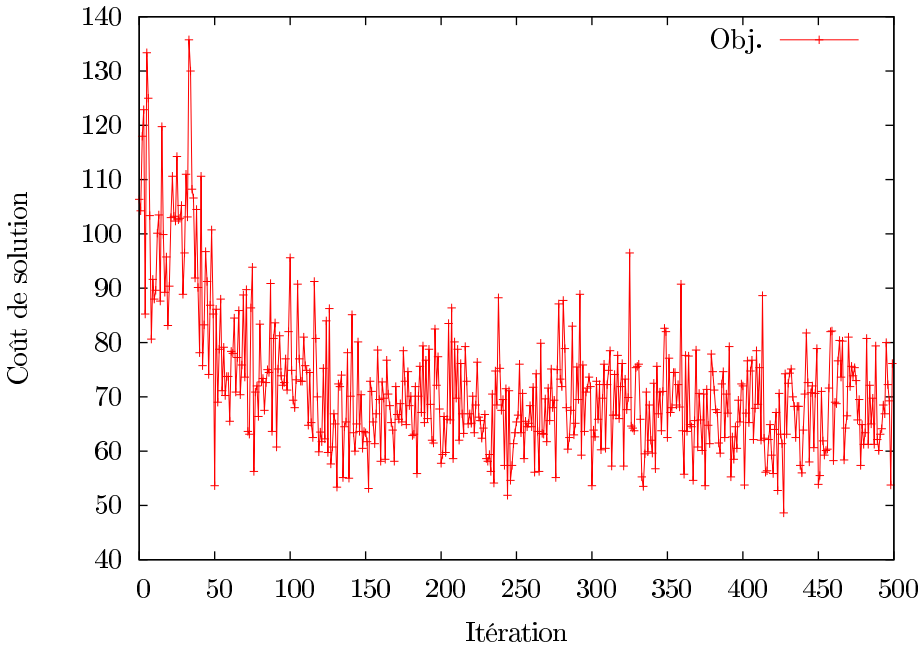

FIGURE 4.10 – Convergence de la méthode PSCRG2 pour l'hôpital de type 1

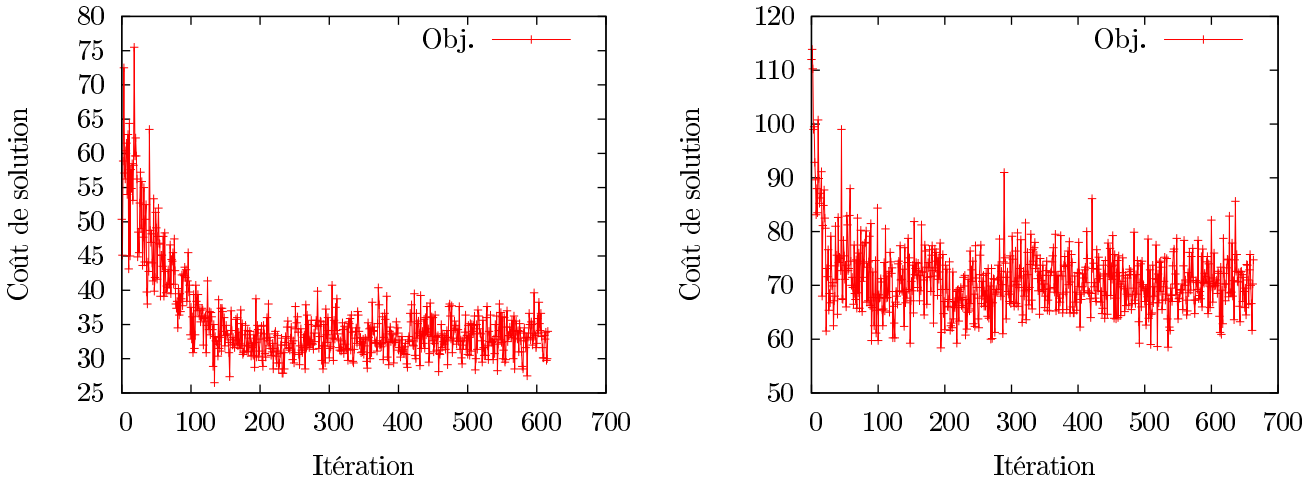

FIGURE 4.11 – Convergences de la méthode PSCRG2 pour les hôpitaux de type 2 et 3

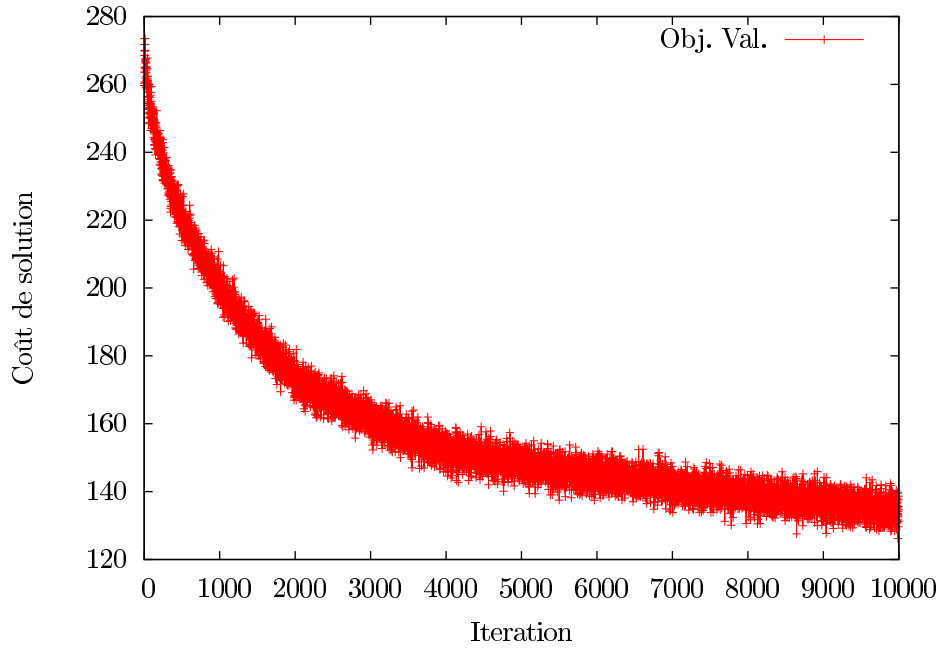

FIGURE 4.12 – Convergence de l'optimisation par analyse de perturbation APCR

Chapitre 5

Affectation des patients aux unités de soins

Sommaire

5.1 Affectation des patients aux unités de soins
5.2 Modélisation
5.3 Affectation des patients aux chambres
5.4 Insertion des urgences
5.5 Expérimentations
5.6 Conclusion

Après avoir traité la planification des admissions des patients réguliers, nous allons maintenant nous intéresser à la gestion du patient, de son affectation dans un lit et donc a fortiori dans une unité de soins. Il est important de séparer les deux problèmes car ils n'ont pas les mêmes objectifs ni les mêmes contraintes. Le problème d'affectation aux unités de soins considère l'hôpital et ses évolutions possibles dans leur globalité. Ainsi l'objectif à rechercher, outre la satisfaction du patient, va être de répartir les demandes afin de gérer, comme précédemment, les dépassements de capacité mais cette fois-ci au niveau de chaque unité de soins. L'affectation du patient à un lit ne va ensuite considérer que le service d'hospitalisation. Les contraintes à prendre en compte sont, dans ce problème, beaucoup plus précises et ne relèvent que de la gestion de l'unité de soins. Le problème dans les deux cas va se résumer à trouver une affectation pour des nouveaux patients même si cela entraîne le déplacement de patients déjà présents. Nous distinguerons, dans chacune de ses deux étapes d'affectation, le problème journalier d'affectation des nouveaux patients électifs et le problème dynamique permettant d'insérer les patients urgents dans le planning existant.

5.1 Affectation des patients aux unités de soins

5.1.1 Introduction

L'étape précédente de notre démarche de gestion du flux de patients a défini un planning des dates d'admission des patients en tenant compte d'une capacité agrégée des lits d'hospitalisation, tous services confondus. En tenant compte des patients présents, des patients dont l'admission est programmée et des

futurs patients inconnus à l'heure de la planification, nous allons à présent nous intéresser au déroulement des séjours des patients.

La gestion de l'affectation des patients à une unité de soins est une étape importante pour les hôpitaux qui mutualisent leurs lits. Par l'intermédiaire de cette gestion, la charge de travail due aux patients se présentant dans l'hôpital va être répartie entre les différents services d'hospitalisation. Cette affectation doit, comme pour le choix des dates d'admission, être réalisée en tenant compte des évolutions de l'occupation des services dans les périodes à venir. C'est également à ce niveau que les patients provenant des urgences sont pris en compte. En effet, pour les patients électifs comme pour les patients urgents, leur admission étant en cours, leur affectation à un lit commence par l'affectation à une unité de soins. Logiquement un patient est affecté dans l'unité où la pathologie dont il souffre relève de la spécialité traitée. Nous dénommons cette unité « *unité nominale* ». Cependant, il peut apparaître judicieux de le placer dans une autre unité si les prévisions le suggèrent.

Nous cherchons au travers de ce problème à connaître le déroulement du séjour des patients. Autrement dit, la solution aux problèmes d'affectation des patients aux unité de soins va nous donner, pour chaque patient, dans quel service et à quel instant il a séjourné. Cette succession d'affectation va découler d'une réponse apportée faisant face aux différentes charges de travail rencontrées et/ou prévisibles.

5.1.2 Description du problème

A chaque instant, nous considérons un état courant de l'hôpital défini par un ensemble de patients présents, un ensemble de patients dont l'admission dans une unité de soins est programmée et un ensemble de patients nouvellement programmés dont la date d'admission a été choisie mais dont l'unité d'affectation n'est pas encore définie. A chaque mise à jour du planning d'affectation des patients aux unités de soins, nous remettons en cause le planning existant. Cela signifie que l'on va choisir non seulement les unités d'accueil des patients programmés mais aussi éventuellement déplacer des patients présents vers une autre unité.

Nous considérons quatre types de patients pour qui la décision à prendre va varier :

- les patients présents et donc déjà affectés
- les patients dont la date d'admission est connue et qui sont déjà pré-affectés
- les patients nouvellement connus qu'il va falloir pré-affecter
- les patients inconnus dont il faut tenir compte

Pour le premier type de patients, deux décisions peuvent être prises. Soit on laisse le patient là où il est affecté, soit on choisit de l'affecter dans un autre service. Bien entendu le choix de l'action à entreprendre, ou non, est régi par un ensemble de contraintes et/ou de préférences. Par exemple, il paraît au premier abord peu logique de déplacer un patient séjournant dans son unité nominale vers une unité autre alors que la réciproque semble intéressante.

Pour les deuxième et troisième type de patients, nous retiendrons leur caractère commun de ne pas être physiquement présent dans l'enceinte de l'hôpital. Il s'agit de patients dont la date d'admission, qui a été définie par la politique d'admission, peut encore être modifiée lors des périodes à venir. Cependant, il serait dommage de ne pas utiliser cette information. En effet, lors du choix de l'affectation des patients à une période donnée, les coûts immédiats sont ceux des affectations des patients déjà présents. Cependant, les charges de travail prévisibles ou connues doivent impérativement être prises en considération au risque de dégrader le planning des périodes à venir suite à de mauvais choix. Pour ces patients, nous emploierons donc le terme de pré-affectation. Leur prise en compte lors de la construction du planning d'affectation ne sera utile que pour guider les choix pour les patients présents. L'affectation résultante pourra être

remise en cause lors de la prochaine phase d'affectation sans induire de pénalités supplémentaires.

Enfin, il paraît, comme pour la planification des admissions, indispensable de prendre en compte les patients inconnus à l'heure de la construction du planning d'affectation mais qui nécessiteront un lit dans le futur. Le turn-over dans les services de court séjour est comme nous l'avons déjà vu tel qu'entre l'entrée et la sortie d'un patient, la majeure partie du planning des patients présents n'a plus rien à voir avec ce qu'il était lors de son entrée. Pour ces patients inconnus et comme précédemment, nous choisissons d'agréger leur demande afin d'obtenir une mesure fiable. En effet, il pourrait parfois apparaître plus pertinent de considérer plus de détails afin de mener la prise en compte de ces patients jusqu'au niveau d'une pré-affectation. Cependant, la variabilité inhérente est telle qu'il est peu probable qu'un patient d'un type particulier soit réellement présent le jour estimé. C'est pour cette raison que nous considérons un niveau d'abstraction plus élevé.

L'objectif que nous nous fixons va être de chercher à minimiser trois types de pénalités :

- Affecter un patient dans un service autre que le service nominal
- Déplacer un patient d'un service vers un autre
- Dépasser la capacité d'accueil d'un service

La première pénalité est liée aux désagréments qui peuvent être causés par ce genre d'affectation. En effet, et en premier lieu, placer un patient dont la pathologie ne relève pas de la spécialité traitée dans l'unité, et donc où le personnel n'est pas formé à ce type de traitement, augmente le risque d'erreurs médicales. En second lieu, le fait que le personnel ne soit pas formé, ou pas habitué, aux protocoles de traitement requis va être consommateur de temps infirmier. D'autres désagréments pour le personnel soignant, comme les tournées des médecins qui doivent alors parcourir plusieurs unités, pourraient être cités. Enfin, côté patient, une telle affectation peut être source de stress et d'une sensation d'insécurité.

La mutation d'un service vers un autre est en général quelque chose de plus contraignant pour le patient. Cependant, cette action est tout de même source d'activités supplémentaires pour le personnel des unités de soins. Même si ce type de tâches fait partie de leurs attributions, elles sont consommatrices de temps qui pourrait être utilisé à d'autres finalités. Notons que dans certains établissements ayant des ressources nécessaires dont la mission est de s'occuper des transferts de patients, une mutation n'est pas forcément pénalisante mis à part pour le confort du patient.

Enfin, nous revenons très brièvement sur les désagréments causés par un service où le nombre de patients est supérieur au nombre de lits théoriques. Dégradation des conditions de travail, augmentation des risques sanitaires, inconfort des patients et difficulté à recevoir de nouveaux patients sont parmi les problèmes soulevés par de telles situations. Les pénalités liées au dépassement de capacité permettent de représenter ces risques.

Le choix de la structure de pénalités est important car de par cette structure, le type de solution proposée va grandement varier. En effet, comme souvent dans les problèmes où l'orientation vers telle ou telle solution est grandement guidée par ce type de structure, il convient d'étudier quelque peu les impacts que peuvent avoir certaines décisions lors de la conception d'un modèle d'optimisation. La première question à se poser concerne bien entendu l'importance relative que doivent avoir les différentes pénalités entre elles.

Dans le cas de notre problème d'affectation, pour un établissement mutualisant ses lits, il est par exemple impensable d'avoir un service en état de dépassement de capacité alors que des places sont disponibles[1] ailleurs. En termes de structure de pénalités, cette supposition peut se traduire en fixant une pénalité de dépassement de capacité supérieure à la somme des coûts de mutation et d'affectation

1. Un lit peut être vacant sans être pour autant disponible. En effet, les pré-affectations des patients programmés dans un futur proche doivent être prises en considération.

dans un service différent du service nominal.

Dans le cas d'un hôpital mutualisant ses lits de manière exceptionnelle, il faut favoriser le retour d'un patient affecté hors de son unité nominale vers celle-ci. On peut dans ce cas de figure imaginer que la pénalité liée au séjour du patient soit négative et ce afin d'annuler la pénalité due au transfert du patient. D'une manière plus générale, il semble intéressant de toujours chercher à ramener les patients dans leur unité nominale. Des aménagements comme celui que nous venons de citer peuvent être proposés. Cependant, le premier avantage menant à affecter un patient dans cette unité restera toujours le coût nul de son séjour au sein de ce service d'hospitalisation.

Le coût d'affectation d'un patient dans un service autre que son service nominal peut varier au cours du temps. Par exemple, en début de soins, le patient est dans une phase aiguë de sa pathologie. Il serait donc préférable de le placer directement dans l'unité où il recevra les meilleurs soins. A l'inverse, en fin de séjour, la technicité des soins est généralement moindre, ou du moins le protocole de soins à appliquer est bien rodé. L'impact de placer ce patient dans une unité quelconque est donc moindre. Selon cette réflexion, nous proposons une structure de pénalités liée à l'affectation dans un service quelconque décroissante au fil du séjour. Notons bien que cette proposition est valable uniquement si l'hypothèse que nous avons formulée se vérifie. Dans d'autres configurations, il serait nécessaire de chercher à formuler différemment la construction de cette structure de pénalités.

Avant de modéliser formellement le problème d'affectation des patients aux unités de soins, intéressons nous à une dernière spécificité due à la capacité d'accueil des services. Chaque service est composé d'un certain nombre de lits. Ce nombre de lit définit la capacité régulière du service. Cependant, comme nous le relatons depuis maintenant quelques chapitres, il arrive qu'exceptionnellement cette capacité soit dépassée pour cause d'afflux de patients. Dans ce cas, les services disposent de marges de manoeuvre et peuvent ajouter des lits dans certaines chambres. Cependant, il demeure un niveau au-delà duquel les ressources disponibles, lits ou soignantes, ne permettent plus de prendre en charge de nouveaux patients. Nous ferons donc la distinction entre la limite théorique de capacité d'un service et sa limite haute au-delà de laquelle nous ne pouvons aller.

5.2 Modélisation

Le problème d'affectation des patients aux unités de soins peut, tout comme le problème de planification des admissions, se formuler sous la forme d'un problème dynamique stochastique de recherche d'une séquence de politiques d'affectation permettant d'obtenir un planning optimal. Une formulation équivalente à celle donnée en section 3.2.1 pourrait être exposée ici. Cependant, les problèmes dynamiques étant relativement proches, nous allons directement nous focaliser sur la description du problème journalier, le problème qui consiste à actualiser le planning d'affectation existant en tenant compte des nouvelles informations disponibles.

5.2.1 Modèle journalier

Nous considérons un ensemble de I patients connus, présents ou dont l'admission est planifiée. Ces patients devront être affectés à une des J unités de soins. Le planning des affectations est calculé en tenant compte des évolutions possibles sur un horizon de H périodes. Chaque unité de soins est définie à chaque instant par une capacité régulière (nombre de lits ouverts), T_{tj}, et un nombre maximum de patients pouvant exceptionnellement être pris en charge, $T_{tj} + T^*_{tj}$ où T^*_{tj} est le nombre maximum de patients supplémentaires pouvant être reçus. Soit Old une matrice représentant l'ancien planning d'affectation des patients, i.e. Old_{ij} vaut 1 si le patient i séjournait jusqu'à présent dans l'unité j. Pour les patients qui ne sont pas encore présents, $Old_{ij} = 0$ quel que soit j. Nous définissons un coût c_{ij} de séjour dans

une unité pour chaque patient. Typiquement ce coût vaut 0 si j est l'unité dans laquelle le patient i doit théoriquement séjourner. Les coûts sont ensuite proportionnels à l'éloignement thématique de l'unité j avec l'unité nominale du patient. Notons qu'en fonction du planning courant, le coût lié à l'unité nominale peut être négatif pour ainsi favoriser le retour du patient dans ce service. Enfin, nous désignons par N_{tj} la variable aléatoire du nombre de patients inconnus à l'instant de planification mais qui seront présents à l'instant t dans l'unité j.
Pour ce modèle, les variables de décisions sont les suivantes :

- $X_{ij} = \begin{cases} 1 & \text{si le patient } i \text{ est affecté à l'unité } j \\ 0 & \text{sinon} \end{cases}$
- $Y_i = \begin{cases} 1 & \text{si le patient } i \text{ a changé de service} \\ 0 & \text{sinon} \end{cases}$
- Δ_{tj} la différence entre la capacité disponible et la capacité requise à l'instant t pour l'unité j
- O_{tj} le coût estimé de dépassement de capacité à l'instant t pour l'unité j

Considérant ces variables de décisions, notre objectif se formule de la manière suivante :

$$\min \sum_{i,j} c_{ij} X_{ij} + \alpha \sum_{i} Y_i + \beta \sum_{j,t} O_{tj} \tag{5.1}$$

s.c.

$$\Delta_{tj} = \sum_{i, a_i \leq t \leq a_i + l_i} X_{ij} + N_{tj} - T_{tj} \qquad \forall t \in H, \forall j \in J \tag{5.2}$$

$$O_{tj} = E\left[g_{tj}(\Delta_{tj})\right] \tag{5.3}$$

$$\sum_{j} X_{ij} = 1 \qquad \forall i \in I \tag{5.4}$$

$$Y_i \geq X_{ij} - Old_{ij} \qquad \forall i \in I, \forall j \in J \tag{5.5}$$

$$\tag{5.6}$$

Où $g_{tj}(\Delta_{tj})$ est une fonction évaluant la pénalité liée au dépassement de capacité dans une unité et définie comme suit :

$$g_{tj}(\Delta_{tj}) = \begin{cases} 0 & \text{si } \Delta_{tj} \leq 0 \\ \alpha_1 \Delta_{tj} & \text{si } 0 < \Delta_{tj} \leq T^*_{tj} \\ \alpha_1 \Delta_{tj} + \alpha_2 (\Delta_{tj} - T^*_{tj}) & \text{si } \Delta_{tj} > T^*_{tj} \end{cases} \tag{5.7}$$

avec $\alpha_1 < \alpha_2$. Cette définition permet de pénaliser plus fortement un dépassement de capacité qui va au-delà du maximum autorisé et ainsi favoriser des solutions respectant cette limite haute.

Dans cette formulation l'objectif (équation (5.1)) cherche à minimiser une somme pondérée des coûts d'affectation des patients, des coûts liés à une mutation d'un patient d'un service vers un autre et des coûts liés aux dépassements de capacité qui peuvent survenir dans les différentes unités. Dans la suite du modèle, l'ensemble de contraintes (5.2) estime le dépassement de capacité de chaque unité en fonction des patients présents, des patients admis au cours de l'horizon considéré et des patients inconnus estimés par la variable N_{tj}. L'ensemble de contraintes (5.3) calcule le coût moyen estimé lié aux dépassements de capacité. L'équation (5.4) assure que chaque patient connu soit affecté à une et une seule unité. Enfin, l'ensemble de contraintes (5.5) permet de définir si un patient est muté lors du calcul du nouveau planning ; i.e. il est affecté à une unité différente de celle où il séjournait.

Ouvrons une petite aparté sur l'évaluation de la fonction objectif et sur les pénalités liées à l'affectation d'un patient hors de son unité nominale. Dans la définition du problème journalier, être dans cette

situation est pénalisé à chaque replanification. Ainsi, nous représentons qu'être dans une unité quelconque n'est pas juste une pénalité prise en compte lors de l'affectation du patient mais. C'est une pénalité de tous les jours.

5.2.2 Optimisation de Monte Carlo

Dans le chapitre 3, l'optimisation de Monte Carlo s'est révélée être la meilleure approche pour prendre en compte les besoins futurs en termes de lits. Nous reprenons cette même approche afin de résoudre le problème journalier d'affectation des patients aux unités de soins. Le concept demeure le même. En générant K échantillons de demandes pour chaque instant et chaque unité, N_{tj}^k, nous approximons ensuite l'espérance mathématique du coût moyen de dépassement de capacité par la moyenne empirique sur le nombre d'échantillons. Nous redéfinissons les variables associées au dépassement de capacité en

- Δ_{tj}^k l'écart à la période t par rapport à la capacité régulière du service j dans l'échantillon k
- O_{tj}^k le coût de dépassement de capacité à la période t du service j pour l'échantillon k

La fonction objectif se réécrit donc comme suit :

$$\min \sum_{i,j} c_{ij} X_{ij} + \alpha \sum_{i} Y_i + \beta \frac{1}{K} \sum_{j,t,k} O_{tj}^k \tag{5.8}$$

L'équation (5.2) devient

$$\Delta_{tj}^k = \sum_{i, a_i \leq t \leq d_i} X_{ij} + N_{tj}^k - T_{tj} \qquad \forall t \in H, \forall j \in J \tag{5.9}$$

et l'équation (5.3) se réécrit sous la forme de deux équations ; l'une pour le dépassement autorisé, l'autre pour les dépassements au-delà du maximum autorisé :

$$O_{tj}^k \geq \alpha_1 \Delta_{tj}^k \qquad \forall t \in H, \forall j \in J \tag{5.10}$$

$$O_{tj}^k \geq \alpha_1 \Delta_{tj}^k + \alpha_2 (\Delta_{tj}^k - T_{tj}^*) \qquad \forall t \in H, \forall j \in J \tag{5.11}$$

Notons que l'ensemble de contraintes (5.11), donnant une valeur à la pénalité de dépassement de la limite haute de capacité du service, n'a en théorie jamais à être franchie. En effet, le dépassement de capacité au-delà de la limite haute suppose que tous les autres services sont saturés également. Ce type d'événement est très rare dans la pratique. Quand il survient, l'hôpital n'a d'autres choix que d'anticiper la sortie d'un patient présent ou bien de refuser une admission en détournant un patient vers un autre hôpital. Si l'évènement devient fréquent, il sera alors nécessaire d'envisager d'autres solutions comme le redimensionnement de l'hôpital. Dans notre modèle, cette contrainte est présente pour garantir la bonne tenue de la méthode de résolution dans le cas où un échantillon mènerait à une telle situation.

5.2.3 Estimations des besoins futurs par unité

Le modèle de résolution par optimisation de Monte Carlo que nous venons de proposer est dépendant de la génération de scénarios de besoins en lits pour chaque unité. Pour le problème de planification des admissions, nous avions fait le choix d'utiliser une clé de répartition des patients sur les jours suivant leur décision d'hospitalisation. Cette clé devait être choisie en adéquation avec la politique d'admission mise en place. Nous conservons cette approche afin de représenter les admissions des patients électifs. Le problème se posant ensuite est de savoir de quelle unité de soins relève la nouvelle demande. Comment représenter dans quel service séjourne le patient ?

Il serait possible de définir une nouvelle clé de répartition des patients dans les unités de soins en accord avec la politique d'affectation que nous voyons dans ce chapitre. La démarche du chapitre précédent

pourrait être mise en oeuvre de manière similaire afin de chercher la meilleure clé de répartition des patients dans les services. Bien entendu, la supposition liée à une telle clé est que chaque patient une fois affecté à un service ne sera plus déplacé.

Une autre approche pourrait être inspirée des résultats proposés dans la littérature par Thomson *et al.* (2009). Des seuils pourraient être définis au-delà desquels les nouveaux patients seraient dirigés vers un service où l'occupation est inférieure à un deuxième seuil. Cette approche présenterait l'intérêt de bien représenter la dynamique des affectations des patients qui peuvent parfois être déplacés afin de libérer des places pour recevoir de nouveaux patients. Cependant, la difficulté est ici dans le choix des seuils à utiliser. Un nouveau problème d'optimisation difficile à résoudre ferait alors son apparition.

Nous proposons une troisième approche qui consiste à considérer que le jeu des consultations et des demandes d'hospitalisation sont suffisamment bien répartis entre les services, en fonction de leur taille, pour permettre aux nouveaux patients d'aller directement dans leur unité nominale. La répartition des nouveaux patients est alors simple à définir. Il suffit en regardant dans l'historique des données la probabilité qu'un nouveau patient aille vers tel service. Afin de tenir compte des différences pouvant exister entre la répartition des patients urgents et la répartition des patients électifs, une clé peut être définie pour chacun de ces types de patients. Sous cet aspect simple, cette clé présente néanmoins un avantage. En choisissant de représenter l'affectation d'une nouvelle demande dans son service nominal, on représente bien lors de la construction du planning d'affectation des patients présents que l'on veut tendre vers un système où chaque nouvelle hospitalisation est affectée là où elle est censée l'être. Les patients se retrouvant affectés en pratique dans une autre unité ne seraient alors que des cas marginaux dus aux difficultés opérationnelles rencontrées.

Nous définissons donc deux clés de répartition des patients dans les unités de soins : η^r pour les patients réguliers et η^u pour les patients provenant des urgences. Chaque composante $\eta^{\cdot}_j$ est donc la probabilité pour un patient d'être affecté dans l'unité de soins j ; i.e. l'unité j est son unité nominale. En couplant l'utilisation de cette clé avec la clé de répartition γ des admissions, nous pouvons alors construire un algorithme (Algo. 5.1) nous permettant d'estimer des séries de valeurs pour la variable aléatoire N_{tj}.

5.3 Affectation des patients aux chambres

Dans le cadre de notre démarche de gestion du flux de patients nous disposons à présent pour chaque unité de soins d'une liste de patients qui y sont affectés. Le problème que nous allons traiter à présent fait suite à cette affectation de patients au différents services et se place dans la continuité du niveau opérationnel de notre approche de gestion du flux de patients.

5.3.1 Description du problème

Le problème d'affectation aux chambres diffère de ceux précédemment rencontrés. Sa résolution est soumise au respect de nombreuses contraintes là où les problèmes précédemment traités ne se devaient que de respecter une contrainte de capacité. Parmi les contraintes à respecter, la non mixité des chambres ou plus généralement la compatibilité entre les patients va se révéler être une contrainte forte. Outre l'incompatibilité des patients entre deux sexes différents, des incompatibilités peuvent concerner des patients ne pouvant séjourner dans la même chambre pour cause de pathologies incompatibles. D'autres incompatibilités peuvent être considérées de manière moins stricte. Par exemple, il est peu souhaitable de voir séjourner ensemble des patients avec un écart d'âge important. Nous pouvons donc voir la compatibilité entre patients de manière stricte(sexe) où de manière préférentielle (âge). Si les contraintes strictes se modélisent facilement dans un programme mathématique, les contraintes exprimant des préférences nécessitent la mise en place, une fois n'est pas coutume, d'un système de pénalités permettant de valoriser le non respect de ces contraintes.

Algorithme 5.1 : Génération de demandes en lits des patients inconnus pour les unités de soins

Soit $list_j$ la liste des patients présents dans l'unité j;
Soit $prog_j$ la liste des patients programmés dans l'unité j **for** $t = 1$ **to** H **do**
  $A \leftarrow$ Nombre de nouveau patients urgent arrivant en t;
  foreach *nouveau patient* i **do**
    $a_i \leftarrow t$;
    $u_i \leftarrow$ Choix d'une unité de soins selon η^u;
    Ajouter i à $list_{u_i}$;
  end
  $A \leftarrow$ Nombre de nouveaux patients électifs connus en t;
  foreach *nouveau patient* i **do**
    $a_i \leftarrow$ Choix de la date d'admission selon γ;
    $u_i \leftarrow$ Choix d'une unité de soins selon η^r;
    Ajouter i à $prog_{u_i}$;
  end
  foreach *unité* j **do**
    retirer de $list_j$ les patients sortant en t;
    ajouter dans $list_j$ les patients programmés en t de $prog_j$;
    $N_{tj} \leftarrow$ Nombre de patients dans $list_j$;
  end
end

Des préférences peuvent également être prononcées dans le choix du type de chambre. Typiquement, un hôpital est composé de chambres simples et de chambres doubles. Un patient peut choisir de vouloir séjourner dans l'un des deux types de chambre. Le personnel soignant peut également, pour des raisons médicales, émettre le souhait d'affecter un patient particulier de préférence dans un type de chambre bien particulier.

Enfin, le dernier type de contraintes à prendre en compte est le respect de la capacité des chambres. Chaque chambre possède une capacité régulière et une capacité maximale dans le cas où l'on y ajoute des lits pour permettre la prise en charge de patients au-delà de la capacité régulière du service.

La minimisation des pénalités liées aux préférences que nous venons d'exposer va entrer en compte dans la recherche d'une solution optimale du planning d'affectation des patients aux chambres des services. Le deuxième élément que nous allons considérer dans la définition de la fonction objectif de ce problème est le nombre de mutations de patients d'une chambre à une autre. En effet, comme pour l'affectation des patients aux unités de soins, ce type d'action provoque un inconfort pour le patient et monopolise des ressources. Cependant, il arrive qu'il soit inévitable de devoir muter un patient. Typiquement, cela survient quand l'occupation du service est proche de sa capacité théorique et que les seuls lits disponibles sont dans une chambre double dont le deuxième lit est déjà occupé. Si le nouveau patient se présentant est du sexe opposé aux occupants de ces chambre, alors il est nécessaire de transférer un patient incompatible dans une autre chambre afin de libérer une place pouvant recevoir le nouveau patient. Notons qu'une autre solution aurait pu être l'ajout d'un lit supplémentaire dans une chambre. Cependant, comme dans le cas des dépassements de capacité de l'unité, on préfèrera déplacer des patients avant d'envisager de mettre en dépassement de capacité une chambre. En pratique pour notre modèle, cela se traduit par une pénalité de dépassement de capacité supérieure à celle de la mutation d'un patient.

Une question se pose. Au moment de l'affectation des patients aux chambres, faut-il ou non prendre en compte les futurs patients pouvant se présenter, qu'ils soient connus ou inconnus ? Si l'on reprend la

définition que nous venons de donner du problème, et surtout de son objectif, alors la réponse est oui. En effet, le problème proposé dans cette section peut se reformuler comme la recherche d'un planning respectant les contraintes de compatibilité des patients et minimisant principalement le nombre de mutations par patient. A ce titre, si nous pouvions connaître tous les futurs patients qui seront admis, alors nous pourrions trouver une affectation bannissant tout changement pendant le séjour. C'est ce qui a été démontré dans Wang *et al.* (2008). En général, il n'est pas possible dans les services de courts séjours généraux d'avoir cette information. Nous pouvons alors être tenté de construire comme dans les approches précédentes une méthode de simulation de Monte Carlo permettant d'estimer les besoins futurs pour des patients inconnus. Cependant, autant dans les problèmes précédents nous pouvions nous contenter d'une valeur agrégée autant dans ce problème d'affectation aux chambre, nous devons considérer les patients en temps que tel. En se rapportant à une unité de soins, le nombre de nouveaux patients se présentant chaque jour est relativement faible. Qui plus est, il est nécessaire d'estimer les arrivées féminines et masculines. Statistiquement, sauf pour des services typés comme la gynécologie, nous pouvons estimer qu'il y a autant de femmes que d'hommes se présentant dans les services de soins. L'estimation deviendrait alors très rapidement imprécise alors que l'ensemble de ces données sont très importantes pour évaluer dans quelle chambre placer quel patient.

Nous choisissons pour ce problème de ne pas prendre en compte une estimation des évolutions futures. Nous nous concentrerons, à chaque replanification des affectations aux chambres, uniquement sur les patients présents ainsi que sur les nouveaux patients à admettre pour la première fois. Cependant nous allons faire en sorte de construire un planning robuste face aux futures évolutions du système. Un planning robuste est, à notre sens, un planning dont les affectations faites ne seront que très peu modifiées lors des phases de replanification à venir. Pour ce faire, nous définissons une notion de stabilité entre patients. Cette notion va nous aider à la construction du planning. Si deux patients sont « stables » alors les placer dans la même chambre contribuera à la robustesse du planning calculé puisqu'il n'y aura a priori pas de raisons de les déplacer par la suite. Le problème va donc intégrer une dimension supplémentaire dans son objectif qui est la stabilité. Plus la somme des couplages effectués entre patients est faible plus le planning sera alors considéré comme efficace et robuste.

La notion de stabilité entre deux patients peut s'estimer assez simplement. Si deux patients ont la même date de sortie estimée, et s'ils sont compatibles, alors en les plaçant dans la même chambre, nous n'aurons plus à toucher à cette affectation durant tout leur séjour. La mesure de stabilité entre deux patients compatibles est alors définie comme étant proportionnelle à la différence de leur date de départ.

5.3.2 Modélisation mathématique

Nous allons formuler le problème d'affectation des patients aux chambres des unités de soins sous la forme d'un programme linéaire en nombres entiers. Dans cette modélisation, nous allons considérer un ensemble I de patients à affecter parmi un ensemble K de chambres. Chaque chambre est définie par sa capacité régulière en lit, T_k et le nombre de lits maximum pouvant être ajoutés, T_k^*. Pour chaque patient i, nous définissons une pénalité c_{ik} d'être affecté dans la chambre k. Cette pénalité représente le non respect des préférences pouvant être formulées aussi bien par le patient lui même que par le personnel soignant. Bien entendu, une chambre répondant aux exigences formulées a une pénalité associée de 0. La compatibilité entre deux patients i et j est indiquée par la donnée a_{ij} qui vaut 1 si les patients sont incompatibles ; 0 s'ils sont compatibles. La mesure de stabilité entre deux patients i et j est dénotée par $f(i,j)$. Nous définissons ici cette mesure de stabilité comme étant l'écart algébrique entre les dates de départ estimées des deux patients plus un terme, ζ_{ij}, permettant de tenir compte d'éventuels autres paramètres de moindre importance régissant la qualité du couplage de deux patients entre eux comme l'âge par exemple.

$$f(i,j) = |d_i - d_j| + \zeta_{ij} \tag{5.12}$$

Enfin, l'ancien planning des affectations est représenté par la donnée Old_{ik} qui vaut 1 si le patient i séjournait jusqu'à présent dans la chambre k.
Les variables de décisions pour ce problème seront les suivantes :

- $X_{ik} = \begin{cases} 1 & \text{si le patient } i \text{ est affecté à la chambre } k \\ 0 & \text{sinon} \end{cases}$
- $Y_i = \begin{cases} 1 & \text{si le patient } i \text{ a changé de chambre} \\ 0 & \text{sinon} \end{cases}$
- S_{ij} la valeur de stabilité entre le patient i et j s'ils sont dans la même chambre
- O_k le nombre de patients au delà de T_k dans la chambre k

A l'aide de ces variables et des données nous pouvons alors formuler l'objectif :

$$\min \sum_{i,k} c_{ik} X_{ik} + \alpha \sum_{i} Y_i + \beta \sum_{k} O_k + \gamma \sum_{i,j} S_{i,j} \tag{5.13}$$

s.c.

$$O_k \geq \sum_{i} X_{ik} - T_k \qquad \forall k \in K \tag{5.14}$$

$$O_k \geq 0 \qquad \forall k \in K \tag{5.15}$$

$$O_k \leq T_k^* \qquad \forall k \in K \tag{5.16}$$

$$S_{ij} \geq f(i,j) * (X_{ik} + X_{jk} - 1) \qquad \forall i,j \in I^2 / a_{ij} = 0, \forall k \in K \tag{5.17}$$

$$S_{ij} \geq 0 \qquad \forall i,j \in I^2 \tag{5.18}$$

$$X_{ik} + X_{jk} \leq 1 \qquad \forall i,j \in I^2 / a_{ij} = 1 \tag{5.19}$$

$$Y_i \geq X_{ik} - Old_{ik} \qquad \forall i \in I, \forall k \in K \tag{5.20}$$

$$\sum_{k} X_{ik} = 1 \qquad \forall i \in I \tag{5.21}$$

L'objectif (Eq. (5.13)) est une somme pondérée par les scalaires α, β et γ entre, dans l'ordre, les pénalités de non respect des préférences, les pénalités de mutation d'un patient, les dépassements de capacité des chambres et la stabilité du planning. Les contraintes (5.16) et (5.15) permettent de mesurer le dépassement de capacité pour la chambre k. Les contraintes (5.16) assurent que la limite haute n'est pas dépassée. Les deux ensembles de contraintes (5.17) et (5.18) permettent de calculer la stabilité liée au couplage de deux patients dans la même chambre. L'équation (5.19) permet de s'assurer que deux patients incompatibles ne séjournent pas dans la même chambre. Notons que ces contraintes permettent d'expliciter formellement cette interdiction. Lors de la résolution du problème, avec cette formulation, il y a un risque de perte de performance dû au nombre important de contraintes induites. En fixant à deux patients incompatibles i et j une valeur élevée à la mesure de stabilité $f(i,j)$, cela reviendrait également à empêcher une telle violation de contraintes et donc à remplacer des contraintes explicites par une structure exprimant des préférences élargies. Les contraintes (5.20) évaluent si le patient i a été affecté dans une chambre autre que celle qu'il occupait. Finalement, le dernier ensemble de contrainte (5.21) assure que chaque patient est bien affecté à une chambre une et une seule fois.

5.4 Insertion des urgences

Les deux sections précédentes viennent de traiter des problèmes généraux d'affectation aux unités de soins puis à une chambre de l'unité d'accueil. Les approches de résolutions que nous venons de voir

traitent des versions journalières de ces problèmes. Autrement dit, ces méthodes permettent de mettre à jour d'une manière générale l'ensemble des affectations des patients et ainsi redistribuer les coûts/pénalités de planning en fonction des nouveaux événements survenus depuis la dernière phase d'affectation. Nous allons à présent nous intéresser aux versions dynamiques de ces problèmes et ce afin de traiter la prise en charge, l'insertion, des patients urgents.

5.4.1 Considérations

Nous allons débuter en considérant quelques hypothèses que nous émettons sur la manière dont doit être gérée l'insertion des patients urgents. La première question à laquelle il nous faut répondre est de décider qui des nouveaux patients urgents et des patients présents doit être prioritaire lors de l'affectation. En d'autres termes, est ce que le patient urgent, lors de son insertion, a la priorité sur les patients déjà présents pour être affecté dans son unité nominale ? Bien entendu, si des places sont disponibles dans le service nominal, la question ne se pose pas. Nous pensons que l'insertion d'un patient en cours de journée dans un planning préétabli est source de perturbation dans l'organisation des services. Si de plus cette insertion doit causer un transfert de patient vers une autre unité, la perturbation en est d'autant plus grande car ce transfert, contrairement à ceux pouvant survenir lors de la version journalière des affectations aux unités de soins, n'est non seulement pas prévu initialement mais, doit de plus se faire rapidement. La notion de rapidité peut ici poser problème. Quelle garantie avons nous que le personnel requis lors d'un transfert soit disponible immédiatement et ce afin d'admettre le patient urgent dans les plus brefs délais ? A priori aucune car ces employés, exerçant souvent dans des unités en sous-effectifs, sont très certainement déjà occupés à de nombreuses autres tâches. Il serait donc préférable vis-à-vis des délais d'admission d'un patient depuis les urgences que celui ci aille directement dans un service où des places sont disponibles. La perturbation due à l'insertion d'un patient urgent dans l'activité du service d'accueil est suffisamment grande sans avoir à s'occuper additionnellement d'un transfert supplémentaire.

Nous relativisons cependant cette affirmation. L'affectation d'un patient, même urgent, à une unité de soins se fait toujours en tenant compte des flux futurs connus et prévisionnels. Aussi, en fonction de ces éléments, il peut se révéler judicieux de transférer un patient d'un service vers un autre afin d'admettre un patient urgent. Nous donnons donc priorité aux patients déjà présents, et par extension, aux patients programmés, pour des raisons de performance des services d'urgences. D'autres motifs peuvent faire pencher vers cette décision. On peut considérer, à un niveau plus politique, que pour l'image d'un établissement de soins, il est préférable de procéder ainsi. En effet, d'un point de vue externe, et notamment du point de vue du patient dont l'hospitalisation est prévue plusieurs semaines à l'avance, il paraît normal de penser que plus une demande d'hospitalisation arrive tôt plus il y a de chances de voir ses souhaits être respectés. Il s'ensuit un esprit de « place réservée » entraînant le placement des urgences dans les lits restants.

Le deuxième point sur lequel nous devons mener une réflexion concerne l'heure de sortie des patients en fin de séjour. A quel moment la chambre qu'ils occupaient sera laissée libre pour un éventuel nouveau patient ? Quand un patient est déclaré sortant, il faut alors pouvoir estimer son heure de sortie. Cette donnée peut être obtenue de différentes manières selon le cas et selon le mode de sortie. Si le départ du patient s'effectue par ambulance, alors l'heure de sortie est celle à laquelle l'ambulance est prévue. Dans le cas où un proche vient chercher le patient sortant, il est nécessaire de voir avec lui à quelle heure il pourra venir. Enfin, si le patient est transféré vers un service de moyen ou long séjour de l'hôpital, il semble également possible de pouvoir convenir d'une heure avec les brancardiers. Malheureusement, ces heures ne sont toutes qu'approximatives. Rien n'empêche que le départ d'un patient soit retardé d'une ou deux heures. Ces retards sont autant de temps durant lequel des patients devront attendre aux urgences.

L'estimation d'une heure de sortie des patients ne semble donc pas être une donnée nous permettant, avec précision, de connaître en temps réel, et même par anticipation, la disponibilité des chambres. De

plus, le transfert des patients depuis les urgences se retrouve dépendant de ce processus de décharge. Nous décidons de reprendre une proposition faite par la MeaH (Cauterman et Engel, 2007) visant à favoriser la libération des lits des patients sortants le plus tôt possible. Cette proposition, qui est probablement l'hypothèse la plus forte que nous faisons sur l'organisation et les moeurs de l'hôpital, consiste en la mise en place d'un salon de sortie dans lequel les patients, dont le départ a lieu dans la journée, iront attendre leur moyen de retour à domicile. L'impact de cette proposition est important puisque dès lors, les lits sont libérés dès le matin offrant ainsi la possibilité d'admettre des patients provenant des urgences tout au long de la journée et non plus uniquement l'après midi, après les départs, comme cela est trop souvent le cas.

Nous venons de voir quelques considérations à prendre en compte lors de la gestion dynamique des affectations des patients aux unités de soins. Tous nos propos tenus dans cette sous-section sont transposables au problème d'affectation des patients aux chambres.

5.4.2 Adaptation du modèle général

Le problème d'insertion de patients urgents est en théorie résolu chaque fois qu'un nouveau patient est en attente d'admission depuis les urgences. Dans la pratique de tous les jours, cela n'a pas forcément de sens en tant que tel. La vitesse de réaction des services d'hospitalisation fait que nous pouvons nous restreindre à résoudre ce problème toutes les heures en fonction de s'il y a ou non des patients en attente aux urgences. Pour résoudre ce problème d'insertion des patients urgents, nous reprenons les problèmes de base, les problèmes journaliers, auxquels nous appliquons quelques modifications.

La première modification est commune aux problèmes d'affectation des patients aux unités de soins et aux problèmes d'affectations aux chambre d'hospitalisation. Elle va concerner la priorité que nous donnons aux patients déjà présents. En fait, il s'agit de guider le choix de la solution vers un planning favorisant le placement des patients urgents là où il y a de la place et évitant de ce fait des transferts non prévus de patients présents et affectés. Le planning général, considérant notamment qu'un patient en début de séjour doit être placé de préférence dans son unité nominale, est réétabli en considérant d'égal à égal les patients électifs et ceux provenant des urgences lors de la prochaine phase de résolution du problème journalier. L'affectation des urgences en cours de journée peut donc être vue comme une solution immédiate permettant de délester le SAU. Cette affectation sera ensuite intégrée dans l'optimisation globale de la répartition des patients entre les différentes unités de soins.

Pour ce faire, nous modifions lors de la résolution du problème d'insertion les pénalités liées au transfert d'un patient de manière à les rendre plus pénalisante qu'une affectation dans un service, ou une chambre, différent(e) de celui, ou celle, où il est préférable d'affecter le patient. En pratique, et dans les deux modèles que nous avons proposés en sous-section 5.3.2 et 5.2.1, cela revient à augmenter la valeur du paramètre de pondération α (Eq. (5.1) et (5.13)).

Il n'y a pas d'autres modifications à apporter au problème d'affectation aux chambres. Pour le problème d'affectation aux unités de soins, il est nécessaire d'adapter l'algorithme d'estimation des besoins futurs afin de prendre en compte les patients urgents pouvant encore arriver pendant la fin de la journée. Pour ce faire, il est nécessaire de définir un taux moyen d'arrivées horaires de patients à hospitaliser depuis les urgences. Dans le cas que nous avons considéré jusqu'à présent, d'un processus d'arrivées suivant une loi de Poisson de paramètre λ arrivées par jour, le processus d'arrivées horaires suivrait alors une loi de Poisson de paramètre $\lambda/24$. On peut cependant apporter quelques modifications suivant l'horaire de la journée en considérant par exemple un taux moindre durant la nuit. Dans l'algorithme 5.1, le nombre d'arrivées urgentes en période $t = 0$ à prendre en compte serait alors la somme des arrivées durant toutes les heures restantes de la journée.

Pour récapituler, nous considérons l'insertion des patients urgents comme suit. Chaque heure, si des patients urgents sont en attente, alors nous résolvons les versions modifiées des problèmes d'affectation

aux unités de soins et aux chambres. Ces versions modifiées donnent la priorité aux patients déjà présents en favorisant des solutions ou chacun reste là où il est actuellement affecté. Ensuite, à la fin de la journée, les problèmes journaliers de planifications sont exécutés afin de prendre en compte toutes les modifications qui ont pu avoir lieu. Cela comprend les arrivées urgentes et les départs mais également les modifications de dates d'admission des patients réguliers.

Terminons cette section en mentionnant que le nombre d'approches pouvant être proposées pour traiter d'un tel problème peut être important. Ainsi lors de l'insertion de patients urgents, il pourrait être retenu de ne pas déplacer les patients déjà présents et de ne bénéficier de quelques souplesses que sur les patients programmés, patients non encore physiquement présents dans l'établissement. Nous avons discuté du choix de donner une priorité à tel ou tel type de patients. Une autre solution possible aurait été de traiter les différents patients d'égal à égal et donc résoudre le problème d'insertion comme le problème journalier. Chacun en fonction de ces opinions peut argumenter en faveur d'un choix ou d'un autre.

5.5 Expérimentations

Nous choisissons de mener nos expérimentations sur les quatres unités de médecine qui ont déjà servi de support au chapitre sur la planification des admissions. Pour mémoire, il s'agit des unités de cardiologie, gastro-entérologie, pneumologie et médecine générale (dont endocrinologie). Ces services faisaient état d'un total de 96 lits d'hospitalisation pouvant être augmenté jusqu'à 116 lits en ajoutant un troisième lit dans certaines chambres doubles. Nous allons décomposer nos expérimentations en trois phases. La première phase va s'intéresser à la performance de l'approche d'affectation aux unités de soins sur différentes instances journalières. Nous comparerons notamment cette approche avec une heuristique de type « best fit ». Nous nous intéresserons ensuite aux performances atteintes sur le long terme en utilisant le support d'une simulation. Finalement, nous intégrerons la gestion de l'affectation à une chambre dans une simulation globale de la gestion du flux de patients telle que nous la sous-entendons depuis le début de ce mémoire : Planification des admissions, planification des affectations aux unités de soins et prise en charge des urgences et finalement affectation aux chambres. Chaque phase d'expérimentation sera menée sur les trois types d'hôpitaux que nous avions défini chapitre 3 : un hôpital où le flux urgent est plus important que le flux électif (hôpital de type 1, CH de Firminy), un hôpital où les deux flux sont équilibrés (hôpital de type 2) et un hôpital ayant un flux électif plus important que son flux urgent (hôpital de type 3).

5.5.1 Performance de l'approche

Dans un premier temps, comme l'approche d'affectation des patients aux unités de soins utilise une optimisation de Monte Carlo, nous allons nous intéresser à l'étude de la convergence de cette approche. En particulier, nous allons chercher le nombre d'échantillons devant être générés afin d'obtenir un résultat le plus proche possible de l'optimal. Notons que nous n'avons pas, comme dans le chapitre sur la planification des admissions, démontré la convergence formelle de cette approche. Néanmoins, les données aléatoires générées en entrée de l'algorithme de génération des échantillons étant toutes indépendantes et identiquements distribuées, alors la loi des grands nombres s'applique. Cela nous permettrait de dérouler une démonstration similaire en tout point à celle qui a déjà été faite. Nous ne regardons ici la convergence que d'un point de vue expérimental. Pour ce faire, nous considérons une instance du problème d'affectation des patients aux unités de soins que nous résolvons 10 fois pour différentes tailles d'échantillon, K, variant entre 2 et 1000. La table 5.1 donne en détail, pour l'hôpital de type 1 et pour chaque taille d'échantillon, la valeur minimale trouvée, la moyenne, la valeur maximale, l'écart type ainsi que la durée moyenne de calcul. Les valeurs de ce tableau (min,moy et max) sont représentées graphiquement par la

figure 5.1. Pour les hôpitaux de type 2 et 3, nous ne donnons que la représentation graphique des résultats associés en figure 5.2.

TABLE 5.1 – Convergence de l'optimisation de Monte Carlo pour l'hôpital de type 1

K	Min.	Moy.	Max.	écart type	CPU(s)
2	2946,95	3055,57	3265,33	108,81	0,00
5	2916,64	2956,49	3098,16	52,90	0,10
10	2912,14	2928,60	2954,91	14,44	0,30
20	2912,64	2917,58	2931,50	5,90	0,60
50	2908,01	2911,52	2915,58	2,76	2,30
100	2906,87	2909,19	2912,41	1,90	8,80
200	2906,87	2907,77	2909,02	0,81	31,90
500	2906,93	2907,25	2908,07	0,36	164,50
750	2906,87	2907,17	2907,99	0,37	331,50
1000	2906,90	2907,13	2907,35	0,17	600,00

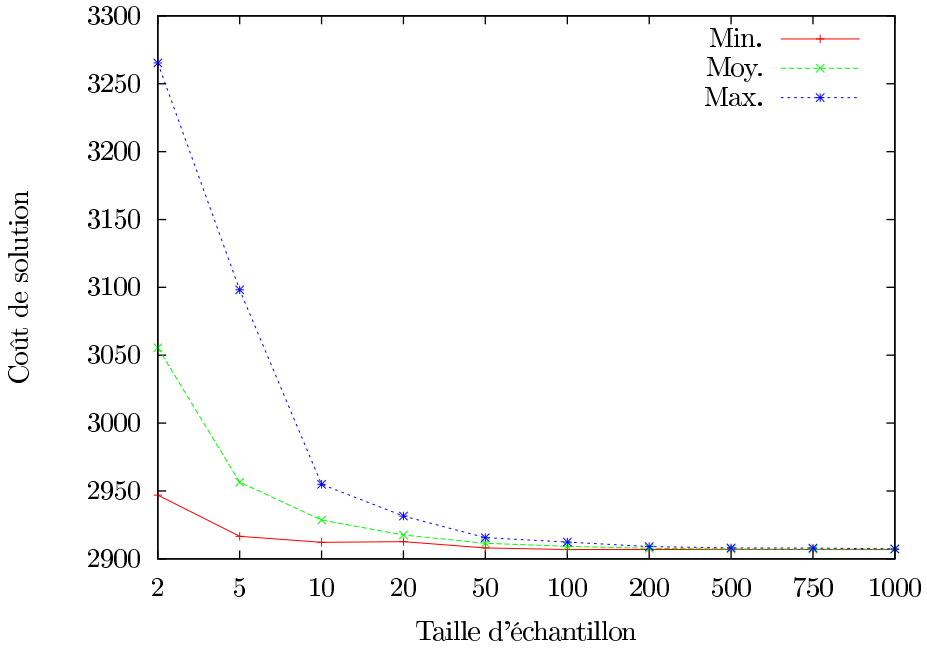

FIGURE 5.1 – Convergence de l'optimisation de Monte Carlo pour l'hôpital de type 1

Pour les trois types d'hôpitaux la convergence est quasi-parfaite à partir d'une taille d'échantillon de 200. Cependant, en prenant en considération le temps de calcul nécessaire, nous retiendrons pour la suite de nos expériences une taille d'échantillon de 50. La convergence pour les trois types d'hôpitaux est suffisamment prononcée pour cette taille et ne justifie donc pas un temps de calcul 15 fois supérieur.

Pour la suite, pour nos besoins de comparaison de performance, nous définissons une heuristique de type « best fit » permettant de résoudre le problème de planification des patients aux unités de soins. Pour chaque patient, cette heuristique regarde en premier lieu si une place est disponible dans l'unité nominale. Dans l'affirmative, le patient est affecté à cette unité. Dans la négative, le patient est affecté à l'unité ayant le plus grand nombre de places disponibles. Les patients sont affectés dans l'ordre où ils sont connus. Ils ne sont pas réaffectés par la suite. Notons que cette heuristique représente des politiques

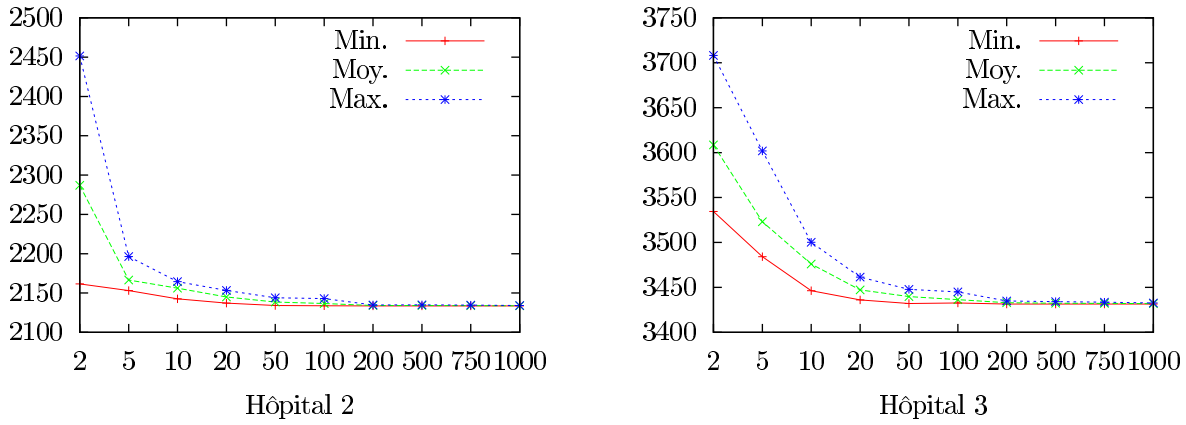

FIGURE 5.2 – Convergence de l'optimisation de Monte Carlo pour les hôpitaux de type 2 et 3

que nous avons pu constater dans la pratique où les patients, une fois affectés à un service, ne sont jamais transférés vers une autre unité sauf en cas de force majeure.

Nous pourrions nous poser la question du choix d'une heuristique offrant de meilleurs performances ou tout au plus moins simpliste que celle proposée. Il faudrait alors plutôt que choisir une unité qui à un instant donné possède le plus de lit disponible regarder sur du plus long terme en fonction du planning de programmé et des prévisions de sortie quelle serait la meilleure unité possible. Malheureusement, compte tenu du caractère de haute variabilité des systèmes hospitaliers, ce type de prévision se révelerait fort probablement peu fiable. Nous conserverons donc notre heuristique de type *best fit* qui possède au moins l'intéret de représenter assez fidèlement les pratiques observées sur le terrain.

La première comparaison que nous allons faire entre notre approche d'optimisation et l'heuristique que nous venons de définir porte sur la comparaison de performances sur la résolution d'instances journalières. Pour chaque type d'hôpital, nous avons généré aléatoirement 10 instances de problèmes d'affectation aux unités de soins. La table 5.2 pour l'hôpital de type1, la table 5.3 pour l'hôpital de type 2 et la table 5.4 pour l'hôpital de type 3 comparent pour chacune des instances les valeurs de l'objectif des solutions fournies et donnent le pourcentage de gain apporté par l'approche d'optimisation par rapport à l'heuristique. Ces gains sont importants puisqu'ils peuvent aller jusqu'à 35% d'amélioration de la solution obtenue par l'heuristique et se situent en moyenne entre 15 et 25 % d'amélioration. Nous remarquons que par deux fois pour l'hôpital de type 1, l'heuristique a fourni une solution située à moins de 5% de la solution optimisée. Il s'agit de deux instances qui peuvent être considérées comme faciles de par le nombre de patients entrant en jeu et comme en atteste les valeurs des objectifs. Ce sont les deux instances pour lesquelles les solutions, pour cet hôpital, sont les moins coûteuses.

5.5.2 Couplage avec le modèle de planification des admissions

Nous introduisons à présent la planification des affectations dans un système réel afin de mesurer la performance obtenue lors des utilisations successives jours après jours mais également lorsque l'on doit insérer en cours de journée les patients provenant des urgences. Pour ce faire, nous reprenons la base de l'algorithme d'évaluation de la performance d'une politique d'admission déjà plusieurs fois réutilisée dans cette thèse. Les modifications que nous apportons pour ce chapitre se placent après l'actualisation du planning des admissions et une fois les entrées/sorties du jour effectuées. Nous calculons alors pour chaque patient présent et pour chaque patient programmé son unité d'affectation par la méthode que nous avons exposée dans les sections précédentes. Ensuite, et afin de prendre en charge les patients urgents tout au long de la journée, nous vérifions toutes les heures si un ou des patients sont en attente au SAU. Dans

TABLE 5.2 – Performances de l'approche d'affectation journalière pour l'hôpital de type 1

	Monte Carlo		Heuristique		
Instance	Obj.	CPU (s)	Obj.	CPU(s)	Gain (%)
1 :1	3198	3	4186	0	23,61
1 :2	2761	2	2858	0	3,40
1 :3	2926	2	3534	0	17,19
1 :4	2804	1	3083	0	9,06
1 :5	3503	3	5408	0	35,24
1 :6	2787	1	2850	0	2,21
1 :7	3076	2	3836	0	19,83
1 :8	2905	2	3433	0	15,37
1 :9	2901	2	3296	0	11,98
1 :10	2823	2	3201	0	11,82
Moy.	2968,29	2,00	3568,59	0,00	14,97

TABLE 5.3 – Performances de l'approche d'affectation journalière pour l'hôpital de type 2

	Monte Carlo		Heuristique		
Instance	Obj.	CPU (s)	Obj.	CPU(s)	Gain (%)
2 :1	2118	2	2991	0	29,17
2 :2	2266	2	3084	0	26,51
2 :3	3573	6	5322	0	32,87
2 :4	1856	2	2659	0	30,18
2 :5	2203	4	3245	0	32,12
2 :6	2239	3	2587	0	13,43
2 :7	2178	3	2532	0	13,99
2 :8	2808	3	3412	0	17,69
2 :9	2272	3	2942	0	22,77
2 :10	1966	2	2234	0	12,02
Moy.	2347,92	3,00	3100,72	0,00	23,08

l'affirmative, ces patients sont insérés dans le planning des unités de soins.

Nous supposons dans cette expérimentation que tous les patients sont mutables et peuvent donc être transférés sans restriction entre les différentes unités composant l'hôpital. Nous comparons trois politiques d'affectation différentes. La première politique, dénotée AUS1 pour Affectation aux Unités de Soins, utilise l'heuristique « best fit » formulée précédemment pour l'affectation journalière et pour l'insertion des patients urgents. La deuxième approche, AUS2, combine l'optimisation proposée pour le problème journalier et l'heuristique pour l'insertion des urgences. L'intérêt d'une telle approche serait de fournir une règle simple pour l'insertion immédiate des urgences et ainsi s'affranchir d'une étape d'optimisation. Enfin l'approche AUS3 utilise la méthode d'optimisation pour les deux problèmes que sont l'affectation journalière aux unité de soins et l'insertion des patients urgents.

La table 5.5 présente les résultats obtenus pour chacune des approches et pour les trois types d'hôpitaux considérés. Les indicateurs que nous considérons sont la valeur de l'objectif tel qu'il est formulé dans ce chapitre, le nombre de patients par jour affectés dans une unité autre que leur unité nominale, le nombre total de mutation ayant lieu par jour, le nombre de mutation par jour permettant de ramener un patient vers son unité nominale (inclu dans le nombre total de mutation), le dépassement moyen de

TABLE 5.4 – Performances de l'approche d'affectation journalière pour l'hôpital de type 3

	Monte Carlo		Heuristique		
Instance	Obj.	CPU (s)	Obj.	CPU(s)	Gain (%)
3 :1	2331	3	2950	0	20,97
3 :2	2623	3	3300	0	20,52
3 :3	2279	3	2730	0	16,52
3 :4	2666	4	3477	0	23,33
3 :5	2003	2	2450	0	18,27
3 :6	2453	4	2900	0	15,40
3 :7	2685	4	3335	0	19,49
3 :8	2004	2	2771	0	27,67
3 :9	1981	2	2520	0	21,40
3 :10	2328	3	3232	0	27,95
Moy.	2335,28	3,00	2966,42	0,00	21,15

capacité par jour[2], le nombre de dépassement de capacité au-delà de la limite haute et le nombre de patients moyens considérés durant la simulation. 2000 réplications de 30 jours sont effectuées.

TABLE 5.5 – Simulation des affectations aux unités de soins dans un contexte réel

	Méthode	Obj	#wrong	#mut	#mutok	Dép. Cap.	Dep. Extra	#patients
Hôpital 1	AUS1	39,22	53,90	0,00	0,00	4,27	0,00	656,55
	AUS2	42,36	24,32	10,91	7,17	4,28	0,00	627,51
	AUS3	36,60	24,41	15,33	9,37	4,75	0,00	657,31
Hôpital 2	AUS1	18,35	44,11	0,00	0,00	2,01	0,00	632,68
	AUS2	19,92	19,93	9,38	6,27	1,89	0,00	629,99
	AUS3	17,95	20,86	12,92	8,29	2,03	0,00	633,47
Hôpital 3	AUS1	22,10	49,27	0,00	0,00	2,43	0,00	658,21
	AUS2	24,74	18,25	10,02	7,11	2,47	0,00	657,66
	AUS3	22,03	18,71	12,68	8,21	2,58	0,00	654,93

Aussi surprenant que cela puisse paraître, la méthode AUS1 fournit des résultats proches, en termes d'objectifs, de ceux de la méthode AUS3 avec tout de même un léger avantage pour la dernière. Pourtant les coûts de ne sont pas répartis de la même manière. Pour AUS1, on dénombre un total, pour l'hôpital de type 1, d'environ 54 patients par jour affectés dans une mauvaise unité. C'est deux fois moins pour AUS3 qui compense la différence de coûts par de nombreux transferts. Pour rappel, le pôle médecine servant de support à nos données d'expérimentation possède 96 lits réguliers. Avec plus de la moitié des patients affectés dans une unité différente de leur unité nominal par jour, nous pouvons tout de même qualifier le planning fourni par AUS1 de médiocre. En fait ce résultat fait apparaître une faiblesse liée à l'indicateur "*nombre de patients par jour affecté dans une unité autre que l'unité nominale*" utilisé dans le calcul de l'objectif. En effet, du point de vue de la méthode AUS1 et de l'heuristique best-fit, ce caractère se retrouve être évalué seulement une fois par jour. Dans l'approche par optimisation, la pénalité liée à la présence d'un patient en dehors de son unité nominal est prise en compte à chaque modification du planning d'affectation ; à chaque arrivée de patients urgents. En ce sens, l'approche AUS3 est meilleure que AUS1.

2. Ce nombre n'implique par obligatoirement que l'hôpital dans sa globalité soit en situation de dépassement de capacité. Une unité peut se retrouver dans une telle situation malgré le fait que des places soient disponibles ailleurs (voir la précédente note relative à la notion de disponibilité des lits).

Concernant les trois approches, nous constatons un taux de dépassement de capacité similaire dans les trois types d'hôpitaux considérés. On peut en fait avancer l'hypothèse que ces dépassements sont ceux que l'on pouvait constater lors de la planification des admissions des patients électifs. Les méthodes employées pour l'affectation aux unités de soins ont, quant à elles, été développées afin de bien répartir la charge entre les services. La capacité maximale d'accueil n'a jamais été dépassée.

Saluons, pour les approches AUS2 et 3, le nombre de transferts visant à ramener un patient dans son unité nominale : près des deux-tiers tous types d'hôpitaux confondus. En ôtant ce nombre de transferts qui peuvent être considérés comme bénéfiques, on arrive à une moyenne de proche de 1 transfert de patient par unité et à destination d'une unité différente de son unité nominale. Pour des unités de 22 à 26 lits, ce nombre est raisonnable et n'entraînera donc pas de grandes perturbations dans les services.

D'un point de vue globale ces deux approches AUS2 et 3 ont des performances similaires. Le choix d'insérer les patients urgents par heuristique avant d'optimiser le planning général semble cohérent. Seule demeure la difficulté à représenter les coûts réels d'affectation dans une mauvaise unité ce qui fausse quelque peu la mesure d'objectif. Quoi qu'il en soit, le choix des pénalités que nous avons utilisées dans cette expérimentation reste arbitraire. Nous reviendrons, d'un point de vue très général sur les difficultés qui font parties intégrantes de la formulation de ces pénalités.

Nous allons à présent nous intéresser à une deuxième série d'expérimentations que nous avons menées sur le problème des affectations des patients aux unités de soins. Cette série d'expérimentation part d'une hypothèse de base qui se distingue de tout ce que nous avons avancé jusqu'à maintenant. Dans ce mémoire de thèse en général, mais plus particulièrement dans ce chapitre, nous avons pris comme piste d'amélioration de la situation d'engorgement chronique des urgences de chercher à faciliter l'hospitalisation des patients provenant de ce service tout au long de la journée. Nous avons notamment fait une hypothèse forte dans ce but qui consiste à libérer les chambres des patients sortant dès le matin. Nous allons cette fois-ci supposer que tous les patients en attente aux urgences sont admis d'un bloc, une fois par jour. En fait, cela revient à supprimer la phase d'insertion des urgences et de ne garder que l'optimisation journalière des affectations aux unités de soins. Les résultats d'une telle approche pour l'heuristique « best fit » et l'optimisation de Monte Carlo, auxquels nous ajoutons les résultats précédents de l'approche AUS3 afin de pouvoir les comparer, sont présentés table 5.6.

TABLE 5.6 – Simulation avec admission des urgences une fois par jour

	Méthode	Obj	#wrong	#mut	#mutok	Dép. Cap.	Dep. Extra
Hôpital 1	Heuristique	61,90	14,90	0,00	0,00	7,14	0,00
	Monte Carlo	12,62	21,61	3,56	2,43	1,81	0,00
	AUS3	36,60	24,41	15,33	9,37	4,75	0,00
Hôpital 2	Heuristique	27,26	17,07	0,00	0,00	3,59	0,00
	Monte Carlo	5,24	18,52	3,47	2,53	0,65	0,00
	AUS3	17,95	20,86	12,92	8,29	2,03	0,00
Hôpital 3	Heuristique	21,06	26,55	0,00	0,00	2,67	0,00
	Monte Carlo	8,80	16,70	3,38	2,62	1,17	0,00
	AUS3	22,03	18,71	12,68	8,21	2,58	0,00

Les résultats sont presque sans commentaire. Nous pouvons constater, au passage, que la méthode d'évaluation de l'objectif se prête mieux au problème journalier qu'au problème avec insertion des urgences. La performance qui pourrait être atteinte pour les unités de soins est très appréciable et est largement due à la suppression des perturbations liées à l'insertion des urgences au fil de la journée.

La question à laquelle nous allons répondre à présent est : comment une telle approche peut avoir un

sens permettant de désengorger les urgences en pratique ? En fait, ceci est possible si l'on revenait à la définition originelle des UHCD. A la base, ces unités devaient permettre de servir de zone tampon entre les urgences, en déchargeant les patients des salles d'urgences, et les services d'hospitalisation, en leur laissant plus de temps pour planifier l'admission de ces patients. Dans les pratiques actuelles, les UHCD sont devenues des zones de soins annexées à part entière aux SAU qui avaient besoin de s'agrandir. Elles ne permettent donc plus de servir de zone tampon. Nous en revenons à un problème soulevé en introduction de ce mémoire : le flux de patients a augmenté, la taille des bâtiments est restée la même. Cependant, comme le montre ces résultats, s'il était possible de créer une zone UHCD destinée uniquement à son rôle propre, la gestion des services d'urgences et d'hospitalisation s'en trouverait améliorée par l'utilisation d'outils tels ceux que nous avons proposés.

5.5.3 Une simulation du processus de gestion du flux de patients

Cette dernière partie de la section dédiée aux expérimentations va s'intéresser à l'affectation des patients aux chambres des unités de soins. Pour ce faire, nous allons intégrer la résolution de ce problème au sein d'un processus de simulation globale de l'approche de gestion du flux de patients que nous avons proposée. La planification des dates d'admission des patients réguliers va être régie par l'approche utilisant l'optimisation de Monte Carlo du chapitre 3. Ensuite, une fois admis, les patients seront affectés aux unités de soins à l'aide de l'approche présentée dans ce chapitre puis affectés aux chambres d'hospitalisation. Entre-temps les patients provenant des urgences seront insérés dans le planning des unités de soins également avec l'approche d'optimisation présentée dans ce chapitre. Finalement les affectations de chacun aux chambres seront déterminées. Cela revient à évaluer la performance globale de notre ambition première qui était de favoriser la prise en charge des patients urgents tout au long de la journée.

Nous ne jugeons pas pertinent dans cette partie de comparer notre approche avec des heuristiques comme cela a pu être le cas pour d'autres expérimentations. En effet, étant donné que nous considérons une succession de décisions résultantes d'algorithmes d'optimisation, définir une même succession venant d'heuristique ne serait fort vraisemblablement pas significative. Aussi, quand cela est possible, nous nous comparerons avec des systèmes réels lors de l'analyse des résultats.

En premier lieu, nous nous intéressons à l'utilisation faite de la capacité en lit de chaque unité. La table 5.7 commence par rappeler pour chaque unité sa capacité en lit théorique et sa limite haute. Ensuite, pour les trois types d'hôpitaux servant de support à nos évaluations, nous indiquons le nombre moyen de patients présents dans chaque unité en considérant un horizon de un mois. Nous indiquons également pour chaque unité le nombre moyen de lits supplémentaires utilisés chaque jour.

TABLE 5.7 – Utilisation des lits par unité et par type d'hôpital

		Unité 1	Unité 2	Unité 3	Unité 4
	Nb. lits	22	26	22	26
	Nb. lits max	28	26	28	34
Hôpital 1	Nb. moyen de patients	22,96	24,64	22,91	26,97
	Nb. de lits sup.	1,79	0,04	1,59	2.02
Hôpital 2	Nb. moyen de patients	21,64	24,07	21,39	25,16
	Nb. de lits sup.	0,92	0,02	0,74	0,97
Hôpital 3	Nb. moyen de patients	22,06	24,60	21,51	25,80
	Nb. de lits sup.	1,1	0,09	0,96	1,23

Une des valeurs les plus remarquables est l'utilisation de lits supplémentaires dans l'unité 2 qui pourtant n'est pas habilitée à dépasser son nombre de lits théoriques. De plus, lors des précédentes expérimentations, nous avons vu qu'aucun dépassement de capacité maximale n'était observé. En fait,

cette limite haute du nombre de patients maximum n'est effectivement jamais dépassée. Par contre, il est arrivé lors de nos réplications de simulations que l'on arrive dans la situation où le nombre de patients devant être isolés a été supérieur au nombre de chambres simples. La première conséquence possible a été le placement d'un tel patient seul dans une chambre double, entraînant de ce fait la perte d'une place qui était pourtant requise étant donné que le service était saturé. La deuxième conséquence, qui a été observée, a provoqué le besoin d'un lit supplémentaire pour ce patient devant être isolé étant donné que toutes les chambres étaient occupées.

Ce phénomène est inclu dans l'utilisation des lits supplémentaires que l'on retrouve pour les autres services. Il vient s'ajouter aux cas où la demande dépasse la capacité et où les hospitaliers sont obligés d'ouvrir temporairement des lits supplémentaires.

La table 5.8 nous présente pour chaque service le nombre de mutations qui a été enregistré. Nous détaillons ensuite le nombre de mutations internes au service afin de faire la part entre les mutations vers une autre unité et les mutations décidées par la politique d'affectation des patients aux lits que nous avons proposée.

TABLE 5.8 – Nombre de mutations par service et par type d'hôpital

		Unité 1	Unité 2	Unité 3	Unité 4
Hôpital 1	Nb. de mutations	2,64	2,94	2,44	3,14
	Dont mutations internes	0,62	0,4	0,54	0,69
Hôpital 2	Nb. de mutations	2,3	2,55	2,2	2,75
	Dont mutations internes	0,49	0,38	0,41	0,52
Hôpital 3	Nb. de mutations	2,25	2,52	2,18	2,66
	Dont mutations internes	0,29	0,25	0,3	0,33

Parmi le nombre total de mutations vers des autres services, il faut en premier lieu considérer les transferts que nous avons précédemment qualifiés de bénéfiques : les transferts de patients vers leur unité nominale. Ces transferts représentant comme nous l'avons vu environ deux-tiers du nombre de transferts entre service, nous retrouvons bien le chiffre approximatif de 1 transfert par jour vers une unité quelconque. Le nombre de transferts internes est très satisfaisant. Pour un hôpital comme celui de Firminy (type 1), nous atteignons un taux d'environ une mutation interne tous les deux jours. D'après les dires des cadres de santé des services concernés, le nombre de mutations internes actuel est d'environ un par jour. De ce point de vue, notre approche est meilleure. Cependant, la notion de transfert entre service est quasi-inexistante dans cet établissement. Notre approche en réalise régulièrement mais c'est, à notre avis, un mince prix à payer pour permettre de favoriser l'écoulement du flux de patients urgents.

5.6 Conclusion

Nous avons dans ce chapitre présenté une approche permettant la gestion du niveau opérationnel de notre cadre de travail. Ce niveau considère les affectations des patients, qu'ils proviennent du flux urgent ou programmé, en pratique. Le problème se modélise formellement comme dynamique où, au fil du temps, une suite de décisions vont chercher à permettre la minimisation de différents critères. Parmi les critères que nous avons pris en compte, le respect des préférences patients et médicales étaient au coeur de nos préoccupations et assortie de la volonté de vouloir favoriser une insertion facilitée des patients urgents. Les transferts de patients d'une chambre vers une autre, relevant parfois d'une autre unité de soins, était l'un des éléments que nous avons introduits de manière à atteindre nos objectifs.

Lors de la construction des méthodes présentées dans ce chapitre, nous avons parfois fait des choix que nous avons justifiés. Selon les convictions de chacun, d'autres alternatives auraient pu être envisagées

afin d'accorder une importance plus grande à certains aspects.

La prise en compte des événements pouvant survenir dans le futur a été envisagée selon deux angles. Lors de l'affectation aux unités de soins, nous avons considérer une approche d'optimisation dite de Monte Carlo. Autrement dit, nous avons chercher à estimer les cas possibles d'évolution les plus probables avant de prendre notre décision. La méthode d'affectation aux chambres a fait le choix de la robustesse face aux événements futurs. Pour ce faire, nous avons défini une notion de stabilité de couplage entre patients nous permettant de garantir, si possible, que les affectations décidées ne seront pas remises en cause.

Cette idée, qui diffère des approches classiques d'affectation aux chambres par simple respect des contraintes, mériterait d'être poussée plus loin. En effet, dans l'état actuel de l'approche proposée, un couplage stable de deux patients peut être brisé lors de la phase d'affectation aux unités de soins. il serait intéressant de guider cette méthode en tenant compte de la qualité des affectations existantes dans les chambres au moment du choix des patients à transférer. Cela pourrait se faire de manière très basique en interdisant par exemple le transfert d'un patient impliqué dans un couplage qualifié de stable.

Conclusion générale

La gestion du flux de patients dans les unités de soins est soumise à de nombreux aléas sources d'incertitudes lors des différentes prises de décision. Nous avons proposé dans cette thèse une suite hiérarchique d'outils de planification afin de s'adapter du mieux possible à l'ensemble de ces événements perturbateurs. Nous avons en particulier chercher comment prendre en charge les patients provenant des urgences de manière à satisfaire le double objectif de minimiser l'impact de la prise en charge d'un patient urgent dans un service et de minimiser les difficultés des services d'urgences à faire hospitaliser leurs patients. L'organisation actuelle des services de soins gére la prise en charge des patients urgents en fonction des patients dont l'admission est programmée. Le postulat que nous avons introduit dans cette thèse consiste à programmer l'admission des patients régulier puis les affectations en tenant compte des patients pouvant se présenter ultérieurement.

Gestion des patients dans les unités de soins : synthèse

La démarche proposée dans cette thèse traite en premier lieu de la planification des dates d'admission des patients réguliers tout en tenant compte de l'impact qu'un tel choix va avoir sur le futur et les nouveaux patients qui se présenteront. Les situations où la capacité d'accueil de l'hôpital est dépassée étant les plus néfastes lors de la prise en charge des nouveaux patients, nous avons en premier lieu chercher à éviter ce cas de figure. Nous avons établi qu'une approche de planification par optimisation de Monte Carlo était la plus performante de par sa capacité à estimer les probables besoins futurs.

Cependant, l'estimation des besoins futurs a nécessité la définition d'une clé de répartition afin de représenter la manière dont les patients sont admis selon la politique retenue. Il a été montré au travers d'expérimentations menées sur plusieurs méthodes d'optimisation stochastique que l'approche par analyse de perturbation fournissait la clé permettant à l'optimisation de Monte Carlo de fournir les meilleurs planning d'admissions.

Ensuite est venu la phase de prise en charge du patient en l'affectant successivement à une unité de soins puis à une chambre. L'approche par optimisation de Monte Carlo a montré une fois de plus ses performances dues à sa capacité à estimer les évolutions du système. Cette approche est efficace aussi bien pour le recalcul journalier des affectations que pour l'insertion des patients urgents en cours de journée. Enfin, l'affectation aux unités de soins s'est faite sous forme d'un problème de couplage entre patients. Cette approche est originale comparée aux travaux qui, jusqu'à présent, se contentaient de trouver un planning d'affectation aux chambres respectant uniquement les contraintes fortes.

Perspectives de recherche

Chacune des étapes de planification, que nous avons proposées, peuvent continuer d'être développées. En premier lieu, l'approche de planification des admissions, telle que nous l'avons présentée, fonctionne parfaitement pour les services de médecine. Même si nous avons formulé des adaptations qui peuvent y

être apportées afin de gérer les admissions en chirurgie, les contraintes liées au bloc opératoire mériteraient d'être plus amplement considérées. Il serait très intéressant, en gardant notre postulat de planification des admissions en tenant compte des évolutions possibles, de gérer simultanément le planning du bloc opératoire et les admissions dans les services de chirurgie.

En amont du problème de choix des admissions se trouve un autre problème qui serait intéressant de résoudre : la répartition des consultations médicales. En effet, c'est lors de ces consultations que les décisions d'hospitalisation sont prises. On peut aisément imaginer que si toutes les consultations d'un service ont lieu le lundi, alors le nombre de patients devant être admis avant la fin de semaine serait important et le risque de saturer le système deviendrait également important. La résolution d'un tel problème aurait comme objectif de répartir les demandes d'admission en fonction des services. Cela présenterait l'intérêt de faciliter l'affectation des patients aux unités de soins en équilibrant les demandes. Un flux bien dimensionné pour un service donné présentera alors des avantages pour les autres services qui n'auront en théorie plus à recevoir de patients destinés théoriquement à ce service. L'impact serait alors global sur l'hôpital et minimiserait le nombre de transferts effectués entre les services.

Lors des deux étapes d'affectation, l'idée du couplage stable entre les patients, dans l'optique de garantir la robustesse du planning, pourrait être étendue. Tout d'abord, il serait intéressant de tester des méthodes issues de la littérature sur des problèmes de couplage en lieu et place de la résolution par programmation linéaire que nous avons proposée. En second lieu, la notion de couplage pourrait être incorporée dans l'affectation aux unités de soins. En effet, déplacer un patient qui était affecté dans une chambre avec un autre patient avec qui ils formaient une affectation "stables", va très certainement occasionner des mutations internes afin de reconstruire une nouvelle solution la plus stable possible. L'idée à expérimenter serait de pénaliser un transfert d'un patient d'un service vers un autre en tenant compte de la stabilité du couplage dans lequel il était engagé.

Nous notons, pour ces deux problèmes, que nous avons formulé un certain nombre d'hypothèses visant à diriger une solution vers tel ou tel type de solutions. D'une manière générale, et en vue de l'implémentation d'un système de gestion des patients tel que nous le concevons dans un système d'informations hospitalier, la réalisation d'un guide reprenant un ensemble d'hypothèses pouvant être formulées et permettant, en fonction des hypothèses retenues, de définir le paramètrage, notamment en termes de structure de pénalités, de la méthode d'optimisation serait un atout considérable.

Enfin, l'optimisation de la clé répartition présente sans doute les perspectives de recherche les plus concrètes. Là où la plupart des propositions que nous venons de formuler considèrent des cas appliqués à la pratique, les méthodes d'optimisation que nous avons employées pour la recherche de clé ont montré des lacunes dans la théorie. Les mauvais résultats pouvant être fournis par des approches pourtant très performantes dans des cas particuliers en sont un exemple. Comment faire pour garantir que, quelles que soient les conditions expérimentales, la méthode converge toujours vers une bonne solution ? En entrant plus dans les détails les méthodes d'optimisation que nous avons utilisées ont été historiquement élaborées pour des problèmes dont les paramètres sont continus. Dans notre cas, ils étaient contraints. Même si quelques études ont tenté d'étendre les résultats du domaine continu au domaine contraint ou discret, les résultats théoriques ne sont pas suffisants laissant une large place à la recherche future sur ces problématiques.

Vers une gestion globale de l'hôpital

L'intégration dans un logiciel d'une approche telle que celle de gestion du flux de patients que nous proposons serait une bonne chose pour l'hôpital. Elle permettrait non seulement la centralisation des informations mais aussi et surtout une gestion équitable des unités concernées. Nous insistons sur le fait que l'apport d'une démarche d'optimisation est bénéfique pour les unités concernées. En effet, dans

des systèmes interdépendants comme les hôpitaux, la définition d'une nouvelle politique optimisant le fonctionnement d'une unité peut avoir des conséquences désastreuses sur d'autres unités qui n'étaient pas préparées à recevoir un flux de patients comme celui qu'elles recevront dorénavant.

Par exemple, la modification des politiques d'admission de patients peut avoir un impact sur des services auxquels les unités de soins font appel comme l'imagerie et l'exploration fonctionnelle. Il est important à nos yeux que chaque unité ait son système de planification optimisé afin de garantir que son rendement sera le meilleur possible en tenant compte de ses paramètres. Quoi qu'il en soit, si chaque unité doit avoir son module de planification, il deviendrait alors intéressant d'avoir un "meta-module" de planification qui permettrait d'harmoniser les interfaces entre les différentes entités constituant un hôpital. Toujours est-il qu'en attendant de voir apparaître un module d'optimisation « gestionnaire d'hôpital » la recherche a encore beaucoup à faire sur la gestion des composantes d'un système hospitalier.

Finalement, pour conclure sur les perspectives d'avenir, si nous devions désigner l'action à entreprendre suite à ce travail de thèse, nous choisirions la planification des ressources d'imagerie et d'exploration fonctionnelle. En effet, ces services sont régulièrement sollicités tout au long du séjour d'un patient. Nous jugeons pertinent de proposer des outils de planification de ces ressources qui tiennent compte du planning d'admission que nous proposons afin de garantir que le séjour du patient se déroule sans contretemps.

*

* *

Nous souhaitons conclure cette thèse par deux citations se rapportant à notre travail.

« *Il n'y a rien de plus triste qu'une vie sans hasard* » Honoré de Balzac (1799-1850), romancier. Le hasard, les incertitudes font partie intégrante de la vie de tous les jours. Il n'est pas possible d'y échapper. Nous avons beau élaborer des méthodes permettant d'anticiper les évolutions d'un système, il restera toujours une part d'incertitude. En ce sens, aucune méthode de prévision n'est parfaite. Nous pouvons juste réduire l'impact de certains de ces aléas.

« *La tradition, c'est le progrès dans le passé ; le progrès, c'est la tradition dans le futur* » Edouard Hériot (1872-1957), homme politique, écrivain et académicien français. Cette citation replace la notion de progrès dans un cadre d'améliorations permanentes. En particulier, les recherches scientifiques d'aujourd'hui représentent un progrès pour demain. Mais demain, la véracité des propos avancés sera remise en cause. Les résultats d'une étude, une thèse, le travail de recherche de toute une vie ne sont jamais achevés. Il reste toujours une place pour faire avancer la science à qui veut la prendre.

Annexes

Annexe A

Le flux de patient du CH de Firminy en 2008

L'hôpital de Firminy a largement contribué, de par son support, au travail présenté dans ce mémoire. Cet établissement est composé d'environ 450 lits et 900 employés ce qui en fait un hôpital de taille moyenne. Géographiquement, le CH de Firminy se trouve en périphérie de la métropole stéphanoise à la frontière du département de la Haute-Loire. La forte ruralité de ce département implique un large périmètre de provenance des patients pouvant aller jusqu'à 40 kms. La population de classe ouvrière est largement prédominante dans ce bassin. La présente annexe fournit un recueil de statistiques permettant de caractériser l'établissement par son activité et ses flux de patients.

A.1 Activité des urgences

Le service d'accueil des urgences a été initialement conçu pour recevoir 12000 passages par an dont 2000 en UHCD. En regardant la table A.1 concernant le flux de patients des années 2007, 2008 et 2009, on constate d'une part que ce nombre est largement dépassé et d'autre part que l'activité est en constante augmentation. La construction d'un nouveau bâtiment est en cours pour faire remplacer les locaux actuels devenus exigus. Outre l'augmentation du nombre de box de consultation et de lit d'UHCD, des commodités d'accès aux différentes ressources de l'hôpital (principalement biologie et scanner) sont prévues.

En mettant la pédiatrie de côté, on constate que le tiers des passages journaliers est suivi d'hospitalisation. Pour une meilleure connaissance du flux de patients dans cette unité, cette proportion relativement importante d'hospitalisation serait à mettre en lien avec un autre indicateur qui est le nombre de passages de personnes agées de plus de 65 ans. La collecte des données pour cet indicateur est en cours et semble indiquer un taux voisin de 70% par rapport au nombre de passages adultes.

A.2 Dimensionnement de l'hôpital et taux d'occupation

Les unités de courts séjours de l'établissement de Firminy sont répartis en trois catégories principales classiquement dénotées par l'acronyme MCO (Médecine Chirurgie Obstétrique) auxquels on ajoute une unité de courts séjours en gériatrie. La table A.2 présente pour chaque composante les unités de soins les consituants ainsi que le nombre de places disponibles théoriques et le nombre de places d'accueil maximum.

Cet hôpital est légèrement sous-dimensionné par rapport à son activité. La table A.3 indique pour

TABLE A.1 – Activité du service d'urgences

	Passages Urgences			Passages par jour		
	2007	2008	2009	2007	2008	2009
Adultes + enfants						
Simples	17384	18136	19582	47,5	49,6	53,5
Suivis d'hospitalisations	7203	7231	7646	19,7	19,8	20,9
Départ sans soins et/ou fugues	4	58	33	0,0	0,2	0,1
Total	**24591**	**25425**	**27261**	**67,2**	**69,5**	**74,5**
Adultes (plus de 15 ans et 3 mois)						
Simples	10707	11964	13196	29,3	32,7	36,1
Suivis d'hospitalisations	5926	5838	6212	16,2	16,0	17,0
Départ sans soins et/ou fugues	1	31	16	0,0	0,1	0,0
Total	**16634**	**17833**	**19424**	**45,4**	**48,7**	**53,1**
Enfants						
Simples	6677	6172	6386	18,2	16,9	17,4
Suivis d'hospitalisations	1277	1393	1434	3,5	3,8	3,9
Départ sans soins et/ou fugues	3	27	17	0,0	0,1	0,0
Total	**7957**	**7592**	**7837**	**21,7**	**20,7**	**21,4**

chaque unité de soins ainsi que pour l'hôpital le taux de remplissage moyen annuel et la proportion de remplissage provenant des urgences. Avec une moyenne annuelle de 88% d'occupation et un remplissage réalisé au deux-tiers par des patients provenant du flux des urgences, on peut aisément imaginer les difficultés de gestion pesant sur les services.

Les unités de chirurgie présentent de faible taux d'occupation. Cela s'explique par deux points. Tout d'abord pour l'unité de semaine, du fait de sa fermeture pour le week end, on constate tout au long de la semaine une diminution du nombre de patients présents. Il y a ici une perte d'heure lit due vraisemblablement à l'absence de programmation de chirurgies courtes en fin de semaine. Ensuite pour l'unité d'hospitalisation complète, la statistique présentée ici ne prend pas en compte les patients d'hospitalisation de semaine dont le séjour se prolonge le week end. Ces patients sont mutés physiquement, mais non administrativement, dans cette unité de soins complets.

A.3 Flux de patients dans les services

La table A.4 présente pour l'année 2009 le nombre d'entrées/sorties réalisées chaque mois en MCO. Environ 1400 patients sont traités par mois tout au long de l'année.

Les tables A.5 et A.6 donnent respectivement pour la médecine et la chirurgie la répartition des entrées directes et urgentes entre les différentes unités. Pour la chirurgie, il est clair que l'unité d'hospitalisation de semaine fonctionne principalement sur l'activité programmée tandis que l'unité d'hospitalisation complète prend en charge le flux urgent.

A.4 Durées de séjour

Pour finir, cette section va concerner les distributions des durées de séjours des patients selon qu'ils relèvent de médecine ou de chirurgie et selon leur provenance, flux programmé ou flux urgent.

Chacun des graphiques suivant permettent de voir la distribution des durées de séjour tel qu'elles ont été observées en 2008. Nous superposons les courbes représentatives de lois de distribution usuelles

TABLE A.2 – Répartition de lits et des chambres

Type activité	Ch. simples	Ch. doubles	Lits sup.	Nb. Lits	Nb. lits max.
Pédiatrie et néonatologie	9	4	0	17	17
Gynécologie obstétrique	23	7	8	37	45
I - Mere Enfant	**32**	**11**	**8**	**54**	**62**
Cardiologie	10	8	6	26	32
Alcoologie	2	4	0	10	10
Gastroentérologie	4	6	0	16	16
Endocrinologie	6	8	6	22	28
Pneumologie	6	10	8	26	34
II - Medecine	**28**	**36**	**20**	**100**	**120**
Chir. hosp. de semaine	4	6	0	16	16
Chir. Hosp. complète	8	11	0	30	30
III – Chirurgie	**12**	**17**	**0**	**46**	**46**
IV - Court Sej. Geriatrique	**8**	**0**	**0**	**8**	**8**
MCO (=I+II+III+IV)	**68**	**47**	**28**	**208**	**236**

TABLE A.3 – Taux d'occupation global et taux de remplissage par les urgences pour l'année 2008

Unité	Taux Occup.(%)	Proportion Urgences (%)
Gastro-entérologie	99,2	64,28
Endocrinologie	101,73	64,34
Pneumologie	94,72	79,3
Chir. Complète	77,77	70,51
Chir. Semaine	48,71	38,71
Courts Séj. Gériatriques	124,25	63,22
Cardiologie	97,77	67,46
Hôpital	**88,36**	**67,05**

approximant bien la distribution empirique observée.

Le seul cas particulier à mentionner concerne les durées de séjours de la chirurgie programmée. En partant de l'observation que la moitié des séjours dure 1 jour, la construction de la distribution théorique est réalisée en deux temps. Tout d'abord, par expérience de Bernoulli de probabilité 0,5, nous définissons si la durée de séjour est 1 ou plus. Les durées supérieures à 1 jour sont ensuite approximées comme pour les autres cas.

Remarque : pour tous les graphiques, l'indication de durée de séjour "31" représente en réalité le nombre de séjour dont la durée est supérieure à 30 jours.

TABLE A.4 – Entrées/sorties en unités de courts séjours

	Jan.	Fev.	Mars	Avr.	Mai	Juin	Jui.	Aout	Sep.	Oct.	Nov.	Dec.
Entrées	1415	1348	1554	1577	1434	1531	1382	1225	1376	1491	1402	1252
Sorties	1424	1346	1567	1578	1450	1526	1403	1203	1361	1513	1385	1451

TABLE A.5 – Répartition des flux d'entrée de patients en médecine

	Entrées Directes	Patients Urgents
Nb. Entrées/jour	3,03	5,72
Cardiologie	33,69	30,31
Gastro-Entérologie	31,26	26,44
Endocrinologie	19,01	17
Pneumologie	16,04	26,3

TABLE A.6 – Répartition des flux d'entrée de patients en chirurgie

	Entrées directes	Patients urgents
Nb. Entrées/jour	2,82	2,67
Chir. Semaine(%)	72,26	20,06
Chir. Complète(%)	27,74	79,94

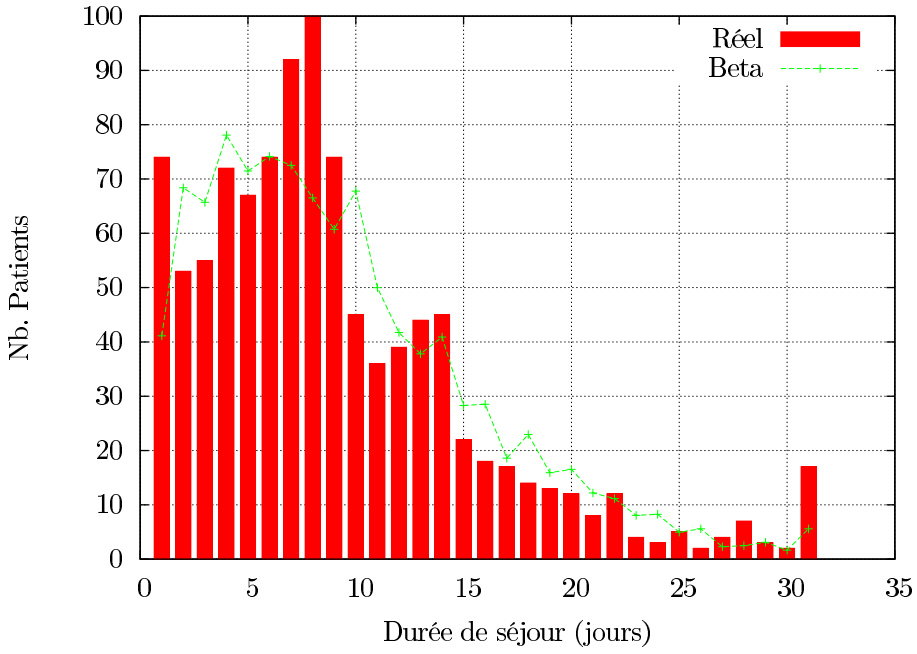

FIGURE A.1 – Distribution des durées de séjour des patients réguliers de médecine

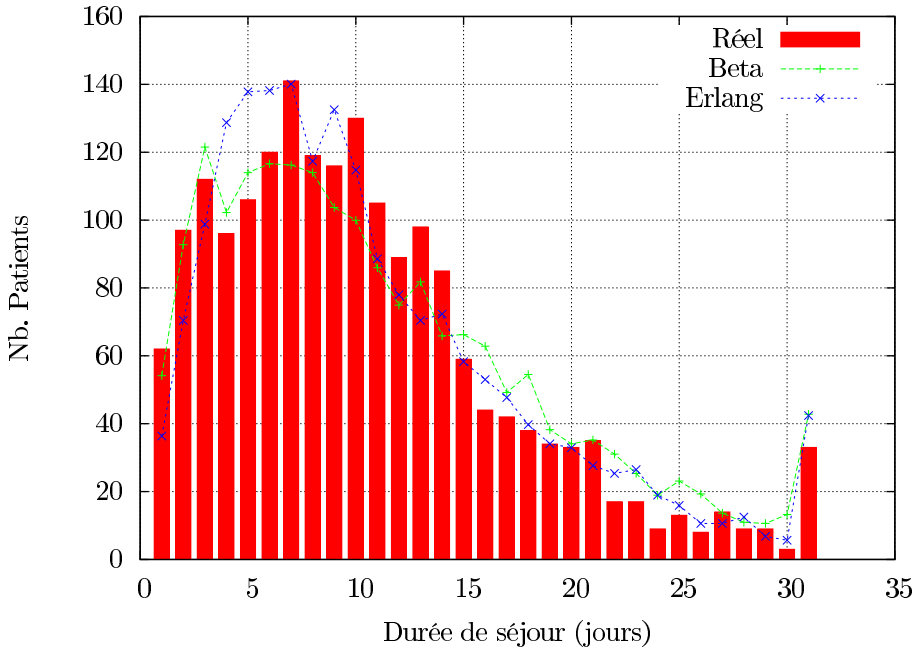

FIGURE A.2 – Distribution des durées de séjour des patients urgents de médecine

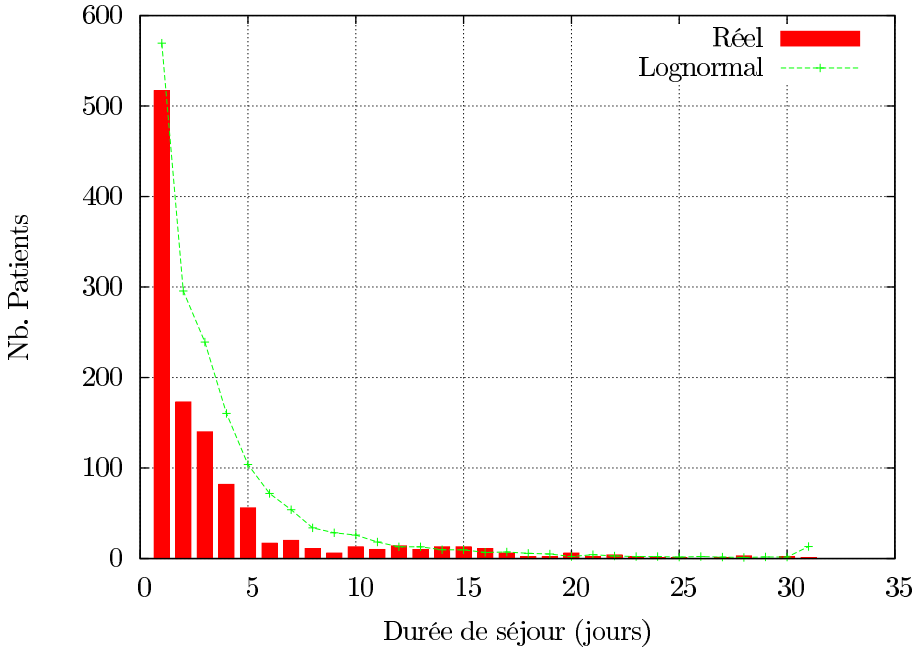

FIGURE A.3 – Distribution des durées de séjour des patients réguliers de chirurgie

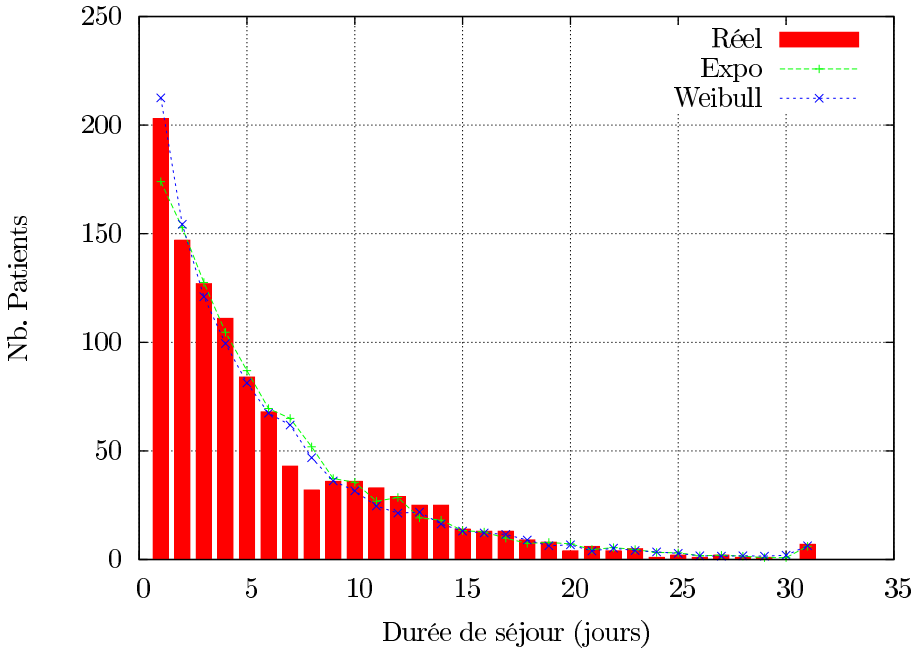

FIGURE A.4 – Distribution des durées de séjour des patients urgents de chirurgie

Annexe B

La tarification à l'activité

Les hôpitaux publics fonctionnaient auparavant sur une dotation financière globale annuelle. Afin de supprimer les différences avec le secteur privé, cette dotation a été supprimée au profit d'une tarification à l'acte. Les revenus sont maintenant liés à la pathologie des patients traités. Les réorganisations nécessaires dans le secteur public ont été nombreuses et sources de beaucoup d'étude scientifiques. La principale conséquence est le besoin d'une gestion des dépenses au plus juste. En 2005, la France consacrait 11,5% de son PIB à la santé ce qui la classe au premier rang européen devant l'Allemagne et au troisième rang mondial après les Etats-Unis (16%) et la Suisse (13%).

B.1 Coûts de production d'un soin

La production de soins en milieu hospitalier utilise différentes ressources dont les coûts d'utilisation viennent s'additionner pour former le coût d'un séjour ou d'un soin. Quatre sources de coûts peuvent être identifiées :

- Temps de personnel soignant ou non
- Frais de pharmacologie
- Utilisation de plateau technique (bloc opératoire, radiologie, laboratoire d'analyse, ...)
- Hôtellerie

Historiquement, le mode financement des hôpitaux a évolué en fonction de ses coûts de production. Au XIX^{eme} siècle, le financement était essentiellement basé sur des fonds propres (fonciers, vignes, ...) et privés (dons, fondations, ...). Les frais de production étaient essentiellement composés de frais de personnel soignant et d'hôtellerie. Durant la première moitié du XX^{eme} siècle, les frais de plateau technique et de pharmacologie ont commencé à venir s'ajouter aux autres. Cette période a vu apparaître le principe de la sécurité sociale comme mode de financement des établissements de santé. Cette tendance s'est ensuite généralisée durant la fin du siècle.

La rétribution faite aux hôpitaux par la sécurité sociale a évolué au court du temps en partant d'un mode de financement par indemnités journalières puis en passant au principe de dotation globale et en aboutissant actuellement au principe de la tarification à l'activité. Chacun des précédents modes de financements présentaient des défauts. L'indemnité journalière poussait les établissements à garder les patients plus que nécessaire. La dotation globale a sous-doté certains établissements et sur-doté d'autres. La tarification à l'activité se veut être plus juste en ne payant que ce qui est nécessaire.

L'activité est définie à partir de ces quatres composantes de coûts de production auxquels viennent s'ajouter des informations liées au patient tel l'âge et la présence éventuelle d'une autre pathologie (diabète, ...). Les établissements de santés, publics et privés, doivent obligatoirement procéder à l'évaluation de leur activité afin de recevoir de la façon la plus juste possible les ressources nécessaires à cette activité

(loi du 31 juillet 1991).

B.2 Groupe Homogène de Malade

Afin d'évaluer l'activité, les patients sont regroupés en groupe homogène de malade (GHM). Ils sont ensuite utilisés pour définir des groupes homogènes de séjours (GHS) auxquels sont associés un tarif. Ce tarif représente le coût global de prise en charge d'un patient. Ainsi ce tarif se base sur des indicateurs tels que le nombre de nuits théoriques passées par un malade, les besoins en examens complémentaires, ... Les actes réalisés sont codifiés par la CCAM. Même si des actes sont prévus dans le tarif initial, rien n'empêche cependant d'en pratiquer d'autres ou d'en remplacer. Ainsi les hôpitaux, les praticiens, restent libre de soigner leur patient comme ils l'entendent. La seule contrainte vis-à-vis de l'établissement est de ne pas dépasser l'allocation budgétaire prévue par la T2A faute de quoi, le financement doit pouvoir être soutenu soit par l'hôpital lui même soit par des fonds propres du patient.

Le tarif est estimé à partir d'un échantillonnage national de séjours hospitaliers équivalents réalisé dans des hôpitaux pilotes.

La classification, le regroupement, de patients en groupes homogènes de malade est décrite dans une arbre de décisions. Les différents éléments de décision sont donnés dans les figures B.1 et B.2. Un extrait de l'arbre de décision pour des pathologies orthopédiques est donné figure B.3.

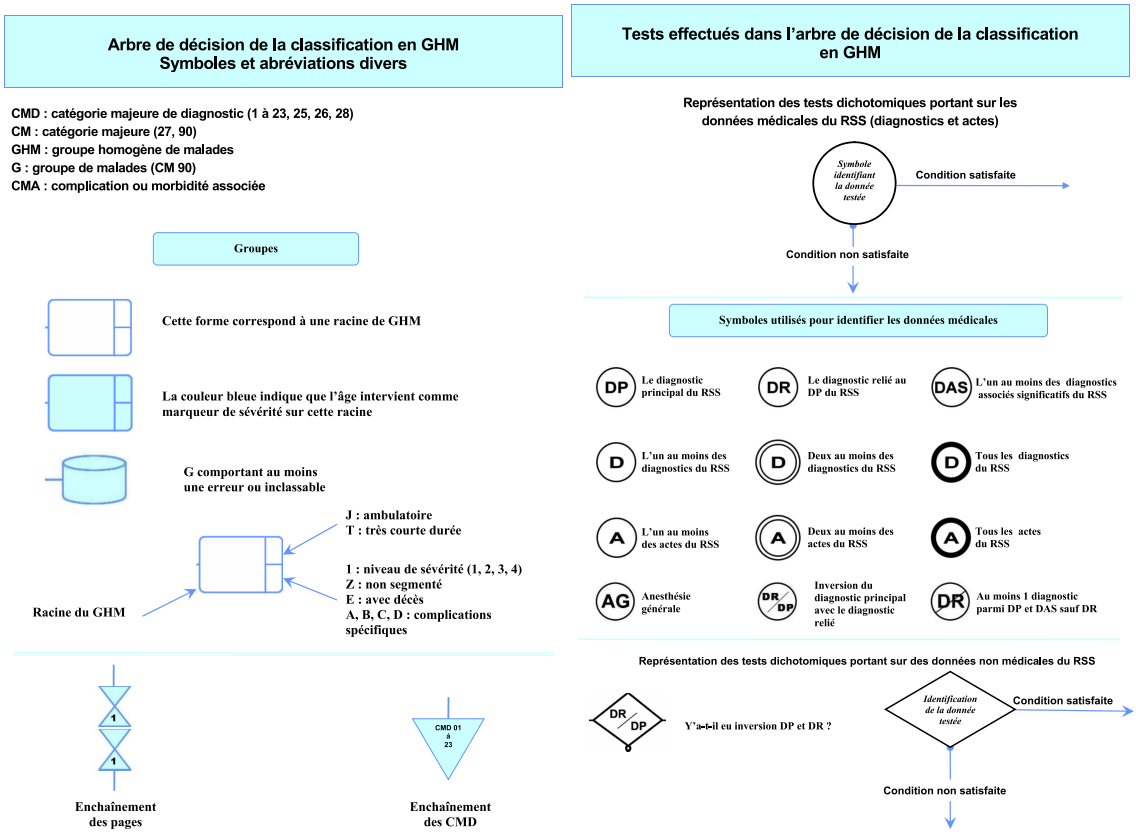

FIGURE B.1 – GHM : Eléments décisionnels

Par exemple, un patient âgé de plus de 69 ans dont le diagnostic principal est un infarctus du myocarde, ayant reçu comme acte technique une coronographie et ayant un diagnostic associé (pathologies compliquant sa prise en charge) de diabète concomitant peut faire partie du GHM Infarctus aigu du myocarde sans complication cardiovasculaire avec CMA de code 05M043. Ce code nous permet de déterminer

le GHS d'appartenance du patient, ici 1657, nous donne le tarif qui était de 3581,35 euros en 2009.

L'avantage essentiel de cette nouvelle politique tarifaire est d'harmoniser les modes de financement entre secteur public et secteur privé. Les établissements sont ainsi incités explicitement à mieux gérer, voire à optimiser, leurs ressources en ayant une meilleure connaissance des coûts. Les hôpitaux performants sont mieux rémunérés.

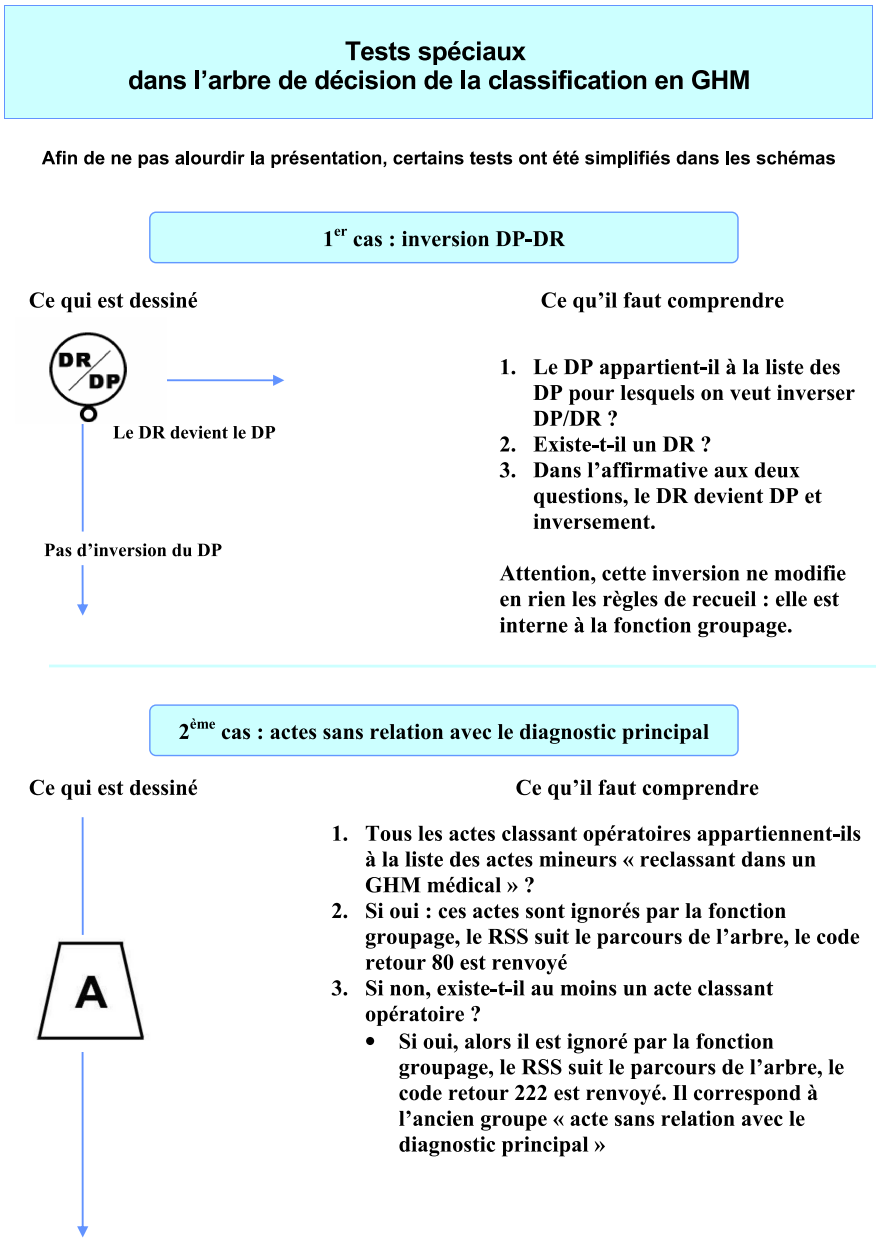

FIGURE B.2 – GHM : Eléments décisionnels (suite)

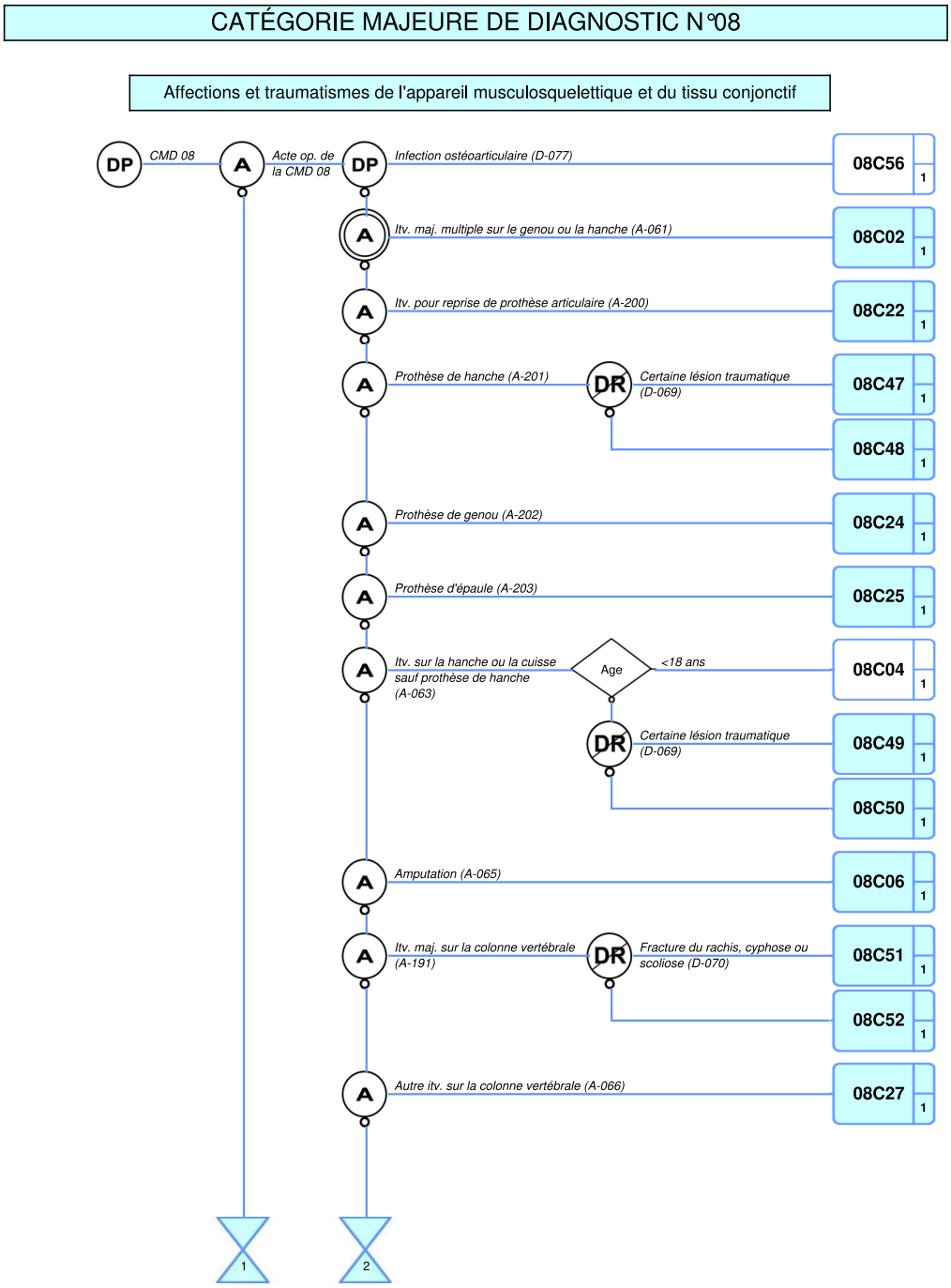

FIGURE B.3 – Exemple d'arbre décisionnel de groupage

Liste de publications

Publications dans des revues internationales

- Alexandre Mazier and Xiaolan Xie. Monte carlo optimization for inpatient admission scheduling with emergency requirement. *European Journal of Operationnal Research*, 2010. submitted.
- Alexandre Mazier, Jean-Charles Billaut, and Jean-François Tournamille. Scheduling preparation of doses for a chemotherapy service. *Annals of Operations Research*, 178 :145-154, 2010.

Conférences et communications internationales

Avec actes

- Alexandre Mazier, Xiaolan Xie, and Marianne Sarazin. Scheduling inpatient admission under high demand of emergency patients. In *6th IEEE Conference on Automation Science and Engineering, CASE 2010, Toronto, Canada*, 2010.
- Alexandre Mazier, Xiaolan Xie, and Marianne Sarazin. Real-time patients assignment : A method for improving emergency department flow. In *Proceedings of the first IEEE Workshop on Health Care Management, Venice, Italy*, 2010.
- Alexandre Mazier and Xiaolan Xie. Scheduling physician working periods of a chemotherapy outpatient unit. In *Proceedings of 13th IFAC symposium on Information Control Problems in Manufacturing, INCOM 2009, Moscow, Russia*, 2009.
- Alexandre Mazier, Jean-Charles Billaut, and Jean-François Tournamille. Scheduling preparation of doses for a chemotherapy service ; a bi-criteria approach. In *Proceedings of the 33rd Conference on Operating Research Applied to Health Services, ORAHS 2007, Saint-Etienne, France*, 2007.
- Alexandre Mazier, Jean-Charles Billaut, and Jean-François Tournamille. Scheduling activities in a chemotherapy service. In *Proceedings of the International Conference on Industrial Engineering and Systems Management, IESM 2007, Beijing, China*, 2007.

Sans actes

- Alexandre Mazier and Xiaolan Xie. Patients assignment with uncertain patient arrivals. *Oral presentation at the 35rd Conference on Operating Research Applied to Health Services, ORAHS 2009, Leuven, Belgium*, 2009.

Conférences et communications francophones

- Alexandre Mazier. Gestion sous incertitude des lits d'hospitalisation *14ème journées STP du GdR MACS, GT Giseh, Paris, France*, 2010
- Alexandre Mazier, Xiaolan Xie, and Marianne Sarazin. Planification des admissions des patients en entrée directe avec prise en charge des urgences. In *5ème Conférence en Gestion et Ingénierie des SystèmEs Hospitaliers, GISEH 2010, Clermont-Ferrand, France*, 2010.
- Pierre-Edouard Reyt, Alexandre Mazier, and Xiaolan Xie. Les méthodes de recherche opérationnelle en santé : Applications aux problèmes d'affectations et d'admissions. In *5ème Conférence en Gestion et Ingénierie des SystèmEs Hospitaliers, GISEH 2010, Clermont-Ferrand, France*, 2010.
- Alexandre Mazier. Gestion en temps réel de l'affectation des patients ; prise en charge des urgences. *Poster à la journée scientifique de l'Institut Fédératif de Recherche Science, Ingénierie et Santé, IFRESIS, Saint Etienne*, 2009.
- Alexandre Mazier. Optimisation d'un hôpital de jour en cancérologie. *Poster à la journée scientifique de l'Institut Fédératif de Recherche Science, Ingénierie et Santé, IFRESIS, Saint Etienne*, 2008.

Bibliographie

Uwe Aickelin et Kathryn Dowsland : An indirect genetic algorithm for a nurse-scheduling problem. *Computers and operations research*, 31:761–778, 2004.

Elif Akcali, Murray J. Côté et Chin Lin : A network flow approach to optimizing hospital bed capacity decisions. *Health Care Management Sciences*, 9:391–404, 2006.

R. Akkerman et M. Knip : Reallocation of beds to reduce waiting time for cardiac surgery. *Health care management science*, 7:119–126, 2006.

Bernard Alesky, Julie Chauvet, Michel Gourgand, Olivier Ponties et Sophie Rodier : Une démarche de modélisation pour les systèmes hospitaliers : Application au nouvel hôpital estaing. *Dans 4ème Conférence Francophone Gestion et Ingénierie des SystèmEs Hospitaliers, GISEH'08, Lausane, Suisse*, 2008.

Md Asaduzzaman, Thierry Chaussalet et Nicola Robertson : A loss network model with overflow for capacity planning of a neonatal unit. *Annals of Operations Research*, 178:67–76, 2010. ISSN 0254-5330. 10.1007/s10479-009-0548-x.

Osei Kwame Asamoah, Steven J. Weiss, Amy A. Ernst, Michael Richards et David P. Sklar : A novel diversion protocol dramatically reduces diversion hours. *The American Journal of Emergency Medicine*, 26:670–675, 2008.

Brent R. Asplin, David J. Magid, Karin V. Rhodes, Leif I. Solberg, Nicole Lurie et Carlos A. Camargo Jr. : A conceptual model of emergency department crowding. *Annals of Emergency Medicine*, 42(2):173 – 180, 2003. ISSN 0196-0644.

Vincent Augusto : *Modélisation, analyse et pilotage de flux en milieu hospitalier à l'aide d'UML et des réseaux de Petri*. Thèse de doctorat, Ecole Nationale Supérieure des Mines de Saint Etienne, 2008.

F. F. Baesler, H. E. Jahnsen et M. Da Costa : The use of simulation and design of experiments for estimating maximum capacity in an emergency room. *Dans Proceedings of the 2003 Winter Simulation Conférence*, 2003.

Adrian Bagust, Michael Place et John W. Posnett : Dynamics of bed use in accomodating emergency admissions : stochastic simulation model. *British Medical Journal*, 319:155–158, 1999.

D. Baubeau, V. Carasco et et. al. : Les passages aux urgences de 1990 à 1998 : une demande de soins non programmés. *DREES*, 72, juillet 2000.

D. Baubeau, V. Carasco et et. al. : Motifs et trajectoires de recours aux urgences. *DREES*, 215, janvier 2003a.

D. Baubeau, V. Carasco et et. al. : Les usagers des urgences, premièrs résultats d'une enquête nationale. *DREES*, 212, janvier 2003b.

Sabrina Bechar et Alain Guinet : Planification des lits d'un établissment de soins. *Dans 3ème Conférence Francophone Gestion et Ingénierie des SystèmEs Hospitaliers, GISEH 06, Luxembourg, Luxembourg*, p. 645–653, 2006.

R. Bekker et A. de Bruin : Time-dependent analysis for refused admissions in clinical wards. *Annals of Operations Research*, 178:45–65, 2010. ISSN 0254-5330. 10.1007/s10479-009-0570-z.

Jeroen Beliën : *Exact and Heuristic Methodologies for Scheduling in Hospitals : Problems, Formulations and Algorithms*. Thèse de doctorat, Faculteit Economishe en Toegepaste Economishe Weteschappen, Katholieke Universiteit Leuven, 2006.

Jeroen Beliën et Erik Demeulemeester : Scheduling trainees at a hospital department using a branch-and-price approach. *European Journal of Operational Research*, 175(1):258 – 278, 2006. ISSN 0377-2217.

Jeroen Beliën et Erik Demeulemeester : Building cyclic master surgery schedules with leveled resulting bed occupancy. *European journal of operational research*, 176:1185–1204, 2007.

Jeroen Beliën et Erik Demeulemeester : A branch-and-price approach for integrating nurse and surgery scheduling. *European Journal of Operational Research*, 189(3):652 – 668, 2008. ISSN 0377-2217.

Rim Ben Bachouch, Alain Guinet et S. Hajri-Gabouj : Gestions des lits mutualisés d'un établissement hospitalier. *Dans 7ème Congrès International de Génie Industriel, Trois-Rivière, Québec, Canada*, 2007.

Nadia Brauner, Gerd Finke, Frederic Prost et Wendelin Serwe : Intégration des méthodes de réécriture et de recherche opérationnelle pour la modélisation et la résolution de contraintes : application à la planification de personnel médical. *Dans Acts of GISEH'03, Lyon, France*, p. 333–339, 2003.

Edmund K. Burke, Peter De Causmaecker, G. Vanden Berge et H. Van Landeghem : The state of the art of nurse rostering. *The Journal of Scheduling*, 7:441–499, 2004.

Xi-Ren Cao : Pertubation analysis of discrete event systems : Concepts, algorithms and applications. *European Journal of Operationnal Research*, 91:1–13, 1996.

Brecht Cardoen : *Operating Room Planning and Scheduling : Solving a Surgical Case Sequencing Problem*. Thèse de doctorat, Faculteit Economishe en Toegepaste Economishe Weteschappen, Katholieke Universiteit Leuven, 2009.

Brecht Cardoen, Erik Demeulmeester et Jeroen Beliën : Operating room planning and scheduling : A literature review. *European Journal of Operational Research*, 201:921–932, 2010.

V. Carrasco : l'activité des services d'urgences en 2004 : une stabilisation du nombre de passages. *DREES*, 524, September 2006.

Ewart R. Carson, Derek G. Cramp, Alastair Morgan et Abdul V. Roudsari : Clinical decision support, systems methodology, and telemedicine : Their role in the management of chronic disease. *IEEE transactions on information technology in biomedicine*, 2(2):80–88, 1998.

M. Carter et S. Ketabi : Bed balancing in surgical wards via block scheduling, a case study at sunnybrook health sciences centre. *Dans Proceedings of the 35^{rd} Operations Research Applied to Health System, ORAHS 2009, Leuven, Belgium*, 2009.

Maxime Cauterman et François Engel : Améliorer la gestion des lits dans les hopitaux et les cliniques. Rapport technique, MeaH, http ://www.meah.sante.gouv.fr, 2007.

Tugba CAYIRLI, Emre VERAL et Harry ROSEN : Designing appointment scheduling systems for ambulatory care services. *Health Care Management Sciences*, 9:47–58, 2006.

M. A. CENTENO, R. GIACHETTI, R. LINN et A. M. ISMAIL : A simulation-ilp based tool for scheduling emergency room staff. *Dans Proceedings of the 2003 Winter Simulation Conférence*, 2003.

Sondes CHAABANE : *Gestion Prédictive des Blocs Opéraroires.* Thèse de doctorat, Institut National des Sciences Appliquées de Lyon, 2004.

Michel CHABROL, Julie CHAUVET et Pierre FÉNIÈS : Un usage multiple du modèle de connaissance unique des processus organisationnels d'une supply chain hospitalière : Application au nouvel hôpital estaing. *Dans 3ème Conférence Francophone Gestion et Ingénierie des SystèmEs Hospitaliers, GISEH 06, Luxembourg, Luxembourg*, 2006.

Julie CHAUVET et Michel GOURGAND : Modélisation et analyse du service des urgences : état de l'art. *Dans 4ème Conférence Francophone Gestion et Ingénierie des SystèmEs Hospitaliers, GISEH'08, Lausane, Suisse*, 2008.

Julie CHAUVET, Michel GOURGAND, Nadine MESKENS et Nikolay TCHERNEV : Modélisation et simulation de la filière des urgences. *Dans 3ème Conférence Francophone Gestion et Ingénierie des SystèmEs Hospitaliers, GISEH 06, Luxembourg, Luxembourg*, 2006.

B. CHEANG, H. LI, A. LIM et B. RODIRGUES : Nurse rostering problems – a bibliographic survey. *European Journal of Operational Research*, 151:447–460, 2003.

J. K. COCHRAN et A. BHARTI : A mutli-stage stochastic methodology for whole hospital bed planning under peak loading. *International Journal of Industrial and Systems Engineering*, 1:8–36, 2006.

Michael M. CONNORS : A stochastic electve admissions scheduling algorithm. Rapport technique, IBM, Health Services Research, 1970.

Stefan CREEMERS et Marc LAMBRECHT : Queueing models for appointment-driven systems. *Annals of Operations Research*, 178:155–172, 2010. ISSN 0254-5330. 10.1007/s10479-009-0646-9.

Amani DAKNOU, Hayfa ZGAYA, Slim HAMMADI et Hervé HUBERT : A dynamic patient scheduling at the emergency department in hospitals. *Dans* 1^{st} *IEEE workshop on health care management, Venice, Italy*, 2010.

H. DARABI, W.L. GALANTER, J.Y.-Y. LIN, U. BUY et R. SAMPATH : Modeling and integration of hospital information systems with petri nets. *Dans Proceedings of IEEE/INFORMS International Conference on Service Operations, Logistics and Informatics, 2009. SOLI '09.*, p. 190 –195, jul. 2009.

Robert W. DAY, Matthew D. DEAN, Robert GARFINKEL et Steven THOMPSON : Improving patient flow in a hospital through dynamic allocation of cardiac diagnostic testing time slots. *Decision Support System*, p. Accepted Manuscript, 2010.

A. de BRUIN, R. BEKKER, L. van ZANTEN et G. KOOLE : Dimensioning hospital wards using the erlang loss model. *Annals of Operations Research*, 178:23–43, 2010. ISSN 0254-5330. 10.1007/s10479-009-0647-8.

H. DE KONING, J. P. S. VERVER, J. van den HEUVEL, S. BISGAARD et R. J. DOES : Lean six sigma in healthcare. *Journal for Healthcare Quality*, 28(2):4–11, 2006.

Peter DEMEESTER, Wouter SOUFFRIAU, Patrick DE CAUSMAECKER et Greet VANDEN BERGHE : A hybrid tabu search algorithm for automatically assigning patients to beds. *Artificial Intelligence in Medecine*, 48:61–70, 2010.

Brian DENTON, James VIAPIANO et Andrea VOGL : Optimization of surgery sequencing and scheduling decisions under uncertainty. *Health Care Management Science*, 10:13–24, 2007.

Melissa DIEFENBACH et Erhan KOZAN : Constructive heuristics for optimum scheduling of emergency departments. *Dans Proceedings of the 35rd Operations Research Applied to Health System, ORAHS 2009, Leuven, Belgium*, 2009.

H. FEI : *Vers un outil d'aide à la planification et l'ordonnancement des blocs opératoires.* Thèse de doctorat, Facultés Universitaires Catholique de Mons, Belgique, 2006.

Michael C. FU et Stacy D. HILL : Optimization of discrete event systems via simultaneous perturbation stochastic approximation. *IIE Transactions*, 29:233–243, 1997.

Micheal FU et Jian-Qiang HU : *Conditional Monte Carlo - Gradient Estimation and Optimization Applications.* Kluwer Academic, 1997.

Pierre FÉNIÈS et Sophie RODIER : Un modèle décisionnel générique pour l'évaluation de lap erformance du processus logistique : Application en contexte hospitalier. *Dans 3ème Conférence Francophone Gestion et Ingénierie des SystèmEs Hospitaliers, GISEH 06, Luxembourg, Luxembourg*, 2006.

Steve GALLIVAN, Martin UTLEY, Tom TREASURE et Oswaldo VALENCIA : Booked inpatient admissions and hospital capacity : mathematical model study. *British Medical Journal*, 324:280–282, 2002.

M. R. GAREY et D. C. JOHNSON : *Computers and Intractability : A Guide to the Theory of NP-Completeness.* Freeman, San Francisco, 1979.

Lalit GARG, Sally MCCLEAN, Brian MEENAN et Peter MILLARD : A non-homogeneous discrete time markov model for admission scheduling and resource planning in a cost or capacity constrained health-care system. *Health Care Management Science*, 13:155–169, 2010. ISSN 1386-9620. 10.1007/s10729-009-9120-0.

Paul GEMMEL et Roland VAN DIERDONCK : Admission scheduling in acute care hospitals : does the practice fit with the theory ? *International Journal of Operations & Production Management*, 19 (9):863–871, 1999.

Na GENG : *Combinatiorial optimization and Markov decision process for planning MRI examinations.* Thèse de doctorat, Ecole Nationale Supérieure des Mines de Saint Etienne, France and Shanghai Jiao Tong University, China, 2010.

Na GENG, Vincent AUGUSTO et Xiaolan XIE : Mri reservation for neurovascular patients. *Dans Proceedings of the fifth annual IEEE international Conference on Automation Science and Engineering, Bangalore, India*, p. 391–396, 2009.

J. GENTLE, W. HÄRDLE et Y. MORI, éditeurs. *Handbook of Computational Statistics*, chapitre 6, Stochastic Optimization, J.C. Spall. Springer, 2004.

Laszlo GERENCSER, Stacy D. HILL, Zsuzsana VAGO et Zoltan VINCZE : Discrete optimization, spsa and markov chain monte carlo methods. *Dans The 2004 American Control Conference*. Boston, Massachusetts, USA, 2004.

B. Glaa, Chirstian Tahon, S. Hammadi et A. O. Daem : Gestion optimisée des infirmiers/infirmières diplomés d'état dans le service des urgences du centre hospitalier de valencienne. *Dans 3ème Conférence Francophone Gestion et Ingénierie des SystèmEs Hospitaliers, GISEH 06, Luxembourg, Luxembourg*, 2006.

Paul Glasserman : *Gradient Estimation Via Perturbation Analysis*. Kluwer Academic, 1991.

F Gorunescu, I McClean, S et H Millard, P : A queuing model for bed-occupency management and planning of hospital. *Journal of the Operational Research Society*, 53:19–24, 2002.

Daisy Gossart, Nadine Meskens et Alain Guinet : Problématique de la gestion des lits d'hospitalisation : état des lieux en belgique et comparaison avec la france. *Dans 5ème Conférence Francophone Gestion et Ingénierie des SystèmEs Hospitaliers, GISEH 2010, Clermont Ferrand, France*, 2010.

Jean Yves Grall : Les maisons médicales de garde. Rapport technique, Ministère de la Santé et des Solidarités, Monsieur le Ministre Xavier Bertrand, juillet 2006.

Linda Green, Sergei Savin et B. Wang : Managing patient demand in a diagnostic medical facility. *Operations Research*, 54(1):11–25, 2006.

Linda Grenn et Sergei Savin : Reducing delays for medical appointments : a queueing approach. *Operations Research*, 56(6):1526–1538, November-December 2008.

Francoise Griveau : Une gestion des lits sous contrôle pour une coordination maîtrisée du flux de patients aux urgences. *Objectif soins*, 161:22–24, October 2007.

Heinz Gröflin, Andreas Klinkert et Nguyen Pham Dinh : Feasible job insertions in the multiprocessor-task job shop. *European Journal of Operational Research*, 185(3):1308 – 1318, 2008. ISSN 0377-2217.

Diwakar Gupta et Brian Denton : Appointment scheduling in health care : Challenges and opportunities. *IIE Transactions*, 40(8):800–819, 2008.

Arnaud Hanset, Nadine Meskens et David Duvivier : Using constraint programming to schedule an operating theatre. *Dans 1^{st} IEEE workshop on health care management, Venice, Italy*, 2010.

P. R. Harper et A. K. Shahani : Modelling for the planning and management of bed capacities in hospitals. *Journal of the Operational Research Society*, 53:11–18, 2002.

Gary Harrison et Gabriel Escobar : Length of stay and imminent discharge probability distributions from multistage models : variation by diagnosis, severity of illness, and hospital. *Health Care Management Science*, 13:268–279, 2010. ISSN 1386-9620. 10.1007/s10729-010-9128-5.

Ruth Hemphill et Bill Nole : Relieving an overcrowded ed and increasing capacity for regional transfers : One hospital's bed management strategies. *Journal of Emergency Nursing*, 31(3):243–246, 2005.

Philip L. Henneman, Michael Lemanski, Howard A. Smithline, Andrew Tomaszewski et Janice A. Mayforth : Emergency department admissions are more profitable than non-emergency department admissions. *Annals of Emergency Medicine*, 53(2):249 – 255.e2, 2009. ISSN 0196-0644.

Joumana Hermassi et Pierre Ladet : Modélisation du processus de prise en charge des appels d'urgence au samu/centre 15. *Dans 4ème Conférence Francophone Gestion et Ingénierie des SystèmEs Hospitaliers, GISEH'08, Lausane, Suisse*, 2008.

Stacy D. Hill : Discrete stochastic approximation with application to resource allocation. *Johns Hopkins APL Technical Digest*, 26(1), 2005.

Yu-Cgi Ho et Xi-Ren Cao : *Pertubation Analysis of Discrete Event Dynamic Systems*. Kluwer Academic, 1991.

L. Jeff Hong et Barry L. Nelson : Discrete optimization via simulation using compass. *Operations Research*, 54(1):115–129, 2006.

L. Jeff Hong et Barry L. Nelson : A framework for locally convergent random-search algorithms for discrete optimization via simulation. *ACM Transaction on Modelling and Computer Simulation*, 17(4), 2007.

Eric Howell, Edward Bessman, Steven Kravet, Ken Kolodner, Robert Marshall et Scott Wright : Active bed management by hospitalists and emergency department throughput. *Annals of Internal Medicine*, 149:804–810, 2008.

Eric Howell, Edward Bessman, Robert Marshall et Scott Wright : Hospitalist bed management effecting throughput from the emergency department to the intensive care unit. *Journal of Critical Care*, 25(2):184 – 189, 2010. ISSN 0883-9441.

X. M. Huang : Decision making support in reshaping hospital medical services. *Health Care Management Science*, 1:165–173, 1998.

XM Huang : A planning model for requirement of emergency beds. *IMA Journal Of Mathematics Applied In Medicine And Biology*, 12(3-4):345–353, SEP-DEC 1995. ISSN 0265-0746.

Anke K. Hutzschenreuter, Peter A. N. Bosman, Ilona Blonk-Altena, Jan van Aarle et Han La Poutré : Agent-based patient admission scheduling in hospitals. *Dans AAMAS '08 : Proceedings of the 7th international joint conference on Autonomous agents and multiagent systems*, p. 45–52, Richland, SC, 2008. International Foundation for Autonomous Agents and Multiagent Systems.

Aida Jebali, Pierre Ladet et A. H. Alouane : Une méthode pour la planification des admissions dans les systèmes hospitaliers. *Dans MOSIM 2003, Toulouse, France*, 2003.

Aïda Jebali : *Vers un outil d'aide à la planification et a l'ordonnancement des ressources dans les services de soins*. Thèse de doctorat, Institut National Polytechnique de Grenoble, 2004.

C Kaandorp, Guido et Ger Koole : Optimal outpatient appointment scheduling. *Health Care Management Science*, 10(3):217–229, 2007.

Gabor D Kelen, Chadd K Kraus, Melissa L McCarthy, Eric Bass, Edbert B Hsu, Guohua Li, James J Scheulen, Judy B Shahan, Justin D Brill et Gary B Green : Inpatient disposition classification for the creation of hospital surge capacity : a multiphase study. *The Lancet*, 368(9551):1984 – 1990, 2006. ISSN 0140-6736.

Jeffrey O. Kephart et David M. Chess : The vision of autonomic computing. *Computer*, 36:41–50, 2003. ISSN 0018-9162.

Yannick Kergosien : *Algorithmes de tournées de véhicules pour l'optimisation des flux de produits et de patients dans un complexe hospitalier*. Thèse de doctorat, Université François Rabelais de Tours, 2010.

Yannick Kergosien, Christophe Lente et Jean-Charles Billaut : Planification de transports de patients entre unités de soins. *Dans 4ème Conférence Francophone Gestion et Ingénierie des SystèmEs Hospitaliers, GISEH'08, Lausane, Suisse*, 2008.

Said KHARRAJA : *Outils d'aide à la planification et l'ordonnancement des plateaux médico-techniques.* Thèse de doctorat, Université Jean Monnet, Saint Etienne, 2003.

J. KIEFFER et J. WOLFOWITZ : Stochastic estimation of the maximum of a regression function. *Annals of Mathematical Statistics*, 23:462–466, 1952.

S. C. KIM, I. HOROWITZ, K. K. YOUNG et T. A. BUCKLEY : Analysis of capacity management of the intensive care unit in a hospital. *European Journal of Operationnal Research*, 115:36–46, 1999.

Seung-Chul KIM, Ira HOROWITZ, Karl K. YOUNG et Thomas A. BUCKLEY : Flexible bed allocation and performance in the intensive care unit. *Journal of Operations Management*, 18:427–443, 2000.

Ali KOKANGUL : A combination of deterministic and stochastic approaches to optimize bed capacity in a hospital unit. *Computer Methods and Programs in Biomedicine*, 90(1):56 – 65, 2008. ISSN 0169-2607.

M. LAGERGREN : What is the role and contribution of models to management and research in health services ? *European Journal of Operational Research*, 105:257–266, 1998.

Mehdi LAMIRI : *Planification des Blocs Opératoires avec Prise en Compte des Aléas.* Thèse de doctorat, Ecole Nationale Supérieure des Mines de Saint Etienne, 2007.

Mehdi LAMIRI, Xiaolan XIE, Alexandre DOLGUI et Frédéric GRIMAUD : A stochastic model for operating room planning with elective and emergency demand for surgery. *European Journal of Operational Research*, 185:1026–1037, 2008.

Sophie LAPIERRE, David GOLDSMAN, Roger COCHRAN et Janice DUBOW : Bed allocation techniques based on census data. *Socio-economic Planning Sciences*, 33:25–38, 1999.

Marek LASKOWSKI, Robert D. MCLEOD, Marcia R. FRIESEN, Blake W. PODAIMA et Attahiru S. ALFA : Models of emergency departments for reducing patient waiting times. *PLoS ONE*, 4(7):e6127, 07 2009.

Matthieu LEDERMAN : La recherche de la performance par l'optimisation des processus. l'exemple du centre hospitalier intercommunal de la côte basque. Mémoire de l'Ecole des Hautes Etudes en Santé Publique, 2008.

Jie LI, Alexandre SAVA et Xiaolan XIE : Simulation-based discrete optimization of stochastic discrete event systems subject to non closed form constraint. *IEEE Transactions on Automatic Control*, 54 (12):2900–2904, 2009.

Jingshan LI et Patricia K. HOWARD : Modeling and analysis of hospital emergency department : an analytical framework and problem formulation. *Dans 6th IEEE Conference on Automation Science and Engineering, CASE 2010, Toronto, Ontario, Canada*, p. 897–902, 2010.

S. J. LITTIG et M. W. ISKEN : Short term hospital occupancy prediction. *Health Care Management Science*, 10:47–66, 2007.

Nelly LITVAK, Marleen van RIJSBERGEN, Richard BOUCHERIE et Mark van HOUDENHOVEN : Managing the overflow of intensive care patients. *European Journal of Operationnal Research*, 185:998–1010, 2008.

Ya LIU, Chengbin CHU et Kanliang WANG : Aggregated state dynamic programming for operating theater planning. *Dans 6th IEEE Conference on Automation Science and Engineering, CASE 2010, Toronto, Ontario, Canada*, p. 1013–1018, 2010.

Peter LUH, Dong CHEN et Lakshman THAKUR : An effective approach for job-shop scheduling with uncertain processing requirements. *IEEE Transactions on Robotics and Automation*, 15(2):328–339, 1999.

Adele MARSHALL, Sally MCCLEAN et Peter MILLARD : Addressing bed costs for elderly : a new methodology for modelling patient outcomes and length of stay. *Health Care Management Science*, 7:27–33, 2004.

J. MATYAS : Random optimization. *Automation and Remote Control*, 26:244–251, 1965.

M. J. MILLER, D. M. FERRIN et M. G. MESSER : Fixing the emergency department : a transformation journey wth edsim. *Dans Proceedings of the 2004 Winter Simulation Conférence*, 2004.

Gergely MINCSOVICS et Nico DELLAERT : Stochastic dynamic nursing service budgeting. *Annals of Operations Research*, 178:5–21, 2010. ISSN 0254-5330. 10.1007/s10479-009-0547-y.

John C. MOSKOP, David P. SKLAR, Joel M. GEIDERMAN, Raquel M. SCHEARS et Kelly J. BOOKMAN : Emergency department crowding, part 1 - concept, causes and moral consequences. *Annals of Emergency Medicine*, 53(5):605–611, 2009a.

John C. MOSKOP, David P. SKLAR, Joel M. GEIDERMAN, Raquel M. SCHEARS et Kelly J. BOOKMAN : Emergency department crowding, part 2 - barriers to reform and strategies to overcome them. *Annals of Emergency Medicine*, 53(5):612–617, 2009b.

M. MURRAY et D.M. BERWICK : Advanced access : reducing waiting and delays in primary care. *Journal of the American Medical Association*, 289:1035–1040, 2003.

J.M. NGUYEN, P. SIX, D. ANTONIOLI, P. GLEMAIN, G. POTEL, P. LOMBRAIL et P. Le BEUX : A simple method to optimize hospital beds capacity. *International Journal of Medical Informatics*, 74(1):39 – 49, 2005. ISSN 1386-5056.

Yasar OZCAN, Marcos LINS, Maria LOBO, Angela da SILVA, Roberto FISZMAN et Basilio PEREIRA : Evaluating the performance of brazilian university hospitals. *Annals of Operations Research*, 178:247–261, 2010. ISSN 0254-5330. 10.1007/s10479-009-0528-1.

C. PAGEL, Martin UTLEY et S. GALLIVAN : The dynamic forecasting of short to medium term workload-based on current case mix. *Dans 33rd Conference on Operational Research Applied to Health Services, ORAHS 2010, Saint Etienne, France*, 2010.

Francoise PAILLASSARD et Linda CASTRO : L'approche lean un outil préalable et indispensable à l'efficacité opérationnelle des systèmes hospitaliers : le cas des services d'urgence. *Dans 5ème Conférence Francophone Gestion et Ingénierie des SystèmEs Hospitaliers, GISEH 2010, Clermont Ferrand, France*, 2010.

Jonathan PATRICK et Martin PUTERMAN : Improving resource utilization for diagnostic services through flexible inpatient scheduling : A method for improving resource utilization. *Journal of the Operations Research Society*, 58:235–245, 2007.

Jonathan PATRICK, Martin PUTERMAN et Maurice QUEYRANNE : Dynamic multipriority patient scheduling for a diagnostic resource. *Operations research*, 56(6):1507–1525, November-December 2008.

J.E. PECORA, A.C. MORETTI, S.D. LAPIERE, P. SORIANO et A. RUIZ : Un algorithme de recherche hybride pour résoudre le problème de confection d'horaires des médecins en salle d'urgence. *Dans 4ème Conférence Francophone Gestion et Ingénierie des SystèmEs Hospitaliers, GISEH'08, Lausane, Suisse*, 2008.

Dinh-Nguyen PHAM et Andreas KLINKERT : Surgical case scheduling as a generalized job shop scheduling problem. *European Journal of Operational Research*, 185(3):1011 – 1025, 2008. ISSN 0377-2217.

J.M. Pines : Moving closer to an operational definition for ed crowdding [letter]. *Academic Emergency Medicine*, 14:382–383, 2007.

Gilda Pour : Prospects for expanding telehealth : Multi-agent autonomic architectures. *Dans International Conference on Computational Intelligence for Modelling Control and Automation, and International Conference on Intelligent Agents, Web Technologies and Internet Commerce (CIMCA-IAWTIC)*. IEEE Computer Society, 2006.

F. J. Ramis, G. L. Palma et F. F. Baesler : The use of simulation for process improvement at an ambulatory surgery center. *Dans Winter Simulation Conference*, 2001.

A. Ramudhin, E. Chan et A. Mokadem : A framework for the modelling, analysis and optimization of pathways in healthcare. *Dans Proceedings of the IEEE Conference on Service Systems and Service Management*, p. 698–702, 2006.

J. C. Ridge, S. K. Jones, M. S. Nielsen et A. K. Shahani : Capacity planning for intensive care units. *European Journal of Operationnal Research*, 105:346–355, 1998.

H. Robbins et S Monro : A stochastic approximation method. *Annals of Mathematical Statistics*, 22:400–407, 1951.

Payman Sadegh : Constrained optimization via stochastic approximation with a simultaneous perturbation gradient approximation. *Automatica*, 33(5):889–892, 1997.

S. Samaha, W. S. Armel et D. W. Starks : The use of simulation to reduce the length of stay in an emergency department. *Dans Proceedings of the 2003 Winter Simulation Conférence*, 2003.

Pablo Santibáñez, Vincent Chow, John French, Martin Puterman et Scott Tyldesley : Reducing patient wait times and improving resource utilization at british columbia cancer agency's ambulatory care unit through simulation. *Health Care Management Science*, 12:392–407, 2009. ISSN 1386-9620. 10.1007/s10729-009-9103-1.

Scott M. Shafer et Timothy L. Smunt : Empirical simulation studies in operations management : context, trends and research opportunities. *Journal of Operations Management*, 22:345–354, 2004.

Sabine Sickinger et Rainer Kolisch : The performance of a generalized bailey–welch rule for outpatient appointment scheduling under inpatient and emergency demand. *Health Care Management Science*, 12:408–419, 2009. ISSN 1386-9620. 10.1007/s10729-009-9098-7.

V. L. Smith-Daniel, S. B. Schweikart et D. E. Smith-Daniel : Capacity management in health care services : review and future research directions. *Decision Sciences*, 19(4):889–919, 1988.

Leif I. Solberg, Brent R. Asplin, Robin M. Weinick et David J. Magid : Emergency department crowding : Consensus development of potential measures. *Annals of Emergency Medicine*, 42(6):824 – 834, 2003. ISSN 0196-0644.

James C. Spall : Multivariate stochastic approximation using a simultaneous perturbation gradient approximation. *IEEE Transactions on Automatic Control*, 37(3):332–341, march 1992.

James C. Spall : Implementation of the simultaneous perturbation algorithm for stochastic optimization. *IEEE Transaction on Aerospace and Electronic Systems*, 34(3):817–823, July 1998a.

James C. Spall : An overview of the simultaneous perturbation method for efficient optimization. *Johns Hopkins APL Technical Digest*, 19(4), 1998b.

M. D. Stevenson, P. A. Oakley, S. M. Beard, A. Brennan et A. L. Cook : Triaging patients with serious head injury : results of a simulation evaluating strategies to bypass hospital without neurosurgical facilities. *Injury*, 32:267–274, 2001.

S. Su et C.-L. Shih : Modeling an emergency medical services system using computer simulation. *International Journal of Medical Informatics*, 72:57–72, 2003.

Duraikannan Sundaramoorthi, Victoria C. P. Chen, Jay M. Rosenberger, Seoung Bum Kim et Deborah F. Buckley-Behan : A data-integrated simulation-based optimization for assigning nurses to patient admissions. *Health Care Management Science*, 13:210–221, 2010.

B. A. Tan, A. Gubaras et N. Phojanamongkolkij : Simulation study of dreyer urgent care facility. *Dans Proceedings of the 2002 Winter Simulation Conférence*, 2002.

Jean-Sébastien Tancrez, Benoit Roland, Jean-Philippe Cordier et Fouad Riane : Etude de la pertubation par les urgences du planning opératoire. *Dans 4ème Conférence Francophone Gestion et Ingénierie des SystèmEs Hospitaliers, GISEH'08, Lausane, Suisse*, 2008.

G Teasdale et B Jennett : *Assessment of coma and impaired consciousness : a practical scale*, vol. 2. Lancet, 1974.

Steven Thomson, Manuel Nunez, Robert Garfinkel et Mathew D. Dean : Efficient short term allocation and reallocation of patients to floor of a hospital during demand surges. *Operations Research*, 57(2):261–273, 2009.

Jiang Tian et Huaglory Tianfield : *Multiagent System Technologies*, chapitre A Multi-Agent Approach to the Design of an E-medicine System, p. 85–94. Springer Verlag, 2003.

Houda Tlahig : *Vers un outil d'aide à la décision pour le choix entre internalisation/externalisation ou mutualisation des activités logistiques au sein d'un établissement hospitalier : cas du service de stérilisation.* Thèse de doctorat, Institut Polytechnique de Grenoble, 2009.

Loraine Trilling : *Aide à la décision pour le dimensionnement et le pilotage de ressources humaines mutualisées en milieu hospitalier.* Thèse de doctorat, Institut National des Sciences Appliquées de Lyon, 2006.

Martin Utley, Steve Gallivan, Kate Davis, Patricia Daniel, Paula Reeves et Jennifer Worrall : Estimating bed requirements for an intermediate care facility. *European Journal of Operational Research*, 150:92–100, 2003.

Martin Utley, Steve Gallivan et Mark Jit : *Health Operations Management*, chapitre 9-How to take variability into account when planning the capacity for a new hospital unit, p. 146–161. Routledge, 2005.

R Van Dierdonck et A. Roth : Hospital ressource planning : concepts, feasability and framework. *Production and Operations Management Journal*, 4(1):2–29, 1995.

Ivan B. Vermeulen, Sander M. Bohte, Sylvia G. Elkhuizen, Han Lameris, Piet J. M. Bakker et Han La Poutré : Adaptive resource allocation for efficient patient scheduling. *Artificial Intelligence in Medicine*, 46:67–80, 2009.

Florent Verstavel : Améliorer la gestion du flux de patients au service des urgences du centre hospitalier de bretagne sud. Mémoire de l'Ecole Nationale de la santé publique, 2007.

Gary M. VILKE, Patty BROWN, Lana an Skogland, Charles SIMMONS et David A. GUSS : Approach to decreasing emergency department ambulance diversion hours. *The Journal of Emergency Medicine*, 26 (2):189–192, 2004.

Jan VISSERS : Patient flow-based allocation of inpatient resources : a case study. *European journal of operational research*, 105:356–370, 1998.

Jan M.H. VISSERS, Ivo J.B.F ADAN et Nico P. DELLAERT : Developing a platform for comparison of hospital admission systems : an illustration. *European Journal of Operational Research*, 180:1290–1301, 2007.

Steven WALCZAK, Walter POFAHL et Ronald SCORPIO : A decision support tool for allocating hospital bed resources and determining required acuity of care. *Decision Support System*, 34:445–456, 2002.

Steven WALCZAK et John SCHARF : Reducing surgical patient costs through use of an artificial neural network to predict transfusion requirements. *Decision Support System*, 30:125–138, 2000.

I.-Jeng WANG et James C. SPALL : Stochastic optimisation with inequality constraintes using simultaneous perturbations and penality functions. *Dans Proceedings of the 42nd IEEE Conference on Decision and Control*, p. 3808–3813. Maui, Hawaii USA, 2003.

I.-Jeng WANG et James C. SPALL : Stochastic optimisation with inequality constraintes using simultaneous perturbations and penality functions. *International Journal of Control*, 81(8):1232–1238, August 2008.

Tao WANG, Alain GUINET, Aissam BELAIDI et Beatrix BESOMBES : Modeling and simulation of emergency services with ARIS and Arena - Case study : the Emergency department of Saint Joseph and Saint Luc Hospital. *Production Planning & Control*, 20(6):484–495, 2009.

Tao WANG, Alain GUINET et Beatrix BESOMBES : A sizing tool for allocation planning of hospital bed ressources. *Dans Proceedings of the 33rd Operations Research Applied to Health System, Saint Etienne, France*, p. 241–250, 2007.

Tao WANG, Alain GUINET et Beatrix BESOMBES : Analyse de la diminution du nombre de changements de lit dans la planification des séjours hospitaliers. *Dans 4ème Conférence Francophone Gestion et Ingénierie des SystèmEs Hospitaliers, GISEH'08, Lausane, Suisse*, 2008.

A. WIINAMAKI et R. DRONZEK : Using simulation in the architectural concept phase of an emergency department design. *Dans Proceedings of the 2003 Winter Simulation Conférence*, 2003.

Peter WILLIAMS, Guangfu TAI et Yiming LEI : Simulation based analysis of patient arrival to health care systems and evaluation of an operations improvement scheme. *Annals of Operations Research*, 178:263–279, 2010. ISSN 0254-5330. 10.1007/s10479-009-0580-x.

Jie XU, Barry L. NELSON et L. Jeff HONG : Industrial strength compass : A comprehensive algorithm and software for optimization via simulation. *ACM Transaction on Modelling and Computer Simulation*, 20(1), 2010.

Bo ZENG, Ayten TURKCAN, Ji LIN et Mark LAWLEY : Clinic scheduling models with overbooking for patients with heterogeneous no-show probabilities. *Annals of Operations Research*, 178:121–144, 2010. ISSN 0254-5330. 10.1007/s10479-009-0569-5.

École Nationale Supérieure des Mines
de Saint-Étienne

NNT : 2010 EMSE 0586

Alexandre Mazier

Stochastic optimization for hospital beds management under uncertainties

Speciality : Industrial Engineering

Keywords : control, emergency, random events, scheduling, optimization, stochastic programming, Monte Carlo optimization

Abstract :

Hospitals have to deals with a lot of random events making their management hard to realize. Those difficulties are mainly due to the uncertainty relative to future evolutions of demand, in particular in term of future arrivals and departures. Despite those difficulties, a fast and efficient hospitalization is required especially for some units like the emergency department. This department has to find quick solution to the problem of hospitalized of their patients. This can only be possible if (i) emergency arrivals are forecasted and so a bed is remaining free for them and/or (ii) the planning of beds occupation is made in a way allowing easy allocations of emergency patients.

Our purpose is going to manage the patient flow in short stay unit (medicine and surgery) starting form the choice of an admission date for each patient until their discharge by keeping in mind the two previous assumptions. By using some stochastic optimization models, we solve a succession of decision problems in order to grant the good state of hospitals. Three level of decision are solved : 1. Admission scheduling for elective patients, 2. Patient assignment to hospital floors, 3. Patient assignment to rooms.

These three levels are firstly tackled independently of the others and secondly used jointly in order to evaluate the performance of the approach. Cases of study are based on data provided by a french hospital partner of this work, Firminy's Hospital Center. We based our experiments on this medium-sized hospital in order to improve the approach of beds management on a real case.

École Nationale Supérieure des Mines
de Saint-Étienne

NNT : 2010 EMSE 0586

Alexandre Mazier

Optimisation stochastique pour la gestion des lits d'hospitalisation sous incertitudes

Spécialité : Génie Industriel

Mots-clefs : pilotage, urgences, imprévus, planification, ordonnancement, optimisation, programmation stochastique, Monte Carlo

Résumé :

Les services de soins hospitaliers sont soumis à de nombreux évènements de natures aléatoires rendant leur gestion et leur pilotage difficiles. Ces difficultés organisationnelles reposent essentiellement sur l'incertitude permanente pesant sur les évolutions futurs, principalement en termes d'arrivées et de départs de patients. Pourtant, une prise en charge rapide et efficace des patients est primordiale pour des services tels que les urgences. Ces services doivent pouvoir placer rapidement leurs patients ce qui n'est possible uniquement si (i) les arrivées ont été anticipées et des places sont laissées vacantes dans les services pour recevoir les patients urgents et/ou (ii) le planning d'occupation des services est construit de telle manière que l'insertion d'un nouveau patient est facilitée.

Notre objectif va être de gérer les flux de patients séjournant dans les services de courts-séjours de l'hôpital, depuis le choix d'admission d'un nouveau patient jusqu'à sa sortie, et ce, en s'inspirant des deux postulats précédant. A l'aide de modèles d'optimisation stochastique, une succession de problèmes de décisions, ayant pour but de garantir le bon fonctionnement des structures hospitalières, est résolue. Une hiérarchie en trois niveaux est appliquée pour résoudre le problème de gestion : 1. Planification des admissions des patients réguliers, 2. Affectation des patients aux unités de soins et insertion des urgences, 3. Affectation des patients d'un service aux chambres.

Ces trois niveaux sont résolus indépendamment les uns des autres puis mis en commun afin d'évaluer la pertinence de l'approche proposée. Les études de cas sont basées sur les données d'un établissement partenaire, le Centre Hospitalier de Firminy (France). Nous nous inspirons de l'exemple de cet hôpital de taille moyenne afin d'évaluer l'impact de la méthode de gestion des flux de patients proposée sur un cas réel.