



**HAL**  
open science

# Accélération des projets de fabrication et modélisation de l'impact de la main d'œuvre additionnelle sur la qualité

Mayassa Bou Orm

► **To cite this version:**

Mayassa Bou Orm. Accélération des projets de fabrication et modélisation de l'impact de la main d'œuvre additionnelle sur la qualité. Gestion et management. Université Paris sciences et lettres, 2017. Français. NNT : 2017PSLED084 . tel-02160073

**HAL Id: tel-02160073**

**<https://theses.hal.science/tel-02160073>**

Submitted on 19 Jun 2019

**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.

# THÈSE DE DOCTORAT

de l'Université de recherche Paris Sciences et Lettres  
PSL Research University

Préparée à l'Université Paris-Dauphine

## Accélération des projets de fabrication et modélisation de l'impact de la main d'oeuvre additionnelle sur la qualité

École Doctorale de Dauphine — ED 543

Spécialité Sciences de gestion

### COMPOSITION DU JURY

#### Directeur de thèse

Jully JEUNET  
Chargée de recherche au CNRS  
HDR en Sciences de gestion  
Université Paris-Dauphine

#### Président du jury

M. Albert DAVID  
Professeur des Universités  
HDR en Sciences de gestion  
Université Paris-Dauphine

#### Rapporteur

Mme Stéphanie BUISINE  
Professeur HDR en Ergonomie et Innovation  
Ecole d'ingénieurs EI.CESI Paris Nanterre

#### Rapporteur

M. Samir LAMOURI  
Professeur des Universités  
Arts et Métiers ParisTech ENSAM

#### Membre du jury

M. Brice MAYAG  
Maître de Conférences en Informatique  
Université Paris-Dauphine

#### Membre du jury

M. Adrien LELIEVRE  
Docteur en Génie Industriel  
Président Directeur Général  
Solutechnic Engineering Electronic

Soutenue le 04.07.2017  
par Mayassa BOU ORM

Dirigée par Jully JEUNET



# Remerciements

En premier lieu, je tiens à exprimer ma profonde reconnaissance à mon directeur de thèse, Jully Jeunet, pour m'avoir guidée et donné la chance de réaliser ces travaux.

Ma gratitude va également à tous les membres du jury. J'exprime mes remerciements à Madame Stéphanie Buisine et à Monsieur Samir Lamouri qui ont accepté d'être rapporteurs de cette thèse. Je remercie également Monsieur Albert David, Monsieur Brice Mayag, et Monsieur Adrien Lelièvre qui me font l'honneur de composer ce jury.

Mes remerciements vont également à Monsieur Mustapha Sali pour ses conseils, à Monsieur Nicolas Bertrand, Directeur Industriel d'Alstom Transport à Belfort et à tous mes anciens collègues d'Alstom Transport qui m'ont aidée à concrétiser ce travail.

Je remercie tous mes collègues doctorants ainsi que les membres du personnel de l'université.

Enfin, je remercie toutes celles et tous ceux qui m'ont encouragée et soutenue tout au long de mon projet de thèse.

*Je dédie cette thèse à l'âme de mon frère Houssam, à mes parents, mes frères, ma soeur et surtout à mon mari et mes enfants qui se sentiront honorés par l'aboutissement de ce travail.*

# Table des matières

<b>Introduction générale</b>	<b>9</b>
<b>1 Notions de base en gestion de projet</b>	<b>15</b>
1.1 Caractéristiques d'un projet . . . . .	17
1.1.1 Liens de précedence entre les activités d'un projet . . . . .	17
1.1.2 Représentation et complexité d'un projet . . . . .	19
1.1.3 Définitions et dépendance des phases . . . . .	21
1.1.4 Typologie . . . . .	22
1.1.4.1 Projets de conception . . . . .	22
1.1.4.2 Projets de fabrication . . . . .	23
1.1.4.3 Stratégie modulaire dans la conception et la fabrication	24
1.2 Durée d'un projet . . . . .	25
1.2.1 Durée d'exécution d'une activité . . . . .	26
1.2.2 Méthode du chemin critique . . . . .	27
1.3 Ressources affectées à un projet . . . . .	27
1.3.1 Typologie et contraintes des ressources . . . . .	28
1.3.2 Coût des ressources . . . . .	29
1.3.3 Détermination du budget d'un projet . . . . .	30
1.4 Accélération d'un projet . . . . .	31
1.4.1 Accélération et compression . . . . .	31
1.4.2 Arbitrage durée - ressource . . . . .	33
1.4.3 Arbitrage durée - coût . . . . .	35
1.5 Conclusion . . . . .	36
<b>2 Le concept de qualité en gestion de projet</b>	<b>37</b>
2.1 Evolution historique de la qualité . . . . .	38
2.2 Evaluation de la qualité . . . . .	40
2.2.1 Qualité subjective et qualité objective . . . . .	41
2.2.2 Indicateurs quantitatifs . . . . .	42
2.2.2.1 Performance d'un produit . . . . .	42
2.2.2.2 Coût de la qualité et de la non qualité . . . . .	44
2.2.3 Indicateurs qualitatifs . . . . .	46
2.2.3.1 Check-list . . . . .	46
2.2.3.2 Descriptif des niveaux de qualité . . . . .	49
2.3 Mesures quantitatives de la qualité individuelle et globale . . . . .	50
2.3.1 Qualité individuelle d'une activité . . . . .	50

2.3.1.1	Taux de conformité à une check-list . . . . .	50
2.3.1.2	Estimation de la qualité selon plusieurs critères . . . . .	51
2.3.1.3	Niveaux discrets de qualité et seuils de qualité minimum . . . . .	51
2.3.1.4	Concept de qualité cumulée . . . . .	52
2.3.1.5	Incertitude de mesure de la qualité . . . . .	52
2.3.2	Qualité globale d'un projet . . . . .	53
2.4	Gestion et amélioration de la qualité . . . . .	54
2.4.1	Management par la qualité totale (TQM) . . . . .	54
2.4.1.1	Définition et origine de la méthode TQM . . . . .	54
2.4.1.2	Démarche TQM . . . . .	55
2.4.1.3	Champs d'application . . . . .	56
2.4.2	Six Sigma . . . . .	56
2.4.2.1	Définition de la méthode Six Sigma . . . . .	56
2.4.2.2	Révolution Six Sigma et facteurs clés de succès . . . . .	57
2.4.2.3	Applications et limites . . . . .	58
2.4.3	Lean Management . . . . .	58
2.4.3.1	Définition et concepts de Lean Management . . . . .	58
2.4.3.2	Mise en oeuvre d'une démarche Lean Management . . . . .	59
2.4.3.3	Lean management en dehors des zones de production . . . . .	59
2.5	Conclusion . . . . .	60
<b>3</b>	<b>Modèles d'arbitrage durée - coût en ordonnancement de projets</b>	<b>61</b>
3.1	Typologie des modèles d'arbitrage durée - coût . . . . .	62
3.1.1	Cas d'une ressource unique . . . . .	62
3.1.2	Cas de plusieurs ressources . . . . .	63
3.2	Modèles d'arbitrage avec une ressource unique non renouvelable . . . . .	65
3.2.1	Arbitrage coût - durée en temps continu (CTCTP) . . . . .	66
3.2.1.1	Fonction linéaire entre le coût et la durée . . . . .	67
3.2.1.2	Fonction convexe du coût d'accélération . . . . .	67
3.2.1.3	Fonction concave . . . . .	68
3.2.2	Modèles d'arbitrage coût - durée en temps discret (DTCTP) . . . . .	68
3.2.2.1	Modèles avec un objectif unique et variantes du problème DTCTP . . . . .	69
3.2.2.2	Détermination de la courbe efficiente coût - durée . . . . .	71
3.3	Modèles à contraintes de ressources multiples . . . . .	72
3.3.1	Méthodes exactes . . . . .	73
3.3.2	Heuristiques constructives fondées sur des règles de priorité . . . . .	76
3.3.3	Méthodes de recherche locale . . . . .	79
3.3.4	Métaheuristiques à population de solutions . . . . .	80
3.3.5	Approches multi-objectif . . . . .	85
3.4	Conclusion . . . . .	88
<b>4</b>	<b>Modèles d'arbitrage durée, coût et qualité</b>	<b>90</b>
4.1	Modèles mono-objectif et approches agrégées . . . . .	91
4.1.1	Qualité et reprise . . . . .	92
4.1.2	Modélisations en temps continu (CTCTP et Qualité) . . . . .	96

4.1.3	Modélisations multi-mode avec une seule ressource non renouvelable (DTCTP et Qualité) . . . . .	100
4.2	Optimisation multi objectif par des méthodes non agrégées fondées sur Pareto . . . . .	102
4.2.1	Algorithmes génétiques multi-objectif . . . . .	103
4.2.2	Autres algorithmes évolutionnaires multi-objectif . . . . .	106
4.3	Conclusion . . . . .	108
<b>5</b>	<b>Impact de l'augmentation des ressources sur la productivité et la qualité</b>	<b>110</b>
5.1	Typologie des ressources . . . . .	110
5.1.1	Ressources renouvelables : la main d'oeuvre . . . . .	110
5.1.2	Hypothèses relatives à la main d'oeuvre en termes de compétences et d'utilisation . . . . .	112
5.1.2.1	Efficiences homogènes et statiques . . . . .	113
5.1.2.2	Efficiences hétérogènes et dynamiques . . . . .	115
5.1.2.3	Hypothèses sur la main d'oeuvre retenues pour la modélisation . . . . .	116
5.2	Relation entre la productivité et la main d'oeuvre additionnelle . . . . .	117
5.2.1	Productivité des travailleurs temporaires . . . . .	118
5.2.2	Productivité et sureffectif des travailleurs en heures supplémentaires . . . . .	120
5.3	Impact de l'accélération sur la qualité . . . . .	121
5.3.1	Qualité et adjonction de main d'oeuvre . . . . .	121
5.3.2	Niveaux de tolérance aux défauts . . . . .	122
5.4	Conclusion . . . . .	123
<b>6</b>	<b>Un modèle d'arbitrage durée, coût et qualité d'un projet avec adjonction de travailleurs temporaires et heures supplémentaires</b>	<b>124</b>
6.1	Objectif et positionnement du modèle dans la littérature . . . . .	124
6.2	Notations et description du problème . . . . .	127
6.3	Modélisation de base avec travailleurs temporaires . . . . .	129
6.3.1	Objectif et contraintes usuelles . . . . .	129
6.3.2	Contraintes sur la consommation de ressources . . . . .	130
6.4	Modélisation de l'impact de l'accélération sur la productivité . . . . .	132
6.4.1	Relation entre sureffectif et productivité . . . . .	132
6.4.2	Etude des bornes sur la main d'oeuvre en heures régulières . . . . .	135
6.4.3	Linéarisation de l'équation de la main d'oeuvre corrigée . . . . .	137
6.5	Ajout des heures supplémentaires . . . . .	138
6.5.1	Analyse du sureffectif de travailleurs en heures supplémentaires	139
6.5.2	Contraintes additionnelles relatives à l'utilisation des heures supplémentaires . . . . .	140
6.6	Modélisation de l'impact de l'accélération sur la qualité . . . . .	142
6.6.1	Accélération avec heures supplémentaires uniquement . . . . .	143
6.6.2	Accélération avec sureffectif uniquement . . . . .	144
6.6.3	Accélération en combinant heures supplémentaires et sureffectif avec pertes de productivité . . . . .	147

6.6.4	Linéarisation et formulation de la contrainte d'un seuil de qualité minimum à respecter . . . . .	152
6.7	Coûts et contrainte de budget . . . . .	153
6.8	Pré-traitement, paramétrage et résumé du programme d'optimisation	154
6.8.1	Pré-traitement . . . . .	155
6.8.2	Paramétrage du modèle . . . . .	156
6.8.3	Résumé du programme d'optimisation . . . . .	158
6.9	Conclusion . . . . .	159
<b>7</b>	<b>Etude de cas relative au projet de fabrication d'une motrice TGV</b>	<b>161</b>
7.1	Description et données du projet de fabrication de la motrice . . . . .	162
7.2	Niveaux de tolérance aux défauts et seuils de qualité minimum . . . . .	166
7.3	Valeurs des paramètres dans l'expérience de simulation . . . . .	170
7.4	Résultat des simulations . . . . .	175
7.5	Conclusion . . . . .	181
	<b>Conclusion générale</b>	<b>183</b>
<b>A</b>	<b>Annexe technique - méthodes de résolution pour les problèmes combinatoires</b>	<b>185</b>
A.1	Optimisation mono-objectif . . . . .	185
A.1.1	Classement des problèmes et typologie des méthodes de résolution . . . . .	185
A.1.2	Méthodes exactes pour les problèmes linéaires en nombres entiers	187
A.1.2.1	La méthode par séparation et évaluation (Branch and Bound) . . . . .	189
A.1.2.2	La méthode des plans sécants (Cutting-Plane) . . . . .	192
A.1.2.3	Algorithmes de coupes et branchements (Branch and Cut) . . . . .	192
A.1.2.4	Génération de colonnes . . . . .	193
A.1.3	Heuristiques dédiées et méthodes de recherche locale . . . . .	193
A.1.3.1	Les heuristiques dédiées . . . . .	193
A.1.3.2	Méthodes de descente . . . . .	193
A.1.3.3	Recuit simulé (Simulated Annealing, SA) . . . . .	194
A.1.3.4	Recherche Taboue (Tabu Search, TS) . . . . .	194
A.1.4	Métaheuristiques à population de solutions . . . . .	194
A.1.4.1	Algorithme génétique (Genetic Algorithm, GA) . . . . .	194
A.1.4.2	Optimisation par colonies de fourmis (Ant Colony Optimization, ACO) . . . . .	196
A.1.4.3	Optimisation par essaims particulaires (Particle Swarm Optimization, PSO) . . . . .	197
A.1.4.4	Optimisation par colonies d'abeilles (Artificial Bee Colony, ABC) . . . . .	198
A.1.4.5	Compétition impérialiste (Imperialist Competitive Algorithm, ICA) . . . . .	199
A.2	Optimisation multi-objectif . . . . .	200
A.2.1	Présentation générale des approches de résolution . . . . .	200

A.2.2	Les méthodes agrégées dites méthodes scalaires . . . . .	202
A.2.2.1	La méthode $\epsilon$ -contrainte . . . . .	202
A.2.2.2	Somme pondérée . . . . .	202
A.2.2.3	Goal Attainment et Goal Programming . . . . .	202
A.2.3	Méthodes non agrégées non Pareto . . . . .	202
A.2.3.1	Méthode lexicographique . . . . .	202
A.2.3.2	Vector Evaluated Genetic Algorithm (VEGA) . . . . .	203
A.2.4	Les méthodes non agrégées fondées sur Pareto . . . . .	203
A.2.4.1	Dominance, Pareto optimalité, frontière de Pareto et classement Pareto . . . . .	204
A.2.4.2	Algorithme évolutionnaire multi-objectif générique (Generic MOEA, Multi Objective Evolutionary Al- gorithm) . . . . .	206
A.2.4.3	Stratégies de préservation de la diversité . . . . .	209
A.2.4.4	Elitisme . . . . .	210
A.2.4.5	Pseudo code et couplage avec principales métaheu- ristiques . . . . .	212
<b>B</b>	<b>Résumé du programme d'optimisation</b>	<b>213</b>
<b>C</b>	<b>Glossaire des notations</b>	<b>220</b>
<b>D</b>	<b>Liste des abréviations</b>	<b>222</b>
	<b>Bibliographie</b>	<b>223</b>

# Introduction générale

« Le XXème siècle est le siècle de la productivité, le XXIème sera celui de la qualité. », Joseph Juran, 1993, Harvard Business Review.

En 2007, Alstom Transport a battu le record du monde de vitesse sur rail, en atteignant 574,8 km/h lors des essais menés sur son TGV V150, sur la ligne à grande vitesse Est-Européenne. La vitesse étant un indicateur de qualité, un but affiché par Alstom est de vendre le TGV à l'étranger face à la concurrence du japonais Shinkansen (fabriqué par Hitachi) et de l'allemand ICE (Siemens). Avec une pression concurrentielle toujours plus forte, les entreprises cherchent non seulement à développer des produits nouveaux mais aussi à en améliorer la qualité pour accroître leur longévité et occuper un segment de marché particulier avec l'avantage concurrentiel qui lui est associé.

**Histoire de la qualité** Au cours du XXème siècle, la qualité a évolué de l'inspection unitaire du produit (début des années 1900) au contrôle par échantillonnage statistique (1930 à 1945). Jusqu'aux environs du milieu du XXème siècle, la qualité était considérée comme « de la belle ouvrage » ou du « bon travail ». Après la seconde guerre mondiale, l'assurance qualité est apparue au début des années 1950 au Japon dans le but d'améliorer la qualité et de créer la confiance en ce qu'une entité satisfera aux exigences pour la qualité. À cette époque de l'histoire de la qualité, un produit de qualité était un produit respectant certaines normes. Aujourd'hui la pression concurrentielle oblige les entreprises à satisfaire les exigences croissantes des consommateurs en termes de prix et qualité, de sorte la qualité constitue une préoccupation stratégique majeure des gestionnaires. Depuis 1990 à nos jours, le Management par la Qualité Totale (TQM) devient la méthode de gestion de la qualité la plus appliquée dans les entreprises visant l'excellence et la satisfaction « totale » du client. Cette méthode est « basée sur la participation de toute la population d'une organisation et visant la réussite à long terme par la satisfaction de la clientèle » (ISO 8402 [79]). Ce type de stratégie s'applique à la fois aux processus de fabrication et à la qualité globale des projets eux-mêmes. Par qualité globale, on désigne la qualité d'un projet qui résulte d'une agrégation des qualités individuelles de ses activités (souvent la moyenne), différente donc de la qualité totale qui désigne une approche stratégique. Nous adoptons le concept de qualité objective qui consiste à évaluer la conformité du produit à des spécifications écrites et tangibles contrairement à la qualité subjective qui vise à satisfaire des exigences intangibles.

**Typologie des projets** Nous nous intéressons aux projets de fabrication consistant à produire à la demande un bien mature et stable qui ne requiert pas ou peu de modifications. A la différence de ce type de projets, les projets de conception incluent une phase d'étude qui précède une phase de développement d'un prototype. Ces deux phases peuvent se chevaucher non seulement pour réduire la durée du projet mais également pour permettre des retours de la production afin de modifier itérativement le prototype en cours d'exécution et inversement. Ces deux phases sont ainsi liées par des contraintes de précédence que l'on peut qualifier de souples. Les projets de fabrication que nous considérons dans cette thèse sont décrits par une gamme opératoire qui liste les activités à exécuter dans un certain ordre traduisant des relations de précédence strictes entre elles. Si une activité possède plusieurs activités qui la précèdent, celles-ci peuvent s'exécuter en parallèle mais ce n'est pas du chevauchement qui consiste symboliquement à casser des relations de précédence.

### **Pourquoi cherche t-on à réduire la durée d'un projet et comment ?**

Pour tout projet, qu'il s'agisse d'un projet de conception ou de fabrication, on cherche à en minimiser la durée pour se garantir un avantage concurrentiel ou pour respecter un délai de livraison. Pour les concepteurs, l'objectif principal est de réduire le délai de mise sur le marché d'un nouveau produit (Time to Market). Dans les projets de fabrication, on cherche à minimiser la durée pour se positionner dans le cadre d'un appel d'offre ou pour pouvoir commencer au plus tôt un autre projet dans un objectif de maximisation du profit. Il est aussi possible de raccourcir la durée d'un projet à la demande du client pour avancer le délai de livraison si le produit fait l'objet de primes d'avance. Dans d'autres cas, la minimisation de la durée n'est pas planifiée mais peut devenir un objectif à atteindre si l'on s'aperçoit en cours d'exécution que le projet est en retard en raison de l'occurrence d'aléas tels que l'absentéisme, les retards de livraison de sous-ensembles achetés ou le traitement de non conformités externes. Ces aléas sont généralement intégrés au niveau tactique en planifiant par exemple le recrutement d'un nombre de temporaires suffisants pour absorber le risque d'absentéisme ou en constituant des stocks de sécurité. En revanche, d'autres aléas difficilement prévisibles comme les grèves ou les conditions climatiques ne sont pas pris en compte en amont et leur survenue peut obliger les gestionnaires à accélérer les activités restantes à exécuter pour pouvoir livrer le produit dans le délai convenu ; le respect des délais de livraison permettant à moyen terme à l'entreprise de préserver la confiance que ses clients lui accordent. Cette thèse se situe dans un cadre déterministe et tactique où l'on raisonne en moyenne de sorte que seuls certains aléas comme l'absentéisme sont incorporés dans la planification à moyen terme de la production.

Il existe plusieurs stratégies pour réduire la durée d'un projet. S'il s'agit d'un projet de conception, on utilise le chevauchement (overlapping) qui consiste à exécuter en parallèle (totalement ou partiellement) deux ou plusieurs activités initialement planifiées d'une manière séquentielle. Le chevauchement se caractérise par le fait qu'une activité aval démarre avant la fin d'une activité amont (avec laquelle elle a un lien de précédence), sur la base d'une information incomplète qui comporte

alors de l'incertitude (Tyagi et al., 2011 [143]). Certains travaux considèrent que cette incertitude peut être réduite par une politique de communication destinée à affiner l'information ; une telle politique nécessitant toutefois une durée et un coût additionnels (voir par exemple Lin et al., 2010 [100]). Le chevauchement peut être comparé à une stratégie de fabrication modulaire caractérisée par sa divisibilité en plusieurs modules (sous-projets) exécutables séparément et en parallèle. Cette stratégie modulaire, initialement développée pour faire face à la complexité des grands projets de conception, a récemment été déployée dans l'industrie automobile et dans le secteur de la construction immobilière, notamment. En revanche, certains projets de fabrication complexes comme un avion ou un train ne peuvent pas être totalement modulaires. Un TGV par exemple est composé de plusieurs modules (voitures, moteurs, motrices, etc.) qui sont fabriquées et assemblées dans différents sites industriels mais chacun de ces modules est un projet de fabrication en soi où les liens de précédence sont stricts et donc non modulaires. Il en est de même pour un immeuble, on peut fabriquer des modules (blocs) qu'on assemble ensuite sur site mais toujours avec certaines contraintes de précédence qu'on ne peut pas violer. Le moyen employé pour réduire la durée de ce type de projets est l'adjonction de ressources additionnelles ou plus performantes, c'est ce que l'on appelle l'accélération d'un projet (Mokhtari et al., 2011 [107]). Ainsi, dans cette thèse nous nous intéressons à l'accélération des projets de fabrication non modulaires où les activités sont donc liées par des contraintes de précédence strictes et où l'on considère la non-préemption c'est-à-dire l'impossibilité d'interrompre l'exécution d'une activité en cours pour l'achever ultérieurement. Dans les structures d'assemblage la non-préemption est fondamentale, la préemption impliquant des arrêts de chaîne.

**Impacts de la réduction de la durée** L'accélération réduit la durée d'un projet mais à un coût additionnel relatif à l'augmentation des ressources engagées. Il existe donc un arbitrage entre la durée et le coût. Ce problème d'arbitrage durée - coût (Time Cost Trade off Problem, TCTP) a fait l'objet de nombreuses contributions dans la littérature, avec cependant une rareté des modélisations qui considèrent explicitement la relation entre la durée et la quantité de ressources utilisés ; celle-ci se résumant la plupart du temps à une expression en termes de coût. Par ailleurs, il est communément admis que l'accélération implique une dégradation de la qualité. Si l'on ajoute le critère de qualité aux côtés de la durée et du coût, on obtient le problème d'arbitrage durée, coût et qualité (Time Cost Trade Off Problem, TCQTP). En effet, respecter les délais, maîtriser les coûts et produire un bien de qualité sont trois objectifs contradictoires et interdépendants. Dans les contributions qui considèrent ce problème, la cause de cette dégradation de qualité n'est pas explicitement modélisée, chaque activité peut être exécutée selon plusieurs modes d'exécution exprimés par des triplets durée, coût et qualité associés à des options d'utilisation des ressources. La plupart du temps, la qualité individuelle d'une activité dans un mode donné fait l'objet d'une estimation, au même titre que le coût et la durée dans ce mode, sans qu'aucune relation directe avec la consommation des ressources ne soit formulée. Ces modélisations cherchent à maximiser la qualité globale d'un projet exprimée le plus souvent comme une moyenne des qualités des activités qui le composent, sans considération de seuils de qualité minimum à atteindre pour toute activité. Cette agrégation par la moyenne implique que certaines activités peuvent

présenter des défauts inacceptables qui seront compensés globalement par des qualités très élevées pour d'autres activités. Ce constat nous a amenés à considérer des seuils de qualité minimum à atteindre pour chacune des activités. Chez Alstom transport par exemple, des mesures de la gravité des défauts sont faites par une cotation issue d'une AMDEC (Analyse des Modes de Défaillance, de leurs Effets et de leur Criticité). Selon cette cotation, le défaut est acceptable ou non. Il en est de même dans l'industrie automobile chez Renault et Nissan. Depuis 2001, l'AVES (Alliance Vehicle Evaluation Standard) est le standard d'évaluation de la qualité globale du véhicule fini, utilisé pour classer les défauts du point de vue du client en défauts graves, très gênants, assez gênants, peu gênants. Cette classification signifie que, pour toute activité individuelle, certains défauts peuvent être tolérés, tandis que d'autres nécessitent une reprise.

**Problématique** Pour les raisons que nous venons d'analyser, nous formulons les interrogations suivantes. *Comment peut-on planifier l'adjonction de ressources additionnelles pour accélérer un projet de fabrication tout en maîtrisant les coûts et en maintenant pour chaque activité un niveau acceptable de qualité ? Comment peut-on modéliser l'impact de l'accélération sur la qualité, le coût et la durée d'un projet ?* Notre problématique consiste ainsi à formuler une relation directe entre l'utilisation des ressources et la qualité et entre l'adjonction des ressources et la perte de productivité, jusqu'à présent absente de la littérature.

**Proposition d'un modèle de planification de la main d'oeuvre additionnelle** Nous proposons une modélisation en temps discret de la relation d'arbitrage entre la durée, le coût et la qualité d'un projet lorsque l'accélération résulte de l'adjonction d'heures supplémentaires et de travailleurs temporaires. Ce modèle s'inscrit dans le cadre des décisions tactiques. Nous considérons uniquement la main d'oeuvre comme ressource car c'est la seule ressource que l'on peut faire varier à moyen terme ; le raisonnement est mené à capacité de production pré-déterminée. L'investissement dans de nouveaux équipements ou technologies peut permettre une accélération d'un projet mais il s'agit d'une décision stratégique qui sort du cadre de cette thèse. Par ailleurs, comme le soulignent Pour et al. (2010) [119] ce type d'investissement peut s'accompagner d'une augmentation de la qualité tandis qu'à moyen terme, l'hypothèse d'une relation de décroissance entre la durée et la qualité est pertinente. Par ailleurs, c'est un accroissement de la main d'oeuvre qui implique le risque le plus important de détérioration de la qualité et de la productivité. Les heures supplémentaires peuvent être mises en oeuvre rapidement mais impliquent une fatigue et donc une baisse de la qualité et de la productivité du travail (Mohamad et al., 2012 [105]). Le recours aux travailleurs temporaires quant à lui est susceptible d'entraîner un sureffectif caractérisé par un nombre de travailleurs supérieur à la capacité normale d'un poste et peut également dégrader la qualité (Gunduz, 2004 [65]).

Comme nous nous intéressons aux projets de fabrication, nous supposons que l'on dispose d'une main d'oeuvre répartie en plusieurs spécialisations professionnelles

avec une productivité identique des travailleurs sans phénomène d'apprentissage ou d'oubli car nous excluons la polyvalence. Nous considérons donc le système de production à un état stable en termes de productivité. Les pertes de productivité sont supposées croître linéairement avec le taux de sureffectif (sureffectif rapporté à l'effectif en mode normal, c'est-à-dire sans accélération). Il s'agit d'une hypothèse en cohérence avec les résultats de plusieurs études (voir Gunduz, 2004 [65]). Nous formulons la qualité comme une fonction linéaire décroissante du sureffectif et des heures supplémentaires et montrons que la relation entre la qualité et la durée n'est pas linéaire en raison des pertes de productivité liées au sureffectif. En effet, ces pertes de productivité impliquent une utilisation accrue du sureffectif entraînant mécaniquement un surcroît de dégradation de la qualité. Dans le modèle que nous proposons chaque activité possède un niveau de tolérance aux défauts qui se traduit par l'existence d'un seuil minimum de qualité à respecter en cas d'accélération. Ce seuil est celui en dessous duquel l'activité présente des défauts qui doivent être repris. Ainsi, le modèle planifie la main d'oeuvre additionnelle en évitant les reprises qui sont généralement couteuses en temps et en matériaux. Notons que la reprise peut, en plus de son coût et sa durée, impliquer des défauts supplémentaires susceptibles de retarder la fabrication comme par exemple l'endommagement d'une pièce lors du démontage.

Nous formulons le problème de minimisation de la durée d'un projet sous contrainte de budget et de qualité comme un programme linéaire mixte où les variables sont le nombre de travailleurs permanents et temporaires et le nombre d'heures supplémentaires affectés à l'exécution de chaque activité à chaque période de l'horizon de planification du projet. Cette optimisation mono-objectif nous permet de nous affranchir de l'utilisation de méthodes de recherche de la frontière de Pareto, avec pour résultat, un ensemble souvent très grand de solutions non dominées parmi lesquelles il est difficile de choisir même en faisant intervenir les préférences des décideurs. Considérer comme objectif principal celui de la minimisation de la durée a d'autant plus de sens au regard à l'ordre de grandeur des pénalités de retard pour les projets industriels. Un retard de production d'une journée dans une usine peut coûter à l'entreprise jusqu'à plusieurs millions d'euros (Bubshait, 2003 [29]). Si ces pénalités sont aussi importantes, c'est parce que le commanditaire a lui même pris des engagements auprès de sa propre clientèle qui lui impose aussi le paiement de pénalités de retard. C'est ainsi qu'en 2014, la Société Nationale des Chemins de Fer (SNCF) a réclamé plus de cent millions d'euros à Alstom et Bombardier suite à des retards de livraison de l'ordre d'une année. Un tel retard empêche la SNCF de mettre en circulation les rames à la date prévue et ne peut donc pas atteindre ses objectifs de ventes planifiées pour rentabiliser ses investissements. Finalement un an de retard représente des pertes d'opportunité (vente) sur un an pour la SNCF. Si la qualité d'un projet se place au dernier rang des objectifs à optimiser, c'est parce que cette qualité est plus difficile à définir et à évaluer bien que ce critère commence à recevoir une attention croissante dans la littérature récente (Basu et al., 2014 [17]). En effet, avec une hausse des exigences en termes de qualité, les nouveaux contrats obligent les décideurs à l'intégrer dans leurs objectifs (Mungle et al., 2013 [109]).

Le présent document est organisé en sept chapitres de la manière suivante. Le chapitre 1 a pour objet de rappeler les notions de base en gestion de projet. Nous donnons plusieurs définitions et méthodes relatives à la durée, aux ressources et au budget d'un projet. Nous introduisons ensuite le problème d'arbitrage durée - ressource et durée - coût. Dans le chapitre 2 nous analysons le concept de qualité et son évolution. Nous présentons également les différentes façons d'évaluer la qualité et ses mesures quantitatives. Nous distinguons le concept de qualité objective que nous adoptons dans cette thèse du concept de qualité subjective. Enfin nous résolvons les différentes stratégies d'amélioration de la qualité. Le chapitre 3 analyse la littérature sur les modèles d'arbitrage coût et durée. Nous présentons d'abord les modèles qui n'intègrent qu'une seule ressource puis les différentes approches pour modéliser et résoudre les problèmes d'ordonnancement de projets avec des ressources multiples et contraintes. Le chapitre 4 est consacré à une revue de la littérature sur le problème d'arbitrage coût - durée - qualité où l'on cherche soit à ramener ce problème à un seul objectif par des méthodes agrégées, soit à trouver les solutions Pareto optimales au problème multi-objectif. Les hypothèses de notre modèle sont discutées dans le chapitre 5. Le modèle est ensuite développé dans le chapitre 6 où nous analysons l'impact du sureffectif et des heures supplémentaires sur la durée et la qualité. Ce modèle de planification de la main d'oeuvre additionnelle est appliqué dans le chapitre 7 au cas d'un projet de fabrication d'une motrice TGV. Cette thèse s'achève par une conclusion générale qui donne les limites de notre approche et fournit plusieurs pistes possibles d'amélioration.

# Chapitre 1

## Notions de base en gestion de projet

Un projet est défini par la norme X50-105<sup>1</sup> comme « une démarche spécifique qui permet de structurer méthodiquement et progressivement une réalité à venir ». Cette norme s'attache ensuite au livrable, en ajoutant « un projet est mis en œuvre pour élaborer une réponse au besoin d'un utilisateur, d'un client ou d'une clientèle. Il implique un objectif et des actions à entreprendre avec des ressources données ». Un projet se déroule ainsi entre un maître d'ouvrage (le client) et un maître d'œuvre (le fournisseur) pour aboutir à un produit ou un service. Les actions à entreprendre dans un projet sont des activités reliées entre elles par des liens de précedence plus ou moins souples.

Une activité est un ensemble d'opérations effectuées sur un même poste de travail. Par exemple une activité d'assemblage peut consister à effectuer plusieurs opérations élémentaires d'assemblage. Ainsi, pour assembler un moteur (activité d'assemblage d'un moteur), on assemble successivement plusieurs composants (opération unitaire d'assemblage d'un et un seul composant sur le moteur).

Chaque activité se caractérise par un contenu en travail c'est-à-dire une quantité de ressources nécessaires à son exécution (Fundeling et Trautmann, 2010 [62]), des dates au plus tôt (début et fin au plus tôt) et des dates au plus tard (début et fin au plus tard). Une date de début au plus tôt d'une activité est égale à la longueur du plus long chemin de l'activité « début » du projet à cette activité, c'est-à-dire la première date possible qui respecte tout les liens de précedence de cette activité. Une date de début au plus tard est la date maximum à laquelle on peut exécuter une activité sans retarder le projet, avec une date limite d'achèvement imposée. La différence entre la date de début au plus tôt et la date de début au plus tard d'une activité est sa « marge totale ». Cette marge représente le retard qu'on peut se permettre sur une activité sans remettre en cause la date de fin du projet. Une marge totale nulle signifie que l'activité est une activité critique c'est-à-dire toute modification de sa durée implique une modification de la durée totale du projet.

L'ordonnancement d'un projet consiste à considérer, en plus des contraintes de précedence entre les activités, les contraintes des ressources. Une ressource néces-

---

1. Norme X50-105 : Le management de projet - Concepts, de l'association française de normalisation (AFNOR)

saire à l'exécution d'une activité peut être un moyen humain ou matériel mais ces moyens ne sont pas illimités. On note que la notion « ressource » existe depuis des milliers d'années, bien avant l'apparition de la notion « projet ». Citons par exemple les premiers projets du monde tels que les pyramides égyptiennes, la grande Muraille de Chine, les cathédrales qui ont nécessité de nombreuses ressources avec des disponibilités surement contraintes (ouvriers, esclaves, architectes, etc.) et ce pendant plusieurs années (durée de projets). Parmi les ressources d'un projet, nous citons le chef de projet qui assure la fonction de direction de projet (project management pour les anglo-saxons) et le contrôleur de projet qui assure la fonction de gestion de projet (project control). Ces sont des acteurs qui ne rentrent pas dans l'exécution du projet c'est-à-dire qui n'impactent pas directement sa durée d'exécution mais qui sont indispensables pour son déroulement.

La direction ou le management de projet consiste à fixer les objectifs du projet (délais, qualité, sécurité d'emploi, etc.) et les moyens mise en oeuvre (ressources matérielles et humaines) qui conduisent à attribuer un budget au projet, ainsi qu'à l'évaluation des risques et la direction des objectifs des travailleurs. La gestion de projet consiste à suivre l'avancement du projet par rapport aux prévisions, à maîtriser les objectifs en s'assurant d'une bonne gestion de ressources et à apporter à la direction tout élément qui peut impacter l'un des objectifs (Giard, 1991 [64]). La notion de management de projet est donc plus large que celle de gestion de projet. Dans cette thèse on s'intéresse aux critères de performance d'un projet dans le cadre de la gestion de projet.

Respecter les délais, maîtriser les coûts et produire de la qualité, connus par le « triangle QCD », sont les trois objectifs interdépendants de tout projet. Le respect du délai d'un projet est un critère primordial de performance en gestion de projet. Même si le planning est rarement respecté, intervient alors le respect du délai relatif aux concurrents et aux clauses de calculs de pénalités contractuelles. La durée et le coût sont souvent liés. Parfois on cherche à respecter le délai pour éviter des pertes économiques liées aux pénalités contractuelles ou au lancement tardif d'un nouveau produit sur le marché. On peut autrement investir dans la réduction de la durée d'un projet pour introduire un autre projet et augmenter le profit. Les stratégies de réduction de la durée dépendent en fait de la typologie des projets comme nous allons le clarifier dans ce chapitre.

Ce chapitre est composé de trois sections et a pour objet de rappeler les notions de base en gestion de projet. La section 1.1 présente les caractéristiques de base d'un projet, y compris les liens de précedence, la représentation graphique et les phases traditionnelles. Dans cette section nous donnons aussi une typologie des projets fondée sur la distinction entre les projets de conception et les projets de fabrication. Dans les sections 1.2 et 1.3 nous présentons la façon dont on calcule la durée et le coût d'un projet, respectivement. Nous abordons ensuite la stratégie d'accélération d'un projet dans la section 1.4 où l'on introduit les problèmes d'arbitrage durée-ressource (sous-section 1.4.2) et durée-coût (sous-section 1.4.3) qui seront détaillés dans le chapitre 3. Enfin, la section 1.5 conclut ce chapitre en relevant notamment

les méthodes appropriées pour déterminer la durée et le coût d'un projet dans un cadre déterministe.

## 1.1 Caractéristiques d'un projet

Un projet se caractérise principalement par les liens de précédence ou les dépendances entre ses activités. Il est généralement représenté par « un graphe de précédence » qui décrit ces liens. Néanmoins, ces liens de précédence peuvent être plus ou moins souples voir inexistantes selon la typologie de projet (projet de fabrication ou de conception). Etant donnée qu'une phase est un ensemble d'activités reliées par des contraintes de précédence. Chaque projet n'a pas les mêmes phases d'exécution. Prenons par exemple un projet de conception d'un nouveau produit, la phase primordial dans ce projet est la phase de conception mais elle ne l'est pas dans un projet de production en série d'un produit stable. Une présentation précise des différents types de dépendance entre les activités d'un projet est donnée dans la sous-section 1.1.1. Dans la sous-section 1.1.2 nous illustrons la représentation graphique d'un projet en abordant la notion de « complexité ». Nous rappelons ensuite les phases traditionnelles d'un projet (sous-section 1.1.3) sans distinction entre les projets à ce stade. Dans la sous-section 1.1.4, nous proposons une typologie de projets en distinguant les projets de conception des projets de fabrication par l'intermédiaire des phases constituant le projet et la nature (souple ou dure) des liens de précédence entre ses activités. Dans cette dernière sous-section, nous abordons également la notion de « modularité » en gestion de projet en clarifiant surtout pourquoi cette notion n'entre pas dans le cadre de notre thèse.

### 1.1.1 Liens de précédence entre les activités d'un projet

Par défaut, une relation de précédence signifie qu'une activité ne peut débiter que si certaines autres activités sont finies. Or, nous distinguons quatre types de dépendance standards entre les activités d'un projet :  $SS$ ,  $FF$ ,  $FS$ ,  $SF$  correspondant aux relations Start to Start (début à début), Finish to finish (fin à fin), Finish to Start (fin à début) et Start to Finish (début à fin), respectivement. Une relation  $SS_{ij}$  signifie que l'activité  $j$  ne peut pas commencer tant que l'activité  $i$  n'a pas démarré. Exemple, tirer une chape en béton ne peut pas se faire avant de commencer à couler le béton. Une relation  $FF_{ij}$  signifie que l'activité  $j$  ne peut pas terminer tant que l'activité  $i$  n'est pas terminée. Elles ne doivent pas forcément terminer en même temps :  $j$  peut se terminer à n'importe quel moment une fois l'activité  $i$  terminée. Exemple : une équipe installe des câbles dans un bâtiment et inspecte le système électrique en même temps. Tant que l'installation de câble (activité  $i$ ) est en cours d'exécution, l'inspection du système électrique (activité  $j$ ) ne peut pas se terminer. Dans le cas d'une relation  $FS_{ij}$ , l'activité  $j$  ne peut débiter que lorsque l'activité  $i$  se soit achevée. Il s'agit du type de relation par défaut dans un projet, et celui qui est le plus souvent utilisé. Exemple : grattage des murs ( $i$ ) doit être terminé avant que le peintre puisse commencer l'enduit peinture ( $j$ ). Et enfin, la relation  $SF_{ij}$  inversement à la relation  $FS_{ij}$  signifie que l'activité  $j$  ne peut se terminer tant

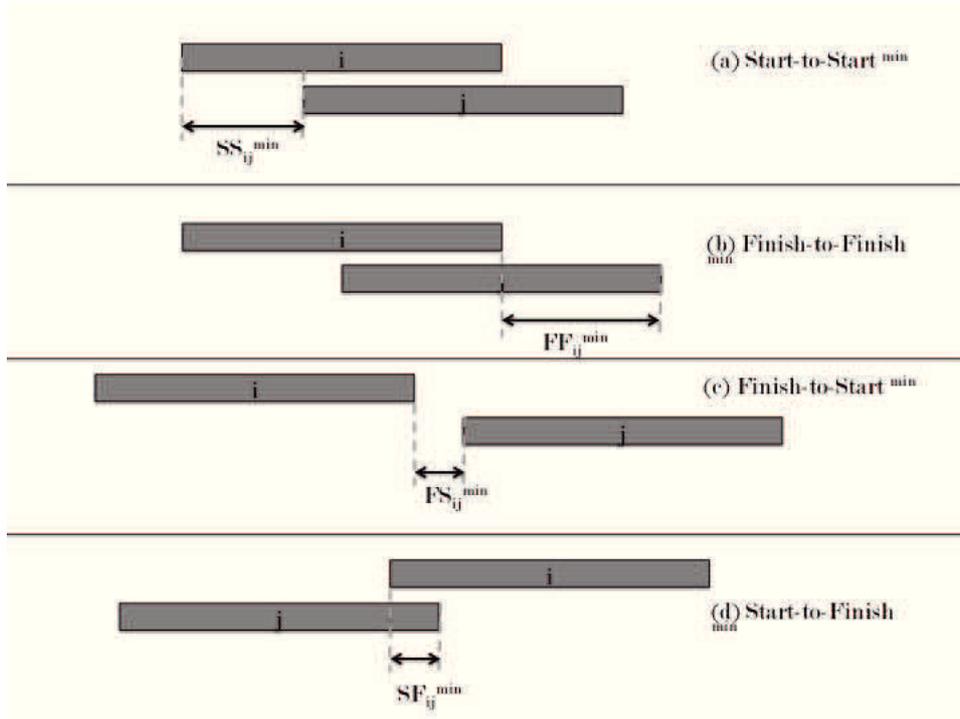


FIGURE 1.1 – Dépendance entre les activités et temps de décalage minimal

que l'activité  $i$  n'a pas commencé. Exemple, on ne peut pas terminer l'assemblage des composants d'un produit ( $j$ ) tant que la livraison n'a pas commencé ( $i$ ).

En plus de ces liens de base, il existe ce que l'on appelle « le temps de décalage minimal » c'est-à-dire le nombre d'unité de temps minimum à décaler entre deux activités ( $SS^{min}$ ,  $FF^{min}$ ,  $FS^{min}$ ,  $SF^{min}$ ).

Un temps de décalage minimal  $SS_{ij}^{min}$  représentée à la figure 1.1(a) signifie que l'activité  $j$  peut débuter une fois que  $SS_{ij}^{min}$  unités de temps se sont écoulées au minimum depuis le début de l'activité  $i$ . Exemple, une activité de plâtrage ( $j$ ) ne peut débuter qu'un certain temps écoulé après une activité d'isolation ( $i$ ). On note ici, qu'un projet peut contenir des non-activités c'est-à-dire des opérations outre que les opérations de production, comme le temps de séchage de plâtre ou de la peinture, qui impliquent également un temps de décalage minimal entre deux activités.

Dans la figure 1.1(b), l'activité  $j$  ne peut se terminer que  $FF_{ij}^{min}$  unités de temps après que  $i$  se soit finie. Exemple, l'installation électriques ( $j$ ) d'une chaîne de production ne peut se terminer que certain temps après l'installation des équipements mécaniques ( $i$ ).

Dans la figure 1.1(c),  $j$  ne peut débuter que  $FS_{ij}^{min}$  unités de temps minimum après que  $i$  se soit achevée. Exemple, une activité de plâtrage ( $i$ ) et de finition de plâtre ( $j$ ).

Dans la figure 1.1(d),  $j$  ne peut se terminer que si  $SF_{ij}^{min}$  unités de temps minimum se sont écoulées depuis le début de  $i$ . Exemple, une livraison ( $i$ ) et une

facturation ( $j$ ).

### 1.1.2 Représentation et complexité d'un projet

Un projet est généralement représenté par un graphe acyclique, appelé « graphe de précedence », qui décrit les relations de précedence entre ses activités. On distingue deux types de graphes de précedence : AOA (Activity On Arc) et AON (Activity On Node). Dans un graphe AOA comme il est représenté à la figure 1.2(a), deux activités successives  $j$  et  $k$  sont présentées sur les arcs, le réseau est plutôt orienté vers les événements (jalons, activités terminées, etc. ). Un graphe AON est plutôt orienté vers les activités et les dépendances et c'est la représentation la plus simple généralement appliquée par les industriels. La figure 1.2(b) illustre un graphe AON où l'activité  $j$  précède l'activité  $k$  et l'arc qui les relie présente le lien de précedence classique de type  $FS$ .

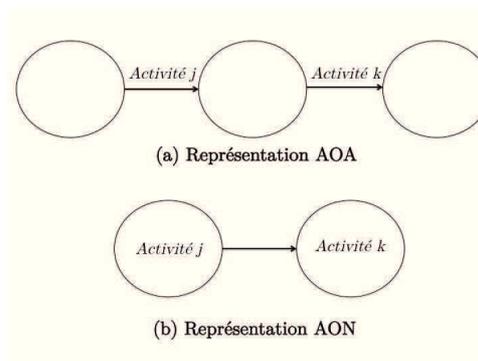


FIGURE 1.2 – Représentations AOA et AON

Nous illustrons ces deux représentations sur un exemple d'un projet de mise en place d'une unité de production inspiré de Pinedo (2005) [118], et que nous référons par « Exercice UP » pour simplifier la lecture. Le tableau 1.1 décrit les huit activités de ce projet (colonne 2) ainsi que leurs durées d'exécution en semaines (colonne 3) et les prédécesseurs et successeurs de chaque activité (colonne 4 et 5).

#	Dénomination des activités	Durée en semaines	Pred.	Succ.
1	Conception des outils de production	4	—	4
2	Préparation des dessins de fabrication	6	—	5
3	Préparation de l'unité de production	10	—	6, 7
4	Achat des outils	12	1	6, 7
5	Achat des pièces de production	10	2	6
6	Préparation des kits de pièces	2	3, 4, 5	8
7	Installation des outils	4	3, 4	8
8	Test	2	6, 7	—

Tableau 1.1 – Description « Exercice UP » (Pinedo, 2005)

Le graphe AOA de ce projet est représenté à la figure 1.3(a). Dans ce graphe l'activité 0 est une activité fictive ayant une durée égale à zéro mais nécessaire pour visualiser la relation de précédence qui existe entre l'activité 4 et l'activité 6, autrement cette relation aurait pas été prise en considération dans la planification ni dans l'exécution du projet. Le graphe AON du projet est représentée à la figure (b). On note que dans la représentation AON que nous adoptons dans cette thèse, les relations de précédence sont plus visibles. Pour un réseau complet du projet, tel qu'il est représenté à la figure 1.4, il suffit de rajouter au graphe de précédence une activité fictive de début qui précède les activités 1 et 2 et une activité fictive de fin qui succède l'activité 8.

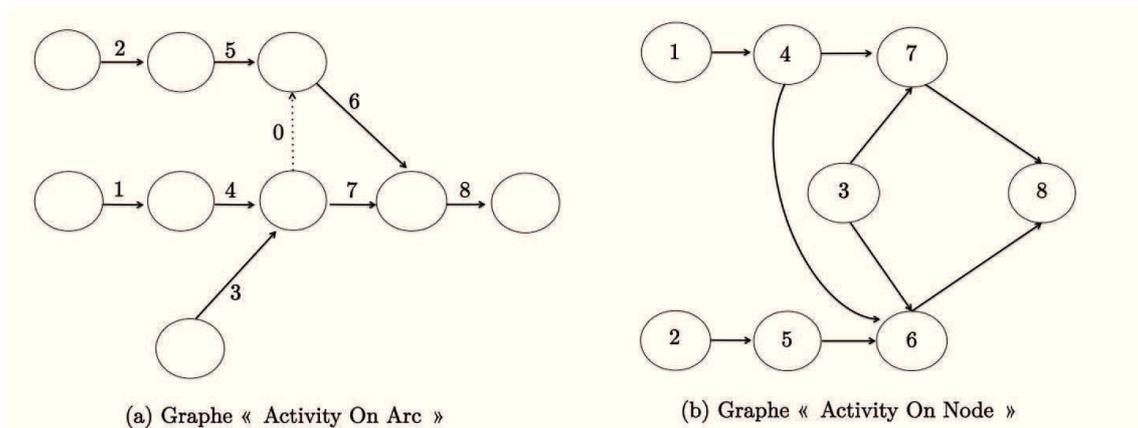


FIGURE 1.3 – Graphes de précédence « Exercice UP »

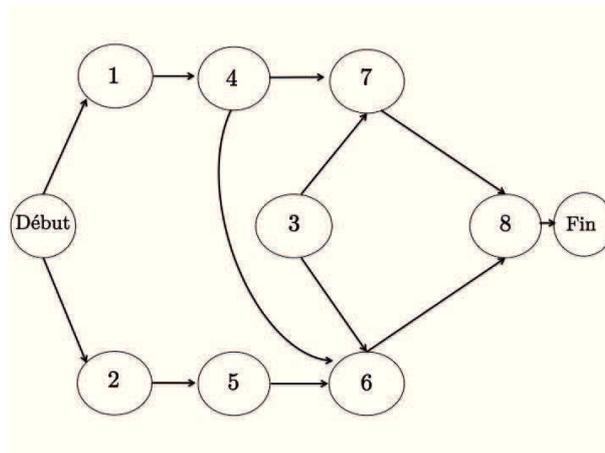


FIGURE 1.4 – Réseau AoN « Exercice UP »

Cependant, quand on parle des graphes, il convient d'évoquer la notion de complexité du projet traduite souvent par un indice de complexité. Cet indice introduit par Bein et al. (1992) [18] représente le nombre minimum de noeuds après une réduction série/parallèle du réseau de projet. Par exemple, pour une réduction série, dans la figure 1.3(b), les arcs qui relient les activité 2, 5 et 6 peuvent être remplacés

par un seul arc qui relie 2 et 6 et donc on réduit le nombre de noeuds en passant de 3 à 2 pour cette série d'activités. Cet indice est donc en croissance avec la croissance de la taille du projet ou le nombre d'activités. Un autre indicateur de complexité utilisé dans la littérature est le coefficient de complexité qui représente, selon Akkan et al. (2005) [7], le rapport entre le nombre d'arcs et le nombre de noeuds d'un graphe, il reflète donc la complexité des liens de précedence entre les activités.

### 1.1.3 Définitions et dépendance des phases

Un projet se caractérise principalement par son cycle de vie, qui est constitué de plusieurs phases. Le nombre et la description des phases peuvent varier d'un domaine à un autre et d'une culture d'entreprise à une autre. Par exemple, les phases d'un projet de construction d'une autoroute ne peuvent pas être identiques aux phases d'un projet de développement d'un programme informatique. Or, le cycle de vie d'un projet peut se résumer par les cinq phases traditionnelles suivantes : développement d'un concept, planification de projet, conception de produit et des processus, production et gestion du service après vente.

1. Développement d'un concept : l'objectif principale de cette première phase est d'identifier les besoins d'un client implicites ou explicites, d'étudier la faisabilité de points de vue technique, financier, social, environnemental et concurrentiel et enfin de transformer l'idée (ou le concept) en projet.
2. Planification de projet : à l'issue de cette phase, le budget et le planning de projet sont fixés, y compris la planification et l'identification des ressources nécessaires à l'exécution de ses différentes activités. Cette phase consiste également à planifier les démarches qualité à mettre en oeuvre (gestion de qualité, méthodes de contrôles, indicateurs, etc.,...).
3. Conception de produit et des processus : par conception de produit (ou product design) on désigne les spécifications techniques et les plans définis par le bureau d'études ou R&D pour tout produit manufacturé. La conception des processus (process design) consiste ensuite à définir les processus du projet ainsi que les relevés de contrôles techniques comme le contrôle dimensionnel par exemple.
4. Exécution : c'est la phase opérationnelle du projet qui consiste en l'exécution de ses activités et au pilotage des critères de performance (durée, coût et qualité) fixés en phase de planification. On note que, particulièrement, dans les grands projets industriels, on passe souvent par une phase de prototype puis une phase de pré-série avant de lancer la production en série. Un prototype est un produit techniquement identique à celui qui doit être fabriqué en série, mais il est réalisé en quantité unitaire ou limitée par des ressources différentes ou dans des sites de production différents. Une pré-série est une quantité limitée des unités identiques à la série et fabriqués à l'aide de ressources identiques à celles qui seront utilisées pour la série. L'objectif d'une telle stratégie est de tester la fonctionnalité du produit (par le prototype) et le processus de production (par la pré-série).
5. Gestion du service après vente : c'est la phase de post-production qui ne rentre pas techniquement dans le lancement d'un produit ou d'un service

mais qui reste indispensable dans le cycle de vie d'un projet. Cette phase consiste à la mise en oeuvre de la gestion et l'applicabilité des garanties. Elle comprend également la production et le stockage des pièces de rechange après la fin de production du produit concerné. Elle peut être considérée comme une phase génératrice de coûts à intégrer immédiatement dans le prix de vente des produits commercialisés, ou inversement comme une source de gains financiers.

### 1.1.4 Typologie

Les caractéristiques de base que nous avons introduites dans les sections précédentes et dans l'introduction du chapitre sont homogènes pour tout type de projet. Or, les projets peuvent se différer en termes de processus, phases, critères de performance, souplesse des liens de précedence, nature de client etc. Par exemple, un grand projet de construction comme le tunnel sous la manche diffère, par son processus, son objectif et son niveau d'exigence, d'un projet de développement d'un logiciel informatique. Sur la base de ces différences, nous proposons une classification de projets en projets de conception ( sous-section 1.1.4.1) et projets de fabrication (sous-section 1.1.4.2). Pour approfondir cette distinction, nous abordons dans la sous-section 1.1.4.3 la stratégie modulaire en conception et en fabrication afin de relever ses particularités dans le cas de projets de fabrication.

#### 1.1.4.1 Projets de conception

L'objet principal d'un projet de conception est le développement d'un produit nouveau ou d'un service, nécessitant donc les trois premières phases traditionnelles d'un projet décrites dans la section précédente. Ces projets peuvent également avoir une phase de production de prototype qui nécessiterait ou non des modifications sur la conception initiale. Le client d'un projet de conception peut être connu ou non. Dans les deux cas, l'objectif principal de ce type de projet est la date de mise à disposition sur le marché (Time to Market). Dans ce type de projets les liens de précedence sont dites « souples » c'est-à-dire les activités peuvent s'exécuter indépendamment, totalement ou partiellement en parallèle, dans un ou plusieurs lieux physiques différents. Par exemple, la conception de la carrosserie d'une voiture et la conception de processus de l'assemblage du moteur d'un même projet peuvent s'exécuter parallèlement par deux sous traitants, deux bureaux d'études ou deux concepteurs du projet, en tenant compte bien évidemment des contraintes techniques. La réduction du Time to Market de ce type de projet fait donc classiquement recours à la stratégie de chevauchement qui consiste à exécuter des activités (ou des phases) en parallèle tandis qu'elle sont initialement planifiées d'une manière séquentielle. Il convient de préciser que le chevauchement n'est qu'un aspect du concept de l'ingénierie concourante. Dans la littérature existante sur le chevauchement, ces deux termes sont généralement confondus et notamment lorsqu'il s'agit de projets de conception. Néanmoins l'ingénierie concourante est une stratégie globale plus large qui comprend des principes culturels et comportementaux, des principes de structuration des activités où on retrouve le chevauchement et les principes de développement de produits dans le sens d'amélioration de la qualité.

Un projet de conception peut contenir une phase d'exécution de prototype avant de lancer une grande série d'un produit techniquement compliqué. Cette phase permet la remise en cause de la conception initiale (ou *redesign*) pour corriger les anomalies détectées.

Parmi les projets de conception, nous distinguons les projets de conception réglée et les projets de conception innovante.

La conception réglée consiste à changer la conception d'un produit existant selon un cahier des charges bien défini et pour un client connu mais pas forcément déterminé c'est-à-dire la conception ne se fait pas forcément sur commande. L'objectif principal de ce type de projet est de développer des produits plus satisfaisants pour leur clientèle pour augmenter les bénéfices. Inversement, il peut parfois être pertinent économiquement de délivrer une satisfaction similaire au client pour un prix inférieur. Prenons l'exemple des nouvelles véhicules DACIA du groupe Renault (Logan, Sandero, Duster, Fokker. . .) sont plus simples que les véhicules Renault haute gamme. Ils intègrent moins de technologies et de fonctionnalités. Par conséquent, ils sont moins chers à produire et donc plus accessibles à certaines catégories de clients.

Un produit ou service innovant, dans le cadre de projets de conception innovante implique la révision de l'identité même d'un objet. Un Iphone par exemple revisite l'identité du téléphone. Ce type de projet nécessite principalement une phase de développement de concept. On note qu'il existe deux types d'innovation, l'innovation « incrémentale » et « radicale » (Tyagi et al., 2011 [143]). L'innovation incrémentale se réfère à l'invention de nouvelles fonctionnalités ou à l'amélioration des fonctionnalités existantes, alors que l'innovation radicale se réfère à l'introduction d'un produit « inconnu » avec un ensemble de caractéristiques originales et un degré de rupture élevé. Il convient de préciser qu'un projet de conception innovante est un projet à rentabilité contrôlée c'est-à-dire qui concerne des produits nouveaux qui peuvent être vendus à un client non connu (potentiel) et sur un marché concurrentiel.

#### 1.1.4.2 Projets de fabrication

Un projet de fabrication se caractérise principalement par l'absence de phases de développement et de conception. Il s'agit d'un projet d'exécution à la demande d'un produit mature et stable qui ne requière pas des modifications conceptuelles (on est plus dans la phase de prototypage). On exclue de cette typologie les projets qui nécessitent une phase de conception stratégique c'est-à-dire lorsqu'il s'agit d'un produit qui va suivre des évolutions adaptatives même pendant sa phase d'exécution. Par exemple, Renault Scenic est un produit qui connaît une évolution permanente, un travail de conception est systématiquement effectué pour intégrer des nouvelles exigences au produit. On note que ce cas peut être classé dans la typologie de projet d'innovation incrémentale. On ajoute qu'un projet de fabrication est un projet à coûts contrôlés qui se caractérise par la négociation des spécifications techniques, d'un budget et d'un délai avec un client bien déterminé. Ces négociations deviennent donc contractuelles et ne changent pas tout au long du projet. Les bénéfices du

maître d'oeuvre dépendront de sa bonne maîtrise de ces objectifs. Un retard ou avance de livraison peuvent aboutir à des pénalités de retard ou des primes d'avance, respectivement.

Le projet de fabrication aboutit à un produit manufacturé comme un bâtiment, un pont, une voiture, etc. En effet, par projets de fabrication nous désignons non seulement les projets de fabrication industrielle mais aussi les projets de construction immobilière en exécution. Un projet de fabrication industrielle est un projet en chaîne, exécuté dans un lieu physique unique et qui produit selon un Takt Time ou un rythme de production (cadence). Chaque activité du projet est exécutée sur un poste de travail bien défini.

Nous citons les projets de fabrication de produit unique comme un chantier naval (chaque navire est unique) et les projets de fabrication série à la demande dépendante c'est-à-dire sur commande d'un client bien précis. Dans les projets de construction, le produit est lancé en construction lorsque l'architecture et les caractéristiques d'ingénierie civil sont converties en spécifications bien définies, il peut être unique (comme un pont, une maison) ou non (ensemble d'immeubles identiques ou avec des caractéristiques identiques).

Dans cette typologie de projet, les contraintes de précédence sont dites « dures » c'est-à-dire toute activité ne peut commencer qu'après l'achèvement de ces prédécesseurs. Par exemple, dans la fabrication d'une motrice TGV, on ne peut pas poser une baie vitrée avant d'avoir assemblé le nez de carénage (tête de la motrice). Dans une construction d'un pont, on ne peut pas procéder au dallage béton avant d'avoir installé la structure métallique. Ainsi la réduction de la durée d'un tel projet ne peut pas se faire par un simple recours au chevauchement. En pratique, la durée d'un projet de fabrication est réduite par le recours aux ressources additionnelles, ce que l'on appelle l'accélération. Nous présentons quelques notions liées à cette stratégie dans la section 1.4.

#### **1.1.4.3 Stratégie modulaire dans la conception et la fabrication**

La modularité est une « orientation stratégique dont l'objectif est de réduire simultanément la complexité, les temps et les coûts de fabrication, puise ses fondements théoriques dans les perspectives tracées dès 1962 par Herbert Simon » (Cohendet et al., 2005 [34]). Cette stratégie consiste à décomposer un produit en plusieurs modules (ou sous-composants) constituant des entités fonctionnelles autonomes (une fonction = un module). Les modules d'un produit sont indépendamment développés dans différents sites avant d'être intégrées au système global. Cette stratégie est plus aisée lorsque les liens de précédence sont souples comme c'est le cas dans les projets de conception.

Néanmoins, la modularité transforme aujourd'hui le monde de l'industrie notamment en automobile (Frigant et Bernard, 2014 [59]). Cette stratégie est récemment largement utilisée dans la production de différents secteurs comme l'automobile mais aussi l'aéronautique, le ferroviaire et la construction immobilière. La production modulaire signifie l'externalisation d'une partie de processus global d'un produit,

c'est-à-dire une famille de composants est fabriquée et assemblée à l'intérieur d'un module autonome dans un site industriel indépendant. En aéronautique comme chez Airbus, les modules de l'avion, différenciés par leur fonctionnalité technique, sont fabriqués dans plusieurs sites avant d'être assemblés. Néanmoins, toujours en aéronautique, Boeing n'est pas un produit modulaire, il est fabriqué et assemblé à 100% aux états unis pour plusieurs raisons techniques et autres (confidentialité, politique, etc.). Toutefois, les produits de ces secteurs (avion, voiture, train, etc.) ne peuvent pas être totalement modulaires. Prenons l'exemple d'un TGV, on note qu'un train TGV est décomposé en plusieurs modules (voitures, moteurs, motrices, etc.) qui sont fabriqués et assemblés dans différents sites industriels mais chacun de ces modules reste un produit complexe en soi ou autrement dit un projet de fabrication complexe qui nécessite de centaines d'activités liées entre elles par des liens de précedence dites « dures » et sur lesquelles la souplesse n'est pas envisagée.

Un autre contre exemple, dans le bâtiment, est l'utilisation des blocs modulaires prêts à être assemblés qui certes impliquent une économie de temps et de l'argent (moins de fournisseurs à gérer, moins de risque qualité,..) mais cette modularité ne supprime pas le besoin en mains d'oeuvre pour l'assemblage, l'ajustement des blocs ni les activités de finitions à exécuter sous certaines contraintes de précedence. Comme dans une chaine de production classique, la non conformité sur un assemblage d'un bloc peut entraîner un retard du projet complet. Donc on ne peut pas parler d'une indépendance absolue entre les différents éléments dans le monde de fabrication (construction ou production industrielle). Dans cette thèse, nous nous intéressons aux projets de fabrication non modulaire. Par conséquent, la modularité ne sera pas considéré comme stratégie d'accélération applicable à notre cas.

## 1.2 Durée d'un projet

La durée d'un projet dépend naturellement des durées individuelles de ses activités. Dans une structure séquentielle, la durée d'un projet est simplement la somme des durées individuelles de ses activités. Dans une structure non-séquentielle (mais toujours acyclique), la méthode de détermination de la durée d'un projet est différente s'il s'agit d'un cadre aléatoire ou déterministe. Lorsque les durées individuelles sont variables, la durée du projet est évaluée par la méthode PERT (Program Evaluation and Review Technology) (Fold et Soumis, 1993 [58]). Cette technique de nature aléatoire ne représente pas la durée ni la date de fin du projet. Nous ne détaillons donc pas cette méthode qui ne rentre pas dans le cadre déterministe de notre thèse. Dans un cadre déterministe, les durées et les liens de précedence entre les activités sont déterminées et fixées au début du projet. La durée d'un projet est calculée par la somme des durées des activités de son chemin critique ou le chemin le plus long d'exécution. Dans la sous-section 1.2.1, nous présentons la manière dont on calcule la durée individuelle d'une activité avec certitude. Nous nous intéressons ensuite à la détermination de la durée d'un projet par la méthode du chemin critique dans la sous-section 1.2.2.

### 1.2.1 Durée d'exécution d'une activité

La durée d'une activité est généralement estimée en fonction des allocations de ressources. Par exemple, on sait qu'une fraiseuse automatique met tant de minutes à faire une opération de taraudage mais lorsqu'ils s'agit de temps humain (exprimé en unité de temps-homme) l'estimation est plus difficile. Dans ce cas, l'estimation de cette durée peut être fournie par des experts qui ont déjà travaillé sur des activités similaires dans d'autres projets ou sur la base d'observations passées.

Dans le cadre déterministe de notre thèse, nous considérons que les durées individuelles sont connues en « mode normal » d'exécution (pas d'accélération). En effet, exécuter une activité en mode normal consiste à exploiter la capacité normale du poste qu'elle nécessite pour son exécution. La capacité normale d'un poste peut se définir comme le nombre d'heures régulières par jour (ou par semaine) effectuées par les travailleurs permanents qui occupent ce poste dans de conditions normales de production c'est-à-dire lorsque la chaîne de production est équilibrée (bonne répartition des activités entre les travailleurs pour éviter pertes de temps et goulets) et sans occurrence d'aléas tels que l'absentéisme, les pannes ou les rebuts. En effet, une situation où il y a un absentéisme qui va ralentir la cadence de production ou des pannes et goulets qui peuvent provoquer des arrêts de chaîne ne sont pas des conditions normales de production/construction. Dans une situation normale de production, le nombre de travailleurs permanents sur chaque poste est généralement constant d'une période à l'autre, autrement dit il y a lissage des ressources.

La durée normale d'exécution d'une activité correspond au rapport entre le nombre d'heures qu'elle nécessite sur un poste (son contenu en travail) et le nombre d'heures régulières effectuées par les travailleurs permanents affectés à ce poste dans les conditions normales de production décrites précédemment. Dans la suite, on note  $d_j^{\max}$  la durée normale d'exécution d'une activité  $j$ . Cette durée correspond bien à la durée maximum d'exécution dans des conditions normales de fonctionnement du poste. Cette durée maximum  $d_j^{\max}$  doit être distinguée de la borne supérieure sur la durée qui peut excéder  $d_j^{\max}$  si le poste fonctionne en sous-capacité, autrement dit s'il y a moins de travailleurs que dans les conditions normales. Mais comme dans cette thèse nous nous intéressons aux décisions tactiques, nous considérerons donc qu'à moyen terme les conditions normales de fonctionnement sont stables de sorte que la durée d'exécution d'une activité ne pourra jamais excéder  $d_j^{\max}$ . Si la capacité normale d'un poste est ajustée à la baisse ou à la hausse en termes de nombre de travailleurs qu'il accueille, cet ajustement ne se fera non pas au niveau tactique (moyen terme), mais sur la base d'observations sur une période plus longue de la production de l'entreprise elle-même conditionnée par la demande.

On note que ce raisonnement s'applique également sur les projets de construction où l'on considère qu'un mode normal consiste à exploiter la capacité normale d'un lieu physique en nombre de personnes par  $m^2$ , dans des conditions normales de déroulement de construction c'est-à-dire sans tenir compte des aléas tels que les conditions climatiques, l'absentéisme, le retard des livraisons, etc.

## 1.2.2 Méthode du chemin critique

La Méthode du chemin critique ou Critical Path Method (CPM) est une technique d'analyse dont le but est d'identifier le chemin critique d'un projet c'est-à-dire l'ensemble des activités à marge nulle, dites critiques (Kelley, 1961 [87]). Le chemin critique est le chemin le plus long du projet. Les durées de ces activités critiques conditionnent la durée totale minimale du projet. Notons que la durée totale d'exécution d'un projet en mode normal est calculée par la méthode CPM sur la base des durées normales de ses activités, c'est-à-dire les durées associées à une allocation minimum (normale) des ressources (voir Weglarz et al. 2011, [156]).

Cette méthode se base principalement sur le diagramme de Gantt, une technique de planification qui permet au chef de projet de représenter graphiquement l'avancement ainsi que les liens de précédence entre les activités d'un projet. Cette méthode est appliquée pour tout type de projet (informatique, industriel, construction, etc) et ce dans un cadre déterministe où les durées individuelles des activités sont connues.

Reprenons notre exemple du projet UP représentée à la figure 1.4. Le diagramme de Gantt de ce projet est représenté à la figure 1.5. Ce diagramme de Gantt permet de visualiser rapidement le chemin critique (en rayé) qui est, dans cet exemple, composé des activités 1, 4, 7 et 8. Il est claire que tout retard de l'une de ces activités critiques implique un retard de la date de fin du projet.

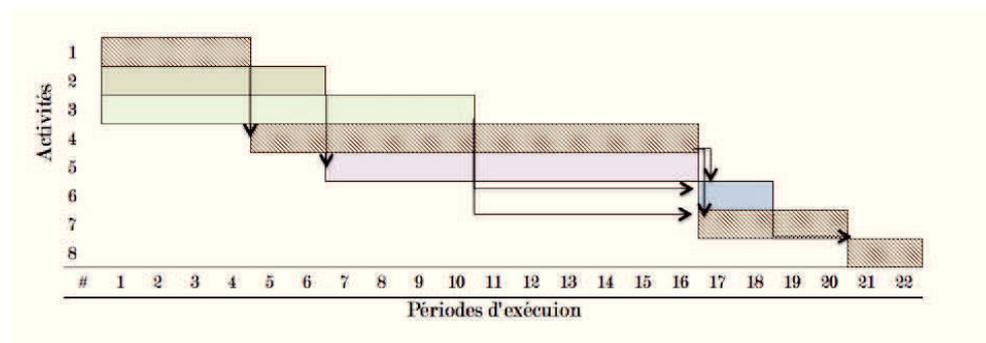


FIGURE 1.5 – Diagramme de Gantt « Exercice UP »

## 1.3 Ressources affectées à un projet

Dans l'exemple d'ordonnancement de la section précédente (figure 1.4), nous considérons simplement les contraintes de précédence entre les activités. Néanmoins, un problème d'ordonnancement classique consiste à tenir compte également de contraintes portant sur l'utilisation et la disponibilité des ressources requises. Une ressource peut être matérielle ou humaine : un équipement, une main d'oeuvre, de la matière première, de l'énergie, etc. Tout type de ressources peut être représentée par un coût. Ces différents coûts sont les éléments principaux nécessaires à l'estimation du budget d'un projet. Dans la sous-section 1.3.1 nous proposons une typologie de ressources fondée sur leurs contraintes. Dans la sous-section 1.3.2, nous analysons la

manière dont on évalue le coût d'une ressource pour aboutir à la détermination du budget d'un projet (sous-section 1.3.3).

### 1.3.1 Typologie et contraintes des ressources

L'utilisation d'une ressource par une activité implique l'allocation d'une partie de son plan de disponibilité à cette activité. Néanmoins, les disponibilités ne sont pas illimitées. Selon Slowinski (1980) [128] et Weglarz (1980a) [154], il existe trois catégories de base de ressources contraintes par leur disponibilité : les ressources renouvelables, non renouvelables et doublement contraintes.

La disponibilité d'une ressource renouvelable est limitée et constante à chaque période de l'horizon de planification. Cette catégorie de ressources est dites « non stockables », c'est-à-dire des ressources comme la main d'oeuvre, les équipements ou l'énergie qu'elles sont perdues si elles ne sont pas utilisées lors de la période dans laquelle elles sont réservées (autrement dit c'est de l'argent perdu). Ces ressources sont souvent disponibles immédiatement après l'exécution de l'activité en cours, et peuvent être utilisées par d'autres activités (nécessitant la même spécialisation professionnelle) dans une autre période.

Les ressources non renouvelables sont limitées pour l'ensemble du projet. Les unités de ces ressources consommées pendant l'exécution d'une activité, ne peuvent pas être réutilisées. En revanche, ces ressources sont « stockables » c'est-à-dire une ressource non renouvelable non consommée, comme l'argent ou la matière première, peut être considérée comme une économie et peut être consommée sur un autre projet (identique ou non).

Une ressource doublement contrainte est à la fois limitée par période et pour l'ensemble du projet. Ces ressources, comme la main d'oeuvre, peuvent être utilisées par un projet sur un nombre limité de périodes. Par exemple, un travailleur spécialisé en un corps de métier en bâtiment ne peut être disponible pour un projet que pendant un nombre limité de périodes, pour être ensuite affecté à un autre projet. Il s'agit d'une combinaison entre une contrainte sur le nombre de personne (ressource renouvelable) et une contrainte sur le nombre total de jours de travail dédiés au projet entier (ressource non renouvelable). Les ressources doublement contraintes ne nécessitent donc pas d'être considérées explicitement parce que chaque ressource doublement contrainte peut être remplacée par une ressource renouvelable et une ressource non renouvelable (Talbot, 1982 [135]).

Cependant, de nombreuses catégories de ressources rencontrées sur le terrain ont été introduites dans la littérature. Artigues et Demassy (2005) [12] distinguent un cas particulier de ressources renouvelables appelé ressources disjonctives avec des disponibilités limitées à une unité par période (pont roulant, par exemple). Cette contrainte liée à la nature de ressources impose la non-réalisation simultanée de deux activités nécessitant la même ressource. Dans cette classe de contrainte, Giard (1991) [64] range également les contraintes des interdictions de réalisation simultanée liées à des raisons de sécurité (protection électrique,...) ou de manque de place (par exemple, aménagement simultané de plusieurs zones à l'intérieur d'une voiture TGV).

Nous citons aussi les contraintes cumulatives. Ces contraintes sont considérées comme une généralisation des contraintes disjonctives quand la disponibilité d'une

ressource dans une période de temps dépend cumulativement de son utilisation (Weglarz et al., 2011 [156]). Par exemple, on a deux personnes spécialisées en peinture et 5 activités nécessitant une personne sont à exécuter au même temps. Si la somme de nombre d'heures nécessaires pour l'exécution de ces 5 activités ne dépassent pas la disponibilité des deux personnes disponibles, le problème est résolu. Mais si le besoin en nombre d'heures dépasse la capacité totale des deux personnes, les activités sont en disjonction et ne peuvent pas être exécutées simultanément. Dans notre thèse, on s'intéresse à cette catégorie de contraintes cumulatives de la main d'oeuvre.

A ces contraintes s'ajoutent également les contraintes de divisibilité et de préemption. Une ressource est divisible si ces unités peuvent être utilisées, séparément, par plusieurs activités. Une ressource peut être discrètement divisible (comme une personne, une machine, etc.) ou continûment divisible (en quantité comme un montant d'argent ou une quantité d'énergie). Une ressource est préemptable si chacune de ses unités peut être préemptée, retirée lors de l'exécution d'une activité en cours, affectée à une autre activité, et ensuite retournée à l'activité précédemment interrompue dont l'exécution peut être reprise, comme si la préemption de ressources n'avait pas eu lieu. Exemple une ressource qualifiée en une spécialisation professionnelle bien précise.

### 1.3.2 Coût des ressources

Qu'il s'agisse d'une ressource renouvelable ou non renouvelable, son coût unitaire se mesure par la somme d'argent versé pour l'acquérir (salaire ou coût horaire d'une main d'oeuvre, coût d'achat d'une machine, prix de matière première, etc.).

Le coût total d'une ressource renouvelable est calculé tout simplement par le produit de son coût unitaire par la durée de son occupation. Exemple : un homme a un coût unitaire de 60 euros (y compris charges directes et indirectes) et qui est alloué à une activité pendant 4 jours a un coût total de 240 euros imputé sur cette activité.

Pour une ressource non renouvelable, le coût se mesure par le prix d'achat ou le montant d'argent versé pour acquérir la quantité consommée. Ce coût augmente naturellement avec l'augmentation de la quantité consommée. Pourtant, le prix unitaire d'une ressource non renouvelable comme la matière première peut suivre une proportion inverse de la quantité utilisée grâce à l'effet volume impliquant un pouvoir de négociation plus fort vis à vis du fournisseur et un abaissement des coûts. Il convient ainsi de préciser que la relation croissante entre le coût et la quantité de ressources consommée est surtout valable dans le cas de ressources renouvelables ou lorsque la ressource non renouvelable prise en compte est l'argent, ce qui est généralement le cas dans la littérature. Notons ainsi qu'une ressource non renouvelable (souvent l'argent) peut être considérée comme une composite d'une ressource renouvelable représentant le coût de cette dernière.

Les ressources représentent les charges d'un projet divisées en charges directes et indirectes. Les charges directes comprennent les matériels, les équipements et la main d'oeuvre directement liés au projet et nécessaires pour l'exécution de ses activités ainsi que la sous traitance ou les intérimaires recrutés pour un projet particulier. Les charges indirects sont ceux qui ne peuvent être imputés directement à un projet

particulier, elles doivent faire l'objet d'une répartition préalable. Il peut s'agir par exemple des assurances, des amortissements et des intérêts sur les investissements en infrastructures ou du salaire des cadres chargés de la supervision de plusieurs projets simultanément.

On rajoute ici que ces charges peuvent être décomposées en charges variables qui varient avec la charge de travail et en charges fixes liées à l'existence de l'entreprise. Afin de clarifier cette différence, nous fournissons dans le tableau 1.2 quelques exemples pour chaque type de charge.

	<b>Charges variables</b>	<b>Charges fixes</b>
<b>Charges directes</b>	Main d'oeuvre directe	Amortissement d'une machine
	Sous traitance	utilisée pour un seul projet
	Matières premières	Budget publicitaire
<b>Charges indirectes</b>	Energie	Loyers
	Cadre de plusieurs projets	Contrats d'assurance

Tableau 1.2 – Exemples des différents types de charges

### 1.3.3 Détermination du budget d'un projet

L'exécution des activités du projet est consommatrice de ressources et donc génératrice de charges. Ces charges sont prises en compte dans le budget du projet et impose la connaissance du besoin en ressources pour chaque activité. En effet, le coût d'une activité peut être simplement interprété comme la somme d'argent associée à une consommation agrégée de toutes les ressources allouées à cette activité. Les charges visées sont tout d'abord les charges directes (Giard, 1991 [64]). On peut également prendre en compte les charges indirectes mais cela nécessite un calcul préalable de la quote-part imputée à l'activité.

Le coût total du projet est la somme de ses coûts directs et indirects. Une réserve de gestion destinée à faire face aux imprévus d'un projet est généralement ajouté à ces coûts (Giard, 1991 [64]). L'estimation de cette réserve peut se faire sur la base des données précises des gestionnaires des projets similaires ou des experts. On note que certains projets mondiaux ont connus des dépassements des coûts énormes à causes des circonstances imprévues. Flyvbjerg (2007) [57] cite plusieurs exemples de ces dépassements budgétaires. Citons ici le projet de Central Artery/Tunnel de Boston qui a dépassé son budget initial de 275% soit l'équivalent de 11 milliards \$ pour des défauts de construction. Un autre exemple est celui du tunnel sous la Manche entre le Royaume-Uni et de la France qui a dépassé son budget de 80% pour la construction et 140% de dépassement de financement global. La réserve de gestion initialement prévue pour ce projet était de 10%. Nous citons également les dépassements très importants des budgets de construction des réacteurs nucléaires EPR en Finlande et à Flamanville (budget multiplié par 3).

## 1.4 Accélération d'un projet

Dans tout type de projet, les gestionnaires cherchent à minimiser la durée totale d'exécution pour garantir un avantage concurrentiel. Dans les projets de conception c'est en général pour réduire la date de mise sur le marché ou le « *Time to market* » qu'on cherche à réduire la durée d'un projet. Dans les projets de fabrication non modulaire que nous considérons, les entreprises cherchent à minimiser la durée de projet pour garantir la satisfaction du client en respectant le délai de livraison exigé par le commanditaire. Dans certains cas, les entreprises cherchent à minimiser la durée d'un projet dans le cadre d'un appel d'offre pour se garantir un avantage concurrentiel. Dans d'autres cas, l'entreprise a un intérêt économique à réduire la durée d'un projet pour avancer un autre projet dans l'objectif de lissage de charge ou pour prendre une nouvelle commande dans un objectif de maximisation du profit. En gestion de projet, une des techniques possibles pour réduire la durée d'un projet de fabrication consiste à accélérer ses activités en leur allouant des ressources supplémentaires renouvelables (main d'oeuvre, équipements) ou non renouvelables (l'argent), ce que l'on appelle « l'accélération » (Mokhtari et al., 2011 [107] ; Roemer et Ahmadi, 2004 [124] ; Demeulemeester et al., 1996 [46]). Par exemple, la main d'oeuvre peut être augmentée par le recours aux heures supplémentaires, aux travailleurs temporaires ou aux équipes (Mohamad et al., 2012 [105]). On note qu'il est également possible d'accélérer une activité en lui affectant des ressources plus productives ce qui implique généralement un investissement dans des outillages plus performants (tour intégrant fraisage et tournage pour éviter les changements d'outils par exemple), une décision qui relève du niveau stratégique et non du niveau tactique auquel nous nous intéressons et qui par définition suppose la constance des moyens de production à moyen terme.

L'accélération implique une augmentation des coûts liés à l'augmentation de ressources conduisant à considérer le problème d'arbitrage durée-coût. Or, comme nous l'avons déjà évoqué, la ressource non renouvelable (le coût) peut simplement représenter le coût de la quantité consommée d'une ressource renouvelable. Lorsqu'une ressource renouvelable unique est considérée, on a affaire au problème d'arbitrage durée-ressource où l'on peut considérer la durée en fonction de la quantité consommée de la ressource et non pas son coût. Dans la sous-section 1.4.1, nous distinguons deux notions utilisées dans la littérature relative à l'accélération du projet : l'accélération et la compression. Nous introduisons ensuite le problème d'arbitrage durée - ressource et durée- coût dans les sous-sections 1.4.2 et 1.4.3, respectivement. Ces problèmes seront détaillés en chapitre 3.

### 1.4.1 Accélération et compression

Prenons l'exemple d'une activité  $j$  qui nécessite pour son exécution 105 heures de travail sur un poste de montage. Ce poste accueille 3 travailleurs permanents dans des conditions normales ce qui confère au poste une capacité normale de  $3 \cdot 7 = 21$  heures régulières par jour (7 heures par jour). Alors l'activité a une durée normale d'exécution  $d_j^{\max}$  égale à  $105/21 = 5$  jours. On note que dans cet exemple introductif, nous considérons une ressource unique, la main d'oeuvre exprimée en

nombre d'heures, mais en réalité, les ressources multiples existent (comme le cas de plusieurs spécialisations professionnelles ou métiers). Le problème d'ordonnement à ressources multiples contraintes sera détaillé en chapitre 3.

Accélérer l'exécution d'une activité consiste à lui allouer des ressources additionnelles par rapport aux ressources disponibles en mode normal. La compression consiste à affecter à l'exécution d'une activité le maximum possible de ressources additionnelles par rapport au mode normal. La notion de maximum est une notion relative, spécifique à une entreprise donnée car une allocation maximum de ressources dépend notamment de la façon dont les postes de travail sont conçus (nombre maximum de travailleurs qu'ils peuvent accueillir) et aux choix de l'entreprise en matière de recrutement de temporaires ou de recours aux heures supplémentaires. Le mode compressé est donc un mode particulier d'accélération correspondant à l'allocation maximum possible des ressources nécessaires à l'exécution d'une activité. Dans la suite, on note  $d_j^{\min}$  la durée d'exécution d'une activité dans le mode compressé puisqu'elle correspond à la plus petite durée possible d'exécution.

Dans notre exemple, l'accélération va consister à ajouter des travailleurs sur le poste ou à recourir à des heures supplémentaires. Imaginons que le poste de montage soit conçu pour accueillir au maximum 6 travailleurs (permanents ou temporaires) et que la politique de l'entreprise consiste à limiter le nombre des heures supplémentaires à 2 heures par jour et par travailleur. Nous présentons dans le tableau 1.3 les différentes combinaisons possibles de ressources au sein de cette entreprise où l'on suppose que le recours aux heures supplémentaires implique systématiquement tous les travailleurs. Autrement dit tous les travailleurs font des heures supplémentaires ou aucun n'en fait. La raison de ce choix réside dans le fait que pour toute heure supplémentaire effectuée, un manager doit être présent de sorte qu'il y a des économies d'échelle en termes d'encadrement à faire travailler tous les opérateurs en heures supplémentaires (ou aucun). Dans la dernière colonne du tableau 1.3 on donne la durée d'exécution de l'activité qui, rappelons-le, requiert un total de 105 heures d'assemblage sur le poste. Cette durée peut être compressée c'est-à-dire réduite à un minimum de 1,94 jours ( $d_j^{\min} = 1,94$ ) avec une allocation maximale de 54 heures sur le poste et qui correspond à la situation où le nombre maximum de 6 travailleurs y est affecté et où tous les travailleurs font 2 heures supplémentaires par jour, c'est-à-dire le maximum autorisé par l'entreprise.

Effectif	Nombre d'heures par jour			Mode	Durée
	Régulières	Sup.	Total		
3	21	0	21	<b>Normal</b>	5,00
3	21	6	27	Accéléré	3,89
4	28	0	28	Accéléré	3,75
4	28	8	36	Accéléré	2,92
5	35	0	35	Accéléré	3,00
5	35	10	45	Accéléré	2,33
6	42	0	42	Accéléré	2,50
6	42	12	54	<b>Compressé</b>	1,94

Tableau 1.3 – Allocation de main d'oeuvre et types de mode

### 1.4.2 Arbitrage durée - ressource

Dans le cas d'une ressource unique renouvelable comme la main d'oeuvre par exemple, la durée  $d_j$  d'une activité  $j$  est exprimée comme une fonction non croissante de la quantité de la ressource  $r$  qui lui est allouée. On parle du problème d'arbitrage durée - ressource (Time Resource Trade off Problem, TRTP).

Or, sur le terrain et pour tout type de projet, lorsqu'on parle de la durée d'un projet ou d'exécution d'une activité, on fait systématiquement le lien avec la productivité des ressources. On rappelle que la productivité du travail se définit comme le rapport entre le nombre d'unités produites (ou nombre de fois que l'on exécute une activité donnée) et le nombre d'heures de travail qui est alloué à la production (ou à l'exécution d'une activité). De manière alternative, on peut remplacer dans ce rapport le nombre d'unités produites par la durée d'exécution d'une activité et définir la productivité marginale du travail comme le rapport  $\Delta d_j / \Delta r$  qui mesure la réduction de durée  $\Delta d_j$  de l'activité résultant d'une augmentation  $\Delta r$  de la ressource  $r$  (nombre d'heures de travail alloué à l'exécution de l'activité).

Afin d'illustrer ce concept, considérons l'exemple précédent d'une activité  $j$  avec un contenu en travail  $W_j = 105$  heures et un nombre de travailleurs permanents  $P_j = 3$  alloué à son exécution et constant d'un jour à l'autre. Chaque opérateur travaille un nombre d'heures par jour  $H_R = 7$ . On suppose que la ressource que l'on peut faire varier  $r$  est constituée d'heures de travail temporaire infiniment divisibles ( $r$  est une variable continue) et l'on note  $d_j = \mathcal{D}_j(r)$  la durée d'exécution de l'activité  $j$  comme une fonction continue  $\mathcal{D}_j$  de la ressource additionnelle  $r$ . La productivité marginale de la ressource  $r$  peut aussi se définir comme la dérivée  $d\mathcal{D}_j/dr$  qui donne la pente de la tangente à la fonction  $\mathcal{D}_j$  au point  $r$ . Une productivité marginale constante se traduit par une dérivée  $d\mathcal{D}_j/dr$  constante, autrement dit la fonction  $\mathcal{D}_j$  est linéaire.

Dans notre exemple, si l'on suppose une productivité marginale constante alors on a

$$d_j = \mathcal{D}_j(r) = d_j^{\max} - \frac{1}{P_j H_R} r, \quad \text{avec} \quad d_j^{\max} = \frac{W_j}{P_j H_R}, \quad (1.1)$$

$$\frac{d\mathcal{D}_j}{dr} = -\frac{1}{P_j H_R}.$$

L'application à notre exemple donne  $d_j^{\max} = 105/(3 \cdot 7) = 5$  et  $\mathcal{D}_j(r) = 5 - \frac{1}{21}r$ . La productivité marginale  $\frac{d\mathcal{D}_j}{dr} = -\frac{1}{21}$  signifie que si l'on ajoute une heure de travail temporaire alors la durée d'exécution de l'activité diminue de  $1/21 = 0,0476$  jour. Ainsi, on peut réduire la durée de l'activité à 2 jours en ajoutant 63 heures de travail temporaire.

Imaginons qu'en raison de l'encombrement du poste de travail la productivité marginale soit décroissante, c'est-à-dire que la contribution d'une heure de travail temporaire à la réduction de la durée diminue au fur et à mesure que l'on ajoute des heures de travail temporaire. Si l'on désigne par  $\gamma$  une constante prenant une petite valeur positive, alors la forme de la fonction peut être du type

$$\begin{aligned} \mathcal{D}_j(r) &= d_j^{\max} - \frac{1}{P_j H_R} r + \frac{\gamma}{P_j H_R} r^2, \\ \frac{d\mathcal{D}_j}{dr} &= -\frac{1}{P_j H_R} + \frac{2\gamma}{P_j H_R} r, \\ \frac{d^2\mathcal{D}_j}{dr^2} &= \frac{2\gamma}{P_j H_R}. \end{aligned} \tag{1.2}$$

La fonction  $\mathcal{D}_j$  est décroissante car sa dérivée première est négative pour des valeurs de  $r < 1/2\gamma$  qui sont des valeurs raisonnables car  $\gamma$  est petit. Cette fonction est convexe puisque sa dérivée seconde est positive. A la figure 1.6 nous représentons ces fonctions dans le cas linéaire et convexe pour  $\gamma$  fixé arbitrairement à 0,0025 (fonction  $\mathcal{D}_j$  donnée respectivement dans les équations (1.1) et (1.2)), où  $r$  est placé en abscisse et  $\mathcal{D}_j(r)$  en ordonnée. Dans le cas linéaire (courbe en trait plein) partant d'une durée de 5 jours il faut 21 heures additionnelles de travail temporaire pour diminuer de 1 jour la durée (passage d'une durée de 5 à 4). Cette augmentation du nombre d'heures est la même si, partant d'une durée de 3 jours on souhaite une réduction de 1 jour de la durée. En effet pour passer de 3 à 2 jours, il faut augmenter le travail temporaire de 42 à 63 heures. C'est ce qui caractérise une productivité marginale constante. En revanche dans le cas convexe (courbe en pointillés), il faut ajouter 22,24 heures temporaires pour réduire la durée de 5 à 4 jours et ajouter 30,66 heures pour une réduction de la durée de 3 à 2 jours. La productivité marginale décroissante traduit donc le fait qu'au fur et à mesure qu'on ajoute du travail temporaire, il faut de plus en plus d'heures pour obtenir une même réduction de 1 jour de la durée.

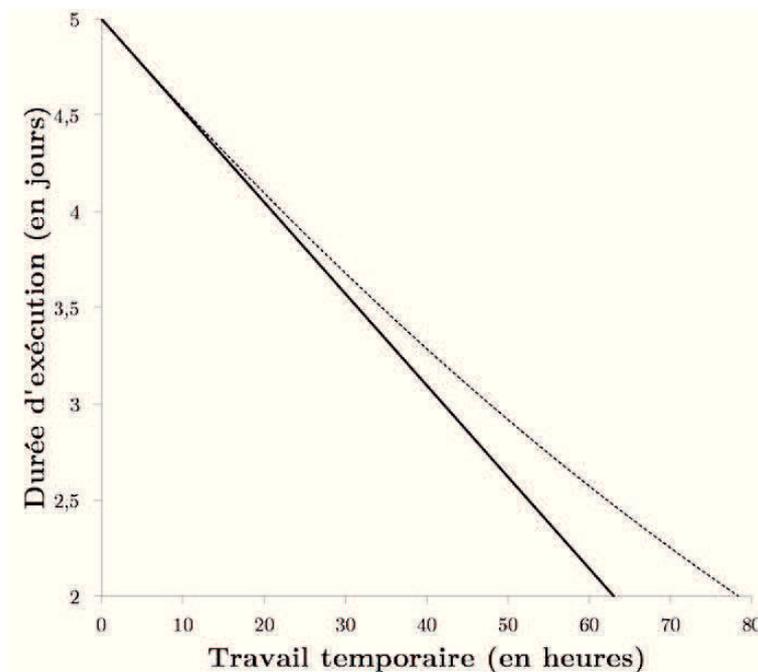


FIGURE 1.6 – Durée d'exécution comme une fonction linéaire ou convexe d'une ressource additionnelle

### 1.4.3 Arbitrage durée - coût

Dans les deux cas de diminution de la durée d'un projet : par l'ajout de ressources additionnelles ou le recours à des ressources plus productives, ces décisions se traduisent par des coûts supplémentaires. Ainsi, dans un problème d'arbitrage durée - coût, on considère le coût de la ressource au lieu de considérer sa quantité.

Si  $c_r$  désigne le coût d'une heure de travail, alors le coût qui résulte de l'ajout de  $r$  heures supplémentaires est simplement égal à  $c_r \cdot r$ . Comme nous avons la relation  $d_j = \mathcal{D}_j(r)$  alors  $r = \mathcal{D}_j^{-1}(d_j)$  et le coût  $c_j$  associé à  $d_j$  s'écrit

$$c_j = c_r \cdot r = c_r \cdot \mathcal{D}_j^{-1}(d_j) = \mathcal{C}_j(d_j). \quad (1.3)$$

Dans notre exemple pour le cas linéaire on obtient  $\mathcal{C}_j(d_j) = c_r \cdot P_j \cdot H_R(d_j^{\max} - d_j)$  qui est bien évidemment une fonction linéaire décroissante de la durée. Dans le cas convexe, pour obtenir  $r = \mathcal{D}_j^{-1}(d_j)$  il faut résoudre  $d_j = d_j^{\max} - \frac{1}{P_j H_R} r + \frac{\gamma}{P_j H_R} r^2$  soit l'équation du second degré  $\frac{\gamma}{P_j H_R} r^2 - \frac{1}{P_j H_R} r + d_j^{\max} - d_j = 0$  ce qui va nous permettre d'obtenir une relation entre  $r$  et  $d_j$  (une seule des deux racines est réalisable). En posant  $\frac{\gamma}{P_j H_R} = a$  et  $-\frac{1}{P_j H_R} = b$  pour simplifier l'écriture, on obtient la fonction de coût  $\mathcal{C}_j(d_j) = c_r \cdot r = c_r \cdot \frac{1}{2a} \left[ -b - \sqrt{b^2 - 4a(d_j^{\max} - d_j)} \right]$ . Le tracé de ces deux fonctions de coût est donné à la figure 1.7 avec un coût unitaire  $c_r = 72$  euros. Ainsi si la productivité marginale de la ressource est décroissante, c'est-à-dire si  $\mathcal{D}$  est une fonction décroissante convexe de la ressource  $r$  alors le coût  $\mathcal{C}$  est aussi une fonction décroissante convexe de la durée (courbe en pointillés sur la figure).

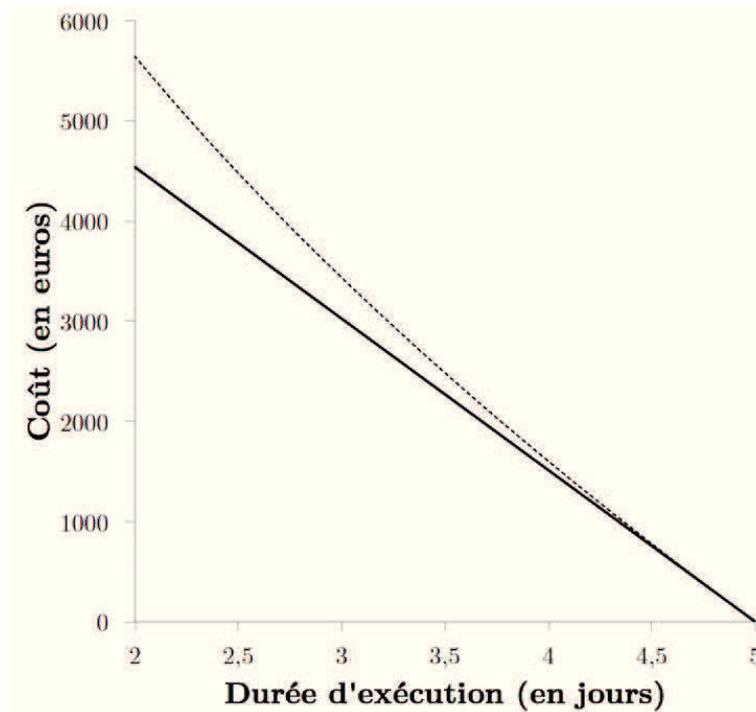


FIGURE 1.7 – Le coût comme une fonction linéaire ou convexe de la durée

L'analyse que nous venons de mener s'applique également au cas d'une productivité marginale croissante dont les exemples concrets se rencontrent plus volontiers dans les projets de conception où le travail en équipe est d'autant plus efficace (jusqu'à un certain point) que la taille de l'équipe est importante. Ainsi, une activité de conception s'effectuera d'autant plus rapidement que l'on ajoute des ressources. Dans ce cas les fonctions  $\mathcal{D}$  et  $\mathcal{C}$  sont des fonctions concaves respectivement de la ressource et de la durée.

## 1.5 Conclusion

La durée d'un projet, telle que nous l'avons étudié dans ce chapitre, se base notamment sur la détermination des durées individuelles de ses activités. De l'autre côté, les durées individuelles sont dépendantes du contenu en travail des activités et de la disponibilité des ressources ou de la capacité normale du poste qu'elle nécessite pour son exécution. En mode normal, la durée d'une activité est supposée comme une fonction décroissante de la quantité allouée de ressources (notamment renouvelables). Par ailleurs, le coût d'une ressource est croissant linéairement ou non avec la quantité consommée.

Les aléas classiques d'un projet, tels que l'absentéisme, les retards d'approvisionnement, les grèves, etc., peuvent fortement impacter le coût et la durée d'un projet. Ces aléas sont généralement gérés au niveau opérationnel. L'absentéisme par exemple est généralement contrôlé par la présence d'un certain pourcentage de travailleurs flexibles entre postes de travail (multi-compétents, polyvalents ou intérimaires). Les retards de livraison peuvent être gérés au niveau opérationnel par la mise en place d'un stock de sécurité interne ou dans un magasin avancé fournisseur (MAF). En outre, la direction d'une entreprise doit savoir réagir (par le dialogue social) face à une grève ou autre situation bloquante afin de limiter les conséquences sur les projets en exécution. Notons que la gestion de ces aléas n'entre pas dans le cadre tactique de notre thèse. Nous relevons ainsi que, lorsqu'on dispose de la capacité normale de chaque poste de travail (en supposant qu'un poste = une activité), la durée et le coût d'un projet peuvent être déterminés avec certitude par la méthode CPM et la méthode des coûts directs, respectivement.

Néanmoins, il convient de préciser que les aléas peuvent être implicitement les causes d'une accélération non planifiée d'un projet. Après l'achèvement d'un certain pourcentage du projet, si le projet est en retard, le gestionnaire peut prendre la décision d'accélérer les activités restantes afin d'éviter les pénalités contractuelles (Mohamad et al. 2012 [105]). Dans tous les cas, l'accélération d'un projet implique un coût additionnel relatif à l'augmentation des ressources engagées appelé coût d'accélération qui peut être une fonction linéaire, convexe ou concave de la durée.

Le troisième critère de performance (la qualité) faisant l'objet principal de notre thèse est détaillé dans le chapitre 2 suivant dédié au concept de qualité. Cette notion est très large et parfois vaste. Dans ce chapitre, nous allons rappeler la définition du concept de qualité et son évolution historique. Ensuite, nous abordons la manière dont on évalue et mesure la qualité. Nous rappellerons enfin les méthodes de gestion et d'amélioration de la qualité.

## Chapitre 2

# Le concept de qualité en gestion de projet

La qualité est décrite comme « l'ensemble des propriétés et caractéristiques d'un produit ou service qui lui confère l'aptitude à satisfaire des besoins exprimés ou implicites »<sup>1</sup>.

Au travers du XXème siècle, la qualité a connu une évolution historique importante qui se traduit par le passage de l'inspection unitaire de pièces fabriquées au management par la qualité totale ou Total Quality Management (TQM). TQM est « une approche de gestion centrée sur la qualité, basée sur la participation de toute la population d'une organisation et visant la réussite à long terme par la satisfaction de la clientèle, et les profits à tous les membres de l'organisation et de la société » (ISO 8402 [79]). Une démarche TQM est avant tout un système de gestion stratégique qui pousse à la maîtrise de la qualité globale de projets en passant de la réduction des non-conformités à leur prévention (Valmohammadi et Roshanzamir, 2015 [147]). Par qualité globale, on désigne la qualité d'un projet qui résulte d'une agrégation des qualités individuelles de ses activités (souvent la moyenne). En résumé, la qualité totale implique une amélioration de la qualité globale. Ces deux concepts totalement différents sont connexes du point de vue managériale. Dans cette thèse c'est la qualité globale d'un projet qui nous intéresse.

Cependant, l'évaluation de la qualité n'est pas toujours simple notamment dans le cas du développement d'un produit nouveau. En revanche, lorsque le client est connu et les spécifications d'un produit sont bien définies, comme dans le cas de projets de fabrication que nous considérons, la qualité est plus simple à évaluer. Sur la base de la typologie de projets que nous avons introduite dans le chapitre 1, nous distinguons l'évaluation objective de l'évaluation subjective de la qualité. En effet, la qualité subjective (perçue) est surtout considérée dans les projets de conception et repose sur des critères généralement intangibles liés aux réactions subjectives, émotionnelles ou cognitives des consommateurs comme l'appréciation de la conception ou de l'esthétique d'un produit. La qualité objective est considérée dans les projets de fabrication et repose sur des critères tangibles, exprimés dans des spécifications contractuelles ou internes et dont la conformité est mesurable par

---

1. Définition donnée par la norme ISO 8402 : Management de la qualité et assurance de la qualité - Vocabulaire

rapport à des référentiels précis.

Dans ce chapitre, nous analysons le concept de qualité et son évolution (section 2.1). Nous présentons également les différentes façons d'évaluer la qualité (section 2.2) en distinguant le concept de qualité subjective du concept de qualité objective (sous-section 2.2.1). Comme nous nous intéressons aux projets de fabrication, c'est l'évaluation objective de la qualité qui est retenue dans cette thèse et sera détaillée dans les sous-sections 2.2.2 et 2.2.3. La section 2.3 introduit la mesure quantitative de la qualité individuelle d'une activité et globale d'un projet. Dans la section 2.4, nous résumons les méthodes stratégiques de gestion et d'amélioration de la qualité. La section 2.5 conclut le chapitre en mettant l'accent sur le fait que toute évaluation objective de la qualité peut être ramenée à une mesure quantitative par un indicateur compris entre 0 et 1.

## 2.1 Evolution historique de la qualité

Au travers du XXème siècle, comme le montre la figure 2.1, l'histoire de la qualité a connu trois périodes remarquables. La première période est celle de « la belle ouvrage » suivie par la période de « conformité aux normes » et la période « d'excellence et satisfaction totale du client ».

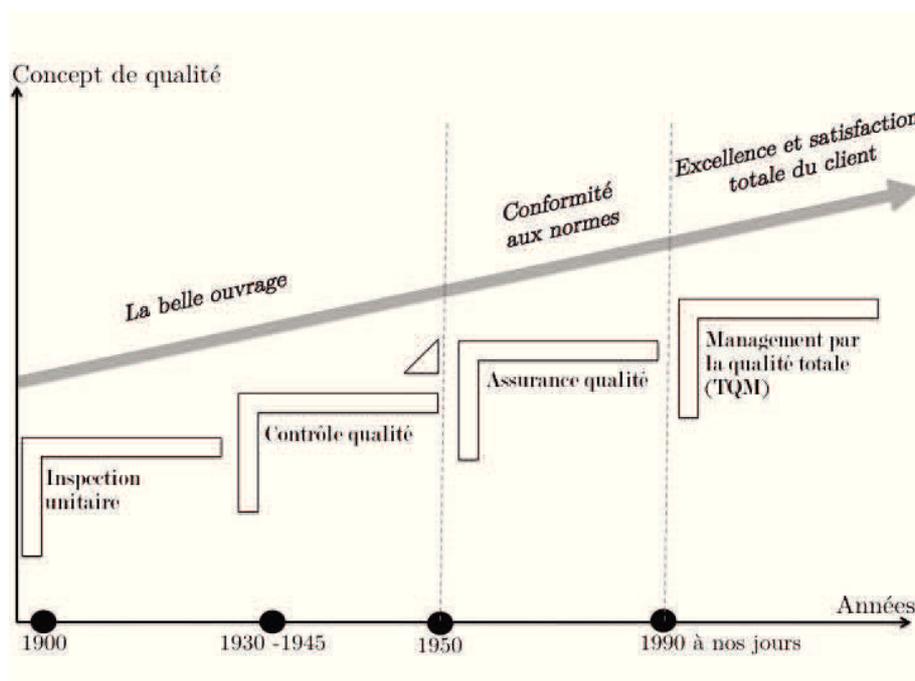


FIGURE 2.1 – Evolution historique de la qualité

Lors de la première période, la qualité a évolué de l'inspection unitaire du produit (début des années 1900) au contrôle par échantillonnage statistique des produits à

la réception et à l'expédition (1930 à 1945). A cette époque la qualité consistait notamment à la vérification dimensionnelle des pièces fabriquées.

Aux environs du milieu du XXème siècle, la qualité est devenue un vrai concept qui correspond à « la conformité aux normes », c'est l'époque de l'assurance qualité apparue au Japon au début des années 1950. L'assurance qualité est « l'ensemble des activités préétablies et systématiques mises en oeuvre dans le cadre du système qualité et démontrées en tant que besoin pour donner la confiance appropriée en ce qu'une entité satisfera aux exigences pour la qualité »<sup>2</sup>. Elle se décline généralement sous la forme d'un référentiel documentaire qui décrit notamment les procédures qualité, les normes et les réglementations appliquées. Par définition<sup>3</sup>, les normes ne constituent que des instructions auxquelles producteurs et utilisateurs sont libres d'adhérer ou non. Le respect d'une norme est donc « volontaire ». Une norme est obligatoire seulement si elle est exigée par une réglementation ou dans le cahier des charges d'un client. Notons qu'une réglementation résulte d'une décision législative d'un état rendant obligatoire le respect des dites « exigences essentielles », en vue de l'intérêt général de la nation (Cochet et Merlet, 2000 [33]). Par exemple, la réglementation thermique RT2012 est une des exigences essentielles dans le secteur de construction dans le cadre de mesures nécessaires à la mise en place de la politique énergétique pour protéger la planète.

Il convient de préciser qu'à cette époque et toujours au Japon, sont apparus les premiers outils d'amélioration de la qualité tels que la roue de Deming (du statisticien anglais William Edwards Deming) dite PDCA (*Plan, Do, Check, Act*) ainsi que le concept du contrôle de la qualité totale ou Total Quality Control (TQC). La roue de Deming (ou PDCA) est un modèle d'amélioration continue utilisé en management de la qualité permettant de repérer avec simplicité les étapes « planifier, exécuter, contrôler et agir » à suivre pour améliorer la qualité dans une organisation. TQC est le système d'amélioration de la qualité le plus ancien décrit par Feigenbaum (1951) [54] comme « un réseau de gestion, contrôle, et de procédures nécessaires pour produire et livrer un produit avec un standard qualité spécifique ». Mais l'objectif principal des entreprises était la conformité aux normes.

Notons qu'il ne faut pas confondre le respect des normes (obligatoires ou volontaires) avec la notion de qualité intrinsèque qui sont deux notions très différentes. En effet ce n'est pas le respect d'une norme qui implique qu'une organisation produit/fabrique de la bonne qualité. Par exemple, une voiture LOGAN de marque DACIA respecte les mêmes normes en France qu'une AUDI. Pourtant ces véhicules n'ont aucun rapport qualitatif entre eux et n'ont pas la même clientèle non plus. Certains commanditaires (aisés et solvables) ne réclament plus des simples produits techniquement efficaces, mais accompagnés des options de confort, sécurité, plaisir, rapidité,

---

2. Définition donnée par la norme ISO 8402, Management de la qualité et assurance de la qualité - Vocabulaire.

3. Selon ISO 9001, la norme est « un document, établi par consensus et approuvé par un organisme reconnu, qui fournit, pour des usages communs et répétés, des règles, des lignes directrices ou des caractéristiques, pour des activités ou leurs résultats, garantissant un niveau d'ordre optimal dans un contexte donné ».

etc. La qualité est ainsi appréhendée comme un facteur premium différenciant. La définition de la qualité, selon la volonté du décideur, peut être donc de dépasser la simple application des normes obligatoires ou contractuelles pour atteindre un niveau d'excellence<sup>4</sup>. L'augmentation croissante des exigences des consommateurs et la concurrence entre les producteurs ont laissé plus de choix aux consommateurs et la qualité est devenue une préoccupation stratégique des entreprises. Depuis les années 1990, nous sommes à l'époque de « l'excellence et la satisfaction totale du client ». On ne cherche plus seulement à produire un produit de qualité mais à lui associer des fonctionnalités supplémentaires qui attirent le client. Pour atteindre ce niveau, des méthodes stratégiques de gestion et d'amélioration de la qualité sont de plus en plus appliquées dans les entreprises. La méthode la plus largement appliquée aujourd'hui est le management par la qualité totale (TQM).

On peut s'interroger ici sur la très forte croissance du marché *providers*, comme la compagnie Low cost Ryanair ou les voitures DACIA plus « simple » que des Renault et vendus à des prix inférieurs, où la qualité n'est pas associée aux fonctions ou services additionnels. Au contraire, ces prestations garantissent le juste nécessaire pour sa clientèle qui ne veulent (ou ne peuvent) pas payer des fonctionnalités inutiles ou considérés par eux comme superflus. L'évolution des exigences peut être donc à la hausse ou à la baisse. Mais cela ne remet pas en cause le besoin d'amélioration de la qualité avec une pondération relative aux exigences. Une compagnie Low cost investit dans les dispositifs qualité liés à la sécurité plus que ceux liés au confort par exemple.

D'autres méthodes d'amélioration continue plus ou moins centrées sur la qualité sont également déployées dans les entreprises, telles que le Six Sigma et le Lean Management. Ces méthodes ainsi que le TQM sont résumées dans la section 2.4 (page 54) après avoir étudié la façon dont on évalue et mesure la qualité.

## 2.2 Evaluation de la qualité

La qualité est un critère de performance indissociable de la notion de projet (Basu, 2014 [17]). Ce critère, difficile à définir, nécessite une étude approfondie de la façon dont on l'évalue en réalité. Selon Sri Yogi (2016) [161] et Das et al. (2008) [40], le client se base surtout sur la performance du fonctionnement d'un produit (*l'output* d'un projet), les caractéristiques de ses fonctionnalités supplémentaires, sa fiabilité, sa durée de vie et autres critères économiques (comme le prix de vente et la garantie). Sur la base de ces éléments, nous distinguons la qualité subjective de la qualité objective (sous-section 2.2.1). Nous classons ensuite les indicateurs d'évaluation objective de la qualité par leur degré de qualification en indicateurs quantitatifs (sous-section 2.2.2) et qualitatifs (sous-section 2.2.3). On note que les indicateurs sont suivis dans un tableau de bord qui concourt à l'amélioration de la qualité dans

---

4. La fondation européenne pour la gestion de la qualité (EFQM) définit l'excellence comme « une pratique remarquable de management d'une organisation et de réalisation des résultats se fondant sur un ensemble de concepts fondamentaux incluant l'orientation résultats, l'orientation clients, le leadership et la constance dans la poursuite des objectifs, le management par les processus et les faits, l'implication des personnels, l'amélioration continue et l'innovation, les partenariats fondés sur des avantages réciproques, la responsabilité sociale de l'entreprise ».

l'entreprise, d'où le lien avec les méthodes de gestion et d'amélioration de la qualité.

### 2.2.1 Qualité subjective et qualité objective

La qualité perçue par le client est l'un des facteurs les plus importants qui sous-tendent le succès d'un produit. Elle est interprétée comme une dimension de la qualité d'un produit, soit comme une perception ou un jugement subjectif du client se rapportant à sa personnalité, à son histoire ou expérience. Styliadis et al. (2015) [134] soulignent qu'en industrie automobile par exemple, la qualité perçue devient un facteur premier dont la compétitivité est une dimension principale. Outre les spécifications techniques basiques (ou de fonctionnalité) d'un produit, on cherche à satisfaire le client sur des points subjectifs tels que l'esthétique visuelle ou l'ergonomie d'utilisation. Les auteurs distinguent la qualité perçue par la valeur (VPQ ou Value based Perceived Quality) de la qualité technique perçue (TPQ ou Technical Perceived Quality). Dans les deux cas la qualité est subjective. La qualité perçue par la valeur dépend des facteurs externes comme l'expérience du client et le patrimoine de la marque, elle est évaluée à travers les sens et la cognition. La qualité technique perçue d'une voiture par exemple est composée de la qualité visuelle, la qualité de la sensation, la qualité sonore et la qualité de l'odeur du véhicule. La qualité visuelle peut être évaluée selon des dimensions telles que l'esthétique et la finition d'une surface. La qualité de l'odeur peut être évaluée selon l'odeur interne et l'intensité de l'odeur d'un véhicule. On note ici que la qualité perçue de nature subjective et difficile à mesurer est surtout prise en compte dans les projets de conception quand on cherche à développer un produit nouveau ou à améliorer un produit existant pour des enjeux de marketing et de concurrence. Une fois le produit conçu, les spécifications sont établies sur la base des exigences clients.

Dans un projet de fabrication fondé sur des spécifications préétablies le subjectif n'entre plus dans l'évaluation de la qualité, une spécification est respectée ou elle ne l'est pas, ce que l'on appelle le concept de qualité objective. Par exemple, un contrôle qualité d'une activité ne varie pas en fonction de la personnalité de l'évaluateur même lorsqu'il s'agit d'un contrôle visuel. Un contrôleur qui va évaluer la conformité d'une peinture par exemple, ne va pas appliquer son goût personnel dans l'évaluation mais va donner une appréciation de type bon, moyen ou mauvais par rapport à des critères tangibles et des spécifications provenant de la direction de l'entreprise ou de la clientèle (une teinte bien précise par rapport à un nuancier de couleurs de peinture ou une composition chimique). Autrement dit un contrôleur visuel de qualité est un technicien qui se sert de ses yeux au même titre qu'un assembleur se sert de ses mains, pour exécuter une activité bien spécifiée (le contrôle visuel). Un exemple concret de l'évaluation visuelle de la qualité par l'association européenne des industries du plâtre « Eurogypsum » est donnée à la sous-section 2.2.3, où on verra clairement que détecter et compter les défauts est toujours possible selon un descriptif bien détaillé. Comme nous nous intéressons aux projets de fabrication et non aux projets de conception, c'est le concept de qualité objective qui est retenue dans cette thèse. Les sous-sections qui suivent ne concernent que cette qualité dite objective et qui ne se rapporte qu'au respect des critères tangibles.

## 2.2.2 Indicateurs quantitatifs

Les indicateurs quantitatifs sont souvent de type « taux de..., nombre de..., indice de..., probabilité de... etc. » Ces indicateurs fournissent des indications chiffrées sur le niveau de qualité. Un indicateur quantitatif évalue donc la qualité par rapport à un seuil exprimé en nombre ou pourcentage. Prenons l'exemple d'une pièce peinte, si le nombre de rayures acceptable est respecté, la qualité est jugée bonne et inversement, s'il l'est pas une reprise de la peinture est nécessaire. Parmi ces indicateurs, on distingue les indicateurs qui portent sur la performance du produit (sous-section 2.2.2.1), et les indicateurs de coûts (sous-section 2.2.2.2) qui sont en lien direct avec le management de la qualité.

### 2.2.2.1 Performance d'un produit

Nous citons ici les différents indicateurs d'évaluation de la performance d'un produit en suivant son flux dans un processus de projet. Nous commençons donc par le taux de rebut qui se limite aux opérations élémentaires d'une activité (type usinage de composants) ou à la fabrication de grande série, nous passons ensuite aux non conformités et aux besoins de reprises des activités avant d'aboutir au produit fini et nous terminons par la fiabilité et la durabilité d'un produit lancé sur le marché.

#### Taux de rebut et taux de vente

Le taux de rebut est le pourcentage de pièces jetées par rapport au nombre de pièces fabriquées. Par exemple, si on a lancé une fabrication de 1000 pièces dont 250 sont jetées, le taux de rebut est de  $(250/1000) \times 100 = 25\%$ . L'application de cet indicateur permet de valoriser le coût de non qualité. Dans notre exemple, si une pièce coûte 20 euros, cela donne un taux de rebut valorisé à 7,5 euros soit un coût de non qualité qui s'établit à 7,5 euros par pièce jetée. La mise au rebut a des causes nombreuses. Citons notamment l'externalisation de la conception ou de la fabrication de certains composants dans le cadre de la modularité. En effet, les modifications de dernière minute qui risquent d'être retardées ou non communiquées, impliquent une fabrication de produits hors d'usage, par conséquent le taux de rebut augmente. On note que cet indicateur n'est pas applicable sur des activités de construction ou d'assemblage qui concernent un sous-ensemble de composants dont la qualité ne peut être vérifiée que sur un point comme par exemple la vérification de l'alignement des phares d'une voiture. Ce type d'activité peut être évalué notamment par une check-list que nous présentons dans la section suivante.

Le taux de rebut est également associé au nombre de produits défectueux et faisant l'objet d'une réclamation de garantie (Das et al., 2008 [40]). On note ici que le taux de vente des pièces de rechange peut être aussi considéré comme un indicateur de la non qualité d'un produit, mais au même temps une source de gain comme pour les constructeurs automobiles. Prenons un exemple ancien mais connu du défaut de fabrication de la timonerie de direction (bielle pendante) à General Motors pour les Cadillac 1959-1960. Ce défaut a impliqué la vente de six fois plus de pièces de rechange pour la bielle pendante au cours de ces deux années que pendant toutes les années qui ont précédé et suivi (Mashaw et Harfst, 1991 [103]).

#### Non-conformité et taux de reprise

Parfois une pièce peut être non conforme mais en état de fonctionnement, c'est-à-dire non rebutée. Dans ce cas, l'évaluation de la qualité peut être faite par le nombre de défauts détectés. Par exemple, le nombre de plis dans une plaque métallique. Selon ce nombre de défauts et selon le type de défauts, les pièces peuvent être reprises c'est-à-dire retravaillées pour devenir conformes. Ce raisonnement est également valable pour toute activité dans un projet de fabrication ou de conception. Ainsi, la qualité d'une activité peut être évaluée par le taux de reprise.

Le taux de reprise se traduit donc comme le taux des activités (opérations de fabrication ou activités de construction) à retravailler partiellement ou totalement suite à une non conformité. Il convient de préciser que le taux de reprise n'est pas proportionnel au nombre de non conformités. Une activité nécessite ou non une reprise selon le niveau de qualité recherché par rapport aux exigences internes ou externes, à un niveau de normes ou niveau d'excellence, etc. L'évaluation de la qualité par ce taux nécessite donc un suivi des causes de reprises. Un taux de reprise élevé peut à la fois signifier un niveau élevé d'exigence ou beaucoup de mauvaise qualité qui nécessite des reprises pour atteindre un niveau de fonctionnalité. Le niveau de fonctionnalité est plutôt définie par l'entreprise et signifie le niveau de qualité nécessaire pour que le produit soit fonctionnel, ce niveau intègre généralement les exigences réglementaires. Par ailleurs, une non conformité peut être acceptable pour atteindre ce niveau de fonctionnalité mais pas acceptable par rapport au niveau d'exigence demandé, nécessitant donc une reprise. On note que le niveau d'exigence correspond par la suite au « seuil de qualité minimum ». Les causes de reprise peuvent être aussi externes c'est-à-dire à la demande de client, comme un changement de plan ou de choix de peinture, de carrelages, etc.

On rajoute ici les non conformités qui peuvent être détectées après le lancement d'un produit dans le marché, nécessitant une reprise du produit final. Ces non conformités sont souvent liées à des décisions managériales pendant le déroulement de projet comme l'accélération d'un projet qui remet en cause son niveau de qualité. Dans ce chapitre, nous citons quelques exemples réels qui relèvent l'impact de l'accélération d'un projet sur la qualité mais la littérature sur cette relation sera discutée dans le chapitre 4. Pour un premier exemple, prenons la société SAMSUNG courant 2016, afin de parer à la sortie annoncée du produit concurrent (iPhone 7) a fait l'impasse sur certaines recherches et sur le respect des procédures internes de contrôle qualité. La conséquence ne s'est pas faite attendre, puisque qu'après avoir annoncé la mise sur le marché de son nouveau téléphone, le GALAXIE 7, le groupe SAMSUNG a dû le retirer immédiatement du marché courant du 2ème Semestre 2016 compte tenu du risque d'explosion et d'incendie de la batterie du téléphone. Il en est de même pour le lanceur Space X du groupe d'Elon MUSK qui a explosé en Septembre 2016 du fait d'un défaut du système de remplissage de carburant dans la fusée. La conséquence a été un arrêt complet de tous les lancements de SpaceX pendant plusieurs mois entraînant des pertes considérables pour le groupe d'Elon MUSK et surtout un impact à moyen terme sur l'image du groupe. Enfin, un 3ème exemple dans l'industrie aéronautique. Le lancement du Boeing 787, bien qu'ayant obtenu toutes les certifications légales requises tant par les autorités américaines que les autorités européennes, a connu lui aussi des déboires graves. Suite à de nombreux incendies inexplicables des batteries en cours de vol, le 787 s'est vu retiré toutes ses

autorisations de vol pendant pratiquement une année entraînant des annulations de commandes en cascades de la part de certaines compagnies aériennes clients. Enfin, ces exemples nombreux démontrent que les conséquences à moyen terme d'un défaut de produit peuvent largement impacter la totalité de la survie de la société incriminée d'où la sensibilité de la notion de qualité.

### **Durabilité et fiabilité du produit**

La qualité d'un produit ou d'un projet peut être également évaluée en association avec le taux de reprise, la durabilité et la fiabilité du produit. La durabilité est définie comme la durée d'utilisation d'un produit avant de se détériorer physiquement ou jusqu'à ce que le remplacement soit préférable (Sri Yogi, 2016 [161]). En effet, la reprise pendant l'exécution du projet peut être dédiée à l'amélioration de la durabilité de son *output*. Notons qu'en réalité et selon la stratégie de l'entreprise et son business model (façon dont la société réalise sa valeur ajoutée et donc sa marge) le maintien d'une certaine non qualité peut être délibéré. Prenons l'exemple cité par Bascoul et Moutot (2009) [15] sur la société Michelin. À travers sa solution Michelin flette Service, Michelin est devenu loueur de pneus (facturation du pneu au kilomètre) et plus vendeur de pneus. Avec ce nouveau business model, Michelin est directement impacté par la durée de vie grandissante de ses pneus. On note que la durée de vie d'un pneu est mesurée en kilométrage qui s'étend de 20 000 à 90 000 km selon plusieurs paramètres influant sa vitesse d'usure comme la conduite, la géographie et la catégorie de pneu (à gomme dure, tendre ou pneus toutes saisons). Plus les pneus durent longtemps, plus Michelin gagne d'argent. Il est donc très exigeant sur la reprise des défauts. Dans l'ancien business model, c'est la logique inverse qui prévaut : plus les pneus tiennent longtemps, moins Michelin en vend et donc moins le chiffre d'affaires est élevé.

D'un autre côté, la fréquence de réparation ou reprise après le lancement d'un produit sur le marché reflète le taux de fiabilité du produit (*reliability* en anglais) définie comme la probabilité de n'avoir aucune défaillance dans une période de temps définie (Sri Yogi, 2016 [161]), c'est-à-dire la probabilité que le produit fonctionne correctement sur une période de temps déterminée dans des conditions déterminées d'utilisation.

Dans ce contexte de fiabilité, nous classons aussi le nombre de problèmes résolus dans une période de temps. Par exemple le nombre de bugs informatique suite au développement d'un nouveau logiciel. Signalons néanmoins que certaines défaillances infimes peuvent ne pas être perçues immédiatement mais à long terme. Exemple : micro fissures dans les cuves acier de certains réacteurs nucléaires en France et en Belgique qui n'ont été décelées que récemment alors que les centrales nucléaires fonctionnent depuis plus de 20 ans.

#### **2.2.2.2 Coût de la qualité et de la non qualité**

Les indicateurs cités précédemment sont les plus couramment utilisés dans les tableaux de bord des entreprises (Das et al., 2008 [40]). Or, la qualité obtenue d'un

produit dépend souvent de l'investissement qu'une entreprise engage pour fabriquer la bonne qualité. Ainsi, l'évaluation de la qualité par son coût (Cost of Quality, CoQ) comprend l'ensemble des coûts destinés à l'obtention de la qualité. La méthode classique la plus largement considérée dans la littérature est celle de la prévention-évaluation-défaillance (Prevention-Appraisal-Failure, P-A-F) introduite par Feigenbaum (1956) [55] et connue aussi sous le nom du « modèle de Crosby ». Selon cette méthode, le coût de la qualité rassemble les coûts de prévention, d'évaluation et de défaillance. Les coûts de prévention représentent toutes les dépenses réalisées pour diminuer les non conformités, ce sont donc les investissements humains et matériels nécessaires pour garantir la bonne qualité de produits. Par exemple, les coûts de formations liées à la qualité et les coûts associés à la maintenance (main d'oeuvre et pièces de rechanges). Dans ces coûts, on classe également les coûts associés au processus d'itération présent systématiquement dans les projets de conception. Ce processus se caractérise par des interactions permanentes entre les activités en amont et en aval (Heisig et al., 2010 [73]), phénomène absent dans les projets de fabrication.

Les coûts d'évaluation rassemblent les coûts dédiés à déterminer le degré de conformité à des spécifications préétablies. Ils comprennent notamment les contrôles qualité en cours de fabrication y compris le coût de la main d'oeuvre occupée pendant la durée des opérations de contrôle et les coûts des équipements ou outils de mesures utilisés (comme le PokaYoke ou détromper pour le contrôle dimensionnel des pièces).

Les coûts de défaillance représentent réellement les coûts de la non qualité (CoNQ) qui font finalement partie des coûts d'obtention de la qualité. Ces coûts sont engagés pour couvrir des dysfonctionnements internes ou externes (comme le nombre de dérogations chez le client ou de non conformités fournisseur). Ils comprennent les coût de mauvaise qualité (Cost Of Poor Quality, COPQ), définis par Chiarini (2012) [32] comme les coûts de non conformité et de produits rebuté, et les coût de perte potentielle de qualité (Potential Quality Loss Cost, PQLC) associés aux reprises des activités en cas de non conformité (Kim et al., 2012 [89]). Afin de pouvoir calculer et gérer le CoNQ, les entreprises se basent généralement sur une classification des non conformités entre non conformités internes et externes (de chez le fournisseur).

Pour l'évaluation de la qualité selon ces coûts, nous présentons à la figure 2.2 le modèle P-A-F provenant de Schiffauerova et Thomson (2006) [125] où le niveau de qualité est compris entre 0% (pour 100% de produits défectueux) et 100% (pour 0% de produits défectueux). Cela veut dire que l'évaluation de CoQ selon le modèle P-A-F peut implicitement traduire une mesure objective de la qualité comprise entre 0% et 100%. En outre, il est clair dans cette figure que le CoQ diminue avec la baisse du coût de défaillance ou le CoNQ, ce qui est logiquement associée à une amélioration de niveau de qualité. D'un autre côté, c'est l'investissement en coût de prévention et évaluation qui implique fortement une augmentation du niveau de qualité. A titre d'exemple sur le manque d'investissement en ces coûts, citons ici la catastrophe pétrolière sans précédent et sans équivalent en terme de gravité, la fuite de pétrole constatée dans le Golfe du Mexique provoquée par un défaut de la station de forage pétrolier en offshore de BP Deepwater, dont l'origine repose notamment sur une défaillance des outils et mesures de protection et de contrôle, un manque

d'investissement sur la qualité.

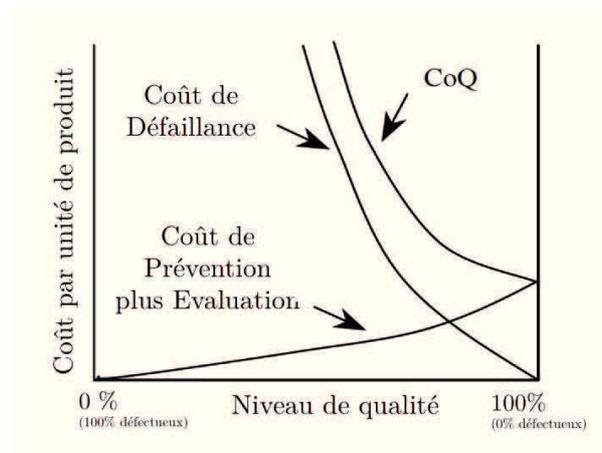


FIGURE 2.2 – Modèle P-A-F et niveau de qualité (Schiffauerova et Thomson, 2006)

L'origine de la non qualité peut être aussi extérieure à l'entreprise mais non détectable malgré un contrôle qualité à la réception. Elle peut se trouver en amont du cycle de production dont l'origine peut se trouver chez les fournisseurs ou sous-traitants, qu'il s'agisse de matières premières ou de produits intermédiaires non conformes, par conséquent incorporés en l'état au cycle de production de produits finis. Prenons l'exemple de TAKATA, un équipementier automobile japonais qui fournit notamment les airbags à General Motor aux USA. Bien que le processus de contrôle qualité de General Motor soit efficace et ayant toutes les qualifications requises, certains conducteurs ont trouvé la mort lors d'accidents aux USA, parce que les airbags ne se sont pas déclenchés au bon moment. Il y a eu des failles dans la détection des malfaçons externes (en amont de la production de véhicules, externalisés chez un fournisseur). On note que dans notre thèse, nous ne considérons pas ces conditions particulières de risque (aléatoire) de non conformité provenant de l'extérieur.

### 2.2.3 Indicateurs qualitatifs

Les indicateurs qualitatifs sont décrits pas des faits. Même s'ils ne sont pas toujours mesurables, ils donnent une appréciation claire de la qualité réalisée qui peut être convertie en mesure quantitative. Ces indicateurs prennent souvent la forme d'une check-list ou d'un descriptif des différents niveaux de la qualité, où l'on cherche à vérifier la conformité des activités à des critères d'inspection ou des niveaux souhaités décrits littéralement et non pas par des chiffres. Dans cette sous-section nous donnons des exemples concrets sur la façon dont évalue la qualité par une check-list ou un descriptif qualité.

#### 2.2.3.1 Check-list

La check-list de contrôle qualité intègre plusieurs items qui considèrent non seulement les spécifications techniques du client mais aussi les exigences internes d'une

entreprise. La check-list est un indicateur approprié pour tout domaine qu'il s'agisse d'un projet de construction d'autoroute, d'un projet informatique ou même d'un développement de service. Les items d'une check-list peuvent être observés par une appréciation de type niveau élevé/moyen/bas (exemple le degré de conformité d'une peinture), par échelonnement (par exemple un échelle de 1 à 5 : excellent, bon, moyen, pas bon, inacceptable) ou même par une appréciation binaire (bon/pas bon).

À Alstom Transport par exemple, la check-list est construite selon une « gamme de contrôle ». Cette gamme décrit pour chaque activité, une liste des éléments ou items à inspecter, le critère d'inspection ou type de défaut recherché et le moyen de contrôle à utiliser. Elle est basée sur des spécifications techniques et des exigences internes et externes. Pour l'évaluation de la qualité, l'évaluateur (le contrôleur qualité) s'appuie toujours sur un critère d'acceptation du client basé sur la gravité de défaut dont la cotation issue d'une AMDEC (Analyse des Modes de Défaillance, de leurs Effets et de leur Criticité) comme nous allons le détailler dans notre étude de cas dans le chapitre 7. Selon cette gravité, le défaut est acceptable ou non et nécessite une reprise obligatoire. Une certaine baisse de qualité est ainsi autorisée pour certaines activités, ce qui signifie que certains items de la check-list peuvent ne pas être validés. Par exemple dans une check-list de 10 items, si 1 élément n'est pas validé mais qu'il respecte le critère d'acceptation, la qualité sera de 0,9 sans que ça affecte la fonctionnalité du produit.

Un autre exemple d'évaluation de la qualité par une check-list est celui de chez Faurecia où on peut accepter certains défauts d'aspect selon leurs zones d'apparition (voir Cracan, 2013 [36]). La figure 2.3 présente les différentes zones de la garniture latérale du coffre de voiture où l'on distingue les zones : A-visible par le client final, B-visible difficilement/occasionnellement par le client final et C-zone non-visible par le client.



FIGURE 2.3 – Défauthèque - Zones A, B, C d'une garniture d'un coffre de voiture - Faurecia

En fonction du défaut observé et de la zone de détection, le critère d'acceptation présenté dans la figure 2.4 indique si la non-conformité est acceptée, refusée ou non applicable (vert, rouge et gris, respectivement). Par exemple, le premier défaut (plis) est acceptable en zone C mais pas en zone A. Le troisième défaut (manque décor) est refusée même dans la zone C, etc.

faurecia		DEFAUTHEQUE GARNITURE LATÉRALE DE COFFRE DROITE ET GAUCHE			PROJET
Site: St Quentin					Date:
					Page: 22
<p>Zone A : Visible par le client final                      Zone B : Visible difficilement / Occasionnellement par le client final                      Zone C : Zone non visible par le client final</p>					
Liste des défauts et critère d'acceptation					
Zone aspect		REFUSE	ACCEPTÉ	NA	
Défaut	Photo	Zone A	Zone B	Zone C	
PLIS		REFUSE	NA	ACCEPTÉ	
DECOLLEMENT		NA	REFUSE	NA	
MANQUE DECOR		NA	NA	REFUSE	
MAUVAISE DECOUPE		REFUSE	NA	NA	

FIGURE 2.4 – Défauthèque – Tableau défauts et critères d'acceptation - Faurecia

### 2.2.3.2 Descriptif des niveaux de qualité

Une autre évaluation de la qualité que celle par la check-list peut être effectuée selon un descriptif détaillé de différents niveaux de la qualité attendue. Prenons l'exemple de l'association européenne des industries du plâtre « Eurogypsum »<sup>5</sup> qui fournit un descriptif pour les critères d'évaluation de la finition et l'enduisage de plaques de plâtre. Comme présenté à la figure 2.5, on distingue quatre niveaux de qualité : Niveau 1 (Q1) basique, Niveau 2 (Q2) standard, Niveau 3 (Q3) supérieur et Niveau 4 (Q4) haut de gamme. Au delà du niveau basique, chaque niveau nécessite au moins les exigences du niveau inférieur. Par exemple l'enduisage (ligne 4 du tableau) selon Q2 comporte la finition de base (Q1) plus un deuxième enduisage pour obtenir une transition continue. Lorsque le cahier des charges ne précise pas le niveau de qualité exigé, le niveau Q2 s'applique automatiquement.

Niveau	Q1	Q2	Q3	Q4
<b>Niveau de finition</b>	Plaques jointées.	Surface plane et lisse. Exigence de qualité visuelle normale.	Surface lisse et plane pour critères de qualité de surface plus exigeant	Aspect visuel demandé de grande qualité.
<b>Qualité apparente à la lumière</b>	Non	normal	Amélioré. Quantité de marques et de traces réduite sous lumière directe. Quelques ombres peuvent être visibles sous lumière rasante.	Niveau supérieur. Quasi absence de marques et de traces. Quasiment aucune ombre en lumière rasante.
<b>Besoins attendus pour l'application</b>	Joints réalisés avec l'enduit adéquat.	Joints réalisés, assurant la continuité de la surface des plaques, si nécessaire poncés.	Joints réalisés avec le niveau de qualité Q2. Une couche d'enduit supplémentaire est appliquée pour élargir la zone des joints. La porosité du papier est bouchée par une couche d'enduit très mince.	Joints réalisés avec le niveau de qualité Q2. Une couche d'enduit avec une épaisseur minimale de 1mm est appliquée sur toute la surface des plaques jointées.
<b>Applications possibles avec ce niveau de qualité</b>	Seulement approprié pour les applications telles que stabilité, résistance au feu et isolation acoustique. Ce niveau de qualité est suffisant pour réaliser les joints de la première rangée de plaques (recouvertes par une autre rangée) ou pour plaques recouvertes par du carrelage ou par une forte épaisseur de plâtre.	Approprié pour les revêtements mats, les peintures ou enduits contenant des grains moyens ou gros (plus grande particule de diamètre supérieur à 1mm).	Approprié pour les revêtements mats, les peintures ou enduits contenant des grains fins (plus grande particule de diamètre inférieur à 1mm).	Approprié pour les finitions lisses ou brillantes telles que papiers peints vinyl ou métalliques, peintures brillantes, stucco lisse ou toute autre technique de finition de haute qualité.
<b>Exigences de planéité</b>	Se référer aux exigences nationales			

FIGURE 2.5 – Descriptif des niveaux de qualité - Eurogypsum

Cette évaluation décrit l'aspect du produit par des mots et non par des chiffres mais selon un descriptif bien détaillé indépendamment de la subjectivité du contrôleur, et donc d'une manière relativement objective. D'ailleurs, les effets d'éclairage

5. Eurogypsum : Produits pour l'enduisage des joints de plaques de plâtre – Définitions, exigences et essais

(par ex. la lumière rasante) peuvent influencer sur l'aspect de la surface et sur l'évaluation du technicien. Pour cela, les conditions d'éclairages sont définis selon le niveau de qualité demandé. La qualité apparente à la lumière (ligne 3 du tableau) pour un niveau Q4 par exemple est meilleure que celle pour un Q3. Les conclusions interprétatives dépendent du niveau recherché et de la nature et modalité des tests effectués.

## 2.3 Mesures quantitatives de la qualité individuelle et globale

Pour satisfaire les éléments d'évaluation d'un produit cités dans la section précédente, non seulement la qualité globale du projet qui aboutit à ce produit doit être contrôlée mais également les qualités individuelles de ses activités. Dans les différents modes d'évaluation de la qualité précédemment abordés, qu'il s'agisse des indicateurs quantitatifs ou qualitatifs, la qualité individuelle d'une activité peut être mesurée quantitativement par une valeur comprise entre 0 et 1 avec ou sans besoin de reprise selon des seuils de qualité minimum. Dans la sous-section 2.3.1, nous introduisons la façon dont on mesure cette qualité individuelle dans la littérature. Nous nous intéressons ensuite, dans la section 2.3.2 aux différentes agrégations proposées pour mesurer la qualité globale d'un projet. Notons que les informations données dans cette section reste introductives et se focalisent sur la mesure de la qualité. Les modélisations quantitatives qui considèrent la qualité sont détaillées dans le chapitre 4.

### 2.3.1 Qualité individuelle d'une activité

#### 2.3.1.1 Taux de conformité à une check-list

Dans les modélisations quantitatives qui considèrent le critère de qualité, la qualité individuelle d'une activité est souvent mesurée par un indicateur variant entre 0 et 1. Cet indicateur traduit le taux de conformité aux spécifications techniques (Icmeli-Tukel et Rom, 1997 [78]) et représente en réalité la proportion d'items validés dans une check-list de contrôle qualité (Mohammadipour et Sadjadi, 2016 [106]). En effet, en mode normal, à toute activité est associé un indicateur de qualité égal à la qualité maximum correspondant à 100% des items validés de la check-list. En revanche, il est communément admis qu'en mode accéléré c'est-à-dire en cas d'accélération d'un projet, on obtient une qualité non plus égale à la valeur maximale mais à une valeur inférieure. Un niveau de qualité de 0,6 par exemple, veut dire qu'on estime à 60% le nombre d'items validés dans la check-list. Pour une activité avec 5 items d'appréciation de sa qualité, cela veut dire que  $0,6 \times 5 = 3$  items sont validés.

La mesure par un indicateur compris entre 0 et 1 peut être également applicable au taux de rebut et au nombre de défauts. Par exemple un taux de rebut de 25% peut être traduit par un niveau de qualité de 0,75. Par ailleurs, on note que certains projets comportent des activités dont l'évaluation de la qualité est binaire : bonne (défauts acceptés, qualité égale à 1) ou mauvaise (défauts refusés, qualité égale 0)

c'est-à-dire qu'il n'y a pas de tolérance aux baisses de qualité. Ou en tout cas si la qualité n'est pas maximale (relativement au niveau attendu ou le seuil de qualité minimum), alors il faut une reprise. Dans ces deux cas, la qualité est binaire et vaut 0 ou 1.

### 2.3.1.2 Estimation de la qualité selon plusieurs critères

Un approche plus fine que la mesure par la check-list est la mesure de la qualité individuelle d'une activité par une somme pondérée d'indicateurs de qualité de cette activité. En effet, le concept de qualité est mesurable en décomposant l'objectif global de qualité en ses principaux attributs et critères relatifs aux activités du projet (El Rayes et Kandil, 2005 [49]). Cette approche est notamment considérée dans la mesure de la qualité de construction où les spécifications techniques sont assez réglementaires et les critères de qualité peuvent être définis de façon à pouvoir les mesurer objectivement et facilement. Pour l'affectation des poids à ces critères de qualité Mungle et al. (2013) [109] proposent d'appliquer l'Analyse Hiérarchique pour les classer par ordre d'importance. A titre d'exemple, prenons une activité d'enrobés bitumineux. La qualité des enrobés est évaluée à l'aide de plusieurs critères comme sa résistance au glissement et son contenu en asphalte. Ces deux critères sont pondérés par l'importance de chacun dans la qualité des enrobés.

### 2.3.1.3 Niveaux discrets de qualité et seuils de qualité minimum

Dans les deux approches précédentes, il y a toujours une possibilité d'établir un seuil de tolérance aux baisses de qualité selon les exigences du client. Dans l'évaluation de Eurogypsum par exemple, les quatre niveaux (Q1, Q2, Q3 et Q4) peuvent être traduits par des seuils de qualité à atteindre. Le niveau Q2 est le niveau standard qui s'applique automatiquement lorsque le client ne spécifie pas d'exigences particulières. Il peut alors être considéré comme le seuil de reprise et quantifié à 0,5. Mais si le commanditaire est très exigeant sur la qualité des finitions, comme cela peut être le cas pour la rénovation ou la construction d'un bâtiment public prestigieux (opéra national de Hambourg par exemple), alors le niveau requis sera Q4 et le seuil de reprise sera fixé à 0,9 voire 0,99. Dans le cas d'une accélération, la réduction de la durée de l'activité va aboutir à l'impossibilité d'effectuer certaines opérations nécessaires à l'obtention de la qualité maximale et dans ce cas le seuil de Q4 ne pourra pas être atteint. Mais si la réduction de la durée est telle que toutes les opérations nécessaires à Q4 sont faites mais pas de manière aussi soignées qu'en mode normal, un expert attribuera une note (comprise entre 0 et 1) de façon relativement objective en respectant certains points de vérification : nombre de rayures (rares rayures, quelques rayures, etc), planéité (totalement lisse, quelques reliefs, etc), homogénéité de la peinture (aucune différence de teinte, quelques différences etc.). A titre d'exemple, il est aussi possible d'effectuer une évaluation dite linguistique avec un "très bien" signifiant une valeur de 1, un "bien", une valeur de 0,95, un "correct" une valeur de 0,9. Et en dessous de 0,9, cela devient insatisfaisant pour une spécification à Q4, et il faudra faire une reprise. Ou alors le client s'est montré flexible et a accepté d'emblée que l'accélération du projet puisse entraîner le non respect de certaines de ses exigences. Si tel est le cas, le seuil de reprise pourra être

fixé à 0,5 (le client est prêt à accepter un niveau standard) ou alors à 0,75 (le client ne veut pas descendre en dessous de Q3).

#### 2.3.1.4 Concept de qualité cumulée

Dans certains cas, les activités sont dépendantes c'est-à-dire que le cumul des défauts individuellement acceptables peut occasionner le dysfonctionnement d'une activité successeur. Fu et Zhang (2016) [60] introduisent le concept de qualité cumulée en partant de l'idée selon laquelle la qualité d'une activité dépend non seulement de sa qualité individuelle mais également de la qualité de ses prédécesseurs. Un regroupement d'activité c'est-à-dire une agrégation des qualités individuelles est nécessaire dans ce cas où se pose alors le problème de la définition même de l'activité dont on évalue la qualité (voir section 4.1.1 du chapitre 4, page 92).

#### 2.3.1.5 Incertitude de mesure de la qualité

La qualité des activités ne peut pas toujours s'évaluer de façon précise par une mesure ou une check-list. Lorsqu'une incertitude pèse sur la mesure de la qualité, celle-ci peut être évaluée par des termes linguistiques comme par exemple élevée, moyenne ou faible. Il est possible de traduire ce type d'évaluation par des nombres flous (voir Zhang et Xing, 2010 [163]). Une représentation de la qualité individuelle  $q_j$  d'une activité  $j$  par un nombre flou triangulaire  $A = \{a_1, a_2, a_3\}$  permet d'exprimer le fait que la qualité individuelle est vraisemblablement de l'ordre de  $a_2$  et assurément plus grande que  $a_1$  mais pas plus élevée que  $a_3$  où  $a_1$ ,  $a_2$  et  $a_3$  prennent leur valeur entre 0 et 1. Ce nombre  $\mu_A$  possède les fonctions membre suivantes

$$\mu_A(q_j) = \begin{cases} \frac{q_j - a_1}{a_2 - a_1} & \text{pour } a_1 \leq q_j \leq a_2 \\ \frac{a_3 - q_j}{a_3 - a_2} & \text{pour } a_2 \leq q_j \leq a_3 \\ 0 & \text{sinon} \end{cases} \quad (2.1)$$

La qualité associée à un mode particulier d'exécution pour l'activité  $j$  peut être décrite par les termes : la plus élevée (The Highest ou TH), élevée (High ou H), très élevée (Very High ou VH), moyenne (Medium, M), faible (Low, L), très faible (Very Low, VL), la plus faible (The Lowest, TL). Chacun de ces termes linguistiques peut être représenté par une suite de nombres flous triangulaires comme sur la figure 2.6 (extraite de Zhang et Xing, 2010 [163]) où  $\mu$  est placé en ordonnée et la qualité en abscisses.

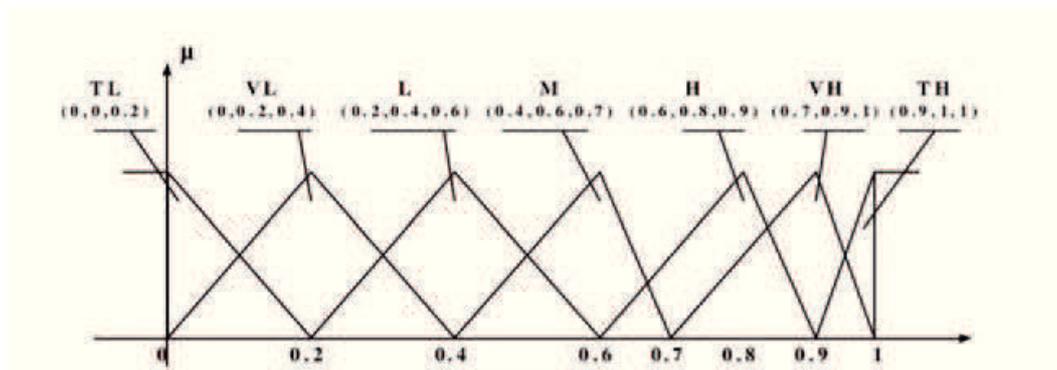


FIGURE 2.6 – Représentation de termes linguistiques pour la qualité par des nombres flous

### 2.3.2 Qualité globale d'un projet

La qualité globale d'un projet résulte d'une agrégation des qualités individuelles de ses activités qui est souvent la moyenne arithmétique (Babu et Suresh, 1996 [14]; Khang et Myint, 1999 [88]; Tareghian et Taheri, 2006 [136]; Rahimi et Iranmanesh, 2008 [120] et Tavana et al. 2014 [138]). Or, dans le cas des projets complexes où les qualités individuelles sont dispersées, c'est la moyenne géométrique des qualités individuelles qui est retenue (Tareghian et Taheri, 2007 [137] et Shankar et al. 2011 [126]).

Cette agrégation présente son inconvénient intrinsèque : la qualité globale d'un projet peut être atteinte mais avec des qualités individuelles qui ne satisfont pas toutes les exigences du maître d'ouvrage. Cette limite est relevée par Khang et Myint (1999) [88] dans leur étude de cas d'un projet de construction d'une usine de ciment. Prenons l'exemple du projet de palettisation automatique (extrait de Kim et al., 2012 [89]) comportant 4 activités principales qui sont la mise à niveau de la dalle en béton ( $i1$ ), l'installation électrique ( $i2$ ), l'installation pneumatique ( $i3$ ) et l'assemblage du système ( $i4$ ). Supposons que les qualités individuelles sont de 0,65 pour ( $i1$ ) et 0,90, pour chacune des trois autres activités. En appliquant la moyenne arithmétique, la qualité globale du projet est de 0,83. Si le seuil de qualité globale minimum est de 0,75 par exemple, le projet est considéré comme conforme aux exigences. En revanche, si le seuil de qualité minimum pour chacune des qualités individuelles est également fixé à 0,75, alors le dallage du béton (activité  $i1$ ) devra être repris.

Une autre mesure de la qualité globale, qui peut répondre à cette limite, est la somme pondérée des qualités individuelles proposée par Afruzi et al. (2014) [2] et Johnson et Liberatore (2006) [83]). Cependant, le poids d'une activité représentant l'importance et la contribution de cette dernière à la qualité globale du projet est un paramètre difficile à estimer. Mungle et al. (2013) [109] et Monghasemi et al. (2015) [108] supposent que la qualité globale du projet est exprimée comme une somme pondérée de la qualité moyenne et de la qualité minimum (minimum des qualités individuelles). Un poids élevé sur la qualité minimum permet de s'assurer qu'aucune activité atteindra un niveau individuel de qualité trop bas. Tandis qu'un poids élevé sur la qualité moyenne permet la recherche d'une plus grande qualité moyenne et on

retrouve alors la même limite de l'agrégation par la moyenne arithmétique. Dans le cas d'une incertitude de mesure de la qualité, la qualité globale peut être exprimée par une somme pondérée des nombres flous, ce qui donne alors un autre nombre flou qui va être représenté par une surface plus grande (additivité des paramètres  $a_i$  dans la sous-section 2.3.1.5).

Au niveau tactique, comme c'est le cas de notre thèse, l'amélioration de la qualité globale d'un projet réside dans l'amélioration et la maîtrise des qualités individuelles de ses activités. En revanche, pour mettre en place les dispositifs nécessaires afin d'atteindre la qualité maximale en mode normal, les entreprises mettent en oeuvre des démarches stratégiques dédiées à la gestion et à l'amélioration de la qualité. Nous présentons ces méthodes dans la section suivante.

## 2.4 Gestion et amélioration de la qualité

La méthode stratégique de gestion de la qualité la plus connue et appliquée dans les différents secteurs industriels et non industriels est le TQM ou le management par la qualité totale. Cette démarche amène à une implication de l'ensemble de personnel visant la satisfaction totale du client. D'autres méthodes telles que Six Sigma et Lean Management sont également largement appliquées. Les principes communs entre ces méthodes résident dans l'orientation client et l'implication de l'ensemble du personnel. Néanmoins, celles-ci présentent également des caractéristiques différentes en raison de leurs origines différentes et de l'historique de leur déploiement en entreprises. Dans cette section, nous portons une attention particulière sur le rôle du management dans l'amélioration de la qualité à travers ces méthodes et sur l'intérêt des indicateurs d'évaluation de la qualité dans de telles démarches. Dans la sous-section 2.4.1, nous fournissons une description complète de la démarche TQM. Nous abordons ensuite le Six Sigma et le Lean Management dans les sous-sections 2.4.2 et 2.4.3, respectivement.

### 2.4.1 Management par la qualité totale (TQM)

#### 2.4.1.1 Définition et origine de la méthode TQM

Rappelons que l'ISO 8402 [79] décrit le concept connu actuellement sous le nom de TQM comme « une approche de gestion centrée sur la qualité, basée sur la participation de toute la population d'une organisation et visant la réussite à long terme par la satisfaction de la clientèle, et les profits à tous les membres de l'organisation et de la société ». Cette méthode n'est donc pas une action ponctuelle. Elle consiste à intégrer dans le temps l'ensemble d'une organisation et des processus (du marketing à la finance, de la conception et de l'ingénierie à la fabrication, du service à la clientèle, etc.), afin d'orienter la culture de l'organisation vers l'amélioration de la qualité. L'industrie automobile japonaise a été la pionnière de cette méthode. TQM est née dans les années 1990 en extension du TQC introduit dans la section 2.1.

### 2.4.1.2 Démarche TQM

Les théories de Deming constituent le fondement de la démarche TQM (Shiba et al., 2003 [127]). Ces théories sont connues sous le nom de PDCA (*Plan, Do, Check, Act*) correspondant aux étapes : planifier, exécuter, contrôler et agir, respectivement. En effet, dans un premier temps, la direction d'une organisation planifie une démarche ou un projet TQM en identifiant les clients de l'organisation et leurs besoins pour établir les objectifs qualité. Suite à cette planification, l'organisation développe des processus qui permettent d'obtenir des produits et services qui répondent aux besoins des clients. Elle met également en place le plan de contrôle et les moyens et outils nécessaires aux opérationnels pour effectuer ce contrôle. L'étape de contrôle permet d'évaluer les résultats par rapport aux objectifs qualité et d'agir c'est-à-dire identifier les projets d'améliorations et mettre en place les équipes et les formations nécessaires aux ressources pour maintenir les objectifs.

D'une manière plus générale, une démarche TQM nécessite la mise en oeuvre de nombreux outils de la qualité. Un outil essentiel dans une démarche TQM est aussi le 5S. Cet outil inspiré du système de production de Toyota permet d'optimiser l'organisation et l'efficacité d'un poste de travail afin d'améliorer la productivité et la qualité. Le nom de l'outil vient des initiales des mots suivants : Seiri / Débarrasser (éliminer ce qui est inutile), Seiton / Ranger (classer, ordonner ce qui est utile), Seiso / Nettoyer (tenir propre les outils, les équipements, l'atelier, etc.), Seiketsu / Organiser (établir et formaliser des règles), Shitsuke/ Maintenir la rigueur (respecter les règles). L'application de ces 5S implique une motivation et une sensibilisation de l'ensemble des employeurs ainsi qu'une communication forte sur la mise en place de la démarche TQM.

Un deuxième levier dans cette démarche est l'auto-évaluation à l'aide des audits internes notamment dédiés à définir les axes et les projets d'amélioration et le besoin en formations spécifiques à la qualité. Un audit permet d'identifier les écarts par rapport à un référentiel ou seuil donné. Il peut être effectué par des auditeurs internes du service qualité ou externes appartenant au groupe de l'entreprise. Par exemple : Alstom Production Systems APSYS est un audit interne à Alstom Transport qui vise à évaluer les bonnes pratiques dans les différents sites d'Alstom transport en cohérence avec la stratégie globale de l'entreprise (qualité, production, supply chain, etc.). La figure 2.7 présente un exemple d'évaluation de la gestion de coût de non qualité dans le cadre d'un audit APSYS.

Theme	N°	Action	Level 1	Level 2	Level 3 : Master Execution	Level 4	Level 5 : Best-in class
Management du CONQ	Q15		Le CONQ est capté par projet.	Les causes racines et les zones d'origine des défauts du CONQ sont identifiés. Un plan d'action est mis en place.	Level 2 + Le qualité valide les points solides. La consolidation du CONQ par plateforme/sub system/Produit/site est disponible.	Le département Finance est impliqué dans les analyses du CONQ Une prévision du CONQ est disponible. Il provient d'un plan d'action "efficace". Les données du CONQ sont connues dans les ateliers.	Une réduction du CONQ S1.1, au minimum de 20 % à comparer avec l'année précédente. Le coût du CONQ S1.1 qui est créé par une équipe (équipe autonome) est connu par cette équipe

FIGURE 2.7 – Exemple d'auto-évaluation par un audit interne- Alstom

Une fois les objectifs et les plans d'actions sont fixés par rapport à des objectifs cibles (bonnes pratiques ou exigences internes et externes), ceux-ci sont complétés par les indicateurs que nous avons présenté dans la section 2.2 (page 40). L'ensemble

des indicateurs est intégré dans un tableau de bord qui est suivi par l'équipe de pilotage de la démarche TQM pour identifier les besoins d'amélioration.

Dans le cadre de cette démarche on rajoute le questionnaire de satisfaction client, qui est systématiquement envoyé par voie électronique lors de l'achat d'un nouveau produit ou même à l'utilisation d'un service comme le transport aérien ou ferroviaire, pour évaluer le *gap* entre la qualité attendue et perçue par le consommateur selon certains critères tels que le confort, la rapidité, la ponctualité, etc. D'après Basu, 2014 [17], ces données sont ensuite utilisées dans des méthodes statistiques pour évaluer le pourcentage de satisfaction aux attentes de client. L'établissement d'un tel indice de référence qui représente le point de vue extérieur permet ainsi d'évaluer la qualité du produit ou d'un service.

### 2.4.1.3 Champs d'application

Bien qu'initialement appliquée aux opérations de fabrication, la méthode TQM est maintenant reconnue comme une méthode de gestion générique, applicable dans les entreprises de services et dans le secteur public comme la gestion hospitalière (Curatolo et al. 2015 [37]). Cette méthode a été largement déployée dans les entreprises au cours des 50 dernières années. Cependant, les trois quarts des projets TQM ont complètement échoué ou ont créé des problèmes assez sérieux dans les organisations souvent en raison d'une négligence du changement culturel dans l'organisation (Valmohammadi et Roshanzamir, 2015 [147]). Selon Brun (2011) [28], à cause de cette faible probabilité de succès, de nombreuses organisations ont adopté à nouveau la norme ISO 9000 [1] qui désigne un ensemble de normes standards fournissant des critères clairs.

## 2.4.2 Six Sigma

### 2.4.2.1 Définition de la méthode Six Sigma

La méthode Six Sigma est à la base une notion statique en référence à la courbe de Gauss. Elle signifie que l'écart entre la borne inférieure et supérieure d'une spécification (exigence d'un client ou d'un contrôle qualité) peut contenir six fois l'écart-type. Pour illustrer cette méthode, prenons l'exemple cité dans Aldowaisan (2015) [10] de fabrication des arbres cylindriques de  $2,00 \pm 0,60$  cm de diamètre (spécification à contrôler). En général, le processus est réglé afin que l'on ait un  $3\sigma$ . En appliquant la loi normale avec  $\sigma = 0,02$ , ceci conduit à un taux de rebut de 2700 pièces pour une production de 1 million de pièces. Si la qualité des processus de fabrication est améliorée, réduisant ainsi la variation  $\sigma$  à 0,01 cm, il en résulte un processus  $6\sigma$ . On aura un taux de rebut de  $2 \times 10^{-9}$  soit 0,0000002%, soit deux pièces au rebut pour un milliard de pièces produites.

Cette méthode orientée qualité vise donc à réduire la variabilité et les défauts d'un processus, dans le but de tendre vers le zéro défaut. Elle vise également à réduire les coûts de toutes sortes, à satisfaire le client et à améliorer la qualité globale des processus, des produits et des services (De Mast et Lokkerbol, 2012 [104]).

### 2.4.2.2 Révolution Six Sigma et facteurs clés de succès

La méthode Six Sigma a été développée en 1986 par Motorola en vue d'accroître sa compétitivité contre les entreprises japonaises dans l'industrie de l'électronique. Cette méthode a été ensuite profondément révisée par General Electric au milieu des années 1990, d'abord sous la forme d'un programme de qualité totale, pour être ensuite classée comme une approche managériale permettant de gérer l'ensemble de l'organisation (Weinstein et al., 2008 [157] ; Brun, 2011 [28]).

Selon Aldowaisan et al. (2015) [10], la première composante stratégique de la méthode Six Sigma est le volet management des processus qui nécessite l'engagement et l'implication totale de la haute direction. La deuxième composante de Six Sigma est la méthode d'amélioration des processus qui comporte deux volets, l'un pour l'amélioration des processus par la méthode de DMAIC (Define, Measure, Analyze, Improve, Control), l'autre pour la création ou la re-conception de processus par la méthode Design For Six Sigma, DFSS. La méthode DMAIC consiste à définir l'objectif, mesurer les attentes des clients, analyser les problèmes, les erreurs et les faiblesses, améliorer et innover et enfin contrôler pour garantir la qualité durable. Cette approche conduit à la satisfaction des clients et à des résultats opérationnellement satisfaisants. La méthode DFSS est une méthode de conception de nouveaux produits ou processus. Celle ci permet d'obtenir une qualité similaire à celle obtenue par l'approche Six Sigma.

Il est clair que Six Sigma s'appuie sur les outils de la méthode TQM. Certains auteurs considèrent que la méthode Six Sigma est une évolution de la méthode TQM (comme Chiarini, 2012 [32]). D'autres considèrent qu'elle est une méthodologie à adopter (Brun, 2011 [28]). Néanmoins les principales différences entre ces deux approches se résument comme suit.

1. Orientation générale et non précise de la méthode TQM vers le résultat final d'un processus, tandis que la méthode Six Sigma met en oeuvre des méthodes d'application précises telles que DMAIC et DFSS qui visent à prévenir les erreurs et se concentrent sur la mesure des économies associées aux diverses actions d'amélioration.
2. Dans la méthode Six Sigma, le départ d'une décision " descendante " ou top-down Leadership, venue du haut, joue un rôle essentiel dans la réussite du déploiement d'outils et de techniques - beaucoup moins dans la méthode TQM - et cela permet d'assurer l'alignement des projets avec les objectifs stratégiques de l'organisation (Brun, 2011 [28]).

De plus, la réussite et la continuité de la méthode Six Sigma dépend de plusieurs facteurs clés en lien avec la qualité d'organisation et des processus tels que :

- l'intégration de la qualité dans l'ensemble des fonctions et départements de l'entreprise, plutôt que de la maintenir comme une entité distincte et spécifique au département de la qualité ;
- la concentration sur des objectifs mesurables et bien définis ;
- la création d'une culture Six Sigma à travers le recrutement des ressources. Le département des ressources humaines doit être impliqué pour la jonction des plans des carrières à des réalisations au sein du programme Six Sigma

pour accroître la motivation et l'engagement du personnel.

### 2.4.2.3 Applications et limites

De nombreuses études ont été menées sur la mise en œuvre de la méthode Six Sigma dans les grandes entreprises de fabrication (Jacobs et al., 2015 [80]; Chiarini, 2012 [32] et John et al., 2008 [82]). En effet, cette méthode trouve une application partout où la variation et le gaspillage existent. Dans Jacobs et al. (2015) [80], les auteurs considèrent la méthode Six Sigma comme une innovation administrative qui peut avoir un impact significatif sur la position concurrentielle d'une organisation, car elle influe sur la capacité de l'organisme à créer un avantage concurrentiel. Elle a été déployée récemment dans les fonctions administratives et de services, notamment dans l'objectif d'améliorer l'efficacité opérationnelle et la qualité de traitement dans les services d'urgence et les salles d'opération (Agarwal et al., 2015 [5]). La rentabilité de cette démarche dans les hôpitaux peut se mesurer en nombre d'heures gagnées. L'évolution de la satisfaction des clients dans ce cas (les patients) est difficile à suivre en raison de l'absence d'une population témoin. Ainsi, certaines composantes de la méthode Six Sigma comme les outils statistiques restent de nature technique plus qu'administrative (Jacobs et al., 2015 [80]).

On note qu'une nouvelle modélisation dite BPM (Business Process Management) est utilisée pour automatiser et gérer les processus améliorés par Six Sigma. Elle rend donc la méthode Six Sigma plus efficace.

## 2.4.3 Lean Management

### 2.4.3.1 Définition et concepts de Lean Management

La méthode Lean Management trouve son origine dans le système de production Toyota (Toyota Production System ou TPS) et se définit comme une méthode stratégique de l'entreprise pour réduire systématiquement les gaspillages et maximiser la valeur dans les processus de fabrication (Bortolotti et al., 2016 [25]). Les concepts globaux sur lesquels s'appuie cette méthode permettent de gérer, d'améliorer et de développer les produits plus rapidement tout en améliorant le niveau de la qualité. Ces concepts sont les suivants :

- analyse de la valeur ajoutée pour le client ;
- élimination des 7 gaspillages (ou Muda en japonais). Les 7 mudas suivants sont identifiés : surproduction, attentes, transport, processus de fabrication sans VA, stocks (perte d'espace et immobilisations financières), déplacements et non qualité ;
- le passage de flux poussés aux flux tirés. Pour rappel, dans le cas de flux poussés, la production est organisée en fonction des prévisions de vente, par lots selon les capacités locales de production. Dans le cas de flux tirés, c'est la demande réelle du client qui déclenche la production ;
- la production en Juste à Temps (Just In Time) selon le Takt Time c'est-à-dire la cadence alignée sur le besoin client.

### 2.4.3.2 Mise en oeuvre d'une démarche Lean Management

La mise en oeuvre d'une démarche Lean Management se fonde principalement sur le Value Stream Map ou VSM (Bortolotti et al., 2016 [25]), ou la représentation du flux de valeur. C'est une description détaillée de l'ensemble des activités nécessaires dans un processus de développement ou de fabrication d'un produit ou service. Cette description permet de détecter les points problématiques, de faire la distinction entre valeur ajoutée ou non et d'éliminer ainsi les gaspillages. C'est en construisant le VSM qu'on mesure notamment le Takt Time dans le but de produire en Juste à Temps et de ne pas former de stock. Ensuite, la description du futur VSM permet d'élaborer un plan d'actions de mise en oeuvre du processus amélioré. La démarche Lean Management ne s'arrête pas à la reconstruction du processus, elle inclut également l'utilisation de techniques d'amélioration continue comme l'outil 5S, la méthode de gestion de stock FIFO (First In First Out), le diagramme de Spaghetti pour la gestion de déplacement des personnes, BPM, le diagramme Ishikawa, etc.

Lean Management est une stratégie managériale qui est focalisée sur la satisfaction du client, comme les autres méthodes d'amélioration de la qualité. En revanche, contrairement à Six Sigma et TQM qui visent le résultat final, cette méthode permet de mesurer avec précision les gains exacts qui lui sont associés à travers le management visuel de flux et la description détaillée d'un processus avant et après la mise en oeuvre des améliorations. En effet, Lean Management est une méthode générale qui peut être utilisée pour l'amélioration de la qualité (dans un processus de gestion de non conformité par exemple), du coût (de stock par exemple), de la durée ou d'autres critères de performance, tandis que TQM et Six Sigma sont des méthodes plus focalisées sur les dimensions de la qualité.

### 2.4.3.3 Lean management en dehors des zones de production

Initialement conçue pour les zones de production, la méthode Lean Management est maintenant adoptée aussi dans les zones hors-production. Par exemple, Bortolotti et al. (2016) [25] étudient l'interaction entre les caractéristiques d'un réseau d'approvisionnement et les pratiques d'un programme Lean. Les auteurs montrent que le déploiement d'une démarche Lean Management dépend beaucoup de la distance entre les fournisseurs et les zones concernées par le Lean. Ainsi, la mise en place d'une démarche Lean Management peut avoir des impacts sur les caractéristiques du réseau d'approvisionnement. Par exemple, une chaîne de production qui doit fonctionner en mode Lean avec une fabrication en Juste à Temps exige également une livraison des matières premières en Juste à Temps avec les quantités nécessaires. Néanmoins, certains fournisseurs ne peuvent pas suivre cette cadence parce qu'ils ont des lots économiques de fabrication par exemple, ils ne peuvent ainsi s'impliquer dans cette démarche sans avoir mis en place des pratiques Lean dans leurs zones également, c'est donc une influence récursive.

Aujourd'hui, du fait que la méthode Lean Management soit née dans les milieux de la production, la majorité de ses outils viennent du secteur industriel. Néanmoins, ils sont également applicables dans d'autres secteurs tels qu'en hôtellerie et en gestion

hospitalière. Cependant, certaines méthodes ne sont pas appropriées à ce type de processus. Par conséquent, une procédure d'évaluation adéquate doit se faire pour définir les outils Lean les plus appropriés pour le champ d'application (Agrawal et al., 2015 [5]; Rauch et al., 2016 [122]).

## 2.5 Conclusion

Dans ce chapitre, nous avons souligné l'existence de deux concepts de la qualité, la qualité objective et subjective. Le concept de qualité subjective est plus pertinent dans le cas des projets de conception. Dans les projets de fabrication que nous considérons c'est le concept de la qualité objective qui est retenu. Nous relevons que l'évaluation objective de la qualité individuelle d'une activité d'un projet, qu'ils s'agissent des indicateurs quantitatifs ou qualitatifs, peut toujours être convertie en une seule mesure quantitative sous la forme d'un score compris entre 0 et 1. Par exemple 90% d'items validés dans une check-list peut présenter un niveau de qualité de 0,9. Même dans le cas d'incertitude lorsque la qualité est linguistiquement évaluée par « bien, moyen, pas bien, etc. », un niveau « moyen » par exemple peut être converti en un niveau de qualité égal à 0,9.

Nous constatons également, comme c'est le cas chez Faurecia, Eurogypsum et Alstom dont des exemples sont présentés à la section 2.2.3, qu'il existe certains « seuils » de qualité minimum sous la forme d'un critère d'acceptation. En dessous de ces seuils, la reprise est obligatoire pour atteindre le niveau souhaité de qualité.

Dans le chapitre suivant, nous analysons les contributions sur le problème d'arbitrage durée - coût en gestion de projets. Cette analyse peut être considérée comme un travail préliminaire à l'étude des modèles d'arbitrage durée - coût qui intègrent de plus la qualité.

## Chapitre 3

# Modèles d'arbitrage durée - coût en ordonnancement de projets

Dans les projets de fabrication à la demande tels que les projets de construction de bâtiments et les grands projets industriels (construction d'une usine d'assemblage ou de process par exemple), les entreprises peuvent chercher à réduire la durée totale d'exécution pour se garantir un avantage concurrentiel en termes de délai de livraison notamment lors des appels d'offre. L'accélération peut être aussi planifiée si le délai proposé ne convient pas au commanditaire ou si ce dernier donne une incitation monétaire pour que le projet se termine plus tôt. Parfois, le gestionnaire de projet a un intérêt économique à accélérer le projet pour démarrer une autre commande plus tôt dans un objectif de maximisation du profit. Dans d'autres cas, l'accélération n'est pas planifiée. Après le démarrage et l'achèvement d'un certain pourcentage du projet, le gestionnaire peut se rendre compte que le projet est en retard en raison d'aléas et prendre la décision d'accélérer les activités restantes afin d'éviter les pénalités contractuelles (Mohamad et al. 2012 [105]). Dans tous les cas, l'accélération d'un projet implique un coût additionnel relatif à l'augmentation des ressources engagées appelé coût d'accélération.

Ce chapitre a pour objet d'analyser la littérature existante sur les problèmes d'arbitrage durée - coût en gestion de projets. Nous commençons par donner une typologie de ces modèles d'arbitrage dans la section 3.1 pour aboutir à leur classification afin de structurer leur présentation. En nous appuyant sur cette typologie, nous présentons d'abord les modèles qui n'intègrent qu'une seule ressource (section 3.2) puis les différentes approches pour modéliser et résoudre les problèmes d'ordonnancement de projets avec des ressources multiples et contraintes (section 3.3) où il s'agit de trouver une séquence de dates de début et de modes d'exécution de sorte à minimiser la durée totale d'exécution sous contraintes de précédence et de respect des capacités disponibles. Il existe plusieurs types d'approches de résolution. Lorsqu'un seul objectif est considéré (minimisation du coût ou de la durée), le problème peut être résolu avec des méthodes exactes, des heuristiques constructives (heuristiques dites dédiées ou spécialisées), des métaheuristiques qui cherchent à améliorer une seule solution à la fois ou encore des métaheuristiques opérant sur une population de solutions. D'autres approches de résolution multi-objectif ont été développées lorsque l'on cherche à minimiser la durée et le coût simultanément. Nous

fournissons dans l'annexe technique A (page 185) une description plus complète de certaines de ces approches et renvoyons opportunément à cette annexe au cours de ce chapitre. Enfin, la section 3.4 conclut le chapitre en relevant notamment la paucité des modèles d'arbitrage durée - ressources.

### 3.1 Typologie des modèles d'arbitrage durée - coût

Pour réduire la durée d'un projet il est possible d'accélérer certaines activités en leur allouant des ressources supplémentaires ou plus productives. Par exemple, la main d'oeuvre peut être augmentée par le recours aux heures supplémentaires, aux travailleurs temporaires ou aux shifts (équipes). En termes de ressources plus productives, on peut citer l'exemple du remplacement d'une fraiseuse simple par une double fraiseuse intégrée avec une table rotative qui permet d'accueillir deux pièces à usiner simultanément.

Nous rappelons que la durée totale d'exécution d'un projet en mode normal est calculée par la méthode CPM (Critical Path Method, méthode du chemin critique) sur la base des durées normales des activités du projet, c'est-à-dire les durées associées à une allocation minimum (normale) des ressources (voir Weglarz et al., 2011 [156]). L'accélération permet de réduire la durée mais à un coût supplémentaire, ce qui conduit au problème d'arbitrage durée-coût ou Time Cost Trade off Problem (TCTP dans la suite). Nous proposons une typologie de ces modèles fondée sur la distinction entre ressource unique et ressources multiples.

#### 3.1.1 Cas d'une ressource unique

Dans le cas d'une ressource unique renouvelable comme la main d'oeuvre par exemple, la durée  $d_j$  d'une activité  $j$  est exprimée comme une fonction non croissante  $\mathcal{D}_j$  de la quantité de ressource  $r_j$  qui lui est allouée. On parle du problème d'arbitrage durée - ressource (Time Resource Trade off Problem, TRTP). La version continue de ce problème (Continuous TRTP ou CTRTP) où la fonction  $\mathcal{D}_j$  est continue n'a jamais été considérée explicitement dans la littérature. Cependant, une variété de problèmes d'ordonnancement de tâches sur plusieurs machines avec des durées continues d'exécution ont été largement étudiés dans de nombreuses contributions dont on peut trouver une revue de la littérature dans Blazewicz et al. (2007) [20]. Le CTRTP est une généralisation du cas discret (Discrete TRTP ou DTRTP) où la fonction  $\mathcal{D}_j$  est discrète, ce qui donne un ensemble fini de modes d'exécution  $m = (d_{jm}, r_{jm})$  pour toute activité  $j$ . Ces modes d'exécution sont tels que le produit  $d_{jm} \cdot r_{jm}$  est égal au contenu total en ressource que l'activité nécessite pour son exécution. Le problème consiste alors à trouver un ordonnancement et à sélectionner un mode d'exécution par activité de sorte à minimiser la durée totale du projet sous contrainte de disponibilité de la ressource. Ce problème a été considéré pour la première fois par De Reyck et al. (1998) [123] et demeure peu étudié dans la littérature. Plusieurs approches de résolution ont été proposées dont on peut trouver une revue dans Demeulemeester and Herroelen (2002) [45].

Les contributions qui considèrent une ressource unique non renouvelable sont de

loin les plus nombreuses. La ressource unique non renouvelable généralement prise en compte est l'argent dont la consommation représente le coût d'une activité. Ce coût lui-même peut-être le coût associé à une consommation agrégée de toutes les ressources allouées à cette activité. Dans les problèmes d'arbitrage coût - durée en temps continu (Continuous TCTP, CTCTP), pour toute activité  $j$  on a une fonction continue  $\mathcal{C}_j$  associant à la durée  $d_j$  de cette activité un coût  $c_j$ , soit  $c_j = \mathcal{C}_j(d_j)$  avec plusieurs formes pour la fonction  $\mathcal{C}_j$  de sorte à refléter différentes hypothèses relatives à la productivité marginale des ressources. Le CTCTP représente le modèle d'arbitrage de base où il s'agit le plus souvent de minimiser le coût sous contraintes de précédence et d'une durée totale d'exécution du projet  $D_{\max}$  à ne pas dépasser.

Dans les problèmes d'arbitrage coût - durée en temps discret (Discrete Time Cost Trade-off Problem, DTCTP), toute activité  $j$  peut être exécutée selon  $M_j$  modes qui correspondent à des combinaisons possibles coût - durée  $(c_{jm}, d_{jm})$  avec  $m = 1, \dots, M_j$ . Un mode correspond donc à une consommation particulière de ressource non renouvelable, l'argent. Le DTCTP consiste à sélectionner pour toutes les activités d'un projet un mode d'exécution par activité pour atteindre l'un des trois objectifs principaux : (i) minimiser la durée totale du projet sous contrainte de budget, auquel cas le coût s'apparente à une ressource contrainte ; (ii) minimiser le coût total du projet sous contrainte d'une durée totale d'exécution à ne pas excéder ; (iii) déterminer la courbe de coût efficiente en fonction de toutes les durées totales possibles du projet.

Qu'il s'agisse du cas continu ou discret, la durée individuelle d'une activité est bornée inférieurement par la durée compressée obtenue avec une allocation maximale de ressource correspondant au mode d'exécution dit compressé. La durée individuelle d'une activité est bornée supérieurement par sa durée dite normale obtenue avec une allocation minimum de ressources.

### 3.1.2 Cas de plusieurs ressources

Lorsqu'il y a plusieurs ressources, il existe généralement pour toute activité un ensemble de modes d'exécution c'est-à-dire un ensemble d'allocations possibles des ressources qui peuvent être affectées à son exécution. Ainsi un mode d'exécution  $m \in M_j$  pour une activité donnée  $j$  est représenté par un vecteur de consommation des différentes ressources et une durée associée  $d_{jm}$ , c'est-à-dire  $(d_{jm}, r_{jm1}, \dots, r_{jmk}, \dots, r_{jmK}, r'_{jm1}, \dots, r'_{jml}, \dots, r'_{jmL})$  où  $r_{jmk}$  désigne la consommation par l'activité  $j$  de la ressource renouvelable  $k$  dans le mode  $m$  avec  $k = 1..K$  et  $r'_{jml}$  désigne la consommation de ressource non renouvelable  $l$  avec  $l = 1..L$ . Toute ressource renouvelable  $k$  est disponible en quantité limitée  $R_k$  à chaque période et toute ressource non renouvelable est disponible en quantité  $R'_l$  sur l'ensemble de l'horizon. La disponibilité par période d'une ressource renouvelable  $k$  est supposée constante d'une période à l'autre. Il s'agit par exemple de la main d'oeuvre ou des machines. Les ressources non renouvelables sont quant à elles disponibles en quantité limitée pour l'ensemble du projet. L'argent (budget), l'énergie et les matières premières appartiennent à cette catégorie. Les ressources qui sont limitées à la fois par période et pour l'exécution du projet sont dites doublement contraintes. La main d'oeuvre peut être une ressource doublement contrainte si par exemple un travailleur

qualifié ne peut être affecté à l'exécution du projet que sur un nombre limité de périodes. Les ressources doublement contraintes ne nécessitent pas d'être considérées explicitement parce que chaque ressource doublement contrainte peut être remplacée par une ressource renouvelable et une ressource non renouvelable (Talbot, 1982 [135]). Le coût correspondant à un mode donné est simplement la somme des coûts unitaires des ressources multipliés par les quantités consommées de ressources dans ce mode.

Lorsqu'il est possible de spécifier la relation entre la durée d'une activité et sa consommation de ressources par une fonction généralement discrète, alors on a affaire à une généralisation du problème DTRTP avec ressources multiples (MDTRTP ou DTRTP with multiple resources). Comme le DTRTP, le MDTRTP a été peu étudié dans la littérature. La spécification d'une telle relation est relativement aisée lorsqu'il s'agit de ressources comme la main d'oeuvre ou les équipements. Cependant, lorsque les ressources sont nombreuses ou hétérogènes, une telle relation peut ne pas exister de sorte que l'on définit des modes d'exécution pour chaque activité où sa durée est estimée en fonction des allocations de ressources. Cette estimation de la durée d'une activité peut être fournie par des experts ou sur la base d'observations passées (collecte des durées observées en fonction des différentes allocations de ressources). L'ensemble des modes d'exécution pour chaque activité est alors relativement réduit et le problème consiste à déterminer dates de début et modes d'exécution pour les activités de sorte à minimiser la durée totale d'exécution du projet en respectant les contraintes de précédence, de non-préemption et de capacité des ressources. Ce problème est connu sous le nom de problème multi-mode d'ordonnancement à contraintes de ressources (Multi-mode Resource Constrained Project Scheduling Problem, MRCPSP) et intègre des ressources généralement discrètes pouvant être renouvelables, non renouvelables et doublement contraintes. Si toutes les ressources sont renouvelables, comme les machines et la main-d'oeuvre et s'il n'existe qu'un seul mode d'exécution par activité, on a le problème standard d'ordonnancement de projet avec plusieurs ressources contraintes (Resource Constrained Project Scheduling Problem, RCPSP). Ce problème n'est pas un problème d'arbitrage coût - durée puisqu'il n'y a qu'un seul mode possible d'exécution pour chaque activité c'est-à-dire une combinaison unique des quantités de chaque ressource qu'elle consomme et à laquelle lui est associée une durée fixe. Ainsi, le MRCPSP est une généralisation du RCPSP dont les approches de résolution ont inspiré celles pour le MRCPSP.

La figure 3.1 résume la typologie des modèles d'arbitrage durée - ressource et durée - coût telle que nous l'avons décrite dans cette sous-section et celle qui précède.

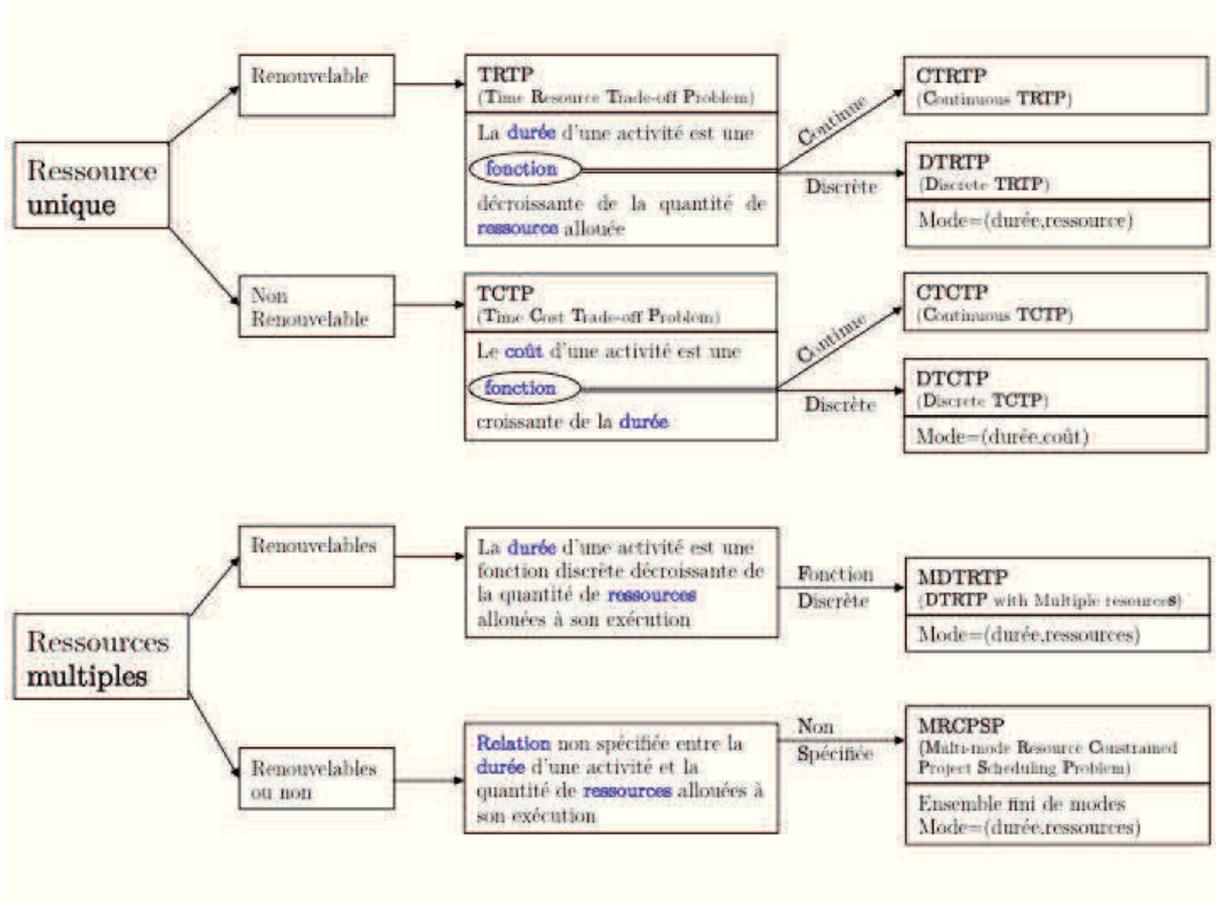


FIGURE 3.1 – Classification des modèles d’arbitrage durée - ressource ou durée - coût

La section 3.2 qui suit ne présente que les modèles d’arbitrage coût - durée avec une ressource unique non renouvelable en raison de la paucité des contributions sur le problème d’arbitrage durée - ressource renouvelable (CTRTP et DTRTP). Dans la section 3.3, nous présentons les modèles à contraintes de ressources multiples en distinguant les différentes approches de résolution.

### 3.2 Modèles d’arbitrage avec une ressource unique non renouvelable

Ces modèles ont pour objectif principal la minimisation du coût total du projet sous contrainte d’une durée maximum à ne pas dépasser ou alors la minimisation de la durée totale du projet sous contrainte de budget. Nous rappelons que les coûts directs d’un projet comprennent les coûts de matériels, des équipements et de la main-d’œuvre directement liés au projet et nécessaires pour l’exécution de ses activités ainsi que le coût de sous traitance pris en compte dans certaines modélisations. Les coûts indirects sont ceux qui ne peuvent être attribués spécifiquement à l’exécution d’un projet particulier et dépendent de la durée de celui-ci. Il peut s’agir par exemple du coût des assurances, des amortissements et des intérêts sur les inves-

tissements en infrastructures ou du salaire des cadres chargés de la supervision de plusieurs projets simultanément. Le coût total du projet est la somme de ses coûts directs et indirects. Une réduction de la durée d'un projet conduit normalement à une réduction de ses coûts indirects dépendant du temps. En revanche, les coûts directs seront d'autant plus importants que l'on cherche à réduire la durée totale d'exécution.

### 3.2.1 Arbitrage coût - durée en temps continu (CTCTP)

Le TCTP en temps continu (Continuous Time Cost Trade-off Problem, CTCTP) est un problème d'arbitrage durée - coût dans lequel le coût d'accélération de toute activité est une fonction continue décroissante de sa durée d'exécution. Nous nous référons à Demeulemeester et al. (1993) [43] et Demeulemeester et Herroelen (2002) [45] pour un état de l'art sur les premières contributions sur ce problème telles que Fulkerson (1961) [61] ou Kelley (1961) [87]) ainsi que les modélisations et les méthodes de résolution pour chaque forme de la fonction coût - durée (linéaire, convexe ou concave).

On considère un ensemble  $V = \{0, 1, \dots, n, n + 1\}$  des activités du projet où les activités fictives 0 et  $n + 1$  désignent respectivement le début et la fin du projet. L'ensemble  $\Gamma^{-1}(j)$  est l'ensemble des prédécesseurs directs de l'activité  $j$ . On note  $t_j$  la date de début de l'activité  $j$ . Pour toute activité  $j = 1..n$ , son coût  $c_j$  est une fonction continue et décroissante de sa durée  $d_j$ , soit  $c_j = \mathcal{C}_j(d_j)$ . La formulation générale du CTCTP s'écrit comme un problème de minimisation du coût

$$\min \sum_{j=1..n} \mathcal{C}_j(d_j) \quad (3.1)$$

sous les contraintes de précédence

$$t_i + d_i \leq t_j, \forall j = 0..n + 1, i \in \Gamma^{-1}(j) \quad (3.2)$$

et d'une durée totale  $D_{\max}$  à respecter soit

$$t_{n+1} \leq D_{\max}, \quad (3.3)$$

où  $t_{n+1}$  est la date de début de la dernière activité (activité fictive de fin  $n + 1$ ) ce qui donne la durée totale d'exécution du projet (makespan).

Comme nous l'avons vu au chapitre 1 (voir sous-section 1.4.3, page 35), la forme de la fonction  $\mathcal{C}_j$  reflète les différentes hypothèses relatives à la productivité marginale des ressources. La figure 3.2 donne les trois formes principales de cette fonction dont vont dépendre les techniques de résolution employées et que nous présentons brièvement dans les sous-sections suivantes. On donne dans cette figure 3.2 la durée en mode normal  $d_j^{\max}$  auquel correspond le coût le plus faible  $c_j^{\min}$  qui est le même dans tous les cas puisque l'on n'ajoute pas de ressources. La durée compressée  $d_j^{\min}$  s'obtient avec un coût de compression  $c_j^{\max}$  qui prend la valeur la plus grande dans le cas convexe et la valeur la plus faible dans le cas concave.

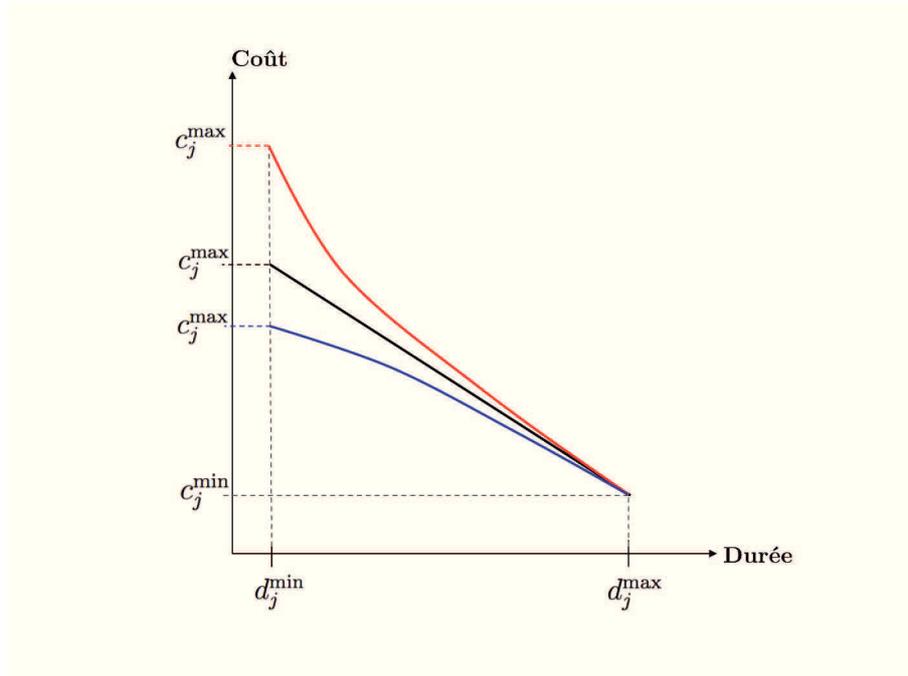


FIGURE 3.2 – Fonctions continues de coût d’une activité linéaire ou convexe ou concave de sa durée

### 3.2.1.1 Fonction linéaire entre le coût et la durée

L’hypothèse de linéarité (voir courbe en noir à la figure 3.2) traduit une productivité marginale constante : la réduction de la durée d’une activité est proportionnelle à l’augmentation de la ressource qu’elle consomme et donc au coût associé à cette augmentation. Cette hypothèse est pertinente lorsque la ressource est un équipement ; elle est en revanche peu réaliste si la ressource considérée est la main d’oeuvre car elle ne permet pas de rendre compte d’une baisse de productivité liée à une utilisation croissante de main d’oeuvre.

Ce cas linéaire peut être formulé comme un programme linéaire et résolu à l’optimum par l’algorithme de Ford-Fulkerson (1961) [61]. D’autres auteurs ont également considéré cette hypothèse de linéarité dans des modèles d’arbitrage faisant intervenir également la qualité et que nous étudions dans le chapitre 4 (voir par exemple Babu et Suresh, 1996 [14])

### 3.2.1.2 Fonction convexe du coût d’accélération

Lorsque le coût d’accélération d’une activité n’est pas une fonction linéaire de sa durée, il est souvent considéré comme une fonction convexe et décroissante (Foldes et Soumis, 1993 [58] et Demeulemeester et al., 1993 [43]), car la convexité reflète une productivité marginale décroissante (voir courbe en rouge à la figure 3.2). L’hypothèse de convexité est particulièrement pertinente lorsque la ressource considérée est la main d’oeuvre car en général, lorsque le nombre de travailleurs augmente, la productivité marginale d’un travailleur supplémentaire est inférieure à celle des travailleurs précédents. Il est clair par exemple qu’une activité pouvant être exécuté

tée par une seule personne en 24 mois ne peut pas être exécutée en un mois par 24 personnes parce que l'accélération peut devenir contre-productive notamment en raison de l'encombrement de l'espace de travail.

Lorsque  $\mathcal{C}_j$  est strictement convexe, l'optimum est trouvé par Elmaghraby (1967) [50]. Lamberson et Hocking (1970) [95] développent une méthode de résolution à l'optimum lorsque la fonction de coût est convexe et différentiable. Kapur (1973) [86] adapte l'algorithme de Ford-Fulkerson (1961) [61] initialement développé pour le cas linéaire au cas particulier où les fonctions  $\mathcal{C}_j$  sont quadratiques et convexes. L'approximation de la fonction convexe par des segments linéaires est aussi une méthode de résolution du CTCTP (Vrat et Kriengkrairut, 1986 [151]). Pour une revue détaillée sur les travaux qui considèrent cette hypothèse de convexité, on peut se référer à Demeulemeester et al. (1993) [43].

### 3.2.1.3 Fonction concave

La concavité de la fonction du coût d'accélération d'une activité traduit une productivité marginale croissante, c'est-à-dire que plus on introduit des ressources additionnelles, plus les ressources sont productives (voir courbe en bleu à la figure 3.2). En d'autres termes, si une unité de ressource permet la production d'une unité de produit en une unité de temps, alors avec deux unités de ressources on obtient plus de deux unités de produit en une unité de temps. C'est le cas par exemple lorsque deux travailleurs peuvent s'entraider ou lorsqu'il y a des synergies dans un groupe de travail. Dans ce cas concave, le problème peut être résolu par la méthode exacte de Branch and Bound (séparation et évaluation) après une approximation linéaire de la courbe (Falk & Horowitz, 1972 [53]). L'algorithme de Branch and Bound est présenté dans l'annexe technique à la sous-section A.1.2.3 (page 189).

## 3.2.2 Modèles d'arbitrage coût - durée en temps discret (DTCTP)

Dans le DTCTP, on considère une ressource unique non renouvelable affectée à l'exécution de toute activité. Le coût  $c_j = \mathcal{C}_j(d_j)$  d'une activité  $j$  est cette fois-ci une fonction discrète  $\mathcal{C}_j$  non croissante de sa durée de sorte que chaque activité  $j$  peut être exécutée selon un ensemble fini de modes  $m = 1..M_j$  correspondant aux allocations possibles de la ressource. Le problème consiste alors à minimiser la durée totale du projet ou le coût total du projet en déterminant pour chaque activité  $j$  un mode d'exécution  $m$  auquel correspond un couple durée - coût  $(d_{jm}, c_{jm})$  en respectant les contraintes de précédence. On note  $x_{jmt}$  la variable binaire qui vaut 1 si l'activité  $j$  exécutée dans le mode  $m$  s'achève à la fin de la période  $t$ . On note de plus  $EF_j$  (Earliest Finish time) la date de fin au plus tôt de l'activité  $j$  qui se calcule en ordonnant le projet au plus tôt et en choisissant pour chaque activité son mode d'exécution le plus rapide. La date de fin au plus tard  $LF_j$  (Latest Finish time) s'obtient en ordonnant le projet avec les modes les plus lents pour toutes les activités. Le problème de minimisation de la durée totale d'exécution d'un projet sous contrainte d'un budget  $B_{\max}$  à ne pas dépasser s'écrit

$$\min \sum_{t=EF_{n+1}}^{LF_{n+1}} t \cdot x_{n+1,m,t} \quad (3.4)$$

Sous la contrainte d'un mode unique par activité

$$\sum_{m=1}^{M_j} \sum_{t=EF_j}^{LF_j} x_{jmt} = 1, \quad j = 0..n + 1, \quad (3.5)$$

avec les contraintes de précédence

$$\sum_{m=1}^{M_j} \sum_{t=EF_i}^{LF_i} t \cdot x_{imt} \leq \sum_{m=1}^{M_j} \sum_{t=EF_j}^{LF_j} t \cdot x_{jmt} - d_{jm}, \quad \forall j = 1..n \quad i \in \Gamma^{-1}(j), \quad (3.6)$$

en respectant la contrainte de budget

$$\sum_{j=1}^n \sum_{m=1}^{M_j} \sum_{t=EF_j}^{LF_j} c_{jm} \cdot x_{jmt} \leq B_{\max}. \quad (3.7)$$

Dans ce problème de budget (appelé B-DTCTP), la contrainte de budget peut être interprétée comme une contrainte de ressource. Un deuxième problème consiste à minimiser le coût total du projet sous contrainte d'une durée totale  $D_{\max}$  à respecter (problème de délai D-DTCTP). Dans le troisième problème, on détermine l'ensemble des points  $(D, B)$  de la frontière d'efficacité tels qu'avec un budget  $B$ , il n'est pas possible d'obtenir une durée totale moindre que  $D$  ou alors la durée  $D$  ne peut pas être atteinte avec un budget inférieur à  $B$ . C'est le problème de la courbe de coût efficace en fonction de toutes les durées possibles (C-DTCTP).

Pour une revue détaillée sur le DTCTP, nous nous référons à Demeulemeester et al. (1993) [43]; De et al. (1995) [41]; Demeulemeester et Herroelen (2002) [45] et Weglarz et al. (2011) [156].

### 3.2.2.1 Modèles avec un objectif unique et variantes du problème DTCTP

Dans ce paragraphe, nous donnons brièvement quelques méthodes de résolution exacte et approchée des problèmes D-DTCTP et B-DTCTP, c'est-à-dire des problèmes avec une fonction objectif unique (minimisation de la durée ou du coût). Nous présentons ensuite les développements récents dans la littérature qui s'intéressent à plusieurs variantes du problème.

En raison de l'importance des problèmes B-DTCTP et D-DTCTP d'un point de vue pragmatique, plusieurs approches de résolution exacte ont été développées. Parmi elles, on peut citer les approches par la programmation dynamique (Hindelang and Muth, 1979 [75]) ainsi qu'un algorithme d'énumération proposé par Harvey and Patterson (1979) [70]. Akkan et al. (2005) [7] relèvent néanmoins l'inaptitude de

ces approches pour résoudre des instances de grande taille. En effet, seuls des problèmes comportant entre 20 et 50 activités peuvent être résolus à l'optimum dans un temps raisonnable. Ces auteurs considèrent alors le problème D-DTCTP pour lequel ils fournissent des bornes supérieures et inférieures sur les durées d'exécution des activités en commençant par utiliser une relaxation des contraintes d'intégralité (relaxation continue). Puis ces bornes sont resserrées en recourant à des techniques de génération de colonnes (voir annexe technique, sous-section A.1.2.4, page 193). L'approche qui en résulte permet de résoudre le D-DTCTP avec un nombre maximum de 138 activités et jusqu'à 20 modes d'exécution par activité, avec un temps de résolution relativement court.

Plusieurs auteurs se sont intéressés plus spécifiquement au problème B-DTCTP de minimisation de la durée sous contrainte de budget (voir par exemple Hazir et al., 2010 [71] ou Hadjiconstantinou et Klerides, 2009 [66]). Ainsi, Hazir et al. (2010) [71] proposent un algorithme exacte de coupes et branchements (voir annexe technique, sous-section A.1.2.3, page 192) pour résoudre des instances de taille réaliste et parviennent ainsi à obtenir la solution optimale pour des instances de 136 activités dans un temps raisonnable.

Une contrainte qu'il est possible de relâcher dans le DTCTP est celle de la non-préemption. Dans le monde réel, la préemption se produit souvent à la suite de restrictions sur les ressources, des contrôles qualité non prévus ou des problèmes techniques. Elle peut être discrète ou continue. Dans le cas discret, la préemption est autorisée à la fin d'une période de temps alors que dans le cas continu, la préemption peut se produire à tout instant. Ainsi, Vanhoucke et al. (2002) [150] rajoutent au DTCTP des contraintes de préemption discrète appelées Time-Switch Constraints (TSC) pour obtenir le problème DTCTP-TSC. Ces contraintes imposent des temps de début spécifiques aux activités et les forcent à être actives ou inactives pendant des fenêtres de temps spécifiques (fenêtres de temps de travail ou de repos, work/rest patterns). Trois types de contraintes de préemption sont considérées selon trois possibilités de fenêtres de temps. Le "day-pattern" c'est-à-dire l'activité peut être seulement exécutée pendant la journée, du lundi au vendredi. Le "dn-pattern" lorsque l'activité peut être exécutée pendant la journée ou la nuit, du lundi au vendredi. Enfin, le "dnw-pattern" qui signifie que l'activité peut être exécutée la journée ou la nuit tous les jours y compris le week-end. Les auteurs proposent un algorithme de Branch and Bound et une procédure heuristique qui utilisent tous deux des calculs de borne inférieure obtenue en résolvant le problème DTCTP sans contraintes de préemption. Les auteurs appliquent leurs méthodes de résolution sur un ensemble de problèmes générés aléatoirement ainsi que sur un exemple illustratif fondé sur un cas réel. Dans Vanhoucke (2005) [148] un nouvel algorithme de Branch and Bound est développé dont la performance est supérieure à celle de l'algorithme précédent.

Parfois, l'objectif de minimisation du coût total du projet est remplacé par la maximisation de la valeur actuelle nette du projet (ou Max-NPV, Net Present Value) c'est-à-dire la différence entre les entrées et les sorties de trésorerie actualisées. Les entrées de trésorerie comprennent les fonds nécessaires pour l'exécution des activités et les sorties comprennent toutes les dépenses liées aux ressources. Ce critère de NPV

est très souvent ignoré dans le DTCTP classique. Cependant, sans flux de trésorerie positif, les obligations de base tels que les paiements aux fournisseurs et aux salariés ne peuvent pas être assurées. Un projet peut échouer si son flux de trésorerie est négatif ou en baisse. Pour ces raisons, la gestion des flux de trésorerie au niveau du projet a attiré l'attention de plusieurs auteurs tels que Erenguc et al. (1993) [51], Icmeli et Erenguc (1996) [77] et Chen et al. (2010) [30]. Ainsi, ces auteurs considèrent les problèmes de planification de projet en intégrant les flux de trésorerie tout au long de la durée de vie du projet. L'objectif est donc de déterminer les dates de début et les durées des activités de sorte que la NPV soit maximisée sous contrainte de ressource non renouvelable.

Vanhoucke et Debels (2007) [149] considèrent également l'objectif de maximisation de la NPV et intègrent en plus des contraintes de TSC, la contrainte de continuité de travail WC (Work Continuity) qui traduit une minimisation de l'inactivité des ressources. Cette contrainte a été initialement définie par El-Rayes et Moselhi (1998) [52] afin de modéliser le déplacement rapide des ressources au sein d'un projet et donc de maintenir une continuité du travail. En effet, cette contrainte est valable dans les projets impliquant la répétition d'activités le long des étapes du projet. Prenons l'exemple de la construction d'un bâtiment à deux étages, avec une activité de menuiserie aux deux étages. La planification doit être faite de sorte à ce que les menuisiers soient continuellement en activité, c'est-à-dire que la menuiserie au deuxième étage doit immédiatement succéder sans interruption à la menuiserie au premier. Ce problème, appelé DTCTP-WC consiste à minimiser le coût total d'un projet c'est-à-dire la somme des coûts directs de ses activités (résultant de la sélection d'un mode pour chaque activité) et du coût de continuité de travail de chaque sous-ensemble d'activités  $V' \subset V$ , qui nécessitent un ensemble commun de ressources soumises à des contraintes de continuité de travail, a priori une spécialité de ressource. Le coût de continuité de travail représente donc le coût par unité de temps des ressources communes d'un groupe d'activité  $V'$ . Ce coût peut être réduit en minimisant le délai entre la première et la dernière activité de  $V'$ . La contrainte de continuité de travail introduit implicitement les ressources, sans distinction entre ressources renouvelable ou non renouvelables, via leurs coûts unitaires, mais sans limites sur la disponibilité de celles-ci. Pour la résolution du problème, les auteurs développent une heuristique qui s'appuie sur les principes de base de la recherche taboue (voir annexe technique, sous-section A.1.3.4, page 194). Cette heuristique est appliquée aux mêmes instances que celles utilisées par Demeulemeester et al. (1998) [44]. Les résultats montrent que les instances de projet avec plus de 50 activités peuvent être résolus à l'optimalité dans 33,93% des cas.

### 3.2.2.2 Détermination de la courbe efficiente coût - durée

La détermination de la courbe efficiente coût - durée revient à rechercher l'ensemble des solutions non dominées en termes de durée totale du projet et de coût total du projet, c'est-à-dire les solutions telles qu'il n'existe pas d'autres solutions réalisables avec une durée totale moindre pour un même coût total ou un coût total moindre pour une même durée totale. Feng et al. (1997) [56] développent un algorithme génétique pour la résolution du problème où toute solution est représentée un chromosome qui, pour l'ensemble des activités, associe un mode d'exécution pour

chacune d'entre elles. Une présentation générale d'un algorithme génétique peut être trouvée dans l'annexe technique (sous-section A.1.4.1, page 194). Le coût total du projet et la durée totale du projet associés à chaque chromosome-solution sont calculés et la fonction de fitness repose sur la distance de chaque chromosome par rapport aux meilleures solutions non dominées trouvées au cours de la recherche et qui forment la courbe efficiente. Cette courbe efficiente est actualisée en intégrant les nouvelles meilleures solutions. L'algorithme est testé sur un problème comprenant 18 activités avec un nombre de modes compris entre 3 et 5 par activité.

Plus récemment, Hadjiconstantinou et Klerides (2009) [66] proposent une approche de résolution du problème D-DTCTP fondée sur la méthode des plans sécants (voir annexe technique, sous-section A.1.2.2, page 192). L'idée de l'approche étant de raisonner sur un sous-ensemble de chemins dans le graphe de précedence qui sont le plus susceptibles de devenir critiques. Pour déterminer la courbe de coût efficiente, les auteurs appliquent leur méthode à un problème de minimisation du coût sous contrainte de délai, en affectant à ce délai plusieurs valeurs possibles (plus ces valeurs sont nombreuses, plus la courbe possède de points). L'algorithme permet de résoudre à l'optimum 89% de problèmes avec 30, 40 et 50 activités en moins d'une minute.

### 3.3 Modèles à contraintes de ressources multiples

Le traitement des contraintes de ressources multiples se fait essentiellement de deux manières dans la littérature. Une première façon de traiter les ressources multiples consiste à considérer qu'il existe pour chaque activité  $j$  un ensemble fini  $M_j$  de modes d'exécution correspondant à des allocations différentes des ressources auxquelles sont associées différentes durées d'exécution de l'activité. L'objectif de ce problème MRCPSP (Multi-mode Resource Constrained Project Scheduling Problem) est de déterminer la date de début de chacune des activités et son mode d'exécution de sorte à minimiser la durée totale d'exécution du projet sous contraintes de précedence, de non-préemption et de capacité des ressources. D'autres objectifs sont aussi possibles mais l'objectif de minimisation de la durée est le plus couramment recherché. Le MRCPSP est une généralisation du RCPSP où les activités sont exécutées dans un mode unique et comme le montrent Kolisch et Drexel (1997) [91], lorsqu'il y a plus d'une ressource non renouvelable, la question de savoir s'il existe une solution réalisable pour le MRCPSP est déjà un problème NP-complet en soi.

Talbot (1982) [135] formule le problème MRCPSP comme un programme linéaire en nombre entier où à chaque activité  $j$  ( $j = 1..n$ ) est associée une variable binaire  $x_{jmt}$  qui vaut 1 si l'activité  $j$  est exécutée dans le mode  $m$  ( $m = 1..M_j$ ) et achevée à la période  $t$  et zéro sinon. L'exécution d'une activité  $j$  dans le mode  $m$  requiert  $r_{jmk}$  unités de ressources renouvelables  $k$  ( $k = 1..K$ ) et  $r'_{jml}$  unités de ressources non renouvelables  $l$  ( $l = 1..L$ ) avec une durée correspondante  $d_{jm}$ . À chaque activité  $j$  est associée une date de fin au plutôt  $EF_j$  (EF, Earliest Finish time) et une date de fin au plus tard  $LF_j$  (LF, Latest Finish time) qui sont calculées en utilisant les allocations maximales et minimales de ressources.

L'objectif principal du MRCPSP est de minimiser la date d'achèvement du projet

c'est-à-dire sa durée totale d'exécution. Le problème se formule à l'identique du DTCTP (équations (3.4) à (3.6) auquel on ajoute les contraintes de capacité sur les ressources renouvelables

$$\sum_{j=1}^n \sum_{m=1}^{M_j} r_{jmk} \sum_{\tau=\max\{t, EF_j\}}^{\min\{t+d_{jm}-1, LF_j\}} x_{jm\tau} \leq R_k, \quad k = 1..K, \quad t = 1..T, \quad (3.8)$$

ainsi que les contraintes de capacité sur les ressources non renouvelables

$$\sum_{j=1}^n \sum_{m=1}^{M_j} \sum_{t=EF_j}^{LF_j} r'_{jmt} x_{jmt} \leq R'_l, \quad l = 1..L, \quad (3.9)$$

où  $T$  est l'horizon calculé avec les modes ayant les plus longues durées. Les ressources doublement contraintes apparaissent dans les deux contraintes sur les ressources renouvelables et non renouvelables.

Une deuxième façon de traiter les contraintes de ressources multiples est de supposer que la durée d'une activité est une fonction discrète de l'allocation des ressources (problème MDTRTP). Ce problème étant nettement moins étudié dans la littérature que le MRCPSP, il ne fait pas l'objet d'un traitement séparé dans ce chapitre. Nous présentons toutefois deux contributions sur ce problème.

Les approches de résolution du MRCPSP ont été consignées dans plusieurs états de l'art récents (Van Peteghem et Vanhoucke, 2014 [115] et Weglarz et al., 2011 [156]). Le MRCPSP peut être résolu optimalement par les méthodes exactes (sous-section 3.3.1). Mais pour des instances de grande taille, il est préférable de recourir à des approches heuristiques en raison du temps de résolution à l'optimum. Plusieurs types d'heuristiques ont été développées pour résoudre le MRCPSP : les méthodes constructives (sous-section 3.3.2) ; les méthodes de recherche locale (sous-section 3.3.3) et les métaheuristiques (sous-section 3.3.4). Ces travaux considèrent tous le problème avec une fonction objectif unique de minimisation de la durée totale du projet ou du coût total du projet et intègrent des ressources renouvelables et non renouvelables ou seulement des ressources renouvelables. En outre, la majorité de ces travaux supposent que les activités ne sont pas préemptables. Des approches multi-objectif ont été développées pour le problème où seules des ressources renouvelables sont prises en compte (sous-section 3.3.4).

### 3.3.1 Méthodes exactes

Les approches exactes de résolution du problème MRCPSP sont proposées par Talbot (1982) [135], Patterson et al. (1989) [114], Speranza (1993) [130], Sprecher (1994) [131], Sprecher et al. (1997) [133], Hartmann et Drexl (1998) [69], Sprecher et Drexl (1998) [132], et Zhu et al. (2006) [165]. Toutes ces approches, à l'exception de celle développée par Zhu et al. (2006) [165] sont fondées sur la méthode de Branch and Bound et sur l'idée d'énumérer partiellement les ordonnancements possibles. Une comparaison détaillée de ces méthodes est fournie dans Hartmann et Drexl (1998) [69].

Patterson et al. (1989) proposent un algorithme dont les étapes sont guidées par l'arbre de précédence qui est l'arbre d'énumération des solutions dans la méthode de Branch and Bound et dont les niveaux correspondent à ceux des activités dans un graphe de précédence (les activités du niveau 2 ne peuvent être ordonnancées avant celles du niveau 1, etc.). Les idées de base de cette approche se trouvent déjà dans Talbot (1982). En restructurant la méthodologie, Sprecher (1994) et Sprecher et Drexl (1998) ont développé des méthodes de calcul des bornes afin d'en accélérer la résolution. Cette méthodologie considère chaque niveau de l'arbre d'énumération, en séparant l'ensemble des activités déjà ordonnancées de l'ensemble des activités éligibles pour l'ordonnancement. Une activité est éligible si toutes ses activités-prédécesseurs ont été ordonnancées. On choisit l'une de ces activités et on lui affecte un mode d'exécution. On calcule sa date de début au plus tôt en fonction des dates de fin de ses prédécesseurs et on vérifie le respect de la contrainte de capacité des ressources. On traite ainsi toutes les activités éligibles d'un niveau avant de passer au niveau supérieur. Une fois que l'on a traité la dernière activité, on a une solution pour le problème. On applique alors un principe de retour en arrière (backtracking) où l'on considère les activités du niveau précédent et on sélectionne pour chacune d'elles un mode d'exécution que l'on n'a pas encore testé. On recalcule chaque fois la date de début au plus tôt et on réactualise la consommation de ressources.

Sprecher et al. (1997) proposent une version différente de cette méthodologie où à chaque niveau  $l$  de l'arbre d'énumération est associé une date fixe  $t_l$  à laquelle les activités peuvent débiter, ce qui modifie la définition d'éligibilité pour les activités. Une activité non ordonnancée est éligible si tous ses prédécesseurs sont ordonnancés avec une date de fin d'exécution antérieure à  $t_l$ . La date  $t_l$  correspond à la date de fin au plus tôt de toutes les activités qui sont en cours d'exécution. Ainsi, à chaque niveau de l'arbre, on procède comme suit. On commence par déterminer  $t_l$  puis l'ensemble des activités éligibles. Certaines d'entre elles ont déjà un mode d'exécution qui leur a été affecté à un niveau précédent mais leur exécution n'a pu débiter en raison d'un problème de disponibilité des ressources. Pour les autres activités, il faut sélectionner un mode avant de lancer leur exécution. Si l'on débute l'exécution de toutes les activités éligibles à la date  $t_l$  et que l'on constate un conflit de ressource, alors on détermine l'ensemble minimal des activités dont on retarde le début d'exécution et on lance les autres. La procédure est ensuite répétée au niveau suivant. Une fois que l'on complète un ordonnancement, on procède à un retour en arrière pour tester les alternatives qui n'ont pas encore été testées (d'autres activités à retarder et d'autres modes d'exécution).

Zhu et al. (2006) [165] ont développé un algorithme exact de coupes et branchements fondé sur la formulation du problème en un programme linéaire en nombres entiers. La relaxation des contraintes d'intégralité sur les variables binaires (qui expriment l'achèvement à une certaine date d'une activité exécutée dans un certain mode) permet le calcul rapide d'une borne sur la durée pour chaque noeud de l'arbre des solutions. Pour chaque solution ne respectant pas les contraintes d'intégralité, on détermine alors un ensemble de contraintes d'inégalité qui sont violées par la solution non entière mais respectées par les solutions entières associées aux noeuds. Lors d'une phase de pré-traitement, des bornes inférieures sur la distance entre chaque paire d'activités  $(i, j)$  liées par la précédence sont calculées (la distance entre 2 ac-

tivités est la durée qui s'écoule entre la fin de l'une d'elles et le début de l'autre). Ces bornes sont utilisées pour réduire le nombre de variables du problème et pour générer des coupes. L'approche est appliquée sur des projets de 20 et 30 activités générés par le logiciel PROGEN développé par Kolisch et al. (1995) [93]. Les auteurs parviennent à obtenir des solutions optimales pour tous les projets comprenant 20 activités et pour la plupart des projets à 30 activités mais pas pour tous. Ainsi, malgré les progrès des méthodes exactes, les projets de plus de 30 activités ne peuvent être tous résolus à l'optimum dans un temps de calcul raisonnable.

Plus récemment, Cheng et al. (2015) [31] considèrent non seulement la préemption mais également le fractionnement. Le fractionnement d'une activité non préemptive autorise l'arrêt momentané de son exécution lorsque le niveau de ressources est temporairement insuffisant. L'exécution doit ensuite être reprise à la prochaine période possible. Dans la version classique du MRCPSp, la quantité de ressources renouvelables par période est supposée constante. Mais en réalité, il est souvent possible que cette disponibilité ne soit pas constante et ce, pour plusieurs raisons. Ces raisons peuvent être prévisibles comme les weekends et les jours fériés pour la main d'oeuvre ou le planning de maintenance pour les machines. Elles peuvent être imprévisibles comme les arrêts maladie ou les pannes des machines. Les auteurs considèrent que la disponibilité d'une ressource varie entre 0 et sa limite supérieure. Ils distinguent trois problèmes de MRCPSp : sans fractionnement des activités, avec fractionnement des activités non préemptives et avec préemption, et examinent la différence entre ces 3 problèmes en termes de formulation mathématique et de solution optimale. Les auteurs développent un algorithme de B&B basé sur l'arbre de précedence conjointement à des heuristiques simples. Les contraintes de ressources sont incluses dans les règles de fenêtres de temps et les règles de priorité. Ils testent leur procédure sur des instances de PSPLIB et PROGEN avec 12 à 16 activités, deux ressources non renouvelables et deux ressources renouvelables et trois alternatives de modes. Les résultats indiquent qu'il y a des réductions significatives de makespan lorsque le fractionnement des activités non préemptives ou la préemption sont autorisés.

Naber et Kolisch (2014) [110] relèvent que le problème MDTRTP a été peu étudié dans la littérature et s'intéressent plus particulièrement au problème d'ordonnancement de projets avec des profils de ressources flexibles où toute activité peut consommer des ressources en quantités variables d'une période à l'autre, la contrainte étant que la somme de ces consommations sur les périodes d'exécution couvre le besoin total en ressources de l'activité. Ainsi, un contenu en travail de 10 hommes-jour pour une activité donnée peut être alloué en un profil constant de 2 hommes sur 5 jours (comme dans le RCPSp) ou en un profil flexible de 3 hommes sur 2 jours et de 2 hommes sur 2 jours. En outre, les auteurs identifient trois types de ressources renouvelables : une ressource principale (la main d'oeuvre) dont la quantité consommée conditionne la quantité consommée de ressources secondaires dite ressources dépendantes (deuxième type de ressources) et le troisième type de ressources, les ressources indépendantes, sont telles que leur quantité consommée est indépendante de la quantité de main d'oeuvre utilisée. Imaginons par exemple un technicien de laboratoire requis sur un jour (ressource principale) qui nécessite un

microscope durant un jour (ressource dépendante). Si le technicien est alloué à 30% sur une activité et à 70% sur une autre activité, ces deux activités étant exécutées en parallèle, alors le microscope est partagé entre ces deux activités selon ces mêmes pourcentages. Si de plus le technicien occupe une salle entière dès lors qu'il est affecté à l'exécution de l'une ou l'autre de ces activités, la salle sera considérée comme une ressource indépendante car son utilisation ne dépend pas du temps de travail alloué à chaque activité. Le problème consiste à déterminer pour chaque activité une date de début d'exécution et une allocation de ressources variables d'une période à l'autre sur la durée d'exécution de l'activité, l'objectif étant de minimiser la durée totale d'exécution du projet sous contraintes de précédence, de couverture des besoins en ressources des activités et de capacité de ces ressources. La durée d'exécution de toute activité va dépendre des quantités de ressources qui lui sont allouées et ces quantités peuvent prendre une infinité de valeurs, de sorte que le problème peut être considéré comme un problème multi mode avec une infinité de modes. Les auteurs proposent plusieurs formulations alternatives dont ils testent l'efficacité en termes de temps de résolution à l'optimum.

### 3.3.2 Heuristiques constructives fondées sur des règles de priorité

Pour le MRCPSP, si l'on suppose que les modes d'exécution sont préalablement affectés aux activités, alors on peut utiliser les heuristiques de résolution du RCPSP pour générer un ordonnancement. Ainsi, les heuristiques constructives pour le MRCPSP résultent d'une adaptation des heuristiques pour le RCPSP que nous commençons par présenter dans ce paragraphe.

Les heuristiques les plus couramment employées pour le RCPSP sont des algorithmes de construction progressive d'une solution qui respecte les contraintes de précédence et de capacité des ressources. Ces algorithmes sont appelés schémas de génération d'un ordonnancement (Schedule Generation Schemes, SGS).

Le SGS sériel pour ordonnancer un projet de  $n$  activités comprend  $n$  itérations. A chaque itération, on détermine l'ensemble des activités éligibles pour être ordonnancées. Ces activités sont telles que tous leurs prédécesseurs ont déjà été ordonnancés (l'exécution a débuté mais n'est pas nécessairement achevée). Sur cet ensemble, on applique une règle de priorité. Il peut s'agir par exemple d'une règle qui sélectionne l'activité avec la marge totale minimum (Minimum Total Slack, ou MIN SLK) c'est-à-dire l'activité qui possède la plus petite différence entre sa durée et la durée totale de tous ses prédécesseurs. L'activité sélectionnée est ordonnancée à sa plus petite date possible de début au plus tôt qui respecte à la fois les contraintes de précédence et de ressources. L'activité est enlevée de la liste des activités restant à ordonnancer et l'algorithme se termine lorsque cette liste est vide.

Dans la version parallèle des SGS, les itérations ne se font non pas sur les activités mais sur les périodes. Ainsi, à chaque itération on commence par déterminer la première période à laquelle des ressources sont disponibles et on repère les activités qu'il est possible d'ordonnancer à cette période. Les activités éligibles sont telles que tous leurs prédécesseurs sont achevés et leurs consommations de ressources n'excèdent pas les capacités restantes. Sur cet ensemble d'activités éligibles on applique

une règle de priorité afin de les ordonner par ordre décroissant de priorité et l'on ordonnance toutes celles qu'il est possible d'ordonnancer tant que les contraintes de capacité sont satisfaites. Les capacités restantes sont actualisées chaque fois qu'une activité est ordonnancée. On avance ensuite à la période suivante où des ressources sont disponibles.

Outre les versions sérielle et parallèle des SGS, il est possible d'exécuter une passe ou plusieurs passes. Les heuristiques les plus anciennes sont des méthodes à une passe qui emploient une version de SGS (sérielle ou parallèle) et une seule règle de priorité pour obtenir un seul ordonnancement réalisable. Les méthodes multi-passe combinent les versions (sérielle ou parallèle) et les règles de priorité pour générer plusieurs ordonnancements. Les méthodes multi-passe les plus courantes sont les méthodes multi-règle de priorité, les méthodes forward-backward et les méthodes de sampling.

Les méthodes multi-règle de priorité consistent à générer plusieurs ordonnancements parmi lesquels on choisit le meilleur en appliquant  $I$  règles différentes de priorité pour obtenir  $I$  ordonnancements (voir par exemple Boctor, 1990 [21]). Une autre possibilité est de combiner  $I$  règles de priorité de sorte à attribuer à chaque activité  $j$  une priorité  $v(j)$  définie comme une somme pondérée des  $I$  indicateurs de priorité avec  $v(j) = \sum_{i=1..I} w_i \cdot v_i(j)$  et  $\sum_{i=1..I} w_i = 1$ , ces combinaisons pouvant aboutir à un nombre virtuellement illimité d'ordonnements. Ce type d'approche est notamment employé dans Ulusoy et Ozdamar (1989) [144] et Thomas et Shali (1997) [140].

Le principe des approches Forward-Backward a été initialement proposé par Li et Willis (1992) [98], l'idée étant d'exploiter les marges libres des activités pour améliorer un ordonnancement initial forward où les activités sont toutes démarrées le plus tôt possible, comme nous l'avons décrit plus haut pour les algorithmes SGS sériel et parallèle. Cette solution initiale obtenue en forward donne une borne supérieure sur la durée d'exécution du projet et cette borne constitue une date de fin pour les activités qui n'ont pas de successeur dans l'application backward où l'on commence par l'activité  $n+1$  pour remonter progressivement à l'activité 0. A chaque itération de l'algorithme backward, on traite l'ensemble des activités dont tous les successeurs ont déjà été ordonnancés (activités éligibles). Pour ces activités, on détermine leur date de fin possible qui correspond à la plus petite date de fin de tous leurs successeurs. La date de début possible est obtenue en retranchant la durée d'exécution de l'activité. On ordonnance toutes les activités éligibles qui, entre ces deux dates, ne violent ni les contraintes de précédence, ni les contraintes de ressource. Pour toute activité éligible qui ne respecte pas les contraintes de précédence avec la date de début ainsi calculée, on recule juste d'une période sa date de début enregistrée dans la solution forward initiale et l'on regarde si les contraintes de précédence sont respectées. Si c'est le cas, on recule encore d'une période cette date, jusqu'à ce que cela ne soit plus possible. Une fois que ces deux solutions ont été générées (la première passe en forward et la deuxième passe en backward), on applique un algorithme d'amélioration forward backward qui consiste à répéter itérativement une passe forward sur la dernière solution obtenue en backward suivie d'une passe backward sur la dernière solution obtenue en forward, et ce, jusqu'à ce que plus aucune amélioration ne soit possible. Dans l'application de l'algorithme forward sur la solution backward, les activités sont

classées par ordre croissant de dates de début collectées dans la solution backward et traitées dans cet ordre, en commençant par l'activité 0. Dans l'application de l'algorithme backward sur la solution forward, les activités sont classées par ordre croissant de dates de fin collectées dans la solution backward en commençant par l'activité  $n + 1$ .

Les méthodes de sampling utilisent généralement un type de SGS et une seule règle de priorité. Plusieurs ordonnancements sont obtenus en randomisant les priorités, c'est-à-dire en associant à toute activité non pas un indicateur de priorité mais une probabilité de sélection d'autant plus élevée que l'indicateur l'est également.

Avec un objectif de minimisation de la durée totale du projet, Boctor (1993) [22] suggère une adaptation des algorithmes SGS parallèles et à une passe au MRCPSP avec les étapes suivantes. Tant qu'il reste des activités à ordonnancer alors (1) déterminer la première période où il existe des ressources renouvelables disponibles ; (2) établir la liste des activités éligibles ; (3) sélectionner l'activité ayant la priorité la plus forte et pour cette activité, choisir le mode d'exécution réalisable ayant la priorité la plus grande ; (4) actualiser les capacités restantes. L'auteur utilise plusieurs règles de priorité sur les activités parmi lesquelles on peut citer les suivantes.

- MIN SLK (Minimum Total Slack) ou marge totale minimum dont le calcul repose sur la différence entre la durée la plus petite des activités non ordonnancées et la durée des activités déjà ordonnancées ;
- MAX NIS (Maximum Number of Immediate Successors) ou sélection de l'activité ayant le plus grand nombre de successeurs immédiats ;
- MAX RWK (Maximum Remaining Work) ou somme, pour chaque activité, de sa durée d'exécution la plus faible ainsi que de celles de tous ses successeurs ;
- MAX PTM (Maximum Processing Time) soit la durée la plus longue dans le mode le plus défavorable ;
- MIN PTM (Minimum Processing Time) c'est-à-dire la plus petite durée possible d'exécution.

Pour les modes, on a notamment

- SFM (Shortest Feasible Mode) ou mode d'exécution réalisable ayant la plus faible durée ;
- LCR (Least Critical Ratio) ou ratio critique le plus faible défini comme le rapport entre la consommation de ressource par un mode et la capacité disponible de cette ressource.

En combinant les règles de priorité sur les activités et sur les modes, l'auteur obtient 21 heuristiques dont la performance est testée sur 120 problèmes à 50 activités et 120 problèmes à 100 activités avec un, deux ou quatre types de ressources.

Dans cette approche, la sélection d'une activité et d'un mode se fait de manière déterministe. Dans l'esprit de l'approche multi passe par randomisation, Drexel et Grunewald (1993) [48] proposent une version aléatoire de ce type d'heuristique constructive où la probabilité de sélectionner un couple activité - mode est proportionnelle à sa durée d'exécution relativement aux durées d'exécution de tous les autres couples activité - mode éligibles.

### 3.3.3 Méthodes de recherche locale

Une présentation générale des méthodes de recherche locale qui sont utilisées ici pour la résolution du problème d'arbitrage durée - coût peut être trouvée dans l'annexe technique (sous-section A.1.3, page 193).

Une stratégie de recherche locale par voisinage pour le MRCPSP a été proposée par Kolisch et Drexl (1997) où une solution est représentée par une liste  $M$  qui associe à chaque activité un mode unique d'exécution et à laquelle on fait correspondre une liste  $S$  de dates d'achèvement pour ces activités dans ces modes d'exécution (ordonnancement). Lors d'une première phase dite de construction de la solution, on assigne à chaque activité le mode d'exécution pour lequel la consommation relative de ressources est la plus faible (obtention de la liste  $M$ ). Ensuite, la capacité restante de ressources non renouvelables est allouée aux activités les plus susceptibles de réduire la durée totale du projet (makespan). Si  $M$  est une solution qui respecte les contraintes de ressources non renouvelables, alors on génère un ordonnancement  $S$  à l'aide d'une heuristique rapide pour résoudre le RCPSP. La solution  $(M, S)$  est enregistrée comme étant la meilleure solution courante  $(M^*, S^*)$ . Lors d'une deuxième phase, on explore le voisinage de  $M$  en générant une solution  $M'$  dans laquelle on change le mode d'exécution d'une seule activité. Si  $M'$  est réalisable alors on génère  $S'$  et si la durée d'exécution de cet ordonnancement est inférieur à celle de la meilleure solution enregistrée alors on pose  $(M^*, S^*) = (M', S')$ . Cette deuxième phase d'exploration du voisinage est itérée un nombre fixe de fois. Lors d'une troisième phase dite d'intensification, un meilleur ordonnancement est recherché pour la meilleure solution courante à l'aide d'une heuristique plus performante pour le RCPSP. La performance de cet algorithme est testée sur deux ensembles d'instances de 10 et 30 activités générées par PROGEN. Chaque ensemble est constitué de 640 instances, les auteurs considèrent deux ressources renouvelables et deux ressources non renouvelables dans chacune des instances. Les résultats obtenus par cet algorithme sont notamment comparés à ceux de Drexl et Grunewald (1993) [48]. Cette heuristique est la seule qui trouve des solutions admissibles pour toutes les instances à 10 activités et pour la plupart des instances à 30 activités.

Boctor (1996) [24] propose une procédure de recuit simulé (Simulated Annealing, SA) pour le MRCPSP où seules des ressources renouvelables sont intégrées. Toute solution est représentée par une liste d'activités à laquelle est associée une liste de modes d'exécution (un seul mode par activité). La liste d'activités est ordonnée de telle sorte que toute activité est précédée par l'ensemble de ses prédécesseurs dans le graphe de précedence et succédée dans la liste par l'ensemble de ses successeurs. Le voisinage d'une solution est obtenue en permutant des activités de la liste. L'auteur montre que la performance de l'algorithme est supérieure à celle des meilleures heuristiques qu'il propose dans deux contributions précédentes (Boctor, 1993 [22] et Boctor, 1994 [23]).

Un algorithme de SA a également été proposé par Jozefowska et al. (2001) [84] qui reprennent le même codage de la solution que celui employé par Boctor (1996) [24]. Le voisinage d'une solution est obtenue en permutant des activités de la liste ou en sélectionnant aléatoirement un autre mode d'exécution pour une activité donnée dans la liste. Deux versions de cet algorithme sont proposées. L'une ne laisse pas-

ser aucune solution non réalisable. L'autre version pénalise ce type de solutions. La performance de l'algorithme est examinée sur des instances de PSPLIB comportant jusqu'à 30 activités, 3 modes par activité, deux ressources renouvelables et non renouvelables. Les résultats obtenus indiquent que la version avec fonction de pénalité des solutions non faisables est la meilleure.

Bouleimen et Lecocq (2003) [26] implémentent également une procédure de recuit simulé (SA) en utilisant la même représentation de la solution, mais un voisinage différent. Les solutions voisines sont générées en deux phases. Pour une activité sélectionnée aléatoirement, son mode d'exécution est remplacé par un autre qui respecte les contraintes de ressource non renouvelables. Si l'on obtient une meilleure solution ainsi, on lui applique une permutation des activités. L'algorithme est testé sur les mêmes instances que celles de Jozefowska et al. (2001) mais aucune comparaison entre les deux méthodes n'est effectuée.

Nonobe et Ibaraki (2002) [112] utilisent la recherche taboue (Tabu Search, TS) où la solution courante est modifiée soit par un changement dans la liste d'affectation des modes soit en modifiant l'ordre des activités dans la liste des activités. La procédure est testée sur des données PSPLIB avec 30 activités mais aucune comparaison n'est faite avec d'autres approches.

### 3.3.4 Métaheuristiques à population de solutions

Les métaheuristiques à population de solutions utilisées pour la résolution du MRCPSP font l'objet d'une présentation plus détaillée dans la sous-section A.1.4 de l'annexe technique (page 194).

Ozdamar (1999) [113] propose un algorithme génétique hybride (Hybrid Genetic Algorithm, HGA) qui repose sur un codage particulier où toute solution est représentée par une liste de modes d'exécution réalisables en termes de consommation des ressources non renouvelables et une liste de règles de priorité (comme MIN SLK, MAX NIS, etc.) qui permet d'ordonner les activités en utilisant successivement les règles de la liste sur l'ensemble des activités éligibles à chaque itération de la procédure SGS. L'opérateur de mutation consiste à changer un mode d'exécution dans la liste des modes ainsi que la position d'une règle de priorité dans la liste des règles. L'opérateur de cross-over s'appuie sur un seul point de croisement et combine une partie des deux listes de l'un des parents avec la partie complémentaire des deux listes de l'autre parent pour générer deux descendants. L'auteur compare la performance de HGA avec celle d'un GA classique où la solution est simplement codée par une liste de modes et une liste d'activités ordonnancées. Les deux algorithmes sont appliqués sur 536 instances générées par ProGen avec 10 activités, 3 modes d'exécution par activité, deux ressources renouvelables et deux ressources non renouvelables, ainsi que 32 instances à 90 activités, deux modes par activité, deux ressources renouvelables et deux ressources non renouvelables. Les résultats obtenus montrent l'efficacité du HGA comparé avec la méthode de recherche locale de Kolisch et Drexler (1997) [91].

Hartmann (2001) [68] propose un GA où toute solution est codée par une liste d'activités ordonnées de sorte à respecter les contraintes de précédence et une liste

de modes d'exécution. Les opérateurs génétiques habituels sont appliqués avec un cross-over utilisant des points de croisement différents sur chacune des deux listes qui composent les parents solution et un opérateur de mutation qui échange la position de deux activités si les contraintes de précédence sont respectées et change aléatoirement un mode. Les solutions qui ne respectent pas les contraintes de ressources sont pénalisées dans la fonction de fitness. L'auteur montre que la performance du GA est supérieure à celles des approches développées par Kolisch et Drexl (1997), Ozdamar (1999) et Bouleiman et Lecoq (2003) sur des instances entre 10 et 30 activités extraites de la bibliothèque de problèmes PSPLIB.

Alcaraz et al. (2003) [9] développent un GA en adoptant aussi le codage en double liste mais en y ajoutant un gène supplémentaire désignant le type de SGS sériel utilisé pour construire l'ordonnancement : forward ou backward. Les opérateurs de cross-over et de mutation opèrent également sur ce gène supplémentaire. Les auteurs utilisent notamment les instances de la bibliothèque PSPLIB et montrent que la performance de leur algorithme dépasse celle de la méthode de recherche locale de Kolisch et Drexl (1997) [91] ainsi que celle du GA de Ozdamar (1999) [113] mais donne de moins bons résultats que le GA développé par Hartmann (2001) [68], et ce, pour toutes les instances à 10 activités.

Plus récemment, Lova et al. (2009) [101] reprennent le codage de Alcaraz et al. (2003) dans un GA hybride avec une fonction de Fitness pénalisant les solutions qui ne respectent pas les contraintes de ressources non renouvelables. Chaque solution possède alors deux indicateurs de performance, la durée totale d'exécution d'une part, et sa pénalité d'autre part. Les solutions sont rassemblées dans plusieurs groupes de performance différente, l'idée étant de séparer les individus efficaces des autres. La fitness d'un individu sera d'autant plus faible que le nombre d'individus dans le groupe auquel il appartient sera grand. Ce principe de pénalisation des solutions ne respectant pas les contraintes de ressources non renouvelables est également adopté par Damak et al. (2009) [39] dans un algorithme à évolution différentielle.

Wuliang et Chengen (2009) [159] développent également un GA mais pour une extension du MRCPSP où chaque mode d'exécution (combinaison durée-consommation ressources) peut utiliser une stratégie d'accélération ou non, ce qui conduit au problème d'arbitrage coût - durée multi-mode à contraintes de ressources (Multi-Mode Resource-Constrained Discrete Time-Cost Tradeoff, MRC-DTCTP). Toute activité  $j$  possède donc une durée normale  $d_{jm}^{\max}$  et une durée compressée  $d_{jm}^{\min}$ , dans chaque mode  $m$ . Deux types de variables binaires sont intégrées dans le modèle :  $y_j$  qui vaut 1 si l'option de compression est activée et 0 sinon, et  $x_{jm}$  qui vaut 1 si l'activité  $j$  est exécutée en mode  $m$ . Le coût total du projet s'écrit alors

$$\sum_{j \in V} \sum_{m \in M_j} x_{jm} \cdot y_j \sum_{k \in K_{jm}} c_k^{\text{dir}} \cdot d_{jm}^{\min} + \sum_{k \in K} c_k^{\text{ind}} \cdot t_{n+1}, \quad (3.10)$$

où  $c_k^{\text{dir}}$  et  $c_k^{\text{ind}}$  désignent respectivement les coûts directs et indirects par unité de temps d'une ressource renouvelable  $k$  et  $K_{jm}$  est l'ensemble des ressources renouvelables utilisées par l'activité  $j$  dans le mode  $m$ . Le coût indirect est celui de la main d'oeuvre permanente nécessaire pour mener le projet jusqu'à la date d'achèvement de la dernière activité  $t_{n+1}$ . Le coût direct comprend uniquement les heures supplémentaires dédiées à la compression d'activités. Le coût de compression associé à la

ressource  $k$  est ainsi calculé par le produit de ce coût direct avec la durée compressée ( $c_k^{\text{dir}} \cdot d_{jm}^{\text{min}}$ ). Les auteurs considèrent non seulement l'objectif de minimisation de la durée mais également l'objectif de minimisation du coût. Dans le GA qu'ils proposent, toute solution est codée par une liste de priorités d'ordonnement réalisables en termes de consommation des ressources renouvelables, une liste de modes d'exécution et une troisième liste désignant les stratégies d'exécution (compression ou non) dans les modes sélectionnés. Ils évaluent la performance de leur algorithme sur des instances à 12 activités et 2 modes extraites de PSPLIB. Il est à noter que ce problème se ramène à un MRCPSP classique si tout mode  $m$  compressé est considéré comme un mode à part entière avec un surcoût  $\sum_{k \in K_{jm}} c_k^{\text{dir}} \cdot d_{jm}^{\text{min}}$ .

Zhang et al. 2006 [162] développent un algorithme par essais particuliers (Particle Swarm Optimization, PSO) pour le MRCPSP où la solution est représentée par une liste de valeurs d'un indicateur de priorité sur les activités et une liste de modes pour les activités. Une procédure de réparation est appliquée pour les solutions avec des modes infaisables pour non respect des contraintes de ressources non renouvelables. Toute solution réalisable est décodée à l'aide d'un SGS sériel. Les auteurs comparent la performance de leur PSO à celle du SA de Bouleimen et Lecocq (2003) [26] et du GA de Hartmann (2001) [68] avec des instances de 10 à 20 activités provenant de PSPLIB. Les résultats montrent que le PSO est un peu moins puissant que le GA mais meilleur que le SA.

Une variante combinatoire du PSO a été utilisée par Jarboui et al. (2008) [81] pour la résolution de MRCPSP. Les auteurs appliquent leur algorithme à des instances de 10 à 30 activités provenant de PSLIB et montrent que sa performance est supérieure à celle du SA de Bouleimen et Lecoq (2003) [26] et à celle du PSO de Zhang et al. (2006) [162].

Li et Zhang (2013) [97] proposent un algorithme d'optimisation par colonies de fourmis (Ant Colony Optimization, ACO) dont ils comparent la performance à celle des approches incluses dans Jarboui et al. (2008) [81] (SA, GA, PSO). sur des instances de PSPLIB de 10 à 30 activités, 3 modes d'exécution par activité et deux ressources renouvelables et non renouvelables. Les auteurs montrent que leur ACO est plus performant que les autres heuristiques à l'exception de CPSO proposé par Jarboui et al. (2008) [81].

Ranjbar et al. (2009) [121] étudient spécifiquement le MDTRTP qui peut être considérée comme une variante du MRCPSP où seules des ressources renouvelables sont intégrées et où la durée d'exécution de toute activité est exprimée explicitement comme une fonction décroissante de sa consommation de ressources. Avec des ressources renouvelables uniquement, un mode  $m$  pour l'activité  $j$  se définit par le vecteur

$(d_{jm}, r_{jm1}, \dots, r_{jmK})$  où  $r_{mjk}$  est la consommation de ressource renouvelable  $k$  par l'activité  $j$  dans le mode  $m$ . Cette consommation  $r_{mjk}$  est liée à la durée  $d_{jm}$  par l'intermédiaire de la quantité totale  $W_{jk}$  de ressource  $k$  que l'activité  $j$  requiert pour son exécution. Prenons l'exemple d'une activité  $j$  nécessitant deux types de main d'oeuvre pour son exécution, électriciens ( $k = 1$ ) et soudeurs ( $k = 2$ ) avec  $W_{1j} = 8$  et  $W_{2j} = 5$  exprimés en équivalent hommes-jour. Cela signifie que

l'activité peut être théoriquement exécutée en 1 jour si 8 électriciens et 5 soudeurs sont affectés en même temps à l'exécution de l'activité  $j$ . Si par ailleurs on dispose d'une main d'oeuvre par jour de  $R_1 = 6$  électriciens et  $R_2 = 7$  soudeurs alors la durée d'exécution la plus petite possible de l'activité est donnée par  $\max \{ \lceil W_{j1}/R_1 \rceil, \lceil W_{j2}/R_2 \rceil \} = \max \{ \lceil 8/6 \rceil, \lceil 5/7 \rceil \} = \max \{ 2, 1 \} = 2$  jours. C'est la ressource la plus contrainte par rapport à la consommation requise qui donne cette durée minimum. En effet, si l'activité ne nécessitait que de la soudure alors elle pourrait être exécutée en 1 jour puisqu'il y a plus de soudeurs par jour que ce que l'activité n'en nécessite au total. En revanche, comme elle requiert l'équivalent de 8 jours de travaux d'électricité et qu'il n'y a que 6 électriciens disponibles par jour, alors les travaux d'électricité sur l'activité ne pourront pas se faire en moins de 2 jours avec une répartition possible de 6 électriciens le premier jour et 2 électriciens le deuxième jour pour la terminer.

Plus généralement, si l'on note  $m = 1$  le mode d'exécution avec la durée d'exécution la plus faible possible, on a  $d_{j1} = \max_k \{ \lceil W_{jk}/R_k \rceil \}$ . Si l'on prend une répartition égalitaire de la consommation des ressources dans tout mode  $m$  alors on a

$$r_{jmk} = \lceil W_{jk}/d_{jm} \rceil \tag{3.11}$$

Pour obtenir les autres modes d'exécution, il suffit d'incrémenter à chaque fois la durée d'exécution d'une période et de recalculer la consommation des ressources selon l'équation (3.11). Chaque nouveau mode  $m+1$  est accepté si pour une ressource au moins, le besoin  $r_{j,m+1,k}$  est différent de  $r_{jmk}$ .

Pour notre exemple, on obtient de cette façon les modes d'exécution consignés dans le tableau 3.1.

$m$	$d_{jm}$	$r_{jm1}$	$r_{jm2}$
<b>1</b>	<b>2</b>	<b>4</b>	<b>3</b>
<b>2</b>	<b>3</b>	<b>3</b>	<b>2</b>
<b>3</b>	<b>4</b>	<b>2</b>	<b>2</b>
<b>4</b>	<b>5</b>	<b>2</b>	<b>1</b>
	6	2	1
	7	2	1
<b>5</b>	<b>8</b>	<b>1</b>	<b>1</b>

Tableau 3.1 – Modes d'exécution efficients.

Les 5 modes d'exécution efficients apparaissent en gras dans le tableau 3.1. Les modes correspondant à des durées égales à 6 et 7 sont éliminés car obtenus avec les mêmes couples de ressources  $(r_{jm1}, r_{jm2}) = (2, 1)$  que le mode 4 qui a une durée de 5 jours seulement.

Les auteurs supposent que si le nombre de modes n'est pas restreint, alors la durée dans le premier mode peut se calculer comme suit  $d_{j1} = \max_k \{ \lfloor \sqrt{W_{jk}} \rfloor, \lceil W_{jk}/R_k \rceil \}$  ce qui donne dans notre exemple même  $d_{j1} = 2$ .

Les auteurs reprennent la formulation du MRCPSP par Talbot (1982) dont ils suppriment simplement les contraintes sur ressources non renouvelables (3.9). Ils développent une métaheuristique de résolution fondée sur la recherche par dispersion (Scatter Search, SS) combinée avec la stratégie d'intensification Path Relinking (PR). La recherche par dispersion oriente l'exploration sur un petit ensemble de bonnes solutions avec un usage modéré de la randomisation. La stratégie Path-relinking explore les trajectoires qui connectent les meilleures solutions obtenues avec des heuristiques.

Dans une phase d'initialisation, chaque élément de la population initiale est créé en générant aléatoirement une liste ordonnée d'activités et de modes d'exécution qui sont ensuite décodées en utilisant un SGS sériel. L'ordre des activités peut être modifié pour obtenir un ordonnancement qui respecte les contraintes de précédence et de disponibilité en ressources. La probabilité d'affecter l'activité  $j$  en position  $i$  dans la liste des activités dépend du nombre de fois où  $j$  a été placée en position  $i$  dans les solutions générées antérieurement. La probabilité d'affecter un mode  $m$  à une activité  $j$  dépend également de la fréquence de ce mode  $m$  dans les solutions existantes. Dans une deuxième phase dite d'intensification, une procédure de recherche locale qui consiste à effectuer deux changements de mode par solution est appliquée jusqu'à ce qu'il ne soit plus possible d'améliorer la meilleure solution. Une troisième phase affecte les solutions dans deux sous-ensembles, l'un contenant les meilleures, l'autre, celles avec des makespans dispersés. L'affectation s'effectue sur la base d'un calcul de distance entre toute solution et les solutions déjà affectées dans les deux sous-ensembles. Enfin, la méthode Path Relinking est utilisée pour combiner les deux sous-ensembles. Une solution initiale de très bonne qualité  $x^{\text{ini}}$  est prise dans le premier ensemble et une autre solution de moins bonne qualité est sélectionnée comme guide  $x^{\text{gui}}$ . L'idée est alors de générer de nouvelles solutions  $x$  en partant de  $x^{\text{ini}}$  pour aller vers  $x^{\text{gui}}$ , et ce, en repérant leurs différences en termes d'ordre des activités ou de modes d'exécution. Par exemple, si l'on constate une première différence en position  $i$  dans la liste des activités de  $x^{\text{ini}}$  et  $x^{\text{gui}}$  alors on remplacera dans  $x^{\text{ini}}$  en position  $i$  l'activité qu'elle contient par l'activité que contient  $x^{\text{gui}}$  à cette même position. La performance de la métaheuristique est testée sur 1050 instances de 10 et 30 activités provenant de Demeulemeester et al. (2000) [42]. Les résultats montrent que la performance de cette procédure dépasse dans la plupart des cas celle d'autres heuristiques existantes comme le SA proposé par Alcaraz et al., (2003) [9]

Une autre variante du problème MRCPSP consiste à relâcher l'hypothèse de non-préemption. Van Peteghem et Vanhoucke (2010) [116] développent un GA pour le MRCPSP en considérant des activités préemptives et non-préemptives. Les auteurs testent leur algorithme sur des instances issues de PSPLIB de 10 à 30 activités et sur les instances de Boctor (1993) [22]. Les résultats montrent que cet algorithme est plus puissant que les GA proposés par Ozdamar (1999) [113], Hartmann (2001) [68] et Alcaraz et al. (2003) [9], ainsi que la recherche locale de Kolisch et Drexel (1997) [91] et le SA de Jozefowska et al. (2001) [84].

### 3.3.5 Approches multi-objectif

Les approches de résolution multi-objectif développées pour le MRCSPS ne considèrent que des ressources renouvelables et sont fondées sur le principe de Pareto optimalité. Dans l'annexe technique (voir section A.2.4), nous détaillons ces méthodes non agrégées et Pareto dont nous donnons ici le principe général. Ces approches non agrégées cherchent à identifier l'ensemble  $P^*$  des solutions Pareto optimales qui par définition sont telles qu'il n'est pas possible d'améliorer l'un des sous-objectifs (durée ou coût) sans détériorer au moins l'un des deux autres. Considérons les deux sous-objectifs  $f_1$  et  $f_2$  à minimiser (la durée et le coût) sous les contraintes de précédence et de non préemption. Toute solution réalisable  $x$  se définit comme une liste d'activités à ordonnancer et une liste de modes d'exécution (un mode d'exécution par activité) qui satisfait les contraintes de précédence et de non préemption. Il est possible de calculer pour toute solution  $x$  la durée totale d'exécution du projet  $f_1(x)$  et le coût total  $f_2(x)$ . Ainsi, pour ce problème, on peut représenter dans un graphique à deux dimensions (durée, coût), l'ensemble des valeurs des deux fonctions objectif associées à toutes les solutions réalisables. Ce type de graphique est donné à la figure 3.3 où  $P$  désigne l'ensemble des objectifs réalisables et  $FP^*$ , la frontière de Pareto c'est-à-dire l'ensemble des valeurs des fonctions objectif associées à l'ensemble  $P^*$  des solutions Pareto optimales.

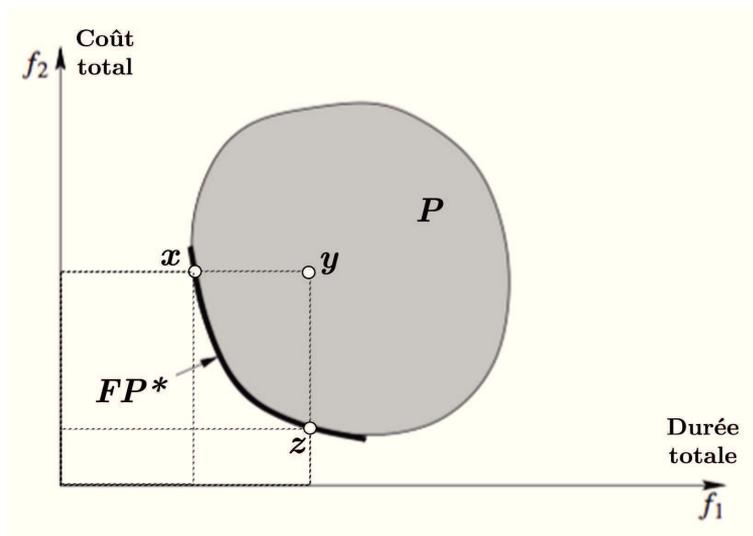


FIGURE 3.3 – Frontière efficiente pour la durée et le coût.

Sur la figure 3.3, on peut observer que les solutions  $x$  et  $z$  situées sur la frontière dominante la solution  $y$  car pour un même coût total, on peut obtenir une durée totale moindre en passant de  $y$  à  $x$ , c'est à dire que l'on a  $f_2(x) = f_2(y)$  mais  $f_1(x) < f_1(y)$ . Ou alors pour une même durée, le coût est moindre en passant de  $y$  à  $z$ , soit  $f_1(y) = f_1(z)$  mais  $f_2(z) < f_2(y)$ . En revanche, partant de  $x$  il est impossible de diminuer le coût sans augmenter la durée (passage de  $x$  à  $z$  par exemple).

En théorie, il est possible de calculer tous les objectifs associés à toutes les solutions réalisables pour obtenir une frontière efficiente continue dite frontière de

Pareto théorique. Mais en pratique, comme il n'est pas possible de calculer cette frontière théorique pour des raisons de temps de calcul, on doit rechercher les solutions non dominées représentatives de cette frontière. Nous présentons à la figure 3.4 un exemple de solutions qui représentent mal la frontière (partie gauche de la figure) ainsi qu'un ensemble de solutions qui la représente bien (partie droite). Les heuristiques de recherche fondées sur Pareto et présentées dans l'annexe technique cherchent précisément à trouver un ensemble de solutions non dominées qui représentent au mieux la frontière théorique en s'appuyant sur des stratégies pour éviter que certaines régions soient surpeuplées et d'autres, sous-peuplées.

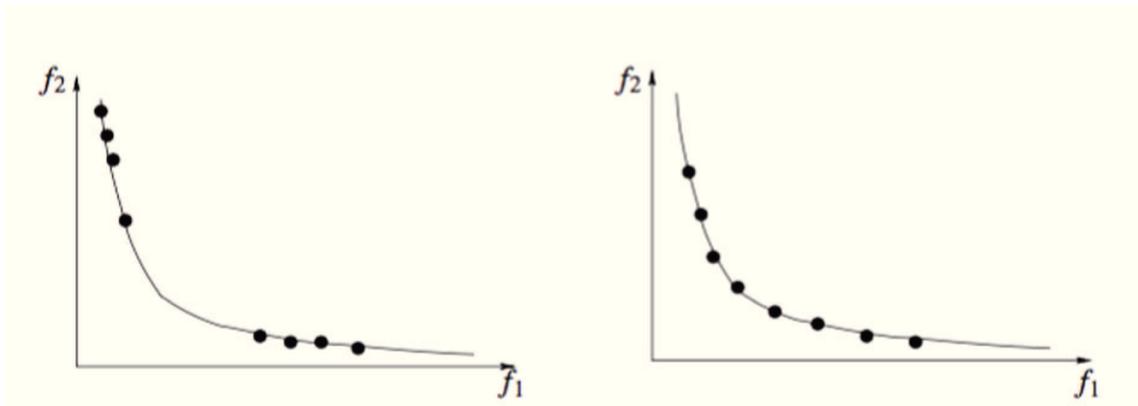


FIGURE 3.4 – Ensemble de solutions fournissant une mauvaise et une bonne représentation de la frontière efficiente.

Toute métaheuristique à population de solutions comme l'algorithme génétique s'adapte naturellement aux problèmes d'optimisation multi-objectif pour lesquels on recherche non pas une solution mais un ensemble de solutions Pareto optimales. Comme les solutions Pareto optimales ne peuvent être connues que si l'on explore toutes les solutions réalisables pour le problème, on recherche alors les meilleures solutions non dominées, c'est-à-dire les solutions explorées par une métaheuristique qui sont telles qu'il n'existe pas de solutions existantes avec des valeurs des fonctions objectif au moins aussi bonnes et une valeur strictement meilleure pour l'un des objectifs au moins. Tout algorithme multi-objectif part d'une population initiale de solutions générées aléatoirement. On détermine pour chacune d'elles son rang de dominance  $r$  qui signifie qu'il existe un nombre  $r - 1$  de solutions qui la dominent. Dans sa version la plus simple, un algorithme génétique multi-objectif affectera une probabilité de sélection pour reproduction d'autant plus grande que le rang de dominance est faible, favorisant ainsi la reproduction des meilleures solutions non dominées (celles qui ont les meilleures valeurs d'objectifs). Cependant une telle stratégie n'encourage pas la diversité de sorte que d'autres algorithmes comme le NSGAI par exemple opèrent sur des sous-populations de rangs de dominance différents. Dans tout sous-ensemble de solutions de rang  $r$  on dit qu'une solution est non dominée car aucune autre solution de rang  $r$  ne la domine. Les solutions potentiellement Pareto optimales ont un rang de dominance égal à 1 car aucune autre solution prise dans l'ensemble de la population ne les domine. Pour chaque

sous-ensemble de solutions non dominées de même rang, on procède à une estimation de leur densité par un calcul de distance (distance de surpeuplement) qui mesure la différence de valeurs des sous-objectifs entre une solution et ses deux plus proches voisins. La distance d'une solution est d'autant plus faible qu'elle est proche de ses voisins, ce qui signifie qu'elle appartient à une région densément peuplée. Inversement, une solution appartenant à une région sous-peuplée aura une distance élevée. L'algorithme NSGAI (Non Sorting GA II) commence par sélectionner aléatoirement deux solutions  $x$  et  $y$  et retient la solution ayant le rang de dominance le plus élevé pour préserver la diversité. Lorsque les deux solutions ont même rang de dominance alors l'algorithme sélectionne celle dont la distance de surpeuplement est la plus grande, c'est-à-dire celle appartenant à une région plus faiblement peuplée. Les opérateurs génétiques qui lui sont alors appliquées permettent d'obtenir de nouvelles solutions susceptibles de peupler cette région sous-peuplée. Les populations parents et enfants sont ensuite fusionnées et le classement Pareto leur est appliqué pour constituer la population du début de l'itération suivante. L'algorithme s'arrête lorsqu'on atteint le nombre maximum d'itérations ou le temps maximum de calcul. Nous détaillons ces étapes dans l'annexe technique à la sous-section A.2.4.2 (page 206). Nous présentons dans la suite de cette sous-section les deux contributions qui s'appuient sur ce type d'approche multi-objectif.

Afruzi et al. (2013) [3] développent un modèle MRC-DTCTP en adoptant une partie des hypothèses de Wuliang et Cheng (2009) [159]. Chaque mode d'exécution est aussi assorti d'une option de compression mais contrairement à Wuliang et Cheng (2009) [159] le coût direct par unité de temps associé à l'utilisation d'une ressource  $k$  donnée n'est pas fixe et peut varier d'une activité à l'autre, d'un mode et d'une stratégie à l'autre. En effet, pour une activité  $j$  exécutée en mode  $m$ , les auteurs distinguent les coûts directs  $c_{jmk}^{\text{dir,max}}$  liés à l'utilisation intensive d'une ressource  $k$  (stratégie de compression) et ceux liés à son utilisation normale, ou une stratégie normale  $c_{jmk}^{\text{dir,min}}$ . Prenons l'exemple du coût de soudage qui varie selon le procédé employé (TIG ou SMAW). Les deux procédés utilisent des outils, des matériaux et une main d'oeuvre différents. Les auteurs supposent de plus que l'occupation par unité de temps d'une ressource  $k$  dépend aussi de la stratégie d'exécution et prend les valeurs ( $r_{jmk}^{\text{max}}$ ) avec l'option de compression et ( $r_{jmk}^{\text{min}}$ ) en mode normal. Les auteurs cherchent à minimiser simultanément la durée totale et le coût du projet sous contrainte de délai ( $D$ ) et de capacité de ressources renouvelables ( $R_k$ ).

Le coût total du projet s'écrit alors

$$\begin{aligned} & \sum_{j \in V} \sum_{m \in M_j} x_{jm} \left( y_j \cdot d_{jm}^{\text{min}} \sum_{k \in K_{jm}} c_{jmk}^{\text{dir,max}} \cdot r_{jmk}^{\text{max}} + (1 - y_j) \cdot d_{jm}^{\text{max}} \sum_{k \in K_{jm}} c_{jmk}^{\text{dir,min}} \cdot r_{jmk}^{\text{min}} \right) \\ & + \sum_{k \in K} c_k^{\text{ind}} \cdot t_{n+1}, \end{aligned} \tag{3.12}$$

et les contraintes de ressources renouvelables s'écrivent

$$\sum_{j \in V} \sum_{m \in M_j} x_{jm} (y_j \cdot r_{jmk}^{\text{max}} + (1 - y_j) \cdot r_{jmk}^{\text{min}}) \leq R_k, \quad k = 1..K. \tag{3.13}$$

Le problème est résolu à l'aide d'un algorithme génétique multi-objectif comme décrit précédemment mais au lieu de s'appuyer sur le concept de Pareto dominance avec les rangs de dominance comme critère de classement, les auteurs utilisent la Pareto dominance floue qui introduit la notion plus fine de dominance de degré  $\alpha$  entre deux solutions car le calcul de ce degré s'appuie sur une différence des valeurs des objectifs entre ces deux solutions. Les auteurs évaluent la performance de leur algorithme sur 10 instances de petite taille et 10 instances de taille plus importante avec 31 activités et 3 modes différents au maximum extraites de PSPLIB avec des paramètres de coût qu'ils génèrent eux mêmes.

Ghoddousi et al. (2013) [63] considèrent le MRCPSP avec ressources renouvelables seulement et cherchent à minimiser simultanément la durée totale d'exécution du projet, son coût total qui comprend des pénalités de retard ainsi que la fluctuation dans l'utilisation des ressources. Cette contribution est donc la seule qui prend en compte un objectif de lissage des consommations de ressources. Les auteurs développent un algorithme NSGA-II pour une recherche de la frontière de Pareto. L'algorithme a été testé sur un projet de construction comprenant 37 activités et une seule ressource renouvelable : la main d'oeuvre avec deux modes d'exécution au maximum par activité.

### 3.4 Conclusion

Les modèles d'arbitrage durée - coût (TCTP) que nous avons analysés dans ce chapitre intègrent une ou plusieurs ressources, renouvelables ou non renouvelables. Le problème d'arbitrage durée - ressource (TRTP) est un cas spécifique où l'on considère une ressource unique renouvelable comme la main d'oeuvre et qui reste très peu étudié dans la littérature. On note que le CTRTP n'a jamais été considéré explicitement mais dans le contexte d'un problème d'ordonnancement de tâches sur plusieurs machines, bien qu'il s'agisse d'une généralisation du DTRTP où chaque activité a un ensemble fini de modes d'exécution.

Comme le DTRTP, le problème avec ressources renouvelables multiples dans un cadre discret (MDTRTP) a été peu étudié dans la littérature. Seuls Naber et Kolisch (2014) [110] et Ranjbar et al. (2009) [121] considèrent ce problème. Dans Ranjbar et al. (2009) [121], les modes d'exécution sont déterminés en reliant la durée de l'exécution d'une activité avec la consommation des ressources par l'intermédiaire de son contenu en travail. Le problème est résolu à l'aide d'une métaheuristique. Contrairement à Ranjbar et al. (2009) [121] et toutes les contributions TCTP, Naber et Kolisch (2014) [110] sont les seuls à considérer des profils de ressources flexibles où, d'une période à l'autre, un nombre différent de travailleurs peut être affecté à l'exécution d'une activité. Cette hypothèse de flexibilité est particulièrement appropriée dans de nombreuses applications réelles. Le problème tel qu'il est présenté par ces auteurs peut être considéré comme un problème multi-mode avec une infinité de modes où chaque mode d'exécution est une combinaison (durée, ressources) car la durée dépend directement de l'allocation des ressources à chaque période d'exécution. Les auteurs proposent plusieurs formulations alternatives pour la résolution du problème.

Lorsque les ressources sont nombreuses ou hétérogènes, le cas le plus traité dans la littérature est le MRCPSP qui, dans sa version classique, intègre à la fois des ressources capacitaires renouvelables et non renouvelables. Pour les instances de projets incluant 20 à 30 activités maximum, ces modélisations sont résolues à l'aide de méthodes exactes. Mais pour des instances de grande taille, des approches heuristiques sont développées ; il s'agit majoritairement de métaheuristiques à population de solutions. On note que ces contributions considèrent toutes l'hypothèse de non-préemption à l'exception de Van Peteghem et Vanhoucke (2010) [117] qui traitent le cas d'activités préemptibles. Tous ces travaux considèrent le problème avec une fonction objectif unique. A notre connaissance seules deux approches de résolution multi-objectif sont développées pour le MRCSPS et l'on relève que ces contributions ne considèrent que des ressources renouvelables. En outre, ces approches multi-objectif sont fondées sur le principe de Pareto optimalité qui implique une multiplicité (et donc une difficulté) de choix entre plusieurs solutions possibles.

Dans le chapitre suivant nous discutons les modèles d'arbitrage durée, coût qui intègrent la qualité ou le Time Cost Quality Trade-off Problem (TCQTP). Nous analysons les différentes hypothèses et limites de ces modèles ainsi que les méthodes de résolution proposées dans la littérature.

# Chapitre 4

## Modèles d'arbitrage durée, coût et qualité

Dans le chapitre 3 nous avons étudié les modèles d'arbitrage durée - coût où l'accélération du projet s'accompagne d'une augmentation des coûts, ce qui définit les termes de l'arbitrage. Mais l'accélération peut également affecter la qualité, en particulier lorsque les ressources additionnelles sont de la main d'oeuvre. En effet, les heures supplémentaires impliquent une fatigue qui entraîne des problèmes qualité. Et le recours aux travailleurs temporaires est susceptible d'entraîner un sureffectif qui peut également faire décroître la qualité. L'impact des heures supplémentaires et du sureffectif sur la qualité a notamment été étudié par Li et al. (2000) [96]. L'incorporation du critère de qualité dans le TCTP conduit au problème d'arbitrage coût, durée et qualité ou Time Cost Quality Trade-off Problem (TCQTP). Dans les modèles TCQTP que nous présentons dans ce chapitre, la qualité individuelle des activités est presque toujours mesurée par un indicateur variant entre 0 et 1 communément interprété comme le pourcentage d'items validés dans une checklist de contrôle qualité. La qualité globale du projet quant à elle est définie par la moyenne arithmétique simple ou pondérée des qualités individuelles, la moyenne géométrique, le minimum des qualités ou par une agrégation combinant moyenne et minimum des qualités individuelles. Le TCQTP se traduit par un lien contradictoire entre les trois critères de performance (durée, coût et qualité) puisqu'une réduction de la durée s'accompagne d'une hausse des coûts et d'une baisse de la qualité. Ainsi, le TCQTP repose sur l'optimisation de trois sous-objectifs et peut être traité comme un problème se ramenant à un objectif unique ou alors par des approches multi-objectif, ce qui justifie le découpage de ce chapitre en deux sections. Ainsi, la première section présente les contributions qui cherchent soit à optimiser un seul objectif en considérant les deux autres comme des contraintes, soit à ramener le problème à un seul objectif en s'appuyant sur des méthodes agrégées (méthode  $\epsilon$ -contrainte, somme pondérée ou méthode du but à atteindre). La seconde section décrit les méthodes non agrégées fondées sur le concept de Pareto optimalité pour la résolution du problème multi-objectif. L'annexe technique détaille ces approches en optimisation multi-objectif (voir section A.2, page 200).

Les modélisations mono-objectif du TCQTP que nous étudions dans la première section de ce chapitre (section 4.1) sont réparties en trois catégories. La première

catégorie (sous-section 4.1.1) groupe les modèles qui considèrent la qualité comme binaire au sens où aucune tolérance aux défauts n'est admise. Si une activité ou tout ou partie des activités d'un projet n'atteint pas la qualité maximale alors des reprises sont nécessaires dont le coût peut être interprété comme un coût de la qualité. Ces modèles cherchent à minimiser le coût de reprise ou la durée de reprise ou une combinaison des deux, en utilisant des approches exactes ou heuristiques. Dans une deuxième catégorie de contributions (sous-section 4.1.2), la qualité est explicitement introduite comme une variable et les relations entre la qualité et la durée d'une part et le coût et la durée d'autre part sont formulées comme des fonctions continues de façon analogue aux modèles d'arbitrage CTCTP étudiés dans le chapitre 3 (sous-section 3.2.1). Une troisième catégorie de modélisations (sous-section 4.1.3) intègre également la qualité comme une variable mais les relations coût, durée et qualité des activités sont discrétisées en supposant l'existence de plusieurs modes d'exécution auxquels sont associés des valeurs pour ces trois critères. On retient l'hypothèse de décroissance du coût par rapport à la durée (plus la durée d'exécution d'une activité est longue, moins elle consomme de ressources additionnelles, et moins le coût est important) et une hypothèse de croissance de la qualité avec la durée (la qualité est d'autant plus faible que la durée est réduite). Les paramètres de durée, coût et qualité dans chaque mode sont généralement supposés être estimés par des experts ou approchés par des fonctions dont l'estimation repose sur des données collectées. Les modélisations de cette troisième catégorie appartiennent à la classe des DTCTP étudiés au chapitre 3 (sous-section 3.2.2) auxquels la qualité est ajoutée pour donner le Discrete Time Cost Quality Trade-off Problem (DTCQTP) qui peut être résolu par des méthodes exactes ou approchées.

En optimisation multi-objectif (section 4.2) c'est ce même DTCQTP qui est traité à une exception près (à notre connaissance, seuls Afruzi et al. 2014 [2] considèrent des ressources multiples dans un modèle MRCPSp avec qualité). Toutes les contributions de la littérature sur le DTCQTP développent des approches de résolution fondées sur Pareto, bien qu'il existe des méthodes non agrégées non Pareto comme par exemple la méthode lexicographique où les objectifs sont classés par ordre de préférence avec une résolution successive des programmes d'optimisation correspondant. Ces approches de Pareto pour le DTCQTP sont relativement récentes (El Rayes et Kandil 2005 [49]). Une première sous-section groupe les approches de résolution s'appuyant sur les algorithmes génétiques multi-objectif (sous-section 4.2.1). Une deuxième sous-section présente les contributions qui exploitent d'autres algorithmes évolutionnaires multi-objectif (sous-section 4.2.2).

Enfin, la section 4.3 conclut ce chapitre en relevant notamment la paucité des modèles avec contraintes de ressources multiples et interroge l'absence dans la littérature des approches non agrégées pour la résolution du problème.

## 4.1 Modèles mono-objectif et approches agrégées

Le seuil de qualité se définit comme la valeur minimum de la qualité en dessous de laquelle on doit obligatoirement faire des reprises. Lorsque le seuil de qualité est fixé à 1, on parle d'évaluation binaire de la qualité au sens où l'activité est évaluée comme conforme aux exigences du commanditaire (ou du maître d'oeuvre) ou non

conforme. La non conformité entraîne des reprises pour rétablir la qualité à son niveau maximum de 1. Mais comme il peut exister une tolérance à de petites baisses de qualité, le seuil peut être fixé à une valeur inférieure à 1 de sorte que la qualité globale d'un projet peut être introduite comme une variable dans les modélisations avec reprise. La sous-section 4.1.1 présente les contributions qui intègrent la reprise dans des modélisations mono-objectif. Les sous-sections suivantes analysent les références de la littérature qui ne considèrent pas la reprise et où la qualité globale est une variable dans des modélisations mono-objectif continues (sous-section 4.1.2) ou discrètes (sous-section 4.1.3).

### 4.1.1 Qualité et reprise

Dans cette sous-section nous étudions les modèles qui intègrent la qualité à travers la reprise. Dans les trois premiers modèles que nous présentons, la qualité d'une activité est considérée comme binaire : l'activité est conforme aux spécifications ou ne l'est pas auquel cas elle nécessite une reprise. Ainsi, la qualité est implicitement prise en compte via la reprise où l'objectif unique est de minimiser le coût de reprise ou la durée ou une combinaison des deux. Le coût de reprise peut être interprété comme le coût de la qualité. Un quatrième modèle admet une tolérance aux baisses de qualité en fixant un seuil inférieur à 1 en dessous duquel les reprises sont toutefois nécessaires. Ces modèles avec reprise ne considèrent qu'un seul objectif avec une approche de résolution exacte ou approchée.

Dans le modèle développé par Icmeli-Tukel et Rom (1997) [78] la qualité est implicitement prise en compte à travers les reprises en cas de non conformité. Les auteurs considèrent plusieurs ressources renouvelables disponibles en quantité limitée et chaque activité  $j$  nécessite pour son exécution un montant  $r_{jk}$  de ressource  $k$ . Les ressources sont allouées à plusieurs intervalles de temps de longueur identique et ponctués par des jalons où les reprises nécessaires sont effectuées. Une fois qu'une activité est achevée, elle requiert à la fois une durée de reprise et un coût de reprise qui est le coût des ressources consommées pour la reprise. Il s'agit de deux paramètres dans le modèle qui font l'objet d'une estimation par des experts. Le problème consiste à déterminer les dates de fin de chaque activité de sorte à minimiser la somme des coûts et des durées de reprise sous les contraintes de précédence, de délai, de budget et de capacité des ressources. Une première formulation s'inspire de celle du RCPSP tandis qu'une deuxième formulation relâche les contraintes d'intégralité sur les dates de fin. Les deux modèles sont résolus à l'optimum pour 110 instances de 32 activités générées aléatoirement avec le générateur d'instances PROGEN développé par Kolisch et al. (1995) [90] avec quatre ressources par instance. Pour obtenir des valeurs pour les paramètres de durée et de coût de reprise, les auteurs utilisent plusieurs fonctions ayant les propriétés suivantes : durée et coût de reprise sont d'autant plus grands que la reprise survient tardivement dans le projet ; le coût de reprise augmente avec la durée de reprise ; la durée de reprise est d'autant plus faible que le nombre de prédécesseurs de l'activité est petit. Ici, les auteurs ne considèrent pas la stratégie d'accélération mais posent le problème de la détermination d'un ordonnancement avec contraintes de ressources où l'objectif est de minimiser la somme des coûts et des durées de reprises. La difficulté d'utiliser ce modèle d'un

point de vue pragmatique réside dans les nombreuses estimations des paramètres de durée et de coût de reprise pour chaque activité et à chaque période possible d'achèvement de ces activités. De plus, l'utilisation comme fonction objectif d'une somme de deux sous-objectifs de dimensions différentes (durée totale de reprise et coût total de reprise) avec un poids identique pour chacun d'entre eux reste discutable en termes de préférences des décideurs et d'interprétation des solutions.

Kim et al. (2012) [89] soulignent que les niveaux de qualité en dessous de 1 ne sont pas acceptables en pratique car les maîtres d'ouvrage (ou les commanditaires) refusent généralement toute dégradation de qualité. Les auteurs partent du constat selon lequel même si la qualité globale d'un projet exprimé comme une moyenne des qualités individuelles atteint une valeur cible, si l'une des activités ne respecte pas les spécifications ou les exigences des commanditaires, alors des reprises doivent nécessairement être engagées pour remettre le niveau de qualité à 1, ce qui impacte la durée et les coûts du projet. Les auteurs parlent d'un coût de perte potentielle de qualité (Potential Quality Loss Cost, PQLC dans la suite). Le PQLC est le coût de reprise en cas de non conformité. Si en revanche, l'activité n'a pas besoin d'être reprise alors on économise ce coût et ce gain d'opportunité constitue indirectement une mesure de la qualité de l'activité. Les auteurs supposent que chaque activité peut être exécutée en mode normal (auquel cas le PQLC est économisé) et en mode compressé auquel est associé le coût PQLC. Toute activité compressée est automatiquement considérée comme une activité à risque de non conformité. Plusieurs types de non conformité lui sont associés ainsi qu'un coût et une durée de reprise. Le PQLC est limité par le coût de compression. La durée de reprise est bornée par la durée compressée. La durée compressée tient implicitement compte de la durée de reprise car elle est inférieure à celle que l'on pourrait atteindre si la reprise n'était pas considérée. Le problème est formulé comme un programme d'optimisation linéaire mixte en nombre entier où l'objectif est de minimiser le coût du projet sous contraintes de précédence, d'une durée totale maximum à respecter et d'un taux limité des activités à risques, c'est-à-dire qu'on n'accélère qu'une partie de toutes les activités du projet pour maintenir un risque global de non conformité acceptable. Le coût du projet est la somme des coûts de compression et des coûts de non conformité de ses activités. En d'autres termes, ces auteurs considèrent que la maximisation de la qualité du projet réside dans la minimisation de ses coûts et notamment de son PQLC. Les variables du modèle sont les durées (continues) des activités  $d_j$  et les variables binaires  $y_j$ , avec  $y_j$  qui vaut 1 si l'activité  $j$  est compressée et zéro sinon. On note  $c_j^{\text{PQL}}$  le PQLC de l'activité  $j$  et  $c_j^{\text{dir}}$  le coût direct d'une unité de temps consacrée à l'exécution de l'activité  $j$ . L'objectif consiste à minimiser la somme des coûts de compression et des coûts de perte potentielle de qualité associés aux non-conformités, ce qui s'écrit

$$\min \sum_{j=1}^n c_j^{\text{dir}} \cdot d_j + \sum_{j=1}^n c_j^{\text{PQL}} \cdot y_j. \quad (4.1)$$

Le modèle est résolu à l'optimum pour un projet de palettisation automatique comportant 18 activités. Ce modèle produit des résultats de type worst-case c'est-à-dire qu'il indique le coût maximum des reprises alors qu'en réalité les risques de non

conformité n'aboutissent pas de manière certaine à la non conformité et donc à la reprise. Il s'agit d'un modèle bi mode (compression ou non) avec une ressource unique non renouvelable et appartient donc à la famille des DTCTP. Les deux références qui suivent considèrent en revanche plusieurs ressources disponibles en quantité limitée (MRCPSP).

Tiwari et al. (2009) [141] développent un modèle de minimisation de la durée totale d'exécution du projet qui comprend la durée de reprise sous contraintes de capacité des ressources renouvelables. Les auteurs s'intéressent à la production de cours et de supports de formation pour des clients qui achètent des services particuliers de téléphonie. Cette production de service est gérée comme un projet constitué de plusieurs activités pouvant être effectuées par des travailleurs ayant des niveaux d'expérience hétérogènes de sorte qu'un mode d'exécution correspond à l'affectation d'un travailleur possédant un niveau particulier d'expérience. Si initialement un travailleur est affecté à la réalisation d'une activité pour laquelle il est sous qualifié (premier mode d'exécution), l'activité devra être reprise par un travailleur plus qualifié ; plusieurs modes d'exécution de cette reprise sont possibles s'il existe plusieurs travailleurs plus qualifiés que le travailleur initial. En revanche, si l'activité est initialement exécutée par le travailleur qualifié pour cela, elle ne nécessitera pas de reprise mais mobilisera ce travailleur pour une durée plus longue que la durée d'exécution de l'activité nécessitée par le travailleur sous qualifié (premier mode). Le problème consiste alors à déterminer les variables binaires exprimant le choix d'un mode d'exécution initial et d'un mode d'exécution pour la reprise si besoin de sorte à minimiser la durée totale d'exécution sachant qu'à chacun de ces modes sont associées une durée et une consommation des ressources. Les contraintes du modèle comprennent donc celles relatives au respect des capacités. La modélisation s'appuie sur celle du MRCPSP. Il est à noter que tous les travailleurs sont capables d'effectuer toutes les activités autrement dit, il ne s'agit pas d'une modélisation multi-compétences (multi skill) mais d'une modélisation avec une main d'oeuvre possédant une compétence unique mais ayant des efficacités hétérogènes qui se traduisent par des durées (productivités) et des quantités de ressources consommées différentes.

Les auteurs utilisent CPLEX pour résoudre à l'optimum leur programme linéaire en 0-1 pour plusieurs instances générées aléatoirement en modifiant le nombre d'activités (jusqu'à 40 activités) et le nombre de modes en initial et en reprise. Le modèle est aussi appliqué à un projet réel de Telecom composé de 14 activités et employant 17 travailleurs (donc 17 modes d'exécution possibles au maximum en initial).

Très récemment, Fu et Zhang (2016) [60] ont également développé une modélisation multi-mode avec ressources contraintes et reprise. Mais à la différence de Tiwari et al. (2009) qui considèrent la qualité comme binaire, les auteurs supposent qu'il existe pour chaque activité un seuil acceptable de qualité au delà duquel les reprises ne sont pas nécessaires. En outre, des ressources non renouvelables sont également incorporées. Dans ce modèle, la qualité individuelle d'une activité se mesure par un indicateur variant entre 0 et 1 traduisant le pourcentage de conformité aux spécifications. En omettant l'indice  $j$  de l'activité pour simplifier l'écriture, La qualité individuelle  $q$  d'une activité  $j$  est supposée dépendre de façon non linéaire de sa

durée d'exécution  $d$  et de son coût direct  $c$ , avec

$$q \equiv \mathcal{Q}(d, c) = q^{\max} \cdot e^{-\frac{1}{2(1-\eta^2)} \left[ \left( \frac{d-\mu_d}{\sigma_d} \right)^2 - 2\eta \left( \frac{d-\mu_d}{\sigma_d} \right) \left( \frac{c-\mu_c}{\sigma_c} \right) + \left( \frac{c-\mu_c}{\sigma_c} \right)^2 \right]}, \quad (4.2)$$

où  $q^{\max}$  est la qualité maximum possible (qualité en mode normal, en général) et les couples de paramètres  $(\mu_d, \sigma_d)$  et  $(\mu_c, \sigma_c)$  représentent respectivement moyenne et variance de la durée et du coût pour une série de valeurs associés à différents modes d'exécution qui permettent leur estimation. Le coefficient  $\eta$  permet d'ajuster la forme de la surface dans le repère  $(d, c, q)$ . Cette expression de la qualité individuelle s'inspire d'hypothèses formulées antérieurement dans les travaux de Liberatore et Pollack-Johnson (2013) [99] que nous détaillons dans la sous-section suivante relative aux modèles d'arbitrage durée, coût et qualité en temps continu (CTCTP et qualité). La forme donnée en (4.2) traduit le fait que la qualité d'une activité est croissante à la fois avec sa durée et son coût et à qualité constante, la relation d'arbitrage durée-coût est décroissante convexe car pour maintenir un même niveau de qualité tout en réduisant la durée, il faut dépenser de plus en plus d'argent pour chaque unité de temps gagnée.

Les auteurs introduisent le concept de qualité cumulée en partant de l'idée selon laquelle la qualité d'une activité dépend non seulement de sa qualité individuelle mais également de la qualité de ses prédécesseurs. Pour déterminer la qualité cumulée d'une activité, les auteurs cherchent une façon pertinente d'agrèger les qualités individuelles. Pour cela, ils introduisent le concept de lien de coordination en plus des liens de précédence sur les activités. Considérons par exemple 3 activités dans la construction d'un bâtiment. L'activité 1 consiste à poser des pieux de fondation, l'activité 2 est la pose des piliers; ces deux activités pouvant s'effectuer en parallèle, leur qualité jointe peut s'évaluer par la moyenne arithmétique de leur qualité individuelle  $1/2(q_1 + q_2)$ . Sur ces 2 structures verticales s'appuie l'activité 3 de pose d'un plancher (structure horizontale en béton armé) qui a donc les activités 1 et 2 comme prédécesseurs. La qualité de l'activité 3 dépend non seulement de sa qualité individuelle  $q_3$  (fonction du coût des matériaux et du temps passé à la pose) mais également de la qualité jointe de ses prédécesseurs. En effet, poser un plancher sur des structures qui ne sont pas suffisamment stabilisées ne peut donner un plancher solide de sorte qu'il existe donc une relation de coordination entre l'activité 3 et les activités 1 et 2, en plus de la contrainte de précédence. Une bonne qualité pour l'activité 3 se traduit comme en probabilité par les événements : "bonne qualité individuelle" et "bonne qualité jointe des prédécesseurs" ce qui s'exprime par le produit  $q_3 \times \frac{1}{2}(q_1 + q_2)$ . Ce produit permet aussi de traduire le fait que la qualité cumulée de l'activité 3 sera faible si la moyenne des qualités des activités 1 et 2 l'est également, et ce, même si  $q_3$  est élevée. Ainsi, les activités d'un projet peuvent être partitionnées en  $G$  groupes d'activités  $V_g$  (avec  $g = 1..G$ ) liées par la coordination et la précédence. Dans chaque sous-ensemble  $V_g$ , la qualité d'une activité est évaluée par sa qualité cumulée et si la qualité cumulée de la dernière activité du groupe n'atteint pas un seuil alors elle est reprise. Soit  $a_g$  la dernière activité du groupe  $V_g$  et soit  $Q(a_g)$  sa qualité cumulée. En notant  $\Gamma^{-1}(j)$  l'ensemble des prédécesseurs de  $j$  alors on a

$$Q(a_g) = \prod_{j \in V_g} \left( q_j \cdot \frac{1}{|\Gamma^{-1}(j)|} \sum_{i \in \Gamma^{-1}(j)} q_i \right). \quad (4.3)$$

L'objectif est de minimiser le coût total de la qualité constitué d'un coût de prévention et d'un coût de défaillance (reprise). Le coût de prévention est défini comme la différence entre le coût direct associés aux modes sélectionnés et le coût direct minimum (en mode normal) de toutes les activités. Le coût de reprise est la somme des coûts des ressources renouvelables mobilisées pendant la durée de la reprise et les ressources non renouvelables consommées par la reprise. Les auteurs formule le problème comme un programme non linéaire en 0-1 dont la fonction objectif est la minimisation du coût de qualité sous contraintes de capacité de ressources renouvelables et non renouvelables, d'un délai d'exécution et sous contrainte que la reprise s'effectue si  $Q(a_g)$  est inférieure à un seuil fixé. Les variables de décisions sont des variables binaires avec  $x_{jmt}$  qui vaut 1 si l'activité  $j$  est exécutée en mode  $m$  à l'instant  $t$  et 0 sinon, et  $z_{jt}$  valant 1 si une reprise de l'activité  $j$  est déclenchée à l'instant  $t$ .

Le problème est résolu de façon approchée par une métaheuristique (Shuffled Frog-Leaping Algorithm) qui combine les principes de l'algorithme génétique et de l'optimisation par essais particuliers (PSO). La métaheuristique est appliquée à un projet de construction d'un viaduc ferroviaire comprenant 14 activités.

### 4.1.2 Modélisations en temps continu (CTCTP et Qualité)

Les trois modèles que nous présentons dans cette sous-section cherchent à déterminer les durées optimales des activités considérées comme des variables continues et ce, pour minimiser la durée totale ou le coût total ou maximiser la qualité globale du projet. Une seule ressource non renouvelable est considérée et les relations entre coût et durée d'une part et qualité et durée d'autre part sont supposées continues. En conséquence ces modèles traitent le CTCTP étudié au chapitre 3 (sous-section 3.2.1) auquel la qualité est ajoutée.

A notre connaissance, Babu et Suresh (1996) [14] sont les premiers auteurs à modéliser l'impact de l'accélération sur la qualité. Comme dans la sous-section 3.2.1.1 du chapitre précédent, le coût  $c_j$  est une fonction affine décroissante de la durée  $d_j$  avec  $c_j = \mathcal{C}_j(d_j)$  et bornée par les points  $(d_j^{\min}, c_j^{\max})$  et  $(d_j^{\max}, c_j^{\min})$  qui représentent respectivement les couples durée et coût en mode compressé et en mode normal. Ces couples de paramètres font généralement l'objet d'une estimation par des experts. Ainsi,  $\mathcal{C}_j(d_j) = a_j + b_j \cdot d_j$  avec une pente  $b_j = (c_j^{\max} - c_j^{\min}) / (d_j^{\max} - d_j^{\min})$  et  $a_j = c_j^{\min} - b_j \cdot d_j^{\max}$ .

De plus, la qualité  $q_j$  est supposée être une fonction affine croissante de la durée  $d_j$  avec  $q_j = \mathcal{Q}_j(d_j)$  et bornée par les points de coordonnées  $(d_j^{\min}, q_j^{\min})$  et  $(d_j^{\max}, q_j^{\max})$  qui désignent respectivement les couples durée - qualité en mode compressé et en mode normal. La qualité en mode compressé généralement estimée par des experts est strictement inférieure à 1 et la qualité en mode normale est souvent supposée

égale à 1. Ainsi on a  $\mathcal{Q}_j(d_j) = a'_j + b'_j \cdot d_j$  avec une pente  $b'_j = (q_j^{\max} - q_j^{\min}) / (d_j^{\max} - d_j^{\min})$  et  $a'_j = q_j^{\min} - b'_j \cdot d_j^{\min}$ .

On rappelle que  $t_j$  désigne la date de début de l'activité  $j$ , avec  $j = 0..n + 1$  où 0 et  $n + 1$  sont respectivement les activités fictives de début et de fin du projet. Les auteurs adoptent une approche scalaire par la méthode  $\epsilon$ -contrainte où l'un des sous-objectifs  $f_k$  est optimisé tandis que les autres s'expriment sous la forme de contraintes du type  $f_j(x) \leq \epsilon_j$  pour tout  $j \neq k$ . Les trois sous-objectifs s'écrivent comme suit.

$$f_1 : t_{n+1} ; \epsilon_1 = D_{\max} \quad (4.4)$$

$$f_2 : \sum_{j=1..n} \mathcal{C}_j(d_j) ; \epsilon_2 = B_{\max} \quad (4.5)$$

$$f_3 : \sum_{j=1..n} \mathcal{Q}_j(d_j) ; \epsilon_3 = Q_{\min} \quad (4.6)$$

Ainsi, un premier programme consiste à minimiser la durée totale  $f_1$  sous les contraintes  $f_2 \leq B_{\max}$  et  $f_3 \geq Q_{\min}$  ainsi que les contraintes de précédence. Dans le deuxième programme on minimise le coût total  $f_2$  sous les contraintes  $f_1 \leq D_{\max}$  et  $f_3 \geq Q_{\min}$ . Enfin, le troisième programme maximise la qualité globale  $f_3$  sous les contraintes  $f_1 \leq D_{\max}$  et  $f_2 \leq B_{\max}$ .

Ces auteurs appliquent les trois modèles à un projet unique incluant 14 activités et inspiré de Hillier et Lieberman (1990) [74]. Les résultats du modèle de minimisation de la durée montrent que pour des seuils faibles de budget à ne pas dépasser, la durée d'exécution du projet est plus élevée et n'est affectée que marginalement par l'exigence de qualité (seuil minimum de qualité). Si le budget augmente, la durée diminue mais cette diminution est fortement tributaire de l'exigence de qualité ; des seuils plus élevés de qualité globale minimum à atteindre impliquent un accroissement substantiel de la durée. Dans la résolution du modèle de minimisation des coûts, une réduction du délai à ne pas dépasser implique une très forte augmentation des coûts et ne peut répondre qu'à des exigences de qualité modestes. Au fur et à mesure que le délai est prolongé, le coût direct du projet est réduit et peut répondre à des exigences de qualité plus strictes. Enfin, l'application du modèle de maximisation de la qualité montre que l'allocation de budgets plus élevés et l'autorisation de délais d'exécution plus longs permettent d'atteindre des niveaux de qualité (moyenne) plus élevés.

Les auteurs considèrent également deux autres agrégations des qualités individuelles des activités pour définir la qualité globale du projet : la moyenne géométrique et le minimum des qualités individuelles. Les résultats sont similaires pour les trois hypothèses d'agrégation des qualités individuelles.

Khang et Myint (1999) [88] appliquent la modélisation proposée par Babu et Suresh (1996) [14] à la construction d'une tour de préchauffage d'une usine de ciment. Le projet comprend 52 activités pour lesquelles les paramètres de durée, coût

et qualité sont aussi estimés par les gestionnaires du projet en mode normal et en mode compressé. Le mode compressé résulte du recours au maximum d'heures supplémentaires et comme celles-ci permettent de multiplier la main d'oeuvre par 1,5 (on passe de 8h à 12h par jour), les durées des activités peuvent être compressées au  $2/3$ . Les auteurs font donc implicitement l'hypothèse d'une productivité marginale constante puisque la relation est linéaire (et décroissante) entre la durée et la quantité de ressource. De plus, les heures supplémentaires coûtant le double des heures normales, on a également une relation linéaire (croissante) entre coût et ressource. Il en résulte que la relation entre le coût et la durée est également linéaire. Les auteurs adoptent donc l'hypothèse de linéarité entre le coût et la durée supposée dans le modèle de Babu et Suresh (1996) [14]. En revanche, l'hypothèse d'une relation décroissante entre la qualité et la durée fut d'abord difficile à admettre par les gestionnaires du projet et les ingénieurs qui étaient convaincus que si la qualité est définie objectivement comme la conformité aux spécifications techniques alors l'utilisation d'heures supplémentaires n'affecterait que marginalement le respect de ces spécifications. En effet, pour eux, la baisse de qualité ne serait que de l'ordre de 2 à 3% et donc négligeable même si l'on utilise le maximum d'heures supplémentaires. Pour chacune des 52 activités, les experts ont finalement fourni des estimations de la qualité compressée associée à une utilisation maximum des heures supplémentaires, avec des baisses de qualité variant de 10 à 30% selon les activités (obtenant ainsi les points  $(c_j^{\max}, q_j^{\min})$ ). Puis les auteurs ont adopté l'hypothèse de linéarité entre la qualité et la durée où le segment de droite croissante est donc borné par les points  $(d_j^{\min}, q_j^{\min})$  et  $(d_j^{\max}, q_j^{\max})$  comme dans Babu et Suresh (1996) [14]. Enfin, les auteurs retiennent la moyenne arithmétique des qualités individuelles comme mesure de la qualité globale du projet. Les trois problèmes d'optimisation sont résolus avec différentes valeurs cible pour le coût, la qualité et la durée et les auteurs parviennent à des conclusions similaires à celles de Babu et Suresh (1996) [14]. En particulier, pour chaque seuil de qualité globale du projet fixé à  $Q_{\min}$ , il existe un seuil de budget au delà duquel un accroissement de budget ne permet pas de réduction significative de la durée. Il existe également un seuil de durée totale en dessous duquel il serait extrêmement coûteux d'obtenir une réduction additionnelle de la durée du projet.

Les auteurs discutent la pertinence des trois modèles et remettent en cause celle du modèle de maximisation de la qualité sous contrainte car il s'agit du modèle le moins intéressant d'un point de vue pragmatique pour les gestionnaires et les ingénieurs. En effet, l'objection principale à l'encontre de ce modèle réside dans les mesures de la qualité qui sont souvent trop subjectives et imprécises pour être raisonnablement considérées dans une fonction objectif. En outre, l'hypothèse de linéarité entre la qualité et la durée reste problématique. De plus, le modèle suppose une évaluation de la qualité en mode compressé, et cette estimation est particulièrement difficile et longue à produire même en impliquant plusieurs experts. Enfin, les auteurs notent que l'usage d'une moyenne arithmétique des qualités individuelles pour mesurer la qualité globale reste critiquable, en particulier pour des problèmes d'agrégation évidents (voir discussion chapitre 2, section 2.3.2, page 53).

Liberatore et Pollack-Johnson (2013) [99] proposent de modéliser la qualité individuelle  $q_j$  d'une activité  $j$  comme une fonction non linéaire de sa durée d'exécution  $d_j$  et de son coût  $c_j$ . Ils partent des hypothèses raisonnables selon lesquelles (i) la

qualité d'une activité est fonction croissante de son coût, c'est-à-dire que dépenser d'avantage d'argent pour l'exécution d'une activité va accroître sa qualité; (ii) à coût constant, la qualité est une fonction croissante de la durée car plus on passe de temps sur une activité, plus on augmente sa qualité. En adoptant ces deux hypothèses, si l'on représente les triplets  $(d_j, c_j, q_j)$  dans un espace à trois dimensions, on doit s'attendre à ce que ce graphe place la plus faible qualité dans le coin du domaine correspondant au coût et à la durée les plus faibles et la qualité la plus haute dans le coin opposé, comme le montre la figure 4.1.

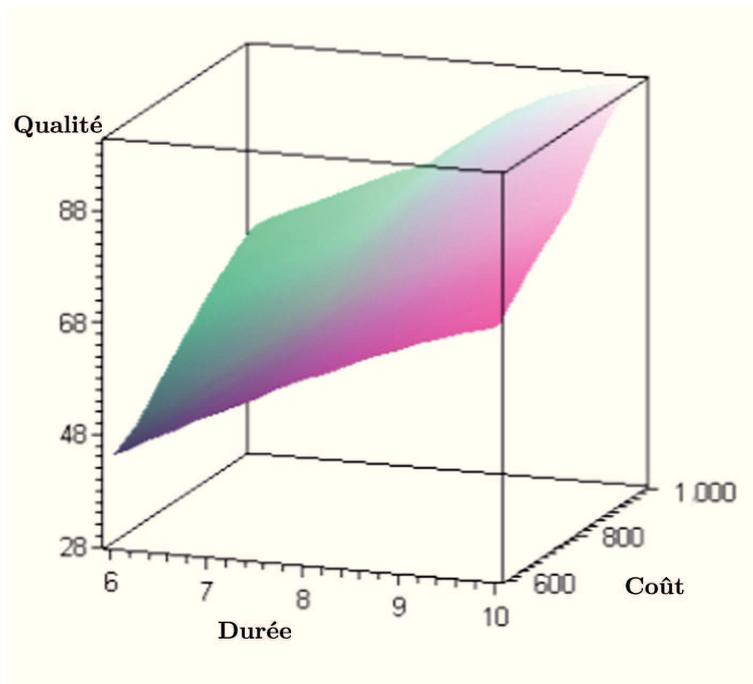


FIGURE 4.1 – La qualité d'une activité comme une fonction continue de sa durée et de son coût.

A qualité fixée, on souhaite que la relation d'arbitrage durée-coût soit décroissante et convexe car pour maintenir un même niveau de qualité tout en réduisant la durée, il faut dépenser de plus en plus d'argent pour chaque unité de temps gagnée. Ce type de relation a en effet été mis en évidence par Babu et Suresh (1996) [14] dans leurs résultats d'expérience. Il est à noter que l'on peut avoir des relations linéaires entre qualité et durée individuelles d'une part et coût et durée individuelles d'autre part, mais lorsqu'on considère l'arbitrage entre ces trois dimensions pour le projet dans sa totalité, c'est différent. On constate en effet qu'à niveaux fixés de qualité globale pour le projet, les graphes de résultats montrent bien que la durée totale d'exécution du projet est décroissante convexe avec le coût total du projet. Autrement dit, c'est la relation d'arbitrage durée totale - coût total du projet à qualité globale fixée qui n'est pas linéaire, même si le coût individuel d'une activité est supposé être une fonction linéaire décroissante de la durée.

La forme mathématique qui a le type d'aspect recherché pour la qualité individuelle est la distribution normale bivariée en probabilité. Les auteurs utilisent ce

type de forme pour exprimer la qualité de toute activité comme une fonction de sa durée et de son coût. En omettant l'indice  $j$  de l'activité pour simplifier l'écriture, on a

$$q \equiv \mathcal{Q}(d, c) = q^{\max} \cdot e^{-\left[\left(\frac{d-\mu_d}{\sigma_d}\right)^2 + \left(\frac{c-\mu_c}{\sigma_c}\right)^2\right]},$$

où  $q^{\max}$  est la qualité maximum possible (généralement obtenue dans le mode normal) et  $(\mu_d, \sigma_d, \mu_c, \sigma_c)$  sont les paramètres de la distribution qu'il convient d'estimer. Les auteurs illustrent la méthode d'estimation sur des données de coût, durée et qualité associées à un projet de traduction linguistique de documents. Pour un traducteur donné et un projet de traduction, ils obtiennent 6 durées différentes en jours, de 2 à 4 jours auxquelles sont associés des coût et des niveaux de qualité exprimant des points de faiblesse ou erreurs dans la traduction. Les quatre paramètres sont estimés par la méthode des moindres carrés non linéaire. Les auteurs utilisent ensuite cette relation pour calculer les qualités individuelles dans un modèle de maximisation non linéaire de la qualité globale d'un projet sous contraintes de précédence, de budget et de délai. La qualité globale du projet est définie par le minimum des qualités individuelles. Les variables du modèles sont les qualités, coûts et durées individuelles des activités.

### 4.1.3 Modélisations multi-mode avec une seule ressource non renouvelable (DTCTP et Qualité)

A notre connaissance Tareghian et Taheri (2006) [136] sont les premiers auteurs à intégrer la qualité dans un modèle d'arbitrage coût - durée en temps discret pour obtenir le problème DTCQTP (Discret Time Cost Quality Trade-off Problem) où une seule ressource non renouvelable, l'argent, est considérée. A chaque activité  $j$  exécutée dans le mode  $m$  ( $m = 1..M_j$ ) correspond un triplet durée, coût, qualité  $(d_{jm}, c_{jm}, q_{jm})$  estimé par des experts. Les auteurs reprennent l'hypothèse d'un coût décroissant avec la durée d'exécution de l'activité, et d'une qualité croissante avec cette durée.

Le problème DTCQTP se formule comme le DTCTP que nous avons étudié à la sous-section 3.2.2 du chapitre 3 (page 61) où l'objectif de maximisation de la qualité s'écrit

$$\max \frac{1}{n} \sum_{j=1}^n \sum_{m=1}^{M_j} q_{jm} \sum_{t=EF_j}^{LF_j} x_{jmt}, \quad (4.7)$$

avec  $x_{jmt}$ , une variable binaire qui vaut 1 si l'activité  $j$  exécutée dans le mode  $m$  s'achève à la fin de la période  $t$ . La somme  $\sum_{t=EF_j..LF_j} x_{jmt}$  représente finalement

le choix d'un mode d'exécution particulier pour l'activité  $j$ . Les auteurs reprennent donc la moyenne arithmétique des qualités individuelles comme mesure de la qualité globale d'un projet et adoptent l'approche scalaire par la méthode  $\epsilon$ -contrainte comme Babu et Suresh (1996) [14] le font pour le cas continu (voir équations (4.4), (4.5) et (4.6)).

Les problèmes sont formulés comme des programmes d'optimisation linéaire en nombre entier où les variables de décisions sont binaires pour la sélection du mode

et entières pour les dates de fin des activités. Ces programmes d'optimisation sont appliqués et résolus à l'optimum pour un projet de 45 activités obtenu avec le générateur d'instances DAGEN développé par Agrawal et al. (1996) [6]. Le nombre de modes par activité est tiré aléatoirement entre 2 et 10. Les paramètres de durée, coût et qualité pour chaque activité dans chaque mode sont générés aléatoirement en respectant les hypothèses de croissance de la qualité avec la durée et de décroissance du coût avec la durée. Les résultats des simulations montrent qu'une qualité globale élevée est évidemment incompatible avec une petite durée d'exécution totale du projet. Lorsque le budget est faible, les exigences de qualité n'affectent pas la durée totale mais en augmentant le budget, il est possible de réduire la durée pour tous les seuils de qualité à atteindre. Les résultats obtenus sont sensiblement les mêmes en utilisant pour calculer la qualité globale du projet non pas la moyenne arithmétique mais la moyenne géométrique car les valeurs des qualités individuelles ne sont pas très dispersées.

Plus tard, ces mêmes auteurs proposent une métaheuristique pour ce même DTC-QTP où la moyenne géométrique est adoptée pour calculer la qualité globale (Tareghian et Taheri, 2007 [137]). Ils appliquent la méthode de recherche par dispersion électromagnétique (Electromagnetic Scatter Search) pour résoudre ce problème. Une population initiale de solutions réalisables est générée en utilisant une mémoire de fréquence pour bien échantillonner la région réalisable. Un certain nombre de ces solutions sont ensuite sélectionnées et améliorées localement. Les solutions améliorées sont ensuite combinées pour générer de nouvelles solutions. Le processus de combinaison utilise des mécanismes d'attraction-répulsion provenant de la théorie de l'électromagnétisme. La recherche s'arrête lorsqu'aucune amélioration significative de l'ensemble des solutions n'est observée. L'approche est testée sur des problèmes générés par les auteurs de façon aléatoire avec 500, 1000, 5000, 10000 et 20000 activités et 2 à 15 modes d'exécution par activité.

Zhang et al. (2014) [164] utilisent la somme pondérée comme autre méthode pour agréger les trois sous-objectifs et proposent une résolution heuristique du problème. Les auteurs partent du constat qu'une durée plus longue ne conduit pas nécessairement à une meilleure qualité. Au delà d'une certaine durée, la qualité peut même décroître ce qui est le cas pour certaines activités dans la construction comme le cimentage dont la maniabilité s'altère assez rapidement. Les auteurs expriment alors la qualité individuelle comme une fonction quadratique de sa durée

$$q_j = \mathcal{Q}_j(d_j) = a_j d_j^2 + b_j d_j + e_j, \quad (4.8)$$

où les paramètres  $a_j, b_j$  et  $e_j$  sont estimés par les ingénieurs du projet. Cette définition a été reprise ultérieurement par Tran et al. (2015) [142]. La qualité globale du projet est mesurée par la somme des qualités individuelles. Les coûts du projet sont les coûts directs et les coûts de retard. L'objectif est de minimiser une somme pondérée des variations de coût, durée et qualité par rapport à leurs valeurs contractuelles. Les auteurs supposent un poids identique pour chaque sous-objectif et résolvent le problème à l'aide d'un algorithme basé sur l'optimisation par essais particuliers (Particle Swarm Optimization, PSO). L'approche est appliquée à un

projet de construction d'un bâtiment de trois étages comprenant 19 activités.

Mohammadipour et Sadjadi (2016) [106] considèrent un DTCQTP où les modes d'exécution de chaque activité correspondent à des réductions possibles de leur durée d'une ou plusieurs périodes de sorte que les variables de décision sont des variables binaires  $y_{jt}$  valant 1 si l'activité  $j$  est réduite de  $t$  périodes et zéro sinon. Pour chaque activité, une réduction de la durée d'une unité s'accompagne d'une augmentation du coût qui est estimée. Sont estimés également les baisses de qualité et un facteur de risque d'impact global de l'accélération sur le projet, pour chaque réduction possible de durées. Ce facteur de risque traduit la possibilité que l'accélération d'une activité donnée impacte le coût total, la durée totale et la qualité des livrables finaux notamment parce que des besoins de reprise peuvent être constatés à la fin. Les auteurs considèrent alors trois sous-objectifs à minimiser : l'augmentation du coût total  $f_1$  ; la baisse totale de qualité  $f_2$  et la somme des risques d'impact sur le projet  $f_3$ . Chacun de ces objectifs est fonction des variables  $y_{jt}$  exprimant le nombre de périodes de réduction de la durée. Pour ramener le problème à un objectif unique, les auteurs utilisent la méthode du but à atteindre (Goal Attainment) où le décideur choisit des valeurs cible  $\Delta_k$  à atteindre pour chacun des  $k$  sous-objectifs et les poids  $w_k$  qu'il leur accorde. La solution optimale est trouvée en minimisant une variable continue  $x$  telle que  $\Delta_k + w_k \cdot x \geq f_k$ . Ce modèle nécessite donc de connaître les préférences du décideur. Le modèle est appliqué à un projet de construction d'un bâtiment de 2 étages comprenant 18 activités réparties en 3 phases : charpente, finitions, équipements, etc. Les auteurs génèrent et résolvent à l'optimum 50 problèmes obtenus en faisant varier les poids  $w_k$  et les valeurs cible  $\Delta_k$  (augmentation du coût, baisse de la qualité et somme des risques).

## 4.2 Optimisation multi objectif par des méthodes non agrégées fondées sur Pareto

Cette section étudie les approches non agrégées et Pareto pour le problème d'arbitrage coût, durée et qualité, contrairement aux approches agrégées que nous avons détaillées dans la section précédente. Le principe de fonctionnement de ces approches non agrégées et Pareto a été décrit dans la sous-section 3.3.5 du chapitre 3. Ces approches sont étudiées de façon plus approfondie dans la sous-section A.2.4 de l'annexe technique.

Une solution au problème d'arbitrage durée, coût et qualité (TCQTP) est une double liste, la première donnant les dates de début (ou de fin) des activités et la seconde, les modes d'exécution des activités. Pour toute solution, il est possible de calculer la durée totale d'exécution du projet, son coût total et sa qualité globale. Ces valeurs des fonctions objectif pour toutes les solutions explorées peuvent se représenter dans un espace à trois dimensions comme le montre la figure 4.2, où les solutions non dominées par aucune autre (éventuellement Pareto optimales) ont été trouvées par un algorithme multi-objectif particulier. La difficulté de ces approches de résolution réside dans la possibilité que la frontière de Pareto (une surface) ne soit pas très bien représentée, comme c'est le cas à la figure 4.2 où la surface est plutôt irrégulière avec des amas de solutions dans certaines régions (comme c'est le

cas pour une durée totale faible, un coût total élevé et une faible qualité globale) et peu de solutions dans d'autres régions.

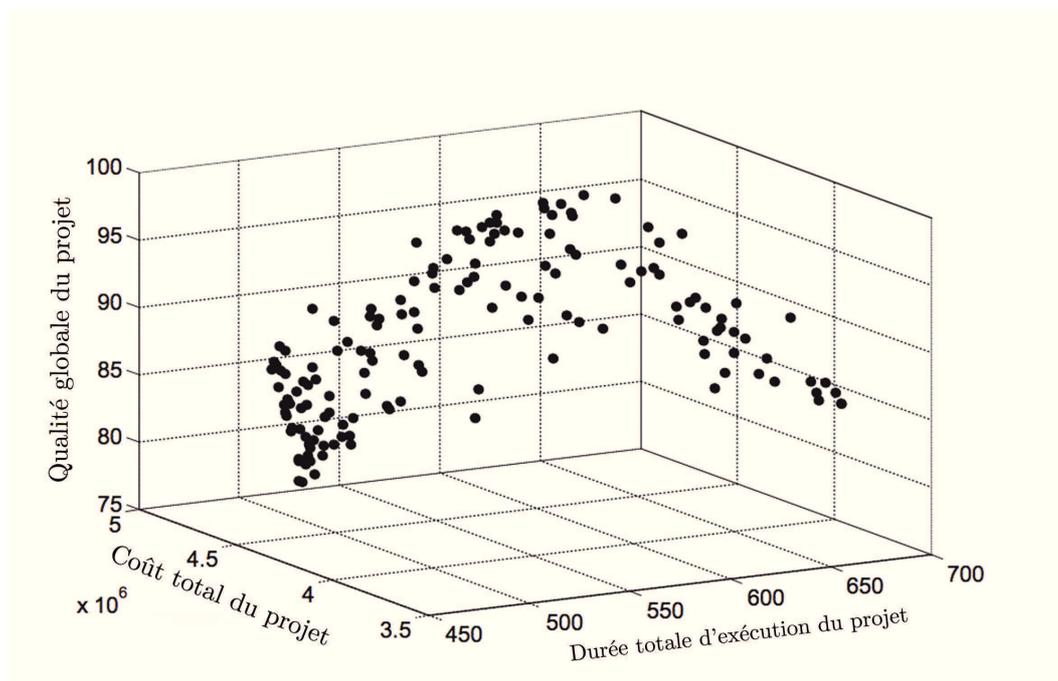


FIGURE 4.2 – Surface de Pareto associée aux solutions du TCQTP explorées par un algorithme multi-objectif particulier.

A l'exception d'une seule contribution qui considère plusieurs ressources renouvelables dans un modèle de type MRCPSPP (Afruzi et al., 2014 [2]), le problème d'arbitrage coût, durée et qualité traité en optimisation multi objectif n'intègre qu'une seule ressource non renouvelable de sorte qu'il s'agit du DTCTQP où l'hypothèse de non préemption est quasiment toujours adoptée, sauf dans la contribution de Tavana et al. (2014) qui sort par conséquent du cadre de cette thèse.

Les contributions qui adoptent des approches multi-objectif non agrégées supposent généralement que la qualité individuelle est estimée par des experts dans chacun des modes, c'est-à-dire que la qualité individuelle n'est pas explicitement exprimée comme une fonction de la durée. La qualité globale du projet est définie le plus souvent comme une moyenne arithmétique simple ou pondérée des qualités individuelles à l'exception de deux contributions qui combinent linéairement la moyenne arithmétique et le minimum des qualités individuelles (Mungle et al., 2013 [109] et Monghasemi et al., 2015 [108]).

#### 4.2.1 Algorithmes génétiques multi-objectif

Ce type d'algorithmes est détaillé dans l'annexe technique, à la sous-section A.2.4.2 (voir page 206).

El Rayes et Kandil (2005) [49] considèrent le DTCTQP en s'appuyant sur un projet réel de construction d'une autoroute. Chaque mode d'exécution d'une acti-

tivité résulte de la combinaison de différentes options d'utilisation de ressources mais sans contraintes sur la capacité de ces ressources de sorte que les auteurs associent finalement un coût à un mode d'exécution. A chaque activité  $j$  est associée une liste d'indicateurs de qualité et chaque indicateur  $l$  prend une valeur  $q_{ljm}$  lorsque l'activité  $j$  est exécutée dans le mode  $m$ . Chaque indicateur  $q_{ljm}$  possède un poids  $w_{lj}$  qui représente sa contribution à la qualité individuelle de l'activité. Prenons l'exemple d'une activité de dallage en béton exécutée selon un mode où l'on va utiliser un matériau et une machine particulière et un certain volume d'heures supplémentaires. La qualité du dallage est évaluée à l'aide de plusieurs critères comme sa résistance à la compression ( $l = 1$ ) et sa qualité de roulement ( $l = 2$ ). Ces deux critères sont pondérés par l'importance de chacun dans la qualité du dallage.

Ces indicateurs (ou critères) de qualité doivent être définis de façon à pouvoir les mesurer objectivement et facilement. Par exemple la qualité de roulement se mesure par l'IRI (Indice de Rugosité International) qui donne le nombre de mètres de débâtements (mouvement vertical de la roue par rapport à la caisse) que la suspension d'une voiture subit lorsqu'elle parcourt un kilomètre à une vitesse de 80 km/h. Pour chaque option d'utilisation des ressources (mode), les résultats de ces tests de qualité peuvent être facilement obtenus et stockés de la même façon que les durées et coûts associés à des projets déjà réalisés. Ces données sont collectées et enregistrées dans des bases gouvernementales et donc accessibles à l'analyse et au traitement statistique de façon relativement simple et immédiate.

Les auteurs supposent que la qualité globale de l'autoroute est définie comme la somme pondérée des qualités individuelles des activités du projet, le poids d'une activité représentant l'importance et la contribution de cette dernière à la qualité globale du projet. La qualité globale s'écrit ainsi

$$\sum_{j=1..n} w_j \sum_{l=1..L} w_{lj} \cdot q_{ljm} \quad (4.9)$$

Le coût du projet est une fonction des coûts de matériels  $H_{jm}$  nécessaires à l'exécution de l'activité  $j$  dans le mode  $m$  et du coût journalier  $c_{jm}$  d'utilisation de la main d'oeuvre multiplié par la durée  $d_{jm}$ . A ces coûts s'ajoutent les coûts de sous traitance  $O_{jm}$ . Le coût total s'écrit

$$\sum_{j=1..n} (H_{jm} + c_{jm} \cdot d_{jm} + O_{jm}) \quad (4.10)$$

La durée du projet est la longueur du chemin critique correspondant à une sélection particulière de modes pour les activités. La recherche de l'ensemble des solutions Pareto optimales s'effectue à l'aide d'un NSGAI (Non Sorted Genetic Algorithm II) que les auteurs appliquent à un projet de construction d'une autoroute comprenant 18 activités inspiré de Feng et al. (1997) [56]. Les résultats montrent que le NSGAI permet d'obtenir les solutions optimales au seul problème d'arbitrage durée - coût ; solutions qui avaient été trouvées initialement par Feng et al. et al. (1997) [56].

Mungle et al. (2013) [109] s'intéressent à la problématique des choix de sous traitance dans les projets de construction d'autoroutes. Comme El Rayes et Kandil

(2005) [49] le suggèrent, ces auteurs exploitent les bases de données d'indicateurs de qualité, durée et coût collectés pour des autoroutes déjà construites. D'un point de vue méthodologique, ils proposent d'appliquer l'Analyse Hiérarchique pour classer ces indicateurs de qualité par ordre d'importance et pour leur donner un poids afin d'aboutir à une expression de la qualité individuelle comme une somme pondérée de ces critères (indicateurs  $l = 1..L$ ). La qualité globale  $Q$  du projet est exprimée comme une somme pondérée de la qualité moyenne et de la qualité minimum (minimum des qualités individuelles), c'est-à-dire  $Q = \alpha \cdot \min_{j=1..n} \{q_j\} + (1-\alpha) \cdot \frac{1}{n} \sum_{j=1..n} q_j$ . Maximiser la qualité globale  $Q$  avec une grande valeur de  $\alpha$  permet de s'assurer qu'aucune activité n'atteindra un niveau individuel de qualité trop bas. Avec une valeur faible de  $\alpha$ , l'accent est mis sur la recherche d'une plus grande qualité moyenne.

Les auteurs modélisent le DTCQTP comme un programme linéaire où les activités peuvent être exécutées par des sous-traitants différents, chaque sous-traitant propose une certaine qualité, à un certain coût et dans un certain délai. Les sous-traitants s'apparentent à des modes d'exécution pour toute activité. L'objectif du modèle est donc de sélectionner les sous-traitants pour les activités de sorte à minimiser la durée d'exécution et le coût du projet tout en maximisant la qualité globale du projet. Le coût du projet est la somme des coûts directs de sous traitance, des coûts indirects, des pénalités si le projet est en retard et des primes si le projet est en avance.

La recherche des solutions Pareto optimales se fait au moyen d'un algorithme génétique multi-objectif avec une stratégie élitiste pour réduire la liste archivée des meilleures solutions non dominées (voir annexe technique, sous-section A.2.4.4) qui repose sur un partitionnement de cette liste en plusieurs clusters de solutions proches. La proximité de deux solutions se calcule au moyen d'une distance Euclidienne entre les valeurs de leurs différents sous-objectifs (voir équation (A.15) dans l'annexe technique). Puis on retient dans la liste archivée une seule solution par cluster. Au lieu de retenir la solution qui constitue le centre de gravité du cluster (solution possédant la plus petite distance moyenne avec toutes les autres solutions du même cluster), les auteurs proposent de sélectionner dans chaque cluster  $c_i$  la solution  $x$  ayant la valeur d'appartenance (membership value)  $\mu(x)$  la plus élevée, avec

$$\mu(x) = \frac{\sum_{k=1,2,3} \mu_k(x)}{\sum_{y \in c_i} \sum_{k=1,2,3} \mu_k(y)} \quad (4.11)$$

et la fonction d'appartenance

$$\mu_k(x) = \left\{ \begin{array}{ll} 1 & \text{si } f_k(x) \leq f_k^{\min} \\ \frac{f_k^{\max} - f_k(x)}{f_k^{\max} - f_k^{\min}} & \text{pour } f_k^{\min} < f_k(x) < f_k^{\max} \\ 0 & \text{si } f_k(x) \geq f_k^{\max} \end{array} \right\} \quad (4.12)$$

Pour cette raison, l'approche de recherche des solutions Pareto optimales est appelée Fuzzy Clustering Genetic Algorithm (FCGA). Les auteurs traitent un problème de construction d'une autoroute à 18 activités et montrent la supériorité de

leur approche en termes de convergence vers la frontière de Pareto par rapport à d'autres algorithmes utilisant le centre de gravité comme solution représentative d'un cluster.

Monghasemi et al. (2015) [108] relèvent l'inaptitude de l'approche de Mungle et al. (2013) [109] à guider les décideurs dans leur choix d'une solution non dominée dans un espace souvent très grand de solutions non dominées. Ils proposent alors de combiner le NSGAI avec une stratégie d'affectation de poids à chacun des sous-objectifs sans pour autant les agréger, de sorte à les hiérarchiser. Ainsi les solutions non dominées pourront être choisies en adéquation avec une structure de préférences des décideurs.

## 4.2.2 Autres algorithmes évolutionnaires multi-objectif

L'annexe technique donne le principe général de fonctionnement de ces algorithmes (voir sous-section A.2.4.5, page 212).

Afshar et al. (2007) [4] s'intéressent au problème DTCQTP en reprenant la même définition des qualités individuelles et globale que celle proposée par El Rayes et Kandil (2005) [49] (voir 4.9). Ils développent une métaheuristique multi-objectif fondée sur l'optimisation par colonies de fourmis (ACO, décrite dans l'annexe technique, sous-section A.1.4.2, page 196), où il existe une colonie par sous-objectif. Les sous-objectifs de minimisation de la durée, du coût et de maximisation de la qualité sont considérés dans cet ordre. Ainsi, les fourmis de la première colonie vont rechercher une solution qui minimise la durée. Ces solutions sont transférées dans la deuxième colonie dont l'objectif est de minimiser le coût ; elles y sont donc évaluées en termes de coût et le taux global de phéromone de cette colonie est actualisé. Les nouvelles solutions obtenues sur la base de ce taux de phéromone sont transférées à leur tour dans la troisième colonie et sont donc évaluées en termes de qualité, avec une actualisation du taux global de phéromone dans cette colonie. Ce processus itératif de transfert successif des solutions d'une colonie vers une autre se poursuit jusqu'à atteindre un certain nombre d'itérations que l'on appelle un cycle. A la fin d'un cycle, les solutions évaluées en termes de qualité dans la troisième colonie sont aussi évaluées en termes de durée et de coût et les solutions non dominées parmi elles sont enregistrées dans une archive. Le taux global de phéromone est réinitialisé à sa valeur de départ au début du cycle suivant. L'algorithme s'arrête au bout d'un nombre prédéfini de cycles. L'algorithme est appliqué sur le même projet de construction d'un bâtiment que celui développé par El Rayes et Kandil (2005) [49] mais n'est pas comparé au NSGAI de ces auteurs.

Afruzi et al. (2014) [2] sont les seuls auteurs qui à notre connaissance introduisent la qualité dans une modélisation avec contraintes de ressources (MRCPSP) où les modes d'exécution peuvent ne pas être indépendants c'est-à-dire que l'utilisation d'un mode particulier pour une activité implique une restriction de choix pour les modes des activités qui lui succèdent. Prenons l'exemple de deux activités principales d'un projet de fabrication d'une tour de distillation : l'assemblage de la tour (activité A) et le détensionnement de ses articulations (activité B). Supposons qu'il existe deux modes d'exécution pour chaque activité de ce projet. Pour l'activité A, l'as-

semblage peut se faire horizontalement (mode A1) ou transversalement (mode A2). L'activité B de détensionnement peut se faire avec un four à chaleur (mode B1) ou par résistance électrique (mode B2). Lorsque l'assemblage (activité A) se fait à l'horizontal (mode A1), le détensionnement (activité B) se fait obligatoirement avec un four à chaleur (mode B1), et lorsque l'assemblage est transversal (mode A2 pour l'activité A), le détensionnement se fait habituellement par résistance électrique (mode B2). Ainsi un mode d'exécution particulier fait appel à une technique spécifique et consomme des quantités données de ressources renouvelables. De plus, chaque mode peut être compressé par adjonction de ressources renouvelables mais on note qu'il n'y a pas de lien entre les consommations de ressources dans le mode normal et dans le mode compressé. Au mode  $m$  normal correspond une durée maximum  $d_{jm}^{\max}$  et une qualité maximum  $q_{jm}^{\max}$  avec un vecteur de consommation maximum de ressources  $(r_{jmk}^{\max})_{k=1..K}$  auxquelles sont associés les coûts unitaires  $(c_{jmk}^{\min})_{k=1..K}$ . Ces paramètres résultent d'une estimation tant dans le mode normal que dans le mode compressé. La qualité globale s'exprime comme une somme pondérée des qualités individuelles avec des poids qui doivent être également estimés. Le coût total est la somme des coûts directs des ressources renouvelables consommées dans chaque mode et chaque option sélectionnés pour chaque activité. Les variables de décision sont des variables binaires avec  $x_{jmt}$  la variable binaire qui vaut 1 si l'activité  $j$  est exécuté dans le mode  $m$  et s'achève à la fin de la période  $t$  et  $y_j$  qui vaut 1 si l'exécution est accélérée. Le choix du mode accéléré pour l'activité  $j$  avec une date de fin en  $t$  se traduit par  $x_{jmt} \cdot y_j = 1$ . Leur formulation du problème s'appuie sur celle que ces mêmes auteurs ont développé pour le MRC-DTCTP et étudié au chapitre 3 (Afruzi et al., 2013 [3]). Les objectifs de maximisation de la qualité, de minimisation de la durée et du coût sont donc formulés non linéairement.

Le problème est résolu à l'aide de l'algorithme MOICA (Multi-Objective Imperialist Competitive Algorithm), un algorithme évolutionnaire multi-objectif qui repose sur l'algorithme ICA (Imperialist Competitive Algorithm) décrit dans l'annexe technique (voir sous-section A.1.4.5, page 199). Chaque solution est codée par trois vecteurs : l'ordre dans lequel les activités sont ordonnancées, les modes sélectionnés et l'option de compression. Dans chaque empire, les colonies initiales ainsi que les populations obtenues par application de plusieurs opérateurs sur ces colonies initiales sont fusionnées et les solutions non dominées sont archivées. Les meilleures colonies sont sélectionnées en fonction des techniques de classement Pareto et d'estimation de la densité ; les autres colonies sont supprimées. L'algorithme est appliqué à 25 instances de PSPLIB avec un nombre d'activités entre 13 et 31 et un nombre de ressources compris entre 2 et 4. Les résultats montrent que l'algorithme MOICA permet d'obtenir des solutions non dominées meilleures en termes des trois critères (durée, coût, qualité) que celles obtenues avec le NSGAI auquel il est notamment comparé, mais pour un temps de calcul plus important.

Tran et al. (2015) [142] reprennent la définition de la qualité individuelle initialement proposée par Zhang et al. (2014) [164] comme une fonction quadratique de la durée (voir équation (4.8)). La qualité globale est une moyenne arithmétique des qualités individuelles. Le coût total comprend les coûts directs, indirects et les pénalités de retard. Mais contrairement à Zhang et al. (2014) qui adoptent une approche de résolution mono-objectif, le DTCQTP est cette fois ci résolu à l'aide

d'un algorithme multi-objectif fondé sur la méthode de colonies d'abeilles artificielles (Artificial Bee Colony, ABC, voir annexe technique, sous-section A.1.4.4, page 198). Cet algorithme multi-objectif reprend dans ses grandes lignes la structure d'un algorithme évolutionnaire multi objectif tel que nous l'étudions dans l'annexe technique. L'algorithme est appliqué à un projet de construction comprenant 60 activités et 5 modes d'exécution au maximum. Les résultats montrent que cet algorithme permet d'obtenir de meilleures solutions non dominées que le NSGAI1 auquel il est notamment comparé.

### 4.3 Conclusion

Les modélisations de l'arbitrage durée, coût et qualité que nous avons étudiées dans ce chapitre considèrent pour la plupart une seule ressource non renouvelable dans un cadre continu (CTCTP et qualité) ou discret (DTCTP et qualité). Les modèles qui considèrent plusieurs ressources sont en effet plutôt rares. On recense ainsi un modèle RCPSP avec reprise où la qualité n'est pas considérée explicitement (Icmeli-Tukel et Rom, 1997 [78]) et deux modèles MRCPSP où la qualité globale est traitée comme une variable dont l'un considère la reprise (Fu et Zhang, 2016 [60]) et l'autre, non (Afruzi et al. 2014 [2]). Ainsi, aucune de ces contributions ne traite l'arbitrage durée, ressources et qualité dans des modélisations de type DTRTP ou MDTRTP que nous avons étudiées au chapitre précédent, de sorte que la qualité n'est jamais mise en relation directe avec les ressources utilisées. On peut néanmoins admettre qu'il peut être difficile d'établir une telle relation lorsque les ressources sont nombreuses ou hétérogènes de sorte que certaines combinaisons de ressources peuvent avoir des effets contraires sur la qualité (voir Tiwari et al., 2009 [141]).

La qualité individuelle d'une activité n'est jamais formulée comme une fonction des quantités de ressources utilisées. Au mieux, elle est exprimée comme une fonction de la durée seule ou de la durée et du coût. Ainsi, la qualité individuelle est une fonction continue linéaire et croissante de la durée dans Babu et Suresh (1996) [14] et Khang et Myint (1999) [88] qui s'intéressent au CTCTP et qualité (CTCQTP). La qualité individuelle est une fonction quadratique de la durée pour Zhang et al. (2014) [164] et Tran et al. (2015) [142] qui traitent le DTCTP avec qualité (DTCQTP). Connaissant la durée d'exécution d'une activité pour un mode d'exécution donné, il est ainsi possible de calculer la qualité individuelle dans ce mode grâce à cette fonction quadratique. Enfin, la qualité individuelle s'exprime comme une fonction continue et non linéaire de la durée et du coût dans le modèle CTCTP et qualité de Liberatore et Pollack Johnson (2013) [99] et dans le modèle MRCPSP de Fu et Zhang (2016) [60] où à nouveau, la qualité individuelle associée à un mode donné se calcule à partir de la durée et du coût de ce mode qui se définit aussi par une consommation de plusieurs ressources. Dans tous les autres modèles DTCTP et qualité, la qualité d'une activité associée à un mode d'exécution particulier fait l'objet d'une estimation, au même titre que le coût et la durée dans ce mode.

La qualité globale d'un projet est définie le plus souvent comme une moyenne arithmétique ou géométrique, simple ou pondérée des qualités individuelles des activités. La moyenne géométrique est utilisée lorsque les qualités individuelles sont dispersées. Plus rarement, la qualité globale combine le minimum et la moyenne des

qualités individuelles (voir Mungle et al., 2013 [109]) ou alors elle est mesurée par le minimum des qualités individuelles (Babu et Suresh, 1996 [14] et Liberatore et Pollack Johnson, 2013 [99]). Une définition plus sophistiquée de la qualité globale se trouve dans le travail de Fu et Zhang (2016) [60] qui introduisent le concept de qualité cumulée pour un sous-groupe d'activités avec des qualités interdépendantes ; la qualité cumulée pour chaque sous-groupe devant atteindre un seuil minimum. Hormis la qualité globale définie comme le minimum des qualités individuelles, les autres types d'agrégation soulèvent un problème de représentativité puisqu'une qualité globale élevée peut inclure des qualités individuelles très insatisfaisantes, voire inacceptables. On peut alors s'étonner du fait que ces contributions ne considèrent pas des seuils de qualité à atteindre pour toute qualité individuelle, ou du moins des sous-groupes rassemblant de façon cohérente un petit nombre d'activités individuelles. Si ce type d'agrégation par la moyenne par exemple peut sembler naturelle voir incontournable lorsque les qualités individuelles sont des paramètres estimés dans chaque mode de façon indépendante de la durée et du coût, la mise en place de seuils de qualité par petits groupes d'activités nous semble plus pertinente dans le cas où les qualités individuelles sont liées par une fonction à la durée ou au coût.

Nous avons vu qu'il existe deux types d'approches pour la résolution de ce problème d'arbitrage durée, coût et qualité. Une approche dite mono-objectif où l'on considère soit un seul objectif principal, soit plusieurs objectifs qui sont ramenés à un seul par différentes méthodes (méthode  $\epsilon$ -contrainte, but à atteindre ou somme pondérée). Les approches multi-objectifs qui traitent chaque objectif séparément sont des méthodes que l'on qualifie de non agrégée et peuvent reposer ou non sur le principe de Pareto optimalité. Toutes les contributions étudiées dans ce chapitre utilisent des algorithmes évolutionnaires qui s'appuient sur le principe de Pareto optimalité. De manière surprenante, aucun de ces travaux n'utilise une approche lexicographique où il s'agit de minimiser successivement chacun des sous-objectifs, s'affranchissant ainsi du concept de Pareto optimalité. Pourtant, de nombreux auteurs s'accordent pour dire que l'objectif prioritaire reste celui de la minimisation de la durée totale, suivi de celui de la minimisation du coût total (Mungle et al., 2013 [109] ; Wang et al., 2013 [153] ; Atkinson, 1999 [13]). Le dernier sous-objectif à considérer étant la maximisation de la qualité globale car plus difficile à évaluer (Basu, 2014[17]). Il existe donc en quelque sorte un ordre de préférence pour ces sous-objectifs et cet ordre rend pertinent l'emploi d'une approche lexicographique, pourtant absente de ces contributions.

# Chapitre 5

## Impact de l'augmentation des ressources sur la productivité et la qualité

Dans cette thèse nous cherchons à formuler analytiquement la relation entre la durée et la main d'oeuvre mobilisée pour un projet ainsi que la relation entre la qualité et l'utilisation de cette main d'oeuvre. Le but de ce chapitre est de présenter et discuter les hypothèses que nous adoptons pour formuler ces relations. En premier lieu, nous discutons notre choix de la main d'oeuvre seulement comme type de ressource (section 5.1) et précisément les heures supplémentaires et le travail temporaire (sous-section 5.1.1). Nous présentons ensuite une revue de la littérature relative au problème d'ordonnancement avec ressources multi-compétences (sous-section 5.1.2) afin de justifier notre choix d'intégrer des ressources flexibles. Dans les sections 5.2 et 5.3 nous étudions la relation entre la main d'oeuvre additionnelle et la productivité d'une part et analysons l'impact de cette main d'oeuvre sur la qualité d'autre part. La section 5.4 conclut le chapitre avec une attention particulière sur la paucité des modélisations qui considèrent les heures supplémentaires et le travail temporaire où la relation est faite avec la qualité.

### 5.1 Typologie des ressources

#### 5.1.1 Ressources renouvelables : la main d'oeuvre

La stratégie d'accélération par adjonction des ressources renouvelables ou non renouvelables est souvent associée aux problèmes d'arbitrage durée - coût (Time Cost Trade off Problem, TCTP) et durée - coût et qualité (Time Cost Trade off Problem, TCQTP).

Le TCTP a été étudié dans le chapitre 3 où nous avons conclu que peu de contributions considèrent ce problème avec des ressources renouvelables où le lien est fait entre la durée d'exécution d'une activité et sa consommation en ressources. Des auteurs comme Naber et Kolisch (2014) [110] et Ranjbar et al. (2009) [121] font ce lien par l'intermédiaire du contenu en travail des activités, dans des modèles qui

considèrent le problème avec ressources renouvelables multiples dans un cadre discret (MDTRTP). On note qu'aucune des contributions existantes sur le TCTP n'aborde l'idée selon laquelle l'accélération peut résulter d'un accroissement de la cadence de travail des opérateurs permanents, même si certains papiers plus généraux évoquent un système de primes pour motiver les travailleurs à travailler plus vite. Il s'agit d'une forme d'accélération sans ajout de ressources que nous ne prendrons pas en compte, en particulier parce que cette solution des primes ne peut être réellement efficace à long terme. En effet, les travailleurs ne peuvent soutenir un rythme intensif pendant très longtemps, primes ou non.

Les modèles TCQTP ont été étudiés dans le chapitre 4 où nous avons trouvé que la qualité est introduite implicitement (par le besoin de reprise) ou explicitement (comme variable du modèle) avec une hypothèse systématique de croissance entre la qualité et la durée. Ces modélisations considèrent pour la plupart une seule ressource non renouvelable (l'argent ou le coût). Il n'existe pas dans cette littérature de modélisation de la relation entre la qualité et les ressources utilisées de type MDTRTP.

Contrairement à ces modélisations, dans cette thèse, nous cherchons à formuler une relation directe entre durée, coût, qualité et quantité de ressources renouvelables utilisées dans un cadre discret, absente jusqu'à présent dans la littérature. Comme nous nous intéressons aux projets de fabrication où les besoins en équipements et matériaux sont généralement assez basiques et peu contraints, on considère uniquement la main d'oeuvre comme ressource. Par rapport aux autres ressources, c'est un accroissement de la main d'oeuvre qui implique le risque le plus important de détérioration de la qualité et la productivité. En effet, la main d'oeuvre peut être augmentée selon trois options : les heures supplémentaires, le sureffectif qui se caractérise par un nombre de travailleurs supérieur à la capacité normale d'un poste (ou d'un lieu de travail) et le travail par quart (équipes). Le travail pendant les heures supplémentaires peut généralement être mis en place assez rapidement mais implique une fatigue et donc une baisse de la qualité et de la productivité du travail (Mohamad et al., 2012 [105]). D'autres problèmes comme la congestion peuvent se poser en cas de sureffectif et impacter également la qualité et la productivité. Le recours aux équipes peut éviter ce problème de congestion. Néanmoins, une étude récente de Jun et El-rayes (2010) [85] montre que le travail l'après-midi et pendant la nuit augmente le taux d'absentéisme et diminue ainsi la productivité du travail. En outre, selon Gunduz (2004) [65] la coordination entre les changements d'équipes est très compliquée à mettre en oeuvre, d'où la solution retenue souvent en pratique d'exploiter la solution d'un sureffectif en conjonction avec les heures supplémentaires. C'est ce que nous retiendrons pour notre modélisation, **on accélère le projet par adjonction de travailleurs temporaires et d'heures supplémentaires.**

### 5.1.2 Hypothèses relatives à la main d'oeuvre en termes de compétences et d'utilisation

Dans la littérature relative à l'ordonnancement (project scheduling) et à l'affectation de la main d'oeuvre (staffing problem), le problème le plus couramment rencontré est le problème d'ordonnancement avec ressources multi-compétences contraintes ou Multi-Skill Resource Constrained Project Scheduling Problem (MSRCPSP). Comme le présentent Correia et al. (2012) [35], dans le MSRCPSP, chaque ressource (la main d'oeuvre) possède une ou plusieurs compétences et chaque activité nécessite un sous-ensemble de ressources maîtrisant les compétences requises à son exécution. On cherche à trouver un ordonnancement des activités d'un seul projet sous contraintes de précédence et des ressources multi-compétences. Les contraintes de ressources dans ce problème sont naturellement plus complexes que ceux du problème classique d'ordonnancement avec ressources contraintes (RCPSP), que nous avons introduit dans le chapitre 3. Les besoins en ressources par compétence et par période dépendent de l'ordonnancement des activités. A partir de ces besoins, une affectation de la main d'oeuvre est décidée. Mais comme la main d'oeuvre par compétence est disponible en quantité limitée, alors on ne peut pas mettre en oeuvre n'importe lequel des ordonnancements possibles. Ce problème d'affectation de la main d'oeuvre dépend donc du problème d'ordonnancement des activités. Les contraintes de disponibilités de ressources influencent l'ordonnancement de sorte qu'il est difficile de considérer les deux types de problèmes de façon indépendante.

Beaucoup de travaux s'intéressent au problème MSRCPSP et peuvent être distingués selon les questions suivantes :

1. A quel niveau de décision s'intéresse-t-on, tactique ou opérationnel? Plus précisément, s'agit-il du problème d'affectation de la main d'oeuvre à un projet particulier lorsque plusieurs projets peuvent être exécutés en parallèle? ou de l'ordonnancement des activités qui composent un projet donné?
2. Est ce que l'efficacité, définie comme la productivité du travail, varie d'une ressource à l'autre (efficacités hétérogènes) ou non (efficacités homogènes)?
3. Est ce qu'il y a un effet d'apprentissage ou d'oubli sur l'efficacité d'une ressource (efficacité dynamique) ou non (statique)? L'effet d'apprentissage opère lorsqu'on doit former un travailleur à chaque changement de poste ou d'activité. L'effet d'oubli a lieu lorsqu'un travailleur a tendance à oublier ses connaissances requises pour une activité donnée quand il passe un certain temps sur une autre activité.

Dans cette thèse, nous considérons un seul projet de fabrication (mono-projet). Le problème de la sélection de projets à traiter simultanément survient surtout dans les entreprises spécialisées dans le développement et sort de notre cadre de recherche. En revanche, dans notre cas, il y a aussi cette interdépendance entre le problème d'ordonnancement et le problème d'affectation de ressources, car les projets de fabrication nécessitent également plusieurs spécialisations professionnelles (ou compétences) contraintes.

Afin de justifier nos hypothèses relatives à la main d'oeuvre en termes de compétence et d'efficacité, nous discutons dans la sous-section 5.1.2.1 plusieurs modélisations qui considèrent les ressources multi-compétence avec efficacités homogènes et statiques ainsi que leur champ d'application. Dans la sous-section 5.1.2.2, nous traitons le cas des ressources avec efficacités hétérogènes et dynamiques.

### 5.1.2.1 Efficacités homogènes et statiques

Les ressources avec efficacités homogènes et statiques sont généralement considérées dans des projets uniques où les activités sont liées par des contraintes de précédence. Bellenguez-Morineau et Néron (2007) [19] s'intéressent au Multi-Skill Resource Constrained Project Scheduling Problem (MSRCPS) où, en plus des contraintes de précédence, les contraintes de ressources multi-compétences sont exprimées de la même façon que Correia et al. (2012) [35]. Les auteurs ajoutent qu'une main d'oeuvre ne peut pas être affectée à une activité qui nécessite une compétence qu'elle ne maîtrise pas. Ces auteurs soulignent que le problème peut être considéré comme un problème multi-mode mais avec un nombre très élevé de modes, en raison du nombre de sous-ensembles de ressources éligibles pour exécuter les activités du projet (c'est-à-dire la main d'oeuvre maîtrisant une compétence requise).

Vairaktarakis (2003) [145] définit une mesure du nombre de sous-ensembles de ressources éligibles pour exécuter une activité nécessitant une ou plusieurs compétences. Cet auteur teste les cas extrêmes de multi-compétences à savoir les cas de ressources pleinement compétentes et des ressources à compétence unique. Les ressources pleinement compétentes sont telles que chacune d'elles possède la totalité des compétences requises par le projet et c'est un cas très spécifique notamment rencontré dans les cabinets d'audits. Les ressources à compétence unique quant à elles sont implicitement considérées dans le RCPSP classique où chaque ressource a une seule compétence spécifique et pour chaque compétence et chaque période, la capacité de toutes les ressources peut être agrégée par une capacité globale. Alfares et Bailey (1997) [11] considèrent les ressources à compétence unique avec efficacité homogène et statique dans le cas des projets de construction. Chaque activité a plusieurs durées possibles d'exécution en fonction de la quantité de main d'oeuvre qui lui est allouée. Chaque durée d'exécution possible est calculée par la division du contenu en travail de l'activité par le nombre de travailleurs permanents qui lui sont alloués, ce qui suppose en réalité que la main d'oeuvre allouée par période est constante. On note que ces auteurs ne considèrent pas les ressources additionnelles. L'objectif est de minimiser la durée totale du projet et le coût de la main d'oeuvre permanente sous contraintes de précédence et de disponibilité globale des ressources. Les auteurs proposent un programme linéaire qui planifie simultanément les activités en intégrant les deux jours de repos pour la main d'oeuvre.

Dans Alba et Chicano (2007) [8] chaque travailleur est caractérisé, en plus du sous-ensemble de compétences qu'il possède, par son salaire et son taux d'implication maximum dans le projet, et qui représente le ratio du nombre d'heures maximum que le travailleur peut consacrer au projet par jour rapporté au nombre d'heures standard d'une journée de travail. Un taux d'implication supérieur à 1 signifie que

le travailleur peut travailler en heures supplémentaires. Par exemple, si le travailleur travaille 10h par jour sur le projet et qu'une journée standard est de 8 heures, alors son taux d'implication est de  $10/8=1,25$ . Les auteurs s'intéressent aux projets de développements des logiciels informatiques où les activités sont liées par des contraintes de précédence. L'objectif est de minimiser le coût et la durée du projet sous contraintes de précédence, de correspondance entre les compétences requises pour l'exécution des activités et celles des travailleurs qui leur sont allouées, et le respect du taux d'implication maximum. Contrairement aux contributions précédentes, ces auteurs évoquent la qualité et relèvent que le gestionnaire de projet doit tenir compte du fait qu'un travailleur surchargé peut effectuer un nombre d'erreurs plus élevé. Cela conduit à une qualité inférieure du produit final et, éventuellement, à la nécessité de reprises. Les auteurs développent un algorithme génétique et le testent sur 48 instances générées par eux-mêmes, avec au maximum 20 travailleurs, 30 activités et 10 compétences par travailleur. Ces auteurs considèrent les heures supplémentaires comme nous le faisons qu'ils expriment non pas en heures mais par un taux d'implication par travailleur. Ce modèle s'inscrit dans le cadre de décisions opérationnelles tandis que le nôtre se situe au niveau tactique (planification à moyen terme de la main d'oeuvre). Et contrairement à notre modèle, le travail temporaire est absent. En outre, le problème multi-objectif durée, coût et qualité n'est pas explicitement modélisé comme un programme mathématique mais seulement qualitativement décrit. La qualité n'est pas donc modélisée, elle est juste évoquée.

Maghsoudlou et al. (2016) [102] modélisent la qualité dans un problème d'affectation de ressources multi-compétences. Il s'agit d'un modèle multi-mode, multi-compétences où la ressource considérée est la main d'oeuvre avec efficacités homogènes et statiques. À chaque mode d'exécution pour chaque activité sont associés une durée, une qualité, un besoin de ressource possédant la compétence requise dans le mode considéré. L'objectif est de minimiser la durée totale d'exécution du projet, le coût total et la qualité globale (somme pondérée des qualités individuelles des activités), sous les contraintes habituelles de précédence, de non préemption et de respect des capacités disponibles en main d'oeuvre et couverture des besoins en compétences requises par chaque activité. Les auteurs développent trois métaheuristiques pour la résolution du problème et comparent les résultats avec 30 problèmes de tailles petite, moyenne et grande.

Bassett (2000) [16] considère les ressources multi-compétente avec efficacités homogènes et statiques dans les projets de recherche et développement (R&D) en industrie chimique, mais contrairement à toutes les contributions précédentes citées dans cette sous-section, des activités indépendantes doivent être planifiées dans des fenêtres de temps avec des ressources internes et externes. Il n'y a pas donc de liens de précédence entre les activités. L'auteur cherche à optimiser l'utilisation du temps et l'expertise des travailleurs mobilisés pour limiter le recours aux ressources externes.

### 5.1.2.2 Efficiences hétérogènes et dynamiques

Valls et al. (2009) [146] s'intéressent au problème multi-compétences avec efficiences hétérogènes dans les centres de services, où ils considèrent que les ressources peuvent avoir différents niveaux d'efficacité qui s'interprètent comme des productivités différentes. En particulier, chaque main d'oeuvre est caractérisée par un niveau d'efficacité qui est un paramètre supérieur ou égal à 1. Plus ce paramètre est élevé, moins le travailleur est productif. On a ainsi une modélisation avec un certain nombre de travailleurs individuels, chacun possède un domaine d'expertise lui permettant d'exécuter un sous-ensemble donné d'activités avec une durée correspondante à son niveau d'efficacité. Les juniors ont un niveau d'efficacité égale à 1, les confirmés ont un niveau 2 et les seniors un niveau 3. Par rapport à la durée d'exécution par un travailleur confirmé (niveau d'efficacité 2), cette durée est augmentée de 25% pour les juniors et diminuée de 25% pour les seniors, on note que ces pourcentages ne sont pas justifiés et restent assez arbitraires. L'objectif principal de la modélisation proposée est d'obtenir rapidement un plan d'affectation des ressources satisfaisant les dates de fin établies par les clients et les contraintes de travail relatives au nombre d'heures de travail et aux pauses réglementaires. Pour la résolution du problème, les auteurs proposent un algorithme génétique hybride qui combine un algorithme génétique classique avec une méthode de recherche locale.

Drexler (1991) [47] considère également les ressources avec efficiences hétérogènes dans le cadre de l'ordonnancement des audits mais avec des ressources entièrement compétentes. Certains auditeurs sont plus compétents pour faire certaines activités que d'autres et la durée de l'activité dépend de la ressource qui lui est allouée. Pour une activité donnée, on a des durées et des coûts différents par auditeur. L'objectif du modèle proposé par cet auteur est de trouver l'affectation des ressources aux activités de sorte à minimiser le coût sous contrainte de délai et de respect de la capacité des ressources.

Tiwari et al. (2009) [141] considèrent les ressources hétérogènes dans un modèle qui intègre la qualité et les auteurs s'intéressent à la production de cours et de supports de formation pour des clients qui achètent des services particuliers de téléphonie ; ce champ d'application étant éloigné du nôtre.

Wu et Sun (2006) [158] considèrent de plus les ressources avec efficiences dynamiques mais dans le cas multi-projet. Dans ce problème, de nombreux projets sont à réaliser sur un horizon de temps discret, sans prise en compte des contraintes de précedence entre les activités. Chaque travailleur peut être assigné à un seul projet à chaque période. À la fin de chaque période, les travailleurs peuvent être réaffectés à des activités différentes. L'objectif est de minimiser le coût des ressources externes. Le problème est formulé comme un programme non linéaire mixte et résolu avec un algorithme génétique.

Walter et Zimmermann (2016) [152] s'intéressent également à des projets indépendants à réaliser sur un horizon de moyen terme (typiquement une années) dé-

coupé en périodes d'un mois. Mais chaque travailleur peut être assigné à différents projets à chaque période. De plus, les profils de ressources sont flexibles c'est-à-dire si un projet nécessite 100 heures de travail par exemple, il pourra être exécuté par un seul travailleur pendant 100 heures, ou par 2 travailleurs pendant 50 heures chacun, etc. Les travailleurs ont des compétences multiples avec quatre niveaux d'efficience (productivité) possibles. L'objectif est de minimiser le nombre de travailleurs affectés à chaque projet ce qui permet de minimiser la taille moyenne de l'équipe affectée à un projet et donc de contourner les inefficiences liées à une taille d'équipe trop importante. Néanmoins, ces auteurs ne proposent pas de modélisation de la durée ni de l'impact de la taille de l'équipe sur la productivité. Le problème est formulé comme un programme linéaire en nombre entiers et résolu avec des instances sur 12 mois, comportant au maximum 100 projets et 50 travailleurs et un maximum de 10 compétences.

Contrairement à toutes les contributions citées précédemment, Heimerl et Kolisch (2010) [72] est le seul modèle qui considère les heures supplémentaires et le travail temporaire, comme nous le faisons. Les auteurs cherchent à minimiser la somme des coûts associés à l'exécution de projets de développement informatique indépendants. Ces coûts comprennent le coût de la main d'oeuvre temporaire et permanente (en heures normales et supplémentaires) nécessaire pour accomplir le projet. Le modèle impose des contraintes de couverture des compétences requises par les projets et de disponibilité de la main d'oeuvre (interne et externe) possédant les dites compétences. Les travailleurs temporaires sont recrutés en unités discrètes et leur proportion ne doit pas dépasser un certain pourcentage de la main d'oeuvre interne pour que les compétences principales restent de l'ordre de la maîtrise de l'entreprise. La contrainte vaut pour chaque projet pris individuellement mais sur l'ensemble des périodes et des compétences. En effet, la durée d'exécution de chaque projet est connue et déterministe, c'est la main d'oeuvre en quantité variable qu'on cherche à allouer à son exécution. Plus cette main d'oeuvre est efficiente (productive), moins on aura besoin d'en allouer à un projet en particulier, et donc on aura davantage de disponibilité de cette main d'oeuvre pour travailler sur d'autres projets.

### 5.1.2.3 Hypothèses sur la main d'oeuvre retenues pour la modélisation

Selon la littérature donnée dans cette section et comme Heimerl et Kolisch (2010) [72] le soulignent, la prise en compte des ressources multi-compétence avec efficacies hétérogènes et dynamiques est particulièrement appropriée pour les entreprises de services, de développement informatique ou dans les projets de conception (Alba et Chicano 2007 [8]; Basset, 2000 [16] et Valls et al., 2009 [146]). Dans ces domaines d'application, et au contraire des entreprises manufacturières et aux projets de fabrication que nous considérons, la ressource principale est une main d'oeuvre multi-compétente dont l'efficience est le plus souvent hétérogène. Les efficacies dynamiques quant à elles sont surtout considérées dans le cas de projets multiples qui sort du cadre de notre thèse où l'on considère un projet unique à la fois.

Dans les projets de fabrication où les ressources sont en effet avec efficacies ho-

mogènes (Tiwari et al., 2009 [141]), considérer la multiplicité de compétences pour chaque ressource n'est pas particulièrement pertinent car chaque spécialisation professionnelle ne possède qu'une compétence spécifique (pas de polyvalence). Cette hypothèse est validée chez Alstom où chaque travailleur est qualifié pour une spécialisation professionnelle (soudure, assemblage, peinture, etc). De même, en construction, il n'est pas restrictif de supposer qu'un maçon n'a pas les mêmes compétences qu'un électricien ou un peintre. On est donc bien dans le cas de compétence unique. Ainsi, à chaque spécialisation professionnelle est associée un groupe de travailleurs possédant une compétence unique (single-skilled resources) avec efficacité homogène (même productivité) et statique (pas d'effet d'apprentissage ou d'oubli).

Ainsi les activités d'un projet de fabrication peuvent être groupées par spécialisation professionnelle ou groupe de compétences, avec une seule compétence au sein de chacune d'elles. Il existe ainsi un groupe d'activités pouvant être effectuées par un groupe de plombiers, maçons ou électriciens, etc. Les travailleurs qualifiés pour un groupe d'activités peuvent exécuter n'importe quelle activité du même groupe avec le même niveau de productivité puisque nous considérons les ressources avec efficacités homogènes. Une compétence unique est requise par activité. Cette hypothèse est validée par la définition même d'une activité comme l'ensemble d'opérations effectuées sur un même poste de travail et nécessitant une même spécialisation professionnelle.

En mode normal d'exécution c'est-à-dire sans adjonction de main d'oeuvre additionnelle, la main d'oeuvre allouée par période et par activité est constante. En cas d'accélération, cette main d'oeuvre peut être flexible c'est-à-dire que, d'une période à l'autre, un nombre différent de travailleurs peut être affectés à l'exécution d'une activité. Nous considérons ainsi la possibilité de profils flexibles de ressources comme le font Naber et Kolisch (2014) [110]. Nous formulons ainsi notre hypothèse sur le profil des ressources qui sera considéré dans notre modèle.

**Hypothèse 1** *Pour le travail permanent, chaque spécialisation professionnelle détient une compétence unique et les travailleurs d'une spécialisation donnée ont une efficacité homogène (même productivité) et statique (pas d'effet d'apprentissage ou oubli). A chaque activité est associée une spécialisation qui est l'unique compétence qu'elle requiert pour son exécution, avec une variation possible du nombre de travailleurs qui lui est affecté d'une période à l'autre c'est-à-dire que les profils de ressources sont flexibles.*

## 5.2 Relation entre la productivité et la main d'oeuvre additionnelle

Rappelons que la productivité se définit comme le nombre d'unités produites par unité de travail (par heure, par exemple), cette unité produite pouvant être une activité. Une baisse de la productivité se traduit ainsi par une durée plus longue pour exécuter une activité donnée. Cette section est dédiée à étudier l'impact d'un recrutement en temporaires (sous-section 5.2.1) et du travail en heures supplémentaires

(sous-section 5.2.2) sur la productivité.

### 5.2.1 Productivité des travailleurs temporaires

Une analyse approfondie de l'impact des travailleurs temporaires sur la productivité a été proposée récemment par Hirsch et Mueller (2012) [76]. Ces auteurs mesurent la productivité de la firme par la valeur ajoutée brute par période (production valorisée au prix de base diminuée des consommations intermédiaires qui n'incluent pas les salaires). Ils supposent que la présence de temporaires dans une entreprise permet une certaine flexibilité nécessaire pour lisser la charge de travail et réduire les sous-utilisations de main d'oeuvre et d'équipement parce qu'ils peuvent remplacer des permanents absents, d'où une augmentation du nombre d'unités produites par unité de travail soit une augmentation de la productivité. Le travail temporaire peut également augmenter la productivité s'il est utilisé comme mode d'évaluation du travailleur pour un recrutement en durée indéterminée, cela fournit une motivation au temporaire pour accroître sa productivité. De plus, la plupart des études empiriques, dont une revue détaillée est donnée dans De Cuyper et al. (2008) [38], montrent que les temporaires sont aussi impliqués que les permanents dans une entreprise.

Cependant, le niveau de compétence (capital humain) d'un travailleur temporaire par rapport à celui d'un permanent peut nuire à sa productivité. En effet, les temporaires ont tendance à être moins productifs car il ont en moyenne moins de connaissances spécifiques à l'entreprise qu'un permanent. L'effet d'entraînement a aussi un impact négatif sur la productivité. En effet, la cohabitation entre permanents et temporaires peut nuire à la productivité de ces deux types de main d'oeuvre en raison de tensions liées au statut. Les permanents peuvent se considérer comme supérieurs aux temporaires en raison de leur expérience ou lorsque les temporaires se voient affecter des activités plus simples ou moins désirables. Ou inversement, comme le soulignent Kraimer et al. (2005) [94], les permanents peuvent craindre une perte de leur statut lorsqu'ils occupent des fonctions similaires à celles des temporaires, parce que permanents et temporaires peuvent se trouver en rivalité pour postuler à des emplois permanents plus qualifiés au sein de l'entreprise. Les permanents se sentent ainsi menacés par la possible embauche en contrat à durée indéterminée des temporaires.

En conclusion, si la flexibilité et l'évaluation ont un impact plutôt positif sur la productivité, le capital humain et l'effet d'entraînement peuvent en revanche nuire à la productivité de la firme, en générant des conflits et une concurrence entre les travailleurs. Nous considérerons donc que les travailleurs temporaires ont une productivité semblable à celle des permanents par compensation des effets contraires ; leur manque d'expérience est compensé par le fait qu'ils travaillent plus dur pour être bien évalués ou éventuellement embauché. Cela exclut les cas où l'on embauche des temporaires plus qualifiés que les permanents et qui peuvent augmenter la productivité ou inversement baisser la productivité par l'effet d'entraînement. Dans le cadre des projets de fabrication, on considère des emplois d'ouvriers qualifiés pour une compétence donnée, soit une seule spécialisation professionnelles (soudure, assemblage, peinture, maçonnerie, etc.). Les temporaires doivent avoir le même niveau de productivité que les permanents pour exécuter toute activité nécessitant la même

spécialisation.

**Hypothèse 2** *La productivité d'un travailleur temporaire est identique à celle d'un travailleur permanent, pour toutes les spécialisations professionnelles.*

Cette hypothèse de productivité identique entre permanents et temporaires est empiriquement confirmée par les gestionnaires Alstom Transport qui nous signalent cependant que la proportion de temporaires dans l'effectif total doit être limitée pour éviter notamment l'effet d'entraînement. En effet, au delà d'un certain nombre de travailleurs (temporaires et permanents), on ne peut plus augmenter la productivité. Broshak et Davis-Blake (2006) [27] montrent que les relations de travail se dégradent entre les travailleurs de façon continue et linéaire avec le pourcentage de temporaires dans l'effectif total, ce qui traduit une baisse de motivation de travail et de productivité. La limitation de la proportion de temporaires a été récemment évoquée par Hirsch et Mueller (2012) [76] qui montrent que l'accroissement de la productivité en fonction du pourcentage de travailleurs temporaires dans l'effectif total suit une courbe en cloche, avec une baisse de la productivité au delà de 15% de temporaires. Pour rapprocher notre modèle de cette étude réelle, nous intégrons la possibilité que le gestionnaire de projet souhaite une proportion maximum de temporaires dans l'effectif alloué à toute activité et à toute période.

Il convient de préciser que c'est le sureffectif qui dégrade la productivité et non le fait d'embaucher des travailleurs temporaires. Si ceux-ci ont même productivité que les permanents et si les postes de travail ne sont pas en sureffectif, il n'y aura de perte de productivité en raison de l'hypothèse 2. Notons aussi que le sureffectif peut également se produire avec les permanents seulement puisque nous autorisons les profils flexibles de ressources. En cas de sureffectif, les travailleurs peuvent se gêner par manque de place et le travail devient plus lent car le fait d'encombrer l'espace est une source d'inefficacité dans le travail. Smith et al. (1987) [129] relèvent qu'une densité de travail supérieure à un homme par 30  $m^2$  entraînerait une baisse de la productivité. La littérature sur la relation entre le sureffectif et la baisse de productivité est limitée et plutôt ancienne. A notre connaissance seul Gunduz (2004) [65] étudie la relation entre la perte de productivité et le taux de sureffectif. L'auteur présente plusieurs estimations de cette relation provenant de ces études relativement anciennes. En nous y référant nous parvenons à l'hypothèse selon laquelle la perte de productivité peut être décrite comme une fonction linéaire croissante du taux de sureffectif. Ainsi on a

**Hypothèse 3** *La perte de productivité est une fonction linéaire croissante du taux de sureffectif.*

Cette hypothèse peut être également justifiée par l'effet de « Ringelmann » évoqué récemment par Walter et Zimmermann (2016) [152] et qui décrit une baisse de productivité individuelle croissante avec l'augmentation de la taille de l'équipe.

Cette baisse augmente en raison de problèmes croissants de coordination et de diminution de la motivation. Le phénomène selon lequel les travailleurs sont moins motivés et font moins d'effort lorsqu'ils travaillent dans un groupe que lorsqu'ils travaillent seuls est appelé « social loafing » ou flânerie sociale. Ce phénomène s'amplifie avec la taille du groupe en raison d'une augmentation des opportunités de « se cacher dans la foule », mais aussi de se sentir « perdu dans la foule » ; en raison d'une dilution de l'influence perçue sur le résultat du groupe, par exemple. Cette hypothèse implique qu'en cas de sureffectif, la durée d'exécution d'une activité est une fonction non linéaire décroissante de la quantité de ressources qui lui est allouée, ainsi que nous le montrons dans la section 6.4 du chapitre 6 où nous développons le modèle que nous proposons.

### 5.2.2 Productivité et sureffectif des travailleurs en heures supplémentaires

Le travail en heures supplémentaires entraîne une baisse de la productivité liée à la fatigue morale et physique des travailleurs et au phénomène d'auto-régulation du rythme de travail décrit par Hanna et Sullivan (2004) [67] où les travailleurs ont tendance à ajuster leur charge de travail et ce, au delà de 10 heures supplémentaires par travailleurs et par semaine. Si les contraintes légales c'est-à-dire le maximum de 8 heures supplémentaires par travailleur et par semaine sont respectées, il n'y a pas de perte de productivité (Thomas et Raynar, 1997 [139]). Les études publiées sur la baisse de productivité liée au travail prolongé en heures supplémentaires et citées dans Mohamad et al. (2012) [105], telles que National Electrical Contractors Association (NECA 1989), Mechanical Contractors Association (MCA 1968) et Construction Industry Institute Thomas et Raynar 1994 (2008), se rejoignent pour parvenir à la conclusion selon laquelle la baisse de productivité est de l'ordre de 15% au-delà de quatre à cinq semaines de travail en heures supplémentaires. Après 10 à 12 semaines de prolongation des heures supplémentaires, la baisse de productivité peut facilement dépasser 35% à 40%. Dans notre cas nous limiterons les heures supplémentaires à 8 heures par semaine et par travailleur, soit 2 demi journées de 4h, ce qui reste inférieur aux 10 heures critiques. Par conséquent, nous considérons que la productivité du travail en heures supplémentaires est identique à celle en heures normales car nous respecterons les contraintes légales.

La perte de productivité liée au travail en heures supplémentaires ne peut donc résulter que d'un sureffectif des travailleurs en heures supplémentaires, une possibilité que nous excluons toutefois à l'aide d'une hypothèse simplificatrice selon laquelle seuls les travailleurs permanents font des heures supplémentaires et au delà des heures normales de travail, les postes ne sont jamais en sureffectif. Nous avons donc l'hypothèse suivante

**Hypothèse 4** *Si les contraintes légales de 8 heures par semaine et par travailleur sont respectées, alors la productivité du travail en heures supplémentaires est identique à celle en heures normales. En outre les travailleurs en heures supplémentaires ne sont jamais en sureffectif de sorte que seul le sureffectif en heures normales peut dégrader la productivité.*

### 5.3 Impact de l'accélération sur la qualité

Dans cette thèse, nous adoptons une mesure de la qualité par le taux des items validés dans une check-list de contrôle qualité comme le font Mohammadipour et Sadjadi (2016) [106]. Cette évaluation par une check-list est réellement utilisée dans plusieurs domaines de fabrication (construction et industrie). En mode normal, la qualité normale de toute activité atteint sa valeur maximum généralement égale à 1, ce qui signifie que 100% des items de la check-list qualité sont validés (Fu et Zhang, 2016 [60] ; Babu et Suresh, 1996 [14] et Khang et Myint, 1999 [88]). En mode accéléré, les contributions TCQTP font toutes l'hypothèse d'une relation décroissante de la qualité avec la réduction de la durée mais ne mettent pas en relation directe la dégradation de qualité avec l'utilisation des ressources.

Dans cette section, nous étudions la relation entre la perte de qualité et la main d'oeuvre additionnelle (sous-section 5.3.1). Nous introduisons ensuite la façon dont nous modélisons la détérioration de la qualité dans le modèle (sous-section 5.3.2).

#### 5.3.1 Qualité et adjonction de main d'oeuvre

En pratique, le recrutement d'un travailleur temporaire se fait toujours selon un descriptif de poste qui nécessite des connaissances et des compétences au moins identiques aux permanents occupant le même poste. Pour les niveaux de qualification que nous considérons, on ne requiert pas des temporaires plus qualifiés que les permanents, ce qui peut-être le cas par exemple lorsqu'on embauche temporairement des consultants. Par ailleurs, il n'existe pas d'études qui montrent que les temporaires font un travail de moins bonne qualité que les permanents, eu égard au type de spécialisations considérées (ouvriers et techniciens). Néanmoins, pour des postes plus qualifiés, la qualité du travail peut augmenter avec le recrutement de temporaires hautement qualifiés qui peuvent apporter de nouvelles pratiques et connaissances. Les questionnaires d'Alstom Transport nous confirment qu'un travailleur temporaire est indépendant lorsqu'il est bien formé pour exécuter un travail de qualité identique à un permanent.

**Hypothèse 5** *A l'issue d'une période de formation relativement courte, les travailleurs temporaires fournissent un travail d'une qualité identique à celle des permanents. Autrement dit, le travail temporaire aboutit à un taux d'items validés dans la check-list de contrôle qualité qui est égal à ce taux pour les permanents.*

S'agissant du sureffectif, seule l'étude de Li et al. (2000) [96] évoque son impact sur la qualité, un impact dont l'existence se justifie naturellement par le risque que les travailleurs se gênent car ayant moins d'espace pour oeuvrer. En outre, comme le soulignent Mohamad et al. (2012) [105] et Hanna et Sullivan (2004) [67], le travail en heures supplémentaires implique une fatigue et donc un risque accru de défauts d'où une baisse de la qualité. A notre connaissance, Li et al. (2000) [96] sont les seuls auteurs qui étudient l'impact sur la qualité de plusieurs combinaisons d'heures supplémentaires et de travailleurs temporaires utilisées pour accélérer un projet de construction de logements standardisés. Cette étude montre en particulier que si les

heures supplémentaires sont utilisées au maximum possible, alors la qualité baisse de 28% tandis que cette baisse n'est que de l'ordre de 15% si seul le sureffectif est employé à son maximum. Mais il n'y a pas de relation analytique de l'arbitrage entre le volume des heures supplémentaires, le taux de sureffectif et la baisse de qualité. Nous aboutissons à la formulation de l'hypothèse suivante.

**Hypothèse 6** *La perte de qualité est proportionnelle au volume du sureffectif et des heures supplémentaires avec un impact plus important de ces dernières sur la qualité.*

### 5.3.2 Niveaux de tolérance aux défauts

En pratique, lorsqu'un technicien qualité déroule une check-list de contrôle qualité, les éléments contrôlés ne sont pas toujours binaires (conforme/non conforme). Certaines non conformités peuvent être tolérées en l'état, notamment lorsqu'il s'agit de défauts d'aspects qui n'affectent pas la fonctionnalité du produit mais son esthétique. Par exemple, quelques grains de peinture dans une activité de peinture d'une voiture est un défaut qui peut être accepté par le client. En revanche, un défaut dans le montage du moteur peut empêcher le client d'utiliser son véhicule ou le mettre en danger, ce type de défaut ne peut pas être toléré. Cette tolérance aux défauts est réellement appliquée dans les projets de fabrication que nous considérons, et ce, selon des niveaux d'exigence ou critères d'acceptation internes et externes comme c'est le cas chez Alstom Transport et Eurogypsum par exemple (voir discussion dans le chapitre 2, sous-section 2.2.3, page 46).

Pour introduire ce type de pratique en termes de traitement de la qualité dans notre modèle, nous définissons des niveaux de tolérance aux défauts auxquels sont associés des seuils de qualité minimum à atteindre; ces seuils correspondant au pourcentage des items d'une check list qui doivent être validés pour éviter les reprises. Les activités peuvent ainsi être groupées dans des ensembles de niveau de tolérance identique.

Cette évaluation de la qualité par rapport à des seuils exclut l'interdépendance des qualités des activités, comme c'est le cas dans le concept de qualité cumulée proposé récemment par Fu et Zhang (2016) [60]. Autrement dit, notre modèle s'adapte aux situations d'indépendance des activités en termes de qualité au sens où la fonctionnalité d'une activité ne peut être remise en cause par des petits défauts acceptables individuellement. Dans le cas contraire (dépendance), les défauts individuellement acceptables finissent par se cumuler occasionnant un dysfonctionnement d'une activité successeur même si pris individuellement, ceux-ci ne dépassent pas le seuil.

**Hypothèse 7** *Les activités sont réparties dans plusieurs ensembles de niveaux de tolérance identique. A chaque ensemble ou niveau de tolérance est associé un seuil de qualité minimum à respecter en cas d'accélération. Ainsi la qualité de toute activité d'un niveau donné de tolérance ne doit pas descendre en dessous de ce seuil pour éviter la reprise.*

## 5.4 Conclusion

Nous avons vu dans ce chapitre que les contributions TCQTP ne modélisent pas la relation entre la qualité, la durée et l'utilisation d'un volume donné de main d'oeuvre. Ces contributions supposent une relation décroissante entre la qualité et la réduction de la durée liée à la consommation des ressources. C'est ainsi à partir d'un autre segment de la littérature que celui sur le TCQTP que nous avons pu construire les hypothèses nécessaires pour notre modèle. Dans le chapitre suivant, nous proposons un modèle d'optimisation mono-objectif où il s'agit de minimiser la durée sous contrainte de budget et de seuils de qualité à respecter. En supposant ainsi que les préférences des décideurs sont classées par ordre décroissant, avec le critère de la durée d'abord puis celui du coût et enfin la qualité, nous nous affranchissons de l'approche multi-objectif souvent utilisée dans la littérature ; cette approche ayant l'inconvénient majeur de produire un ensemble souvent très grand de solutions dominantes parmi lesquelles il est difficile de choisir (voir par exemple Mungle et al., [109]).

# Chapitre 6

## Un modèle d'arbitrage durée, coût et qualité d'un projet avec adjonction de travailleurs temporaires et heures supplémentaires

Ce chapitre présente la modélisation que nous avons développée pour accélérer les projets de fabrication en tenant compte des pertes de productivité et de qualité qu'implique le recours à une main d'oeuvre additionnelle en heures supplémentaires et en travail temporaire. La section 6.1 positionne le modèle par rapport à la littérature existante. La section suivante 6.2 définit les notations utilisées et fournit une description générale du problème. Dans la section 6.3, nous donnons la formulation du problème de minimisation de la durée sous contraintes de précédence, de non préemption et de disponibilité de la main d'oeuvre lorsque l'on considère uniquement le recours au travail temporaire. Cette formulation s'inspire de Naber et Kolisch (2014) [110] qui considèrent le même problème avec des profils de ressource flexible mais sans travail temporaire. Dans la section 6.4 nous étudions l'impact du sureffectif sur la productivité pour obtenir une formulation de l'effectif de travailleurs en heures normales corrigé des pertes de productivité. La section 6.5 introduit les heures supplémentaires dans la modélisation. La section 6.6 exprime la qualité comme une fonction du sureffectif et des heures supplémentaires et analyse la relation entre qualité des activités et durée d'exécution de celles-ci. La section 6.7 discute et formule la contrainte de budget. La section 6.8 fournit des bornes inférieures et supérieures sur la durée du projet et développe une idée de méthodologie pour estimer les paramètres. Enfin, la section 6.9 conclut en donnant les limites de la modélisation et des voies de recherche future.

### 6.1 Objectif et positionnement du modèle dans la littérature

Dans le chapitre 4, nous avons analysé la littérature existante sur le problème d'arbitrage durée, coût et qualité (Time Cost Quality Trade-off Problem, TCQTP)

et nous avons conclu qu'il n'existe pas de modélisation de la relation entre qualité et ressources utilisées. Ainsi, l'objectif de la modélisation que nous proposons dans ce chapitre est de formuler la relation entre qualité, coût, durée et quantité de ressources utilisées. Nous considérons uniquement la main d'oeuvre comme ressource. En effet, par rapport aux autres ressources comme les équipements ou les matériaux, c'est un accroissement de la main d'oeuvre qui implique le risque le plus important de détérioration de la qualité, car augmenter la main d'oeuvre passe par le recours aux heures supplémentaires et au travail temporaire. De plus, comme on s'intéresse aux projets de fabrication, les besoins en outillage et en matières premières sont les mêmes quel que soit le mode d'exécution.

La qualité est affectée par l'adjonction de main d'oeuvre en raison de l'utilisation d'heures supplémentaires et par un possible sureffectif caractérisé par un nombre de travailleurs supérieur à la capacité normale d'un poste. Il est communément admis que le travail en heures supplémentaires implique une fatigue et donc une baisse de la qualité. S'agissant du sureffectif, seule l'étude de Li et al. (2000) [96] évoque son impact sur la qualité, un impact dont l'existence se justifie toutefois naturellement par le risque que les travailleurs se gênent car ayant moins d'espace pour oeuvrer.

Par ailleurs, heures supplémentaires et sureffectif ont un effet sur la productivité qu'il convient d'analyser, s'agissant d'étudier comme nous le faisons la relation entre la durée d'exécution et la quantité de ressource (main d'oeuvre) utilisée. Le travail en heures supplémentaires entraîne une baisse de productivité liée à la fatigue, à une auto-régulation du rythme de travail ou à une possible baisse de conscience professionnelle et ce, au delà de 10 heures par travailleur et par semaine (Thomas et Raynar, 1997 [139]; Hanna et Sullivan, 2004 [67]; Mohamad et al., 2012 [105]). Mais si les contraintes légales sont respectées (au maximum 8 heures supplémentaires par travailleur et par semaine), la perte de productivité liée au travail en heures supplémentaires ne peut résulter que d'un sureffectif des travailleurs en heures supplémentaires, une possibilité que nous excluons toutefois à l'aide d'une hypothèse simplificatrice mais raisonnable (les travailleurs en heures supplémentaires ne sont jamais en sureffectif). Nous évitons ainsi la prise en compte d'un effet croisé des heures supplémentaires sur la qualité et la productivité. Ainsi, dans notre modèle, seul le sureffectif de travailleurs en heures normales impacte la productivité. Si un tel sureffectif est favorisé par l'embauche de travailleurs temporaires, il peut aussi avoir lieu avec des travailleurs permanents uniquement puisque nous considérons la possibilité de profils de ressources flexibles où, d'une période à l'autre, un nombre différent de travailleurs peut être affecté à l'exécution d'une activité (voir notamment Naber et Kolisch, 2014 [110]) tandis que le mode normal d'exécution suppose une quantité constante des ressources utilisées le long de la durée normale d'exécution.

Ainsi, le modèle que nous proposons formule la qualité comme une fonction linéaire décroissante du sureffectif et des heures supplémentaires, permettant alors d'obtenir une relation directe entre l'utilisation de ressources additionnelles et la qualité, jusqu'à présent absente de la littérature. Il incorpore de plus la baisse de productivité résultant du sureffectif, formulant ainsi de façon plus réaliste la relation (non linéaire) entre la durée d'exécution et la quantité de ressources utilisées. Le coût que nous prenons en compte est le coût des heures de travail permanent,

temporaire et supplémentaire. Contrairement aux modèles multi-mode qui à la fois limitent le nombre de modes et requièrent en pratique pour chaque mode d'exécution l'estimation par des experts de la durée, de la qualité et du coût, notre approche permet d'associer à toute combinaison possible des deux types de main d'oeuvre additionnelle (sureffectif et heures supplémentaires) une durée, une qualité et un coût grâce aux relations analytiques que nous développons.

Nous formulons le problème de minimisation de la durée d'un projet sous contrainte de budget et de qualité comme un programme linéaire mixte où les variables sont le nombre de travailleurs permanents et temporaires et le nombre d'heures supplémentaires affectés à l'exécution de chaque activité à chaque période de l'horizon. Dans la littérature relative à l'ordonnancement, il existe peu de modélisations où les heures supplémentaires et le travail temporaire sont considérés. Yang et al. (2004) [160] s'intéressent au problème d'ordonnancement de plusieurs activités utilisant une même main d'oeuvre avec un volume d'heures normales et supplémentaires limité à chaque période. L'objectif est de trouver un ordonnancement qui minimise le coût des heures supplémentaires et les coûts de retard en respectant les contraintes de précedence, de non préemption et de disponibilité en main d'oeuvre. Heimerl et Kolisch (2010) [72] considèrent non seulement les heures supplémentaires mais également le travail temporaire dans un problème qui consiste à minimiser la somme des coûts associés à l'exécution de projets indépendants, c'est-à-dire non liés par des contraintes de précedence. Les travailleurs sont groupés en fonction des compétences qu'ils possèdent et au sein d'un même groupe de compétences, les travailleurs peuvent avoir différents niveaux d'efficience qui s'interprètent comme des productivités différentes. Dans notre modèle, contrairement à Yang et al. (2004) [160] nous incorporons aussi le travail temporaire comme le font Heimerl et Kolisch (2010) [72] mais pour un projet de fabrication où les activités sont liées par des contraintes de précedence. Par ailleurs, nous considérons également plusieurs types de compétences nécessaires à l'exécution d'un projet (des corps de métiers comme la soudure ou l'assemblage) mais chaque spécialisation professionnelle ne possède qu'une compétence spécifique (pas de polyvalence), de sorte que notre modèle s'apparente à un problème d'affectation de ressources humaines à compétence unique (single-skilled resources) avec efficience homogène (pas de différence de productivité entre les travailleurs d'une même spécialisation professionnelle) et statique (pas d'effet d'apprentissage ou d'oubli). Ces hypothèses sont appropriées pour les projets de fabrication que nous considérons car c'est en général dans les sociétés de service (audit par exemple) ou pour les projets de conception que les modèles d'affectation de ressources à compétences multiples sont pertinents puisque chaque type de travailleurs (consultant par exemple) possède plusieurs compétences différentes nécessaires à l'exécution du projet. A notre connaissance, Maghsoudlou et al. (2016) [102] sont les seuls auteurs qui intègrent la qualité dans un problème d'affectation de ressources multi-compétence. Il s'agit plus précisément d'un modèle multi-mode, multi-compétence avec efficience homogènes et statiques où à chaque mode d'exécution pour chaque activité sont associés une durée, une qualité, un besoin de ressource possédant la compétence requise dans le mode considéré.

## 6.2 Notations et description du problème

On considère un projet de fabrication constitué d'un ensemble  $V = \{0, 1, \dots, n, n+1\}$  d'activités où les activités 0 et  $n+1$  sont les activités fictives de début et de fin du projet. Toute activité  $j$  est liée par des contraintes de précédence avec un ensemble  $\Gamma^{-1}(j)$  d'activités prédécesseur. Chaque activité  $j$  possède un ensemble d'activités qui lui succèdent (successeurs) que l'on note  $\Gamma(j)$ .

Toute activité  $j$  est exécutée sur un seul poste de travail qui réunit des travailleurs d'une seule spécialisation professionnelle  $k$ . On note  $V_k$  l'ensemble des activités qui nécessitent pour leur exécution la spécialisation  $k$ , avec  $k = 1..K$ . Si l'on s'intéresse par exemple à un projet de construction d'un immeuble alors  $k = 1$  peut représenter la main d'oeuvre spécialisée en maçonnerie,  $k = 2$ , celle en plomberie, etc. On désigne par  $P_k$  le nombre total de permanents de spécialisation  $k$  dédiés au projet considéré.

En mode normal, un nombre  $P_{kj}$  de permanents de spécialisation  $k$  est alloué à l'exécution de l'activité  $j$  sur un nombre de périodes égal à  $d_j^{\max}$  qui correspond à la durée normale d'exécution de l'activité. Le contenu en travail  $W_j$  d'une activité  $j$  se définit par  $W_j = P_{kj} \cdot d_j^{\max}$  qui représente le besoin total en main d'oeuvre requise pour exécuter l'activité  $j$  en équivalent hommes - jour. Par exemple, si  $P_{kj} = 2$  et  $d_j^{\max} = 3$  cela veut dire que l'activité peut être exécutée en mode normal en 3 jours avec 2 travailleurs de spécialisation  $k$  par jour ; le besoin total de main d'oeuvre c'est-à-dire son contenu en travail est de  $W_j = 6$  hommes-jour. Le paramètre  $P_{kj}$  peut s'interpréter comme la capacité normale en nombre de travailleurs du poste sur lequel l'activité  $j$  est exécutée. On parle de sureffectif sur l'activité  $j$  à une certaine période dès lors qu'un nombre de travailleurs strictement supérieur à  $P_{kj}$  opèrent sur cette activité à cette période. L'effectif maximum  $s_j$  qu'il est possible d'allouer à une activité  $j$  à toute période va dépendre de la capacité du poste  $A_j$  exprimée en nombre de travailleurs sur l'activité  $j$ , du contenu en travail  $W_j$  que l'activité nécessite au total pour son exécution et d'un paramètre  $\gamma_j$  qui mesure l'impact du sureffectif sur la productivité. Lorsque cet impact est important, l'effectif maximum va être naturellement limité puisque l'ajout de travailleurs additionnels peut devenir contre productif. En revanche, lorsque cet impact est nul, alors on peut allouer  $W_j$  travailleurs à l'exécution de l'activité si la capacité maximum  $A_j$  le permet, c'est-à-dire si  $A_j \geq W_j$ . Remarquons que pour deux activités différentes  $i$  et  $j$  exécutées à des moments différents sur un même poste, on a bien évidemment  $A_i = A_j$ .

La durée normale  $d_j^{\max}$  est une borne supérieure sur la durée de l'activité et l'on a par ailleurs  $d_0^{\max} = d_{n+1}^{\max} = 0$ . La borne inférieure  $\underline{d}_j$  sur la durée d'exécution de l'activité  $j$  est une durée théorique obtenue en allouant à son exécution toutes les ressources disponibles.

La longueur de l'horizon  $T$  correspond à la durée totale d'exécution du projet en mode normal, selon un ordonnancement au plus tôt ou au plus tard et se calcule donc sur la base des durées normales d'exécution  $\{d_j^{\max}\}_{j=1..n}$ , en imposant un nombre de permanents égal à  $P_{kj}$  à chaque activité  $j$  et à toute période d'exécution. Ainsi,  $T$  constitue la borne supérieure sur la durée totale du projet. La borne inférieure  $\underline{T}$  sur la durée du projet est calculée à partir des bornes inférieures  $\{\underline{d}_j\}_{j=1..n}$  sur les durées des activités.

On note  $ES_j$  la date de début au plus tôt de l'activité  $j$  calculée sur la base

des bornes inférieures de durées  $\{d_j\}_{j=1..n}$  et l'on note  $LF_j$  sa date de fin au plus tard déterminée à partir des durées normales  $\{d_j^{\max}\}_{j=1..n}$ . Les périodes possibles d'exécution de l'activité  $j$  se situent donc dans l'intervalle  $\{ES_j..LF_j\}$  que l'on note  $\mathcal{T}_j$ . On peut ainsi définir  $V_{kt}$ , l'ensemble des activités qui nécessitent la spécialisation  $k$  à la période  $t$  avec  $V_{kt} = \{j \in V_k | t \in \mathcal{T}_j\}$ .

En mode normal, la qualité de toute activité  $j$  atteint sa valeur maximum  $q_j^{\max}$  généralement égale à 1, ce qui signifie que 100% des items de la check-list qualité sont validés. Lorsque l'on accélère le projet, la qualité d'une activité décroît avec le sureffectif et les heures supplémentaires qui lui sont alloués. On note  $\Delta q_j^O$  la perte de qualité associée à l'utilisation d'une heure supplémentaire et  $\Delta q_j^S$  la perte de qualité résultant de l'emploi d'un travailleur en sureffectif par jour.

Chaque activité possède un niveau  $\theta$  de tolérance aux défauts, ce qui signifie que sa qualité en cas d'accélération ne doit pas descendre en dessous d'un seuil minimum que l'on note  $q_\theta^{\min}$ . Plutôt que de définir un seuil minimum de qualité par activité ce qui nécessiterait un surcroît d'estimation, on groupe les activités par niveaux de tolérance, avec un nombre  $\Theta$  de niveaux de tolérance inférieur à  $n$ . On désigne par  $V^\theta$  le sous-ensemble des activités ayant même niveau de tolérance  $\theta$ , avec  $\theta = 1..\Theta$  avec  $q_1^{\min} > q_2^{\min} > \dots > q_\Theta^{\min}$ . Par exemple, pour  $\theta = 1$  on peut avoir  $q_1^{\min} = 1$  de sorte que toute activité dans  $V^1$  est dite critique au sens où aucun défaut n'est toléré. A l'inverse, les activités dans  $V^\Theta$  sont celles pour lesquelles on tolère le plus de défauts. Il s'agit en général d'activités qui ne remettent pas en cause la fonctionnalité du produit fini et dont les défauts affectent son esthétique.

On peut imposer une proportion maximum  $\lambda$  de temporaires dans l'effectif alloué à toute activité et à toute période. Les heures supplémentaires sont assurées par les permanents uniquement ; elles se font au delà du nombre  $H_R$  d'heures régulières (normales) par jour et sont limitées à un nombre  $H_O$  d'heures par jour et par travailleur. De plus, tout travailleur permanent ne peut excéder un nombre  $H_{OW}$  d'heures supplémentaires par semaine. On note  $c_P$ ,  $c_U$  et  $c_O$ , le coût d'une heure de travail permanent, temporaire et supplémentaire, respectivement.

Enfin, on désigne par  $\bar{w}_j$ ,  $\bar{p}_j$  et  $\bar{u}_j$ , les bornes supérieures sur le nombre de travailleurs total, permanents et temporaires respectivement, alloués à l'activité  $j$  à toute période.

L'objectif est de minimiser la durée  $C_{max}$  du projet en respectant les contraintes de précédence, de non préemption, de besoins en main d'oeuvre et de disponibilité en ressources, tout en maîtrisant les coûts, avec des qualités individuelles au delà des seuils minimum. Les variables de décision sont les suivantes. On note respectivement  $p_{jt}$  et  $u_{jt}$  le nombre de travailleurs permanents et temporaires de spécialisation  $k$  alloués à l'exécution de l'activité  $j$  en  $t$ . La variable  $h_{jt}$  désigne le nombre d'heures supplémentaires affectées à l'activité  $j$  en  $t$  et  $z_{jt}$  est une variable binaire qui prend la valeur 1 si l'activité  $j$  est en cours d'exécution à la période  $t$  et zéro sinon, pour tout  $t \in \mathcal{T}_j$  et  $j = 1..n$ .

Ainsi, les variables  $p_{jt}$  et  $u_{jt}$  sont des variables entières pour des raisons de pertinence en termes de planification des recrutements en temporaires et pour établir un calendrier prévisionnel du personnel. En outre, un raisonnement en volume d'heures impliquerait une modélisation en temps continu, ce qui constitue une approche inhabituelle en gestion de projet dans la pratique. Par ailleurs, les temporaires sont

recrutés pour une journée de travail au moins (beaucoup plus rarement par demi-journées dans les industries manufacturières) et une modélisation en temps continu ou en volume d'heures n'autoriserait pas une grande visibilité sur l'affectation du personnel aux différents postes, et pourrait impliquer en particulier qu'un travailleur puisse passer 3 heures sur un poste puis 5 heures sur un autre, ce qui va à l'encontre de la pratique d'affectation sur les postes où un travailleur n'opère généralement pas sur plusieurs postes différents dans une même journée. En revanche, les heures supplémentaires sont des variables continues car légalement payables dès la première minute.

A l'issue du processus d'optimisation, la durée optimale d'exécution  $d_j \leq d_j^{\max}$  d'une activité  $j$  est donc égale à  $\sum_{t=t_j^*..t_j^*+d_j^*-1} z_{jt}$  où  $t_j^*$  est la période optimale de début d'exécution de l'activité  $j$  et l'allocation optimale des ressources est telle que la somme  $\sum_{t=t_j^*..t_j^*+d_j^*-1} w_{jt}$  couvre le besoin  $W_j$  en main d'oeuvre, où  $w_{jt}$  désigne la main d'oeuvre effective en permanents, temporaires et heures supplémentaires à chaque période. Comme en mode normal nous avons la relation  $W_j = P_{kj} \cdot d_j^{\max}$  soit encore  $d_j^{\max} = W_j/P_{kj}$ , cette relation se généralise de la manière suivante

$$d_j = \left\lceil \frac{W_j}{w_j} \right\rceil \quad \forall j = 1..n, \quad (6.1)$$

où  $w_j$  désigne la main d'oeuvre moyenne allouée à chaque période d'exécution avec  $w_j = \frac{1}{d_j} \sum_{t=t_j^*}^{t_j^*+d_j^*-1} w_{jt}$ . Nous utiliserons à plusieurs reprises cette relation à des fins d'illustration dans la présentation du modèle.

Une liste des notations utilisées est également donnée dans l'annexe B.

## 6.3 Modélisation de base avec travailleurs temporaires

Dans cette section, nous formulons le problème de minimisation de la durée du projet sous les contraintes habituelles de précédence, de non préemption et de respect de la capacité en ressources en incorporant le travail temporaire. Nous nous inspirons de la formulation du problème sous contrainte de ressources multiples avec profils flexibles de ressources donnée par Naber et Kolisch (2014) [110]. Nous ne considérons ici ni l'impact du sureffectif sur la productivité ni la qualité. On suppose dans un premier temps que le recrutement en travailleurs temporaires n'est pas restreint et qu'il n'y a pas de limite  $A_j$  sur le nombre maximum de travailleurs qu'il est possible d'allouer à l'activité  $j$ .

### 6.3.1 Objectif et contraintes usuelles

L'objectif est de minimiser  $C_{max}$  la durée totale d'exécution du projet, ce qui s'exprime sous la forme des contraintes suivantes

$$C_{max} \geq t \cdot z_{jt} \quad \forall j = 1..n \quad t \in \mathcal{T}_j. \quad (6.2)$$

Ainsi, c'est la dernière période d'exécution de l'activité  $n$  qui donne la durée du projet.

**Contraintes de précédence.** Elles garantissent qu'une activité  $j$  ne peut débiter avant que ses prédécesseurs soient tous achevés. Ces contraintes s'écrivent

$$\sum_{\tau=ES_j}^t z_{j\tau} \leq d_j^{\max} (1 - z_{it}) \quad \forall j = 1..n \quad i \in \Gamma^{-1}(j) \quad t \in \mathcal{T}_i \cap \mathcal{T}_j. \quad (6.3)$$

Soit par exemple une activité  $j$  ayant l'activité  $i$  comme prédécesseur, avec  $\mathcal{T}_i = \{2, 3, 4, 5\}$  et  $\mathcal{T}_j = \{4, 5, 6\}$ . On a deux contraintes de précédence, pour  $t = 4, 5$ . Si l'activité  $i$  est en cours d'exécution en  $t = 4$ , c'est-à-dire si  $z_{i4} = 1$  alors on doit avoir  $z_{j4} = 0$  car l'activité  $j$  ne peut être en cours d'exécution en même temps que  $i$ . Si l'activité  $i$  est en cours d'exécution en  $t = 5$ , soit  $z_{i5} = 1$  alors l'activité  $j$  ne peut être en cours d'exécution ni en même temps ni avant, soit  $z_{j4} + z_{j5} \leq 0$  ce qui implique  $z_{j4} = z_{j5} = 0$ . Si en revanche l'activité  $i$  se termine à la période 3 c'est-à-dire si  $z_{i4} = z_{i5} = 0$ , la contrainte (6.3) traduit simplement le fait que la durée d'exécution de l'activité  $j$  ne peut excéder sa durée normale  $d_j^{\max}$ .

**Contraintes de non préemption.** Elles imposent l'impossibilité d'interrompre l'exécution d'une activité en cours pour être reprise ultérieurement et s'appliquent pour toute activité d'une durée normale  $d_j^{\max} \geq 2$  et telle que  $LF_j - ES_j \geq 3$ . En effet, une activité dont la durée n'est que d'une période ne peut être interrompue puisque les périodes sont entières. Pour toute activité dont la durée normale est supérieure ou égale à deux périodes, pour qu'il y ait une possibilité d'interruption il faut alors que la longueur de l'intervalle d'exécution soit au moins égal à trois périodes. Ces contraintes de non préemption s'écrivent

$$\sum_{\tau=t+1}^{LF_j} z_{j\tau} + d_j^{\max} (z_{jt} - z_{jt+1}) \leq d_j^{\max} \quad j = 1..n \quad t = ES_j..LF_j - 3. \quad (6.4)$$

Considérons par exemple une activité  $j$  avec  $d_j^{\max} = 2$  et  $\mathcal{T}_j = \{2, 3, 4, 5\}$ . La non préemption implique que l'on ne peut avoir  $(z_{j2}, z_{j3}, z_{j4}, z_{j5}) = (1, 0, 1, 0)$  ou  $(0, 1, 0, 1)$  ou  $(1, 0, 0, 1)$  mais seulement  $(1, 1, 0, 0)$  ou  $(0, 1, 1, 0)$  ou  $(0, 0, 1, 1)$ . On a 3 contraintes du type précédent. Pour  $t = 2$  on a  $z_{j3} + z_{j4} + z_{j5} + 2(z_{j2} - z_{j3}) \leq 2$ ; pour  $t = 3$  on a  $z_{j4} + z_{j5} + 2(z_{j3} - z_{j4}) \leq 2$  et pour  $t = 4$  on a  $z_{j5} + 2(z_{j4} - z_{j5}) \leq 2$ . Les combinaisons  $(1, 0, 1, 0)$  et  $(1, 0, 0, 1)$  violent la première contrainte. La combinaison  $(0, 1, 0, 1)$  viole la deuxième contrainte.

### 6.3.2 Contraintes sur la consommation de ressources

La contrainte de respect de la quantité de main d'oeuvre permanente disponible à chaque période pour chaque type de spécialisation s'écrit

$$\sum_{j \in V_{kt}} p_{jt} \leq P_k \quad \forall k = 1..K \quad t = \alpha_k.. \beta_k, \quad (6.5)$$

avec

$$\begin{aligned}\alpha_k &= \min_{j \in V_k} ES_j \\ \beta_k &= \max_{j \in V_k} LF_j,\end{aligned}\tag{6.6}$$

et

$$V_{kt} = \{j \in V_k \mid t \in \mathcal{T}_j\}.\tag{6.7}$$

Ainsi, l'intervalle de temps  $\{\alpha_k \dots \beta_k\}$  est la réunion des périodes possibles d'exécution de toute activité nécessitant la spécialisation  $k$ .

La contrainte de satisfaction des besoins en main d'oeuvre pour l'exécution de toute activité s'écrit

$$\sum_{t \in \mathcal{T}_j} w_{jt} = W_j \quad \forall j = 1..n,\tag{6.8}$$

où  $w_{jt}$  est le nombre de travailleurs permanents ou temporaires en heures régulières de spécialisation  $k$  affectés à l'activité  $j$  en  $t$ . Cette contrainte garantit que la consommation de main d'oeuvre par l'activité  $j$  sur l'ensemble de ses périodes d'exécution couvre son contenu (besoin) en travail. Dans le cas où il n'y a pas de perte de productivité liée au recrutement de temporaires susceptibles d'induire un sureffectif, on a

$$w_{jt} = p_{jt} + u_{jt},\tag{6.9}$$

ce qui implique une relation linéaire entre consommation de main d'oeuvre et durée d'exécution de l'activité (si on double la main d'oeuvre alors la durée d'exécution sera divisée par deux). Beaucoup d'auteurs supposent une relation linéaire entre coûts d'accélération et réduction de la durée via cette relation linéaire entre réduction de la durée et augmentation des ressources (voir par exemple Roemer et Ahmadi, 2004 [124]). Cette relation linéaire traduit une productivité marginale constante.

**Bornes supérieures et inférieures sur le nombre de travailleurs en heures régulières alloués à l'exécution d'une activité à une période donnée.** Ces contraintes imposent que le nombre de travailleurs sur une activité en cours d'exécution doit être compris entre deux valeurs et s'écrivent

$$p_{jt} + u_{jt} - \bar{w}_j z_{jt} \leq 0 \quad \forall j = 1..n \quad t \in \mathcal{T}_j,\tag{6.10}$$

$$p_{jt} + u_{jt} - \underline{w}_j z_{jt} \geq 0 \quad \forall j = 1..n \quad t \in \mathcal{T}_j.\tag{6.11}$$

Si la capacité du poste n'est pas limitée alors la borne supérieure  $\bar{w}_j$  est égale à  $W_j$  qui est le nombre de travailleurs qu'il faut pour réduire la durée d'exécution à son minimum, c'est-à-dire à un jour. Comme le recrutement de travailleurs temporaires n'est pas restreint, il est possible d'affecter ce nombre  $W_j$  de travailleurs à l'exécution de l'activité  $j$ , en complétant le nombre limité de permanents de spécialisation  $k$  par des intérimaires. La borne inférieure  $\underline{w}_j$  correspond à l'allocation de travailleurs permanents en mode normal, c'est-à-dire  $\underline{w}_j = P_{kj}$ . On suppose en effet, comme dans la plupart des modélisations, qu'en mode accéléré on n'affecte pas moins de travailleurs à l'exécution d'une activité qu'en mode normal (voir notamment Naber et Kolisch, 2014 [110]).

## 6.4 Modélisation de l'impact de l'accélération sur la productivité

Le sureffectif entraîne une baisse de productivité qui résulte non seulement de l'encombrement de l'espace (les travailleurs se gênent physiquement) mais également d'une dilution de la supervision et des possibles tensions psychologiques entre les travailleurs permanents et temporaires.

### 6.4.1 Relation entre sureffectif et productivité

On note  $w_{jt}^R$  la variable intermédiaire qui exprime le nombre de travailleurs (de spécialisation  $k$ ) en heures régulières (normales) affectés à l'exécution de l'activité  $j$  en  $t$ , et corrigé de la perte de productivité liée au sureffectif.

Considérons par exemple une activité  $j$  qui requiert en mode normal  $P_{kj} = 3$  travailleurs de spécialisation  $k$  et supposons qu'à la période  $t$  on alloue  $p_{jt} = 3$  permanents et  $u_{jt} = 1$  temporaire à cette activité. Il y a donc sureffectif puisque  $p_{jt} + u_{jt} = 4 > P_{kj} = 3$ . Raisonnons en termes d'heures avec  $H_R = 8$  heures de travail régulier par jour et par travailleur. S'il n'y a pas de perte de productivité liée au sureffectif alors la totalité des 32 heures de travail en  $t$  font avancer l'exécution de l'activité. Si en revanche il y a par exemple 2,67 heures de travail inefficace sur ces 32 heures, c'est-à-dire que ces 2,67 heures ne font pas avancer l'activité parce que les travailleurs se gênent (encombrement de l'espace de travail), alors la perte de productivité que l'on note  $\delta_{jt}$  est égale à  $1 - (32 - 2,67)/32 = 0,0833$ . Cela signifie que chaque travailleur produit non pas 8 heures de travail efficace mais  $(1 - \delta_{jt}) \cdot 8 = 7,33$  heures. Compte tenu de cette perte de productivité, c'est comme si l'on avait non pas  $p_{jt} + u_{jt} = 4$  travailleurs mais  $w_{jt}^R = (1 - \delta_{jt}) \cdot (p_{jt} + u_{jt}) = (1 - 0,0833) \cdot 4 = 3,66$  travailleurs efficaces en  $t$  qui font avancer l'exécution de l'activité.

Ainsi, la main d'oeuvre corrigée des pertes de productivité sur l'activité  $j$  à la période  $t$  s'écrit

$$w_{jt}^R = (1 - \delta_{jt}) \cdot (p_{jt} + u_{jt}). \quad (6.12)$$

De manière équivalente, la perte de productivité sur toute activité  $j$  à la période  $t$  se mesure par le taux de variation  $\delta_{jt}$  de la productivité, avec

$$\delta_{jt} = 1 - \frac{w_{jt}^R}{p_{jt} + u_{jt}}. \quad (6.13)$$

A notre connaissance, seul Gunduz (2004) [65] étudie la relation entre la perte de productivité  $\delta_{jt}$  et le taux de sureffectif  $\frac{p_{jt}+u_{jt}}{P_{kj}} - 1$ . L'auteur présente plusieurs estimations de cette relation provenant d'études relativement anciennes qui concernent entre autres la fabrication de chaudières électriques. Ces études montrent toutes que la relation entre la perte de productivité  $\delta_{jt}$  et le taux de sureffectif peut être approchée par une fonction affine par morceaux dont l'équation change donc pour des valeurs de taux de sureffectif compris dans certains intervalles. En prenant l'approximation la moins favorable parmi ces différentes études, on obtient l'équation suivante

$$\delta_{jt} = \begin{cases} 0,5 \cdot \left( \frac{p_{jt}+u_{jt}}{P_{kj}} - 1 \right) & \text{si } \left( \frac{p_{jt}+u_{jt}}{P_{kj}} - 1 \right) \in ]0, 1], \\ 0,1 \cdot \left( \frac{p_{jt}+u_{jt}}{P_{kj}} - 1 \right) + 0,4 & \text{si } \left( \frac{p_{jt}+u_{jt}}{P_{kj}} - 1 \right) \in ]1, 3], \\ 0,025 \cdot \left( \frac{p_{jt}+u_{jt}}{P_{kj}} - 1 \right) + 0,625 & \text{si } \left( \frac{p_{jt}+u_{jt}}{P_{kj}} - 1 \right) \in ]3, 5]. \end{cases} \quad (6.14)$$

Ainsi la perte de productivité varie linéairement entre 0 et 50% pour un taux de sureffectif compris entre 0 et 100%. Pour un taux de sureffectif compris entre 100 et 300%, la perte de productivité est comprise entre 50 et 70% avec une relation linéaire entre les deux. Pour un taux de sureffectif compris entre 300% et 500%, la perte de productivité augmente linéairement entre 70 et 75%. On constate ainsi que ce sont les premières unités de main d'oeuvre additionnelle qui ont davantage d'impact sur la perte de productivité que les dernières, et ce, au delà d'un certain taux de sureffectif. En pratique cependant, le taux de sureffectif excède rarement 100% (doublement de la main d'oeuvre) par décision des gestionnaires ou en raison de la capacité d'accueil des postes de travail, de sorte qu'il est raisonnable de supposer la relation suivante

$$\delta_{jt} = \gamma_j \cdot \left( \frac{p_{jt} + u_{jt}}{P_{kj}} - 1 \right), \quad (6.15)$$

où  $\gamma_j$  représente le coefficient d'impact du sureffectif sur la productivité pour l'activité  $j$ . D'après l'équation (6.12), on a  $w_{jt}^R = (1 - \delta_{jt}) \cdot (p_{jt} + u_{jt})$  et en remplaçant  $\delta_{jt}$  par son expression en fonction de  $\gamma_j$  (équation (6.15)) on obtient  $w_{jt}^R \equiv w_j^R(p_{jt} + u_{jt})$  avec

$$w_j^R(p_{jt} + u_{jt}) = \begin{cases} (1 + \gamma_j) \cdot (p_{jt} + u_{jt}) - \frac{\gamma_j}{P_{kj}} \cdot (p_{jt} + u_{jt})^2 & \text{si } p_{jt} + u_{jt} > P_{kj}, \\ p_{jt} + u_{jt} & \text{sinon.} \end{cases} \quad (6.16)$$

On a donc affaire à une expression qu'il convient de linéariser, ce que nous faisons dans la sous-section 6.4.3.

A titre d'illustration, la figure 6.1 donne le tracé de la main d'oeuvre corrigée  $w_{jt}^R$  en fonction de l'effectif  $p_{jt} + u_{jt}$  (selon l'équation (6.16)) avec  $P_{kj} = 3$  comme dans l'exemple précédent et pour quatre valeurs de  $\gamma_j$ . Les points ont été artificiellement reliés pour rendre la figure plus claire mais les valeurs de l'effectif  $p_{jt} + u_{jt}$  restent discrètes.

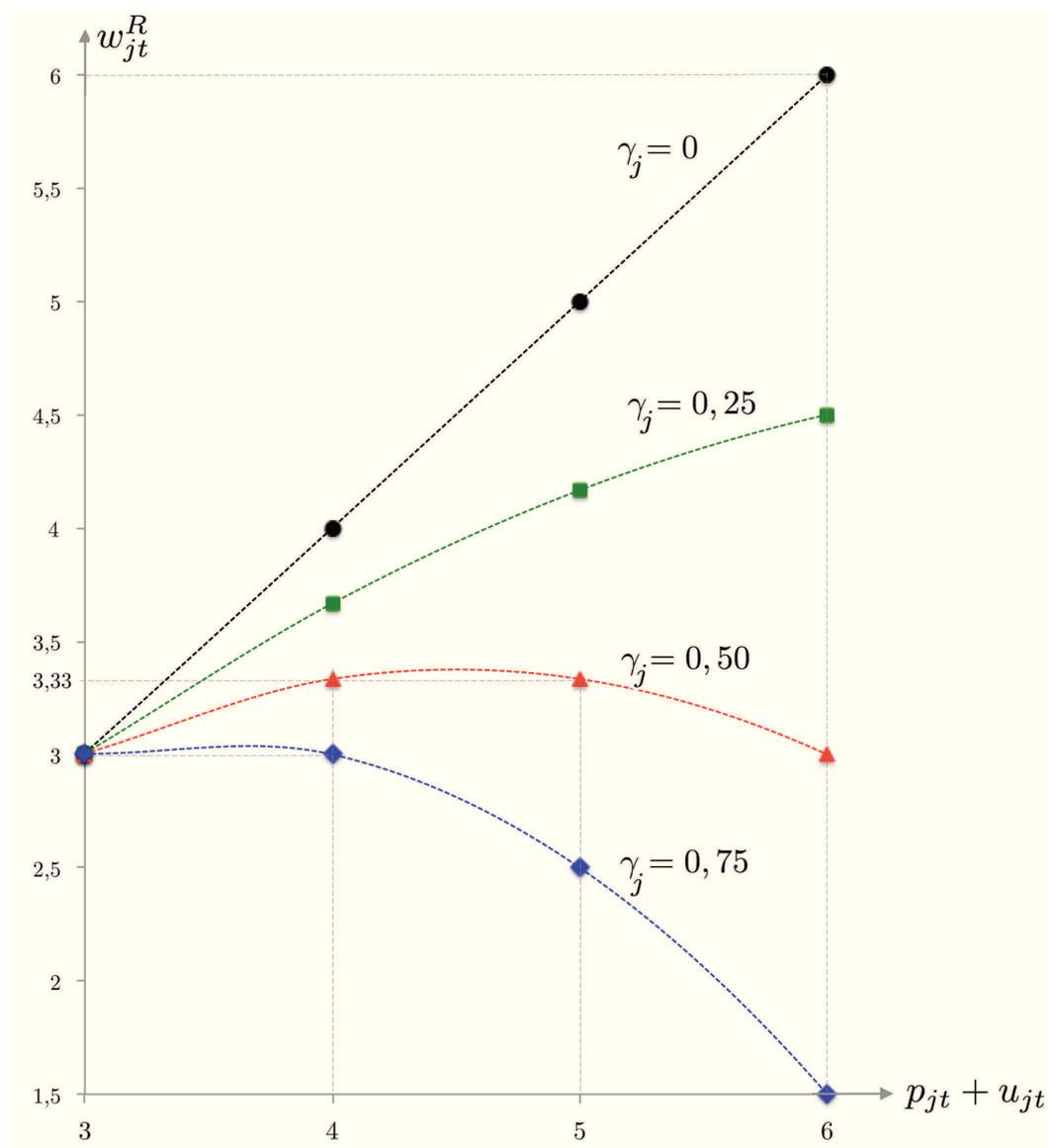


FIGURE 6.1 – Main d'oeuvre corrigée des pertes de productivité en fonction du nombre de travailleurs.

Pour  $\gamma_j = 0$ , le sureffectif n'a pas d'impact de sorte que  $w_{jt}^R = p_{jt} + u_{jt}$ . A mesure que  $\gamma_j$  augmente, la main d'oeuvre corrigée décroît pour une même valeur d'effectif  $p_{jt} + u_{jt}$ . Comme on peut le visualiser pour  $\gamma_j = 0,50$ , il existe un nombre maximum de travailleurs au delà duquel ajouter un travailleur n'apporte rien. On voit en effet

que pour  $p_{jt} + u_{jt} = 4$ , l'équivalent en main d'oeuvre efficace, soit 3,33 travailleurs est le même que pour  $p_{jt} + u_{jt} = 5$ . Du fait des pertes de productivité, on a une borne supérieure naturelle pour  $p_{jt} + u_{jt}$  qui correspond au maximum de la courbe, c'est-à-dire au point  $b_j$  qui annule la dérivée de la fonction  $w_{jt}^R \equiv w_j^R(p_{jt} + u_{jt})$  donnée dans l'équation (6.16). Ce calcul donne  $b_j = (1 + \gamma_j)P_{kj}/2\gamma_j$ , et comme  $b_j$  doit être un nombre entier, on prend  $b_j = \lfloor (1 + \gamma_j)P_{kj}/2\gamma_j \rfloor$ . Le nombre maximum de travailleurs en heures normales sur l'activité  $j$  va être donné en réalité par  $\min(b_j, W_j)$ . En effet, si l'impact  $\gamma_j$  du sureffectif sur la productivité est très faible ( $\gamma_j$  tend vers zéro) alors a priori on peut ajouter presque autant de travailleurs sur l'activité qu'on le souhaite. Mais cela ne sert à rien d'allouer à l'activité plus de travailleurs qu'il n'en faut pour compresser sa durée à un jour, ce qui est le cas si le nombre de travailleurs sur l'activité à une période donnée est égal à son contenu en travail  $W_j = d_j^{\max} \cdot P_{kj}$ . La borne supérieure  $s_j$  qui est donc l'effectif maximum s'écrit alors

$$s_j = \min\left(\left\lfloor \frac{1 + \gamma_j}{2\gamma_j} P_{kj} \right\rfloor, d_j^{\max} \cdot P_{kj}\right) \quad \forall j = 1..n. \quad (6.17)$$

Par ailleurs, la figure 6.1 montre que pour  $\gamma_j = 0,75$ , le sureffectif ne permet pas d'accélérer l'activité. Il existe ainsi une valeur limite de  $\gamma_j$  pour laquelle l'ajout d'un seul travailleur en sureffectif soit  $p_{jt} + u_{jt} = P_{kj} + 1$  produit une main d'oeuvre corrigée qui n'excède pas  $P_{kj}$ , c'est-à-dire  $w_{jt}^R = (1 + \gamma_j) \cdot (P_{kj} + 1) - \frac{\gamma_j}{P_{kj}} \cdot (P_{kj} + 1)^2 \leq P_{kj}$  ce qui donne  $\gamma_j \geq P_{kj}/(P_{kj} + 1)$ . Dans notre exemple, avec  $P_{kj} = 3$ , on obtient bien  $\gamma_j = 0,75$  qui est la valeur pour laquelle même un seul travailleur en sureffectif n'accélère pas l'activité.

Si l'entreprise détermine elle-même un nombre maximum  $A_j$  de travailleurs en heures régulières qu'il est possible d'allouer à l'activité  $j$ , et ce, notamment en raison d'une capacité pré-déterminée du poste de travail sur lequel l'activité  $j$  est exécutée, alors l'effectif maximum  $s_j$  est donné par

$$s_j = \min\left(\left\lfloor \frac{1 + \gamma_j}{2\gamma_j} P_{kj} \right\rfloor, d_j^{\max} \cdot P_{kj}, A_j\right), \quad \gamma_j \in [0, P_{kj}/(P_{kj} + 1)[ \quad \forall j = 1..n. \quad (6.18)$$

Dans la sous-section suivante nous approfondissons l'étude des bornes sur la main d'oeuvre en introduisant notamment une contrainte supplémentaire sur le recrutement de travailleurs temporaires.

#### 6.4.2 Etude des bornes sur la main d'oeuvre en heures régulières

Jusque là, nous avons supposé que le recrutement de travailleurs temporaires n'est pas limité en nombre pas plus qu'il ne l'est en pourcentage maximum dans l'effectif total alloué à l'exécution d'une activité. Il est alors possible que le processus d'optimisation conduise à n'affecter que des temporaires à une activité à un moment donné, ce qui peut être discutable. Avec la contrainte suivante, nous intégrons la possibilité que le gestionnaire de projet souhaite une proportion maximum  $\lambda \in ]0, 1[$  de temporaires dans l'effectif alloué à toute activité et à toute période, soit

$$u_{jt} \leq \frac{\lambda}{1-\lambda} p_{jt} \quad \forall j = 1..n \quad t \in \mathcal{T}_j. \quad (6.19)$$

Si  $\lambda = 1$  la contrainte précédente n'existe pas, on peut recruter autant de temporaires que la borne  $s_j$  le permet. Si  $\lambda = 0$  cela signifie qu'aucun recrutement de temporaire n'est possible de sorte que l'accélération du projet ne peut se faire qu'avec des heures supplémentaires. La contrainte précédente s'écrit alors  $u_{jt} \leq 0$  et permet donc d'affecter une valeur nulle au nombre de temporaires.

**Borne sur la main d'oeuvre permanente.** A toute période possible d'exécution, la main d'oeuvre permanente affectée à une activité est bornée par la main d'oeuvre permanente totale disponible dans l'entreprise ou par l'effectif maximum sur l'activité, c'est-à-dire

$$p_{jt} \leq \min(s_j, P_k) = \bar{p}_j \quad \forall j \in V_k \quad t \in \mathcal{T}_j \quad k = 1..K. \quad (6.20)$$

**Borne sur la main d'oeuvre temporaire.** Si  $\lambda = 0$ , aucun temporaire n'est recruté de sorte que la borne supérieure sur la main d'oeuvre temporaire vaut zéro. Lorsqu'il n'y a pas de contrainte sur une proportion maximum de permanents, c'est-à-dire lorsque  $\lambda = 1$  alors le nombre de travailleurs temporaires est borné par l'effectif maximum  $s_j$  (si à une période donnée et pour une activité donnée il n'y a aucun permanent, le nombre de temporaires ne devra de toute façon pas excéder la limite  $s_j$ ). Pour  $\lambda \in ]0, 1[$  les temporaires sont limités à un pourcentage de permanents qui sont eux-mêmes bornés par  $\bar{p}_j$ ; la borne sur les temporaires est alors obtenue en posant  $p_{jt} = \bar{p}_j$  et en utilisant la contrainte (6.19) pour mettre les temporaires à leur maximum, soit  $\frac{\lambda}{1-\lambda} p_{jt} = u_{jt}$ . On obtient finalement

$$u_{jt} \leq \bar{u}_j \quad \text{avec } \bar{u}_j = \begin{cases} 0 & \text{si } \lambda = 0 \\ \min(s_j, \lfloor \frac{\lambda}{1-\lambda} \bar{p}_j \rfloor) & \text{si } \lambda \in ]0, 1[ \\ s_j & \text{si } \lambda = 1 \end{cases} \quad \forall j = 1..n \quad t \in \mathcal{T}_j. \quad (6.21)$$

**Borne supérieure sur la main d'oeuvre régulière.** Des bornes (6.20) et (6.21) on peut en déduire une nouvelle borne supérieure  $\bar{w}_j$  pour la somme  $p_{jt} + u_{jt}$  qui constitue la main d'oeuvre totale en heures régulières pour l'activité  $j$  en  $t$ . Contrairement à  $s_j$ , cette borne  $\bar{w}_j$  dépend de la proportion maximum de temporaire et s'écrit

$$\bar{w}_j = \min(\bar{p}_j + \bar{u}_j, s_j) \quad \forall j = 1..n \quad t \in \mathcal{T}_j. \quad (6.22)$$

**Etude des différentes situations.** Il est possible d'obtenir une expression plus simple de la borne  $\bar{w}_j = \min(\bar{p}_j + \bar{u}_j, s_j)$  en analysant la valeur qu'elle prend en fonction des valeurs de  $s_j$  par rapport au nombre total de permanents  $P_k$  et

selon les différentes valeurs de proportions de temporaires. On rappelle que l'on a  $\bar{p}_j = \min(s_j, P_k)$  et  $\bar{u}_j = \min(s_j, \lfloor \frac{\lambda}{1-\lambda} \bar{p}_j \rfloor)$ .

Considérons d'abord le cas où  $s_j \leq P_k$ . On a alors  $\bar{p}_j = s_j$  et donc  $\bar{u}_j = \min(s_j, \lfloor \frac{\lambda}{1-\lambda} s_j \rfloor)$  avec  $0 \leq \bar{u}_j \leq s_j$  quelle que soit la proportion  $\lambda$ . On obtient alors  $s_j \leq \bar{p}_j + \bar{u}_j \leq 2s_j$  ce qui implique  $\bar{w}_j = s_j \quad \forall \lambda \in [0, 1]$ .

Dans le cas contraire, c'est-à-dire si  $s_j > P_k$  alors  $f\bar{p}_j = P_k$  ce qui donne  $\bar{u}_j = \min(s_j, \lfloor \frac{\lambda}{1-\lambda} P_k \rfloor)$ . Discutons la valeur de  $\bar{u}_j$  selon celle de  $\lambda$ .

- Si  $\lambda = 0$  alors  $\bar{u}_j = 0$  et donc  $\bar{w}_j = \min(P_k, s_j) = P_k$ .
- Si  $\lambda = 1$  alors  $\bar{u}_j = s_j$  et donc  $\bar{w}_j = \min(P_k + s_j, s_j) = s_j$
- Si  $\lambda \in ]0, 1[$  alors  $\bar{u}_j = \min(s_j, \lfloor \frac{\lambda}{1-\lambda} P_k \rfloor)$  avec  $P_k < \bar{u}_j < s_j$  nécessairement.

On obtient  $\bar{w}_j = \min(P_k + \min(s_j, \lfloor \frac{\lambda}{1-\lambda} P_k \rfloor), s_j)$  avec  $\bar{w}_j \leq s_j$ . En effet puisque  $P_k < \bar{u}_j < s_j$  alors  $2P_k < \bar{p}_j + \bar{u}_j < s_j + P_k$  de sorte que  $\bar{p}_j + \bar{u}_j$  prend pour valeur maximale  $s_j + P_k - 1$  ce qui donne  $\bar{w}_j = \min(s_j + P_k - 1, s_j) = s_j$  quelle que soit la valeur de  $P_k$ , celle-ci étant toujours au moins égale à 1.

Le tableau 6.1 résume les différentes expressions de  $\bar{w}_j$  que nous venons d'obtenir.

$s_j = \min(\lfloor \frac{1+\gamma_j}{2\gamma_j} P_{kj} \rfloor, d_j^{\max} \cdot P_{kj}, A_j)$ d'après l'éq. (6.18)		
$\lambda = 0$	$s_j > P_k$	$s_j \leq P_k$
$\bar{w}_j = P_k$	$\bar{w}_j = \min(P_k + \min(s_j, \lfloor \frac{\lambda}{1-\lambda} P_k \rfloor), s_j)$	$\bar{w}_j = s_j$
$\lambda \in ]0, 1[$		
$\lambda = 1$	$\bar{w}_j = s_j$	

Tableau 6.1 – Expressions de la borne supérieure  $\bar{w}_j$  sur l'effectif dans tous les cas de figure

Dans tous les cas de figure, on en conclut que l'on a nécessairement  $\bar{w}_j \leq s_j$  soit

$$\bar{w}_j \leq \min(\lfloor \frac{1+\gamma_j}{2\gamma_j} P_{kj} \rfloor, d_j^{\max} \cdot P_{kj}, A_j) \quad \forall j = 1..n. \quad (6.23)$$

### 6.4.3 Linéarisation de l'équation de la main d'oeuvre corrigée

Pour linéariser la fonction  $w_{jt}^R \equiv w_j^R(p_{jt} + u_{jt})$  donnée par l'équation (6.16) où les valeurs possibles de l'effectif  $p_{jt} + u_{jt}$  sont dans l'intervalle  $\{0, 1, 2, \dots, \bar{w}_j\}$ , on note  $y_{jtl}$  la variable binaire qui vaut 1 si  $p_{jt} + u_{jt}$  prend la valeur  $l$  avec  $l \in \{0, \dots, \bar{w}_j\}$  et 0 sinon. La variable  $p_{jt} + u_{jt}$  peut alors s'écrire comme suit

$$p_{jt} + u_{jt} = \sum_{l=0}^{\bar{w}_j} l \cdot y_{jtl} \quad \forall j = 1..n \quad t \in \mathcal{T}_j, \quad (6.24)$$

ce qui constitue une nouvelle contrainte à laquelle on ajoute la contrainte suivante pour s'assurer que le nombre de travailleurs en heures régulières prend une et une seule valeur, soit

$$\sum_{l=0}^{\bar{w}_j} y_{jtl} = 1 \quad \forall j = 1..n \quad t \in \mathcal{T}_j. \quad (6.25)$$

Il suffit ensuite de calculer pour chaque valeur possible  $l \in \{0, \dots, \bar{w}_j\}$  que prend la somme des variables  $p_{jt} + u_{jt}$ , la valeur de la fonction de main d'oeuvre corrigée  $w_j^R(l)$  donnée en (6.16). La fonction  $w_{jt}^R \equiv w_j^R(p_{jt} + u_{jt})$  s'écrit désormais sous une forme linéaire comme suit

$$w_{jt}^R = \sum_{l=0}^{\bar{w}_j} w_j^R(l) \cdot y_{jtl} \quad \forall j = 1..n \quad t \in \mathcal{T}_j. \quad (6.26)$$

Cette méthode de linéarisation est appropriée dans notre cas où les valeurs possibles de la main d'oeuvre régulière  $p_{jt} + u_{jt}$  sont discrètes et bornées, avec une borne supérieure relativement petite. Ainsi, pour chaque valeur  $l = 0.. \bar{w}_j$  nous calculerons préalablement  $w_j^R(l)$  conformément à l'équation (6.16). Ces valeurs seront stockées dans un vecteur pour chaque activité  $j$  de sorte que  $w_{jt}^R$  prendra la valeur  $w_j^R(l)$  si  $y_{jtl} = 1$ .

L'expression de  $w_{jt}^R$  dans (6.26) remplace  $w_{jt}$  dans la contrainte (6.8) et la contrainte (6.10) utilise la nouvelle définition de la borne  $\bar{w}_j$  donnée en (6.22).

Notons que la contrainte (6.8) ne s'exprime pas comme une égalité stricte entre le volume total de main d'oeuvre affecté à l'activité et corrigé des pertes de productivité et le contenu requis en main d'oeuvre  $W_j$  parce qu'il y a peu de chance que cette main d'oeuvre corrigée soit exactement égale à  $W_j$ . Pour illustrer, considérons une activité  $j$  avec  $P_{kj} = 3$ ,  $d_j^{\max} = 8$ ,  $\gamma_j = 0,0857$  et un sureffectif de 4 travailleurs qui permettent de réduire la durée à  $d_j = 7$  jours. Imaginons que le sureffectif se concentre sur la première période d'exécution, ce qui correspond à l'allocation  $(p_{jt} + u_{jt})_{t=1..7} = (7, 3, 3, 3, 3, 3, 3)$ . Avec l'équation (6.16), on peut calculer la main d'oeuvre totale corrigée qui en résulte :  $w_j^R(7) + 18 = 6,2 + 18 = 24,2$  même si nous n'avons besoin que de  $W_j = 24$  hommes - jour. Si l'on imposait la stricte égalité entre main d'oeuvre totale corrigée et  $W_j$ , dès lors que  $\gamma_j > 0$ , il n'y aurait que de très rares cas où le programme d'optimisation pourrait accélérer l'activité avec un sureffectif.

## 6.5 Ajout des heures supplémentaires

Cette section débute par une analyse des difficultés que pose la prise en compte d'un sureffectif de travailleurs en heures supplémentaires. Cette analyse nous amène alors à supposer que les travailleurs en heures supplémentaires ne sont jamais en sureffectif. Nous donnons ensuite les contraintes relatives aux heures supplémentaires qui viennent s'ajouter à notre modèle.

On rappelle que  $H_R$  désigne le nombre d'heures normales (régulières) de travail par jour ;  $H_O$  le nombre maximum d'heures supplémentaires par travailleur et par jour et  $H_{OW}$ , le nombre maximum d'heures supplémentaires par travailleur et par semaine au delà duquel la productivité du travail en heures supplémentaires décroît.

### 6.5.1 Analyse du sureffectif de travailleurs en heures supplémentaires

On introduit la variable entière  $o_{jt}$  pour désigner le nombre de travailleurs en heures supplémentaires (de spécialisation  $k$ ) alloués à l'activité  $j$  en  $t$  et on rappelle que  $h_{jt}$  est le nombre d'heures supplémentaires affectées à l'activité  $j$  en  $t$  (variable continue). Le nombre  $o_{jt}$  de travailleurs en heures supplémentaires se déduit du nombre des heures supplémentaires comme suit :  $o_{jt} = \lceil h_{jt}/H_O \rceil$ , ce qui se traduit par

$$h_{jt}/H_O \leq o_{jt} \leq 1 + h_{jt}/H_O \quad \forall j = 1..n \quad t \in \mathcal{T}_j. \quad (6.27)$$

En l'absence de sureffectif des travailleurs en heures supplémentaires, le travail en heures supplémentaires qui s'ajoute à  $w_{jt}^R$  et exprimé dans la même unité, à savoir en équivalent homme-jour est égal à  $h_{jt}/H_R$ . Prenons par exemple  $H_R = 8$  et  $H_O = 4$  et imaginons que  $h_{jt} = 14$ . On a  $o_{jt} = \lceil 14/4 \rceil = 4$  travailleurs en heures supplémentaires et un équivalent homme-jour rapporté en heures régulières de  $14/8 = 1,75$ . Autrement dit, 14 heures de travail en heures supplémentaires sont équivalentes au travail de 1,75 travailleurs en heures régulières.

Il y a sureffectif des travailleurs en heures supplémentaires à la période  $t$  si  $o_{jt} > P_{kj}$  et le taux de sureffectif en  $t$  s'écrit  $\frac{o_{jt}}{P_{kj}} - 1$ . On note alors  $w_{jt}^O$  la variable intermédiaire qui exprime le nombre de travailleurs (de spécialisation  $k$ ) en heures supplémentaires (indice  $O$  pour overtime) affecté à l'exécution de l'activité  $j$  en  $t$ , et corrigé des pertes de productivité liées au sureffectif. On a la relation

$$w_{jt}^O = (1 - \delta_{jt}^O) \frac{1}{H_R} h_{jt}, \quad (6.28)$$

avec  $\delta_{jt}^O = \max\left(0, \gamma_j \cdot \left(\frac{o_{jt}}{P_{kj}} - 1\right)\right)$  si l'on suppose que le sureffectif résultant de travailleurs supplémentaires a même impact  $\gamma_j$  sur la productivité que le sureffectif en travailleurs en heures régulières. De façon analogue à l'équation (6.16), cette relation s'écrit

$$w_{jt}^O = \begin{cases} (1 + \gamma_j) \frac{1}{H_R} h_{jt} - \frac{\gamma_j}{P_{kj}} \frac{1}{H_R} h_{jt} o_{jt} & \text{si } o_{jt} > P_{kj}, \\ \frac{1}{H_R} h_{jt} & \text{sinon.} \end{cases} \quad (6.29)$$

L'introduction de cette équation dans le modèle nécessiterait la linéarisation du produit des variables  $h_{jt} \cdot o_{jt}$  et donc la création de variables et contraintes additionnelles qui viendraient s'ajouter à un programme déjà complexe. De plus, cette estimation des pertes de productivité sur le travail en heures supplémentaires souffre d'un biais lié au fait que chaque travailleur en heures supplémentaires ne travaille pas nécessairement le maximum autorisé à savoir  $H_O$  heures. En reprenant l'exemple précédent avec  $h_{jt} = 14$  et  $P_{kj} = 3$ , la répartition qui minimise le sureffectif consiste à faire travailler le maximum de travailleurs pendant  $H_O = 4$  heures. On

aura alors 3 travailleurs pendant 4 heures et 1 travailleur pendant 2 heures, c'est-à-dire en termes de présence, 4 travailleurs de 16 à 18h et 3 travailleurs de 18 à 20h. Entre 18 et 20h, il n'y a donc pas de sureffectif, seulement entre 16 et 18h. La baisse de productivité de 13,33% va donc s'appliquer non pas sur la totalité des heures supplémentaires mais sur les 8 heures de travail accompli entre 18 et 20h. On aura finalement  $w_{jt}^O = (1 - 0,1332)(\frac{1}{8}8) + \frac{6}{8} = 1,6168$  tandis que l'expression (6.28) indique que  $w_{jt}^O = (1 - 0,1332)(\frac{14}{8}) = 1,5169$ . C'est donc encore plus compliqué à formuler car cela fait intervenir des calculs de durée précise durant laquelle il y a sureffectif.

Ainsi, pour simplifier le modèle, on suppose que les travailleurs en heures supplémentaires ne sont jamais en sureffectif, c'est-à-dire que ce nombre n'excède pas la capacité normale  $P_{kj}$ . Comme il n'y a pas de sureffectif à modéliser, la variable  $o_{jt}$  qui désigne le nombre de travailleurs en heures supplémentaires n'est plus nécessaire. Cette hypothèse a du sens car en raison du nombre restreint de permanents pour chaque type de spécialisation et de la contrainte légale sur les heures supplémentaires, il est peu vraisemblable qu'il y ait du sureffectif en heures supplémentaires. Cette hypothèse se traduit par la contrainte suivante

$$h_{jt} \leq H_O P_{kj} \quad \forall j = 1..n \quad t \in \mathcal{T}_j. \quad (6.30)$$

Sous cette hypothèse  $w_{jt} = w_{jt}^R + w_{jt}^O$  s'écrit finalement

$$w_{jt} = \sum_{l=0}^{\bar{w}_j} w_j^R(l) \cdot y_{jtl} + \frac{1}{H_R} h_{jt} \quad \forall j = 1..n \quad t \in \mathcal{T}_j. \quad (6.31)$$

Cette expression de  $w_{jt}$  est désormais celle qui doit être incluse dans la contrainte (6.8).

### 6.5.2 Contraintes additionnelles relatives à l'utilisation des heures supplémentaires

Les travailleurs en heures supplémentaires ne doivent pas occuper plus d'espace que les travailleurs en heures normales ce qui a du sens car espace, outils, ou postes de travail ont été organisés en début de journée pour accueillir un certain nombre de travailleurs. On impose ainsi la contrainte suivante

$$h_{jt}/H_O \leq p_{jt} + u_{jt} \quad \forall j = 1..n \quad t \in \mathcal{T}_j. \quad (6.32)$$

Cette contrainte garantit surtout que si  $p_{jt} + u_{jt} = 0$  alors  $h_{jt} = 0$  nécessairement, et si  $h_{jt} > 0$  alors  $p_{jt} + u_{jt} \geq 1$ . Cela permet un calcul correct de la durée. Imaginons par exemple que deux travailleurs ayant préalablement travaillé en heures normales sur d'autres activités que  $j$  viennent simultanément travailler pendant  $H_O = 4$  heures supplémentaires sur l'activité  $j$  en  $t$  tandis qu'aucun permanent n'y travaillait durant  $t$ . Cette situation se traduit par  $p_{jt} + u_{jt} = 0$  et  $h_{jt}/H_O = 2$ . Avec  $H_R = 8$ ,

le contenu en travail de l'activité  $j$  en  $t$  est égal  $w_{jt} = 0 + h_{jt}/H_R = 1$  ce qui veut dire que la durée du travail sur l'activité est d'un jour, et pourtant concrètement le temps passé en  $t$  sur cette activité est d'une demi-journée (avec deux travailleurs travaillant simultanément). Donc pour comptabiliser correctement la durée sur une activité, il faut qu'il y ait au moins un travailleur à plein temps sur l'activité en  $t$ , c'est-à-dire si  $h_{jt} > 0$  alors  $p_{jt} + u_{jt} \geq 1$ .

Par hypothèse, seuls les travailleurs permanents font des heures supplémentaires (hypothèse standard que l'on trouve notamment dans Heimerl et Kolisch, 2010 [72]), il faut donc qu'à chaque période, il y ait suffisamment de permanents pour assurer les heures supplémentaires  $h_{jt}$ , c'est-à-dire

$$\frac{1}{H_O} \sum_{j \in V_{kt}} h_{jt} \leq \sum_{j \in V_{kt}} p_{jt} \quad \forall k = 1..K \quad t = \alpha_k.. \beta_k. \quad (6.33)$$

Si une période donnée  $t$  correspond à une période possible d'exécution pour une seule activité  $j$  dans tout le projet alors la contrainte précédente s'écrit  $h_{jt}/H_O \leq p_{jt}$  et même si elle contredit la contrainte (6.32), elle permet d'éviter la situation où  $p_{jt} = 0$  et  $u_{jt} > 0$  qui peut aboutir à  $h_{jt} > 0$  alors qu'aucun permanent n'est actif. C'est une situation rare qui peut toutefois se produire parce que l'on ne cherche pas à minimiser le coût dans un premier temps, alors tant que la contrainte de budget est respectée, l'optimisation peut tout à fait conduire à recruter un temporaire et à ne pas utiliser un permanent disponible.

On note que dans la contrainte précédente, le membre de droite ne peut être remplacé par le paramètre  $P_k$  car il faut considérer le nombre de permanents possiblement actifs durant la période  $t$  puisque pour travailler en heures supplémentaires, le travailleur doit déjà avoir effectué une journée en heures normales, par définition.

Remarquons enfin que la contrainte précédente autorise le déplacement des travailleurs de spécialisation  $k$  d'une activité vers une autre ; ce n'est pas nécessairement le permanent qui aura travaillé sur  $j$  en  $t$  qui fera les heures supplémentaires sur cette activité. Si l'on avait supposé que seuls les permanents qui ont travaillé sur  $j$  en  $t$  peuvent faire les heures supplémentaires sur cette activité, alors la contrainte précédente serait devenue  $h_{jt}/H_O \leq p_{jt}$  pour tout  $j$  et pour tout  $t$  ce qui aurait démultiplié le nombre de contraintes (car  $n > K$  en pratique) tout en rendant le modèle moins flexible. Avec cette contrainte (6.33) on reste plus général avec une flexibilité des permanents sur différentes activités dans la même journée. Pour un travailleur, le fait de changer d'activité dans une journée peut limiter sa fatigue associée à une durée plus longue sur une même activité. D'un autre côté, ce changement à la fin d'une journée de travail régulier peut causer une perte de productivité liée au temps d'adaptation à l'exécution de cette nouvelle activité. L'un dans l'autre, on dira que la mobilité des permanents d'une activité vers une autre est sans effet sur la productivité.

En outre, on impose une contrainte de non dépassement d'un certain nombre d'heures supplémentaires par semaine pour éviter les pertes de productivité liées au simple fait d'utiliser des heures supplémentaires (et non liées au sureffectif). On

impose pas plus de  $H_{OW}$  heures par semaine et par travailleur permanent, sur une base de 5 jours ouvrés. La contrainte s'écrit

$$\sum_{\tau=t}^{t+4} \sum_{j \in V_{k\tau}} h_{j\tau} \leq \frac{1}{5} H_{OW} \sum_{\tau=t}^{t+4} \sum_{j \in V_{k\tau}} p_{j\tau} \quad \forall k = 1..K \quad t = \alpha_k.. \beta_k - 4 \quad (6.34)$$

Pour toute spécialisation  $k$  telle que  $\beta_k - \alpha_k < 5$  alors il n'y a qu'une seule contrainte qui s'écrit

$$\sum_{\tau=\alpha_k}^{\beta_k} \sum_{j \in V_{k\tau}} h_{j\tau} \leq \frac{\beta_k - \alpha_k}{5} H_{OW} \sum_{\tau=\alpha_k}^{\beta_k} \sum_{j \in V_{k\tau}} p_{j\tau}. \quad (6.35)$$

Les deux contraintes précédentes sont globales. Pour respecter plus strictement les contraintes légales qui autorisent tout travailleur à ne travailler qu'un certain nombre d'heures supplémentaires par semaine, il aurait fallu définir la variable  $o_{ljt} = 1$  si le travailleur  $l$  est affecté à l'activité  $j$  en  $t$ , avec  $l = 1..P_k$  où  $P_k$  est le nombre de travailleurs de spécialisation  $k$  mais cela aurait impliqué une multiplication des variables et des ajouts de contraintes pour relier chaque travailleur à l'exécution de chaque activité, en heures normales et en heures supplémentaires. On ne se situe pas au niveau opérationnel dans cette modélisation mais au niveau tactique (planification du nombre d'intérimaires et d'heures supplémentaires).

Il est en réalité fort probable que la moyenne des travailleurs actifs sur une semaine soit égale à l'effectif  $P_k$  car nous calibrerons l'effectif en mode normal pour que toute la main d'oeuvre soit active à chaque période de sorte qu'il suffira finalement d'imposer la contrainte suivante

$$\sum_{\tau=t}^{t+4} \sum_{j \in V_{k\tau}} h_{j\tau} \leq H_{OW} \cdot P_k \quad \forall k = 1..K \quad t = \alpha_k.. \beta_k - 4. \quad (6.36)$$

Une borne supérieure pour le cumul des heures supplémentaires au cours d'une période donnée et pour chaque spécialisation est donnée par  $H_O P_k$ , ainsi on a

$$\sum_{j \in V_{kt}} h_{jt} \leq H_O \cdot P_k \quad \forall k = 1..K \quad t = \alpha_k.. \beta_k. \quad (6.37)$$

## 6.6 Modélisation de l'impact de l'accélération sur la qualité

En mode normal c'est-à-dire lorsque la durée d'exécution de l'activité est égale à  $d_j^{\max}$  on obtient une qualité normale que l'on note  $q_j^{\max}$ . On rappelle que  $\Delta q_j^O$  désigne la perte de qualité associée à l'utilisation d'une heure supplémentaire et  $\Delta q_j^S$  la perte de qualité résultant de l'emploi d'un travailleur en sureffectif par jour.

Nous supposons que l'impact sur la qualité des heures supplémentaires est supérieur à celui des heures en sureffectif de sorte que  $\Delta q_j^S < H_R \cdot \Delta q_j^O$  (voir Li et al., 2000 [96]). La perte de qualité  $\Delta q_{jt}$  pour l'activité  $j$  à la période  $t$  peut se formuler de la manière suivante

$$\Delta q_{jt} = \begin{cases} \Delta q_j^S \cdot (p_{jt} + u_{jt} - P_{kj}) + \Delta q_j^O \cdot h_{jt} & \text{si } p_{jt} + u_{jt} > P_{kj}, \\ \Delta q_j^O \cdot h_{jt} & \text{sinon.} \end{cases} \quad (6.38)$$

Ainsi, la perte de qualité par période d'exécution est croissante linéairement avec le volume du sureffectif et des heures supplémentaires et la qualité  $q_j$  d'une activité  $j$  se définit par

$$q_j = q_j^{\max} - \sum_{t \in \mathcal{T}_j} \Delta q_{jt} \quad \forall j = 1..n. \quad (6.39)$$

La question que l'on se pose est de savoir comment caractériser la relation entre la durée d'exécution d'une activité et sa qualité. Pour répondre à cette question, nous allons analyser successivement plusieurs cas d'accélération, avec des heures supplémentaires et du sureffectif, combinés ou non, en considérant l'absence ou la présence de pertes de productivité. Dans les illustrations numériques nous supposons  $q_j^{\max} = 1$ , comme le font bon nombre d'auteurs (Fu et Zhang, 2016 [60]; Babu et Suresh, 1996 [14] et Khang et Myint, 1999 [88]).

### 6.6.1 Accélération avec heures supplémentaires uniquement

Rappelons tout d'abord que la durée en mode normal se définit par  $d_j^{\max} = W_j/P_{kj}$ . Cette durée s'obtient notamment en allouant à chaque période un même nombre  $P_{kj}$  de permanents sur  $d_j^{\max}$  périodes. En ajoutant un volume  $h_j$  d'heures supplémentaires à chaque période d'exécution de l'activité, il est possible d'atteindre une durée  $d_j$  définie par  $d_j = \lceil W_j/(P_{kj} + h_j/H_R) \rceil$ , soit  $d_j = \lceil d_j^{\max} \cdot P_{kj}/(P_{kj} + h_j/H_R) \rceil$ . De façon analogue si l'on fixe  $d_j$  à une valeur souhaitée alors le volume d'heures supplémentaires que l'on doit allouer à chaque période d'exécution pour atteindre cette durée doit être égal à  $h_j = P_{kj} \cdot H_R \cdot (d_j^{\max} - d_j) / d_j$ , ce qui représente un volume total  $d_j \cdot h_j$  d'heures supplémentaires, soit encore  $P_{kj} \cdot H_R \cdot (d_j^{\max} - d_j)$ . Toute allocation  $(h_{jt})_{t=1..d_j}$  telle que  $\sum_{t=1..d_j} h_{jt} = P_{kj} \cdot H_R \cdot (d_j^{\max} - d_j)$  et respectant les contraintes de disponibilité et de non sureffectif en heures supplémentaires permet d'obtenir une durée  $d_j$  avec une qualité identique  $q_j = 1 - \Delta q_j^O \cdot \sum_{t=1..d_j} h_{jt}$  soit

$$q_j = 1 - \Delta q_j^O \cdot P_{kj} \cdot H_R \cdot (d_j^{\max} - d_j) \quad \text{si } p_{jt} + u_{jt} = P_{kj} \quad \forall t = 1..d_j \quad \forall j = 1..n. \quad (6.40)$$

On observe en effet que la qualité ne dépend pas de la répartition des heures supplémentaires entre les différentes périodes mais seulement du volume total utilisé de ces heures sur l'ensemble des  $d_j$  périodes. La relation (6.40) exprime la qualité comme une fonction linéaire de la durée, comme le supposent Babu et Suresh (1996)

[14] et Khang et Myint (1999) [88]. Cette relation vaut quelle que soit la valeur de  $\gamma_j$  puisqu'il n'y a pas de sureffectif en heures supplémentaires.

Pour illustrer, considérons une activité avec  $P_{kj} = 3$ ,  $d_j^{\max} = 8$  et  $\Delta q_j^O = 0,00625$ . Si l'on souhaite réduire la durée de l'activité à  $d_j = 7$  jours, il faut  $(d_j^{\max} - d_j) \cdot P_{kj} \cdot H_R = (8 - 7) \cdot 3 \cdot 8 = 24$  heures supplémentaires réparties sur 7 jours, en n'excédant pas la limite de  $H_O \cdot P_{kj} = 4 \cdot 3 = 12$  heures par jour. Par exemple, les allocations  $(h_{jt})_{t=1..d_j} = (12, 12, 0, 0, 0, 0, 0)$  ou  $(3.43, 3.43, 3.43, 3.43, 3.43, 3.43, 3.43)$  peuvent convenir car elles sont toutes deux telles que  $\sum_{t=1..7} h_{jt} = 24$ . Elles donnent même qualité  $q_j = q_j = 1 - 24 \cdot 0,00625 = 0,85$ . En accélérant au maximum l'activité en utilisant uniquement des heures supplémentaires, on atteint une durée compressée  $\underline{d}_j^O = \lceil d_j^{\max} / (1 + H_O / H_R) \rceil = \lceil 8 / [(1 + 4/8)] \rceil = 6$  qui s'obtient avec le volume juste nécessaire  $H_R \cdot (W_j - \underline{d}_j^O \cdot P_{kj}) = 8 \cdot (24 - 6 \cdot 3) = 48$  heures supplémentaires à répartir sur les 6 jours. La qualité s'établit à une valeur  $\underline{q}_j^O = 1 - 48 \cdot \Delta q_j^O = 0,70$ .

## 6.6.2 Accélération avec sureffectif uniquement

**Accélération avec sureffectif seul en l'absence de pertes de productivité.** Considérons maintenant le cas d'une accélération de l'activité avec le sureffectif seulement en commençant l'analyse sous l'hypothèse où il n'y a pas de perte de productivité, c'est-à-dire  $\gamma_j = 0$ . Dans notre modèle, il est possible d'affecter un nombre différent  $p_{jt} + u_{jt}$  de travailleurs en heures régulières d'une période à l'autre mais supposons dans un premier temps que ce nombre est identique soit  $p_{jt} + u_{jt} = p_j + u_j$  pour  $t = 1..d_j$ . On a alors la relation  $d_j = \lceil d_j^{\max} \cdot P_{kj} / (p_j + u_j) \rceil$  ce qui donne  $p_j + u_j = P_{kj} \cdot d_j^{\max} / d_j$  en laissant momentanément de côté les contraintes d'intégralité. Pour une durée  $d_j < d_j^{\max}$  on aura nécessairement  $p_j + u_j > P_{kj}$ . Le sureffectif par période est alors  $p_j + u_j - P_{kj}$  ce qui donne un volume de main d'oeuvre additionnelle égale à  $d_j \cdot (p_j + u_j - P_{kj})$  soit encore un nombre total de  $P_{kj} \cdot (d_j^{\max} - d_j)$  travailleurs en sureffectif pour atteindre la durée  $d_j$  avec une qualité

$$q_j = 1 - \Delta q_j^S \cdot P_{kj} \cdot (d_j^{\max} - d_j) \quad \text{si } h_{jt} = 0 \quad \forall t = 1..d_j \quad \text{et } \gamma_j = 0 \quad \forall j = 1..n. \quad (6.41)$$

Dans ce cas, la qualité est une fonction linéaire croissante de la durée. Un tel couple durée qualité  $(d_j, q_j)$  s'obtient pour toute allocation  $(p_{jt} + u_{jt})_{t=1..d_j}$  telle que  $\sum_{t=1..d_j} (p_{jt} + u_{jt} - P_{kj}) = P_{kj} \cdot (d_j^{\max} - d_j)$  à condition qu'elle soit réalisable, c'est-à-dire  $p_{jt} + u_{jt} \leq \bar{w}_j$  et  $\sum_{t=1..d_j} p_{jt} + u_{jt} \leq \bar{w}_j$ .

**Accélération avec sureffectif seul et pertes de productivité.** Analysons maintenant la situation sans heures supplémentaires où il y a cette fois-ci des pertes de productivité liées au sureffectif. Ainsi, si  $\gamma_j > 0$ , on a la relation  $d_j = \lceil d_j^{\max} \cdot P_{kj} / w_j^R(p_j + u_j) \rceil$  avec  $w_j^R(p_j + u_j) = (1 + \gamma_j) \cdot (p_j + u_j) - \frac{\gamma_j}{P_{kj}} \cdot (p_j + u_j)^2$  conformément à l'équation (6.16). En ignorant dans un premier temps la contrainte d'intégralité, pour déterminer  $p_j + u_j$  permettant d'obtenir une durée  $d_j$  on doit résoudre l'équation du second degré  $-d_j \cdot \gamma_j / P_{kj} \cdot (p_j + u_j)^2 + d_j \cdot (1 + \gamma_j) \cdot (p_j + u_j) - d_j^{\max} \cdot P_{kj} = 0$ , ce qui donne comme racine acceptable

$$p_j + u_j = (P_{kj}/2\gamma_j) \cdot \left[ (1 + \gamma_j) - \frac{1}{d_j} \sqrt{(1 + \gamma_j)^2 \cdot d_j^2 - 4\gamma_j d_j^{\max} \cdot d_j} \right]. \quad (6.42)$$

Cette racine a très peu de chance d'être un nombre entier. Notons alors  $S(d_j)$  le nombre entier de travailleurs en sureffectif qu'il faut au total pour atteindre une durée  $d_j$  avec  $S(d_j) = \lceil d_j \cdot (p_j + u_j - P_{kj}) \rceil$ . La qualité associée à cette durée impliquant ce sureffectif est égale à  $q_j = 1 - \Delta q_j^S \cdot S(d_j)$ . En remplaçant dans  $S(d_j)$  l'effectif  $p_j + u_j$  par son expression en fonction de  $d_j$  donnée par l'équation (6.42) on obtient finalement

$$q_j = 1 - \Delta q_j^S \cdot (P_{kj}/2\gamma_j) \cdot \left[ (1 - \gamma_j) \cdot d_j - \sqrt{(1 + \gamma_j)^2 \cdot d_j^2 - 4\gamma_j d_j^{\max} \cdot d_j} \right] \quad (6.43)$$

si  $h_{jt} = 0 \quad \forall t = 1..d_j$  et  $\gamma_j > 0 \quad \forall j = 1..n$ .

Ainsi, lorsqu'il y a des pertes de productivité liées au sureffectif, la qualité varie de façon non linéaire avec la durée. En effet, lorsque  $\gamma_j > 0$  il faut davantage de main d'oeuvre pour atteindre une durée  $d_j$  que dans la situation où  $\gamma_j = 0$ , ce qui signifie un surcroît de travailleurs en sureffectif et donc une baisse d'autant plus importante de la qualité. Comme auparavant, n'importe quelle allocation du sureffectif sur les  $d_j$  périodes produit cette même qualité  $q_j$  car seule la somme des sureffectifs intervient dans le calcul de la qualité, c'est-à-dire  $\sum_{t=1..d_j} (p_{jt} + u_{jt} - P_{kj}) = S(d_j)$ .

Pour illustrer, reprenons l'exemple précédent où l'on accélère la durée seulement avec du sureffectif en supposant que la baisse de qualité  $\Delta q_j^S$  résultant de l'utilisation d'un travailleur en sureffectif sur une période s'élève à  $\Delta q_j^S = 0,025$ . On suppose d'abord  $\gamma_j = 0$ . Pour atteindre une durée  $d_j = 7$  le sureffectif total doit être égal à  $P_{kj} \cdot (d_j^{\max} - d_j) = 3$  ce qui donne une qualité  $q_j = 1 - 3\Delta q_j^S = 0,925$ . Notons que cette qualité de 0,925 est supérieure à la qualité de 0,85 que nous avons obtenue pour une même durée  $d_j$  en accélérant avec des heures supplémentaires seulement. Cela provient de l'hypothèse selon laquelle l'impact des heures supplémentaires sur la qualité est plus important que celui du sureffectif.

Supposons maintenant  $\gamma_j = 0,0857$ . On commence par calculer l'effectif moyen à chaque période  $p_j + u_j$  selon l'équation (6.42) avec  $\gamma_j = 0,0857$ ,  $P_{kj} = 3$ ,  $d_j^{\max} = 8$  et  $d_j = 7$  ce qui donne  $p_j + u_j = 3,4758$ . Ce volume théorique de main d'oeuvre par période permet d'obtenir exactement le contenu requis  $W_j = 24$  hommes - jour. On obtient  $S(d_j) = \lceil d_j \cdot (p_j + u_j - P_{kj}) \rceil = \lceil 7 \cdot (3,4758 - 3) \rceil = 4$ , ce qui implique une qualité  $q_j = 1 - 4\Delta q_j^S = 0,90$  inférieure à la qualité de 0,925 obtenue dans le cas où  $\gamma_j = 0$ .

**Correction de qualité associée au sureffectif avec pertes de productivité.** En réalité, lorsque  $\gamma_j > 0$  la qualité s'établit à une valeur légèrement supérieure à celle donnée par l'équation (6.43) si tout le sureffectif  $S(d_j)$  n'est pas utilisé, ce qui est généralement le cas puisque ce dernier résulte d'un arrondi à l'entier supérieur de  $d_j \cdot (p_j + u_j - P_{kj})$ . Dans notre exemple avec  $\gamma_j = 0,0857$  nous avons obtenu un sureffectif de 4 travailleurs à répartir sur  $d_j = 7$  jours. Imaginons que l'on concentre ce sureffectif sur la première période d'exécution, ce qui correspond à l'allocation

$(p_{jt} + u_{jt})_{t=1..7} = (7, 3, 3, 3, 3, 3, 3)$ . La main d'oeuvre corrigée qui en résulte est égale à  $w_j^R(7) + 18 = 6, 2 + 18 = 24, 2$  mais nous n'avons besoin que de  $W_j = 24$  hommes - jour. Il faut alors déterminer la main d'oeuvre réellement active en période 1 pour en déduire le volume de sureffectif oisif c'est à dire le nombre de personnes qui ne sont pas actives à cette période, et qui n'affectent donc pas la qualité.

Plus généralement, pour une allocation avec sureffectif, c'est à la période où ce sureffectif est le plus faible qu'il faut imposer l'oisiveté car le gain de qualité en résultant sera d'autant plus grand. Notons  $\tau$  cette période. On cherche à déterminer à cette période  $\tau$  la main d'oeuvre réellement active  $p_{j\tau}^{\text{active}} + u_{j\tau}^{\text{active}}$  telle que  $w_j^R(p_{j\tau}^{\text{active}} + u_{j\tau}^{\text{active}}) = w_{j\tau}^{\text{active}}$  avec

$$w_{j\tau}^{\text{active}} = W_j - \sum_{t \neq \tau} w_j^R(p_{jt} + u_{jt}). \quad (6.44)$$

Il nous faut donc résoudre l'équation

$$(1 + \gamma_j) \cdot (p_{j\tau}^{\text{active}} + u_{j\tau}^{\text{active}}) - \frac{\gamma_j}{P_{kj}} \cdot (p_{j\tau}^{\text{active}} + u_{j\tau}^{\text{active}})^2 = w_{j\tau}^{\text{active}} \quad (6.45)$$

ce qui donne comme racine

$$p_{j\tau}^{\text{active}} + u_{j\tau}^{\text{active}} = (P_{kj}/2\gamma_j) \cdot \left[ (1 + \gamma_j) - \sqrt{(1 + \gamma_j)^2 - 4(\gamma_j/P_{kj})w_{j\tau}^{\text{active}}} \right]. \quad (6.46)$$

Si l'on note  $w_{j\tau}^{\text{idle}}$  le volume de main d'oeuvre oisive à la période  $\tau$  avec

$$w_{j\tau}^{\text{idle}} = p_{j\tau} + u_{j\tau} - (p_{j\tau}^{\text{active}} + u_{j\tau}^{\text{active}}), \quad (6.47)$$

alors la qualité s'établit à une valeur corrigée

$$q_j = q_j = 1 - \Delta q_j^S \cdot S(d_j) + \Delta q_j^S \cdot w_{j\tau}^{\text{idle}} \quad (6.48)$$

qui dépend du sureffectif à la période  $\tau$  et donc de l'allocation de la main d'oeuvre le long de la durée d'exécution de l'activité.

Considérons à nouveau l'allocation  $(p_{jt} + u_{jt})_{t=1..7} = (7, 3, 3, 3, 3, 3, 3)$  à des fins d'illustrations. On en déduit  $\tau = 1$  et  $w_{j1}^{\text{active}} = 24 - 6 \cdot 3 = 6$ . On obtient  $p_{j1}^{\text{active}} + u_{j1}^{\text{active}} = 6, 7116$  d'après l'équation (6.46), d'où  $w_{j1}^{\text{idle}} = p_{j1} + u_{j1} - (p_{j1}^{\text{active}} + u_{j1}^{\text{active}}) = 7 - 6, 7116 = 0, 2884$ . La qualité s'établit réellement à  $q_j = 0, 90 + 0, 2884 \cdot 0, 025 = 0, 9072$ .

Si l'on prend au contraire une allocation avec une répartition égalitaire du sureffectif (lissage de charges) c'est-à-dire  $(p_{jt} + u_{jt})_{t=1..d_j} = (4, 4, 4, 4, 3, 3, 3)$  avec  $\tau = 4$  et  $w_{j4}^{\text{active}} = 24 - \sum_{t=1..3} w_j^R(4) - 9 = 3, 3428$  ce qui donne  $p_{j4}^{\text{active}} + u_{j4}^{\text{active}} = 3, 3794$  et donc  $w_{j4}^{\text{idle}} = 0, 6265$ , on obtient une qualité  $q_j = 0, 90 + 0, 6264 \cdot 0, 025 = 0, 9155$  qui est supérieure à la qualité  $q_j = 0, 9072$  correspondant à l'allocation qui concentre tout le sureffectif sur une période. Ceci s'explique par la relation croissante entre la

perte de productivité et le sureffectif qui implique une oisiveté moindre lorsque le sureffectif est plus important puisque les travailleurs se gênent d'autant plus qu'ils sont nombreux ; il faut alors davantage de travail pour atteindre une même cadence de production. Cette moindre oisiveté entraîne mécaniquement une dégradation plus importante de qualité parce que la qualité dépend du sureffectif réellement utilisé.

### 6.6.3 Accélération en combinant heures supplémentaires et sureffectif avec pertes de productivité

Si l'on autorise à la fois les heures supplémentaires et le sureffectif, le processus d'optimisation pourra conduire à une utilisation simultanée de ces deux types de main d'oeuvre additionnelle pour obtenir la qualité la élevée possible. Partant de notre exemple avec  $\gamma_j = 0,0857$ ,  $P_{kj} = 3$ , et  $W_j = 24$  nous donnons dans le tableau 6.2 la qualité associée à plusieurs allocations de main d'oeuvre additionnelle permettant d'atteindre une même durée  $d_j = 7$ . Pour les deux premières allocations qui n'utilisent que du sureffectif, la qualité exacte est la qualité corrigée de l'oisiveté donnée par l'équation (6.48). Celle-ci se calcule en commençant par déterminer  $w_{j\tau}^{\text{active}}$  selon l'équation (6.44) puis la main d'oeuvre active  $p_{j\tau}^{\text{active}} + u_{j\tau}^{\text{active}}$  en utilisant l'équation (6.46). On en déduit  $w_{j\tau}^{\text{idle}}$  défini par l'équation (6.47) et donc la qualité corrigée qui fait intervenir ce volume oisif de main d'oeuvre (équation (6.48)). Ainsi, on reporte dans les deux premières lignes du tableau les calculs que nous avons détaillés précédemment pour une accélération avec un sureffectif seulement et une allocation qui lisse ce sureffectif (première allocation) ou le concentre au contraire sur une période (deuxième allocation). Pour ces deux allocations, la borne inférieure de la qualité est donnée par l'équation (6.38) qui ne tient pas compte de l'oisiveté. Les troisième et quatrième allocations combinent un sureffectif moindre avec des heures supplémentaires, ce sureffectif étant utilisé pleinement (pas d'oisiveté), la qualité exacte s'obtient par l'équation (6.38). On constate que le nombre d'heures supplémentaires pour une allocation qui concentre le sureffectif sur une période est supérieur à celui qui est nécessaire dans l'allocation lissée (3,51 heures contre 2,74). Cela s'explique par des pertes de productivité d'autant plus importantes que le sureffectif périodique est grand. La qualité se trouve par conséquent plus dégradée dans une allocation où le sureffectif n'est pas lissé. Dans la dernière ligne, nous reportons le volume d'heures supplémentaires nécessaires pour atteindre cette même durée  $d_j = 7$  lorsque l'accélération s'effectue sans sureffectif. Nous obtenons dans ce cas la qualité la plus faible en raison de l'hypothèse selon laquelle les heures supplémentaires dégradent davantage la qualité que le sureffectif.

Allocation de main d'oeuvre			Qualité	
#	Sureffectif	Heures sup.	Exacte	Borne inf.
<b>1</b>	(1, 1, 1, 1, 0, 0, 0)	0	0,9155	0,90
<b>2</b>	(4, 0, 0, 0, 0, 0, 0)	0	0,9072	0,90
<b>3</b>	(1, 1, 1, 0, 0, 0, 0)	2,74	0,9078	--
<b>4</b>	(3, 0, 0, 0, 0, 0, 0)	3,51	0,9030	--
<b>5</b>	0	48 ( $h_{jt} \leq 12$ )	0,85	--

Tableau 6.2 – Qualité associée à plusieurs allocations de main d'oeuvre additionnelle pour une même durée  $d_j = 7$

Si l'on raisonne en termes de qualité exacte, en comparant les allocations #1 et #3 qui lissent le sureffectif, on constate qu'il est préférable d'accélérer avec un sureffectif uniquement, quitte à ce que l'un des travailleurs ait un temps d'inactivité (allocation #1) plutôt que de combiner les deux types de main d'oeuvre additionnelle en les utilisant pleinement (allocation #3). Ce constat reste vrai pour les allocations #2 et #4 qui concentrent le sureffectif. La meilleure solution parmi les quatre est celle qui n'utilise que du sureffectif lissé sur les périodes, même si cette main d'oeuvre n'est pas utilisée totalement. Si l'on raisonne en termes d'estimation basse de la qualité (borne inférieure), alors toute allocation qui combine sureffectif et heures supplémentaires est meilleure que celles qui n'utilisent que du sureffectif car la correction d'oisiveté n'est pas appliquée, ce qui pénalise ces allocations #1 et #2.

Pour toutes les allocations avec sureffectif (allocations #1 à #4), on remarque toutefois que les différences de qualité restent faibles que l'on considère ou non les corrections de qualité et il n'est pas dit d'obtenir des valeurs aussi précises des seuils de qualité minimum à respecter. Pour des check-lists de contrôle qualité d'une activité comprenant par exemple 20 items, les seuils de qualité minimum sont des multiples de  $1/20 = 0,05$ . Ainsi, si l'un de ces seuils est égal à 0,90 par exemple toutes les allocations avec sureffectif que nous venons d'analyser sont acceptables, avec ou sans correction de la qualité. On peut raisonnablement supposer que le nombre d'items des check-lists de contrôle qualité dépasse rarement la vingtaine dans les industries manufacturières, chez Alstom ces items sont même limités au nombre de dix. En outre, l'introduction de cette correction de qualité dans l'équation (6.38) serait coûteuse en variables et contraintes additionnelles. Une linéarisation possible consisterait à déterminer pour toute valeur de  $w_{j\tau}^{\text{active}}$  la main d'oeuvre active  $p_{j\tau}^{\text{active}} + u_{j\tau}^{\text{active}}$  correspondante à l'aide de l'équation (6.46) pour en déduire  $w_{j\tau}^{\text{idle}}$  qui autorise le calcul de la qualité corrigée donnée en équation (6.48). Comme la valeur de  $w_{j\tau}^{\text{active}}$  dépend de la somme de la main d'oeuvre corrigée sur les autres périodes, le nombre de valeurs possibles pour  $w_{j\tau}^{\text{active}}$  s'élève à  $(\mathcal{T}_j - 1) \cdot \bar{w}_j$  avec autant de variables binaires à créer. Par conséquent, nous retenons le calcul de la qualité sans correction de l'oisiveté donnée par l'équation (6.38) qui fournit une estimation pessimiste de la qualité associée aux cas d'accélération où les heures supplémentaires sont proscrites (sureffectif seul).

**Bornes inférieures sur la durée et la qualité.** Commençons par définir une borne inférieure  $\underline{d}_j$  sur la durée d'exécution de toute activité  $j$  avec une accéléra-

tion résultant d'un effectif maximum  $\bar{w}_j$  de travailleurs en heures régulières et un maximum de  $H_O P_{kj}$  d'heures supplémentaires à toute période d'exécution. On a

$$\underline{d}_j = \left\lceil \frac{W_j}{w_j^R(\bar{w}_j) + \frac{H_O}{H_R} P_{kj}} \right\rceil, \quad (6.49)$$

où  $w_j^R(\bar{w}_j) = (1 + \gamma_j) \cdot \bar{w}_j - \frac{\gamma_j}{P_{kj}} \cdot \bar{w}_j^2$  si  $\bar{w}_j > P_{kj}$  d'après l'équation (6.16). Il s'agit bien d'une borne inférieure car il n'est pas certain que l'on puisse affecter à chaque période  $H_O \cdot P_{kj}$  heures supplémentaires en raison des contraintes légales hebdomadaires. De plus, si le recrutement en temporaires est limité voire inexistant, il se peut que l'effectif maximum  $\bar{w}_j$  s'établisse à la valeur de  $P_k$  (voir équations (6.22) et (6.20)) et dans ce cas, si deux activités nécessitant la spécialisation professionnelle  $k$  sont exécutées simultanément elles se partageront ce nombre  $P_k$  de permanents disponibles pour le projet de sorte qu'aucune d'elles ne pourra en utiliser la totalité.

En appliquant les équations (6.38) et (6.39) on obtient la borne inférieure  $\underline{q}_j$  sur la qualité associée à la durée  $\underline{d}_j$  donnée en équation (6.49)

$$\underline{q}_j = q_j^{\max} - \underline{d}_j \cdot [\Delta q_j^S \cdot (\bar{w}_j - P_{kj}) + \Delta q_j^O \cdot H_O \cdot P_{kj}], \quad (6.50)$$

Si l'activité est accélérée sans sureffectif et avec une allocation maximale d'heures supplémentaires, alors la durée compressée  $\underline{d}_j^O$  que l'on obtient est définie par  $\underline{d}_j^O = \lceil W_j / (H_O \cdot P_{kj} / H_R) \rceil$  avec un nombre d'heures supplémentaires que l'on suppose identique à chaque période et égal à  $h_{jt} = h_j = H_O \cdot P_{kj} / H_R$  sur  $\underline{d}_j^O$  périodes d'exécution. Pour une durée  $\underline{d}_j^O$ , la qualité s'établit à une valeur  $\underline{q}_j^O = q_j^{\max} - \underline{d}_j^O \cdot \Delta q_j^O \cdot H_O \cdot P_{kj}$ . Un raisonnement similaire peut-être mené pour une compression de la durée avec un sureffectif uniquement avec une affectation à chaque période d'un nombre identique  $\bar{w}_j - P_{kj}$  de travailleurs en sureffectif. Comme nous avons supposé  $\Delta q_j^S < H_R \cdot \Delta q_j^O$  alors pour atteindre une même durée accélérée, la qualité résultant de l'utilisation seule d'heures supplémentaires sera inférieure à celle obtenue avec du sureffectif seulement; elle s'établira entre ces deux valeurs pour toute allocation de main d'oeuvre additionnelle combinant à la fois du sureffectif et des heures supplémentaires.

**Illustration étendue.** La figure 6.2 représente les différentes configurations d'accélération que nous venons d'analyser pour toutes les durées possibles d'exécution d'une activité donnée  $j$ , toujours en prenant les paramètres suivants  $P_{kj} = 3$ ;  $W_j = 24$ ;  $\gamma_j = 0,0857$ ;  $\Delta q_j^S = 0,025$  et  $\Delta q_j^O = 0,00625$ . On suppose de plus une capacité du poste  $A_j = 10$  et un nombre de permanents  $P_k = 5$  avec un recrutement théoriquement illimité en temporaires ( $\lambda = 1$ ). On commence par déterminer l'effectif  $s_j$  selon l'équation (6.18), ce qui donne  $s_j = \min\left(\left\lfloor \frac{1,0857}{0,1714} \cdot 3 \right\rfloor, 24, 10\right) = (19, 24, 10) = 10$ . On utilise ensuite l'équation (6.22) pour obtenir l'effectif maximum  $\bar{w}_j = \min(\bar{p}_j + \bar{u}_j, s_j) = s_j = 10$  car  $\bar{u}_j$  est illimité. Ainsi, on ne pourra jamais affecter plus de  $\bar{w}_j = 10$  travailleurs par jour en heures régulières sur le poste où l'activité est exécutée. Le travail en heures supplémentaires ne peut excéder  $H_O \cdot P_{kj} = 4 \cdot 3 = 12$

heures par jour au total sur l'activité  $j$ . Avec  $H_R = 8$ , d'après l'équation (6.49) on obtient une borne inférieure sur la durée  $\underline{d}_j = \lceil 24/(w_j^R(10) + \frac{4}{8} \cdot 3) \rceil = 3$  où  $w_j^R(10) = 8,0003$  est calculé selon l'équation (6.16).

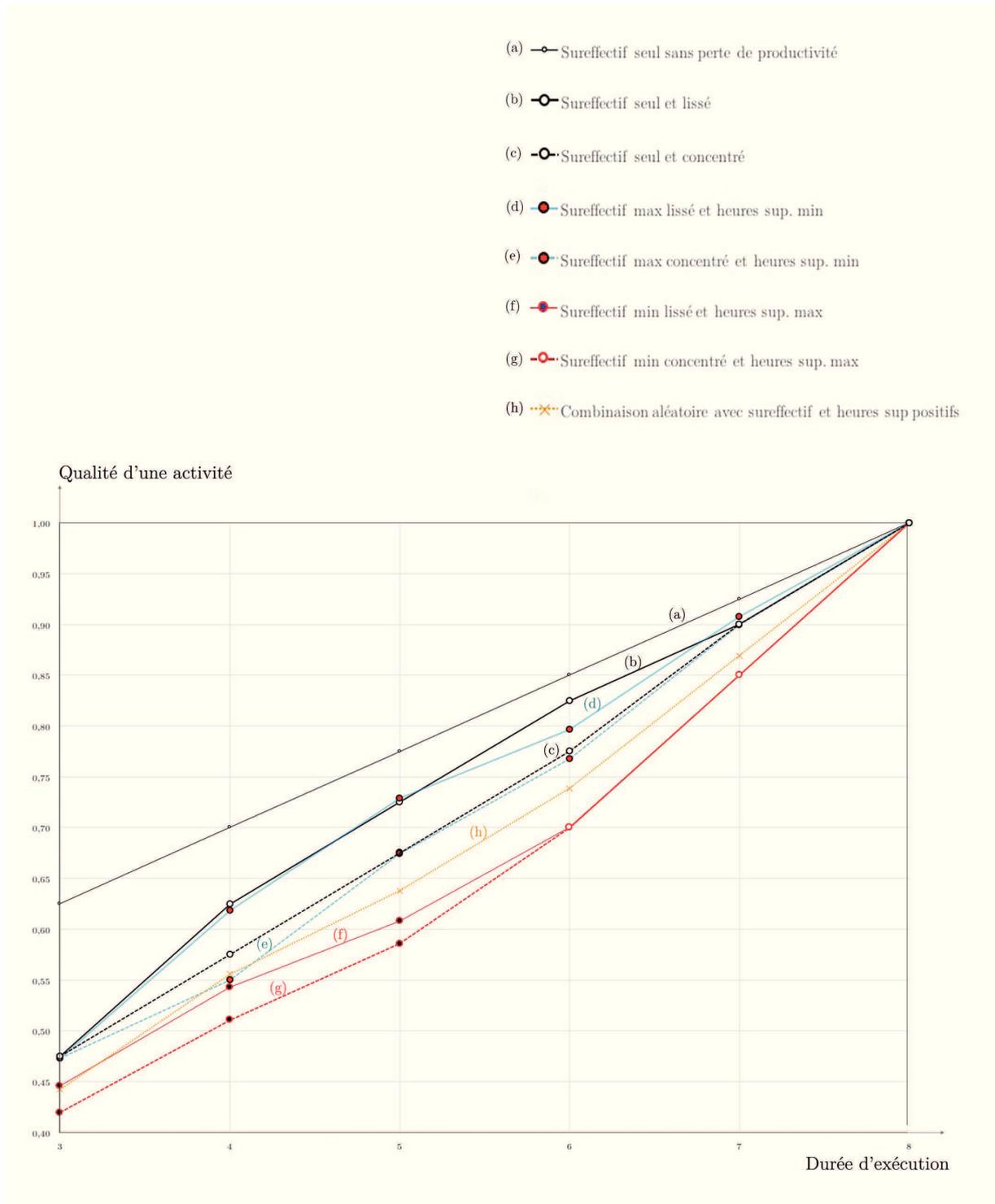


FIGURE 6.2 – Qualité d'une activité en fonction de sa durée d'exécution dans différentes configurations d'utilisation de main d'oeuvre additionnelle

La figure 6.2 relie artificiellement les couples durée - qualité  $(d_j, q_j)$  pour en faciliter la lecture mais le modèle reste en temps discret avec des durées d'exécution entières portées en abscisses. Dans tous les cas de figure examinés, la qualité  $q_j$  est calculée selon l'équation (6.38), c'est-à-dire sans correction d'une oisiveté résultant d'un sureffectif qui ne serait pas complètement utilisé car produisant une main d'oeuvre totale légèrement supérieure à  $W_j$ . La première série (a) donne les couples  $(d_j, q_j)$  obtenus en utilisant uniquement un sureffectif lorsqu'il n'y a pas de pertes de productivité ( $\gamma_j = 0$ ). Toutes les autres séries supposent  $\gamma_j = 0,0857$ . Cette première série (a) fournit la meilleure qualité qu'il est possible d'atteindre, et ce, en ajoutant un sureffectif qui produira toujours un contenu en main d'oeuvre exactement égal à  $W_j = 24$  de sorte que la qualité donnée en ordonnée est "exacte" (pas de correction de qualité à effectuer éventuellement). En outre, La répartition du sureffectif sur les périodes d'exécution n'affecte pas la qualité. On note enfin dans cette série (a) la linéarité entre durée et qualité, résultat de l'absence de perte de productivité liée au sureffectif ( $\gamma_j = 0$ ).

Dès lors qu'il y a des pertes de productivité, en accélérant avec un sureffectif uniquement (série (b)), la qualité se dégrade par rapport à la situation (a) et la relation durée - qualité n'est plus linéaire. La série (b) est construite à partir d'un sureffectif lissé sur les périodes d'exécution et la main d'oeuvre totale corrigée des pertes de productivité excède légèrement  $W_j = 24$  de sorte que la qualité reportée sur la figure est une estimation pessimiste, calculée selon l'équation (6.38). Une correction de l'oisiveté permettrait d'atteindre une qualité exacte supérieure. Cependant, la qualité se dégrade assez fortement si le sureffectif requis pour accélérer l'activité est concentré sur un minimum de période d'exécution, comme le montre la série (c). En effet, l'ajout d'un grand nombre de travailleurs à certaines périodes va entraîner de telles pertes de productivité qu'il faudra davantage de travailleurs pour satisfaire les besoins de l'activité, ce qui implique mécaniquement une baisse de qualité (puisque'elle est proportionnelle au sureffectif).

Partant de la série (b) avec sureffectif seul et lissé, même avec une estimation pessimiste de la qualité, on observe qu'en ajoutant des heures supplémentaires à cette configuration (b), on aboutit à la série (d) qui donne des qualités inférieures. Pour cette série (d), le contenu total en main d'oeuvre est égal à  $W_j$  car on utilise un maximum de sureffectif combiné à un minimum d'heures supplémentaires pour atteindre exactement  $W_j$ . Ainsi, comme nous l'avons analysé antérieurement dans cette section, il est généralement plus avantageux en termes de qualité de ne pas recourir aux heures supplémentaires si on peut obtenir la durée désirée avec un sureffectif uniquement mais lissé, quitte à ce qu'il y ait une légère oisiveté. Cela est dû à un impact négatif sur la qualité plus fort pour les heures supplémentaires que pour le sureffectif. Néanmoins, ces solutions combinant les deux types de main d'oeuvre peuvent permettre une réduction du coût, une autre dimension de l'arbitrage incluse dans cette modélisation. Tout comme la série (d), la série (e) combine un sureffectif maximum et un minimum d'heures supplémentaires mais cette fois, le sureffectif est concentré sur un minimum de périodes, ce qui dégrade la qualité car il faut davantage d'heures supplémentaires pour compenser les pertes de productivité.

Les séries (f) et (g) sont obtenues en limitant au maximum le sureffectif tout en recourant le plus possible aux heures supplémentaires. Ces allocations de main

d'oeuvre sont les moins bonnes en termes de qualité, et l'on constate à nouveau qu'une concentration du sureffectif (série (g)) dégrade davantage la qualité par rapport à la série (f) où il y a un lissage qui implique une utilisation plus restreinte de main d'oeuvre additionnelle. Notons qu'à partir d'une durée  $d_j \geq 6$ , il est possible de n'utiliser que des heures supplémentaires de sorte que les séries (f) et (g) se rejoignent en  $d_j = 6$  et on retrouve alors la relation linéaire entre qualité et durée pour toute durée  $d_j \geq 6$ . Enfin, nous avons produit une série (h) en combinant de façon arbitraire sureffectif et heures supplémentaires pour montrer que la qualité associée se situe à l'intérieur de l'ensemble délimité par les séries (b) et (g) qui donnent respectivement l'enveloppe haute et basse de l'ensemble des qualités.

Le tableau 6.3 complète la figure 6.2 en donnant pour trois configurations particulières parmi les huit que nous venons d'étudier, le nombre de travailleurs en sureffectif (# Sur.), le nombre d'heures supplémentaires (Hrs sup.) et la qualité pour chacune des durées possibles d'exécution. Par exemple, pour la série (b) qui fournit la qualité la plus élevée, seul le sureffectif est utilisé avec un lissage sur les périodes; si l'on souhaite ramener la durée à 3 jours, il faudra 21 opérateurs additionnels ce qui donne une main d'oeuvre corrigée (colonne Contenu travail) de 24,001 et une qualité (pessimiste) de 0,475. On remarque que l'oisiveté est maximum pour une durée de 7 jours puisque l'on dispose de 24,543 hommes - jour tandis qu'ils n'en faut que  $W_j = 24$ . Pourtant la qualité s'établit à 0,90, qui représente sa valeur maximale pour cette durée parmi toutes les configurations possibles. On note que l'on n'indique pas le contenu en travail pour les séries (g) et (h) parce que celui-ci est toujours exactement égal à  $W_j = 24$ , les heures supplémentaires étant autorisées.

Durée	Série (b) - Qualité élevée			Série (g) - Qualité faible			Série (h) - Qualité interméd.		
	<i>Sureffectif max et lissé</i>			<i>Sureff. min concentré et hrs sup.</i>			<i>Sureff. et hrs sup. arbitraires</i>		
	# Sur.	Contenu travail	Qualité	# Sur.	Hrs sup.	Qualité	# Sur.	Hrs sup.	Qualité
3	21	24,001	<b>0,475</b>	15	32,909	<b>0,419</b>	15	29,252	<b>0,442</b>
4	15	24,086	<b>0,625</b>	9	42,283	<b>0,511</b>	10	31,083	<b>0,556</b>
5	11	24,343	<b>0,725</b>	2	58,285	<b>0,586</b>	5	37,942	<b>0,638</b>
6	7	24,143	<b>0,825</b>	0	48,000	<b>0,700</b>	2	33,828	<b>0,739</b>
7	4	24,543	<b>0,900</b>	0	24,000	<b>0,850</b>	1	16,914	<b>0,869</b>
8	0	24,000	<b>1,000</b>	0	0,000	<b>1,000</b>	0	0,000	<b>1,000</b>

Tableau 6.3 – Main d'oeuvre additionnelle utilisée dans trois des configurations présentées à la figure 6.2

#### 6.6.4 Linéarisation et formulation de la contrainte d'un seuil de qualité minimum à respecter

Il convient de linéariser l'expression (6.38) en adoptant une méthode analogue à celle que nous avons employée pour linéariser la fonction de main d'oeuvre corrigée des pertes de productivité  $w_j^R(p_{jt} + u_{jt})$ . On peut ainsi réutiliser les mêmes variables binaires  $y_{jtl}$  que nous avons créées pour cette linéarisation sans besoin d'en ajouter d'autres. On considère la perte de qualité liée au sureffectif uniquement et on la note  $\Delta q_j^S(p_{jt} + u_{jt})$  pour la définir comme une fonction de l'effectif en heures régulières

alloué à la période  $t$ . On a ainsi  $\Delta q_j^S(p_{jt} + u_{jt}) = \Delta q_j^S \cdot (p_{jt} + u_{jt} - P_{kj})$  si  $p_{jt} + u_{jt} > P_{kj}$  et 0 sinon. Il est possible de calculer  $\Delta q_j^S(l)$  pour tout  $l = 0.. \bar{w}_j$ . L'expression (6.38) s'écrit alors sous la forme linéaire

$$\Delta q_{jt} = \sum_{l=0}^{\bar{w}_j} \Delta q_j^S(l) \cdot y_{jlt} + \Delta q_j^O \cdot h_{jt} \quad \forall j = 1..n \quad t \in \mathcal{T}_j. \quad (6.51)$$

avec

$$\Delta q_j^S(l) = \begin{cases} \Delta q_j^S \cdot (l - P_{kj}) & \text{si } l > P_{kj} \\ 0 & \text{sinon} \end{cases} \quad \forall l = 0.. \bar{w}_j \quad j = 1..n. \quad (6.52)$$

Ainsi toutes les variations de qualité associées à toutes les valeurs possibles du sureffectif seront préalablement calculées et stockées dans le code du programme d'optimisation. Par exemple si  $l = 0$  à une période donnée  $t$  on aura  $y_{j0t} = 1$  et donc  $\Delta q_j^S(0) = 0$ . Avec  $P_{kj} = 3$  et  $p_{jt} + u_{jt} = 5$  alors  $y_{j5t} = 1$  et  $\Delta q_j^S(5) = 2\Delta q_j^S$ , une valeur calculée comme paramètre dans le code.

En rappelant que  $V^\theta$  désigne l'ensemble des activités de niveau  $\theta$  de tolérance aux défauts avec un seuil de qualité minimum  $q_\theta^{\min}$ , les contraintes de respect du seuil de qualité minimum s'écrivent

$$q_j^{\max} - \sum_{t \in \mathcal{T}_j} \Delta q_{jt} \geq q_\theta^{\min} \quad \forall j \in V^\theta, \quad \theta = 1.. \Theta, \quad (6.53)$$

où  $\Delta q_{jt}$  est défini par l'équation (6.51). L'avantage de ce type de contraintes c'est qu'on s'affranchit d'une agrégation par la moyenne qui suppose implicitement une compensation des qualités individuelles, avec les critiques qu'on peut habituellement lui adresser. En effet, si l'on considère par exemple deux activités et si l'une d'elles a une qualité de 0,4 inférieure et l'autre possède une qualité égale à 1, la qualité moyenne du projet est égale à 0,7. Si l'on impose pour le projet un seuil de qualité moyenne à 0,7 alors les qualités individuelles vont être acceptées bien qu'en réalité, on ne peut pas tolérer une qualité individuelle de 0,4, pour la première activité si son niveau de tolérance est tel que son seuil de qualité minimum est égal à 0,6 par exemple.

On note que l'on peut définir dans notre modèle autant de niveaux de tolérance qu'il y a d'activités, c'est-à-dire imposer à chaque activité  $j$  un seuil de qualité minimum  $q_j^{\min}$  qui lui est propre, soit  $\Theta = n$ . Cela ne complique pas la résolution du programme d'optimisation mais en revanche, cela impose un surcroît d'estimation des seuils minimum de qualité.

## 6.7 Coûts et contrainte de budget

Pour simplifier, on suppose que le coût du travail est le même quelles que soient les spécialisations professionnelles. Cette hypothèse peut être levée facilement dans

la modélisation sans que cela modifie le nombre des contraintes. En outre elle est relativement réaliste puisque les spécialisations que nous considérons nécessitent toutes un niveau de qualification relativement peu élevé. Cette hypothèse serait inappropriée si les spécialisations étaient substituables pour les activités c'est-à-dire si une même activité pouvait s'exécuter en utilisant l'une ou l'autre des spécialisations, auquel cas un arbitrage entre les différentes spécialisations de coûts différents devrait s'opérer. Mais ici, l'arbitrage ne se fait pas entre les différentes spécialisations mais entre les différents types de main d'oeuvre pour une spécialisation donnée, à savoir travail permanent, temporaire ou heures supplémentaires.

On note  $c_P$  le coût d'une heure de travail permanent (l'hypothèse précédente revient à poser  $c_{k,P} = c_P$  pour tout  $k = 1..K$ ),  $c_U$  le coût d'une heure de travail temporaire et  $c_O$  le coût d'une heure de travail supplémentaire.

En mode normal, le coût total du projet est égal au coût de la main d'oeuvre permanente utilisée, soit  $c_P \cdot H_R \cdot \sum_{j=1}^n d_j^{\max} \cdot P_{kj}$ , ce qui représente le budget normal pour le projet. En tenant compte de la main d'oeuvre temporaire et en heures supplémentaires, le coût total du projet (ou budget total), noté  $B$ , s'écrit plus généralement

$$B = \sum_{j=1}^n \sum_{t \in \mathcal{T}_j} (c_P \cdot H_R \cdot p_{jt} + c_U \cdot H_R \cdot u_{jt} + c_O \cdot h_{jt}). \quad (6.54)$$

Si l'on désigne par  $\rho$  le taux d'accroissement du budget normal que l'on s'autorise pour accélérer le projet, alors la contrainte de budget s'écrit

$$B \leq (1 + \rho) \cdot c_P \cdot H_R \cdot \sum_{j=1}^n d_j^{\max} \cdot P_{kj}, \quad (6.55)$$

où  $B$ , le budget total (ou coût total) du projet est défini par l'équation (6.54).

Remarquons que l'on aurait pu calculer le budget en mode normal sur la base de la main d'oeuvre totale permanente disponible le long de l'horizon  $T$  qui correspond à la durée totale d'exécution du projet en mode normal, c'est-à-dire à partir de  $T \cdot \sum_{k=1}^K P_k$  puisqu'il s'agit d'une main d'oeuvre que l'on doit payer de toute façon. Mais le coût de cette main d'oeuvre totale n'est pas le coût qui peut être imputé au projet puisqu'à certaines périodes il est possible que la main d'oeuvre requise soit inférieure à la main d'oeuvre disponible. Dans ce cas, les permanents qui ne sont pas affectés à l'exécution du projet auquel on s'intéresse sont alloués à un autre projet dont le coût inclura ces permanents-là. De plus, le coût défini comme  $T \cdot \sum_{k=1}^K P_k$  suppose que la durée du projet est égale à  $T$  ce qui n'est plus le cas en mode accéléré.

## 6.8 Pré-traitement, paramétrage et résumé du programme d'optimisation

Dans la première sous-section, nous donnons le calcul des bornes sur la durée totale du projet. Nous montrons ensuite comment se détermine le nombre de per-

manents de chaque spécialisation professionnelle à allouer au projet. La deuxième sous-section explique comment paramétrer le modèle. Enfin, la troisième sous-section résume le programme d'optimisation.

### 6.8.1 Pré-traitement

**Borne supérieure  $T$  sur la durée du projet.** On commence par déterminer les dates de début au plus tôt sur la base des durées normales d'exécution. On note  $\overline{ES}_j$  ces dates avec  $\overline{ES}_j = \max_{i \in \Gamma^{-1}(j)} \{ \overline{ES}_i + d_i^{\max} \}$  et  $\overline{ES}_0 = 0$ . Puis on détermine les dates de début au plus tard, toujours sur la base des durées normales, soit  $\overline{LS}_j = \min_{i \in \Gamma(j)} \{ \overline{LS}_i \} - d_j^{\max}$  avec  $\overline{LS}_{n+1} = \overline{ES}_{n+1}$ . Enfin, on obtient  $LF_j = \overline{LS}_j + d_j^{\max}$ . La borne supérieure sur la durée du projet correspond simplement à sa durée totale d'exécution en mode normal, c'est-à-dire  $T = LF_{n+1}$ .

**Borne inférieure  $\underline{T}$  sur la durée du projet.** Les dates possibles de début au plus tôt,  $ES_j$ , sont simplement données par  $ES_j = \max_{i \in \Gamma^{-1}(j)} \{ ES_i + \underline{d}_i \}$  avec  $ES_0 = 0$  et  $\underline{d}_i$  est la borne inférieure sur la durée de l'activité  $i$  telle que définie dans l'équation (6.49). On obtient ainsi  $\mathcal{T}_j$ , l'ensemble des périodes possibles d'exécution de l'activité  $j$  avec  $\mathcal{T}_j = \{ ES_j..LF_j \}$ . On recalcule ensuite les dates de début au plus tard et de fin au plus tard sur la base des bornes inférieures de durée  $\underline{d}_j$  pour obtenir  $\underline{T}$ , la borne inférieure sur la durée du projet.

La durée totale d'exécution du projet est ainsi bornée

$$\underline{T} \leq C_{max} \leq T. \quad (6.56)$$

**Nombre total de permanents affectés au projet.** Pour toute spécialisation professionnelle  $k$ , le nombre de permanents affectés au projet est tel qu'il permet son exécution en mode normal, en prenant un ordonnancement au plus tôt ou au plus tard. Le calcul de ce nombre  $P_k$  à partir d'un ordonnancement au plus tôt se fait comme suit

$$P_k = \max_{t=1..T} \left( \sum_{j \in V_k | t \in \{ \overline{ES}_j.. \overline{ES}_j + d_j^{\max} \}} P_{kj} \right) \quad \forall k = 1..K. \quad (6.57)$$

On considère ainsi l'ensemble des activités nécessitant la spécialisation  $k$ , puis pour chaque période  $t$  de l'horizon, si cette période est une période d'exécution en mode normal pour une ou plusieurs activités, la somme des permanents qui leur sont alloués définit le besoin en permanents à la période  $t$  pour toutes les activités en cours d'exécution nécessitant la spécialisation  $k$ . On prend ensuite le maximum de ces besoins en permanents sur l'ensemble des périodes de l'horizon pour obtenir  $P_k$ , le nombre de permanents de spécialisation  $k$  affectés au projet.

### 6.8.2 Paramétrage du modèle

**Coefficient  $\gamma_j$  d'impact du sureffectif sur la productivité et dégradation de qualité  $\Delta q_j^S$  associée à l'utilisation d'un travailleur en sureffectif durant une période d'exécution.** En pratique, pour estimer le coefficient d'impact  $\gamma_j$  il suffit d'affecter à l'activité  $j$  un nombre identique d'une période à l'autre de travailleurs en sureffectif pour toute la durée de son exécution et relever cette durée observée  $d_j^{\text{obs}}$ ; les heures supplémentaires devant être fixées à zéro. Ainsi, à chaque période on aura un effectif identique  $p_{jt} + u_{jt} = p_j + u_j > P_{kj}$ . On utilise ensuite la relation  $d_j^{\text{obs}} = W_j/w_j^R(p_j + u_j)$  soit l'équation  $d_j^{\text{obs}} = W_j / \left[ (1 + \gamma_j) \cdot (p_j + u_j) - \frac{\gamma_j}{P_{kj}} \cdot (p_j + u_j)^2 \right]$  à résoudre en  $\gamma_j$  en délaissant délibérément la contrainte d'intégralité pour obtenir une estimation plus précise de  $\gamma_j$ . La résolution de cette équation en  $\gamma_j$  en fournit l'estimation suivante

$$\hat{\gamma}_j = \frac{P_{kj} \cdot [W_j - (p_j + u_j) \cdot d_j^{\text{obs}}]}{d_j^{\text{obs}} \cdot [P_{kj} \cdot (p_j + u_j) - (p_j + u_j)^2]} \quad \forall j = 1..n. \quad (6.58)$$

Par exemple, si l'on ajoute un seul travailleur en sureffectif à chaque période d'exécution d'une activité donnée  $j$ , soit  $p_j + u_j = 4 > P_{kj} = 3$  avec  $W_j = 24$ , si la durée observée d'exécution de l'activité dans ces conditions de sureffectif est égale à  $d_j^{\text{obs}} = 6,1764$  jours, en appliquant l'équation (6.58) on obtient une estimation  $\hat{\gamma}_j = 0,0857$ .

Dans ces mêmes conditions de sureffectif, on évalue la qualité de l'activité à l'issue de son exécution (au bout de  $d_j^{\text{obs}}$  périodes) par validation ou non des items de la check-list de contrôle qualité de sorte à relever une qualité observée  $q_j^{\text{obs}}$ . En combinant les équations (6.38) et (6.39) on obtient la relation  $q_j^{\text{obs}} = q_j^{\text{max}} - d_j^{\text{obs}} \cdot \Delta q_j^S \cdot (p_j + u_j - P_{kj})$  ce qui nous permet d'obtenir une estimation de la dégradation de qualité liée à l'utilisation d'un travailleur en sureffectif durant une période, soit

$$\hat{\Delta q}_j^S = \frac{q_j^{\text{max}} - q_j^{\text{obs}}}{d_j^{\text{obs}} \cdot (p_j + u_j - P_{kj})} \quad \forall j = 1..n. \quad (6.59)$$

Si l'on reprend l'exemple précédent où  $p_j + u_j = 4 > P_{kj} = 3$  et  $d_j^{\text{obs}} = 6,1764$  avec une qualité en mode normal  $q_j^{\text{max}} = 1$  et une qualité observée  $q_j^{\text{obs}} = 0,8462$  exprimant la non validation de deux items d'une check-list en contenant treize ( $q_j^{\text{obs}} = (2/13) - 1 = 0,8462$ ) alors l'équation (6.59) fournit l'estimation  $\hat{\Delta q}_j^S = 0,0249$ . Ainsi, tout travailleur en sureffectif employé durant une journée de travail en heures régulières provoquera une dégradation de la qualité égale à 0,0249 qui correspond au pourcentage d'items non validés par jour. Sur une durée de  $d_j^{\text{obs}} = 6,1764$  jours, cela représente bien une dégradation de qualité égale à  $6,1764 \times 0,0249 = 0,1538$  soit 15,38% d'items non validés (soit 2 items sur 13). En d'autres termes, chaque travailleur en sureffectif durant une journée de  $H_R$  de travail régulier induit un pourcentage d'items non validés égal à 2,49%.

**Dégradation de qualité  $\Delta q_j^O$  associée à l'utilisation d'une heure supplémentaire.** Une procédure d'évaluation de la qualité similaire à la précédente

peut être utilisée. Pour estimer  $\Delta q_j^O$  il convient d'affecter à chaque période d'exécution de toute activité  $j$  un nombre identique  $h_j$  d'heures supplémentaires, en excluant le sureffectif, de sorte qu'à chaque période l'allocation de main d'oeuvre est constante et égale à  $P_{kj} + h_j/H_R$ . La durée d'exécution  $d_j$  qui en résulte est donnée par  $d_j = W_j/(P_{kj} + h_j/H_R)$ . Une fois l'activité achevée, on en évalue la qualité par validation ou non des items de la check-list pour en déduire  $q_j^{\text{obs}}$ . En combinant comme précédemment les équations (6.38) et (6.39) on obtient la relation  $q_j^{\text{obs}} = q_j^{\text{max}} - d_j \cdot \Delta q_j^O \cdot h_j$  ce qui nous permet d'obtenir

$$\hat{\Delta q}_j^O = \frac{q_j^{\text{max}} - q_j^{\text{obs}}}{d_j \cdot h_j} \quad \forall j = 1..n. \quad (6.60)$$

Si l'on reprend l'exemple précédent avec  $P_{kj} = 3$ ;  $h_j = 8$ ;  $H_R = 8$  et toujours  $W_j = 24$  alors on obtient une durée  $d_j = 24/(3 + 8/8) = 6$ . Si la qualité observée de l'activité s'établit à la valeur  $q_j^{\text{obs}} = 0,6923$  exprimant la non validation de quatre items d'une check-list en contenant treize ( $q_j^{\text{obs}} = (4/13) - 1 = 0,6923$ ) alors l'équation (6.59) fournit l'estimation  $\hat{\Delta q}_j^O = 0,0064$ . Cela signifie que chaque heure supplémentaire induit la non validation de 0,64% des items de la check-list.

On obtiendra des estimations de  $\hat{\gamma}_j$ ,  $\hat{\Delta q}_j^S$  et  $\hat{\Delta q}_j^O$  d'autant plus précises que la méthode de relevé d'observations de durée et qualité décrite plus haut sera répétée un grand nombre de fois, idéalement en faisant varier  $h_j$  et  $p_j + u_j - P_{kj}$  tout en conservant leur constance d'une période à l'autre.

**Coûts unitaire de main d'oeuvre et heures de travail.** On rappelle que  $c_P$  désigne le coût d'une heure de travail permanent,  $c_U$  le coût d'une heure de travail temporaire et  $c_O$  le coût d'une heure de travail supplémentaire. On peut supposer comme le font Heimerl et Kolisch (2010) [72] que  $c_P < c_O < c_U$ . Nous aurions pu aussi considérer des coûts du travail différents selon les spécialisations; cela peut-être introduit sans difficulté dans la modélisation en définissant les paramètres  $c_{Pk}$ ,  $c_{Uk}$  et  $c_{Ok}$  pour tout  $k = 1..K$ . Cette hypothèse selon laquelle le coût est le même quelle que soit la spécialisation se justifie par un niveau de qualification relativement homogène pour les différentes spécialisations.

Le nombre d'heures en travail régulier  $H_R$  par jour est supposé ne pas excéder 7,8 heures pour respecter le maximum des 39 heures hebdomadaires. Le nombre d'heures supplémentaires  $H_O$  par travailleur et par jour ne peut excéder 4 heures avec une limite hebdomadaire  $H_{OW} = 8$  heures.

**Niveaux de tolérance et seuils de qualité.** Comme nous l'avons discuté à la section 6.2, les activités sont groupées par niveau  $\theta$  de tolérance aux défauts pour éviter des surcroûts d'estimation des seuils minimum à respecter  $q_\theta^{\text{min}}$ . Tout seuil  $q_\theta^{\text{min}}$  se définit par le nombre d'items dont la non validation ne remet pas en cause les spécifications du client ou la fonctionnalité du produit ou lorsque ces items non validés n'affectent que faiblement l'esthétique des finitions. Un nombre limité de niveaux de tolérance se justifie par le fait que plusieurs activités peuvent être contrôlées par des check-lists présentant de fortes similarités.

**Proportion maximum de temporaires  $\lambda$ .** Si  $\lambda = 1$  alors il n'y a pas de limite à la proportion maximum de temporaires, les contraintes (6.19) ne sont pas activées. On pourra également considérer le cas où seules les heures supplémentaires sont utilisées pour accélérer le projet, excluant ainsi le recours au recrutement de temporaires ( $\lambda = 0$ ). D'après les données d'une étude de l'ESA (Economics and Statistics Administration) datant de 2012 (voir Nicholson, 2015 [111]), on trouve des valeurs de  $\lambda$  qui varient entre 0,05 et 0,25 selon différents secteurs d'activité. Nous pourrions également utiliser ces deux valeurs. On pourra également considérer  $\lambda = 0,50$  car ce n'est pas une proportion rare dans certaines entreprises, notamment dans les entreprises de service.

**Nombre de permanents affectés à chaque activité en mode normal  $P_{kj}$ .** si l'on adopte une approche de simulation extensive et que l'on ne dispose pas du nombre  $P_{kj}$  de permanents de spécialisation  $k$  alloués à toute activité  $j$ , alors on pourra tirer au hasard les  $P_{kj}$  (entre 1 et 5 par exemple) afin d'adapter des instances issues de PSPLIB (voir Kolisch et Sprecher, 1996 [92])

### 6.8.3 Résumé du programme d'optimisation

Dans l'annexe B, nous fournissons un résumé du programme d'optimisation où nous rappelons les notations employées.

**Programme de minimisation de la durée du projet sous contrainte de budget et de qualité minimum.** Un premier programme d'optimisation consiste à minimiser la durée totale d'exécution du projet donnée en (6.2) sous les contraintes de précédence (6.3), de non préemption (6.4) et du respect de la capacité en main d'oeuvre permanente (6.5). A ces contraintes s'ajoutent celle de la satisfaction des besoins en main d'oeuvre pour l'exécution de toute activité (6.8) avec  $w_{jt}$  défini par (6.31) et le respect des bornes inférieures et supérieures sur le nombre de travailleurs en heures régulières lors de l'exécution (6.10) et (6.11), avec les contraintes de linéarisation (6.25) et (6.26). S'ajoutent également les contraintes sur les heures supplémentaires données en (6.33), (6.34), (6.35) et (6.37). On a enfin les contraintes sur la qualité (6.53) avec  $\Delta q_{jt}$  défini par (6.51) et la contrainte de budget (6.55). On prend en compte les bornes supérieures sur la durée totale (6.56), sur la main d'oeuvre permanente (6.20) et temporaire (6.21) et sur les heures supplémentaires (6.30). Si l'on impose une proportion maximum de temporaires, on ajoutera la contrainte (6.19).

Si l'on suppose  $\lambda = 1$  alors les ressources sont en quantité illimitée du fait du recrutement possible de n'importe quel nombre de travailleurs temporaires et dans ce cas, une écriture de la contrainte 6.8) sous la forme d'une inégalité comme dans Naber et Kolisch (2014) [110], c'est-à-dire  $\sum_{t \in \mathcal{T}_j} w_{jt} \geq W_j$  peut conduire à affecter sur l'activité plus de travailleurs que nécessaire, avec des dégradations possibles de la qualité toutefois limitées par les contraintes de seuil de qualité minimum. Pour éviter ce type de solution qui peut être proposé par le programme d'optimisation, il faut définir une valeur pertinente pour  $\rho$ , le coefficient de dépassement du budget.

Dans le cas où  $\lambda = 0$  c'est-à-dire que l'on ne recourt pas au travail temporaire, il peut quand même y avoir sureffectif pour certaines activités puisque l'accélération résultant de l'utilisation des heures supplémentaires permet de libérer des permanents qui pourront alors s'ajouter à l'effectif normal de certaines activités à certaines

périodes.

Si l'on souhaite respecter l'organisation du travail qui prévaut en mode normal, on peut imposer la contrainte suivante

$$p_{jt} \geq P_{kj} \cdot z_{jt} \quad \forall j = 1..n \quad t \in \mathcal{T}_j, \quad (6.61)$$

qui viendrait remplacer la contrainte (6.11) mais qui autorise moins de flexibilité dans l'utilisation des ressources. En effet, la contrainte (6.61) impose au moins autant de permanents en mode accéléré qu'il y en a en mode normal, de sorte que ces permanents ne peuvent pas être remplacés par des temporaires. On pourra toujours examiner les solutions que l'on obtient avec l'une ou l'autre de ces deux contraintes (6.11) ou (6.61). Si, en imposant le respect de la contrainte (6.61), on obtient la même durée, le même coût et la même qualité qu'avec la contrainte (6.11), alors il est évident que d'un point de vue organisationnel, le gestionnaire de projet préférera une solution qui respecte la contrainte (6.61).

## 6.9 Conclusion

Le modèle que nous avons développé dans ce chapitre s'inscrit dans le cadre de décisions tactiques où l'on raisonne à capacité fixe de production (postes de travail et équipements prédéfinis) avec une planification d'un recrutement flexible en main d'oeuvre additionnelle pour accélérer le projet de sorte à obtenir un avantage concurrentiel s'il s'agit d'un appel d'offre, ou dans un objectif de maximisation du profit pour débiter au plus tôt une autre commande. L'accélération résulte d'une utilisation d'heures supplémentaires et de travailleurs en sureffectif et intègre son impact sur la productivité et la qualité. Le modèle permet ainsi de répondre à la décision tactique de l'allocation optimale de l'effectif dédié à un projet de fabrication pour en minimiser la durée tout en maîtrisant les coûts et en imposant à chaque activité une qualité acceptable de sorte qu'il exclut les reprises indésirables car coûteuses en temps et en matériaux.

Ce modèle pourrait être enrichi d'une contrainte de lissage de la main d'oeuvre pour faciliter le travail de supervision et éviter des tensions entre les travailleurs lorsqu'ils se trouvent en grand nombre sur un même poste. Cependant une telle contrainte peut oblitérer des possibilités d'accélération puisque le nombre de travailleurs disponibles varie d'une période à l'autre et un lissage pourrait induire une sous-utilisation de la main d'oeuvre à certaines périodes.

Par ailleurs, nous aurions pu intégrer des coûts fixes de recrutement et de formation des travailleurs temporaires et affecter également des coûts unitaires de main d'oeuvre différents selon les spécialisations.

Il aurait pu être pertinent de considérer également un effectif de supervision car en effet, il n'est pas rare que les entreprises limitent les heures supplémentaires en raison d'une réticence des superviseurs de chaîne à effectuer eux-mêmes des heures supplémentaires. Il en aurait résulté une contrainte additionnelle sur les heures supplémentaires limitées par celles que les superviseurs sont à même d'accepter.

En nous appuyant sur plusieurs résultats de la littérature (voir notamment Gunduz, 2004 [65]), nous avons supposé une relation linéaire entre sureffectif et perte de productivité. Mais cette relation prend peut-être d'autres formes selon l'industrie considérée aussi une validation empirique d'une telle relation aurait-elle été judicieuse. Une approximation linéaire permet néanmoins de produire une estimation pessimiste de l'impact du sureffectif sur les pertes de productivité ce qui peut conduire à un surdimensionnement du nombre de temporaires dont on planifie le recrutement mais cela peut toutefois se corriger au niveau opérationnel, par un recrutement moindre en considérant un horizon temporel plus court (de l'ordre de la semaine par exemple).

Une remarque similaire peut être faite concernant l'hypothèse de linéarité entre perte de qualité et sureffectif et heures supplémentaires, où cette perte de qualité est proportionnelle au nombre de travailleurs en sureffectif et au nombre d'heures supplémentaires. Mais là encore, la linéarité fournit une estimation de type worst-case et sans redimensionner la main d'oeuvre additionnelle au niveau opérationnel, la qualité réellement observée pourra s'établir à des valeurs supérieures à celles produites par le modèle, ce qui ne peut nuire au projet de fabrication.

Ces deux hypothèses de linéarité ont néanmoins l'avantage de limiter le nombre de variables utilisées pour les linéarisations, diminuant de fait la complexité du programme d'optimisation et l'autorisant ainsi à produire plus rapidement une estimation de la main d'oeuvre planifiée pour l'accélération. Dans le chapitre suivant, nous appliquons notre modèle de planification de la main d'oeuvre additionnelle au cas d'un projet de fabrication d'une motrice TGV.

# Chapitre 7

## Etude de cas relative au projet de fabrication d'une motrice TGV

Dans ce chapitre, nous appliquons le modèle d'optimisation de la main d'oeuvre additionnelle au cas d'un projet de fabrication d'une motrice à très grande vitesse (TGV) dont le diagramme simplifié est donné à la figure 7.1. Pour la collecte des données et des informations relatives à ce projet, nous nous sommes adressés aux gestionnaires du site industriel d'Alstom Transport de Belfort et notamment au directeur industriel du site, au responsable qualité production ainsi qu'à des techniciens et des cadres industriels.

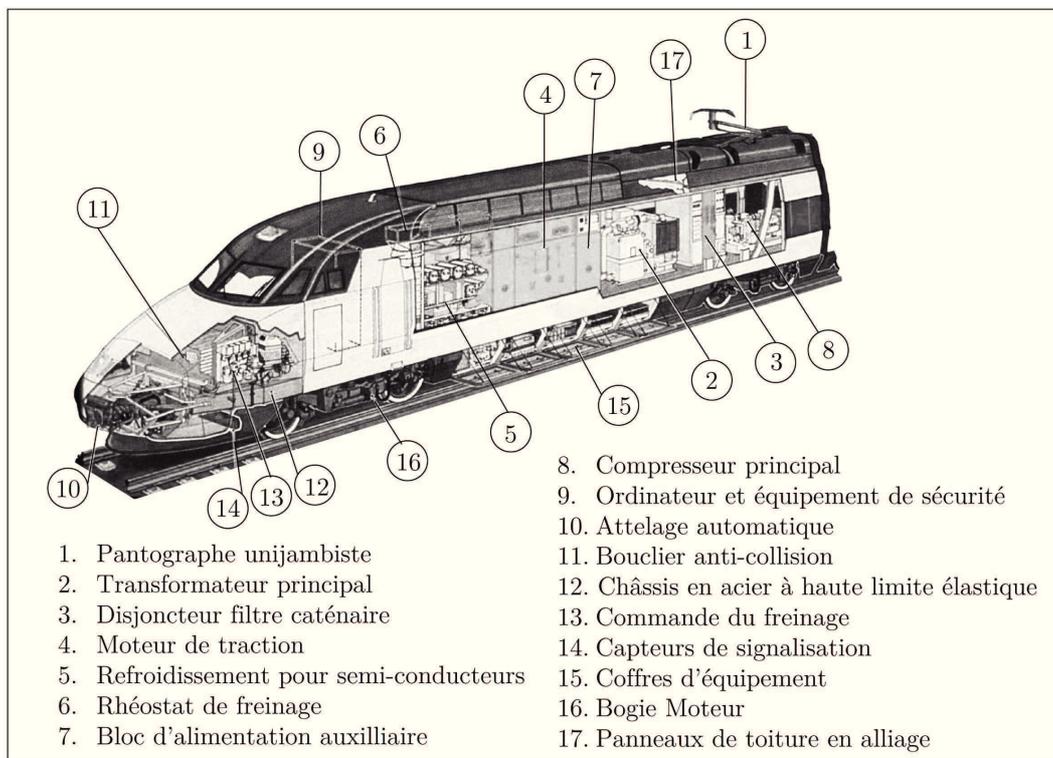


FIGURE 7.1 – Diagramme simplifié d'une motrice TGV

Dans une première section (section 7.1) nous décrivons ce projet de fabrication et en fournissons les données de base comme par exemple les relations de précédence entre les activités ou le nombre de spécialisations professionnelles requises. La section 7.2 propose une méthodologie pour affecter aux activités des niveaux de tolérance aux défauts que nous définissons. Nous développons dans la section 7.3 le raisonnement que nous avons mené pour fixer les paramètres de la simulation dont nous analysons les résultats dans la section 7.4. Enfin, la section 7.5 conclut en indiquant notamment des suggestions de recherche future. Nous rappelons qu'un résumé du programme d'optimisation se trouve dans l'annexe B (voir page 213).

## 7.1 Description et données du projet de fabrication de la motrice

On rappelle qu'une activité est un ensemble d'opérations effectuées sur un même poste de travail. Le projet de fabrication d'une motrice TGV comprend plus de 500 opérations de fabrication réparties en 37 activités. Ces activités sont liées par des contraintes de précédence strictes et sont exécutées sur 13 postes de travail différents en traversant un processus à quatre étapes (ateliers) : la chaudronnerie où l'on exécute le montage des sous-ensembles bruts comme le châssis et la caisse, la peinture où l'on peint la motrice, l'assemblage où l'on exécute les activités de câblage et de montage des pièces achetées (pièces peintes comme le carénage et les portes ou pièces pré-assemblées comme les sièges et les cloisons) et enfin les essais où différents types de contrôle sont effectués (comme les essais dynamiques, pneumatiques et électriques). La figure 7.2 présente ce processus de fabrication à quatre ateliers ainsi que le nombre de postes et la durée d'exécution par atelier de fabrication. La durée normale du projet est de 107 jours ( $T = 107$ ), et calculée sur la base des durées normales des activités. La durée normale d'une activité  $j$  est mesurée depuis son entrée sur le poste où elle est exécutée jusqu'à son entrée sur le poste suivant où s'effectue l'activité  $i$  qui lui succède. Cette durée comprend ainsi le temps opératoire, le temps de contrôle qualité et le temps d'acheminement de l'activité vers le poste suivant.

On considère ainsi un ensemble  $V = \{0, 1, \dots, 37\}$  d'activités où les activités 0 et 37 désignent respectivement les activités fictives de début et de fin du projet. Chacune des activités du projet ne requiert pour son exécution qu'une seule spécialisation  $k$  parmi sept spécialisations professionnelles que le projet mobilise, soit  $k = 1, 2, 3, 4, 5, 6, 7$  qui correspondent respectivement aux spécialisations en soudure, peinture, assemblage, câblage, électricité, pneumatique, automatique-électronique. Nous avons donc sept sous-ensembles  $\{V_k\}_{k=1..7}$  d'activités où le sous-ensemble  $V_k$  groupent les activités qui nécessitent la spécialisation  $k$ . Toute activité n'est exécutée que sur un seul poste. Des postes différents peuvent accueillir une même spécialisation (il existe ainsi deux postes de soudure) et ceux-ci se distinguent par des équipements et une configuration propre à chacun. En revanche un même poste ne réunit jamais simultanément des spécialisations différentes.

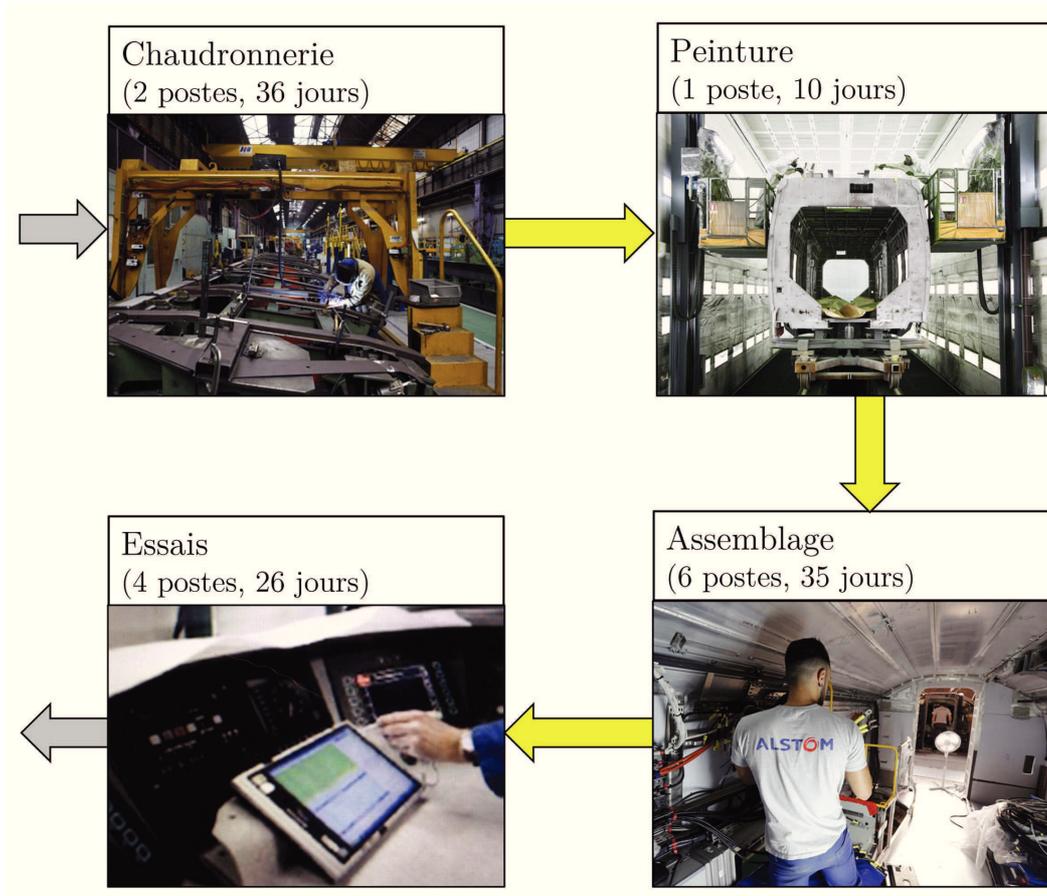


FIGURE 7.2 – Processus de fabrication d'une motrice TGV

Le tableau 7.1 liste les activités du projet de fabrication de la motrice et en donne une description. Pour chaque activité  $j$ , le tableau donne l'ensemble de ses prédécesseurs  $\Gamma^{-1}(j)$ , celui de ses successeurs  $\Gamma(j)$ , la spécialisation professionnelle qu'elle nécessite, sa durée normale  $d_j^{\max}$ , l'effectif  $P_{kj}$  de permanents en mode normal et le nombre maximum  $A_j$  de travailleurs qu'il est possible d'allouer à son exécution et qui dépend de la capacité maximale du poste requis. Pour les activités dont la durée normale est d'un jour, soit  $d_j^{\max} = 1$  alors on a nécessairement  $A_j = P_{kj}$  car ces activités ne peuvent être accélérées puisqu'elles ont déjà une durée minimum. La capacité du poste sur lequel ces activités sont exécutées peut quant à elle excéder  $A_j$ . C'est la raison pour laquelle on fait la distinction entre capacité du poste et nombre maximum de travailleurs  $A_j$ .

$j$	Description de l'activité	Préd.	Succ.	Spécialisation		Durée		
		$\Gamma^{-1}(j)$	$\Gamma(j)$	Nom	$k$	$d_j^{\max}$	$P_{kj}$	$A_j$
0	Début	$\emptyset$	1	–	–	0	–	–
1	Op. 1 Montage châssis	0	2	Soudure	1	7	4	5
2	Op. 2 montage châssis	1	3	”	1	7	4	5
3	Montage cathédrale gauche	2	4	”	1	7	3	4
4	Montage cathédrale droite	3	5	”	1	7	3	4
5	Assemblage caisse	4	6	”	1	8	3	4
6	Peinture	5	7	Peinture	2	10	3	4
7	Tuyauterie parties méca.	6	9	Assembl.	3	4	4	5
8	Pré-câblage haute tension	0	9	Câblage	4	9	1	2
9	Câblage haute tension	7, 8	10	Assembl.	3	1	4	4
10	Plancher + conduits	9	12	”	3	1	4	4
11	Précâblage basse tension	0	12	Câblage	4	24	1	2
12	Câblage basse tension	10, 11	13, 14	Assembl.	3	2	4	5
13	Câblage partie avant	12	15	”	3	1	2	2
14	Equip. carénage + trappes	12	15	”	3	1	3	3
15	Carénage + baie	13, 14	16	”	3	1	5	5
16	Pupitre + habillage cabine	15	17	”	3	1	5	5
17	Cloison et armoire	16	18	”	3	1	5	5
18	Blocs intérieurs	17	19	”	3	2	5	6
19	Moteur et appareils fosse	18	20, 21	”	3	2	5	6
20	Raccordement pour capots	19	22	Eléctric.	5	6	1	2
21	Equipement des capots	19	22	”	5	6	1	2
22	Montage capots	20, 21	23	”	5	2	2	3
23	Câblage/équillage	22	24	”	5	1	2	2
24	Equipement ligne de toiture	23	25	”	5	1	2	2
25	Essai d'air	24	26	Assembl.	3	1	2	2
26	Pose des cales	25	27	”	3	1	2	2
27	Préparation mise sur bogie	26	28	”	3	2	2	3
28	Mise sur Bogie + raccordement	27	29	”	3	2	2	3
29	Persiennes	28	30	”	3	3	2	3
30	Essais basse tension	29	31	Eléctric.	5	11	2	2
31	Essais haute tension	30	32	”	5	2	2	2
32	Essais pneumatique	31	33	Pneum.	6	4	2	2
33	Essais équipements de sécurité	32	34	Autom.	7	5	2	2
34	Essais fournisseur Siemens	33	35	”	7	2	2	2
35	Essais dynamique	34	36	”	7	1	2	2
36	Essais mécanique	35	37	”	7	1	2	2
37	Fin	36	$\emptyset$	–	–	0	–	–

Tableau 7.1 – Données des activités du projet de fabrication de la motrice TGV

Toutes ces données proviennent des plannings réels de différents ateliers du processus de fabrication. Les couleurs attribuées aux indices des activités correspondent à leur niveau de tolérance aux défauts que nous analysons dans la section 7.2. Ainsi on distingue trois niveaux de tolérance avec en rouge les activités qui ne tolèrent aucun défaut, en bleu les activités qui en tolèrent le plus, et en noir les activités pour lesquelles cette tolérance est intermédiaire. Pour toute activité  $j$ , on suppose que sa qualité en mode normal est égale à 1, soit  $q_j^{\max} = 1$ , ce qui revient à admettre une probabilité nulle d'occurrence de défauts lorsque l'activité n'est pas accélérée. C'est une hypothèse standard dont nous n'avons pu faire l'économie, ne disposant pas du suivi des défauts chez Alstom ; ces défauts pouvant également survenir en mode normal, diminuant ainsi la qualité en mode normal  $q_j^{\max}$  à une valeur inférieure à 1.

La figure 7.3 représente le diagramme de Gantt du projet avec un ordonnancement au plus tard des activités qui sont placées en ordonnée, et les jours, en abscisse. A chacun des treize postes est associée une couleur. Ainsi, les deux premières activités sont exécutées sur le premier poste de soudure, les trois activités suivantes sont exécutées sur le deuxième poste de soudure, la sixième activité est réalisée sur le poste unique de peinture, etc.

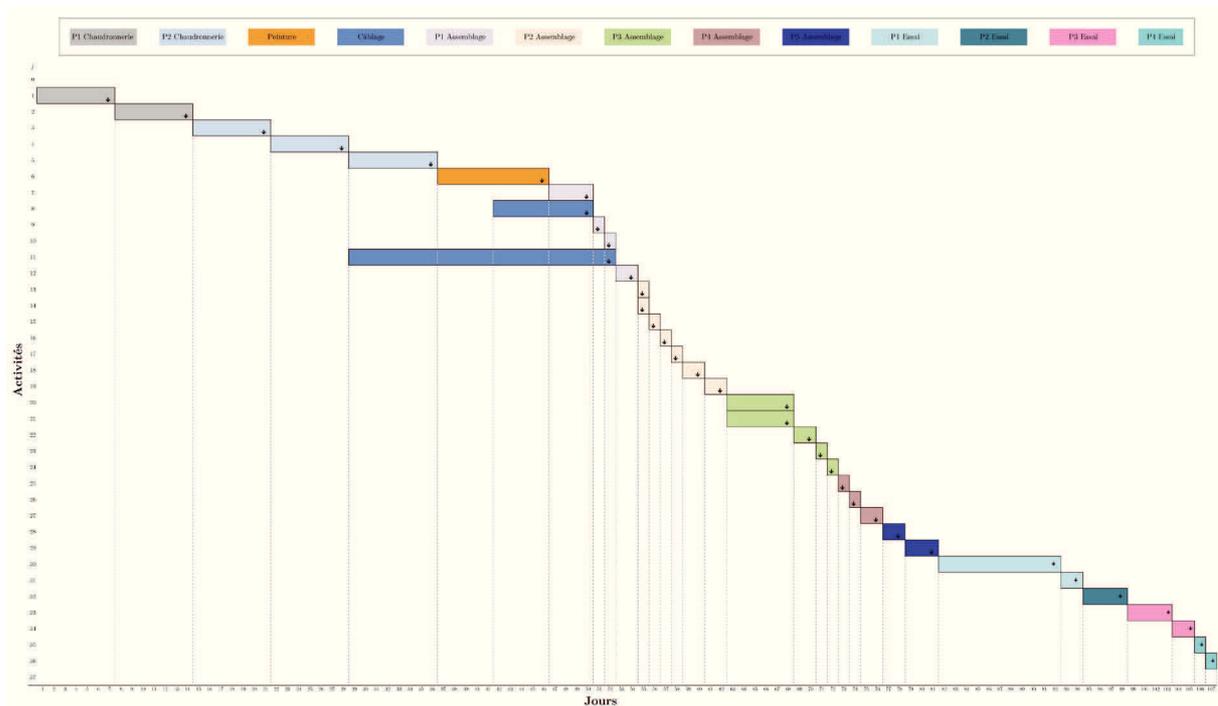


FIGURE 7.3 – Diagramme de Gantt du projet en mode normal - ordonnancement au plus tard

**Coûts unitaires de main d'oeuvre et nombre maximum d'heures de travail.** Un ouvrier permanent qualifié chez Alstom est payé en moyenne 13 euros nets de l'heure, ce qui représente un coût horaire pour l'entreprise  $c_P = 22,52$  euros. Pour le recrutement des temporaires, Alstom passe systématiquement par une agence de sorte que l'heure de travail temporaire lui coûte 20% plus cher qu'une heure de

travail permanent soit  $c_U = 1,2 \cdot c_P = 27,02$  euros. Une heure supplémentaire chez Alstom coûte 25% de plus qu'une heure normale, soit  $c_O = 1,25 \cdot c_P = 28,15$  euros. Une journée de travail régulier comprend  $H_R = 7,8$  heures (soit 39 heures par semaine). Selon la législation en vigueur, les heures supplémentaires par travailleur sont limitées à  $H_O = 4$  par jour et à  $H_{OW} = 8$  par semaine.

**Nombre de permanents affectés au projet.** Dans le tableau 7.1, nous avons consigné le nombre de permanents alloués à l'exécution de chaque activité en mode normal, c'est-à-dire  $P_{kj}$ . Le nombre de permanents  $P_k$  de chaque spécialisation chez Alstom qu'il convient de mobiliser pour le projet est tel qu'il permet d'en assurer la fabrication en mode normal selon un ordonnancement au plus tôt ou au plus tard. Ainsi, le calcul du nombre de permanents  $P_k$  par spécialisation dédié au projet utilise les paramètres  $P_{kj}$  selon l'équation (6.57) donnée au chapitre 6 (voir page 155). Ces nombres de permanents  $P_k$  pour chacune des spécialisation  $k = 1..7$  sont les suivants :  $P_1 = 4$  soudeurs,  $P_2 = 3$  peintres,  $P_3 = 5$  assembleurs,  $P_4 = 2$  câbleurs,  $P_5 = 2$  électriciens,  $P_6 = 2$  pneumaticiens et  $P_7 = 2$  automaticiens-électroniciens.

## 7.2 Niveaux de tolérance aux défauts et seuils de qualité minimum

Chaque activité possède un niveau de tolérance aux défauts auquel correspond un seuil de qualité minimum à atteindre pour éviter les reprises. Les activités peuvent être groupées par niveau semblable de tolérance. Pour le projet considéré, nous avons identifié trois niveaux de tolérance aux défauts  $\theta = 1, 2, 3$ .

1. **Niveau nul de tolérance aux défauts ( $\theta = 1$ ).** Les activités de cet ensemble  $V^1$  permettent de garantir que le produit fonctionne correctement avec une évaluation nécessairement binaire de la qualité (cela fonctionne ou non). Aucun défaut ne peut alors être toléré car le défaut est l'expression en soi de la non fonctionnalité ou non conformité et implique donc une reprise. Pour le projet considéré, cet ensemble regroupe toutes les activités d'essais comme par exemple les essais électriques haute et basse tension, ainsi on a  $V^1 = \{25, 30, 31, 32, 33, 34, 35, 36\}$ ; ces activités étant repérées par leur indice en rouge dans le tableau 7.1.
2. **Niveau intermédiaire de tolérance aux défauts ( $\theta = 2$ ).** L'ensemble  $V^2$  rassemble des activités fonctionnelles uniquement pour lesquelles un petit seuil de tolérance aux baisses de qualité est autorisé sans que cela remette en cause la fonctionnalité du produit fini ni les exigences du client. L'indice de ces activités apparaît en noir dans le tableau 7.1, ainsi  $V^2 = \{1, 2, 3, 4, 5, 7, 8, 9, 10, 11, 12, 13, 15, 16, 17, 18, 19, 20, 22, 23, 28, 29\}$ .
3. **Niveau élevé de tolérance aux défauts ( $\theta = 3$ ).** L'ensemble  $V^3$  comprend non seulement des activités fonctionnelles mais également des activités de finition. Pour les activités de finition comme la peinture, certains défauts peuvent être acceptés par demande de dérogation au client qui revoit alors à la baisse ses exigences de sorte que l'activité n'est pas reprise car conforme aux nouvelles spécifications. Les activités fonctionnelles de ce niveau de tolé-

rance regroupent souvent des opérations de montage de composants sur des sous-assemblages non visibles par le client et pour lesquels on peut alors tolérer des défauts d'aspect, comme c'est le cas par exemple pour le montage d'équipements sous le carénage et à l'intérieur des trappes (activité 14). Ces activités sont indiquées par un indice en bleu dans le tableau 7.1 ; on a ainsi  $V^3 = \{6, 14, 21, 24, 26, 27\}$ .

Toutes les activités hors activité d'essais font l'objet d'un auto-contrôle par l'opérateur lui-même à la fin de certaines opérations. Il s'agit par exemple de vérifier que ces opérations sont conformes aux instructions figurant dans la gamme opératoire fournie par le bureau des méthodes. Ces instructions peuvent se présenter sous la forme d'un relevé dimensionnel notamment en chaudronnerie (on compare la mesure d'une pièce à sa cote assortie d'une tolérance) ou d'un test de continuité électrique pour des opérations de câblage sur des sous-assemblages (cela fonctionne ou non). L'évaluation de la qualité de ces opérations est donc binaire et les non conformités sont systématiquement reprises. Ce processus d'auto-contrôle et les reprises éventuelles sur les opérations font partie d'une évaluation basique de la qualité et les durées normales d'exécution des activités tiennent compte du temps de contrôle et de ce type de reprise sur opérations. Les activités d'essais ne font pas l'objet d'un auto-contrôle car elles consistent à tester la fonctionnalité d'un produit semi-fini ou fini et non à y exécuter des opérations. La qualité des activités d'essais correspond finalement à la qualité du produit testé et est nécessairement évaluée par un technicien qualité.

Outre l'auto-contrôle qui permet une évaluation basique de la qualité des activités (hors activités d'essais) par validation de la conformité d'un certain nombre d'opérations, il existe une deuxième étape plus fine de contrôle final, la visite qualité effectuée par un technicien qualité. Ce contrôle final est fait au moyen d'une check-list pour chaque activité. Cette étape consiste à vérifier par exemple les montages, la présence de toutes les pièces, les aspects des pièces, etc. La check-list de contrôle intègre non seulement les spécifications techniques du client mais aussi les exigences internes du site de Belfort et du groupe Alstom. Ces exigences sont définies par un système d'audit interne appelé APSYS (Alstom Production System) dont l'un des objectifs est l'amélioration du processus qualité pour atteindre un certain niveau d'excellence sans faire de la surqualité. La surqualité, c'est-à-dire une qualité au delà des exigences du client et de l'entreprise, pénalise cette dernière en générant des coûts de reprises non nécessaires eu égard aux spécifications ou des surcoûts de main d'oeuvre si les opérateurs passent du temps à parfaire inutilement le travail.

Des mesures de défauts sont faites par une cotation issue d'une AMDEC (Analyse des Modes de Défaillance, de leurs Effets et de leur Criticité) ; une démarche collective menée par un groupe de travail qui identifie pour chaque activité les défaillances (défauts) possibles et détermine pour chacun d'eux, leurs effets et leur gravité. Chez Alstom, la gravité est mesurée sur une échelle de 1 à 10 et la reprise est obligatoire pour une gravité strictement supérieure à 5. Si la gravité du défaut est inférieure ou égale à 5, le défaut peut être accepté sans effet sur les opérations suivantes ni sur les qualités individuelles des activités suivantes. Les défauts de faible gravité sont en général des défauts d'aspect qui altèrent superficiellement l'esthétique du produit tandis que les défauts graves touchent à la fonctionnalité, au non

respect des spécifications du client ou à l'environnement (composition chimique des peintures par exemple). C'est sur la base de ces check-lists que nous définissons les seuils de qualité minimum  $q_{\theta}^{\min}$ .

**Seuil de qualité minimum pour les activités de niveau nul de tolérance.**

Comme la qualité de ces activités est évaluée de manière binaire, le seuil de qualité minimum correspond tout simplement à la qualité maximum obtenue en mode normal et qui vaut 1 indiquant que l'activité fonctionne correctement ; autrement dit, tous les items de la check-list doivent être validés. Pour ces items, la notion de niveau de gravité devient donc inutile. Ces activités ne sont jamais accélérées car l'accélération entraîne un risque de défaillance trop élevé pour que l'entreprise s'autorise à le prendre. Cela revient à supposer que la probabilité d'occurrence d'un défaut en cas d'accélération est égale à 1. Ainsi, pour toute activité  $j \in V^1$ , le nombre maximum de permanents ( $A_j$ ) qu'il est possible de lui allouer correspond au nombre  $P_{kj}$  de permanents qui lui est affecté en mode normal. C'est le cas par exemple de l'activité  $j = 31$  d'essais électriques haute tension qui requiert la spécialisation électricité ( $k = 5$ ) et pour laquelle on a  $P_{5,31} = A_{31} = 2$  (voir tableau 7.1). En conclusion, les activités de niveau nul de tolérance ( $\theta = 1$ ) ont par conséquent un seuil minimum de qualité  $q_{\theta}^{\min} = q_1^{\min} = 1$ . La figure 7.4 présente un extrait de la check-list pour l'activité 33 d'essais d'équipement de sécurité et qui concerne la vérification du bon fonctionnement de l'éclairage de sécurité.

ABCD RGV		GAMME DE CONTROLE ESSAI EQUIPEMENTS DE SECURITE			
Zone & Montage	Description de la tâche d'inspection	Critère d'inspection	Moyen de contrôle	localisati	Référence fiche de critère
<b>Equipements de sécurité</b>					
<b>Eclairage secours</b>					
	+ En armoire ACBT de la remorque DC :  ✓ Mettre le commutateur « secours radio » Z (SE) RAD sur la position « Soccorsi »  ✓ En armoire ABT de la remorque DC, déclencher CC1 (ECL) ( BA )  ✓ Enclencher CC1 (ECL)  ✓ Répéter l'opération précédente sur les remorques MD et AC	Mise en service de l'éclairage secours dans la quadriplette.  Extinction, dans la remorque DC, de l'éclairage secours alimenté par la ligne 751A → voir synoptique de la remorque. L  Même vérification  Même vérification	Manuel  Manuel  Manuel  Manuel	3-Essai  3-Essai  3-Essai  3-Essai	N.A.  N.A.  N.A.  N.A.
	+En armoire ACBT de la remorque DC :  ✓ Mettre le commutateur « secours radio » Z (SE) RAD sur la position « Normale »	Extinction de l'éclairage secours dans la quadriplette	Manuel	3-Essai	N.A.

FIGURE 7.4 – Exemple d'une check-list pour une activité ayant un niveau nul de tolérance aux défauts

**Seuils de qualité minimum pour les activités de niveau non nul de tolérance.** Pour illustrer le raisonnement, supposons tout d'abord que chaque activité  $j$  possède un seuil de tolérance  $q_j^{\min}$  qui lui est propre. A cette activité est associée une check-list comprenant au total  $L_j$  items parmi lesquels il existe un nombre  $l_j$  d'items traduisant des défauts acceptables (gravité inférieure ou égale à 5). Nous supposerons alors que le seuil de qualité minimum est égal à la proportion de défauts non acceptables dans la liste, c'est-à-dire  $q_j^{\min} = 1 - (l_j/L_j)$ . Implicitement, cela revient à supposer que tout défaut a la même probabilité d'occurrence, ce qui n'est pas nécessairement le cas. Cette hypothèse est néanmoins cohérente avec les pertes de qualité  $\Delta q_j^S$  et  $\Delta q_j^O$  associées à l'utilisation de main d'oeuvre additionnelle parce qu'il s'agit de pertes de qualité globale sur l'activité, sans distinction de l'impact précis de cette main d'oeuvre sur chacun des items de la liste. Autrement dit, la main d'oeuvre additionnelle affecte uniformément n'importe lequel des items de la check-list. L'introduction dans la modélisation de probabilités d'occurrence de chacun des défauts possibles conduirait à un autre type de modélisation (probabiliste) qui nécessiterait de surcroît une évaluation fine des fréquences des défauts dans toutes les situations possibles, avec et sans main d'oeuvre additionnelle, et ce, pour estimer les probabilités d'occurrence des défauts en mode normal et dans les différents modes d'accélération. En outre, nous n'avons pu obtenir d'Alstom le suivi des fréquences de défauts et même si nous l'avions obtenu, il aurait fallu connaître les conditions de collecte et de survenue de ces défauts (mode normal, accéléré et de quelle manière ; lié ou non à la main d'oeuvre, etc.).

Finalement, à partir des données de check-lists dont nous avons pu disposer, nous avons divisé en deux groupes les activités de niveau non nul de tolérance. L'ensemble  $V^2$  contient les activités auxquelles sont associées des check-lists comprenant en moyenne 5 items avec 1 seul défaut acceptable de sorte que  $q_2^{\min} = 4/5$  et l'ensemble  $V^3$  inclut les activités avec 3 défauts acceptables en moyenne sur un total de 9 items d'où  $q_3^{\min} = 2/3$ .

Pour illustrer, le tableau 7.2 présente le contenu des check-lists de trois activités possédant un niveau intermédiaire de tolérance aux défauts ( $\theta = 2$ ). Il s'agit des activités 1 (montage châssis), 12 (montage câblage basse tension) et 29 (montage modules). Nous avons choisi ces activités parce qu'elles nécessitent chacune une spécialisation professionnelle différente et leurs check-lists sont représentatives de celles des autres activités dans  $V^2$ . Chacune de ces trois activités  $j = 1, 12, 30$  possède une check-list avec cinq items à vérifier et un seul défaut acceptable ( $q_2 = 0,8$ ). Pour chacune des check-lists, on trouve dans le tableau 7.2 la description des items, les défauts correspondants, leur effet et leur niveau de gravité. Prenons l'exemple de l'activité 1 dont on évalue la qualité de la soudure (item 1). Si la soudure est collée ou mal pénétrée, le produit pourra présenter des problèmes d'étanchéité (effet). On ne peut donc tolérer pareil défaut dont la gravité est maximale (10). En revanche, si la soudure présente un défaut d'aspect (item 2) telle une porosité, l'effet est un risque très faible de non étanchéité de sorte qu'on peut accepter ce défaut dont la gravité est égale à 2. Pour l'activité 12, si l'état des câbles (item 2) est tel qu'ils présentent quelques rayures peu visibles (défaut d'aspect) et sans conséquences, l'activité reste conforme mais si le défaut est une blessure profonde des câbles (item 3) qui risque fortement d'affecter le fonctionnement du produit fini,

une reprise est nécessaire.

Le tableau 7.3 donne la check-list de l'activité de peinture ( $j = 6$ ) dont le niveau de tolérance aux défauts est élevé ( $\theta = 3$ ). Sur les neuf items à contrôler, on peut tolérer trois défauts parce qu'ils ont seulement des effets esthétiques. On a ainsi  $q_\theta^{\min} = q_3^{\min} = 0,67$ . Si par exemple la peinture présente un défaut mineur de teinte (item 3), ce défaut peut être accepté par le client à qui on adresse préalablement une demande de dérogation. L'activité sera alors considérée comme conforme aux exigences du client qu'il aura revues à la baisse. D'autres défauts de peinture demeurent inacceptables, aucune demande de dérogation n'est envisageable ni souhaitée par l'entreprise, ce qui est le cas par exemple des débordements de peinture qui doivent nécessairement être corrigés (item 1 dans le tableau, zone de peinture non respectée).

### 7.3 Valeurs des paramètres dans l'expérience de simulation

**Coefficient  $\gamma_j$  d'impact du sureffectif sur la productivité.** On note en premier lieu que la définition du sureffectif chez Alstom est la même que celle que nous avons adoptée : l'activité est exécutée en sureffectif dès lors qu'il lui est affecté un nombre de travailleurs strictement supérieur à l'effectif  $P_{kj}$  en mode normal. Par ailleurs, les gestionnaires de production confirment qu'il est raisonnable de supposer que les pertes de productivité sont proportionnelles au sureffectif. Ils affirment également que la productivité d'un temporaire à l'issue de sa formation est égale à celle d'un permanent. En revanche, Alstom ne dispose pas de relevés de durées observées d'exécution des activités lorsqu'il y a sureffectif. Il n'est donc pas possible de produire une estimation des coefficients d'impact  $\gamma_j$ . Pour affecter des valeurs pertinentes à ces coefficients nous avons alors exploité l'inégalité suivante et obtenue dans le chapitre précédent que nous rappelons ici

$$\bar{w}_j \leq \min\left(\left\lfloor \frac{1 + \gamma_j}{2\gamma_j} P_{kj} \right\rfloor, d_j^{\max} \cdot P_{kj}, A_j\right) = s_j \quad \forall j = 1..n. \quad (7.1)$$

Autrement dit, l'effectif maximum ne peut dépasser la valeur de  $s_j$ . On peut montrer facilement que  $s_j = P_{kj}$  quelle que soit la valeur de  $\gamma_j$  pour toutes les activités qui ne peuvent pas être accélérées, c'est-à-dire les activités dont la durée normale est d'un jour ou celles dont le niveau de tolérance aux défauts est nul et pour lesquelles on a par conséquent  $A_j = P_{kj}$ . Ainsi, pour ces activités l'effectif maximum est égal à l'effectif en mode normal, peu importe la valeur de  $\gamma_j$ . Nous fixerons celle-ci arbitrairement à zéro pour toutes ces activités.

<i>j</i>	Item	Description	Défaut possible	Effet	Gravité
1	1	Qualité de la soudure	Soudure collée ou mal pénétrée	Fissures et non étanchéité du produit	10
	2	Aspect de la soudure	Soudure poreuse	Risque très bas de non étanchéité	2
	3	Alignement des composants	Non respect des tolérances	Châssis désaxé	10
	4	Symétrie des composants	Inversion gauche et droite	Gêne dans les opérations suivantes	10
	5	Planéité des composants	Déformation du châssis	Difficulté dans le calage moteur	10
12	1	Test de continuité	Pas de continuité	Gêne dans les opérations suivantes	10
	2	Etat des câbles	Rayure	Esthétique	3
	3	Etat des câbles	Blessure profonde	Risque de dysfonctionnement	10
	4	Longueur des câbles	Ecart dimensionnel	Gêne dans les opérations suivantes	7
	5	Montage des câbles	Câbles mal placés	Gêne dans les opérations suivantes	9
29	1	Réglage du montage	Non conformité géométrique	Esthétique	2
	2	Serrage des persiennes	Vis non serrées (une ou plusieurs)	Risque de chute des persiennes	8
	3	Symétrie des persiennes	Inversion gauche et droite	Dysfonctionnement des persiennes	8
	4	Teinte des persiennes	Defaut de teinte	Non respect des spécifications du client	9
	5	Contact entre persiennes	Déformation et mauvais contact	Dysfonctionnement des persiennes	9

Tableau 7.2 – Check-list qualité de trois activités à niveau intermédiaire de tolérance aux défauts

<i>j</i>	Item	Description	Défaut possible	Effet	Gravité
6	1	Zone de peinture	Zone de peinture non respecté	Non respect des spécifications client	10
	2	Teinte	Défaut majeur de teinte	Non respect des spécifications client	9
	3	Teinte	Défaut mineur de teinte localisé	Esthétique	3
	4	Finition de la résine	Couche de résine non uniforme	Non respect des spécifications client	8
	5	Conformité de la résine	Résine manquante	Non respect des spécifications client	8
	6	Etat de la peinture	Absence de peinture	Non respect des spécifications client	9
	7	Etat de la peinture	Ecaillage invisible	Esthétique	3
	8	Etat de la peinture	Rayures	Esthétique	2
	9	Composition de la peinture	Non respect de % chimiques	Non respect des spécifications client	9

Tableau 7.3 – Check-list qualité pour une activité de tolérance élevée aux défauts (peinture)

Pour les autres activités du projet, on constate par le calcul que  $\min(d_j^{\max} \cdot P_{kj}, A_j) = A_j$  de sorte que  $s_j$  prend l'expression simplifiée  $s_j = \min\left(\left\lfloor \frac{1+\gamma_j}{2\gamma_j} P_{kj} \right\rfloor, A_j\right)$ . On peut déterminer la valeur  $\gamma_j^*$  du coefficient d'impact pour laquelle  $s_j = A_j$  en résolvant simplement  $\frac{1+\gamma_j}{2\gamma_j} P_{kj} = A_j$ , ce qui donne  $\gamma_j^* = \frac{P_{kj}}{2A_j - P_{kj}}$ . Cette valeur  $\gamma_j^*$  est la valeur maximum de l'impact du sureffectif sur la productivité qui autorise encore l'utilisation du sureffectif maximum. Au delà de cette valeur, c'est-à-dire pour  $\gamma_j > \gamma_j^*$  il n'y aura plus la possibilité d'accélérer l'activité avec un sureffectif. Nous poserons alors

$$\gamma_j = \psi \cdot \frac{P_{kj}}{2A_j - P_{kj}}, \quad (7.2)$$

où le paramètre  $\psi$  traduit un niveau moyen ou global de sensibilité de la productivité au sureffectif. Nous considérons quatre valeurs pour ce paramètre, soit  $\psi = \{0,05; 0,25; 0,50; 0,75\}$  pour traduire respectivement un impact quasi négligeable, faible, moyen et fort du sureffectif sur la productivité.

**Perte de qualité  $\Delta q_j^S$  résultant de l'emploi d'un travailleur en sureffectif par jour et perte de qualité  $\Delta q_j^O$  associée à l'utilisation d'une heure supplémentaire.**

Comme Alstom ne dispose pas de relevés de qualité observée en mode accéléré, nous nous appuyons sur l'information suivante que nous avons pu recueillir auprès du responsable qualité production. Ce responsable qualité estime qu'à partir d'un certain pourcentage  $\varphi$  de sureffectif de l'ordre de 40%, la qualité tombe en dessous du seuil minimum acceptable, ce qui implique des reprises. A toute période d'exécution on a alors en moyenne un effectif  $p_{jt} + u_{jt} = p_j + u_j = (1 + \varphi)P_{kj}$ , ce qui donne une durée

$$d_j(\varphi) = \frac{W_j}{w_j^R ((1 + \varphi)P_{kj})}, \quad (7.3)$$

et une qualité associée

$$q_j(\varphi) = 1 - d_j(\varphi) \cdot ((1 + \varphi)P_{kj} - P_{kj}) \cdot \Delta q_j^S. \quad (7.4)$$

Cette qualité  $q_j(\varphi)$  doit être égale à  $q_\theta^{\min}$  puisqu'au delà du pourcentage  $\varphi$  de sureffectif Alstom affirme que le risque de reprise est très élevé. Il suffit alors de résoudre  $q_j(\varphi) = q_\theta^{\min}$  en  $\Delta q_j^S$  pour obtenir une estimation de ce paramètre, ce qui donne

$$\Delta q_j^S = \frac{1 - q_\theta^{\min}}{d_j(\varphi) \cdot \varphi \cdot P_{kj}} \quad \forall j \in V^\theta. \quad (7.5)$$

Cette estimation de la perte de qualité  $\Delta q_j^S$  dépend non seulement de la valeur de  $\varphi$  mais également de celle de  $\gamma_j$  qui entre dans le calcul de la durée  $d_j(\varphi)$  via la main

d'oeuvre corrigée des pertes de productivité. Ainsi, chaque fois que l'on fait varier  $\gamma_j$  par une variation de  $\psi$  (équation (7.2)) on obtient une nouvelle estimation de  $\Delta q_j^S$  à paramètre  $\varphi$  fixé. Par conséquent, dans les simulations, on commence d'abord par calculer  $\gamma_j$  puis  $\Delta q_j^S$  en fonction de la valeur de  $\varphi$ . Pour les simulations, nous prendrons  $\varphi = 0,4$  conformément à l'estimation du responsable qualité production.

Par ailleurs, les gestionnaires du site industriel d'Alstom Transport de Belfort confirment l'hypothèse de linéarité de la relation de décroissance entre qualité et main d'oeuvre additionnelle. De plus, ils affirment qu'à l'issue d'une période de formation, la qualité du travail d'un temporaire rejoint celle d'un permanent. Cette période de formation s'achève après la validation par un technicien qualité de trois à quatre montages que le temporaire a exécuté et contrôlé seul. La longueur de cette période est relativement restreinte car les temporaires travaillent généralement dur et de façon appliquée. Dans la mesure du possible et pour limiter le coût et la durée de la formation, Alstom cherche toujours à reprendre les mêmes temporaires via les agences auxquelles elle s'adresse.

Les gestionnaires d'Alstom confirment également l'hypothèse selon laquelle l'impact des heures supplémentaires sur la qualité est approximativement deux fois plus important que celui du sureffectif, ce qui rejoint la conclusion de l'étude de Li et al., [96]). Ainsi on posera

$$\Delta q_j^O = 2\Delta q_j^S / H_R. \quad (7.6)$$

**Proportion maximum de temporaires  $\lambda$ .** Cette proportion est naturellement limitée par le nombre maximum  $A_j$  de travailleurs qu'il est possible d'affecter à toute activité  $j$ . Etant donné les valeurs de  $A_j$  dans notre étude de cas, nous constatons que le nombre maximum de travailleurs en sureffectif est toujours égal à 1, pour toutes les activités. Cela n'a donc pas de sens de considérer des valeurs intermédiaires de  $\lambda$  entre 0 et 1 exclus puisque ce paramètre s'applique non pas globalement mais à chacune des activités, où l'on impose  $\frac{\lambda}{1-\lambda}p_{jt} - u_{jt} \geq 0$  lorsque  $\lambda > 0$ . Ainsi, nous étudierons le cas où  $\lambda = 0$  et l'accélération ne peut donc être faite qu'avec des heures supplémentaires et le cas où  $\lambda = 1$  c'est-à-dire un recrutement de temporaires théoriquement illimité mais limité en pratique par l'effectif maximum  $A_j$ .

**Coefficient de dépassement du budget  $\rho$  et contrainte organisationnelle.** Les gestionnaires d'Alstom confirment que l'objectif prioritaire est celui de la minimisation de la durée car les pénalités de retard sont considérables, de l'ordre de 125000 euros par jour pour un TGV de sorte que les dépassements de budget supportés en cas d'accélération n'excèdent rarement voire jamais ces pénalités. Nous fixerons alors  $\rho = 2$  qui implique une désactivation de la contrainte de budget dans le programme. Conformément aux contraintes organisationnelles de l'entreprise, nous supposons dans le modèle que le nombre de permanents affectés à une activité en cours d'exécution est au moins égal à ce nombre en mode normal, soit

$$p_{jt} \geq P_{kj} \cdot z_{jt} \quad \forall j = 1..n \quad t \in \mathcal{T}_j. \quad (7.7)$$

## 7.4 Résultat des simulations

En combinant les quatre valeurs pour le coefficient global  $\psi$  de sensibilité de la productivité de l'entreprise au sureffectif avec  $\psi = \{0,05; 0,25; 0,50; 0,75\}$  aux deux valeurs de la proportion de temporaires, soit  $\lambda \in \{0; 1\}$ , nous obtenons huit configurations différentes. Pour chacune d'elles, soit on autorise le nombre maximum légal d'heures par semaine ( $H_{OW} = 8$ ) soit on les restreint à une heure par travailleur et par semaine ( $H_{OW} = 1$ ). Cette restriction sur les heures supplémentaires se justifie par une difficulté de trouver des superviseurs de chaîne qui accepte de travailler au delà des heures régulières, ainsi qu'Alstom nous en a fait part. On obtient finalement 16 instances pour le problème d'optimisation de la planification de la main d'oeuvre additionnelle appliqué au projet de fabrication de la motrice.

Le programme d'optimisation a été codé en C avec un appel au progiciel d'optimisation CPLEX 12.5 et exécuté sur un Intel Xeon E5\_2430 avec 2.20 gigahertz (16 Cores) CPU et 49 gigabytes RAM. Nous avons limité à 30 minutes le temps de résolution de chaque instance. A l'exception d'une seule instance, nous avons obtenu une solution optimale pour toutes les autres et ce, dans un moindre temps, à savoir moins de 15 secondes pour l'une d'elles et environ 17 minutes pour l'instance la plus difficile à résoudre. Le nombre de variables  $z_{jt}$ ,  $p_{jt}$ ,  $u_{jt}$ ,  $h_{jt}$  et  $y_{jtl}$  (binaires utilisées pour les linéarisations) est de l'ordre de 5000 avec approximativement 7000 contraintes en moyenne. Nous rappelons que le résumé du modèle d'optimisation en annexe B donne pour chaque type de contrainte, son nombre.

Le tableau 7.4 donne les résultat des simulations où les premières colonnes décrivent l'instance en termes d'utilisation des heures supplémentaires, de sensibilité globale  $\psi$  de la productivité au sureffectif et de proportion  $\lambda$  de temporaires. La colonne "durée optimale" donne la durée  $C_{max}$  du projet à l'issue de l'optimisation. La variation globale en pour-cent (les trois colonnes suivantes) est calculée pour la durée, la qualité et le coût. Pour la durée, cette variation globale est simplement égale à  $(C_{max}/T) - 1$  où  $T = 107$  jours est la durée du projet en mode normal. La variation globale de la qualité est donnée par  $(\sum_{j=1..n} q_j / \sum_{j=1..n} q_j^{max}) - 1$  soit encore  $(\sum_{j=1..n} q_j / n) - 1$  puisque nous avons supposé  $q_j^{max} = 1, \forall j$ . La variation globale du coût est déterminée par rapport au coût en mode normal, c'est-à-dire

$$\frac{c_P H_R \sum_t p_{jt} + c_U H_R \sum_t u_{jt} + c_O \sum_t h_{jt}}{c_P H_R \cdot d_j^{max} \cdot P_{kj}} - 1, \quad (7.8)$$

qui s'écrit également comme la somme des trois taux

$$\frac{(c_P H_R \sum_t p_{jt} - c_P H_R \cdot d_j^{max} \cdot P_{kj})}{c_P H_R \cdot d_j^{max} \cdot P_{kj}} + \frac{c_U H_R \sum_t u_{jt}}{c_P H_R \cdot d_j^{max} \cdot P_{kj}} + \frac{c_O \sum_t h_{jt}}{c_P H_R \cdot d_j^{max} \cdot P_{kj}}, \quad (7.9)$$

exprimant respectivement la variation du coût associée aux permanents, aux temporaires et aux heures supplémentaires. Les trois dernières colonnes donnent le détail de chacun de ces taux de variation du coût. On note que les coûts salariaux (ou budget salarial) du projet en mode normal s'élèvent à  $c_P H_R \cdot d_j^{max} \cdot P_{kj} = 60952,6$  euros.

Description instance				Durée optimale	Variation globale (%)			Ventilation variation du coût (%)		
#	Heures sup.	Sensibilité sureffectif	Proportion temporaires		Durée	Qualité	Coût	Permanents	Temporaires	Heures sup.
1	Faible	0,05	0	103	-3,74	-2,62	0,28	0,00	0,00	0,28
2	Faible	0,05	1	98	-8,41	-3,53	2,31	-9,22	11,06	0,46
3	Faible	0,25	0	105	-1,87	-1,31	0,73	0,00	0,00	0,73
4	Faible	0,25	1	105	-1,87	-1,31	0,73	0,00	0,00	0,73
5	Faible	0,50	0	104	-2,80	-1,78	2,19	1,73	0,00	0,46
6	Faible	0,50	1	101	-5,61	-2,81	5,98	-5,19	10,37	0,79
7	Faible	0,75	0	104	-2,80	-1,90	4,00	3,17	0,00	0,83
8	Faible	0,75	1	103	-3,74	-2,74	6,37	-0,58	6,22	0,72
9	Max.	0,05	0	100	-6,54	-3,77	1,18	-3,75	0,00	4,92
10	Max.	0,05	1	98	-8,41	-4,49	2,16	-8,36	7,26	3,25
11	Max.	0,25	0	101	-5,61	-3,36	1,79	-3,46	0,00	5,25
12	Max.	0,25	1	99	-7,48	-4,75	3,38	-8,36	7,26	4,47
13	Max.	0,50	0	101	-5,61	-3,42	2,64	-3,17	0,00	5,81
14	Max.	0,50	1	100	-6,54	-4,66	4,18	-7,49	6,22	5,45
15	Max.	0,75	0	101	-5,61	-3,45	2,92	-4,32	0,00	7,24
16	Max.	0,75	1	101	-5,61	-3,52	4,98	-5,48	4,84	5,62

Tableau 7.4 – Résultat des simulations

Etudions la première instance dans le tableau 7.4 où il n'y a pas de recrutement de temporaires. La durée optimale du projet est de 103 jours, soit un gain de 4 jours par rapport au mode normal (baisse de la durée de 3,74%). Cette durée est obtenue avec un budget en heures supplémentaires seulement égal à 0,28% du budget normal. En réalité, cette accélération est obtenue en optimisant l'utilisation de la main d'oeuvre permanente allouée au projet. En effet si l'on considère les cinq premières activités, elles utilisent toutes la même spécialisation de soudure ( $k = 1$ ). Quatre soudeurs sont mobilisés au total ( $P_1 = 4$ ) pour le projet et commencent par exécuter l'activité  $j = 1$  entre le jour 1 et le jour 7 qui les nécessite tous en mode normal ( $P_{1,1} = 4$ , avec  $k = 1$  et  $j = 1$ ). Au début du jour 8, les 4 soudeurs passent à l'activité  $j = 2$  qu'ils terminent au jour 14. Au début du jour 15, l'exécution de l'activité  $j = 3$  commence mais elle ne requiert que 3 soudeurs en mode normal ( $P_{1,3} = 3$ ) pour une durée de 7 jours ( $W_j = 21$ ). Le programme d'optimisation propose alors la solution que nous donnons dans le tableau 7.5 et qui permet d'exécuter l'activité en 6 jours au lieu de 7. Le sureffectif est utilisé autant que possible avec un soudeur additionnel aux jours 16, 17 et 20. Avec un coefficient d'impact  $\gamma_j = 0,03$  faible, le travail de 4 soudeurs équivaut à celui de 3,96 soudeurs qui font effectivement avancer l'exécution de l'activité. Pour atteindre le besoin en main d'oeuvre  $W_j = 21$ , outre le sureffectif, il faut exactement 0,94 heure supplémentaire. La troisième ligne du tableau donne pour chaque jour la main d'oeuvre totale (comprenant les heures supplémentaires) et corrigée des baisses de productivité. Cette main d'oeuvre est exprimée en nombre de travailleurs avec une somme sur la durée totale qui est bien égale à 21. On rappelle que cette ligne s'obtient en calculant  $w_j^R(p_{jt}) + h_{jt}/H_r$ , avec  $w_j^R(p_{jt}) = (1 + \gamma_j) \cdot p_{jt} - \frac{\gamma_j}{P_{kj}} \cdot p_{jt}^2$  où  $u_{jt} = 0, \forall t$  puisque  $\lambda = 0$  (voir équation

(6.16) page 133 dans le chapitre précédent). La qualité de l'activité s'établit à  $q_j = 1 - 0,107 = 0,894$  avec un détail des pertes par jour donné dans la dernière ligne du tableau. On a en effet pour cette activité des pertes unitaires de qualité  $\Delta q_j^S = 0,033$  et  $\Delta q_j^O = 0,008$  ce qui donne bien par exemple une perte de qualité au jour 17 de  $\Delta q_j^S + 0,94\Delta q_j^O \approx 0,041$ . Pour obtenir une durée du projet égale à 103, le programme d'optimisation aboutit à une réduction de 1 jour des activités 3, 4, 5, 8, 11, 20 et 21.

	Jour (t)						
	15	16	17	18	19	20	
Permanents $p_{jt}$	3	<b>4</b>	<b>4</b>	3	3	<b>4</b>	
Heures sup. $h_{jt}$	0	0	<b>0,94</b>	0	0	0	
Eq. main d'oeuvre	3	3,96	4,08	3	3	3,96	$\sum = 21$
Perte de qualité	0	<b>0,033</b>	<b>0,041</b>	0	0	<b>0,033</b>	$\sum = 0,107$

Tableau 7.5 – Solution optimale pour l'activité  $j = 3$  dans l'instance 1

Partant de cette première instance, on obtient une réduction conséquente de la durée du projet qui passe de 103 à 98 jours, en autorisant le recrutement de temporaires (instance #2). Les heures supplémentaires sont peu utilisées puisque fortement limitées ( $H_{OW} = 1$  heure par semaine et par travailleur), de sorte que le sureffectif est employé au maximum car les pertes de productivité en résultant sont faibles ( $\psi = 0,05$ ). Le taux de variation du coût associé aux permanents est négatif ( $-9,22\%$ ) parce que ces permanents sont requis sur un nombre plus petit de périodes qu'en mode normal. La perte globale de qualité associée est toutefois plus importante.

Dans les instances #3 et #4, la durée du projet est la même (105 jours) car la contrainte sur le recrutement de temporaires est sans effet. C'est pour cette instance #4 que nous n'avons pas pu obtenir de solution optimale dans la limite de temps. Cela provient du fait que le sureffectif est déjà exploité au maximum sans temporaires si bien que la possibilité d'un recrutement externe ne permet pas de réduire davantage la durée. Rappelons en effet que dans cette étude de cas, le sureffectif est limité à un seul travailleur sur les activités pouvant être accélérées.

L'instance #5 pour laquelle la durée optimale est de 104 jours présente la particularité d'un taux d'accroissement du budget en main d'oeuvre permanente (1,73%). Cela s'explique par une utilisation plus importante de permanents qu'en mode normal pour certaines activités. Une analyse de la solution pour l'activité 3 dans cette instance #5 montre que contrairement à l'instance #1, la somme des permanents sur la durée d'exécution excède le contenu en main d'oeuvre  $W_j = 21$ , cette somme étant égale à 22. Le premier taux d'accroissement dans la formule (7.9) à savoir  $\frac{(c_P H_R \sum_t p_{jt} - c_P H_R \cdot d_j^{\max} \cdot P_{kj})}{c_P H_R \cdot d_j^{\max} \cdot P_{kj}}$  se simplifie pour donner  $\frac{\sum p_{jt}}{d_j^{\max} \cdot P_{kj}} - 1$  qui devient positif dès lors que la somme des permanents alloués à une activité dépasse cette somme en mode normal. Ce taux, lorsqu'il est positif peut s'interpréter comme le pourcentage de permanents additionnels qu'il faut le long de la durée normale pour atteindre la réduction associée de la durée.

Globalement, à partir du tableau 7.4, on constate que les temporaires lorsqu'il est possible de les recruter ( $\lambda = 1$ ) sont moins utilisés dans les configurations où la sensibilité de la productivité au sureffectif est forte. Dans ces situations, un recours aux heures supplémentaires semble plus important pour accélérer le projet mais en supportant des pertes de qualité plus importantes. La figure 7.5 est construite à partir des données du tableau 7.4 et donne la qualité en fonction de la durée du projet en groupant les instances par valeurs de sensibilité  $\psi$  de la productivité (négligeable, faible, moyenne ou forte). La qualité globale est simplement définie comme la moyenne des qualités des activités, soit encore  $1 - \sum_j (q_j - q_j^{\max}) / \sum_j q_j^{\max}$  où les valeurs de la variation globale de qualité  $\sum_j (q_j - q_j^{\max}) / \sum_j q_j^{\max}$  sont données pour chaque instance dans la septième colonne du tableau 7.4.

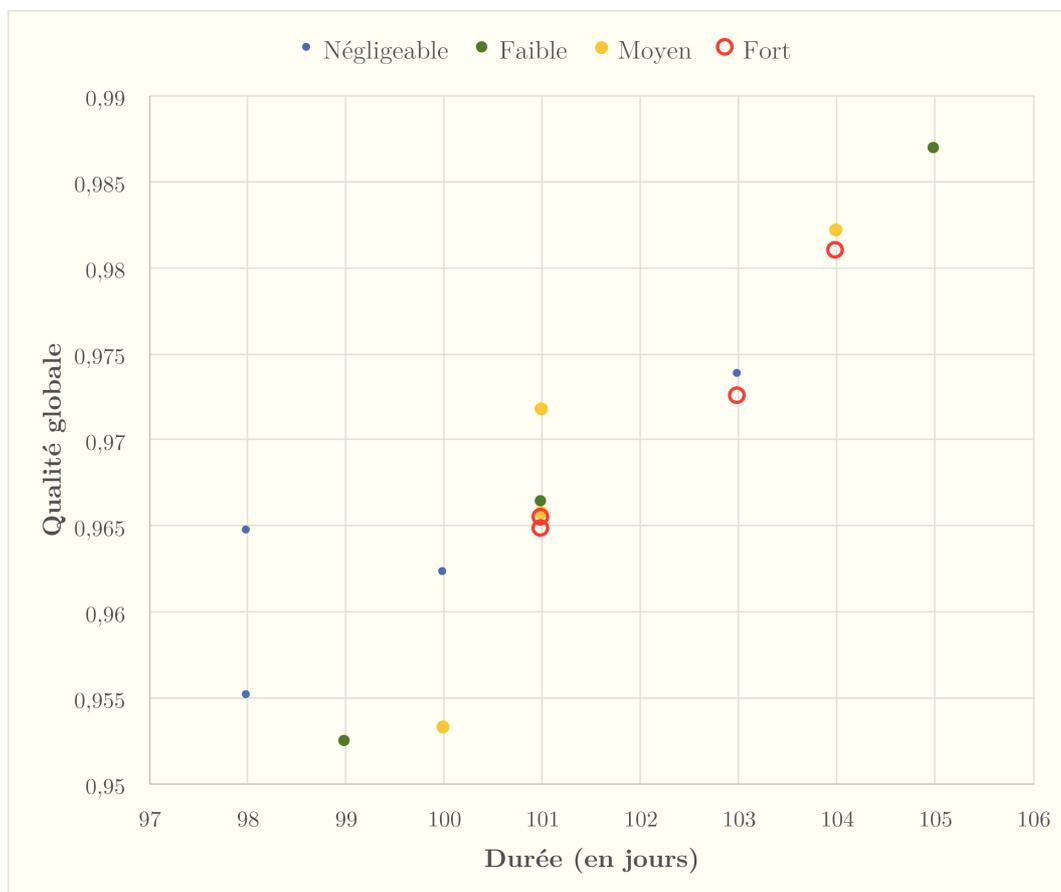


FIGURE 7.5 – Relation qualité - durée par instance selon la sensibilité de la productivité au sureffectif

On remarque tout d'abord que la qualité globale s'établit à des valeurs élevées pour toutes les instances ; elle ne descend pas en dessous de 0,952 (obtenue pour une réduction de 8 jours) et peut atteindre une valeur aussi élevée que 0,987 pour une réduction de 2 jours de la durée du projet. Cela est dû au fait que l'accélération est limitée dans notre étude de cas à la fois par un sureffectif qui ne va jamais au delà de 1 travailleur et par un nombre restreint d'activités dont on peut réduire la durée. En effet, 50% des activités seulement peuvent être accélérées. Les activités qu'il est

impossible d'accélérer, outre celles de niveau zéro de tolérance aux défauts, sont celles dont la durée d'exécution est de 1 jour. On constate sans surprise l'existence d'une relation croissante entre la qualité et la durée et les durées les plus faibles s'obtiennent pour  $\psi = 0,05$  (négligeable) et  $\psi = 0,25$  (faible). En particulier, des durées de projet strictement inférieures à 100 ne sont atteintes que pour ces valeurs de la sensibilité. Pour une même durée, la qualité obtenue est d'autant plus élevée que  $\psi$  est faible puisqu'à durée égale, il faut moins de travailleurs en sureffectif ce qui limite les pertes de qualité. Ainsi, on observe par exemple que pour une durée de 100 jours, la qualité la plus élevée s'obtient pour  $\psi = 0,05$ . Pour une durée de 101 jours, l'une des instances semble dominer l'autre lorsque l'impact sur la productivité est très élevé. Il s'agit des deux dernière instances (#15 et #16) où l'impact  $\psi = 0,75$  est fort ; pour une même durée, l'instance #15 fournit une qualité supérieure (baisse de 3,45% contre 3,52%). Mais si cette instance est dominante en termes de qualité, elle ne l'est pas nécessairement en termes de coûts.

Considérons alors la figure 7.6 qui donne pour chaque durée de projet, l'augmentation du coût qui lui est associée et ce, dans les différentes configuration d'impact  $\psi$ . Nous avons ajouté aux points une courbe de tendance pour gagner en lisibilité sauf pour l'impact fort où la tendance n'était pas représentative.

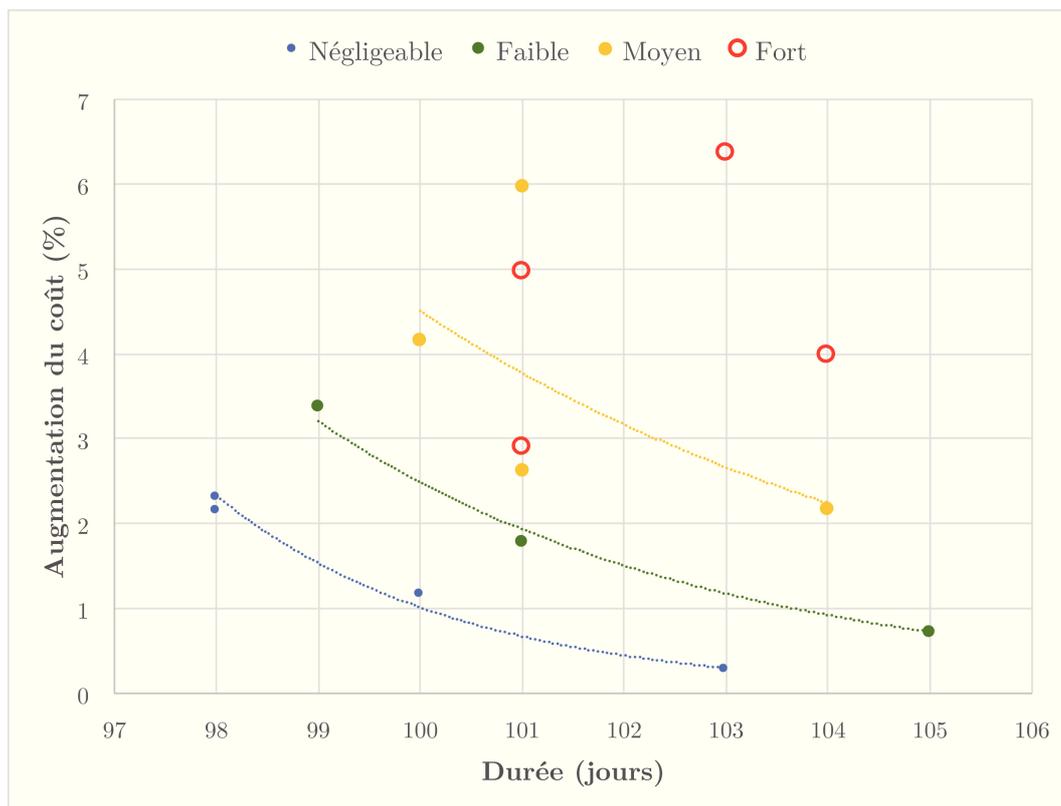


FIGURE 7.6 – Relation durée - coût par instance selon la sensibilité de la productivité au sureffectif

Si l'on étudie les couples durée - coût par valeur de  $\psi$  on constate tout d'abord que globalement, l'augmentation du coût est d'autant plus faible que les durées du projet sont longues ; les meilleures instances du point de vue du coût pouvant être

atteintes avec  $\psi$  négligeable ou faible. A valeur de  $\psi$  identique, certaines instances en dominant d'autres. C'est notamment le cas pour deux instances avec  $\psi$  moyen ( $\psi = 0,50$ ) et pour une durée de 101 jours : l'instance #6 avec une augmentation du coût de 5,98% est dominée par l'instance #13 où la variation du coût n'est que de 2,64%. Néanmoins cette dominance ne vaut qu'au regard du critère de coût parce que dans l'instance #6, les pertes de qualité sont plus faibles que dans l'instance #13. A nouveau, on retrouve une dominance de l'instance #15 par rapport à l'instance #16 en termes de qualité, de sorte que l'on peut conclure que l'instance #16 est dominée par l'instance #5 au regard de tous les critères pour une durée de 101 jours.

Pour obtenir des solutions dominantes, il est possible d'adopter une approche lexicographique. Les gestionnaires d'Alstom confirment la pertinence d'une telle approche avec une minimisation de la durée en premier lieu, puisque les pénalités de retard sont rédhitoires. Ainsi, dans notre cas, la méthode lexicographique va consister à minimiser d'abord la durée du projet sans contrainte de budget mais en respectant les contraintes de seuil de qualité minimum. C'est ce premier programme que nous avons résolu dans notre étude de cas et qui donne la durée optimale du projet  $C_{max}^*$ . Un deuxième programme d'optimisation aurait pour objectif la minimisation du coût du projet  $B$  sous la contrainte  $C_{max} = C_{max}^*$ , en conservant toujours les contraintes de qualité minimum. La résolution de ce deuxième programme donne le budget optimal du projet  $B^*$  pour atteindre la durée minimum  $C_{max}^*$  ainsi que les valeurs des qualités individuelles  $q_j^*$  respectant les seuils. Dans un troisième programme d'optimisation, la fonction objectif  $Q$  de qualité globale à maximiser est une somme pondérée des qualités individuelles des activités, avec

$$Q = \sum_{j=1..n} \eta_j \cdot q_j, \quad (7.10)$$

où  $\eta_j$ , le poids affecté à la qualité de l'activité  $j$  reflète sa contribution (son importance) dans la qualité globale et l'on impose de plus

$$q_j^{\max} - \sum_{t \in \mathcal{T}_j} \Delta q_{jt} \geq q_j^* \quad \forall j = 1..n, \quad (7.11)$$

Plusieurs ensembles de poids  $\{\eta_j\}_{j=1..n}$  pourraient être considérés. Des poids identiques signifient que chaque activité contribue dans la même mesure à la qualité globale. Si l'accent est mis sur les activités de faible niveau de tolérance, on tendra à accorder davantage d'importance à la durée de vie du produit. Si en revanche, des poids plus importants sont placés sur des activités de fort niveau de tolérance, c'est plutôt l'esthétique du produit fini qui est privilégiée sans pour autant que cela remette en cause sa durée de vie standard garantie par les seuils minimum de qualité  $q_{\theta}^{\min}$ .

## 7.5 Conclusion

Cette étude de cas nous a permis de tester la validité de la modélisation développée au chapitre précédent. Nous obtenons ainsi des durées optimales pour le projet de fabrication d'une motrice TGV tout en respectant les seuils de qualité minimum définis par groupes d'activités selon leur niveau de tolérance aux défauts. Même en n'activant pas la contrainte de budget, nous obtenons dans le pire des cas une variation du coût de l'ordre de 6,5% qui correspond à un surcoût salarial de moins de 4000 euros, pour une réduction de 4 jours de la durée du projet (instance #8 où la sensibilité de la productivité au sureffectif est forte). Ce surcoût salarial par jour de réduction soit 1000 euros est négligeable comparativement aux pénalités de retard de livraison d'un TGV qui sont de l'ordre de 125000 euros par jour. Naturellement, à ce surcoût salarial pour la motrice doivent s'ajouter d'autres surcoûts, comme par exemple ceux liés à un approvisionnement anticipé des composants. Ainsi, le modèle peut être utilisé pour accélérer le projet en cours d'exécution si l'on constate un retard à un moment donné. Auquel cas, il suffit d'appliquer le modèle sur l'ensemble des activités restant à exécuter, en créant une nouvelle activité fictive de début reliant les prochaines activités à lancer en fabrication. S'il s'agit d'obtenir un avantage concurrentiel en termes de délai promis au client, le modèle peut-être appliqué ex-ante et donnera le coût et la qualité pour chaque jour de réduction, laissant ainsi la possibilité à la direction de l'entreprise d'arbitrer entre plusieurs solutions, en les comparant notamment aux primes d'avance le cas échéant.

Si l'objectif principal demeure la minimisation de la durée sous les contraintes de qualité, le coût reste une dimension que l'entreprise cherche toujours à maîtriser, aussi une voie d'extension de la modélisation proposée consiste-t-elle à mettre en oeuvre une approche lexicographique telle que nous l'avons décrite dans la section précédente.

Une autre utilisation qui peut être faite de notre modèle de planification est de simuler des extensions de capacité pour un re-dimensionnement des ateliers. En particulier, il est possible avec ce modèle de déterminer l'impact sur la durée, le coût et la qualité de l'ajout d'un ou plusieurs poste de travail. Dans cette étude de cas en effet, le sureffectif maximum est d'un seul travailleur et seule la moitié des activités peut être accélérée, limitant ainsi à la fois la réduction de la durée du projet et les pertes de qualité. Ces pertes de qualité sont en effet faibles, avec une qualité globale qui ne descend pas en dessous de 95,24% pour une réduction de 8 jours de la durée totale et un impact faible du sureffectif sur la productivité ( $\psi = 0,25$ ). Les résultats que nous obtenons sont toutefois en cohérence avec la nature du projet de fabrication considéré qui supporte mal les pertes de qualité, pour des raisons d'image de l'entreprise, de sécurité et d'une faible marge de flexibilité sur les spécifications des clients. Il aurait été intéressant à ce titre de tester le modèle sur d'autres projets de fabrication comme les projets de construction immobilière où les attentes des clients peuvent assez facilement faire l'objet de révisions à la baisse si le délai de livraison est avancé. On peut songer par exemple aux achats d'appartements pour lesquels le client peut revoir ses exigences sur les finitions s'il lui est possible de prendre plus rapidement possession des lieux.

Des simulations extensives nous auraient permis d'étudier plus finement l'impact

sur la durée, le coût et la qualité de la sensibilité de la productivité au sureffectif ou alors l'impact du volume des heures supplémentaires. De même, la structure du produit dans cette étude de cas est une structure d'assemblage où chaque activité ne possède le plus souvent qu'un seul successeur et jamais plus de deux successeurs. Pour cette structure d'assemblage et le nombre restreint d'activités pouvant être accélérées (18 activités) nous obtenons pour toutes les instances à l'exception d'une seule, la durée optimale dans un temps raisonnable. Néanmoins, pour des projets structurellement plus complexes et plus souples en termes de sureffectif possible, la performance de notre modèle d'optimisation risque de se modifier, avec une probable inaptitude à fournir des solutions optimales dans un temps raisonnable.

# Conclusion générale

Dans cette thèse, nous avons proposé un modèle d'arbitrage durée, coût et qualité d'un projet où l'accélération peut se faire par un recours aux heures supplémentaires et au travail temporaire. Notre problématique de départ consistait à formuler une relation directe entre l'utilisation de ressources et la durée, le coût et la qualité d'un projet en cas d'accélération ; l'objectif étant de proposer un outil d'aide à la décision en termes de planification de la main d'oeuvre additionnelle tout en optimisant les trois critères de performance d'un projet (durée, coût et qualité). Nous avons considéré les projets de fabrication où les activités sont liées par des contraintes de précedence et de non-préemption.

En premier lieu, dans le chapitre 1, nous avons rappelé les notions de base en gestion de projet sur la durée et le coût, où nous avons distingué les projets de fabrication des projets de conception en relevant par ailleurs plusieurs particularités pour les projets modulaires. Le concept de qualité, concept-clé dans cette thèse, a fait l'objet d'une analyse dans le chapitre 2 où nous identifions deux types de qualité, la qualité objective et subjective. Le concept de qualité subjective est plus pertinent dans le cas des projets de conception. Pour les projets de fabrication que nous considérons, c'est l'évaluation objective qui est retenue. L'étude de la façon dont on évalue objectivement la qualité nous a permis d'adopter une mesure de la qualité par un indicateur compris entre 0 et 1 qui correspond au pourcentage d'items validés dans une check-list de contrôle qualité.

Dans le chapitre 3 nous avons analysé la littérature relative au problème d'arbitrage durée - coût (TCTP) où nous avons notamment conclu que le problème avec ressources renouvelables multiples dans un cadre discret est très peu étudié. Il en est de même concernant les contributions qui traitent le problème d'arbitrage durée, coût et qualité et présentées dans le chapitre 4 que nous concluons par le constat d'une paucité des modèles avec contraintes de ressources renouvelables multiples et d'une absence d'approches non agrégées pour la résolution du problème. A l'issue de cette revue de la littérature, nous avons vu que la qualité globale d'un projet est définie le plus souvent comme une moyenne des qualités individuelles des activités (moyenne arithmétique ou géométrique, simple ou pondérée). Nous avons pu identifier également qu'aucune des contributions étudiées ne considère des seuils de qualité à atteindre pour toute activité d'un projet.

C'est à partir d'un autre segment de la littérature que celui sur le TCQTP que nous avons pu construire les hypothèses nécessaires pour formuler analytiquement la relation entre la durée et la main d'oeuvre mobilisée pour un projet ainsi que

la relation entre la qualité et l'utilisation de cette main d'oeuvre. Il s'agit d'une part d'études relatives à l'effet du sureffectif et des heures supplémentaires sur la productivité et la qualité, et d'autre part de contributions modélisant le problème de l'ordonnancement avec main d'oeuvre additionnelle et celui de l'affectation des ressources aux activités d'un projet. C'est dans le chapitre 5 que nous avons présenté et discuté les hypothèses que nous avons adoptées. Ces hypothèses sont retranscrites mathématiquement dans le chapitre 6 où nous proposons un modèle de planification des heures supplémentaires et du recrutement en temporaires qui permet d'élaborer un calendrier prévisionnel du personnel à moyen terme. Le problème de minimisation de la durée sous contrainte de budget et de respect des seuils de qualité est formulé comme un programme linéaire mixte où les variables sont le nombre de travailleurs permanents et temporaires et le nombre d'heures supplémentaires affectées à l'exécution de chaque activité à chaque période de l'horizon. Nous avons testé dans le chapitre 7 la validité de notre approche sur un projet de fabrication d'une motrice TGV. Pour ce projet, nous avons simulé plusieurs configurations de pertes de productivité et de recours à la main d'oeuvre additionnelle. Pour toutes ces instances, nous obtenons presque toujours une solution optimale dans un temps raisonnable. Les résultats des simulations montrent que les solutions sont sensibles aux pertes de productivité et l'on retrouve naturellement une croissance entre la qualité et la durée d'une part et une décroissance entre la durée et le coût d'autre part.

Il est possible selon l'industrie considérée que la relation entre les pertes de qualité et la main d'oeuvre additionnelle ne soit pas linéaire de sorte qu'une étude empirique aurait été souhaitable pour valider ou non cette hypothèse de linéarité. Enfin, notre modèle pourrait être enrichi d'une contrainte de lissage de la main d'oeuvre pour faciliter le travail de supervision. Il est également pertinent de considérer un effectif de supervision en dessous duquel les heures supplémentaires ne sont pas autorisées. En outre, nous aurions pu intégrer des coûts fixes de recrutement et de formation des travailleurs temporaires.

# Annexe A

## Annexe technique - méthodes de résolution pour les problèmes combinatoires

L'objectif de cette annexe est de fournir une typologie des méthodes de résolution pour les problèmes d'optimisation combinatoire étudiés dans cette thèse. Ces problèmes sont de nature discrète c'est-à-dire que tout ou partie des variables sont entières. Il s'agit ici de décrire le principe de fonctionnement des approches rencontrées le plus fréquemment dans la littérature relative au problème d'arbitrage durée, coût et qualité (chapitres 3 et 4). Cette présentation des méthodes n'est donc pas exhaustive et possède un niveau de détail approprié pour le champ disciplinaire dont la thèse relève.

### A.1 Optimisation mono-objectif

Un problème d'optimisation combinatoire avec un objectif unique peut être défini par un vecteur de variables  $x = (x_1, x_2, \dots, x_n)$ , une fonction objectif  $f(x)$  à maximiser ou à minimiser sous un ensemble de contraintes où l'ensemble de toutes les solutions réalisables, appelé espace de recherche, est un ensemble de très grande taille. Une solution est dite réalisable si elle respecte toutes les contraintes du problème.

#### A.1.1 Classement des problèmes et typologie des méthodes de résolution

Les approches de résolution exacte garantissent l'obtention d'une solution optimale au problème mais nécessitent parfois un temps de calcul rédhibitoire. Les problèmes d'optimisation combinatoire sont difficiles à résoudre, c'est-à-dire que leur temps de résolution à l'optimum croît de façon exponentielle avec leur taille. C'est par exemple le cas pour le problème du voyageur de commerce qui consiste à trouver un trajet de moins de  $a$  kilomètres pour visiter un nombre  $n$  de villes ( $n$  est la taille du problème) reliées entre elles par des routes. C'est un problème difficile mais si l'on dispose d'un trajet, il est presque immédiat de vérifier si ce trajet fait

moins de  $a$  kilomètres ou non. On dit alors que le problème est dans NP c'est-à-dire qu'il existe un algorithme en temps polynomial pour vérifier qu'une solution donnée convient. Un algorithme est dit en temps polynomial s'il existe deux constantes  $C$  et  $k$  telles que pour tout entier  $n$  (taille du problème), l'algorithme s'exécute en moins de  $C \cdot n^k$  opérations élémentaires. Les opérations élémentaires sont par exemple des additions, des multiplications, des comparaisons etc. On retiendra les définitions suivantes.

**Classe P des problèmes faciles.** Le problème est dit dans P s'il existe un algorithme en temps polynomial (résolution efficiente) pour trouver sa solution optimale et non pour vérifier qu'une solution donnée est réalisable.

**Classe NP des problèmes de décision.** Un problème est dans NP s'il existe un algorithme en temps polynomial pour vérifier qu'une solution donnée est réalisable. Remarquons que les problèmes P sont inclus dans NP. En effet, comme par définition il existe un algorithme polynomial pour résoudre un problème P alors il existe aussi un algorithme polynomial pour vérifier qu'une solution à ce problème est réalisable.

**Problèmes de décision NP-complets.** Les problèmes NP complets sont des problèmes NP pour lesquels on ne peut pas affirmer qu'il existe un algorithme de résolution en temps polynomial (car aucun algorithme polynomial n'a été trouvé pour les résoudre). Ces problèmes sont équivalents au sens où si on trouve un algorithme polynomial pour en résoudre un, alors on peut en déduire des algorithmes polynomiaux pour résoudre tous les autres. On appelle ces problèmes NP-complets. Ainsi, à ce jour, on ne sait toujours pas s'il existe des algorithmes plus rapides que les algorithmes connus pour résoudre les problèmes NP-complets et qui nécessitent un temps superpolynomial (exponentiel, non borné par une fonction polynomiale) par rapport à la taille de l'instance.

On peut montrer qu'un problème est NP-complet si l'on arrive à lui trouver une analogie et donc une formulation équivalente à celle d'un problème NP-complet connu. Par exemple, Kolisch et Drexl (1997) [91] montrent que le MRCPSPP avec ressources renouvelables et non renouvelables est une généralisation d'un problème de job shop où il s'agit d'ordonnancer  $n$  tâches sur plusieurs machines; ces tâches n'ayant pas toutes le même ordre de passage sur les machines. Comme ce problème de job shop est NP-complet alors le MRCPSPP avec ces deux types de ressources l'est également.

On note que la résolution d'un programme linéaire en nombres entiers (PLNE) est un problème NP-complet. Comme certains problèmes dans P peuvent se formuler comme des PLNE, alors leur résolution devient un problème NP-complet. Cette remarque est destinée à relever l'importance de la formulation d'un problème pour sa résolution.

Comme les méthodes exactes nécessitent un temps de calcul qui croît de façon exponentielle avec la taille du problème, on utilise souvent des heuristiques qui permettent d'obtenir rapidement au moins une solution réalisable mais sans en garantir l'optimalité. La figure A.1 résume les approches de résolution pour les problèmes d'optimisation combinatoire avec une fonction objectif unique et distingue ainsi les

méthodes exactes des méthodes heuristiques. Il existe principalement quatre types de méthodes de résolution exacte. L'algorithme de Branch and Bound est de loin le plus utilisé. Les méthodes de coupe (cutting plane) plus rapides ont néanmoins l'inconvénient de ne pas toujours produire une solution réalisable. En revanche elles peuvent être combinées à une méthode de Branch and Bound pour en améliorer la performance ; on parle alors de méthodes Branch and Cut. Enfin, la génération de colonnes constitue une autre approche principale de résolution exacte.

Les approches heuristiques peuvent se répartir en trois grandes catégories. Si les heuristiques développées pour un problème particulier exploitent des connaissances particulières relatives à ce problème, on parle d'heuristiques dédiées ou spécifiques. Dans le cas contraire, il s'agit de métaheuristiques qui par définition sont des méthodes générales de résolution, applicables à un grand nombre de problèmes. Ces métaheuristiques peuvent manipuler une seule solution à la fois (méthodes de recherche locale) ou alors une population de solutions. Les méthodes de recherche locale comme les méthodes de descente, le recuit simulé ou la recherche taboue tendent à intensifier la recherche en exploitant une partie de l'espace des solutions réalisables. Les méthodes à populations de solutions comme les algorithmes génétiques, l'optimisation par colonies de fourmis, par essais particuliers ou par colonies d'abeilles ou encore la compétition impérialiste diversifient la recherche en explorant différentes parties de l'espace de recherche.

Dans la sous-section A.1.2, nous présentons les principales approches de résolution exacte. La sous-section A.1.3 commence par une brève description des heuristiques dédiées pour le problème d'arbitrage durée - coût que nous avons étudiées en détail dans le chapitre 3. Nous présentons ensuite les principales méthodes de recherche locale qui, comme les heuristiques dédiées pour le TCTP opèrent sur une solution à la fois. Enfin, la sous-section A.1.4 décrit le fonctionnement des métaheuristiques à population de solutions rencontrées dans la littérature relative au problème d'arbitrage durée et coût ou durée, coût et qualité.

## A.1.2 Méthodes exactes pour les problèmes linéaires en nombres entiers

Un programme linéaire en nombres entiers consiste à minimiser (ou maximiser) une fonction objectif linéaire qui s'exprime par une somme de variables multipliées par des coefficients (paramètres) sous un ensemble de contraintes linéaires et où les variables doivent prendre des valeurs entières. Un tel programme s'écrit

$$\min \sum_{j=1..n} c_j \cdot x_j \tag{A.1}$$

où chaque contrainte  $k$  peut s'exprimer sous la forme suivante

$$\sum_{j=1..n} a_{jk} \cdot x_j \leq b_k, \tag{A.2}$$

### Résolution exacte

- **Branch and Bound** (Méthode de séparation et évaluation)
- **Cutting-plane** (Méthode des plans sécants)
- **Branch and Cut** (Algorithmes de coupes et branchements)
- **Génération de colonnes** (Column generation)

### Résolution approchée

#### Heuristiques spécifiques (dédiées)

#### Métaheuristiques à solution unique (recherche locale)

- **Méthodes de descente**
- **Recuit simulé** (Simulated Annealing, SA)
- **Recherche Taboue** (Tabu Search, TS)

#### Métaheuristiques à population de solutions

- **Algorithme génétique** (Genetic Algorithm, GA)
- **Optimisation par colonies de fourmis** (Ant Colony Optimisation, ACO)
- **Optimisation par essaims particulaires** (Particle Swarm Optimisation, PSO)
- **Optimisation par colonies d'abeilles** (Artificial Bee Colony, ABC)
- **Compétition impérialiste** (Imperialist Competitive Algorithm, ICA)

FIGURE A.1 – Approches de résolution mono-objectif.

l'ensemble de toutes les contraintes peut s'écrire aussi sous une forme matricielle

$$A \cdot x \leq b, \quad (\text{A.3})$$

avec  $x = (x_1, \dots, x_n)$  et l'on impose de plus les contraintes d'intégralité

$$x_j \in \mathbb{Z}_+, \quad (\text{A.4})$$

où  $\mathbb{Z}_+$  désigne l'ensemble des nombres entiers non négatifs.

Lorsqu'une partie seulement des variables du problème sont entières les autres étant continues, on parle de programme linéaire mixte en nombres entiers. Il existe principalement quatre types de méthodes exactes de résolution.

### A.1.2.1 La méthode par séparation et évaluation (Branch and Bound)

Pour illustrer le principe de la méthode considérons le problème suivant

$$\max z = x_1 + 5x_2 \quad (\text{A.5})$$

sous les contraintes

$$\begin{aligned} x_1 + 10x_2 &\leq 20 \\ x_1 &\leq 2 \end{aligned} \quad (\text{A.6})$$

et où l'on impose de plus les contraintes d'intégralité

$$x_1, x_2 \in \mathbb{Z}_+ \quad (\text{A.7})$$

En ignorant les contraintes d'intégralité, on peut appliquer l'algorithme du simplexe qui donne la solution optimale pour les problèmes linéaires où les variables sont continues. La solution fournie par le simplexe est  $x_1 = 2$  et  $x_2 = 1,8$  ce qui donne  $z = 11$ . La première idée serait d'arrondir  $x_2$  à l'entier immédiatement supérieur, soit  $x_2 = 2$ . Mais cette solution est infaisable car elle ne respecte pas la contrainte  $x_1 + 10x_2 \leq 20$ . En arrondissant alors à l'entier immédiatement inférieur, soit  $x_2 = 1$ , les contraintes sont respectées mais  $z = 7$  n'est pas l'optimum. En effet, la solution optimale de ce problème est  $x_1 = 0$  et  $x_2 = 2$  ce qui donne  $z = 10$ .

Une autre méthode de résolution à laquelle on peut penser consiste à énumérer de toutes les solutions possibles et à choisir la meilleure. Cela peut marcher pour les problèmes de petites tailles, mais ce n'est pas pensable pour les problèmes de moyenne et grande taille. Par exemple, un problème contenant 20 variables binaires donne  $2^{20} = 1048576$  solutions à énumérer. L'explosion combinatoire est encore plus importante lorsqu'on considère des variables entières prenant leur valeur dans un intervalle plus large.

La méthode de Branch and Bound constitue la technique appropriée pour résoudre les problèmes linéaires en nombres entiers. Elle se fonde sur l'observation

selon laquelle l'énumération des solutions entières pour un problème donné prend la forme d'un arbre. Considérons par exemple l'énumération complète des solutions d'un problème de minimisation d'une fonction objectif  $f$  de trois variables entières : une variable entière "générale"  $x_1 \in \{1, 2, 3\}$  et deux variables binaires  $x_2$  et  $x_3$ . La figure A.2 montre l'énumération complète de toutes les solutions.

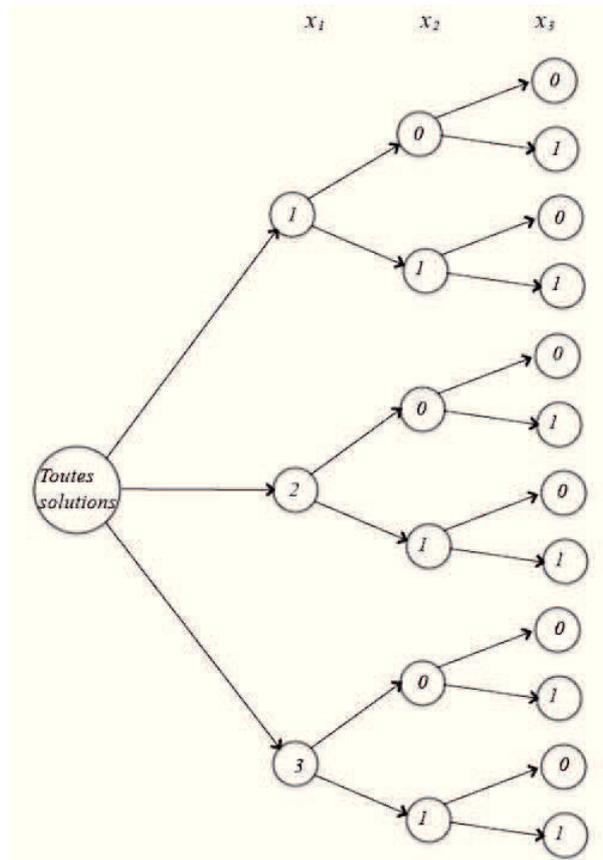


FIGURE A.2 – Arbre d'énumération

Le noeud racine représente toutes les solutions au problème. Le noeud  $x_1 = 2$  représente toutes les solutions avec  $x_1 = 2$ . Ce noeud est le noeud parent du noeud enfant  $x_2 = 0$  qui contient quant à lui toutes les solutions où on a la fois  $x_1 = 2$  et  $x_2 = 0$  etc. Les noeuds feuilles de l'arbre sont ceux pour lesquels les trois variables ont toutes une valeur entière. Il en existe 12 qui représentent toutes les combinaisons possibles de solutions entières au problème. L'idée de base du Branch and Bound est d'éviter autant que possible que l'arbre grandisse, en explorant uniquement les noeuds qui fournissent les solutions les plus prometteuses. Le caractère prometteur de ces solutions est évalué en estimant une borne sur la valeur de la fonction objectif. La séparation (branching) se produit lorsqu'un noeud est sélectionné pour une exploration des solutions issues de ce noeud. On utilise la terminologie suivante.

- Noeud : toute solution, partielle ou complète
- Noeud - feuille (leaf node) ou feuille : solution complète (les valeurs de toutes les variables sont connues)

- Noeud - bourgeon (bud node) : toute solution partielle, réalisable ou non.
- Méthode de détermination d'une borne : elle permet d'estimer la meilleure valeur de la fonction objectif qu'il est possible d'obtenir en laissant grandir un noeud. Cette méthode doit toujours produire une estimation optimiste.
- Méthode de séparation : processus par lequel on crée un noeud enfant à partir d'un noeud bourgeon.
- Incumbent : meilleure solution entière trouvée jusque là.

C'est la méthode de détermination d'une borne qui est le vecteur d'efficacité de l'algorithme de branch and bound.

Plus connu sous son nom anglais Branch and Bound (B&B), cet algorithme repose ainsi sur une méthode arborescente de recherche d'une solution optimale par séparation et évaluation, en représentant les sous-ensembles de solutions par des noeuds. La séparation consiste à décomposer au fur et à mesure de la recherche l'ensemble des solutions réalisables en plusieurs sous-ensembles pour définir des sous-problèmes. Dans notre exemple (voir figure A.2), une première séparation du problème est faite à partir des valeurs de  $x_1$ . Ainsi, le noeud  $x_1 = 2$  par exemple représente le problème de minimisation de  $f$  où  $x_1$  prend une valeur fixée à 2.

L'évaluation permet de réduire l'espace de recherche en éliminant les sous-ensembles de l'arborescence qui ne peuvent pas contenir la solution optimale. Cette élimination se fait de la façon suivante. On calcule une borne inférieure (dans le cas de la minimisation) sur la valeur optimale de la fonction objectif  $f$  associée aux valeurs possibles des solutions issues d'un noeud. Par exemple, une borne inférieure sur  $f$  pour le noeud  $x_1 = 2$  peut s'obtenir en minimisant  $f$  avec  $x_1 = 2$  et en relâchant les contraintes d'intégralité sur  $x_2$  et  $x_3$  (relaxation continue). Une telle borne  $f^{\text{inf}}$  s'obtient en résolvant ce problème avec l'algorithme du simplexe. La valeur optimale  $f^*$  pour le problème où  $x_1$  est fixée à 2 est telle que  $f^* \geq f^{\text{inf}}$ . Pour illustrer, imaginons que l'on cherche à minimiser la durée totale d'exécution  $f$  d'un projet et supposons que  $x_j$  désigne la date de début de l'activité  $j$ . Supposons que la borne inférieure sur la durée totale pour le noeud  $x_1 = 2$  obtenue en appliquant le simplexe soit égale à 30 jours. Cela veut dire que toute solution avec  $x_1 = 2$  et  $x_2$  et  $x_3$  prenant n'importe quelle valeur entière réalisable donnera une durée totale  $f$  qui sera nécessairement supérieure (ou égale) à 30 jours. Si pour le noeud  $x_1 = 3$  la borne inférieure calculée de la même façon est égale à 20 jours alors cela signifie que toute solution entière comprenant  $x_1 = 3$  fournira une durée totale supérieure ou égale à 20 jours. On a ici une meilleure borne pour le noeud  $x_1 = 3$  que pour le noeud  $x_1 = 2$  qui ne peut donc pas contenir la solution optimale parce que donnant lieu à des durées totales qui ne peuvent pas être inférieures à 30 jours. En conséquence, l'exploration du noeud  $x_1 = 2$  est abandonnée. Ainsi la borne minimum est mémorisée au cours de l'exploration et sert de comparaison avec les bornes inférieures calculées pour chaque noeud et permet l'élimination de ceux-ci lorsque leur borne inférieure est supérieure à la borne minimum. On mémorise également la meilleure valeur de la fonction objectif associée à toute solution réalisable découverte au cours de l'exploration, celle-ci constituant une borne supérieure.

Notons qu'il existe autant d'algorithmes de Branch and Bound qu'il existe de méthodes de calcul de bornes ou de stratégies de séparation.

### A.1.2.2 La méthode des plans sécants (Cutting-Plane)

La méthode des plans sécants encore appelée méthode de coupes planes a été développée par Gomory dans les années 1950 pour résoudre des problèmes d'optimisation combinatoire qui s'expriment comme des programmes linéaires. Cette méthode consiste à enlever une partie des contraintes du problème et à résoudre ce problème relaxé. On part du problème initial  $P$  suivant

$$P = \begin{cases} \min \sum_{j=1..n} c_j \cdot x_j \\ A \cdot x \leq b \end{cases} \quad (\text{A.8})$$

et l'on commence par résoudre

$$P_0 = \begin{cases} \min \sum_{j=1..n} c_j \cdot x_j \\ A_0 \cdot x \leq b_0 \end{cases} \quad (\text{A.9})$$

où  $A_0 \cdot x \leq b_0$  représente une partie des contraintes extraites de  $A \cdot x \leq b$ . On obtient ainsi une solution  $x^0$  au problème  $P_0$  qui constitue une relaxation du problème  $P$ . Puis on identifie les contraintes qui sont violées par  $x^0$  parmi celles qui n'ont pas été prises en compte dans  $P_0$ . On choisit l'une de ces contraintes  $\alpha \cdot x \leq \beta$  violée par  $x^0$ ; une telle contrainte est une coupe car elle coupe le point indésirable de l'espace des solutions possibles. On ajout cette contrainte  $P_0$  pour former  $P_1$  avec

$$P_1 = \begin{cases} \min \sum_{j=1..n} c_j \cdot x_j \\ A_0 \cdot x \leq b_0 \\ \alpha \cdot x \leq \beta \end{cases} \quad (\text{A.10})$$

On résout  $P_1$  puis on identifie à nouveau les contraintes violées par  $x^1$  et l'on choisit l'une d'elles que l'on intègre à  $P_1$  pour former le problème  $P_2$ . On réitère jusqu'à ce que la solution courante ne viole plus aucune contrainte de  $P$  et l'on obtient ainsi une solution optimale pour le problème initial  $P$ . La méthode des coupes planes est peu performante mais elle peut-être combinée à une méthode de Branch and Bound pour en améliorer la performance. Cette combinaison constitue la méthode de Branch and Cut que nous présentons dans le paragraphe suivant.

### A.1.2.3 Algorithmes de coupes et branchements (Branch and Cut)

La méthode "Branch and Cut" commence d'abord par une application sur le problème considéré de la méthode des coupes planes telle que nous l'avons décrite au paragraphe précédent. Il se peut cependant que cette méthode ne produise à la fin qu'une solution fractionnaire (non entière) et donc non optimale. Dans ce cas, on effectue une étape de branchement où l'on choisira une des variables fractionnaires  $x_j$  dans la solution pour considérer deux sous-problèmes en affectant la valeur  $x_j = 0$  pour l'un et la valeur  $x_j = 1$  pour l'autre (cas d'une variable  $x_j$  binaire). On applique alors la méthode des coupes pour les deux sous-problèmes. Cette phase de branchement est répétée de manière récursive jusqu'à l'obtention d'une solution entière optimale. Cette méthode Branch and Cut s'est révélée très efficace pour la résolution de problèmes d'optimisation combinatoire difficiles comme le problème du voyageur de commerce.

#### A.1.2.4 Génération de colonnes

Cette méthode repose sur le constat selon lequel les programmes linéaires de grande taille ont souvent beaucoup de variables (colonnes) dont la valeur optimale est nulle, c'est-à-dire des variables qui n'améliorent pas la fonction objectif. L'idée est alors de ne prendre en compte pour la résolution du problème qu'un petit sous-ensemble de variables. Le problème initial est appelé problème maître, et le problème réduit est appelé problème restreint. Ce problème restreint est donc plus simple à résoudre. Cependant, si l'ensemble des variables du problème restreint ne contient pas celles qui donnent la solution optimale pour le problème maître alors pour atteindre la solution optimale du problème maître, il faut rajouter au problème restreint des variables pouvant améliorer la solution. Ainsi, la méthode consiste à générer, au sein d'un algorithme à plusieurs étapes, les variables (colonnes) qui sont susceptibles d'améliorer la solution courante.

#### A.1.3 Heuristiques dédiées et méthodes de recherche locale

Nous présentons dans cette même sous-section les heuristiques dédiées et les principales méthodes de recherche locale parce qu'elles opèrent toutes sur une seule solution à la fois. Les heuristiques dédiées pour le problème d'arbitrage durée - coût sont brièvement rappelées dans un premier paragraphe, puis nous décrivons trois méthodes de recherche locale plus générales et que l'on appelle également métaheuristiques à solution unique. La recherche locale part d'une solution et lui associe un ensemble de solutions voisines en définissant préalablement une relation de voisinage. Une procédure d'exploitation de ce voisinage est ensuite utilisée dans le but d'améliorer la solution courante.

##### A.1.3.1 Les heuristiques dédiées

Comme nous l'avons vu dans la première sous-section de cette annexe, les heuristiques dédiées s'appuient sur l'exploitation de connaissances spécifiques relatives à un problème donné. Pour le problème de l'arbitrage durée - coût étudié au chapitre 3, c'est dans ce chapitre lui-même que nous avons présenté les principales heuristiques spécifiques à ce problème. Les heuristiques les plus couramment employées sont des algorithmes de construction progressive d'une solution qui respecte les contraintes de précédence et de capacité des ressources. Ces algorithmes sont appelés schémas de génération d'un ordonnancement (Schedule Generation Schemes, SGS) et sont plus amplement détaillées encore par Kolish et Hartmann dans le chapitre 7 du livre de Weglarz (1999) [155].

##### A.1.3.2 Méthodes de descente

Les méthodes de descente ont l'avantage d'être particulièrement rapides. Considérons une fonction  $f$  à minimiser (la durée totale d'exécution du projet par exemple). Partant d'une solution initiale  $x$  pour ce problème, on détermine l'ensemble  $N_x$  des solutions qui lui sont voisines. Pour le problème d'arbitrage durée - coût, une relation de voisinage peut consister à changer un petit nombre de fois (voire une seule

fois) les modes d'exécution de chaque activité. On sélectionne une solution voisine  $x' \in N_x$  telle que  $f(x') \leq f(x)$  puis on continue à explorer le voisinage  $N_x$  en retenant chaque fois la solution qui améliore l'objectif. A l'issue de cette exploration, on obtient un minimum local. L'exploration du voisinage peut être exhaustive (Hill Climbing) ou partielle (First Improvement Hill Climbing)

### A.1.3.3 Recuit simulé (Simulated Annealing, SA)

Dans le recuit simulé, la stratégie d'exploitation du voisinage permet de se diriger vers une solution voisine de moins bonne qualité avec une probabilité non nulle de sorte à échapper aux optima locaux. La probabilité d'accepter une solution détériorante dépend d'un paramètre  $T$  (la température) dont la valeur initiale est fixée arbitrairement. Cette probabilité est d'autant plus grande que  $T$  est grand. Au cours des itérations,  $T$  diminue de sorte que l'on accepte de moins en moins des solutions qui dégradent l'objectif. L'efficacité de cet algorithme dépend très fortement de son paramétrage.

### A.1.3.4 Recherche Taboue (Tabu Search, TS)

La recherche taboue est une méthode de recherche locale qui repose sur une mémoire des solutions visitées (liste taboue) pour échapper aux optima locaux. L'exploitation du voisinage permet de se déplacer de la solution courante vers un voisin ayant une valeur de fonction objectif identique ou meilleure mais pour éviter que l'algorithme ne passe de l'un à l'autre à l'infini, la méthode interdit de considérer les solutions récemment explorées. La liste taboue contient généralement un nombre fixe de solutions et est actualisée en enregistrant chaque nouvelle solution récemment visitée et en supprimant la solution la plus ancienne dans la liste. La taille de cette liste est un paramètre critique pour l'efficacité de la méthode qui nécessite par ailleurs la détermination d'un critère d'arrêt.

## A.1.4 Métaheuristiques à population de solutions

Contrairement aux méthodes de recherches locales, les métaheuristiques à population de solutions travaillent sur plusieurs solutions simultanément de sorte à préserver une diversité essentielle pour échapper aux optima locaux et à explorer en parallèle plusieurs régions de l'espace de recherche. Il existe plusieurs stratégies d'évolution de cette population qui peuvent s'inspirer du processus de sélection naturelle où les individus les plus adaptés gagnent la compétition de la reproduction tandis que les moins adaptés meurent avant la reproduction (algorithmes génétiques). D'autres stratégies d'évolution s'appuient sur le comportement social de certains animaux comme les oiseaux en vol ou les fourmis et les abeilles à la recherche de nourriture.

### A.1.4.1 Algorithme génétique (Genetic Algorithm, GA)

L'algorithme génétique part d'une population initiale de solutions (chromosomes) générées aléatoirement ou avec des heuristiques dédiées. On évalue leur performance

(Fitness) relative pour affecter à ces solutions une probabilité de sélection. Cette Fitness détermine en effet la probabilité de survie et de reproduction de chaque solution dans les prochaines générations. Considérons par exemple une fonction  $f$  à minimiser sous un ensemble de contraintes et soit une solution  $x^l$  (vecteur de  $n$  variables) parmi les  $N$  solutions qui forment la population à une itération donnée. La fitness  $\varphi(x^l)$  de cette solution  $x^l$  peut se calculer comme suit

$$\varphi(x^l) = 1 - \frac{f(x^l)}{\sum_{i=1..N} f(x^i)} \quad (\text{A.11})$$

et sera donc d'autant plus grande que la solution  $x^l$  donne une valeur faible de la fonction objectif. La probabilité de sélection de  $x^l$  sera d'autant plus forte que sa Fitness (mesure de la performance) est grande. Les solutions sélectionnées sont modifiées à l'aide des opérateurs de croisement (cross-over) et de mutation. Ainsi, plusieurs individus se reproduisent, certains disparaissent et seuls les individus les mieux adaptés survivent. On recommence ce cycle (itération) jusqu'à satisfaction d'une condition d'arrêt portant sur le temps de calcul ou sur la qualité de la solution (définie en termes de valeur de la fonction objectif). Le pseudo code d'un algorithme génétique est donné dans le tableau A.1.

---

Poser  $t = 0$  et initialiser la population  $P_t = P_0$  en générant  $N$  solutions aléatoirement

Tant que condition d'arrêt non vérifiée (nombre de générations ou temps de calcul)

    Evaluer la Fitness des solutions de  $P_t$

    Sélectionner des individus de  $P_t$  en fonction de leur Fitness

    Appliquer les opérateurs évolutionnaires sur ces individus (mutation et cross-over)

    Evaluer la Fitness des descendants ainsi obtenus

    Remplacer dans la population  $P_t$  les individus les moins performants par les meilleurs descendants pour former  $P_{t+1}$

    Poser  $t = t + 1$

Fin boucle tant que

---

Tableau A.1 – Pseudo code d'un algorithme génétique.

Illustration. Considérons le DTCTP pour un projet de  $n$  activités et soit  $m_j^i$  le mode particulier d'exécution pour l'activité  $j$  dans une solution  $i$ . Cette solution  $i$  au DTCTP peut être codée comme un vecteur de mode  $M^i$  avec  $M^i = (m_1^i, m_2^i, \dots, m_n^i)$ . S'il s'agit de minimiser la durée d'exécution du projet alors la Fitness de chaque solution pourra s'exprimer comme le ratio entre sa durée (longueur du chemin critique obtenu pour  $M^i$ ) et la durée minimum donnée par l'une des solutions de la population (solution la meilleure).

La mutation d'une solution peut simplement correspondre au changement de mode pour l'une des activités. Le cross-over à un seul point sélectionne une activité à partir de laquelle les matériaux sont échangés. Si l'on considère deux solutions  $a$  et  $b$  pour un projet comportant 5 activités et 3 modes d'exécution par activité avec  $M^a = (1, 2, 2, 3, 1)$  et  $M^b = (2, 1, 1, 2, 3)$  avec l'activité 3 comme point de croisement, on obtient les descendants suivants :  $M^{a'} = (1, 2, 1, 2, 3)$  et  $M^{b'} = (2, 1, 2, 3, 1)$ .

#### A.1.4.2 Optimisation par colonies de fourmis (Ant Colony Optimization, ACO)

Le principe de cette métaheuristique repose sur le comportement des fourmis qui utilisent pour communiquer une substance appelée phéromone produite par une glande située dans leur abdomen. Une fourmi éclairceuse commence par explorer au hasard l'environnement autour du nid. Si elle découvre une source de nourriture, elle rentre plus ou moins directement au nid en sécrétant des phéromones sur son chemin. Ces phéromones vont attirer d'autres fourmis passant à proximité qui vont donc suivre cette piste. En revenant au nid, ces mêmes fourmis vont renforcer la piste. Si deux pistes sont possibles pour atteindre la même source de nourriture, la plus courte sera donc parcourue par plus de fourmis que la plus longue. La piste courte sera par conséquent renforcée par un plus grand dépôt de phéromones tandis que la longue piste finira par disparaître car les phéromones sont volatiles.

Considérons le MRCPSP pour lequel chaque fourmi fait face à deux types de décision : l'ordre dans lequel les activités sont exécutées (ordonnancement) et leur mode d'exécution. La figure A.3 illustre la problématique décisionnelle d'une fourmi.

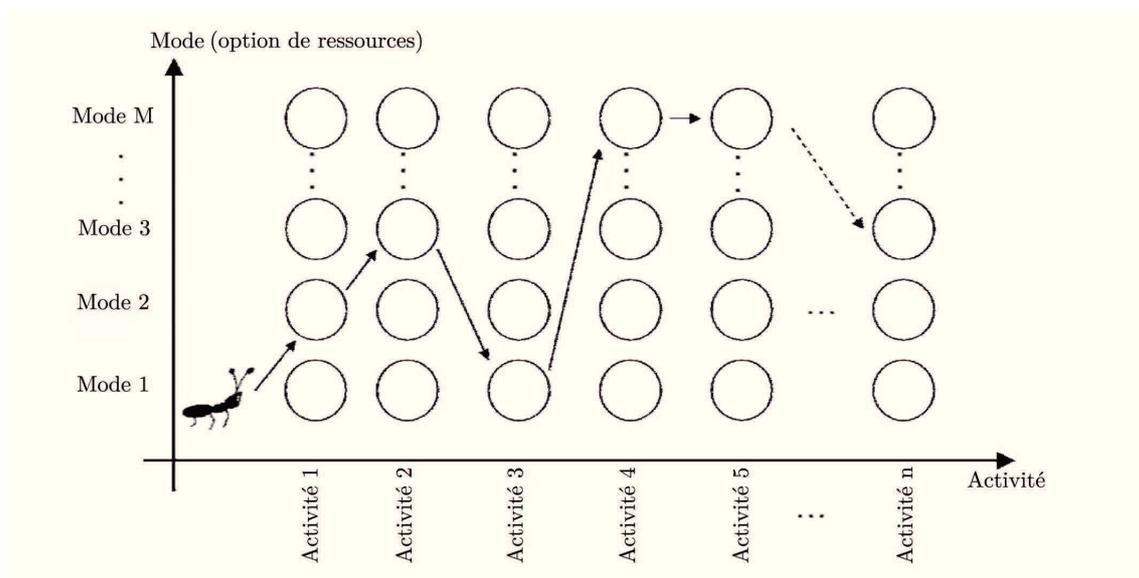


FIGURE A.3 – Trajet d'une fourmi pour le problème MRCPSP

Il existe une probabilité  $p_{ij}$  que la fourmi séquence l'activité  $j$  en position  $i$  dans la liste des activités et une autre probabilité  $q_m$  de sélection du mode  $m$  d'exécution de

l'activité  $j$ . La probabilité  $p_{ij}$  dépend positivement du taux de phéromone déposé pour le couple  $(i, j)$  qui traduit l'affectation de l'activité  $j$  en position  $i$  dans la liste; cette probabilité dépend également de la désirabilité de cette action. Cette désirabilité est mesurée par la différence entre le maximum des marges totales de toutes les activités pouvant être ordonnancées en position  $i$  et la marge totale de  $j$  (différence entre la date de début au plus tard et la date de début au plus tôt). Cela signifie que les activités sont d'autant plus désirables que leur marge totale est faible. Cette probabilité intègre également deux paramètres  $\alpha$  et  $\beta$  qui contrôlent respectivement l'influence du taux de phéromone et l'influence de la désirabilité de l'action. La probabilité  $q_m$  est définie selon le même principe, avec une désirabilité d'affectation du mode  $m$  qui va dépendre de la différence entre le maximum des durées des modes éligibles et la durée du mode  $m$ . Les modes ayant des durées longues sont les moins attractifs.

A chaque cycle de recherche, un nombre  $L$  de solutions sont générées en sélectionnant activités et modes d'exécution en fonction des probabilités  $p_{ij}$  et  $q_m$ . A la fin d'un cycle, les  $L$  solutions sont évaluées en calculant leur durée totale d'exécution (objectif de minimisation de la durée). Les taux de phéromone sont actualisés en tenant compte d'un facteur d'évaporation (diversification des chemins) et de l'accroissement des phéromones déposés sur chaque chemin. Cet accroissement de phéromones sera d'autant plus grand pour  $(i, j)$  que ce choix figure à la fois dans la meilleure solution locale trouvée au cours du cycle précédent et dans la meilleure solution globale (pression de sélection). Les cycles sont répétés jusqu'à atteindre une condition d'arrêt (nombre maximum de cycles ou temps de calcul).

#### A.1.4.3 Optimisation par essais particuliers (Particle Swarm Optimization, PSO)

La métaheuristique PSO s'appuie sur une représentation simplifiée du mouvement des organismes comme les oiseaux par exemple. Cette méthode d'optimisation se base sur la collaboration des individus entre eux et possède par conséquent des similarités avec les algorithmes de colonies de fourmis. Grâce à des règles de déplacement très simples dans l'espace des solutions, les particules (solutions) peuvent converger progressivement vers un minimum local. Cette métaheuristique semble cependant mieux fonctionner pour des espaces en variables continues.

Imaginons le scénario suivant. Un groupe d'oiseaux cherche aléatoirement de la nourriture. Il n'existe qu'une seule source de nourriture dans l'espace exploré par les oiseaux. Aucun oiseau ne connaît l'emplacement de la nourriture mais ils connaissent la distance qui les sépare d'elle. Par conséquent, quelle est la meilleure stratégie pour trouver cette nourriture? Une stratégie efficace consiste à suivre l'oiseau (solution) le plus proche de la nourriture (ayant la meilleure valeur de la fonction objectif).

L'algorithme PSO est initialisé avec un groupe de  $N$  particules avec  $i = 1..N$  et recherche les optima par des actualisations à chaque itération. Considérons par exemple une fonction  $f$  à minimiser de  $n$  variables sous un ensemble de contraintes. A chaque itération  $t$ , à toute particule  $i$  est associé un vecteur solution  $x^i$  de  $n$  variables c'est-à-dire  $x_t^i = (x_{1,t}^i, x_{2,t}^i, \dots, x_{n,t}^i)$  que l'on appelle également vecteur des positions de la particule  $i$ . Pour reprendre l'analogie avec les oiseaux, un vecteur de 3 variables permet de positionner un oiseau dans un espace à 3 di-

mensions. A toute particule  $i$  est aussi associé un vecteur de vélocités en  $t$ , soit  $v_t^i = (v_{1,t}^i, v_{2,t}^i, \dots, v_{n,t}^i)$ . De plus, chaque particule  $i$  conserve en mémoire la meilleure solution  $x_{\text{local}}^i = (x_{1,\text{local}}^i, x_{2,\text{local}}^i, \dots, x_{n,\text{local}}^i)$  qu'elle a visitée au cours de l'ensemble des itérations passées, avec  $f(x_{\text{local}}^i) = \min_{\tau=0..t-1} f(x_\tau^i)$ . L'algorithme conserve par ailleurs en mémoire la meilleure solution  $x_{\text{global}}$  obtenue jusqu'à présent sur l'ensemble des particules et qui est telle que  $f(x_{\text{global}}) = \min_{i=1..N} f(x_{\text{local}}^i)$ . La vélocité  $v_{j,t}^i$  associée à la  $j^{\text{ème}}$  composante du vecteur des vélocités de la particule  $i$  se calcule comme suit

$$v_{j,t}^i = w_t \cdot v_{j,t-1}^i + c_1 \cdot r_1 \cdot (x_{j,\text{local}}^i - x_{j,t-1}^i) + c_2 \cdot r_2 \cdot (x_{j,\text{global}} - x_{j,t-1}^i), \quad (\text{A.12})$$

où  $w_t$  est un facteur d'inertie (la particule tend à suivre sa propre voie),  $c_1$  est un coefficient d'apprentissage cognitif (la particule tend à revenir vers le meilleur site par lequel elle est déjà passée) et  $c_2$  est un coefficient d'apprentissage social (la particule tend à se diriger vers le meilleur site déjà atteint par ses voisins),  $r_1$  et  $r_2$  sont des valeurs entre 0 et 1 tirées aléatoirement et uniformément.

La  $j^{\text{ème}}$  composante du vecteur solution  $x_t^i$  est actualisée comme suit

$$x_{j,t}^i = v_{j,t}^i + x_{j,t-1}^i. \quad (\text{A.13})$$

On note que lorsque les variables doivent prendre des valeurs entières alors la vélocité peut être arrondie à l'entier le plus proche.

L'algorithme s'arrête au bout d'un certain nombre d'itérations ou lorsqu'une limite de temps est atteinte.

#### A.1.4.4 Optimisation par colonies d'abeilles (Artificial Bee Colony, ABC)

Comme les fourmis, les abeilles sont des insectes sociaux et vivent au sein d'une colonie organisée, formée d'une reine unique et d'abeilles de plusieurs types. Une abeille est capable, par une danse ou par la production de phéromones, de communiquer aux autres abeilles l'endroit où elle a découvert de la nourriture. Il y a trois types d'abeilles. Les abeilles actives (employed bees) sont associées à des sources spécifiques de nourriture; les abeilles spectatrices (onlookers) observent la danse des abeilles actives et les abeilles éclaireuses (scout bees) cherchent aléatoirement des sources de nourriture. Les abeilles spectatrices et éclaireuses sont des abeilles que l'on qualifie d'inactives (unemployed). Chaque source de nourriture correspond à une solution au problème. La quantité de nectar associée à une source de nourriture représente la Fitness d'une solution. La solution possédant la meilleure valeur de la fonction objectif est mémorisée.

Dans une phase d'initialisation, on commence par générer  $N$  solutions (sources de nourritures). Au cours d'une itération donnée, les abeilles actives recherchent une nouvelle source (solution) au voisinage d'une autre source trouvée antérieurement. Cette solution voisine est générée en combinant une solution  $i$  avec certaines composantes de solutions  $l$  choisies aléatoirement. La nouvelle solution ainsi obtenue remplacera  $i$  dans la population avec une probabilité dépendant de sa Fitness.

#### A.1.4.5 Compétition impérialiste (Imperialist Competitive Algorithm, ICA)

ICA (Imperialist Competitive Algorithm) est un algorithme évolutionnaire qui reproduit le comportement des empires en mettant en place un mécanisme de compétition afin d'affaiblir puis de supprimer les empires les moins puissants (c'est-à-dire les empires avec les moins bonnes solutions). L'ICA s'appuie sur un modèle d'évolution sociale et humaine tandis que les algorithmes génétiques sont fondés sur l'évolution biologique des espèces. La phase d'initialisation consiste à générer aléatoirement un ensemble de solutions incluses dans l'espace de recherche du problème. Ces solutions initiales sont appelées pays initiaux. Les pays sont l'équivalent des chromosomes dans les GA et des particules dans l'optimisation par essaims particuliers (PSO). La valeur de la fonction objectif associée à ces pays définit leur puissance. Les pays initiaux les plus puissants (ceux avec la meilleure valeur de la fonction objectif) deviennent des Empires en commençant à prendre le contrôle d'autres pays. Les deux opérateurs principaux utilisés dans cet algorithme sont l'assimilation et la révolution. L'assimilation pousse les colonies de chaque empire vers un état impérialiste tandis que la révolution introduit des changements aléatoires dans la position de certains pays. Pendant l'assimilation et la révolution une colonie peut atteindre une meilleure position avec une possibilité de prendre le contrôle de l'empire en entier et donc de lui subtiliser l'état d'impérialisme. Les étapes de l'algorithme sont résumées ci-après.

##### Initialisation

Générer  $N$  colonies (solutions) aléatoirement et calculer la valeur de la fonction objectif (durée totale du projet à minimiser) associée à chaque solution  $i = 1..N$ . Chaque colonie solution  $i$  est codée par une double liste, celle des activités à séquencer et celle des modes d'exécution pour les activités (comme dans la description de l'ACO).

Choisir les  $B$  meilleures solutions ( $b = 1..B$ ) pour en faire des impérialistes (un impérialiste est une solution). La puissance  $p_b$  de l'impérialiste  $b$  peut se calculer comme une mesure relative de la durée totale d'exécution associée à la solution  $b$  par rapport aux meilleures et moins bonnes durées totales obtenues dans la population.

Affecter les  $N - B$  solutions (colonies) aux impérialistes  $b = 1..B$  par tirage aléatoire selon leur puissance (probabilité)  $p_b$ . Un empire  $e$  (avec  $e = 1..E$  et  $E = B$  dans cette phase d'initialisation) va donc être constitué d'un impérialiste  $b$  (solution  $b$ ) et de l'ensemble des colonies (solutions) qui lui ont été affectées. La puissance  $p_e$  de chaque empire  $e$  est actualisée en fonction des durées totales d'exécution des solutions qu'il contient.

##### Assimilation (intensification)

Dans ce mouvement, les colonies-solutions contenues dans un empire donné  $e$  vont se rapprocher de la solution impérialiste  $b$  qu'il contient. Ces colonies vont récupérer de l'information contenues dans  $b$ . Par exemple, une colonie solution va récupérer dans sa liste d'activités à séquencer une partie du séquençement de l'impérialiste  $b$  de l'empire auquel elle appartient. Pour un projet de 5 activités et une liste  $S^b$  des activités à séquencer dans la solution impérialiste  $b$  avec  $S^b = (2, 3, 1, 5, 4)$  et une liste  $S^i = (1, 2, 4, 3, 5)$  pour une colonie solution  $i$ ; après assimilation on peut obtenir par exemple  $S^{i'} = (1, 3, 4, 5, 5)$  qui sera transformée en  $(1, 3, 4, 5, 2)$  pour la

rendre réalisable.

### Révolution (diversification)

La révolution consiste à appliquer un ou plusieurs opérateurs de modification d'une solution de façon aléatoire sur une partie de la population totale des  $N$  solutions. Par exemple on pourra choisir de modifier de la sorte 10% des  $N = 100$  solutions soit 10 solutions et leur appliquer une mutation. Sur la solution  $S^i = (1, 2, 4, 3, 5)$  cela pourra donner  $S^{i'} = (1, 2, 5, 3, 4)$ .

### Actualisation et compétition

Après chaque mouvement d'assimilation ou de révolution, on redéfinit l'impérialiste dans chacun des empires comme la solution ayant la durée totale minimum (celle-ci peut en effet changer). La puissance de chaque empire est actualisée en fonction des durées (actualisées) associées aux solutions qu'il contient. On calcule ensuite une puissance relative pour chaque empire (par rapport à la puissance des autres). Cette puissance relative définit la probabilité qu'un empire absorbe l'empire le plus faible. On recommence les cycles jusqu'à ce qu'il n'y ait plus qu'un seul empire qui gouverne l'ensemble de toutes les solutions-colonies.

## A.2 Optimisation multi-objectif

En optimisation multi-objectif, on considère un ensemble de fonctions objectifs  $f_k(x)$  à optimiser avec  $k = 1..K$  et  $x = (x_1, \dots, x_n)$  sous un ensemble de contraintes et en théorie on cherche une solution  $x^*$  qui optimise simultanément tous les objectifs  $f_k$ . Cependant une telle solution  $x^*$  n'existe pratiquement jamais car dans la plupart des situations réelles, les objectifs à optimiser sont en conflit les uns avec les autres. Il y a alors une difficulté à établir une définition de l'optimum dans ce cas là. Le décideur peut toutefois exprimer le fait qu'il préfère une solution à une autre mais il n'existe pas une solution meilleure que toutes les autres.

### A.2.1 Présentation générale des approches de résolution

La figure A.4 présente les deux grands types d'approches de résolution pour les problèmes multi-objectif. Les approches agrégées dites méthodes scalaires consistent à agréger les différents objectifs pour se ramener à un problème mono-objectif. Les approches non agrégées traitent les objectifs séparément. Parmi ces approches non agrégées on distingue celles qui utilisent le concept de dominance au sens de Pareto de celles qui ne s'appuient pas sur ce concept. On dit d'une solution qu'elle en domine une autre si elle donne des valeurs des fonctions objectif au moins aussi bonnes et une valeur strictement meilleure pour l'un des objectifs au moins. Les approches fondées sur Pareto cherchent à déterminer l'ensemble des solutions Pareto optimales qui par définition sont des solutions telles qu'il n'est pas possible d'améliorer l'un des objectifs sans détériorer au moins l'un des autres objectifs. Considérons par exemple l'objectif  $f_1$  de durée totale et l'objectif  $f_2$  de coût total et deux solutions  $x$  et  $y$  qui correspondent chacune à un ordonnancement particulier (c'est-à-dire l'ensemble des dates de début de toutes les activités d'un projet). Si l'on a  $(f_1(x), f_2(x)) = (15, 50)$  et  $(f_1(y), f_2(y)) = (15, 75)$  alors  $x$  domine  $y$  car pour une même durée de 15, le coût est inférieur strictement. Si l'on peut de plus affirmer qu'il est impossible d'obtenir

une durée inférieure à 15 avec un budget de 50, ou s'il est impossible de diminuer le coût en conservant une durée de 15, alors la solution  $x$  est Pareto optimale.

### Approches agrégées (méthodes scalaires)

- **Méthodes  $\epsilon$ - contrainte**
- **Somme pondérée**
- **Goal attainment** (But à atteindre)
- **Goal programming** (Programmation avec buts)

### Approches non agrégées

#### Non Pareto

- **Optimisation lexicographique**
- **Vector Evaluated Genetic Algorithm (V E G A )**

#### Pareto

- **Algorithmes génétiques multi-objectif**
  - Multi Objective Genetic Algorithm (MOGA)
  - Non Sorted Genetic Algorithm (NSGAI)
  - Fuzzy Clustering Genetic Algorithm (FCGA)
- **Autres algorithmes évolutionnaires multi-objectif**
  - Multi Objective Ant Colony Optimisation (MOACO)
  - Multi Objective Imperialist Competitive Algorithm (MOICA)
  - Multi Objective Artificial Bee Colony (MOABC)

FIGURE A.4 – Approches de résolution multi-objectif

Nous présentons d'abord les méthodes agrégées dans la sous-section A.2.2 puis, dans un souci de clarté, nous consacrons une sous-section aux approches non agrégées et non Pareto (sous-section A.2.3) et une autre sous-section aux approches agrégées fondées sur Pareto (sous-section A.2.4).

## A.2.2 Les méthodes agrégées dites méthodes scalaires

Le principe de ces méthodes est de revenir à un problème mono-objectif avec différentes agrégations qui nécessitent un paramétrage, comme par exemple la somme pondérée qui obligent à la définitions de poids sur les objectifs. La difficulté de ces approches réside donc dans la détermination des paramètres, ce qui implique une connaissance a priori des préférences des décideurs.

### A.2.2.1 La méthode $\epsilon$ -contrainte

Dans cette méthode, une seule fonction objectif  $f_l$  est optimisée tandis que les autres fonctions objectif sont exprimées comme des contraintes, c'est-à-dire  $f_k \leq \epsilon_k$ . L'objectif  $f_l$  représente l'objectif préféré du décideur. Ses préférences sont également représentées par les valeurs affectées aux paramètres  $\epsilon_k$ . Par exemple, l'objectif  $f_l$  préféré peut être la durée et l'objectif  $f_k$ , le coût qui s'exprimera alors sous la forme d'une contrainte de budget ( $\epsilon_k$ ), ce budget devant être prédéterminé par le décideur.

### A.2.2.2 Somme pondérée

Cette approche consiste à transformer un problème multi-objectif en un problème avec une seule fonction objectif qui s'exprime comme une somme pondérée des différents objectifs. Les poids représentent l'importance accordée par le décideur aux différents objectifs. L'avantage de cette méthode est de pouvoir utiliser ensuite une technique d'optimisation mono-objectif pour résoudre le problème agrégé. Mais la difficulté réside dans le paramétrage des poids et dans l'inaptitude de la méthode à atteindre certaines solutions Pareto optimales quels que soient les poids choisis.

### A.2.2.3 Goal Attainment et Goal Programming

La méthode du but à atteindre que nous avons présentée dans le chapitre 4 consiste pour le décideur à spécifier les buts c'est-à-dire les valeurs cible que chacun des objectifs doit atteindre et les poids qu'il accorde à chacun de ces objectifs.

Dans la programmation par buts, la fonction objectif consiste à minimiser une somme pondérée des écarts entre les objectifs et leur valeur cible.

## A.2.3 Méthodes non agrégées non Pareto

Ces méthodes traitent les objectifs séparément. Dans l'approche lexicographique, le décideur classe les objectifs par ordre de préférence. Dans l'algorithme V.E.G.A, l'importance accordée à chacun des objectifs est la même. Ces méthodes exploitent ainsi des relations dérivées de la dominance parce que la relation de dominance elle-même n'intègre pas la notion de préférence et manque de flexibilité à cet égard.

### A.2.3.1 Méthode lexicographique

Dans cette méthode, les objectifs sont classés en fonction d'un ordre d'importance proposé par le décideur. Ensuite, les fonctions objectif sont traitées dans cet ordre pour obtenir l'optimum. Supposons que les fonctions objectif à minimiser  $f_k$  soient

classées par ordre décroissant d'importance. Le premier problème consiste alors à minimiser  $f_1(x)$  et l'on obtient la solution optimale avec  $f_1^*$ . Dans un deuxième problème, il s'agit alors de minimiser  $f_2(x)$  sous la contrainte  $f_1(x) = f_1^*$ . Dans un troisième problème on minimise  $f_3(x)$  sous les contraintes  $f_1(x) = f_1^*$  et  $f_2(x) = f_2^*$ . Ces optimisations successives se poursuivent jusqu'à ce que l'on traite le dernier objectif.

### A.2.3.2 Vector Evaluated Genetic Algorithm (VEGA)

L'algorithme VEGA considère une population de  $N$  individus répartis en  $K$  sous-populations où  $K$  est le nombre de fonctions objectif du problème. A chaque sous-population est associée une seule des fonctions objectifs qui permet de déterminer l'efficacité d'un individu au sein de cette sous-population. Ainsi, chaque sous-population groupe progressivement les meilleurs individus pour l'objectif associé. Les individus sont mélangés et les croisements sont opérés en tenant compte de l'efficacité de chaque individu. Cependant, le risque d'une telle méthode est de se retrouver à la fin avec une population constituée d'individus moyens dans tous les objectifs. En outre, et comme la somme pondérée, cet algorithme ne permet pas d'atteindre certaines solutions Pareto optimales.

### A.2.4 Les méthodes non agrégées fondées sur Pareto

Une façon raisonnable de résoudre un problème multi-objectif consiste à explorer un ensemble de solutions dont chacune satisfait convenablement les objectifs sans être dominée par une autre solution. Toute métaheuristique à population de solutions s'adapte naturellement à ce problème de recherche non pas d'une solution unique mais d'un ensemble de solutions non dominées. Idéalement, on cherche l'ensemble des solutions Pareto optimales dont les valeurs des fonctions objectifs forment ce que l'on appelle la frontière de Pareto. Cette frontière de Pareto ne peut se déterminer que si l'on calcule les valeurs des fonctions objectifs associées à toutes les solutions réalisables, ce qu'il est impossible de faire en pratique. On cherche alors à se rapprocher le plus possible de cette frontière de Pareto dite frontière théorique. Une méthode appropriée pour cela doit être en mesure de guider le processus de recherche vers la frontière de Pareto tout en assurant une bonne répartition des solutions sur la frontière. Pour cela, il est fondamental de maintenir la diversité des solutions au cours de l'exploration. En effet, sans mesures préventives, la population tend à former un petit nombre de clusters (amas) avec le risque que les solutions restent piégées dans des optima locaux. La figure A.5 illustre ce que l'on attend d'une bonne approche de recherche de la frontière de Pareto dans un exemple à deux fonctions objectifs : la durée totale et la qualité globale d'un projet. Il s'agit donc de s'approcher le plus possible de la frontière théorique (partie gauche de la figure) et également de bien la représenter par des solutions uniformément réparties (partie droite de la figure).

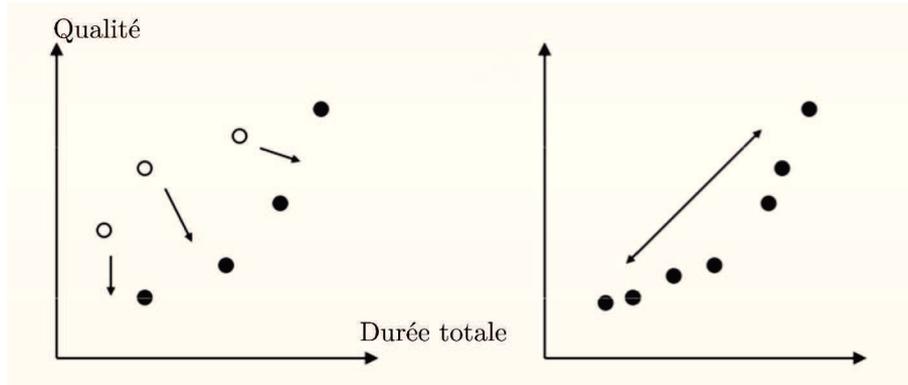


FIGURE A.5 – Recherche satisfaisante d'une frontière efficiente

#### A.2.4.1 Dominance, Pareto optimalité, frontière de Pareto et classement Pareto

Dans cette sous-section nous étudions de façon plus approfondie les notions de base associées au concept de Pareto optimalité.

**Dominance.** Considérons deux solutions  $x$  et  $y$  et les fonctions objectifs  $f_k$  à minimiser avec  $k = 1..K$ . On dit que  $x$  domine  $y$  si elle donne des valeurs des fonctions objectif au moins aussi bonnes que  $y$  et une valeur strictement meilleure pour l'un des objectifs au moins. Plus formellement,  $x$  domine  $y$  si  $f_k(x) \leq f_k(y)$  pour tout  $k = 1..K$  et s'il existe au moins un objectif  $k'$  tel que  $f_{k'}(x) < f_{k'}(y)$ .

Exemple. On considère un codage de la solution pour le DTCQTP en double liste, une liste des activités à ordonnancer et une liste de leur mode d'exécution et les trois critères habituels de durée totale ( $f_1$ ), coût total ( $f_2$ ) et qualité globale ( $f_3$ ). Soit  $x$  et  $y$  deux double listes activités et modes différentes. Imaginons par exemple que  $F(x) = (f_1(x), f_2(x), f_3(x)) = (5, 2, 0.9)$  et  $F(y) = (5, 2, 0.4)$ . Dans ce cas, la solution (liste d'activités et modes d'exécution)  $x$  domine la solution  $y$  car  $x$  et  $y$  ont même durée et même coût mais la qualité globale associée à  $x$  est strictement supérieure à la qualité globale associée à  $y$ .

**Pareto optimalité d'une solution.** Une solution  $z$  est Pareto optimale s'il n'existe aucune autre solution qui la domine du point de vue d'un ou plusieurs critères  $k$  dont la valeur est donnée par  $f_k$ . Autrement dit, une solution est Pareto optimale s'il n'est pas possible d'améliorer l'un des objectifs sans détériorer au moins l'un des autres objectifs.

Exemple. Considérons les deux solutions précédentes  $x$  et  $y$  et une solution  $z$  avec  $F(z) = (5, 2, 0.95)$ . Cette solution  $z$  domine  $x$  et  $y$  mais pour dire d'elle qu'elle est Pareto optimale, il faut affirmer de plus qu'on ne peut pas obtenir une qualité strictement supérieure à 0.95 avec un coût inférieur ou égal à 2 ou une durée inférieure ou égale à 5. Autrement dit, l'amélioration de la qualité au delà de 0.95 passe nécessairement par une augmentation des coûts (utilisation d'une machine plus précise qui coûte plus cher) ou par une augmentation de la durée (qui s'accompagnera alors aussi d'une augmentation des coûts car on utilise les ressources sur une durée plus longue).

**Frontière de Pareto (ou frontière efficiente).** La frontière de Pareto notée  $FP^*$  est l'ensemble des valeurs des fonctions objectifs associées à l'ensemble  $P^*$  des solutions Pareto optimales.

Exemple. Ainsi la solution  $z$  est Pareto optimale et  $F(z) = (5, 2, 0.95)$  constitue un point de la frontière de Pareto.

**Rang de dominance.** Le rang de dominance d'une solution est égal à  $r$  si un nombre  $r - 1$  de solutions la dominent. Ainsi, toutes les solutions non dominées ont un rang de dominance égal à 1 car aucune autre solution les domine; les solutions dominées par une seule solution ont un rang de 2, etc.

**Classement Pareto.** Un ensemble  $P$  (population) constitué de  $N$  solutions peut ainsi être partitionné en  $R$  sous-ensembles  $P^r$  avec  $r = 1..R$  comprenant chacun toutes les solutions ayant un rang de dominance égal à  $r$ . On note que  $P^1 = P^*$  c'est-à-dire que les solutions de rang de 1 sont potentiellement Pareto optimales. Par ailleurs, on peut dire aussi que chaque sous-ensemble  $P^r$  pour  $r \geq 2$  contient des solutions non dominées de rang  $r$ . En effet, la notion de non dominance peut être interprétée en termes relatifs au sens où dans  $P^r$  aucune solution de rang  $r$  n'en domine une autre de rang  $r$ . Cette méthode de classement nécessite de comparer chaque solution aux  $N - 1$  autres sur la base de chacun des critères  $k = 1..K$ , ce qui donne au total  $N \cdot (N - 1) \cdot K$  comparaisons à effectuer. Dans la suite, on note  $r(x)$  le rang de dominance d'une solution  $x$ .

Exemple. Considérons deux fonctions objectifs  $f_1$  et  $f_2$  où  $f_1$  est à maximiser (qualité exprimée en pourcentage d'une check liste) et  $f_2$  est à minimiser (durée exprimée en jours), et imaginons cinq solutions A, B, C, D, E correspondant à des combinaisons particulières d'ordre sur les  $n$  activités d'un projet donné et de modes d'exécution qui leurs sont affectés. On calcule pour ces cinq solutions la qualité globale et la durée totale. Ces cinq solutions sont représentées à la figure A.6 dans l'espace des valeurs des fonctions objectifs avec  $f_1$ , la qualité globale en abscisse et  $f_2$  la durée totale d'exécution du projet en ordonnée.

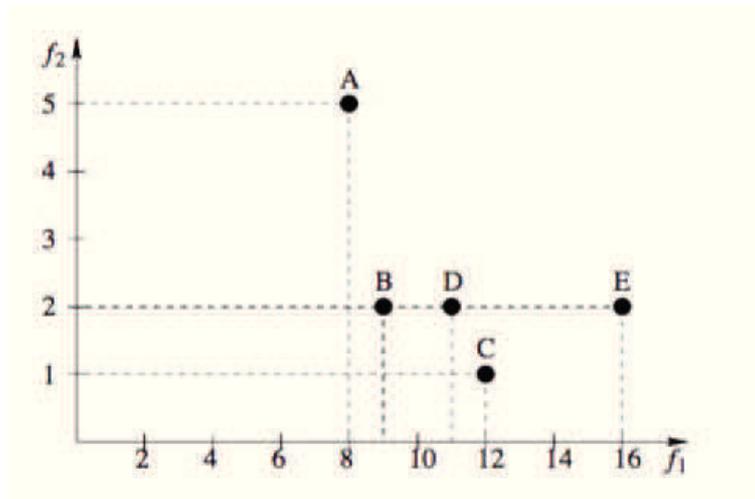


FIGURE A.6 – Représentation de solutions dans l'espace qualité - durée.

On considère d'abord l'ensemble de toutes les solutions et on identifie celles qui ne sont dominées par aucune autre. On observe que la solution A a la qualité la plus faible et la durée la plus élevée, elle est donc dominée par les solutions B,C,D,E. La solution B est dominée par au moins une solution qui est la solution évidente D ayant même durée mais une qualité plus élevée. La solution D est dominée à l'évidence par E (même durée mais qualité plus élevée). En revanche les solutions C et E ne sont dominées par aucune autre et ne se dominent pas non plus entre elles puisque E n'est pas meilleure que C pour tous les critères et inversement. Les solutions C et E sont donc placées dans  $P_1$  et leurs valeurs de fonction objectif dans la frontière  $PF_1$ .

On considère maintenant les solutions restantes A, B et D et on recherche celles qui ne sont dominées par aucune autre. On constate que seule D n'est dominée par aucune autre solution, on lui attribue alors le rang  $r = 2$ .

On considère les solutions restantes c'est-à-dire celles auxquelles on n'a pas encore affecté de rang, à savoir A et B et on voit que B n'est pas dominée par A. On affecte le rang  $r = 3$  à cette solution que l'on range dans  $P_3$ . Il ne reste plus que la solution A qui, toute seule ne peut être dominée par aucune autre et on lui affecte par conséquent le rang 4 tout en la classant dans  $P_4$ . Le résultat de ce classement est donné à la figure A.7.

#### A.2.4.2 Algorithme évolutionnaire multi-objectif générique (Generic MOEA, Multi Objective Evolutionary Algorithm)

Toute métaheuristique à population de solutions s'adapte naturellement aux problèmes d'optimisation multi-objectif pour lesquels on recherche non pas une solution mais un ensemble de solutions Pareto optimales. Comme ces métaheuristicques explorent différentes régions de l'espace des solutions, elles possèdent naturellement la faculté de diversifier les solutions pour des problèmes difficiles. Dans le GA, l'opérateur de cross-over peut exploiter la structure des solutions par rapport aux différents

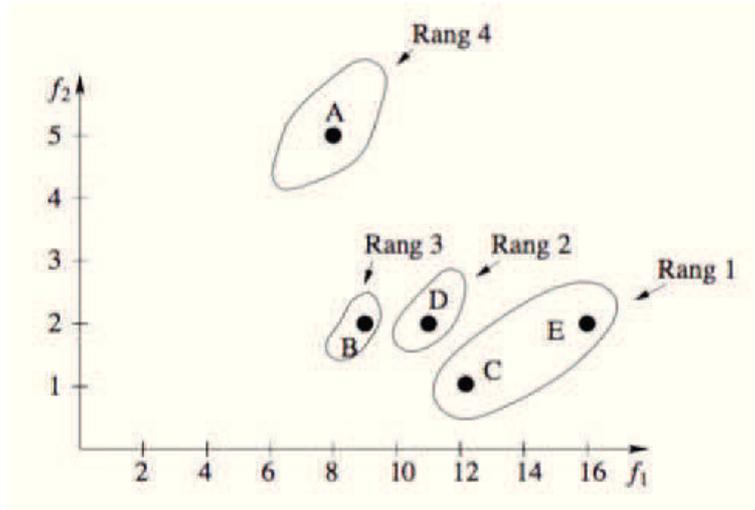


FIGURE A.7 – Rang de dominance des solutions.

objectifs pour générer de nouvelles solutions non dominées dans des régions inexplorées de la frontière de Pareto.

Tout MOEA efficace doit incorporer les étapes génériques suivantes. On note que pour clarifier l'exposé, certaines étapes sont développées en utilisant une stratégie particulière qui appartient à des algorithmes plus spécifiques.

**Initialisation.** Etape d'initialisation de la population  $P_0$  en générant aléatoirement  $N$  solutions et calculer pour chacune d'elles les valeurs des objectifs  $f_k$  avec  $k = 1..K$ . Les solutions peuvent être codées en binaire, entier ou réel.

**Classement Pareto.** Celui-ci est appliqué à la population  $P_0$  et permet d'obtenir la sous-population  $P_0^*$  des solutions non dominées de rang 1 sur laquelle nous raisonnons dans la suite pour simplifier.

**Estimation de la densité.** On note  $P^{**}$  l'ensemble de toutes les solutions non dominées archivées au cours des itérations précédentes et  $PF^{**}$  l'ensemble des valeurs des fonctions objectifs associées à ces solutions ;  $PF^{**}$  est donc l'ensemble des points de la meilleure frontière de Pareto obtenue jusqu'à présent. Naturellement, au début  $P^{**}$  est un ensemble vide. Pour constituer une frontière de Pareto avec des points uniformément répartis, il faut éviter que des régions de cette frontière soient surpeuplées et d'autres, presque désertes. On va alors chercher à transformer par la mutation ou le cross-over, des solutions de faible densité, c'est-à-dire des solutions telles que leurs voisins ont des valeurs de fonction objectifs éloignées. Par cette transformation, on espère obtenir de nouvelles solutions qui iront peupler les régions de densité faible.

Pour évaluer la densité d'une solution, il est possible de lui associer ce qu'on appelle une distance de surpeuplement (crowding distance ou encore distance d'influence). Pour cela, on considère chaque objectif  $f_k$  pris individuellement et on constitue pour chacun d'eux une liste ordonnée de toutes les solutions (non dominées) de  $P_0^*$  par ordre croissant de leur valeur de  $f_k$ . Soit  $x_{i,k}$  l'individu qui occupe la place

$i$  dans la liste ordonnée des individus relativement à l'objectif  $f_k$ . La distance de surpeuplement de l'individu  $x_{i,k}$  par rapport à l'objectif  $f_k$  et notée  $\delta_k(x_{i,k})$  peut se calculer comme suit

$$\delta_k(x_{i,k}) = \begin{cases} \frac{f_k(x_{i+1,k}) - f_k(x_{i-1,k})}{f_k^{\max} - f_k^{\min}} & \text{pour } 2 \leq i \leq |P_t^*| - 1, \\ \infty & \text{pour } i = 1 \text{ et } i = |P_t^*|, \end{cases} \quad (\text{A.14})$$

où  $f_k^{\max}$  et  $f_k^{\min}$  désignent respectivement les valeurs maximale et minimale de l'objectif  $f_k$  trouvées jusqu'à présent et  $t = 0$  puisque l'on traite l'itération initiale. La distance de surpeuplement  $\delta(x)$  de la solution  $x$  s'obtient en additionnant simplement ses distances de surpeuplement par rapport à chaque objectif  $k$ , soit  $\delta(x) = \sum_{k=1..K} \delta_k(x_{i,k})$ . Une valeur faible de  $\delta(x)$  signifie que la solution  $x$  est peu différente de ses voisins ; la solution  $x$  appartient à une région dense. Elle aura alors une faible probabilité d'être sélectionnée pour cross-over et mutation. On note que les deux solutions extrêmes de la liste seront préservées (probabilité nulle de sélection pour transformation).

**Opérateurs évolutionnaires (mutation, cross-over, etc.) et constitution de la population à l'itération suivante.** On sélectionne aléatoirement des individus de  $P_0^*$  en fonction de leur fitness qui s'appuie sur leur distance de surpeuplement. Plus cette distance est élevée, moins la région est dense et plus la probabilité de sélection est élevée. En leur appliquant des opérateurs évolutionnaires, on obtient un ensemble de descendants que l'on appelle  $P_0^{\text{child}}$ . Il s'agit ensuite de combiner l'ensemble  $P_0^*$  de solutions non dominées et l'ensemble  $P_0^{\text{child}}$  pour obtenir une nouvelle population  $P_1$  de  $N$  individus non dominés qui constitue la population au début de l'itération 1. Pour cela, on commencera par appliquer le classement Pareto sur l'ensemble  $P_0^* \cup P_0^{\text{child}}$  ; certaines solutions non dominées de l'ensemble  $P_0^*$  pouvant se retrouver dominées par certaines solutions de  $P_0^{\text{child}}$ . Il se peut qu'après classement Pareto, la nouvelle population  $P_1$  contienne moins de  $N$  solutions, auquel cas, il est possible d'appliquer à nouveau des opérateurs évolutionnaires sur une nouvelle sélection d'individus provenant de  $P_0^*$ . Si au contraire  $P_1$  contient plus de  $N$  solutions, on pourra employer une stratégie élitiste pour la réduire.

**Actualisation de l'archive des meilleures solutions non dominées.** Il s'agit ici de fusionner  $P_1$  avec  $P^{**}$  pour actualiser  $P^{**}$  et pour cela, il faut ré-appliquer le classement Pareto car il se peut certaines solutions de  $P_1$  dominant certaines solutions de  $P^{**}$  et inversement.

**Itérations suivantes.** On note qu'au début de toute itération  $t$  pour  $t \geq 1$ , on a  $P_t = P_t^*$  c'est-à-dire que la population au début de l'itération  $t$  est constituée uniquement de solutions non dominées. Il faut ensuite actualiser  $f_k^{\max}$  et  $f_k^{\min}$  pour calculer les densités des solutions de  $P_t$  et ce, grâce à l'équation (A.14). On applique ensuite les opérateurs évolutionnaires sur des individus issus de  $P_t^*$  sélectionnés aléatoirement selon leur distance  $\delta$  et l'on obtient ainsi  $P_t^{\text{child}}$ . Puis on effectue un classement Pareto sur  $P_t^* \cup P_t^{\text{child}}$  pour obtenir  $P_{t+1}^*$  dont on fait en sorte que sa taille soit égale à  $N$ . L'archivage s'obtient en effectuant un classement Pareto sur  $P_{t+1}^* \cup P^{**}$ , pour obtenir la version actualisée de  $P^{**}$ .

### A.2.4.3 Stratégies de préservation de la diversité

**Distance de surpeuplement (crowding distance).** Dans le MOEA générique précédent, la stratégie de préservation de la diversité employée repose sur le calcul d'une distance de surpeuplement entre les points de la frontière efficiente (voir équation (A.14)). On a aussi supposé que toute population à une itération donnée est uniquement constituée de solutions de rang de dominance égal à 1. Néanmoins, il est possible de travailler également sur une population  $P_t = \cup_{r=1..R} P_t^r$  constitué de sous-populations de solutions de rang de dominance égal à  $r$  et de calculer les distances entre tout couple de points appartenant à une même sous-population  $P_t^r$  (selon l'équation (A.14)). L'algorithme NSGAI (Non Sorting GA II) travaille sur des sous-populations de rangs de dominance différents et utilise la distance de surpeuplement pour départager les cas d'égalité de la façon suivante. Sélectionner aléatoirement deux solutions  $x$  et  $y$ ; si ces deux solutions ont même rang de dominance  $r$  (appartiennent toutes deux à  $P^r$ ) alors conserver pour transformation (par les opérateurs génétiques) la solution ayant la distance de surpeuplement la plus grande. Sinon c'est la solution avec le rang de dominance le plus élevé que l'on retient.

**Partage de Fitness et comptage de niche (Fitness sharing and niche counting).** Dans cette approche, la taille (ou rayon) d'un voisinage (ou niche) est contrôlée par un paramètre de rayon de la niche, noté  $\sigma_{\text{share}}$ . On compte le nombre de solutions situées dans la même niche (comptage de niche) et la Fitness est diminuée proportionnellement au nombre d'individus partageant la même niche. L'objectif est de promouvoir la génération de solutions dans les régions les moins peuplées.

Pour pouvoir appliquer le partage de Fitness, il faut préalablement calculer les distances entre tout couple  $(x, y)$  de solutions de même rang  $r$  c'est-à-dire appartenant à une même sous-population  $P^r$  et ce, pour tout  $r = 1..R$ . On peut utiliser la distance Euclidienne  $d(x, y)$  sur les valeurs des fonctions objectif en normalisant par rapport aux valeurs maximales et minimales (respectivement  $f_k^{\max}$  et  $f_k^{\min}$ ) trouvées jusqu'à présent, c'est-à-dire

$$d(x, y) = \sqrt{\sum_{k=1}^K \left( \frac{f_k(x) - f_k(y)}{f_k^{\max} - f_k^{\min}} \right)^2} \quad \forall (x, y) \in P^r, \quad r = 1..R. \quad (\text{A.15})$$

Le compte de niche  $\nu(x)$  associé à toute solution  $x$  de  $P$  se calcule comme suit

$$\nu(x) = \sum_{\substack{y \in P \\ r(y) = r(x)}} \max \left\{ \frac{\sigma_{\text{share}} - d(x, y)}{\sigma_{\text{share}}}, 0 \right\} \quad (\text{A.16})$$

Une solution ayant un voisinage (niche) très peuplé possèdera un compte de niche  $\nu$  élevé et possèdera alors une faible probabilité de se reproduire. Cette stratégie de niche permet de promouvoir la génération de solutions dans les régions les moins peuplées de la frontière. On note que le MOGA utilise ce type de distance Euclidienne sur les objectifs mais la distance peut aussi se calculer dans l'espace des solutions,

en utilisant par exemple la distance de Hamming qui compte le nombre de bits différents entre deux solutions  $x$  et  $y$  codées en binaire. Le NSGA s'appuie sur une distance de Hamming.

La Figure A.8 représente graphiquement la façon dont on considère le voisinage d'une solution fondée sur la niche (partie gauche de la figure) ou sur le distance de surpeuplement (partie droite de la figure). Visuellement, on constate immédiatement l'importance de bien paramétrer le rayon de niche.

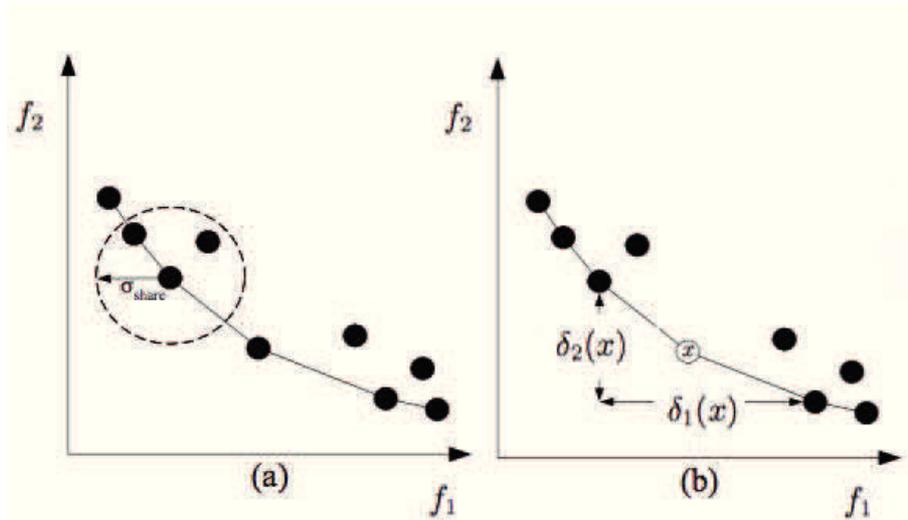


FIGURE A.8 – Partage de Fitness (a) et distance de surpeuplement (b).

On note enfin que la stratégie la plus simple pour calculer une probabilité de sélection consiste à faire une interpolation linéaire sur la base du rang de dominance. Ainsi, les solutions non dominées de rang 1 se voient assigner une efficacité maximale de 1 tandis que les solutions qui sont dominées par un nombre maximum de solutions ont une efficacité nulle. La probabilité de sélection sera d'autant plus élevée que l'efficacité l'est. Ce type de calcul de probabilité de sélection est utilisé dans la version originelle du MOGA.

#### A.2.4.4 Elitisme

Dans un contexte mono objectif, l'élitisme se traduit par la préservation de la meilleure solution trouvée jusqu'à présent ; cette meilleure solution a donc une immunité contre la sélection (pour reproduction) et survivra toujours d'une génération à l'autre. A cet égard, toutes les solutions non dominées trouvées par un GA multi-objectif sont considérées comme des solutions élites. Mais l'implémentation d'une stratégie de préservation des élites (élitisme) n'est pas aussi évidente en multi-objectif en raison du nombre possiblement important de solutions élites.

La sélection aléatoire ne garantit bien évidemment pas que les solutions non dominées survivront dans les prochaines générations de sorte qu'une stratégie élitiste évidente consiste à copier dans  $P_{t+1}$  toutes les solutions non dominées trouvées dans

$P_t$ . Dans ce cas, la taille des populations varie au cours des générations (populations de taille dynamique). Il se peut toutefois que l'on souhaite que la taille de la population demeure la même d'une génération à l'autre, auquel cas un mécanisme de sélection élitiste doit être mis en place. Nous présentons ci-après la stratégie élitiste déployée dans l'algorithme NSGAI.

**Stratégie élitiste du NSGAI.** On rappelle qu'à l'itération  $t$  on sélectionne des individus de la population  $P_t$  selon leur distance de surpeuplement pour leur appliquer les opérateurs de mutation et cross-over. On génère ainsi une population  $P_t^{\text{child}}$  généralement de taille  $N$ . On applique le classement Pareto sur la fusion  $P_t \cup P_t^{\text{child}}$  pour déterminer les sous-ensembles  $P_t^r$  de solutions de même rang  $r$ , avec  $r = 1..R$ . La population de la génération suivante  $P_{t+1}$  est alors remplie d'abord avec les solutions de  $P_t^1$ , puis avec celles de  $P_t^2$  etc. Si, pour achever de remplir  $P_{t+1}$  (de sorte à atteindre une taille  $N$ ) on ne peut prendre qu'une partie des solutions d'un ensemble  $P_t^r$  (avec le rang le plus élevé possible), on tirera les solutions de cet ensemble en fonction de leur distance de surpeuplement.

**Archive et clustering (partitionnement).** La plupart des algorithmes MO utilisent une archive  $P^{**}$  des meilleures solutions non dominées qui est une liste (ensemble) externe à l'algorithme. Cette archive est actualisée en fonction des solutions non dominées que l'on trouve à chaque itération. Néanmoins, cette actualisation nécessite un temps important de calcul de sorte que des procédures efficaces d'actualisation ont été proposées dans la littérature. En outre, comme il peut exister un très grand nombre de solutions non dominées, l'ensemble  $P^{**}$  peut devenir extrêmement grand. Des techniques d'élagage (pruning) de la liste ont alors été développées pour que la liste n'excède pas une certaine taille  $N$ . Certaines sont fondées sur des méthodes de partitionnement (clustering), comme c'est le cas dans le SPEA (Strength Pareto Evolutionary Algorithm). Nous présentons ci-après la méthode d'élagage "average linkage clustering method" utilisée par le SPEA.

**Etape 1.** Affecter chaque solution  $x$  de  $P^{**}$  à un cluster  $c_i$ ,  $C = \{c_1, c_2, \dots, c_M\}$

**Etape 2.** Calculer une distance entre toutes paires de clusters

$$d(c_i, c_j) = (1/(|c_i| \cdot |c_j|)) \sum_{x \in c_i, y \in c_j} d(x, y) \text{ où } d(x, y) \text{ peut être calculée en utilisant l'équation (A.15).}$$

**Etape 3.** Fusionner la paire de clusters qui possède la distance minimum en un seul (et nouveau) cluster

**Etape 4.** Si  $|C| \leq |N|$  aller à l'étape 5, sinon aller à l'étape 2

**Etape 5.** Pour chaque cluster, déterminer la solution qui possède la plus petite distance moyenne avec toutes les autres solutions du même cluster, cette solution est appelée le centre de gravité. Conserver les centres de gravité associés aux clusters dans la liste  $P^{**}$  et enlever les autres solutions de  $P^{**}$ .

La dernière question à régler est celle de la sélection des solutions de  $P^{**}$  qui doivent être réintroduites dans la population. Une façon simple de procéder est de réintroduire dans la population  $P_t$  un nombre fixe de solution de  $P^{**}$ , éventuellement sélectionnées aléatoirement.

### A.2.4.5 Pseudo code et couplage avec principales métaheuristiques

Maintenant que nous avons décrit les principales stratégies utilisées dans les MOEA, nous en donnons le pseudo code dans le tableau A.2.

---

Initialiser $t = 0$ , générer la population $P_0$ et initialiser l'archive $P^{**}$
Evaluer les objectifs $f_k(x)$ pour tout $k = 1..K$ et tout $x \in P_0$
Classement Pareto (affectation des rangs de dominance)
Calcul des probabilités de sélection (dist. de surpeuplement ou partage de Fitness)
Tant que condition d'arrêt non vérifiée (nombre de générations ou temps de calcul)
Sélectionner des individus de $P_t$
Appliquer les opérateurs évolutionnaires sur ces individus
Obtention des descendants qui forment $P_t^{\text{child}}$
Evaluation des objectifs $f_k$ pour tout $x \in P_t^{\text{child}}$
Classement Pareto sur $P_t \cup P_t^{\text{child}}$
Calcul des probabilités de sélection pour tout $x \in P_t \cup P_t^{\text{child}}$
Constituer $P_{t+1}$ (réduction de $P_t \cup P_t^{\text{child}}$ et sélection dans $P^{**}$ )
Archiver $P_t \cup P_t^{\text{child}}$ dans $P^{**}$
Poser $t := t + 1$
Fin boucle tant que

---

Tableau A.2 – Pseudo code MOEA générique.

Utilisé à plusieurs reprises dans la littérature relative aux modèles d'arbitrage durée, coût et qualité, l'algorithme NSGAII se fonde sur le pseudo code donné dans le tableau A.2 où le calcul des probabilités de sélection s'appuie sur celui des distances  $\delta$  de surpeuplement donné par l'équation (A.14) avec  $\delta(x) = \sum_{k=1..K} \delta_k(x_{i,k})$ . De plus, la stratégie élitiste décrite dans la sous-section A.2.4.4 est employée pour former  $P_{t+1}$ . En revanche le NSGAII n'archive pas dans une liste  $P^{**}$  les meilleures solutions non dominées trouvées au cours des générations.

D'autres auteurs ont développé des MOEA pour résoudre le DTCQTP en combinant les métaheuristiques ACO, ICA ou ABC avec les différentes stratégies en MO que nous venons de discuter à savoir classement Pareto, préservation de la diversité et élitisme. De tels algorithmes sont utilisés dans le chapitre 4 où nous les détaillons davantage.

# Annexe B

## Résumé du programme d'optimisation

### Paramètres et variables

#### Paramètres à entrer ou à estimer

$V = \{0, 1, \dots, n, n + 1\}$  Ensemble des activités du projet où les activités 0 et  $n + 1$  sont les activités fictives de début et de fin du projet

$\Gamma^{-1}(j)$  Ensemble des prédécesseurs directs de l'activité  $j$

$\Gamma(j)$  Ensemble de ses successeurs directs

$V_k$  Sous-ensemble des activités qui nécessitent la spécialisation professionnelle  $k$ , avec  $k = 1..K$  où  $K$  désigne le nombre total de spécialisations professionnelles requises pour le projet

$P_{kj}$  Nombre de permanents de spécialisation  $k$  alloués à l'exécution de l'activité  $j$  en mode normal

$A_j$  Nombre maximum de travailleurs qu'il est possible d'affecter à l'activité  $j$  en raison de la capacité du poste

$\gamma_j$  Coefficient d'impact du sureffectif sur la productivité

$\lambda$  Proportion maximum de temporaires souhaitée dans l'effectif total alloué à toute activité et à toute période

$d_j^{\max}$  Durée normale d'exécution de l'activité  $j$  avec  $d_0^{\max} = d_{n+1}^{\max} = 0$

$q_j^{\max}$  Qualité en mode normal (en général égale à 1)

$\Delta q_j^S$  Perte de qualité résultant de l'emploi d'un travailleur en sureffectif par jour

$\Delta q_j^O$  Perte de qualité associée à l'utilisation d'une heure supplémentaire

$V^\theta$  Sous-ensemble des activités ayant même niveau de tolérance  $\theta$ , avec  $\theta =$

1.. $\Theta$  $q_\theta^{\min}$  Seuil de qualité minimum pour un niveau  $\theta$  de tolérance aux défauts $H_R$  Nombre d'heures de travail régulier par jour $H_O$  Nombre maximum d'heures supplémentaires par travailleur permanent et par jour $H_{OW}$  Nombre maximum d'heures supplémentaires par travailleur permanent et par semaine $c_P$  Coût d'une heure de travail permanent $c_U$  Coût d'une heure de travail temporaire $c_O$  Coût d'une heure supplémentaire

### Paramètres à calculer

 $P_k$  Nombre total de permanents de spécialisation  $k$  affectés au projet $\underline{d}_j$  Borne inférieure sur la durée d'exécution de l'activité  $j$  $ES_j$  Date de début au plus tôt de l'activité  $j$  calculée sur la base des bornes inférieures de durées  $\{\underline{d}_j\}_{j=1..n}$  $LF_j$  Date de fin au plus tard déterminée à partir des durées normales  $\{d_j^{\max}\}_{j=1..n}$  $T$  Longueur de l'horizon définie par la durée du projet en mode normal c'est-à-dire sur la base des durées  $d_j^{\max}$  $\underline{T}$  Borne inférieure sur la durée totale d'exécution du projet, calculée à partir des bornes inférieures des durées de chaque activité  $\underline{d}_j$  $\mathcal{T}_j$  Ensemble des périodes réalisables d'exécution de l'activité  $j$ ;  $\mathcal{T}_j = \{ES_j..LF_j\}$  $\underline{w}_j, \bar{w}_j$  Bornes inférieures et supérieures sur le nombre de travailleurs alloués à l'exécution de l'activité  $j$  à chaque période d'exécution $\bar{p}_j, \bar{u}_j$  Bornes supérieures sur la main d'oeuvre permanente et temporaire (respectivement) qui peut être allouée à l'exécution de l'activité  $j$  à toute période

### Autres notations

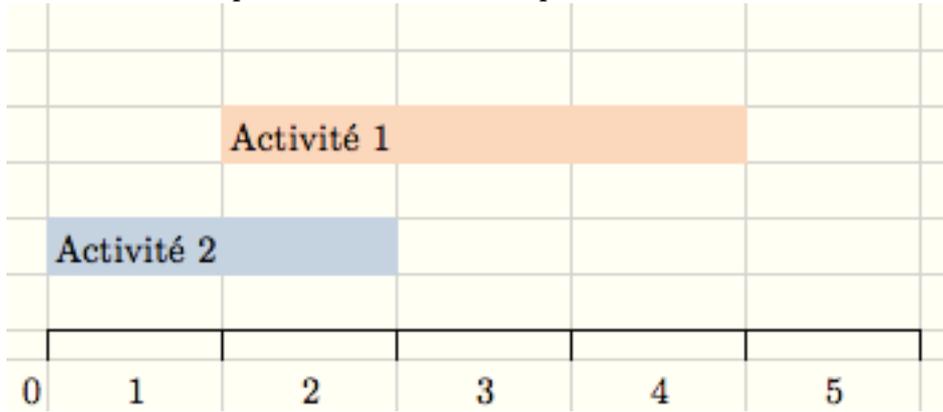
 $\mu = \sum_{j=1}^{n-1} (LF_j - ES_j)$  Notation permettant de simplifier le comptage du nombre de contraintes $V_{kt}$  Ensemble des activités nécessitant la spécialisation  $k$  et exécutables en  $t$ , avec  $V_{kt} = \{j \in V_k \mid t \in \mathcal{T}_j\}$  $\alpha_k$  Notation utilisée pour faciliter l'écriture de certaines contraintes, avec  $\alpha_k = \min_{j \in V_k} ES_j$

$\beta_k$  Notation utilisée pour faciliter l'écriture de certaines contraintes, avec  $\beta_k = \max_{j \in V_k} LF_j$

## Variables et détermination de leur nombre

$z_{jt}$  Variable binaire qui prend la valeur 1 si l'activité  $j$  est en cours d'exécution à la période  $t$  et zéro sinon, pour tout  $t \in \mathcal{T}_j$  et  $j = 1..n$ . Il y a  $\mu$  variables binaires de ce type (BIN)

La figure ci-après illustre ces variables pour deux activités dont les intervalles d'exécution sont respectivement de 2 et 3 périodes.



Pour l'activité 1, on a  $z_{1,1} = 0$ ;  $z_{1,2} = z_{1,3} = z_{1,4} = 1$ ;  $z_{1,5} = 0$

Pour l'activité 2, on a  $z_{2,1} = z_{2,2} = 1$ ;  $z_{2,3} = z_{2,4} = z_{2,5} = 0$

$p_{jt}$  Nombre de travailleurs permanents de la spécialisation  $k$  alloués à l'exécution de l'activité  $j$  en  $t$  avec  $t \in \mathcal{T}_j$ . Il y a  $\mu$  variables entières de ce type (INT)

$u_{jt}$  Nombre de travailleurs temporaires (de spécialisation  $k$ ) alloués à l'exécution de l'activité  $j$  en  $t$ . Il y a  $\mu$  variables entières de ce type (INT)

$h_{jt}$  Nombre d'heures supplémentaires (de spécialisation  $k$ ) alloués à l'exécution de l'activité  $j$  en  $t$ . Il y a  $\mu$  variables continues de ce type (REAL)

$y_{jtl}$  Variable binaire qui vaut 1 si  $p_{jt} + u_{jt}$  prend la valeur  $l \in \{0, 1, 2.. \bar{w}_j\}$ . Il y a  $M = \sum_{j=1}^{n-2} (\bar{w}_j + 1)(LF_j - ES_j)$  variables binaires de ce type (BIN)

$C_{max}$  Durée totale d'exécution du projet, 1 variable entière (INT)

Au total, le nombre de variables est égal à  $4\mu + M + 1$ .

## Pré-traitement et bornes

**Nombre total de permanents affectés au projet.** Pour toute spécialisation professionnelle  $k$ , le nombre de permanents affectés au projet est tel qu'il permet son exécution en mode normal, en prenant un ordonnancement au plus tôt ou au plus tard

$$P_k = \max_{t=1..T} \left( \sum_{j \in V_k | t \in \{\overline{ES}_j.. \overline{ES}_j + d_j^{\max}\}} P_{kj} \right) \quad \forall k = 1..K.$$

**Effectif maximum**

$$s_j = \min\left(\left\lceil \frac{1+\gamma}{2\gamma} P_{kj} \right\rceil, d_j^{\max} P_{kj}, A_j\right) \quad \forall j = 1..n.$$

**Borne sur la main d'oeuvre permanente**

$$p_{jt} \leq \min(s_j, P_k) = \bar{p}_j \quad \forall j = 1..n \quad t \in \mathcal{T}_j.$$

**Borne sur la main d'oeuvre temporaire**

$$u_{jt} \leq \bar{u}_j$$

avec

$$\bar{u}_j = \begin{cases} 0 & \text{si } \lambda = 0 \\ \min(s_j, \lfloor \frac{\lambda}{1-\lambda} \bar{p}_j \rfloor) & \text{si } \lambda \in ]0, 1[ \\ s_j & \text{si } \lambda = 1 \end{cases} \quad \forall j = 1..n \quad t \in \mathcal{T}_j.$$

**Borne supérieure sur les heures supplémentaires.** L'hypothèse selon laquelle les travailleurs en heures supplémentaires ne sont jamais en sureffectif se traduit par la borne supérieure suivante

$$h_{jt} \leq H_O P_{kj} \quad \forall j = 1..n \quad t \in \mathcal{T}_j.$$

**Borne sur la main d'oeuvre régulière**

$$\bar{w}_j = \min(\bar{p}_j + \bar{u}_j, s_j) \quad \forall j = 1..n \quad t \in \mathcal{T}_j.$$

**Borne sur les durées individuelles des activités**

$$\underline{d}_j = \left\lceil \frac{W_j}{\bar{w}_j + \frac{H_O}{H_R} P_{kj}} \right\rceil$$

**Borne supérieure sur la durée totale d'exécution  $T$ .** On commence par déterminer les dates de début au plus tôt sur la base des durées normales d'exécution. On note  $\overline{ES}_j$  ces dates avec  $\overline{ES}_j = \max_{i \in \Gamma^{-1}(j)} \{\overline{ES}_i + d_j^{\max}\}$  et  $\overline{ES}_0 = 0$ . Puis on détermine les dates de début au plus tard, toujours sur la base des durées normales, soit  $\overline{LS}_j = \min_{i \in \Gamma(j)} \{\overline{LS}_i\} - d_j^{\max}$  avec  $\overline{LS}_{n+1} = \overline{ES}_{n+1}$ . Enfin, on obtient  $LF_j = \overline{LS}_j + d_j^{\max}$ . La borne supérieure sur la durée totale d'exécution du projet correspond simplement à la durée totale d'exécution en mode normal, c'est-à-dire  $T = LF_{n+1}$ .

**Borne inférieure sur la durée totale d'exécution  $\underline{T}$ .** Les dates possibles au plus tôt de début d'exécution,  $ES_j$ , sont simplement données par

$$ES_j = \max_{i \in \Gamma^{-1}(j)} \{ES_i + \underline{d}_i\}, \text{ avec } ES_0 = 0.$$

On obtient ainsi  $\mathcal{T}_j$ , l'ensemble des périodes possibles d'exécution de l'activité  $j$  avec  $\mathcal{T}_j = \{ES_j..LF_j\}$ .

On recalcule ensuite les dates de début au plus tard et de fin au plus tard sur la base des durées minimum  $\underline{d}_j$  pour obtenir  $\underline{T}$ .

## Objectif, contraintes et nombre de contraintes

Définition de l'objectif de minimisation de la durée du projet sous forme des contraintes suivantes

$$t \cdot z_{jt} - C_{max} \leq 0 \quad \forall j = 1..n \quad t \in \mathcal{T}_j. \quad (\text{B.1})$$

Nombre de contraintes de ce type :  $\mu$ .

### Contraintes de précédence

$$d_j^{\max} \cdot z_{it} + \sum_{\tau=ES_j}^t z_{j\tau} \leq d_j^{\max} \quad \forall j = 1..n \quad i \in \Gamma^{-1}(j) \mid LF_i \geq ES_j \quad (\text{B.2})$$

avec  $t = \max(ES_i, ES_j) .. \min(LF_i, LF_j) - 1$ .

Nombre de contraintes de précédence :

$$\sum_{j=1}^n \sum_{i \in \Gamma^{-1}(j) \mid LF_i \geq ES_j} \{ \min(LF_i, LF_j) - \max(ES_i, ES_j) \}.$$

**Contraintes de non préemption.** Elles s'appliquent à toute activité telle que  $LF_j - ES_j \geq 3$  et d'une durée normale  $d_j^{\max} \geq 2$ .

$$\sum_{\tau=t+1}^{LF_j} z_{j\tau} + d_j^{\max} (z_{jt} - z_{jt+1}) \leq d_j^{\max} \quad j = 1..n \quad t = ES_j .. LF_j - 3. \quad (\text{B.3})$$

Nombre de contraintes de non préemption :

$$\sum \{ j \mid d_j^{\max} \geq 2; LF_j - ES_j \geq 3 \} \{ LF_j - ES_j - 2 \}$$

**Proportion maximum  $\lambda$  de temporaires (activée si  $\lambda \neq 1$  et  $\lambda \neq 0$ )**

$$\frac{\lambda}{1-\lambda} p_{jt} - u_{jt} \geq 0 \quad \forall j = 1..n \quad t \in \mathcal{T}_j. \quad (\text{B.4})$$

Nombre de contraintes :  $\mu$

**Respect de la quantité de main d'oeuvre permanente disponible à chaque période pour chaque type de spécialisation (pas de contrainte sur les temporaires dont le recrutement n'est pas limité)**

$$\sum_{j \in V_{kt}} p_{jt} \leq P_k \quad \forall k = 1..K \quad t = \alpha_k .. \beta_k. \quad (\text{B.5})$$

Nombre de contraintes :  $\sum_{k=1..K} (\beta_k - \alpha_k)$

**Pas plus de travailleurs en heures supplémentaires que de travailleurs en heures régulières**

$$p_{jt} + u_{jt} - \frac{1}{H_O} h_{jt} \geq 0 \quad \forall j = 1..n \quad t \in \mathcal{T}_j. \quad (\text{B.6})$$

Nombre de contraintes :  $\mu$

**Nombre suffisant de permanents pour assurer les heures supplémentaires**

$$\sum_{j \in V_{kt}} p_{jt} - \frac{1}{H_O} \sum_{j \in V_{kt}} h_{jt} \geq 0 \quad \forall k = 1..K \quad t = \alpha_k.. \beta_k \quad (\text{B.7})$$

Nombre de contraintes :  $\sum_{k=1..K} \sum_{t=\alpha_k.. \beta_k} |V_{kt}|$

**Contrainte de non dépassement d'un certain nombre d'heures supplémentaires par semaine**

$$H_{OW} \frac{1}{5} \sum_{\tau=t}^{t+4} \sum_{j \in V_{k\tau}} p_{j\tau} - \sum_{\tau=t}^{t+4} \sum_{j \in V_{k\tau}} h_{j\tau} \geq 0 \quad \forall k = 1..K \quad t = \alpha_k.. \beta_k - 4. \quad (\text{B.8})$$

Nombre de contraintes :  $\beta_k - \alpha_k - 4$  pour toute spécialisation  $k$  telle que  $\beta_k - \alpha_k \geq 5$

Pour toute spécialisation  $k$  telle que  $\beta_k - \alpha_k < 5$  alors il n'y a qu'une seule contrainte qui s'écrit

$$H_{OW} \frac{\beta_k - \alpha_k}{5} \sum_{\tau=\alpha_k}^{\beta_k-1} \sum_{j \in V_{k\tau}} p_{jt} - \sum_{\tau=\alpha_k}^{\beta_k-1} \sum_{j \in V_{k\tau}} h_{j\tau} \geq 0 \quad (\text{B.9})$$

Nombre de contraintes : une seule pour toute spécialisation  $k$  telle que  $\beta_k - \alpha_k < 7$ .

**Respect de la quantité de main d'oeuvre disponible en heures supplémentaires à chaque période pour chaque type de spécialisation**

$$\sum_{j \in V_{kt}} h_{jt} \leq H_O P_k \quad \forall k = 1..K - 1 \quad t = \alpha_k.. \beta_k. \quad (\text{B.10})$$

Nombre de contraintes :  $\sum_{k=1..K} \beta_k - \alpha_k$

**Linéarisation**  $p_{jt} + u_{jt}$

$$p_{jt} + u_{jt} - \sum_{l=0}^{\bar{w}_j} l \cdot y_{jtl} = 0 \quad \forall j = 1..n \quad t \in \mathcal{T}_j. \quad (\text{B.11})$$

Nombre de contraintes :  $\mu$

$$\sum_{l=0}^{\bar{w}_j} y_{jtl} = 1 \quad \forall j = 1..n \quad t \in \mathcal{T}_j. \quad (\text{B.12})$$

Nombre de contraintes :  $\mu$

**Satisfaction des besoins en main d'oeuvre pour l'exécution de toute activité**

$$\sum_{t \in \mathcal{T}_j} \left( \frac{1}{H_R} h_{jt} + \sum_{l=0}^{\bar{w}_j} w_j^R(l) \cdot y_{jtl} \right) - d_j^{\max} \cdot P_{kj} = 0 \quad \forall j = 1..n. \quad (\text{B.13})$$

avec

$$w_j^R(l) = \begin{cases} (1 + \gamma)l - \frac{\gamma}{P_{kj}} l^2 & \text{si } l > P_{kj} \\ l & \text{sinon} \end{cases} \quad \forall l = 0..\bar{w}_j \quad j = 1..n.$$

Nombre de contraintes :  $n - 2$

**Bornes supérieures et inférieures sur la consommation de main d'oeuvre pour une activité en cours d'exécution**

$$-\bar{w}_j z_{jt} + p_{jt} + u_{jt} \leq 0 \quad \forall j = 1..n \quad t \in \mathcal{T}_j. \quad (\text{B.14})$$

Nombre de contraintes :  $\mu$

$$-P_{kj} z_{jt} + p_{jt} + u_{jt} \geq 0 \quad \forall j = 1..n \quad t \in \mathcal{T}_j. \quad (\text{B.15})$$

Nombre de contraintes :  $\mu$

Avec

$$\bar{w}_j = \min(\bar{p}_j + \bar{u}_j, s_j)$$

**Contraintes sur la qualité**

$$\sum_{t \in \mathcal{T}_j} \left( \sum_{l=0}^{\bar{w}_j} \Delta q_j^S(l) \cdot y_{jtl} + \Delta q_j^O \cdot h_{jt} \right) \leq q_j^{\max} - q_\theta^{\min} \quad \forall j \in V^\theta, \quad \theta = 1..\Theta. \quad (\text{B.16})$$

Nombre de contraintes :  $n - 2$ . Avec le calcul préalable suivant

$$\Delta q_j^S(l) = \begin{cases} \Delta q_j^S \cdot (l - P_{kj}) & \text{si } l > P_{kj} \\ 0 & \text{sinon} \end{cases} \quad \forall l = 0..\bar{w}_j \quad j = 1..n. \quad (\text{B.17})$$

**Une contrainte de budget**

$$\sum_{j=1}^n \sum_{t \in \mathcal{T}_j} (c_P \cdot H_R \cdot p_{jt} + c_U \cdot H_R \cdot u_{jt} + c_O \cdot h_{jt}) \leq (1 + \rho) \cdot c_P \cdot H_R \cdot \sum_{j=1}^n d_j^{\max} \cdot P_{kj}. \quad (\text{B.18})$$

# Annexe C

## Glossaire des notations

### Chapitre 1

- $j = 1..n$     Activité  $j$  où  $n$  est le nombre total d'activités
- $d_j, c_j$     Durée et coût d'une activité  $j$
- $d_j^{\max}$     Durée normale d'exécution de l'activité  $j$
- $d_j^{\min}$     Durée compressée d'exécution de l'activité  $j$
- $\Delta d_j$     Réduction de durée de l'activité  $j$  résultant d'une augmentation  $\Delta r$  de la ressource  $r$
- $W_j$     Contenu en travail de l'activité  $j$
- $P_j$     Nombre total de travailleurs permanents alloué à l'exécution de l'activité  $j$
- $H_R$     Nombre d'heures de travail régulier par jour
- $d_j = \mathcal{D}_j(r)$     Durée d'exécution de l'activité  $j$  comme une fonction continue  $\mathcal{D}_j$  de la ressource additionnelle  $r$
- $d\mathcal{D}_j/dr$     Productivité marginale de la ressource  $r$ , c'est la pente de la tangente à la fonction  $\mathcal{D}_j$  au point  $r$
- $\gamma$     Coefficient d'impact des heures de travail temporaire sur la productivité
- $c_r$     Coût d'une heure de travail
- $c_j = \mathcal{C}_j(d_j)$     Coût de l'activité  $j$  comme une fonction  $\mathcal{C}_j$  de sa durée d'exécution  $\mathcal{C}_j(d_j)$

## Chapitres 3 et 4

$V = \{0, 1, \dots, n, n + 1\}$  Ensemble des activités du projet où les activités 0 et  $n + 1$  sont les activités fictives de début et de fin du projet.

$\Gamma^{-1}(j)$  Ensemble des prédécesseurs directs de l'activité  $j$

$q_j$  Qualité d'une activité  $j$

$M_j$  Ensemble des modes d'exécution de la tâche  $j$

$c_{jm}, d_{jm}, q_{jm}$  Coût, durée et qualité de la tâche  $j$  dans le mode  $m = 1..M_j$

$r_{jm}$  Quantité d'une ressource unique utilisé par l'activité  $j$  dans le mode

$EF_j$  et  $LF_j$  Date de fin au plus tôt et au plus tard de l'activité  $j$

$D_{\max}$  Seuil de durée totale à ne pas dépasser pour contrainte de délai

$B_{\max}$  Seuil de budget max à ne pas dépasser pour contrainte de budget

$x_{jmt}$  Variable binaire qui vaut 1 si l'activité  $j$  exécutée dans le mode  $m$  s'achève en  $t$

$r_{jmk}$  et  $r'_{jml}$  Quantité de ressource  $k$  renouvelable ( $l$  non renouvelable) utilisée par  $j$  dans le mode  $m$

$R_k$  et  $R'_l$  Capacité sur ressources  $k$  renouvelable et  $l$  non renouvelable

$T$  Horizon avec les durées les plus longues

$M$  Une liste de modes d'exécution

$S$  Une liste d'activités à ordonnancer dans l'ordre de la liste

$W_{jk}$  Contenu en travail de la ressource  $k$  affecté à la tâche  $j$

$d_j^{\max}$  et  $d_j^{\min}$  Durées individuelles de l'activité  $j$ , normales et compressées (même type de notation pour coût et qualité normale et compressé)

$d_{jm} = \mathcal{D}_j(r_{jm})$  Durée d'exécution l'activité  $j$  dans le mode  $m$  où une quantité  $r_{jm}$  de ressource  $r$  est utilisée

## Chapitre 4

$q_j = \mathcal{Q}_j(d_j)$  Qualité l'activité  $j$  comme une fonction  $\mathcal{Q}_j$  de la durée

$Q_{\min}$  Seuil de qualité globale minimum à respecter

$Q$  Qualité globale du projet (agrégation des qualités individuelles)

# Annexe D

## Liste des abréviations

AMDEC	Analyse des Modes de Défaillance, de leurs Effets et de leur Criticité
AOA	Activity On Arc
AON	Activity On Node
APSYS	Alstom Production System
CoNQ	Cost of Non Quality
CoQ	Cost of Quality
CPM	Critical Path Method
CTCTP	Continuous Time-Cost Trade-off Problem
DFSS	Design For Six Sigma
DMAIC	Define, Measure, Analyse, Improve, Control
DTCQTP	Discret Time Cost Quality Tradeoff problem
DTCTP	Discrete Time-Cost Trade-off Problem
DTRTP	Discrete Time-Resource Trade-off Problem
MSRCPSP	Multi-Skill Resource Constrained Project Scheduling problem
NPV	Net Present Value
P-A-F	Prevention-Appraisal-Failure
TCQTP	Time Cost Quality Trade off Problem
TCTP	Time Cost Trade off Problem
TPS	Toyota Production System
TQC	Total Quality Control
TQM	Total Quality Management
TRTP	Time Resource Trade off Problem

# Bibliographie

- [1] ISO 9000 : Systèmes de management de la qualité, principes essentiels et vocabulaire.
- [2] E. N. AFRUZI, A.A. NAJAFI, E. ROGHANIAN et M. MAZINANI : A multi-objective imperialist competitive algorithm for solving discrete time, cost and quality trade-off problems with mode-identity and resource-constrained situations. *Computers and Operations Research*, 50:80–96, 2014.
- [3] E. N. AFRUZI, E. ROGHANIAN, A.A. NAJAFI et M. MAZINANI : A multi-mode resource-constrained discrete time-cost tradeoff problem solving using an adjusted fuzzy dominance genetic algorithm. *Scientia Iranica, Transactions E : Industrial Engineering*, 20:931–944, 2013.
- [4] A. AFSHAR, A. KAVEH et O.R. SHOGLI : Multi-objective optimisation of time-cost-quality using multi-colony ant algorithm. *Asian Journal Of Civil Engineering (building and housing)*, 8:113–124, 2007.
- [5] S. AGARWAL, J.J. GALLO, A. PARASHAR, K. K. AGARWAL, S.G. ELLIS, U. N. KHOT, R. SPOONER, E. M. TUZCU et S. R. KAPADIA : Impact of lean six sigma process improvement methodology on cardiac catheterization laboratory efficiency. *Cardiovascular Revascularization Medicine*, 2015.
- [6] M.K. AGRAWAL, S.E. ELMAGHRABY et W. HERROELEN : Dagen : a generator of test sets for project activity nets. *European Journal of Operational Research*, 90:376–382, 1996.
- [7] C. AKKAN, A. DREXL et A. KIMMS : Network decomposition-based benchmark results for the discrete time-cost tradeoff problem. *European Journal of Operational Research*, 165(2):339 – 358, 2005. Project Management and Scheduling.
- [8] E. ALBA et F. CHICANO : Software project management with gas. *Information Sciences*, pages 2380–2401, 2007.
- [9] J. ALCARAZ, C. MAROTO et R. RUIZ : Solving the multi-mode resource-constrained project scheduling problem with genetic algorithms. *Journal of the Operational Research Society*, 54(6):614–628, 2003.
- [10] T. ALDOWAISAN, M. NOURELFATH et J. HASSAN : Six sigma performance for non-normal processes. *European Journal of Operational Research*, 247(3):968–977, 2015.
- [11] H. ALFARES et J. BAILEY : Integrated project task and manpower scheduling. *IIE Transactions*, 29(9):711–717, 1997.

- [12] C. ARTIGUES et S. DEMASSY : Gestion de production et ressources humaines : Méthodes de planification dans les systèmes productifs, 2005.
- [13] R. ATKINSON : Project management : cost, time and quality, two best guesses and a phenomenon, its time to accept other success criteria. *International Journal of Project Management*, 17:337–342, 1999.
- [14] A.J.G. BABU et N. SURESH : Project management with time, cost, and quality considerations. *European Journal of Operational Research*, 88:320–327, 1996.
- [15] G. BASCOUL et J.M. MOUTOT : *Marketing et développement durable*. DUNOD, 2009.
- [16] M. BASSET : Assigning projects to optimize the utilization of employees time and expertise. *Computers and Chemical Engineering*, 24(2):1013–1021, 2000.
- [17] R. BASU : Managing quality in projects an empirical study. *International Journal of Project Management*, 32:178–187, 2014.
- [18] W.W. BEIN, J. KAMBUROWSKI et M.F.M. STALLMAN : Optimal reduction of two-terminal directed acyclic graphs. *SIAM Journal on Computing*, 21:1112–1129, 1992.
- [19] O. BELLENGUEZ-MORINEAU et E. NERON : A branch-and-bound method for solving multi-skill project scheduling problem. *RAIRO Operations Research*, 41(2):155–170, 2007.
- [20] J. BŁAZEWICZ, K. ECKER, E. PESCH, G. SCHMIDT et J. WEGLARZ : *Handbook on Scheduling : From Theory to Applications*. Springer Verlag, Berlin, 2007.
- [21] F. F. BOCTOR : Some efficient multi-heuristic procedures for resource-constrained project scheduling. *European Journal of Operational Research*, 90(1):3–13, 1990.
- [22] F.F. BOCTOR : Heuristics for scheduling projects with resource restrictions and several resource-duration modes. *International Journal of Production Research*, 31(11):2547–2558, 1993.
- [23] F.F. BOCTOR : A new and efficient heuristic for scheduling projects with resource constraints and multiple execution modes. *Proceedings of the 4th international workshop on project management and scheduling*, pages 63–66, 1994.
- [24] F.F. BOCTOR : Resource-constrained project scheduling by simulated annealing. *International Journal of Production Research*, 34, 1996.
- [25] T. BORTOLOTTI, P. ROMANO, P.J. MARTINEZ-JURADO et J. MOYANO-FUENTES : Towards a theory for lean implementation in supply networks. *International Journal of Production Economics*, 175:182–196, 2016.
- [26] K. BOULEIMEN et H. LECOCQ : A new efficient simulated annealing algorithm for the resource-constrained project scheduling problem and its multiple mode version. *European Journal of Operational Research*, 149(2):268 – 281, 2003. Sequencing and Scheduling.
- [27] J.P. BROSHAK et A. DAVIS-BLAKE : Mixing standard work and nonstandard deals : the consequences of heterogeneity in employment arrangements. *Academy of Management Journal*, 49(2):371–393, 2006.

- [28] A. BRUN : Critical success factors of six sigma implementations in italian companies. *International Journal of Production Economics*, 131:158–164, 2011.
- [29] A.A. BUBSHAIT : Incentive disincentive contracts and its effects on industrial projects. *International Journal of Project Management*, 21(63-70), 2003.
- [30] W-N. CHEN, J. ZHANG, H. S-H. CHUNG, R-Z HUANG et O. LIU : Optimizing discounted cash flows in project scheduling-an ant colony optimization approach. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 40(1):64–77, 2010.
- [31] J. CHENG, J. FOWLER, K. KEMPF et S. MASON : Multi-mode resource-constrained project scheduling problems with non-preemptive activity splitting. *Computers and Operations Research*, 53:275 – 287, 2015.
- [32] A. CHIARINI : *From Total Quality Control to Lean Six Sigma*. Springer, Briefs in Business, 2012.
- [33] C. COCHET et J.D. MERLET : *Réglementation, normalisation et innovation*. Annales des ponts et chaussées, 2000.
- [34] P. COHENDET, M. DIANI et C. LERCH : Stratégie modulaire dans la conception. une interprétation en termes de communautés. *Revue française de gestion*, 5(158):121–143, 2005.
- [35] I. CORREIA, L.L. LOURENCO et F. Saldanha da GAMA : Project scheduling with flexible resources : Formulation and inequalities. *Operational Research Spectrum*, 34(3):635–663, 2012.
- [36] V. CRACAN : Amélioration de la performance industrielle. master qualité et performance dans les organisations, université de technologie de compiègne. Mémoire de D.E.A., Université de Technologie de Compiègne, 2013.
- [37] N. CURATOLO, S. LAMOURI, J.-C. HUET et A. RIEUTORD : Démarches d’amélioration en milieu hospitalier : du management de la qualité totale au lean. *Annales Pharmaceutiques Françaises*, 73(4):245 – 256, 2015.
- [38] N. De CUYPER, J. De JONG, H. De WITTE, K. ISAKSSON, T. RIGOTTI et R. SCHALK : Literature review of theory and research on the psychological impact of temporary employment : towards a conceptional model. *International Journal of Management Reviews*, 10(1):25–51, 2008.
- [39] N. DAMAK, B. JARBOUI, P. SIARRY et T. LOUKIL : Differential evolution for solving multi-mode resource-constrained project scheduling problems. *Computers and Operations Research*, 36(9):2653 – 2659, 2009.
- [40] A. DAS, M. PAGELL, M. BEHM et A. VELTRI : Toward a theory of the linkages between safety and quality. *Journal of Operations Management*, 26:521–535, 2008.
- [41] P. DE, E. J. DUNNE, J.B. GHOSH et C.E. WELLS : The discrete time-cost tradeoff problem revisited. *European Journal of Operational Research*, 81:225–238, 1995.
- [42] E. DEMEULEMEESTER, B. DEREYCK et W. HERROELEN : The discrete time/resource trade-off problem in project networks : A branch and bound approach. *IIE Transactions*, 32:1059–1069, 2000.

- [43] E. DEMEULEMEESTER, W. HERROELEN et S. ELMAGHRABY : *Optimal Procedures for the Discrete Time-Cost Trade-off Problem in Project Networks*. Onderzoeksrapport Nr9337, 1993.
- [44] E. DEMEULEMEESTER, B. De REYCK, B. FOUBERT, W. HERROELEN et M. VANHOUCKE : New computational results for the discrete time/cost trade-off problem in project networks. *Journal of the Operational Research Society*, 49:1153–1163, 1998.
- [45] E.L. DEMEULEMEESTER et W.S. HERROELEN : *Project scheduling : a research handbook*. International Series in Operations Research and Management Science, 2002.
- [46] E.L. DEMEULEMEESTER, W.S. HERROELEN et S.E. ELMAGHRABY : Optimal procedures for the discrete time/cost trade-off problem in project networks. *European Journal of Operational Research*, 88:50–68, 1996.
- [47] A. DREXL : Scheduling of project networks by job assignment. *Management Science*, 37(12):1590–1602, 1991.
- [48] A. DREXL et J. GRÜNEWALD : Nonpreemptive multi-mode resource-constrained project scheduling. *IIE Transactions*, 25(5):74–81, 1993.
- [49] K. EL-RAYES et A.KANDIL : Time-cost-quality trade-off analysis for highway construction. *Journal of Construction Engineering and Management*, 131:477–486, 2005.
- [50] S.E. ELMAGHRABY : On the expected duration of pert type networks. *Management Science*, 1967.
- [51] S.S. ERENGUC, S. TUFEKCI et C.J. ZAPPE : Solving time-cost trade-off problems with discounted cash flows using generalized benders decomposition. *Naval Research Logistics*, 40(1):25–50, 1993.
- [52] K. El-Rayes et O. MOSELHI : Resource-driven scheduling of repetitive activities. *Construction Management and Economics*, 16:433–446, 1998.
- [53] JE. FALK et JL. HOROWITZ : Critical path problems with concave cost-time curves. *Management Science*, 19(4):446, 455 1972.
- [54] A.V. FEIGENBAUM : *Quality control : principles, practice and administration*. McGraw-Hill, New York, 1951.
- [55] A.V. FEIGENBAUM : Total quality control. *Harvard Business Review*, 1956.
- [56] C. FENG, L. LIU et S.A. BURNS : Using genetic algorithms to solve construction time-cost trade-off problems. *Journal of Computing in Civil Engineering*, 11(3):184–189, 1997.
- [57] B. FLYVBJERG : Policy and planning for large infrastructure projects. *World Bank Policy Research Working*, 2007.
- [58] S. FOLDES et F. SOUMIS : Pert and crashing revisited : Mathematical generalizations. *European Journal of Operational Research*, 64:286–294, 1993.
- [59] V. FRIGANT et J. BERNARD : Comment la production modulaire transforme l'industrie automobile. *Revue d'économie industrielle*, 1(145):11–44, 2014.

- [60] F. FU et T. ZHANG : A new model for solving time-cost-quality trade-off problems in construction. *School of Economics and Management of China University of Petroleum (East China), Qingdao, Shandong, PR China*, 2016.
- [61] D.R. FULKERSON : A network flow computation for project cost curves. *Management Science*, 7:167–178, 1961.
- [62] C.U. FUNDELING et N. TRAUTMANN : A priority-rule method for project scheduling with work-content constraints. *European Journal of Operational Research*, 203:568–574, 2010.
- [63] P. GHODDOUSI, E. ESHTEHARDIAN, S. JOOYBANPOUR et A. JAVANMARDI : Multi-mode resource-constrained discrete time-cost-resource optimization in project scheduling using non-dominated sorting genetic algorithm. *Automation in Construction*, 30:216–227, 2013.
- [64] V. GIARD : *Gestion de projets*. Economica, Paris, 1991.
- [65] M. GUNDUZ : A quantitative approach for evaluation of negative impact of overmanning on electrical and mechanical projects. *Building and Environment*, 39(5):581 – 587, 2004.
- [66] E. HADJICONSTANTINOUCHE et E. KLERIDES : A new path-based cutting plane approach for the discrete time-cost trade-off problem. *Computational Management Science*, 2009.
- [67] A. S. HANNA et K. T. SULLIVAN : Impact of overtime on construction labor productivity. *Cost Engineering*, 46(4):20–27, 2004.
- [68] S. HARTMANN : Project scheduling with multiple modes : a genetic algorithm. *Annals of Operations Research*, 102(1-4):111–135, 2001.
- [69] S. HARTMANN et A. DREXL : Project scheduling with multiple modes : A comparison of exact algorithms. *Networks*, 32(283-297), 1998.
- [70] R.T. HARVEY et J.H PATTERSON : An implicit enumeration algorithm for the time-cost tradeoff problem in project network analysis. *Foundations of Control Engineering*, 4:107–117, 1979.
- [71] O. HAZIR, M. HAOUARI et E. EREL : Discrete time-cost trade-off problem : A decomposition-based solution algorithm for the budget version. *Computers and Operations Research*, 37(4):649–655, 2010.
- [72] C. HEIMERL et R. KOLISCH : Scheduling and staffing multiple projects with a multi-skilled workforce. *Operations Research*, 32:343–368, 2010.
- [73] P. HEISIG, P. J. CLARKSON et S. VAJNA : *Modelling and Management of Engineering Processes*. Springer, 2010.
- [74] F.S. HILLIER et G.J. LIEBERMAN : Introduction to mathematical programming. *McGraw-Hill, New York*, 1990.
- [75] T.J. HINDELANG et J.F. MUTH : A dynamic programming algorithm for decision cpm networks. *Operations Research*, 27:225–241, 1979.
- [76] B. HIRSCH et S. MUELLER : The productivity effect of temporary agency work : evidence from german panel data. *The Economic Journal*, 2012.

- [77] O. ICMELI et S.S. ERENGUC : The resource constrained time-cost trade-off project scheduling problem with discounted cash flow. *Journal of Operations Management*, 14(3):255–275, 1996.
- [78] O. ICMELI-TUKEL et W. O. ROM : Ensuring quality in resource constrained project scheduling. *European Journal of Operational Research*, 103:493–496, 1997.
- [79] ISO8402 : *Management de la qualité et assurance de la qualité - Vocabulaire*.
- [80] B. W. JACOBS, M. SWINK et K. LINDERMAN : Performance effects of early and late six sigma adoptions. *Journal of Operations Management*, 36:244–257, 2015.
- [81] B. JARBOUI, N. DAMAK, P. SIARRY et A. REBAI : A combinatorial particle swarm optimization for solving multi-mode resource-constrained project scheduling problems. *Applied Mathematics and Computation*, 195(1):299 – 308, 2008.
- [82] A. JOHN, R. MERAN, O. ROENPAGE et C. STAUDTER : *Six Sigma and Lean Toolset*. Springer, 2008.
- [83] B.P. JOHNSON et M.J. LIBERATORE : Incorporating quality considerations into project time cost tradeoff analysis and decision making. *IEEE Transactions on engineering management*, 53:534–542, 2006.
- [84] J. JOZEFOWSKA, M. MIKA, R. ROZYCKI, G. WALIGORA et J. WEGLARZ : Simulated annealing for multi-mode resource-constrained project scheduling. *Annals of Operations Research*, 102(1-4):137–155, 2001.
- [85] D. JUN et K. EL-RAYES : Optimizing labor utilization in multiple shifts for construction projects. *Construction Research Congress*, pages 1185–1193, 2010.
- [86] K.C. KAPUR : An algorithm for the project cost-duration analysis with quadratic and convex cost functions. *IIE Transactions*, 5(4):314–322, 1973.
- [87] J.E. KELLEY : Critical path planning and scheduling mathematical basis. *Operations Research*, 9:296–320, 1961.
- [88] D.B. KHANG et Y.M. MYINT : Time, cost and quality trade-off in project management : a case study. *International Journal of Project Management*, 17:249–256, 1999.
- [89] J.Y. KIM, C.W. KANG et I. HWANG : A practical approach to project scheduling : considering the potential quality loss cost in the time-cost tradeoff problem. *International Journal of Project Management*, 30:264–272, 2012.
- [90] R. KOLISCH : Project scheduling under resource constraints - efficient heuristics for several problem cases. *Physica-Verlag*, 1995.
- [91] R. KOLISCH et A. DREXL : Local search for nonpreemptive multi-mode resource- constrained project scheduling. *IIE Transactions*, 11(987-999), 1997.
- [92] R. KOLISCH et A. SPRECHER : Psplib : A project scheduling library. *European Journal of Operational Research*, pages 205–216, 1996.

- [93] R. KOLISCH, A. SPRECHER et A. DREXL : Characterization and generation of a general class of resource-constrained project scheduling problems. *Management Science*, 41:1693–1703, 1995.
- [94] M.L. KRAIMER, S.J. WAINE, R.C. LIDEN et R.T. SPARROWE : The role of job security in understanding the relation between employees perceptions of temporary workers and employees performance. *Journal of Applied Psychology*, 90(2):389–398, 2005.
- [95] L.R. LAMBERSON et R.R. HOCKING : Optimum time compression in project scheduling. *Management Science*, 16(10):597–606, 1970.
- [96] H. LI, P.E.D. LOVE et D.S. DREW : Effects of overtime work and additional resources on project cost and quality. *Engineering, Construction and Architectural Management*, 7(3):211–220, 2000.
- [97] H. LI et H. ZHANG : Ant colony optimization-based multi-mode scheduling under renewable and nonrenewable resource constraints. *Automation in Construction*, 35:431 – 438, 2013.
- [98] R. Y. LI et J. WILLIS : An iterative scheduling technique for resource constrained project scheduling. *European Journal of Operational Research*, 56:370–379, 1992.
- [99] M.J. LIBERATORE et B. POLLACK-JOHNSON : Improving project management decision making by modeling quality, time, and cost continuously. *Engineering Management, IEEE Transactions on*, 30:518–528, 2013.
- [100] J. LIN, Y. QIAN, W. CUI et Z. MIAO : Overlapping and communication policies in product development. *European Journal of Operational Research*, 201:737–750, 2010.
- [101] A. LOVA, P. TORMOS, M. CERVANTES et F. BARBER : An efficient hybrid genetic algorithm for scheduling projects with resource constraints and multiple execution modes. *International Journal of Production Economics*, 117(2):302–316, 2009.
- [102] H. MAGHSOUDLOU, B. AFSHAR-NADJAFI et S.T.A. NIAKI : A multi-objective invasive weeds optimization algorithm for solving multi-skill multi-mode resource constrained project scheduling problem. *Computers and Chemical Engineering*, 88:157–169, 2016.
- [103] J.L. MASHAW et D.L. HARFST : The struggle for auto safety. *Harvard Journal of Law and Technology*, 4:307–312, 1991.
- [104] J. De MAST et J. LOKKERBOL : An analysis of the six sigma dmaic method from the perspective of problem solving. *International Journal of Production Economics*, 139(2):604–614, 2012. Compassionate Operations.
- [105] S. B. H. S. MOHAMAD, M. H. ADENAN et I. Abdul RAHMAN : Project crashing impact to labour productivity and quality of work in construction project. *The International Conference on Civil and Environmental Engineering Sustainability (IConCEES 2011), Johor Bahru, Malaysia*, 2012.
- [106] F. MOHAMMADIPOUR et S.J. SADJADI : Project cost-quality-risk tradeoff analysis in a time-constrained problem. *Computers and Industrial Engineering*, 95:111–121, 2016.

- [107] H. MOKHTARI, R.B. KAZEMZADEH et A. SALMASNIA : Time-cost tradeoff analysis in project management : An ant system approach. *IEEE Transactions on engineering management*, 58, 2011.
- [108] S. MONGHASEMI, M. R. NIKOO, M. A. K. FASAEI et J. ADAMOWSKI : A novel multi criteria decision making model for optimizing time cost quality trade-off problems in construction projects. *Expert Systems with Applications*, 42(6):3089 – 3104, 2015.
- [109] S. MUNGLE, L. BENYOUCEF, Y-J. SON et M.K.TIWARI : A fuzzy clustering-based genetic algorithm approach for time-cost-quality trade-off problems : A case study of highway construction project. *Engineering Applications of Artificial Intelligence*, 26:1953–1966, 2013.
- [110] A. NABER et R. KOLISCH : Mip models for resource-constrained project scheduling with flexible resource profiles. *European Journal of Operational Research*, 239:335–348, 2014.
- [111] J.R. NICHOLSON : Temporary help workers in the u.s. labor market, 2015.
- [112] K. NONOBE et T. IBARAKI : Formulation and tabu search algorithm for the resource constrained project scheduling problem. In : *Ribeiro, C.C., Hansen, P. (Eds.), Essays and Surveys in Metaheuristics. Kluwer, Boston*, pages 557–588, 2002.
- [113] L. OZDAMAR : A genetic algorithm approach to a general category project scheduling problem. *IEEE Transactions on Systems, Man, and Cybernetics – Part C : Applications and Reviews*, 29(1):44–59, 1999.
- [114] J. PATTERSON, R. SLOWINSKI, F. TALBOT et J. WEGLARZ : An algorithm for a general class of precedence and resource constrained scheduling problem. *Advances in Project Scheduling*, pages 3–28., 1989.
- [115] V. PETEGHEM et M. VANHOUCKE : An experimental investigation of metaheuristics for the multi-mode resource-constrained project scheduling problem on new dataset instances. *European Journal of Operational Research*, 235(62-72), 2014.
- [116] V. Van PETEGHEM : *Multi-mode resource-constrained project scheduling problem : Solution procedures and extensions*. Thèse de doctorat, Ph.D. Thesis, Ghent University, 2010.
- [117] V. Van PETEGHEM et M. VANHOUCKE : A genetic algorithm for the preemptive and nonpreemptive multi-mode resource-constrained project scheduling problem. *European Journal of Operational Research*, 201(2):409–418, 2010.
- [118] M. L. PINEDO : *Planning and Scheduling in Manufacturing Planning and Scheduling in Manufacturing and Services*. Springer Series in Operations Research, 2005.
- [119] N. S. POUR, M. MODARRES, M.B. ARYANEJAD et R. T. MOGHADAM : The discrete time-cost-quality trade-off problem using a novel hybrid genetic algorithm. *Applied Mathematical Sciences*, 4:2081–2094, 2010.
- [120] M. RAHIMI et H. IRANMANESH : Multi objective particle swarm optimization for a discrete time, cost and quality trade -off problem. *World Applied Sciences Journal*, 4:270–276, 2008.

- [121] M. RANJBAR, B.D. REYCK et F.KIANFAR : A hybrid scatter search for the discrete time resource trade-off problem in project scheduling. *European Journal of Operational Research*, 193(1):35–48, 2009.
- [122] E. RAUCH, A. DAMIAN, P. HOLZNER et D. T. MATT : Lean hospitality - application of lean management methods in the hotel sector. *Procedia*, 41:614–619, 2016.
- [123] B. De REYCK, E. DEMEULEMEESTER et W. HERROELEN : Local search methods for the discrete time-resource trade-off problem in project networks. *Naval Research Logistic Quarterly*, 45:553–578, 1998.
- [124] T.A. ROEMER et R. AHMADI : Concurrent crashing and overlapping in product development. *Operations Research*, 52:606–622, 2004.
- [125] A. SCHIFFAUEROVA et V. THOMSON : A review of research on cost of quality models and best practices. *International Journal of Quality and Reliability Management*, 23:647–669, 2006.
- [126] N. R. SHANKAR, M.M.K. RAJU, G. SRIKANTH et P. H. BINDU : Time, cost and quality trade-off analysis in construction of projects. *Contemporary Engineering Sciences*, 4:289–299, 2011.
- [127] S. SHIBA, A. GRAHAM, D. WALDEN et R. PIETRI : *TQM : 4 révolutions du management manuel d'apprentissage et de mise en oeuvre du management par la qualité totale*. Paris : Dunod, 2003.
- [128] R. SLOWINSKI : Two approaches to problems of resource allocation among project activities : a comparative study. *Journal of the Operational Research Society*, 31(8):711–723, 1980.
- [129] A.G. SMITH : Measuring on-site production. In *Transactions of the American Association of Cost Engineers*, 31st Annual Meeting, Atlanta, 1987.
- [130] M. G. SPERANZA : Hierarchical models for multi-project planning and scheduling. *European Journal of Operational Research*, 64:312–325, 1993.
- [131] A. SPRECHER : Resource-constrained project scheduling : exact methods for the multi-mode case. *Lecture Notes in Economics and Mathematical Systems*, 409. Springer, Berlin., 1994.
- [132] A. SPRECHER et A. DREXL : Multi-mode resource-constrained project scheduling by a simple, general and powerful sequencing algorithm. *European Journal of Operational Research*, 107:431–450, 1998.
- [133] A. SPRECHER, A. HARTMANN et S. DREXL : An exact algorithm for project scheduling with multiple modes. *OR Spektrum*, 19, 1997.
- [134] K. STYLIDIS, C. WICKMAN et R. SODERBERG : Defining perceived quality in the automotive industry : an engineering approach. *Procedia CIRP*, 36:165–270, 2015.
- [135] F.B. TALBOT : Resource-constrained project scheduling with time-resource trade-offs : the nonpreemptive case. *Management Science*, 28(10):1197–1210, 1982.
- [136] H.R. TAREGHIAN et S.H. TAHERI : On the discrete time, cost and quality trade-off problem. *Applied Mathematics and Computation*, 181:1305–1312, 2006.

- [137] H.R. TAREGHIAN et S.H. TAHERI : A solution procedure for the discrete time, cost and quality tradeoff problem using electromagnetic scatter search. *Applied Mathematics and Computation*, 190:1136–1145, 2007.
- [138] M. TAVANA, A.R. ABTAHI et K. KHALILI-DAMGHANI : A new multi-objective multi-mode model for solving preemptive time-cost-quality trade-off project scheduling problems. *Expert Systems with Applications*, 41:1830–1846, 2014.
- [139] H. R. THOMAS et K. A. RAYNAR : Scheduled overtime and labor productivity : quantitative analysis. *Journal of Construction Engineering and Management*, 181, 1997.
- [140] P. THOMAS et S. SALHI : An investigation into the relationship of heuristic performance with network resource characteristics. *Journal of the Operational Research Society*, 48(1):34–43, 1997.
- [141] V. TIWARI, J.H PATTERSON et V.A. MABERT : Scheduling projects with heterogeneous resources to meet time and quality objectives. *European Journal of Operational Research*, 193:780–790, 2009.
- [142] D. H. TRAN, M-Y CHENG et M-T CAO : Hybrid multiple objective artificial bee colony with differential evolution for the time-cost-quality tradeoff problem. *Knowledge-Based Systems*, 74:176–186, 2015.
- [143] S.K. TYAGI, K. YANG, A. TYAGI et S.N. DWIVEDI : Development of a fuzzy goal programming model for optimization of lead time and cost in an overlapped product development project using a gaussian adaptive particle swarm optimization-based approach. *Engineering applications of artificial intelligence*, 24:866–879, 2011.
- [144] G. ULUSOY et L. OZDAMAR : Heuristic performance and network resource characteristics in resource constrained project scheduling. *Journal of the Operational Research Society*, 40(12):1145–1152, 1989.
- [145] G. VAIRAKTARAKIS : The value of resource flexibility in the resource-constrained job assignment problem. *Management Science*, 49(6):718–732, 2003.
- [146] V. VALLS et A. Perez ans S. QUINTANILLA : Skilled workforce scheduling in service centres. *European Journal of Operational Research*, 193(3):791 – 804, 2009.
- [147] C. VALMOHAMMADI et S. ROSHANZAMIR : The guidelines of improvement : Relations among organizational culture, tqm and performance. *International Journal of Production Economics*, 164:167–178, 2015.
- [148] M. VANHOUCKE : New computational results for the discrete time/cost trade-off problem with time-switch constraints. *European Journal of Operational Research*, 165(2):359–374, 2005.
- [149] M. VANHOUCKE et D. DEBELS : The discrete time-cost trade-off problem : extensions and heuristic procedure. *Journal of Scheduling*, 10(4-5):311–326, 2007.
- [150] M. VANHOUCKE, E. DEMEULEMEESTER et W. HERROELEN : Discrete time-cost trade-offs in project scheduling with time-switch constraints. *Journal of the Operational Research Society*, 53(7):741–751, 2002.

- [151] P. VRAT et C. KRIENCKRAIRUT : A goal programming model for project crashing with piecewise linear time-cost trade-off. *Engineering Costs and Production Economics*, 10(2):161 – 172, 1986.
- [152] M. WALTER et J. ZIMMERMANN : Minimizing average project team size given multi-skilled workers with heterogeneous skill levels. *Computers and Operations Research*, 70:163–179, 2016.
- [153] T. WANG, S. GUO et Y. LIU : Pareto process optimization of product development project using bi-objective hybrid genetic algorithm. *Advances in Engineering Software*, 65:12–22, 2013.
- [154] J. WEGLARZ : On certain models of resources allocation problems. *Kybernetes*, 9(1):61–66, 1980.
- [155] J. WEGLARZ : *Project Scheduling : Recent Models, Algorithms and Applications*, volume 14. Springer, 1999.
- [156] J. WEGLARZ, J. JOZEFOWSKA, M. MIKA et G. WALIGORA : Project scheduling with finite or infinite number of activity processing modes : A survey. *European Journal of Operational Research*, 208:177–205, 2011.
- [157] L.B. WEINSTEIN, J. CASTELLANO, J. PETRICK et R.J. VOKURKA : Integrating six sigma concepts in an mba quality management class. *Journal of Education for Business*, 83(3):233–238, 2008.
- [158] M-C. WU et S-H. SUN : A project scheduling and staff assignment model considering learning effect. *The International Journal of Advanced Manufacturing Technology*, 28(11):1190–1195, 2006.
- [159] L.P. WULIANG et G.W. CHENG : A multi-mode resource-constrained discrete time-cost trade-off problem and its genetic algorithm based solution. *International Journal of project management*, 27(6):600–609, 2009.
- [160] B. YANG, J. GUENES et W. J. O. BRIEN : A heuristic approach for minimizing weighted tardiness and overtime costs in single resource scheduling. *Computers and Operations Research*, 31:1273–1301, 2004.
- [161] K. S. YOGI : An empirical and fuzzy logic approach to product quality and purchase intention of customers in two wheelers. *Pacific Science Review B : Humanities and Social Sciences*, 2016.
- [162] H. ZHANG, C. TAM et H. LI : Multi-mode project scheduling based on particle swarm optimization. *Computer-Aided Civil and Infrastructure Engineering*, 21(2):93–103, 2006.
- [163] H. ZHANG et F. XING : Fuzzy-multi-objective particle swarm optimization for time-cost-quality trade-off in construction. *Automation in Construction*, 19(8):1067–1075, 2010.
- [164] L. ZHANG, J. DU et S. ZHANG : Solution to the time-cost-quality trade-off problem in construction projects based on immune genetic particle swarm optimization. *Journal of Management in Engineering*, 30:163–172, 2014.
- [165] G. ZHU, J.F. BARD et G. YU : A branch-and-cut procedure for the multi-mode resource-constrained project scheduling problem. *INFORMS Journal on Computing*, 18(3):377–390, 2006.



## Résumé

Cette thèse s'inscrit dans la littérature relative au problème d'arbitrage durée, coût et qualité en gestion de projet. Nous développons un modèle d'accélération des projets de fabrication par le recours au travail temporaire et aux heures supplémentaires. Nous prenons en compte l'impact de la main d'œuvre additionnelle sur la durée via les pertes de productivité et sur la qualité des activités qui composent le projet. La qualité de toute activité est mesurée par le pourcentage d'items validés dans sa check-list de contrôle qualité. Nous proposons une formulation linéaire du problème de minimisation de la durée du projet sous contrainte de budget et de respect d'un seuil de qualité minimum pour chacune des activités que le projet comporte. Le programme d'optimisation mixte qui en résulte est appliqué au cas d'un projet de fabrication d'une motrice TGV et permet d'obtenir une planification prévisionnelle de la main d'œuvre pour atteindre la durée optimale d'exécution. Le modèle peut également être utilisé pour accélérer un projet en cours d'exécution ou à des fins plus stratégiques de dimensionnement des postes de travail.

## Mots Clés

Gestion de projet, accélération, arbitrage durée - coût et qualité, travail temporaire, heures supplémentaires, productivité, programmation linéaire

## Abstract

We address the discrete time resource-constrained project scheduling problem in which each activity has a specified work content and its resource usage may vary from period to period. We consider temporary work and overtime as additional renewable resources for crashing the project. We assume that the project quality may be affected by crashing as well as its completion time through productivity loss due to overmanning. We develop a Mixed-Integer Linear Programming (MILP) model to minimise the makespan subject to a budget constraint and to acceptable quality levels for all activities so as to avoid rework. The proposed approach is applied to an actual manufacturing project of a very high speed train motor coach.

## Keywords

Project management, Time Resource Quality Trade-off, Flexible resource profiles, Overmanning, Overtime, Productivity loss, Linear programming, Case study.