
Chapitre IX

LIGNES DE FABRICATION OU D'ASSEMBLAGE

On observe une utilisation croissante de la structure de ligne de production ou de ligne d'assemblage¹, pour la fabrication de produits manufacturés complexes. Cette structure de production, introduite au [chapitre I, § I-2.2.3, page 53](#), est économiquement plus performante que celle de la configuration en ateliers spécialisés, sous des conditions décrites à la [page 60](#). L'un des facteurs déterminants de cette analyse de point mort est le volume de production susceptible d'être traité par la ligne étudiée. L'évolution des caractéristiques du marché (décrite à la [page 67](#)) se traduit par une variété accrue des produits qui conduirait inéluctablement à condamner la structure en ligne si celle-ci n'était pas capable d'absorber la diversité.

- Cette variété des produits manufacturés, analysée au [§ I-3, page 102](#), du [chapitre II](#), peut s'obtenir notamment par une combinaison appropriée de modules, chaque module étant choisi dans un ensemble limité de modules interchangeables. Dans ce cas, la structure en ligne est capable d'absorber la diversité et sa supériorité économique sur la structure en ateliers spécialisés est facile à vérifier, les flux de production de produits différents pouvant s'agréger sans problème dans l'analyse du point mort. On parlera dans ce chapitre de **produits homogènes**, cette homogénéité s'appréciant du point de vue des problèmes posés par la conception d'une ligne.
- La diversité introduite par la présence de composants optionnels, se traduisant par une variation du temps de travail sur certains postes, a des impacts importants dans la définition de la ligne et dans l'ordonnancement de la production (ce qui rend plus compliquée l'analyse de la supériorité économique de cette structure de production). On parlera alors de **produits hétérogènes**, toujours en retenant le point de vue des problèmes posés par la conception d'une ligne. Deux remarques doivent être faites:
 - On peut, sur une même ligne, fabriquer et assembler des gammes de produits utilisant la même plate-forme (voir [page 103](#)), ce qui induit la même variabilité du temps de travail sur certains postes. L'avantage de cette approche est de mieux pouvoir faire face à des variations de la répartition de la demande globale entre les gammes et donc de rendre le système productif plus flexible. Si le volume de production est assez important, on peut avoir intérêt à panacher des lignes ne traitant que des produits homogènes (chacune traitant la partie à peu près certaine de la

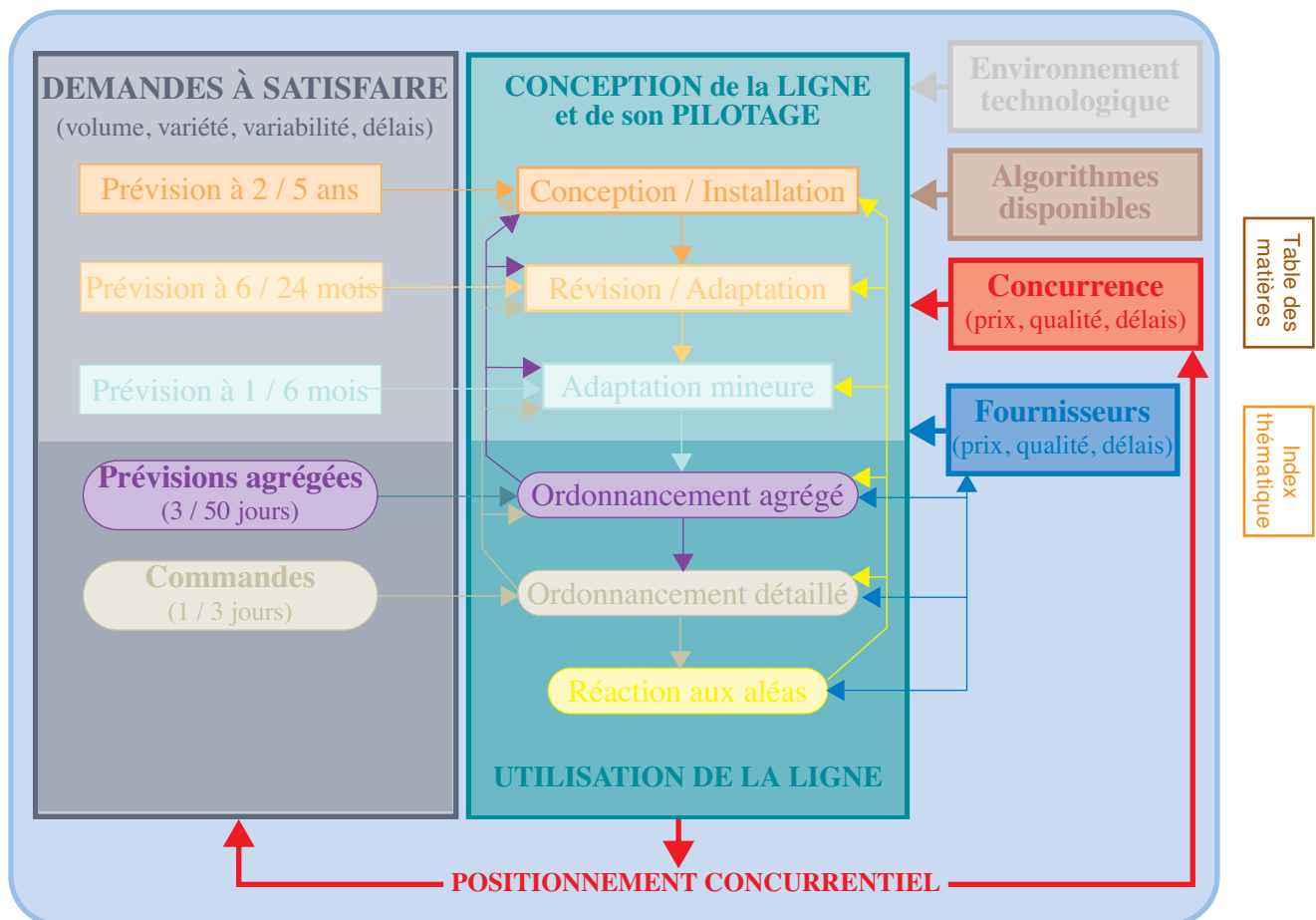
1. On parle encore de **chaînes de production** ou de **chaînes d'assemblage**.

demande d'une gamme) et des lignes polyvalentes, assurant la flexibilité voulue mais subissant l'impact de toutes erreurs de prévision¹.

- Il faut noter que, dans certaines industries, la tendance est d'obtenir des fournisseurs de premier rang, des modules intégrant ou non des composants optionnels, ce qui revient à externaliser le problème chez le fournisseur pour limiter l'hétérogénéité des produits et, ce faisant, la complexité des problèmes rencontrés².

Nous examinerons successivement le cas des lignes dédiées à la fabrication de produits homogènes (section I) et celui des lignes dédiées à la fabrication de produits hétérogènes (section II, page 611). Dans les deux cas, il faudra aborder le problème de la conception de la ligne puis celui de son utilisation. Ces deux problèmes ne sont pas indépendants et ils sont conditionnés par plusieurs facteurs, ce qu'illustre la figure 162.

FIGURE 162
De la conception de la ligne à son pilotage



1. Le coût de telles lignes est généralement plus élevé que celui de lignes spécialisées dans la production de produits homogènes mais la comparaison doit tenir compte du caractère plus ou moins certain de la demande et de l'intérêt économique de la variété de l'offre.

2. Par exemple, dans l'industrie automobile, les boucliers-arrière de certains véhicules haut de gamme peuvent ou non intégrer un capteur de recul. En demandant au fournisseur de livrer des boucliers complets dans un ordre respectant la variété requise au montage (couleur, présence ou non du capteur de recul), on passe alors d'un approvisionnement de composants à un approvisionnement de modules et les problèmes liés à la diversité sont externalisés chez l'équipementier.

L'interdépendance entre le problème de conception et celui de l'utilisation de la ligne tient au fait qu'il est recommandé, au niveau de la conception, de tester la robustesse et la pertinence des choix opérés sur le plan technique et celui de ses règles de pilotage, en simulant l'utilisation du système dans des situations plus ou moins proches de celle retenues dans la conception. Il s'agit alors de vérifier la réponse du système pour faire face à des variations de la demande, tant en volume, qu'en composition (si la ligne est conçue pour absorber une certaine variété), ou à des aléas de production (pannes, ruptures d'approvisionnement et problèmes de qualité¹), compte tenu du degré de flexibilité² physique de la ligne et des règles de pilotage, lesquelles peuvent s'appuyer sur des algorithmes relativement sophistiqués. C'est également en fonction de la conception qu'est pris en compte le degré de transformation que peut supporter la ligne pour faire face à des transformations plus ou moins importantes de la demande (d'où les boucles de rétroaction de la [figure 162](#)).

De nombreux facteurs conditionnent la conception et l'utilisation d'une ligne. Sans sous-estimer les facteurs de contingence internes (orientations stratégiques, capacité de financement...), on s'attachera ici aux facteurs externes. Tout d'abord, les caractéristiques déterministes (tendance, saisonnalité) et stochastiques de la demande jouent un rôle important sur la conception de la ligne, mais aussi celle des produits³ et les choix opérés vont conditionner la position concurrentielle de l'entreprise sur le créneau de demande visé et donc l'importance de la demande à satisfaire. La réponse du système dépend également des fournisseurs, dans une optique de chaîne logistique, d'autant plus fortement que les flux seront tendus. Enfin, les technologies disponibles et leurs coûts jouent un rôle important dans la conception de la ligne, ainsi que les règles de pilotages disponibles (en particulier les algorithmes permettant une meilleure conception ou un meilleur ordonnancement).

Table des
matièresIndex
thématique

SECTION I LIGNES DÉDIÉES À LA FABRICATION DE PRODUITS HOMOGÈNES

On examinera d'abord le problème de la conception d'une ligne de production ou d'assemblage de produits homogènes (§ I-1) puis celui de son utilisation dans son aspect d'ordonnancement (§ I-2, [page 599](#)). La correction au plus tôt de problèmes de qualité induit un certain nombre de perturbations (§ I-3, [page 600](#)) dont il convient de tenir compte; de ce point de vue, il n'y a pas de différence notable entre les produits homogènes et les produits hétérogènes.

1. La ligne pouvant comporter des stocks-tampons permettant d'absorber des aléas. Par ailleurs, même dans le cas d'un approvisionnement à flux tendu (ou approvisionnement synchrone), ce que l'on évoquera après, les composants qui viennent de l'extérieur font l'objet de stocks en bord de ligne, ces stocks incluant un stock de sécurité plus ou moins important. L'importance de ces stocks, contrainte physiquement lors de la conception de la ligne, permet d'assurer un découplage plus ou moins important des problèmes et donc d'en éviter la propagation.

2. Cette problématique de la flexibilité est discutée au [chapitre III](#), § I-1.5, [page 147](#).

3. En particulier par le biais de la standardisation (voir [chapitre III](#), § III, [page 133](#)) et de la conception modulaire (voir [chapitre II](#), [page 103](#)).

I-1 Conception d'une ligne de production ou d'assemblage de produits homogènes

Dans un premier paragraphe (§ I-1.1) nous examinerons le problème de base de l'équilibrage d'une chaîne et introduirons les concepts utiles pour son analyse, avant de présenter les approches utilisées pour les résoudre. On dispose en réalité de deux catégories d'outils. Ceux qui relèvent de la programmation mathématique (§ I-1.2, page 594) permettent une définition claire du problème à résoudre et font appel à des algorithmes garantissant l'optimalité de la solution pour des problèmes de dimension limitée. Les techniques heuristiques d'équilibrage d'une chaîne (§ I-1.3, page 596) ne garantissent pas l'optimalité, mais sont d'une utilisation assez facile et permettent de résoudre des problèmes d'une plus grande dimension.

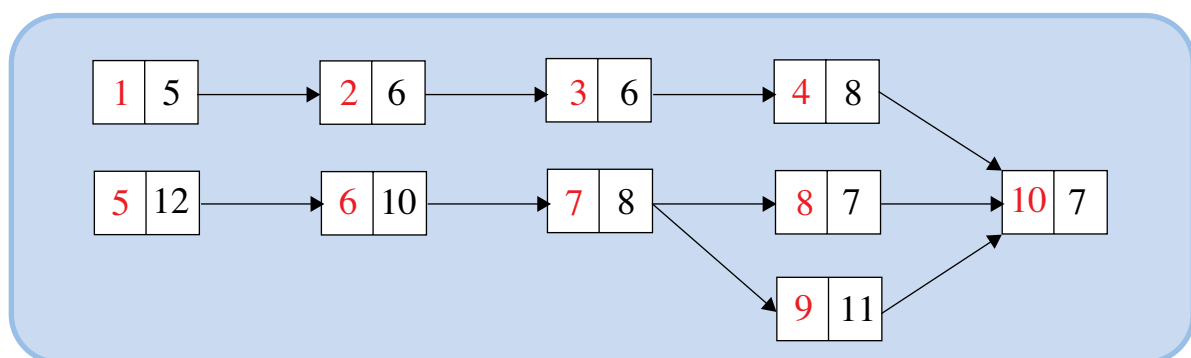
Plusieurs configurations de ligne ont été introduites (voir la figure 6 de la page 54 et l'animation de la figure 7 de la page 56); nous ne traiterons ici que le cas le plus simple: celui d'une ligne unique dans laquelle chaque opérateur ne traite que des opérations consécutives.

I-1.1 La problématique de l'équilibrage d'une ligne de fabrication ou d'assemblage

Moulacaf SA fabrique divers articles électroménagers. La chaîne de fabrication de son nouveau modèle de rôti-ssoire de luxe comporte une cinquantaine d'opérations. Nous n'examinerons ici que les problèmes relatifs à un sous-ensemble bien défini. Les 10 opérations sont visualisées dans le graphe *Potentiel - Tâches* de la figure 163, en utilisant une version simplifiée des conventions d'écriture adoptées dans le chapitre IV: à gauche le code repérant l'opération, à droite son temps opératoire (exprimé en dixièmes de minute):

FIGURE 163

Grappe Potentiel - Tâches des opérations nécessaires à l'exécution du sous-ensemble du nouveau modèle de rôti-ssoire de Moulacaf



Le problème posé est celui de la définition d'une ligne capable de fabriquer 3 500 rôti-ssoires par mois de 164 heures, minimisant le nombre de postes de travail nécessaires, tout en équilibrant les charges de travail de chacun de ces postes. Un poste de travail – on parle encore de **station** dans la littérature des lignes de production ou d'assemblage (cf. page 54) – se définit par la prise en charge par une même personne¹ de plusieurs opérations consécutives ou indépendantes. On

1. Dans le cas d'une ligne traitant des produits hétérogènes, cette personne peut éventuellement être aidée par un renfort occasionnel, comme on le verra à la page 611.

suppose donc ici que le nombre d'opérateurs est variable. On présentera ultérieurement (page 596) une autre formulation dans laquelle le nombre d'opérateurs est fixe et où l'on cherche à maximiser la production, qui devient alors la variable du problème.

Dire que le système doit être en mesure de fabriquer 3 500 rôtissoires par mois, c'est imposer que chaque rôtissoire soit exécutée en un temps de $164 / 3\,500 = 0,04686$ heure $= 0,04686 \times 60 = 2,81$ minutes ($= 28,1$ dixièmes de minutes, pour rester cohérent avec l'unité de temps retenue à la figure 163). C'est cette mesure de la productivité du système, appelée **cycle de fabrication** (noté c) que l'on utilise dans l'analyse d'une chaîne de production. On peut remarquer que l'on retrouve dans la définition de la productivité la distinction classique entre la notion de fréquence (nombre d'unités produites par unité de temps) et la notion de temps unitaire – ici le cycle de fabrication – qui n'est autre que l'inverse de la fréquence.

Il est évident qu'aucun poste de travail ne devra passer plus de 28 dixièmes de minute sur chaque rôtissoire, faute de quoi le niveau minimal de productivité de la chaîne, que l'on s'impose au départ comme contrainte, ne sera pas respecté. À l'inverse, il est tout aussi évident que le poste le plus chargé constitue le goulot d'étranglement du système productif et que c'est lui qui fixera la productivité de la chaîne. Par exemple, si le poste de travail le plus chargé requiert 12 dixièmes de minute, la production maximale horaire sera $60 / 1,2 = 50$ rôtissoires, ou encore $164 \times 50 = 8\,200$ rôtissoires/mois. Si l'on désire accroître en permanence la capacité de production sans avoir à multiplier le nombre de chaînes ou à augmenter les heures travaillées, il faut alors demander au spécialiste de la fonction méthode qui a procédé au découpage des opérations de modifier leur définition pour diminuer le temps opératoire des plus importantes (par transfert de charge de travail sur des opérations adjacentes, ou amélioration de la productivité). Par exemple, si l'on veut atteindre en régime de croisière une production de 10 000 rôtissoires, il faudra que la durée de l'opération élémentaire la plus longue n'excède pas $164 \times 60 / 10\,000 = 9,8$ dixièmes de minute.

Revenons au problème initialement posé. Le temps d'exécution total de toutes les opérations est de 80 dixièmes de minutes. Le nombre de postes de travail idéal serait $\sum_i t_i / c = 80 / 28,1 = 2,86$ postes de travail pour que la chaîne puisse être parfaitement équilibrée. Ce nombre qui n'est pas entier ne peut être réalisé; dans le meilleur des cas, il faudra au moins 3 postes de travail, mais rien ne garantit que le nombre arrondi immédiatement supérieur soit:

- d'une part réalisable du fait des contraintes de succession (supposons par exemple que les opérations 1 à 4 durent chacune 1,5 minute, il ne sera alors pas possible d'avoir moins de 4 postes de travail), ou pour des raisons d'impossibilité physique de regroupement de certaines opérations (outillage nécessaire très différent empêchant l'organisation rationnelle du poste de travail et donc faussant la validité des temps opératoires annoncés),
- et, d'autre part, soit facile à trouver (pour vous en convaincre, il vous suffit de chercher à résoudre le problème posé pour n'obtenir que 3 postes de travail).

Pour un nombre de postes de travail ($\sum_i t_i / c \leq k < n$), le temps mort total, en

$$\text{pourcentage du temps total de travail, sera: } 100 \frac{kc - \sum_i t_i}{kc} = 100\% - \frac{100}{kc} \sum_i t_i.$$

Il est équivalent de chercher à minimiser le nombre de postes de travail ou de minimiser le pourcentage des temps morts. Une ligne sera dite parfaitement équilibrée si ce pourcentage est nul, mais une chaîne peut être théoriquement équilibrée à 90 % (c'est-à-dire avoir 10 % de temps mort) et connaître des fluctuations par rapport à ce taux, d'une part parce que les temps opératoires sont des temps qui ne sont pas certains mais aléatoires et, d'autre part, parce que le niveau d'activité réel peut varier par rapport aux prévisions (sous-utilisation de la ligne).

Nous venons d'évoquer le caractère aléatoire des temps opératoires. L'approche stochastique de cette classe de problèmes est particulièrement délicate. Aussi raisonne-t-on en avenir certain, en majorant d'une marge de sécurité les différents temps opératoires¹. Examinons maintenant les solutions proposées pour résoudre ces problèmes.

I-1.2 Formalisation du problème par la programmation mathématique

On dispose de deux batteries d'outils : ceux qui relèvent de la programmation mathématique (voir le [chapitre XVI, page 589](#)), qui garantissent l'optimalité de la solution trouvée, et les procédures heuristiques d'équilibrage de chaîne, qui sont rapides mais ne garantissent pas l'optimalité de la solution. En réalité, cette distinction est moins nette qu'il ne le paraît car, dans la première approche, les méthodes de résolution numérique disponibles ne permettent pas de traiter, dans un temps raisonnable et avec une garantie d'optimalité, des problèmes de grande dimension.

Le recours à la programmation mathématique² s'effectue dans deux directions, la programmation dynamique et la programmation linéaire en nombres entiers. Les performances de la programmation dynamique dans ce domaine sont médiocres ; la formulation par la programmation linéaire en nombres entiers est en effet plus rapide de 50 %. Celle-ci, due³ à Thangavelu et Shetty (1971, [415]), doit son efficacité, d'une part à une astuce de présentation et, d'autre part, au fait que

1. On a intérêt à calculer cette marge de sécurité (coefficient $h > 1$) en tenant compte des compensations des temps autour d'une valeur moyenne \bar{t} de l'opération, pour k objets passant successivement sur ce poste (k , variant de 1 à K), en faisant en sorte que la probabilité d'un dépassement (événement $\sum_k t_k > h k \bar{t}$) soit inférieure à une probabilité donnée $P\left(\sum_k t_k > h k \bar{t}\right) < \alpha$. On peut généraliser le raisonnement en cas de compensation possible sur plusieurs postes consécutifs (que l'on notera i) : $P\left(\sum_k \sum_i t_{ik} > h_i k \bar{t}_i\right) < \alpha$. En l'absence de solution analytique, l'appel à la simulation via des add-in d'Excel, comme @Risk (Palisade.com), permet facilement de trouver une solution (h_i) par tâtonnement ; dans le cas où le problème se limite à un seul poste, la détermination du h optimal s'effectue immédiatement avec un logiciel comme *Risk Optimizer*.

2. Voir Johnson et Montgomery (1974, [245]), p. 365-368.

3. Voir Johnson et Montgomery (1974, [245]), p. 374-375. Un historique de l'évolution de ces formulations peut être trouvé dans Scholl (1999, [373], p. 29). Deux autres formulations du problème peuvent être trouvées dans cet ouvrage, lesquelles sont plus performantes, sous certaines conditions.

les variables utilisées ne pouvant prendre que les valeurs 0 ou 1, des algorithmes performants peuvent être utilisés.

La formulation repose sur les variables binaires x_{ik} qui valent 1 si l'opération i est réalisée sur le poste k , et 0 dans le cas contraire. Il faut que l'opération i soit réalisée sur l'un des postes k , ce qui conduit à la [relation 112](#):

$$\sum_{k=1}^K x_{ik} = 1, \text{ pour } i = 1 \text{ à } N \quad \text{relation 112}$$

Il convient ensuite de respecter des contraintes d'antériorité des opérations¹, lesquelles se répercutent sur les postes. Sachant que le numéro de poste auquel une opération i est affectée est $\sum_k kx_{ik}$, le fait que l'opération i doit précéder l'opération j , se répercute sur leurs postes d'affectation²:

$$\sum_k kx_{ik} \leq \sum_k kx_{jk}, \text{ pour tout } i, \text{ ancêtre de } j \quad \text{relation 113}$$

Par ailleurs, la charge de travail ($\sum_i t_i x_{ik}$) d'un poste k ne doit pas dépasser le temps de cycle c visé:

$$\sum_i t_i x_{ik} \leq c, \text{ pour } k = 1, \dots, K \quad \text{relation 114}$$

Pour déterminer le **nombre minimal de postes de travail**³, il suffit de partir d'une borne supérieure K (qui ne peut être supérieure au nombre d'opérations) et d'utiliser la variable indicatrice y_k qui vaut 1 si le poste k est utilisé. Dans ces conditions, la fonction-objectif est:

$$\text{Min} \sum_{k=1}^K y_k \quad \text{relation 115}$$

Elle tend à faire prendre une valeur nulle aux y_k . Pour forcer la variable indicatrice y_k à prendre la valeur 1 si le poste est utilisé, il suffit d'introduire la contrainte décrite par la [relation 116](#) qui force y_k à 1 dès qu'une opération est affectée au poste k :

$$\sum_i x_{ik} \leq Ny_k, \text{ pour } k = 1, \dots, K \quad \text{relation 116}$$

La solution de ce problème n'a pas de raison de lisser la charge de travail entre les postes. Il peut être alors souhaitable de recommencer la recherche d'une solu-

1. Cette «propagation de contraintes» des opérations sur les postes peut ne pas être fondée, en particulier si la ligne est organisée physiquement en serpent (voir l'animation de la [figure 7, page 56](#)) permettant à un opérateur de ne pas traiter que des opérations contiguës. Cela étant, les contraintes de déplacement doivent alors être prises en compte, un opérateur ne pouvant prendre en compte n'importe quel groupe d'opérations dont le cumul des temps opératoires est inférieur au temps de cycle; cette prise en compte est possible (avec utilisation d'une matrice de temps de transport d'un poste à un autre) mais complique singulièrement le problème. Ajoutons que le § III-2.1.3, page 319, du [chapitre IV](#) traite de la formulation du problème d'ordonnancement de projet par la programmation mathématique.

2. Il est évident que cette contrainte doit être remplacée lorsque l'on est en présence de lignes en serpent, permettant à un opérateur de ne pas traiter que des opérations contiguës.

3. Dans l'ouvrage de Guéret, Prins & Sevaux (1992, [210], p. 125-130), une variante (illustrée par l'utilisation d'un modèleur) part d'une formulation sans fonction-objectif, dans laquelle on décrémente progressivement le nombre de postes dans le problème posé, jusqu'au moment où l'on ne trouve plus de solution.

tion, en retenant le nombre de postes $\sum_{k=1}^K y_k$ qui vient d'être trouvé et de le considérer comme fixé dans la seconde formulation du problème qui, par l'abaissement du temps de cycle, doit contribuer au lissage.

Cette seconde formulation du problème vise à **minimiser le temps de cycle c** pour un nombre K prédéterminé de postes. Les relations 115 et 116 sont remplacées par la relation 117 qui décrit la nouvelle fonction-objectif :

$$\text{Min } c \quad \text{relation 117}$$

Ces problèmes sont NP-difficiles (voir page 361). Des méthodes de résolution de type *branch and bound* (voir page 390) ont été mises au point pour résoudre des problèmes pouvant comporter jusqu'à un millier d'opérations, de manière d'autant plus efficace que chaque poste traite un nombre suffisant d'opérations¹.

Ces modèles peuvent être adaptés pour prendre en compte de nouvelles contraintes, en s'appuyant sur les principes décrits au § II-2, page 1135 du chapitre XVI. Par exemple, il peut être indispensable que les opérations i_1 et i_2 soient exécutées sur le même poste de travail (par exemple, parce que le travail est le même et porte sur des pièces symétriques); on peut alors mobiliser la règle 4 de la page 1140, qui conduit à créer la relation 118:

$$x_{i_1 k} = x_{i_2 k}, \text{ pour } k = 1, \dots, K \quad \text{relation 118}$$

I-1.3 Les procédures heuristiques de résolution du problème d'équilibrage de chaîne

Plusieurs méthodes de recherche empiriques d'une solution ont été mises au point. Certaines sont basées sur un algorithme de type déterministe et peuvent résoudre des problèmes de quelques centaines d'opérations et dizaines de postes de travail, tandis qu'une autre méthode est basée sur une simulation aléatoire. Nous n'en présenterons que quelques-unes pour le seul intérêt de la description de démarches de résolution car il existe plusieurs logiciels commerciaux qui s'appuient sur des combinaisons d'heuristiques sophistiquées.

I-1.3.1 Les algorithmes déterministes

De nombreux algorithmes ont été imaginés. Celui proposé en 1961 par Helgeson et Birnie² se focalise sur les tâches en les classant selon une méthode de pondération de leurs positions (*Ranked Positional Weight Method*):

- il ordonne les tâches par ordre décroissant d'un coefficient de position défini comme le cumul des temps opératoires de la tâche considérée et des tâches qui la suivent,
- puis il affecte les postes de travail (définis pour un cycle de production) aux tâches dans cet ordre en respectant les antécédences.

1. Voir, en particulier, l'algorithme FABLE (*Fast Algorithm for Balancing Lines Effectively*) mis au point par Johnson en 1988 (voir Nof, Wilhelm & Warnecke, 1997, [318], p. 214-227); cet algorithme a une bonne efficacité pour des problèmes conduisant à charger en moyenne les postes à 6 opérations.

2. Voir Nof, Wilhelm & Warnecke (1997, [318], p. 210-211).

Nous ne nous étendrons pas sur cette méthode qui présente comme avantage principal de fournir rapidement une solution (et donc une borne supérieure au nombre de postes), mais qui peut être assez éloignée de la solution optimale.

La seconde méthode, plutôt focalisée sur les postes, a été proposée par Kilbridge et Wester¹. Elle consiste à partir du **tableau 171** où l'on note, pour chaque opération, le nombre d'opérations qui la précèdent (ce qui correspond à la notion de niveau dans le tableau d'antériorité que nous avons établi dans les techniques d'ordonnancement de la série unitaire; voir **tableau 55**, page 290).

TABLEAU 171
Préparation de l'application de l'algorithme de Kilbridge et Wester

Opération i	Niveau	t_i		Opération i	Niveau	t_i
1	0	5	\Rightarrow	1	0	5
2	1	6		5	0	12
3	2	6		2	1	6
4	3	8		6	1	10
5	0	12		3	2	6
6	1	10		7	2	8
7	2	8		4	3	8
8	3	7		8	3	7
9	3	11		9	3	11
10	4	7		10	4	7

Table des
matières

Index
thématique

La productivité minimale imposée est celle d'un cycle de production le plus proche possible de 28 dixièmes de minute, sans pouvoir dépasser cette valeur. Ce cycle est réalisable puisqu'aucun temps opératoire élémentaire n'excède cette durée. La procédure proposée consiste à affecter des opérations à un premier poste, puis à recommencer pour un deuxième poste en partant des opérations restantes, etc. jusqu'à épuisement des opérations à affecter. La règle d'affectation revient à attribuer à un poste en priorité des opérations restantes du niveau le plus bas, en complétant progressivement son «service». Lorsque plusieurs opérations d'un même niveau sont susceptibles de compléter le service d'un poste, sans que toutes ces opérations puissent lui être attribuées, on retiendra la combinaison d'opérations qui sature juste le poste² et à défaut celle qui a le temps opératoire le plus grand possible, afin de laisser le maximum de marge de manœuvre aux postes suivants. Appliquons cet algorithme.

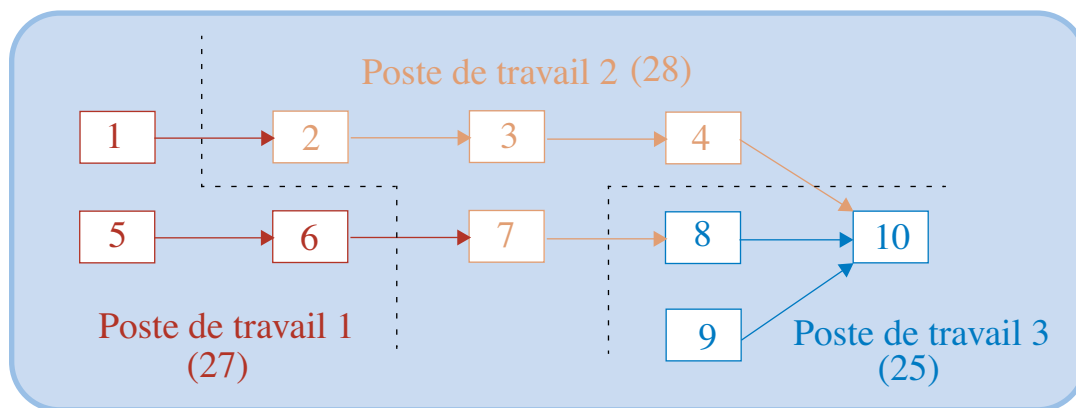
- Il conduit à attribuer au premier poste toutes les opérations de niveau 0, c'est-à-dire les opérations 1 et 5, ce qui correspond à un service de $t_1 + t_5 = 5 + 12 = 17$. Il reste donc un complément de service possible de $28 - 17 = 11$ que l'on cherchera dans les opérations de niveau 1. Deux opérations sont candidates (2 et 6) mais ne peuvent être simultanément retenues; on retient donc celle qui a le temps opératoire le plus fort sans dépasser le

1. Voir Starr (1978, [395]), p. 202-204. Starr montre sur un exemple numérique simple la supériorité de la méthode de Kilbridge et Wester sur celle de Helgeson et Birnie. Voir aussi Buffa et Taubert (1972, [73]), p. 335-347.

- reliquat de 11 (opération 6). Le premier poste a donc une charge de travail de 27 dixièmes de minute et réalise les opérations 1, 5 et 6.
- Le deuxième poste se voit attribuer la dernière opération de niveau 1 qui reste, l'opération 2 ($t_2 = 6$), les deux opérations de niveau 2 peuvent être retenues car le cumul des temps opératoires reste inférieur à 28 ($t_2 + t_3 + t_7 = 6 + 6 + 8 = 20$). Trois opérations de niveau 3 sont candidates, mais une seule peut être retenue, celle qui a le temps opératoire le plus long possible, c'est-à-dire l'opération 4 qui « charge » le poste à 28 dixièmes (on peut remarquer qu'aucun temps n'est perdu dans ce poste).
 - Le troisième poste se voit affecter les trois opérations restantes (8, 9 et 10), avec une charge de travail de 25. Au total, cette répartition donne un travail effectif de $27 + 28 + 25 = 80$, c'est-à-dire le cumul de tous les temps opératoires pour un travail potentiel de $3 \times 28 = 84$, le temps mort correspond à $4 / 84 = 4,8\%$ du temps total, autrement dit la chaîne est équilibrée à $80 / 84 = 95,2\%$.

FIGURE 164

Définition d'une chaîne équilibrée en application de l'algorithme de Kilbridge et Wester



On peut remarquer qu'un meilleur équilibre entre les postes peut être obtenu en permutant les tâches 4 et 8, ce qui donne une charge de travail de 27 au poste 2 et de 26 au poste 3 ; cette solution permet en outre une production supérieure à celle requise (3644 contre 3500).

Les performances de la méthode de Kilbridge et Wester sont en général très bonnes lorsque le temps de cycle est relativement important, c'est-à-dire qu'un poste de travail comporte des opérations de plusieurs niveaux. Lorsque le cycle est

2. *Note de la page précédente.* L'exemple retenu est trop simple pour montrer toutes les implications de la méthode proposée. En effet, une opération A d'un niveau i , par exemple $i = 2$, n'est pas forcément suivie d'une opération B de niveau $i + 1$; B peut être par exemple de niveau 5, parce que sa réalisation est conditionnée non seulement par la réalisation préalable de A, mais aussi de celle d'autres opérations. Lorsque ce cas se produit, on mentionne dans le tableau de préparation de l'application de l'algorithme, la plage de niveau 2 - 4 au lieu du seul niveau 2, en face de l'opération A. Ceci implique que A se trouvera affectée à un poste comportant des opérations dont le niveau le plus faible est 2, 3, ou 4. L'intérêt de cette mention est le suivant. Supposons que l'on étudie le deuxième poste de travail et que celui-ci comporte les opérations de niveau 2 non prises en compte dans le premier poste de travail, plus des opérations de niveau 3, mais que l'on ne trouve pas de combinaison d'opérations de niveau 3 susceptible de compléter exactement le service de ce deuxième poste de travail ; on pourra examiner alors si un échange de l'opération A avec une opération de niveau 4 permet de saturer ce deuxième poste de travail. Si tel est le cas, cet échange peut être réalisé sans violer les contraintes d'antériorité. Rien ne s'oppose à ce que d'autres opérations que l'opération A soient dans le même cas, ce qui augmente la probabilité de saturer le deuxième poste de travail, ... et également l'exploration combinatoire.

faible et qu'il est nécessaire de faire appel à plusieurs postes de travail pour traiter des opérations d'un même niveau, davantage d'essais doivent être tentés (problème combinatoire pour un même niveau) sans qu'on ait la certitude de trouver de bons résultats.

I-1.3.2 Les algorithmes stochastiques

Arcus¹ a proposé en 1966 une procédure heuristique d'équilibrage de chaîne, connue sous le nom de COMSOAL (*Computer Method of Sequencing Operations for Assembly Lines*) qui est dotée d'assez bonnes performances. Elle illustre l'amorce d'un courant de résolution de problèmes hautement combinatoires, en s'appuyant sur la génération contrôlée d'un grand nombre de solutions faisables en conservant, tout au cours de ce processus, la meilleure solution trouvée. Dans la génération d'une solution, il y a construction progressive d'une solution, sans remise en cause des décisions antérieurement prises dans cette construction de solution; lorsque la décision que l'on doit prendre pour compléter la solution comporte plusieurs alternatives, un choix est effectué aléatoirement sur la base de probabilités calculées en fonction d'un critère prédéfini. Cette approche reprend une partie des principes qui fondent les algorithmes déterministes. L'optimalité de la solution n'est pas garantie mais comme dans ce type de problèmes, il existe généralement un très grand nombre de solutions ayant la même performance, on espère par ce biais atteindre l'une d'entre elles ou s'en approcher suffisamment.

La construction d'une séquence réalisable s'effectue en modifiant la procédure de Kilbridge et Wester comme suit: lorsque plusieurs opérations d'un même niveau sont «candidates à un poste de travail», on remplace la recherche d'une combinaison complétant exactement le service du poste étudié par une sélection aléatoire de cette opération, compatible avec le cycle imposé. Reprenons notre exemple pour illustrer cette procédure. Le choix entre les opérations 2 et 6, toutes deux du niveau 1, s'effectue aléatoirement en leur affectant, par exemple, la même probabilité de 50%. Si le tirage au sort sélectionne l'opération 2, le poste de travail suivant commencera par inclure l'opération 6 qui est celle de niveau le plus faible parmi les opérations non encore affectées. Dans une version plus performante, des probabilités inégales sont attribuées aux opérations candidates selon des principes plus ou moins sophistiqués et l'on peut tenir compte de contraintes autres que celles d'antériorité, qui sont les seules que nous avons prises en compte ici².

I-2 Ordonnancement sur ligne de production ou d'assemblage de produits homogènes

Par définition, l'ordonnancement choisi n'a pas d'incidence sur la charge de travail de chacun des postes, lorsque l'on est en présence de produits homogènes. Il joue cependant un rôle lorsque, sur un poste donné, on a le choix entre plusieurs modules ou composants, pour personnaliser le produit. En cas d'approvisionnement synchrone (défini en [page 509](#)), il est généralement demandé aux fournis-

1. Voir Arcus (1966, [19]); les performances de cette approche sont discutées dans Johnson et Montgomery (1974, [245], p. 173-174) ou Buffa et Taubert (1972, [73], p. 348-352), décrivent avec détail l'organigramme de traitement (ce qui ne fait que formaliser le raisonnement rapidement décrit ici). Voir également Nof, Wilhelm & Warnecke (1997, [318], p. 211-213)

2. Par exemple, le fait qu'il soit préférable qu'un même opérateur prenne en charge deux opérations portant sur deux pièces symétriques.

seurs de livrer certains modules assurant la variété (sièges automobiles personnalisés par leurs housses) dans l'ordre de montage de ces composants ou modules (on parle d'encyclage; voir [page 509](#)), ce qui présente un intérêt en cas de composants ou modules encombrants, coûteux ou très variés. La réquisition du fournisseur dans un approvisionnement synchrone lui laisse un délai très court pour livrer les modules dans l'ordre demandé, d'autant qu'il reste souvent une dernière opération pour différencier le produit (montage de la housse) suivant le principe de différenciation retardée (voir [pages 105 et 464](#)), dans le cadre d'un assemblage à la commande pour le fournisseur. La séquence définie dans l'**ordre de réquisition** transmis au fournisseur a souvent intérêt à lisser la consommation des références, ce qui implique que les solutions d'ordonnement ne sont pas équivalentes. S'il n'existe qu'un poste de travail pour lequel un ordre de réquisition est périodiquement passé, la solution est triviale. Si plusieurs postes de travail sont concernés, le problème devient rapidement complexe dans la mesure où le lissage ne peut être le meilleur pour tous les ordres de réquisition.

Dès 1983, Monden (1998, [\[307\]](#), chapitre XVI et annexe III) présente l'approche Toyota qui vise à définir un ordonnancement lissant la consommation des composants optionnels et propose un algorithme de résolution (*Goal chasing method I*, avec version simplifiée, le *Goal chasing method II*) qui, ayant tendance à repousser certains problèmes en fin de journée, a conduit à une modification intégrant une préoccupation de lissage de la charge de travail (1998, [\[307\]](#), chapitre XVII). L'approche initiale, orientée «juste à temps», a fait l'objet de nombreux articles, certains proposant des formulations alternatives, dont certaines intègrent une préoccupation de lissage de la charge de travail, ou des algorithmes plus performants¹.

Notons, enfin, que cette variété peut impliquer des changements de réglage de certaines stations de travail, différents selon le réglage antérieur mais sans incidence sur le temps opératoire lui-même. Une solution à ce problème peut être trouvée en utilisant l'approche décrite au [§ I-2.1, page 390](#), du [chapitre V](#).

I-3 Incidence des aléas sur le fonctionnement de la ligne

On examinera successivement l'incidence de l'arrêt d'un poste de travail sur le fonctionnement de la ligne ([§ I-3.1](#)) et celle du traitement immédiat de problèmes de qualité ([§ I-3.2, page 602](#)). Ces différentes observations sont valables également lorsque la ligne est dédiée à la fabrication ou l'assemblage de produits hétérogènes.

I-3.1 Incidence de l'arrêt d'un poste de travail sur le fonctionnement de la ligne

Quel que soit le système productif, d'inévitables incidents font que tout ne se passe pas comme prévu: temps opératoires réalisés différents des temps prévus, rupture d'approvisionnement, pannes de machines, absence inopinée d'un opérateur, problèmes de qualité en production... S'agissant d'une ligne de production ou d'assemblage, l'arrêt d'un poste de travail de la ligne conduit à l'arrêt de la

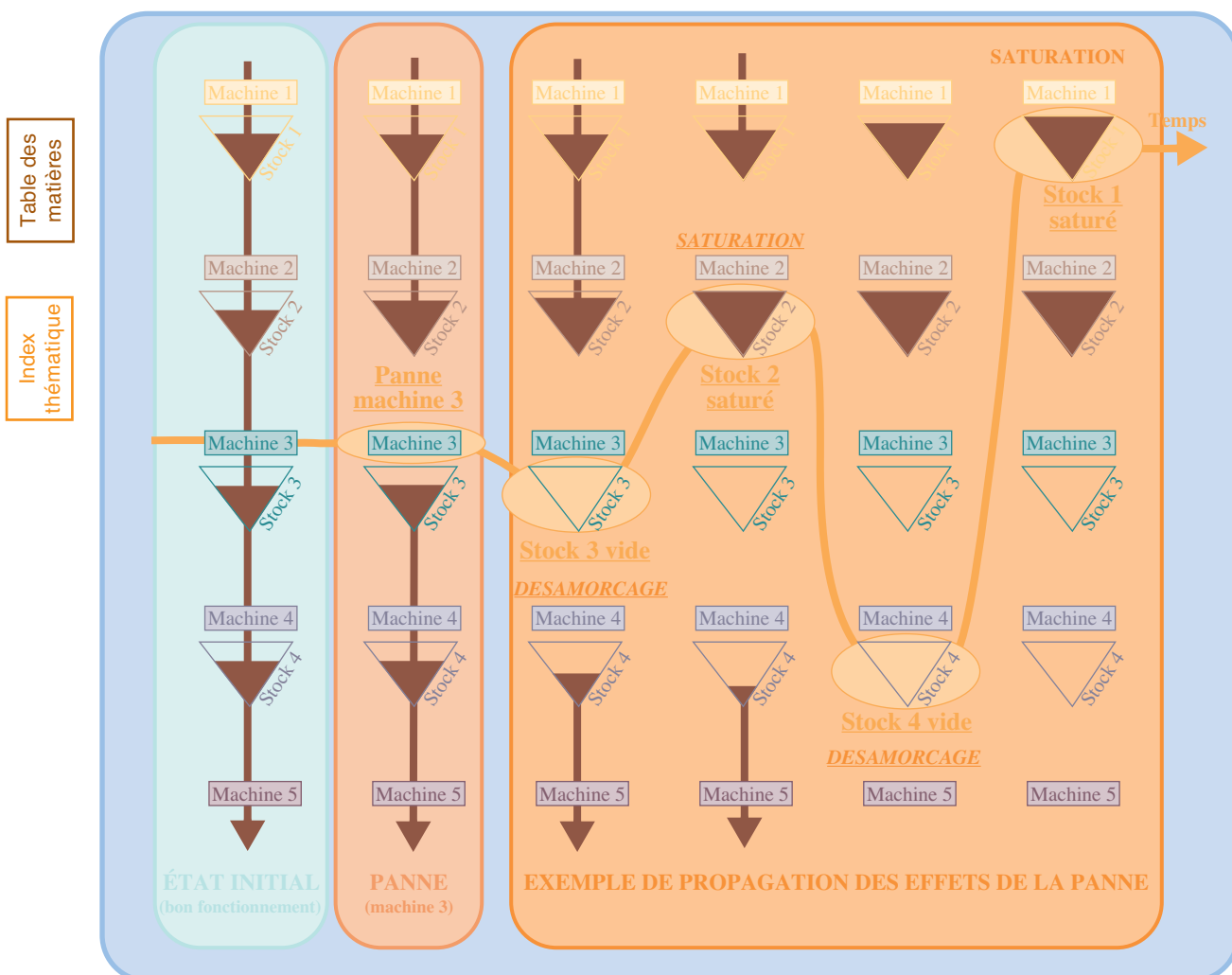
1. Par exemple Mintelburg (1989, [\[303\]](#)), Inman & Bulfin (1991, [\[238\]](#)), Kubiak (1993, [\[266\]](#)), Kubiak et Sethi (1991, [\[267\]](#)), Bautista, Companys & Corominas (1996, [\[40\]](#)), Joo & Wilhelm (1993, [\[252\]](#)), d'autres montrant que cette approche est un cas particulier d'une approche plus générale (Mintelberg & Sinnamon (1989, [\[304\]](#))).

ligne plus ou moins rapidement, en fonction de l'importance des stocks qui existe entre les postes de travail. Ces stocks-tampons permettent un certain découplage retardant la propagation d'un incident (on reviendra sur ce point à la page 624) mais il ne s'agit pas du seul mécanisme de découplage (comme cela a été déjà discuté à la page 512). Le mécanisme de paralysie progressif de la ligne est décrit dans la figure 165¹: la panne de la machine 3 empêche la poursuite de l'approvisionnement du stock-aval, provoquant:

- un vidage progressif du stock-aval, qui, une fois vide conduit à l'arrêt du poste suivant; ce mécanisme de **désamorçage** de la ligne se propage en aval chaque fois qu'un stock-tampon se vide; à chaque nouveau désamorçage, un nouveau poste de travail se trouve au chômage technique;
- une **saturation** de stock-amont qui conduit à un arrêt du poste-amont qui ne peut plus stocker sa production; il y a propagation de la saturation en amont, chaque fois qu'un stock-tampon devient plein; à chaque nouvelle saturation, un nouveau poste de travail se trouve au chômage technique.

FIGURE 165

Exemple de propagation des effets d'une panne



1. Repris de Giard et Pellegrin (1992, [198]).

Il découle de ce qui précède qu'en cas d'arrêt du travail sur l'un des postes de la ligne, l'absence de stocks-tampons conduit inéluctablement à la paralysie immédiate de toute la ligne. On peut ajouter que, dans le cadre d'une organisation en ateliers spécialisés de type *flow shop* (voir [page 51](#) et [page 362](#)), l'arrêt d'un poste de travail se traduit également par une paralysie progressive du système par désamorçage et saturation, dans des conditions voisines de celles décrites ici, à condition de tenir compte du fait que la propagation est retardée d'une part par les plages d'inutilisation des postes de travail liée à l'ordonnancement (voir [page 378](#)) et, d'autre part, par le fait que les commandes peuvent « sauter » certains postes. Dans le cadre d'une organisation en ateliers spécialisés de type *job shop* (voir [page 51](#)), ces mécanismes de paralysie progressive du système par désamorçage et saturation se produisent également mais sont conditionnés par la variété des gammes des commandes lancées en production.

I-3.2 Incidence de la prise en compte immédiate des problèmes de qualité sur une ligne de production

La prise de conscience de l'importance de la gestion de la qualité (qui sera abordée en [section IV, page 971](#), du [chapitre XIV](#)) a conduit les entreprises à passer progressivement du contrôle du produit à celui des processus. Le contrôle statistique des processus peut amener à arrêter un poste de travail pour effectuer des réglages, ce qui conduit à la paralysie progressive de la ligne, évoquée au paragraphe précédent.

Pour certains objets manufacturés complexes, comme les véhicules automobiles (que l'on prendra comme exemple pour illustrer cette classe de problèmes), un contrôle systématique est effectué à certaines étapes de production sur la ligne de production, pour s'assurer du respect de certaines caractéristiques qui jouent sur la sûreté de fonctionnement (opération de ferrage en tôlerie automobile) ou sur certaines caractéristiques techniques auxquelles le client est particulièrement sensible (peinture automobile). Ces contrôles systématiques ne sont pas effectués en fin de ligne mais juste après la dernière opération de transformation du produit aboutissant normalement au respect d'un certain nombre de spécifications (contrôle des caisses peintes à la sortie d'une cabine de peinture). La détection d'un problème de qualité conduit à rectifier les défauts constatés sur le produit avant qu'il ne continue à progresser sur la ligne de production. Souvent, ces retouches mobilisent des équipements spécifiques qui conduisent à retirer de la ligne le produit défectueux pour rectification des défauts constatés, le plus souvent sur une petite ligne dédiée¹ ou un atelier spécialisé. Une fois retouché et jugé conforme, le produit est réinjecté sur la ligne, juste après son point de dérivation (ce qu'illustre l'animation de la [page 603](#)²).

Ce processus de traitement de la qualité conduit à une perturbation du séquençement, appelée décyclage, dont on analysera les conséquences (§ [I-3.2.1](#)), avant de présenter des résultats analytiques permettant de mieux cerner le problème et de mieux évaluer les réponses à apporter (§ [I-3.2.2, page 605](#))³.

1. On peut aussi repartir en arrière sur la ligne si cette rectification implique la même séquence de travail, ce qui est plus rare.

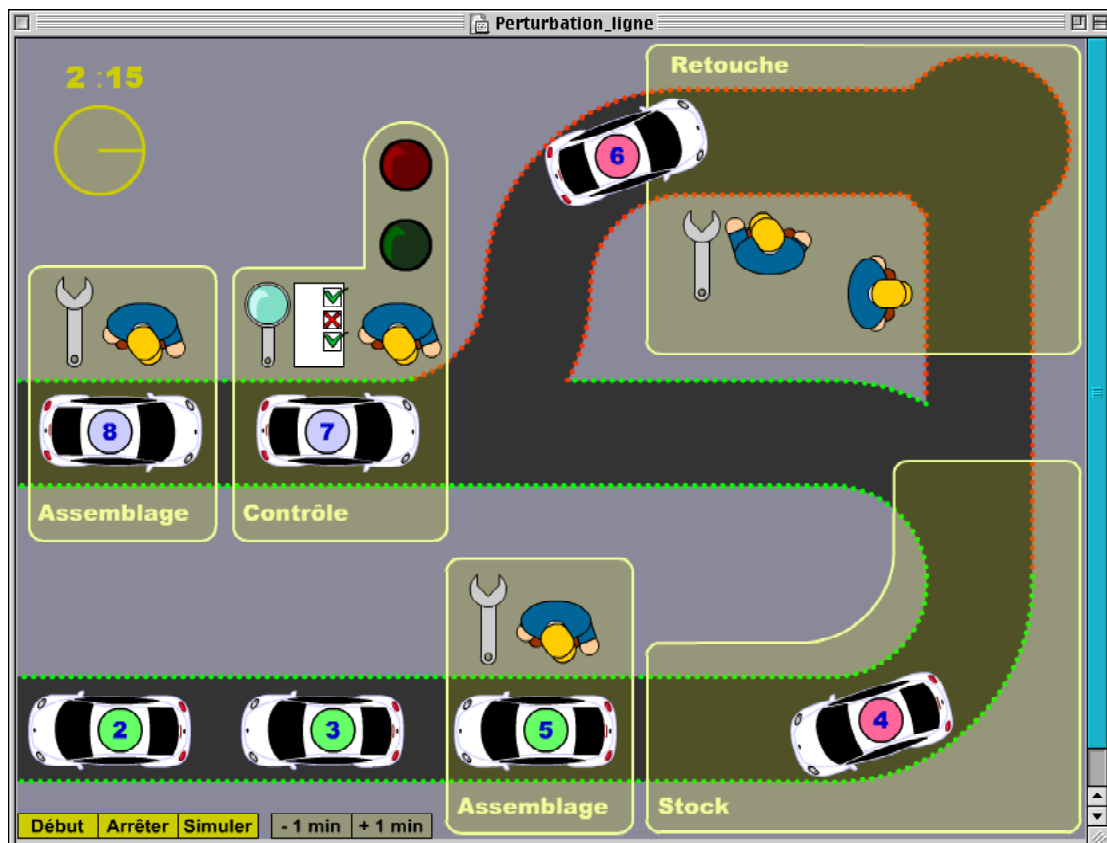
2. Voir avertissement de la note du bas de la [page 8](#).

I-3.2.1 Génération de décyclages lors des retouches en dérivation de la ligne de production

Ce traitement en ligne des problèmes de qualité a deux conséquences, illustrées par le didacticiel associé à la [figure 166](#) : il crée un décyclage et nécessite la présence d'un stock-tampon pour éviter tout désamorçage.

FIGURE 166

Incidence sur le séquençement, des retraits momentanés de produits pour rectification de problèmes de qualité



En effet, le séquençement des produits est perturbé : a priori, un produit qui fait l'objet d'un travail de retouche pendant deux cycles perd deux rangs, tandis que les deux produits suivants en gagnent un. Ce n'est pas si simple puisque chacun des deux produits suivants peut à son tour partir en retouche. Cette perturbation du séquençement, conduisant à des retards ou avances des produits en aval d'un point de contrôle de la qualité pouvant nécessiter des retouches « hors ligne », conduit à un **décyclage** de l'ordonnancement¹. Celui-ci est sans conséquence notable si, dans la phase d'assemble final, le produit est sans variété ou si la variété porte sur des modules interchangeables stockables facilement « en bord de chaîne ». Ceci implique une variété très limitée et un encombrement restreint des modules, pour

3. *Note de la page précédente.* Une analyse plus complète de ces processus peut être trouvée dans l'article de Danjou, Le Roy & Giard (2000, [117]).

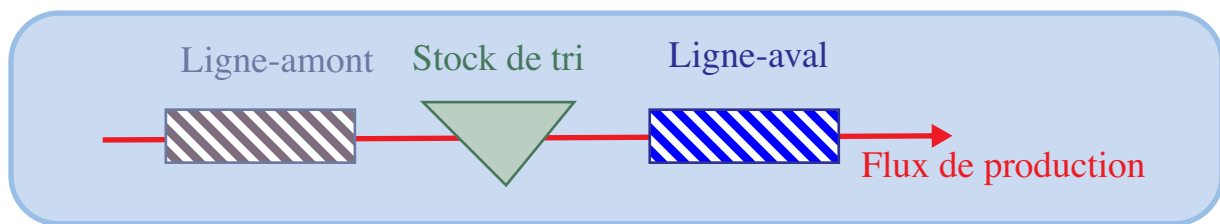
1. On parle de décyclage du film (défini à la [page 509](#)), dans l'industrie automobile. Dans cette métaphore, chaque image du film correspond à un véhicule répondant à une demande précise, caractérisée par une combinaison d'options de personnalisation. Certaines sont à prendre nécessairement dans une liste fermée (de moteurs, de couleurs, etc.), d'autres correspondent à un élément facultatif (toit ouvrant, etc.). La présence de ces dernières fait que l'on est en présence d'un produit hétérogène selon la terminologie définie en [page 589](#).

permettre la création de stocks de sécurité de ces modules interchangeables, montables à ce poste.

La séquence définie dans l'ordre de réquisition transmis au fournisseur n'a aucune chance d'être respectée si cet ordre est adressé avant passage au dernier point de contrôle qualité pouvant conduire à un retrait d'un produit pour mise en conformité; il s'ensuit que l'encyclage n'est pas correct, d'où une certaine désorganisation sur la ligne d'assemblage. Pour limiter l'importance du décyclage et de ses conséquences, certaines entreprises, en particulier celles de l'industrie automobile, positionnent entre la ligne de production et la ligne d'assemblage un stock-tampon particulier, qualifié de **stock de tri** (ou **tri-stock**) qui se caractérise par le remplacement de la règle du «premier arrivé - premier sorti» par d'autres règles (heuristiques) qui tendent principalement à restaurer le séquençement initial¹ et à conserver une certaine pertinence aux ordres de réquisition pilotant les approvisionnements de l'aval et passés aux fournisseurs avant perturbation de l'ordonnancement (figure 167). Ce réordonnancement est glissant, puisque

FIGURE 167

Le système productif



chaque entrée ou sortie modifie la liste des candidats, et doit être résolu en temps réel, en raison de la taille du stock de tri et de la faiblesse du temps de cycle². La résolution de ce problème, rencontré par tous les constructeurs automobiles³, s'appuie sur un ensemble d'heuristiques⁴. Le paramétrage de ces heuristiques est délicat et la robustesse de ce module de pilotage dépend non seulement de l'importance des perturbations (*sensibilité aux aléas*) mais aussi de la structure

Table des
matières

Index
thématique

1. Dans l'industrie automobile, plusieurs types de stocks de tri existent : certains donnent un accès immédiat à tous les véhicules du stock (stock à accès direct), d'autres correspondent à un ensemble contigu de stocks parallèles fonctionnant en mode FIFO (stock à accès partiel, appelés « tiroirs »). Cette dernière configuration est utilisée de manière très différente selon les constructeurs, Renault utilise cette configuration avec stock de tri de 30 tiroirs de profondeur 4, alors que PSA utilise des stocks de tri de 4 tiroirs de profondeur 25 (Castelain et al., [87]); le AS/RS (*automated storage/retrieval system*) de l'usine d'assemblage d'Arlington de la General Motors comporte 192 positions (Feare, 1998, [147]). Dans les deux cas, ce stock intermédiaire permet donc non seulement de découpler les deux sous-systèmes adjacents en absorbant la désynchronisation des flux reçus et émis mais aussi de modifier l'ordre des véhicules sortant du stock, par rapport à celui des véhicules entrant dans le stock, pour rectifier l'incidence des inéluctables perturbations du séquençement initial.
2. Un temps de calcul de n cycles oblige en effet à exécuter l'ordonnancement des n premiers véhicules du réordonnancement; les arbitrages effectués sur les $n - 1$ derniers véhicules de cette liste exécutable ne peuvent pas prendre en compte les véhicules qui rentreront en stock postérieurement au départ du premier véhicule de la liste. Ce biais est sans doute faible si la taille du stock de tri est grande; il ne l'est sûrement pas en cas de stock de tri de faible taille. Dans ce dernier cas, l'arbitrage entre la qualité de la solution et la non prise en compte correcte de l'ensemble des possibles milite en faveur d'un temps de calcul d'un ou de deux cycles. Cela étant, la taille importante de certains stocks de tri à accès direct (cf. Feare, 1998, [87]) change considérablement les possibilités calculatoires; reste cependant à prouver que ce surdimensionnement est économiquement justifié.
3. Voir, par exemple, Castelain et al. [87] pour PSA et [307], p. 270 et sq. pour Toyota.
4. La combinaison d'heuristiques sophistiquées peut même donner naissance à un véritable système-expert (Monden, 1998, [307], p. 273).

des contraintes et de leur degré de saturation pour un ordonnancement donné (*sensibilité au mix des produits*).

Par ailleurs, pour éviter tout désamorçage de la ligne, il est indispensable d'avoir en aval du poste de contrôle de qualité un **stock-tampon** comportant initialement un nombre de produits ayant franchi avec succès cette étape de contrôle, égal au nombre maximal de cycles nécessaire aux rectifications. Dans la simulation, la retouche dure toujours deux minutes, ce qui conduit à un stock tampon initial de deux véhicules (temps de cycle égal à une minute); au cours de la simulation, vous pouvez constater que l'effectif du stock-tampon varie entre 0 et 2. Un temps de retouche aléatoire conduit habituellement à dimensionner le stock-tampon à un niveau inférieur au nombre maximal de cycles nécessaires à une retouche, ce qui revient à accepter une certaine probabilité d'occurrence d'un désamorçage¹.

I-3.2.2 L'approche analytique des conséquences du décyclage

On peut déterminer analytiquement la distribution de probabilité du décyclage (§ I-3.2.2.1) et ses conséquences sur les indicateurs de suivi de production (§ I-3.2.2.2, page 608) et sur les stocks de sécurité (§ I-3.2.2.3, page 609). Les résultats analytiques et exemples présentés dans ce § I-3.2.2, sont tirés de Giard, Danjou, Boctor (2001, [192] ou [193]) et de Giard & Danjou (1999, [194]).

I-3.2.2.1 Détermination de la distribution de probabilité du décyclage

Sous un certain nombre d'hypothèses peu restrictives², on peut déterminer la relation qui lie la distribution de probabilité $P(R = r)$ du nombre R entier, positif ou nul, de places perdues par un véhicule à un poste de contrôle ($0 \leq r \leq \rho$) et la distribution de probabilité³ $P(\Delta = \delta)$ du nombre δ de places gagnées ou perdues par un véhicule quelconque, compte tenu de l'ensemble des retards pris par les véhicules ordonnancés. Le raisonnement permettant d'aboutir à cette distribution (*relation 119, page 606*), en l'absence de transformation de l'ordonnancement par un stock de tri (voir *page 604*), est le suivant.

1. On peut déterminer le niveau optimal S^* de ce stock-tampon en explicitant l'arbitrage entre le coût de possession de l'encours dans ce stock et celui de l'arrêt de la chaîne, ce qui conduit, en première approximation reposant sur l'hypothèse implicite d'un « retour à la normale » entre deux désamorçages (hypothèse réaliste si la probabilité d'un désamorçage est faible), à utiliser un modèle de gestion calendaire, celui du marchand de journaux présenté à la *page 672*. Pour un coût unitaire c_p associé à la détention d'une unité détenue dans le stock tampon et un coût unitaire c_r associé au désamorçage pendant un cycle, le niveau optimal est tel que

$$P(X > S^*) = \frac{c_p}{c_p + c_r}, \text{ où } X \text{ est la durée (exprimée en nombre de cycles) de rectification du problème de qualité.}$$

2. - on est en régime de croisière, c'est-à-dire que l'on ne tient pas compte des problèmes rencontrés à la mise en place ou à l'arrêt du processus de production;
 - la durée de traitement dans les processeurs de la ligne est constante, le temps de cycle de la ligne permettant d'établir la correspondance entre le temps (date), le rang d'un véhicule et sa localisation;
 - la distribution de probabilité du nombre de rangs perdus est indépendante du nombre d'objets en retouche (pas d'accélération ou de ralentissement du rythme de travail en fonction de la charge de travail);
 - les événements « retouche d'un véhicule » sont indépendants, autrement dit le fait qu'un véhicule soit envoyé en retouche ne modifie pas la probabilité de passage en retouche des véhicules suivants;
 - en aval du poste de contrôle, le stock-tampon est suffisant pour éviter tout désamorçage.
3. La représentation de cette distribution est connue sous le nom de **courbe de décyclage**.

D'une manière générale, c'est-à-dire pour une distribution quelconque $P(R = r)$ dans laquelle r peut prendre n'importe quelle valeur entière comprise entre 0 et ρ , le véhicule qui occupe le rang j dans l'ordonnancement se présente devant le poste de contrôle au début de la période j . Après contrôle, il subit un retard r (positif ou nul), qui le fait sortir déclaré « sans défaut » et rentrer dans le stock-tampon à la fin de la période $j + r$, avec le rang $j + r$ si aucun des véhicules qui le précèdent n'est passé derrière lui. Les véhicules susceptibles de passer derrière lui ont nécessairement un rang à l'entrée du poste de contrôle compris entre $h = (j + r) - 1$ et $h = (j + r) - \rho$, que r soit positif ou nul. Dans le cas contraire, son rang sera égal à $j + r - x$, où x est le nombre de véhicules qui le précédaient et qui sont passés après. Une même valeur de décyclage δ peut être obtenue par différentes combinaisons de valeurs prises par r et x , δ pouvant varier entre $-\rho$ (cas où $R = 0$ et où tous les véhicules qui le précèdent partent en retouche) et $+\rho$ (cas où $R = \rho$ et où aucun des véhicules qui lui succèdent ne part en retouche). Sachant que l'on a nécessairement $0 \leq x \leq \rho$ et $0 \leq r \leq \rho$, les différentes combinaisons de valeurs prises par r et x sont données par $x = r - \delta$, pour $\text{Max}(0, \delta) \leq r \leq \text{Min}(\rho, \rho + \delta)$, et $-\rho \leq \delta \leq \rho$. La probabilité du décyclage Δ est alors donnée par la [relation 119](#):

$$P(\Delta = \delta) = \sum_{\text{Max}(0, \delta)}^{\text{Min}(\rho, \rho + \delta)} P(R = r) \cdot P(X = r - \delta), \text{ pour } -\rho \leq \delta \leq \rho \quad \text{relation 119}$$

qui utilise la relation récurrente¹ [120](#):

$$P(X = x) = \frac{N_x}{x!} \cdot \Pi, \text{ où } N_x = \sum_{t=1}^x (-1)^{t-1} \cdot \frac{(x-1)!}{(x-t)!} \cdot s_t \cdot N_{x-t}, \text{ avec } N_0 = 1,$$

$$s_k = \sum_{i=1}^{\rho} \left[\frac{P(R \geq i)}{P(R < i)} \right]^k \text{ et } \Pi = \prod_{i=1}^{\rho} P(R < i) \quad \text{relation 120}$$

On peut illustrer ces relations par un exemple numérique s'appuyant sur une probabilité de retouche $P(R > 0) = p = 10\%$ ou 15% ou 20% et les deux distributions de probabilités conditionnelles (dissymétriques) du [tableau 172](#), ce qui conduit aux distributions de décyclage du [tableau 172](#), illustrées à la [figure 168](#) (le caractère bimodal de ces courbes de décyclage est caractéristique de ce type de courbe, pour des distributions conditionnelles unimodales).

TABLEAU 172
Distributions conditionnelles $P(R = r / R > 0)$ utilisées

Probabilité conditionnelle $P(R = r / R > 0)$	$r = 1$	$r = 2$	$r = 3$	$r = 4$	$r = 5$	$r = 6$	$r = 7$	$r = 8$	$r = 9$	$r = 10$
Dissymétrique «gauche»	10%	29%	22%	14%	10%	5%	4%	3%	2%	1%
Dissymétrique «droite»	1%	2%	3%	4%	5%	10%	14%	22%	29%	10%

1. On a $P(X = 0) = \Pi$, $P(X = 1) = \Pi \cdot N_1$, avec $N_1 = s_1$, $P(X = 2) = \frac{N_2}{2!} \cdot \Pi$, avec $N_2 = (N_1)^2 - s_2$,

$P(X = 3) = \frac{N_3}{3!} \cdot \Pi$, avec $N_3 = (s_1)^3 - 3s_2s_1 + 2s_3$.

FIGURE 168

Distribution de probabilités du décyclage d'un véhicule, pour $p = 10$ et différentes valeurs de probabilité de retouche

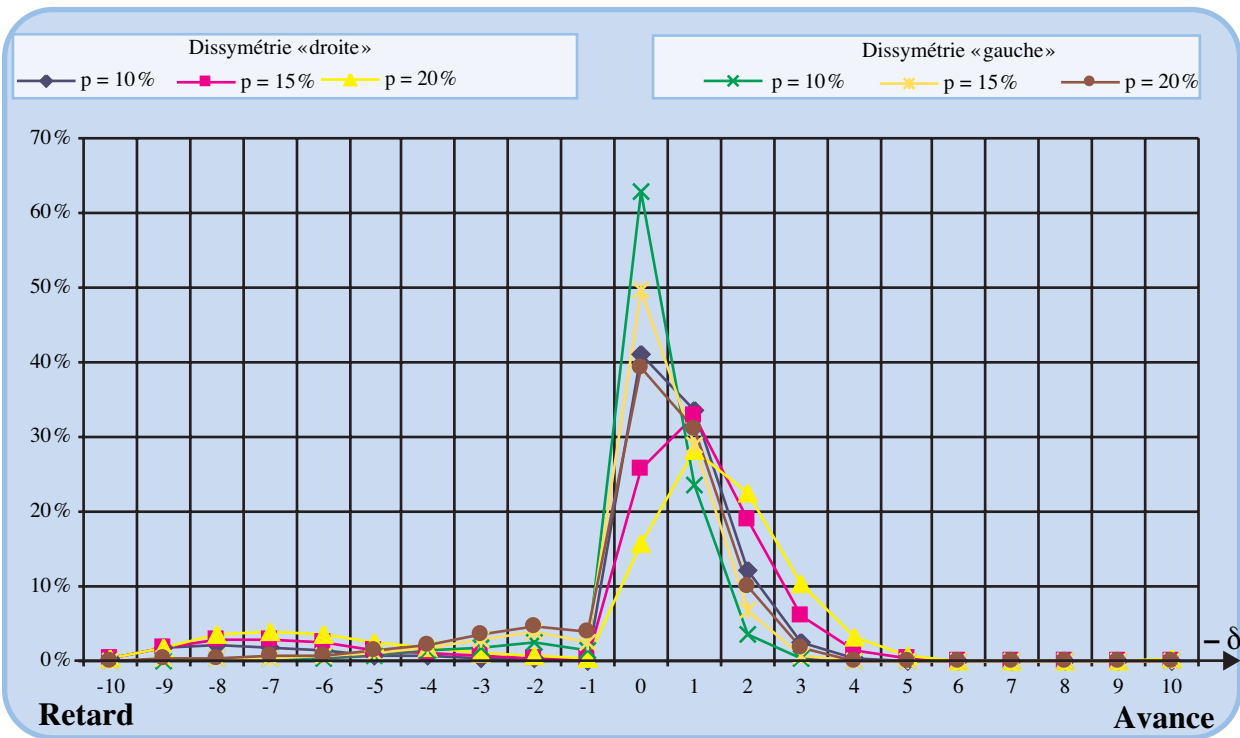


TABLEAU 173

Distribution de probabilités du décyclage d'un véhicule, pour $p = 10$ et différentes valeurs de probabilité de retouche

		Dissymétrie «droite» de la distribution conditionnelle			Dissymétrie «gauche» de la distribution conditionnelle		
Probabilité de retouche p		10%	15%	20%	10%	15%	20%
Avance	- 10	0,46%	0,45%	0,39%	0,07%	0,09%	0,09%
	- 9	1,70%	1,89%	1,82%	0,17%	0,22%	0,26%
	- 8	2,23%	3,01%	3,44%	0,27%	0,37%	0,46%
	- 7	1,89%	2,98%	3,96%	0,37%	0,52%	0,66%
	- 6	1,37%	2,34%	3,43%	0,47%	0,67%	0,86%
	- 5	0,87%	1,58%	2,49%	0,84%	1,17%	1,44%
	- 4	0,56%	1,02%	1,65%	1,26%	1,78%	2,23%
	- 3	0,39%	0,68%	1,08%	1,94%	2,73%	3,41%
	- 2	0,28%	0,48%	0,74%	2,65%	3,79%	4,81%
Retard	- 1	0,18%	0,32%	0,51%	1,55%	2,61%	3,77%
	0	41,16%	25,65%	15,73%	62,99%	49,68%	39,18%
	1	33,74%	33,00%	28,13%	23,63%	28,84%	31,10%
	2	12,24%	18,83%	22,48%	3,53%	6,69%	9,95%
	3	2,58%	6,24%	10,44%	0,27%	0,79%	1,62%
	4	0,35%	1,33%	3,11%	0,01%	0,05%	0,15%
	5	0,03%	0,19%	0,62%	0,00%	0,00%	0,01%
	6	0,00%	0,02%	0,08%	0,00%	0,00%	0,00%
	7	0,00%	0,00%	0,01%	0,00%	0,00%	0,00%
	8	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%
	9	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%
	10	0,00%	0,02%	0,37%	0,00%	0,00%	0,00%

Cette distribution théorique du décyclage est importante en pratique pour plusieurs raisons :

- elle permet de mieux comprendre les perturbations induites par le décyclage sur la pertinence de l'approvisionnement de modules respectant un encyclage établi avant passage sur les postes de la ligne pouvant nécessiter des retouches, puisque cette distribution de probabilité est celle du retard ou de l'avance du produit sur lequel un module spécifique doit être monté ;
- elle permet d'établir une relation de cause à effet entre les améliorations de la qualité (se traduisant par une baisse en moyenne mais, encore plus, en dispersion de la distribution des retards pris dans les retouches) et la diminution des perturbations induite par les décyclages sur la pertinence des approvisionnements respectant un encyclage donné ;
- elle permet de porter un jugement pertinent sur certains indicateurs de performance utilisés dans le suivi en production (§ I-3.2.2.2) et donne des indications sur le niveau du stock de sécurité à constituer pour un module sur la ligne d'assemblage, en aval de ce point de contrôle (§ I-3.2.2.3, page 609).

L'existence d'un stock de tri permettant de modifier un ordonnancement perturbé ne permet plus de relier simplement la distribution des retards en retouche et celle du décyclage, puisqu'interviennent les algorithmes du stock de tri qui se substituent à la règle du « premier entré - premier sorti » (voir page 604). Dans ce cas, ce qui sera dit sur les indicateurs de performance et les stocks de sécurité ne sera pas pertinent.

I-3.2.2.2 Incidence du décyclage sur les indicateurs de suivi de l'ordonnancement

L'un des indicateurs de performance classiquement utilisé dans l'environnement de production en ligne est celui du pourcentage de produits effectivement achevés au cours d'une période donnée, par exemple la journée, par rapport à l'ensemble des produits que l'ordonnancement initial avait désigné comme devant sortir au cours de cette période. En l'absence de transformation de l'ordonnancement par un stock de tri, la relation analytique liant la distribution de probabilités de cet indicateur à celle du nombre de rangs perdus par un véhicule faisant l'objet de retouches, permet de porter un jugement sur la performance obtenue, tenant compte du caractère aléatoire des valeurs prises par cet indicateur ; c'est donc dans une perspective d'étalonnage de tableau de bord que cette relation est établie.

L'indicateur est calculé à la sortie d'un poste d'une ligne (le dernier poste de la ligne ou tout autre poste), pour une fenêtre de temps donnée (la journée, par exemple). Cette période se définit par une durée de H périodes élémentaires définies par le cycle de base de la ligne. Ce poste traite donc H produits au cours de cette période : le premier produit quittant le poste à la fin de la première période élémentaire et le $H^{\text{ème}}$ produit, quittant le poste à la fin de la $H^{\text{ème}}$ période. La liste Ω des produits devant initialement sortir au cours de cette fenêtre de temps diffèrera de la liste Ω' des produits effectivement sortis : seul le sous-ensemble $\Omega \cap \Omega'$ de l'ensemble Ω' , respectant les prévisions. Si l'on note Z , le nombre de produits de Ω' n'appartenant pas à Ω , l'indicateur $(H - Z) / H$ est un bon indicateur du respect de la prévision. Le poste sur lequel la mesure est effectuée étant supposé ne pas subir de perturbations (pannes, ruptures d'approvisionnement, etc.), l'indi-

cateur mesure la performance du sous-système productif amont délimité par le poste ω auquel on s'intéresse et le poste α en amont du poste ω , qui est dernier poste, pour lequel on est certain, d'une part, que l'ordonnancement utilisé pour déterminer la liste Ω est nécessairement respecté et, d'autre part, que ce poste ne peut être à l'origine de problèmes de qualité qui pourront être détectés par l'un des postes de contrôle susceptibles d'envoyer des produits en retouche, postes nécessairement situés entre les postes α et ω ¹. En l'absence de perturbation, le temps qui s'écoule entre le passage d'un véhicule dans le poste α et celui dans le poste ω est constant et peut être important par rapport à la fenêtre de temps retenue. Dans ces conditions, l'indicateur de performance est doté d'une certaine inertie, dans la mesure où il intègre les conséquences d'événements qui se sont produits en dehors de la fenêtre de temps. Par ailleurs, la comparaison au cours du temps d'indicateurs calculés pour des fenêtres de temps de longueurs différentes s'en trouve affectée.

Le nombre de véhicules sortant de la fenêtre de temps est $Z = X + Y$, somme de deux variables aléatoires indépendantes (si $H > 2\rho$), où X est le nombre de «véhicules devant être traité en début de fenêtre de temps et sortant de cette fenêtre» et Y , le nombre de «véhicules devant être traité en fin de fenêtre de temps et sortant de cette fenêtre». Dans le cas général ayant conduit à la [relation 120 de la page 606](#), on démontre la [relation 121](#) qui lie la distribution de probabilité de l'indicateur à celle des retouches. La [figure 169 de la page 610](#) illustre l'utilisation de cette [relation 121](#), en reprenant l'exemple précédent ($\rho = 10$, $p = 20\%$, distribution de probabilité conditionnelle dissymétrique à gauche du [tableau 172 de la page 606](#)), pour une durée $H = 120$.

$$\mathcal{L}\left(\frac{H-Z}{H}\right) = \mathcal{L}(Z) = \Pi \sum_{\text{Max}(0, z-\rho)}^{\text{Min}(\rho, z)} \frac{N_x}{x!} \left[P(R=0) \left\{ \frac{N_{z-x}}{(z-x)!} \right\} + P(R>0) \left\{ \frac{N_{z-x-1}}{(z-x-1)!} \right\} \right]$$

relation 121

I-3.2.2.3 Incidence du décyclage sur les stocks de sécurité des modules à assembler

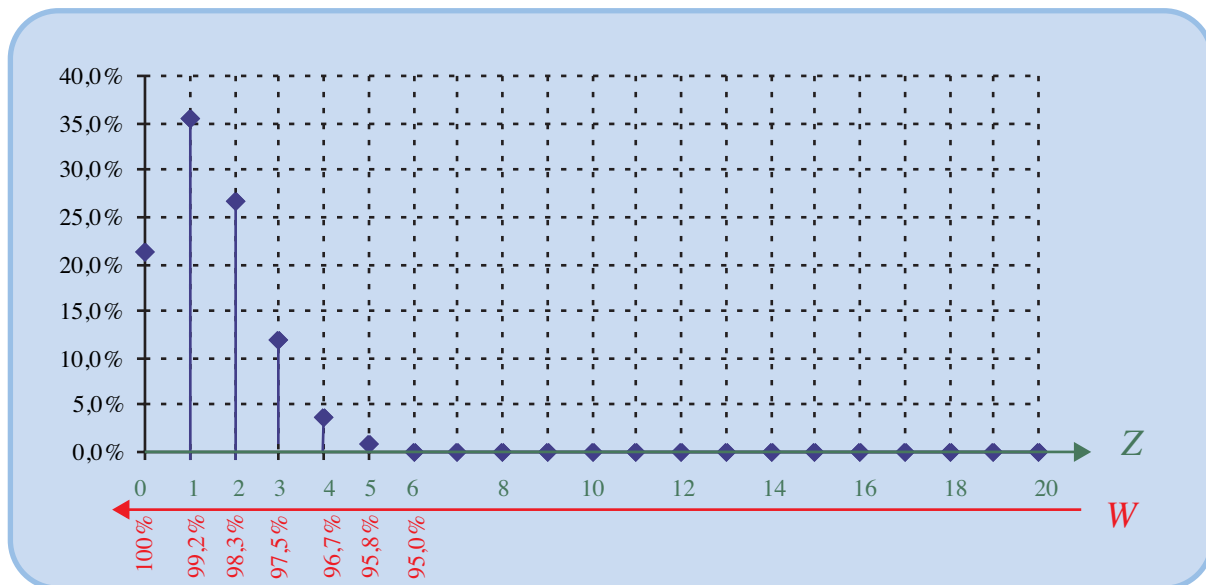
Le problème du stock de sécurité semble relever d'une généralisation de la démarche précédente, dans la mesure où l'on s'intéresse à la demande d'une pièce pendant une fenêtre de temps $\{t_1; t_2\}$ sur un poste donné. Les choses ne sont pas si simples pour deux raisons :

- les véhicules qui rentrent dans la fenêtre de temps (qui sépare deux approvisionnements périodiques) ou en sortent ne sont pas tous consommateurs du composant à gérer (l'approche retenue a banalisé ces véhicules, dans le sens où leur position initiale est une information perdue);
- les échanges de véhicules entre la fenêtre de temps $\{t_1; t_2\}$ et les fenêtres antérieures à t_1 ne sont pas à prendre en compte; en effet, les livraisons cumulées jusqu'à celle effectuée en t_1 sont censées couvrir les demandes allant jusqu'en t_2 ; dans ces conditions, à supposer que le stock de sécurité se

1. En cas de multiplicité des postes de contrôle pouvant conduire à des retouches, on se ramène sans perte de généralité à un poste de contrôle unique caractérisé par une distribution des retards combinant les distributions de retard de chacun des postes de contrôle (problème classique de convolution).

FIGURE 169

Distribution de probabilités de l'indicateur de performance $W = (120 - Z)/120$ pour $\rho = 10$ et $p = 20\%$ et distribution de probabilité dissymétrique à gauche



soit avéré suffisant pour éviter tout désamorçage, les véhicules devant consommer le composant considéré avant t_2 et s'étant présentés avant t_2 sont normalement pris en compte par les livraisons effectuées. Seules les permutations de véhicules autour de cette date t_2 peuvent conduire à un désamorçage à contrer par un dimensionnement approprié du stock de sécurité.

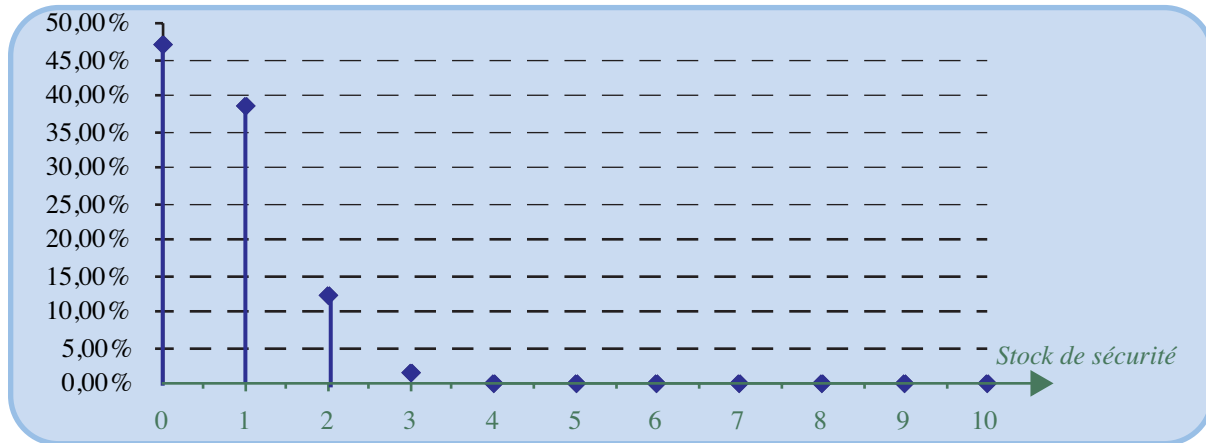
La [relation 120 de la page 606](#) donnant la probabilité que x véhicules devant être traités avant t_2 soient traités postérieurement à cette date fournit les bases d'une solution. Sur les x véhicules qui passent après t_2 , x_1 consomment le composant optionnel tandis que sur les x véhicules qui passent avant t_2 , x_2 consomment ce composant. Si $x_1 > x_2$ aucun problème ne se pose, la consommation effective étant inférieure à la consommation prévisionnelle. Si $x_2 > x_1$ il est nécessaire d'avoir un stock de sécurité de $x_2 - x_1$ pour éviter tout désamorçage. Dans les pires conditions, on aura $x_2 = x$ et $x_1 = 0$, ce qui conduit à dimensionner le stock de sécurité à x qui constitue une borne supérieure. La [figure 170, page 611](#), donne la distribution de probabilité du stock de sécurité **maximum** pour l'exemple précédent ($\rho = 10$, $p = 20\%$, distribution de probabilité conditionnelle dissymétrique à gauche du [tableau 172 de la page 606](#)). On constate qu'un stock de sécurité de 3 unités suffit à faire face aux aléas, tant que les perturbations restent de caractéristiques stables. Les hypothèses que l'on pourra formuler sur le fait que le composant optionnel est ou non consommé de façon régulière joue sur le niveau de compensation entre les consommations du composant par les véhicules qui font l'objet de la permutation étudiée.

On notera que le *niveau du stock de sécurité est indépendant* de l'intervalle de temps séparant deux livraisons, et donc *du niveau de la demande* couverte par cette livraison, contrairement à ce qui se passe dans le cas d'un approvisionnement destiné à couvrir des demandes aléatoires indépendantes¹. Ce stock de sécurité dépend avant tout de l'importance des perturbations de l'ordonnancement initial

qui résultent du niveau de qualité en production et de l'organisation mise en place pour les retouches jugées nécessaires.

FIGURE 170

Distribution de probabilité du stock de sécurité maximal pour $\rho = 10$ et $p_0 = 20\%$ et distribution de probabilité dissymétrique à droite



SECTION II LIGNES DÉDIÉES À LA FABRICATION DE PRODUITS HÉTÉROGÈNES

Comme dans la section précédente, on examinera d'abord le problème de la conception d'une ligne de production ou d'assemblage de produits hétérogènes (§ II-1) puis celui de son utilisation dans son aspect d'ordonnancement (§ II-2, page 614).

Table des
matières

II-1 Conception d'une ligne de production ou d'assemblage de produits hétérogènes

Index
thématique

La conception de ce type de ligne est beaucoup plus complexe car, par définition (page 589), l'hétérogénéité conduit à ce que le temps opératoire de certains postes varie en fonction du type de produits (indice p). Plusieurs voies, non exclusives, peuvent être explorées :

- Le travail supplémentaire à effectuer sur certaines opérations peut être réalisé par du personnel en **renfort**, pour faire en sorte que le temps de cycle soit respecté, auquel cas, on se ramène au problème précédent, sur la base du travail minimal à effectuer sur ce poste. La gestion de ce renfort de personnel peut être difficile, car il doit être partagé entre plusieurs postes susceptibles de connaître des surcharges épisodiques.

1. *Note de la page précédente.* Dans les différents modèles du chapitre XI, on montre que le stock de sécurité, pour une probabilité de rupture donnée et dans le cas d'une demande suivant une loi Normale, dépend de l'écart-type de la demande. On sait par ailleurs que la variance de la demande d'une somme de demandes aléatoires indépendantes (portant sur des périodes consécutives) est égale à la somme des variances de ces demandes indépendantes ; il s'ensuit que le stock de sécurité croît quand la période calendaire croît. Ceci reste vrai lorsque les variables aléatoires ne sont pas normales. L'une des caractéristiques du problème traité ici est que, contrairement à ceux classiques de gestion des stocks, les demandes d'un composant sur une période ne peuvent respecter l'hypothèse d'indépendance des demandes élémentaires puisque le retard de n rang d'un véhicule implique l'avance de 1 rang des $n - 1$ véhicules qui le suivent. On est là en présence d'une problématique méconnue par la littérature spécialisée du domaine. La recherche du « zéro stock » se traduit donc par la création d'une nouvelle problématique de gestion des stocks.

- On peut faire suivre un poste sur lequel le travail d'une opération varie en fonction du type de produits, d'un **stock-tampon** d'une taille suffisante pour éviter les phénomènes de saturation et de désamorçage. Dans ce cas, il faut que, dans la conception de la ligne, le travail moyen associé à cette opération soit défini au-dessus de la charge moyenne pour non seulement permettre de compenser les surcharges de travail par des sous-charges de travail mais encore de pouvoir faire face au problème posé, par une succession de produits impliquant une surcharge de travail. Ce stock-tampon ne pouvant être extensible, il est évident que l'ordonnancement ne peut ignorer totalement ce problème d'espacement entre produits demandant plus de travail que les autres sur certains postes.
- Lors de la conception de la ligne, on peut faire l'hypothèse que l'on trouvera toujours un ordonnancement évitant l'arrêt de la ligne par saturation ou désamorçage, et travailler à partir de charges moyennes. Dans cette perspective, on peut limiter, voire faire disparaître, le stock-tampon. C'est ce type d'approche, impliquant une bonne prise en compte des problèmes d'ordonnancement au stade de la conception de la ligne, que l'on privilégie de nos jours.

Lorsque l'opérateur se déplace sur la ligne avec le produit, l'amplitude de ses déplacements varie selon la quantité de travail à effectuer. Dans l'animation de la [figure 171 de la page 613](#), la ligne travaille avec un temps de cycle de 60 secondes et effectue des opérations d'assemblage sur des véhicules comportant ou non un toit ouvrant. Sur le second poste de cette ligne, le travail à effectuer sur un véhicule à toit ouvrant est de 90 secondes, contre 45 secondes pour un véhicule n'en comportant pas. Le dépassement de 30 secondes se traduit par un déplacement additionnel de l'opérateur; il est compensé si au moins les deux véhicules qui suivent un véhicule à toit ouvrant n'ont pas de toit ouvrant, comme l'illustre l'animation dans laquelle il a été choisi de ne pas mettre de stock-tampon en aval de ce poste à temps de travail variable. On peut ajouter que si les arrivées dans chaque station sont cadencées à la minute, celles du poste 3 s'effectuent 90 secondes après celles du poste 2 (ce qui implique une attente des véhicules sans toit ouvrant devant la troisième station).

La conception de la ligne dépend alors étroitement de la structure de la demande devant être traitée par la ligne (α_p , étant la part que représente le produit p dans la demande totale en volume).

La formulation du problème est une adaptation de celle proposée au [§ I-1.2, page 594](#). Une approche possible consiste à formuler le problème en travaillant sur une période donnée¹ mais les résultats peuvent varier en fonction de la période. Il est aussi simple de reprendre le principe des nomenclatures de planification (voir [page 465](#)) et de remplacer le temps opératoire de i qui varie maintenant en fonction du produit p (t_{ip}) par sa moyenne pondérée de ces temps opératoires,

1. Voir Nof et al. 1997, [318], p. 233 & sq. et Scholl, 1999, [373], p. 79 & sq.

FIGURE 171

Espacement entre produits hétérogènes permettant la compensation de la charge de travail sur un poste à durée de travail variable selon les produits

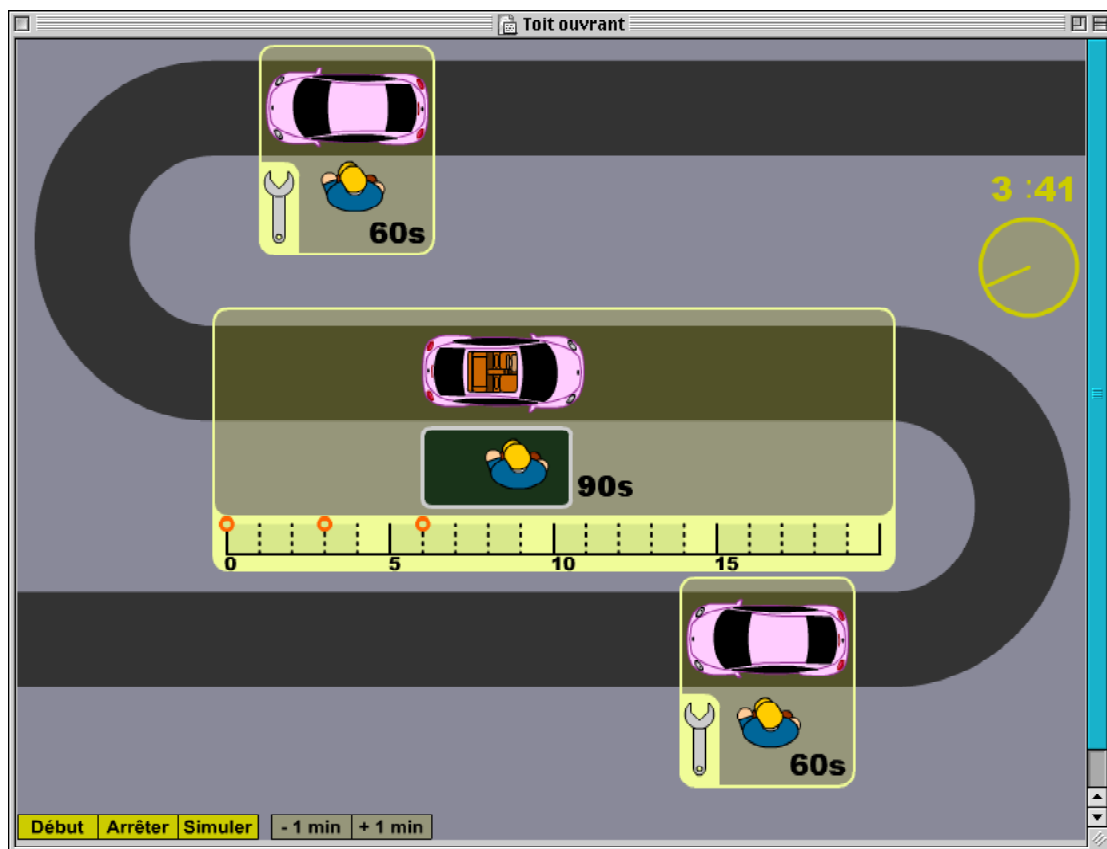


Table des matières

Index thématique

$t_i = \sum_p \alpha_p t_{ip}$, ce qui revient à remplacer $(\sum_i t_i x_{ik})$ dans la relation 114 de la page 595 par $(\sum_i (\sum_p \alpha_p t_{ip}) x_{ik})$ et conduit à la relation 122.

$$\sum_i (\sum_p \alpha_p t_{ip}) x_{ik} \leq c, \text{ pour } k = 1, \dots, K \quad \text{relation 122}$$

Au remplacement près de la relation 114 par la relation 122, les deux formulations du § I-1.2, page 594 sont inchangées. La pertinence de cette adaptation est conditionnée par le respect de deux hypothèses : l'absence de variation de la composition de la demande (stabilité des α_p) et l'existence d'ordonnancements permettant d'espacer suffisamment les produits à surcharge de travail.

Les hypothèses de travail formulées lors de la conception d'une ligne, c'est-à-dire plusieurs mois avant sa mise en service, sont rarement respectées. Pour limiter les risques de saturation et de désamorçage qu'implique le non-respect de ces hypothèses, on est nécessairement amené :

- à prendre des marges (par surestimation de la part de produits impliquant une surcharge de travail, dans la demande et donc à sous-estimer celle des autres, afin de respecter la contrainte $\sum_p \alpha_p = 1$) qui génèrent une sous-utilisation des postes à contenu de travail variable ; en tout état de cause, cette saturation totale de ces contraintes est difficilement envisageable¹ en raison de leurs

interdépendances (un même véhicule peut être concerné par plusieurs contraintes) et de l'existence de perturbations conduisant à une modification du séquençement initial;

- à introduire un stock-tampon minimal qui peut avoir aussi pour objet de retarder la nécessité de redéfinir la ligne, en cas de dérive dans la structure de la demande;
- à prendre des options techniques, dans la conception de la ligne, qui en permettent une redéfinition rapide (boucles de rétroaction de la [figure 162](#)), ce qui implique des choix laissant une certaine marge de manœuvre et généralement plus onéreux;
- à fiabiliser le système productif pour réduire le décyclage et son incidence (on reviendra sur ce point au [§ II-3, page 619](#));
- à porter une grande attention aux techniques d'ordonnancement.

II-2 Ordonnancement sur ligne de production ou d'assemblage de produits hétérogènes

Dans le cas d'une fabrication / assemblage de produits hétérogènes sur une ligne, l'ordonnancement (appelé parfois **engagement** ou **programme de production**) est plus complexe que si l'on a affaire à des produits homogènes, dans la mesure où la variabilité du temps de travail sur certaines stations conduit à respecter un certain espacement entre les produits (comme on l'a illustré avec l'animation de la [figure 171, page 613](#)). Pour rendre plus concret notre raisonnement, nous nous appuyerons encore sur l'exemple de la construction automobile.

Dans de nombreuses industries travaillant à la commande, sur lignes de production ou d'assemblage, on dispose de carnets de commandes sur plusieurs jours de travail (dont une partie importante peut ne pas correspondre à des commandes de client final identifié) et il est évident que l'on ne fait pas d'ordonnancement sur la totalité des commandes. Le travail s'effectue en deux temps.

- On procède d'abord à une répartition globale des commandes en les affectant à des journées de production, en tenant compte de la date de livraison et de leurs caractéristiques. Cette procédure a été qualifiée d'**ordonnancement agrégé** dans la [figure 164 de la page 598](#); aucun séquençement des commandes n'est effectué à ce niveau. L'ensemble affecté à chacune des journées à venir tient compte de critères commerciaux mais aussi de critères de production, puisqu'il s'agit d'éviter que le problème d'ordonnancement détaillé ne soit trop difficile à résoudre et/ou trop coûteux.
- Seul l'ensemble du premier jour fait ensuite l'objet d'un **ordonnancement détaillé**. C'est à ce type d'ordonnancement que nous nous intéresserons dans ce [§ II-2](#).

Le jour suivant, l'ordonnancement agrégé est adapté pour tenir compte de nouvelles commandes, dans le cadre de l'horizon retenu pour cet ordonnancement. Dans la mesure du possible, on cherche à ne pas trop perturber la répartition des tout premiers jours car les informations correspondantes auront été transmises aux fournisseurs. Par exemple, le fournisseur de sièges de voiture est avisé du

1. Toutes les entreprises du secteur automobile gardent d'importantes marges de sécurité pour les raisons évoquées ici (Rachamadugu & Yano, 1994, [\[355\]](#)).

nombre prévisionnel de housses d'un certain type qui seront montées sur des sièges à livrer dans 3 jours mais, à ce moment-là, les informations permettant de définir l'encyclage ne sont pas encore disponibles.

L'ordonnancement doit tenir compte de contraintes techniques et de considérations économiques.

- Les contraintes techniques d'ordonnancement résultent de la variabilité du temps opératoire de certains postes de la ligne de montage, en raison d'un travail spécifique lié à une option à monter (par exemple, présence d'un toit ouvrant). Pour ces postes, qualifiés de **critiques en variabilité de temps opératoires**, un véhicule doté d'une option qui conduit à une quantité de travail, supérieure au temps de cycle, doit être suivi de véhicules ayant une quantité de travail inférieure au temps de cycle, jusqu'à ce que la surcharge de travail du premier véhicule soit compensée par le cumul des marges inutilisées par les véhicules suivants (ce qui a été illustré avec l'animation de la [figure 171, page 613](#)). Certaines contraintes sont techniquement intangibles tandis que d'autres sont levables au prix de renforts momentanés en personnel, sur les postes de la ligne réalisant ce travail supplémentaire optionnel.
- Sur certains postes de la ligne de production, qualifiés de **postes critiques en séquençement**, le changement d'une option de personnalisation du véhicule (couleur dans un poste de peinture, par exemple), conduit à supporter un coût de réglage. Par exemple, le type d'options «couleur du véhicule» peut comporter quelques dizaines d'options (les options d'un même type d'options étant nécessairement exclusives); si le nouveau véhicule arrivant en cabine de peinture a l'option «bleue» du type d'option «peinture» alors que la séquence (plus ou moins importante) de véhicules précédents avait l'option «rouge», il faut procéder au nettoyage de pistolet en cabine de peinture. Pour des raisons économiques, on cherche à limiter cette opération qui correspond à un temps de lancement, éventuellement masqué, fonction du séquençement retenu (voir [page 388](#)). Le temps opératoire n'est pas affecté par cet ordre de passage.

Plusieurs formulations de cet ordonnancement détaillé sont possibles et correspondent à des configurations différentes¹. Celle que nous retiendrons (adaptée de Giard, 1997, [\[184\]](#)) ne traite que les lignes cadencées par un temps de cycle pré-déterminé, l'arrivée des produits dans chaque station s'effectuant avec la même cadence (ce qui n'est pas forcément vrai pour leurs sorties), et dans lesquelles l'espace de travail de l'opérateur sur une station n'empiète pas sur celui des stations adjacentes.

La détermination de l'ordonnancement de la production d'une journée (ou une demi-journée) peut faire appel à la programmation mathématique. Cette formali-

1. Scholl (1999, [\[373\]](#), p. 94 et sq.) ainsi que Nof, Wilhelm & Warnecke (1997, [\[318\]](#), p. 340 et sq.) présentent des modèles orientés vers le traitement de lignes non cadencées, ainsi qu'un modèle proposé par Yano et Rachamadugu (1991, [\[355\]](#)) concernant l'ordonnancement de lignes cadencées et intégrant la détermination des temps de renforts nécessaires pour passer outre des contraintes d'espacement; cette approche s'appuie sur une description des déplacements des opérateurs mais celle des renforts, entre les stations qui les sollicitent, ne sont pas pris en compte. Une autre formulation est tournée vers la définition d'une séquence-type lançant au moins une fois tous les types de produits et qui est susceptible d'être dupliquée au cours de la journée (*Minimum Part Set*); cette approche est pertinente lorsque la variété est faible.

sation permet de bien se mettre d'accord sur le problème à résoudre et facilite, si nécessaire, la mise au point d'heuristiques permettant de résoudre un problème d'ordonnancement, habituellement de grande dimension, en un temps acceptable.

Soit x_{ij} une variable binaire valant 1 si le véhicule i ($i = 1, \dots, N$) a le rang j . Un seul véhicule i peut avoir le rang j , ce que traduit la [relation 123](#):

$$\sum_{i=1}^N x_{ij} = 1, \text{ pour } j = 1, \dots, N \quad \text{relation 123}$$

et le véhicule i se voit attribuer nécessairement un seul rang j , ce que traduit la [relation 124](#):

$$\sum_{j=1}^N x_{ij} = 1, \text{ pour } i = 1, \dots, N \quad \text{relation 124}$$

On note k l'un des postes de la ligne d'assemblage, en ne nous intéressant en fait qu'aux seuls postes affectés par cette variabilité de la charge de travail (sous-ensemble \mathcal{K}). La durée du travail requis par le véhicule i sur le poste de travail k est notée θ_{ik} et le temps de cycle est noté $\bar{\theta}$, l'arrivée des véhicules sur ce poste étant cadencée par ce temps de cycle. Si l'on note h , le rang du véhicule positionné sur le poste de travail k (avec $h = 1, \dots, N$), le temps de travail à exécuter sur ce

véhicule de rang h est $\sum_{i=1}^N \theta_{ik} x_{ih}$. L'écart entre cette durée et le temps de cycle

$\sum_{i=1}^N \theta_{ik} x_{ih} - \bar{\theta}$ correspond à un dépassement s'il est positif et, dans le cas contraire,

à une éventuelle possibilité de rattrapage d'une surcharge de travail antérieure.

Notons R_{kh} l'excédent de charge de travail à résorber après traitement du véhicule de rang h ($R_{k0} = 0$), sachant que ce report est limité par une quantité de travail $R_{k, \max}$, liée au déplacement maximum du véhicule sur le convoyeur (comme illustré par l'animation de la [figure 171, page 613](#)). Ce report, qui ne peut être négatif, intègre, si elle existe, la charge de travail non résorbée après traitement du véhicule de rang $h - 1$. Il s'ensuit que cet excédent est :

$$R_{kh} = \text{Max} \left\{ 0, \text{Min} \left[R_{k, h-1} + \sum_{i=1}^N \theta_{ik} x_{ih} - \bar{\theta}, R_{k, \max} \right] \right\}, \text{ ce que l'on traduit par la}$$

[relation 125](#).

$$W_{kh} \leq R_{k, \max}; W_{kh} \leq R_{k, h-1} + \sum_{i=1}^N \theta_{ik} x_{ih} - \bar{\theta}; R_{kh} \geq 0; R_{kh} \geq W_{kh}$$

$$\text{pour } h = 1, \dots, N \text{ et } k \in K \quad \text{relation 125}$$

Dans cette formulation, on suppose que ce poste k est indépendant des postes adjacents, en ce sens qu'aucune compensation sur les temps opératoires n'est possible. Ceci implique qu'en régime de croisière, le poste $k + 1$ commencera¹

avec un décalage, par rapport au poste k , égal à $\bar{\theta}$, augmenté du plus grand écart possible $\theta_{ik} - \bar{\theta}$.

Cette formulation comporte $4N$ contraintes pour chaque poste $k \in K$, ce qui conduit à chercher à limiter la taille N de l'ensemble de véhicules à ordonnancer et le nombre K de postes critiques. Si plusieurs postes ont des contraintes « voisines », il suffit alors de ne s'intéresser qu'au plus pénalisant. Ce nombre élevé de contraintes pour un poste critique s'explique par la très grande variété postulée de temps opératoires sur ce poste qui oblige à vérifier que le temps opératoire de chacun des véhicules ne compromet pas un éventuel rattrapage de retards imputables à un ou plusieurs véhicules de rang inférieur. Le nombre de contraintes diminue très fortement si la variété du nombre de temps opératoire se limite à 2 que l'on notera $\tau_{k_{max}}$ (montage d'une option) et $\tau_{k_{min}}$ (pas de montage d'une option). Dans ce cas, l'espacement minimal τ_k entre deux véhicules ayant le temps opératoire $\tau_{k_{max}}$ est fourni par la [relation 126](#)¹:

$$\tau_k = (\tau_{k_{max}} - \bar{\theta}) / (\bar{\theta} - \tau_{k_{min}}) \quad \text{relation 126}$$

Bien évidemment τ_k doit être arrondi à la valeur immédiatement supérieure, en cas de résultat non entier. On peut alors remplacer la [relation 125 de la page 616](#) par la [relation 127](#) qui fait passer le nombre de contraintes d'espacement de $4N$ à $N - \tau_k$, pour $k \in K$.

$$\sum_{j=h}^{h+\tau_k} \sum_{i=1}^N \theta_{ik} x_{ij} < h \cdot \bar{\theta}, \text{ pour } h = 1, \dots, N - \tau_k \text{ et } k \in K \quad \text{relation 127}$$

Cette dernière solution amène deux remarques:

- Si la variété des temps opératoires est supérieure à deux pour un poste critique, on peut éventuellement se ramener à ce cas en retenant comme temps opératoire de toutes les options le temps le plus élevé $\tau_{k_{max}}$, ce qui permet de limiter la taille du problème mais peut être gênant si l'on s'approche des contraintes utilisées lors de l'équilibrage de la ligne.
- Les contraintes d'espacement peuvent être plus complexes. Si la zone de déplacement de l'opérateur permet de traiter deux véhicules en dépassement du temps de cycle, il faut adapter la [relation 126](#), pour exploiter une capacité de rattrapage accrue, comme cela est illustré dans l'animation de la [figure](#)

[171, page 613](#): $\tau_k = \frac{2(\tau_{k_{max}} - \bar{\theta})}{\bar{\theta} - \tau_{k_{min}}} + 1$, la [relation 127](#) restant inchangée. L'expres-

1. Note de la page précédente. C'est la solution retenue dans la simulation. On peut également faire commencer le travail sur la station $k+1$ aussitôt achevé le travail sur la station k (décalage θ_{ik}) mais se pose alors la question de savoir s'il faut alors propager ces variations, c'est-à-dire à faire commencer le travail sur la station $k+2$ avec un décalage $\theta_{ik} + \bar{\theta}$, etc. La solution la moins perturbante est celle retenue ici.

1. Exemple numérique: $\bar{\theta} = 60'$, $\tau_{k_{max}} = 75'$, $\tau_{k_{min}} = 55'$, $\tau_k = \frac{75 - 60}{60 - 55} = 3$, ce qui signifie que, pour rattraper les 15 secondes de dépassement dues au montage de l'option, il faut faire suivre ce véhicule de 3 véhicules sans option.

sion formelle de cette contrainte devient, pour l'exemple retenu, «pas plus de 2 véhicules avec option, sur tout ensemble de 6 véhicules consécutifs».

Il faut maintenant introduire une fonction-objectif pour achever une formulation exploitable en programmation linéaire. L'introduction d'un point de vue économique ne doit prendre en compte que ce qui est lié à l'ordonnancement retenu. Celui-ci influe sur des coûts associés à des opérations (réglages, nettoyages...) réalisées sur certains postes de travail en raison de variations de spécifications des véhicules qui se suivent (par exemple, nettoyage de pistolets nécessités par un changement de couleur de peinture). D'autres coûts associés à l'ordonnancement peuvent être également utilisés comme, par exemple, les coûts liés à l'irrégularité de certains besoins en composants qui perturbe l'organisation de l'approvisionnement retenu. On se contentera ici de la première famille de coûts évoquée, sachant que des extensions sont possibles mais qu'elles risquent vite de buter sur la disponibilité de données fiables et pertinentes. On se limitera ici au cas d'un coût de réglage c_s lié au seul nombre de réglages effectué sur un poste s concerné par ce type de problème (sous-ensemble \mathcal{S}). Ceci implique que ce coût soit à peu près indépendant de la séquence de réglages associée à l'ordonnancement¹. Notons y_{js} la variable dichotomique valant 1 si le véhicule de rang j a un réglage différent de celui du véhicule précédent ($j-1$) sur le poste s . Dans ces conditions, la minimisation des coûts de lancement sur le sous-ensemble \mathcal{S} de postes affectés par la séquence retenue de véhicules est :

$$\text{Min} \left(\sum_{s \in \mathcal{S}} c_s \sum_{j=2}^N y_{js} \right) \quad \text{relation 128}$$

Il faut maintenant forcer la variable y_{js} à prendre la valeur 1 lorsqu'il y a changement de réglages sur le poste s . Notons β_{is} le numéro de réglage utilisé par le véhicule i sur le poste s . Le numéro de réglage du poste s pris par le véhicule de rang j dans l'ordonnancement est $\sum_{i=1}^N \beta_{is} \cdot x_{ij}$. Il s'ensuit que changement de

réglage entre le véhicule $j-1$ et le véhicule j arrive lorsque la différence $\sum_{i=1}^N \beta_{is} \cdot x_{ij} - \sum_{i=1}^N \beta_{is} \cdot x_{i,j-1}$ n'est pas nulle. Pour forcer y_{js} à prendre la valeur 1 dans ce cas, il suffit d'appliquer les principes décrits au [tableau 350 de la page 1137](#) qui se traduisent ici par la [relation 129](#) :

$$-Ny_j \leq \sum_{i=1}^N \beta_{is} \cdot x_{ij} - \sum_{i=1}^N \beta_{is} \cdot x_{i,j-1} \leq Ny_{js}, \text{ pour } j=2, \dots, N \text{ et } s \in \mathcal{S} \quad \text{relation 129}$$

1. Il est évident que le coût de nettoyage nécessaire pour passer du noir au blanc est supérieur à celui pour passer du noir au blanc. Ce problème d'ordonnancement (partiel) a été abordé au § I-2, page 388 du [chapitre V](#) et son inclusion dans le problème traité ici ne pose pas de problème particulier. Cela étant, pour un nombre suffisamment élevé de réglages de changement de peinture, l'usage d'un coût standard judicieusement calculé doit donner, après multiplication par le nombre de réglages, une valeur pas trop éloignée de celle obtenue avec l'usage de coûts dépendant de la séquence de couleurs. On peut ajouter qu'il n'est pas certain qu'un raffinement du modèle soit économiquement judicieux, le gain marginal obtenu risquant fort d'être inférieur au supplément de coût de recherche d'une solution, induit par l'accroissement de taille du modèle.

qui conduit à l'effet recherché parce que la fonction-objectif (relation 128) tend à rendre nuls le plus possible de y_{js} et que les y_{js} de la relation 129 ne peuvent être nuls que si numéro de réglage du poste s ne change pas et que dans les autres cas (second terme strictement positif ou négatif), y_j est nécessairement égal à 1.

Pour terminer, on peut prendre en considération la gestion de renforts qu'il faut introduire si le séquençement ne respecte pas l'espacement minimal induit par cette variabilité de (pris en compte à la relation 127, page 617). On peut imaginer, dans une formulation simplifiée, de mettre un renfort sur le poste k pendant le traitement du véhicule h , ce qui se traduit par un accroissement Z_{hk} du temps d'opérateurs disponible sur le poste k , qui permet de traiter tout ou partie de l'excédent

$R_{k,h-1} + \sum_{i=1}^N \theta_{ik} x_{ih} - \bar{\theta}$, si cette valeur est positive. Il suffit alors de remplacer cette expression par $R_{k,h-1} + \sum_{i=1}^N \theta_{ik} x_{ih} - \bar{\theta} - Z_{hk}$. Il s'ensuit que cet excédent

devient : $R_{kh} = \text{Max} \left\{ 0, \text{Min} \left[R_{k,h-1} + \sum_{i=1}^N \theta_{ik} x_{ih} - \bar{\theta} - Z_{hk}, R_{k,\max} \right] \right\}$, ce qui conduit à la relation 130 :

$$W_{kh} \leq R_{k,\max}; W_{kh} \leq R_{k,h-1} + \sum_{i=1}^N \theta_{ik} x_{ih} - \bar{\theta} - Z_{hk}; R_{kh} \geq 0; R_{kh} \geq W_{kh}$$

pour $h = 1, \dots, N$ et $k \in K$

relation 130

Le coût de ces renforts est alors intégré dans la fonction-objectif de la relation 128, ce qui conduit à la relation 131, dans laquelle c_k est le coût d'une unité de temps de renfort sur le poste k :

$$\text{Min} \left(\sum_{s \in S} c_s \sum_{j=2}^N y_{js} + \sum_{k \in K} c_k \sum_{h=1}^N Z_{hk} \right)$$

relation 131

Dans cette formulation, la gestion de ces renforts n'est pas prise en compte mais peut l'être en adaptant la démarche décrite au § I-3.1.2, page 558, du chapitre VIII, pour tenir compte de problèmes de transports entre les stations.

II-3 Organisation de la réactivité aux aléas de production

Les mécanismes de propagation du désamorçage et de la saturation d'une ligne, décrits au § I-3, page 600, restent valables dans le cas d'une ligne dédiée à la production de produits hétérogènes.

On a vu que, pour les produits homogènes, la reprise au plus tôt des problèmes de qualité conduit à un décyclage qui remet en cause la pertinence de l'approvisionnement de certains modules. Ceci reste vrai pour les produits hétérogènes mais s'ajoute en outre le risque du non-respect de contraintes d'espacement, qui provoque saturation et désamorçage. Pour limiter ce risque, on peut augmenter l'espacement requis entre certains produits, en tenant compte de la connaissance que l'on a des distributions de décyclage mais ceci implique une structure de la demande moins contraignante en ordonnancement que celle utilisée dans la

conception de la ligne. On peut également utiliser un stock de tri (page 604) avec des algorithmes privilégiant les contraintes d'espacement lorsque la restauration de l'ordonnancement initial n'est pas possible.

Pour terminer, il faut évoquer une autre approche de l'ordonnancement, s'appuyant sur le principe de différenciation retardée et utilisée par certains constructeurs automobiles (comme PSA), qui permet aussi de limiter l'impact du décyclage. Rappelons que, dans la métaphore du film utilisé dans cette industrie pour décrire le séquençage initial des véhicules, un film est une suite ordonnée d'images dans laquelle chaque image du film correspond à un véhicule répondant à une demande précise, caractérisée par une combinaison d'options de personnalisation, certaines sont à prendre nécessairement dans une liste fermée (de moteurs, de couleurs, etc.), d'autres correspondent à un élément facultatif (toit ouvrant, etc.). Dans la conception du film reposant sur le principe de différenciation retardée (évoquée ci-dessus), on définit le film lancé pour une journée par une suite d'images inachevées, les détails retenus dans ces images correspondant à des caractéristiques partagées par tous les véhicules, plus des caractéristiques de différenciation d'une première partie de la ligne (éventuellement délimitée par un stock de tri), en s'efforçant de respecter les contraintes des postes critiques en variabilité de temps opératoires de cette partie de ligne. À ce point, sur un film inévitablement perturbé, on complète les images en ajoutant de nouvelles caractéristiques de différenciation utilisées sur la seconde partie de la ligne, en s'efforçant toujours de respecter les contraintes des postes critiques en variabilité de temps opératoires de cette partie de ligne et, bien sûr, en tenant compte des caractéristiques de la demande à satisfaire pour la journée; ce processus se poursuit tant que toutes les images ne sont pas complétées. Cette approche est connue sous le nom de **baptême progressif**.