

# **RECHERCHE OPERATIONNELLE ET GESTION DE LA PRODUCTION**

**Yves CRAMA<sup>1</sup>, Lionel DUPONT<sup>2</sup> et Gerd FINKE<sup>3</sup>**

<sup>1</sup> Ecole d'Administration des Affaires  
Université de Liège  
Boulevard du Rectorat 7 (B31)  
4000 Liège Belgique

<sup>2</sup> Ecole Nationale Supérieure de Génie Industriel  
46 Avenue Félix Viallet  
38031 Grenoble Cedex France

<sup>3</sup> Laboratoire Leibniz - Institut IMAG  
46 Avenue Félix Viallet  
38031 Grenoble Cedex France

Juin 1997

Article préparé pour publication dans la revue  
*Nouvelles de la Science et des Technologies*

## Introduction

Depuis leurs origines, la recherche opérationnelle et la gestion de la production entretiennent des relations privilégiées. Très tôt, les problèmes de gestion industrielle ont intéressé les mathématiciens: citons les écrits de Monge relatifs à l'organisation efficace des transports de déblais et de remblais ou l'ouvrage de Kantorovich sur les « méthodes mathématiques dans l'organisation et la planification de la production », publié en 1939, ouvrage qui contenait en germe certains des développements fulgurants de la programmation linéaire enregistrés dans les décennies suivantes. Sur le terrain, de nombreuses méthodes issues de ces travaux ont été appliquées de manière effective. Citons pour mémoire:

- la gestion des approvisionnements par la quantité économique de commande décrite par Harris (1913) dans son analyse des arbitrages fondamentaux de la gestion des stocks;
- la gestion de projet par les méthodes PERT/CPM (années 50);
- les nombreuses applications industrielles de la programmation linéaire dès les années 50 dans le secteur de l'industrie pétrolière (planification de la production et de la distribution), agro-alimentaire (transport, mélanges), sidérurgique (modélisation de processus), mécanique (ordonnancement);
- la planification des besoins en composants (MRP1 puis MRP2).

La gestion de la production et les domaines connexes de la logistique industrielle ont donc rapidement supplanté la gestion des opérations militaires comme champ d'action principal et, simultanément, comme catalyseur du développement de la recherche opérationnelle.

La gestion industrielle a considérablement évolué depuis les années 1960. Schématiquement, on peut dire que l'on est passé du modèle Ford au modèle Toyota et d'une économie de production à une économie de marché. Les principales causes de cette évolution sont:

- le renversement du rapport entre l'offre et la demande: pour la plupart des produits, les capacités de production actuelles dépassent très largement ce que le marché peut absorber;
- la mondialisation de l'économie;
- l'accélération du renouvellement des produits.

Pour survivre, les entreprises ont dû se refocaliser sur ce qui constitue le coeur de l'activité industrielle: la création de valeur ajoutée. Pour ce faire, elles s'efforcent de supprimer toutes les activités superflues par la recherche de production en juste à temps, ou de production au plus juste, et par un recentrage sur les concepts de qualité, de coûts et de délais. La recherche opérationnelle continue à jouer un rôle actif dans cette croisade. Il suffit pour s'en convaincre de consulter la liste des finalistes de la « Franz Edelman Award » (le prix de la société américaine de recherche opérationnelle – INFORMS – qui couronne chaque année une implantation particulièrement réussie et significative de méthodes quantitatives en milieu industriel) et de constater que les applications en planification de la production, gestion des stocks, distribution et logistique se taillent la part du lion dans cette liste (voir par exemple Assad, Lilien et Wasil 1992 et la revue *Interfaces*).

Par nécessité, on se contentera d'offrir ici un très bref aperçu des applications de la recherche opérationnelle en gestion de la production. L'exposé sera organisé autour de trois thèmes principaux: planification de la production et programmation mathématique, gestion des stocks et conception de systèmes. A titre d'illustration, certains modèles fondamentaux et

particulièrement représentatifs seront décrits de façon détaillée. En général, cependant, le portrait esquissé permettra seulement de mettre en évidence quelques-uns des succès ou difficultés rencontrés dans les différents domaines, ainsi que certaines pistes de développement récemment explorées. Pour une couverture plus complète du sujet, le lecteur pourra consulter les références mentionnées en annexe. En particulier, l'ouvrage édité par Graves, Rinnooy Kan et Zipkin (1993) fournit un panorama très large de l'état de l'art sur le plan théorique. Pour un aperçu de la mise en pratique de ces théories en milieu industriel, on renverra au recueil édité par Assad, Wasil et Lilien (1992), à la revue *Interfaces* et à la section « OR Practice » de la revue *Operations Research*.

## **Planification de la production et programmation mathématique**

Considérons le cas d'un constructeur automobile et supposons que la demande commerciale lui impose le montage de 500 voitures le lendemain. Pour y arriver, il va falloir que les pièces nécessaires soient en stock, que la chaîne de montage soit cadencée pour une production de 500 voitures/jour, et que la main d'oeuvre requise soit présente. Ceci suppose que de nombreuses décisions aient été prises dans les semaines, voire les mois précédents, en un mot que l'on ait planifié cette production du lendemain.

Le besoin de planification existe dans la plupart des entreprises et ce pour deux raisons principales:

1. le délai de livraison acceptable par le client est généralement plus court que le délai de fabrication, ce qui oblige les entreprises à anticiper la demande pour prendre les décisions relatives aux approvisionnements ou au lancement de tout ou partie de la fabrication;
2. les capacités d'adaptation des entreprises vis à vis des variations du marché sont limitées. Bien qu'elles cherchent à augmenter leur flexibilité (on parle aussi d'agilité) afin d'améliorer leur réactivité, les entreprises restent des structures lourdes, ce qui les amène à produire certains articles de manière anticipée.

Une bonne planification doit essayer de concilier au mieux des objectifs souvent contradictoires:

- assurer que le client trouve le produit qu'il désire au moment et à l'endroit voulus,
- fournir à la production proprement dite (ateliers, opérateurs, etc.) les conditions de travail les plus correctes possibles, ce qui demande au minimum qu'à tout moment les ressources nécessaires à la production soient présentes, qu'il s'agisse des équipements, des ressources humaines ou des matières;
- assurer une production au moindre coût ou de rentabilité maximale.

Ces objectifs sont assez flous. Dans de nombreux cas, les deux premiers objectifs seront traduits en contraintes quantitatives: on fixera la demande à satisfaire au cours de chaque période et on imposera que les besoins requis pour la production prévue soient couverts par les capacités. On se ramène ainsi à des modèles de minimisation de coût sous contraintes de demande et de capacité.

Avant de mettre en place un système de planification, une première question se pose: sur quelles données faut-il travailler? Si nous fabriquons des voitures du modèle X, en version 3

ou 5 portes, avec 5 motorisations et 20 couleurs possibles, jusqu'à quel degré de détail faut-il descendre dans la planification? En pratique, la caractéristique couleur peut être écartée d'emblée: on peut en effet estimer que ce détail se règle dans la semaine précédant la fabrication. La motorisation elle, est une caractéristique importante; mais si le délai de fabrication d'un moteur est de 3 mois, est-il intéressant d'en tenir compte pour la planification à un an? Peut-on s'en passer pour la planification à 3 mois? Même sur ce petit exemple, on constate que plusieurs questions se posent: quel horizon de temps faut-il considérer, quels produits faut-il prendre en compte, faut-il estimer la capacité de production au niveau de l'usine, des ateliers ou des machines? La réponse la plus habituelle est de mettre en place un système hiérarchique correspondant:

- aux décisions stratégiques sur le long terme,
- aux décisions tactiques sur le moyen terme,
- aux décisions opérationnelles sur le court terme.

Le terme hiérarchique signifie qu'un niveau travaille impérativement dans le cadre défini par le ou les niveaux supérieurs. Parallèlement, le degré de finesse de l'information prise en compte ne sera pas le même. Plus le niveau est bas, et plus l'information retenue sera fine; plus le niveau est élevé, plus l'information requise sera agrégée.

Dans ce cadre, les modèles de programmation mathématique occupent une place importante dans l'arsenal des techniques de planification de production, en particulier pour l'établissement de plans à moyen ou à court terme.

### **Planification à moyen terme**

La planification à moyen terme prend place dans un cadre de décision où le portefeuille de produits et le processus de production peuvent être considérés comme des données irrévocables, fixés par la stratégie de l'entreprise. Les questions qui se posent à ce niveau de décision tactique portent donc sur l'utilisation optimale du système productif dans le but de satisfaire la demande prévisionnelle sur l'horizon du plan.

A ce niveau de décision, il règne trop d'incertitudes pour s'encombrer de détails qui compliqueraient inutilement la prise de décision. Le gestionnaire cherche donc habituellement à planifier « dans les grandes lignes ». Les produits sont agrégés par familles, les ressources en grandes catégories (de personnel, d'équipements, etc.), les périodes sont relativement longues (typiquement, un mois). On retient essentiellement les contraintes sur les ressources critiques (approvisionnements limités, équipements goulets). Pour cette raison, on appelle parfois plan agrégé le plan de production à moyen terme. L'agrégation des décisions permet de simplifier considérablement la formulation, la résolution et l'interprétation des modèles (moins de données à collecter, moins de calculs à effectuer, moins de résultats à analyser). Par ailleurs, elle augmente également la qualité des prévisions de demande ainsi que l'estimation d'autres paramètres.

Les variables contrôlées par le planificateur sont essentiellement de deux types: les variables attachées aux produits et les variables attachées aux ressources. Pour comprendre ceci, considérons une entreprise ayant à fabriquer  $N$  familles de produits (indiqués par  $i$ ) sur un horizon de  $T$  périodes (indiquées par  $j$ ). La fabrication de ces produits utilise  $M$  ressources

(indicées par k: matières, équipements, main d'oeuvre). La demande  $d_{i,j}$  du produit i pour la période j peut être satisfaite de différentes façons: par la production  $X_{i,j}$  du produit i effectuée au cours de cette période j, par le stock  $S_{i,j}$  du produit i disponible en début de la période j, ou par le report sur les périodes ultérieures d'une partie  $P_{i,j}$  de la demande (livraisons différées). En combinant ces diverses possibilités, la firme peut niveler son niveau d'activité et donc lisser l'utilisation des ressources. Les équations de passage d'une période j-1 à la période j sont alors de la forme:

$$(X_{i,j} + S_{i,j}) - (d_{i,j} + P_{i,j-1}) = S_{i,j+1} - P_{i,j}.$$

Ce dont on dispose (fabrication plus stock:  $X_{i,j} + S_{i,j}$ ) moins la demande totale (la demande de la période augmentée de la demande différée:  $d_{i,j} + P_{i,j-1}$ ) donne le stock initial de la période suivante ( $S_{i,j+1}$ ) ou la demande différée ( $-P_{i,j}$ ).

A ce niveau de planification, on peut considérer que les besoins sont proportionnels aux niveaux de production. Dans le cas le plus simple, la disponibilité  $b_{k,j}$  de la ressource k pour la période j est connue. Soit  $a_{i,k}$  la quantité de k nécessaire pour produire une unité de i. L'équation exprimant le fait que le besoin doit rester inférieur ou égal à la capacité disponible s'écrit alors:

$$\sum_i a_{i,k} X_{i,j} \leq b_{k,j}.$$

Lorsqu'il est possible de stocker la ressource k, en posant  $R_{k,j}$  le stock de k disponible en début de période j, l'équation devient:

$$\sum_i a_{i,k} X_{i,j} + R_{k,j+1} = b_{k,j} + R_{k,j}.$$

Il est parfois possible d'ajuster le niveau de certaines ressources: on peut recourir à une seconde source pour les approvisionnements, sous-traiter des opérations lorsque les équipements sont sous-capacitaires, ou faire varier le niveau de main d'oeuvre (heures supplémentaires, chômage technique, intérim, flexibilité du temps de travail, embauche, licenciement). Ces ajustements peuvent avoir leur portée limitée a priori par des contraintes techniques (capacité du sous-traitant), légales (taux maximum d'heures supplémentaires) ou par d'autres considérations (politique de maintien de l'emploi, conventions sectorielles, etc.). Ceci implique l'addition de nouvelles variables de décision aux modèles, ainsi que de nouvelles contraintes linéaires. Vu la diversité des situations possibles, il n'est pas possible de donner des équations génériques applicables à tous les cas. A titre d'illustration, prenons deux exemples où nous supposons que la ressource k représente des heures de main d'oeuvre.

**Exemple 1.** Il est possible de faire des heures supplémentaires dans la limite de 10% des heures normales. Soit  $Y_{k,j}$  les heures supplémentaires effectuées en période j:

$$\begin{aligned} \sum_i a_{i,k} X_{i,j} &\leq b_{k,j} + Y_{k,j} \\ Y_{k,j} &\leq 0,10 b_{k,j} \end{aligned}$$

**Exemple 2.** La variation de main d'oeuvre d'une période à l'autre se fait par embauche ou licenciement. Mais les variations sont limitées à 10%.

Posons  $M_{k,j}$  le niveau de la main d'oeuvre en période  $j$ ,  $E_{k,j}$  et  $L_{k,j}$  les embauches et licenciements en début de période  $j$ . On a:

$$\begin{aligned} \sum_i a_{i,k} X_{i,j} &\leq M_{k,j} \\ M_{k,j} &= M_{k,j-1} + E_{k,j} - L_{k,j} \\ 0,9 M_{k,j-1} &\leq M_{k,j} \leq 1,1 M_{k,j-1}. \end{aligned}$$

Dans ces modèles, l'objectif sera de minimiser l'ensemble des coûts: production, stockage, sous traitance, main d'oeuvre, etc. Ces coûts sont le plus souvent linéarisés, ce qui conduit à la formulation de modèles de programmation linéaire. La souplesse de cette modélisation, alliée aux autres avantages bien connus de la programmation linéaire (disponibilité des logiciels, efficacité des calculs, analyse de sensibilité) font de la programmation linéaire un instrument de choix pour la planification de la production agrégée.

### **Planification à court terme**

Le plan de production à court terme peut être vu comme une version détaillée et désagrégée du plan de production à moyen terme: il désagrège les familles de produits en articles individuels et il planifie la production sur un horizon plus court en tenant compte d'une découpe plus fine en sous-périodes.

Une des caractéristiques essentielles du plan à court terme est la prise en compte explicite des particularités du processus de production: nomenclature des produits (matières, composants, sous-assemblages), enchaînement et durée des opérations aux différents centres de production (gammes opératoires, délais de livraison, ...), temps de préparation et de lancement des opérations (nettoyage des installations, réglage des machines, changements d'outils, ...), etc.

De nombreux modèles d'optimisation ont été proposés pour l'élaboration de plans de production à court terme. La formulation de ces modèles dépend évidemment, dans une certaine mesure, du processus technologique considéré. On retrouve ainsi des modèles distincts selon que l'on analyse des industries de process ou d'assemblage. Dans la plupart des cas, cependant, l'existence des temps de lancement (set-ups) introduit des discontinuités ou non linéarités qui se traduisent par l'apparition de variables de décision binaires dans les formulations. Les modèles multi-produits multi-niveaux qui en résultent sont donc généralement des problèmes de programmation linéaire mixte de grande taille, souvent difficiles à résoudre de façon exacte.

Un des modèles les plus simples et les mieux connus dans cette catégorie est celui de détermination de tailles de lots optimales proposé par Wagner et Whitin. Ce modèle considère un seul article pour lequel on connaît la demande  $d_j$  (besoin net) à la période  $j$ . Le coût de production d'un lot de  $x$  articles est exprimé par la fonction  $f(x)$  et le coût de possession de  $y$  articles pendant une période est donné par  $g(y)$ . On fait l'hypothèse que les fonctions  $f(x)$  et  $g(y)$  sont concaves, ce qui permet de modéliser d'éventuelles économies d'échelle. Le

problème est de déterminer un plan de production satisfaisant la demande de chaque période et de coût total minimum.

Sous les hypothèses énoncées, on peut montrer que parmi les solutions optimales, il en existe toujours une pour laquelle la production  $X_j$  de la période  $j$  est ou bien nulle ou bien égale à la demande cumulée des périodes  $j, j+1, \dots, j+k$  pour un certain  $k \geq 0$ . Il suffit donc de se restreindre à la recherche des solutions présentant la structure décrite par ce résultat. Cette structure peut se traduire par le graphe suivant.

- Un sommet  $V_j$  représente le fait qu'on lance la production à la période  $j$ . Un sommet  $V_{T+1}$  est ajouté pour borner l'horizon de planification.
- Un arc de  $j$  à  $j+k+1$  représente la situation où l'on lance la production en  $j$  et à nouveau en  $j+k+1$ , mais pas entre ces deux périodes. Dans ce cas, la production en  $j$  doit permettre de satisfaire la demande des périodes  $j$  à  $j+k$ . La valeur de cet arc est par conséquent posée égale à:  $f(d_j + \dots + d_{j+k}) + g(d_{j+1}) + 2g(d_{j+2}) + \dots + kg(d_{j+k})$ .

Le problème consiste alors à déterminer un chemin du sommet  $V_1$  au sommet  $V_{T+1}$  dont la valeur totale soit la plus faible possible. Il s'ensuit que la solution optimale du modèle de Wagner-Whitin peut être facilement calculée par un algorithme de plus court chemin (programmation dynamique). Des modèles plus généraux sont dérivés assez naturellement du modèle de base pour tenir compte de produits multiples, des relations entre ces produits (relations traduites dans les nomenclatures), de contraintes de capacité sur les différents centres de production, etc. Comme mentionné plus haut, cependant, les modèles ainsi formulés deviennent rapidement complexes, ce qui justifie les approches de décomposition hiérarchique souvent utilisées pour les attaquer.

## Ordonnancement

Dans le très court terme, le département production se trouve confronté à une collection d'ordres de fabrication (OF) à exécuter. Chaque OF consiste en une liste d'opérations à effectuer, mais ne précise pas nécessairement l'ordre dans lequel ces opérations doivent être exécutées, ni l'instant auquel elles doivent être entamées, ni les postes de production auxquels elles doivent être affectées. Ordonnancer la production, c'est planifier les dates de lancement et de fin des opérations ainsi que l'affectation de ces opérations aux différents postes de production susceptibles de les exécuter. On classe généralement la fonction d'ordonnancement parmi les activités de planification à très court terme, quoique dans le cas de l'ordonnancement de projets elle puisse parfois tenir compte d'un horizon de plusieurs années.

L'ordonnancement est une branche de l'optimisation combinatoire et la plupart de ses problèmes peuvent être modélisés sous la forme de programmes linéaires en variables entières (nous ignorerons ici les aspects stochastiques qui viennent parfois compliquer les modèles de base). Pour illustrer cette approche et ses limites, considérons  $n$  OFs (pièces) 1, 2, ...,  $n$  de durée opératoire  $p_1, p_2, \dots, p_n$  qui doivent être ordonnancés sur une machine unique. Les seules variables de décision sont les dates de lancement  $s_1, s_2, \dots, s_n$  de chacun des OFs sur la

machine. Un grand nombre de contraintes sont modélisables, indépendamment du critère retenu:

- l'OF n'est pas disponible avant la date  $r_j$ :  $s_j \geq r_j$
- l'OF doit être achevé pour la date  $d_j$ :  $s_j + p_j \leq d_j$
- l'OF j doit précéder l'OF i:  $s_j + p_j \leq s_i$ .

Lorsqu'il n'y a pas de contraintes sur l'ordre de passage des OFs i et j, il faut également formuler l'alternative:

soit i est lancé avant j et la contrainte doit être:  $s_i + p_i \leq s_j$

soit j est lancé avant i et la contrainte doit être:  $s_j + p_j \leq s_i$ .

Cette alternative peut être modélisée en introduisant une variable  $y_{ij}$  de valeur 0-1 (avec l'interprétation que  $y_{ij}$  vaut 1 si i est lancé avant j) et les deux contraintes:

$$p_i + s_i \leq M(1-y_{ij}) + s_j$$

$$p_j + s_j \leq My_{ij} + s_i$$

où M est une constante suffisamment grande.

Les modèles de programmation linéaire en nombres entiers ainsi formulés sont souvent trop complexes pour pouvoir être résolus directement par des algorithmes génériques. La théorie de l'ordonnement s'est donc développée comme une branche individuelle de la recherche opérationnelle, avec ses propres méthodologies et outils mathématiques. Aussi ses domaines d'application débordent-ils souvent le cadre de la productique (ils empiètent en particulier largement sur l'informatique).

La littérature sur les problèmes d'ordonnement est vaste et variée. D'un point de vue théorique, elle est consacrée pour une bonne part à l'analyse de la complexité algorithmique de ces problèmes. Cet effort s'est organisé autour d'une classification des problèmes articulée sur plusieurs critères.

1. *L'environnement de production*: on distingue ainsi les situations où chaque OF requiert une seule opération de celles où chaque OF doit subir plusieurs opérations. Dans le premier cas, une ou plusieurs machines parallèles peuvent être disponibles. Dans le second cas, les machines peuvent être organisées en ligne de production (flow shop) ou en atelier (job shop).

2. *Les caractéristiques des OFs et des opérations*: opérations morcelables ou non, existence éventuelle de relations de précedence entre opérations, de dates de mise à disposition et de dates d'exigibilité pour les OFs, temps de lancement des opérations, etc.

3. *Le critère de performance*: les critères utilisés pour juger de la qualité d'un plan d'ordonnement peuvent être nombreux et contradictoires. Les critères purement économiques (coût, profit) sont mal adaptés dans le cadre du très court terme. On utilise donc plutôt des critères d'efficacité organisationnelle: respect des délais (nombre d'OFs en retard  $U$ , retard moyen  $L$ , plus grand retard  $L_{max}$ ), productivité (date d'achèvement de la dernière opération  $C_{max}$ , date d'achèvement moyen des opérations  $C$ , temps de cycle), taux d'utilisation des machines, etc.

Les premiers travaux théoriques sur ce type de modèles ont pu établir la solvabilité polynomiale de quelques-uns des modèles parmi les plus simples:

- deux machines en ligne, OFs disponibles en début d'horizon, critère  $C_{max}$  (Johnson 1954);

- machine unique, OFs disponibles en début d’horizon, critère  $L_{max}$ : règle de la date d’exigibilité minimum (EDD; Jackson 1955);
- machine unique, OFs disponibles en début d’horizon, critère  $C$  (éventuellement pondéré): règle du temps opératoire minimum (SPT) généralisée (Smith 1956);
- machine unique, OFs disponibles en début d’horizon, critère  $U$  (Moore 1968);
- deux machines, chaque OF doit subir une opération sur chacune des deux machines dans un ordre quelconque, critère  $C_{max}$  (Gonzalez et Sahni 1976).

On pourrait ajouter à cette liste la méthode de chemin critique rendue populaire dans les années 50 par le développement des modèles PERT/CPM et leur application à l’ordonnancement de projet: il s’agit en fait d’un algorithme de minimisation du critère  $C_{max}$  dans une situation impliquant un nombre infini de machines, une seule opération par OF et des relations de précedence entre opérations.

Rapidement, cependant, il est apparu que la plupart des modèles d’ordonnancement sont difficiles à résoudre de façon exacte. Une illustration typique est fournie par la minimisation de  $C_{max}$  sur une machine unique lorsque le passage d’une opération  $i$  à une opération  $j$  requiert un temps de réglage  $s_{ij}$ : en effet, il est aisé de vérifier que ce problème est équivalent au célèbre problème du voyageur de commerce. En termes de complexité algorithmique, ceci implique qu’il est *NP-difficile*, et donc peu susceptible de pouvoir être résolu efficacement. Dans les années 70, un nombre impressionnant de modèles d’ordonnancement sont venus enrichir la liste des problèmes connus pour être NP-difficiles.

Notons toutefois que dans le contexte industriel, il est parfois possible de trouver la solution optimale du problème rencontré, bien qu’il soit en théorie NP-difficile. La complexité est une propriété asymptotique qui ne prend tout son sens que dans des situations impliquant un nombre élevé d’OFs et/ou de machines. Si de telles situations sont courantes dans la pratique, elles ne constituent cependant pas nécessairement la règle, loin s’en faut. L’exemple (réel) suivant souligne ce point. Une ligne de production standard produit 1500 boîtes de boisson en aluminium par minute. Les temps de transition  $s_{ij}$  (par exemple, entre la production d’une boîte de 0,33 litre et celle d’une boîte de 0,5 litre) peuvent exiger jusqu’à 4 jours. En conséquence, les boîtes sont produites en lots importants et, sur une période de production de 15 jours, on ne produit pas plus de deux ou trois catégories de boîtes différentes. Le problème du voyageur de commerce correspondant est donc trivial.

Les efforts de recherche sur les modèles fondamentaux de l’ordonnancement s’orientent actuellement dans plusieurs directions distinctes:

- le développement d’algorithmes d’optimisation exacts, principalement de type branch-and-bound ou branch-and-cut; ces algorithmes n’autorisent la résolution que de problèmes d’assez petite taille;
- le développement d’heuristiques, fréquemment basées sur les principes généraux des métaheuristiques de type recuit simulé, exploration tabou, algorithmes génétiques, etc; ces approches permettent d’obtenir des solutions pour des problèmes de très grande taille, mais la qualité des ordonnancements ainsi calculés ne peut généralement pas être garantie;
- l’analyse de la performance théorique de certaines heuristiques simples: l’idée est ici de démontrer qu’une heuristique particulière fournit toujours une solution dont la valeur ne s’éloigne de la valeur optimale que de  $R$  % au plus, dans le pire des cas (idéalement,  $R$  est

une constante absolue et de faible valeur); il s'agit là d'un domaine de recherche extrêmement actif à l'heure actuelle.

Nous avons uniquement parlé ici de l'ordonnancement des OFs sur les machines. Bien que les machines constituent évidemment une ressource essentielle, d'autres ressources interviennent aussi dans les processus de production: moyens de manutention (ponts roulants, chariots filoguidés, grues), outillage ou outils (en particulier dans les centres d'usinage automatisés), opérateurs spécialisés, etc. Ces ressources additionnelles doivent être disponibles pour que les opérations puissent s'effectuer. Leur considération complexifie encore les modèles considérés.

Notons enfin que diverses perturbations se produisent en permanence dans les ateliers: pannes, retards, commandes urgentes, etc. L'ordonnancement en temps réel consiste à contrôler l'exécution des opérations selon l'ordonnancement planifié en prévoyant des réactions rapides en cas de perturbation.

D'un point de vue pratique, de nombreuses études ont établi le bénéfice que pouvaient retirer les entreprises d'une approche formalisée et optimisée de leur processus d'ordonnancement. Il reste, cependant, que les logiciels d'ordonnancement commerciaux n'intègrent généralement qu'une très faible partie des connaissances théoriques disponibles dans ce domaine. Les ordonnancements implantés en atelier sont souvent basés sur des règles de dispatching simples, qui permettent de décider de la prochaine opération à exécuter sur chaque poste de travail en ne tenant compte que des caractéristiques des OFs disponibles à ce poste (et donc, en négligeant les interactions entre ces OFs et les autres OFs en cours dans l'atelier, ainsi que les interactions entre postes de travail). Le problème d'ordonnancement (ou de pilotage) est ainsi ramené, pour l'essentiel, à une série de problèmes d'ordonnancement à une machine. Quelques règles de dispatching souvent utilisées en pratique sont les suivantes: parmi les opérations prêtes à être exécutées sur le poste considéré, sélectionner

- la première opération arrivée à ce poste (premier arrivé premier servi, FIFO), ou
- l'opération de temps opératoire le plus faible (SPT), ou
- l'OF dont la date d'exigibilité est la plus précoce (EDD), etc.

Dans le cas d'ateliers fortement automatisés ou pilotés par ordinateur, il est possible de disposer d'une vue générale de l'atelier et des règles plus élaborées peuvent être utilisées: par exemple, sélectionner d'abord l'OF qui, par la suite, ira sur la machine la moins chargée. Mais fondamentalement, ces méthodes gèrent des priorités et restent sous-optimales.

## **Gestion des stocks**

La gestion des stocks est indissociable de la problématique de la planification de la production et aurait donc pu être traitée dans la section précédente. Les questions de détermination de taille de lots évoquées dans le cadre de la planification à court terme, par exemple, sont clairement du ressort de la gestion des stocks. Du point de vue du chercheur opérationnel, cependant, la gestion des stocks (et plus particulièrement des approvisionnements) a développé un ensemble de problèmes, de modèles et de méthodes qui lui sont propres et qui en font un champ d'investigation séparément identifiable.

L'objectif de la gestion des stocks est d'assurer la disponibilité des articles considérés sur un horizon donné tout en minimisant les coûts encourus. Trois types de coûts sont généralement pris (plus ou moins) explicitement en compte pour évaluer la qualité d'une politique de gestion des stocks: coûts de commande, coûts de possession et coûts de rupture de stock. Dans la plupart des modèles classiques, ces coûts sont considérés comme des paramètres fixes. Par contre, les approches plus récentes, telles celles développées dans le cadre de la gestion Juste-A-Temps ou de la gestion de la Qualité Totale, envisagent les coûts de stockage comme des variables dont la valeur peut (doit) être influencée par des investissements appropriés (visant, par exemple, la réduction des temps de lancement). Par ailleurs, certains auteurs (et de nombreux gestionnaires) observent que les coûts énumérés peuvent être difficilement mesurables. Ceci est vrai, en particulier, pour les coûts de rupture de stock (ou pénurie). Pour cette raison, le risque de rupture est fréquemment modélisé à travers une contrainte de maintien d'un niveau de service prédéterminé (probabilité de rupture, nombre attendu de ruptures, temps écoulé entre ruptures successives, taux d'articles fournis en temps et en heure etc.).

Diverses hypothèses peuvent également être posées concernant les caractéristiques de la demande. En particulier, celle-ci peut être déterministe ou aléatoire, de distribution connue ou inconnue, stationnaire ou dynamique. Lorsque la demande est aléatoire, il est nécessaire de définir plus précisément la mesure de performance à optimiser. Le critère usuel est celui de l'espérance des coûts (par période, dans le long terme), éventuellement sous contrainte de niveau de service. Les modèles d'optimisation stochastique ainsi obtenus sont souvent très complexes et ne peuvent être résolus que de façon approximative.

Enfin, de nombreuses autres caractéristiques peuvent venir enrichir les modèles considérés: délai de livraison connu ou aléatoire, capacités de stockage et de production limitées, ristournes accordées en fonction de la quantité commandée, quantités livrées ou rendements aléatoires, stocks multi-échelons et/ou à localisations multiples, etc.

Le modèle de Harris, ou modèle de la quantité économique de commande, ou modèle EOQ (Economic Order Quantity) permet d'illustrer simplement certaines des questions fondamentales de la gestion des stocks. Ce modèle s'applique à la situation suivante: il s'agit de gérer un seul article, pour lequel toute quantité commandée est livrée instantanément, et pour lequel on connaît:

D = demande par unité de temps, supposée constante et connue avec certitude,

C = coût de passation d'une commande,

P = coût de possession d'une unité pendant une période.

Le problème est de déterminer un plan de production satisfaisant la demande à chaque instant (les ruptures de stock ne sont pas autorisées) et de coût total minimum. Il est intéressant de remarquer que ce modèle est parfaitement analogue au modèle de détermination des tailles de lots de Wagner-Whitin, à ceci près que la demande  $y$  est supposée continue et constante plutôt que discrète et variable.

Il est facile de vérifier que, sous ces hypothèses, la politique optimale consiste à réapprovisionner une quantité constante  $Q$  lorsque le stock s'annule. Le coût total par article vaut alors:

$$f(Q) = C/Q + PQ/2D.$$

La valeur  $Q^*$  qui minimise la fonction de coût est donc obtenue facilement:

$$Q^* = \sqrt{2DC/P}.$$

Cette quantité  $Q^*$  est appelée quantité économique de commande. L'intervalle de temps écoulé entre deux réapprovisionnements, ou période économique entre commandes, est donné par :

$$T^* = \sqrt{2C/DP}.$$

Il est évidemment hors de question de décrire ici les nombreuses extensions du modèle EOQ traitées dans la littérature. Nous mentionnerons seulement, à titre d'exemple, les travaux récents portant sur la gestion simultanée de plusieurs articles liés par des relations de dépendance: pièces et sous-assemblages d'un produit final (cf. MRP) ou produits évoluant dans un réseau de distribution (usines → dépôts → détaillants). De tels modèles sont considérablement plus complexes que le modèle EOQ de base et ne peuvent généralement pas être résolus de façon optimale. Cependant, des résultats très intéressants ont pu être obtenus en restreignant a priori la classe des politiques d'approvisionnement considérées et en résolvant ces modèles de façon approximative, mais quasi-optimale. Une restriction imposée dans certains travaux consiste ainsi à imposer que la longueur de l'intervalle écoulé entre approvisionnements successifs soit nécessairement une puissance de 2 (par exemple, 1, 2, 4 ou 8 semaines) et que la fréquence de réapprovisionnement d'un article soit toujours supérieure à celle de tous ses composants. Pour étranges qu'elles puissent paraître, ces hypothèses sont néanmoins inspirées de principes adoptés par certaines entreprises. De nombreuses entreprises imposent des livraisons sur des créneaux fixes. Ainsi tel fournisseur doit livrer le jeudi matin toutes les 2 semaines. Ces principes permettent en fait de lisser la charge de travail des magasiniers et de réguler le flux des camions ou des trains sur les aires de déchargement.

Prises simultanément, les hypothèses ainsi formulées impliquent que chaque période d'approvisionnement d'un composant correspond également à une période d'approvisionnement de l'article dans la composition duquel il entre: cette stratégie évite donc l'accumulation de stocks d'encours de composants. Par ailleurs, la première hypothèse implique également que la durée de l'intervalle de réapprovisionnement est un nombre entier « raisonnable », plutôt que le nombre irrationnel bizarre qui pourrait émerger du modèle EOQ. Sous des hypothèses assez générales, on démontre qu'il est possible de calculer efficacement de telles politiques « puissances de 2 » dont le coût est étonnamment proche (souvent à quelques pour-cents) du coût de la politique d'approvisionnement optimale.

La littérature regorge de « success stories » relatives à l'implantation de méthodes scientifiques de gestion des stocks développées sur mesure pour des entreprises particulières (cf. Assad, Wasil et Lilien 1992). Par ailleurs, cependant, plusieurs études semblent indiquer que l'écrasante majorité des entreprises et des logiciels commerciaux continuent de se fier essentiellement aux bonnes vieilles formules de Harris pour concevoir leurs politiques de

gestion des stocks. A l'heure où l'on ne parle que de zéro stock, ceci peut sembler paradoxal. Si l'on établit un diagramme de Pareto ou une classification en catégorie ABC, on constate que 20% des pièces achetées par une entreprise représentent typiquement 80% des coûts d'achat (catégorie A). Ces pièces font l'objet d'une gestion très fine. A l'opposé, on trouve de très nombreuses pièces de faible valeur (catégorie C: rondelles, vis, boulons, écrous etc.) pour lesquelles des livraisons fréquentes en petites quantités ne présentent pas d'avantage. Outre le coût payé pour le transport proprement dit, toute commande représente un coût administratif et comptable non négligeable pour l'entreprise. Pour de telles pièces, réduire les quantités livrées conduirait inmanquablement à augmenter les coûts, si bien que les méthodes classiques de réapprovisionnement leur restent applicables dans une large mesure.

## **Conception des systèmes productifs**

Les méthodes de recherche opérationnelle ont également trouvé de nombreuses applications dans le domaine de l'aide à la décision en matière de stratégie de production et, en particulier, de conception de systèmes de production. Des méthodologies aussi diverses que l'optimisation combinatoire (localisation/distribution, layout d'ateliers, ...), la modélisation multicritère (localisation, décisions d'investissement, ...), la programmation stochastique (modifications de capacité), la théorie des files d'attente ou la simulation (conception d'ateliers, de systèmes de distribution, ...) y sont utilisées.

Les modèles de localisation d'entreprises comptent parmi les mieux connus et les plus populaires dans ce domaine. Les modèles fondamentaux se répartissent essentiellement en deux grandes classes, selon que l'ensemble des localisations envisagées est représenté par un espace continu ou discret. Le premier type de modélisation conduit à la formulation de problèmes d'optimisation non linéaires, tandis que le second donne naissance à des problèmes d'optimisation mixtes (variables entières et continues). Malgré leur simplicité (ou grâce à elle), de tels modèles suffisent souvent à capter l'essence des décisions auxquelles sont confrontés les décideurs et fournissent donc de précieux éléments d'aide à la décision.

La rationalisation des flux matières exigée par les principes de gestion industrielle moderne a fait surgir de nouveaux problèmes, en particulier, celui de la réimplantation des ateliers fonctionnels. Lorsque l'on imagine un atelier, on se représente souvent une ligne de montage organisée pour la production répétitive d'un seul type de produit, telle qu'a pu l'immortaliser Charlie Chaplin dans « Les Temps Modernes ». En fait dans la majorité des ateliers, les équipements sont regroupés en sections fonctionnelles (section tours, fraiseuses, perceuses) plutôt que selon une logique dictée par le produit. Ce type d'atelier est utilisé pour les petites et moyennes séries et permet de traiter des produits très diversifiés. Mais en contrepartie, cette organisation présente un taux faible de productivité. Les principaux problèmes posés tiennent à la difficulté de rationaliser les déplacements entre machines, ce qui implique la constitution de stocks d'encours importants. Les conséquences directes de ces encours se mesurent sur les temps d'attente des pièces qui dépassent largement les temps d'usinage. Dès les années 60, de nombreux modèles mathématiques ont été développés pour optimiser l'implantation des systèmes de production. Ces modèles ont été à la base de logiciels commerciaux encore utilisés de nos jours:

– COMSOAL (Arcus 1966) pour l'équilibrage des lignes de production,

- CRAFT (Buffa, Armour, Vollmann 1964), ALDEP (Seehof, Evans 1967), COFAD (Tompkins, Reed 1976), PLANET (Deisenroth, Apple 1972) pour l'implantation d'ateliers.

Ces logiciels ne sont bien entendu pas figés et évoluent parallèlement aux avancées de la recherche en ce domaine.

Afin d'améliorer le fonctionnement des ateliers et de les faire bénéficier des avantages combinés de l'organisation par produit et de l'organisation fonctionnelle, on tend depuis quelques années à réimplanter les ateliers sous forme d'îlots de production. Un îlot de production est constitué par le regroupement de quelques machines capables d'assurer la production d'une variété limitée de pièces. Un îlot de production est donc moins souple qu'un atelier traditionnel, mais en échange, il est plus rationnel et productif.

La première étape de la réimplantation est d'identifier des groupes de pièces liées spécifiquement à certains groupes de machines (redécoupage logique). Pour cette étape, on utilise des méthodes de type technologie de groupe. De manière formelle, on part d'une matrice à éléments 0,1 dont on cherche à réarranger les lignes et colonnes de manière à faire apparaître des sous-matrices diagonales denses en "1" tout en laissant subsister peu de "1" en dehors de ces blocs diagonaux ("1" résiduels). Plus spécifiquement, dans le cadre de la décomposition en îlots, on partira d'une matrice à n lignes et m colonnes où:

- les lignes correspondent aux machines,
- les colonnes correspondent aux pièces à usiner.

On pose  $a_{ij} = 1$  si la machine  $i$  doit usiner la pièce  $j$ , et  $a_{ij} = 0$  sinon. A titre d'exemple, considérons la matrice suivante où les machines sont numérotées de 1 à 6 et les pièces de 1 à 10, et où les 0 sont omis:

	1	2	3	4	5	6	7	8	9	10
1	1	1					1			
2			1			1			1	1
3	1	1		1				1		
4				1	1	1			1	
5			1		1	1				1
6		1		1			1	1		

Après permutation de lignes et colonnes, cette matrice peut s'écrire:

	3	5	6	9	10	1	2	4	7	8
2	1		1	1	1					
4		1	1	1				1		
5	1	1	1		1					
1						1	1		1	
3						1	1	1		1
6							1	1	1	1

Les sous-matrices obtenues par la technologie de groupe permettent d'identifier les ensembles de machines qui constituent un îlot de production et les pièces qui doivent être fabriquées par cet îlot. Sur l'exemple, on voit que l'on peut séparer l'atelier initial en deux îlots:

- un îlot constitué des machines 2, 4 et 5 sur lesquelles on usinera les pièces 3, 5, 6, 9 et 10;
- un îlot constitué des machines 1, 3 et 6 sur lesquelles on usinera les pièces 1, 2, 4, 7 et 8.

Il reste à traiter le cas des "1" résiduels, c'est-à-dire dans notre exemple le passage de la pièce 4 sur la machine 4. Plusieurs possibilités peuvent être envisagées:

- accepter que des pièces transitent d'un îlot à l'autre; cette solution doit rester exceptionnelle, sous peine de complexifier les flux de production et d'aller ainsi à l'encontre du but recherché;
- investir dans une nouvelle machine 4bis apte à réaliser la pièce 4; le suréquipement permet ici d'améliorer le flux global de l'atelier;
- revoir la conception de la pièce 4;
- sous-traiter l'opération.

La question fondamentale de la technologie de groupe, c'est-à-dire la partition optimale de la matrice d'incidence machines-pièces, s'applique également à d'autres situations industrielle. Cette question générique mène à la formulation de problèmes d'optimisation NP-complets. De nombreuses heuristiques ont été utilisées pour aborder ces problèmes: algorithme ROC (Rank Order Clustering) dû à King (1980), algorithme BEA de McCormick, Schweitzer et White (1972), etc.

La conception des systèmes de production utilise aussi la théorie des files d'attente et les méthodes de simulation numérique. Ces techniques font partie de la panoplie de base du chercheur opérationnel, en particulier lorsqu'il s'agit de conception de systèmes (mais leur champ d'application s'étend évidemment à bien d'autres domaines de la gestion de la production; il comprend notamment l'évaluation des décisions de planification tactiques ou opérationnelles). Les réseaux de files d'attente permettent d'analyser le comportement de systèmes de production interconnectés (machines - matériel de manutention - palettes). Les mesures de performance dégagées de cette analyse sont par exemple la quantité moyenne d'encours accumulée aux différentes stations de production, le taux d'utilisation des ressources, le temps de passage des ordres de fabrication dans l'atelier, etc. Typiquement, cependant, ces différentes mesures ne concernent que le comportement stationnaire, ou à long terme, du système. Par ailleurs, la théorie des files d'attente ne fournit généralement pas d'informations détaillées sur la distribution de probabilité des paramètres considérés. La simulation numérique peut venir ici à la rescousse, au prix d'un certain alourdissement des moyens de calcul requis et d'un manque de transparence des modèles. L'utilisation des méthodes de simulation en gestion de la production devrait être favorisée par l'apparition récente de logiciels conviviaux et puissants dont certains sont spécifiquement destinés à l'analyse de systèmes de production.

## **Conclusion**

Comme nous l'avions annoncé en introduction, cet article n'a fait qu'esquisser quelques domaines d'application de la recherche opérationnelle en gestion de production. En

particulier, nous n'avons pas abordé les problèmes liés à l'optimisation des ressources (détermination des mélanges, placement de pièces, optimisation des découpes), le séquençement sur les lignes de production ou les problèmes soulevés par la logistique amont (localisation des centres de groupage/dégroupage, organisation des tournées).

A l'heure actuelle, la recherche opérationnelle offre une formidable panoplie d'outils, trop peu connus hélas, capables d'assister les décideurs industriels dans leurs prises de décisions. Le terme « outil » révèle bien les limites de ce que peut apporter la recherche opérationnelle au monde industriel. Son ambition n'est pas de remplacer le décideur, mais de lui indiquer la meilleure décision (plan de production, ordonnancement, politique de stockage) possible dans le cadre du modèle connu par le système. On sait cependant qu'un tel modèle ne capte qu'une partie restreinte des informations dont dispose l'utilisateur et n'intègre que partiellement les objectifs ou contraintes dont il doit tenir compte. Or, les informations délaissées peuvent dans certaines circonstances devenir primordiales. Ici comme ailleurs, la prise de décisions repose sur une articulation raisonnée entre l'homme et les systèmes artificiels et non pas sur le remplacement du premier par les seconds.

## Références

- Askin R.G., Standridge C.R.: *Modeling and Analysis of Manufacturing Systems*. John Wiley & Sons, New York, 1993.
- Assad A.A., Wasil E.A., Lilien G.L.: *Excellence in Management Science Practice*. Prentice-Hall, Englewood Cliffs, N.J., 1992.
- Buzacott J.A., Shanthikumar J.G.: *Stochastic Models of Manufacturing Systems*. Prentice-Hall, Englewood Cliffs, N.J., 1993.
- Carlier J., Chrétienne Ph.: *Problèmes d'Ordonnancement*. Masson, Paris, 1988.
- Crama Y., Oerlemans A.G., Spieksma F.C.R.: *Production Planning in Automated Manufacturing*. Springer, Berlin, 1996, 2ème édition.
- Giard V.: *Gestion de la Production*. Economica, Paris, 1988, 2ème édition.
- Graves S.C., Rinnooy Kan A.H.G., Zipkin P.H., eds.: *Logistics of Production and Inventory*. Elsevier Science Publishers, Amsterdam (1993).
- Silver E., Peterson R.: *Decision Systems for Inventory Management and Production Planning*. John Wiley & Sons, New York, 1985.