
Etude préliminaire à un modèle de prévision à court terme de l'activité d'un transporteur sous température dirigée

Wilfried Despagne

Laboratoire Lab-STICC,
Agrostar (le pôle systèmes d'information du Groupe STEF-TFE)
Wilfried.Despagne@stef-tfe.com

RÉSUMÉ. Cet article décrit une problématique de recherche opérationnelle. Un transporteur sous température dirigée cherche à optimiser la planification de ses ressources humaines et matérielles par la prévision à très court terme de son activité. Le challenge réside dans le fait de trouver un modèle de prévision unique s'adaptant, sans intervention humaine, aux spécificités des 57 agences du transporteur. La matière première est l'information récoltée par le transporteur depuis plus de six ans. Les outils sont des algorithmes mathématiques utilisés pour la prévision des séries temporelles. Le travail décrit ici, vise à combiner ces outils pour qu'ils extraient le maximum d'information déterministe capable d'être anticipée. L'introduction pose la problématique et son contexte économique. Elle est suivie d'un descriptif des procédures utilisées et d'un argumentaire pour défendre leur choix. Les solutions informatiques adoptées sont inventoriées. Enfin, la conclusion renvoie à des pistes d'études.

ABSTRACT. This article describes an operational research problem. A firm specialized in temperature controlled transportation wants to optimize the planning of its human and material resources using short term activity forecasting. The challenge is to find an unique forecasting model adapted, without any human intervention, to the specific needs of 57 company's offices. To do it, the company has been collecting data for five years. To analyse them, mathematical algorithms for forecasting time series are used. The work is to combine these tools to extract the maximum of determinist information that should be anticipated. The introduction presents the problem and its economic context. This is followed by a description of the process used and arguments to defend choices done. The adopted solutions are inventoried. Finally, the conclusion refers to courses of study.

MOTS-CLÉS : série chronologique, prévision, chaîne logistique, transport.

KEYWORDS: time series, forecasting, supply-chain, carriage.

1. Introduction

Dans le cadre de ce travail, nous présentons un état des lieux de la prévision d'activité chez un transporteur sous température dirigée. Cette démarche a permis d'acquérir un point de vue formalisé du problème et une automatisation des procédures.

Le transport sous température dirigée est l'activité qui consiste à transporter de la marchandise soumise à une température définie entre -25°C et $+15^{\circ}\text{C}$. Supposons que ce transporteur soit un grand groupe et qu'il dispose d'un réseau d'une cinquantaine d'agences. Pour une définition du réseau nous référons à (Branche, 10/2006, §1). Une agence comporte une plate-forme (ou quai) sur laquelle est réceptionnée la marchandise. Ces quais permettent au transporteur de préparer les commandes, les trier et les étiqueter pour les dépêcher à travers l'Hexagone en moins de 48h. L'agence est en charge d'enlever la marchandise chez ses clients régionaux pour l'injecter dans le réseau qui s'occupe de la livrer à bon port. Ces trois transports sont appelés dans l'ordre « ramasse », « expédition » et « livraison ».

Dans plus de 90% des cas, l'agence reçoit du client expéditeur ses ordres de transports moins de 3h avant l'enlèvement de la marchandise. Pour résumer en quelques mots la problématique, les ordres de transports tombent par fax ou EDI (Échange de Données Informatisé) à 7h du matin pour un départ entre 8h et 12H en fonction de la destination. La livraison doit se faire entre 18h et 22h le soir sur toute la France. À cause des opérations spéciales (promotions), il arrive que les quantités à transporter varient de 1 à 10 d'un jour à l'autre. Dans ces cas, comment faire face si les clients n'en communiquent pas la période ?

Le transporteur doit donc s'adapter aux exigences de ses clients. Pour l'aider, il souhaite mettre en place un système de prévision pour l'aide à la gestion des ressources (Seiersen, 04/2006, §2.2). Le système doit permettre d'anticiper le poids des marchandises à transporter et le nombre de lettres de transports¹ à couvrir. Ces deux informations, prévues à un horizon de 15 jours et à périodicité journalière, lui permettent d'anticiper l'effectif humain et matériel sur le quai ainsi que le nombre de semi-remorques à mettre à disposition. Par souci de simplification, nous appellerons prévision de l'activité, l'anticipation du poids et du nombre de lettres de transports.

Ayant comme souci d'homogénéiser les procédures de traitement, le transporteur convoite un système de prévision capable de s'adapter aux spécificités de ses différentes agences. Les prévisions doivent être pertinentes, facilement consultables, conviviales et accessibles par une interface web sur l'intranet (Cluzel, 04/2006, §1,2,3). Enfin, les objectifs de la direction sont d'atteindre une erreur de prévision quotidienne (en valeur absolue) inférieure à 5%. Pour atteindre les objectifs, le transporteur propose au prévisionniste un historique des masses transportées depuis l'année 2000.

La suite de cet article propose une méthode de prévision qui combine méthodes endogènes et exogènes (Bourbonnais *et al.*, 2007, chap. 9). La méthode cherche à anticiper, à un rythme quotidien et à un horizon de 15 jours, les valeurs de trois chroniques dont la somme de deux d'entre elles est égale à la troisième. Deux algorithmes de décomposition de la chronique sont utilisés. Les estimations obtenues suivant les deux algorithmes sont combinées.

1. Document matérialisant le contrat de transport passé entre le transporteur et l'expéditeur

2. Le Modèle

2.1. Vue d'ensemble

Nous cherchons à modéliser trois séries temporelles. Elles reflètent les flux des marchandises sorties d'un quai. Les marchandises sont soit transportées vers un autre quai, c'est alors de « l'expédition », soit livrées chez le destinataire final, c'est de la « distribution ». La somme des deux donne le « total » des denrées traitées à quai. Pour une définition précise de ces trois notions nous renvoyons à l'article de (Fadde-Costes, 10/1999, §2.2). Les deux quantités à prévoir, le nombre de lettres de transport et le poids des marchandises associées, sont relevés quotidiennement par le transporteur. La taille de l'historique est ainsi de cinq ans d'observations.

Le modèle statistique choisi, cherche à se rapprocher le plus possible de l'activité du transporteur en la décomposant. Supposons que l'activité d'une agence est déterminée par trois composantes : les facteurs déterministes, les facteurs stochastiques et les facteurs non observés. Les facteurs déterministes (la saisonnalité, les jours fériés, les promotions) ont la particularité d'être observés dans le passé et leur date d'arrivée dans le futur est connue. Les facteurs stochastiques se décomposent également en variables observées (grèves, perte ou gain d'un portefeuille client) mais leurs dates d'apparition futures sont inconnues. Les facteurs non observés sont des événements exceptionnels inédits (liquidation judiciaire du principal concurrent) dont les valeurs passées, présentes et à venir sont inconnues. Le modèle proposé crée une copie de la chronique. Pour chacune des deux copies, il utilise une méthode appropriée pour extraire les facteurs déterministes avant d'appliquer un modèle « autoprojectif » pour estimer les variables stochastiques. Les deux prévisions obtenues après recompositions, sont soumises à une combinaison linéaire conduisant à l'erreur de prévision minimale.

Soit le triplet $(X_t, Y_t, Z_t) \in \mathbb{R}^3$ représente les valeurs à la date t des chroniques en « expédition », « distribution » et « total ». Elles sont soumises à la contrainte $X + Y = Z$. Ces valeurs sont calculées à partir d'autres chroniques $(U_{X,t}, U_{Y,t}, U_{Z,t})$ par $X_t = \omega_1 U_{X,t}$, $Y_t = \omega_2 U_{Y,t}$ et $Z_t = U_{Z,t}$. ω_1 et ω_2 sont déterminés afin de satisfaire la contrainte.

$$U_{E,t}^1 = T_t S_t^1 (\beta^1 F)_t V_t^1 \epsilon_t^1 \quad [1]$$

$$U_{E,t}^2 = T_t S_t^2 (\beta^2 F)_t V_t^2 \epsilon_t^2 \quad [2]$$

avec $E = \{X, Y, Z\}$, T la tendance, F un vecteur binaire correspondant à des événements calendaires, β leurs pondérations, S les coefficients saisonniers, V les processus stationnaires, ϵ les bruits blancs. Pour étudier les composantes indépendamment les unes des autres, nous utilisons la fonction logarithme népérien qui permet de passer d'un modèle multiplicatif à un modèle additif.

$$\ln(U_{E,t}^1) = \ln(T_t) + \ln(S_t^1) + \ln((\beta^1 F)_t) + \ln(V_t^1) + \ln(\epsilon_t^1) \quad [3]$$

$$\ln(U_{E,t}^2) = \ln(T_t) + \ln(S_t^2) + \ln((\beta^2 F)_t) + \ln(V_t^2) + \ln(\epsilon_t^2) \quad [4]$$

Les résultats obtenus par les équations (3) et (4) sont combinée dans l'équation (5).

$$U_{E,t} = \lambda \ln(U_{E,t}^1) + (1 - \lambda) \ln(U_{E,t}^2) \quad [5]$$

2.2. Modélisation des éléments déterministes

2.2.1. La tendance

L'activité des agences connaît deux tendances, une tendance intra-annuelle et une tendance inter-annuelle. La tendance intra-annuelle décrit l'activité d'une agence entre janvier et décembre. La tendance inter-annuelle est celle à long terme. Après les fortes dépenses de fin d'année, les ménages se remettent à économiser en janvier. C'est pourquoi l'activité est forte en décembre et chute en janvier. D'autre part, dans le milieu économique, les tendances sont lentes et progressives (Burtschy, 1980). Ces constatations nous font choisir une tendance linéaire par morceau de périodicité annuelle (Vaté, 1993, chap. 5,6). Son équation s'écrit :

$$T_t = \theta t + \varphi An(t) + cste$$

avec $An(t)$ l'année correspondante à la date t ; θt représente la tendance inter-annuelle, $\varphi An(t)$ la tendance intra-annuelle.

2.2.2. La saisonnalité

Les quantités transportées forment une superposition de mouvements oscillatoires de périodes hebdomadaires et journalières. Ainsi, ils admettent une double saisonnalité que nous allons estimer pour l'effacer de la chronique. La saisonnalité hebdomadaire comporte 53 coefficients et la saisonnalité journalière comporte 313 jours ouvrés. La saisonnalité hebdomadaire est causées par des événements extérieurs tels que la météo, les vacances scolaires, les périodes de fête. Inclus dans la saisonnalité, il ne sera plus nécessaire de les analyser individuellement.

La saisonnalité journalière, très marquée, est due à une répartition de l'activité sur les 6 jours ouvrés de la semaine. Cette répartition dépend de l'agence en question. Le samedi par exemple, l'activité est réduite à son strict minimum, elle varie ensuite suivant les jours d'approvisionnements des grossistes et des GMS (Grandes et Moyennes Surfaces).

Il existe de nombreuses méthodes de désaisonnalisation. Elles ont l'avantage de décrire l'activité aux décideurs. Les coefficients saisonniers montrent l'écart de la valeur moyenne constatée pour une semaine i et un jour j par rapport à la tendance. En matière de prévision des ventes, la méthode idéale n'existe pas. Partant du principe que deux valent mieux qu'une, pourquoi ne pas appliquer deux méthodes pour ne garder qu'une combinaison des résultats, comme le suggèrent (Bates *et al.*, 1969), suivant le critère de minimisation de la variance des erreurs (cf § 2.4). La première méthode de désaisonnalisation est celle des moyennes mobiles. La deuxième est une décomposition par régression linéaire.

La méthode des moyennes mobiles (Brockwell *et al.*, 1996, chap. 1) permet d'estimer des coefficients saisonniers selon les 3 étapes suivantes,

- calculer la série des moyennes mobiles centrées,

- calculer l'écart entre les valeurs observées et la moyenne mobile,
- normaliser les écarts, pour aboutir aux coefficients saisonniers.

Cette méthode est appliquée une première fois pour corriger les variations hebdomadaires et une deuxième fois pour corriger les variations journalières. Les résultats obtenus sont les coefficients saisonniers S^1 de l'équation (1).

La deuxième méthode de décomposition est celle proposée par (Buys-Ballot, 1847). Elle consiste à trouver les coefficients S^2 de l'équation (2) par MCO (Moindre Carrés Ordinaire).

$$\ln(U_{E,t}^2) - \ln(T_t) = \gamma_1 S_t^{2,1} + \gamma_2 S_t^{2,2} + \gamma_3 S_t^{2,3} + \gamma_4 S_t^{2,4} + \gamma_5 S_t^{2,5} + \gamma_6 S_t^{2,6} + \Phi_0 S S_t^{2,0} + \dots + \Phi_{52} S S_t^{2,52} + \zeta_t$$

La chronique diminuée de sa tendance, se décompose en une suite de composantes saisonnières correspondantes aux 6 jours de la semaine, aux 53 semaines de l'année et d'un processus ζ_t . Les p composantes saisonnières sont des variables binaires pour p saisons dans l'année. La variable binaire est égale à 1 lorsque la donnée se rapporte à la saison envisagée et 0 partout ailleurs.

2.2.3. Évènements calendaires

Un jour férié provoque une récupération de l'activité perdu sur les jours voisins. Par exemple, un jeudi férié peut conduire à une augmentation de l'activité le lundi par anticipation ou le vendredi par retard. Souvent, un jour férié a des conséquences prévisibles sur une période de 9 jours (J-4, J, J+4). Les conséquences sont différentes selon le jour férié, le jour de la semaine, l'agence en question.

La récupération de la perte d'un jour de travail sera variable selon que ce jour est un lundi, un mardi, ou un autre. Si c'est un samedi, il y a peu d'activité à rattraper, alors que le lundi est une journée chargée. Si le jour férié tombe un vendredi, les GMS anticipent et demandent à être livrées en double quantité, le précédent jeudi. Les livraisons explosent le jeudi et les expéditions gonflent le mercredi, et ainsi de suite.

Un jour férié est souvent synonyme de fête ce qui engendre une augmentation de la consommation des ménages et par conséquent un renforcement de l'activité du transporteur. Mais cette hausse est variable selon qu'il s'agisse de la Toussaint ou de Noël. Les agences ne sont pas égales face à la hausse d'activité engendrée par un jour férié. Prenons une agence située à côté d'une usine de chocolat. La distribution des chocolats de Pâques monopolise une grande partie des ressources de l'agence.

Pour tenter de séparer les phénomènes engendrés par la tombée d'un jour férié nous suivons une méthode inspirée de (Longstaff *et al.*, 2004). Pour ce faire nous retenons les informations suivantes :

- nom du jour férié,
- jour de la semaine (lundi, ..., samedi),
- jour de la semaine des 4 jours précédents et des 4 jours suivants,
- éloignement des 8 jours encadrant le jour férié ($-4, -3, -2, -1, 1, 2, 3, 4$).

Il en résulte un tableau binaire de 20 colonnes, une pour identifier le jour de la semaine, huit nommées entre -4 et 4 pour indiquer la distance du jour impacté au jour férié et 11 autres pour reconnaître les jours fériés. Les jours fériés recensés sont : Jour de l'An, Pâques, Fête du Travail, Victoire 1945, Ascension, Pentecôte, Fête National, Assomption, Toussaint, Armistice, Noël. Les T lignes représentent le nombre d'enregistrements de l'historique. Cette matrice, appelée F est composée de variables binaires permettant d'ajuster un modèle de régression sur la série désaisonnalisée : $W = \beta F + \xi$, avec W la chronique désaisonnalisée. Les 20 variables binaires ne sont pas toutes significatives. Une comparaison entre la variance des estimations et la variance des erreurs (test de Fischer) permet de retenir les variables les plus discriminantes.

2.3. Modélisation des éléments stochastiques

Les procédures précédentes ont permis d'extraire de la chronique la tendance, les composantes saisonnières et les événements prévisibles. La série temporelle obtenue (V) présente une relation de cause à effet entre l'observation à une date t et les observations précédentes ($t - 1, t - 2, t - 3, t - 4, t - 5, t - 6$). Pour modéliser et prévoir cette série, nous optons pour le lissage exponentiel simple (Box *et al.*, 1970) qui a l'avantage d'être automatisable.

$$\hat{V}_{T+1} = (1 - \alpha) \sum_{j=0}^{T-1} \alpha^j V_{T-j}$$

Il prend en compte les observations passées (de $T - 1$ à $T - 6$) et les pondère par la constante de lissage α . Cette dernière est estimée de sorte à minimiser la différence carrée entre la chronique et les estimations du modèle.

Les valeurs prédites sont additionnées aux composantes déterministes que nous avons soustraites précédemment. Après application de la fonction exponentielle, nous obtenons l'estimation de la chronique d'origine.

2.4. Combinaison des prévisions

Du fait de deux procédures de décomposition, nous obtenons deux prévisions. L'erreur de prévision obtenue par les deux méthodes ne permet pas de conclure à la supériorité d'une sur l'autre. Le schéma de combinaison choisi est proposé par (Bourbonnais *et al.*, 2007, chap. 11). Il cherche à minimiser la variance de l'erreur de prévision résultant de la performance antérieure des prévisions individuelles. Ne dépendant pas de la spécification d'un seul modèle, les prévisions combinées tentent à être plus robustes. La prévision combinée PC , est une moyenne pondérée des deux prévisions individuelles PU^1 et PU^2 ; $PC = \lambda PU^1 + (1 - \lambda) PU^2$, λ est le coefficient pondérateur, $0 < \lambda < 1$. Soit EPC , l'erreur de prévision combinée, $EPC = \lambda EPU^1 + (1 - \lambda) EPU^2$, la variance de l'erreur de prévision est $V(EPC) = \lambda^2 V(EPU^1) + (1 - \lambda)^2 V(EPU^2) + 2(1 - \lambda)\lambda COV(EPU^1, EPU^2)$. Nous cherchons λ qui minimise $V(EPC)$, en annulant la dérivée première par rapport à λ . La solution est, si les erreurs de prévisions sont corrélées

$$\lambda = \frac{V(EPU^2) - COV(EPU^1, EPU^2)}{V(EPU^1) + V(EPU^2) - 2COV(EPU^1, EPU^2)}$$

,sinon

$$\lambda = \frac{V(EPU^2)}{V(EPU^1) + V(EPU^2)}$$

3. Applications et résultats

Soit une agence de transport dont 60% de l'activité se fait en distribution, contre 40% en expédition. Entre janvier et mars, l'activité est stable autour de 750 tonnes (du lundi au vendredi). Les jours fériés d'avril et mai, provoquent de brusques changements de régime. Le poids des marchandises transportées peut passer de moins de 10 tonnes un jour férié à plus de 1 110 tonnes deux jours avant. Les samedis et jours fériés exclus, l'activité d'avril et mai frôle les 1000 tonnes/jour. Juin est une période d'accalmie. Avec une activité autour de 900 tonnes/jours et quelques jours fériés, juillet et août sont des mois chargés. Septembre, octobre sont une période qui retrouve le calme de début d'année avec une moyenne inférieure à 800 tonnes/jour. Enfin, l'activité progresse régulièrement tout le long de décembre pour atteindre un pic de 1 114 tonnes cinq jours avant Noël.

Le tableau 1 compare les résultats obtenus par le modèle avec les observations réelles. L'échantillon d'apprentissage est l'historique des poids précédent la date T . T varie entre le 01/01/07 et le 31/07/07 avec un pas de six jours. Nous prédisons les dates $T + 1$ à $T + 6$. Nous comparons $\hat{X}_{T+i}$, $\hat{Y}_{T+i}$, $\hat{Z}_{T+i}$ avec X_{T+i} , Y_{T+i} , Z_{T+i} pour $i = \{1, \dots, 6\}$.

Expédition		Distribution		Total	
BAR	DR	BAR	DR	BAR	DR
8%	10%	9%	12%	7%	9%

Tableau 1. Indicateurs de confiance.

Les indicateurs d'évaluations sont, le biais absolu relatif (BAR) et la dispersion relative (DR). Les indicateurs se réfèrent à la moyenne des valeurs observées. Les résultats du modèle se trompent de 7% par rapport à la moyenne de la série « total des marchandises passées à quai ». Les erreurs sont susceptibles de varier entre plus ou moins 9%. Cependant, l'objectif d'une erreur de prévision annuelle moyenne, quelque soit l'agence, inférieure à 5%, n'est pas atteint. Mais on s'aperçoit que selon les agences et les jours de la semaine, une prévision juste à plus de 95% sur les 3 dernières semaines glissantes est réalisable.

4. Conclusion et recherches futures

Les calculs de prévision sont centralisés au service d'informatique décisionnelle du transporteur. Les résultats sont publiés sur une interface web spécialement créée pour ce besoin. Ils sont présentés sous forme de tableaux de bords et de graphiques. Les prévisions étant juste à plus de 90% en moyenne, elles sont une sérieuse aide à la décision.

Le point faible de cette modélisation est le nombre trop important de paramètres au regard de la taille de l'historique (5 ans). En effet, pour chacune des décompositions il y a 2 paramètres pour la tendance, 313 pour la saisonnalité quotidienne, 53 pour l'hebdomadaire, 20 pour les événements calendaires et un dernier pour le lissage exponentiel. L'avantage est de pouvoir expliquer au décideur l'effet de chacun des paramètres sur la quantité de marchandises sortie du quai.

Une toute autre approche sera de désaisonnaliser par des filtres de Kalman emboîtés et introduire dans le modèle, sous forme d'impulsions, les effets des événements calendaires. C'est une méthode préconisée par (Martin, 1999) et appliquée sur la prévision de la consommation électrique.

Néanmoins, les économistes misent plus sur ce qu'ils appellent les « prévisions collaboratives ». Les statistiques ne suffisent pas pour obtenir une prévision fiable. Grâce au développement de la gestion partagée, il faut enrichir les résultats par la validation ou les commentaires des différents services de l'entreprise (la logistique, le marketing) et même des clients.

5. Bibliographie

- Bates J., Granger C., « The combination of forecasts », *Operational Research Quarterly*, 1969.
- Bourbonnais R., Usunier J. C., *Prévision des ventes, théorie et pratique*, Economica, 2007.
- Box G., Jenkins G., *Time series analysis : Forecasting and control*, San Francisco : Holden-Day Inc, 1970.
- Branche F., « Transport de messagerie », *revue technique de l'ingénieur*, 10/2006.
- Brockwell J. P., A. Davis R., *Introduction to Time Series and Forecasting*, Springer, 1996.
- Burtschy B. e. M. C., « A propos de prévision à court terme de la production industrielle », *Revue De Statistique Appliquée*, vol. tome 28 n° 2, p. 5-24, 1980.
- Buys-Ballot C. H. D., « Les changements périodiques de temperature », 1847.
- Cluzel G., « Rentabilité d'un système d'information. Approche théorique », *revue technique de l'ingénieur*, 04/2006.
- Fadde-Costes N., « Système d'information logistique et transport », *revue technique de l'ingénieur*, 10/1999.
- Longstaff F. A., Wang W. A., « Les changements périodiques de temperature », *The Journal of Finance*, vol. Vol. 59, No. 4, p. 1877-1900, 2004.
- Martin M.-M., « Filtrage de Kalman d'une série saisonnière, Application à la prévision de la consommation d'électricité. », *Revue De Statistique Appliquée*, vol. tome 47, n° 4, p. 69-86, 1999.
- Seiersen N., « Systèmes d'information logistique », *revue technique de l'ingénieur*, 04/2006.
- Vaté M., *Statistique Chronologique et Prévisions*, Economica, 1993.