

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

Wilfried Despagne^{1,2}

¹ Université Européenne de Bretagne, Lab-STICC (UMR CNRS 3192),
Université de Bretagne-Sud, Centre Yves Coppens, Campus de Tohannic,
F-56017, Vannes

² STEF-TFE, la logistique du froid made in Europe
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.

Mots clés : série chronologique, prévision, chaîne logistique, transport.

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 goal of 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. It is followed by a description of the chosen process and arguments to defend those choices. The adopted solutions are inventoried. Finally, the conclusion refers to courses of study.

Key-words : time series, forecasting, supply-chain, carriage.

1 Introduction

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 $-25\text{ }^{\circ}\text{C}$ et $+15\text{ }^{\circ}\text{C}$. Les marchandises sont essentiellement des denrées alimentaires périssables, les produits carnés, les produits de la mer, les fruits et légumes, les produits laitiers, les surgelés, mais aussi des plantes ou médicaments. Elles ont en commun d'être soumises à un cahier des charges très strict qui définit la « chaîne du froid ». La « chaîne du froid » est le processus qui permet de maintenir un produit à basse température. Le froid ralentit la propagation des microorganismes. La loi impose des règles en cette matière. Les arrêtés du 9 mai 1955 (réglementant l'hygiène des aliments remis directement au consommateur) et du 20 juillet 1998 (fixant les conditions techniques et hygiéniques applicables au transport des aliments) soumettent les industriels à une obligation de résultat. À la contrainte du maintien de la température liée à la chaîne du froid, s'ajoutent celles liées au métier du transporteur. Là encore, le législateur régleme la durée du travail d'un chauffeur routier et l'autorisation de circulation. Ce ne sont ici que des exemples de lois parmi d'autres permettant de percevoir l'ampleur des contraintes législatives en matière de transport et plus encore de transport frigorifique. Sans une maîtrise de gestion des ressources matérielles et humaines, ces contraintes font déborder les coûts de fonctionnement.

Le domaine de l'alimentaire met en relation producteurs, industriels, distributeurs et transporteurs. Ils forment une chaîne appelée « Supply Chain » (chaîne logistique, [Aya05]). La contrainte consistant à garder la fraîcheur des produits entre leur lieu de fabrication et celui de distribution, impose aux différents maillons de la chaîne de travailler en flux tendus. La « Date Limite de Consommation » (DLC) est de quelques jours pour des découpes de volaille ou de 21 jours pour des yaourts. Elle gouverne la « Supply Chain » des produits frais et imprime un rythme rapide à la chaîne de distribution. Dans son mémoire, Charlotte [Ter04] souligne que les industriels répondent aux commandes quotidiennes des GMS (Grandes et Moyennes Surfaces) pour approvisionner leurs entrepôts. Les quantités de ravitaillements fluctuent quotidiennement selon les sorties de caisse des GMS. Tous les soirs, toutes les données de vente en magasin sont centralisées à l'entrepôt pour que les commandes soient expédiées dès le lendemain. Ce réapprovisionnement automatique représente : 80% des ventes pour l'épicerie et 98% des produits frais ([Ter04]). Pour soutenir ce rythme et alimenter les linéaires sans pour autant avoir de stocks, les acteurs doivent prévoir leurs ventes. La logique voudrait que les GMS partagent leurs prévisions de ventes, aussi bien à un rythme mensuel que quotidien, avec les industriels de l'agroalimentaire et que ces derniers les relient aux transporteurs. Mais la loi de la concurrence ne permet pas cette pratique. Les GMS craignent de donner trop d'informations aux industriels qui, se voyant dans une position de force, pourraient en profiter pour augmenter leurs prix. Ainsi, les acteurs de la chaîne logistique établissent des prévisions chacun de leur côté.

Dans le cadre de ce travail, nous proposons un système de prévision des ventes d'un transporteur sous température dirigée. 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 nous référons à [Bra06]. 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. Dans le jargon métier c'est un transport en A pour B. Grossièrement, il existe une agence par région. D'une part, elle 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 ». D'autre part, l'agence reçoit de la marchandise du réseau pour la livrer aux points de vente de sa zone de chalandise.

Comme les autres maillons de la « Supply Chain » alimentaire, le transporteur travaille suivant la méthode JAT (Juste A Temps). 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, [Sei06]) 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 ([Sei06]). Le système doit permettre d'anticiper le poids des marchandises à transporter et le nombre de lettres de transports ³ 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. Dès lors, les prévisions de l'activité visent à fournir des éléments cruciaux pour :

- la planification des ressources matérielles et humaines,
- optimiser les règles de ramasse, d'expéditions et de distribution,
- formaliser le comportement à court, moyen et long terme des clients,
- atteindre un niveau de service élevé,
- limiter la dépendance vis-à-vis de l'incertitude.

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 facilement consultables, conviviales et accessibles par une interface web sur l'intranet ([Clu06]). Enfin, les objectifs de la direction sont d'atteindre une erreur de prévision quotidienne inférieure à 5%.

Pour répondre aux exigences de traçabilités imposées par l'Europe (texte CE n° 178/2002), améliorer la rentabilité et du même coup le service rendu au client, le groupe utilise un Data Warehouse. C'est un entrepôt de données Oracle© dans lequel sont stockées toutes les informations relatives aux colis transportés et entreposés par le groupe. Les données proviennent d'un système d'information appelé GTI (Gestion Intégrée du Transport). Il regroupe les applications informatiques du groupe et alimente la base de données, soit plusieurs giga-octets d'information. L'information de référence est l'ordre de transport. Il donne accès aux informations de chargement et déchargement : produit, unité logistique, lieu de déchargement-chargeement, date, heure, tiers expéditeur, tiers destinataire, nombre de colis, type d'emballage, poids, catégories d'emballages ... La base

³Document matérialisant le contrat de transport passé entre le transporteur et l'expéditeur

de données est mise à jour quotidiennement avec les données de la veille. Cette masse d'information permet de créer un historique des volumes transportés depuis l'année 2000. Il constitue la population sur laquelle vont se construire les prévisions.

La suite de cet article propose une méthode de prévision qui combine méthodes endogènes et exogènes ([BU07]). 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. Enfin une technique appelée « top-down aggregation » est mise en œuvre (§ 2.5) pour satisfaire la contrainte qui lie les trois chroniques.

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 [FC99]. 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 d'elle, il utilise une méthode appropriée pour extraire les facteurs déterministes avant d'appliquer un modèle « auto-projectif » pour estimer les variables stochastiques. Les deux prévisions obtenues après recompositions, sont soumises à une combinaison linéaire conduisant à l'erreur de prévision minimale. La différence entre les résultats obtenus par le modèle et les observations résulte des facteurs non observés.

Soit le triplet $(X_t, Y_t, Z_t) \in \mathbb{R}^3$ représentant les valeurs à la date t des chroniques en « expédition », « distribution » et « total ». Les variables 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. Ces autres chroniques sont calculées par :

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

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

avec $E \in \{X, Y, Z\}$, T la tendance, F un vecteur binaire correspondant à des événements calendaires, β leurs pondérations, S les coefficients saisonniers, V les processus stationnaires et ϵ les bruits blancs. Pour étudier les composantes indépendamment les unes des autres, nous utilisons la fonction logarithme népérien.

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

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

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 ([BM80]). Ces constatations nous font choisir une tendance linéaire par morceau de périodicité annuelle ([Vat93]). 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 due à des périodes d'activités fluctuantes. Elles sont 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. L'idéal serait de s'affranchir du compteur des semaines pour ne retenir que les distances par rapport à des événements calendaires. Ainsi, nous n'aurions plus à nous préoccuper du fait que pendant l'année A tel jour férié est tombé la semaine $s13$ alors que l'année suivante il est tombé la semaine $s14$. C'est une perspective à approfondir. Pour l'instant, nous retenirons 53 semaines dont les deux extrêmes sont corrigés suivant le nombre de jours qu'elles comportent. La semaine 1 est celle qui contient le premier lundi de l'année. Il en découle que si le premier janvier ne tombe pas un lundi alors lui et les jours suivant jusqu'au prochain lundi sont numérotés de la semaine 0. Ainsi, la semaine 0 et la dernière semaine de l'année (52) sont les seules susceptibles de ne pas compter 6 jours ouvrables. Elles sont pondérées suivant le nombre de jours qu'elles contiennent.

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. Rappelons que les GMS représentent 80% des volumes transportés.

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 [BG69, Sch06], suivant le critère de minimisation de la variance des erreurs (§ 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 ([BR96]) 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 [BBD47]. 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

Les évènements suivants sont appliqués sur les deux séries $((\ln(U_E^i) - \ln(T) - \ln(S^i), i = \{1, 2\}))$, corrigées des variations saisonnières et de la tendance, obtenues par moyennes mobiles et par la méthode de Buys-Ballot. La perte d'un jour d'activité provoque la récupération de cette activité 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 pour des livraisons en A pour B. S'il tombe

un lundi les GMS anticipent un peu sur le vendredi précédent et récupèrent surtout le mardi. Si le jour férié est un jeudi, il y a de fortes chances que l'activité du vendredi soit réduite, car les salariés font le pont. Le mercredi précédent sera d'autant plus chargé.

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. L'usine approvisionne tous les magasins de France en chocolat de Pâques. Cet approvisionnement commence des mois avant l'évènement et monopolise une grande partie des ressources de l'agence. D'autres n'ont pas ce client et subissent moins de poussée ou elles la subissent à d'autres moments. Elles peuvent par exemple transporter le muguet du 1^{er} mai sur l'Hexagone.

Parfois, le 4^{eme} jour suivant un jour férié est aussi le 2^{eme} jour précédent un autre jour férié. C'est ce qui arrive en mai entre la fête du Travail (1^{er} mai) et la Victoire 1945 (8 mai). Dans ce cas, il est difficile de séparer les effets issus des deux jours fériés.

Pour tenter de séparer ces 4 phénomènes engendrés par la tombée d'un jour férié nous suivons une méthode inspirée de [LW04]. 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 évènements 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 [BJ70] qui a l'avantage d'être automatisable.

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

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 le carré de la différence 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 [BU07]. 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

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

, et sinon

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

2.5 Redressement des prévisions et intervalle de prévision

Les chroniques du poids passé à quai en « expédition », « distribution » et au « total » sont prévues séparément. Or le total doit être égal à la somme des « expéditions » et « distributions ». Pour rendre cohérentes les prévisions, il faut redresser les chiffres. Par leur importance, les chiffres du « total » sont moins variants, donc plus aisés à prévoir. Partant du principe que l'erreur de prévision de la somme est moins élevée que les erreurs cumulées des « expéditions » et « distributions », nous choisissons de garder la prévision de la série « total des marchandises passées à quai » pour corriger les deux autres. Cette technique est appelée « top-down agrégation ». La procédure d'ajustement est la suivante, soit

- \hat{X}_{t+h} le poids des expéditions prévues pour la date $t + h$,
- \hat{Y}_{t+h} le poids des distributions prévues pour la date $t + h$,
- \hat{Z}_{t+h} le poids total prévu pour la date $t + h$.

Nous cherchons α et β tel que $[\hat{Z}_{t+h} - (\alpha\hat{X}_{t+h} + \beta\hat{Y}_{t+h})]^2$ soit minimum, sous contrainte $1,5 > \alpha > 0,5$ et $1,5 > \beta > 0,5$. Les contraintes indiquent que les prévisions en « expédition » ou « distribution » ne peuvent pas se tromper de plus de 50%. La méthode Quasi-Newton a été choisie pour résoudre le problème d'optimisation.

Pour accompagner les valeurs prévisionnelles, nous calculons les intervalles de prévision bilatéraux au niveau de probabilité de 95%. L'intervalle de prévision peut être utile lorsque qu'une décision est difficile à prendre. Il peut par exemple décider l'attribution d'un camion en plus sur une tournée. D'autre part, l'intervalle de prévision peut donner une idée de la confiance à accorder à la prévision. Plus l'intervalle s'écarte moins la prévision est stable.

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 est 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 précédents la date T . T varie entre le 01/01/07 et le 31/07/07 avec un pas de six jours. Nous prédisons l'activité \hat{X}_{T+i} , \hat{Y}_{T+i} , \hat{Z}_{T+i} aux dates $T + 1$ à $T + 6$ et nous comparons cette prévision avec les séries chronologiques X_{T+i} , Y_{T+i} , Z_{T+i} effectivement réalisées pour $i = \{1, \dots, 6\}$.

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

TAB. 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%.

Un moyen de vérifier l'efficacité du modèle est d'étudier l'erreur de prévision. Elle doit suivre un comportement de processus bruit blanc. [Mél06] préconise d'utiliser le test de [LB78] pour tester l'autocorrélation des erreurs de prévision. Sur l'exemple précédent, les résultats du test (tableau 2)⁴, pour des retards variant de 1 à 6, nous laisse penser que le modèle a capturé la dépendance des 6 premiers retards sur la chronique « Total » et des 5 premiers retards sur la chronique « Distribution ». A contrario, les erreurs de prévision en « Expédition » gardent l'information permettant de prévoir leurs valeurs suivantes à partir des 6 précédentes. Nous ne pouvons donc pas conclure à l'hypothèse d'indépendance des erreurs de prévision en « Expédition ».

Retards	1	2	3	4	5	6
Expédition	0.003	0.0003	0.0009	0.0016	0.0039	0.0081
Distribution	0.896	0.742	0.5382	0.6284	0.1983	0.0007
Total	0.453	0.4718	0.6579	0.8027	0.4224	0.0538

TAB. 2 – p-value Ljung&Box Test

⁴avec un risque de 5%, l'hypothèse nulle est rejetée quand la p-value est inférieure à 0,05

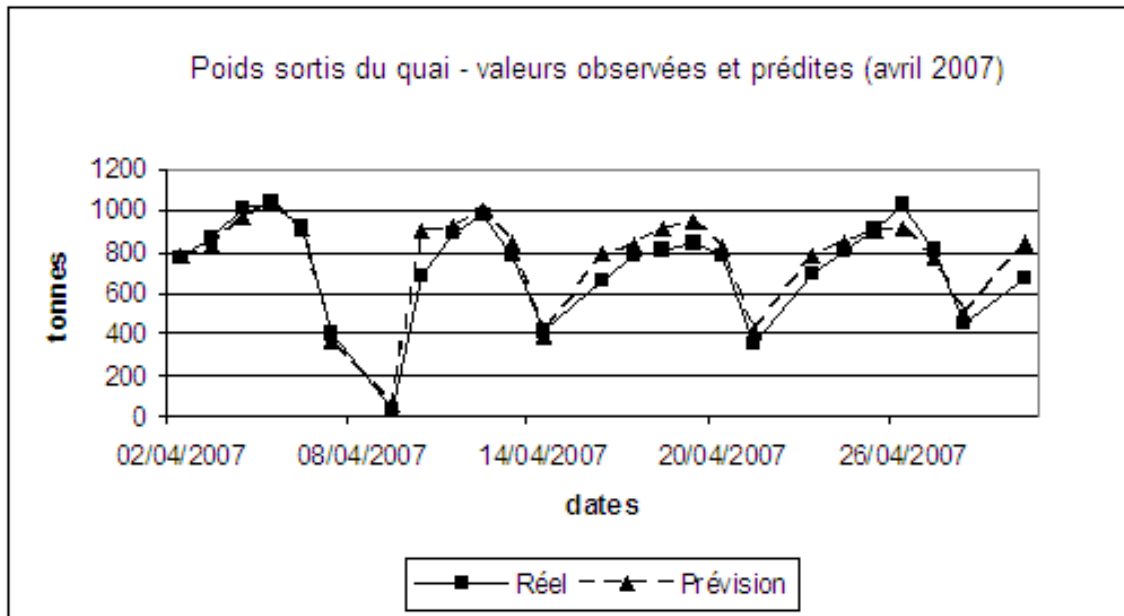


FIG. 1 – Comparaison observations réelles et prédites

Notons que ces résultats sont au moins aussi bons, sinon meilleurs que ceux obtenus par les modèles existants chez le transporteur (voir [MJ05]). Le gain du modèle réside dans le fait qu'il s'adapte aux chroniques des 57 agences du groupe. Il est dit intelligent dans le sens où, pour chaque prévision calculée, ses paramètres sont réestimés à partir de l'historique. Cependant, l'objectif d'une erreur de prévision annuelle moyenne, quelle que 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.

Ajoutons que la qualité des prévisions varie suivant la volatilité du comportement de l'activité. Une période d'activité stable (janvier à mars), permet d'extrapoler la chronique sans dégager d'erreurs importantes. Par contre, les périodes de turbulence (avril, mai) posent encore des problèmes (voir Fig. 1). Il apparaît que la méthode choisie pour mesurer l'effet des jours fériés sur l'activité n'est pas la plus adéquate. Pour être performante, la méthode nécessite plus d'historique. Rappelons qu'elle estime un coefficient de perte ou gain par MCO pour chacun des 4 jours passés et suivants un jour férié. Le cycle permettant de retrouver un même jour de la semaine pour un jour férié est 28 ans. Cependant, un cycle de 11 ans, voir 7 ans, pourrait être admissible. Le lecteur intéressé par la répartition des jours fériés peut se référer à la définition donnée sur le site de [Wik]. D'autre part, les observations sont éloignées dans le temps et le comportement des ménages change au cours du temps.

4 Conclusion et recherches futures

Les calculs de prévision sont centralisés au service d'informatique décisionnelle du transporteur. L'algorithme présenté dans cet article met à jour hebdomadairement, à horizons de 28 jours, les 6 séries chronologiques pour chacune des 57 agences du transporteur.

Les résultats sont publiés sur une interface web spécialement créée pour ce besoin. Ils sont présentés sous forme de tableaux de bord et de graphiques. Les prévisions étant justes 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 façon de réduire le nombre de paramètres est d'estimer la saisonnalité par une transformée de Fourier et de classer des événements calendaires en fonction de leurs effets. Il est également envisagé d'utiliser la corrélation qui peut exister entre les séries temporelles de 57 agences pour améliorer la modélisation. Les corrélations proviennent du fait que les agences forment un réseau et s'expédient de la marchandise entre elles.

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 [Mar99] 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.

Références

- [Aya05] Sofiane Ayadi. *Le Supply Chain Management : Vers une optimisation globale des flux*. Working paper, Université Catholique de Lyon, 2005.
- [BBD47] Buys-Ballot and C. Heinrich Diedrich. Les changements périodiques de température. 1847.
- [BG69] J.M Bates and C.W.J Granger. The combination of forecasts. *Operational Research Quarterly*, 20(4), 1969.
- [BJ70] George Box and Gwilym Jenkins. *Time series analysis : Forecasting and control*. San Francisco : Holden-Day Inc, 1970.
- [BM80] B. Burtschy and C. Menendian. A propos de prévision à court terme de la production industrielle. *Revue De Statistique Appliquée*, 28(2) :5–24, 1980.
- [BR96] J. Peter Brockwell and A. Davis Richard. *Introduction to Time Series and Forecasting*. Springer, 1996.
- [Bra06] François Branche. Transport de messagerie. *Revue technique de l'ingénieur*, dossier n°AG8151 V2, 2006.
- [BU07] Régis. Bourbonnais and Jean Claude Usunier. *Prévision des ventes, théorie et pratique*. Economica, 2007.

- [Clu06] Geoffroy Cluzel. Rentabilité d'un système d'information. approche théorique. *Revue technique de l'ingénieur*, dossier n°AG5310, 2006.
- [FC99] Nathalie Fadde-Costes. Système d'information logistique et transport. *Revue technique de l'ingénieur*, dossier n°AG8030 V2, 1999.
- [LB78] G.M. Ljung and G.E.P. Box. On a measure of lack of fit in time series models. *Biometrika*, 65 :297–303, 1978.
- [LW04] Francis A. Longstaff and W. Ashley Wang. Les changements périodiques de temperature. *The Journal of Finance*, 59(4) :1877–1900, 2004.
- [Mar99] M.-M. Martin. Filtrage de kalman d'une série saisonnière, application à la prévision de la consommation d'électricité. *Revue De Statistique Appliquée*, 47(4) :69–86, 1999.
- [MJ05] C. Métivier and B. Jaffrès. Maintenance de l'outil de prévision tonnages. Rapport interne, STEF-TFE, 07 2005.
- [Mél06] G. Mélard. Initiation à l'analyse des séries temporelles et à la prévision. *Modulad*, 35, 2006.
- [Sch06] S.P. Schnaars. An evaluation of rules for selecting an extrapolation model on yearly sales forecasts. *Interfaces*, 16 :100–1007, 2006.
- [Sei06] Nicholas Seiersen. Systèmes d'information logistique. *Revue technique de l'ingénieur*, dossier n°AG5300, 2006.
- [Ter04] Charlotte Terrolle. *Évolution des rapports entre industriels et grande distribution : du partenariat à la satisfaction clients, vers l'émergence de nouvelles stratégies d'achat*. Mémoire de fin d'études, Université Paris I Pantéhon-Sorbonne, 2004.
- [Vat93] Michel Vaté. *Statistique Chronologique et Prévisions*. Economica, 1993.
- [Wik] Wikipédia. Fêtes et jours fériés en france. <http://fr.wikipedia.org/wiki/F>