



**HAL**  
open science

**Construction, analyse et implémentation d'un modèle de prévision. Déploiement sous forme d'un système de prévision chez un opérateur européen du transport et de la logistique.**

Wilfried Despagne

► **To cite this version:**

Wilfried Despagne. Construction, analyse et implémentation d'un modèle de prévision. Déploiement sous forme d'un système de prévision chez un opérateur européen du transport et de la logistique.. Mathématiques [math]. Université Européenne de Bretagne, 2010. Français. NNT: . tel-00487327

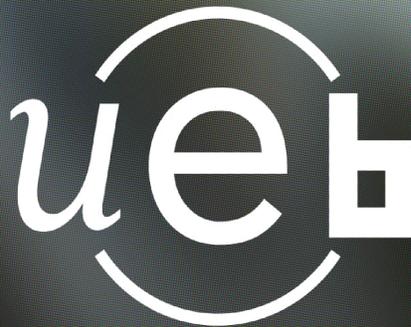
**HAL Id: tel-00487327**

**<https://theses.hal.science/tel-00487327>**

Submitted on 28 May 2010

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



THÈSE / UNIVERSITÉ DE BRETAGNE SUD

UFR Sciences et Sciences de l'Ingénieur  
sous le sceau de l'Université Européenne de Bretagne

Pour obtenir le grade de :  
DOCTEUR DE L'UNIVERSITÉ DE BRETAGNE SUD  
Mention : Mathématiques, Spécialité: Statistiques appliquées  
École Doctorale SICMA

présentée par

**Wilfried DESPAGNE**

Laboratoire des Sciences et Techniques de l'Information,  
de la Communication et de la Connaissance  
CNRS (UMR 3192)

# Construction, analyse et implémentation d'un modèle de prévision.

*Déploiement sous forme d'un système de  
prévision chez un opérateur européen du  
transport et de la logistique.*

Thèse soutenue le 1<sup>er</sup> Avril 2010,  
devant la commission d'examen composée de :

**M. Stéphane LALLICH**  
Rapporteur

**M. Régis BOURBONNAIS**  
Rapporteur

**M. Alain HILLION**  
Examinateur

**M. Serge CAPITAINE**  
Examinateur

**M. Emmanuel FRENOD**  
Directeur de thèse

**Mme. Valérie MONBET**  
Co-Directrice de thèse

## Construction, analyse et implémentation d'un modèle de prévision.

*Déploiement sous forme d'un système de prévision chez un opérateur européen du transport et de la logistique.*

*La thèse est le fruit d'un projet collaboratif. Ce projet innovant a réuni des acteurs majeurs de la formation et de la recherche (UBS, UEB, Lab-STICC) avec un professionnel du transport et de la logistique (STEF-TFE). L'ANRT a participé au financement par le biais d'une convention industrielle de formation par la recherche (CIFRE).*



FIGURE 1 – Les partenaires

# Résumé

Cette thèse montre que pour innover, le secteur privé peut faire appel à la connaissance académique. En effet, un partenariat entre un établissement universitaire et un professionnel du transport de marchandises a permis de répondre à des problématiques industrielles par l'application de mathématiques de pointe. Cette thèse porte sur la conception, l'implémentation et la mise en production d'un système de prévision d'activité pour l'optimisation de la planification des ressources. Elle propose également un modèle mathématique pour modéliser les flux de marchandises qui transitent sur un quai de messagerie.

Pour comprendre la problématique posée par l'industriel, le contexte industriel dans lequel s'insèrent les travaux de recherche est posé. Il s'en suit une réflexion sur la manière d'aborder le problème de mise en place d'un système de prévision en entreprise. Une synthèse de l'existant en matière de production d'information prévisionnelle est menée dans l'entreprise. En s'appuyant sur les informations récoltées, une méthodologie pour intégrer, analyser et prévoir des indicateurs économiques, est avancée. Cette méthodologie est appliquée avec succès dans le groupe STEF-TFE, leader français du transport sous température dirigée. Dans le but de planifier les ressources nécessaires pour faire face à l'activité prévue, une recherche a été menée pour modéliser les flux de marchandises en transit sur un quai de messagerie.

Le résultat de la thèse est, qu'en 2010, 70 agences de transports du groupe STEF-TFE ont accès aux prévisions d'activités. Aussi, la méthodologie avancée est susceptible d'être utilisée dans divers secteurs industriels et notamment ceux appartenant à la chaîne logistique.

# Abstract

This thesis shows for innovation purposes, the private sector can use academic knowledge. Indeed, a partnership between a university and a professional merchandise transportation business has forwarded responses to industrial issues through the application of advanced mathematics. This thesis is about the development, implementation and setting up of an activity forecasting system for planning and resource enhancement. It also offers a mathematical model for the merchandise flow that passes in transit through express parcel docks.

To better understand industrial issues, the industrial context of this work is set forth. A thought process on how to best approach the issue of setting up a forecasting system in a firm is then presented. A synthesis of the existing production means of forecasting information is carried out in the firm. Using the information gathered, a methodology to integrate, analyse and forecast economic indicators is set forth. This methodology is successfully applied in the STEF-TFE group, a French leader in cold logistic. To plan the necessary recourses to deal with the forecasted activity, research was carried out on modelling the flow of goods in transit on a express parcel dock.

The result of this thesis is that in 2010, 70 transportation agencies of the STEF-TFE group will have access to forecasts of the activity. Furthermore, the methodology set forth here is likely to be used in various industrial domains.

# Remerciements

A l'origine de ce projet de recherche il y a l'intervention de l'**OGGAM**<sup>1</sup> et une relation de longue durée entre l'**UBS**<sup>2</sup> et **STEF-TFE**. Grâce cette coopération un poste Cifre<sup>3</sup>, financé pour partie par l'**ANRT**<sup>4</sup> et pour une autre partie par STEF-TFE, a été créé.

Je tiens à remercier en tout premier lieu **Emmanuel Frénod** qui a dirigé cette thèse. Tout au long de ces trois années, il m'a accordé sa confiance. Tout en me laissant une liberté d'action, il a su orienter mes recherches dans le bon sens. Il a su me motiver dans les moments de doutes et a toujours été disponible pour d'intenses discussions. Je le remercie vivement.

Je remercie **Valérie Monbet**, co-directrice de la thèse, qui m'a enseigné la rigueur du vocabulaire statistique. Elle a aussi été d'une grande aide dans la rédaction du modèle statistique de prévision.

Je remercie les rapporteurs de cette thèse **Régis Bourbonnais** et **Stéphane Lallich** pour leurs réflexions avisées. Elles ont contribué à la lisibilité du document et enrichissent le fond des idées soulevées.

Merci également aux autres membres du jury qui ont accepté de juger ce travail : **Alain Hillion**, directeur du laboratoire Lab-STICC, et **Serge Capitaine**, Directeur Général Adjoint, Commerce et Marketing STEF-TFE. J'ajouterais que ce projet de recherche n'aurait pas pu voir le jour sans l'appui de Serge Capitaine.

Je voudrais remercier l'équipe Business Intelligence avec laquelle j'ai travaillé. Je commencerais par remercier les responsables, **Yann Morin** (avant 2008) et **Benoît Colas** (depuis 2008). Pour reprendre un ordre alphabétique, je remercie les membres de

---

1. Un réseau de compétences en TIC au service de l'innovation dans le monde de la recherche et de l'économie

2. Université de Bretagne Sud

3. Conventions industrielles de formation par la recherche

4. Association Nationale de la Recherche et de la Technologie

l'équipe : **Gaël Bernier, Vincent Cauchepin, Christophe Cuzin, Fabrice Le-Roux, Stéphane Oberlechner, Jean-Ghislain Robert, Vincent Thongdy** et **Yoann Touffait** pour leur aide. Je n'oublierai pas les aides permanentes reçues de l'équipe DEX<sup>5</sup> et tout spécialement de **Jérôme Leroyer** et **Eric Songis**.

Je tiens aussi à remercier **Marie-Claude Boyaval** et **Jean-Luc Despagne** pour avoir contrôlé l'orthographe et la grammaire du document.

Je remercie Agrostar<sup>6</sup>, mon employeur, de m'avoir accordé le temps nécessaire à la finalisation du manuscrit et à la préparation de la soutenance.

---

5. Direction d'exploitation

6. Service informatique du groupe STEF-TFE

# Table des matières

<b>Résumé</b>	<b>iii</b>
<b>Abstract</b>	<b>iv</b>
<b>Remerciements</b>	<b>v</b>
<b>Table des matières</b>	<b>vii</b>
<b>Liste des tableaux</b>	<b>xi</b>
<b>Table des figures</b>	<b>xiii</b>
<b>Introduction</b>	<b>1</b>
<b>1 Contexte général</b>	<b>6</b>
1.1 Préambule . . . . .	6
1.1.1 Transport sous température dirigée . . . . .	6
1.1.2 Groupe STEF-TFE . . . . .	16
1.2 A l'écoute de l'industriel . . . . .	25
1.2.1 Enjeux de la prévision d'activité chez TFE . . . . .	26
1.2.2 Suivi des marchandises dans le réseau TFE . . . . .	29
1.2.3 Méthodologie et prérequis à l'installation d'un système de prévision	30
1.2.4 Définition des besoins de TFE en information prévisionnelle . . . . .	32
1.2.5 Intégration du système de prévision au système d'information de TFE	36
1.3 Une démarche de projet industriel et scientifique . . . . .	38
1.3.1 Analyse et découpage du projet de recherche . . . . .	39
1.3.2 Risque et estimation des coûts du projet de recherche . . . . .	48
<b>2 Etude d'un système de prévision et de reporting</b>	<b>51</b>
2.1 Les concepts autour de la prévision des ventes . . . . .	51
2.1.1 De la demande aux ventes . . . . .	51
2.1.2 De la prévision au Système de prévision . . . . .	52
2.1.3 Le système de prévision : un outil transversal . . . . .	55

2.1.4	Le système de prévision : un outil d'aide à la décision et de productivité . . . . .	56
2.2	Les questions préalables à la mise en place d'un système de prévision de la demande . . . . .	59
2.2.1	Que doit permettre la prévision de la demande ? . . . . .	59
2.2.2	Quelles sont les données sources de la prévision des ventes ? . . . . .	60
2.2.3	Quels sont les paramètres de prévision des ventes ? . . . . .	62
2.2.4	Quelles sont les techniques à choisir pour prévoir les ventes ? . . . . .	64
2.2.5	Pourquoi prendre en compte l'expertise terrain pour valider la prévision des ventes ? . . . . .	67
2.2.6	Quel budget est à accorder à la prévision des ventes ? . . . . .	68
2.2.7	Quelle est la taille du portefeuille client ? . . . . .	68
2.2.8	Quel est le nombre de nouveaux produits ? . . . . .	68
2.2.9	Existe-t-il des différences régionales ? . . . . .	69
2.2.10	Quelles sont les compétences du prévisionniste ? . . . . .	69
2.2.11	Quelle est l'erreur de prévision acceptable ? . . . . .	70
2.2.12	Quel est le format de restitution de la prévision des ventes ? . . . . .	71
2.2.13	Quelle procédure suivre pour aboutir à une prévision des ventes juste ? . . . . .	72
2.3	Les composants du système de prévision . . . . .	74
2.3.1	Environnement . . . . .	75
2.3.2	Organisation . . . . .	76
2.3.3	Techniques . . . . .	78
2.3.4	Système informatique . . . . .	83
2.3.5	Mesures de performance . . . . .	90
2.4	Le processus collaboratif . . . . .	92
<b>3</b>	<b>Analyse de l'existant avant le projet de recherche Horizons</b>	<b>96</b>
3.1	La structure de l'information chez TFE et son interrogation . . . . .	97
3.1.1	Interaction entre le système d'information et le système de prévision	97
3.1.2	Segmentation des flux de marchandises . . . . .	100
3.1.3	Source des heures productives sur le quai . . . . .	105
3.1.4	Structure et agencement des données sources . . . . .	106
3.2	Les modèles de prévision utilisés chez TFE . . . . .	107
3.2.1	Notions autour des séries chronologiques . . . . .	108
3.2.2	Modèle économétrique de TFE . . . . .	110
3.2.3	Modèle endogène de TFE - lissage par moyennes mobiles . . . . .	115
3.2.4	Modèle endogène de TFE - méthode de Holt-Winters . . . . .	120
3.2.5	Outil informatique utilisé . . . . .	123
<b>4</b>	<b>Le système de prévision Horizons : Conception, développement et mise en</b>	

<b>œuvre</b>	<b>125</b>
4.1 Nouvelle organisation de la prévision chez TFE . . . . .	125
4.2 Harmonisation et centralisation de l'information . . . . .	126
4.3 Modèles de prévision proposés en analyse statistique . . . . .	128
4.4 Modèle statistique de prévision proposés par « Horizons » . . . . .	133
4.4.1 Historique . . . . .	134
4.4.2 Nettoyage des données . . . . .	135
4.4.3 Schéma de décomposition . . . . .	139
4.4.4 Décomposition d'une chronique : méthode n°1 . . . . .	140
4.4.5 Décomposition d'une chronique : méthode n°2 . . . . .	146
4.4.6 Evénements calendaires . . . . .	150
4.4.7 Cycle . . . . .	153
4.4.8 Modèle complet . . . . .	153
4.4.9 Prévision . . . . .	154
4.4.10 Recomposition de la prévision . . . . .	156
4.4.11 Combinaison des prévisions . . . . .	157
4.4.12 Redressement des prévisions . . . . .	158
4.4.13 Contrôle de la qualité d'ajustement par le test du bruit blanc . . . . .	159
4.4.14 Avantages et inconvénients . . . . .	161
4.5 Validation du modèle mathématique proposé . . . . .	162
4.5.1 Critères de comparaisons . . . . .	162
4.5.2 Exemple : L'agence de transport TFE à Nantes . . . . .	164
4.5.3 Etude des variations des erreurs de prévision sur une trentaine d'agences . . . . .	168
4.6 Choix et cartographie des outils informatiques . . . . .	168
4.6.1 Outil ETL de Pentaho™ . . . . .	169
4.6.2 Logiciel de traitement statistique . . . . .	172
4.6.3 Représentation des résultats . . . . .	176
4.6.4 Interfaçage Kettle / R / site web . . . . .	177
4.7 Développement des modules du système de prévision . . . . .	178
4.7.1 Rédaction du modèle mathématique en R . . . . .	178
4.7.2 Alimentation des tables sources (oracle©) par Kettle . . . . .	183
4.7.3 Interface graphique : partie émergée de l'iceberg . . . . .	186
4.8 Mode d'emploi du système de prévision « Horizons » . . . . .	193
4.9 Retour sur investissement . . . . .	197
4.10 Industrialisation : Prévision à moyen terme . . . . .	200
<b>5 Simulation des flux de marchandises sur un quai de messagerie</b>	<b>201</b>
5.1 Modèle « fin » . . . . .	202
5.2 Modèle « agrégé » . . . . .	208
5.3 Modèle de l'organisation des quais en réseau . . . . .	210

5.4	Tests de validation . . . . .	212
5.4.1	Résultats obtenus par le modèle « fin » . . . . .	213
5.4.2	Résultats obtenus par le modèle « agrégé » . . . . .	218
5.5	Perspectives . . . . .	220
5.5.1	Calcul de la quantité de main d'œuvre . . . . .	220
5.5.2	Allocation de la main d'œuvre à chaque position . . . . .	221
5.5.3	Optimiser le temps $\tilde{\nu}(p)$ . . . . .	222
	<b>Conclusion</b>	<b>229</b>
	<b>Bibliographie</b>	<b>231</b>
	<b>Glossaire</b>	<b>240</b>

# Liste des tableaux

1.1	<i>Palmarès des principales sociétés de transport frigorifique</i> <i>Sources : L'officiel des transporteurs magazine, Novembre 2004</i> <i>(1) ensemble routier multi-températures</i> <i>(2) activité frigo seule . . . . .</i>	<i>11</i>
1.2	<i>tableau extrait de l'interface web . . . . .</i>	<i>38</i>
1.3	<i>Le cycle de R&amp;D aboutissant au modèle mathématique et à son programme informatique en quelques chiffres . . . . .</i>	<i>42</i>
1.4	<i>procédure ETL . . . . .</i>	<i>43</i>
1.5	<i>Le cycle de conception et de mise en œuvre du lot 2 en quelques chiffres</i>	<i>43</i>
1.6	<i>Le cycle de conception et mise en œuvre du lot 3 en quelques chiffres . .</i>	<i>45</i>
1.7	<i>Planning grosse maille du projet « Horizons » . . . . .</i>	<i>47</i>
1.8	<i>Analyse du risque lié au projet « Horizons » . . . . .</i>	<i>49</i>
2.1	<i>Schéma additif . . . . .</i>	<i>80</i>
2.2	<i>Schéma multiplicatif . . . . .</i>	<i>81</i>
3.1	<i>Avantages et inconvénients du modèle économétrique de TFE . . . . .</i>	<i>115</i>
3.2	<i>Exemple des poids jours calculés par trimestre . . . . .</i>	<i>118</i>
3.3	<i>Exemple des poids jours des semaines avec jour férié . . . . .</i>	<i>119</i>
3.4	<i>Avantages et inconvénients du modèle endogène de TFE - lissage par moyennes mobiles . . . . .</i>	<i>120</i>
3.5	<i>Coefficients journaliers du modèle endogène de TFE - méthode de Holt-Winters . . . . .</i>	<i>122</i>
3.6	<i>Avantages et inconvénients du modèle endogène de TFE - méthode de Holt-Winters . . . . .</i>	<i>123</i>
4.1	<i>Tables créées dans l'Infocentre pour le besoin du projet Horizons . . . . .</i>	<i>129</i>
4.2	<i>Matrice des événements calendaires - Un champ date, un champ pour chaque évènement calendaire analysé, un champ pour la distante entre l'enregistrement et le jour férié. . . . .</i>	<i>152</i>
4.3	<i>Avantages et inconvénients du modèle de prévision « Horizons » . . . . .</i>	<i>161</i>
4.4	<i>Moyenne des erreur de prévision . . . . .</i>	<i>164</i>
4.5	<i>Ecart-type des erreur de prévision . . . . .</i>	<i>165</i>

4.6	<i>Indicateurs de performance</i>	165
4.7	<i>Indicateurs de performance, suite</i>	166
4.8	<i>Pourcentage de prévisions justes</i>	166
4.9	<i>Pourcentage de prévisions justes, suite</i>	167
5.1	<i>Paramètres de position</i>	213
5.2	<i>Paramètres de position, suite</i>	214
5.3	<i>Paramètres de quai</i>	214
5.4	<i>Paramètres d'équipe de quai</i>	215
5.5	<i>Paramètres d'équipe de quai, suite</i>	215
5.6	<i>Paramètres du modèle</i>	216
5.7	<i>Paramètres de position</i>	218
5.8	<i>Paramètres de quai</i>	218
5.9	<i>Paramètres de quai, suite</i>	219
5.10	<i>Paramètres d'équipe de quai</i>	219
5.11	<i>Paramètres du modèle</i>	219

# Table des figures

1	<i>Les partenaires</i> . . . . .	ii
1.1	<i>La chaîne logistique</i> . . . . .	14
1.2	<i>Localisation des sites TFE en Europe en 2007</i> . . . . .	17
1.3	<i>Situation géographique des agences TFE</i> . . . . .	18
1.4	<i>Voyage des flux physiques chez TFE</i> . . . . .	20
1.5	<i>circuit de la marchandise de l'expéditeur au destinataire</i> . . . . .	21
1.6	<i>Schéma d'un quai TFE</i> . . . . .	23
1.7	<i>Flux engendrés par le réseau</i> . . . . .	29
1.8	<i>Organigramme d'un système de prévision</i> <i>Source : R. Bourbonnais et JC Usunier</i> . . . . .	30
1.9	<i>Matrice de l'activité d'une agence TFE</i> . . . . .	34
1.10	<i>Les horizons de prévision</i> . . . . .	35
1.11	<i>Cycle de vie d'un projet en SI</i> . . . . .	39
2.1	<i>Interaction entre système d'information et prévision des ventes : Démarche transversale (Mentzer et Moon, 2005) (85)</i> . . . . .	56
2.2	<i>Pyramide produit</i> . . . . .	61
2.3	<i>Arbre de sélection pour les méthodes de prévision Armstrong, 1999 (8)</i> . . . . .	66
2.4	<i>Processus de prévision des ventes, Mentzer et Moon, 2005 (85)</i> . . . . .	75
2.5	<i>Chaîne logistique dont le stock de sûreté cumulé est de 110,5%</i> . . . . .	94
3.1	<i>Architecture du Système d'Information chez STEF-TFE</i> . . . . .	98
3.2	<i>Architecture du Réseau Informatique chez STEF-TFE</i> . . . . .	99
3.3	<i>Traduction des flux physiques en flux informatiques</i> . . . . .	102
3.4	<i>Le découpage de l'activité transport en segments</i> . . . . .	103
3.5	<i>Une partie du modèle relationnel de la base GTI</i> . . . . .	107
3.6	<i>Représentation de la tendance, la saisonnalité et l'effet d'une variable explicative</i> . . . . .	113
4.1	<i>Serie chronologique du poids quotidien, en kg, total, des marchandises passées à quai avant et après nettoyage des valeurs extrêmes</i> . . . . .	137
4.2	<i>Tendance par palier</i> . . . . .	141
4.3	<i>Poids hebdomadaire en tonne passé à quai entre 2001 et 2007</i> . . . . .	142

4.4	<i>Superposition des poids quotidiens, en tonne, passé à quai, entre 1999 et 2007, d'une agence</i>	143
4.5	<i>Représentation graphique de la saisonnalité</i>	146
4.6	<i>Découpage de l'historique en échantillon d'Apprentissage et échantillon Test</i>	162
4.7	<i>Représentation graphique de la chronique des poids quotidiens passés à quai entre Janvier et Février 2007</i>	166
4.8	<i>Représentation graphique de la chronique des poids quotidiens passés à quai entre Avril et Mai 2007</i>	167
4.9	<i>Dispersion des pourcentages d'erreur quotidien commis sur la prévision du poids des marchandises passées à quai dans 36 agences</i>	169
4.10	<i>Kettle gère des flux ETL (gauche) par une interface graphique simple (droite)</i>	171
4.11	<i>Cartographie des applications retenues dans le système de prévision « Horizons »</i>	177
4.12	<i>Arbre du package R regroupant les fonctions de prévisions « Horizons »</i>	179
4.13	<i>Impression écran d'une tâche Kettle de mise à jour quotidienne des indicateurs</i>	186
4.14	<i>Impression écran d'une transformation Kettle d'extraction des données et de leur transformation pour chargement</i>	187
4.15	<i>Impression écran de la tranformation Kettle d'extraction des données et de leur transformation pour chargement</i>	188
4.16	<i>Page d'accueil du site web « Horizons »</i>	188
4.17	<i>Propositions de prévision journalière sur le site web « Horizons »</i>	189
4.18	<i>Indicateurs de confiance aux prévisions sur le site web « Horizons »</i>	191
4.19	<i>Interface pour ajuster les propositions de prévision sur le site web « Horizons »</i>	192
4.20	<i>Process d'utilisation du système de prévision « Horizons »</i>	194
4.21	<i>Valeur extrême identifié rapidement grâce à l'interface web « Horizons »</i>	197
5.1	<i>Décomposition du quai en 5 compartiments</i>	202
5.2	<i>Répartition de la masse des marchandises sur les différentes zones du quai, suivie du détail pour chaque position</i>	224
5.3	<i>Répartition de la masse des marchandises sur les différentes zones du quai, avec <math>n^{Res} = 0,08</math></i>	225
5.4	<i>Répartition de la masse des marchandises sur les différentes zones du quai, avec <math>S_{max}^{Res} = 1</math></i>	226
5.5	<i>Répartition de la masse des marchandises sur les différentes zones du quai, avec <math>n^{In} = 0,5</math></i>	227
5.6	<i>Incohérence dans la répartition de la masse des marchandises sur les différentes zones du quai</i>	227
5.7	<i>Superposition des résultats du modèle « fin » avec ceux du modèle « agrégé »</i>	228

# Introduction

« Gouverner c'est prévoir », cette maxime se justifie également dans le milieu économique. Pour adapter leur stratégie face à la crise économique de 2008 et 2009, les entrepreneurs cherchent à anticiper l'activité de leur entreprise. C'est un enjeu stratégique. Pour piloter l'entreprise, les dirigeants ont besoin d'informations prospectives comme des indicateurs de croissance, de rentabilité et de performance opérationnelle. L'objectif annoncé des prévisions est un programme de maîtrise des coûts par le pilotage des ressources humaines et matérielles au service des besoins prévus.

Prévoir n'est pas chose simple. La crise ajoute, à la difficulté de prévoir, un manque de visibilité, une incertitude économique et financière, et pousse les opérateurs économiques à la prudence. Malgré cela, et le fait que la publication d'information prospective reste facultative, 83%<sup>7</sup> des entreprises du CAC 40 ont fourni des prévisions pour l'année 2009. Pourtant, en communiquant la vision qu'ils ont du futur, les dirigeants engagent leur crédibilité. Mais les prévisions sont de nature à rassurer. Elles fournissent une vision prospective permettant d'agir. Si les prévisions sont de nature pessimiste, elles conduisent à un programme de réduction des coûts alors que des prévisions optimistes conduisent à des investissements.

Cette thèse se propose d'étudier la conception, le développement et la mise en œuvre d'un système de prévision. A travers le cas de STEF-TFE, leader français de la logistique du froid, entreprise cotée en bourse<sup>8</sup>, la thèse expose un processus rigoureux d'élaboration des prévisions. Les résultats obtenus ont permis de valider ce processus. L'originalité de cette thèse est d'utiliser une démarche scientifique pour répondre à une problématique industrielle.

Le projet a pris forme grâce à une relation de longue durée entre l'Université de Bretagne Sud et le groupe STEF-TFE. La volonté des dirigeants du groupe d'investir dans la recherche opérationnelle, et la détermination des chercheurs universitaire à accompagner le groupe, ont permis de définir un sujet de recherche et de donner les moyens

---

7. Prévisions 2009, Philippe Kubisa, Associé, Elodie Fornas, Global Capital Markets Group, PricewaterhouseCoopers

8. STEF-TFE, valeur boursière cotée sur le marché du Compartiment B

d'y répondre. Pour résumer à l'extrême, la formulation de l'objectif initial du projet était de fournir une prévision des quantités de marchandises à transporter quotidiennement par la branche transport du groupe. Cette branche concentre son expertise autour du **groupage**, de la **messagerie**, du transport de **lot complet** et des flux internationaux. Le besoin de prévisions fiables est d'autant plus sensible que le secteur du transport frigorifique fonctionne en flux tendu à cause du caractère périssable de certains produits. La branche transport du groupe STEF-TFE regroupe plus de 70 **agences** de groupage en France et en Europe. Les camions des agences ramassent des marchandises chez les clients pour les livrer chez leur destinataire ou les expédier vers d'autres agences de groupage qui se chargent alors de la livraison au destinataire final. Pour aboutir à un système de prévision homogène aux 70 agences, il faut harmoniser les données à partir desquelles sont élaborées les prévisions, construire un système de prévision commun, convaincre de son efficacité et former le personnel.

L'impact d'une mauvaise prévision des quantités de marchandises à transporter se traduit soit par des manutentionnaires en surnombre sur les quais d'approvisionnement ou de déchargement, ce qui représente des coûts inutiles ; soit par un manque de personnel, ce qui contrarie le client. En somme, la prévision est un élément essentiel de la maîtrise de la qualité des services et d'une gestion optimisée des ressources. Elle est aussi le préalable incontournable de toute prise de décision (voir section 1.2).

Mettre en œuvre la prévision des quantités de marchandises à transporter s'avère difficile. Ces quantités sont en effet le résultat de séries chronologiques qui concernent non seulement des facteurs ayant un comportement stochastique, mais aussi des facteurs explicatifs (météo, jours fériés, jours de grèves, impact sur le consommateur de l'actualité : grippe aviaire et vache folle). La prévision demande donc de faire appel à des méthodes de modélisation statistique, probabiliste et déterministe (voir section 2.3.3).

Prévoir n'est pas tout, il faut que la prévision puisse servir dans le quotidien. Elle doit être utilisée pour dimensionner les besoins en ressources humaines. Le personnel doit être suffisamment flexible pour répondre aux fluctuations d'activité et le budget doit permettre d'allouer des ressources supplémentaires (voir sections 2.1.3 et 2.1.4).

D'autre part pour qu'une prévision soit envisageable il faut que la grandeur à prévoir soit une reproduction de situations s'étant produites dans le passé. Un historique des événements passés est donc nécessaire. S'il existe, l'historique n'est souvent pas uniformisé. La définition de la canicule n'est pas la même suivant que l'on se trouve à Lille ou à Madrid. La notion de qualité de service attendu par les clients est aussi très différente selon que l'on se trouve en France ou en Espagne. Cette confusion peut être lourde de conséquences si nous utilisons des grandeurs sans les uniformiser dans le processus de

décision (voir section 2.2.2 et 2.2.9).

Il faut aussi cibler les facteurs qui ont une influence significative sur la quantité de marchandises transportée. Prévoir suppose de pouvoir discerner des corrélations entre ces données ; par exemple entre une augmentation de la température en été de deux degrés de plus que la moyenne et la hausse des quantités de glaces (sorbets) à transporter vers les supermarchés. Nous devons également dissocier les facteurs influant sur le long terme de ceux liés au rythme saisonnier (fête, vacance, solde) ou encore des facteurs événementiels prévisibles ou non comme une épidémie de grippe aviaire (voir section 2.4 et 2.3.1).

Le processus de prévision nécessite d'estimer des paramètres (complexes) du modèle en opérant des calculs statistiques. Cela ne peut pas se faire à l'aide d'outils bureautiques simples. Il faut alors envisager de programmer un logiciel permettant de prendre en compte de nombreux composantes et qui puisse être implémenté au niveau industriel dans toutes les succursales du groupe (voir section 2.3.4 et 4.6).

Par la suite il faut prendre en compte l'efficacité du modèle. Il est capital de contrôler l'écart entre le prévu et le réel ; lui attribuer un ordre de grandeur et le cas échéant améliorer la capacité prédictive du modèle en détectant ses dérives. Pour chacune des prévisions, il faut mesurer sa fiabilité (voir section 2.3.5 et 4.5).

La prévision en entreprise n'est ni une décision prise par un manager, ni un chiffre issu d'une boîte noire, c'est un ensemble d'outils, de procédures et d'organisations permettant de parvenir à un résultat prévisionnel fiable. En effet, une bonne prévision est plus qu'un chiffre (voir section 2.2.11), elle doit être utilisée en concordance avec d'autres informations disponibles. C'est ce que nous appelons un système de prévision.

Le processus de prévision imaginé commence par l'identification de l'environnement dans lequel se place l'entreprise. Ainsi, le chapitre 1 présente le contexte industriel dans lequel s'insèrent les travaux de recherche. Il fait ressortir les objectifs industriels, les moyens accordés et les décisions prises pour utiliser ces moyens afin de répondre aux objectifs. Le chapitre 1 introduit aussi les éléments théoriques qui sont développés dans le chapitre 2. Ces premiers éléments permettent de mieux comprendre le contexte. Enfin, le chapitre 1 dévoile un aperçu de ce qui a été réalisé et les obstacles franchis pour y parvenir.

Le chapitre 2 est la bibliographie pour présenter le contexte scientifique. Le chapitre 2 s'appuie, entre autres, sur deux grandes références qui sont l'ouvrage de R. Bourbonnais et J.C. Usunier s'intitulant « Prévision des ventes » (18), et l'ouvrage de T.J. Mentzer

et A.M. Moon s'intitulant « Sales Forecasting Management » (85). Le chapitre 2 énumère les points importants à retenir pour la mise en place d'un système de prévision. Notons la fiabilité des données sources, la déclaration des hypothèses sur lesquelles reposent les prévisions, la validation des informations prévisionnelles par les organes dirigeants, la documentation des méthodes de calcul des prévisions, l'horizon de temps, l'agrégation ou désagrégation des prévisions, la sensibilité des prévisions aux variations, les supports de communication, etc.

Le chapitre 3 est un audit. Il présente l'existant en matière de prévision d'activité et système d'information présent dans l'entreprise. Cette synthèse de l'existant, inédite jusqu'à présent, a contribué grandement à la définition précise des besoins de prévision, Le bilan de l'existant présenté dans le chapitre 3 a été un élément moteur à la définition précise des besoins de prévision, à la normalisation des indicateurs, à la rédaction d'un cahier des charges et à l'élaboration des recommandations.

Le chapitre 4 présente les apports personnels. Nous documentons les outils mathématiques et informatiques choisis. Par une démarche scientifique, nous apportons une méthodologie pour intégrer, analyser et prévoir des indicateurs économiques. Cette méthodologie a été testée à grande échelle. Elle a été approuvée par les résultats obtenus. Elle a aussi été réutilisée pour prévoir des grandeurs mensuelles. La méthodologie est développée et les choix qui ont été faits sont argumentés. Le produit final est présenté et les recommandations d'utilisation sont données. Ainsi, du développement à l'industrialisation, en passant par la formation du personnel et la communication, la mise en place d'un système de prévision à l'échelle du groupe est entièrement documentée.

Suite aux prévisions d'activité, l'objectif suivant de la thèse est d'effectuer un premier pas dans la modélisation des flux de marchandises circulant sur le quai.

Le réseau d'agences de groupage et messagerie du groupe STEF-TFE permet de proposer à ses clients des livraisons sur tout l'Hexagone en moins de 48h. Le fonctionnement d'un tel réseau nécessite un pilotage rigoureux à l'aide d'un ensemble de procédures et des logiciels permettant de gérer les flux d'information, les flux physiques et des interfaces entre les différents acteurs. Cet ensemble est connu sous le nom de supply chain management (SCM) ou gestion de la chaîne logistique (GCL) en français. Parvenir à la maîtrise des flux (physique et informatique) suppose d'industrialiser les processus : optimisation des plans de transport, synergie des flux, introduction de l'informatique, etc. Les outils SCM ont considérablement aidé les transporteurs à améliorer leur productivité. L'outil phare du transporteur est le TMS (Transport Management System). Il est l'équivalent du WMS (Warehouse Management System) utilisé pour la gestion des entrepôts (voir section 3.1.1). Le TMS permet de planifier et optimiser les voyages, d'exécuter des opérations courantes, de dialoguer avec le client et par l'intermédiaire du reporting,

d'analyser les capacités du transporteur.

Une activité de recherche a été menée pour transporter le maximum de produits au moindre coût. Des logiciels académiques et commerciaux, basés sur des algorithmes de méthodes numériques et d'optimisation, permettent d'optimiser les plans de transport tout en satisfaisant la grande majorité des contraintes imposées. En revanche, la compréhension et la modélisation d'un quai de messagerie sont un domaine peu exploré. L'objectif du dernier chapitre (5) de la thèse est de contribuer de manière significative à son exploration.

Dans un premier temps, les flux physiques et informatiques occasionnés par le transport d'une marchandise sont décrits. Ces informations nous permettront de définir les entrées et sorties du quai de messagerie. Ensuite deux modèles décrivant le comportement d'un quai de messagerie, à deux échelles différentes, à l'aide d'équations différentielles ordinaires, sont construits. Ces deux modèles sont ensuite implantés dans l'environnement R<sup>9</sup>. Quelques tests, sur des exemples simples, démontrent la pertinence d'utiliser ces modèles pour simuler le temps de traitement de la marchandise sur un quai ou pour répartir la main d'œuvre aux différentes tâches.

Bien que non implémentée, une méthodologie pour construire une mise en réseau de modèles de quai est décrite. Plusieurs questions industrielles sont également posées en termes de problèmes d'optimisation utilisant les modèles construits.

---

9. langage de programmation pour le traitement de données et l'analyse statistique

# Chapitre 1

## Contexte général

### 1.1 Préambule

#### 1.1.1 Transport sous température dirigée

Pour prévoir les ventes, il faut s'intéresser à l'environnement de l'entreprise (voir section 2.3.1). C'est aussi l'objet des quatre sections suivantes. Elles introduisent l'activité qui consiste à transporter de la marchandise sous température dirigée, le milieu géographique dans lequel est implantée cette activité, son contexte économique et son fonctionnement.

##### 1.1.1.1 L'activité transport sous température dirigée

Les notions abordées ci-dessous nous permettent d'introduire l'importance de la prévision des ventes. Nous indiquons que les ventes d'un transporteur frigorifique, sont les transports réalisés entre un point  $A$  et un point  $B$  tout en maintenant la chaîne du froid. Selon la désignation de ces deux points, le transport est appelé **ramasses**, **expédition** ou distribution (voir section 3.1.2). Il nécessite du personnel routier, manutentionnaire, administratif et une infrastructure importante (camions, remorques, quais de messagerie, bureaux). Le transport est régi par des lois. Ainsi, pour respecter les contraintes, il est nécessaire d'anticiper. C'est par la prévision des ventes, que le transporteur sera à même d'anticiper son activité et dès lors planifier ses ressources humaines et matérielles. Commençons par introduire l'activité d'un transporteur frigorifique.

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°C. Il est nécessaire de distinguer le froid positif qui correspond à la logistique des produits frais, du froid négatif traitant celle des produits surgelés (56). La marchandise est acheminée d'un point *A* vers un point *B*. 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 des médicaments. Elles ont en commun d'être soumises à un cahier des charges très strict qui est défini dans la « chaîne du froid ».

La « chaîne du froid » (113) est le processus qui permet de maintenir un produit à basse température. Le froid ralentit la propagation des micro-organismes. 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. Tous les moyens de préventions et d'autocontrôle doivent être mis en œuvre pour vendre des denrées ne représentant aucun risque pour la santé. Suivant la catégorie des produits, des températures sont imposées. Les fruits et légumes doivent être maintenus de 7°C à 8°C, les produits frais et ultra frais de 2°C à 3°C et les surgelés à -20°C.

A 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 impose des règles en matière de durée de travail d'un chauffeur routier et d'autorisation de circulation. Le temps de travail journalier d'un chauffeur ne peut pas excéder 10 heures, son temps de conduite sans interruption est limité à 4h30. La circulation des véhicules de plus de 7,5 tonnes affectés au transport routier de marchandises est interdite en France les samedis et veilles de jours fériés à partir de 22 heures, jusqu'à 22 heures les dimanches et jours fériés. Le transporteur frigorifique bénéficie d'une dérogation quand il transporte des denrées alimentaires, il a l'autorisation de circuler dans un rayon de 150 km du premier point de livraison ou de collecte.

Ce ne sont ici que des exemples de lois parmi d'autres pour percevoir l'ampleur des contraintes législatives en matière de transport et plus encore de transport frigorifique. Les obstacles ne s'arrêtent pourtant pas là. Le transporteur frigorifique doit faire face à d'importants investissements d'infrastructures. Les équipements spécifiques à un transporteur sous température dirigée sont, entre autres, un parc de véhicules et un parc immobilier frigorifique.

Les camions frigorifiques doivent obtenir l'agrément de CAMEFROID (organisme des services sanitaires nationaux) (88). Ces véhicules se déclinent en trois catégories :

- les véhicules utilitaires légers permettant de livrer les commerçants du centre-ville,

- les camions porteurs destinés à transporter des charges lourdes. Ils possèdent sur le même châssis la cabine et le frigo. Ils sont des véhicules de livraisons, urbains ou régionaux.
- les tracteurs ou semi-remorques sont des véhicules de plus de 7,5 tonnes. Les tracteurs sont composés d'une partie motrice comprenant la cabine et d'un châssis auxquels une caisse frigorifique est attelée. Ils sont destinés à des longs trajets.

Quel que soit le véhicule, il est équipé d'une cellule isotherme. Elle permet de maintenir la température à un niveau constant entre -25°C et +25°C. Certaines cellules sont capables de se diviser pour proposer des compartiments à différentes températures. Elles sont appelées bi-températures. Elles fonctionnent en frais/surgelés ou suivant deux températures de frais. Les chauffeurs prennent la température des denrées à leur chargement. La température est contrôlée au déchargement à quai. La procédure stipule qu'à chaque manquement de la marchandise, sa température doit être sondée. Ainsi, en cas de litige, il est possible de revenir au dernier contrôle. Dans les véhicules, un mécanisme enregistre à intervalles réguliers la température du compartiment. Ainsi, en cas de problème alimentaire, comme la listériose, il est facile de tracer les températures<sup>1</sup>.

Le métier du transporteur consiste à collecter la marchandise chez les clients **expéditeurs** (souvent ce sont des producteurs industriels), de la regrouper selon les clients destinataires (souvent ce sont les points de vente) puis de les livrer. Pour regrouper et dégroupier la marchandise, le transporteur a besoin d'un quai frigorifique (1) (voir section 1.1.2.2.3). C'est un entrepôt équipé d'une ventilation lui permettant d'assurer des températures inférieures à 8°C. Le **quai** permet d'accueillir des camions en entrée comme en sortie. Les uns déchargent pendant que les autres chargent la marchandise. Ce flux d'entrées - sorties résume l'activité d'un quai d'un transporteur frigorifique. En entrée, nous définissons les **ramasses** et les **expéditions** reçues alors qu'en sortie nous distinguons les expéditions et les livraisons 1.1.2.2.3,5.1.

Les **ramasses** désignent les flux des expéditeurs vers les transporteurs. Les **expéditions** (ou messagerie) sont les flux internes au réseau du transporteur. Les **livraisons** ou **distributions** sont les flux du transporteur vers les destinataires finaux. Un client peut expédier des colis en groupage ou en lots complets. Le **groupage** signifie que ces colis seront expédiés avec d'autres venant de clients différents. Le **lot complet** désigne le transport d'un lot en direct depuis le lieu de chargement jusqu'au lieu de déchargement et justifiant l'usage d'un véhicule exclusif (20).

---

1. Transporteur & technologies, n°27, nov. 2001

### 1.1.1.2 Le contexte géographique

Comprendre l'interaction entre les acteurs, leur site géographique et leur rôle (distributeur ou industriel) permet d'enrichir la qualité des prévisions. A l'inverse, anticiper les comportements des acteurs à long terme permet de prendre des décisions stratégiques comme l'implantation d'un nouveau quai de messagerie par exemple. C'est pourquoi nous analysons les relations géographiques entre industriels et distributeurs.

Le transporteur frigorifique doit s'adapter à son milieu. Il s'implante dans un tissu industriel. En amont, les expéditeurs sont pour une grande part des IAA (Industriels de l'Agro-Alimentaire). Leurs usines de conditionnement sont proches de l'agriculteur. En aval, les points de vente tenus par les GMS (Grandes et Moyennes Surfaces) sont installés dans le milieu urbain. Ainsi, le transporteur frigorifique est dans l'obligation de desservir aussi bien les régions agricoles qu'urbaines, soit tout le territoire. Dans un territoire régional, les industriels de l'agro-alimentaire, lui sous-traitent les liaisons inter-usines, l'approvisionnement des plates-formes, les échanges inter-plates-formes et les livraisons aux points de vente (26).

Pour livrer les points de vente au niveau national, le transporteur a besoin d'un réseau (ou d'une toile) (20). Prenons l'exemple de la Bretagne. C'est une région agricole riche en industriels de l'agro-alimentaire. Le transport frigorifique tient une place importante dans l'activité économique de la région. Mais la région est excentrée par rapport au reste de la France et plus encore de l'Europe. Il n'est pas possible pour un routier d'effectuer le trajet Brest-Marseille d'une seule traite. Un réseau de plates-formes localisées à des points stratégiques devient alors un avantage considérable pour le transporteur. Il raccourcit les distances pour le chauffeur. Le trajet Brest-Marseille devient un trajet Brest-Rennes-Lyon-Marseille. Le transport fait alors intervenir trois chauffeurs ce qui augmente son coût. Mais le bénéfice, en contrepartie, est considérable. Le trajet est accompli en *A pour A* dans le jargon transport. C'est-à-dire qu'il est parti le matin pour arriver le soir même. Les lois sur les temps de conduite des chauffeurs (voir section 1.1.1.1) n'auraient pas permis d'effectuer le trajet en une traite avec seulement un chauffeur. Il aurait également fallu payer l'hébergement du chauffeur. A savoir aussi, qu'un chauffeur national coûte plus cher qu'un chauffeur régional. Enfin, le réseau permet aux camions de l'aller de revenir avec une cargaison pleine de fruits de la région Provence Alpes Côte d'Azur par exemple.

Le transporteur a pour seul souci d'économiser du temps. Pour éviter les bouchons de la circulation urbaine et pour accéder rapidement aux grands axes routiers, il plante ses **quais** ou plates-formes logistiques dans des zones industrielles situées en bordure de voies rapides (46).

Nous venons de montrer que le transporteur est un maillon qui relie non seulement

les industriels avec les distributeurs, mais aussi les régions entre elles. Il contribue au développement du tissu économique régional.

### 1.1.1.3 Le contexte économique

Prévoir les ventes au plus juste, c'est aussi comprendre le contexte économique de l'entreprise. Il contient des éléments clés pour la performance des prévisions. Une entreprise doit être capable d'agir sur son environnement économique pour ne pas le subir. D'où l'importance de dresser un état des lieux.

Le transport frigorifique est un maillon de l'activité « Transport & Logistique ». Cette activité consiste à organiser les flux de biens et d'informations de manière transparente pour le client. Le transport frigorifique est étroitement lié à la santé des industriels de l'agro-alimentaire et des grandes et moyennes surfaces. Ils sont les premiers utilisateurs des services proposés. Les lois Royer 1973 et Raffarin 1996, visant à réduire la surface autorisée d'un magasin en centre ville, ont obligé les grandes et moyennes surfaces à délocaliser leurs entrepôts de stockage en périphérie. Dans la foulée, leur gestion a été sous-traitée à des entreprises spécialisées dans la logistique. Afin que la concurrence s'exerce et que les prix de vente baissent, les grandes et moyennes surfaces multiplient leurs références produits (111). Ainsi, elles font pression sur les industriels en leur montrant qu'elles n'ont aucune difficulté de se passer de leurs produits. Cette tactique est bénéfique au transporteur, elle augmente d'autant les trajets et la gestion des entrepôts. Nous verrons dans le paragraphe suivant comment le transporteur s'organise autour de ces deux acteurs (IAA et GMS). Le transport frigorifique a également bénéficié du développement des produits alimentaires surgelés de ces 15 dernières années. La consommation de ces produits a augmenté en volume de 2,5% en 2005 (79). Pour situer le transport frigorifique parmi l'ensemble des marchandises convoyées, ajoutons qu'il représente 15% du volume transporté sur les routes françaises en 2005<sup>2</sup>.

Par sa nature, le transport routier frigorifique national, de produits frais, ne craint pas la concurrence fluviale ou ferroviaire. Les impératifs de temps dus à la date limite de consommation (DLC) des produits frais empêchent ces deux secteurs de se positionner. L'avion restant trop onéreux, ils sont réservés à l'international. Les transporteurs frigorifiques routiers sont les seuls capables de répondre aux exigences du cahier des charges des produits frais. Par contre, les produits surgelés peuvent être transportés par voie maritime dans des navires réfrigérés ou par rail.

La France compte un grand nombre de transporteurs régionaux, mais une vingtaine seulement réalisent un chiffre d'affaires (CA) supérieur à 20 millions d'euro et se po-

---

2. DAEI/SESP - Enquête TRM 2005

sitionnent à l'international. Les chiffres du tableau 1.1 mettent en avant la domination de STEF-TFE dans le paysage du froid et du frais. Suivent derrière Nexia et TNT. D'autres groupes spécialistes du froid résistent à l'image de STG, Olano, Transport Lahaye et le groupe Gringore. Le GRPD (Groupement National des Transporteurs Routiers de Denrées et Produits Périssables) estime la croissance du transport sous température dirigée entre 2 et 3% par an. Elle dépend étroitement de la consommation alimentaire des ménages.

Rang	Raison Sociale	Effectif	CA 2003 en K€	Flotte frigo (2)
1	STEF-TFE	11986	1347000	2300
2	NEXIA FROID	2245	199586	300
3	TNT LOGISTICS	1500	150000	300
4	STG	1329	122691	1150
5	GROUPE OLANO	600	75000	300
6	TRANSPORT LAHAYE	867	75000	502
7	GROUPE GRINGORE	580	56165	280
8	GROUPE DENTRESSANGLE (1)	nc	52700	230
9	TRANSPORT CHABAS	377	52246	130
10	BSA INTERNATIONAL	620	51358	136

TABLE 1.1 – *Palmarès des principales sociétés de transport frigorifique*

Sources : *L'officiel des transporteurs magazine*, Novembre 2004

(1) ensemble routier multi-températures

(2) activité frigo seule

Le transport frigorifique n'échappe pas à la tendance mondiale des fusions / acquisitions (111). Pour répondre aux attentes toujours croissantes de leurs clients, en matière de performances et de garanties les entreprises doivent s'agrandir. Pour croître une entreprise a deux possibilités :

- soit, elle développe de nouvelles capacités de production,
- soit, elle prend le contrôle d'une entreprise existante. Il est alors question d'un processus de fusion / acquisition.

Nombreuses sont les sociétés qui ont été rachetées par de grands groupes. La plus récente en date est l'acquisition de Cryologic par le groupe STEF-TFE en Janvier 2006. Cryologic était un leader français de la logistique sous température contrôlée. Avec 12.000 employés et un chiffre d'affaires de 1,4 milliard d'euros en 2004, STEF-TFE est le premier groupe français de la logistique et du transport de produits alimentaires frais et surgelés. Il se place parmi les 20 premiers groupes de transport français après Aéroports de Paris. A la différence de ses concurrents, STEF-TFE est le seul groupe français

centré uniquement sur le secteur de la **messagerie** du froid.

La concentration des transporteurs permet de disposer des infrastructures logistiques (parc immobilier & parc véhicules) indispensables pour créer un réseau national. Rachetés par des grands groupes comme STEF-TFE, les transporteurs locaux (Mélédou, Le Bas Froid, Le Coz et Mahé en Bretagne) continuent à organiser le réseau régional. Les grands groupes sous-traitent également une part de leur activité transport aux PME régionales. Elles assurent les **ramasses** et **livraisons**.

Malgré les fusions / acquisitions, des groupes comme Nexia peuvent ne pas survivre. Nexia était le deuxième groupe de transport et de logistique frigorifique en France. Il pesait 200 millions d'euros de chiffre d'affaires annuel réparti sur 28 **agences** dans l'Hexagone. Pourtant, ce groupe est en redressement judiciaire depuis janvier 2007.

Mais STEF-TFE ne reste pas seul sur le marché, son premier concurrent est la société suisse Kuehne et Nagel. Ce groupe est un des leaders mondiaux de la logistique et du « supply chain management » (voir section 1.1.1.4), avec un effectif de 47.000 personnes, 830 implantations réparties sur plus de 100 pays et plus de 6 millions de mètres carrés de surfaces d'entreposage. Pour entrer dans le marché français, le groupe a fait l'acquisition de ACR Logistics. Basée à Paris, ACR Logistics est une importante structure britannique fortement implantée en France.

Le marché du transport frigorifique semble attirer des transporteurs d'autres milieux. Norbert Dentressangle est une puissante société de transport et logistique française. Pour entrer dans le marché du transport sous température dirigée elle s'empare début 2006 de TNT logistics. Cette dernière possède, entre autres, un savoir-faire dans les secteurs d'activités de la température dirigée. En Octobre 2007, elle fait une proposition d'achat sur le groupe Christian Salvesen, qui lui aussi possède des activités de transport et d'entreposage frigorifique.

Sur un autre secteur que celui des produits alimentaires, nous repérons Chronopost et DHL. Ils se spécialisent dans le transport des produits de santé sous température dirigée. Avec un volume de trois milliards d'unités distribuées chaque année en France, le marché du médicament reste attractif pour les transporteurs<sup>3</sup>.

Une pénurie de camions en 2007 augmente les coûts de la sous-traitance transport. La flambée constante du pétrole pousse les prix du transport vers la hausse. Concernant le marché des produits alimentaires, la loi Dutreil (Janvier 2006), qui vise à faire baisser les prix des produits de grandes marques, engendre des tensions dans les négociations tarifaires entre les chargeurs et les prestataires. Alors que la consommation des ménages croît entre 2001 et 2008 à un rythme annuel de 2,1%, les achats des produits alimentaires progressent deux fois plus lentement<sup>4</sup>.

Le transporteur a la charge, quelle que soit la distance, d'acheminer des produits frais avec la contrainte de préserver leur saveur et d'informer le client par des systèmes de

---

3. L'officiel des transporteurs, n°2283, janvier 2005

4. INSEE : La consommation des ménages en 2005

suivi informatique, du lieu de son colis. Pour y parvenir, il doit programmer une logistique sans faille.

#### 1.1.1.4 La logistique

Le maintien de la chaîne du froid lors du transport des marchandises, les contraintes législatives, la multiplicité des acteurs économiques et leur éloignement géographique impose une logistique complexe. Maîtriser les mécanismes de la chaîne logistique permet d'anticiper les besoins afin de renforcer les maillons faibles.

Dans la définition officielle de la norme AFNOR (norme X 50-600), la logistique est une fonction « dont la finalité est la satisfaction des besoins exprimés ou latents, aux meilleures conditions économiques pour l'entreprise et pour un niveau de service déterminé. Les besoins sont de nature interne (approvisionnement de biens et de services pour assurer le fonctionnement de l'entreprise) ou externe (satisfaction des clients). La logistique fait appel à plusieurs métiers et savoir-faire qui concourent à la gestion et à la maîtrise des flux physiques et d'informations ainsi que des moyens »<sup>5</sup>. A en croire cette définition, tous les services de l'entreprise seraient susceptibles d'être impliqués par la logistique. Le spécialiste du transport et de la logistique se restreint à la « chaîne logistique » (« supply chain » en anglais, voir la définition dans la même section) qui va des fournisseurs aux clients.

Aujourd'hui, l'économie de consommation, vise à satisfaire avant tout les attentes du consommateur final. Ce dernier a des exigences d'hygiène alimentaire, il montre un engouement pour des produits frais et réclame des prix bas. Il en découle des changements de comportements de la part des producteurs. Les problèmes de sécurité alimentaire poussent à la conservation des produits à des températures constantes et contrôlées (100). De plus en plus de produits alimentaires sont ainsi transportés et stockés sous température dirigée. Il en est de même de certains produits non alimentaires destinés à la santé et aux cosmétiques. Pour conserver la fraîcheur des produits, il faut en produire juste assez de façon à ne pas les stocker et réduire les délais d'acheminement. C'est ce qui est appelé la méthode JAT (Juste à Temps) (111). La durée moyenne des stocks de produits frais est de 2 à 3 jours alors que les surgelés peuvent dépasser l'année. L'autre cause du JAT est la Date Limite de Consommation (DLC). Elle est de quelques jours pour des coupes de volaille ou de 21 jours pour des yaourts. La DLC gouverne la Supply Chain des produits frais et imprime un rythme rapide à la chaîne de distribution<sup>6</sup>.

La vision moderne de la gestion industrielle vise à éliminer de façon systématique tous

---

5. agrojob.com

6. Stratégie logistique n°66, mai 2004

les gaspillages et à rechercher l'amélioration continue de la flexibilité, de la productivité et de la réactivité. L'entreprise doit partir des besoins du client et s'organiser pour répondre à la demande de façon rapide et efficace (111). C'est ce que proposent les spécialistes de la logistique. Contrairement à certaines idées reçues, le JAT n'est pas synonyme de vitesse et précipitation, mais bien d'organisation globale et de respect des délais convenus<sup>7</sup>. Par leur métier d'origine, le transport, les logisticiens sont le lien entre le producteur et le distributeur. Ils présentent des services à forte valeur ajoutée qui sont : hygiène, traçabilité, sécurité, rapidité et zéro stock.

D'après la définition de l'encyclopédie « e-business », la chaîne logistique ou « supply chain », en anglais, est définie comme un « ensemble de procédures et de logiciels permettant de gérer de façon optimale la totalité des flux d'information, des flux physiques et des interfaces entre les différents acteurs, producteurs et fournisseurs qu'implique la fabrication d'un produit ou l'offre d'un service. Ils se basent sur les renseignements concernant la demande jusqu'aux données nécessaires à la distribution, en passant par la conception et la production proprement dite ».

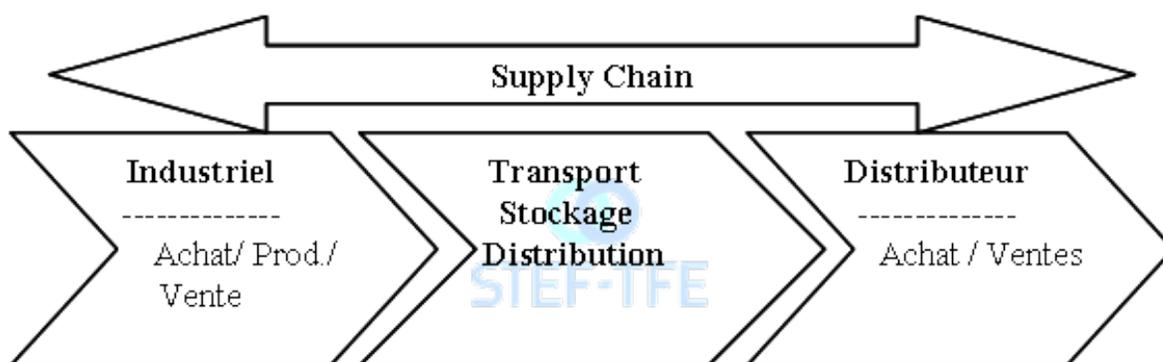


FIGURE 1.1 – La chaîne logistique

La figure 1.1 illustre la définition de la chaîne logistique. Elle présente les 3 principaux acteurs qui sont l'industriel, le transporteur et le distributeur. Les industriels prennent en charge l'acheminement de l'usine vers les entrepôts de distributions, tandis que les grandes et moyennes surfaces supervisent la distribution de leurs entrepôts vers les magasins. La tendance actuelle est d'externaliser les services de la logistique et du transport pour se recentrer sur leur domaine de compétence. Les investissements de départ pour prétendre faire de la logistique du froid sont d'ailleurs trop élevés. Ainsi, les grands groupes de transporteurs proposent la location et la gestion des entrepôts, l'étiquetage, le co-packing et le transport. Cette prestation est appelée « supply chain management » (98). Le transporteur se trouve néanmoins dans une situation délicate. Il doit à la fois répondre aux exigences du producteur et du distributeur. « Nos trois contraintes sont la

7. La performance logistique, J-Philippe Guillaume, 1993, chez Nathan

*tension des flux, le respect de la chaîne du froid et la traçabilité* » confie Jean-Pierre Sancier, directeur général adjoint de STEF, 2004.

Dans un contexte de forte concurrence, les industriels des produits frais font preuve d'une extrême réactivité pour coller aux exigences des distributeurs. Mais leurs relations restent tendues. Un rapport de force commercial entre eux a pour conséquence de tirer les prix du transport vers le bas. La rentabilité devient très faible. Il est coutume de dire qu'un transporteur fait son bénéfice le 31<sup>ème</sup> jour du mois.

Pour répondre aux exigences imposées par les chargeurs et les distributeurs, d'après une étude réalisée par « Eurostaf » en 2007<sup>8</sup>, les prestataires logistiques et de transport s'attachent à :

- *Approvisionner les distributeurs en flux tendus,*  
Les industriels de l'agro-alimentaire et de la grande distribution cherchent à optimiser les DLUO (Date Limite d'Utilisation Optimale) et à réduire les délais de livraison de façon à accélérer la rotation de leurs stocks.
- *Maîtriser les coûts,*  
Les carburants représentent 20% des coûts de revient et environ 5% du prix final d'un produit. L'augmentation constante du prix des carburants pousse les prix du transport vers la hausse. Les frais de personnel et ceux du foncier qui subissent une forte progression compriment également les marges.
- *Optimiser les coûts de transport et de la logistique,*  
De nouvelles méthodes d'organisation destinées à améliorer la productivité et le taux de service sont adoptées pour gérer les entrepôts, comme la préparation de commande par reconnaissance vocale. De même, les prestataires recourent à la sous-traitance de la traction ou au pooling pour rationaliser les coûts de transport et gagner en flexibilité.
- *Investir pour améliorer la sécurité et la traçabilité,*  
Le respect de la chaîne du froid est un pré-requis et une composante essentielle de la qualité du service. Les prestataires doivent donc posséder des entrepôts et des véhicules spécialement équipés.
- *Densifier leur réseau,*  
Cette stratégie peut impliquer des acquisitions, comme STEF-TFE l'a fait avec Cryologic en 2007. Elle peut passer par des implantations directes, comme le fait STEF -TFE (ouverture de 7 nouveaux sites logistiques ces deux dernières années).
- *Proposer des prestations à plus forte valeur ajoutée,*  
Ceci passe par des prestations de différenciation retardée (co-packing et co-manufacturing) ou par la gestion de la traçabilité.

---

8. Etude Eurostaf : Le transport et la logistique sous température dirigée en Europe

La maîtrise des coûts passe par une optimisation des plannings et donc par la prévision des ventes (voir section 2.2.1). Les nouvelles implantations et autres investissements se font qu'après une étude de marché et donc d'une prévision des ventes. De même, pour proposer de nouvelles prestations, il faut anticiper les besoins des clients.

Plus généralement, les performances du transporteur dépendent de la façon dont il maîtrise les flux. Nous avons souligné que les flux physiques des marchandises étaient accompagnés de flux d'informations. Ces derniers composent la matière première des prévisions de vente chez un transporteur (voir section 3.1).

Le transporteur est le maillon qui unit producteur et distributeur. Ces derniers lui accordent leur confiance et se reposent sur lui pour être livrés à l'heure. Pour y parvenir, le transporteur doit veiller à être constamment à la pointe dans les domaines techniques et technologiques (voir section 2.3.3 et 2.3.4). Il doit également entretenir une relation professionnelle saine avec les acteurs de la chaîne logistique. Ils apportent les informations dont le transporteur a besoin pour améliorer ces prévisions (voir section 2.4).

Pour illustrer la théorie par un cas concret, nous entrons au sein du groupe STEF-TFE.

## **1.1.2 Groupe STEF-TFE**

### **1.1.2.1 Carte d'identité**

STEF-TFE est le premier logisticien du froid en France. Il opère sur les secteurs du transport, de la prestation logistique et des systèmes d'information. Il pèse 1,7 milliard d'euros de chiffres d'affaires en 2006 et compte plus de 11.900 salariés. Comme l'illustre la figure 1.2, le groupe et ses partenaires disposent de plus de 331 plates-formes et entrepôts et 23.000 camions à travers toute l'Europe. Depuis son introduction en bourse (1999), le cours du titre STEF-TFE a été multiplié par 3,5. Dans le même temps, son CA a progressé de 40% et son résultat net de 300%<sup>9</sup>.

Le groupe STEF-TFE confie la prestation logistique (surgelés, frais et restauration rapide hors foyer) à sa branche STEF. Celle-ci gère les entrepôts en température positive et négative. Le transport des produits frais et surgelés, relève de la compétence de TFE. Avant 2009, la logistique et le transport des produits de la mer sont la spécialité de la filiale Tradimar. En 2009 Tradimar disparaît et l'activité est confiée à TFE. Les systèmes d'information liés au pilotage des flux stockés ou tendus sont regroupés dans la filiale Agrostar. Le groupe détient également des parts dans la CMN spécialiste dans le fret

---

9. L'officiel des transporteurs, n°2302, mars 2005

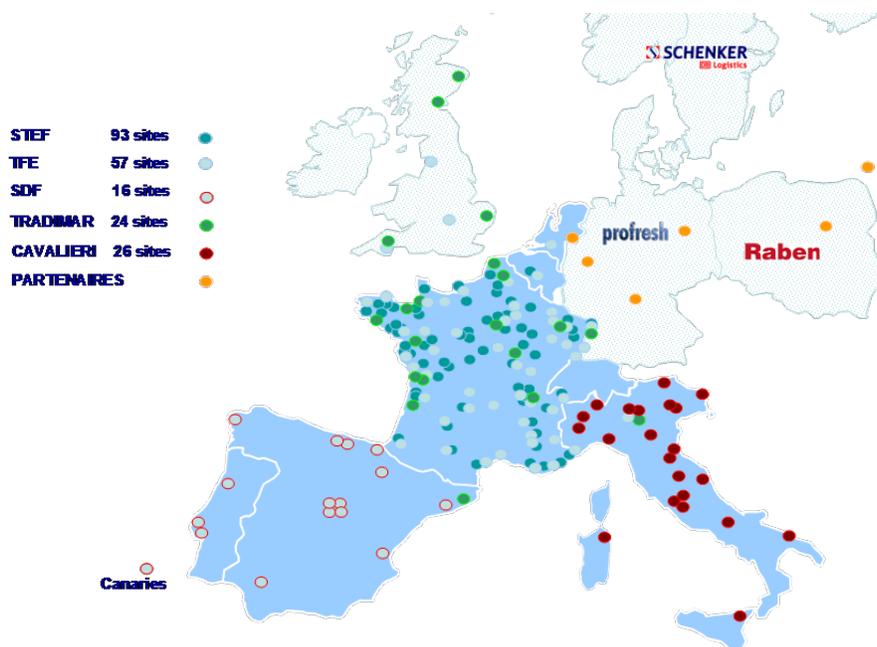


FIGURE 1.2 – Localisation des sites TFE en Europe en 2007

entre la Corse et le continent.

Le pôle transport TFE livre 100.000 destinataires réguliers par semaine. Il dispose, en France, de 70 plates-formes de groupage sous température dirigée (0/4 °C), situées tant dans les zones de production que dans les zones de consommation et de distribution<sup>10</sup>. Toutes les régions françaises sont représentées (voir carte, figure 1.3). TFE est en mesure de prendre en charge le transport de camions complets, le **groupage** et la **messagerie**. Grâce à son réseau, TFE assure ses livraisons en moins de 24 heures (livraison en A pour B, selon le jargon en usage dans le groupe) quelle que soit la destination en France.

Le pôle STEF est le spécialiste de la prestation logistique en frais et surgelé. Il exploite dans toute la France des entrepôts surgelés, en flux stockés, et des plates-formes, en flux tendus. Compte tenu des contraintes spécifiques liées au grand froid (voir section 1.1.1.1), les grandes et moyennes surfaces sous-traitent la logistique du froid. Ainsi, chacune des plates-formes STEF sert d'entrepôt pour le compte d'industriels, de distributeurs et de restaurateurs hors domicile. Les prestations associées au service d'entrepôt sont la congélation et la surgélation des produits alimentaires, le conditionnement et le traitement par froid de fruits et légumes, le calibrage, la préparation de matière première, le co-packing. STEF propose également la gestion des approvisionnements, la préparation des commandes, et l'organisation des **livraisons** aux points de vente. Les concurrents directs de STEF sont Frigoscandia et Sofrino-Sogena.

10. Rapport annuel 2006 de STEF-TFE

Tradimar est le numéro 1 du transport des produits de la mer. Il assure 80% des trafics au départ de Boulogne-sur-Mer, premier port de pêche de France. Tradimar réalise 30% de son activité à l'international (Angleterre, Écosse, Italie, Espagne). Il transporte 500.000 tonnes de produits de la mer issus des 150 ports européens, pour alimenter 5.000 clients, des grandes et moyennes surfaces aux grossistes <sup>11</sup>.

Agrostar, est le service informatique du groupe STEF-TFE. Par ses solutions informatiques, il optimise la logistique globale et les performances économiques du groupe et de ses clients. Chaque opération de la chaîne logistique (commande, préparation des commandes, livraisons, transports, manutentions, stockages), nécessite une application spécifique. Elles peuvent être regroupées dans un système d'information voué au métier de l'entreprise. Les plus répandus dans le transport et la logistique sont les modules TMS et WMS. Le TMS (Transport Management System) est un outil informatique dédié à la gestion du transport alors que le WMS (Warehouse Management System) couvre les besoins liés à la gestion des stocks et à l'approvisionnement des magasins.

### 1.1.2.2 Le transporteur TFE

**1.1.2.2.1 Le réseau** Seuls des groupes constitués de réseaux et disposant de moyens importants sont à même d'assurer des trafics internationaux.

A la différence d'autres transporteurs, TFE s'est consacré à transporter uniquement des denrées sous température dirigée. C'est une contrainte importante. Dans un territoire segmenté par activité de production, le transporteur part d'une région  $R1$  pour livrer les produits de cette région dans la région  $R2$ . Or l'activité de la région  $R2$  est différente de celle de  $R1$ . Un camion qui revient vide, (sans être chargé des produits d'une activité de  $R2$ ), ne produit aucune recette. C'est pourquoi le transporteur doit être capable de se diversifier. Dans la pratique, c'est moins évident. La législation impose de respecter des règles suivant les produits transportés. Pour préserver la qualité sanitaire des frigos, une remorque isotherme n'a

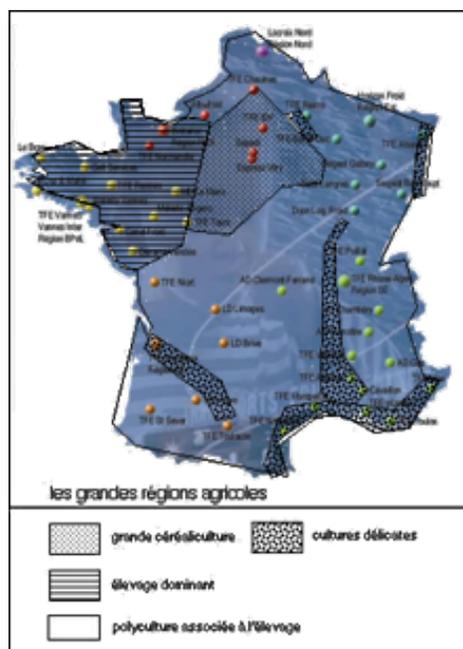


FIGURE 1.3 – Situation géographique des agences TFE

11. L'officiel des transporteurs, n°2120, juin 2001

pas le droit de transporter du sable par exemple. D'autre part, de par leur nature (liquide, solide, fragile) les produits se transportent différemment. Parmi les transports de marchandises spécialisées, l'INSEE recense 7 sous catégories :

- transport routier frigorifique,
- transport routier de produits pétroliers,
- transport routier d'autres vracs liquides ou gazeux,
- transport routier de marchandises conteneurisées,
- transport routier de meubles,
- transport routier de marchandises en vrac sec,
- transports routiers spécialisés n.c.a (non classés ailleurs).

Pour augmenter les chances de retour à plein, TFE innove et met en place un réseau national de **messagerie** sous température contrôlée. Avec des commerciaux dans chaque site et le rachat de transporteurs locaux, TFE construit un portefeuille client local. C'est le plus sûr moyen de faire circuler les camions à plein. Le deuxième avantage du réseau est qu'il permet de relier les sites de production et les points de vente dans un délai de moins de 24 heures.

Le réseau TFE découpe la France en 6 régions, Bretagne/Pays de Loire, Est, Nord, Normandie/Ile de France, Sud-Ouest, Sud-Est. Chaque région comporte un ensemble d'**agences** dont une direction régionale. En moyenne, les plates-formes se trouvent à moins de 150 km l'une de l'autre. L'emplacement des sites n'est pas anodin. La carte de la figure 1.3, fait apparaître que les régions agricoles du Nord-Ouest (Bretagne, Normandie), du Bordelais, de l'Alsace, de la Champagne, des Pays de la Loire et celles du Sud-Ouest ont été bien implantées par les agences de transport TFE. Les régions urbaines (le Nord, la région parisienne, l'Ouest) comportent également de nombreux sites. D'autre part, certaines agences sont plus importantes par leur taille. Elles sont appelées « plate-forme étoile ». Elles sont situées sur des axes stratégiques. Leur rôle est de recevoir la marchandise de toutes les agences de la région et de la desservir sur l'Hexagone. Les plates-formes étoiles sont : Paris, Rennes, Lyon, Avignon, Dijon, Niort, Langres, Toulouse, Le Mans, Nantes (voir section 1.2.2).

STEF-TFE a fait construire un grand nombre de ses plates-formes. Deux raisons pour cela, la première est qu'il existe peu d'offres disponibles dans ce secteur, la deuxième est liée à la localisation. Pascal Wagner, secrétaire général du groupe, relève que « *les choses se jouent à 10 ou 20 km pour trouver le bon barycentre et les nœuds de communication qui permettent d'économiser des kilomètres ou du temps* »<sup>12</sup>.

Le réseau engendre des flux physiques et d'information. C'est pourquoi nous étudions les deux types de flux.

---

12. Stratégie logistique, n°57, juin 2003

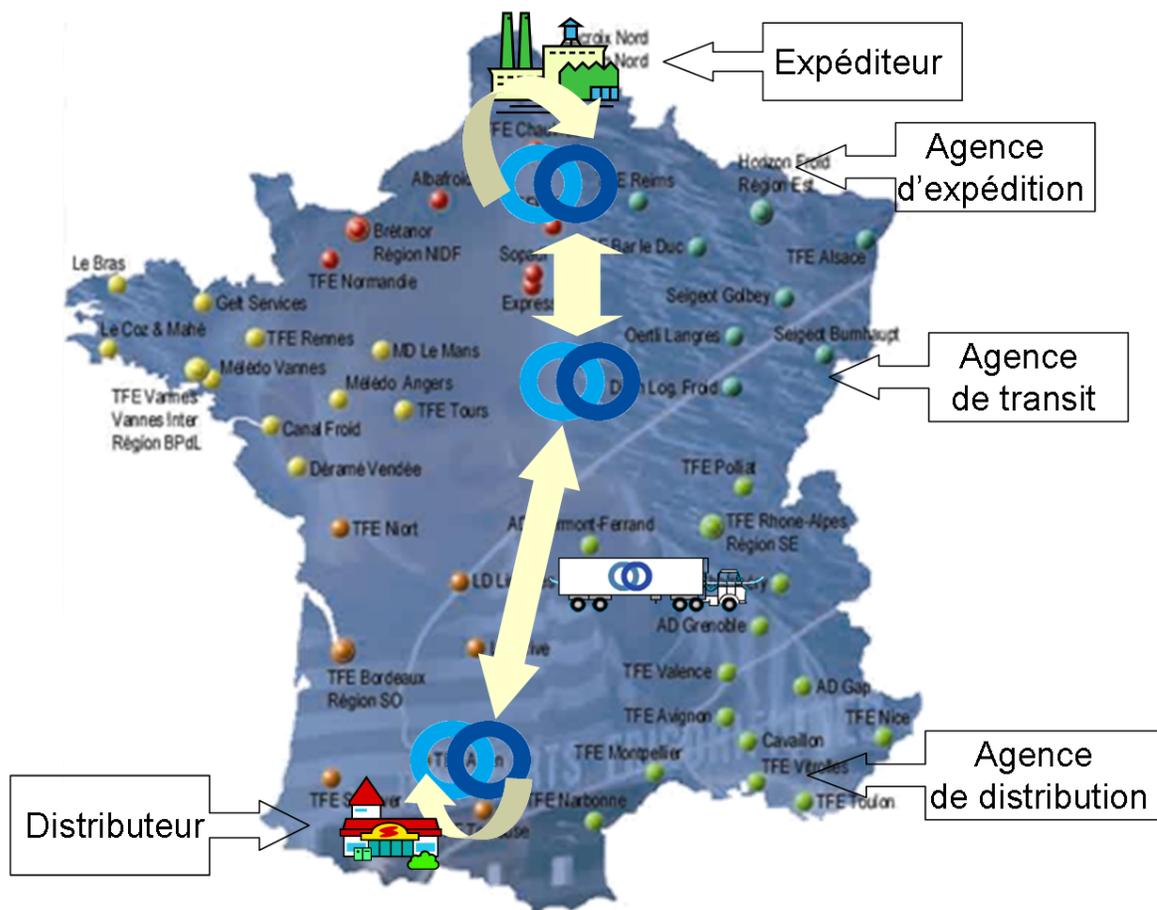


FIGURE 1.4 – Voyage des flux physiques chez TFE

**1.1.2.2.2 Les flux physiques** Les plates-formes (**agence**) du groupe travaillent en étroite collaboration. Pour réaliser les chaînes de transport en A pour A ou A pour B, les flux de transport fonctionnent suivant la figure 1.4. Rappelons qu'un transport A pour A, signifie ramasser la marchandise chez l'expéditeur le jour A pour la livrer chez le destinataire moins de 24H après. Entre 8h et 12h, les camions de l'agence partent en **tournée** pour distribuer les marchandises reçues la nuit par le réseau TFE (distribution régionale). Avant de rentrer, ils partent faire le tour des industriels pour ramasser la marchandise à destination des distributeurs (**ramasses**).

Ces dernières sont déposées sur la plate-forme TFE pour être triées suivant leur destination. Elles sont ensuite expédiées entre 14h et 19h vers les agences du réseau (**expédition**). Le même cycle se reproduit dans l'agence réceptrice.

Les flux physiques sont décomposés par segments (voir figure 1.5) pour ensuite être traduits en objets informatiques (voir section 3.1.2). Chaque transport est un segment. Suivant son lieu de départ et sa destination le segment porte un nom différent. C'est le

moyen choisi par TFE pour tracer le circuit qu'emprunte la marchandise depuis l'expéditeur au destinataire. Chacun de ces segments est informatisé et enregistré dans une banque de données.

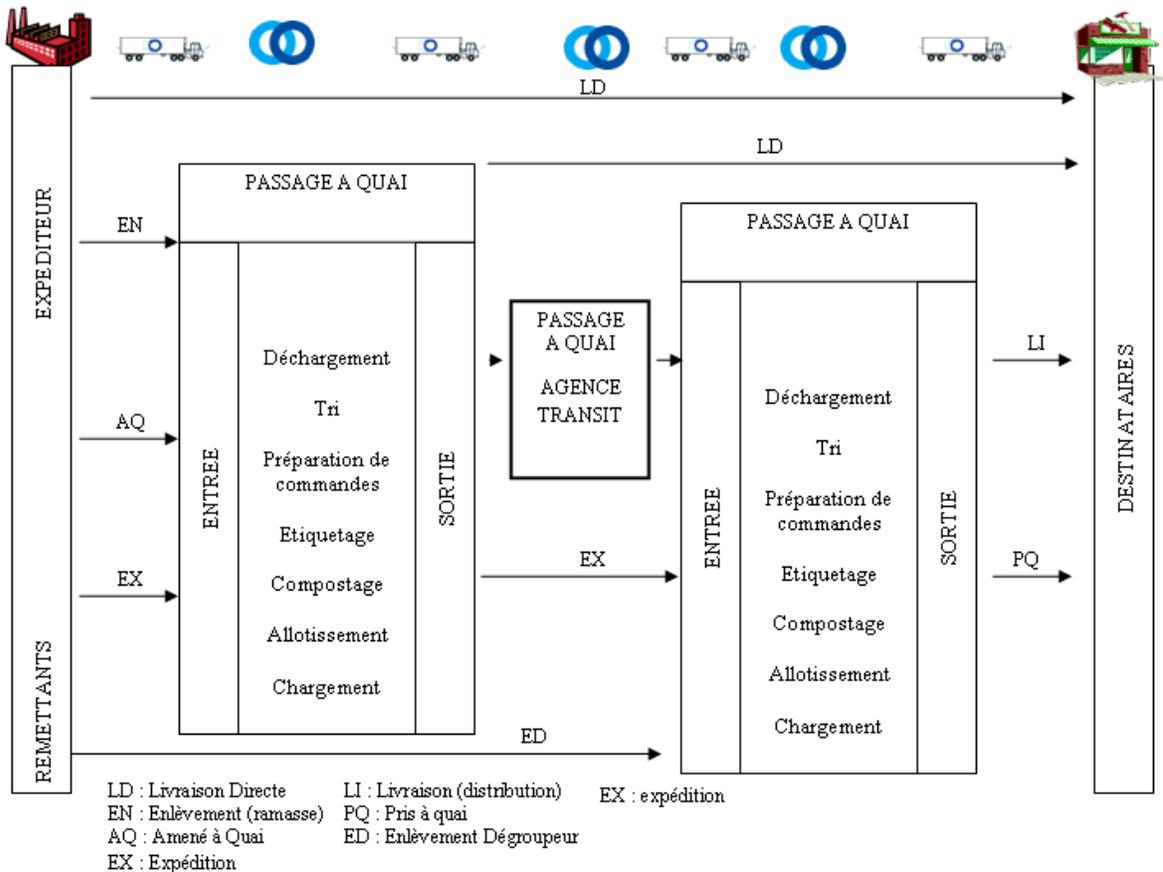


FIGURE 1.5 – circuit de la marchandise de l'expéditeur au destinataire

La figure 1.5 désigne les différents types de transfert de marchandises possibles entre deux acteurs. Les segments d'enlèvement sont les opérations consistant à prendre en charge la marchandise chez un expéditeur pour l'amener au quai de l'agence de transport. Les ramasses désignent les enlèvements réguliers ne nécessitant plus de confirmation.

Les flux d'entrées à quai appartiennent à l'une des trois catégories suivantes :

- remises à quai effectuées par les expéditeurs ou remettants,
- retours de livraison / ramasses / enlèvements effectués par les véhicules de l'agence,
- arrivages en provenance des autres agences.

La remise des marchandises aux destinataires peut être effectuée de deux manières différentes :

- livraison chez le destinataire par TFE,
- prise à quai TFE par le destinataire.

La livraison est toujours faite au destinataire désigné sur le document de transport. Les flux de sortie de **quai** appartiennent à l'une des deux catégories suivantes :

- expédition vers les autres agences,
- livraison vers un destinataire final.

Les segments « livraison », « livraison Directe » (ou « distribution ») et « expédition » sont les principaux flux en termes de poids de marchandises en sortie du **quai**. C'est pourquoi ils décrivent, à eux seuls, l'activité d'un quai. Les enseignes de grandes et moyennes surfaces imposent le rythme de l'activité **livraison** (distribution) des agences de transport. Les producteurs et l'industrie agro-alimentaire imposent, pour leur part, le rythme de l'activité **expédition** des agences de transport.

Le transporteur TFE fait la distinction entre livraison locale (LI) et livraison directe (LD). La dernière survient quand une **agence** livre un destinataire final situé hors de sa zone de chalandise. A contrario, une livraison (locale) se produit dans la zone commerciale de l'agence.

Les « expéditions » sont des flux inter-réseaux. Pour augmenter la productivité, les agences se partagent leurs ressources. Les marchandises en provenance des agences Le Bras et Le Coz et Mahé, dans le Finistère, vont transiter par l'agence de Rennes pour être soit livrées sur place soit expédiées vers une autre agence. Dans le premier cas, l'agence de Rennes est appelée agence de « dégroupage » dans le deuxième cas elle est une agence de « transit ». Les agences du Finistère sont les « remettants ». Elles seront facturées par Rennes pour la prestation rendue.

Quand le **voyage** est trop long pour être effectué en une seule traite (aller-retour dans la journée), les agences *A* et *B* s'organisent. Elles envoient chacune leur chauffeur sur la route. Ces derniers se croisent à un point de rendez-vous fixé à mi-chemin. Ils échangent leur remorque et font demi-tour. Cette astuce permet d'économiser des frais de déplacement pour l'agence et au chauffeur de rentrer chez lui après sa journée de travail.

**1.1.2.2.3 Schéma du passage à quai** Le véhicule et l'entrepôt sont les outils de travail de référence. Les **quais** (ou plates-formes ou **agences**) TFE sont installés dans des régions à fort potentiel économique. D'un point de vue financier, l'investissement qu'ils représentent est considérable. Les bâtiments ont une surface de 5.000 à 6.000 m<sup>2</sup>. Ils sont équipés de climatisations fonctionnant 24 heures sur 24. Les températures sont affichées en temps réel via un système d'écran de contrôle. L'hygiène doit y être de qualité irréprochable.

Contrairement aux entrepôts de la STFE, ceux de TFE ne sont pas destinés au stockage.

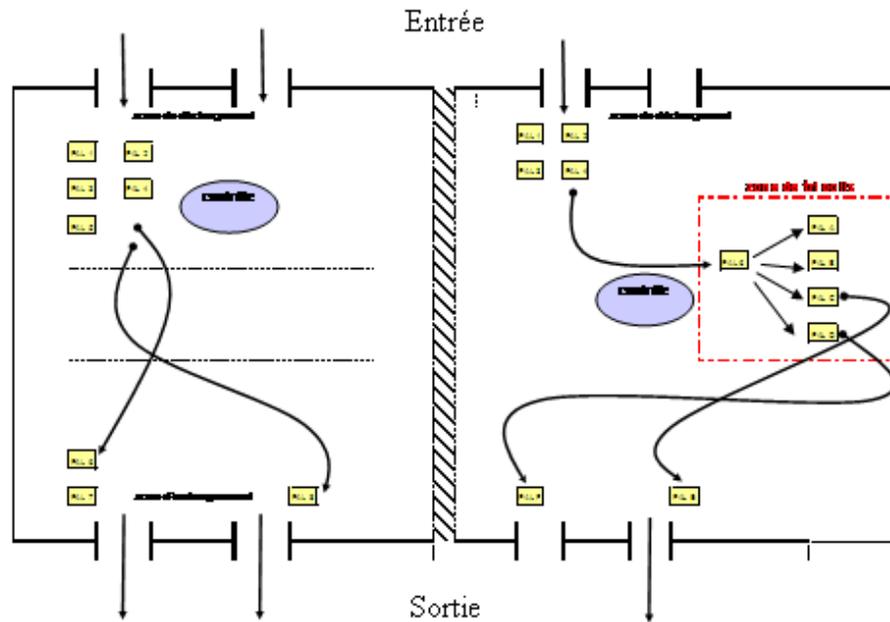


FIGURE 1.6 – Schéma d'un quai TFE

Ils ont pour vocation d'orienter la marchandise vers sa destination. Ils jouent le rôle de nœuds du réseau.

La figure 1.6 montre le fonctionnement d'un quai de messagerie. En entrée, le quai reçoit des marchandises expédiées par les autres agences du réseau, les marchandises apportées à quai par les clients et celles enlevées pendant une tournée de ramasses. Habituellement, les marchandises sont posées sur des palettes standard appelées « Europe ». Elles supportent une charge de 300 kg sur un volume de 1,728 m<sup>3</sup> (80×120×180). Un véhicule supporte jusqu'à 33 palettes.

Sur un quai, les manipulations de produits sont réduites au minimum. Selon les cas, les palettes passent directement du quai d'entrée à celui de sortie (cross docking), ou elles sont reconditionnées dans la zone de tri. Dans ce cas, les colis sur la palette d'entrée sont éclatés sur différentes palettes selon leurs destinations. Une palette au sol peut ainsi contenir plusieurs palettes empilées les unes sur les autres. Elle est appelée palette gerbée. Une agence comme celle de Rennes traite une moyenne de 3.000 palettes jour. C'est une centaine de camions en provenance de plus de 500 **expéditeurs** qui arrivent à quai de l'agence de Rennes. Après passage à quai, les marchandises repartent vers plus de mille destinations.

La productivité sur le **quai** se calcule par le taux de productivité qui est le ratio tonne / heure par manutentionnaire. Il se situe entre 2,5 et 3 T/H.

**1.1.2.2.4 Les flux d'information** Début des années 2000, le groupe STEF-TFE investit massivement en moyens informatiques. Une des clés du succès repose sur la mise en place d'un système d'information particulièrement avancé. Pour répondre aux exigences de traçabilité 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 entrepôt de données (« Data Warehouse » en anglais). Cet entrepôt de données stocke toutes les informations relatives aux colis transportés et entreposés par le groupe, soit plusieurs giga-octets d'information.

Le règlement européen soumet les transporteurs sous température dirigée à une obligation générale de traçabilité. Chaque maillon de la chaîne du froid doit tenir à jour les entrées et les sorties des denrées ou produits avec la liste des fournisseurs et clients correspondants et disposer de systèmes permettant de mettre ces informations à la disposition des autorités compétentes <sup>13</sup>.

La banque de données du groupe STEF-TFE met en ligne quelque 100 millions de contrats qui sont passés au crible de l'analyse multidimensionnelle avec l'aide d'outils décisionnels. Avec ces outils, les collaborateurs accèdent de manière simplifiée, via une interface web, à un ensemble de statistiques pour la prise de décision. Présentées sous forme de tableaux de bord ou de cubes dynamiques, ces analyses rendent compte, grâce à des indicateurs de rentabilité, de la performance d'une agence, d'un client ou d'une tournée <sup>14</sup>. Pour diminuer le risque d'utilisation abusive, les collaborateurs se voient attribuer un mot de passe avec lequel ils ont accès à l'information restreinte à leur domaine de compétence. Les contrôleurs de gestion par exemple accèdent aux informations leur permettant de surveiller la bonne santé de leur agence (grille tarifaire, niveau de production, nombre de colis, retard de livraison, ...). Les commerciaux, quant à eux, accèdent à des informations clients (activité, rentabilité, ...).

L'outil web développé est appelé Infomanager. Le catalogue Infomanager est partagé en rubriques telles que « Contrôle de gestion opérationnel », « Ordre de transport », « Statistiques commerciales », « Domaine d'activité », « Heure quartz » (relatif aux disques des camions), « Tableau de bord de pilotage », « Ressources humaines », « Suivi des horaires navettes »... (voir section **3.1.1**).

Suite à cette expérience concluante, le groupe a continué sa progression dans le partage de l'information. Il propose aujourd'hui d'en faire bénéficier ses clients. A travers le portail TFE-online, des outils web comme e-sent, e-track, e-plan, permettent aux clients de passer leurs ordres de transport, de suivre géographiquement leurs expéditions en temps réel, de connaître le nombre de contrôles température effectués, les dates et heures de ramasses de leur localité, les tarifs et temps de transport pour une destination, les litiges, les retards, ... D'autres outils proposent à de grandes enseignes des tableaux de bord dynamiques sur leur propre activité. Elles peuvent notamment savoir où sont vendus

---

13. L'officiel des transports magazine, Novembre 2004

14. Transports Actualités, n°751, juillet 2002

leurs produits et qui les achète.

D'autres flux d'informations sont générés. La réglementation de janvier 2005 sur la traçabilité, oblige les différents acteurs de la filière à communiquer entre eux en faisant dialoguer les systèmes d'information. D'où la nécessité de maîtriser l'EDI (Echanges de Données Informatisées)<sup>15</sup>. L'EDI permet, à travers un protocole de normes internationales, d'échanger des données d'ordinateurs à ordinateurs. Les informations échangées sont diverses et variées. Par exemple, les poulets sont pesés, étiquetés avec le poids, le numéro de lot, le prix au kilo, la température, etc. Ces données circulent entre chaque maillon de la chaîne logistique.

« A Givors, nous avons particulièrement mis l'accent sur la traçabilité, grâce à une intégration complète de notre informatique avec celle des clients et via des systèmes radio, une gestion en temps réel de l'entrepôt et de la production automatique d'étiquettes code-barre aux normes EAN-Gencod qui permettent de tracer les produits sortant de l'entrepôt jusqu'aux destinataires ». commente Daniel Page, directeur de l'entrepôt Givors (Rhône Alpes) en 2001<sup>16</sup>.

## 1.2 A l'écoute de l'industriel

L'entreprise TFE vend du transport de marchandises. Ce transport est facturé soit au poids, soit au nombre de palettes. Une vente équivaut à une **lettre de voiture**. La lettre de voiture ou **position** (voir section 3.1.2) dans le jargon du transporteur est le document qui scelle le contrat passé entre le transporteur et son client. La lettre de voiture contient entre autres comme informations, le nom de l'expéditeur, le poids, le lieu de chargement et le lieu de livraison des marchandises. Les ventes des agences de transport TFE sont donc résumées par deux indicateurs, le **poids des marchandises** à transporter et le **nombre de lettres de voiture** (ou positions) associées.

Le nombre de lettres de voiture est fortement corrélé avec le nombre d'arrêts pendant les **tournées** de **ramasses** ou de **livraison**. « On conçoit aisément qu'enlever plusieurs centaines de colis par jour auprès d'un seul client coûte moins cher au commissionnaire qu'enlever le même nombre de colis auprès de dizaines de clients différents. Le même raisonnement s'applique à la livraison pour la concentration des domiciles livrés », Branche, 2006 (20). C'est pourquoi TFE souhaite prévoir le nombre de lettres de voiture quotidien par agence du réseau.

Historiquement chez TFE, le poids des marchandises transportées est l'indicateur de référence. Un grand nombre d'indicateurs de productivité et rentabilité en découle. Le poids peut, par exemple, être converti en nombre de palettes au sol. C'est un indicateur

---

15. L'officiel des transporteurs magazine, Mars 2004

16. L'officiel des transporteurs magazine, Mars 2004

précieux quand l'objectif est de remplir les camions de 33 palettes.

Ainsi, une importante partie de la thèse se consacre à prévoir les ventes des agences de transport TFE. Par la suite, dans le contexte de l'entreprise TFE, nous préférons parler de prévision d'activité au lieu de prévision des ventes. L'activité est définie comme la combinaison des facteurs poids et nombre de lettres de voiture. La section 1.2.1 introduit la demande des clients et l'organisation que TFE met en place pour y répondre. Elle montre également comment la prévision d'activité joue un rôle déterminant pour augmenter la productivité tout en satisfaisant les attentes des clients. La section 1.2.2 souligne les échanges de marchandises entre les agences de transport du réseau. La section 1.2.3 introduit une suite d'actions à exécuter pour installer à un système de prévision. La section 1.2.4 définit le besoin de TFE en information prévisionnelle. Elle précise aussi à quel échelon de la pyramide et à quel horizon l'activité des agences de transport TFE doit être prévue. La section 1.2.5 pose

## 1.2.1 Enjeux de la prévision d'activité chez TFE

La tendance des magasins est de travailler en flux les plus tirés possibles. Il faut alimenter les linéaires en évitant de stocker. Les industriels répondent aux commandes quotidiennes des grandes et moyennes surfaces pour approvisionner leurs entrepôts. Les quantités de ravitaillement fluctuent quotidiennement selon les sorties de caisse des grandes et moyennes surfaces. « *Dans le cas de Carrefour, tous les soirs, toutes les données de vente en magasin sont centralisées à l'entrepôt, les commandes sont expédiées dès le lendemain. Ce réapprovisionnement automatique représente : 80% des ventes pour l'épicerie et 98% des produits frais* », Terrolle, 2004 (111).

Pour faire face, les gros industriels de l'agro-alimentaire établissent des prévisions de vente aussi bien à un rythme mensuel que quotidien. « *Le premier maillon de la chaîne logistique, c'est la prévision des ventes. Et cette prévision doit être la plus proche possible de la réalité. Elle constitue un avantage concurrentiel décisif sur son marché où le maître mot est « le juste-à-temps »* », Ayadi, 2005 (12). Ainsi, les industriels utilisent la prévision des ventes comme un outil permettant d'approvisionner les distributeurs en flux tendus.

Qu'en est-il du transporteur ? Nous avons mentionné dans la section 1.1.1.4 que la chaîne logistique le plaçait entre l'industriel et le distributeur. Pour satisfaire la demande de transport des marchandises entre les différents acteurs, le transporteur a également intérêt à anticiper ces transports. Or les industriels ne sont pas encore prêts à partager les informations prévisionnelles avec le transporteur. Ils craignent de donner trop d'information au transporteur qui, se voyant dans une position de force, pourraient en profiter

pour augmenter ses prix. C'est pourquoi le transporteur doit également prévoir la demande de son secteur d'activité. La section 2.1.1 montre que la prévision de la demande passe par la prévision des ventes, c'est-à-dire, pour TFE, la prévision du nombre de lettres de voiture et du poids des marchandises à transporter.

Une brève description de l'organisation du transporteur permet d'appréhender les enjeux de la prévision d'activité chez TFE.

Les chargeurs font leurs commandes tous les jours pour un transport le même jour.

- Les ordres de transports sont passés par fax ou EDI à 7h du matin. La marchandise est collectée chez le donneur d'ordre entre 8h et 12h. C'est le circuit de ramasse (voir section 1.1.2.2.2).
- La marchandise récoltée est déchargée sur le quai de l'agence TFE pour être regroupée par destination. C'est l'activité de groupage (voir section 1.1.1.1).
- Les marchandises groupées, sont expédiées entre 18h et 22h sur toute la France.
- Les agences du réseau qui reçoivent les marchandises expédiées et qui les livrent chez le destinataire final sont appelées agences de dégroupage. Tous les jours, après 19h, elles reçoivent, par EDI, les lettres de voiture à livrer.
- Suite au dégroupage, la tournée de livraison (ou distribution) se fait entre 8h et 12h, avant la tournée de ramasse.

Ainsi, dans 90% des cas, l'agence de transport reçoit les ordres de transport de l'industriel moins de 3 heures avant l'enlèvement de la marchandise. Elle a donc peu de temps pour s'organiser. Pour constamment répondre présent à ses clients, l'agence propose un plan de transport. Il comprend les horaires de collectes chez les chargeurs et les horaires d'arrivées en fonction des destinations. De cette façon, les clients utilisent le réseau de TFE aussi facilement qu'ils prennent un train. Par contre, il reste pour TFE la difficulté d'apprécier suffisamment à l'avance :

- le nombre de camions à planifier sur une tournée de ramasse, d'expédition ou de distribution,
- le nombre de manutentionnaires sur le quai pour traiter les marchandises collectées lors des tournées de ramasses ou reçues des expéditions du réseau.

En outre, il arrive que les quantités à transporter varient de 1 à 10 d'un jour à l'autre. Les opérations spéciales, comme les promotions, en sont la cause. Elles désorganisent le plan de transport. Comme les clients n'en communiquent pas la période, c'est à TFE de les prévoir. Ajoutons que TFE utilise l'intérim pour répondre à la variation d'activité. Or, l'intérim représente un coût important qu'il faut réduire. Pour ne plus subir mais anticiper l'activité, une prévision du **poids des marchandises passées à quai** et du **nombre de lettres de voiture journalier passé à quai** sera un réel outil d'aide à la décision. D'une part, TFE est capable de déterminer le nombre de camions en fonction des noms des

chargeurs, du poids des marchandises et du nombre de **positions** (ou **lettre de voiture**). D'autre part, TFE déduit du poids des marchandises passées à quai et du nombre de position, le nombre **d'heures productives sur le quai**.

Le transport pour les grandes et moyennes surfaces représente à peu près 12% du chiffre d'affaire (CA) de TFE. La règle des 80/20 ne s'applique donc pas forcément dans toutes les agences de transport du groupe. Quand bien même, si 20% des clients génèrent 80% du CA, il faut s'efforcer de satisfaire les 80% de clients restants. Du petit agriculteur ou épicier à l'industriel ou GMS, ils sont nombreux. Leurs demandes évoluent très rapidement. Pour tirer cette demande et ainsi être plus réactif, il faut établir des prévisions fiables de la demande. Or prévoir le nombre de lettres de voiture et le poids passés à **quai** quotidiennement équivaut à anticiper le besoin du client, à améliorer le niveau de service et optimiser la chaîne logistique (voir section 2.1.1).

Les **agences** sont aussi des entreprises indépendantes. Elles gèrent leur comptabilité et leurs propres clients. Elles se facturent les prestations et font jouer la concurrence. Elles parlent de clients internes quand il s'agit d'agences du groupe et de clients externes dans les autres cas. Dans le but de réduire les coûts de transport, de multiples combinaisons sont envisageables pour acheminer un produit. L'optimisation du chargement d'un camion ne consiste pas uniquement à le remplir. Quelle que soit la distance, il est parfois plus rentable pour une agence de ramasser un demi-camion de marchandises chez l'industriel et de le livrer directement chez le destinataire final, que de faire passer cette marchandise par le réseau, pour remplir les camions. Cet exemple montre que le déchargement, le tri et le chargement sur un quai sont une part importante de la facture. C'est pourquoi il est devenu urgent pour TFE de baisser les coûts de fonctionnement du quai. La prévision du nombre de lettres de voiture et du poids des marchandises passées à quai va l'y aider. Ajoutons que cette prévision est la matière première pour optimiser les ressources matérielles et humaines nécessaires à l'activité sur le quai.

« *Plus généralement, les exigences actuelles de qualité de service (rapidité, fiabilité, sécurité . . .) imposent aux prestataires et aux entreprises de transport de devenir des organisations réactives et flexibles, capables de faire face aux inévitables aléas et fluctuations de l'activité* », Fabbe-Costes, 1999 (49). La prévision de la demande est un outil pour y parvenir.

Nous venons de montrer que parmi le nombre de lettres de voiture et le poids des marchandises transportées, nous retenons, pour la prévision, que les marchandises passées à quai. Ces prévisions visent à fournir des éléments cruciaux pour :

- planifier des ressources matérielles et humaines,
- optimiser les règles de **ramasse**, d'**expéditions** et distribution,

- faire baisser les coûts de fonctionnement,
- formaliser le comportement à court, moyen et long terme des clients,
- atteindre une qualité de service élevée,
- limiter la dépendance vis-à-vis de l'incertitude.

Prévoir les indicateurs cités nécessite de maîtriser l'acheminement des marchandises entre l'expéditeur et le destinataire.

### 1.2.2 Suivi des marchandises dans le réseau TFE

Dans un but prévisionnel, le suivi des marchandises dans le réseau est un élément clé. Par définition, le réseau relie les agences du groupe entre elles par des échanges de marchandises. En théorie, cette connexion engendre des lois de cause à effet. Prenons une agence A, son activité en expédition et distribution dépend de ses ramasses, mais aussi des expéditions reçues du réseau. Ces dernières dépendent des ramasses du réseau. En pratique, il existe dans le réseau TFE plusieurs sous réseaux. Nous avons précédemment défini la plate-forme étoile comme une agence située sur un axe de communication routier stratégique et, de par cette situation, elle recevait les marchandises des différentes agences excentrées de la région pour les réexpédier à travers l'Hexagone.

Ainsi, comme l'indique la figure 1.7, chaque plate-forme étoile comporte son propre réseau d'agences. Les nœuds qu'elles représentent, permettent la connexion entre le réseau local et le réseau national. Mais les agences décidant ne pas passer par une plate-forme étoile pour rejoindre un autre réseau sont, dans certains cas, autorisées à le faire. Cette complexité ne permet pas de dire que toutes les expéditions d'une agence A vont vers l'agence B ou que toutes les livraisons de l'agence B proviennent des expéditions de l'agence A. Un moyen de suivre les flux est de prendre les lettres de voiture une par une. Sachant qu'il en existe plus de 100 millions, il paraît difficile, même avec les outils informatiques d'aujourd'hui, de les analyser individuellement. De même, suivre les différents

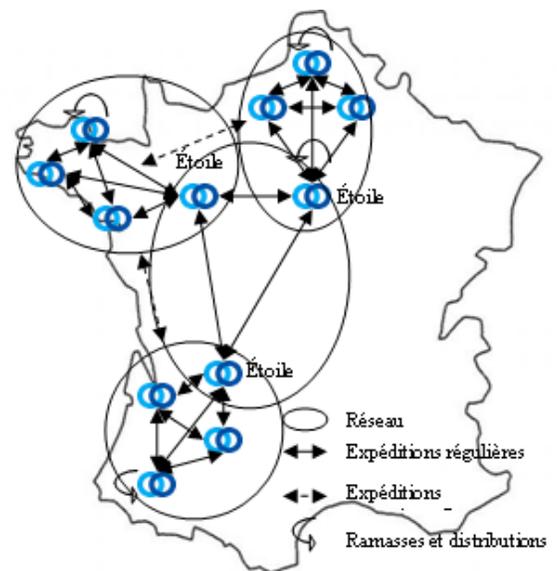


FIGURE 1.7 – Flux engendrés par le réseau

flux constitués de plus de deux points d'arrêts s'avère difficile à cause du nombre de possibilités illimitées entre les 70 agences du groupe. En revanche, TFE souhaite analyser les flux entre deux acteurs, soit entre une agence et son client, soit entre deux agences (voir matrice de la figure 1.9). Notons que deux entités forment un réseau minimal. La section 1.2.4 montre que les prévisions du nombre de **positions** et du poids des marchandises transportées entre deux acteurs de la chaîne logistique sont inutilisables. La maille choisie, c'est-à-dire entre deux acteurs, est trop fine (voir section 2.2.2). C'est pourquoi la thèse se concentre à prévoir **le nombre de positions et le poids des marchandises passées à quai et partant vers une tournée d'expédition ou une tournée de distribution**. Une perspective sera de prévoir le poids des marchandises transportées entre une agence et ses 5 à 10 plus gros clients ou agences du groupe.

### 1.2.3 Méthodologie et prérequis à l'installation d'un système de prévision

Hubert Carpentier, directeur de l'organisation et des systèmes d'information des supermarchés Match, déclare : « *il ne fait aucun doute que des prévisions de vente très précises engendrent d'importants gains de productivité* »<sup>17</sup>.

A l'échelle d'une grande entreprise comme TFE qui souhaite un système de prévision opérationnel commun à ses 70 **agences**, il est primordial de mettre en place une chaîne de processus. Comme l'indique la figure 1.8 un système de prévision repose sur :

- des données d'entrée issues d'un système d'information et du résultat des précédentes prévisions,
- des processus de correction et validation des prévisions fournies,
- un modèle mathématique construit à partir d'hypothèses statistiques et d'informations décrivant le phénomène à prévoir,

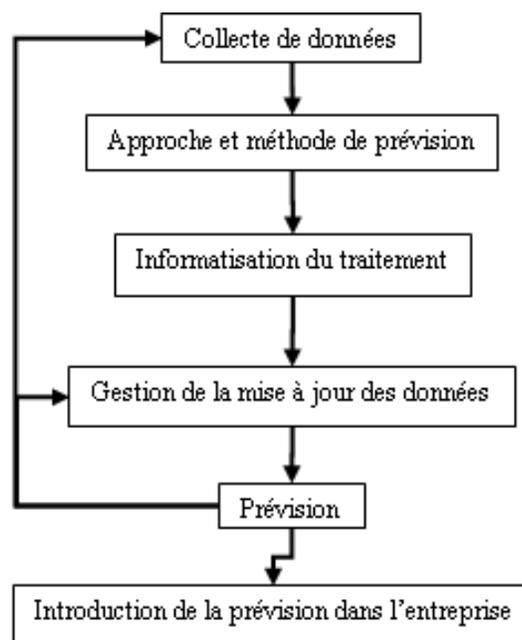


FIGURE 1.8 – Organigramme d'un système de prévision  
Source : R. Bourbonnais et JC Usunier

17. Success Story : Le système de prévision des ventes chez Match, Aldata Solution 2008

- un développement informatique du système de prévision permettant une mise à jour régulière et automatisé des prévisions,
- des pratiques de management pour introduire la prévision en entreprise.

Ainsi, avant même de se lancer dans une suite de traitement, d'analyse de données et de calculs des prévisions, il faut répondre à des questions comme : Quelles grandeurs sont à prévoir ? Quel doit être l'horizon de prévision ? Quelle est la fréquence des mises à jour ? Quelle erreur de prévision peut être perçue comme acceptable ? Après avoir répondu aux questions (voir section 2.2), les données internes à l'entreprise seront analysées et transformées dans le but de constituer un historique fiable (voir section 4.4.1 et 4.4.2). Il s'en suivra la recherche d'un modèle mathématique de prévision compatible avec les données disponibles (voir section 4.4.3). Il sera informatisé pour être exécuté à intervalle régulier selon les mises à jour des données (section 4.7). L'information prévisionnelle pourra être corrigée le cas échéant. Au final, elle sera diffusée par un système de reporting (voir section 4.7.3).

Hubert d'Hondt<sup>18</sup>, met l'accent sur le fait que « *de nombreuses organisations ont dépensé des sommes importantes à mettre en place des systèmes logistiques à base de progiciels de gestion intégrés et d'optimisation (ERPs, APS) pour réduire les stocks et les délais sans atteindre les objectifs visés et donc sans obtenir le retour sur investissement escompté* »<sup>19</sup>. Il n'est donc pas question pour TFE de faire l'achat d'un progiciel de prévision et de commettre la même erreur. Au contraire, le challenge de la thèse est de construire autour des préconisations de R. Bourbonnais, JC Usunier (18) et J. T. Mentzer, M. A. Moon (85) une démarche permettant de monter un système de prévision qui colle aux besoins réels de l'entreprise, avec un petit budget.

Mettre en place un système de prévision demande de la part de l'entreprise qu'elle possède les connaissances et les compétences couvrant au moins les quatre domaines suivants :

- la capacité de définir ses réels besoins en termes de prévision et les contraintes qui y sont liées,
- l'application de différentes méthodes de prévision,
- les procédures lui permettant de sélectionner les méthodes de prévision adéquates à des situations spécifiques,

---

18. Vice-président Global Services chez Cartesis

19. Tribune publiée par Hubert d'Hondt le 06 Novembre 2003 dans /www.journaldunet.com

- un support organisationnel pour l'application et l'utilisation des méthodes de prévision (87).

La suite de la thèse fournit à l'entreprise une réponse à ces différents points.

## 1.2.4 Définition des besoins de TFE en information prévisionnelle

Comme il a déjà été dit, historiquement, l'unité de référence chez TFE est le poids des marchandises transportées. Les marchandises étaient de la matière première transportée en sac de poids équivalent. Les contrats avec les chargeurs étaient signés sur la base du poids à transporter. La législation impose des poids de charge maximum aux camions. Avant l'arrivée des chariots de manutention et chariots élévateurs, la main-d'œuvre était proportionnelle au poids de la marchandise à manipuler. D'autre part, les sources officielles<sup>20</sup> utilisent comme indicateur de l'activité du transport, le poids des marchandises transportées.

Aujourd'hui, le poids ne semble plus être un indicateur prépondérant. Les produits sont très dissemblables. Du fait de leur packaging (commercial) ils prennent de plus en plus de volume. Une tonne de pots de yaourt n'est pas analogue en volume à une tonne de carcasses de viande crochétées. Admettons que la tonne de yaourt représente une dizaine de palettes europe au sol (PES) alors que la tonne de viande n'en prend qu'une. Grâce aux chariots de manutention, qu'il s'agisse de déplacer la palette de yaourt ou la palette de viande, l'effort physique est équivalent. Par contre, la tonne de yaourt est plus longue à déplacer, car elle nécessite plus de vas et vient pour traiter la dizaine de palettes. Ainsi, l'indicateur de l'activité du quai est relatif au nombre de palettes à déplacer. TFE étant capable de convertir les prévisions du poids des marchandises transportées pour un certain chargeur en quantité de palettes au sol, le **service quai** est un consommateur de ces prévisions.

Le **service exploitation** d'une agence de transport gère l'affrètement des camions. Il décide du nombre de camions à mettre à disposition sur une **tournée** de **ramasse**, d'**expédition** ou de distribution (voir section 1.2.1). Ce nombre dépend de la quantité de palettes à charger et du nombre de clients à desservir, soit le nombre de destinataires. Or, une **position** satisfait un **expéditeur** et un destinataire. Le nombre de clients à desservir est donc corrélé au nombre de positions journalier. Dès lors, le service exploitation nécessite des prévisions en nombre de positions. Le service d'exploitation décide également des marchandises qui remplissent un camion. Les camions supportent des charges maximales rarement atteintes. Par contre, ils sont limités à 33 palettes au sol. C'est pour-

---

20. Le Service de l'observation et des statistiques (SOeS), rattaché au Commissariat général au développement durable (CGDD)

quoi le service d'exploitation nécessite aussi des prévisions du poids des marchandises passées à quai. Après conversion en nombre de palettes au sol il estime le nombre de camions à affréter.

La **direction des ressources humaines** (DRH) a pour objectif la planification du personnel de **quai** et des chauffeurs. Actuellement l'activité quai est partagée en quai nuit et quai jour. Le quai nuit est chargé de remplir les camions de distribution avec la marchandise réceptionnée par les expéditions du réseau. Le quai jour réceptionne la marchandise ramassée dans la zone de chalandise de l'agence pour la charger dans les camions d'expédition. Suite à des expériences menées dans des agences TFE, il s'avère que cette organisation n'est plus nécessairement l'optimale. Pour l'améliorer et planifier les horaires de travail, la DRH souhaite obtenir des prévisions d'activité, non plus par expédition ou distribution, mais par tranches horaires.

La **direction d'exploitation** (DEX) est l'administration qui supervise l'ensemble des agences de la branche transport du groupe STEF-TFE. C'est aussi le donneur d'ordre du projet de recherche qu'est la thèse. Dans le but de satisfaire l'ensemble des utilisateurs potentiels, la DEX a transcrit les besoins en prévisions dans une matrice (voir figure 1.9). Les lignes de cette matrice représentent les provenances des marchandises passées à quai. Les colonnes représentent les destinations de ces mêmes marchandises. Le bloc du bas de la matrice croise les provenances et les destinations des marchandises qui ne passent pas à quai. Les cellules de la matrice représentent un poids ou un nombre de **positions** à une date.

La figure 1.5 de la section 1.1.2.2.2 montre que les marchandises passées à quai proviennent soit d'une **tournée d'expédition** (**segments EX**), soit d'une tournée de **ramasse** (**segments EN**), soit d'un amené à quai (**segments AQ**).

Les marchandises chargées sur le quai partent, soit à destination d'un client pour livraison, soit vers une agence de transport, appelée dégroupier. Le client destinataire peut se trouver dans la zone de chalandise du transporteur, comme à l'extérieur. Dans le premier cas nous parlons de « destinataire final local », dans le second cas, nous le nommons « destinataire final hors local ». Ils sont représentés dans la matrice de la figure 1.9 par les deux derniers blocs de colonnes.

Ainsi, cette matrice représente l'activité quotidienne d'une agence de transport TFE. L'ensemble des marchandises transportées par une agence y est représenté. La somme de leurs poids, ou du nombre de **positions** qu'elles représentent, est à lire dans la cellule nommée « 1 ». La DEX souhaite prévoir quotidiennement chacune des cellules de la matrice. Elle demande également à ce que cette matrice soit construite pour deux catégories de marchandises : « frais et non réfrigérés » et « surgelés ».

	Date	Sortie des marchandises (segments de sortie)													
		Segments EX				Segments LI, PQ									
		Agence A	Dégroupeur 1	Dégroupeur 2	...	Somme	Destinataire final (local) 1	Destinataire final (local) 2	... + Autres	Somme	Destinataire final (hors local) 1	Destinataire final (hors local) 2	... + Autres	Somme	Somme
Horaires	23h-24h	01h-02h			8h-9h	9h-10h			10h-11h	11h-12h					
Entrée des Marchandises passées à quai (segments d'entrée)	Segment EN, AO	Cliant remettant 1	Détail				Détail				Détail				5.1
		Cliant remettant 2	Détail				Détail				Détail				5.1
		...	Détail				Détail				Détail				5.1
		Somme	7			4.1				4.2				4.3	
	Segment EX, AO	Agence remettant 1	Détail				Détail				Détail				5.2
		Agence remettant 2	Détail				Détail				Détail				5.2
		...	Détail				Détail				Détail				5.2
		Somme				4.4				4.5				4.6	
	Somme		6.1	6.1	6.1	3.1	6.2	6.2	3.2	6.3	6.3	3.2	2.1		
			Dégroupeur 1	Dégroupeur 2	...	Somme	Destinataire final (local) 1	Destinataire final (local) 2	...	Somme	Destinataire final (hors local) 1	Destinataire final (hors local) 2	...	Somme	Somme
		23h-24h	01h-02h			8h-9h	9h-10h			10h-11h	11h-12h				
Entrée des Marchandises ne passant pas à quai	Client remettant	Segments ED				Segments LD				Segments LD				5.3	
		Segments ED				Segments LD				Segments LD				5.3	
		Segments ED				Segments LD				Segments LD				5.3	
	Somme				4.7				4.8				4.9		
Somme		6.4	6.4	6.4	3.3	6.5	6.5	3.4	6.6	6.6	3.4	2.2			
Somme		5.5	5.5	5.5		5.6	5.6		5.7	5.7		1			

FIGURE 1.9 – Matrice de l'activité d'une agence TFE

La figure 1.10 témoigne de différents horizons de prévision. La DEX souhaite des horizons de prévision au cas par cas. Certaines cellules de la matrice doivent être renseignées sur les 3 horizons de la figure 1.10, alors que d'autres seulement sur le long terme. Les cellules marquées « détail », « segment ED » et « segment LD » présentent des informations opérationnelles. Leurs prévisions sont attendues à un horizon à court terme (jour  $J$  plus 6 à 15 jours). Les cellules de la matrice représentant des sommes portent des informations plutôt considérées comme stratégiques. La DEX en attend des prévisions hebdomadaires à horizons de 26 semaines.

De plus, les objectifs de la DEX sont d'atteindre une erreur de prévision inférieure à 5% à

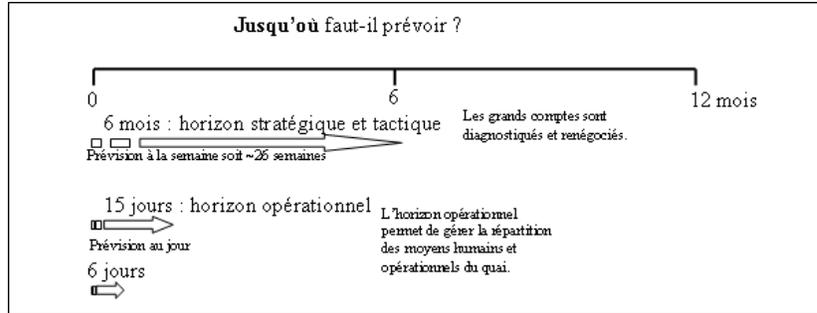


FIGURE 1.10 – *Les horizons de prévision*

la journée. Suivant l'activité des agences et des jours concernés, les 5% peuvent représenter entre 5 et plus de 50 tonnes pour une journée qui voit passer 100 à 1.000 tonnes au quai. L'idéal serait de ne pas se tromper de plus ou moins une semi-remorque soit 25 tonnes.

Les premières investigations ont vite démontré que la réalisation de cette matrice serait difficile. Généralement, une **agence** de transport TFE collecte des marchandises chez plus d'une cinquantaine de chargeurs et livre chez plus d'une cinquantaine de destinataires. Dans un souci de lisibilité, la matrice ne pourrait pas afficher l'ensemble de ces acteurs. Elle devrait donc ne retenir que les 10 premières provenances et les 10 premières destinations. Mais rien ne laisse envisager que les 10 premières provenances s'orientent vers les 10 premières destinations et vice versa. Donc même ce choix oblige à afficher un bien plus grand nombre de provenances et destinations. Au final, nous risquons de nous retrouver avec une matrice creuse.

D'autre part, au regard de la pyramide produit (voir figure 2.2 de la section 2.2.2), les cellules marquées « détail » sont des indicateurs résidant à la base de la pyramide. Or la section 2.2.2 nous apprend que, de par leur grand nombre, il est périlleux de prévoir ces indicateurs. Il est conseillé de les regrouper pour atteindre un palier supérieur de la pyramide. Une façon d'y parvenir est par exemple de rassembler les **segments d'expédition** suivant la région administrative des **expéditeurs** (voir nomenclature INSEE).

Enfin, l'erreur de prévision attendue ne peut pas être fixée seulement suivant des impératifs économiques. R. Bourbonnais (18) explique que l'erreur de prévision attendue est la moitié du coefficient de variation de la série à prévoir, corrigé des variations saisonnières (voir section 4.7.1). La section 2.2.11 argumente le choix d'une erreur de prévision acceptable.

Rappelons que le système de prévision a pour finalité d'apporter un outil d'aide à la décision pour les centres de production (Quai, Parc camion) et de logistique (Exploitation, DRH). Parallèlement, les prévisions pourront être utilisées à d'autres fins comme la

détermination des objectifs par la direction.

Aussi, le choix retenu est de commencer par prévoir :

- le nombre de positions et le poids des marchandises passées à quai. C'est la cellule « 2.1 » de la matrice,
- le nombre de positions et le poids des marchandises passées à quai et à destination d'une tournée de distribution. C'est la cellule « 4 .5 » de la matrice,
- le nombre de positions et le poids des marchandises passées à quai et à destination d'une tournée d'expédition. C'est la somme des cellules « 4 .4 » et « 4.6 » de la matrice.

Pour répondre aux attentes de planification des ressources, nous prévoyons :

- le nombre total d'heures productives sur le quai,
- le nombre d'heures productives sur le quai pour traiter les marchandises destinées à la distribution,
- le nombre d'heures productives sur le quai pour traiter les marchandises destinées à l'expédition.

Le tout représente 9 séries temporelles communes aux 70 agences de transport du groupe, soit 630 séries temporelles à prévoir toutes les semaines. L'historique des chroniques débute pour beaucoup d'entre elles le 1<sup>er</sup> Janvier 2002. La maille de temps entre deux observations est le jour. Les indicateurs cités seront prévus :

- sur un intervalle de temps quotidien, à un horizon de 24 jours,
- sur un intervalle de temps mensuel, à un horizon de 24 mois.

La prévision des autres cellules de la matrice sera envisagée lorsque l'entreprise TFE maîtrisera l'ensemble des contraintes liées à un système de prévision. Comme le disait Jean de la Fontaine, « rien ne sert de courir, il faut partir à point ».

### **1.2.5 Intégration du système de prévision au système d'information de TFE**

Comme le mentionne la section 1.2.3, un système de prévision nécessite un environnement et un développement informatique pour une mise à jour régulière et automatisée des données réelles et prévues. Voici quelques aspects informatiques nécessaires au système de prévision à installer.

Le système de prévision doit communiquer avec le système d'information d'Agrostar (voir section 3.1). L'information de référence est l'ordre de transport. L'ordre de transport donne

accès aux informations de chargement et déchargement : produit, unité logistique, lieu de déchargement-chargement, date, heure, tiers **expéditeur**, tiers destinataire, nombre de colis, type d'emballage, poids, catégories d'emballages, etc. La base de données est mise à jour quotidiennement avec les données de la veille. « *Aujourd'hui, ce référentiel qui compte plus de 300.000 fichiers tiers est un élément clé de notre système de traçabilité* »<sup>21</sup> indique Serge Capitaine, directeur général adjoint de TFE. D'autres informations, provenant des outils cités dans la section 1.1.2.2.4 seront utiles pour affiner les prévisions. Nous recensons les ordres de transport passés plus de deux jours avant la date effective de départ, les tarifications, les distances des trajets, les effectifs, etc.

Le système de prévision doit fonctionner sur le principe d'un ETL (Extract Transform Load) (voir section 4.6.1). Les données nécessaires sont extraites d'une base de données « métier » pour être sauvegardées sous un format lisible par le logiciel de traitement statistique choisi. Elles sont traitées par ce même logiciel et exportées vers un système de reporting. Pour agencer le tout, une application permettant de faire l'interface entre le logiciel de traitement statistique et le reporting doit être trouvée (voir section 4.6.4). Les traitements doivent être complètement automatisés. L'intervention humaine en bout de chaîne c'est-à-dire le destinataire des prévisions, peut être envisagée pour modifier les résultats en fonction des dernières informations conjoncturelles non prises en compte.

Concrètement, depuis plus de cinq ans, TFE collecte, pour chacune de ses **agences**, un grand nombre de données. Pour répondre à l'objectif des besoins prévisionnels, l'historique du poids des marchandises, leurs provenances et destinations est un point de départ. S'ajoute des données calendaires (jours fériés, vacances, fêtes), des données internes (perte ou gain d'un portefeuille client), des données externes (conjoncture économique, concurrence). Leurs sources doivent être informatisées.

L'activité du transporteur TFE est liée à la grande consommation (alimentation, pharmacie, horticulture). L'action marketing des producteurs et des distributeurs provoque des variations importantes de la consommation et donc du transport. Il s'en suit un bouleversement de la répartition des provenances et **expéditions** des marchandises du transporteur. Il devient alors intéressant d'intégrer dans le modèle de prévision des variables explicatives traduisant l'action marketing. Il est envisagé de récolter ces informations par l'utilisation de l'interface graphique des prévisions. Au cours du temps, l'utilisateur détecte des erreurs de prévision qu'il explique par un événement (voir section 4.8). Le prévisionniste prend alors en compte ce fait et l'intègre dans le modèle.

La fréquence de mise à jour est hebdomadaire. L'horizon de prévision est fonction de la réactivité de l'agence à s'adapter. Il ne sert à rien de prévoir pour le lendemain si la

---

21. L'officiel des transporteurs magazine, Avril 2005

planification des ressources se fait à la semaine.

Les résultats prévisionnels sont présentés sous forme de tableau accessible par une interface web (voir section 4.7.3). L'utilisateur se connecte avec un mot de passe, il choisit une agence TFE et voit apparaître le tableau suivant :

Ce tableau fournit les tonnages réalisés suivant les critères sélectionnés par l'utilisateur.

semaine 12	Lundi	Mardi	Mercredi	Jeudi	Vendredi	Samedi
Tonnage 2008	253	265	292	246	231	126
Prévision 2009	255	260	298	250	220	112
Ton. Réel 2009	259	258	NaN	NaN	NaN	NaN

TABLE 1.2 – tableau extrait de l'interface web

Il peut choisir entre domaines d'activités (Expédition, Distribution, Ramasse, Total) ou par client de l'agence.

### 1.3 Une démarche de projet industriel et scientifique

La section 1.2 a posé, avec l'entreprise TFE, les objectifs, les attentes, les contraintes et l'organisation de l'information prévisionnelle. La section 1.3 décrit la méthode utilisée pour conduire la thèse.

« *L'ingénierie du logiciel est l'application pratique des connaissances scientifiques à la conception et à l'élaboration des programmes informatiques et de la documentation nécessaire à leur développement, leur fonctionnement et leur maintenance* », Boehm, 1988 (16).

Cette définition de l'ingénierie du logiciel résume bien la démarche suivie dans cette thèse. Les premiers temps ont été consacrés à la formation et l'acquisition de connaissance. La compréhension qui en a résulté est retranscrite dans le deuxième chapitre de la thèse. En suivant une démarche scientifique, nous aboutissons non seulement à un modèle mathématique de prévision, mais aussi à un plan de développement du système de prévision (voir section 1.3.1). C'est ce que la définition entend par « connaissances scientifiques ». Le quatrième chapitre de la thèse centre ses efforts sur « la conception et l'élaboration » d'un outil informatique pour mise en production à échelle industrielle. La formation des utilisateurs, la maintenance du logiciel et son évolution font partie de l'industrialisation. Ainsi le système de prévision mis en place chez STEF-TFE est passé d'une version bêta à une version V1 et V2.1.

### 1.3.1 Analyse et découpage du projet de recherche

« Un projet correspond à la situation dans laquelle on se trouve quand on doit atteindre un objectif avec des moyens et dans un délai donné », Morley, 1999 (93). C'est-à-dire que dans la gestion de projet il faut chercher à équilibrer les objectifs en fonction des moyens et du délai imparti. Beynon-Davies, 1989 (14), décompose un projet de développement logiciel en trois phases : l'analyse, la conception, et la mise en œuvre. Le CNRS (32), dans sa recommandation de conduite de projet en Système d'Information (SI), ajoute la phase d'exploitation et la phase de maintenance et évolution.

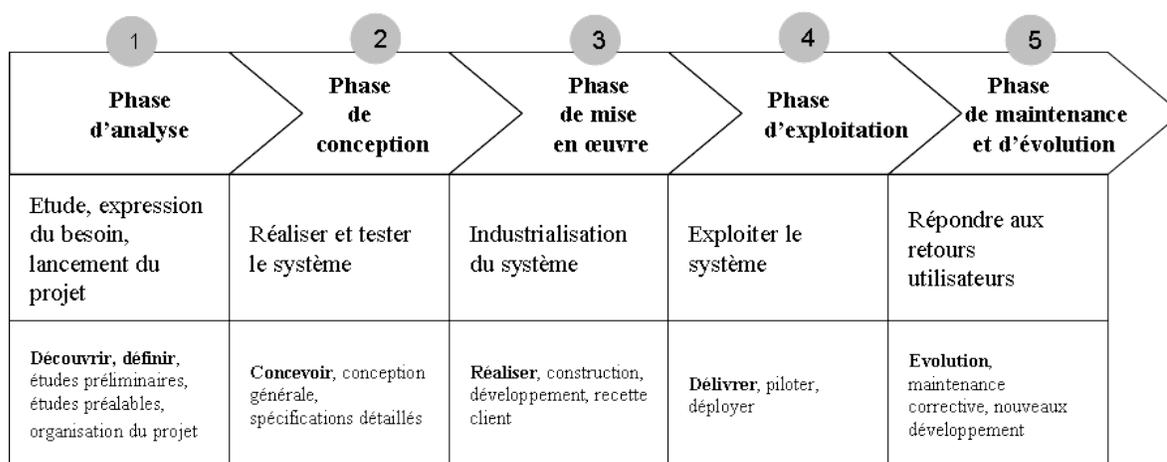


FIGURE 1.11 – Cycle de vie d'un projet en SI

La **phase d'analyse** (voir figure 1.11) cherche à clarifier les besoins, vérifier la faisabilité, évaluer les impacts et la rentabilité du projet.

Le projet de mise en place d'un système de prévision pour les agences TFE est confié à Agrostar, le pôle informatique du groupe. Le projet se nomme « **Horizons** ». Lors de son lancement, deux groupes de travail ont été mis en place : un COPIL (Comité de pilotage) et une équipe projet. Le COPIL est chargé du pilotage stratégique du projet, d'analyser les besoins et de définir le champ de l'étude. Suivant les directives internes d'Agrostar, le COPIL doit être constitué de personnes ayant le pouvoir décisionnel nécessaire au bon déroulement du projet. En cas de non consensus, le COPIL a recours au commanditaire du projet pour prendre une décision. L'équipe projet est présidée par un directeur opération de projet. Il est garant des objectifs opérationnels contractés dans le cadre du projet. Il est chargé du pilotage opérationnel du projet et des ressources. Il gère la relation contractuelle avec les prestataires et le client. Il reporte au comité de pilotage l'avancement du projet, ainsi que les risques et problèmes rencontrés.

Le COPIL du projet « Horizons », regroupe le responsable du service d'information déci-

sionnelle<sup>22</sup>, le responsable de l'ingénierie métiers à la direction d'exploitation (DEX)<sup>23</sup>, le directeur de thèse à l'UBS<sup>24</sup>. L'équipe projet est constituée de moi-même, de mon directeur de thèse et d'un stagiaire<sup>25</sup> pour 6 mois.

La phase d'analyse du projet « Horizons » a rassemblé des experts métiers aussi bien du côté des concepteurs (Agrostar) que du côté des commanditaires (Direction d'Exploitation TFE). Suite aux réunions de travail, plusieurs livrables ont été produits (voir annexes).

- Le « contrat de collaboration de recherche dans le cadre d'une convention cifre » analyse le budget et le retour sur investissement du projet. Il précise l'environnement du projet, les acteurs, les commanditaires et les maîtres d'œuvre. Il évalue une durée globale.
- Un « cahier des besoins » comme son nom l'indique recueille une première estimation des besoins pour réaliser le projet. Il établit un inventaire de l'existant. Il définit également différents scénarios possibles pour mener à bien le projet.
- Le « sujet de thèse » délimite les contours du projet et spécifie le bénéfice apporté par le côté « Recherche » du projet.
- Le « cahier des charges » exprime les objectifs et les besoins du client. Il définit les exigences fonctionnelles (fonctions de base) et les techniques (performances, aptitudes, interfaces, environnements, etc.) du produit ou service à réaliser. Il précise également le budget et le planning à respecter pour la réalisation du projet.

Cette phase d'analyse, qui peut également être qualifiée de découverte et de définition du projet, a permis de découper le projet « Horizons » en trois lots :

- Lot 1 : Recherche et développement d'un modèle mathématique de prévision d'activité,
- Lot 2 : Développement de requêtes d'extraction, de transformation et d'enregistrement dédiées au système de prévision,
- Lot 3 : Développement d'une interface web.

Les lots se suivent dans le temps et sont dépendants les uns des autres. Ce qui signifie que pour finir le lot 2 il faut terminer le lot 1 et il faut finir le lot 2 avant de terminer le lot 3. Chaque lot passe par les phases de conception et de mise en œuvre. Ce n'est qu'après la conception et les tests des lots que nous pourrons passer à la phase d'exploitation de l'ensemble.

Avec le lot 1 nous rentrons dans la **phase de conception**. L'objectif de la phase de

---

22. Benoît Colas

23. Jérôme Leroyer

24. Emmanuel Frénod

25. Maria Jaimes

conception est de savoir où nous mettons les pieds. Il faut éviter de se précipiter à rédiger le code et s'apercevoir trop tard que le sujet a été abordé du mauvais côté. C'est pourquoi l'écriture théorique d'un modèle mathématique de prévision est considérée comme de la conception. Ce n'est qu'après sa validation scientifique que le modèle mathématique sera développé et testé.

Alors que le projet débute en novembre 2006, la phase de conception démarre en janvier 2007. Le modèle mathématique de prévision spécifique à l'activité du transporteur n'est pas connu. C'est là qu'intervient la recherche. La recherche est le processus qui consiste à découvrir ce qui n'est pas encore connu. C'est pourquoi la recherche du modèle mathématique est menée suivant une **démarche scientifique**. Conventionnellement, cette démarche propose cinq paliers : l'observation, la formulation du problème, l'expérimentation, le traitement des résultats et la généralisation et développement.

Ainsi, la recherche du modèle mathématique de prévision d'activité a débuté par l'observation des données. Pour ce faire, il a fallu saisir la structure des bases de données « métiers » et de l'entrepôt de données (voir section 3.1.1). Ces notions ont permis d'extraire, dans un format adapté, les données représentant l'activité des agences TFE et les données pouvant influencer cette activité. L'ensemble des données récoltées a été analysé par des méthodes de data-mining et de statistique descriptive (voir annexes - Statistique descriptive, Procédure d'extraction).

Suite à l'observation, nous avons cherché à formuler le problème. Cette étape a conduit à la rédaction d'un document intitulé « la problématique » (voir annexes). Des premières possibilités de réponses avaient déjà vu le jour dans l'étape précédente, mais le choix de privilégier les méthodes endogènes au détriment des méthodes exogènes (voir section 2.3.3) s'est réellement fait après avoir posé le problème et soulevé les bonnes questions (voir section 2.2). Nous verrons tout de même dans la section 3.2.2 que des modèles de prévision basés sur les méthodes exogènes peuvent être viables pour anticiper l'activité des agences de transport du groupe STEF-TFE. Le système de prévision mis en place permet la minimisation des erreurs de prévision tout en minimisant les coûts : le respect des délais de mise à jour, les coûts financiers et les coûts en termes de temps.

La recherche du modèle de prévision a nécessité plusieurs phases d'expérimentation. Les premières sont apparues à la lecture de la bibliographie sur le sujet (voir section 4.3) et suite à l'analyse statistique des séries chronologiques. La formulation du problème a débouché sur la piste la plus pragmatique. Elle repose sur des méthodes connues et décrites dans l'ouvrage de Bourbonnais, en 2007 (18). La formulation finale du modèle auquel nous avons abouti découle d'une suite d'expérimentations. Il y a donc un cycle itératif dans la démarche scientifique. Nous entrons dans la boucle de l'itération en posant un modèle mathématique, nous vérifions ses hypothèses de base, son déroulement et ses résultats pour les comparer aux modèles précédents. La condition de sortie de la boucle est de trouver le modèle le mieux adapté pour résoudre le problème posé.

C'est en août 2007 que commence la généralisation et développement à proprement dit. Cette phase est aussi appelée **phase de mise en œuvre** du lot 1. La rédaction du modèle mathématique, noir sur blanc est achevée (voir section 4.4). La validation de la théorie du modèle mathématique de prévision a été faite dans Wilfried Despagne, (37; 41; 38; 39; 40; 42; 43) en 2008, 2009. Un programme informatique est développé. Il subit une panoplie de tests aussi bien d'un point de vue informatique que du point de vue des résultats obtenus (voir annexes - fiabilité résultats). Il en résulte 22 scripts rédigés dans l'environnement de développement R (voir section 4.7.1).

Le coût d'un projet est fonction du temps qui y est consacré. C'est pourquoi la table 1.3 relève le nombre de jours passés sur les 5 paliers d'une démarche scientifique. Ces chiffres sont à comparer aux 217 jours d'activité dans une année complète. Ajoutons que le développeur a saisi 1.740 lignes de code fonctionnelles.

Observation	Formulation du problème	Expérimentation et Traitement des résultats	Généralisation et développement
11 jours	11 jours	56 jours	19 jours

TABLE 1.3 – *Le cycle de R&D aboutissant au modèle mathématique et à son programme informatique en quelques chiffres*

De la même façon que le lot 1, le lot 2 va passer par les phases de conception et de mise en œuvre. Suite à la construction du modèle mathématique, nous connaissons les informations qu'il faut extraire pour calculer la prévision d'activité. Par contre, le système de prévision doit également afficher un certain nombre d'autres informations permettant à l'utilisateur de se faire sa propre opinion sur les prévisions ou valoriser la prévision. Ainsi, le développement des requêtes d'extraction, de transformation et de chargement (ETL) dédiées au système de prévision, a nécessité au préalable de réfléchir sur les informations que le système de prévision allait partager. Les méthodes de développement des requêtes ETL ou d'un site web sont connues. La difficulté réside dans l'architecture de l'ensemble et dans la maîtrise de ces outils. Les premières réunions ont soulevé l'enthousiasme et les idées ont fusé. Il était question d'extraire des données géographiques pour développer un site web avec application cartographique. L'idée étant d'intégrer à terme une notion de distancier entre les agences, ou encore de visualiser graphiquement, sur un fond de carte nationale, les flux expédiés entre agences. Une autre proposition était de construire un cube avec l'aide des outils de « reporting » et des « requeteurs ». Les axes d'analyse du cube étant le temps, les régions, les clients, la ramasse, l'expédition et la livraison (voir section 3.1.2). Après avoir exploré ces pistes, et d'autres, nous avons fini par choisir de présenter l'information prévisionnelle au travers d'un site web (voir section

#### 4.7.3).

La phase de conception du lot 2 a ainsi débouché sur la structure que porte un ETL. La table 1.4 montre qu'il existe plusieurs sources de données (HP,GTI, GTA, Infocentre). Elles sont décrites dans la section 3.1.

	<b>Extract</b>	<b>Transform</b>	<b>Load</b>
sources {	HP	Uniformisation des données	Insertion dans les tables dédiées entreposées dans l'infocentre
	GTI	Nettoyage des données	
	GTA	Prévision	
	Infocentre		

TABLE 1.4 – *procédure ETL*

Pour afficher les informations demandées dans le site web, il faut les produire (Extract + Transform) et les stocker (Load) dans une base de données. Les tables dédiées au projet « Horizons » sont entreposées dans un entrepôt de données (voir section 4.2). Les tables spécifiques au projet « Horizons » sont alimentées par des procédures d'uniformisation des données, de nettoyage des données et de prévision (voir schéma 1.4 et section 4.4). Nous reviendrons en détail sur les procédures d'alimentation de mise à jour de données et de leur prévision dans le chapitre 4.

La table 1.5 présente les temps passés sur les tâches de conception et développement du lot 2.

Conception	Développement		
	Historique expe/distri	Historique heures productives	Prévision 2007 / 2008
23 jours	25 jours	10 jours	6 jours

TABLE 1.5 – *Le cycle de conception et de mise en œuvre du lot 2 en quelques chiffres*

Le lot 3 consiste à présenter les données prévisionnelles. Une partie de la réflexion sur l'information à présenter et la façon de la présenter a été menée lors de la conception du lot 2. L'idée est de développer une interface web simple, ergonomique et conviviale permettant à la fois :

- d'analyser visuellement et de manière plus approfondie l'activité d'une agence,
- de saisir et de s'approprier les prévisions proposées.

La description de l'interface graphique est décrite dans la section 4.7.3. Mais avant d'aboutir à ce niveau de précision, un ensemble d'étapes a dû être franchi. Pour préparer les utilisateurs à l'arrivée d'une nouvelle application et pour les faire participer au projet, nous avons commencé par leur envoyer un questionnaire (voir annexes - Questionnaire). L'ensemble des questions cherche à recueillir la perception qu'ont les utilisateurs sur la prévision d'activité. Le nombre de retour n'a pas été satisfaisant pour en déduire une réponse collective. Nous avons alors pensé qu'il serait plus simple d'aller vers les utilisateurs avec une démonstration du site que nous cherchons à développer. Cette démarche a eu un certain succès et a permis de faire évoluer nos premières réflexions. Ainsi, nous avons affiné le détail de nos 4 fenêtres principales (Proposition de prévision, Indicateurs de confiance, Correction des prévisions, Indice de productivité) et ajouté de nouvelles fenêtres. Cette démarche a aussi permis d'impliquer l'utilisateur final à la conception du produit qu'il sera amené à utiliser.

Les lots 1, 2 et 3 sont programmés sur un serveur de développement. Ils passent les tests d'intégration pour s'assurer du bon fonctionnement de chaque composant, des liens qui les unissent et de leurs performances face à des données volumineuses. La véracité des données est maintes fois contrôlée. Les échantillons sont envoyés à des spécialistes métiers pour qu'ils valident les indicateurs.

La mise en production se fait suite à un test à taille réelle. En Juin 2008, 5 agences pilotes (Dijon, Nantes, Mâcon, Lille, Niort) testent pendant quatre mois le système de prévision. Les chefs d'agences, les CGO (Contrôleur de Gestion Opérationnel) et les chefs de quai reçoivent une formation pour utiliser l'application. Le but étant qu'ils se l'approprient et qu'ils en tirent le meilleur pour résoudre des problématiques métiers courantes. Le retour des utilisateurs est étudié dans la section 4.9.

La mise en production du site « Horizons » se fait en Septembre 2008. Les composants formant le système de prévision, sont basculés sur des serveurs de production. Pour ce faire, il faut installer les programmes adéquats, les paramétrer, les faire communiquer, installer les packages et les scripts développés pour le système de prévision, configurer les mises à jour automatiques, vérifier les temps de traitement, ajuster des leviers si nécessaire.

En Octobre 2008, les lots 1, 2 et 3 passent en **phase d'exploitation**. Ce qui signifie que le système de prévision est livré à la DEX. Les chefs de sites sont avertis par un mail de la DEX que les prévisions d'activités sont disponibles sur le site « Horizons ». Une communication pour présenter le projet et l'outil qui en découle, est publiée sur l'intranet du groupe.

La formation des utilisateurs va faire l'objet d'un lot spécifique. C'est une activité à part

entière qui nécessite la rédaction de documents pour le support de la formation (voir annexes - Support de formation), une prise de contact avec les utilisateurs, une organisation des plannings de formation et des déplacements pour se rapprocher des utilisateurs.

L'utilisation quotidienne du site web provoque le début de la **phase de maintenance**. De nouvelles tâches apparaissent :

- le suivi des mises à jour,
- la vérification des données affichées,
- l'assistance aux utilisateurs,
- la gestion des dysfonctionnements qui subsistent.

A cela s'ajoute :

- ce que nous avons déjà mentionné, la formation des utilisateurs,
- l'ajout d'agences n'ayant pas été prises en compte dans la première version et sur demande de la Dex,
- des évolutions mineures sur le choix et la définition des indicateurs,
- des évolutions mineures sur la mise en page de l'interface web.

En Janvier 2009, le site compte plus de 300 connexions. Les connexions sont tracées en récupérant l'adresse IP du poste client. Une même adresse IP n'est comptée qu'une seule fois par jour. Il est possible de retrouver l'agence à laquelle appartient l'adresse IP. Ainsi, des statistiques d'utilisation du site sont produites tous les mois.

La table 1.6 présente les temps passés sur les tâches de conception et développement du lot 3.

Conception	Développement
6 jours	28 jours

TABLE 1.6 – *Le cycle de conception et mise en œuvre du lot 3 en quelques chiffres*

La phase d'exploitation comme la phase de maintenance des lots 1, 2 et 3 est en cours à l'heure de la rédaction de cette thèse. Le transporteur est équipé d'un système de prévision qui fonctionne quasiment en autonomie. Les utilisateurs sont chaque mois plus nombreux à y récolter de l'information prévisionnelle. Ils font remonter leurs attentes pour le développement des versions futures du système de prévision. Des projets locaux de planifications sont montés. Ils prennent comme sources les prévisions fournies par « Horizons ».

La maintenance du système de prévision de l'activité quotidienne consiste à surveiller le

bon déroulement des développements informatiques ; à ajouter, si besoin, une agence de transport dans le système de prévision ; à répondre à des modifications mineures ; à assister les utilisateurs, répondre à leurs interrogations et les conseiller, etc.

Nous pourrions penser qu'à ce stade le projet est terminé. Mais voilà, le succès du projet et surtout l'enthousiasme qu'il a soulevé vis-à-vis des prévisions fournies, a poussé la direction à continuer. Nous avons ainsi défini 5 lots :

- Lot 4 : Communication interne et formation des utilisateurs,
- Lot 5 : Prévision du tonnage par demi-journées en entrée et sortie de quai,
- Lot 6 : Prévision du tonnage par région en sortie de quai,
- Lot 7 : Prévision mensuelle des indicateurs à horizon de 12 mois pour le budget,
- Lot 8 : Modélisation mathématique des flux circulant sur un quai de messagerie TFE.

Pour être complet dans l'inventaire des lots, nous ajoutons deux lots supplémentaires :

- Lot 9 : Promotion de la thèse du côté académique (articles, conférences, rédaction),
- Lot 10 : Management (gestion de projet, réunions).

Le planning de la figure 1.7 nous permet de résumer les temps de charges (voir section 1.3.2) et de leur ordonnancement dans le temps sur ce projet à long terme. C'est un macro-planning dans le sens où, dans un mois donné, il est possible de réaliser plusieurs tâches. Le planning montre que les développements sous R (lot 1) ont dû être réajustés en Avril 2008, en fonction des besoins de la conception des requêtes ETL (lot 2). Le développement de l'interface Web (lot 3) n'a pu commencer que lorsque le lot 2 rendait ses premiers résultats en Novembre/Décembre 2007. La première formation a été donnée en Mars 2008 à Dijon. C'est un site pilote. Suite à ce retour d'expérience, des améliorations du système de prévision ont été possibles. Suite à la mise en production en Août 2008, les formations ont pu reprendre.

Le lot 5 est en phase de développement. Des indicateurs sont calculés. Nous attendons la recette du commanditaire. Plusieurs aller-retours entre le commanditaire et le développeur sont à envisager. Le lot 6 est en cours de développement. Les premiers résultats sont extraits en Mai 2009. Le lot 7 est réalisé par une étudiante stagiaire de 6 mois (entre Avril et Septembre 2009) en master de statistique et marketing. Sur l'exemple de la méthode de prévision journalière, elle produit un modèle mathématique de prévision mensuelle. Elle développe également les requêtes d'ETL et l'interface web dédiées (voir section 4.10).

La modélisation mathématique des flux circulant sur le quai (lot 8) a fait l'objet de la rédaction du modèle théorique (voir section 5.1). Une application informatique a été développée pour démontrer sa faisabilité (voir section 5.4).

		2007											
		1	2	3	4	5	6	7	8	9	10	11	12
lot 1		■	■	■	■	■	■	■	■	■	■		
lot 2				■	■	■			■	■		■	■
lot 3													
lot 4													
lot 5													
lot 6													
lot 7													
lot 8													
lot 9			■	■	■								
		2008											
		1	2	3	4	5	6	7	8	9	10	11	12
lot 1					■								
lot 2		■	■	■	■	■			■				
lot 3		■	■		■			■		■	■		■
lot 4				■				■		■	■	■	■
lot 5						■				■	■	■	
lot 6													
lot 7													
lot 8						■	■						
lot 9		■	■	■		■				■	■		
		2009											
		1	2	3	4	5	6	7	8	9	10	11	12
lot 1													
lot 2													
lot 3		■	■	■	■	■							
lot 4		■	■	■	■	■							
lot 5													
lot 6				■		■							
lot 7				■	■	■	■	■	■	■			
lot 8		■	■										
lot 9				■	■	■				■	■	■	■

TABLE 1.7 – *Planning grosse maille du projet « Horizons »*

Le lot 9 englobe tout ce qui concerne la présentation des travaux de recherche devant la communauté scientifique française et internationale. Ceux sont des tâches ponctuelles

qui nous amènent à prendre du recul sur les travaux pour les publier dans des revues scientifiques (37; 38; 39; 40; 42; 43) ou pour les rendre accessibles au plus grand nombre (fêtes des sciences par exemple). Le lot 10 ne figure pas dans le planning, car il nécessite de passer du temps quotidiennement. Ce sont les réunions à organiser, les réponses à donner rapidement à des problèmes ponctuels, l'encadrement des stagiaires, le suivi du projet et son encadrement, etc.

### 1.3.2 Risque et estimation des coûts du projet de recherche

Les risques liés à un projet de recherche sont nombreux. Il y a le risque de ne pas déboucher sur un résultat ; les résultats ne sont pas à la hauteur des attentes ; les utilisateurs rejettent le système ; le système développé est trop compliqué ou non adapté au besoin des utilisateurs ; les budgets sont dépassés ; le projet est abandonné en cours de route.

Le profil de risque (Morley, 1999) (93), est identifié par 6 critères mesurés sur une échelle de 0 à 5. La table 1.8 montre les résultats liés au projet qui nous concerne. Pour réduire le risque lié à la taille (voir table 1.8), nous avons découpé le projet en sous-projets que nous avons nommés des lots (voir section 1.3.1).

Le risque attaché à la difficulté technique (voir table 1.8) correspond à la nouveauté technologique et scientifique, à l'échelle de l'entreprise. Ni les techniques scientifiques, ni les outils informatiques utilisés n'ont fait l'objet de recherches précédentes. Il va de soit que les outils informatiques choisis ont fait l'objet d'une analyse approfondie pour vérifier leur capacité d'intégration dans le système d'information du groupe. C'est ainsi que nous avons baissé le risque d'intégration (voir table 1.8). Cependant, des ajustements peuvent toujours être nécessaires.

La configuration organisationnelle (voir table 1.8) est liée au nombre de décideurs. Plus il y a de services ou entités liés au projet, moins les décisions sont faciles et rapides à prendre. En ce qui concerne le projet de mise en production d'un système de prévision, les orientations stratégiques sont présentées par le chef de projet (moi-même) au commanditaire (la DEX). Il prend les décisions qui lui semblent les plus justes. Les décisions techniques sont prises par le chef de projet. Souvent, les difficultés résident dans le choix des priorités à prendre. A la vue des premiers résultats, les utilisateurs captivés, font de nombreuses demandes d'évolutions. Leur développement est alors à hiérarchiser.

Le changement (voir table 1.8) désigne le fait que nous proposons aux utilisateurs de repenser la gestion de leur activité. Nous leur demandons à l'avenir de planifier la gestion de leurs ressources. Pour ce faire, nous leur proposons de se familiariser avec un nouvel outil de prise de décision. Le fait de passer d'une analyse a posteriori à une anticipation

Nature du risque	Degré du risque pour le projet					
	0	1	2	3	4	5
taille du projet					X	
difficulté technique					X	
degré d'intégration		X				
configuration organisationnelle		X				
changement			X			
instabilité de l'équipe projet	X					

TABLE 1.8 – Analyse du risque lié au projet « Horizons »

de l'activité est un changement profond. D'autre part, les utilisateurs ont encore souvent peur que la machine remplace leur savoir-faire. Pour limiter ce risque, l'accompagnement par la formation est primordial. Un autre changement provient des outils qui n'ont pas été utilisés auparavant dans l'entreprise. C'est le cas du logiciel de traitement statistique R. L'intégration de ces nouveaux outils dans le système d'information du groupe nécessite du temps de développement et un effort de conception/innovation important.

L'instabilité (voir table 1.8) de l'équipe projet sur une durée de trois ans est théoriquement forte. Agrostar fait appel à des prestataires externes (comme les SSII) pour le développement des applications. Or, le turn-over des Sociétés de Services en Ingénierie Informatique (SSII) est important. Le transfert de connaissance n'est jamais assuré à 100% lors d'un changement de développeur. Le risque d'erreur de développement dû à des erreurs d'interprétation est alors élevé. Ce qui a un impact sur les délais et sur la cohérence de la conception. Mais l'équipe du sujet de thèse est restreinte. Elle ne fait pas appel à des prestataires externes. Elle est passionnée par les travaux qu'elle mène. Le risque zéro n'existant pas, nous l'avons estimé à 1 sur une échelle de 5.

La charge représente la quantité de travail nécessaire, indépendamment du nombre de personnes qui vont effectivement réaliser ce travail. La charge peut s'exprimer en jour/homme, mois/homme ou année/homme. Elle permet d'estimer le coût des ressources humaines du projet (voir tables 1.3, 1.5, 1.6). La durée dépend du nombre de personnes. Si la charge est estimée à 30 jours/homme et qu'on lui affecte 2 personnes, la durée sera de 15 jours. Dans le cas de ce projet de thèse, nous estimons qu'une charge est réalisée par une seule personne (voir planning de la figure 1.7).

Le coût financier du projet pour l'employeur Agrostar est estimé à moins de 200 K€ répartis 3 ans. Ce coût peut être comparé au coût d'une licence d'un progiciel de prévision. Pour la taille d'un groupe comme STFE-TFE, l'entrée de gamme pour l'achat d'un progiciel de prévision se situe entre 30 et 80 K€ annuel. Sur ce créneau nous trouvons les progiciels FuturMaster, Predicast et Walter's. Les éditeurs de ces logiciels sont, dans

le même ordre, FuturMaster, Aperia, BSK (voir section 2.3.4.3).

Les éditeurs cités ont développé des outils bien adaptés aux produits de la grande consommation, agroalimentaire, pharmacie, mode. Dans le haut de gamme, nous pouvons citer les éditeurs, Gold-Forecast de la société Aldata, n.skep demand planning de chez Dynasys et SAP SCM Demand Planning de chez SAP. Ces progiciels coûtent entre 50 et 1.000 K€<sup>26</sup>.

---

26. Supply Chain Mag n°17 Septembre 2007 Progiciels de prévision : Des outils très sollicités

# Chapitre 2

## Etude d'un système de prévision et de reporting

### 2.1 Les concepts autour de la prévision des ventes

#### 2.1.1 De la demande aux ventes

Nous avons mentionné que pour planifier son activité, l'entreprise souhaite connaître la demande de ses clients, pour y répondre au mieux. La demande, c'est ce que le client achèterait s'il le pouvait. C'est son envie d'accéder à un bien ou service proposé. Cette envie et ce pouvoir sont influencés par le contexte socio-économique qui l'entoure. S. Meunier Martins, 2003 (87), souligne que les ouvrages traitant du sujet de prévision font souvent, par abus de langage, l'amalgame entre la prévision de la demande et la prévision des ventes, c'est pourquoi nous définissons les deux notions.

En marketing, la demande d'un bien ou service est une fonction qui dépend du prix du bien ou service, du prix des biens ou services concurrents, du revenu des consommateurs et de leur goût. Nous percevons immédiatement que les variables de cette fonction sont difficiles, voire impossibles, à mesurer. En effet, le goût d'un consommateur est une notion abstraite et plus encore lorsque qu'il est question du goût d'un panel de consommateurs. Il semble également difficile de prévoir les goûts futurs des consommateurs. L'estimation du revenu des consommateurs reste souvent approximative. Quels biens ou services concurrents choisir ? Il est souvent difficile de mettre à la même enseigne deux biens ou services tellement ils ont leurs propres caractéristiques.

Les ventes sont liées à la capacité de l'entreprise à accepter les ordres d'achat des consommateurs. En pratique, les ventes sont les quantités de biens ou services pro-

duites et vendues par l'entreprise à ses clients. Cette donnée est facile à calculer. Elle sert pour une multitude d'indicateurs (CA, prix, etc.). C'est pourquoi toutes les entreprises savent l'obtenir, et mieux encore, elles l'ont calculée au fil du temps, à intervalle régulier. Il en découle un historique des ventes.

Nous souhaitons prévoir la demande, mais celle-ci est difficilement mesurable. S. Meunier Martins, 2003 (87), rappelle dans sa thèse, que sous certaines conditions, les ventes sont une bonne approximation de la demande. Il faut notamment que la demande soit parfaitement alimentée par les ventes. C'est pourquoi nous nous rabattons sur les ventes. Elles sont quantifiables et historisées. Cet historique va composer la source des prévisions de vente issues d'un modèle mathématique. En effet, le principe des modèles mathématiques est d'extrapoler la donnée passée dans le futur.

Le nombre de biens ou services vendus reste le meilleur indicateur pour apprécier la demande du marché. La prévision des ventes doit permettre d'anticiper la production future et ainsi de s'y préparer en planifiant ses ressources par exemple. En anticipant les ventes, l'entreprise cherche également à tirer les ventes pour être plus réactive et répondre aux nouveaux besoins des clients.

Notons tout de même, qu'en règle générale, les ventes ne permettent pas de faire ressortir les situations de marché dans lesquelles la demande est supérieure à l'offre et vice versa. Les chiffres permettant de donner l'écart entre ce que le consommateur aurait voulu acheter et ce qu'il a réellement pu acheter sont rares. Cet écart peut être dû à de nombreuses causes, rupture de stock, délai de livraison trop long, capacité de production insuffisante, etc. En se fondant sur des historiques de vente, le risque est de se priver de ventes potentielles en n'adaptant pas correctement l'offre à la demande.

### **2.1.2 De la prévision au Système de prévision**

C'est un fait, l'homme cherche à explorer l'avenir. Nous sommes passés du temps des prophéties au temps des sciences. La prévision n'est pas une science exacte, ce qui n'empêche pas les chercheurs d'avoir posé des fondements pour avancer sur le sujet. Attardons nous sur la vision qu'en ont les économistes et les mathématiciens. A son époque déjà, le scientifique, Léonard de Vinci (1452-1519), déclare « *Ne pas prévoir c'est déjà gémir* ». Deux siècles plus tard, l'économiste Adam Smith (1723-1790) précise que ce n'est pas le prévisionniste qui donne le rythme du marché mais il cherche à l'anticiper. Dans la première moitié du siècle dernier, l'économiste Jan Tinbergen (1903-1994), suite à la crise économique de 1929, cherche des solutions pour prédire les crises futures. A la fin des années 1960, Robert Klein construit le « Wharton Econometric Forecasting Model ». Ce modèle analyse les fluctuations économiques à l'échelle d'un pays pour prédire les récessions. Plus récemment, et en nous recentrant sur l'échelle de l'entreprise,

citons les ouvrages Granger, 1986 (58), Ansion, 1990 (6), Mélard, 1990 (89), Wei, 1994 (118), Brockwell et Richard, 1996 (21), Markridakis, et al., 1998 (81), Armstrong, 2001 (7), Bourbonnais et Usunier, 2007 (18) pour regrouper différentes méthodes de prévision mathématique utilisées en économie et notamment en prévision des ventes. Les mathématiques ne sont pas le seul outil pour aboutir à une prévision. Helmer et Dalkey, 1963 (61) expérimentent la méthode Delphi. Elle consiste à demander l'opinion d'un expert. Enfin, les auteurs Moon et Mentzer, 2000 (91), Fildes, et al., 2006 (51) se sont penchés sur l'insertion de la prévision des ventes en entreprise, et plus précisément ce qui est appelé aujourd'hui les systèmes de prévision (ou FSS : Forecasting Support System en anglais).

Avant d'introduire ce qu'est un système de prévision, définissons ce que nous entendons par prévision. La prévision est un scénario, très probable si le contexte reste inchangé dans le futur, et connaissant la situation présente et passée. La prévision consiste à projeter les récurrences. Selon les techniques utilisées, les chiffres prévisionnels seront différents. Ainsi, il existe une infinité de prévision. La prévision des ventes est l'image la plus probable de ce que les ventes futures seront. Elle utilise des méthodes statistiques pour déduire des informations présentes, une vision future.

Fitoussi, 1997 (52), nous met en garde, pour que nous ne soyons pas tenté d'accorder une confiance aveugle aux prévisions. Il explique qu'il « *est utile de comprendre les limites des prévisions pour ne point s'en servir comme argument d'autorité* ». Rappelons qu'une condition fondamentale des prévisions est de considérer la continuité dans le futur de ce qu'ont été les événements passés. Or, d'une part, (Fitoussi, 1997) fait remarquer que certaines informations cruciales ne sont pas vraiment disponibles (dans le présent), « *car, pour l'essentiel elles dépendent de décisions à venir et qu'il n'existe pas vraiment de théorie permettant de déduire des données existantes ce que seront ces décisions* » (futures). D'autre part, les ventes sont sujettes à l'aléa. Les ventes sont influencées par de nombreux facteurs externes, ce qui les rend volatiles. Nous en concluons que les conditions posées par les modèles de prévision, sur les données utilisées et sur le cadre théorique, ne sont jamais respectées. Ainsi, par définition, une prévision est une donnée erronée. C'est pourquoi il n'est pas facile de faire des prévisions, c'est même souvent risqué d'en fournir car notre compétence est remise en cause. Mais les prévisions, de manière générale, sont fondamentales.

Sans prévision il n'est pas possible de planifier. Chacun d'entre nous se fie aux prévisions météorologiques pour s'habiller le matin. Si nous prenons le parapluie pour sortir, c'est qu'il pleut ou que nous prévoyons qu'il va pleuvoir. Nicholas Seiersen, 2006 (105), rappelle que le but de l'entreprise est de piloter les ressources matérielles et humaines vers les objectifs prévus.

Attardons-nous sur les objectifs des ventes. Il semble parfois exister une confusion de la part des entreprises, entre prévision de vente et objectif des ventes. Si nous pouvons simplement poser des objectifs de vente, et attendre des commerciaux et du marketing qu'ils

les atteignent, pourquoi faire des prévisions de vente ? Dans la prévision, nous cherchons à réduire l'erreur entre le réalisé et le prévu. En fixant des objectifs, nous cherchons à les atteindre, voire les dépasser. Les fonctions de planification et objectif devraient être informées par les prévisions de vente. Les objectifs et la planification devraient être déduits des prévisions. La réciproque est fautive. C'est-à-dire que les objectifs et la planification ne peuvent pas déterminer les prévisions. Et c'est pourquoi Moon et Mentzer, 2000 (91), témoignent qu'il n'est pas possible de faire de la planification ou de dresser des objectifs sans faire de la prévision au préalable et que les termes ne sont pas synonymes, il y a un ordre. Enfin, il doit exister une forte connexion entre le business plan et la prévision des ventes.

Puisque la prévision est une condition « sine qua non » pour piloter l'entreprise, alors il vaut mieux de mauvaises prévisions que pas de prévisions du tout. Mais tout l'art de l'entreprise est d'optimiser l'infrastructure existante en se basant sur les prévisions les plus justes. (Rigoureux de chez Catalliances, 2001) confirme, « *les prévisions sont un point clé de l'optimisation des ventes et de la qualité de service. Elles sont calculées à partir des historiques de vente, corrigés par un certain nombre de paramètres maîtrisés par les entreprises* ».

Des résultats d'études, repris par Armstrong et Brodie, 1999 (8), montrent qu'en 1987, 99% des entreprises préparent des prévisions quand elles développent un plan marketing, 93% des entreprises (1975) indiquent que les prévisions sont un des aspects les plus critiques, mais très important pour le succès de l'entreprise. En 1985, une étude montre que la prévision est le domaine le plus commun au travers de la diversité des activités des entreprises. Ainsi, la prévision fait partie du Business Performance Management (BPM). Ce sont de puissants environnements qui ont pour but d'unifier les aspects traditionnels de restitution et de reporting des performances de l'entreprise avec de nouvelles fonctionnalités permettant la simulation et l'anticipation de décisions stratégiques pour l'entreprise. Le management de la Supply Chain (chaîne logistique) n'est pas une exception. Fildes et al., 2006 (51), confie que la prévision y joue donc un rôle clé. Rappelons que la Supply Chain désigne l'ensemble des intervenants entre la fabrication d'un produit chez l'industriel et sa mise en vente au consommateur final

En somme, la prévision en entreprise n'est ni une décision prise par un manager, ni un chiffre issu d'une boîte noire, c'est un ensemble d'outils, de procédures et d'organisations permettant de parvenir à un résultat prévisionnel fiable. En effet, une bonne prévision est plus qu'un chiffre, elle doit être utilisée en concordance avec d'autres informations disponibles. C'est ce que nous appelons un système de prévision.

L'analyse statistique sert de base aux systèmes de prévision de vente. La prévision peut être considérée comme le résultat issu d'un modèle prévisionnel. Le système de prévision, quant à lui, inclut dans une démarche verticale et transversale, un modèle prévisionnel (voir section 4.4), la mise en production d'outils informatiques (voir section 4.7) et une procédure d'utilisation (voir section 4.8). La démarche verticale désigne l'ensemble des

liaisons existantes entre la consolidation de l'information de sources différentes, la simulation, le calcul de plusieurs variantes de l'activité future d'une entreprise et la restitution de l'information. Le but d'un système de prévision étant, avant tout, d'alerter l'entreprise. La démarche transversale vise à comprendre la relation qui existe entre le service de prévision et le reste de l'organisation (voir section 2.1.3). Non seulement la prévision en sera enrichie, mais elle se traduira aussi en opérations concrètes. Le système de prévision doit s'adosser sur le système d'information de l'entreprise. Le système de prévision doit impérativement interagir avec le système d'information et les différentes fonctions transversales de l'entreprise (finance, marketing, production, ventes logistiques) (voir section 2.1.3). Nous verrons comment mettre en production l'outil informatique de prévision (voir section 4.7). Le système de prévision doit communiquer auprès des différents utilisateurs et ERP (Enterprise Resource Planning) de planification (voir section 2.3.4). Enfin, les utilisateurs doivent apprendre à utiliser ce nouvel outil d'aide à la décision pour bénéficier de l'ensemble des atouts proposés (voir section 4.8). Soulignons qu'un système de prévision vise à obtenir et analyser les données, puis à produire la prévision et à la communiquer (voir section 4.6.4).

### 2.1.3 Le système de prévision : un outil transversal

Après avoir introduit comment passer d'une simple prévision à un « système de prévision », il semble important de définir les besoins de chacun en matière de prévision. La figure 2.1 montre que tous les services d'une entreprise (finance, marketing, production, commande, ventes, logistique) peuvent être concernés par le besoin de prévision. C'est pourquoi Mentzer et Moon, 2005 (85), indiquent que le système de prévision est transversal et communique avec le système d'information. Pour Cluzel, 2006 (31), « *le système de prévision doit être un outil pour les unités opérationnelles, comme pour les structures de direction. La mise en place d'une application de prévision des ventes permet de fournir les renseignements nécessaires :*

- *à la chaîne production/distribution pour mettre le bon produit, au bon moment, au bon endroit et au moindre coût ;*
- *aux acheteurs pour s'approvisionner (. . .) en optimisant le couple coûts des stocks de matières premières/conditions consenties par les fournisseurs en fonction des volumes ;*
- *aux directions du marketing et de la recherche et développement pour s'orienter vers les produits dont la rentabilité est la plus grande au détriment des articles à faible marge (. . .) ;*
- *à la direction générale pour prendre des décisions tactiques et stratégiques impliquant l'ensemble de l'organisation ».*

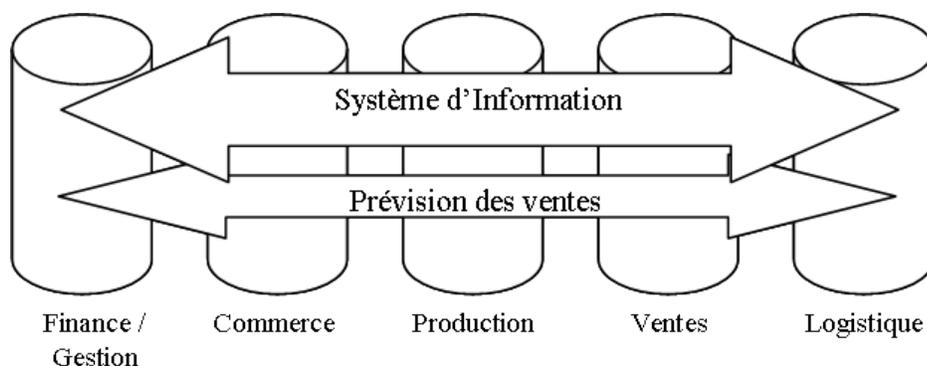


FIGURE 2.1 – *Interaction entre système d'information et prévision des ventes : Démarche transversale (Mentzer et Moon, 2005) (85)*

Suite aux prévisions, le service des finances construit son budget ; le service du commerce mise plus ou moins d'effort sur tel ou tel client ; le service de la production augmente ou réduit ses capacités de production ; le service des ventes cherche à dépasser les prévisions et le service de la logistique pilote les ressources humaines et matérielles au service des besoins prévus. Chaque service établit ainsi un plan d'affaires (plan d'approvisionnement, plan de fabrication, plan de vente, plan de distribution, plan RH, plan financier, plan de développement). Ils peuvent être à court terme, moyen ou long terme. Ils sont regroupés dans un plan général appelé S&OP (Sales & Operation Planning). A travers ces plans, le S&OP gère les ressources à partir des prévisions de la demande. C'est un moyen de prioriser les plans, les choix stratégiques et les risques à prendre. Charles Pomerleau <sup>1</sup> ajoute : « le S&OP a aussi pour vocation de suivre et contrôler la performance et la variance de l'entreprise ».

#### **2.1.4 Le système de prévision : un outil d'aide à la décision et de productivité**

Nous venons de présenter la genèse d'un système de prévision et le cadre dans lequel il s'implante. L'investissement dans un système de prévision doit être vu comme rentable dès qu'il permet un accroissement de la pertinence des décisions. Les décisions prises en entreprise ont pour finalité l'accroissement de la productivité. Ainsi, la prévision est aussi un outil de productivité. L'objet de cette section est d'étudier les décisions que nous pouvons prendre suite à la connaissance des prévisions de vente et les gains de

1. Consortium de recherche For@c, Août 2004

productivité que ces décisions engendrent. Les promoteurs d'outil de prévision<sup>2</sup> vantent un meilleur niveau de réponse au client. Ils ajoutent que la prévision des ventes permet de renforcer l'attention portée aux fournisseurs et aux partenaires commerciaux en leur donnant une meilleure visibilité des tendances de la demande des clients. En effet, la prévision des ventes peut permettre de négocier avec le fournisseur et le transporteur. Si nous connaissons à l'avance les marchandises dont nous aurons besoin chez le fournisseur ou à faire transporter sur l'année à venir, nous pouvons négocier des prix sur de gros volumes et favoriser les transports de lots.

Lors d'une interview pour le site internet Décideo.fr, M. Bruley<sup>3</sup> donne « sept bonnes raisons de mettre en place une solution de prévision de la demande.

- *Amélioration du service proposé aux clients : comprendre la demande pour améliorer les niveaux de service, notamment sur les produits les plus rentables,*
- *Optimisation de l'assortiment du stock : comprendre comment chaque référence contribue aux résultats afin d'identifier et de corriger aisément les problèmes liés au stock,*
- *Optimisation de la gestion des promotions : prévoir avec précision des promotions afin de sélectionner la meilleure d'entre elles pour obtenir le résultat voulu,*
- *Retour sur investissement rapide : les projets de prévision de la demande type se concrétisent par un retour sur investissement de plusieurs millions d'euros et sont rentabilisés en moins de 12 mois après la mise en œuvre,*
- *Anticipations précises : hausse des ventes et des rotations de stocks,*
- *Gestion des exceptions : des alertes proactives permettent de transformer rapidement un problème potentiel en opportunité,*
- *Impacts sur les démarques et les vols : réduction des démarques de 5% à 10% et baisse des vols de 20% ou plus ».*

Ne soyons pas dupes, les commerciaux vantent les mérites des outils de prévision pour mieux les vendre. Tous les problèmes ne vont pas se volatiliser. En règle générale, l'entreprise demande que l'investissement dans un bien ou service apporte un bénéfice, c'est le retour sur investissement ou ROI (Return Of Investment). Le ROI est la somme gagnée par rapport au capital initialement investi. Le montant investi est souvent facilement identifiable. En revanche, l'argent gagné l'est moins. Dans le cas d'un investissement pour un système de prévision, c'est la somme des économies liées à la réduction des opérations et marketing. Fildes et al., 2006 (51), fait remarquer que les décisions d'achat, marketing, fabrication, direction, budget et logistique sont toutes dépendantes de la prévision. En effet, en marketing, les décisions fondées sur les prévisions de la demande permettent de mieux planifier les opérations commerciales, les points de vente, les efforts à faire sur le

---

2. SAP Demand Planning, 2007

3. Directeur Marketing de Teradata France (2007)

packaging. La prévision de la demande peut également servir à établir les prix de vente en anticipant le volume et les coûts engendrés, et en déduire le bénéfice. De par sa prévision, la demande tire la production et ne la pousse plus. C'est-à-dire que les quantités à produire sont déduites des prévisions de vente et non plus des ventes effectives. De plus, certaines industries de pointe ont besoin, pour la fabrication de leur produit, de composants qui mettent du temps à être produits. Si les approvisionnements ne sont pas prévus correctement toute la chaîne de production peut en subir les conséquences. Ensuite, le budget est par définition un état prévisionnel des dépenses et des recettes de l'entreprise. La logistique utilise les prévisions pour gérer les stocks. Il faut constamment trouver un équilibre de la gestion des stocks dans un contexte de zéro rupture. Enfin, le graphique de la série chronologique des ventes analyse la santé du business. Ce même graphique peut montrer la différence entre une activité rentable et une qui ne l'est pas. La prévision permet alors des actions correctives et un service au client amélioré, indiquent Forslund et Jonsson, 2007 (53).

L'entreprise est constamment à la recherche de la performance. En se dotant d'un système de prévision, elle acquiert un outil supplémentaire de pilotage. Le système vise à augmenter la réactivité de l'entreprise pour un meilleur niveau de réponse. Il ambitionne d'augmenter le contrôle de la production par la mise en œuvre d'une planification maîtrisée. Marien, 2001 (80), soutiennent que la prévision de la demande et la planification des ressources le long de la chaîne logistique peut apporter de gros bénéfices. McCarthy, Tereza M. et Golcic, 2002 (83), ajoutent que la prévision des ventes a été reconnue comme contribuant à l'amélioration des performances. Enfin, Hubert Carpentier, directeur de l'organisation et des Systèmes d'Information chez Match, enchérit en déclarant « *qu'il ne fait pas de doute que les prévisions de vente très précises engendrent d'importants gains de productivité* »<sup>4</sup>. Par contre, Cluzel, 2006 (31), montre que pour être à la hauteur des attentes, le système de prévision doit « *donner à l'entreprise des informations utiles, à jour, fiables, cohérentes et précises, dans un monde en évolution permanente* ».

La société anonyme ModelEdition a interrogé 41 grosses et moyennes entreprises européennes pour connaître leur besoin en matière de prévision. Il ressort que les applications de business plan représentent 32% de l'ensemble des modèles de prévision réalisés par les entreprises. Les applications budgétaires représentent 20%, l'établissement des prix 18%, l'estimation des coûts 15% et les allocations de production 5%. Ce qui montre que, pour l'entreprise, l'objectif principal des prévisions de vente est la planification de son activité.

Nous venons de montrer que le système de prévision est un outil d'aide à la prise de décision. Ces décisions ont pour but l'accroissement de la productivité et/ou de la qualité des processus. Le système de prévision fournit un avantage concurrentiel. La prévision à court terme a pour objectif de faciliter les activités de production alors que la prévision à long terme a pour but de donner des pistes marketing ou financières expliquent Mentzer

---

4. Success Story : Le système de prévision des ventes chez Match, Aldata Solution 2008

et Schroeter, 1994 (86) et Seiersen, 2006 (105). Pour illustrer cette série de bénéfices, reprenons l'exemple de Deceuninck dans le (Journal de la logistique n°14, Avril 2004). A partir des prévisions établies à l'horizon d'un an, l'entreprise réalloue les capacités saisonnières deux fois par an afin d'éviter les goulets d'étranglement, tant sur les outillages que sur les machines. Toujours à partir des prévisions sur un horizon de 4 à 10 semaines, une fois par semaine, l'entreprise gère les nouveaux stocks et les commandes, les délais de livraison des commandes des clients, tout en assurant la planification optimale des ressources. Grâce à l'outil mis en place, Deceuninck observe une réduction des investissements et des coûts opérationnels, donc des économies.

## **2.2 Les questions préalables à la mise en place d'un système de prévision de la demande**

Etant donné le coût qu'engendre la mise en place d'un système de prévision, l'entreprise doit faire l'effort préalable de répondre à une série de questions. Ce sont 13 questions posées et argumentées dans les sections 2.2.1 à 2.2.13. Ces mêmes questions ont été posées durant le projet de recherche « Horizons ». Ainsi, à la suite de chaque argumentaire nous référons aux sections permettant d'obtenir la réponse à la question posée dans le cadre du projet « Horizons ».

### **2.2.1 Que doit permettre la prévision de la demande ?**

Moon et Mentzer, 2000 (91), distinguent trois préoccupations pivotant autour des prévisions :

- la prévision de la demande, définie comme une projection dans le futur de la demande, sachant un certain nombre de conditions environnementales ;
- le planning opérationnel, défini comme un ensemble de décisions et actions managériales prises pour atteindre ou surpasser les prévisions de la demande ;
- les objectifs de vente, définis comme les ventes à atteindre, pour satisfaire la demande et motiver les commerciaux et le marketing.

Nous avons longuement écrit, dans la section 2.1, sur les liaisons de dépendance qui existent entre prévision des ventes, planning opérationnel et objectif des ventes. Nous rappelons ici que la prévision des ventes est la condition « sine qua non » à toute planification et que, de par ce fait, elle est aussi un outil essentiel pour le pilotage de l'entreprise. Il n'en reste pas moins qu'il faut être plus précis pour répondre à la question « que

doit permettre la prévision des ventes ? ». En effet, en fonction de l'utilisation que nous en faisons, elle sera calculée par article ou par gamme d'article ; elle sera une prévision journalière, mensuelle ou annuelle ; l'erreur acceptable sera plus ou moins importante, et ainsi de suite.

La réponse à la question « que doit permettre la prévision d'activité chez TFE ? » est examinée dans la section [1.2.1](#).

Les trois questions suivantes sont reprises par Abraham et Ledolther, 1983 (3). Ils identifient cinq paramètres qui influencent le choix d'un modèle particulier de prévision : la disponibilité des données, le degré de fiabilité des données, l'horizon de prévision, le budget à accorder, la complexité attendue, Tan, 2007 (109).

## **2.2.2 Quelles sont les données sources de la prévision des ventes ?**

Pour établir des prévisions à partir de méthodes statistiques, il faut être en possession d'un historique des grandeurs à prévoir. Cela peut paraître évident, mais il y a encore quelques années les entreprises ne stockaient pas systématiquement l'information. D'autre part, même si elles le font aujourd'hui, les données ne sont pas toujours propres. Soit, il manque des informations, soit elles ne reflètent pas la réalité, soit la requête d'extraction est erronée. Quoi qu'il en soit, l'opération la plus importante dans le processus de prévision est la récupération des historiques et leur assainissement, avoue Robert Husset, Directeur Général d'Aldata Solution<sup>5</sup>.

Quelles sont les données à disposition de l'entreprise qui peuvent être utilisées dans le modèle mathématique de prévision ? Quel âge ont les données ? Quelles sont les données externes qui expliquent la grandeur à prévoir ? Quelle est la qualité des données ? Quelle est la fiabilité ? Quel est le niveau de détail ? Quelle est la fréquence ?

L'âge des données, ou encore l'historique des données va déterminer les modèles mathématiques à utiliser. Si seulement un an d'historique est à portée, alors il est conseillé d'utiliser les « simplistic fixed model » (FMTS). Les OMTS (voir section [2.3.5](#)) ont besoin d'au moins quatre ans d'historique.

Les ventes sont souvent influencées par des phénomènes externes (promotion, météo, mode, confiance des ménages, . . .). Si l'entreprise est capable de maîtriser ces variables, de les prévoir et même de les provoquer, ou influencer, elle a intérêt à utiliser la régression comme outil mathématique de prévision.

Les données ne sont pas toujours aussi fiables qu'on le voudrait. Des erreurs de manipulation, saisie, transformation, mise à jour, insertion, etc, sont toujours possibles. Si

---

5. Success Story : Le système de prévision des ventes chez Match, Aldata Solution 2008

les mises à jour ne sont pas faites à temps, cela a nécessairement un impact sur les prévisions.

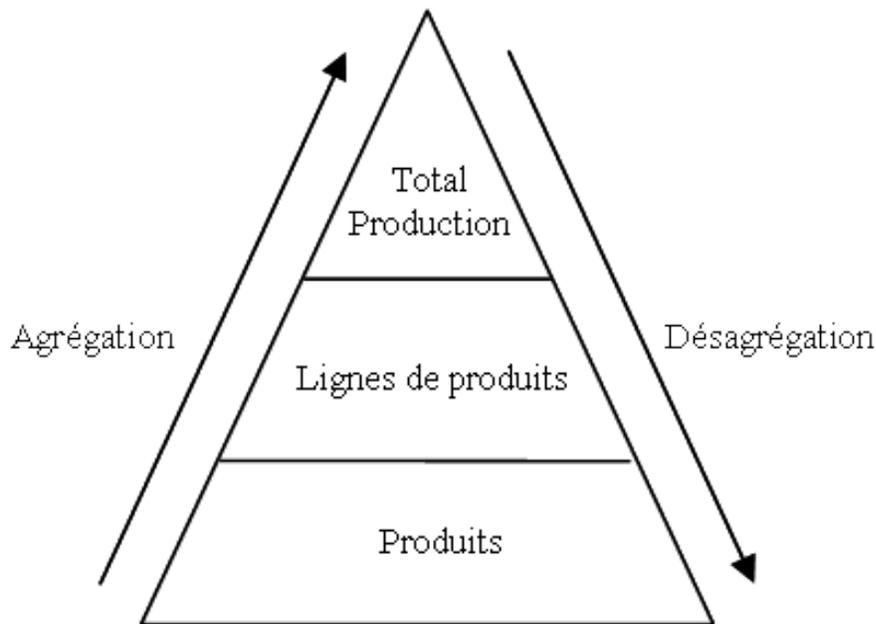


FIGURE 2.2 – *Pyramide produit*

A l'image de la figure 2.2, la hiérarchie des prévisions est représentée par une pyramide, car le bas nécessite beaucoup plus de prévisions que le haut. Quand il s'agit de prévoir la vente de 30.000 articles différents dans un supermarché, la prévision devient laborieuse (bas de la pyramide). Certains articles sont nouveaux et ne possèdent pas d'historique de vente. D'autres ont des ventes tellement volatiles qu'il semble impossible de les prévoir. Il est alors important d'être capable de regrouper ces articles en familles de produits, voire en segments de marchés (milieu de la pyramide). L'idée est qu'il sera toujours plus juste de prévoir leur regroupement, car la baisse des ventes des uns est amortie par la vente des autres. Ainsi, la série chronologique des familles de produits est moins volatile. Suite à leur débat sur les techniques de prévision à utiliser pour les larges gammes d'articles, Narasimhan, McLeavy et Billington, 1995 (96), conseillent d'agréger les articles en familles de produits, avant de les prévoir. La prévision obtenue peut ensuite être désagrégée au niveau des produits en utilisant des coefficients d'affectation (top-down, en anglais), Widiarta, Viswanathan et Piplani, 2009 (119).

Une autre méthode est l'agrégation du bas vers le haut (bottom-up, en anglais). Elle consiste à sommer les prévisions des différents produits pour parvenir à la ligne de produits. La somme des prévisions par ligne de produits fournit la production totale. Cette méthode est lourde à maintenir, car elle implique un modèle de prévision pour chacun des produits. D'autre part, prévoir l'échelon du haut comme une somme des prévisions

des échelons précédents a l'inconvénient d'additionner les erreurs de prévision. Or la somme des erreurs est supérieure à l'erreur d'une somme.

Une troisième solution préconisée est de calculer des prévisions à chaque niveau et de les corriger par un algorithme d'optimisation de sorte qu'elles soient cohérentes entre elles.

La méthode la moins coûteuse semble de calculer une prévision au niveau d'agrégation qui permet de répartir la prévision de manière homogène dans les niveaux du dessus (bottom-up) comme dans ceux du dessous (top-down). Il faut alors distinguer le calcul de la prévision au niveau d'agrégation adéquat et la restitution de cette prévision aux niveaux restants.

Comme pour la pyramide produit, il existe une pyramide temps. Les données sont associées à une unité de temps. Certaines sont disponibles à l'heure, d'autre au jour, au mois, au trimestre, à l'année. Les prévisions suivent la même unité de temps que la donnée. Si nous cherchons à faire des prévisions de vente quotidiennes, il faut que l'historique des ventes soit connu au jour. Le niveau des données, annuel, mensuel ou journalier, est à mettre en relation avec la planification que nous cherchons à établir. C'est ce qui est appelé en anglais « forecasting hierarchy ».

Les données utilisées pour la prévision d'activité, dans le cadre du projet de recherche « Horizons » sont étudiées dans la section 4.4.1. Leur fiabilité est testée dans la section 4.4.2. Le système de prévision « Horizons » prend, aussi, en compte la notion de hiérarchie des prévisions dans la section 4.4.12.

### 2.2.3 Quels sont les paramètres de prévision des ventes ?

Les concepts de base des prévisions de vente sont :

- composante,
- saisonnalité,
- horizon de prévision,
- fréquence de rafraîchissement.

La prévision étant en amont de la planification, les concepts de base de la prévision reposent essentiellement sur le planning.

La **composante** est de savoir quelle donnée doit être prévue. Elle se décide suivant ce que l'entreprise cherche à en déduire (voir section 2.2.1). Pour planifier ses ressources, la production a besoin de prévision en unité de production. Le marketing a besoin de prévision de la demande pour ajuster les prix. La direction souhaite des prévisions de

vente pour finaliser son budget.

Le degré de **saisonnalité** du produit affecte le choix des méthodes mathématiques à utiliser (voir section 2.3.3).

L'**horizon de prévision** est en relation avec la durée du plan. L'horizon de prévision est défini par la taille du planning que nous voulons en déduire (en unités temporelles). Nous comptons trois horizons parmi les séries temporelles

- prévisions à court terme (< 3 mois). C'est un horizon de prévision opérationnel. Les prévisions sont utilisées pour prendre des décisions continues sur le planning, l'inventaire, le management, l'activité logistique,
- prévisions à moyen terme (3 mois à 2 ans). C'est un horizon tactique. Les prévisions sont utilisées pour le budget, le contrôle de coût, le marketing, la mise sur le marché de nouveaux produits, le plan de restructuration des forces de vente, le plan de distribution, etc.
- prévision à long terme (> 2 ans). C'est un horizon stratégique. Les prévisions sont utilisées pour décider de rentrer dans un nouveau marché, développer de nouveaux produits ou services, établir un contact à long terme.

Les temps donnés (3 mois ou 2 ans) sont à titre indicatif, car la représentation du court, moyen, long terme dépend des projets et de la conception qu'en a l'entreprise. Des prévisions de capacité à long terme, par exemple, seront considérées comme très importantes pour des entreprises qui font de la production sur commande (ou make-to-order, en anglais) alors que des prévisions à court terme seront utiles pour des entreprises qui produisent pour alimenter le stock (make-to-stock), expliquent Forslund et Jonsson, 2007 (53).

Il n'est pas nécessaire de faire des prévisions à très court terme si l'entreprise n'a pas le temps de réagir à celles-ci. Prenons l'exemple d'une **agence** de transport à laquelle le prévisionniste annonce une hausse d'activité dans 2 jours, de l'ordre de 5%. Le planning convertit cette hausse en 24 heures productives, soit 3 personnes à plein-temps. La question est : « où trouver ces 3 personnes de plus en moins de 48H ? ». La réactivité de l'entreprise ne sera pas au rendez-vous. C'est pourquoi, l'horizon de prévision dépend aussi de la réactivité de l'entreprise face aux annonces.

Si l'entreprise décide d'utiliser un modèle de prévision différent pour chaque niveau temporel de la pyramide des temps (voir section 2.2.2), l'**intervalle de temps** ou la maille de temps entre deux périodes dépend de l'unité temps des plannings. Si les plannings se font par jour de semaine, les données prévisionnelles sont au jour. Si l'entreprise décide d'utiliser qu'une seule prévision pour la restituer sur différentes échelles de temps, il lui faut trouver le niveau d'agrégation temporel lui permettant de répartir la prévision de façon homogène sur les mailles de temps supérieures et inférieures à celle choisie pour la

prévision (voir section 2.2.2).

Plus élevée est la **fréquence de rafraîchissement**, meilleures seront les prévisions. Si le planning est mis à jour une fois par mois, les prévisions devront l'être au moins une fois par mois. L'important est qu'elles le soient avant la révision du planning.

Dans le cadre de la mise place d'un système de prévision chez TFE, les paramètres évoqués ci-dessus sont discutés dans la section 1.2.4.

## 2.2.4 Quelles sont les techniques à choisir pour prévoir les ventes ?

Mentzer et Gomez, 1994 (84), montrent qu'il n'existe pas de technique réellement supérieure à l'autre. La raison vient de la grande variété des séries chronologiques et des horizons de prévision. S'il vient à l'esprit par exemple de comptabiliser le nombre de voitures passant sur le pont de Kerino à Vannes, entre 12h et 13h, jour après jour, en 2009, le résultat est une série de valeurs exprimées en nombre de véhicules, chacune d'entre elles correspondant à un jour de cette année. Une machine, cependant, pourrait réaliser un enregistrement du nombre de véhicules en continu et nous pourrions ainsi disposer d'informations (infiniment) plus nombreuses. Les prévisions de ces deux séries chronologiques pourraient être estimées à divers horizons. Cela pour souligner que les séries chronologiques entourent notre quotidien et qu'elles « occupent une place importante dans tous les domaines de l'observation ou de la collection de données »(48). Elles contiennent toutes des spécificités propres et, par conséquent, il semble difficile de retenir une seule technique comme la meilleure qui soit.

D'autre part, les techniques les plus complexes ne sont pas toujours les plus fructueuses. Beaucoup de techniques utilisent la même approche et ont, par conséquent, la même efficacité. Par exemple, la méthode de lissage exponentiel double, la méthode de Holt Winters et un modèle ARIMA peuvent être identiques sous certaines conditions (21). Ce qui nous amène à constater que beaucoup de techniques sont similaires alors qu'elles portent des noms différents. Enfin, Mentzer et Gomez, 1994, préconisent d'en choisir une ou deux avec lesquelles le prévisionniste est familier et qui se prêtent bien aux séries chronologiques étudiées.

La section 2.3.3 classe des techniques de prévision. A savoir que les méthodes statistiques sont toujours plus performantes que les méthodes simplistes, pour la bonne raison qu'elles généralisent ces méthodes. En revanche, il est plus difficile de comparer les méthodes statistiques aux méthodes empiriques, car il faudrait étudier chaque cas individuellement Marien, 2001 (80). Schnaars, 2006 (103), Armstrong et Collopy, 1998 (9), Armstrong et Roderick, 1999 (8), Mentzer et Schroeter, 1994 (86), Mentzer et

Moon, 2005 (85) pensent que la combinaison de prévision engendre une amélioration des résultats. Ainsi, il est bien de choisir plusieurs méthodes. Schnaars examine quelles méthodes sont les plus efficaces pour des prévisions annuelles pour divers produits à la consommation. Il conclut à une combinaison de méthodes alternatives. Mentzer et Moon soutiennent qu'il est important de reconnaître le rôle que peut apporter une réflexion empirique pour ajuster des prévisions obtenues par un processus mécanique. Seul le choix de la technique mathématique de prévision utilisée, fait appel à une réflexion de la part du prévisionniste. Une réflexion est aussi menée sur le choix des données, le choix des indicateurs de performance. Mais plus concrètement, il est recommandé d'utiliser l'expérience, la connaissance du métier, le savoir-faire, pour aboutir à des prévisions formelles. Un premier avantage est la capacité à prévoir un changement brutal (un palier) dans le comportement de la série chronologique. Ceci n'est pas possible avec les modèles mathématiques qui ne se basent que sur l'historique. Ou alors avec l'analyse de régression, si une des variables explicatives représente ce changement brusque. Le second avantage est celui d'utiliser la masse d'informations représentées par le savoir-faire d'un expert, son intuition, son jugement pour valider les prévisions.

Comment, parmi toutes ces méthodes, choisir la plus appropriée ? Armstrong, 2001 (7), tente d'y répondre et recommande de choisir la méthode de prévision en fonction d'un nombre de questions auxquelles il est aisé de répondre. Elles sont représentées par la figure 2.3.

Selon cette décomposition du processus de choix, la première question qui se pose au prévisionniste est de savoir s'il existe un historique. Si ce n'est pas le cas, il faudra avoir recours à un processus d'estimation empirique.

« *Pour ces procédures d'estimation empirique, la question suivante est de savoir si la situation implique une interaction entre les décideurs, et si des changements importants sont envisagés* ». Ces changements concernent aussi bien l'environnement interne qu'externe de l'entreprise. « *Pour d'importants changements, est-il question de faire une analyse de la politique actuelle, et si c'est le cas, quelle est la meilleure source pour cerner la situation ?* » Allons nous consulter des experts externes, faire des études d'opinions, lancer une enquête sur les intentions d'achat, demander une étude de marché ? S'il existe un grand nombre de données, s'agit-il de séries temporelles ? La question suivante est de savoir si nous avons des informations sur les relations empiriques que nous pouvons nous attendre à rencontrer comme, par exemple, l'élasticité des prix, Tellis, 1988 (110). Lorsque le prix augmente, la demande diminue. Si une telle connaissance empirique des relations est disponible, il faut se servir de modèles économétriques. De plus, tant que possible, il est bon d'utiliser le savoir-faire de l'entreprise pour maîtriser les ventes et ainsi prévoir au plus juste.

Quand les séries temporelles sont à disposition, mais qu'il manque des données causales, il est préférable d'extrapoler. Si nous ne disposons d'aucune connaissance préalable des relations, mais d'une connaissance du domaine, (par exemple : un commercial

annonce l'arrivée d'un nouveau portefeuille client, ce qui va avoir pour effet une augmentation des ventes), il faut faire appel aux prévisions basées sur les règles.

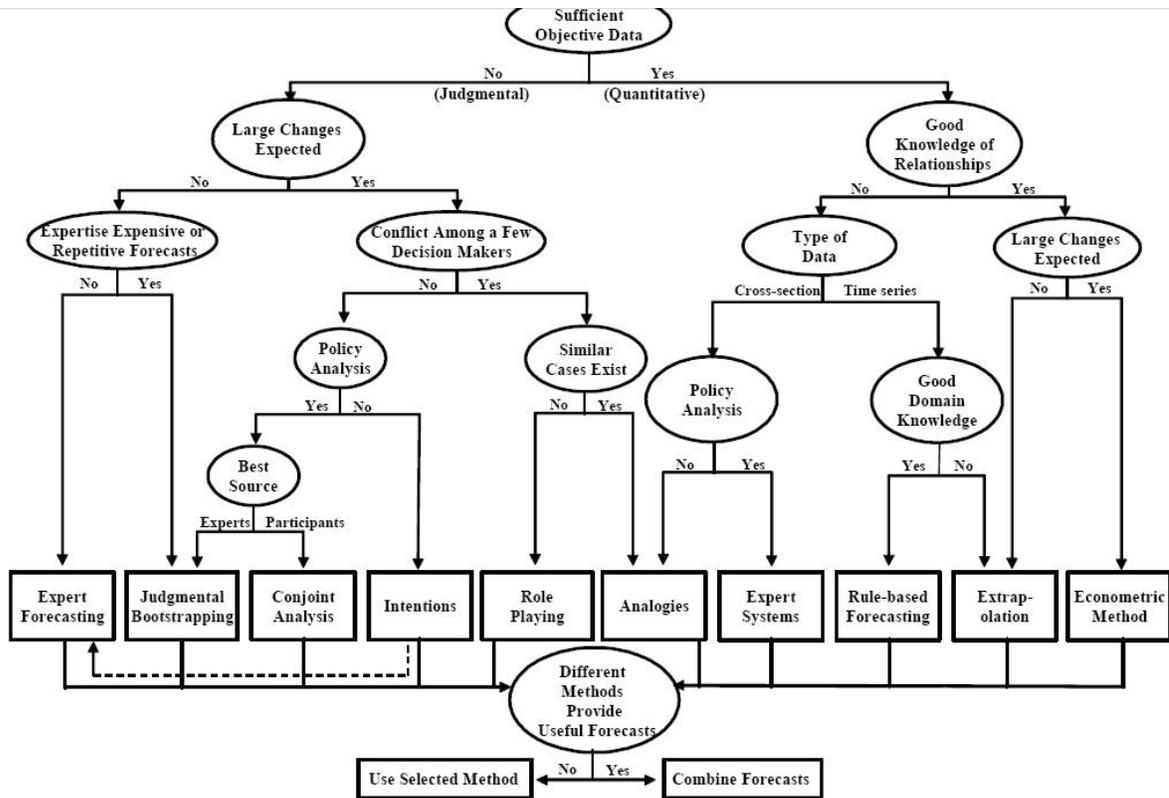


FIGURE 2.3 – Arbre de sélection pour les méthodes de prévision Armstrong, 1999 (8)

Ajoutons que ce schéma n'est pas complet, en ce sens où l'horizon de prévision est un critère important dans la sélection du choix du modèle. Il permet de distinguer les méthodes de prévision à long terme, de celles à moyen et court terme.

De manière générale, quand l'historique est conséquent et qu'un grand nombre de prévision est demandé, la méthode la plus répandue est l'extrapolation des séries chronologiques. Ces méthodes sont bon marché alors que les méthodes économétriques sont plus coûteuses. Le coût provient essentiellement de l'achat des données externes à l'entreprise et du coût humain, pour l'entretien et la surveillance de ces données.

Le choix des techniques à utiliser pour prévoir l'activité des agences de transport TFE est débattu dans la section 4.3.

## 2.2.5 Pourquoi prendre en compte l'expertise terrain pour valider la prévision des ventes ?

Mentzer et Moon, 2005 (85), Keen et Scott Morton, 1978 (70), mettent en évidence que la combinaison du verdict des managers et la prévision statistique fournissent de meilleurs résultats que les deux indépendamment. Fildes, Goodwin et Lawrence, 2006 (51), confirment et précisent que le modèle prend en compte des événements réguliers globaux alors que le manager prend en compte des événements plus fins tels qu'une promotion. Ces facteurs sont souvent analysés comme du bruit dans les modèles mathématiques. Ils ajoutent qu'il est évident que ce potentiel issu de la combinaison du manager et du moteur de prévision est sous exploité par les systèmes de prévision. D'une part, les managers ont tendance à ne faire confiance qu'à leur propre jugement où les données quantitatives seraient plus judicieuses. D'autre part, quand le jugement du manager est approprié, les systèmes de prévision du commerce n'offrent pas la possibilité de prendre en compte ce jugement. Rappelons l'hypothèse de base sur les prévisions : la trame ou le comportement passé va continuer de la même manière dans le futur. Or si le comportement passé cesse, les statistiques vont faillir et il n'y a plus que le discernement humain qui sera à prédire, écrit Markridakis, et al., 1998. C'est pourquoi il est conseillé de ne pas utiliser uniquement les prévisions statistiques. D'un autre côté, l'humain fait également des erreurs et il ne faut pas non plus s'appuyer uniquement sur son verdict. Goodwin, 2002, évoque la combinaison des deux prévisions (statistique et empirique) en remarquant que la pensée humaine s'adapte et prend en compte un événement, mais son inconscient ne peut prendre en compte qu'un nombre limité de données et souffre d'erreur cognitive. Il va donc donner plus de poids à l'évènement considéré au détriment des événements mis de côté. A l'inverse, les méthodes statistiques seront rigides, mais peuvent prendre en compte un grand nombre de données et vont en maximiser l'utilité. Nicholas Seiersen, 2006 (105), ajoute : « *quand bien même la technologie est devenue puissante dans le soutien aux prévisions et à la planification, les entreprises performantes constatent qu'il est impératif de mettre en œuvre un forum où la validité des prévisions et des plans soit vérifiée par les principaux responsables marketing, ventes, distribution, fabrication, finances. A l'issue de cette réunion S&OP (sales and operation planning) le plan d'approvisionnement des marchés est accepté et mis en œuvre* ». Forslund et Jonsson, 2007 (53), concluent, les prévisions doivent être contrôlées par l'homme avant d'être injectées dans un système de planification automatique.

Le système de prévision « Horizons » propose de prendre en compte l'expertise terrain en permettant à certain utilisateur de corriger les prévisions proposées. Le système leur demande dans ce cas d'expliquer le motif de la correction apportée (voir section 4.7.3 et 4.8).

## 2.2.6 Quel budget est à accorder à la prévision des ventes ?

Peu d'entreprises fixent un budget de prévision adéquat ou réaliste. L'erreur souvent commise est de considérer que le système de prévision existe dans l'entreprise dès que le logiciel a été facturé. Il faut être conscient que l'introduction d'un nouvel outil informatique dans un service quelconque fait souvent face à une résistance au changement. D'autre part, l'image qu'ont les managers de la prévision n'est pas toujours fameuse. Ils ont entendu dire quelque part que la prévision est toujours fausse. C'est pourquoi il faut accompagner la mise en production de l'outil par une formation. Elle doit persuader l'utilisateur de l'intérêt qu'il y a à utiliser l'outil de prévision. Suite à la formation, il s'écoule un temps assez long pour que tout fonctionne et que l'utilisateur s'approprie l'outil de prévision.

Le budget accordé au projet de recherche « Horizons » est révélé dans la section 1.3.2.

## 2.2.7 Quelle est la taille du portefeuille client ?

Moins grand est le portefeuille client, plus l'entreprise peut espérer des prévisions pour chaque client. L'entreprise a également plus de facilités à discuter avec ses clients. Alors qu'une entreprise qui a 65 millions de foyers comme clients ne peut pas s'imaginer des prévisions par client. Par contre, elle peut les classer sur le modèle de la pyramide présentée en section 2.2.2. Suite à cela, la loi des grands nombres lui laisse espérer de bonnes prévisions.

Dans le cas de la prévision d'activité chez TFE, les segments de sortie de quai ont été agrégés en deux classes : les expéditions et les distributions (voir 1.2.4).

## 2.2.8 Quel est le nombre de nouveaux produits ?

La vente des nouveaux produits est difficile à prévoir, car par définition, il n'existe pas d'historique des ventes. Si le nouveau produit ressemble à un connu, il sera toujours possible de reprendre l'historique des ventes du produit connu. Dans le cas contraire, il sera utile de faire des études de marché ou de passer par des prévisions d'expert (voir section 2.2.5).

## 2.2.9 Existe-t-il des différences régionales ?

Si les ventes sont très différentes selon les régions, cela complique les prévisions de vente. Il faut alors passer par les historiques de vente par région et multiplier les facteurs exogènes par le nombre de régions. Un même produit dont les ventes se comportent différemment selon deux régions équivaut à prévoir deux produits différents.

C'est le cas des **agences** de transport TFE. La section 1.1.2.2.1 décrit l'éclatement du réseau de transport et la section 4.3 souligne la technique retenue pour contourner le problème.

## 2.2.10 Quelles sont les compétences du prévisionniste ?

Dans son numéro de Juin 2008, le Supply Chain Magazine<sup>6</sup> présente le prévisionniste comme un chasseur d'information. Son objectif est de « *réaliser des prévisions de ventes les plus justes possible. Et surtout, il est en mesure d'analyser les résultats effectifs par rapport aux prévisions et de proposer les solutions et moyens propres à réduire les écarts. Il relie l'analyse statistique avec la réalité de l'activité de l'entreprise, dont il est un fin connaisseur* ». Ainsi, le prévisionniste est polyvalent. Non seulement il doit avoir des notions de statistique et d'économétrie, mais il doit également « acquérir une bonne connaissance de l'ensemble du fonctionnement de l'entreprise ». Comme son outil (voir section 2.1.3), le prévisionniste occupe un poste transversal. Pour aboutir à une prévision fiable et pérenne, le prévisionniste doit être à l'écoute des managers, il doit savoir utiliser les outils de prévision théorique et informatique, et il doit rester humble pour accepter de manière récurrente l'erreur de prévision, Fildes et al., 2006 (51).

Fildes et al., 2006 (51), signalent qu'il est dangereux de laisser le manager faire ses propres prévisions. D'une part, il ne possède pas les notions de bases en statistiques, d'autre part sa prévision sera automatiquement influencée de la vision qu'il a du futur. Le problème étant qu'il publiera les prévisions comme provenant d'un modèle scientifique alors que ce ne sera pas réellement le cas. Enfin quand un utilisateur a la possibilité de créer son modèle de prévision avec le système de prévision, ses choix statistiques sont souvent très pauvres. Il entre les paramètres par défaut et appuie sur le bouton « best forecast » pour que la machine cherche un résultat par le biais d'une boîte noire. La justesse des prévisions est alors souvent à remettre en cause.

Le système de prévision « Horizons » a été conçu de sorte que la donnée prévision-

---

6. Supply Chain Magazine, n°25, Juin 2008

nelle soit comprise et utilisée par des non statisticiens. L'initiation à sa lecture est donnée dans la section 4.8.

### 2.2.11 Quelle est l'erreur de prévision acceptable ?

D'un coté, les experts affirment qu'il vaut mieux des prévisions fausses que pas de prévisions du tout, mais, d'un autre coté, soulignons que les prévisions médiocres peuvent avoir une influence négative sur l'activité de l'entreprise. Seiersen, 2006 (105), mentionne que « *les erreurs dans les prévisions de vente se détectent au fil des ventes réelles et mettent en cause les enchaînements et les quantités à fabriquer, à entreposer, à transporter, à livrer, à installer* ». Bien entendu, nous ne pouvons pas éliminer l'incertitude. Le challenge du prévisionniste est de réduire cette incertitude. Lorsqu'il fait des prévisions, il se trompe plus ou moins. Si les prévisions sont trop optimistes, il y a trop de dépenses pour des ventes qui ne seront pas réalisées, le profit est réduit. D'autre part, des prévisions trop pessimistes, engendrent des performances médiocres et font que le client est insatisfait de la prestation rendue, car il y a un manque de ressources allouées. D'un autre côté, une prévision trop pessimiste peut avoir comme effet de faire réagir l'entreprise. Elle met en œuvre des actions volontaristes pour que la prévision pessimiste ne se réalise pas. Qu'est-ce qu'une bonne prévision ?

Nombreux sont ceux qui pensent que plus les prévisions sont justes et mieux c'est. Or le bénéfice gagné par de meilleures prévisions ne vaut souvent pas les coûts qui y ont été injectés. Chercher une erreur toujours plus faible coûte cher. Pourquoi utiliser du temps et de l'argent à chercher des prévisions plus justes si l'erreur de prévision n'est pas pénalisante ? Autrement dit, si le risque de perdre un client est faible et que le coût de stockage est faible pourquoi augmenter la justesse des prévisions de vente ? Bourbonnais, 2007 (18), pense que « *c'est à l'entreprise de mettre en place l'organisation qui convient pour remédier à ses erreurs, en fonction de ses objectifs* ». Un moyen, par exemple, d'augmenter la justesse des prévisions des ventes, est de connaître les commandes des clients, par avance. Pour les pousser à anticiper leurs commandes, l'entrepreneur peut leur proposer un rabais de 3% si la commande dépasse trois tonnes et qu'elle est émise 3 semaines à l'avance. Les erreurs des prévisions passées peuvent également permettre de corriger les prévisions à venir, mais personne n'est capable de rendre des prévisions fiables à 100%.

Selon Fitoussi, 1997 (52), « *les erreurs de prévision peuvent avoir au moins 3 origines : une insuffisance d'information sur le présent, une mauvaise spécification théorique, la non-réalisation de certaines hypothèses. Une incertitude irréductible au sens où certains évènements sont imprévisibles, alors même que leurs conséquences sur l'activité sont déterminantes* ».

Une analyse de retour sur investissement devrait être faite pour déterminer les réels besoins sur le pourcentage d'erreur acceptable, l'horizon et l'intervalle. Néanmoins, le but du prévisionniste reste de se rapprocher au mieux de la réalité en cherchant à réduire l'erreur de prévision au plus juste, à un niveau acceptable pour le management. L'importance de l'exactitude des prévisions dépend de la manière dont elles sont utilisées dans la planification et le contrôle des processus. Il est souvent difficile, pour l'entreprise, de juger de l'erreur acceptable. Elle dépend de l'utilité que l'entreprise fait des prévisions. Et pourtant l'entreprise doit fixer un niveau d'erreurs admissibles pour ne pas chercher inutilement la performance théorique. D'autre part, comme le mentionne Cluzel, 2006 (31), « *les prévisions permettent de mesurer les écarts, puis de les analyser afin de prendre des mesures correctives et ainsi s'engager dans une démarche de progrès continu* ». Ainsi, une entreprise doit améliorer ses prévisions avec l'expérience qu'elle accumule. Mais l'erreur de prévision n'est pas le seul indicateur qui permette au manager d'accorder sa confiance à la donnée prévisionnelle. Il surveille que la prévision n'arrive pas en retard par rapport à son utilité. Par exemple, si la prévision apparaît après que le planning a été décidé, elle ne sert plus à rien. La donnée prévisionnelle doit aussi être stable. Un manager, qui voit sa prévision du mois prochain changer tous les jours, va manquer de confiance. Le format de la prévision doit aussi convenir à celui qui l'utilise. Si elle arrive par fax et nécessite trop d'opérations avant de pouvoir utiliser l'information, elle sera mise de côté.

L'erreur de prévision acceptable du projet « Horizons » est mentionnée dans les sections 1.2.4 et 4.5.

### **2.2.12 Quel est le format de restitution de la prévision des ventes ?**

L'environnement de restitution doit être convivial, simple et clair. Tous les services doivent y accéder car les prévisions de vente sont aussi là pour motiver le personnel. Ainsi, les prévisions de vente pourront même être dépassées. L'environnement de restitution des prévisions doit rappeler à l'utilisateur ce qu'il a l'habitude de voir. Il ne doit pas sembler compliqué et du coup repousser l'utilisateur. C'est pourquoi les données qui accompagnent la prévision doivent s'afficher selon le niveau de compréhension de chacun. Un commercial ne regarde pas les mêmes données qu'un planificateur.

Suivant l'utilisateur, la prévision peut être donnée sous une forme différente. L'horizon diffère selon que nous nous adressons à un patron d'agence ou à un chef d'équipe. Le premier a une vision à long terme. Il a des objectifs à tenir, un budget à gérer. Son horizon est au-delà des 6 mois. Alors que le chef d'équipe doit faire face aux fluctuations quotidiennes. Il veut des prévisions journalières pour planifier son équipe.

La prévision peut être accompagnée d'indicateurs qui permettent de faciliter sa lecture. Elle peut être envoyée par mail, par EDI, fichier plat ou être consultable grâce aux outils de reporting proposés par la plateforme BI.

L'utilisateur peut également être un ERP de planification. Dans ce cas, il faut définir le format informatique de restitution de la donnée prévisionnelle. Elle doit être stockée à un endroit précis du système d'information. Elle doit être mise à jour selon les besoins du module de planification.

Le système de prévision, mis en place chez TFE, choisit le mode web (voir section 4.6.3) et une interface commune à tous (voir section 4.7.3) pour diffuser l'information. La confidentialité est préservée par un système d'identification et une limitation des droits selon l'utilisateur.

### **2.2.13 Quelle procédure suivre pour aboutir à une prévision des ventes juste ?**

Nous avons plusieurs fois mentionné qu'un système de prévision n'était pas seulement un outil informatique ou seulement un modèle mathématique. C'est aussi toute une organisation et un ensemble de procédures à mettre en œuvre au sein de l'entreprise. « *La qualité d'un système de prévision est d'abord une question de management* »<sup>7</sup>. Il faut définir une procédure par laquelle seront développées les prévisions. Mentzer et Moon, 2005 (85), rappellent que trop souvent, l'entreprise tente de développer un système de prévision sans en comprendre le management. Souvent dans les entreprises, peu de personnes maîtrisent l'ensemble des procédures de prévision, c'est pourquoi il est difficile de corriger les prévisions. Beaucoup connaissent un maillon, mais pas la chaîne entière.

Prenons l'exemple d'une entreprise qui, pour le même produit, a des commerciaux dans différents pays, des clients de natures différentes et qui utilisent le produit à des fins différentes. L'entreprise veut un système de prévision qui permette de calculer une prévision quantitative, avec des ajustements en fonction de la localité géographique, du client et de la force de vente. Elle veut aussi des ajustements par gamme de produits et enfin un planning.

Pour la mise en place d'un système de prévision, conseillons à cette entreprise la procédure suivante : commencer par un calcul informatique des prévisions. Les prévisions de vente sont éclatées par gamme de produits, localité géographique et client, pour être envoyées aux commerciaux. Chaque force de vente a en possession un grand nombre d'indicateurs marché avec lesquels elle peut faire un ajustement qualitatif des prévisions.

---

7. Les Echos n°19043, Décembre 2003

Les commerciaux renvoient les prévisions en central. Les corrections venant des commerciaux sont regroupées pour être ré-agrégées, les managers marketing reçoivent les prévisions agrégées pour la ligne de produits qu'ils ont en charge. Ils corrigent, s'il y a lieu, et renvoient au groupe de prévision. Une nouvelle boucle est formée. Les prévisions sont envoyées au niveau plus haut pour validation et entrent dans le processus planning. Notez que cette procédure a besoin d'une transmission de données informatiques, l'agrégation et désagrégation des données, et de l'horizon de prévision et la comparaison des prévisions avec le réel à chaque instant.

Dans de nombreuses compagnies, une automatisation des procédures de prévision est mise en place pour que les prévisions soient mises à jour à intervalle régulier (hebdomadaire, quotidien). Cette automatisation raccourcit le temps donné aux managers pour étudier et valider les prévisions. C'est pourquoi il est primordial que les prévisions soient accompagnées d'un ensemble d'indicateurs ayant pour but de retourner un « feed-back » sur les erreurs de prévision. Si les utilisateurs comprennent les erreurs de prévision, ils sont à même de les corriger. Ce qui permet aussi au prévisionniste de conclure que l'erreur n'est pas juste du bruit mais qu'elle contient des causes déterministes.

D. Bourgoïn, Directrice Générale d'Azap, souligne que : « *souvent, les projets sont plus globaux que les prévisions, ils incluent les approvisionnements et la planification* »<sup>8</sup>. Moon et Mentzer, 2000 (91), complètent que les entreprises souhaitant implanter un système de prévision et planification doivent se préparer à être confrontées à une nouvelle façon de faire du business. Ce ne sera pas toujours évident de faire accepter une nouvelle méthode de travail ou d'organisation. « *Les managers opérationnels doivent acquérir le réflexe de faire remonter les informations vers le prévisionniste. Les logisticiens doivent abandonner un peu de leur autonomie* »<sup>9</sup>. Des résistances vont se créer. C'est pourquoi il est recommandé de ne pas tout installer en un seul coup, mais plutôt d'attendre que les utilisateurs acceptent et s'approprient les outils au fur et à mesure. Il est plus important de gagner de petites batailles, et de démontrer l'efficacité des prévisions que d'installer une usine à gaz.

### 2.2.13.1 Retour d'expérience

Mentionnons les résultats d'un audit chez Lucent Technologie fourni par Moon et Mentzer, 2000 (91), et un exemple de démarche chez Rayovac apporté par Marien, 2001 (80).

L'audit sur le CDP (Customer Demand Planning) chez Lucent Technologie a recommandé :

---

8. Supply Chain Magazine, n°17, Septembre 2007

9. Les Echos n°19043, Décembre 2003

- une formation plus approfondie sur la manière de faire les prévisions et sur la manière de les utiliser ;
- récompenser les meilleures prévisions, ce qui permettra aussi de faire en sorte que les utilisateurs fassent attention aux prévisions ;
- préconiser le suivi des prévisions et récompenser les personnes qui ajoutent de l'information aux prévisions en leur montrant ce que cette information a apporté de positif ;
- mettre en place un service de prévision qui propose des prévisions (initiales) par des méthodes statistiques. Cette prévision de base pourra être ensuite diffusée au sein des différents services qui les ajusteront ;
- recommander que le service de prévision soit implanté au plus proche du service marketing, car ce dernier a plus de visibilité sur l'avenir.

La démarche de prévision chez Rayovac est un processus itératif. En début d'année, un plan opérationnel annuel, basé sur des mesures de résultats, est produit. Il est utilisé comme baromètre pour suivre l'évolution de la prévision. Périodiquement, une prévision basée sur l'historique des ventes est calculée par des modèles mathématiques. Une réunion regroupant les ventes, le marketing, la finance, les fournisseurs, les distributeurs, le transport, est organisée. Pendant cette réunion, les différents partis comparent leurs prévisions issues du modèle mathématique et les erreurs de prévision passées, pour s'entendre sur les nouvelles prévisions. Si ces nouvelles prévisions diffèrent trop du plan d'action annuel, soit ils corrigent leur copie, soit ils réorientent le plan d'action. Les prévisions définies et approuvées, sont injectées dans les ERP de planification.

Un autre retour d'expérience est présenté en annexes. Il concerne le site logistique « Bain de Bretagne » (STEF-TFE). C'est un résumé du rapport d'audit fourni par la société Eurodécision dans le cadre d'un projet d'automatisation des plannings des équipes opérationnelles.

## **2.3 Les composants du système de prévision**

Comment assurer une grande qualité des prévisions afin d'induire des décisions sereines et augmenter les performances de l'entreprise ? Toute entreprise qui mesure l'importance des différents environnements qui l'entourent et qui cherche à les coordonner peut prétendre réaliser des prévisions performantes. Mentzer et Schroeter, 1994 (86), distinguent à l'époque, trois aspects dans le processus de prévision : logiciel (système informatique), technique de prévision (technique), gestion des services (organisation). Quelques années plus tard, Mentzer et Moon, 2005 (85), reprennent les composants

d'un système de prévision pour les illustrer par la figure 2.4.

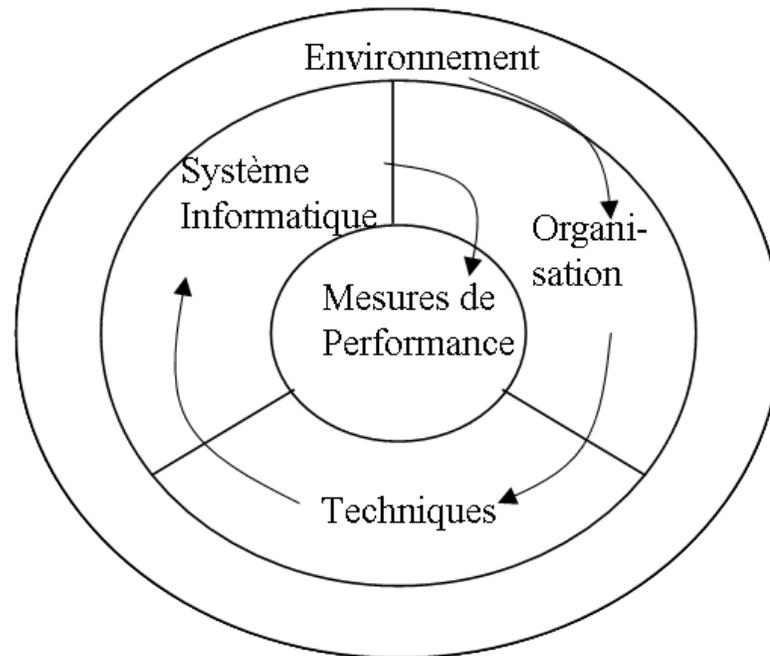


FIGURE 2.4 – *Processus de prévision des ventes, Mentzer et Moon, 2005 (85)*

Le processus de prévision commence par l'identification de l'environnement dans lequel se place l'entreprise. Elle définit alors une organisation, c'est-à-dire les services responsables de la production des prévisions. Le prévisionniste étudie les techniques. En fonction de l'environnement, de l'organisation et des techniques retenus, il choisit un logiciel de prévision. Nous développons chacun des composants dans les sections suivantes.

### 2.3.1 Environnement

La couronne extérieure (voir figure 2.4) représente l'environnement qui entoure l'entreprise. C'est l'ensemble des facteurs externes à l'entreprise qui influencent son activité. L'environnement désigne généralement le marché sur lequel l'entreprise se positionne et, en particulier, la demande des clients. Pour préciser, l'environnement peut être découpé en deux parties. Une première partie regroupe l'ensemble des facteurs sur lesquels l'entreprise peut agir (2) :

- la publicité, pour faire connaître le produit,

- les types points de vente, leur localisation et leur nombre, pour rendre le produit accessible au plus grand nombre,
- la mise en valeur du produit dans les points de vente,
- les ambassadeurs du produit pour renforcer sa crédibilité,
- les services liés au produit (livraison, garanties, etc),
- les politiques d'investissement, de délocalisation pour rendre le produit innovant et bon marché,
- la motivation du personnel pour accroître la rentabilité . . .

Par opposition, la deuxième partie est l'ensemble des facteurs sur lesquels l'entreprise n'a pas d'influence :

- les facteurs juridiques et institutionnels (réglementation juridique, sociale, fiscale),
- les relances de l'Etat,
- l'inflation,
- la croissance économique,
- l'évolution du taux de change,
- l'évolution du prix des matières premières.

C'est dans cette dernière partie que nous allons trouver l'aléa. Pour réduire l'aléa, le prévisionniste cherche à recueillir de l'information sur les 2 types de facteurs. Il est ainsi en mesure de d'identifier les facteurs les plus discriminants et de proposer des solutions pour agir en conséquent. Ainsi, l'entreprise dicte son activité future. Elle n'est plus prise au dépourvu. L'aléa induit par le futur est d'autant plus réduit.

L'environnement qui entoure l'entreprise TFE est décrit dans l'introduction de la section **1.1.1.**

## **2.3.2 Organisation**

D'après l'étude menée par Mentzer et Moon, 2005 (85), les entreprises abordent les prévisions de vente de quatre manières différentes : approche indépendante, approche concentrée, approche négociée, approche consensuelle. Ces approches décrivent la gestion des relations entre services. Elles sont analysées suivant trois axes : la coordination, la communication et la collaboration. La communication est l'échange écrit, verbal ou informatique entre les services. La coordination est la structure entre deux ou plusieurs services. La collaboration est une orientation commune des différents services et de l'entreprise avec ses partenaires qui décident d'un objectif à atteindre.

Les entreprises préconisant **l'approche indépendante** visent à ce que chaque service de l'entreprise utilise son propre moyen de prévision, ses propres techniques, ses propres données, pour ses propres valeurs à prévoir. Les uns prévoient à court terme alors que les autres prévoient à moyen ou long terme. Le point faible de cette approche est que, d'une part, l'entreprise utilise beaucoup de ressources pour aboutir à un même résultat et, d'autre part, qu'il manque de la communication entre les services. L'échange d'informations entre eux est nul. Sous prétexte que ce sont des experts de la vente, il est, par exemple, souvent demandé aux commerciaux de prévoir les ventes. Dans ce cas, il faut faire extrêmement attention aux questions qui leur sont posées et comment sont agrégées leurs réponses, soulignent Armstrong et Brodie, 1999 (8). Un autre exemple pose le problème de savoir comment, la production ou la logistique, peuvent aboutir à des prévisions fiables pour le planning de production ou distribution, sans connaître les opérations événementielles du marketing ?

En revanche, les entreprises, privilégiant **l'approche concentrée**, donnent la responsabilité des prévisions à un seul département, exemple logistique ou marketing. Cette approche managériale permet une meilleure communication sur les prévisions, car elle est de la responsabilité du service. En revanche, elle ne prend pas en compte l'aspect de collaboration, car le service concerné sera influencé par ses propres données pour établir les prévisions. Si c'est le service en charge du commerce, ils feront des prévisions au niveau du client à un horizon mensuel, ou à plusieurs années. Alors que la production a besoin de prévision au niveau produit et à horizon plus court.

Les entreprises, favorisant **l'approche négociée**, développent des prévisions dans chacun des grands départements de l'entreprise. Puis elles rassemblent les prévisions à intervalle régulier pour construire, via une négociation une prévision commune et officielle au niveau de l'entreprise. En termes de communication entre services, l'objectif est atteint ; En revanche la collaboration n'est qu'un leurre puisque chaque département apporte ses propres prévisions. La prévision n'est donc pas guidée par un objectif commun et une information commune, mais par des objectifs et informations séparés des différents départements.

Enfin, les entreprises qui embrassent **l'approche consensuelle** nomment un délégué de chaque département et un responsable prévisionnel pour former un groupe de travail afin de rendre des prévisions issues de données provenant de chaque département. L'entreprise leur demande d'aboutir à une prévision commune tout en respectant les trois axes. Bien entendu, il faut que les personnes du groupe aient le même niveau de responsabilité et qu'ils communiquent au niveau de leur département respectif.

Pour illustrer, citons la revue « Les Echos »<sup>10</sup> : « *Les acheteurs et les logisticiens doivent abandonner un peu de leur autonomie ; quant aux managers opérationnels, ils doivent acquérir le réflexe de faire remonter les informations vers le prévisionniste* ». « *Une or-*

---

10. Les Echos n°19043, Décembre 2003

ganisation centrée autour de la prévision des ventes ne fonctionne que si elle est vraiment voulue et portée par la direction générale », confirme Martine Grosse<sup>11</sup>. « (...) pour être pertinente, une prévision ne peut être faite que localement. Il faut surtout que le prévisionniste puisse travailler en toute indépendance, quel que soit son rattachement hiérarchique », ajoute Fabien Siguier<sup>12</sup>. Nous retrouvons dans cet extrait, l'idée que la prévision doit être proche du terrain, qu'elle doit faire l'objet de communication et collaboration entre les services, et qu'en aucun cas elle ne doit être influencée par des conflits d'intérêts.

L'organisation retenue chez TFE est décrite dans l'introduction de la section 3.1.

### 2.3.3 Techniques

A partir des années 60, une véritable mutation s'est produite quant au fond et à l'expansion de la pensée prospective précise François Hetman, 1969 (62). Depuis, une très large gamme de méthodes sont apparues. Les entreprises se sont emparées des méthodes et des résultats obtenus par les statisticiens pour tenter de décrire, analyser, et trouver des relations de cause à effet pour expliquer la production. D'après Mentzer et Moon, 2005 (85), il existe environ 70 techniques pour simuler des séries chronologiques. La littérature sur le sujet classe ces techniques en trois groupes. Marien, 2001 (80), les nomme « Judgmental approach », « Simplistic models », « Statistical approach ». Chatfield, 2001 (27), préfère les termes « Judgmental forecast », « Univariate methods », « Multivariate methods ». Nous allons les appeler « prévision empirique », « prévision simpliste » et « prévision statistique ».

La **prévision empirique** consiste à demander à des personnes réputées compétentes dans un domaine leur avis sur son évolution. Leurs avis résultent de leur connaissance du domaine considéré, ou de celle de domaines analogues. La prévision empirique peut être utilisée quand il manque des données historiques, mais aussi pour ajuster une proposition de prévision mathématique. Des méthodes éprouvées telles que « Delphi » décomposent le processus de réflexion et offrent une large gamme d'outils pour aboutir à une prévision optimale. La méthode Delphi, repose sur la recherche de consensus de la part d'un groupe d'experts. La méthode Delphi propose de répondre à une question dont la réponse sera connue dans un futur lointain ; « Que sera le transport frigorifique ? ». Les experts se penchent sur la question, rédigent leurs conclusions et se les partagent. C'est une méthode itérative, après chaque rencontre les experts réfléchissent individuellement en prenant en compte l'avis des autres. En fin de boucle, l'opinion des experts

---

11. directrice marketing de Manugistics, fournisseur de solutions de supply chain management

12. contrôleur de gestion de formation, chez fromagerie Bel

doit converger vers une « prédiction ». La prédiction est basée sur un raisonnement plus ou moins intuitif conduisant à annoncer un évènement probable. Alors que la prévision suppose que nous disposons d'une série d'observations et d'hypothèses conduisant à une modélisation du phénomène étudié clarifient M. David et J.-C. Michaud, 1989 (36). Les avantages les plus souvent mentionnés sont le fait que la prévision empirique incorpore le savoir du terrain (produit, retour sur expérience, expertise) et que la prévision ne nécessite ni historique, ni calcul savant. Les inconvénients sont que la prévision est subjective et très coûteuse. D'une part, les prévisions sont forcément biaisées par les politiques de l'entreprise, les buts fixés, les souhaits, l'optimisme, le conservatisme et la persuasion. D'autre part, l'expert chargé de la prévision est toujours influencé par des facteurs externes et il ne peut pas croiser de manière optimale toutes les connaissances. En particulier, les prévisions ne sont pas générées avec toutes les informations jugées importantes, mais seulement avec celles qui ont marqué l'esprit de l'expert, c'est à dire les dernières en date ou les plus utilisées. Enfin, le coût des prévisions qualitatives est élevé, car elles engendrent un temps d'expertise important.

A l'opposé, des méthodes qualitatives, représentées ci-dessus par les prévisions empiriques, se situent les méthodes quantitatives. Elles sont utilisées pour prévoir à un horizon à court ou moyen terme. Elles reposent sur des modèles mathématiques. Leur principe de base est d'estimer la valeur future d'une variable, à partir de la connaissance d'un historique. L'historique peut comprendre le phénomène à étudier ainsi que les variables ayant une influence sur lui. Généralement, les valeurs sont observées à des dates fixées et équidistantes. L'hypothèse est qu'il est possible de déceler une loi dans les données du passé et qu'elle restera vraie dans un avenir proche. Ces techniques sont partagées entre la prévision simpliste et la prévision statistique.

La **prévision simpliste** est un ensemble d'algorithmes simples de prévision tel que les moyennes mobiles, la reprise des données de l'année précédente, la régression linéaire au premier degré, etc. Ces méthodes sont appréciées, car elles sont faciles à comprendre et à utiliser. Elles sont souvent incorporées dans des logiciels de prévision les plus anciens, souvent associées à un algorithme qui choisit le meilleur modèle pour un historique donné. Bien qu'éprouvées, ces méthodes ne sont pas recommandées si nous voulons faire de la prévision à partir de bases de données. De nouveaux logiciels utilisent des algorithmes plus performants pour prévoir plusieurs séries.

La **prévision statistique** se décompose en deux catégories : les techniques endogènes et les techniques exogènes.

Les **techniques endogènes** sont des processus univariés qui utilisent comme seule source d'information celle contenue dans la **série temporelle** (ou chronique). Un type de modélisation très répandu dans le milieu de l'analyse temporelle, et présenté pour la première fois par Warren Persons en 1919, est celui qui isole 4 composantes :

- la tendance est l'allure (croissance ou décroissance) que prend la chronique sur le long terme. Pour la représenter, il est classique de l'exprimer par une droite dite de régression linéaire. Bien qu'il soit parfois tentant de l'approximer par une fonction non linéaire (exponentiel, logarithme), des précautions doivent être prises quant à l'utilisation de ce type de fonction à des fins prévisionnelles. En effet, ces courbes ont une branche parabolique verticale (pour l'exponentielle) ou horizontale (pour le logarithme) qui peut alors symboliser une croissance extrêmement rapide et infinie ou au contraire une saturation. D'après B.Burtschy et C.Menendian, 1980 (24), ces cas de figure sont rares dans le milieu économique.
- la saisonnalité est caractérisée par une variation du phénomène étudié qui se répète à intervalles réguliers sur une période annuelle. Le mouvement saisonnier se répète d'une année à l'autre.
- le cycle est un mouvement comportant une phase de croissance et une phase de décroissance. La périodicité d'un cycle est au-delà d'une année. Au contraire de la saisonnalité, l'intervalle de temps entre les phases n'est pas forcément régulier. C'est pourquoi dans la plupart des études, la tendance et le cycle sont regroupés en une seule composante.
- le résidu, aussi appelé fluctuation accidentelle, est ce qui reste de la chronique après avoir isolé la tendance, le cycle et la saisonnalité. Il peut être en partie expliqué par les facteurs corrélés à la chronique (la météo, offre promotionnelle, grève, ...).

Après avoir été isolées, les 4 composantes fondamentales sont extrapolées individuellement les unes des autres, et réagrégées, pour aboutir à la prévision.

Il existe deux grands types de décomposition, le schéma additif et le schéma multiplicatif.

Le schéma additif, considère que les composantes sont indépendantes les unes des autres. Autrement dit, la variation saisonnière ne fluctue pas avec le temps.

Le schéma multiplicatif, considère que les composantes sont liées et interagissent entre

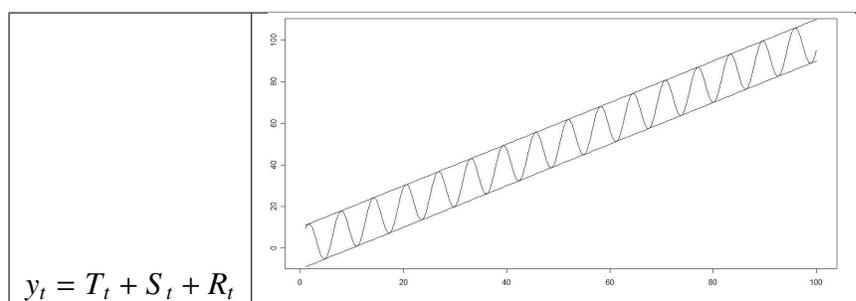


TABLE 2.1 – Schéma additif

elles.

Passer d'un schéma multiplicatif à un schéma additif se fait par la fonction logarithmique.

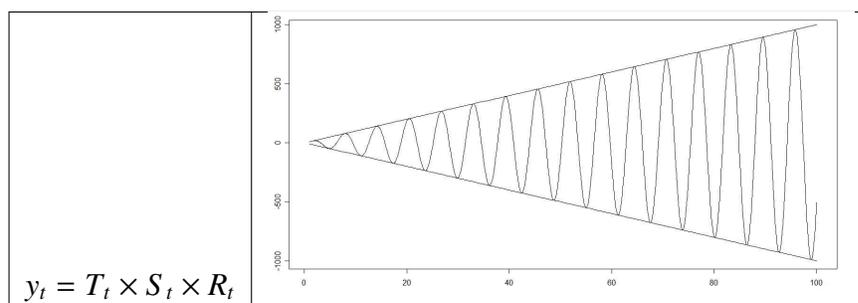


TABLE 2.2 – Schéma multiplicatif

Les techniques d'estimation des composantes sont nombreuses. Nous avons vu que la tendance pouvait prendre plusieurs formes (linéaire, exponentielle, polynomiale). La méthode la plus souvent utilisée en économétrie pour estimer les coefficients est celle des moindres carrés ordinaires (MCO). Les procédures de calcul des coefficients saisonniers sont aussi vastes. La plus populaire est l'estimation de la composante saisonnière par moyenne mobile. R.Bourbonnais et J.C.Usunier, 2007 (18), font remarquer que quelle que soit la méthode retenue, le risque d'incorporer dans la saisonnalité des fluctuations dues à des valeurs erratiques est toujours présent. C'est pourquoi il est conseillé de « nettoyer » la série (correction des jours ouvrables, des valeurs dues à une activité jugée anormale, etc.) avant de lancer les calculs.

D'après Mentzer et Moon, 2005 (85), les techniques endogènes sont partagées entre les « Open-Model Time Series Techniques » (OMTS) et les « Fixed Model TS techniques » (FMTS).

Les OMTS analysent la série chronologique passée pour en déterminer quatre composantes et les projeter dans le futur. Les quatre composantes sont celles identifiées pour la première fois par Warren Persons en 1919. Nous faisons allusion à la tendance, la saisonnalité, le cycle et le bruit (voir section 4.4.3). Les OMTS posent l'hypothèse que les 4 composantes existent dans la série chronologique à prévoir. Ces techniques utilisent une suite de formules pour analyser les composantes individuellement. Les OMTS ont besoin d'un historique conséquent. D'autre part, elles peuvent comporter un grand nombre de paramètres délicats à manipuler. C'est pourquoi elles nécessitent une expérience et un savoir-faire du prévisionniste. Les modèles proposés par Box & Jenkins, 1970 (19), ou l'analyse spectrale dont une introduction est présentée par Graps, 1995 (59), font partie des OMTS.

Les FMTS supposent que seulement certaines des quatre composantes de la chronique existent. Les techniques se servent d'équations en posant des hypothèses de départ pour déterminer si les composantes existent ou non dans la série chronologique. Les FMTS sont utilisées quand il y a peu d'historique, quand le comportement de la série

chronologique change considérablement et régulièrement. Elles sont connues sous les appellations : lissage simple, lissage exponentiel, Holt-Winters (voir section 3.2.4).

Les **techniques exogènes**, aussi appelées modèles économétriques, font appel à l'analyse par régression. Elles ajoutent des sources d'informations provenant de facteurs plus ou moins corrélés avec la chronique à prévoir. Ces sources d'informations sont appelées facteurs exogènes. Ainsi, la fonction principale du modèle économétrique est l'explication de la **série temporelle** à travers d'autres facteurs. Par des calculs de corrélation, le modèle économétrique entraîne le prévisionniste à rechercher d'autres séries temporelles pouvant expliquer celles à prévoir. Dans le contexte de la prévision des ventes, les facteurs exogènes décrivent l'environnement de l'entreprise (voir section 2.3.1). Dans un environnement économique, les quantités de vente d'un produit dépendent étroitement de la force de ventes de l'entreprise, des changements conjoncturels du marché, du comportement des consommateurs, de l'attractivité du produit, de la situation économique du pays, etc. Lorsque les facteurs explicatifs des ventes sont identifiés, leurs valeurs futures sont utilisées pour prédire les ventes.

Dans une perspective de prévision, cette démarche est opérationnelle dans la mesure où nous connaissons, par avance, les valeurs des variables explicatives. C'est rarement le cas. Pour pallier à cet inconvénient, il faut, soit instaurer un décalage de temps entre la variable à prévoir et les variables exogènes, soit prévoir les variables explicatives. Enfin, il ne faut pas perdre de vue que deux séries chronologiques sont comparables seulement si elles sont corrigées des variations saisonnières. Autrement, si nous ne désaisonnalisons pas, il n'est pas possible de déterminer si l'effet d'interactions entre les séries est dû à la saison ou non. D'autre part, deux séries temporelles n'ont pas forcément la même saisonnalité (amplitude et période). Prenons l'exemple d'une agence de transport basée à Nice. Son activité de distribution augmente significativement en été (juillet, août). Cette hausse est due aux nombreux touristes pendant la période des vacances scolaires. Le nombre de touristes en Côte d'Azur, comme l'activité de l'agence de transport, sont des séries à forte saisonnalité. Les deux augmentent tout naturellement en juillet-août, mais pas de la même intensité car les touristes n'ont pas tous les mêmes comportements d'achat. Pour nous affranchir de cet effet, nous considérons des valeurs corrigées des variations saisonnières (CVS). Ainsi, nous pouvons non seulement comparer n'importe quel mois à n'importe quel autre, mais aussi apprécier le véritable impacte du tourisme sur l'activité de l'agence (106).

Souvent, les changements dans les valeurs des facteurs explicatifs retenues ne sont pas dus au hasard. Ils sont la réponse à une action. Par exemple, l'action d'une campagne publicitaire a pour effet une augmentation des ventes. De cette façon, l'entreprise peut influencer les valeurs futures des facteurs exogènes retenus, et ainsi, orienter la prévision des ventes. En stimulant les variables explicatives retenues, l'entreprise simule un ensemble de scénarios et choisi celui qui lui convient le mieux. Mais dans la pratique,

les choses ne sont pas aussi simples. Armstrong et Brodie, 1999 (8), expliquent que la difficulté dans le modèle économétrique est de trouver les bonnes variables explicatives qui décrivent à 100% la chronique à prévoir. De plus, à cause de l'incertitude de l'environnement qui entoure les ventes, les facteurs explicatifs du modèle économétrique sont susceptibles de changer dans le temps. Leur effet sur les ventes est également variable dans le temps. Les entreprises doivent donc scruter périodiquement l'environnement pour ne pas passer à côté de variables explicatives fondamentales. Enfin, les variables explicatives sélectionnées permettent d'ajuster un modèle de régression. L'équation suivante décrit la **série temporelle** ( $y_t$ ) comme une combinaison linéaire de 4 séries explicatives, d'une constante et d'un bruit.

$$Y_t = a_1X_{1,t} + a_2X_{2,t} + a_3X_{3,t} + a_4X_{4,t} + a_0 + \varepsilon_t$$

Les modèles économétriques permettent d'anticiper les changements structurels (au niveau de l'entreprise), les variations macro-économiques, les politiques gouvernementales, etc. L'avantage du modèle économétrique est de quantifier l'influence des facteurs sur ce qui est à prévoir et ainsi proposer à l'entreprise des actions correctives.

Les techniques utilisés dans les modèles de prévision TFE sont décrites dans les sections 3.2 et 4.4.

### 2.3.4 Système informatique

Personne ne peut prévoir sans information préalable. Pour obtenir les informations du passé et du présent, pour cibler les variables exogènes, pour publier les résultats, le logiciel de prévision doit communiquer et interagir avec le système d'information. Les données doivent être numériques et fiables pour pouvoir être traitées aisément et rapidement. Pour extraire de l'information des données récoltées durant les années passées, les entreprises utilisent l'informatique décisionnelle ou BI (Business Intelligence), pour le terme anglais. Dans sa forme la plus simple la BI a pour objectif de donner à la bonne personne, l'information la plus juste, au moment le plus opportun, pour prendre la plus pertinente des décisions. Techniquement, la plateforme BI désigne les outils informatiques permettant l'intégration des données, leur analyse et leur restitution. Ces outils se nomment : système d'information, entrepôt de données, application métiers (ou data mart en anglais), restitution de données (ou reporting en anglais), tableau de bord, fouille des données (ou data mining en anglais).

Les entreprises se sont rendues compte que l'historisation de leurs données pouvait leur rapporter plus que des tableaux de bords à posteriori. Elles peuvent, en effet, utiliser l'expérience acquise en BI pour y intégrer une analyse prédictive de leur activité.

Le logiciel de prévision n'a pas pour vocation de refaire ce que la plateforme BI sait déjà si bien faire. Au contraire, le logiciel de prévision doit s'adosser sur les outils et les résultats de BI existants dans l'entreprise. Ce sont eux qui permettent de collecter, stocker et exploiter les données d'origines diverses. C'est pourquoi nous listons des architectures de système d'information existant (voir section 2.3.4.1). Nous poursuivons par l'étude d'une interface entre le système d'information et le logiciel de prévision (voir section 2.3.4.2). Nous proposons une liste de critères auxquels il est souhaitable que l'outil de prévision réponde favorablement (voir section 2.3.4.3). Enfin nous apportons des éléments pour faciliter la lecture des prévisions par l'utilisateur (voir section 2.2.12).

En considérant que le logiciel de prévision fait partie de l'informatique décisionnelle, il désigne non seulement le programme informatique permettant de prévoir, mais aussi l'ensemble des systèmes de communication utilisés pour interroger le système d'information et restituer l'information créée (voir l'application chez TFE dans la section 4.6.4).

#### 2.3.4.1 Le système d'information ou MIS (Management Information System)

Un système d'information stocke les informations nécessaires au pilotage de l'entreprise. La difficulté est, d'une part, de sélectionner l'information pertinente et d'autre part, de trouver cette information dans le grand éventail des sources de données disponibles. Celles-ci sont des bases de données « métiers » internes à l'entreprise, des bases de données en provenance des clients et des bases de données achetées chez des prestataires. Mentzer et Moon, 2005 (85), recensent trois architectures de données possibles :

- « *System disconnects* », désigne des bases de données « métiers » sans connexion entre elles. Ce qui signifie qu'il n'y a ni échange, ni croisement d'information.
- « *Internal Open Systems Architecture* », dans cette architecture chaque branche de l'entreprise partage ses informations dans une boîte commune appelée entrepôt de données (ou data warehouse en anglais). Le service production, le service des ventes, le service de la logistique, le service de la finance, le service du marketing y déposent leurs données et accèdent à l'ensemble. Ainsi, toute information utile pour les prévisions est facilement accessible et est électronique. Toute personne voulant accéder à une information passe par le système central et, ainsi, peut produire ses propres prévisions.
- « *External Open System Architecture* », est basé sur le même principe mais implique en plus les clients et collaborateurs extérieurs. Les informations apportées par ces derniers sont susceptibles d'améliorer la qualité des prévisions de vente. Cette démarche amène à la notion de prévision collaborative appelée CPFR (Collaborative Planning Forecasting Replenishment)

qui sera discutée ultérieurement (voir section 2.4).

Les aspects d'architecture du système d'information chez TFE sont abordés dans les sections 3.1 et 3.1.1.

#### **2.3.4.2 Le système d'information - le système de prévision**

Les grandes entreprises, cumulent des milliers d'indicateurs. Ces indicateurs ne sont pas toujours donnés pour la même unité de temps, ils sont mis à jour à des périodes différentes, ils sont stockés sur des serveurs différents, ils ne sont pas bruts, certains se construisent à partir d'autres indicateurs. Le logiciel de prévision doit avoir accès à toutes les données susceptibles d'avoir une influence sur la grandeur à prévoir.

Suivant la structure du système d'information disponible et la façon dont le logiciel de prévision est implanté, l'accès aux données sera différent. Si le SI est en mode « System disconnects » le logiciel de prévision accède au mieux à une base de données « métiers » et, au pire, à une base de données enregistrée par le prévisionniste sur son propre poste de travail. Dans tous les cas il ne peut pas regrouper des informations d'origines différentes. Mentzer et Moon, 2005 (85), font remarquer que c'est le meilleur moyen pour que le prévisionniste développe, par une analyse individuelle, ses prévisions pour son besoin personnel. Le résultat est la duplication des efforts. Chacun fait ses prévisions de son côté avec ses propres données sans partager l'information.

Si l'entreprise possède un « Internal » ou « External Open System Architecture », le logiciel de prévision doit pouvoir interroger l'infocentre. Par ce biais, le logiciel de prévision accède à l'intégralité de l'information disponible au niveau de l'entreprise. Souvent ces mêmes entreprises ont implanté un système de communication EDI ou des interfaces web permettant de faciliter la prévision collaborative aussi bien au sein de l'entreprise qu'avec ses partenaires.

Les communications entre le logiciel de prévision et le système d'information chez TFE sont étudiées dans les sections 4.6.1 et 4.7.2.

#### **2.3.4.3 Caractéristiques requises d'un logiciel de prévision**

D'après le magazine « Stratégie Logistique »<sup>13</sup>, le marché français compte une trentaine de progiciels de prévision de la demande. Citons les éditeurs comme APERIA, FUTURMASTER, JDA SOFTWARE, SAP, TXT E-SOLUTION. Ils se spécialisent dans

---

13. Progiciels de prévision de la demande (...), Stratégie Logistique, Hors Série n°5, 2007

des secteurs d'activité de la grande distribution, l'industrie, la pharmacie, l'agroalimentaire, la mode, les produits de grande consommation (CPG). « *Les progiciels de prévision ne cessent d'évoluer pour mieux répondre à la demande, gestion des promotions, processus collaboratif, budget élaboré à partir des prévisions valorisées, intégrations de variables exogènes, prise en compte du réseau de distribution etc.* »<sup>14</sup>.

Cependant, « *les grands groupes sont encore souvent sous Excel® ou Access® en raison du déploiement des ERP* ». D'après une étude de ModelEdition SA, réalisée sur 41 grosses et moyennes entreprises européennes sur leur besoin en matière de prévision, il en ressort qu'Excel® est utilisé pour 64% des prévisions. Les entreprises « *ont plus développé des interfaces de saisies simplifiées, que, déployé des outils de calcul de prévision élaborés. Leur souhait de baisser les stocks les pousse néanmoins à adopter des solutions qui prennent en compte des spécificités* » plus complexes, indique Rémy Le Moigne<sup>15</sup>. Comme pour tous les autres outils, il est difficile de faire un choix parmi la vaste panoplie proposée. L'objet de cette section est de donner des pistes de réflexion pour adapter l'outil informatique de prévision au besoin de l'entreprise.

Mentzer et Moon, 2005 (85), énumèrent les principes de bases d'un logiciel de prévision. Le logiciel doit :

1. répondre aux attentes du prévisionniste,
2. communiquer avec son environnement et proposer une gamme complète de méthodes de prévision,
3. permettre de croiser les méthodes de prévision empirique et statistique,
4. permettre de comparer des techniques pour en retenir une,
5. mettre en avant le « feed-back »,
6. permettre de sélectionner des catégories d'articles en fonction de leur importance stratégique.

Développons ces 6 critères :

1/ Le logiciel de prévision doit pouvoir répondre aux attentes du prévisionniste. Si le prévisionniste veut des prévisions à horizon de trois semaines avec une maille de temps à la journée, l'outil doit pouvoir les lui fournir. Si le prévisionniste veut pouvoir faire des ajustements, il doit pouvoir les faire, etc. L'important est que ce soit l'outil de prévision qui s'adapte aux protocoles mis en place par le prévisionniste et non le contraire. Trop souvent, lorsque l'entreprise a dans l'idée de faire des prévisions de vente, elle se précipite pour acquérir un logiciel. Or, la démarche à suivre est d'abord de poser la méthode de travail, pour ensuite acheter un outil de prévision qui sache s'y ajuster.

---

14. Supply Chain Magazine, n°17, Septembre 2007

15. associé Supply Chain d'Infineo

2/ Nous avons vu dans la section 2.3.4.2 que, pour accéder aux données, l'outil de prévision doit communiquer avec le système d'information de l'entreprise. C'est très souvent une étape complexe à mettre en œuvre. Les bases de données sont stockées sur différents serveurs. Les connexions aux bases de données doivent être paramétrées pour une utilisation optimale. L'accès aux données doit être sécurisé. Les requêtes permettant d'interroger les bases de données génèrent un besoin important en ressources informatiques. C'est pourquoi, le système de prévision doit intégrer un module de gestion et de communication. C'est-à-dire qu'il doit être compatible avec d'autres systèmes afin de communiquer avec ces derniers dans le but de récupérer les données utiles et de diffuser la prévision dans les ERP de planification. Parallèlement, il faut s'assurer que le système de prévision puisse intégrer des événements exceptionnels (promotion, prix qui changent, ...) et qu'il dispose des documents faisant état des hypothèses de départ des prévisions.

Ainsi, le prévisionniste peut avoir accès à l'ensemble des données de l'entreprise pour les analyser et en extraire une prévision des ventes. Le prévisionniste ne doit cependant pas se préoccuper de la complexité du système qui se cache derrière le logiciel de prévision, pour accéder aux données. En d'autres termes, l'accès aux données doit être transparent pour le prévisionniste.

Le logiciel doit fournir un ensemble d'algorithmes permettant d'exécuter les méthodes de prévision mathématique usuelles. Leur programmation ne doit pas être à refaire. Le logiciel doit restituer, sous une forme compréhensible, l'information nécessaire pour permettre au prévisionniste d'ajuster ou de valider un modèle de prévision. Ce qui signifie que le prévisionniste doit savoir paramétrer un modèle mathématique de prévision et interpréter les résultats obtenus.

3/ Les logiciels de prévision ne permettent pas de combiner les techniques de prévision statistiques et les techniques de prévision empirique. Or les avantages et inconvénients des différentes techniques sont très complémentaires. Il semble qu'une bonne prévision commence par une approche des techniques endogènes. La prévision est alors ajustée grâce à des techniques exogènes. Elle est donnée aux experts pour qu'ils corrigent suivant les techniques empiriques. Autrement dit, c'est la prise en compte des données qualitatives. Elle permet d'interroger les experts métiers pour les associer à la prévision. Ainsi, Moon, Mentzer et Thomas, 2000 (91), soulignent que l'outil de prévision doit donner la possibilité à l'expert de rentrer des informations subjectives comme la prévision à risques. Il doit également lui donner la possibilité de valider ou corriger les propositions de prévision. Le cycle entre les propositions de prévision mathématiques, obtenues par les prévisionnistes, et les différents services de l'entreprise représentés par les experts métiers, doit être bien défini. A quel moment du processus les experts peuvent-ils faire les ajustements et quand les prévisions seront-elles validées pour être injectées dans le DPR ? (voir section 2.2.13).

4/ Le logiciel de prévision doit être capable, parmi plusieurs modèles, de sélectionner le meilleur modèle de prévision, suivant des critères de comparaison définis. C'est le fameux bouton « best fit » (ou meilleure prévision, en français). A partir des données récoltées dans le système d'information, le logiciel de prévision sélectionne les facteurs les plus discriminants, analyse l'historique de la série chronologique à prévoir, simule différentes techniques statistiques sur l'historique fourni et applique celle qui donne les meilleures prévisions suivant un ou plusieurs critères paramétrés. Seiersen, 2006 (105) mentionne qu'il n'est pas rare de trouver des dizaines de calculs des prévisions de vente avec sélection automatique de la plus appropriée. Lorsqu'il s'agit de prévoir les ventes de nombreux produits, ce travail préalable qui se fait automatiquement, peut être très utile au prévisionniste. Il ne lui reste plus qu'à ajuster les techniques pour chacun des produits. Par contre, soulignons comme le fait Stan, 2005 (108), un problème de confiance dans les prévisions réalisées automatiquement. De même, Hondt, 2007 (44), note que « *les algorithmes d'optimisation sont souvent trop complexes et ne sont pas compris par les planificateurs générant ainsi des syndromes de boîtes noires, prise en compte partielle des recommandations de l'outil ou application systématique sans analyse de cohérence. D'où le risque de non-utilisation ou d'utilisation malencontreuse* ». D'autre part, Stan, 2005 (108) observe que « *les situations sont fréquentes dans lesquelles les modèles complexes ne donnent pas de meilleurs résultats que les modèles simples* ».

Soulignons, qu'une étude menée par Stan, 2005 (108), a comparé des indicateurs fondés sur les erreurs de prévision, obtenus par des algorithmes communs à chacun des 6 logiciels comparés. L'étude a montré que la performance des modèles mathématiques n'est pas significativement différente d'un logiciel à l'autre.

5/ Il est arrivé fréquemment que les utilisateurs des systèmes de prévision n'aient pas de notions élémentaires en statistique pour créer, avec l'aide de l'outil un modèle de prévision. C'est pourquoi le système de prévision doit être capable de leur retourner des indicateurs d'erreurs et/ou de visualisation graphique leur permettant de juger de la validité du modèle. C'est ce qui est appelé le « feed-back ». Ces informations permettent au prévisionniste d'apprendre les incertitudes liées aux prévisions. Le « feed-back » peut être un graphique superposant les courbes du réel et du prévu. Il peut afficher des indicateurs relatifs à l'erreur de prévision et à la stratégie de prévision. Enfin, le « feed-back » doit révéler la corrélation entre la variable dépendante et les variables exogènes.

6/ Prenons l'exemple d'une grande surface. Elle vend des milliers de produits dans ses linéaires. Pour son réapprovisionnement, elle souhaite prévoir les ventes de chacun de ces produits. Par contre, pour diverses raisons (nouveaux produits, réapprovisionnement difficile, forte marge) certains produits doivent faire l'objet d'un suivi plus rigoureux que d'autres. Le logiciel doit alors permettre au prévisionniste de classer ces produits

comme stratégiques. Ainsi, l'attention du prévisionniste doit se focaliser ces produits en particulier pour lesquels une faible erreur de prévision est attendue. Les autres produits continueront à être prévus par le logiciel, de manière automatique.

Au six critères retenus par Mentzer et Moon, 2005 (85), ajoutons celui de « easy-to-use ». Comme un système décisionnel, le système de prévision doit guider l'utilisateur à construire et exécuter un processus de décisions en l'assistant dans le choix de l'usage des opérateurs. Pour ce faire, Fildes et al., 2006 (51), suggère que le système de prévision limite l'accès au choix des modèles et des données, ainsi qu'au choix de l'ordre des opérations à effectuer et des variables exogènes à sélectionner. Toujours dans l'optique de guider l'utilisateur dans le meilleur choix du modèle, un système de prévision doit aussi être facile à utiliser. Les méthodes mathématiques offertes doivent être diversifiées et paramétrées par défaut, pour en faciliter l'utilisation.

Retenons également les critères de performances proposés par un éditeur de logiciel <sup>16</sup> :

- rapidité et facilité travail décideur / utilisateur,
- capacité de simulation stratégique,
- travail collaboratif,
- gestion de paramètres du modèle,
- gestion de scénario et traçage modification,
- capacité d'agrégation et de désagrégation de données.

D. Bourgoin, Directrice Générale d'Azap, note que les éditeurs ont un vrai rôle d'éducation quant aux capacités réelles des logiciels de prévision et de planification. Bien souvent, les maquettes réalisées lors des processus avant vente polluent le discours en faisant rêver inconsidérément les responsables pour mieux les séduire. Derrière viennent les désillusions. Il y a encore d'importants projets APS (Advanced Planning System) qui sont des échecs <sup>17</sup>.

Finalement, Marien, 2001 (80), estime que le marché propose quatre catégories de logiciels de prévisions :

- « *spreadsheet* »,  
ils sont souvent utilisés pour générer des prévisions. Ils permettent d'implémenter facilement des modèles simplistes. Ils proposent des environnements commodes pour saisir l'historique des données, pour visualiser les prévisions et générer des rapports. Citons les logiciels tels qu'Excel© et les « add\_on » associés, comme Eview.

---

16. ModelEdition SA : [www.modeledition.com](http://www.modeledition.com)

17. Supply Chain Magazine, n°17, Septembre 2007

- « *broad-based statistics packages* », ils sont conçus pour des statisticiens. Ils offrent une longue gamme de méthodes d'analyses, dont les prévisions. Ils demandent à ce que l'utilisateur ait un fort bagage théorique, les paramètres ne sont pas toujours donnés par défaut. Mentionnons les logiciels SPSS, SAS.
- « *business forecasting packages* », ils sont conçus pour des commerciaux. Ils intègrent des méthodes de prévision connues (régression, lissage, Box & Jenkins) Ils offrent, un degré d'automatisation dans la sélection des paramètres de modèles pour que les non spécialistes puissent l'utiliser. Ils permettent souvent d'ajouter des commentaires aux prévisions. Faisons référence aux progiciels Forecast Pro, Autocast.
- « *Forecasting engines* », ils sont conçus pour générer automatiquement un grand nombre de prévisions. Ils sont capables de communiquer et donc de travailler avec le système de planification de type ERP, MRP, Supply Chain. Similaires aux « *business forecasting packages* » en terme de fonctions, ils sont conçus pour traiter de grands volumes. Comme exemple, nous trouvons Futurmaster, JDA-Software.

L'application, chez TFE, des critères cités est à consulter dans la section [3.1](#).

### 2.3.5 Mesures de performance

Nous avons mentionné qu'il existait une infinité de prévision. Partant de ce principe, toute personne est susceptible de donner une prévision. Il n'est donc pas tout de donner une prévision, faut-il encore que nous puissions mesurer sa crédibilité au cours du temps. L'utilisateur fera d'autant plus confiance aux prévisions fournies, qu'il peut vérifier leurs performance dès les résultats du réalisé connus. Toutefois, il existe tellement d'indicateurs statistiques qui définissent la justesse des prévisions qu'il est difficile de choisir lesquels fournir et comment les interpréter.

Mentzer et Moon, 2005 ([85](#)), proposent de les classer sous trois catégories :

- mesures actuelles,
- mesures relatives à une prévision parfaite,
- mesures relatives au modèle de prévision.

Les **mesures actuelles** de performance sont fondées sur l'erreur de prévision. L'erreur de prévision est le résultat de la donnée prévisionnelle moins la donnée réelle à la même période. Mathématiquement, ça pourrait être l'inverse, mais pas d'un point de vue pédagogique. Il faut toujours garder en mémoire que les chiffres doivent être une aide à la décision pour ceux qui les interprètent. Présentée ainsi, l'erreur de prévision nous indique

que, quand elle est positive, nous avons surestimé le réel et quand elle est négative, nous avons sous-estimé le réel. C'est une interprétation didactique et compréhensible de l'erreur. A partir des erreurs de prévision réparties à intervalle régulier dans le temps, une moyenne peut être calculée sur une période donnée. L'erreur moyenne (ou Mean Error en anglais),

$$ME = \frac{1}{N} \sum e, \text{ avec } e = \text{donnée prévue} - \text{donnée réalisée}$$

Il faut toutefois garder à l'esprit, que cette moyenne efface une erreur positive sur une période par une erreur négative d'une autre période. Ainsi, un des critères de performance du modèle serait de trouver une erreur moyenne proche de zéro. L'interprétation sera de dire que la prévision a aussi bien tendance à surestimer qu'à sous-estimer le réel. Par contre, si l'erreur est négative, la prévision aura plutôt tendance à sous-estimer et vice versa.

Pour pallier le problème qu'une erreur positive soit effacée par une erreur négative, l'indicateur erreur moyenne absolue (MAE : mean absolute error, en anglais) utilise la valeur absolue de l'erreur. Il permet de fournir l'erreur moyenne de prévision sans se préoccuper du signe. C'est un indicateur plus révélateur que l'erreur moyenne. Par contre, il ne nous donne pas d'indications sur la taille de l'erreur par rapport aux périodes. Pour ce faire, il faut la rapporter au nombre de périodes qui nous fournira un pourcentage d'erreurs plus facile à interpréter. D'autre part, il donne une importance égale aussi bien aux petites qu'aux grandes erreurs. Or les grandes erreurs sont plus néfastes à l'entreprise que les plus petites. Pour faire la distinction entre grandes et petites erreurs, il est préconisé d'utiliser la somme des erreurs quadratiques. En effet, la fonction quadratique bénéficie de l'avantage de rester constamment positive et celui de pondérer moins fortement les faibles erreurs et plus fortement les erreurs importantes. Comparons une erreur de 2 pour un jour  $J$  et une erreur de 4 pour le jour  $J + 1$ . L'écart entre les deux erreurs est de 2, mais il passe à 12 ( $16 - 4$ ) lorsque que nous comparons les erreurs quadratiques. Pour faciliter l'interprétation, il faut ramener l'erreur quadratique à un référentiel connu. Pour ce faire, elle est divisée par le nombre de périodes sur lesquelles elle a été calculée. Nous obtenons ainsi une MSE (Mean Square Error) par période.

Les **mesures de performances relatives** à une prévision parfaite sont les erreurs de prévision rapportées aux valeurs réalisées.

$$PE = \left( \frac{\text{prev-reel}}{\text{reel}} \right) \times 100$$

Cet indicateur rend un pourcentage de sous ou surestimation du réalisé sur la période du calcul. Son défaut apparaît quand le réel est nul. Informatiquement, la division par zéro n'est pas permise. Il faut alors créer une règle qui affecte à l'indicateur une valeur (exemple : 100%) quand les ventes sont nulles. La représentation graphique de cet indicateur indique sur quelle période la prévision a été majorée et sur quelle autre elle a été

dépréciée. Suite à cette analyse, le prévisionniste peut chercher à améliorer les performances du modèle.

Pour prendre plus de recul et ne pas rentrer trop dans le détail, l'indicateur : moyenne des erreurs absolues en pourcentage (ou MAPE, Mean Absolute Percentage Error en anglais) fournit une vision de l'erreur de prévision sur un ensemble de périodes appelées  $N$ .

$$\text{MAPE} = \frac{1}{N} \sum |\text{PE}|$$

Cet indicateur a le défaut de donner la même importance aux erreurs anciennes qu'aux erreurs récentes. En d'autres termes, il ne montre pas la justesse des dernières prévisions. Pour corriger ce défaut, nous utilisons le pourcentage d'erreur moyenne absolue mobile (ou Year to Date Mean Absolute Percent Error, YTD en anglais). C'est une moyenne mobile sur les 12 derniers mois ou les 4 dernières semaines, etc., selon les besoins.

Les **mesures de performance du modèle**, comparent les résultats du modèle utilisé avec les résultats d'une prévision naïve. La prévision naïve est le réel du mois passé, par exemple, utilisé comme prévision pour le mois d'après. Si le modèle adopté ne donne pas de meilleures performances que la prévision naïve, il est stérile. L'indice de Theil, 1966, est la statistique qui rapporte les prévisions du modèle adopté aux prévisions naïves. Si le résultat est supérieur à 1 alors le modèle adopté est à écarter. Nous chercherons le modèle de prévision qui rapprochera l'indice de Theil de 0. Pour simplifier, la formule suivante peut être utilisée :

$$\text{SFTAB (Sales Forecasting Technique Accuracy Benchmark)} = \frac{\text{MAPE(modèle)}}{\text{MAPE(Naïve)}}$$

C'est au prévisionniste, en concertation avec les utilisateurs, de décider des mesures de performance à afficher.

La section 4.5 décrit les mesures de performance retenues dans le système de prévision développé pour TFE.

## 2.4 Le processus collaboratif

Ce qui est nommé « collaboratif », en informatique, est un principe qui offre aux systèmes d'informations, ou plus globalement aux architectures, la capacité de dialoguer ensemble en temps réel, et ce, sur la base d'un référentiel d'échange commun, explique Léon de SAHB, Directeur Général d'Agrostar. Rassembler les hommes pour qu'ils

échantent leurs connaissances métiers pour le bien de la prévision est un processus collaboratif. « *Les processus collaboratifs servent à mettre en regard les connaissances détenues à divers endroits d'une organisation pour assurer que toutes les données clés vont bien être prises en compte dans la décision finale* »<sup>18</sup>. Seiersen, 2006 (105), définit le processus collaboratif dans le contexte des prévisions de vente comme « *une réunion mensuelle structurée, pour arriver à une prévision de consensus tenant compte de toutes les perspectives, ainsi qu'un plan d'opération pour servir les ventes* ».

« *Pour ce faire, les systèmes doivent être capables d'une part de recevoir un certain nombre de données, de les intégrer et d'autre part, d'émettre des données qui seront intégrées dans le système du partenaire. Cela suppose un travail préalable dit « collaboratif » et un travail d'intégration en interne chez chaque partenaire, au sein de sa chaîne logistique et au sein de son système d'information, ainsi qu'un travail d'intégration externe avec les partenaires amont et aval des chaînes logistiques et des systèmes d'information* »<sup>19</sup>.

C'est en 1998, aux États-Unis, que pour la première fois un consortium de fournisseurs et détaillants (VICS : Voluntary Interindustry Commerce Standart Association) publie un guide pour la prévision et la planification collaborative (CPFR : Collaborative planning forecasting et replenishment). Depuis McCarthy et Golobic, 2002 (83), recensent deux approches de la prévision collaborative vue par la littérature :

- la prévision collaborative intra-entreprise : Diehn (2000/2001) (45), Lapide (1999) (73), Reese (2000/2001) (102), Wilson (2001) (121)
- la prévision collaborative inter-entreprise : Alkerman (2000) (4), Andraski (1999) (5), Barratt et Oliveira (2001) (13), Ireland et Bruce (2000) (67), VICS (1999) (11)

La prévision collaborative intra-entreprise est le processus de prévision sur lequel insistent de nombreux auteurs pour conjuguer la puissance de l'outil informatique et l'expérience des hommes, afin d'aboutir aux prévisions les plus exactes possible (voir section 2.2.5). Ce processus se déroule au sein d'une même entreprise. Elle doit être capable de maîtriser la communication entre les managers et le système. Elle doit avoir rodé les techniques de prévision et ses mesures de performance. Ainsi, une entreprise doit maîtriser ces 4 composantes dans son système de prévision avant de prétendre faire de la prévision collaborative.

La prévision collaborative inter-entreprise est une extension de la prévision collaborative intra-entreprise. Elle inclut, en plus, les partenaires externes face à face ou par le biais de l'informatique, Burgin et al., 2000 (23).

Pour démontrer les avantages d'une prévision collaborative inter-entreprises, citons l'exemple donné par Mentzer et Moon, 2005 (85). La figure 2.5 désigne une chaîne logistique dont les acteurs ne communiquent pas leurs prévisions.

---

18. Supply Chain Magazine, n°17, Septembre 2007

19. Léon de Sahb, La traçabilité, Pratique logistique, Grés d'Or 2005

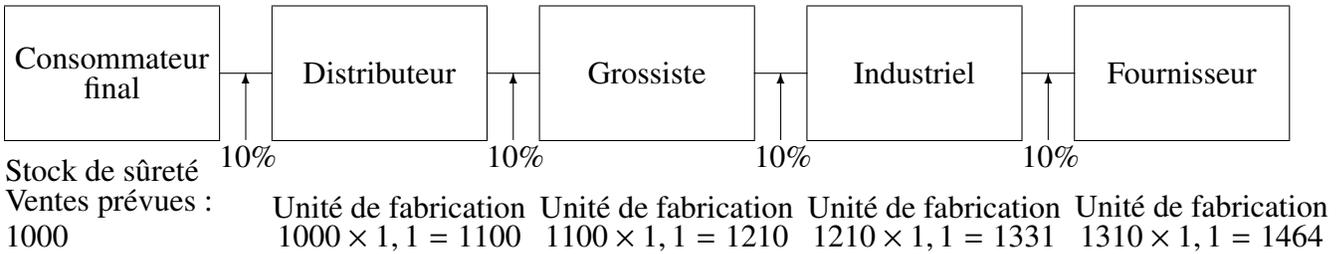


FIGURE 2.5 – Chaîne logistique dont le stock de sûreté cumulé est de 110,5%

Le distributeur se donne une marge de 10% pour le consommateur final, mais ne la communique pas au grossiste. Du coup le grossiste se donne une marge de 10% sur la commande du distributeur ( $1100 \times 1,1$ ) pour faire face à l'imprévu. L'industriel fait de même sur la commande du grossiste et ainsi de suite. Au total la marge de 10%, soit 100 unités de plus, qui était destinée à remédier à une erreur de prévision des ventes au niveau du consommateur final se transforme en  $(100 + 210 + 331 + 464) = 1105$  unités de surproduction soit une erreur de 110,5%.

Avec une relation de communications entre tous les acteurs, il est possible de baisser cette erreur à  $(100 + 100 + 100 + 100) \div 1000 = 40\%$ . C'est-à-dire que chaque acteur de la chaîne communique la marge calculée à l'échelon précédent (dans cet exemple, 100 chez le distributeur) et n'en ajoute pas inutilement.

Pour atteindre ce niveau de confiance entre les partenaires, pour qu'ils échangent de l'information et établissent des processus communs, le CPFR, présenté par le VICS, recommande la feuille de route suivante :

- développement et approbation des objectifs,
- rédaction d'un plan commun pour aboutir aux objectifs,
- création d'une prévision commune,
- identification les exceptions,
- attribution des exceptions à tous,
- création et remplissage de la commande.

Il est à noter que ce ne sont que des outils et qu'aucune prévision ne sera fiable si les partenaires n'ont pas forgé des liens forts et durables entre eux. C'est l'aspect le plus difficile : passer de l'adversaire traditionnel à partenaire ayant le même objectif, la satisfaction du consommateur final. Le point commun dans les entreprises qui réussissent la prévision collaborative, expliquent McCarthy et Golicic, 2002 (83), est la relation de long terme qui existe entre elles. Mentzer et al. ,2000 (91), relate qu'à la question posée aux entrepreneurs « quel est le facteur de succès d'une prévision collaborative ? », ils répondent que les partenaires doivent travailler comme une seule entité pour des objectifs communs. La prévision n'est qu'un aspect de cette chaîne logistique collaborative.

Concrètement, dans les faits, Smâros, 2003 (107), souligne un manque de demande de prévision au niveau des entreprises européennes. C'est une des raisons pour lesquelles l'amélioration promise des performances de la chaîne logistique, par la prévision collaborative (CPFR : Collaborative Planning Forecasting and Replenishment), a été plus lente qu'attendu. D'autre part, Zhao et al., 2002 (123), montrent que la valeur de l'information partagée est significativement influencée par le modèle de la demande, le modèle de prévision utilisé et la capacité des fournisseurs à satisfaire la demande. Les acteurs ont, pour des raisons tangibles, ou non, de grandes craintes à partager des informations telles que le niveau de leur stock, les commandes futures, les temps d'approvisionnement, etc. qui sont nécessaires aux prévisions collaboratives. L'influence des facteurs environnementaux (contexte de concurrence, de négociation des prix, ...) ne présage pas une prévision collaborative objective. Trop souvent, les partenaires ne se connaissent pas assez. Il faut informer, motiver, convaincre les partenaires des avantages que les deux partis peuvent obtenir par le biais de la prévision collaborative. Il faut encourager les partenaires à communiquer les informations nécessaires pour augmenter la fiabilité des prévisions. Pour aider les partenaires et leur apprendre comment améliorer leurs propres prévisions, il est nécessaire de mettre en place un consultant en prévision.

# Chapitre 3

## Analyse de l'existant avant le projet de recherche Horizons

Imaginez la gestion d'un réseau de **messagerie**, sans équivalence en France, composé de plus 70 plateformes et 1.900 véhicules, pour satisfaire chaque jour 15.000 clients en moins de 24 heures. Le réseau TFE compte 100.000 destinataires réguliers par mois. Tous les jours, des milliers d'articles changent de main ; entre une plateforme et ses clients, des centaines de camions arrivent et repartent avec un chargement différent. Les plates-formes sont toutes sous température dirigée.

Pour maîtriser les flux physiques et informatiques qui découlent de la gestion d'un tel réseau, l'entreprise a besoin d'un système d'information (SI). TFE utilise un entrepôt de données dans lequel sont collectées toutes les informations provenant des solutions informatiques métiers (Transport, Logistique, ...). Les informations ainsi récoltées permettent à l'entreprise d'assurer la gestion de ses activités. Fabbe-Costes, 1999 (49), explique que le SI « *intervient à tous les stades du processus décisionnel : la prévision et la planification d'activité, le déclenchement de la circulation, le suivi et le pilotage des mouvements, le contrôle et l'évaluation des opérations et de l'organisation. Jusqu'à présent, le système d'information ne pouvait donner qu'une vision a posteriori des données, constatant un état de fait* ». Pour aller plus loin, le système de prévision à implanter, va fournir les moyens d'anticiper l'activité future. Justement, pour ne plus subir, mais anticiper son activité, TFE souhaite approfondir son savoir-faire en matière de prévision et de planification. Le transporteur cherche par ce moyen à :

- comprendre la demande et l'anticiper pour améliorer sa prestation au client ;
- optimiser les flux de marchandises, comprendre comment chaque flux de marchandises (**expédition, livraison, ramasse**) contribue aux résultats afin d'identifier et de corriger aisément les problèmes liés aux litiges ;

- optimiser la gestion des événements ponctuels (jours fériés, promotion, ...) et prévoir avec précision l'impact afin d'y faire face avec sérénité ;
- gérer des exceptions ou mettre en place des alertes proactives permettant de transformer rapidement un problème potentiel en opportunité ;
- anticiper de manière précise les fluctuations de l'activité et ainsi augmenter sa sérénité face à ces fluctuations ;
- Bénéficier d'un retour sur investissement rapide. Les éditeurs de logiciels promettent que les projets de prévision de la demande se concrétisent par un retour sur investissement de plusieurs dizaines de milliers d'euros et sont rentabilisés en moins de 12 mois après leur mise en œuvre.

La prévision de l'activité dans les sites de transport TFE ne date pas d'aujourd'hui. Avant le projet de recherche « Horizons », la responsabilité de la prévision d'activité était donnée aux **agences**. Ainsi, certains sites de transport utilisent déjà des tableurs pour faire de la prévision sur leurs propres données (voir section 3.2) et non sur des données centralisées (voir section 3.1). Le souci est que cette organisation s'assimile à une « approche indépendante » (voir section 2.3.2), avec les inconvénients qu'elle renferme. L'objectif de ce chapitre est d'analyser l'existant chez TFE dans l'optique de ne pas répéter les mêmes erreurs et surtout d'utiliser ce qui fonctionne.

## **3.1 La structure de l'information chez TFE et son interrogation**

La section 3.1.1 révèle l'interaction entre les bases de données « métiers » et l'entrepôt de données. Elle décrit la source des données et la façon dont elles sont récoltées. La section 3.1.2 décrit les procédures d'extraction et leurs règles de calcul pour alimenter l'historique et extrapoler les données prévisionnelles. La section 3.1.4 décrit la partie de l'entrepôt de données utiles pour l'extraction des données et leur transformation pour aboutir aux séries chronologiques décrivant l'activité des sites de transport.

### **3.1.1 Interaction entre le système d'information et le système de prévision**

L'entrepôt de données rassemble les données provenant de diverses applications métiers. Il fournit une vision transversale à travers plusieurs métiers de l'entreprise. La

figure 3.1 indique les sources d'alimentation de l'entrepôt de données. Ce sont des bases de données Oracle® dédiées aux activités transversales de l'entreprise (Transport, Logistique, Contrôle de gestion et Ressources Humaines, Direction Technique Véhicules, Direction des Achats). Les bases de données sont issues des applications TMS (Transport Management System), WMS (Warehouse Management System), SAP (Systems, Applications, and Products for data processing), GTA (Gestion des Temps et Activité). L'entrepôt de données enregistre l'historique des transactions menées par ces bases de données. Les chaînes de chargement sont quotidiennes, hebdomadaires, mensuelles et trimestrielles. Seiersen, 1998 (104), signale que l'entrepôt de données peut être interrogé par « une palette d'outils d'aide à la décision, comprenant des outils de prévision, de simulation, ainsi que des outils data mining, de reporting ou encore des algorithmes de recherche opérationnelle ». Ces outils sont présents sur la couche inférieure de la figure 3.1.

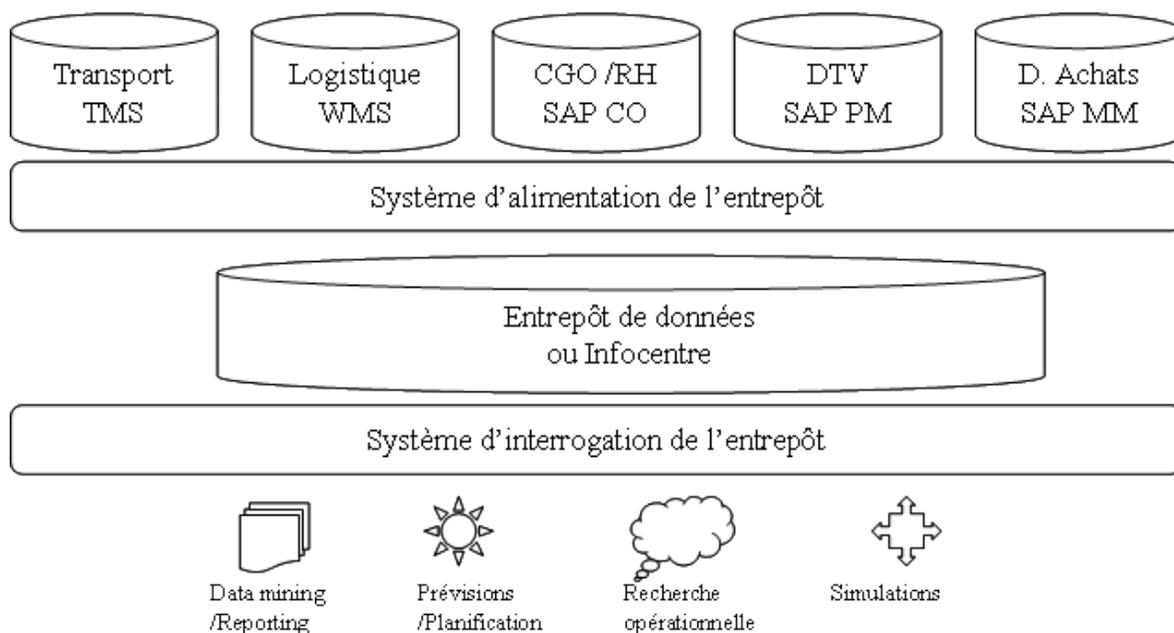


FIGURE 3.1 – Architecture du Système d'Information chez STEF-TFE

La plate-forme décisionnelle développée par Agrostar se nomme Infomanager. Cette plate-forme est développée sous Cognos® V7 et C8. L'objectif d'Infomanager est de fournir au groupe STEF-TFE un système d'information décisionnel permettant :

- l'accès direct aux indicateurs clés de performance par domaine fonctionnel,
- une navigation naturelle vers les rapports,

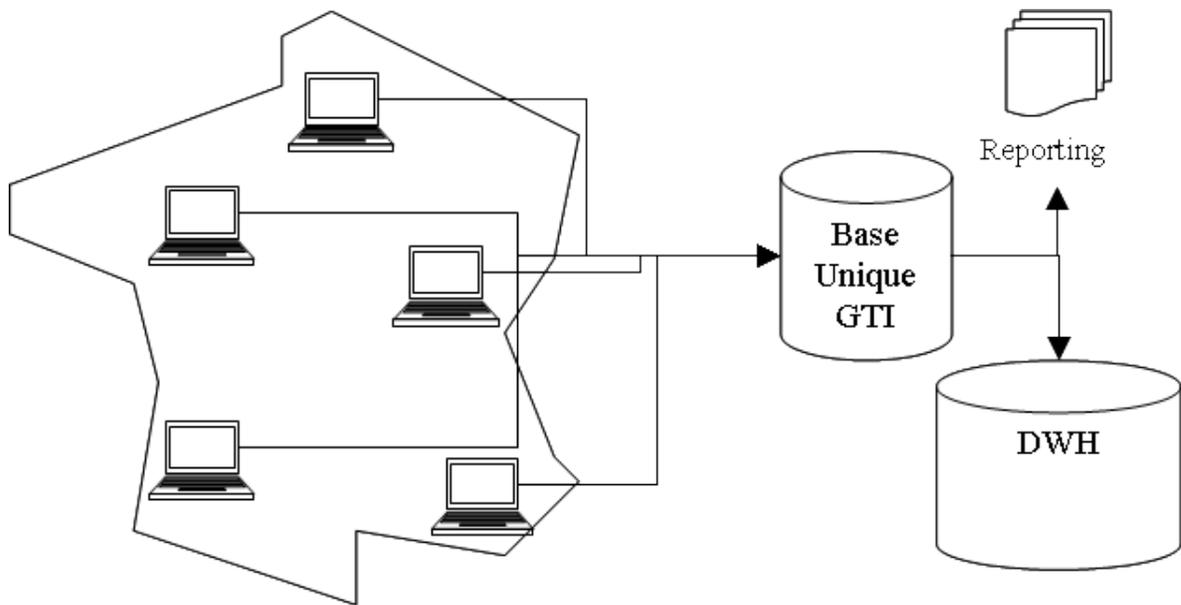


FIGURE 3.2 – Architecture du Réseau Informatique chez STEF-TFE

- l'utilisation d'outils de requêtage, d'analyse ou d'anticipation collaboratifs,
- la mise en place d'alertes et la détection d'erreurs dans les données du référentiel.

Infomanager est un outil à disposition de l'ensemble des services et métiers du groupe. Citons le CGO (Contrôle de Gestion Opérationnel), la DHR (Direction Ressources Humaine), la DEX (Direction d'Exploitation), la DTV (Direction Technique Véhicules), la DC (Direction Commerciale), ADV (Administrateur Des Ventas), les Directeurs de site, les Directeurs Région, les services Opérationnels (Quai, Exploitation, SAV).

Parmi cet ensemble d'outils de présentation de l'information, d'aide à la décision, d'analyse des données en aval de l'entrepôt de données (ou Infocentre, voir figure 3.1), notons que Fabbe-Costes, 1999 (49), retient la « prévision » comme module essentiel. Le système de prévision à installer chez le transporteur a pour vocation d'anticiper l'activité des agences de transport. Historiquement, l'activité est mesurée par le poids des marchandises transportées et par le nombre de lettres de voiture signées (voir section 1.2.1). Branche, 2006 (20), signale que ce sont deux « *indicateurs essentiels. Ils conditionnent une partie des moyens matériels mobilisés* ». La prestation de transport est encore souvent facturée au poids, sinon à la palette. Une semi-remorque pleine contient 33 palettes Europe pour un poids de 25 tonnes.

L'information requise pour construire l'historique des séries chronologiques à prévoir provient de la base de données du TMS (Transport Management System). Le TMS est l'ensemble des solutions informatiques dédié à la gestion du transport des marchandises. Avant 2006, le réseau informatique du transporteur était constitué d'une multitude

de serveurs régionaux reliés à un serveur central appelé « HP 3000 ». Les postes informatiques régionaux se connectaient au serveur régional. L'avantage étant que quand le serveur central tombait en panne, les serveurs régionaux et par conséquent les flux d'informations continuaient à fonctionner. L'inconvénient majeur était la réplication des données nécessaires pour partager l'information. Pour remédier à sa vétusté et dans un souci de modernisation, le transporteur change en 2007/2008 de réseau informatique. Il passe à GTI (Gestion du Transport Intégré). C'est une base de données unique (voir figure 3.2) qui permet un partage de l'information en temps réel. Lorsque qu'un camion prend son départ à Limoges, par exemple, l'ensemble du réseau du transporteur est au courant.

Le réseau du transporteur étant par « *nature multi-acteurs et multi-sites, le SIC (Système d'Information et de Communication) a pour principale mission d'assurer la cohérence de ces ensembles complexes d'opérations que le transporteur cherche à synchroniser* » explique Fabbe-Costes, 1999 (49). Le SIC au niveau transport est le TMS. « *Il a pour mission de réaliser le couplage des flux d'information (technique, administratif, stratégique), avec des flux physiques (lot, emballage, marchandise) avec des actions (manutention, transport) et les acteurs (chauffeurs, manutentionnaire) qui les réalisent* » (49). Il contient des modules de planification et d'organisation du transport. Ces modules permettent de planifier les voyages, d'allouer les ressources (véhicules, chauffeurs) aux voyages, d'organiser et gérer les transports (ramasses, livraisons, expéditions) et la gestion du groupage de marchandises. L'ensemble des informations saisies dans ces modules est enregistré dans la base de données GTI et historisé dans l'Infocentre. Ainsi, le transporteur répond à la volonté d'avoir une traçabilité totale des processus logistiques (voir section 1.1.1.1). Il est en mesure de proposer à ses clients, un suivi informatique des expéditions (SIE) et d'amener « *des preuves de bonne réalisation de la prestation* » (49). Pour appréhender l'organisation des tables dans cette base, il faut avant tout comprendre la segmentation des flux de marchandises indique Seiersen, 1998 (104).

### 3.1.2 Segmentation des flux de marchandises

Pour qu'un transport existe, il faut qu'un donneur d'ordres fasse appel à un transporteur en lui envoyant une lettre de voiture (ou bon de commande). Le donneur d'ordres est aussi, dans la plupart des cas, le tiers facturé. La lettre de voiture comporte, au minimum, un client expéditeur chez lequel le transporteur vient ramasser la marchandise et un client destinataire, chez lequel le transporteur vient livrer la marchandise. Nous distinguons d'ores et déjà « *les flux amont, des usines aux entrepôts, et les flux aval des entrepôts aux destinataires finaux* » souligne Artous, 2006 (10). Seront aussi indiqués

sur la lettre de voiture, un certain nombre d'indicateurs caractérisant la marchandise à transporter (nom, poids, nombre de colis, nombre de palettes ...). L'ensemble de ces informations permet de créer une « **position** ». C'est la traduction informatique de la lettre de voiture. Chaque position est enregistrée dans la table POS (voir figure 3.5). Cette table comporte plus de 55 millions d'enregistrements. Un enregistrement équivaut à une position. Deux millions de nouvelles positions sont créés tous les mois.

A partir du lieu de chargement de la marchandise chez le client expéditeur, et du lieu de déchargement chez le client destinataire, les solutions TMS avec l'aide des exploitants, définissent des **voyages** que vont emprunter les marchandises. Le but étant, bien entendu, d'optimiser le coût du transport. Les voyages sont enregistrés dans la table VOY (voir figure 3.5). La table VOYAGE comprend 5 millions d'enregistrements. Un voyage est décomposé en **bordereaux**. Les bordereaux regroupent des segments.

Ainsi, une position va effectuer un ensemble de voyages. Pour la suivre géographiquement et en temps réel tout le long de son transport, la position est découpée en segments. L'intitulé des segments dépend de certaines caractéristiques du lieu de chargement de la marchandise et de son lieu de déchargement (voir figure 3.4). Le découpage des positions par segments est disponible dans la table SEG (voir figure 3.5). Elle compte 68 millions d'enregistrements. Un segment a une date de chargement et une date de déchargement qui sont issues du PTO (Plan de Transport Opérationnel). Le PTO est une « moulinette » à segments. Grâce à lui, chaque segment de position trouve automatiquement sa place dans le bon voyage planifié explique Branche, 2006 (20).

Sur le terrain (figure 3.3), le lieu et la date de chargement et de déchargement du segment sont saisis manuellement ou à l'aide d'un scanner optique permettant de lire des codes barres. De cette façon, le système d'information compare le planifié avec le réalisé. La figure 3.4 affiche une présentation générale des segments d'entrées et de sorties existants. Deux cas sont possibles, soit la marchandise, en provenance de l'**expéditeur**, est amenée à **quai** d'une agence TFE pour être conditionnée et repartir vers une nouvelle destination, soit l'expéditeur affrète un camion complet pour transporter la marchandise de son lieu de chargement vers le destinataire final. C'est du transport de lot.

Un segment est identifié par un évènement. Certains se retrouvent en entrée et en sortie du quai de l'agence. C'est le cas d'évènements EX.

- EX : L'évènement associé au segment est dit d'expédition. Lorsque la marchandise est transportée entre deux agences du réseau TFE, il peut être en sortie comme en entrée de quai (agence).

Parmi les évènements d'entrée, nous comptons

- ED : lorsque qu'une agence *B* est mandatée par une agence *A* pour ramasser la marchandise chez un expéditeur appartenant à la zone de chalandise de l'agence *A*, l'évènement associé au segment entre l'expéditeur et l'agence *B* est appelé « enlèvement dégroupier ».

- EN : lorsque la marchandise est ramassée chez l'expéditeur, par les véhicules de l'agence TFE, l'évènement associé au segment est dit « enlèvement ».
- AQ : lorsque c'est l'expéditeur qui se charge de transporter la marchandise au quai de l'agence TFE, l'évènement associé au segment est dit « amené à quai ».
- RP : lorsque le destinataire de la marchandise n'a pas été trouvé ou lorsqu'il a refusé la marchandise, celle-ci est renvoyée à l'expéditeur. L'évènement associé à ce segment est appelé « reprise ».
- RE : Pour faciliter le transport des marchandises, elles sont posées sur des palettes, dans des rolls ou des bacs. Ces emballages appartiennent à l'expéditeur et doivent lui être renvoyés après livraison de la marchandise chez le destinataire. L'évènement associé au segment de retour est « remise emballage ».

Parmi les évènements de sortie, se trouvent les :

- LI : lorsqu'une agence TFE s'occupe de transporter la marchandise entre une agence de transport et son destinataire final, l'évènement associé au segment est appelé « Livraison ».
- PQ : lorsque le destinataire final vient prendre la marchandise au quai de l'agence TFE, l'évènement associé au segment est appelé « Prise à quai ».

Lorsque la marchandise est transportée entre l'expéditeur et le destinataire final sans qu'elle ne subisse de déchargement / chargement intermédiaire, l'évènement associé au segment est dit « Livraison Directe » (LD).

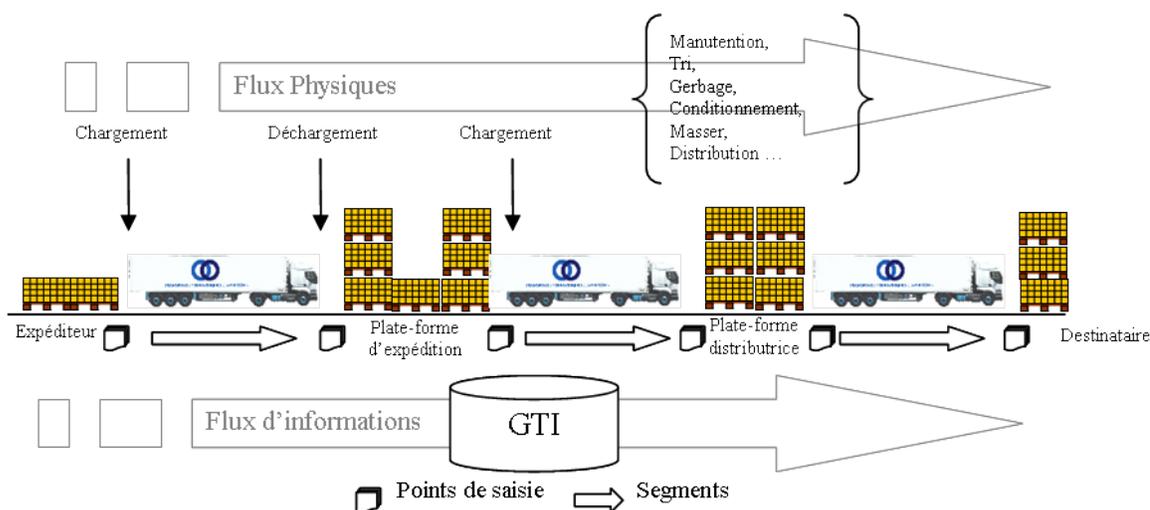


FIGURE 3.3 – Traduction des flux physiques en flux informatiques

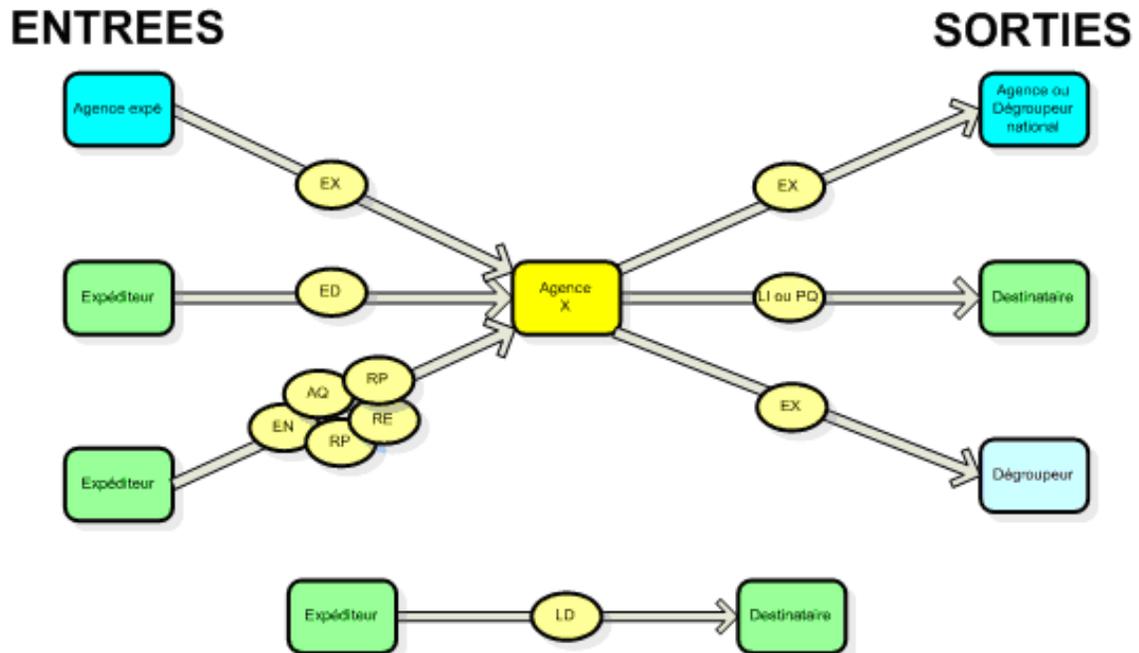


FIGURE 3.4 – Le découpage de l'activité transport en segments

Les segments de positions voyageant ensemble et qui répondent au même critère d'unicité (même catégorie de froid, même lieu de chargement, même date) sont regroupés dans un **bordereau**. Pour simplifier, un camion transporte des bordereaux. Les bordereaux sont saisis dans la table BORDEREAU (voir figure 3.5). Elle compte 12 millions d'enregistrements.

A partir des différents évènements, une typologie des flux de sortie est définie (voir annexes - Indicateurs journaliers). Le logisticien des flux distingue l'activité d'expédition et l'activité de distribution. L'expédition est l'ensemble des marchandises qui ont pour destination un lieu de déchargement hors de la zone de chalandise du lieu de chargement. La direction d'**exploitation** du transporteur, cherche à typer ses transports de marchandises sur une maille de suivi assez fine, pour prendre en compte les spécificités des 70 agences du réseau et pour que le pilotage des flux de marchandises reflète au mieux la réalité du terrain. Ainsi, les expéditions regroupent 16 catégories indivisibles. Elles peuvent être agrégées pour former de nouveaux indicateurs. C'est le cas des indicateurs à prévoir.

Pour une agence donnée, le poids des marchandises passées à quai et qui partent en

expédition est donné par la formule suivante,

$$\begin{aligned} \text{Tonnage expédié avec passage à quai} = & \text{TDE} + \text{TEX} + \text{TEXTT} + \text{TLMHgr} \\ & + \text{TLHgr} + \text{TLHct} + \text{TLMHct} \end{aligned} \quad (3.1)$$

Avec les catégories de poids indivisibles suivantes :

- *Tonnage Expédié par Dégroupeur (TDE) =*  
Somme des poids des positions expédiées pour livraison par un dégroupéur local Hors zone.
- *Tonnage Expédié Traction (TEX) =*  
Somme des poids des positions du périmètre venant des EX.
- *Tonnage Expédié Transit (TEXTT) =*  
Somme des poids des positions TT du périmètre venant des EX.
- *Tonnage Livraisons bordereaux Manuels Hors zone Groupage (TLMHgr) =*  
Somme des poids des positions bordereaux manuels et dont le destinataire ne fait pas partie de la zone de livraison de l'agence (e-plan) et dont le voyage a un centre de coût 89.xx.
- *Tonnage Livraisons Hors zone Groupage (TLHgr) =*  
Somme des poids des positions sur une tournée LD appartenant à un voyage dont le centre de profit commence par 89.
- *Tonnage Livraisons Hors zone CT (TLHct) =*  
Somme des poids des positions sur une tournée LD appartenant à un voyage dont le centre de profit commence par 88.
- *Tonnage Livraisons bordereaux Manuels Hors zone CT (TLMHct) =*  
Somme des poids des positions bordereaux manuels dont le destinataire ne fait pas partie de la zone de livraison de l'agence (e-plan) et dont le voyage a un centre de coût 88.xx.

De la même façon que pour l'expédition, les flux de marchandises en distributions sont regroupés pour former une typologie. L'indicateur à prévoir est le poids des marchandises passées à quai et qui partent en distribution. Cet indicateur est obtenu par la formule suivante :

$$\text{Tonnage distribué avec passage à quai} = \text{TDL} + \text{TLL} + \text{TPQ} + \text{TLMLocal} \quad (3.2)$$

Avec les catégories de poids indivisibles suivantes :

- *Tonnage Livraisons par dégroupéur (TDL) =*  
Somme des poids des positions expédiées pour livraisons par un dégroupéur local.
- *Tonnage Livraisons Locales (TLL) =*  
Somme des poids des positions bordereaux LI sur une tournée LI.

- *Tonnage Pris à Quai (TPQ) =*  
Somme des poids des positions avec un segment de sortie PQ.
- *Tonnage Livraisons bordereaux Manuels Locales (TLM Local) =*  
Somme des poids des positions bordereaux manuels et dont le destinataire fait partie de la zone de livraison de l'agence (e-plan).

La séquence classique des coûts dans un réseau est : enlèvement, passage à quai, traction, tri et passage à quai, traction, passage à quai et livraison (voir figure 3.4). L'enlèvement, et la livraison, ainsi que les coûts de passage à quai, constituent l'essentiel des coûts explique Branche, 2006 (20).

Le quai fonctionne jour et nuit. Sa mission est de réceptionner, contrôler, zoner, stocker, préparer les commandes et charger la marchandise dans les délais prévus en respectant l'intégrité de la marchandise et les règles d'hygiène et de sécurité. L'équipe de jour est chargée de l'expédition alors que celle de nuit prend en charge les colis pour la distribution. Anticiper le poids des marchandises, le nombre de positions passées à quai, en expédition et en distribution, permet de fournir une donnée essentielle pour planifier, contrôler les coûts matériels et humains engendrés par ces marchandises. Mais la typologie des flux de marchandises n'existe pas en tant que telle dans la base de données GTI. Elle est le résultat d'un processus de croisement de données issues de la base de données GTI. Ce résultat est stocké dans l'Infocentre (voir section 4.2).

### 3.1.3 Source des heures productives sur le quai

Les heures productives quai sont les heures travaillées par le personnel employé sur le quai. Le personnel badge une première fois en arrivant, et une seconde fois en quittant le travail. Ainsi, les heures sont comptabilisées et une productivité quai peut être déduite. L'indice de productivité quai est le ratio Tonnes / Heures.

Par le biais d'un progiciel de gestion de ressources humaines, appelé eTemptation, la badgeuse envoie les temps dans la base de données GTA (Gestion de Temps et Activité). eTemptation est l'outil RH du groupe STEF-TFE. Il permet de gérer les calendriers, de construire des plannings, d'administrer les employés et leurs contrats, d'affecter des badges et de recueillir les temps badgés, etc. Les heures badgées sur le quai sont affectées à des centres de coûts. Ces derniers permettent d'identifier les heures travaillées pour l'activité expédition et celles travaillées pour l'activité distribution. Une grande partie de l'expédition est réalisée par l'équipe de jour entre 13h et 21h. Alors que la distribution est réalisée par l'équipe de nuit entre 22h et 6h. Cette corrélation entre expédition / distribution et activité quai jour / nuit permet de rapprocher l'indice de productivité en expédition à la productivité du quai jour et l'indice de productivité en distribution à la productivité du quai nuit. Par contre, les indices de productivité ne sont pas comparables.

L'activité du quai jour est avant tout destinée à massifier les flux. Le quai reçoit les colis enlevés lors des tournées de ramasses et procède à des opérations de **groupage** avant expédition vers les sites de **messagerie** du réseau. L'activité du quai jour consiste à ventiler les colis reçus la nuit, en provenance du réseau et à destination de la livraison des « *GMS jusqu'aux boucheries de proximité, en passant par la restauration rapide ou la restauration collective* » explique Serge Capitaine, Dg adjoint de TFE. L'activité du quai jour n'est donc pas la même que l'activité du quai nuit.

L'indice de productivité, mais aussi le poids et le nombre de positions passées à quai, ne sont pas des indicateurs pour comparer les agences de transport entre elles. Ils ne « *visent pas nécessairement à établir un palmarès, la typologie de chacun des sites pouvant donner des résultats très différents pour un même indicateur. Un bon chiffre pour un site ne l'est pas obligatoirement pour le site voisin et inversement. L'analyse la plus pertinente consiste donc à comparer une performance individuelle dans le temps. Ces indicateurs doivent permettre d'avoir une vision de l'exploitation et de faire des comparaisons site à site. Cela doit encourager à échanger sur les meilleures pratiques à appliquer* »<sup>1</sup>.

### 3.1.4 Structure et agencement des données sources

Nous venons d'indiquer que la source des données à prévoir est extraite de la base de données « métier » appelée GTI. Une partie du schéma logique en modèle relationnel (ou schéma conceptuel) est tracée dans la figure 3.5. Il indique les tables utilisées pour typer les flux, et les relations entre ces tables.

Un plan de ramasses et de livraison (table LIVPLAN) est proposé aux clients par l'agence de transport (table DOM\_ACT). Le client envoie une **lettre de voiture** à l'agence. Le récépissé est archivé (table FICH). La lettre de voiture est traduite informatiquement (table POS). Elle indique un tiers **expéditeur** et un tiers destinataire (table TIERS). Le PTO divise la position en segments (table SEG). Les segments ayant le même lieu et même date de chargement et de déchargement sont rangés dans un **bordereau** (table BORDE-REAU). Le bordereau réalise un voyage (table VOY). Le **voyage** rassemble des tournées planifiées par le PTO (table TOUR\_PLAN). Le **voyage** est composé de points d'arrêts (table ARRET). Les tournées et les arrêts desservent des tiers (table TIERS) et des dégrouppeurs (table DEGR\_PTO). Ces derniers font part d'évènements (table EVT\_SIE) s'il y a lieu. Un évènement peut être un litige sur la marchandise, par exemple, dans ce cas la marchandise n'est pas acceptée par le destinataire et un processus de négociation est entamé entre les sites du transporteur et l'expéditeur pour déterminer les responsabi-

---

1. Marc Vettard, Directeur d'Exploitation TFE, 2009

lités. Si l'expéditeur est responsable, la marchandise se voit attribuer un nouveau numéro de position et repart vers ce dernier par le circuit de transport.

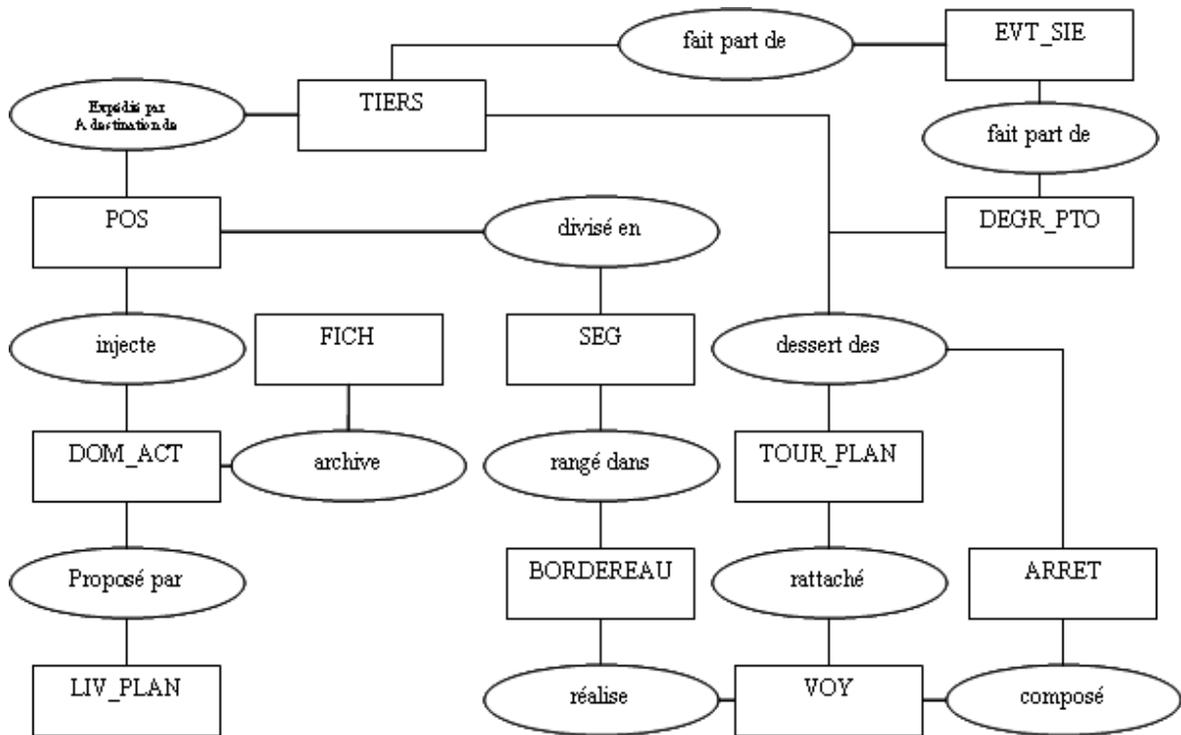


FIGURE 3.5 – Une partie du modèle relationnel de la base GTI

Un croisement des informations obtenues par les tables du schéma logique (figure 3.5) permet d'extraire un historique des indicateurs décrits dans la section 3.1.2. Le résultat quotidien des 16 indicateurs en **expédition** et des 12 indicateurs en distribution alimente la table INDIC\_AGE\_Q dans l'Infocentre.

## 3.2 Les modèles de prévision utilisés chez TFE

L'idée de se doter d'un système de prévision, chez TFE, n'est pas nouvelle. Des agences TFE ont réfléchi sur l'action des facteurs temporels et économiques. Elles ont missionné leurs experts terrains pour qu'ils se penchent sur le sujet et qu'ils développent des outils de prévision spécifiques à l'agence. Ainsi, elles ont construit, chacune de leur côté, un modèle les caractérisant. Dans ce sens, elles ont suivi une approche indépendante et non concentrée ou même consensuelle (voir section 2.3.2).

Avant le projet de recherche « Horizons », une vingtaine d'agences, sur les 70 du réseau

TFE, utilisaient l'un ou l'autre des trois outils présentés dans cette section. Leurs points communs :

Ils sont écrits pour prévoir l'activité des agences TFE, c'est-à-dire les séries temporelles suivantes :

- poids total en kg des marchandises passées à quai, sur une journée d'activité,
- poids expédié en kg des marchandises passées à quai, sur une journée d'activité,
- poids distribué en kg des marchandises passées à quai, sur une journée d'activité,
- nombre de positions totales des marchandises passées à quai, sur une journée d'activité,
- nombre de positions expédiées des marchandises passées à quai, sur une journée d'activité,
- nombre de positions distribuées des marchandises passées à quai, sur une journée d'activité.

L'intervalle de temps entre deux observations d'une chronique est le jour. Pour une majorité des agences de transport étudiées, l'historique des chroniques débute le 1<sup>er</sup> janvier 2002. Les valeurs des séries chronologiques à prévoir varient entre 0 kg et 3.200.000 kg. Les algorithmes des modèles statistiques développés par les agences de transport ont la particularité

- d'être implantés avec un tableur ;
- de prévoir des valeurs mensuelles ou hebdomadaires qu'ils distribuent sur les 6 jours ouvrés de la semaine ;
- d'être spécifiques à un ensemble d'agences restreint, à cause de paramètres fixés et non calculés ;
- d'obtenir une erreur en moyenne absolue variant autour de 5% et 10%.

Leur différence réside dans le choix du modèle statistique :

- deux d'entre eux s'appuient sur des modèles extrapolatifs : lissage par moyenne mobile et lissage exponentiel ;
- le troisième s'appuie sur un modèle économétrique, et donc sur des variables explicatives.

### **3.2.1 Notions autour des séries chronologiques**

Avant d'étudier les modèles de prévision utilisés chez TFE, attardons-nous à définir certaines notions liées à l'analyse de séries chronologiques.

Une **série chronologique**, aussi appelée chronique ou **série temporelle**, se définit comme

une suite d'observations d'une grandeur homogène, espacées dans le temps. C'est un ensemble de couples dont la première composante est le temps et la deuxième, la valeur qui y est associée. La série temporelle est ordonnée de façon chronologique. Le pas entre deux valeurs successives n'est pas nécessairement constant. Nous appelons série chronologique régulière une série dont les dates d'observation sont à périodicité constante. Pour être prévisible, une chronique doit se composer d'une superposition de mouvements oscillatoires (périodiques) auxquels s'ajoutent une tendance et une erreur.

L'objectif des méthodes de prévision est de déterminer la valeur qu'une chronique peut prendre à un horizon donné. L'**horizon de prévision** (noté  $h$ ) varie de l'immédiat au long terme. Il ne peut pas être inférieur à l'intervalle de temps s'écoulant entre deux observations successives. En revanche, il peut lui être supérieur et nous parlerons d'horizon à long terme. La notion d'horizon est étroitement liée à l'utilité que nous voulons donner à une prévision. Dans le milieu économique, un horizon stratégique concerne le long terme (>12 mois), le niveau tactique concerne le moyen terme (entre 2 semaines à 6 mois) et le niveau opérationnel concerne le court terme et les événements journaliers (voir section 1.2.4).

Les modèles de prévision d'une série chronologique posent l'hypothèse forte que le passé est garant de l'avenir. Cela veut dire que le phénomène étudié continuera à se comporter comme il l'a fait dans le passé, explique Claude Olivier, 2002<sup>2</sup>. Comme dans toute étude statistique, il faut donc passer par une phase exploratoire permettant de comprendre et d'apprécier les phénomènes temporels influant sur la grandeur étudiée. Elle permettra de répondre à la règle d'or de toute prévision. G. Ansion, 1990 (6), nous apprend que cette règle repose sur la pondération des données du passé. C'est une perception qualitative que nous avons sur la série à étudier. Suite à l'analyse descriptive, il faut pouvoir décider de l'importance que nous accordons aux données suivant leur âge d'observation. Les phénomènes observés récemment sont-ils plus significatifs que ceux constatés dans un passé plus lointain ou vice versa ?

La **saisonnalité** est une variation de l'activité de l'entreprise à certaines périodes de l'année (106). Cette variation se répète tous les ans, à la même période (voir section 2.2.3). La saisonnalité est souvent liée au type d'activité. Celle de TFE est conditionnée, entre autres, par les fêtes de fin d'année. Tous les ans l'activité augmente considérablement en décembre pour rechuter en janvier.

La **période** est la durée d'une observation saisonnière. Il y a, par exemple, 52 variations saisonnières correspondant chacune à une période d'une semaine.

---

2. Claude Olivier, Université du Québec, 2002

L'**historique** ou « ensemble d'informations » sont les éléments dont le prévisionniste dispose afin de réaliser une prévision. Suivant la granularité de l'horizon de prévision (heure, jour, mois, année) et de la méthode de prévision utilisée, l'historique devra être plus ou moins long. Les méthodes endogènes (décomposition, lissage, Box & Jenkins) se nourrissent uniquement du passé de la variable à prévoir. Alors que les méthodes exogènes utilisent également les valeurs passées d'autres variables.

### 3.2.2 Modèle économétrique de TFE

Le modèle économétrique a été développé pour les agences de la région Bretagne et Pays de la Loire. Le principe du modèle est d'expliquer les variations des chroniques à prévoir par des variables économiques étroitement liées. Un exemple marquant est la relation qui existe entre le poids mensuel des marchandises transportées par TFE et l'emploi salarié du transport. Si l'activité du transport se porte bien, elle va augmenter sa masse salariale, et vice versa.

Les variables économiques qui dirigent l'activité de distribution sont différentes de celles qui dirigent l'activité d'**expédition**. Le tonnage distribué reflète l'activité des grandes et moyennes surfaces, à l'échelle du département et du niveau des importations en produit frais et surgelés. Alors que le tonnage expédié est relatif à la santé des exportations nationales de la production locale.

Nous l'avons souligné, le niveau d'information est important pour les modèles économétriques. Ils nécessitent des informations précises et représentatives du phénomène observé. Certains indicateurs de conjoncture économique se trouvent sur des sites publics (INSEE ou de la Banque de France).

Le prévisionniste a, ainsi, fait l'inventaire d'un ensemble d'indicateurs économiques généraux externes pour expliquer le poids mensuel des marchandises transportées par une agence. A cet ensemble, il ajoute, d'une part, des facteurs exogènes internes à l'activité TFE. Ce sont des variables explicatives telles que la perte ou le gain d'un portefeuille client. Elles expliquent les ruptures de tendance. D'autre part, le prévisionniste complète le modèle par 12 variables binaires, en référence aux mois de l'année. Elles ont pour rôle d'extraire la saisonnalité mensuelle de la chronique.

La prévision de la valeur mensuelle s'écrit par l'équation de régression linéaire suivante :

$$Y_t = \sum_{i=1:m} a_i X_{i,t-1}^{eco} + \sum_{i=1:p} b_i X_{i,t-1}^{int} + \sum_{i=1:12} c_i X_{i,t}^{mois} + d + \varepsilon_t \quad (3.3)$$

Soit,

- $T$  est le nombre de valeurs mensuelles historisées,
- $t = 1 : T$  est le temps en mois,

- $Y_t$  est la valeur de la chronique à la date  $t$ ,
- $a$  est le vecteur  $(1 \times m)$  des coefficients des variables économiques externes,
- $b$  est le vecteur  $(1 \times p)$  des coefficients des variables économiques internes,
- $c$  est le vecteur  $(1 \times 12)$  des coefficients saisonniers mensuels,
- $X^{eco}$  est une matrice  $(m \times T)$  des  $m$  variables économiques externes,
- $X_{i,t-1}^{eco}$  est la valeur de la variable économique  $i$  à la date  $t - 1$ ,
- $X^{int}$  est une matrice  $(p \times T)$  des  $p$  variables économiques internes,
- $X_{i,t-1}^{int}$  est la valeur de la variable économique interne  $i$  à la date  $t - 1$ ,
- $X^{mois}$  est une matrice  $(12 \times T)$  des variables binaires relatives aux 12 mois de l'année,
- $X_{i,t}^{mois}$  est la valeur de la  $i^{\text{ème}}$  variable à la date  $t$ , de la matrice  $X^{mois}$ ,
- $d$  est une constante,
- $\varepsilon_t$  est un résidu.

La première somme de l'équation 3.3 représente l'effet des facteurs économiques externes à l'entreprise. La deuxième somme représente l'effet des facteurs économiques internes à l'entreprise. La troisième somme représente la saisonnalité mensuelle. Ainsi, la valeur de la chronique, à la date  $t$ , est une fonction des indicateurs économiques externes à la date  $t - 1$ , des indicateurs économiques internes à la date  $t - 1$  et du coefficient saisonnier mensuel à la date  $t$ . Le décalage temporel entre la valeur de la chronique à la date  $t$  et les valeurs des facteurs économiques à la date  $t - 1$  existent, car ces derniers ne sont pas encore connus à la date  $t$ . Par contre, le coefficient saisonnier à la date  $t$  est connu. Le prévisionniste estime que les valeurs des facteurs explicatifs en  $t - 1$  sont une bonne approximation des valeurs à la date  $t$ . Le décalage ainsi créé permet de prévoir à la date  $t$ , la valeur inconnue de la chronique à la date  $t + 1$ , alors que les valeurs des facteurs économiques à la date  $t + 1$  sont inconnues.

Les paramètres du modèle 3.3 sont estimés par la méthode des moindres carrés. Ils sont validés par le ratio de Student. Le ratio de Student est le rapport de l'estimation du coefficient de régression ( $\hat{a}_i$ ) et de son écart type ( $\sigma_{\hat{a}_i}$ ). Sous l'hypothèse de résidus gaussiens, ce ratio suit une loi de Student, ce qui permet de tester si le coefficient est significativement différent de 0 (voir cours d'économétrie de G. Chevillon (30)).

L'équation 3.3 rend une valeur mensuelle de la chronique à prévoir. L'étape suivante consiste à ventiler cette prévision pour aboutir à une valeur journalière. Pour ce faire, le prévisionniste répartit la valeur mensuelle de la chronique sur quatre semaines. Il calcule donc pour chaque mois de l'année, quatre poids hebdomadaires. Dans un deuxième temps, il répartit la valeur hebdomadaire ainsi obtenue sur six jours de la semaine (du

lundi au samedi). Les poids sont définis comme une moyenne des évènements passés.

$\beta_{i,m}$  est le poids moyen annuel de la semaine  $i$  dans le mois  $m$ .

$$\beta_{i,m} = \frac{1}{N} \sum_{t=1:N} \frac{y_{i,m,t}}{y_{m,t}} \quad (3.4)$$

- $i = 1, \dots, 4$  est le numéro de semaine du mois,
- $m = 1, \dots, 12$  est le numéro du mois de l'année,
- $N$  est le nombre d'années de l'historique,
- $y_{i,m,t}$  est la valeur de la chronique pour la semaine  $i$ , du mois  $m$ , de l'année  $t$ ,
- $y_{m,t}$  est la valeur de la chronique pour le mois  $m$ , de l'année  $t$ .

L'analyse des poids des jours montre qu'ils varient au cours des différents mois de l'année. Pour cette raison, le prévisionniste calcule des poids jour pour chacun des 12 mois de l'année.

$\psi_{j,m}$  est le poids moyen annuel du jour  $j$  dans le mois  $m$ .

$$\psi_{j,m} = \frac{1}{N} \sum_{t=1:N} \frac{y_{j,m,t}}{y_{m,t}} \quad (3.5)$$

- $j = \{\text{lundi, mardi, mercredi, jeudi, vendredi, samedi}\}$  est le jour de la semaine,
- $m = 1, \dots, 12$  est le numéro du mois de l'année,
- $N$  est le nombre d'années de l'historique,
- $y_{j,m,t}$  est la valeur de la chronique pour le du jour  $j$  du mois  $m$  de l'année  $t$ ,
- $y_{m,t}$  est la valeur de la chronique pour le mois  $m$  de l'année  $t$ .

Ainsi, la valeur prévue pour le jour  $j$  de la semaine  $i$  du mois  $m$  s'écrit :

$$\hat{y}_{j,i,m} = \hat{y}_m \times \beta_{i,m} \times \psi_{j,m} \quad (3.6)$$

avec,

- $\hat{y}_m$  est la prévision de la valeur de la chronique pour le mois  $m$  (voir formule 3.3),
- $\beta_{i,m}$  est le poids de la semaine  $i$  dans le mois  $m$  (voir formule 3.4),
- $\psi_{j,m}$  est le poids du jour  $j$  dans le mois  $m$  (voir formule 3.5).

Le schéma de la figure 3.6 illustre la décomposition de la chronique tel que nous venons de la formuler.

Les jours fériés ont aussi une forte influence sur l'activité d'une agence TFE (voir section 4.3 et 4.4.6.1). Le prévisionniste a dénombré 10 jours fériés (lundi de Pâques, jeudi de l'Ascension, Pentecôte, 14 juillet, 15 août, 1<sup>er</sup> novembre, 11 novembre, 1<sup>er</sup> mai, 8 mai,

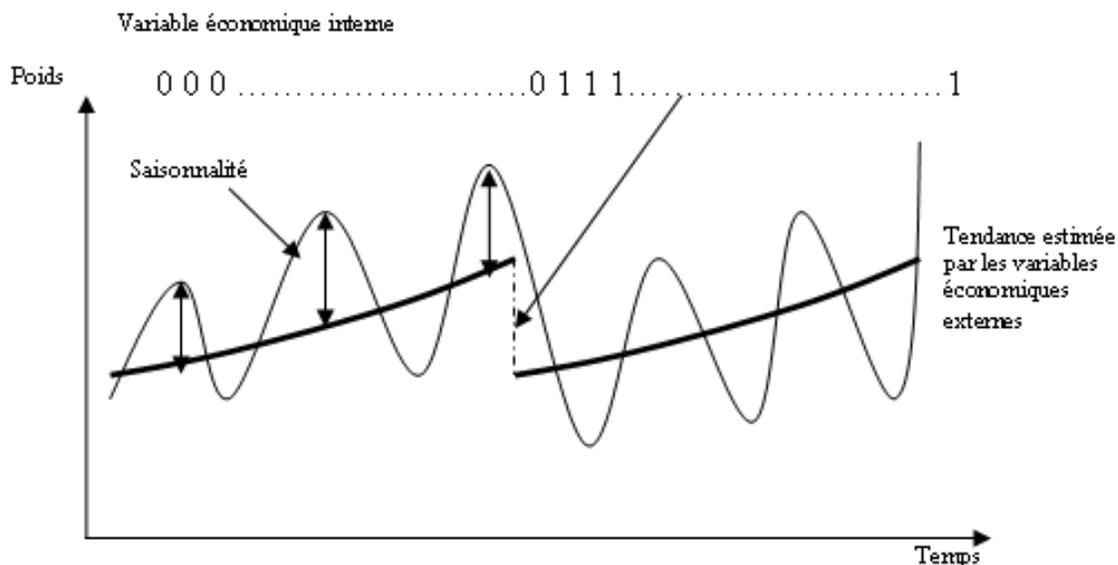


FIGURE 3.6 – Représentation de la tendance, la saisonnalité et l'effet d'une variable explicative

noël). Sachant que l'effet d'un 25 décembre n'est pas comparable à celui d'un jeudi de l'Ascension, il les traite indépendamment. Il considère, par ailleurs, qu'ils n'ont un effet que sur la semaine à laquelle ils appartiennent. Comme pour les jours « normaux », il calcule des poids hebdomadaires et journaliers pour les fériés.

$\beta_{i,f}$  est le poids moyen annuel de la semaine  $i$  comportant le férié  $f$  dans le mois  $m$ .

$$\beta_{i,f} = \frac{1}{N} \sum_{t=1:N} \frac{y_{i,f,t}}{y_{m,f,t}} \quad (3.7)$$

- $i = 1, \dots, 4$  est le numéro de semaine du mois à jour férié,
- $f = 1, \dots, 10$  est le numéro du jour férié,
- $m = 1, \dots, 12$  est le numéro du mois à jour férié,
- $N$  est le nombre d'années de l'historique,
- $y_{i,f,t}$  est la valeur de la chronique pour la semaine  $i$ , comportant le férié  $f$ , de l'année  $t$ ,
- $y_{m,f,t}$  est la valeur de la chronique pour le mois  $m$ , comportant le férié  $f$ , de l'année  $t$ .

Le prévisionniste pose l'hypothèse que la perte d'activité du jour férié est rattrapée par les autres jours de la semaine. Il en résulte une nouvelle répartition de la valeur hebdomadaire sur les 5 jours restants de la semaine. Selon le jour chômé de la semaine, cette

répartition ne sera pas la même.

$\psi_{j,i,f}$  est le poids moyen annuel du jour  $j$  dans la semaine  $i$  comportant le férié  $f$  dans le mois  $m$ .

$$\psi_{j,i,f} = \frac{1}{N} \sum_{t=1:N} \frac{y_{j,i,m,f,t}}{y_{m,f,i,t}} \quad (3.8)$$

- $j = \{\text{lundi, mardi, mercredi, jeudi, vendredi, samedi}\}$  est le jour de la semaine,
- $i = 1, \dots, 4$  est le numéro de semaine comportant le jour férié dans le mois  $m$ ,
- $f = 1, \dots, 10$  est le numéro du jour férié,
- $m = 1, \dots, 12$  est numéro du mois à jour férié,
- $N$  est le nombre d'années de l'historique,
- $y_{j,i,m,f,t}$  est la valeur de la chronique pour le jour  $j$ , de la semaine  $i$ , du mois  $m$ , comportant le férié  $f$ , de l'année  $t$ ,
- $y_{i,m,f,t}$  est la valeur de la chronique pour la semaine  $i$ , du mois  $m$ , comportant le férié  $f$ , de l'année  $t$ .

Le modèle économétrique s'écrit :

Si la valeur journalière à prévoir,  $\hat{y}_{j,i,m}$ , n'est pas un jour férié et n'est pas incluse dans une semaine à jour férié alors :

$$\hat{y}_{j,i,m} = \hat{y}_m \times \beta_{i,m} \times \psi_{j,i,m} \quad (3.9)$$

Si la valeur journalière à prévoir,  $\hat{y}_{j,i,m}$ , est incluse dans une semaine à jour férié alors :

$$\hat{y}_{j,i,m} = \hat{y}_m \times \beta_{i,f} \times \psi_{j,i,f} \quad (3.10)$$

Si la valeur à prévoir,  $\hat{y}_{j,i,m}$ , est celle d'un jour férié alors :

$$\hat{y}_{j,i,m} = 0 \quad (3.11)$$

Pour résumer, le concepteur de ce modèle combine variables externes et internes à l'entreprise. Les premières sont des indicateurs économiques généraux pour caractériser la tendance de la série chronologique à prévoir, les secondes sont destinées à maîtriser des fluctuations inhérentes à l'activité d'une agence TFE. L'équation de cette tendance permet d'estimer une valeur mensuelle à un horizon  $h$ . Cette valeur est ventilée sur quatre semaines par pondération par rapport à la moyenne des poids des semaines du même mois des années précédentes. De la même manière, il ventile les quatre valeurs qui en découlent en 24 ( $4 \times 6$ ) valeurs journalières. Pour prendre en compte l'effet des jours fériés, il sépare les semaines impactées de celles qui ne le sont pas, et les étudie séparément.

Avantages	Inconvénients
la prise en compte de variables internes pour expliquer les ruptures de tendance,	le manque de variables internes pour expliquer des événements exceptionnels au niveau de la semaine ou, plus bas, du jour,
le fait d'appréhender l'activité de TFE par des facteurs externes à faible variance,	les variables internes comme externes sont différentes suivant l'agence TFE pour laquelle le modèle est établi, (le poids mensuel transporté par l'agence de Nantes est corrélé à l'emploi salarié du transport du département « Loire-Atlantique » et non au niveau national), ce qui implique une importante recherche d'informations parfois payantes et des modèles de prévision différents par région,
l'anticipation de l'activité à long terme en surveillant les indicateurs économiques externes.	le processus de ventilation conduit à ne pas valoriser les valeurs les plus récentes, de la chronique à prévoir.

TABLE 3.1 – *Avantages et inconvénients du modèle économétrique de TFE*

### 3.2.3 Modèle endogène de TFE - lissage par moyennes mobiles

Un deuxième modèle développé chez TFE s'inspire des techniques descriptives « simples » (75). Il a été pensé pour prévoir l'activité des agences de la région Nord-Ouest. Le modèle statistique décompose la chronique en une somme de la tendance, d'un cycle et d'une saisonnalité. La prévision de la valeur hebdomadaire s'écrit selon le schéma additif suivant :

$$Y_t = T_t + C_t + S_t + \varepsilon_t \quad (3.12)$$

Ainsi, l'unité de temps de la chronique est la semaine. La valeur hebdomadaire prévue,  $\hat{y}_t$ , est ensuite ventilée sur 6 jours ouvrables.

#### 3.2.3.1 Prévision de la chronique à valeurs hebdomadaires

Soit  $y$  l'historique, depuis 2002, des valeurs de la chronique hebdomadaire à prévoir. Le prévisionniste pose l'hypothèse d'une saisonnalité de 52 semaines, ce qui l'amène à calculer 52 coefficients saisonniers. Ils sont déterminés par la moyenne des différences entre les moyennes mobiles et les valeurs réelles de la chronique à prévoir. Les étapes du calcul de la prévision d'une valeur hebdomadaire de la chronique sont les suivantes :

- La moyenne mobile est une estimation de la tendance. Différentes techniques de calcul sont envisageables selon ce que nous cherchons à montrer. Par exemple, une moyenne glissante sur les 52 dernières périodes lisse la série en effaçant l'effet de la saisonnalité.

$$MM_s = \frac{1}{52} \sum_{i=(s-52)}^s y_i \quad (3.13)$$

$y_i$  est la valeur réelle de la chronique à la semaine  $i$ .

Pour effacer l'effet que peut avoir une semaine sur la suivante, le prévisionniste ajoute une moyenne mobile de deux périodes.

$$MMC_{s-1} = \frac{MM_{s-1} + MM_s}{2} \quad (3.14)$$

- Les coefficients saisonniers permettent d'estimer la saisonnalité sur l'année, de la chronique à prévoir. Dans le cas d'un schéma additif (voir section 2.3.3), si le coefficient saisonnier d'une semaine  $s$  est supérieur à 0 alors la valeur moyenne de la chronique pour la semaine  $s$  est supérieure à la valeur moyenne des valeurs de la chronique comprises dans l'intervalle  $[s - 52; s + 52]$ . Inversement si la valeur du coefficient saisonnier d'une semaine  $s$  est négatif, c'est qu'en moyenne la valeur de la semaine  $s$  est inférieur à la moyenne de l'ensemble des valeurs de la chronique comprises dans l'intervalle  $[s - 52; s + 52]$ . Les étapes du calcul sont :

- l'écart entre la série observée et la moyenne mobile :

$$e_i = y_i - MMC_i \quad (3.15)$$

$i$  est un indice pour parcourir toute les valeurs de la chronique à prévoir.

- la moyenne des écarts relatifs à la même semaine :

$$\bar{e}_s = \frac{1}{N} \sum_{t=1}^N e_{s,t} \quad (3.16)$$

$s$  est l'indice des semaines  $s = 1, \dots, 52$ ,  $t$  est le numéro de l'année dans l'historique de la chronique et  $N$  est le nombre d'années de ce même historique.

- le calcul de la somme des moyennes des écarts :

$$S_s = \bar{e}_s - \frac{S}{52} \quad (3.17)$$

- le calcul des coefficients saisonniers, en vertu du principe de conservation des aires, la somme des fluctuations saisonnières (coefficients saisonniers) est nulle sur une année. Ce qui veut dire que le prévisionniste normalise des écarts moyens :

$$S = \sum_{s=1}^{52} \bar{e}_s \quad (3.18)$$

– La tendance correspond à la série des moyennes mobiles.

$$T_i = MMC_i \quad (3.19)$$

– Le cycle est une saisonnalité dont la périodicité dépasse l'année. Il est obtenu par la différence entre la tendance et la moyenne mobile.

$$C_i = MMC_i - T_i \quad (3.20)$$

– La prévision d'une valeur hebdomadaire de la chronique est le résultat de la combinaison entre la saisonnalité, la tendance et le cycle.

$$\hat{y}_{n+1} = S_s + T_{n+1} + C_{n-2} \quad (3.21)$$

Soit  $y_n$  la dernière valeur connue de la chronique et  $y_{n+1}$  la valeur à prévoir.  $\hat{y}_{n+1}$  est une prévision de  $y_{n+1}$ .  $S_{n+1}$  correspond au coefficient saisonnier de la semaine relative à  $n+1$ ,  $T_{n+1}$  est l'extrapolation de la tendance  $T_{n+1} = T_n + (T_n - T_{n-1})$ .  $C_{n-2}$  est la dernière valeur connue du cycle. Le décalage de 2 unités de temps est celui retrouvé pendant le calcul de la moyenne mobile additionné à celui de l'horizon de prévision ( $h = 1$ ).

### 3.2.3.2 Redistribution sur 6 jours

La valeur prédite d'une semaine est ventilée sur les 6 jours ouvrables (du lundi au samedi). Les coefficients de pondération sont estimés en fonction des événements passés. La méthodologie utilisée se résume en 5 étapes :

– Etape 1

Connaissant la valeur réelle quotidienne, le prévisionniste calcul le ratio entre cette dernière et celle de la semaine associée.  $y_{a,s}$  est la valeur réelle que prend la chronique, hebdomadaire, la semaine  $s$  de l'année  $a$ .  $y_{a,s}^j$  est la valeur réelle que prend la chronique, journalière, le jour  $j$  de la semaine  $s$  de l'année  $a$ .

$$y_{a,s} = y_{a,s}^{lundi} + \dots + y_{a,s}^{samedi} \quad (3.22)$$

$s = \{1, \dots, 52\}$  est le numéro de la semaine dans l'année  $a$ ,  $a = \{2002, \dots, 2007\}$ .

$R_{a,s}^j$  est le poids de la valeur prise par la chronique, le jour  $j$  de la semaine  $s$  de l'année  $a$ , dans la semaine  $s$  de l'année  $a$ .

$$R_{a,s}^j = \frac{y_{a,s}^j}{y_{a,s}} \quad (3.23)$$

$j = \{lundi, \dots, samedi\}$  est le jour de la semaine.

– Etape 2

Pour chaque année de l'historique, le prévisionniste mesure une moyenne trimestrielle des quotients calculés à l'étape 1. Il distingue les semaines « normales » (voir tableau 3.2) des semaines encadrant ou incluant un jour férié (voir tableau 3.3).

$R_{a,t}^j$  est le poids moyen dans le trimestre  $t$  de l'année  $a$  du jour  $j$  dans une semaine « normale ».

$$R_{a,t}^j = \frac{1}{3 \times 4} \sum_{s=1}^{3 \times 4} R_{a,s}^j \quad (3.24)$$

Dans un trimestre  $t$ ,  $t = \{1, \dots, 4\}$ , le prévisionniste compte  $3 \times 4$  semaines « normales » et 6 jours dans une semaine,  $j = \{\text{lundi}, \dots, \text{samedi}\}$ .

Le tableau 3.2 illustre un exemple des poids calculés avec la formule 3.24.

		Lundi	Mardi	Mercredi	Jeudi	Vendredi	Samedi
2003	Trim 4	17,21%	19,48%	18,75%	20,46%	17,56%	6,54%
2004	Trim 4	17,62%	18,51%	18,11%	19,11%	17,84%	8,81%
2005	Trim 1	18,84%	18,08%	19,24%	17,19%	18,18%	8,46%
2005	Trim 2	17,69%	17,89%	19,68%	17,53%	18,49%	8,72%
2005	Trim 3	17,06%	17,67%	19,62%	17,73%	18,57%	9,35%
2005	Trim 4	17,86%	18,17%	19,02%	17,69%	18,44%	8,82%
2006	Trim 1	17,88%	17,41%	19,81%	16,82%	19,09%	8,99%
2006	Trim 2	17,87%	16,84%	19,49%	16,99%	19,69%	9,12%
2006	Trim 3	17,71%	17,40%	19,19%	17,73%	18,61%	9,36%
2006	Trim 4	18,04%	17,11%	19,29%	17,35%	19,17%	9,04%

TABLE 3.2 – Exemple des poids jours calculés par trimestre

Les poids des jours dans les semaines à jours fériés sont calculés non plus au trimestre, mais sur tout l'historique.

$$RF_f^j = \frac{\sum_{k=\text{semaine à jour } f \text{ férié}} R_k^j}{\text{nb de semaines à jours } f \text{ fériés}} \quad (3.25)$$

$RF_f^j$  est le poids du jour  $j$  dans la semaine  $f$  qui a comme férié le jour  $f$ .

Le tableau 3.3 illustre les poids calculés avec la formule 3.25. Remarquons que la diagonale correspond aux jours fériés. Par exemple, la valeur prévue, par la formule 3.21, pour une semaine dont le jeudi est férié, sera multipliée par 0,2242 pour obtenir une prévision du lundi, par 0,2519 pour obtenir une prévision du mardi, etc.

	Lundi	Mardi	Mercredi	Jeudi	Vendredi	Samedi
Lundi	<b>1,12%</b>	21,35%	23,96%	23,05%	21,93%	8,60%
Mardi	21,17%	<b>0,95%</b>	23,07%	22,52%	23,07%	9,21%
Mercredi	22,46%	23,14%	<b>1,57%</b>	22,23%	22,76%	7,85%
Jeudi	22,42%	25,19%	21,72%	<b>1,47%</b>	20,82%	8,37%
Vendredi	20,41%	23,64%	25,44%	20,75%	<b>0,82%</b>	8,95%
Samedi	20,12%	20,44%	21,68%	20,65%	16,00%	<b>1,10%</b>

TABLE 3.3 – Exemple des poids jours des semaines avec jour férié

Remarquons que le prévisionniste ne distingue pas, dans le calcul des poids journaliers, les semaines voisinant une semaine à jour férié. Les semaines qui encadrent une semaine à jour férié se voient attribuer les mêmes poids journaliers que les semaines normales. Néanmoins, les semaines qui encadrent une semaine à jour férié sont prises en compte dans le modèle à l'étape suivante.

– Etape 3

Le prévisionniste corrige la prévision hebdomadaire dans le cas où la semaine prévue contient un jour férié ou est voisine d'une semaine à jour férié. La règle à suivre est la suivante :

- si le jour férié de la semaine à prévoir est un lundi, la valeur prévue est multipliée par 0,9. Le coefficient correcteur est noté  $CY = 0,9$  ;
- si la semaine à prévoir précède une semaine dont le jour férié tombe un lundi, la valeur prévue est multipliée par 1,05,  $CY = 1,05$
- sinon,  $CY = 1$ .

– Etape 4

Le prévisionniste dispose d'une série de variables représentant des événements exceptionnels (campagne promotionnelle, gain ou perte de dossier, ...). Les variables seront notées  $VE$ . S'il le juge nécessaire, il additionne, ou soustrait, la valeur apportée par ces événements à la valeur prévue.

– Etape 5

Le prévisionniste corrige la valeur hebdomadaire prévue suivant les étapes 3 et 4, puis avec l'aide des coefficients de l'étape 2, il la ventile sur les 6 jours ouvrables.

$$\hat{y}_{n+1} = \hat{y}_{n+1} \times CY_{n+1} + VE_n \quad (3.26)$$

$\hat{y}_{n+1}$  est la valeur corrigée de la prévision obtenue par la formule 3.21.

La ventilation de la valeur prévu  $\hat{y}_{n+1}$  sur les 6 jours de la semaine, est :

- o Dans le cas d'une semaine normale,

$$\hat{y}_{j,n+1} = \hat{y}_{n+1} \times R_{a,t}^j \quad (3.27)$$

- o Dans le cas d'une semaine incluant un jour férié, le jour  $f$ ,

$$\hat{y}_{j,n+1} = \hat{y}_{n+1} \times RF_f^j \quad (3.28)$$

$\hat{y}_{j,n+1}$  est la prévision du jour  $j$  de la semaine  $n + 1$ ,

$\hat{y}_{n+1}$  est la prévision de la semaine  $n + 1$ ,

$R_{a,t}^j$  est le poids moyen dans le trimestre  $t$  de l'année  $a$  du jour  $j$  dans une semaine  $n + 1$ .

$RF_f^j$  est le poids du jour  $j$  dans la semaine  $n + 1$ , qui a comme férié le jour  $f$ .

Les avantages et inconvénients de ce modèle sont :

Avantages	Inconvénients
facilité de la compréhension de l'activité d'une plate-forme,	non prise en compte de l'effet du jour férié suivant que c'est un lundi de Pentecôte ou un Noël.
mise en évidence des périodes critiques et difficiles à prévoir,	
le bruit du modèle pointe l'influence des phénomènes extérieurs.	

TABLE 3.4 – Avantages et inconvénients du modèle endogène de TFE - lissage par moyennes mobiles

### 3.2.4 Modèle endogène de TFE - méthode de Holt-Winters

Ce modèle a été pensé et développé pour les agences de la région Sud-Ouest. Comme pour le modèle précédent, l'unité de temps de la série chronologique est la semaine. Le prévisionniste fait l'hypothèse d'une saisonnalité de 52 semaines. La prévision journalière découle de la prévision hebdomadaire par l'utilisation de coefficients de répartition.

A partir de la série hebdomadaire, le prévisionniste utilise les formules de lissage exponentiel proposées par Holt-Winters (63; 122). Il décide d'utiliser un schéma additif. En

effet, de manière générale sur les agences étudiées, l'amplitude entre les poids minimums (le samedi) et les poids maximums ne varie pas avec le temps.

La méthode de Holt-Winters améliore la prévision par lissage exponentiel. C'est une technique dite « autoprojective ». Le principe est de privilégier les observations les plus récentes par l'utilisation de coefficients de pondération. D'une manière récursive, les formules de lissage sont mises à jour et permettent d'obtenir une prévision. Dans le cas du lissage exponentiel simple la formule de lissage a pour but d'estimer la tendance sur les  $n$  dernières observations. Suivant l'importance donnée aux observations les plus récentes par rapport aux plus anciennes, elles se voient affecter des poids. A l'estimation de la tendance, Holt-Winters introduit des coefficients pour modéliser la saisonnalité.

Dans le cas d'un schéma additif avec saisonnalité, les formules de lissage s'écrivent :  
Le lissage de la moyenne,

$$a1_{s,a} = \alpha((y_{s,a} - VC_{s,a}) - S_{s,a-1}) + (1 - \alpha)(a1_{s-1,a} + a2_{s-1,a}) \quad (3.29)$$

Le lissage de la tendance,

$$a2_{s,a} = \beta(a1_{s,a} - a1_{s-1,a}) + (1 - \beta)a2_{s-1,a} \quad (3.30)$$

Le lissage de la saisonnalité,

$$S_{s,a} = \gamma((y_{s,a} - VC_{s,a}) - a1_{s,a}) + (1 - \gamma)S_{s,a-1} \quad (3.31)$$

avec

- $y_{s,a}$  est la valeur prise par la chronique la semaine  $s$  de l'année  $a$ ,
- $\hat{y}_{s,a}$  est l'estimation de la valeur prise par la chronique la semaine  $s$  de l'année  $a$ ,
- $\bar{y}_a$  est la moyenne des valeurs hebdomadaires prises par la chronique l'année  $a$ ,
- $VC_{s,a}$  est la variable de correction de la semaine  $s$  de l'année  $a$ ,
- $Pe$  est la période (ici 52 semaines ou une année)
- $h$  est l'horizon de prévision en nombre de semaines (inférieur à un an),
- $\beta, \alpha, \gamma$  : les constantes de lissage.

Le choix des paramètres de lissage est arbitraire. Le paramètre  $\alpha$  conditionne les résultats de prévision. Si nous voulons une prévision réactive, c'est-à-dire qui accorde une plus grande importance au passé récent, nous choisirons un paramètre  $\alpha$ , proche de 1. Les paramètres  $\beta$  et  $\gamma$  ont la même propriété. Le prévisionniste de TFE a choisi d'octroyer de l'importance aux données lointaines ( $\alpha$  proche de 0), d'être plus réactif sur le coefficient de lissage de la tendance ( $\beta$  proche de 0.4) et sur celui du coefficient de saisonnalité ( $\gamma = 0.2$ ).

L'initialisation des valeurs initiales permet de débiter le calcul récursif à partir de la deuxième année.

$$a1_{s=1,a=2} = \bar{y}_{a=1} \quad (3.32)$$

$$a2_{s=1,a=2} = 0 \quad (3.33)$$

$$S_{s=1,a=2} = y_{s=1,a=2} - \bar{y}_{a=1} \quad (3.34)$$

La prévision à partir de la semaine  $s$  pour un horizon  $h$  s'écrit :

$$\hat{y}_{s+h,a} = a1_{s+h,a-1} + h \times a2_{s+h,a-1} + S_{s+h,a-1} + VC_{s+h,a} \quad (3.35)$$

Ayant obtenu une prévision à la semaine,  $\hat{y}_{s+h,a}$ , le prévisionniste cherche des coefficients journaliers de répartition.

$$\phi_j = \frac{\sum_{i=1}^n y_{j,i}}{\sum_{i=1}^n y_i} \quad (3.36)$$

$n$  est le nombre de valeurs de la chronique journalière. Rappelons que pour l'une ou l'autre des six séries temporelles décrivant l'activité des agences de transport TFE (voir section 3.2),  $n$  est approximativement égale à 6 jours  $\times$  52 semaines  $\times$  7 ans d'historique, soit 2.184 valeurs.  $\phi_j$  est poids du jour  $j$ , ( $j = \{lundi, \dots, samedi\}$ ) dans la semaine. Il est égal à la somme des valeurs prises par la chronique du jour  $j$  rapportée à la somme des valeurs de la chronique journalière. Ne sont pris en compte que les valeurs des semaines sans jours fériés.

Les poids des jours des semaines à jours fériés sont calculés de la même façon, mais en ne prenant en compte que les semaines ayant comme jour férié celui étudié. Ainsi, le prévisionniste aboutit au tableau 3.5.

Férié	Lundi	Mardi	Mercredi	Jeudi	Vendredi	Samedi
Aucun	17,23%	21,05%	19,34%	19,17%	16,79%	6,42%
Lundi	0,00%	23,29%	24,54%	23,16%	19,72%	9,29%
Mardi	20,40%	0,00%	25,52%	24,54%	20,15%	9,39%
Mercredi	23,80%	26,00%	0,00%	21,31%	18,02%	10,88%
Jeudi	22,08%	26,87%	24,74%	0,00%	17,92%	8,39%
Vendredi	19,95%	24,78%	26,35%	21,02%	0,00%	7,89%
Samedi	19,72%	21,42%	21,35%	18,49%	19,02%	0,00%

TABLE 3.5 – Coefficients journaliers du modèle endogène de TFE - méthode de Holt-Winters

La ventilation de la prévision hebdomadaire sur les 6 jours ouvrés de la semaine devient :

$$\hat{y}_{j,s+h,a} = \hat{y}_{s+h,a} \times \phi_j \quad (3.37)$$

$\hat{y}_{j,s+h,a}$  est la prévision du jour  $j$  de la semaine  $s + h$  de l'année  $a$ . C'est le résultat d'une ventilation de la prévision hebdomadaire,  $\hat{y}_{s+h,a}$  (3.35), par le poids,  $\phi_j$  (3.36).

Les avantages et inconvénients de ce modèle sont :

Avantages	Inconvénients
une grande facilité de mise en œuvre due à des calculs simples,	non prise en compte de l'effet du jour férié suivant que c'est un lundi de Pentecôte ou un Noël,
une réactivité aux événements récents si $\alpha \approx 1$ ,	le choix arbitraire des coefficients n'est pas facile et influence fortement les résultats,
une facilité d'implémentation sur tableur.	non prise en compte de l'effet des événements extérieurs,
	système réagit avec un temps de retard à une modification de la chronique (« course poursuite » entre réalisation et prévision quand $\alpha \approx 1$ ) (24).

TABLE 3.6 – Avantages et inconvénients du modèle endogène de TFE - méthode de Holt-Winters

### 3.2.5 Outil informatique utilisé

Ces modèles ont été ajusté et implanté avec l'aide d'un tableur. Or le tableur est peu adapté aux tâches de prévision en raison de ses limites,

- d'automatisation des traitements,
- de mise à jour des données,
- de changement de scénario,
- de contrôles des données,
- de rapidité.

Ainsi, les modèles statistiques n'ont pas pu être généralisés et utilisés sur un serveur pour déployer les résultats sur l'ensemble des agences. D'après une étude de « mode-IEdition SA » réalisée auprès de 41 grosses et moyennes entreprises, 64% utilisent un

tableur pour leurs prévisions et seulement 11% utilisent des outils de simulation spécialisés (voir section 2.3.4.3). Pourtant, les prises de décisions pourraient être sensiblement améliorées par ces derniers. C'est le choix fait par le projet de recherche « Horizons ». Il a permis d'homogénéiser la définition du calcul de l'historique des chroniques à prévoir, de centraliser les traitements sur un serveur et de diffuser l'information prévisionnelle sur l'ensemble du réseau.

# Chapitre 4

## Le système de prévision Horizons : Conception, développement et mise en œuvre

Le challenge du projet « Horizons » est de proposer un modèle statistique de prévision de l'activité du **quai** adapté à toutes les **agences** TFE, et d'implémenter ce modèle au travers d'une application informatisée pour l'intégrer dans la chaîne opérationnelle. C'est un pas vers une approche concentrée, c'est-à-dire que les prévisions d'activité seront centralisées dans un seul département. Pour réduire les coûts de licences, les outils informatiques utilisés doivent être open source. Il faut aussi convaincre l'ensemble des dirigeants du bien-fondé du projet. Il faut gagner la confiance des hommes et des femmes du terrain pour qu'ils utilisent le système de prévision proposé.

### 4.1 Nouvelle organisation de la prévision chez TFE

La section **3.2** rappelle qu'avant le projet Horizons, l'organisation de la prévision d'activité s'est faite suivant une approche indépendante et non concentrée (voir section **2.3.2**). Pour harmoniser les procédures, TFE a choisi de créer une cellule de prévision centralisée au siège. Ce choix cherche à répondre aux préconisations mentionnées dans la section **2.3.2**. L'avantage de la centralisation est de maîtriser les effectifs, de limiter les achats de licences logiciels, d'harmoniser l'extraction des données (voir section **4.2**), de gérer une seule application (voir section **4.7**) et de diffuser par client léger le résultat des recherches (voir section **4.7.3**). Un système de prévision ainsi centralisé autorise

une interaction avec le système d'information, une comparaison des résultats entre sites, l'accès rapide à l'information et sa correction, si nécessaire, en une seule fois. Pour être au cœur du système d'information, la cellule de prévision est placée physiquement à côté du service d'information décisionnel. Le service d'information décisionnel propose, crée et diffuse les indicateurs pertinents, permettant aux équipes managériales et opérationnelles de prendre les bonnes décisions et ainsi, de piloter leur activité sereinement. Ce service est le plus à même de fournir, en toute impartialité, l'information dont la cellule de prévision a besoin. Le service d'information décisionnel fournit à la cellule de prévision un accès sans limites à l'entrepôt de données et aux outils permettant de l'interroger (voir section 3.1.1). La cellule de prévision a pour mission :

- d'extraire les données historiques décrivant l'activité à prévoir,
- de chercher un modèle mathématique de prévision unique, compatible à l'activité des 70 sites de transport,
- de restituer de façon compréhensible les résultats prévisionnels,
- d'automatiser les procédures (extraction, transformation, prévision, indicateur de qualité, restitution),
- de mettre à jour quotidiennement les données,
- de former les managers à l'utilisation du système de prévision.

Ces missions sont développées dans les sections suivantes.

## 4.2 Harmonisation et centralisation de l'information

La procédure d'uniformisation des données est développée suivant la définition donnée à la valeur à prévoir. Rappelons qu'un objectif du projet de recherche « Horizons » est de prévoir les chroniques suivantes (voir section 1.2.4) :

- le nombre quotidien de **position**s et le poids des marchandises passées à quai,
- le nombre quotidien de positions et le poids des marchandises passées à quai et à destination d'une **tournée** de distribution (**livraison**),
- le nombre quotidien de positions et le poids des marchandises passées à quai et à destination d'une tournée d'**expédition**.

Pour répondre aux attentes de planification des ressources, il faut aussi prévoir :

- le nombre total d'heures productives quotidiennes sur le quai,
- le nombre d'heures productives quotidiennes sur le quai pour traiter les marchandises destinées à la distribution,

- le nombre d’heures productives quotidiennes sur le quai pour traiter les marchandises destinées à l’expédition.

Ces 9 séries chronologiques sont communes aux 70 agences de transport du groupe STEF-TFE. Le modèle statistique de prévision se focalise sur les 6 premières chroniques. Ce qui signifie que le modèle statistique doit convenir à la prévision de 420 ( $6 \times 70$ ) chroniques au niveau du groupe. La prévision des trois dernières chroniques, relatives aux heures productives par agence, est déduite de la prévision du poids quotidien des marchandises passées à **quai**.

Les valeurs prises par ces 9 séries chronologiques ont une définition qui a évolué entre le début du projet et sa mise en production. Jusqu’en Janvier 2008, le transporteur n’avait pas de définition commune pour l’ensemble des agences de ce qu’étaient le poids en distribution et celui en expédition. Les agences avaient toutes des procédures d’extraction du poids et du nombre de positions en expédition et en distribution différentes. Il était ainsi impossible de comparer des agences entre elles. Il semblait également impossible d’inventorier les différentes règles d’extraction et de les développer toutes pour coller aux besoins spécifiques de chacune des agences.

Au cours de l’année 2007 se développe au sein du transporteur, une politique d’harmonisation des outils et des processus métiers. Entre 2007 et 2009, toutes les agences du transporteur migrent leur TMS de HP vers GTI (voir section **3.1.1**). Ce nouveau TMS impose à tous de travailler de façon uniforme. La Direction d’Exploitation (DEX) pose alors en Janvier 2008 une définition commune des indicateurs en expédition et en distribution (voir section **3.1.2**). Les indicateurs sont calculés quotidiennement et ont pour vocation d’analyser l’exploitation de façon quotidienne et mensuelle. Le système de prévision peut alors devenir un outil métier à l’envergure du groupe. Non seulement il prévoit une partie des indicateurs définis par la DEX, mais il permet également d’analyser leur évolution à travers le temps et de comparer leur performance selon les agences. Il faut tout de même garder à l’esprit que les agences de transport du groupe STEF-TFE ne sont pas toutes comparables, leur taille et leur activité varient. Par contre l’évolution dans le temps des indicateurs est comparable entre agences.

Nous avons souligné que la base de données GTI est alimentée depuis 2007. Les données d’avant 2007 sont à récolter dans la base de données HP3000. Les tables issues de la base de données GTI et celles issues de HP3000 ne sont pas identiques. C’est pourquoi les procédures d’extraction et de transformation des données, pour coller à définition des indicateurs à prévoir, donnée par la DEX, sont différentes selon les bases de données sources (HP3000 ou GTI).

Dans le but de centraliser l’information, les tables recueillant l’information nécessaire à la prévision des indicateurs d’activité sont stockées dans l’entrepôt de données du groupe

STEF-TFE. Pour accueillir l'historique des données à prévoir, trois nouvelles tables sont créées dans l'infocentre. La table HORIZONS (voir table 4.1) contient un enregistrement quotidien par site de transport. Ne sont présents que les sites pour lesquels il a été possible d'extraire un historique d'au moins trois ans. Néanmoins, la première date de la table est le 1<sup>er</sup> Janvier 2000, car certaines agences cumulent des données depuis cette date. Les données historisées sont le poids et le nombre de positions en expédition et en distribution passées à quai. La somme des deux donne le total du poids ou du nombre de positions passées à quai. La table HORIZONS ajoute à chacune de ces informations deux informations supplémentaires : la prévision et la correction de la prévision, apportée par l'utilisateur, le cas échéant. Enfin, elle historise le motif de la correction apportée (voir section 4.7.3).

La table AGE\_HORIZONS (voir table 4.1) contient un enregistrement par agence. Elle permet d'obtenir, à partir de son code, le libellé d'une agence en toutes lettres. Elle conserve, pour chacune des variables à prévoir, les coefficients pondérateurs utilisés dans le modèle mathématique de prévision, pour combiner les prévisions issues des différentes méthodes (voir section 4.4.11). Elle recueille également un champ permettant d'activer, ou non, l'agence pour qu'elle soit visible sur l'interface web.

La table HQUAI\_HORIZONS (voir table 4.1) rassemble les informations sur les heures productives du quai. Elle est construite de la même façon que la table HORIZONS mais dans le but de répondre à la prévision du nombre d'heures productives et du taux de productivité (voir section 3.1.3).

### 4.3 Modèles de prévision proposés en analyse statistique

De nombreuses méthodes statistiques de prévision ont été étudiées avant de trouver celle qui convenait le mieux. Pour être dans la continuité de ce que l'entreprise de transport avait déjà réalisé en la matière, il va de soi que les modèles de la section 3.2 ont servi de base pour chercher à ne garder que les avantages et tenter d'éliminer les inconvénients. Rappelons que ces modèles utilisent les techniques de :

- décomposition de la série chronologique en tendance, saisonnalité, erreur,
- lissage exponentiel,
- moyenne mobile.

Ce sont les méthodes les plus répandues dans les ouvrages statistiques. Elles sont également mentionnées par G. Mélard (90). Il propose un tour d'horizon des méthodes statistiques utilisées par les entreprises dans le but de prévoir les ventes, « *le rendement d'un investissement, la pénétration d'un marché ou l'effet du passage aux 35 heures* », à court et moyen terme. Il rappelle les méthodes endogènes et exogènes, qu'il appelle

<b>HORIZONS</b>	<b>AGE_HORIZONS</b>	<b>HQUAI_HORIZONS</b>
DDATE	AGENCE	DDATE
AGENCE	COEFFDISTPDS	AGENCE
ANNEE	COEFFEXPPDS	ANNEE
MOIS	COEFFDISTNBPOS	MOIS
JDUMOIS	COEFFEXPNBPOS	JDUMOIS
JDESEM	COEFFTOTPDS	JDESEM
SDEANNEE	COEFFTOTNBPOS	SDEANNEE
EXPPDS_PAST	OK	H_QUAIDIST
EXPPDS_REEL	DESCRIPTION	NBMAT_QUAIDIST
EXPPDS_PREV	CNUIPDS_CHG	H_QUAIEXPE
EXPPDS_CORR	CJOUPDS_CHG	NBMAT_QUAIEXPE
DISTPDS_PAST	CTOTPDS_CHG	TXH_PRODDIST
DISTPDS_REEL	CJOUNBPOS_CHG	TXH_PRODEXPE
DISTPDS_PREV	CNUINBPOS_CHG	H_QUAIDIST_PLAN
DISTPDS_CORR	CTOTNBPOS_CHG	H_QUAIEXPE_PLAN
TOTPDS_PAST	CNUIPDS_DECHG	TXM_PRODDIST
TOTPDS_REEL	CJOUPDS_DECHG	TXM_PRODEXPE
TOTPDS_PREV	CTOTPDS_DECHG	NBMAT_QUAIDIST_PLAN
TOTPDS_CORR	CJOUNBPOS_DECHG	NBMAT_QUAIEXPE_PLAN
EXPNBPOS_PAST	CNUINBPOS_DECHG	PREV_TX_PRODDIST
EXPNBPOS_REEL	CTOTNBPOS_DECHG	PREV_NBMATDIST
EXPNBPOS_PREV		PREV_TX_PRODEXPE
EXPNBPOS_CORR		PREV_NBMATEXPE
DISTNBPOS_PAST		H_QUAITOT
DISTNBPOS_REEL		NBMAT_QUAITOT
DISTNBPOS_PREV		TXH_PRODTOT
DISTNBPOS_CORR		H_QUAITOT_PLAN
TOTNBPOS_PAST		TXM_PRODTOT
TOTNBPOS_REEL		NBMAT_QUAITOT_PLAN
TOTNBPOS_PREV		PREV_TX_PRODTOT
TOTNBPOS_CORR		PREV_NBMATTOT
MOTIF_CORR		

TABLE 4.1 – Tables créées dans l'Infocentre pour le besoin du projet Horizons

« méthodes extrapolatives » et « méthodes explicatives ». Comme méthodes les plus fréquemment utilisées, il cite le lissage exponentiel, la moyenne mobile d'ordre  $k$ , le modèle de Brown et Holt-Winters, la régression linéaire multiple, la méthode de Box et Jenkins (ARIMA) et l'analyse spectrale.

Avant de commenter certaines de ces méthodes, rappelons brièvement la problématique. L'entreprise de transport souhaite trouver un modèle de prévision à horizon de 45 jours pour six séries temporelles : poids et nombre de lettres de voiture en expédition, distribution et total passés à quai. Nous constaterons par la suite que la chronique du nombre de lettres de voiture journalier suit approximativement la même saisonnalité que celle du poids transporté. L'entreprise de transport cherche à se doter d'un outil de prévision unique s'adaptant à toutes les agences de son réseau, et qui soit complètement automatisé. Le modèle doit fournir des résultats dans un temps d'exécution jugé raisonnable. L'écart type des erreurs de prévision ne doit pas dépasser 10% de la valeur prédite. L'entreprise ne souhaite pas investir dans l'achat de données externes pouvant servir de facteurs explicatifs de l'activité. Il est prouvé que les jours fériés ont une influence déterministe et quantifiable sur l'activité. Les séries chronologiques discrètes à prévoir ont au moins 3 ans d'historique. L'unité d'observation est la journée.

Suite à ces contraintes, le modèle exogène seul était exclu. Les facteurs susceptibles d'expliquer l'activité d'une agence du Nord de la France sont différents de ceux qui expliquent l'activité d'une agence située dans l'ouest ou le sud. Il y aurait donc fallu écrire plusieurs modèles statistiques de prévision. D'autre part, ces facteurs sont liés soit aux clients (promotion des ventes), soit à l'activité économique du marché (indicateurs INSEE). Dans un cas comme dans l'autre, les données sont onéreuses et le risque qu'elles ne soient pas mises à jour à temps les rend inutilisables.

Ayant, à peu de chose près, comme unique information l'historique de la chronique à prévoir, il est légitime de vouloir se retourner vers les modèles « Auto-Regressive - Integrated - Moving Average » ARIMA (21). Les méthodes pour utiliser ces modèles (AR, MA, ARMA, ARIMA, SARIMA) ont été développées par Box et Jenkins, en 1976. Les modèles sont appelés autorégressifs car ils cherchent à déterminer une valeur de la série, en fonction d'une combinaison linéaire des valeurs qui la précèdent. Le cours (35), écrit par J.J. Daudin, C. Duby, S. Robin et P. Trécourt, explique que pour appliquer les processus AR et MA il faut que la série chronologique soit stationnaire. Une suite de variables aléatoires est stationnaire quand leurs moyennes et leurs covariances sont constantes dans le temps.

*Soit un processus temporel à valeurs réelles et en temps discret  $Z_1, Z_2, \dots, Z_t$ . Il est dit stationnaire de second ordre si*

$$\begin{aligned} E(Z_i) &= \mu & \forall i = 1 \dots t \\ \text{Var}(Z_i) &= \sigma^2 & \forall i = 1 \dots t \\ \text{Cov}(Z_i, Z_{i-k}) &= f(k) & \forall i = 1 \dots t \text{ et } \forall k = 1 \dots t \end{aligned}$$

La tendance et la saisonnalité sont des facteurs de non stationnarité. Un processus d'intégration permet d'éliminer la tendance par différenciation. La saisonnalité, quant à elle, peut être éliminée par différenciation saisonnière (116).

Bien que ces modèles soient fréquemment utilisés pour prévoir des séries économiques (99; 24; 117), ils ont dans notre cas précis deux inconvénients. Le premier est que les modèles de la classe ARIMA ne prennent pas en compte l'effet d'événements calendaires : les jours fériés, les vacances, les fêtes. Ces événements affectent environ 1/4 des semaines et provoquent une forte variation de l'activité des agences de transport. Le deuxième inconvénient est que les paramètres  $(p,d,q)(P,D,Q)$  des modèles de la classe SARIMA sont généralement à déterminer manuellement. Les algorithmes capables de les déterminer automatiquement sont lents et coûteux en ressources machine. Ils doivent comparer un grand nombre de combinaisons de modèles possibles entre  $p, d, q, P, D$  et  $Q$  pour minimiser les trois critères de performances qui sont : le critère d'information d'Akaike (1974), de Schwarz (1978) et d'Hannan-Quinn (1979). *Chacun de ces critères est fondé d'une part sur une mesure de vraisemblance et d'autre part sur une mesure de pénalité, cette dernière étant une fonction croissante avec le nombre de paramètres à estimer. Il n'existe pas de modèle optimal. Un modèle est optimal seulement relativement au critère que l'on a utilisé pour le sélectionner, (33).*

Pour résoudre le premier inconvénient, des chercheurs comme Ferrara, Guegan, Indjehagopian, Mustafa, Chen, Vidakovic et Mavris (50; 66; 95; 28) utilisent les modèles de la classe ARIMAX. C'est une généralisation du modèle ARIMA dans lequel une variable externe est incorporée (eXternal inputs). Ainsi la prévision de la chronique ne dépend pas uniquement de son historique. Aussi appelée « analyse d'intervention », cette théorie a fait l'objet d'une grande activité de recherche. Box et Tio (1975) sont à l'origine de la théorie, mais par la suite d'autres auteurs s'y sont intéressés : Granger (1980), Brokwell et Davis (1987), Chang et al. (1988), Box, Jenkins et Reinsel (1994).

Il est intéressant de constater qu'une entreprise de transport de voyageurs (la RATP) bute sur les mêmes problématiques qu'une entreprise de transport sous température dirigée. Ferrara et Guegan (50) expliquent comment ils construisent un modèle ARMAX pour prévoir le trafic et le nombre de titres de transport vendus. « *L'approche de Box et Jenkins (1970) utilise simplement l'information quantitative contenue dans les données, alors que l'analyse d'intervention (ARIMAX) permet d'ajouter de manière additive une information de type qualitative, par le biais de variables binaires exogènes* ». Une Entreprise d'électricité au Etats-Unis fait également face aux mêmes problèmes (22). Comme dans le cas de l'activité du transporteur, la consommation d'électricité montre une double saisonnalité. La saisonnalité hebdomadaire, la saisonnalité journalière et la tendance sont éliminées par différenciation. Comme pour l'activité du transport sous température dirigée, la consommation d'électricité subit une forte variation lors des jours fériés. L'analyse d'intervention permet alors de prendre en compte, dans l'approche de Box et Jenkins, une information qualitative par le biais de variables binaires exogènes (50). L'avantage

de l'analyse d'intervention par rapport à un modèle économétrique, est qu'elle modélise de manière continue et non ponctuelle l'effet de l'intervention (exp : jours fériés) sur les valeurs environnantes.

Une approche similaire mais toutefois différente, pour prévoir la consommation d'électricité, est proposée par M. Martin (82). Ses travaux suggèrent de partir d'un modèle SARIMA, d'en déduire un « modèle d'état saisonnier » pour le filtrer à l'aide de filtres de Kalman emboîtés. Par la suite, il utilise les modèles d'intervention pour modéliser les effets des événements exceptionnels qui perturbent les valeurs de série chronologique à prévoir. Mentionnons sur ce sujet les travaux plus récents de Lemoine et Pelgrin (76), ainsi que l'article (124) décrivant une manière de mettre en pratique ces méthodes à l'aide de logiciel statistique.

Bien que ces techniques soient très prometteuses dans les cas de l'analyse d'une chronique décrivant l'activité d'une agence de transport particulière, nous ne les avons pas retenues. La raison qui prévaut est qu'il semble difficile d'automatiser le calcul des paramètres d'un modèle SARIMA(X) pour répondre aux particularités des quelque 342 séries chronologiques à prévoir. Bien que ces chroniques se ressemblent du point de vue de la saisonnalité, des différences subsistent. D'autre part, dans le cadre de l'activité du transporteur, nous avons remarqué que les différentes différenciations ne permettaient pas de rendre les séries chronologiques systématiquement stationnaires. Or, c'est une hypothèse fondamentale dans les modèles AR et MA.

Pour prévoir un grand nombre de séries chronologiques, Morineau et al. (92) proposent de commencer par les classer à l'aide d'une analyse par correspondance. A l'aide d'une analyse par composantes principales sur la matrice de covariance de chaque classe, ils extraient autant de nouvelles chroniques. Ces dernières sont prévues par les modèles de la classe SARIMA. Sur ce sujet, D. Ladiray a publié un article en 2006 (72) pour faire le point sur les méthodes utilisées dans le but de combiner l'analyse des données et l'analyse des séries temporelles. Mais ces techniques semblent plus adaptées à la prévision d'un très grand nombre de séries temporelles comme, par exemple, la vente de milliers de références produits dans un Hypermarché. Elles permettent également de prévoir la vente d'articles nouveaux pour lesquels il n'existe pas d'historique. Il est alors possible de les affecter à une classe de produits et d'utiliser la chronique associée à cette classe.

Pour tenter de mesurer l'effet des jours fériés sur l'activité, une méthode non paramétrique a retenu notre attention. Elle est proposée par Poggi (99). La consommation d'électricité en France est, à l'époque, en 1994, prévue par des modèles de la famille SARIMA. Poggi remarque que ces modèles ne sont pas adaptés pour prévoir l'effet des jours fériés sur la consommation électrique. Il propose un modèle non paramétrique. L'idée est de construire une fenêtre d'observation de taille déterminée, et centrée au-

tour de la dernière valeur connue de la chronique. C'est la fenêtre témoin. Il analyse la chronique observée dans la fenêtre témoin pour la comparer aux périodes précédentes, suivant un indice de similarité. La prévision est déduite des observations de la fenêtre retenue (celle au plus fort indice de similarité avec la fenêtre témoin). Cette méthode a l'avantage de prévoir des séries temporelles diverses, à l'aide d'un seul et même outil élémentaire. L'inconvénient est qu'elle est gourmande en ressource informatique puisqu'il faut constamment parcourir la chronique pour la comparer à la fenêtre témoin.

#### **4.4 Modèle statistique de prévision proposés par « Horizons »**

Mettre en œuvre la prévision des indicateurs d'activité des agences de transport TFE s'avère difficile. Ces indicateurs sont, en effet, le résultat d'événements chronologiques qui concernent non seulement des facteurs ayant un comportement déterministe (jours, congés scolaires, zone géographique) mais aussi des facteurs stochastiques (météo, jours fériés, jours de grève, impact de l'actualité sur le consommateur : grippe aviaire et vache folle, marché économique). La prévision demande ainsi de faire appel à des méthodes de modélisation statistiques et probabilistes. « *Les méthodes endogènes utilisent l'historique des données de la variable à prévoir. Les méthodes exogènes utilisent des données corrélées à celles que l'on s'efforce à prévoir* » (18).

La problématique posée par TFE étant de prédire l'activité à des fins opérationnelles (horizons de prévision à très courts termes), il semble logique de privilégier les méthodes quantitatives. Elles ont également le mérite de pouvoir être programmées informatiquement et ainsi être automatisables.

Supposons que l'activité d'une agence TFE est déterminée par trois composantes, une composante déterministe (saisonnalité, jour férié) dont les valeurs futures sont connues, une composante stochastique (grève, perte d'un portefeuille client) dont les valeurs futures sont inconnues, et une composante non observée (liquidation judiciaire du principal concurrent - Nexia) dont les valeurs passées, présentes et à venir sont inconnues (29). Le modèle proposé extrait de la chronique la composante déterministe avant d'appliquer un modèle « autoprojectif » pour caractériser la composante stochastique. L'écart entre le modèle et les observations résulte de la composante non observée ou d'une composante mal estimée. Ce modèle trouve sa valeur ajoutée dans le fait qu'il s'adapte aux spécificités de chaque agence. Il est implémenté pour être exécuté sans intervention humaine. Comme les modèles précédents, il s'appuie sur des algorithmes mathématiques avancés, ayant pour principe d'extrapoler les événements passés sur le futur. Pour ne

pas perdre l'information du passé inférieur à une semaine, l'algorithme de prévision ne va pas chercher à ventiler les prévisions hebdomadaires en prévisions journalières. Au contraire, il s'appuie sur la donnée journalière, car c'est l'information la plus fine présente dans le système d'information de l'entreprise.

L'algorithme de prévision présenté a fait l'objet d'une publication (37) et a été présenté dans des congrès (38; 39; 40; 42; 43).

#### 4.4.1 Historique

L'historique d'une **série temporelle** est l'ensemble des valeurs de la série temporelle, incluses dans un intervalle de temps et utilisées pour générer une prévision. La première valeur prévue d'une série temporelle fait généralement suite à la dernière valeur de l'historique de cette même série temporelle. Rappelons qu'un modèle de prévision cherche à décrire certaines régularités dans l'historique pour les extrapoler et ainsi proposer une prévision de la chronique.

L'uniformisation des données en provenance des bases de données HP3000 et GTI (voir section 4.2), suivant les définitions des indicateurs à prévoir (voir section 3.1.2), a permis de créer les historiques nécessaires à la prévision des indicateurs d'activité.

L'algorithme de prévision, développé dans le cadre de cette thèse, prévoit trois séries temporelles liées par une relation linéaire. Prenons l'exemple des trois chroniques suivantes,

- poids quotidien, en kg, total, des marchandises passées à quai,
- le poids quotidien, en kg, en expédition, des marchandises passées à quai,
- le poids quotidien, en kg, en distribution, des marchandises passées à quai.

La première série temporelle, appelée par abus de langage, poids total, est le résultat de la somme des deux autres. Il en va de même quand il s'agit de prévoir le nombre de positions quotidien passé à **quai** d'une agence de transport. Par la suite, pour désigner les 6 chroniques, nous parlerons de poids ou nombre de positions, total, en expédition, en distribution. L'algorithme de prévision prévoit les valeurs des trois séries (total, expédition, distribution) individuellement et les redresse de sorte à garder la contrainte de linéarité.

Les résultats observés montrent qu'un historique de 5 ans d'observations permet d'anticiper correctement les variations des indicateurs d'activité des agences de transport. Aussi, l'algorithme de prévision proposé fonctionne avec un minimum de trois ans d'historique.

## 4.4.2 Nettoyage des données

Les valeurs numériques qui constituent l'historique des trois séries temporelles (poids ou nombre de positions, au total, en expédition, en distribution), données en paramètre d'entrée de l'algorithme, sont souvent entachées de valeurs dites extrêmes ou manquantes. Les valeurs extrêmes peuvent être, ou non, réellement observées. Lorsqu'elles ne le sont pas, c'est qu'elles sont issues d'une erreur de saisie ou d'une erreur de traitement de l'information. Dans tous les cas de figure, elles ont un impact néfaste sur le calcul des prévisions. C'est pourquoi il est nécessaire de parcourir les valeurs de l'historique pour les corriger le cas échéant et, ainsi, aboutir à une série temporelle régulière et corrigée des valeurs extrêmes.

Nettoyer les données, signifie trier les données dans le temps, vérifier qu'il ne manque pas de jours, vérifier que les dimanches sont effacés et corriger les données extrêmes (voir 3 prochaines section [4.4.2.1](#), [4.4.2.2](#), [4.4.2.3](#)).

### 4.4.2.1 Rendre la série régulière

Dans le cas présent, pour être traitée, une série temporelle doit être **régulière**. C'est-à-dire qu'entre deux valeurs il existe toujours le même intervalle de temps. L'intervalle de temps entre deux valeurs consécutives des séries temporelles étudiées est le jour. Par contre, les dimanches ne sont pas compris. Ainsi, la semaine comprend 6 jours ouvrables. Autrement dit, la valeur qui suit celle d'un samedi doit impérativement être celle d'un lundi. Une année comprend au minimum 312 jours ouvrés, (52 semaines fois 6 jours ouvrés) et au maximum 318 jours ouvrés (53 semaines fois 6 jours ouvrés).

L'algorithme programmé avec le logiciel R (voir section [4.7.1](#)), parcourt l'historique des séries temporelles fournies en paramètre d'entrée. Lorsqu'une date est manquante, l'algorithme l'insère et lui attribue, dans un premier temps, une valeur nulle. Lorsqu'une date correspond à un dimanche, la valeur associée est supprimée. Suite à ce traitement, les séries temporelles sont régulières.

### 4.4.2.2 Correction des valeurs nulles

Mis à part les jours fériés, nous posons l'hypothèse qu'il n'existe pas d'activité nulle, aussi bien pour les séries temporelles décrivant les indicateurs en distribution que pour celles en expédition. D'autre part, les valeurs observées montrent que l'activité des agences de transport est fortement liée au jour de la semaine (lundi, mardi, etc). C'est

pourquoi, quand une valeur est nulle, et qu'elle ne correspond pas à un jour férié, la moyenne des valeurs observées pour le jour de la semaine, le mois et l'année en question, lui est attribuée. Pour ne pas fausser les résultats, la moyenne ne prend pas en compte les valeurs nulles. Ainsi, il arrive que cette moyenne ne puisse pas être calculée par manque de données. Dans ce cas, nous remplaçons la moyenne du jour de la semaine dans le mois et l'année, par la moyenne du jour de la semaine dans l'année. Une autre solution aurait été d'attribuer à la valeur nulle, la valeur prévue par le modèle statistique de prévision. Cette solution n'a pas été choisie, car les valeurs nulles en début d'historique (inférieur à 3 ans) n'auraient pas pu être corrigées par manque d'historique.

#### 4.4.2.3 Recherche et correction des valeurs extrêmes

Une valeur est aberrante quand elle s'écarte anormalement des valeurs moyennes de la chronique. Il est difficile de savoir si la valeur atypique s'est réalisée ou si elle provient d'une erreur de saisie. Si la valeur s'est effectivement réalisée, nous devrions pouvoir l'expliquer et insérer une composante déterministe dans le modèle de prévision. Dans ce cas, l'idéal serait de faire appel au jugement d'un expert métier pour identifier les causes de chacune des valeurs extrêmes retenues. Mais face au nombre important de séries et de dates lointaines dans le passé, la tâche semble ardue. Même l'expert ne se souvient pas précisément de ce qui est arrivé un certain lundi 3 janvier d'il y a 3 ans par exemple. C'est pourquoi, nous faisons appel à des méthodes statistiques pour déterminer qu'une valeur est aberrante et la remplacer par une estimation plus réaliste, à l'exception près des valeurs relatives à un jour férié. Ces valeurs pourraient être désignées comme extrêmes car l'activité y est très faible. Or, la faiblesse de l'activité s'explique justement par le jour férié. Nous étudions les jours fériés et leurs effets sur les jours voisins dans la section 4.4.6.1. En attendant, pour que ces valeurs ne soient pas détectées comme extrêmes, nous les modifions par la moyenne du jour de la semaine et de l'année en question. Après le traitement statistique, consistant à rechercher les valeurs extrêmes, nous leur réattribuons leurs vraies valeurs.

Pour détecter les valeurs extrêmes, nous créons un intervalle de valeurs admissibles. Toute valeur non comprise dans l'intervalle sera désignée comme aberrante. Cet intervalle se définit pour un mois de l'année. Ainsi si nous sommes face à un historique d'une longueur de trois ans, nous construisons  $12 \times 3$  intervalles.

Construction de l'intervalle « d'admissibilité ». Il découle d'un outil de statistique descriptive appelé la boîte à moustache (ou boxplot en anglais). C'est une invention de TUKEY (1977) pour représenter schématiquement une distribution. L'outil définit un intervalle de valeurs comprises entre une extrémité inférieure et extrémité supérieure. Les valeurs sont alors définies comme atypiques lorsqu'elles sont situées au-delà de ces frontières

(74).

L'intervalle « d'admissibilité » s'écrit :

$$[Q1 - 1,5(Q3 - Q1); Q3 + 1,5(Q3 - Q1)] \quad (4.1)$$

$Q1$  et  $Q3$  sont les valeurs du premier et du troisième quartile. 25% des valeurs de la distribution sont inférieurs à  $Q1$  et 25% sont supérieurs à  $Q3$ .  $(Q3 - Q1)$  est l'écart interquartile (InterQuartile Range). Il est utilisé comme indicateur de dispersion. Il correspond à 50% des observations situées dans la partie centrale de la distribution.

Toute observation non comprise dans l'intervalle est corrigée par la valeur que prend l'extrémité la plus proche de l'intervalle « d'admissibilité ».

L'algorithme met en œuvre cette méthode pour corriger les valeurs extrêmes des deux premières séries temporelles fournies en paramètres d'entrée. Suite à cela, il redéfinit la troisième série temporelle comme la somme des deux premières.

Remarquons que ce processus de nettoyage des données donne des résultats raisonnables quand, dans un même mois, un trop grand nombre de termes de la série n'est pas nul. Dans le cas contraire, les valeurs nulles, désignées comme extrêmes, se verront attribuer toute la même valeur moyenne. D'autre part, le processus n'efface pas les sauts que prennent les valeurs d'une année sur l'autre. Pour illustrer ces deux phénomènes évoqués, étudions la figure 4.1.

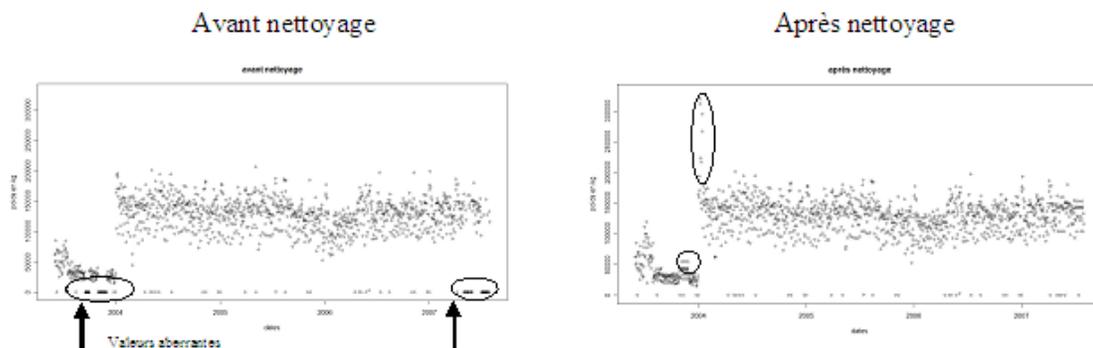


FIGURE 4.1 – Serie chronologique du poids quotidien, en kg, total, des marchandises passées à quai avant et après nettoyage des valeurs extrêmes

La figure 4.1 a pour but de montrer un cas extrême. Mais ces cas se répètent assez fréquemment pour être soulignés. La figure 4.1 montre les valeurs prises par une série temporelle, entre le 1<sup>er</sup> Janvier 2003 et le 31 Décembre 2007, avant et après le processus de nettoyage des données. Nous observons un palier entre la fin de l'année 2003 et le début de l'année 2004. Il peut être dû à diverses causes, par exemple, une erreur dans le traitement informatique des données. Rappelons que la source des données n'est pas

la même entre les données avant et après adoption du nouveau TMS (voir section 3.1.4). Les procédures de transformation des données, selon leur base de données d'origine, sont différentes. Mais elles sont pensés de sorte qu'elles définissent la même grandeur (voir section 3.1.2). Le palier observé peut également être dû à une hausse avérée de l'activité. Cette hausse fait suite à l'arrivée d'un nouveau portefeuille client, ou même d'un marché. La hausse d'activité peut aussi être due à un regroupement d'agences de transport. Par exemple, les agences Mélédo et TFE Vannes se sont regroupées pour n'en former plus qu'une. Dans tous les cas de figure, ce palier risque de fausser les calculs de prévision. C'est pourquoi, quand l'algorithme détecte un palier important, il efface de l'historique des valeurs observées antérieurement à la date du palier.

Précisons que la figure 4.1 représente les valeurs prises par la série temporelle « poids quotidien, en kg, total, des marchandises passées à quai d'une agence de transport TFE ». Ces valeurs sont égales à la somme des valeurs prises par la série « poids quotidien, en kg, des marchandises passées à quai d'une agence de transport TFE et partant en expédition » et des valeurs prises par la série « poids quotidien, en kg, des marchandises passées à quai d'une agence de transport TFE et partant en distribution ». Par abus de langage, nous appelons les séries respectivement, « poids total », « poids en expédition », « poids en distribution ». La figure 4.1 montre que, suite au processus de nettoyage des données, la série « poids total » fait apparaître des valeurs extrêmes début de l'année 2004. Ce n'était pas l'objectif recherché. Ces valeurs s'expliquent par le phénomène suivant. Début de l'année 2004, la série temporelle « poids en distribution » montrait des valeurs très fortes alors que la série temporelle « poids en expédition » montrait des valeurs nulles aux mêmes dates. Suite au processus de nettoyage des données, ces dernières ont été redressées. Or, après le nettoyage des données des séries « poids en expédition » et « poids en distribution », la série « poids total » est recalculée pour qu'elle soit égale à la somme des deux premières. C'est ainsi que des valeurs extrêmes sont apparues dans la série temporelle « poids total ».

La figure 4.1 montre également que les valeurs nulles d'un grand nombre de termes de la série temporelle, observées fin 2003 et fin 2007, ont toutes été corrigées par la même valeur. C'est parce qu'en Novembre 2003 et en Décembre 2007, il manquait un grand nombre d'observations, qu'elles ont été estimées par la valeur moyenne des observations de l'année. Par exemple, une valeur nulle, observée un lundi du mois de Décembre 2007, est corrigée par la moyenne des valeurs observées les lundis de l'année 2007. Si les valeurs observées les lundis de novembre 2007 n'avaient pas manqué, nous aurions retenu comme correction, leur moyenne.

Les étapes d'extraction, d'uniformisation et le nettoyage des données ont permis d'aboutissent à l'édifice et à la consolidation d'un historique des séries temporelles à prévoir. Un historique propre établit les fondations permettant d'exécuter une suite de procédures pour parvenir à des prévisions.

### 4.4.3 Schéma de décomposition

Pour prévoir les séries temporelles décrivant l'activité des agences de transport, nous choisissons, dans un premier temps, de les décrire par les composantes qui les constituent. Chaque série temporelle étudiée peut être décrite par la somme ou le produit de trois composantes : la tendance, la saisonnalité et le bruit. Comme il a été expliqué dans la section 2.3.3, il faut choisir entre un schéma de décomposition additif ou multiplicatif. Une méthode pour déterminer le schéma à garder est celle de la « bande ». Elle consiste à tracer deux droites des MCO, passant respectivement par les maxima et les minima de chaque période de la chronique. Si les droites sont parallèles, il faut décider d'un schéma additif, sinon d'un schéma multiplicatif. Le résultat de l'application de cette méthode sur les séries temporelles étudiées est systématiquement le choix d'un schéma multiplicatif. La raison en est simple. L'activité des agences de transport a fortement progressé au cours du temps. Cette progression est plus accentuée du lundi au vendredi que le samedi. En d'autres termes, l'activité au cours du temps du samedi progresse moins vite que celle des autres jours de la semaine. Cet écart crée l'accroissement de l'amplitude entre la droite passant par les minima et la droite passant par les maxima de chaque période de la chronique.

Citons R. Bourbonnais et JC. Usunier (18) qui font remarquer que le schéma de décomposition multiplicatif est actuellement le plus utilisé en économie. Ils ajoutent : le schéma multiplicatif « *est commode puisque le logarithme de la chronique conduit au schéma additif* ».

Ainsi, nous posons l'hypothèse que les séries temporelles étudiées peuvent être décomposées sur le schéma multiplicatif. Pour étudier les composantes du schéma, indépendamment les unes des autres, nous nous ramenons à un schéma additif, par le passage au logarithme.

Soit  $Y_t$  une série temporelle, sa décomposition, suivant le schéma multiplicatif, est :

$$Y_t = T_t \times S_t \times R_t \quad (4.2)$$

avec

- $T_t$  la valeur de la tendance à la date  $t$ ,
- $S_t$  la valeur de la saisonnalité à la date  $t$ ,
- $R_t$  la valeur du résidu à la date  $t$ .

Le passage au schéma additif se fait par la formule suivante,

$$Z_t = \ln(Y_t) = \ln(T_t) + \ln(S_t) + \ln(R_t) \quad (4.3)$$

Les sections 4.4.4 et 4.4.5 expliquent comment la tendance et la saisonnalité des séries à étudier sont estimées. Nous utilisons deux méthodes pour le faire. La première

consiste d'abord à déduire la tendance, pour ensuite déduire la saisonnalité, par la méthode des moyennes mobiles. La deuxième méthode est celle proposée par Buy-Ballot (1847). Suite à l'application des deux méthodes, nous obtenons deux séries pour représenter le bruit. Dans la section 4.4.6, nous déduisons, des deux séries, l'effet des événements calendaires. Dans la section 4.4.9, nous détaillons le calcul d'une prévision pour chacune des deux séries. Nous ajoutons aux valeurs prévues l'effet des événements calendaires, la saisonnalité et la tendance. Nous obtenons ainsi deux propositions de prévision que nous combinons avec la condition de minimiser la variance de l'erreur entre la valeur réelle et la prévision.

#### 4.4.4 Décomposition d'une chronique : méthode n°1

##### 4.4.4.1 Traitement de la tendance

Plusieurs raisons nous font préférer une modélisation de la tendance par une fonction linéaire. D'une part comme le font remarquer à juste titre B. Burtschy et C. Menéndian (24), dans le milieu économique les tendances sont progressives. D'autre part, nous avons mentionné dans la section 2.3.3 qu'il était risqué d'utiliser les fonctions non linéaires. Et surtout, l'observation des valeurs décrivant l'activité des agences de transport TFE dans le temps se prête à une modélisation de la tendance, par une fonction linéaire. Pour exemple, la figure 4.2 montre le logarithme du poids quotidien de toutes les marchandises passées à quai d'une certaine agence TFE, entre 2002 et 2009. Le nuage de points indique, qu'en début d'année, les poids sont moins importants qu'en fin d'année. Une explication possible est la baisse de consommation des ménages en produits alimentaires après les fêtes de fin d'année. Michel Vaté (115), propose d'approcher la tendance d'une telle suite d'observations par une fonction linéaire par morceaux de périodicité annuelle.

Ainsi, nous posons l'hypothèse que la tendance des séries temporelles à étudier peut être décrite par la combinaison linéaire d'une tendance intra-annuelle et d'une tendance inter-annuelle. La tendance intra-annuelle est comprise entre janvier et décembre. La tendance inter-annuelle est comprise entre la première date et la dernière date de l'historique étudié.

L'équation de la tendance est une fonction du temps, elle s'écrit :

$$f(t) = T_t = \alpha t + \beta An(t) + \eta \quad (4.4)$$

avec  $t$ , le temps, et  $An(t)$  l'année correspondante au temps  $t$ ,  $\alpha$ ,  $\beta$  et  $\eta$  sont les inconnues à estimer. L'algorithme, programmé sous R, utilise la méthode de décomposition QR (57).

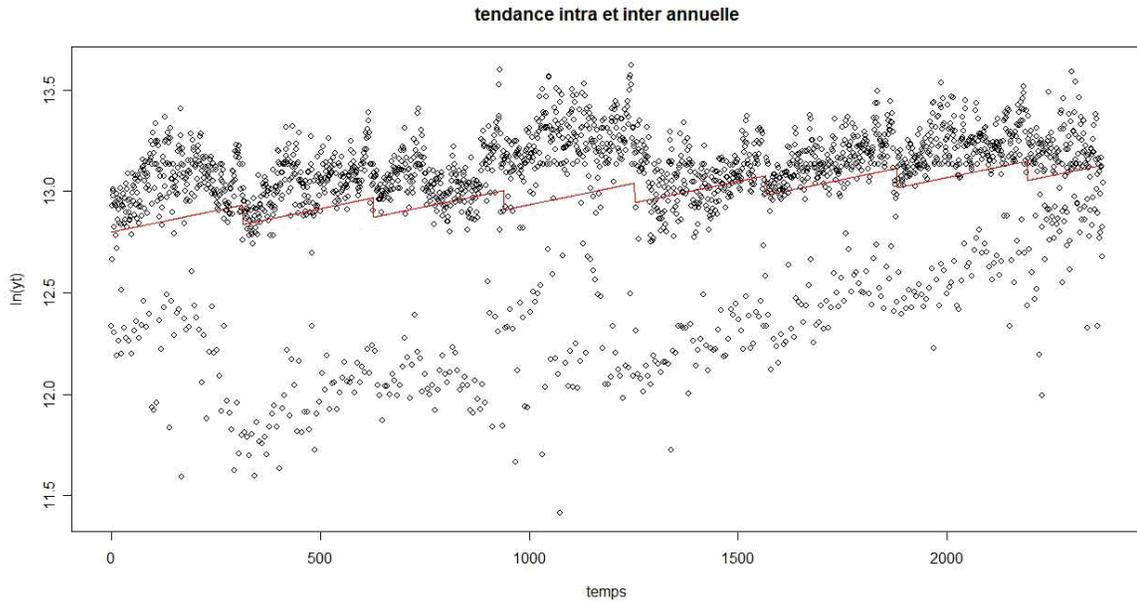


FIGURE 4.2 – *Tendance par palier*

Le terme  $\beta An(t)$  permet de créer le palier entre deux années consécutives. Toujours pour être plus proche des données observées, nous aurions pu choisir un palier différent pour chaque année comprise dans l'historique. Mais, dans ce cas, quels paliers choisir dans une perspective de prévision des années futures ? La méthode retenue a le mérite de créer un palier moyen. Il s'accroît très légèrement avec les années, mais l'impact sur les données déduites de leur tendance reste négligeable. La méthode retenue conduit aussi à un modèle parcimonieux.

Nous venons de l'évoquer, l'étape suivante est de retrancher cette tendance à la chronique. Suite à cette opération, nous obtenons une nouvelle chronique dont les valeurs varient autour d'une moyenne proche de zéro. La nouvelle chronique compte donc des valeurs négatives. Or les valeurs négatives n'appartiennent pas à l'ensemble de définitions de la fonction logarithme. Rappelons que cette dernière permet de passer d'un schéma multiplicatif à un schéma additif (voir formule 4.3). Pour redresser les valeurs négatives, nous soustrayons aux valeurs de la nouvelle chronique, sa valeur minimale. Il en résulte une chronique appelée  $z'_t$  dont l'équation s'écrit :

$$Z'_t = Z_t - T_t - \min(Z_t - T_t) \quad (4.5)$$

#### 4.4.4.2 Les composantes saisonnières

Pour se rendre compte de la saisonnalité de l'activité des agences de transport, il faut superposer sur un même graphique les variations des activités annuelles. Pour exemple, la figure 4.3 superpose le poids hebdomadaire passé à quai d'une certaine agence TFE entre 2001 et 2007.

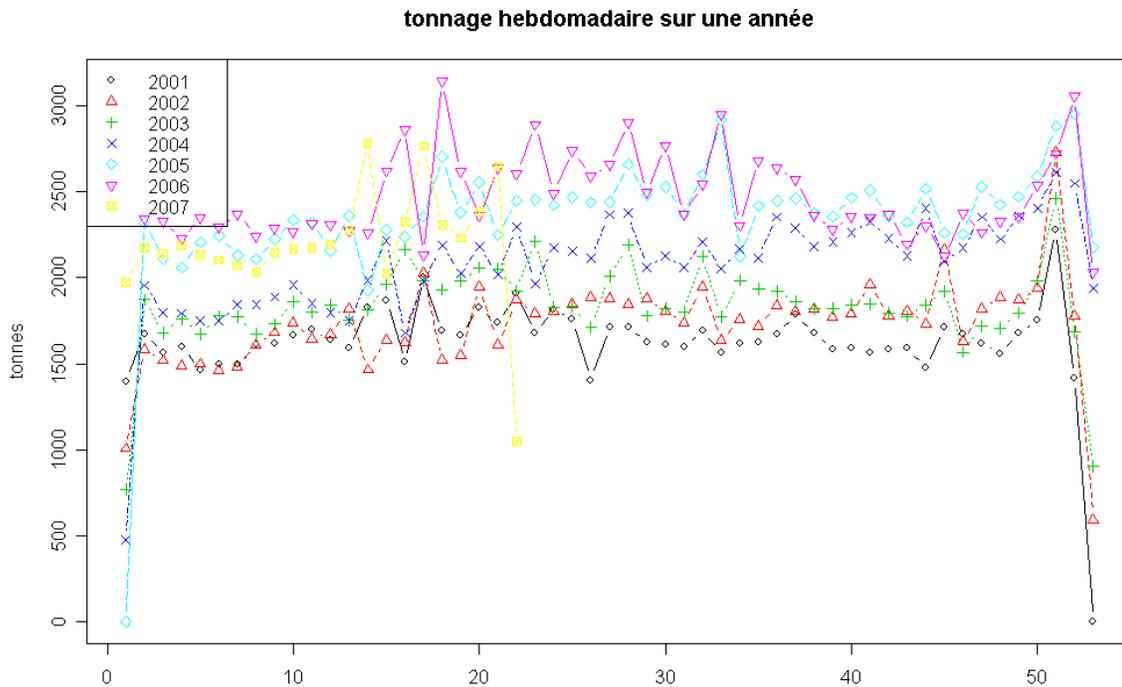


FIGURE 4.3 – Poids hebdomadaire en tonne passé à quai entre 2001 et 2007

Nous nous intéressons directement à la saisonnalité hebdomadaire, et non mensuelle, car nous cherchons à prévoir le jour. L'avantage de la semaine par rapport au mois est qu'elle contient toujours 6 jours ouvrables. Alors que le nombre de jours ouvrables, par mois, peut varier. Par contre, le graphique de la figure 4.3 pointe le problème des numéros de semaines 1 et 52. Elles ne comptent pas 6 jours ouvrables et c'est pourquoi le poids est plus faible. Elles devront être redressées.

Les valeurs prises par les indicateurs poids et nombre de positions passées à quai d'une agence TFE, admettent une double saisonnalité. La saisonnalité hebdomadaire comporte 53 coefficients (semaine 0 à semaine 52) et la saisonnalité journalière en comporte 318 (53 sem.  $\times$  6 jours).

La saisonnalité hebdomadaire est due à des périodes d'activité fluctuante. Ces périodes

peuvent durer plus d'une semaine, mais comme elles sont parfois mitoyennes sur deux mois consécutifs, c'est une raison de plus pour ne pas retenir la saisonnalité mensuelle. Ces fluctuations d'activité sont causées par des événements extérieurs tels que la météo, les vacances scolaires, les périodes de fête. Ces événements, et notamment un certain nombre de jours fériés, varient de numéro de semaine d'une année à l'autre. C'est ce qui explique les pics d'activité décalés sur le graphique de la figure 4.3.

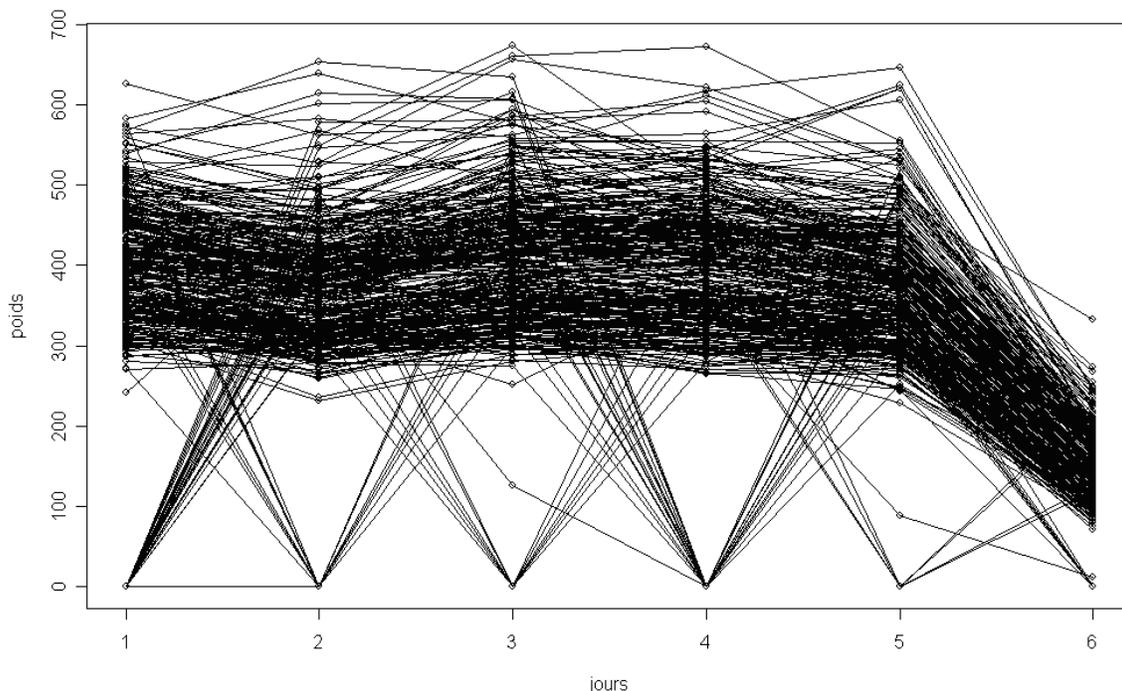


FIGURE 4.4 – *Superposition des poids quotidiens, en tonne, passé à quai, entre 1999 et 2007, d'une agence*

Comme l'atteste la figure 4.4, la saisonnalité journalière est très marquée. Elle est due à une répartition de l'activité sur les 6 jours ouvrés de la semaine. Cette répartition dépend de l'agence en question. Par contre, le point commun entre toutes les agences est une très faible activité le samedi. Elles cherchent, dans la mesure du possible à ventiler les transports sur le reste des jours de la semaine.

Certains facteurs sont intégrés explicitement dans la saisonnalité. Nous pensons aux vacances scolaires, aux promotions qui se répètent à la même période chaque année. C'est la raison pour laquelle nous ne nous attardons pas à les analyser.

Pour définir les coefficients saisonniers hebdomadaires, il faut partir d'une chronique dont

le laps de temps entre deux observations est la semaine. Pour le système de numérotation des semaines nous nous sommes inspirés de la norme ISO 8601. Nous définissons la règle comme suit :

- la semaine commence le lundi,
- les jours de chaque semaine sont numérotés de 1 pour le lundi à 6 pour le samedi,
- la semaine 1 est celle qui contient le premier lundi de l'année (c'est contraire à la norme ISO qui dit que la semaine 1 est celle qui contient le premier jeudi de l'année).

Il s'en suit que si le 1<sup>er</sup> janvier ne tombe pas un lundi, ce jour et les suivants, jusqu'au prochain lundi, appartiennent à la semaine 0, qui, avec la dernière semaine de l'année (52) sont les seules susceptibles de ne pas compter 6 jours ouvrables.

L'algorithme programmé sous R corrige la valeur des semaines 0 et 52. Si elles comportent moins de 6 jours et si le nombre de jours est supérieur à 3, sans compter le samedi (valeur faible), la valeur hebdomadaire est pondérée par

$$\frac{\text{nombre de jours ouvrables (6)}}{\text{nombre de jours de la semaine à corriger}} \quad (4.6)$$

Sinon, elle prend comme valeur 0,9 fois la valeur de la semaine suivante (semaine 1) ou précédente (semaine 51).

#### 4.4.4.3 Méthode de décomposition saisonnière par moyennes mobiles

Après avoir éliminé la tendance, nous cherchons à estimer des coefficients saisonniers des chroniques. La méthode des moyennes mobiles est une technique souvent utilisée pour lisser les valeurs prises par une série chronologique. Cette méthode permet ainsi d'effacer les fluctuations saisonnières. Des moyennes mobiles d'ordre 12 sont, par exemple, calculées pour éliminer la saisonnalité mensuelle d'une série temporelle, dont l'unité de temps entre deux valeurs est le mois. Comme il a été expliqué dans la section [3.2.3.1](#), les coefficients saisonniers sont calculés par la moyenne des différences (ou des rapports) entre les moyennes mobiles et les valeurs prises par la série temporelle.

Dans le cas des chroniques décrivant l'activité des agences de transport, nous commençons par calculer des moyennes mobiles d'ordre 53 pour effacer la saisonnalité hebdomadaire. La série des moyennes mobiles s'écrit :

$$MM_t = \frac{1}{53} \sum_{t-26}^{t+26} z'_t \quad (4.7)$$

Imaginons que la série temporelle  $z'_t$  contienne une valeur pour chaque semaine comprise entre le 1<sup>er</sup> Janvier 2002 et le 31 Décembre 2008. Cet intervalle de temps comprend 371 ( $7 \times 53$ ) semaines. Le calcul des moyennes mobiles n'est possible qu'à partir

de la 27<sup>e</sup> semaine de la série temporelle, et jusqu'à la semaine 345 (371 – 26). Il y a donc une perte d'information. Or, nous souhaitons obtenir 371 moyennes mobiles et non 318 (371 – 53) pour les soustraire aux 371 valeurs de la chronique d'origine. Pour ce faire, nous ajoutons des valeurs en tête et en queue de la série des moyennes mobiles. Les 26 premières valeurs adjointes sont celles de l'année suivante à la même époque, soit, dans notre exemple, les semaines 1 à 26 de l'année 2003. Les 26 dernières sont celles de l'année précédente à la même époque, soit, dans notre exemple, les semaines 27 à 53 de l'année 2007.

Suite à cet exercice, dans le cas d'un schéma additif, nous pouvons calculer l'écart entre les valeurs observées et les moyennes mobiles. Dans le cas d'un schéma multiplicatif nous calculons le rapport entre les valeurs observées et les moyennes mobiles. Pour les 53 numéros de semaines, une moyenne des écarts, ou rapports, est calculée. Les 53 coefficients, ainsi obtenus, sont normalisés pour aboutir aux 53 coefficients saisonniers notés  $S_i$ , ( $i = 0, \dots, 52$ ).

A ce stade, nous avons calculé 53 coefficients saisonniers. Or, la série à étudier est une série temporelle dont le pas de temps entre deux valeurs est le jour, et non la semaine. Pour extraire la saisonnalité hebdomadaire de la série journalière, nous commençons par calculer des poids journaliers moyens, pour répartir le coefficient saisonnier hebdomadaire sur les 6 jours ouvrables de la semaine. L'équation s'écrit :

$$p_j = \frac{\sum_{t=1}^T z'_{j,t}}{\sum_{t=1}^T z'_t} \quad (4.8)$$

$p_j$  est la somme de toutes les observations du jour  $j$  divisée par la somme des observations.

Dans le cas de l'exemple cité plus haut,  $p_{(j=lundi)}$  est la somme des valeurs observées tous les lundis compris entre le 1<sup>er</sup> Janvier 2002 et le 31 Décembre 2008, divisée par la somme de toutes les valeurs observées durant cette même période.

Pour déduire la saisonnalité hebdomadaire de la série journalière, nous utilisons la formule suivante :

$$z''_{a,i,j} = z'_{a,i,j} - (S_i \times p_j) \quad (4.9)$$

avec  $a$  l'indice de l'année,  $i$  celui de la semaine et  $j$  du jour.

Après avoir retiré l'effet de la saisonnalité hebdomadaire, il reste à estimer l'effet de la saisonnalité journalière, pour le retirer à son tour. La même méthodologie est utilisée. A partir de la série temporelle  $z''_{a,i,j}$ , les 318 ( $53 \times 6$ ) coefficients saisonniers journaliers sont calculés. La série temporelle  $\ln(R_{a,i,j}^{(1)})$  obtenue par la formule 4.10, est dépourvue de tendance, de saisonnalité hebdomadaire et journalière (voir equation 4.3).

$$\ln(R_{a,i,j}^{(1)}) = z_{a,i,j}'' - S_{i,j} \quad (4.10)$$

$S_{i,j}$ , est le coefficient saisonnier du jour  $j$  de la semaine  $i$ .

Le graphique de la figure 4.5 illustre un exemple de variation des coefficients saisonniers d'un schéma multiplicatif.

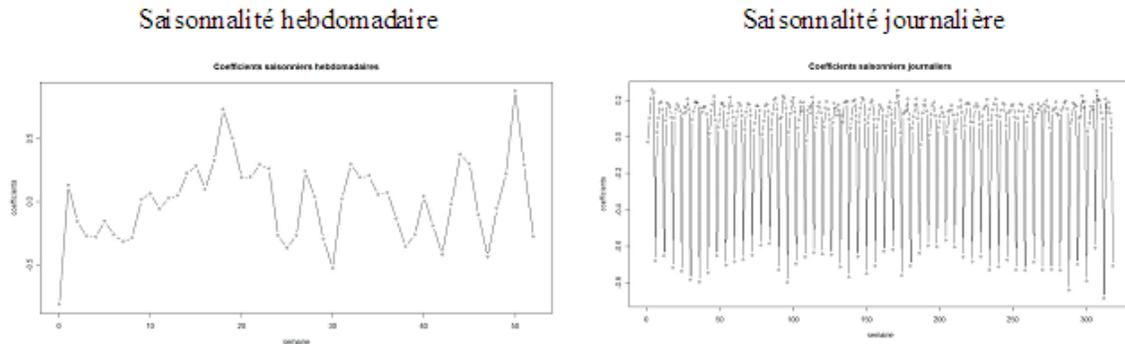


FIGURE 4.5 – Représentation graphique de la saisonnalité

Les coefficients saisonniers montrent l'écart de la valeur moyenne constatée par rapport à la moyenne annuelle représentée par l'axe  $y = 0$ . Un coefficient de 0.5 signifie que l'activité est 1.5 fois supérieure à l'activité moyenne annuelle. A la lecture des coefficients hebdomadaires, nous constatons que l'activité est :

- plus faible que l'activité moyenne annuelle, les semaines 2 à 9,
- moyenne, les semaines 10 à 14,
- au-dessus de la moyenne annuelle, les semaines 19 et 52.

## 4.4.5 Décomposition d'une chronique : méthode n°2

### 4.4.5.1 Méthode de décomposition linéaire - formules de Buys-Ballot

En matière de prévision des ventes, la méthode idéale n'existe pas. Partant du principe que deux valent mieux qu'une, pourquoi ne pas utiliser une deuxième technique et ne garder que le meilleur des deux. La deuxième technique choisie est celle proposée par Buys-Ballot (25). C'est une dessaisonalisation par régression linéaire.

Soit le processus suivant,

$$Z_t = at + b + \gamma_1 S_t^{(1)} + \gamma_2 S_t^{(2)} + \gamma_3 S_t^{(3)} + \gamma_4 S_t^{(4)} + \gamma_5 S_t^{(5)} + \gamma_6 S_t^{(6)} + \ln(R_t) \quad (4.11)$$

La chronique  $z_t$  se décompose en une tendance, une suite de composantes saisonnières  $S_t^j$  et d'un résidu  $\ln(R_t)$ . 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.

Reprenons notre exemple. L'historique de la chronique à prévoir comprend une valeur pour chaque jour compris entre le 1<sup>er</sup> Janvier 2002 et le 31 Décembre 2008. Le premier janvier 2002 étant un mardi, nous commencerons l'historique au 7 Janvier 2002 pour le premier jour, soit un lundi. Le modèle s'écrit,

$$\begin{bmatrix} 7641 \\ 7815 \\ 7812 \\ 7774 \\ 6430 \\ 3527 \\ 7133 \\ \vdots \\ z_t \end{bmatrix} = b \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} + a \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \\ \vdots \\ t \end{bmatrix} + \gamma_1 \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ \vdots \\ S_t^1 \end{bmatrix} + \dots + \gamma_6 \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ \vdots \\ S_t^6 \end{bmatrix} + \phi_1 \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 0 \\ \vdots \\ S_t^7 \end{bmatrix} + \dots + \phi_{53} \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ \vdots \\ S_t^{59} \end{bmatrix} + \begin{bmatrix} \ln(R_1) \\ \ln(R_2) \\ \ln(R_3) \\ \ln(R_4) \\ \ln(R_5) \\ \ln(R_6) \\ \ln(R_7) \\ \vdots \\ \ln(R_t) \end{bmatrix} \quad (4.12)$$

Les paramètres  $\gamma$  représentent la saisonnalité journalière au sein d'une semaine et les paramètres  $\phi$ , la saisonnalité hebdomadaire dans une année.

Sous la forme matricielle nous obtenons,

$$\begin{bmatrix} 7641 \\ 7815 \\ 7812 \\ 7774 \\ 6430 \\ 3527 \\ 7133 \\ \vdots \\ z_t \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & \dots & 0 & 1 & 0 & \dots & 0 \\ 1 & 2 & 0 & \dots & 0 & 1 & 0 & \dots & 0 \\ 1 & 3 & 0 & \dots & 0 & 1 & 0 & \dots & 0 \\ 1 & 4 & 0 & \dots & 0 & 1 & 0 & \dots & 0 \\ 1 & 5 & 0 & \dots & 0 & 1 & 0 & \dots & 0 \\ 1 & 6 & 0 & \dots & 1 & 1 & 0 & \dots & 0 \\ 1 & 7 & 1 & \dots & 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots \\ 1 & t & S_t^1 & \dots & S_t^6 & S_t^7 & S_t^8 & \dots & S_t^{59} \end{bmatrix} \begin{bmatrix} b \\ a \\ \gamma_1 \\ \vdots \\ \gamma_6 \\ \phi_1 \\ \vdots \\ \phi_{53} \end{bmatrix} + \begin{bmatrix} \ln(R_1) \\ \ln(R_2) \\ \ln(R_3) \\ \ln(R_4) \\ \ln(R_5) \\ \ln(R_6) \\ \ln(R_7) \\ \vdots \\ \ln(R_t) \end{bmatrix} \quad (4.13)$$

soit,

$$Z = X\theta + \ln(R) \quad (4.14)$$

Le paramètre  $\theta$  ( $\theta = [b, a, \gamma_1, \dots, \gamma_6, \phi_1, \dots, \phi_{53}]'$ ) s'estime par MCO,  $\hat{\theta} = (X'X)^{-1}X'Z$ . D'autre part, nous voulons que la somme des fluctuations saisonnières soit nulle sur

l'année, nous imposons alors la contrainte :  $\sum_{i=1}^6 \gamma_i + \sum_{i=1}^{53} \phi_i = 0$

Le problème d'estimation par moindres carrés devient,

$$\left\{ \begin{array}{l} \min_{\beta, \gamma, \phi} \sum_{t=1}^T \left[ z_t - \beta_1 - \beta_2 t - \sum_{j=1}^6 S_t^j \gamma_j - \sum_{j=1}^{53} S_t^j \phi_j \right]^2 \\ \text{sous contrainte } \sum_{j=1}^6 \gamma_j + \sum_{j=1}^{53} \phi_j = 0 \end{array} \right. \quad (4.15)$$

Ainsi, pour chacune des 53 semaines et chacun des 6 jours travaillés de la semaine, nous obtenons des coefficients saisonniers.

La prévision à la date  $t$  et à horizons  $t + h$  s'écrit,

$$\hat{z}_t = at + b + \sum_{j=1}^6 \gamma_j S_t^j + \sum_{j=1}^{53} \phi_j S_t^j \quad (4.16)$$

La série corrigée de sa tendance et des variations saisonnières (CVS) s'écrit,

$$\ln(R_t^{(2)}) = z_t - \hat{z}_t \quad (4.17)$$

Suite aux deux méthodes de décomposition, nous obtenons deux séries  $\ln(R_t^{(1)})$  et  $\ln(R_t^{(2)})$  (voir équation 4.10 et 4.17). Ces séries contiennent encore de l'information déterministe, que nous pouvons caractériser. Cette information décrit l'effet des événements calendaires sur l'activité des agences de transport. Nous détaillons la méthode d'estimation dans la section 4.4.6.

Notons au préalable que d'autres méthodes de lissage existent.

#### 4.4.5.2 Autres méthodes de lissage

Les coefficients saisonniers sont obtenus en calculant la moyenne des rapports entre la série lissée et la série réelle, sur les différentes périodes. Après normalisation, nous obtenons les coefficients saisonniers. La littérature propose d'autres méthodes de lissage, comme EVF, LOESS et filtrage par régression.

La méthode EVF (Eigen Vector Filtering ou Filtrage par les vecteurs propres) (65), décompose le signal en composantes orthogonales. Elle débute par l'écriture d'une matrice de  $r$  séries décalées d'un multiple de  $\tau$ , à partir de l'historique de  $n$  observations  $x_i$ .  $\tau$  est

le décalage temporel,  $r$  est le décalage maximum entre les deux séries extrêmes.

$$X = \begin{bmatrix} X_{(r-1)(\tau+1)} & \cdots & X_{(2\tau+1)} & X_{(\tau+1)} & X_1 \\ X_{(r-1)(\tau+2)} & \cdots & X_{(2\tau+2)} & X_{(\tau+2)} & X_2 \\ X_{(r-1)(\tau+3)} & \cdots & X_{(2\tau+3)} & X_{(\tau+3)} & X_3 \\ X_{(r-1)(\tau+4)} & \cdots & X_{(2\tau+4)} & X_{(\tau+4)} & X_4 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ X_n & \cdots & X_{n-(r-3)\tau} & X_{n-(r-2)\tau} & X_{n-(r-1)\tau} \end{bmatrix} \quad (4.18)$$

Une analyse en composantes principales (ACP) est réalisée sur la matrice de covariance entre les  $r$  descripteurs de  $X$ , ce qui permet d'extraire les valeurs propres  $\lambda$  et les vecteurs propres  $V$ . Les valeurs propres expliquent la part de variance associée aux éléments de la décomposition. Le premier vecteur propre, associé à la première valeur propre, est assimilé à la tendance de la série temporelle. En multipliant le premier vecteur propre par sa composante nous obtenons une estimation de la chronique.

La difficulté de cette méthode est le choix du décalage maximum entre les 2 séries extrêmes. Soit  $\Delta t = (r - 1)\tau$  le décalage entre 2 séries successives, si  $\Delta t$  est trop petit, la série prédite à partir du premier axe factoriel suivra, à peu de chose près, la courbe originale. Si le décalage est trop grand, elle va s'ajuster à une droite de tendance.

La méthode LOESS, est une régression polynomiale sur les  $k$  plus proches voisins. Nous considérons les  $k$  voisins à gauche et à droite d'une observation. Nous effectuons une régression par MCO d'un polynôme d'ordres 1 ou 2 pour récupérer la valeur prédite au temps  $t$ . Quand l'ordre du polynôme est 0, nous nous ramenons à une décomposition par moyenne mobile. Le paramètre à contrôler dans cette méthode est le nombre de voisins  $k$ . Comme pour les moyennes mobiles, nous pouvons considérer une fenêtre égale à une semaine, et donc prendre 6 termes consécutifs.

Le filtrage par régression consiste à ajuster une sinusoïdale sur la chronique. Le rapport des données à la fonction nous amène aux coefficients saisonniers.

Remarquons que chacune de ces méthodes nécessite de fixer un paramètre. Pour aboutir à une décomposition optimale, nous pourrions proposer de choisir les paramètres des différentes méthodes de sorte que le coefficient de variation de la série corrigée des variations saisonnières et de sa tendance, soit minimal. Le coefficient de variation est le rapport entre l'écart-type et à la moyenne. Il exprime la dispersion de la chronique autour de sa moyenne. Plus celle-ci est faible, plus la prévision sera précise. Pour estimer les paramètres des méthodes de lissage sous la contrainte évoquée, nous pourrions utiliser les méthodes d'optimisation vues dans la section 4.4.9.

L'usage de cette procédure montre que la saisonnalité des séries étudiées est bien d'ordre 6. Par contre, l'algorithme consistant à tester les ordres de 2 à 318 ( $6 \times 53$ ),

pour en définir l'optimal, est trop coûteux en temps d'exécution.

La comparaison des différentes méthodes sur un échantillon d'agences révèle que la méthode des moyennes mobiles est retenue 9 fois sur 10. D'autre part, des procédures de détermination de la saisonnalité plus appropriées existent : analyse spectrale, test de périodicité de Whittaker Robinson. Par contre, elles nécessitent une intervention humaine pour l'interprétation, alors que nous cherchons, ici, à automatiser complètement la procédure de modélisation.

#### **4.4.6 Evénements calendaires**

La méthode qui suit est appliquée sur les deux séries corrigées des valeurs extrêmes, de la tendance et des variations saisonnières obtenues par moyenne mobile et par la méthode de Buys-Ballot.

##### **4.4.6.1 Caractériser l'impact des jours fériés (97)**

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 8 jours ( $J - 4, \dots, J, \dots, J + 4$ ). Les conséquences sont différentes selon le type du jour férié, le jour de la semaine dans lequel il apparaît, et l'agence en question.

La récupération de la perte d'un jour de travail sera variable selon que ce jour est un lundi ou un samedi. 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 grandes et moyennes surfaces anticipent et demandent que la livraison soit doublée le jeudi. Les livraisons explosent le jeudi, et les expéditions gonflent le mercredi pour des livraisons en  $A$  pour  $B$ . Si le jour férié tombe un lundi, les grandes et moyennes surfaces anticipent un peu sur le vendredi précédent et récupèrent surtout le mardi. Si le 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 sera d'autant plus chargé.

Un jour férié est souvent synonyme de fête, ce qui engendre une augmentation de la consommation des ménages. Par effet de boule de neige, l'activité du transporteur se renforce. Mais cette hausse est variable selon qu'il s'agit de la Toussaint ou de Noël.

Les agences ne sont pas égales face à la hausse d'activité engendrée par un jour férié. L'agence de Bretanor, par exemple, est située à côté d'une usine « Côte-d'Or ». L'usine approvisionne tous les magasins de France en chocolat de Pâques. Cet approvisionnement commence des mois avant l'événement et monopolise une grande partie des ressources de l'agence TFE. D'autres n'ont pas ce client et subissent moins de poussées. Ou elles les subissent à d'autres moments. Canal froid à Nantes, par exemple, transporte le muguet du 1<sup>er</sup> avril sur l'Hexagone. La hausse d'activité n'est, alors, pas répartie sur les mois qui précèdent, mais sur la semaine qui précède. Le rush est alors d'autant plus violent.

Parfois, le 4<sup>e</sup> jour suivant un férié est aussi le 2<sup>e</sup> jour précédent un autre férié. C'est ce qui arrive en mai, entre la fête du Travail (1<sup>er</sup> mai) et la victoire 1945 (8 mai). Dans ce cas, il est difficile de séparer l'effet des deux jours fériés.

Pour tenter de séparer ces quatre phénomènes engendrés par la survenance d'un jour férié nous retenons, pour chacun d'eux, les informations suivantes :

- jour de la semaine (lundi, . . . , samedi) du férié,
- jour de la semaine des 4 jours précédents le férié,
- jour de la semaine des 4 jours suivants le férié,
- nom du jour férié.

La table 4.2 montre la matrice ( $T \times 12$ ) à laquelle nous aboutissons.

- 1 colonne pour le temps en dates,
- 10 colonnes pour le nom des jours fériés (lundi de Pâques, jeudi de l'Ascension, Pentecôte, 14 juillet, 15 août, 1<sup>er</sup> novembre, 11 novembre, 1<sup>er</sup> mai, 8 mai, Noël),
- 1 variable codée entre  $-4$  et  $4$  pour indiquer l'éloignement du jour impacté au jour férié,
- $T$  lignes pour le nombre d'enregistrement de l'historique.

Cette matrice (table 4.2) est convertie en tableau disjonctif complet pour servir de variables binaires afin d'ajuster un modèle de régression sur la série désaisonnalisée. La matrice binaire est appelée ( $X_1$ ).

#### 4.4.6.2 L'année comme variable calendaire

Le modèle de tendance attribue la même pente chaque année pour modéliser l'allure que prend l'activité sur une longue période (plus d'un an). Pour faire varier la pente en

Dates	Lundi de Pâques	Fête du Travail	...	pos
		Vendredi précédent un samedi 1er mai		-1
		Samedi 1er mai		
		Lundi suivant un samedi 1er mai		+1
...	...	...		
04/04/2007	Mercredi précédent			-4
05/04/2007	Jeudi précédent			-3
06/04/2007	Vendredi précédent			-2
07/04/2007	Samedi précédent			-1
08/04/2007	Lundi de Pâques			
09/04/2007	Mardi suivant			+1
10/04/2007	Mercredi suivant			+2
11/04/2007	Jeudi suivant			+3
12/04/2007	Vendredi suivant			+4
...	...	...		
28/04/2007		Jeudi précédent un mardi 1er mai		-4
29/04/2007		Vendredi précédent un mardi 1er mai		-3
30/04/2007		Samedi précédent un mardi 1er mai		-2
01/05/2007		Lundi précédent un mardi 1er mai		-1
02/05/2007		Mardi 1er mai		
03/05/2007		Mercredi suivant un mardi 1er mai		+1
...	...	...	...	...

TABLE 4.2 – Matrice des événements calendaires - Un champ date, un champ pour chaque évènement calendaire analysé, un champ pour la distante entre l'enregistrement et le jour férié.

fonction des années comprises dans l'historique de la série chronologique, nous construisons une matrice binaire, avec comme intitulé des colonnes, les années de l'historique. Cette matrice est appelée  $(X_2)$ . Elle est concaténée à la matrice  $(X_1)$  pour être utilisée dans le modèle de régression linéaire, et ainsi éliminer la tendance annuelle. Nous nous heurtons alors au même problème que celui évoqué lors de la détermination du modèle de la tendance. Comment, lorsque qu'on se place en année  $A$  pour prévoir les valeurs de l'année  $A + 1$ , déterminer la pente pour l'année  $A + 1$ . Pour y remédier, nous établissons la règle suivante.

Lorsque nous cherchons à prévoir des valeurs de janvier  $A + 1$ , la pente retenue est  $\alpha t + \beta An(t) + \eta$  dans l'équation 4.4 ou  $at + b$  dans l'équation 4.16. Après janvier, nous considérons avoir assez d'observations ( $\approx 24$  jours d'observations) pour estimer une pente de l'année  $A + 1$ . Au fur et à mesure que nous avançons dans l'année  $A + 1$ , la

pende se modifiera en fonction des nouvelles observations enregistrées.

#### 4.4.6.3 Le modèle de régression des événements calendaires

Nous utilisons les séries temporelles obtenues après déduction de la tendance et de la saisonnalité suivant les méthodes de décomposition de la section 4.4.4 et de la section 4.4.5. Les équations de régression s'écrivent :

$$\ln(R_t^{(1)}) = \lambda_1 X_t^{(EC)} + b_1 + U_t^1 \quad (4.19)$$

$$\ln(R_t^{(2)}) = \lambda_2 X_t^{(EC)} + b_2 + U_t^2 \quad (4.20)$$

$X^{(EC)}$  est la matrice des événements calendaires. Elle contient côte à côte les tableaux disjonctifs  $X_1$  et  $X_2$  évoqués respectivement dans la section 4.4.6.1 et la section 4.4.6.2. Les paramètres  $\lambda_1$ ,  $\lambda_2$ ,  $b_1$  et  $b_2$  sont estimés par la méthode *QR* (57).  $U_t^1$  et  $U_t^2$  sont les composantes stochastiques des séries  $\ln(R_t^{(1)})$  et  $\ln(R_t^{(2)})$ . Nous posons l'hypothèse que les séries  $U_t^1$  et  $U_t^2$  sont stationnaires, car elles sont issues de la série chronologique  $Z_t$ , de l'équation 4.3, dont nous avons déduit la tendance, la saisonnalité et l'effet des événements calendaires.

#### 4.4.7 Cycle

Les ouvrages décrivant les séries chronologiques mentionnent souvent le cycle comme composante de la série. Le cycle est un phénomène itératif de longue durée qui varie autour de la tendance à long terme et qui est lié à la conjoncture économique (prospérité, crise, dépression, reprise). Les chroniques étudiées ne comportent pas de cycles en tant que tels, car le phénomène de croissance ou de décroissance ne se répète pas sur une période constante. Il se peut aussi que les historiques des chroniques étudiées ne soient pas assez long pour repérer un cycle. L'article de Jean-Marie Dufour (47) évoque des cycles de 7 ans dans le loyer d'un terrain. Ce cycle a été trouvé par William Petty (1623-1687). William Stanley Jevons (1984) identifie un cycle de onze ans, dans l'activité commerciale.

#### 4.4.8 Modèle complet

Comme il a été dit dans la section 4.4.3, nous posons l'hypothèse que le schéma des séries temporelles étudiées est multiplicatif. Dans la mesure où les valeurs prises par

les chroniques étudiées sont positives, il est aisé de passer du schéma multiplicatif à un schéma additif par passage au logarithme.

Notons,

- $t$ , l'indice de temps, il correspond à une année  $a$ , une semaine  $i$ , et un jour  $j$ ,
- $Y_t$ , la chronique corrigée des valeurs extrêmes,
- $T_t$ , la tendance,  $S_t$ , la composante périodique,  $R_t$ , les résidus.

Alors,

$$\ln(Y_t) = \ln(T_t) + \ln(S_t) + \ln(R_t) \quad (4.21)$$

Et en utilisant la méthode préconisée dans la section 4.4.4, les composantes se développent sous la forme suivante :

$$\ln(T_t) = \alpha t + \beta An(t) + \eta_1 + \min(\alpha t + \beta An(t) + \eta_1) \quad (4.22)$$

$$\ln(S_t) = S_i^1 \times p_j + S_{i,j}^2 \quad (4.23)$$

$$\ln(R_t) = \lambda_1 X_t^{(EC)} + b_1 + U_t^1 \quad (4.24)$$

En utilisant la méthode préconisée dans la section 4.4.5, les composantes se développent sous la forme suivante :

$$\ln(T_t) = at + b \quad (4.25)$$

$$\ln(S_t) = \sum_{j=1}^6 \gamma_j S_t^j + \sum_{j=1}^{53} \phi_j S_t^j \quad (4.26)$$

$$\ln(R_t) = \lambda_2 X_t^{(EC)} + b_2 + U_t^2 \quad (4.27)$$

Nous avons montré que les séries temporelles décrivant l'activité des agences de transport peuvent être décomposées suivant au moins, deux méthodes. La décomposition nous permet de retirer de la série des valeurs observées, les composantes déterministes. Suite à cette action, il reste une série de valeurs aléatoires ou, du moins, inexplicées. Selon la méthode de décomposition utilisée, les séries des valeurs inexplicées ( $U_t^1, U_t^2$ ) diffèrent. La section 4.4.9 montre comment modéliser les séries des valeurs inexplicées dans le but de les extrapoler.

## 4.4.9 Prévision

Les procédures précédentes nous ont permis de séparer la tendance, la composante saisonnière et les événements prévisibles, des résidus.

Suite à la génération des graphiques présentant les corrélogrammes<sup>1</sup> de la série des

1. Mesure l'association entre une variable est ses valeurs décalées de  $k$  périodes.

résidus, il est à noter qu'il reste, selon les cas, 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 prévoir une telle série, nous avons le choix entre un lissage exponentiel simple, un lissage exponentiel de Holt-Winters et un lissage autorégressif à moyenne mobile. Ces derniers permettent d'exprimer la valeur d'une série temporelle à l'instant présent par une somme pondérée de ses valeurs passées et d'un bruit blanc. Les auteurs Box et Jenkins ont fortement contribué à développer ces modèles dans les années 1970 pour définir les modèles ARMA. Par contre, ces modèles nécessitent l'intervention d'un expert sachant interpréter les corrélogrammes, ou l'utilisation d'un algorithme coûteux de ressource informatique, pour identifier les paramètres  $p$  et  $q$  des propriétés autorégressives et de moyenne mobile. C'est pourquoi nous préférons les modèles exponentiels qui ont l'avantage d'être automatisables. Bien qu'elle est sensée ne plus exister dans le résidu, dans certains cas seulement, une saisonnalité journalière persiste. Nous favorisons donc le lissage de Holt-Winters (voir section 3.2.4) car il intègre un paramètre de saisonnalité.

#### 4.4.9.1 Recherche des paramètres « optimaux » d'un lissage de Holt-Winters

Rappelons les équations du lissage de Holt-Winters pour une chronique  $x_t$  quelconque (voir section 3.2.4).

Lissage de la moyenne :

$$a1_t = \alpha(x_t - S_{t-Pe}) + (1 - \alpha)(a1_{t-1} + a2_{t-1}) \quad (4.28)$$

Le lissage de la tendance,

$$a2_t = \beta(a1_t - a1_{t-1}) + (1 - \beta)a2_{t-1} \quad (4.29)$$

Le lissage de la saisonnalité,

$$S_t = \gamma x_t - a1_t + (1 - \gamma)S_{t-Pe} \quad (4.30)$$

$Pe$  désigne la période. Notons  $U_t^1$  et  $U_t^2$ , les séries temporelles des valeurs inexplicées obtenues suite à la déduction des composantes déterministes, suivant les deux méthodes de décomposition proposées. Le résultat des autocorrélogrammes des séries  $U_t^1$  et  $U_t^2$  nous apprend qu'il reste généralement une corrélation entre l'observation à la date  $t$  et celles aux dates  $t - 1$  et  $t - 6$ . Nous pouvons en conclure que la composante périodique d'ordre 6 n'a pas été suffisamment bien estimée. C'est pourquoi nous gardons le paramètre  $\gamma$  du modèle de Holt-Winters, pour lisser la saisonnalité. Par contre, nous annulons le paramètre  $\beta$ , car nous posons l'hypothèse que la tendance des séries  $U_t^1$  et  $U_t^2$  est nulle.

Pour parvenir à un lissage de Holt-Winter optimum, nous choisissons les paramètres  $\alpha$  et  $\gamma$  pour minimiser l'erreur quadratique entre les chroniques  $U_t^1, U_t^2$  et leurs estimations :  $\hat{U}_t^1, \hat{U}_t^2$ .

Les traitements suivants s'appliquent aux deux séries  $U_t^1, U_t^2$ . Mais pour ne pas répéter inutilement les équations, nous utilisons le caractère  $U_t$  pour désigner l'une ou l'autre des séries  $U_t^1$  et  $U_t^2$ .

Rappelons que la prévision des valeurs futures s'écrit, en remplaçant la série  $x_t$  par la série  $U_t$ , (voir section 3.2.4) :

$$\hat{U}_{t+h} = a1_t + h \times a2_t + S_{t-Pe+h} \quad \text{si } 1 \leq h \leq Pe \quad (4.31)$$

$$\hat{U}_{t+h} = a1_t + h \times a2_t + S_{t-Pe+2h} \quad \text{si } Pe + 1 \leq h \leq 2Pe \quad (4.32)$$

Par optimisation, nous cherchons à trouver les valeurs des paramètres  $\alpha$  et  $\gamma$  qui minimisent la fonction coût. Dans notre cas, la fonction coût est la somme des erreurs quadratiques :

$$f(U, \hat{U}) = \min \left( \sum_{t=1}^T (U_t - \hat{U}_t)^2 \right) \quad (4.33)$$

avec  $T$  le nombre d'observations de l'historique.

La procédure de minimisation se fait par l'algorithme L-BFGS-B (1997). Il adapte la méthode d'optimisation Quasi-Newton aux séries de grandes tailles et en ajoutant des contraintes d'encadrement « bound constraints ». Les contraintes sont que  $\alpha$  et  $\gamma$  soient compris entre 0 et 1. En partant d'une valeur initiale, l'algorithme itératif parcourt les points de l'espace des solutions, selon une méthode de gradient, défini par les contraintes, jusqu'à trouver les valeurs qui répondent au critère imposé. Le critère étant de trouver le minimum de la fonction coût. Pour une documentation plus complète sur les méthodes numériques utilisées pour résoudre des systèmes d'équations (voir (69)).

L'estimation des paramètres  $\alpha$  et  $\gamma$ , permet d'utiliser l'algorithme de Holt-Winters pour prévoir les valeurs futures des séries temporelles  $U_t^1$  et  $U_t^2$ . Pour ce faire, les valeurs initiales  $a1_0$  et  $S_0$  sont initialisées respectivement avec l'ordonnée à l'origine de la droite de tendance,  $MM_6(U_t) = At + B$  ( $MM_6$  désigne la moyenne mobile centrée d'ordre 6) et aux coefficients saisonniers, suite à la décomposition par moyennes mobiles.

#### 4.4.10 Recomposition de la prévision

Suite à l'extrapolation de la série  $U_t$ , il reste à recomposer la série prédite. Pour ce faire, les valeurs prédites par la méthode de Holt-Winters sont additionnées aux composantes déterministes que nous avons soustraites précédemment.

Soit,

$$\hat{U}_{t+h} = a1_t + h \times a2_t + S_{t-Pe+h} \quad \text{si } 1 \leq h \leq Pe \quad (4.34)$$

$$\hat{U}_{t+h} = a1_t + h \times a2_t + S_{t-Pe+2h} \quad \text{si } Pe + 1 \leq h \leq 2Pe \quad (4.35)$$

la prévision du jour  $t + h$ , calculée le jour  $t$ , donnée par Holt-Winters.

Nous y ajoutons,

- l'impact des événements calendaires ( $\lambda X_{t+h}^{(EC)} + b$ ).

Selon la première méthode de décomposition, nous ajoutons aussi,

- les coefficients saisonniers journaliers à  $t + h$  :  $S_{i,j}^2$ ,
- les coefficients saisonniers hebdomadaires à  $t + h$  :  $S_i^1 \times p_j$ ,
- la tendance à  $t + h$  :  $\alpha(t + h) + \beta An(t + h) + \eta_1 + \min(\alpha(t + h) + \beta An(t + h) + \eta_1)$ ,

ou selon la deuxième méthode de décomposition, nous ajoutons

- la tendance à  $t + h$  :  $a(t + h) + b$
- les coefficients saisonniers à  $t + h$  :  $\sum_{j=1}^6 \gamma_j S_{t+h}^j + \sum_{j=1}^{53} \phi_j S_{t+h}^j$ .

La recomposition amène à la série des valeurs prédites  $\ln(Y_{t+h})$ . En appliquant la fonction exponentielle, nous obtenons la prévision de la série des valeurs observées :

$$Y_{t+h} = \exp(\ln(Y_{t+h})) \quad (4.36)$$

#### 4.4.11 Combinaison des prévisions

Du fait des deux procédures de décomposition, nous obtenons deux résultats de prévision. Les erreurs de prévision obtenues par les deux méthodes ne permettent pas de conclure à la supériorité de l'une sur l'autre. C'est pourquoi nous combinons les résultats. L'objectif est de minimiser la variance des erreurs de prévision. La méthode choisie a été proposée par R. Bourbonnais et JC Usunier dans leur ouvrage « Prévision des ventes, théorie et pratique » (18).

L'erreur de prévision est la différence entre la valeur observée et sa prévision.

$$EP_{t+h} = y_{t+h} - \hat{y}_{t+h} \quad (4.37)$$

La prévision combinée ( $PC$ ) est une moyenne pondérée des deux prévisions individuelles  $P1$  et  $P2$ ,

$$PC = kP1 + (1 - k)P2, \quad k \text{ est le coefficient pondérateur, } 0 < k < 1. \quad (4.38)$$

L'erreur de prévision combinée,

$$EPC = kEP1 + (1 - k)EP2 \quad (4.39)$$

La variance de l'erreur de prévision,

$$V(EPC) = k^2V(EP1) + (1 - k)^2V(EP2) + 2k(1 - k)Cov(EP1, EP2) \quad (4.40)$$

Nous cherchons  $k$  qui minimise  $V(EPC)$ , en annulant la dérivée, la solution est :

$$k = \frac{V(EP2) - Cov(EP1, EP2)}{V(EP1) + V(EP2) - 2Cov(EP1, EP2)} \quad (4.41)$$

Si les erreurs de prévision sont corrélées alors,  $Cov(EP1, EP2) = 0$

Si  $k > 1$  ou  $k < 0$

alors, si  $V(EPC, k = 0) > V(EPC, k = 1)$

alors,  $k = 1$

sinon  $k = 0$

Dans la pratique, avant chaque prévision à une date  $t$ , nous recalculons les coefficients  $k$  sur un historique d'une année précédant la date  $t$ . L'historique comprend les données observées et prévues.

Clive Granger (58) démontre également que la variance de l'erreur de prévision combinée est inférieure aux variances respectives des deux erreurs issues de  $P1$  et  $P2$ . Par ailleurs, les prévisions combinées devraient être plus robustes, puisqu'elles ne dépendent pas fortement de la spécification d'un seul modèle ajoutent T.J. Jordan et M.R. Savioz (112).

Notons qu'il existe d'autres méthodes de combinaison de prévision issues de modèles différents. F. Li et G. Tkacz (113) comparent des méthodes de combinaison linéaire et non linéaire. Ils distinguent le cas où les erreurs de prévision des différents modèles sont corrélées, du cas où elles ne le sont pas.

Mentionnons aussi les méthodes ensemblistes de prévision utilisées plus particulièrement en météorologies. La méthode appelée « Ensemble Transform Kalman Filter » (15) analyse la covariance des erreurs de prévision.

#### 4.4.12 Redressement des prévisions

Les chroniques du poids ou du nombre de positions quotidiens passés à quai, en expédition, en distribution et au total, sont prévues séparément. Or l'observation à la date  $t$  de la chronique « total » doit être égale à la somme de l'observation de la chronique

« expéditions » à la date  $t$  et de l'observation de la chronique « distributions » à la date  $t$ . Pour répondre à cette contrainte, il faut redresser les valeurs prévues. Nous posons l'hypothèse que l'erreur de prévision de la somme est moins élevée que les erreurs cumulées des prévisions des séries temporelles « expéditions » et « distributions ». Ainsi, nous choisissons de garder inchangée la prévision de la série temporelle « total des marchandises passées à quai », pour corriger les deux autres.

Rappelons que les séries temporelles à prévoir sont :

- poids total en kg des marchandises passées à quai, sur une journée d'activité,
- poids expédié en kg des marchandises passées à quai, sur une journée d'activité,
- poids distribué en kg des marchandises passées à quai, sur une journée d'activité,
- nombre de positions totales des marchandises passées à quai, sur une journée d'activité,
- nombre de positions expédiées des marchandises passées à quai, sur une journée d'activité,
- nombre de positions distribuées des marchandises passées à quai, sur une journée d'activité.

La procédure d'ajustement est la suivante. Soit,

- $exp_{t+h}$  l'indicateur prévu en « expéditions » avant redressement pour la date  $t + h$ ,
- $dis_{t+h}$  l'indicateur prévu en « distributions » avant redressement pour la date  $t + h$ ,
- $tot_{t+h}$  l'indicateur prévu au « total » avant redressement pour la date  $t + h$ ,
- $\widetilde{exp}_{t+h}$  l'indicateur prévu en « expéditions » après redressement pour la date  $t + h$ ,
- $\widetilde{dis}_{t+h}$  l'indicateur prévu en « distributions » après redressement pour la date  $t + h$ ,
- $\widetilde{tot}_{t+h}$  l'indicateur prévu au « total » après redressement pour la date  $t + h$ ,

Alors,

$$\widetilde{exp}_{t+h} = tot_{t+h} \frac{exp_{t+h}}{exp_{t+h} + dis_{t+h}} \quad (4.42)$$

$$\widetilde{dis}_{t+h} = tot_{t+h} \frac{dis_{t+h}}{exp_{t+h} + dis_{t+h}} \quad (4.43)$$

$$\widetilde{tot}_{t+h} = tot_{t+h} \quad (4.44)$$

Cette méthode a l'avantage de conserver les proportions prévues entre la valeur de l'indicateur en expédition et sa valeur en distribution.

#### 4.4.13 Contrôle de la qualité d'ajustement par le test du bruit blanc

Une manière de vérifier que l'information déterministe contenue dans la chronique d'origine a bien été extraite consiste à tester l'indépendance des résidus. Le test choisi

est celui de Ljung et Box.

L'historique des séries temporelles étudiées est divisé en un échantillon d'apprentissage et un échantillon test (voir section 4.5.1). L'échantillon d'apprentissage permet d'estimer les paramètres du modèle. L'échantillon test permet de comparer la valeur réellement observée avec la valeur prédite. Les résidus sont donc calculés en comparant la valeur observée ( $y_t$ ) dans l'échantillon test et la valeur prédite ( $\hat{y}_t$ ) pour la même date.

$$\varepsilon_t = y_t - \hat{y}_t \quad (4.45)$$

$$(4.46)$$

Le test de Ljung permet de tester si une variable suit un bruit blanc. Un bruit blanc est la réalisation d'un processus aléatoire stationnaire.

Ainsi,  $u_t$  est un bruit blanc si pour tout  $t \in T : E[u_t] = 0, E[u_t^2] = \sigma^2$ , avec  $u_t$  et  $u_{t-h}$  indépendants si  $h \neq 0, t$  et  $(t-h) \in T$ <sup>2</sup>.

L'hypothèse nulle du test de Ljung est  $H_0$  : la série suit un bruit blanc, autrement dit les  $\kappa$  coefficients d'autocorrélation sont nuls.

La statistique  $Q = T(T+2) \sum_{\kappa=1}^h \frac{1}{T-\kappa} \tau_\varepsilon^2(\kappa)$  suit une loi du  $\chi^2$  à  $\kappa$  degrés de liberté.

$$\tau_\varepsilon(\kappa) = \frac{\frac{1}{T} \sum_{t=\kappa+1}^T (\varepsilon_t - \bar{\varepsilon})(\varepsilon_{t-\kappa} - \bar{\varepsilon})}{\frac{1}{T} \sum_{t=1}^T (\varepsilon_t - \bar{\varepsilon})^2} = \frac{cov(\varepsilon_t, \varepsilon_{t-\kappa})}{var(\varepsilon_t)} \quad (4.47)$$

est la fonction d'autocorrélation d'ordre  $\kappa$ . En d'autres termes, c'est la moyenne des corrélations entre  $\varepsilon_{t-\kappa}$  et  $\varepsilon_t$  pour  $t = \{1, \dots, T - \kappa\}$ . La valeur de  $Q$  dépend du nombre de retard  $\kappa$  sur lequel elle est calculée (36). Nous pensons que dans le cas de l'activité d'une agence TFE, un retard supérieur à 6 jours n'est plus significatif. C'est-à-dire que le résultat d'un lundi peut dépendre des 6 jours de la semaine précédente et donc du lundi précédent. Au delà et après dessaisonalisation, cela n'a plus de sens.

Si  $Q > \chi_{\kappa,0.95}^2$ ,  $H_0$ , l'hypothèse que la série est un bruit blanc, est rejetée au risque 5%.

Le test manque de puissance. Supposons qu'on ait vraiment un processus bruit blanc et admettons, en première approximation, que les  $\kappa$  statistiques de test soient mutuellement indépendantes. La probabilité de rejeter au moins une hypothèse  $H_\kappa$  est égale à

2. Les Séries Temporelles, Emmanuel César & Bruno Richard, Université de Versailles Saint-Quentin-en-Yvelines, Module XML et Data Mining - Mars 2006

$1 - (0.95)^\kappa$  et peut donc être bien plus grande que 0.05. Dans notre cas,  $\kappa = 8$ , la probabilité de détecter de l'autocorrélation qui n'existe pas est de l'ordre de 33%. A contrario, il arrive que l'hypothèse de bruit blanc soit acceptée alors qu'il y a de l'autocorrélation pour l'un ou l'autre retard. Pour bien faire, G. Mélard (89) conseille

- d'effectuer des tests individuels pour tous les retards de 1 à  $\kappa$  ; la statistique de test est  $\sqrt{T}\tau_\varepsilon(\kappa) \approx N(0, 1)$ ,
- d'effectuer le test global et de considérer son résultat en même temps que le nombre de rejets lors de tests individuels.

Dans le cadre de la prévision d'activité des agences de transport, le test du bruit blanc est un moyen de contrôler que le modèle de prévision utilisé, a retiré toute l'information déterministe des séries chronologiques étudiées. C'est un indicateur pour apprécier la justesse du modèle de prévision. D'autres indicateurs de performance des propositions de prévision seront abordés dans la section 4.5.

#### 4.4.14 Avantages et inconvénients

Les avantages et inconvénients de ce modèle sont :

Avantages	Inconvénients
prend en compte les spécificités d'un jour férié,	reste rigide dans l'attribution des périodes de saisonnalité (journalière et hebdomadaire),
corrige l'historique de sa tendance annuelle et celle à long terme,	méthode fruste qui considère que la saisonnalité ne peut pas évoluer,
automatise la recherche des coefficients optimaux,	ne prend pas en compte l'effet de facteurs externes relatifs à la conjoncture économique,
ajuste les 3 chroniques pour rendre des résultats cohérents,	compte un nombre de paramètres à estimer trop important au regard de la taille de l'historique (~ 5 ans). 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.
explique au décideur l'effet des paramètres sur la série à prévoir.	

TABLE 4.3 – Avantages et inconvénients du modèle de prévision « Horizons »

## 4.5 Validation du modèle mathématique proposé

### 4.5.1 Critères de comparaisons

Pour apprécier les prévisions, nous nous remettons à des critères d'évaluation. Ce sont des statistiques permettant de mesurer la qualité d'une prévision. Elles se basent sur l'écart à la réalisation (la moyenne d'erreurs de prévision antérieures), et l'incertitude entourant le modèle de prévision (la variance de l'erreur). Pour ce faire, il faut constituer une base d'analyse. Comme l'illustre le schéma de la figure 4.6, la base d'analyse sépare l'historique des valeurs observées de la série temporelle à prédire, en deux échantillons. Dans le cas de la description de l'activité des agences de transport, l'observation des données est souvent effectuée sur une durée de 5 ans (au mieux 7 ans). Les 12 mois les plus récents constituent l'échantillon test, lequel servira de référence pour évaluer la qualité des prévisions. Les 48 mois précédents constituent l'échantillon d'apprentissage sur lequel est formulé le modèle pour estimer les observations de la période récente.

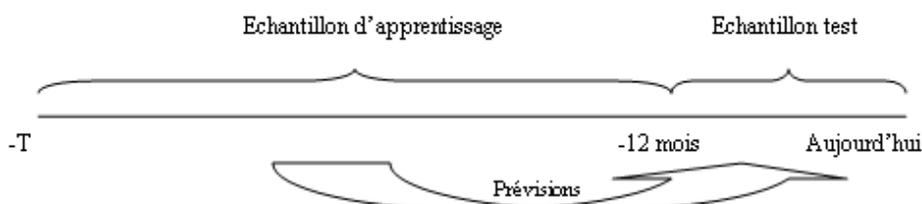


FIGURE 4.6 – Découpage de l'historique en échantillon d'Apprentissage et échantillon Test

Il est évident que les observations de l'échantillon test ne doivent pas servir à estimer les paramètres du modèle de prévision, sans quoi les valeurs prédites pourraient être la conséquence, et non la cause, de la valeur du résultat à prévoir, explique Tufféry (114). La performance du modèle de prévision se mesure en comparant les valeurs prédites par le modèle pour les dates de l'échantillon test ( $\hat{y}_t$ ), avec les valeurs observées à ces mêmes dates ( $(y_t)$ ). La comparaison de valeurs prédites et observées se fait par le biais des indicateurs de mesure suivants :

- *ME (Erreur Moyenne) : c'est la différence moyenne entre la valeur observée et sa prévision. Cette mesure donne la tendance d'un modèle à fournir des prévisions supérieures ou inférieures aux réalisations.*

$$ME = \frac{1}{T} \sum_{t=1}^T (y_t - \hat{y}_t)$$

- *MAE (Erreur Moyenne Absolue) : c'est la moyenne des différences absolues entre la valeur observée et sa prévision. Cette mesure considère l'importance plutôt que*

le sens des erreurs de prévision.

$$MAE = \frac{1}{T} \sum_{t=1}^T |y_t - \hat{y}_t|$$

- RMSE (Root Mean Square Error) : c'est la racine de l'erreur de prévision quadratique moyenne. En élevant au carré l'erreur de prévision moyenne, les grosses erreurs de prévision sont d'avantage pondérées que les petites.

$$RMSE = \sqrt{\frac{1}{T} \sum_{t=1}^T (y_t - \hat{y}_t)^2}$$

- BAR (Biais Absolu Relatif) : c'est la moyenne absolue des erreurs rapportée à la moyenne absolue des valeurs réelles. Une méthode de faible biais relatif sera dite juste.

$$BAR = MAE/\bar{y}$$

- DR (Dispersion Relative) : c'est la dispersion de l'erreur autour de la moyenne des valeurs réelles. Une méthode de faible dispersion relative sera dite précise explique R. Loutfoullah (78).

$$DR = \frac{\sigma_{\text{erreur}}}{\bar{y}}$$

- $r^2$  (coefficient de détermination) : c'est un indice statistique qui mesure la qualité de l'ajustement linéaire des estimations du modèle. Il donne la proportion de la variation des observations qui s'expliquent par les variations de la variable obtenue par le modèle. Un coefficient proche de 0 signifie un mauvais ajustement, alors qu'un coefficient proche de 1 montre un bon ajustement linéaire entre le réel et la prévision.

$$r^2 = \frac{\sum_{t=1}^T (\hat{y}_t - \bar{y}_t)^2}{\sum_{t=1}^T (y_t - \bar{y}_t)^2} = \frac{\text{variance expliquée}}{\text{variance totale}}$$

- Pourcentage d'erreur absolue inférieure à 5% et 10% : cet indicateur donne la proportion d'erreurs absolues comprises entre 0 et 5% et la proportion d'erreurs absolues dans l'intervalle [0% – 10%]. L'objectif étant de se rapprocher des 100% d'erreurs absolues incluses dans l'intervalle [0% – 5%].

- - Erreur minimal et maximal : ces mesures donnent une idée de la dispersion des erreurs. Notons que ces mesures peuvent être dues à des valeurs extrêmes. Il faut les manier avec précaution.

Il est difficile pour l'utilisateur des prévisions d'utiliser l'ensemble de ces mesures pour se faire une opinion sur la justesse des prévisions proposées. Dans la section 2.3.5, nous avons souligné comment classifier les mesures de performance selon le but recherché.

Suite au choix des mesures de performance, la section 2.2.11, définit l'erreur acceptable. C'est un juste milieu entre, le coût à investir pour améliorer les performances prévisionnelles, et l'impact des décisions prises sur la base d'une prévision erronée.

#### 4.5.2 Exemple : L'agence de transport TFE à Nantes

60% de l'activité de l'agence de transport de Nantes se font en distribution, contre 40% en expédition. L'importance de la distribution vis-à-vis de l'expédition s'explique par le fait que Nantes est une métropole de 250.000 habitants. L'agence de Nantes est une plate-forme étoile pour les expéditions en provenance de Bretagne, Loire Atlantique et à destination des régions du Sud-Ouest. 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 le 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 par jour. Juin est une période d'accalmie. Avec une activité autour de 900 tonnes/jour 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.

Les tableaux suivants comparent les résultats de prévision à un horizon d'une semaine, sur la période du 1er Janvier au 31 Juillet 2007. Trois chroniques sont étudiées : le poids quotidien passé à **quai** en expédition, en distribution et le total. Les modèles sont vulgairement appelés « TFE 2 » pour désigner celui illustré dans la section 3.2.3, « TFE 3 » dans la section 3.2.4 , et nouveau dans la section 4.4.

	Moyenne des erreurs de prévision : ME		
Modèle	Expédition	Distribution	Total
TFE 3	-11	-4	-7
TFE 2	-7	25	18
Nouveaux	-11	1	-10

TABLE 4.4 – Moyenne des erreur de prévision

L'erreur moyenne indiquée dans les tableaux 4.4,4.5 indiquent une sous- ou surestimation des prévisions. Rappelons que l'erreur de prévision s'écrit  $e_t = y_t - \hat{y}_t$ . Le « nouveau » modèle surestime en moyenne de 10 tonnes le poids quotidien total passé à **quai**,

Modèle	Ecart-type des erreurs de prévision : VE		
	Expédition	Distribution	Total
TFE 3	36	77	80
TFE 2	35	55	66
Nouveaux	23	54	60

TABLE 4.5 – *Ecart-type des erreur de prévision*

Modèle	Expédition			Distribution		
	RMSE	BAR	DR	RMSE	BAR	DR
TFE 3	38T	11%	15%	76T	12%	17%
TFE 2	36T	12%	15%	60T	11%	12%
Nouveaux	25T	8%	10%	54T	9%	12%

TABLE 4.6 – *Indicateurs de performance*

alors que le modèle « TFE 2 » sous-estime le même indicateur, de 18 tonnes. Une erreur moyenne nulle est un indicateur permettant de conclure à un modèle de prévision équilibré. En effet, une erreur moyenne nulle signifie que le modèle de prévision sur-estime autant qu'il sous-estime les valeurs réellement observées. Les tableaux 4.4,4.5 indiquent également l'écart-type des erreurs calculées. L'écart-type mesure la dispersion des erreurs autour de leur moyenne. Ce qui signifie que plus l'écart-type est faible, plus l'intervalle de confiance autour de la prévision est serré, et meilleure sera la confiance à accorder à la prévision. Ainsi, un modèle de prévision doit chercher à rendre l'erreur moyenne nulle et réduire écart-type des erreurs. Les tableaux 4.4,4.5 montrent que, selon ces deux mesures, le nouveau modèle de prévision remporte la victoire sur les deux autres modèles. Il n'y a que l'erreur moyenne obtenue par le modèle de prévision « TFE 2 », sur la prévision de la série temporelle appelée « expédition », qui soit meilleure que l'erreur moyenne obtenue par le nouveau modèle de prévision.

Les tableaux 4.6, 4.7 comparent les 3 modèles suivant le biais absolu relatif et la dispersion relative. Ces indicateurs se réfèrent à la moyenne des valeurs observées. Le nouveau modèle se trompe en moyenne de 7% par rapport à leur moyenne des valeurs observées de la série temporelle : « poids total des marchandises passées à quai ». Les erreurs varient de 9% autour de cette moyenne (estimée à 7%). Ces résultats sont meilleurs que ceux obtenus par les deux autres modèles. Cependant, l'objectif d'une erreur de prévision inférieure à 5% n'est pas atteint.

Les figures 4.7, 4.8 donnent un aperçu détaillé des poids quotidiens passés au quai de l'agence de transport de Nantes. Ils affichent les valeurs quotidiennes observées, suivies des estimations obtenues par les 3 modèles.

	Total		
Modèle	RMSE	BAR	DR
TFE 3	80T	9%	12%
TFE 2	68T	8%	10%
Nouveaux	61T	7%	9%

TABLE 4.7 – Indicateurs de performance, suite

Modèle	Expédition			Distribution		
	$r^2$	%EA <5%	%EA <10%	$r^2$	%EA <5%	%EA <10%
TFE 3	0.87	29%	51%	0.75	35%	57%
TFE 2	0.86	21%	40%	0.82	27%	58%
Nouveaux	0.94	39%	62%	0.84	43%	69%

TABLE 4.8 – Pourcentage de prévisions justes

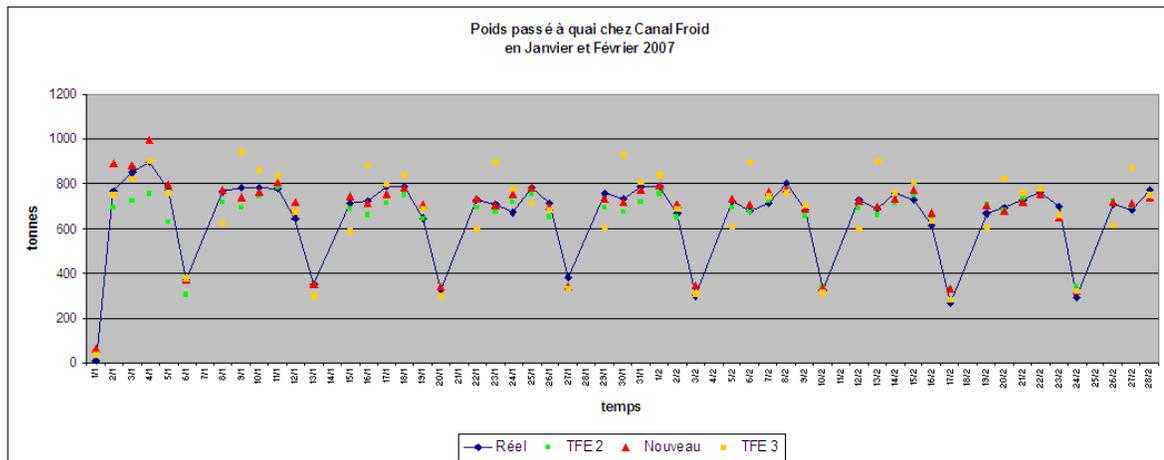


FIGURE 4.7 – Représentation graphique de la chronique des poids quotidiens passés à quai entre Janvier et Février 2007

Observons que le modèle « TFE 2 » ne considère plus le lundi de Pentecôte comme un jour férié. Or la réalité est différente. Suivant les politiques des agences et de leurs clients, le lundi de Pentecôte est chômé, ou non.

Les courbes des figures 4.7,4.8 révèlent que la qualité des prévisions varie suivant la bonne stabilité 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. Il apparaît que la

Modèle	Total		
	$r^2$	%EA <5%	%EA <10%
TFE 3	0.87	41%	67%
TFE 2	0.90	44%	69%
Nouveaux	0.92	52%	76%

TABLE 4.9 – Pourcentage de prévisions justes, suite

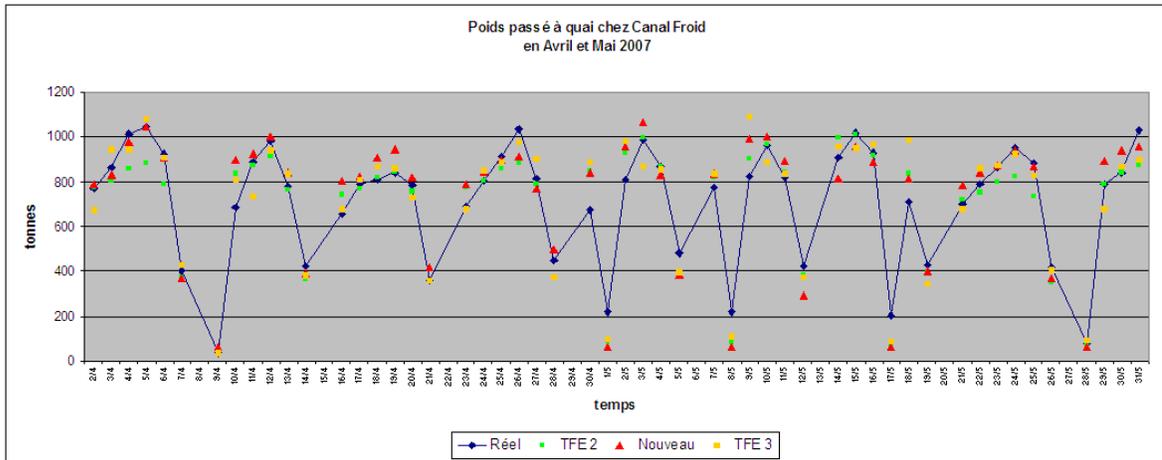


FIGURE 4.8 – Représentation graphique de la chronique des poids quotidiens passés à quai entre Avril et Mai 2007

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 de données historiques. Rappelons qu'elle estime un coefficient de perte ou gain par MCO pour chacun des 5 jours précédents et suivants un férié. Cette estimation se base sur le jour de la semaine du férié, or, en 7 ans, Noël a peu de chance de tomber plus de 2 fois un lundi, par exemple. Le cycle permettant de retrouver un même jour de la semaine pour un jour férié est de 28 ans. Cependant, un cycle de 11 ans, voire 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 Wikipédia (120). Ainsi pour calculer les pondérations, le processus se base sur échantillon d'au mieux 2 à 3 observations. D'autre part, ces observations sont éloignées dans le temps. Améliorer l'analyse des effets provoqués par les jours fériés est un objectif à atteindre.

Ajoutons que les tests de normalité (Shapiro-Wilk, droite d'Henry) et les tests d'absence d'autocorrélation (Ljung-Box) sont positifs sur la série des erreurs du nouveau modèle. D'après Kozicki et Tinsley (71), le non-rejet du test d'absence d'autocorrélation est interprété comme une indication que le modèle décrit bien la chronique.

Enfin, la comparaison des modèles devrait se poursuivre suivant des critères non statistiques comme la rapidité d'exécution des algorithmes.

### **4.5.3 Etude des variations des erreurs de prévision sur une trentaine d'agences**

Citons Méraud et Tymem (94) qui affirment qu'une méthode doit être jugée par les résultats auxquels elle conduit quand elle est appliquée à un nombre important de séries. La section 2.3.5 nous apprend qu'un indicateur de performance est le pourcentage d'erreur (PE). Le graphique de la figure 4.9 résume la distribution annuelle des PE quotidiens commis sur la prévision du poids des marchandises passées à quai dans 36 agences de transport en 2009. Pour chacune des agences, les PE quotidiens sont partagés en 4 parts égales. Le rectangle central comprend 50% des PE quotidiens. Il est partagé en deux par la médiane de la série. 25% des valeurs sont comprises entre les extrémités du segment, en pointillés, au-dessus du rectangle et 25% des valeurs sont comprises entre les extrémités du segment au-dessous du rectangle. Notons que, pour l'ensemble des agences représentées, la médiane des pourcentages d'erreur quotidiens est proche de 0. Dans la grande majorité des cas, 50% des pourcentages d'erreurs quotidiens sont compris entre -10% et 10%. L'écart moyen entre l'extrémité du segment supérieur et l'extrémité du segment inférieur est de 45%. En moyenne, sur l'année 2009, le pourcentage d'erreur fait sur la prévision du poids des marchandises passées quotidiennement à quai est de 2%. L'écart type de cette même série est de 17%.

Retenons que les écarts entre la prévision, du poids quotidien des marchandises passées à quai dans 36 agences de transport en 2009, et la réalisation se dispersent de manière homogène autour d'une moyenne proche de 0.

## **4.6 Choix et cartographie des outils informatiques**

Comme le préconise l'organigramme de la figure 1.8 dans la section 1.2.3, les données nécessaires à la prévision des indicateurs sont collectées et un modèle statistique de prévision est défini. L'étape suivante est l'informatisation des traitements. C'est l'objet des sections 4.6 à 4.7

Nous l'avons mentionné, les indicateurs à prévoir n'existent pas tels quels dans les bases

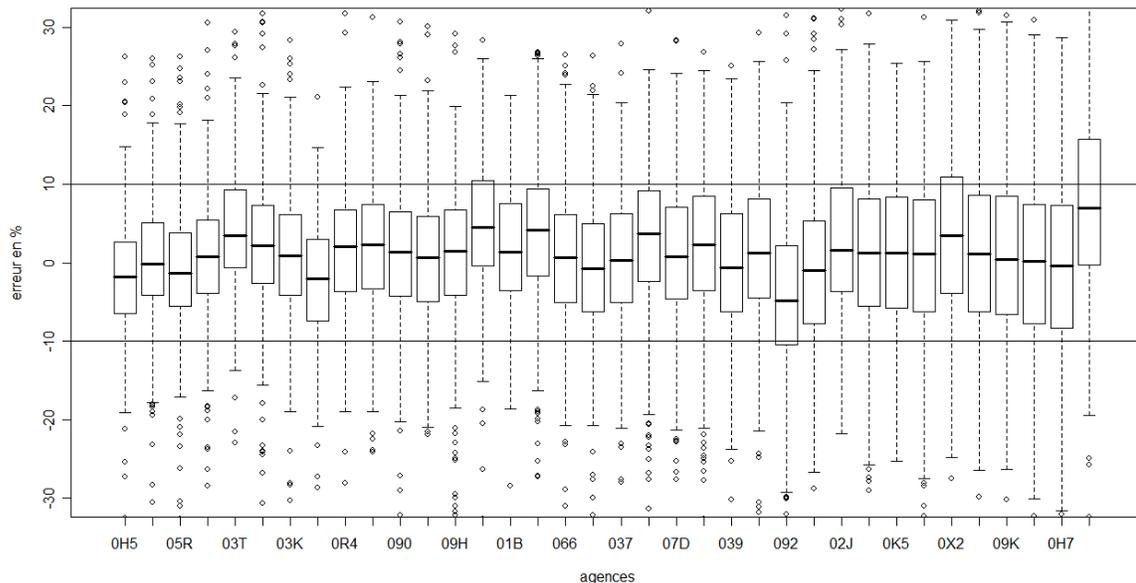


FIGURE 4.9 – Dispersion des pourcentages d’erreur quotidien commis sur la prévision du poids des marchandises passées à quai dans 36 agences

de données « métiers ». Il faut croiser des données provenant de diverses tables pour aboutir aux indicateurs voulus. Les outils informatiques permettant d’interroger des tables dans des bases de données, pour extraire de l’information, la manipuler afin de l’exploiter dans une autre base de données, se nomment ETL (Extract Load Transform). Pour traduire le modèle mathématique de prévision en programme informatique, il existe une diversité de logiciels. Dans cette section, nous allons nous attarder à argumenter le choix des outils informatiques utilisés pour mettre en place le système de prévision.

#### 4.6.1 Outil ETL de Pentaho™

ETL est l’acronyme pour Extract, Transform, Load. Ce sont des outils permettant de passer les informations générées dans les bases de données « métiers » vers l’infocentre. Les bases de données « métiers » sont structurées de telle sorte qu’elles répondent aux besoins opérationnels (application métiers). Alors que les tables et leurs liaisons, dans l’infocentre, sont agencées de manière à accueillir des informations de base de données « métiers » diverses, pour les regrouper, les ordonner, les croiser, les synthétiser et les restituer suivant des axes d’analyse. Franco et Lignerolles, 2000 (54), écrivent : « L’acquisition de données se déroule en trois phases : l’extraction, la prépara-

tion, et le chargement ».

L'**extraction** des données des bases « métiers », consiste à collecter les données utiles. Il est important de ne pas extraire de superflu, ce qui a pour conséquence de polluer inutilement le l'entrepôt de données et de ralentir les bases de données « métiers ». Il est également important de contrôler la qualité des données extraites. La qualité des données est un des piliers essentiels à prendre en considération dans toute approche d'un système d'information. Enfin, la performance des requêtes d'extraction est considérablement importante. Dans le souci d'éviter la saturation du réseau, des entrées / sorties de données ou de l'unité centrale, les requêtes d'extraction sont planifiées dans le temps et insérées dans une liste pour être ordonnées. Une requête ne peut par exemple, être exécutée qu'à partir du moment où la requête précédente dans la liste est terminée. La recherche de la performance en termes de rapidité d'exécution se fait en utilisant les caractéristiques de la structure physique des bases des données (taille, index, partitionnement) et les particularités du requêteur (MySQL, SQL Serveur, Oracle PL/SQL).

La **préparation** des données tient à respecter les définitions des indicateurs que nous cherchons à historiser dans l'infocentre. C'est la transformation des données métiers en données décisionnelles (pilotage des achats et des ventes, reporting financier, consolidation budgétaire, marketing, ...). Cette transformation peut s'avérer complexe et coûteuse en ressources machine. C'est pourquoi, comme lors de l'extraction, il est essentiel de chercher la performance des requêtes de transformation. Plus d'informations sur le sujet sont à trouver dans l'article de Fresnais, 1999 (55). La transformation consiste également à analyser la qualité des données sources, à vérifier, à éliminer les doublons, à contrôler la cohérence des données en les croisant, à examiner les raisons des valeurs manquantes, ...

Le **chargement** des données transformées nécessite de s'assurer que la structure de ces données correspond à celles des tables qui vont la recueillir (type, taille, unicité des champs). C'est une phase délicate et assez longue notamment en raison du volume des données. Il faut s'assurer des bonnes conditions de sa réalisation. Par exemple, nous veillerons à ne pas écraser des données existantes, mais à procéder à une mise à jour correcte de l'historique dans l'infocentre. Pour suivre le déroulement des procédures informatiques, il est vivement conseillé de générer des fichiers « log » ou d'ordonner des insertions dans une table de la base de données pour suivre le chargement pas à pas. Un fichier « log » est un fichier texte regroupant l'ensemble des événements survenus lors du chargement des données. Le but étant de reprendre le chargement à l'endroit où il s'est arrêté, si cela devait arriver.

Le processus d'extraction de transformation et de chargement peut être planifié et ainsi s'exécuter régulièrement. « *Il est important de savoir que la réalisation de l'ETL constitue 70% d'un projet décisionnel en moyenne. Et ce n'est pas pour rien, ce système est complexe et ne doit rien laisser s'échapper, sous peine d'avoir une mauvaise information*

dans l'entrepôt, donc des données fausses, donc inutilisables »<sup>3</sup>.

Le service d'Information Décisionnel auquel est rattaché le prévisionniste chez TFE utilise l'ETL attaché à Pentaho™. Pentaho™ est une solution d'informatique décisionnelle « open source » entièrement développée en Java. Pentaho™ utilise des outils et composants pour couvrir toute l'informatique décisionnelle dont la collecte et l'intégration des données qui se fait par l'ETL Kettle.

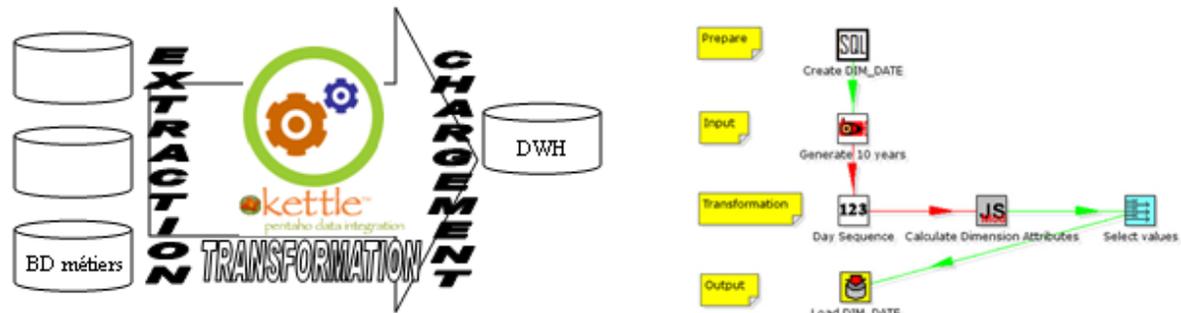


FIGURE 4.10 – Kettle gère des flux ETL (gauche) par une interface graphique simple (droite)

Comme le montre la figure 4.10, Kettle permet l'extraction, la transformation et le chargement des données. Kettle propose une interface graphique conviviale mettant en avant une boîte à outils complète, permettant d'interroger n'importe quelle source de données (bases de données, fichiers plats, tableur, . . .), de les manipuler en appliquant des règles de transformation complexes, d'afficher et d'intégrer dans un nouvel environnement. « Kettle fonctionne sur des bases de données de production possédant des millions d'enregistrements »<sup>4</sup>.

La figure 4.10 donne un exemple de l'interface graphique. Kettle donne des utilitaires nommés étapes, sous forme de « drag and drop ». L'exemple montre une étape d'extraction des données par une requête SQL, la flèche verte indique vers quelle étape est dirigé le flux. Suite à la première étape, une succession de différentes étapes se suivent (création de ligne, script java, sélection de valeurs) pour finir par une étape de chargement (voir <http://www.pentaho.com/index.php>).

Le système de prévision intègre cet outil pour créer et mettre à jour l'historique des données à prévoir. Il sert également à lancer l'exécutable du modèle mathématique de prévision. Nous décrivons en détail les développements dans la section 4.7.2.

3. Yazid Grim, conseiller d'Information Décisionnel (2008)

4. [www.pentaho.com](http://www.pentaho.com)

## 4.6.2 Logiciel de traitement statistique

Rappelons qu'une des conditions de départ pour ce projet de mise en production d'un système de prévision chez le transporteur TFE est d'utiliser des logiciels libres. Les logiciels libres sont dotés de licences diverses (GNU/GPL). Généralement, les logiciels libres sont gratuits et permettent un accès au code source, mentionne Guay, 2002 (60). Les logiciels libres publient régulièrement de nouvelles versions et regroupent un grand nombre de développeurs bénévoles à travers le monde. L'accès au code source est gage d'honnêteté scientifique, argumentent les défenseurs du logiciel libre. Il donne la possibilité pour tous de vérifier ce qui est réellement implémenté et de contrôler les résultats publiés.

Parmi les logiciels libres de traitement de données statistique (ou datamining), mentionnons RAPIDMINER, WEKA, TANAGRA, R, ORANGE.

**RAPIDMINER** (ou YALE) est développé par l'université de Dortmund. Le logiciel a son propre langage. Il propose une interface utilisateur graphique pour une exploitation statistique des données.

**WEKA** est développé par l'université de Waikato en Nouvelle-Zélande. Le logiciel est implémenté en Java et propose une plate-forme de programmation visuelle.

**TANAGRA** est développé par Ricco Rakotomalala, 2005 (101). C'est une plate-forme destinée à l'enseignement et à la recherche. Il est implémenté en Delphi. Il propose une plate-forme facile d'accès.

**ORANGE** est développé par l'université de Lubiana en Slovénie. Le code source est programmé en C++. Il propose des méthodes d'apprentissage automatique et de visualisation.

**R** est un système d'analyse statistique et graphique créé en 1996 par Ross Ihaka et Robert Gentleman. Le logiciel est généralement présenté comme la copie du logiciel commercial S-PLUS. Le code est principalement en C. R comporte de nombreuses fonctions pour l'analyse statistique et graphique.

RAPIDMINER, WEKA, TANAGRA, ORANGE sont très orientés vers les processus d'exploration de données. Franco et Lignerolles, 2000 (54), rappellent que l'objectif du data-mining est « *de tirer parti de l'information disponible pour constituer des modèles, pour découvrir des tendances ou pour anticiper l'avenir* ». Les logiciels de data-mining évoqués permettent l'utilisation des arbres de classification, des réseaux de neurones, la régression logistique, la segmentation, les règles d'association, etc. R propose les concepts de data-mining mais aussi une multitude d'autres fonctions statistiques. En cela il est plus généraliste.

Quand s'est posé le choix de l'outil pour faire les premières explorations de données, le modèle de prévision mathématique qui allait être utilisé par le système de prévision chez TFE n'était pas encore connu. Les techniques de fouilles de données exposées ci-dessus semblaient dans un premier temps répondre au besoin de prévision. Elles

permettent de classer les agences selon des critères communs, de déterminer une typologie, de construire des arbres de décision permettant ainsi d'anticiper l'activité, de lancer des ACP (Analyse en Composante Principale) et pour certains logiciels, d'utiliser des techniques de prévision. Morineau et al., 1994 (92), proposent cet enchaînement de techniques pour faire de l'analyse et de la modélisation des séries temporelles.

Le choix du logiciel s'est fait selon des critères mis en avant dans la section 2.3.4 et des souhaits formulés par le transporteur dans la section 1.2.4. Nous avons également pris en compte les avis résultant des enquêtes menées par Rakotomalala, 2005 (101) et DirectMarketing.net<sup>5</sup>. Il en découle les éléments suivants :

L'installation du logiciel sur le poste de travail ou serveur doit pouvoir se faire simplement. Elle ne doit pas nécessiter l'installation de serveur lourd au préalable. C'est le cas des 5 logiciels testés. Avec une petite nuance, WEKA nécessite l'installation des bibliothèques Java, ORANGE nécessite un interpréteur Python et R exige un compilateur C et un compilateur Fortran.

Les interfaces graphiques des logiciels RAPIDMINER, WEKA, TANAGRA et ORANGE permettent une utilisation intuitive du système, dès lors que l'utilisateur a de solides connaissances en data-mining. Ils proposent une panoplie d'outils sous forme d'icônes à glisser dans la fenêtre principale. C'est une interface drag & drop (en français, glisser - déposer). Dans ce type d'interfaces, l'utilisateur saisit un objet, puis le déplace pour l'amener à un autre endroit. En revanche R est un logiciel de programmation procédurale, il n'offre qu'un éditeur de texte pour saisir les commandes.

Le mode client-serveur permet à plusieurs utilisateurs d'utiliser le logiciel installé sur un serveur. Les logiciels cités, à l'exception de R, sont tous mono-utilisateurs c'est-à-dire qu'ils ne fonctionnent pas en architecture client-serveur.

La possibilité d'exécuter le programme en mode batch est primordiale. Un batch est un fichier texte dans lequel sont inscrites des commandes interprétées par le bash (ou shell). Un bash est un interpréteur de commande. Il joue le rôle de jonction entre l'utilisateur et le système d'exploitation. C'est l'avant-dernière couche du système d'exploitation. Vient ensuite le noyau. Il existe une multitude d'interpréteurs de commande sous Unix. Citons le Bourne shell (sh), le plus ancien et le plus utilisé, le Bourne Again Shell (Csh), le Korn shell (ksh), Perl, Python etc. Les commandes du fichier batch sont vérifiées et exécutées l'une à la suite de l'autre. L'une des commandes doit permettre d'exécuter le programme de prévision développé sur l'un des logiciels libres. En mode batch le programme s'exécute sans l'ouverture de l'interface graphique. L'avantage du mode batch est de pouvoir l'exécuter à partir de différents environnements, comme Pentaho™ pour ce

---

5. <http://www.dmnews.net/freeware.php>

qui concerne l'implantation du système de prévision chez TFE. L'ensemble des logiciels comparés permet d'être exécuté en mode batch.

Nombreux sont les serveurs de production du groupe STEF-TFE qui tournent sous Unix. Ils utilisent Bourne shell comme interpréteur de commande. Les logiciels doivent donc supporter les deux plateformes (UNIX et WINDOWS). C'est le cas des 5 logiciels testés.

L'historique des données à prévoir est entreposé dans des bases de données Oracle®. Les logiciels utilisés pour la prévision doivent donc permettre un accès et un interfaçage avec les SGBD (Système de Gestion de Base de Données). Les logiciels mentionnés présentent tous une interface avec SQL.

L'historique des données d'une agence contient un enregistrement par jour depuis l'année 2000, soit environ 3650 enregistrements fin 2009. Ce chiffre est à multiplier par le nombre d'agences (70) pour aboutir au nombre d'enregistrements (255.500) contenus dans la table de références du système de prévision. Le nombre d'enregistrements augmente avec le temps, c'est pourquoi les logiciels utilisés doivent permettre de traiter un grand volume de données, en un laps de temps estimé raisonnable. Rakotomalala, 2005 (101) compare les performances des logiciels TANAGRA, RAPIDMINER, WEKA, ORANGE pour charger et traiter un tableau composé de 2 millions d'observations et 41 variables. Le système de prévision à mettre en place n'a pas besoin de lire des données de cette ampleur, mais nous pouvons imaginer que le logiciel de traitement statistique retenu serve à d'autres projets. Les résultats des tests menés par Rakotomalala, 2005, concluent que TANAGRA et ORANGE sont les seuls à avoir pu charger et traiter le jeu de données. Nous avons chargé le même jeu de données avec R et avons appliqué les mêmes traitements. Il s'avère que dans la version de base, R n'est pas capable de traiter ce jeu de données en un temps estimé raisonnable. Mais d'après les développements des sociétés AT&T Labs, logi.cals et REvolution il semble que cela soit possible.

Les logiciels testés ne diffèrent pas significativement au regard de la qualité des algorithmes. Les résultats obtenus sont conformes aux attentes. Une étude de cas menée par Rakotomalala (101) a comparé les performances de méthodes de régression logistique des logiciels WEKA, TANAGRA, ORANGE, R, RAPIDMINER. Mise à part RAPIDMINER qui fournit « *une qualité d'optimisation très en deçà des autres logiciels* », il en résulte des résultats légèrement dissemblables d'un logiciel à l'autre. L'étude note que R est au-dessus du lot en termes de temps de calcul.

Rappelons que nous souhaitons mener, à l'aide de ces logiciels, des calculs de prévision. Par le biais de la régression, ils permettent tous de faire de la prévision éco-

nométrique. Ils permettent également l'analyse des séries temporelles. RAPIDMINER intègre des plugins pour traiter les séries temporelles. Il sera à même d'exécuter le modèle de prévision suggéré par Liu et al., 2001 (77). Les auteurs cherchent à automatiser des procédures d'extraction de connaissances et la fouille de données pour les appliquer simultanément sur grand nombre de séries temporelles. Les résultats permettent le paramétrage des méthodes de prévision de Box-Jenkins. Enfin, R est un logiciel de statistique complet. Il propose des bibliothèques incluant un grand nombre de méthodes statistiques aussi bien pour la fouille de données que pour le traitement des séries temporelles mentionnent Cryer et Chan, 2008 (34).

Un critère de choix est également l'existence d'un langage de programmation évolué. Celui-ci n'existe que dans R. C'est-à-dire qu'il est possible de programmer ses propres fonctions ou de reprogrammer des fonctions existantes pour les adapter au contexte. Dans la version de base, R dispose de la plupart des fonctionnalités utiles pour la statistique. Les possibilités deviennent énormes dès que nous utilisons les bibliothèques mises librement à disposition de tous. Elles sont développées par les utilisateurs et leur code source en R est lisible. Ces bibliothèques couvrent la statistique descriptive, unie et multivariée, l'économétrie, la finance, la biométrie, la prévision, etc. Grâce à son langage de programmation évolué, R permet l'enchaînement programmé de plusieurs algorithmes, ce dont aura besoin le système de prévision. La section 4.7.1 décrit un enchaînement de prévisions suivant plusieurs méthodes et combinaisons dans l'optique de réduire la variance des erreurs.

Il ressort de cette comparaison que les logiciels ORANGE, WEKA, TANAGRA et RAPIDMINER auraient été satisfaisants et surtout plus conviviaux et « easy-to-use » s'il s'agissait de rechercher un modèle de prévision pour une seule agence du réseau de transport TFE. Or il est question d'un modèle mathématique et de sa traduction algorithmique pour une utilisation dans un processus industriel. Ceci ne semble possible qu'avec le langage de programmation R. D'autre part, les techniques de prévision statistique endogènes et exogènes implantées dans les différents logiciels ne convenaient pas pleinement aux séries temporelles à prévoir.

A l'encontre des autres logiciels de data-mining testés, R ne peut pas être utilisé sans expérience de programmation et, encore moins, sans notions statistiques. Citons les propos de Christophe Pallier (chercheur au CNRS), Christophe Lalanne (chercheur à l'INSERM) : « *R comprend de nombreuses commandes, il est illusoire d'envisager utiliser ce logiciel sans lire un minimum de documentations. Notre expérience est que les premières heures d'analyse de données avec R nécessitent de fréquents recours aux documentations, mais lorsqu'on est devenu à l'aise, alors il n'y a pratiquement plus besoin de s'y référer* ». Le web offre une multitude de documentations pour bien débuter avec R et notamment le site principal ([www.r-project.org](http://www.r-project.org)).

### 4.6.3 Représentation des résultats

Le système de prévision doit être une application groupe. C'est-à-dire que toute personne du groupe STEF-TFE est susceptible de l'utiliser. Pour des raisons de coût, de disponibilité, de temps, de contrainte, de sécurité, les entreprises, et même les particuliers aujourd'hui, ne veulent plus installer de logiciels sur leur poste de travail, ni même de stocker des fichiers dessus. C'est pourquoi il est imposé de développer des applications accessibles en ligne. Utiliser des applications en ligne présente de multiples avantages. Il s'agit « *de pouvoir utiliser l'application directement dans un navigateur et plus uniquement localement* » précise David Bénard<sup>6</sup>. Il n'est plus besoin d'installer de logiciels, l'application est disponible à partir de n'importe quel poste de travail connecté au réseau et les travaux réalisés sont sauvegardés sur un serveur sécurisé.

L'interface utilisateur, créée à l'occasion de la mise en place du système de prévision, est en elle-même un projet de recherche pour le groupe STEF-TFE et sa branche informatique AGROSTAR. L'interface utilisateur doit être une application client-serveur mettant en œuvre les nouveautés qu'apporte le Web 2.0. Le Web 2.0 permet de créer des interfaces donnant la possibilité « *aux internautes d'interagir à la fois avec le contenu des pages, mais aussi entre eux* » souligne Huyghe, 2009 (64).

L'application web du système de prévision fait appel à des langages de programmation tels que MXML, HTML, PHP, Javascript. La mise en place des effets visuels est développée par une application internet riche (RIA, Rich Internet Application). Adobe©Flex Builder permet d'améliorer considérablement la consultation des données dans une interface Web et de créer des graphiques adaptés aux prises de décisions. Adobe©Flex Builder est un environnement de développement (IDE, Integrated Development Environment) fondé sur Eclipse<sup>7</sup>. Il permet de développer des applications utilisant les technologies Adobe©Flex, MXML, Adobe©Flash player, ActionScript. Après avoir rédigé son programme, le développeur le compile dans un fichier (.swf) lisible par n'importe quel serveur web. Le principe est simple d'utilisation et universel. Par contre, il faut se former aux langages cités.

Une page web est un simple fichier contenant du texte formaté avec des balises HTML. Les technologies CSS, PHP, Javascript sont interprétées et exécutées directement par le navigateur internet. Il n'a donc pas besoin de compiler les programmes. Dans le contexte de l'interface utilisateur du système de prévision,

- les codes HTML et CSS structurent et affichent les pages,
- le Javascript permet une interaction entre l'utilisateur et la page web,
- le PHP autorise l'interaction entre la page web, l'utilisateur et les bases de données.

---

6. <http://www.infos-du-net.com/actualite/dossiers/56-applications-internet.html>

7. [http://fr.wikipedia.org/wiki/Eclipse\\_\(logiciel\)](http://fr.wikipedia.org/wiki/Eclipse_(logiciel))

Nous n'avons pas utilisé d'interface graphique pour développer sous ces technologies, mais juste un éditeur de texte.

#### 4.6.4 Interfaçage Kettle / R / site web

La cartographie des applications du système de prévision tourne autour de l'ETL Kettle, de l'outil de développement statistique R, des bases de données Oracle, de l'entrepôt de données et de l'interface graphique en client léger. Ces applications ne sont pas installées sur les mêmes serveurs, mais sont pourtant dépendants les unes des autres. La figure 4.11 schématise la cartographie des applications utilisées pour mettre en place le système de prévision.

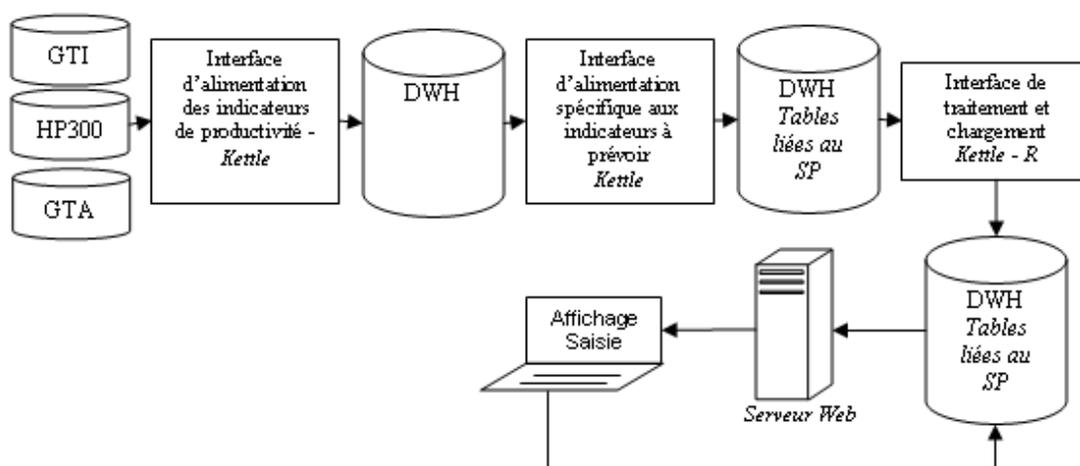


FIGURE 4.11 – Cartographie des applications retenues dans le système de prévision « Horizons »

L'application d'alimentation, de traitement et de chargement est hébergée sur le serveur ETLVIP02. Les tables liées au SP (Système de prévision) sont conservées dans la base de production de l'équipe décisionnelle (le DWH sur pkgwidprd.stef-tfe.fr). Les interfaces d'alimentation sont développées sous Kettle (voir figure 4.11). Elles fonctionnent quotidiennement (à 11h) pour mettre à jour l'historique des indicateurs à prévoir et, plus spécifiquement, les indicateurs de productivité de la veille. Les données métiers peuvent faire l'objet d'une modification pendant trois semaines. C'est pourquoi une mise à jour hebdomadaire fonctionne tous les lundis, à 15h00, pour prendre en compte les dernières modifications des bases de données « métiers » sur 3 semaines glissantes. Tous les

lundis, à 12h00, une application Kettle extrait, de la base de données, l'historique des indicateurs à prévoir et des données de paramétrage pour les charger dans un fichier texte, par agence. L'application développée sur R lit ces fichiers et pour chacun d'eux crée un nouveau fichier texte avec les données prévisionnelles. Des fichiers log sont également produits pour suivre le bon déroulement des opérations. Les fichiers texte produits par R sont lus par une application Kettle et chargés dans les tables de l'entrepôt décisionnel. Le site web créé est hébergé sur un serveur dédié : Cognai01. L'utilisateur du site doit s'authentifier avec son compte LDAP (Lightweight Directory Access Protocol). Il peut consulter, mais aussi saisir des données. Elles sont alors enregistrées directement dans les tables de l'infocentre, dédiées au système de prévision.

Par le biais des fichiers batch l'ensemble de ces processus sont lancés à la suite les uns des autres. « *Le Cron est le programmeur de tâches d'Unix/Linux. Il utilise les fichiers crontab pour déterminer la date et l'heure auxquelles une tâche va être exécutée* »<sup>8</sup>. Ainsi, le « Cron » permet d'automatiser périodiquement l'exécution des batchs.

## 4.7 Développement des modules du système de prévision

Suite au choix et l'assemblage des outils informatiques, il reste à écrire les programmes.

### 4.7.1 Rédaction du modèle mathématique en R

Nous n'allons pas, dans cette section, donner un cours sur l'utilisation du logiciel R pour pouvoir lire le code qui a été écrit. Par contre, il semble intéressant de se pencher sur ce que permettent les différentes fonctions développées, et la façon dont elles sont imbriquées dans le programme principal. Les 22 fonctions écrites sous R (voir figure 4.12) peuvent être utilisées indépendamment les unes des autres, par contre elles sont toutes utiles pour coller au modèle mathématique illustré en section 4.4. Elles sont regroupés dans un package. Un package en R, se présente sous la forme d'un fichier compressé qui regroupe des fonctions (ou scripts), leur documentation et un jeu de données servant à illustrer l'utilisation des fonctions. L'avantage d'un package est qu'il est facile à installer sur tous les postes équipés d'une version de R.

---

8. <http://blog.fedora-fr.org/paquet-fedora-du-jour/post/Le-pourquoi-du-Mercredi:-Crontab>

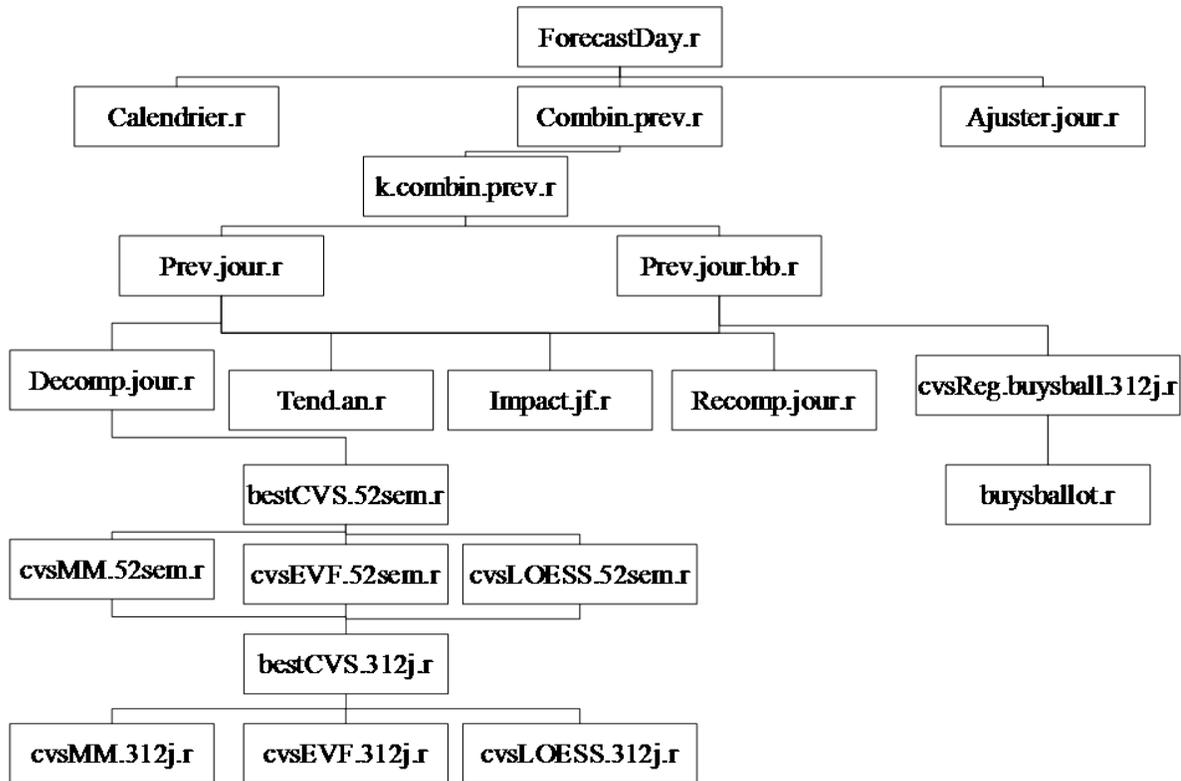


FIGURE 4.12 – Arbre du package R regroupant les fonctions de prévisions « Horizons »

Le programme principal (ForecastDay.r) est exécuté pour obtenir les prévisions des trois séries chronologiques ( $TS1$ ,  $TS2$ ,  $TS3$ ) avec la contrainte que la somme des deux premières soit égale à la troisième ( $TS1 + TS2 = TS3$ ). Il prend en argument d'entrée :

- Le chemin du fichier d'entrée contenant l'historique des séries chronologiques. Le fichier texte doit comprendre un enregistrement par ligne. Les valeurs, (code de l'agence, date,  $TS1$ ,  $TS2$ ,  $TS3$ ) sont séparées par un point virgule.
- Le chemin du fichier d'entrée contenant les coefficients de pondération pour les deux méthodes de prévision. Il comporte le code de l'agence, le coefficient de  $TS1$ , celui de  $TS2$  et de  $TS3$ . Ces coefficients sont recalculés régulièrement. Ils permettent de conjuguer les prévisions issues de deux méthodes différentes, en minimisant la variance des erreurs de prévision (voir section 4.4.11).
- Le chemin du fichier de sortie.
- L'horizon de prévision en nombre de jours.

Le programme commence par vérifier si les données d'entrée sont conformes aux attentes. Y a-t-il bien trois séries chronologiques, trois coefficients, un chemin de sortie ? Les formats date et numérique sont-ils valides ?

Il s'en suit un nettoyage des séries chronologiques (voir section 4.4.2). Les dimanches

sont supprimés. Les dates de la chronique à prévoir sont comparées avec celles de la suite ordonnée de dates obtenues après exécution de la fonction « calendrier ». Si des dates manquent dans la chronique à prévoir, elles sont ajoutées et une valeur théorique leur est attribuée. Les données atypiques sont localisées pour être corrigées.

La première date de prévision est celle qui suit la dernière date de la série chronologique fournie en paramètre d'entrée.

Les étapes suivantes vont se dérouler pour chacune des trois séries temporelles.

- Le programme vérifie que la moyenne des 50 dernières valeurs soit supérieure à 100. C'est pour éviter qu'il n'y ait trop de valeurs extrêmes. Rappelons que nous sommes dans un contexte de poids et nombre de positions. Il semble impossible pour une agence TFE de réaliser moins de 100 kg, ou même 100 positions, la journée, même pour un samedi. Cette valeur a été fixée de manière empirique, suite à plusieurs tests.
- La condition précédente réussie, le programme exécute alors les fonctions R « prev.jours » et « prev.jours.bb ». La première permet une prévision en utilisant la décomposition par moyenne mobile, alors que la deuxième fonction utilise la décomposition de Buys-Ballot (voir section 4.4.4 et 4.4.5).
- Le programme attribue à la prévision des jours fériés, la moyenne des jours fériés contenus dans l'historique.
- Les prévisions issues des deux méthodes sont combinées par la fonction « combin.prev ». Elle prend en arguments d'entrée les prévisions issues des deux méthodes et le coefficient de pondération associé à la série chronologique (passé en argument d'entrée du programme principal). Notons qu'aux prévisions sont associés les intervalles de confiance et qu'ils sont combinés de la même façon.

Ayant obtenu une prévision pour chacune des trois séries chronologiques, il reste à ajuster les prévisions pour qu'elles répondent à la contrainte  $TS1 + TS2 = TS3$ . Cet ajustement se fait par la fonction « ajuster.jour ».

A chaque date prévisionnelle, le programme ajoute la valeur de la chronique observée l'année précédente, pour la même semaine, et le même jour de la semaine.

L'exécution du programme génère un fichier texte à l'endroit donné, en paramètre d'entrée (path). Ce fichier texte contient le code de l'agence, la date, la prévision des trois chroniques ( $TS1$ ,  $TS2$ ,  $TS3$ ) et les valeurs en  $A - 1$ . Notons qu'il ne mentionne pas les intervalles de confiance. C'est, qu'à l'heure actuelle, ils ne sont pas valorisés. Nous avons, toutefois, pensé à les calculer pour répondre à une demande future.

Nous avons précédemment évoqué deux fonctions (« prev.jours », « prev.jours.bb ») sans approfondir le raisonnement qui a permis leur écriture. Elles prennent toutes deux comme arguments d'entrée, une suite de dates ordonnées dans le temps, des valeurs associées, un horizon de prévision en jours et le type de modèle (additif ou multiplicatif). Elles rendent un tableau contenant les dates et les valeurs prévues avec l'intervalle de

confiance, ainsi que l'erreur de prévision attendue.

L'erreur de prévision attendue est  $\left[ \sigma_{CVS} / C\bar{V}S \right] / 2$  (18), avec CVS pour désigner la série corrigée des variations saisonnières.

Bourbonnais et Usunier, 2007 (18), montrent, par cette formule, que l'erreur de prévision attendue peut être estimée par la moitié de la variation de la série CVS autour de sa moyenne. Il sera difficile d'atteindre une erreur de prévision inférieure à celle-ci.

La fonction « prev.jours » rend une prévision en utilisant la méthode des moyennes mobiles pour obtenir une saisonnalité. Elle procède dans l'ordre :

- Extraction des années à partir des dates.
- Exécution de la fonction « decomp.jours ».
  - Définir les jours fériés entre la date de début et la date de fin de la série chronologique, grâce à la fonction « calendrier ».
  - Attribuer une valeur moyenne aux jours fériés, en fonction du jour de la semaine.
  - Si le modèle choisi est additif, appliquer la fonction logarithmique sur la série chronologique.
  - Éliminer la tendance avec la fonction « tend.an ».
  - Pour éviter les valeurs négatives, soustraire de la série chronologique sa valeur minimum.
  - Appliquer une désaisonnalisation sur 52 semaines, avec l'aide de la fonction « cvsMM.52sem ». Cette fonction prend en compte que certaines années comptent 53 semaines alors que d'autres n'en comptent que 52. Elle corrige les semaines extrêmes. Elle rend ainsi la série régulière, suivant une saisonnalité de 52 semaines. Elle lisse la série en utilisant la moyenne ou la médiane mobile, selon le choix qui a été fait. Il en résulte la série corrigée des variations saisonnières (CVS), des coefficients saisonniers, une tendance issue de la moyenne, ou médiane mobile, et des résidus (voir section 4.4.4).
  - La deuxième désaisonnalisation fonctionne de la même façon. La fonction « cvsMM.312j » rend la série régulière de 312 jours (364j - 52 dimanches). Cette série commence par un lundi pour finir par un samedi. L'algorithme applique une désaisonnalisation par moyenne ou médiane mobile, suivant un schéma additif ou multiplicatif sur un pas de 312 jours.
  - Retourner dans un tableau, la série chronologique entrée en paramètre de fonction et sa CVS ainsi que l'ensemble des données utiles pour recomposer la série chronologique.
- Éliminer l'impact des jours fériés. Pour ce faire, utiliser la fonction « impact.jf ». Elle construit un tableau avec un enregistrement par date contenue dans la série

chronologique. Un enregistrement contient autant de champs que de jours fériés et fêtes. Les champs sont renseignés de façon binaire selon que la date de l'enregistrement est, oui ou non, le jour férié en question. Un autre champ indique la distance de la date vis-à-vis d'un férié. Cette distance va jusqu'à 6 jours avant ou après le férié. Le tableau contient également un champ indiquant le jour de la semaine de chacune des dates. Ces variables permettent d'appliquer une régression linéaire pour estimer leur impact sur la série chronologique. Cet impact est éliminé, par soustraction, dans le cas d'un modèle additif ou, par division, dans le cas d'un modèle multiplicatif.

- Calculer l'erreur de prévision attendue.
- Calculer une prévision à horizon  $h$ , donné en paramètre d'entrée de la fonction, à l'aide de la fonction `HoltWinters`. A la différence des autres fonctions citées, celle-ci fait partie des fonctions présentes dans la bibliothèque R. En fixant les paramètres  $\beta$  et  $\gamma$  à 0, nous obtenons la fonction de prévision par lissage exponentiel simple.
- Recomposer les valeurs prévisionnelles à partir de la prévision obtenue par lissage exponentiel simple. C'est-à-dire, qu'il faut ajouter ou multiplier selon le cas, les différentes composantes retranchées précédemment (effet jour férié, saisonnalité, tendance). Cette suite d'instructions s'exécute avec la fonction « `recomp.jours` ».
- Construire un tableau prévisionnel. Il comporte les dates à prévoir, les champs, année, mois, jour du mois, jour de la semaine, férié, prévision et intervalle de confiance. Retourner ce tableau.

La fonction « `prev.jours` », décrite ci-dessus, contient des variantes. Elle peut, par exemple, choisir la meilleure décomposition entre moyenne mobile, médiane mobile, `evf`, `loess` (voir section 4.4.5.2). Mais des tests empiriques ont montré que le lissage par moyenne mobile est la meilleure technique pour révéler la saisonnalité des séries chronologiques à prévoir (poids et nombre de positions quotidien d'une agence TFE). D'autre part, en termes de temps de calcul, tester les différentes possibilités s'avère trop long.

La fonction « `prev.jours.bb` » se comporte comme « `prev.jours` », à l'exception de la méthode pour éliminer la composante saisonnière. Elle utilise pour cela la fonction « `cvs-Reg.buysball.312j` ». Cette fonction rend la série chronologique **régulière** sur un cycle de  $52 \times 6 = 312$  jours l'année. Elle fait appel à la fonction « `buys.ballot` » pour calculer 52 coefficients saisonniers hebdomadaires et 6 coefficients saisonniers relatifs aux jours de la semaine, suivant la méthode de Buys-Ballot. De retour dans la fonction « `cvs-Reg.buysball.312j` », les coefficients sont soustraits, ou divisés, selon que le schéma est additif ou multiplicatif. Le résultat de la fonction est une série chronologique corrigée des variations saisonnières.

Pour pouvoir être transférables du serveur de développement au serveur de production, les fonctions sont regroupées dans un package. Après une installation de R (Windows ou Linux), le package est installé de manière simple pour que les fonctions soient immédiatement utilisables. Le programme principal a été écrit de telle sorte qu'il puisse être

exécuté à partir d'un fichier batch. Dans le fichier batch sont enregistrés les paramètres d'entrée et de sortie de la fonction « prev.jour » :

- le chemin (ou path) du fichier texte comprenant l'historique des séries chronologiques,
- le chemin du fichier comprenant les coefficients de pondération entre les prévisions issues des deux modèles,
- l'horizon de prévision
- les chemins des fichiers de sorties

Par le biais du fichier batch, les fonctions R vont être exécutées par l'ETL Open Source Kettle.

#### 4.7.2 Alimentation des tables sources (oracle©) par Kettle

Kettle est l'outil ETL de la solution open source de BI (Business Intelligence) Pentaho™. Pentaho™ est une solution d'informatique décisionnelle entièrement développée en Java. Cette solution a été choisie par Agrostar en 2007 pour remplacer le tout PL/SQL (voir section <sup>9</sup>).

Comme pour la section précédente, nous n'avons pas, ici, l'intention de donner un cours pratique sur l'utilisation de Kettle. Mais il semble important d'appréhender la démarche du développement. Elle nécessite une réflexion pour répondre à des problèmes sous contraintes tels que :

- la création d'un historique entre 2000 et 2005 à partir des données H3000, puis entre 2006 et 2009 à partir de GTI,
- le calcul des prévisions 2007, 2008, 2009, pour permettre le calcul d'indicateurs sur la fiabilité du modèle de prévision,
- la mise à jour régulière des coefficients de pondération entre les deux méthodes de prévision,
- la mise à jour hebdomadaire des prévisions quotidiennes, à horizon de 9 semaines,
- la mise à jour quotidienne du réel de la veille,
- l'insertion d'une nouvelle agence (création historique, prévision 2007 à 2009, prévision futures).

Les six problèmes ont pour contrainte commune la rapidité d'exécution, sachant que pour certaines requêtes, il s'agit d'interroger des tables de plus de 6 millions d'enregistrements. Il s'agit également de coder les programmes de telle sorte qu'il soit possible

---

9. Procedural Language/SQL

de changer rapidement et simplement des paramètres communs (agences, dates, indicateur à prévoir). L'exécution de la mise à jour des données réelles et prévisionnelles doit être automatisable et s'exécuter par un cron<sup>10</sup>. L'ETL Kettle joue un rôle de pont entre les bases de données « métiers », l'infocentre, le moteur de prévision développé sous R, les messages mail vers l'administrateur et les utilisateurs.

L'étape 1 consiste à extraire des bases HP3000 et GTI les données nécessaires pour aboutir à la définition des indicateurs à prévoir (poids et nombre de positions), comme nous l'avons exposé dans la section 3.1. Le projet « Horizons » a permis de pousser le transporteur à poser une définition commune au niveau de l'ensemble des agences pour décrire les indicateurs à prévoir. Cette définition a été traduite en requêtes SQL, issues de la BD HP3000, et en requêtes SQL, issues de la BD GTI. Les deux bases ne comportent pas les mêmes tables et encore moins les mêmes champs. Un gros travail a été réalisé avec les personnes de la Dex, et celles des agences pilotes, pour contrôler la véracité des données extraites. Les indicateurs ainsi extraits sont injectés dans la table HORIZONS de l'Infocentre. Cette table comporte pour chaque agence et pour chaque jour entre 2002 et aujourd'hui, le poids et le nombre de positions passées à **quai**, en expédition et en distribution.

L'étape 2 consiste à calculer des prévisions sur les années passées afin d'obtenir des statistiques sur la qualité des prévisions. Considérons que nous voulions calculer les prévisions de l'année 2008. L'algorithme commence par extraire l'historique des séries chronologiques à prévoir entre 2002 et 2007. Il exécute le programme de prévision, développé sous R, pour enregistrer les prévisions de la semaine 1 en 2008, dans la table HORIZONS. La boucle suivante ajoute à l'historique le réalisé de la semaine 1, pour prévoir la semaine 2, et ainsi de suite jusqu'à atteindre la semaine 52. Cette étape doit calculer, pour 70 agences de transport, les prévisions des deux années antérieures à l'année en cours.

Les étapes 1 et 2 sont des procédures qui durent plusieurs jours en temps de traitement. Pour donner un ordre de grandeur, il faut calculer 6 indicateurs quotidiens pour 70 agences sur 7 années, soit 917.280 données. C'est une charge de travail non négligeable pour des serveurs qui ont aussi d'autres procédures à faire tourner. C'est pourquoi il est impératif que le mécanisme fonctionne pour ne pas avoir à le relancer à de multiples reprises.

L'étape 3 est la mise à jour des coefficients de pondération se fait une fois par an. Cette fréquence est empirique. Elle ne résulte pas d'un calcul savant mais d'une gestion quotidienne des données. Une mise à jour hebdomadaire des coefficients de pondération ne montre pas une variation significative de leurs valeurs. C'est le cumul d'au moins 12 mois de données qui est susceptible de les faire changer. Le programme développé sous R calcule les prévisions des deux années qui précèdent la date en cours, en utilisant les

---

10. nom donné au programme qui permet, sur les systèmes Unix, d'exécuter automatiquement des commandes à un cycle de dates spécifiées à l'avance

deux méthodes de prévision, et suivant le principe décrit dans l'étape 2. A partir des erreurs de prévision issues des deux méthodes, le programme utilise l'équation décrite dans la section 4.4.11, pour mesurer les coefficients de pondération. Ils permettent de combiner les deux méthodes de prévision.

L'étape 4, consiste à mettre à jour, tous les matins à 11h, les indicateurs réalisés la veille. L'étape 5, est une mise à jour des prévisions. Elle se déclenche tous les lundis à 15h00 pour calculer de nouvelles prévisions journalières à horizons de 6 semaines en prenant en compte les derniers résultats réalisés (voir étape 4).

Pour donner au lecteur un aperçu des fonctionnalités de l'ETL Kettle les étapes 4 et 5 sont décrites d'un point de vue technique.

D'abord, Kettle donne la possibilité de créer deux types de processus :

- Les transformations sont une suite de traitements effectués sur les champs et les observations des données extraites d'une ou plusieurs BD. Les traitements sont des opérations de tri, de filtre, de conversion de données, etc.
- Les tâches, permettent d'exécuter une suite de transformations et de tâches. Ainsi une tâche peut en exécuter une autre. Elle autorise l'envoi de mails, l'importation de données ou le lancement de programmes externes par l'intermédiaire d'exécution de commandes shell. L'exécution de chaque étape peut être contrôlée, suivant le résultat des étapes précédentes.

La figure 4.13 illustre la tâche Kettle qui réalise l'étape 4. Elle montre une succession de transformations (carrés verts) et de tâches (carrés orange). Les deux premières transformations affectent un nom, une description et la date à laquelle s'exécute la tâche. Les trois dernières transformations créent un message d'alerte pour avertir du bon ou du mauvais déroulement de la tâche ainsi que l'heure à laquelle elle s'est terminée. Ces informations sont stockées dans une table de suivi des flux Kettle et ont pour but de tracer la tâche dès lors que le « cron » l'exécute.

La figure 4.14 illustre la tâche Kettle qui réalise l'étape 5. Cette tâche, appelée « hor\_maj\_hebdo », exécute une transformation nommée « maj\_horizons\_from\_indic\_age ». Elle crée 14 enregistrements relatifs aux 14 dates précédant la date en cours. Les enregistrements correspondant aux dimanches sont supprimés. Pour chacune des agences du groupe, et chacune des 14 dates, une requête SQL extraite d'une table appelée « indic\_age », les indicateurs nécessaires (voir 3.1.2) pour aboutir au poids et au nombre de positions passées à **quai**, en expédition et distribution. La donnée ainsi obtenue est insérée dans la table HORIZONS de l'infocentre. Relevons que la table « indic\_age » est elle-même mise à jour quotidiennement et que le traitement que nous décrivons est lancé après à la fin de cette étape. Cette condition est prise en compte dans le « cron » qui lance la tâche « hor\_maj\_reel\_quotidien ».

Suite à l'exécution de la tâche « hor\_maj\_hebdo\_sie », la transformation « hor\_alim\_hquai\_distri »

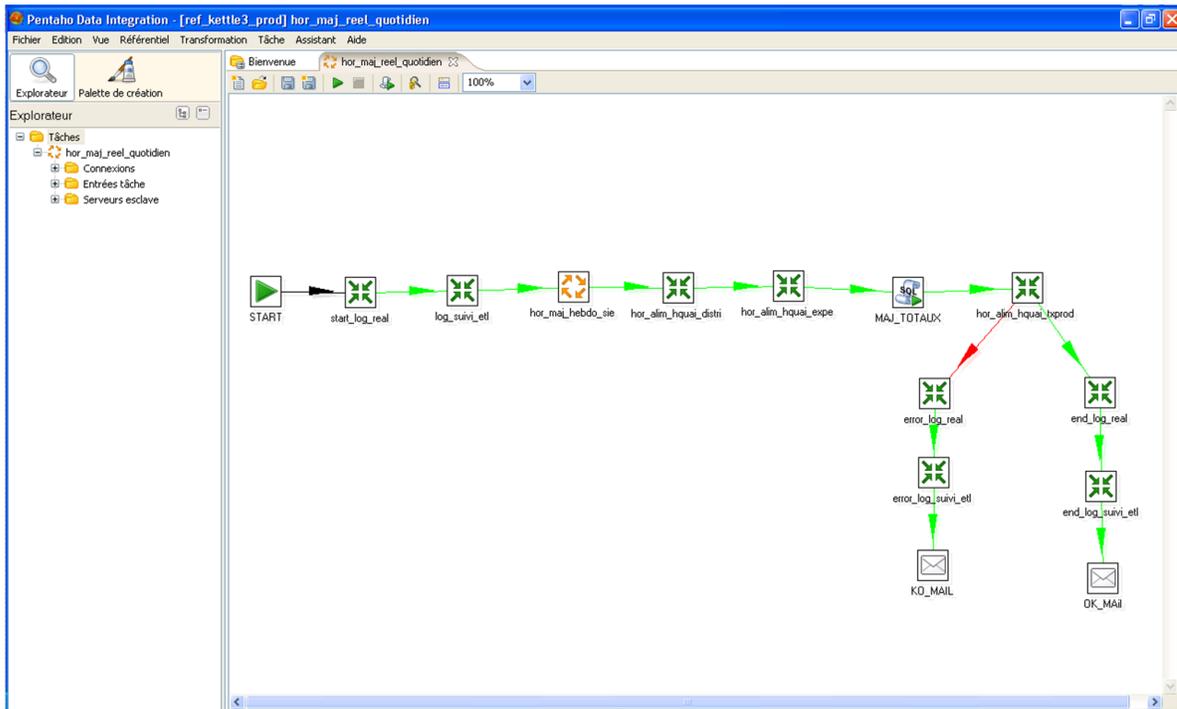


FIGURE 4.13 – Impression écran d'une tâche Kettle de mise à jour quotidienne des indicateurs

se charge (voir figure 4.15). Elle récupère les 14 dates précédant la date en cours et les codes de l'ensemble des agences du groupe. Pour chacune, elle fait tourner une requête SQL pour extraire les heures productives du **quai**. Les heures productives en distribution correspondent à l'activité nuit et sont badgées dans la nuit du jour  $J - 1$ , 22h30 à  $J$ , 7h00. N'affichant pas l'activité du dimanche dans le reporting final, les heures du dimanche en distribution sont ajoutées aux heures du lundi. La transformation « hor\_alim\_hquai\_distri » se termine en alimentant la table « heure\_quai » des valeurs calculés.

La transformation « hor\_alim\_hquai\_expe » fonctionne de la même façon que la transformation « hor\_alim\_hquai\_distri ». Mais les heures productives en expédition collent à l'activité jour, et sont badgées le jour  $J$ , 7h00 à 22h30. L'activité d'expédition du dimanche est comptée le samedi.

### 4.7.3 Interface graphique : partie émergée de l'iceberg

L'interface graphique développée en client léger a pour but de restituer les résultats prévisionnels de manière claire et utile. L'objectif est de faire de cette application un outil

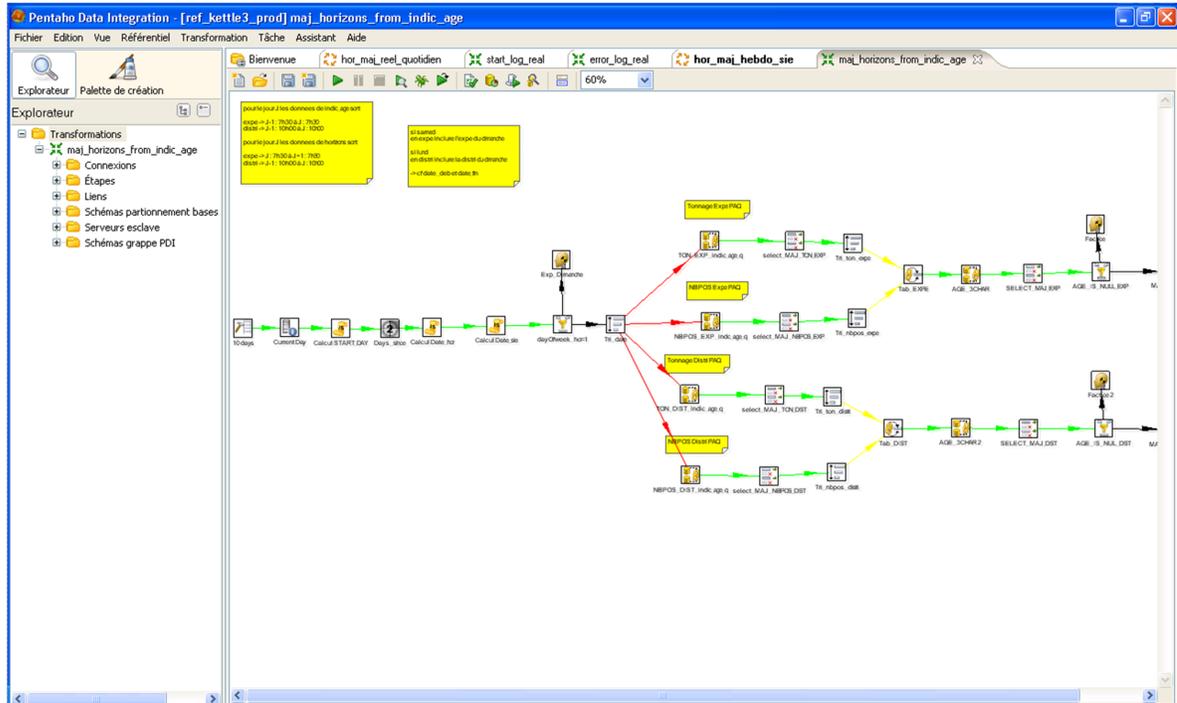


FIGURE 4.14 – Impression écran d’une transformation Kettle d’extraction des données et de leur transformation pour chargement

d’aide à la décision. La solution proposée apporte un volet scientifique au travers de la statistique, pour anticiper la demande. Plus qu’un outil de prévision, l’outil permet également d’analyser l’activité d’une agence en naviguant dans le temps. Ces options graphiques permettent de visualiser en un clin œil les données atypiques, les baisses et les hausses d’activité. Les tableaux de contrôle ont vocation d’entrer dans le détail. Des tableaux de bord renseignent sur la justesse des prévisions passées. La page planification offre une visualisation graphique et des éléments de planification, dont une proposition d’heures productives quotidiennes, à horizon de 5 semaines. L’interface graphique du système de prévision fournit une solution complète, puissante et facile à interpréter pour anticiper et planifier l’activité d’une agence de transport TFE. L’outil « Horizons », doit rendre le travail de planification plus simple, plus rapide et plus fiable.

Le développement de l’interface graphique est réalisé en HTML, PHP et Adobe Flex. L’interface doit être capable d’afficher dans un délai acceptable pour l’utilisateur, des informations extraites de tables Oracle® pesant parfois plus de 6 millions d’enregistrements. A l’image de la figure 4.16, l’interface graphique « Horizons » s’est inspirée de celle de Goggle Analytics®. La page d’accueil présente l’outil de prévision « Horizons ». Un texte reprend l’objectif de l’outil et les informations que l’utilisateur va trouver, en entrant. Des liens hypertextes conduisent à des pages d’aide et de présentation du projet. Ces pages répondent aux questions suivantes : Quelle est la genèse du projet ? Quelles

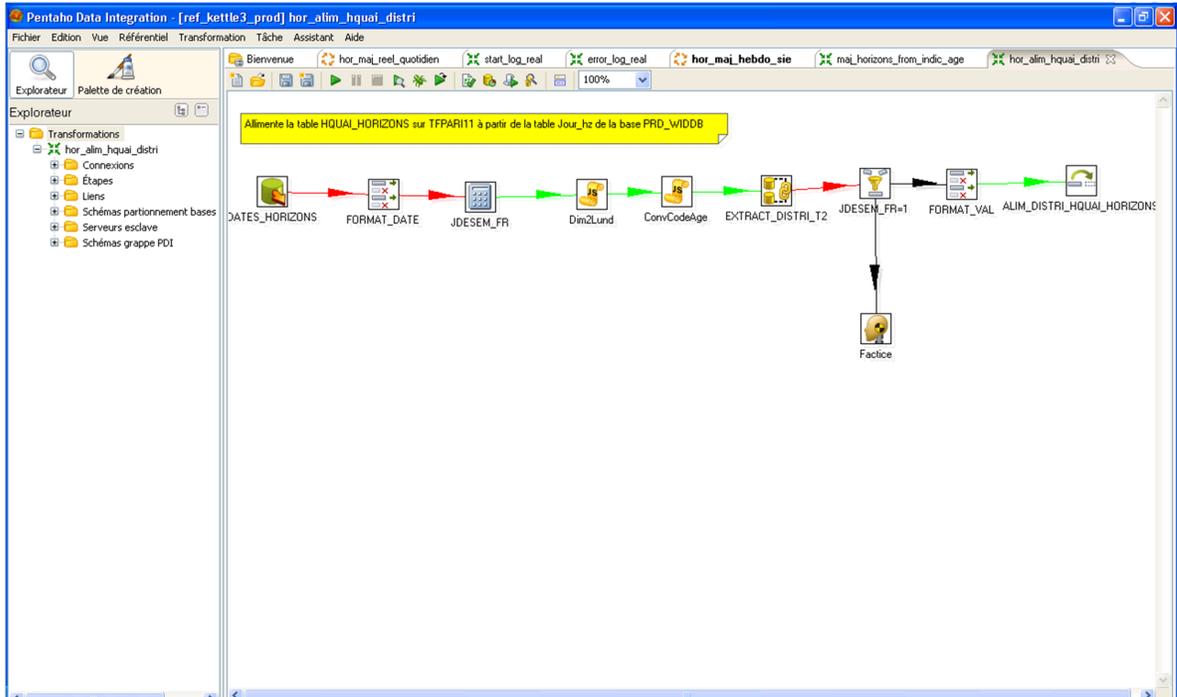


FIGURE 4.15 – Impression écran de la tranformation Kettle d’extraction des données et de leur transformation pour chargement



**Anticipez votre activité**  
 AGROSTAR Horizons vous fourni une prévision de vos expéditions et distributions en tonnes et en nombre de positions

Nouveauté : Pour vous connecter à Horizons, utilisez votre login AD / Contactez [W.Despaigne](#) en cas de problèmes

#### AGROSTAR Horizons

Vous permet en un clin d’œil, de visualiser votre activité quotidienne. Les indicateurs sont les poids et les positions des marchandises passées à quai de votre agence. Les valeurs passées et à venir vous sont présentées sous forme de graphiques et tableaux de bords.



[Dossier Horizons Intranet](#)

Quand bien même la technologie est devenue puissante dans le soutien aux prévisions, il est impératif de faire valider les résultats par les principaux responsables - marketing, ventes, exploitation, quai, RH, contrôle de gestion, ...

Quatre fenêtres vous sont proposées :

##### [Proposition de prévision](#)

Vous fournis les prévisions des deux semaines à venir, les résultats des semaines passées, et ceux de l’année précédente à la même époque.

##### [Indicateurs de confiance](#)

Pour chaque jour, vous accédez aux erreurs de prévisions moyennes et aux pourcentages de prévisions justes sur les trois et cinq semaines passées.

##### [Correction des prévisions](#)

Vous propose de corriger les prévisions proposées pour la semaine à venir. Ainsi votre expérience du terrain, contribue à optimiser le moteur de calcul.

##### [Indice de productivité](#)

Vous propose un indice de productivité (tonnes/heure) et les heures productives à la journée. L’indice sert à mesurer l’efficacité de votre quai.

Nom d’Utilisateur:   
 Mot de Passe:

Utilisez votre login et mot de passe Windows  
 Demande d’accès : [Despaigne](#)  
 (précisez nom et agence)

#### En savoir plus sur Horizons

- [Le projet](#)
- [Les prévisions](#)
- [Les cas d’études](#)
- [Les mathématiques](#)
- [Source des données](#)
- [La technologie Horizons](#)
- [Propriétés des prévisions](#)

FIGURE 4.16 – Page d’accueil du site web « Horizons »

sont les principales propriétés des prévisions ? Quelles méthodes mathématiques et algorithmiques ont été utilisées ? Quels sont les outils informatiques utilisés ? Quelle est

la source des données ? Quelles sont les agences pilotes ? Pour accéder aux données réelles et prévisionnelles, l'utilisateur doit s'identifier. Il saisi son « login » et son mot de passe de l'Active Directory (AD). C'est un service composé d'un annuaire qui recense l'ensemble des collaborateurs du groupe STEF-TFE. Le service AD vérifie la présence de l'utilisateur dans l'annuaire du groupe, et l'authentifie.

Après authentification, le modèle de page est découpé en 3 parties (voir figure 4.17).

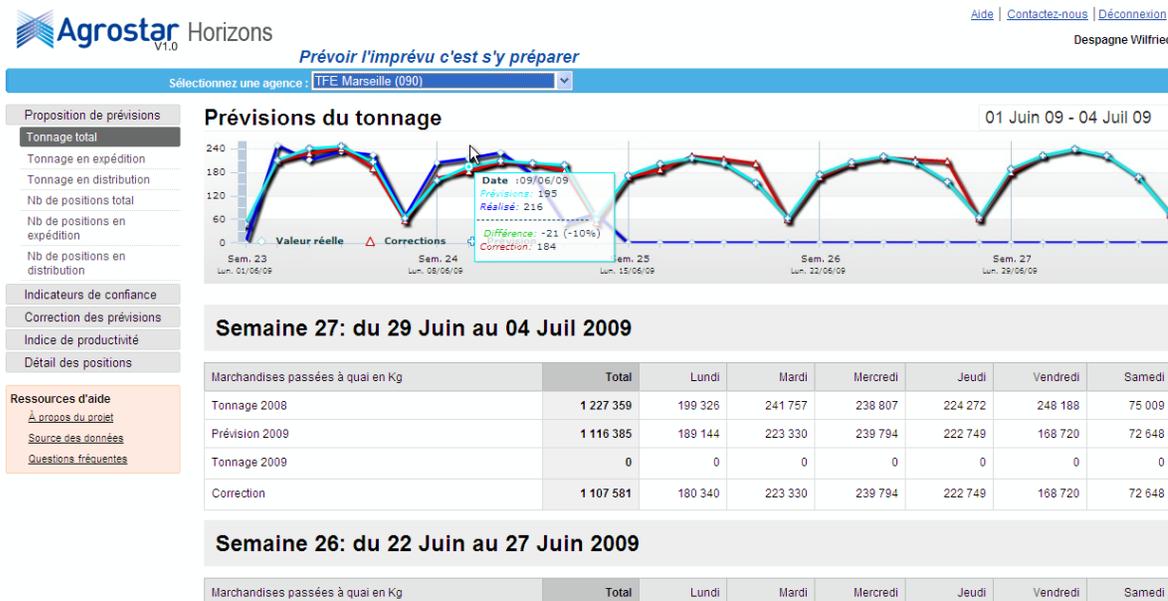


FIGURE 4.17 – Propositions de prévision journalière sur le site web « Horizons »

Une colonne gauche présente les menus de navigation. Une partie graphique, en tête de page, permet d'obtenir une vision d'ensemble rapide et synthétique des données détaillées au-dessous. La partie présentant les données détaillées est encadrée à droite du menu de navigation et en dessous de l'encadré graphique.

Les pages propositions de prévision (tonnage passé à quai, en expédition, distribution, nombre de positions passées à quai, en expédition, en distribution) comportent, dans l'encadré graphique, des courbes représentant la donnée réelle, prévisionnelle et corrigée. Grâce aux outils d'Adobe flex builder©, le graphique est animé. Lorsque l'utilisateur s'approche d'un point (intersection entre jours, en abscisse, et valeur de la donnée, en ordonnée) à l'aide du curseur de la souris, une boîte rectangulaire s'affiche pour indiquer le détail relatif à ce point (date, donnée réelle, prévision, correction, différence entre prévision et donnée réelle, différence en pourcentage par rapport au réel) (voir figure 4.17). Cette boîte apparaît en rouge lorsque l'erreur de prévision est supérieure à 10%, en vert dans d'autres cas. Les données détaillées sont rangées dans des tableaux. Un tableau est composé de 6 colonnes (les jours de la semaine du lundi au samedi) et de 4 lignes (donnée en A – 1, donnée réalisée, donnée prévue, donnée corrigée). L'utilisa-

teur a le choix de visionner entre 2, 5 et 10 semaines de données. Les 10 semaines lui donnent une tendance à long terme de son activité. Les 5 semaines permettent de visualiser les erreurs de prévision des deux dernières semaines et d'anticiper l'activité des 3 prochaines, par exemple. L'affichage de deux semaines est utile pour communiquer autour des prévisions. Le graphique et les deux tableaux prennent une page d'impression. Il est conseillé d'imprimer les résultats de la semaine précédente et les prévisions de la semaine à venir. Le rapport, ainsi affiché à la vue de tous, se veut transparent sur les erreurs de prévision passées, mais motive aussi les équipes en leur communiquant l'activité qu'elles ont soutenue. Connaissant le rythme de la semaine précédente, à la vue du rapport prévisionnel, elles estiment le rythme auquel elles doivent s'attendre, la semaine en cours.

Les tableaux des données détaillées fournissent, pour chaque jour de la semaine, le résultat de l'année précédente à la même époque (semaine, jour). Ce chiffre est essentiel. C'est un référent pour valider la prévision. En effet, le moteur de calcul reste mécanique, il peut s'enrayer. C'est pourquoi l'utilisateur doit valider les prévisions (voir section 4.8). Le résultat en  $A - 1$  est un référent qui le lui permet. Cet indicateur permet également de se positionner par rapport à l'année précédente. L'activité est-elle en baisse, stable ou en hausse ?

La deuxième information des tableaux est la prévision. La troisième ligne fournit le réalisé. Tous les jours, le réel de la veille est disponible. Il est ainsi automatiquement comparé au prévisionnel et contribue aux statistiques de confiance (voir section 2.3.5). La quatrième ligne affiche la valeur corrigée et validée par le manager.

Toujours dans un souci de transparence, l'interface graphique de l'outil « Horizons » donne un ensemble de statistiques pour décrire la performance des prévisions passées. « *Les prévisions comportent toujours des incertitudes et des imprécisions. Il importe d'en évaluer l'ordre de grandeur et de le communiquer aux utilisateurs* », (17). Les pages web « indicateur de confiance » fournissent un diagnostic pour l'évaluation des prévisions du poids et du nombre de positions passées à quai, en expédition ou en distribution. Ce diagnostic est possible pour n'importe quel intervalle de temps et pour toute période entre le 1er Janvier 2007 et aujourd'hui. Il s'agit, par des moyens graphiques (histogramme des erreurs), de visualiser rapidement les jours pour lesquels le modèle est le plus performant. Pour un non-statisticien, la variance des erreurs est plus accessible par l'histogramme que par la donnée brute. Des statistiques numériques (moyennes, écart-type, pourcentage de prévision juste) attestent les performances de la prévision (voir figure 4.18).

Rappelons que le modèle mathématique a été conçu pour s'adapter à l'activité de plus de 70 agences de transports. Ces agences ont chacune leur spécificité. Il est raisonnable de penser que le modèle ne les prend pas toutes en compte. C'est pourquoi il sera plus efficace pour les unes et moins pour les autres. Le diagnostic des performances est créé à cet effet. Il cherche à rendre les indicateurs statistiques compréhensibles et faciles à

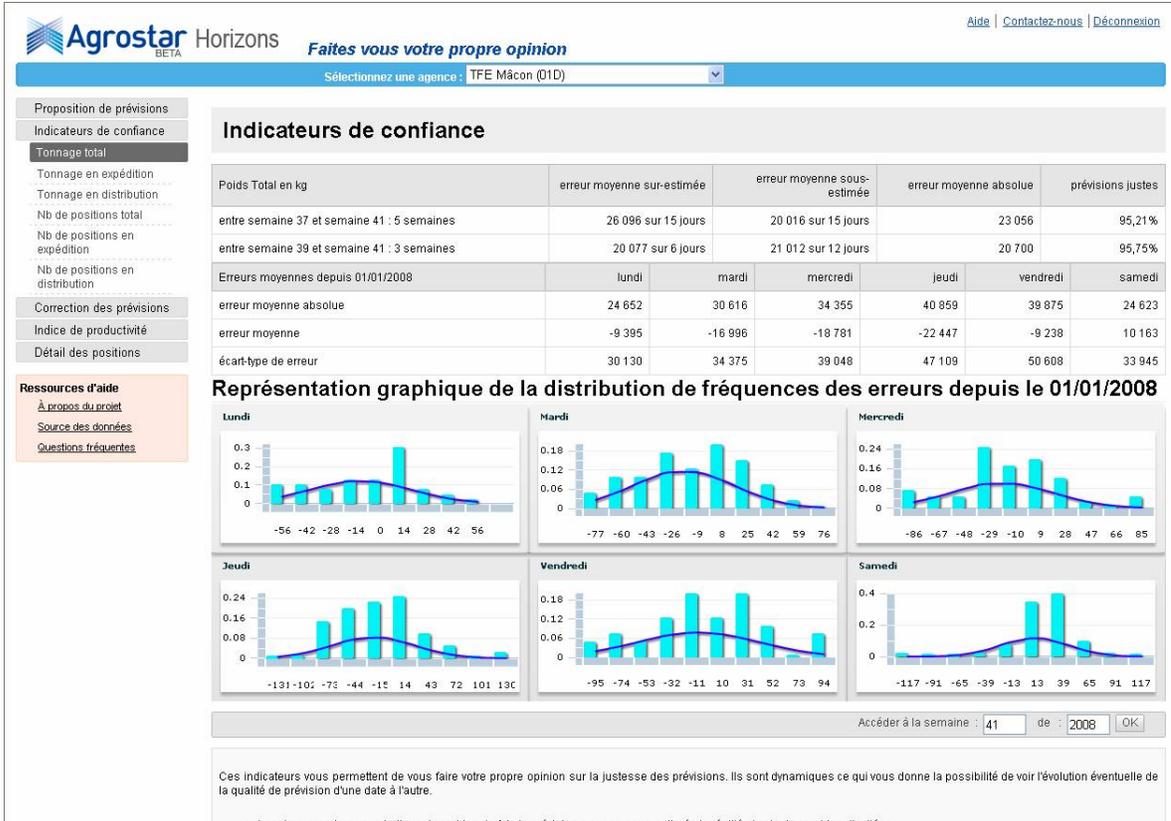


FIGURE 4.18 – Indicateurs de confiance aux prévisions sur le site web « Horizons »

interpréter. Suite à cette consultation, le manager est à même de se faire une opinion approfondie sur la confiance qu'il peut accorder aux prévisions proposées.

L'interface graphique de l'outil « Horizons » propose une fenêtre de prévision collaborative (voir figure 4.19). Elle permet de d'ajuster ou de valider chacune des prévisions proposées à horizon d'une semaine. A chaque ajustement, un motif doit être apporté. Ainsi, avec le temps, TFE recueille des informations cruciales pour décrire les raisons des pics d'activités. Les motifs d'ajustement sont à choisir dans une liste : gain portefeuille, perte portefeuille, événement commercial, férié, début du mois, fin du mois, opération saisonnière, prévision « Horizons », facteur météo, autres. Cette liste a été discutée avec les équipes du terrain. Toutefois, elle n'est pas exhaustive et peut s'allonger. Après six mois à un an de récolte d'informations, le prévisionniste pourra améliorer ses prévisions. En effet, sachant qu'à la même époque, un an auparavant, une promotion était traitée, le système devra envoyer une alerte pour demander si cette promotion se répétera et, le cas échéant, il faudra la prendre en compte dans le calcul des prévisions.

Un ajustement sur la proposition de prévision nombre de position ou poids total passé à quai, se répercute à 50% sur la prévision donnée en expédition et à 50% sur celle en distribution. Alors que si l'ajustement est fait sur la proposition de prévision en expédition

Sélectionnez une agence : TFE Marseille (090)

Choisissez un motif de correction

**Semaine 26: du 22 Juin au 27 Juin 2009 - Total**

Marchandises passées à quai en Kg	Total	Lundi	Mardi	Mercredi	Jeudi	Vendredi	Samedi
Tonnage 2008	1 791 910		302 159	359 981	380 304	325 673	118 828
Prévision 2009	1 545 128		275 951	302 673	307 296	263 871	87 413
Correction	1 777 414		307659	337364	350668	336496	93568

**Semaine 26: du 22 Juin au 27 Juin 2009 - Poids en expédition**

Marchandises passées à quai en Kg	Total	Lundi	Mardi	Mercredi	Jeudi	Vendredi	Samedi
Tonnage 2008	464 839	88 008	92 413	103 549	95 870	74 222	10 777
Prévision 2009	419 327	79 541	87 832	88 698	61 814	93 604	7 838
Correction	477 179	100000	96741	98681	86420	85995	9342

**Semaine 26: du 22 Juin au 27 Juin 2009 - Poids en distribution**

FIGURE 4.19 – Interface pour ajuster les propositions de prévision sur le site web « Horizons »

ou en distribution, la différence entre l’ajustement et la proposition de prévision est répercutée sur la prévision de l’indicateur total passé à quai. Dès qu’un ajustement est saisi, il apparaît dans la partie graphique et l’ensemble des utilisateurs peut le consulter.

L’outil veut être une ébauche de la planification. C’est pourquoi il intègre un volet « taux de productivité ». Ces pages fournissent, sur le même modèle que la prévision des poids, le réalisé et la prévision des heures productives et le taux de productivité. Le taux de productivité est le ratio entre le poids et le nombre d’heures productives. C’est un indicateur de productivité pour les responsables de site.

Après une première présentation de l’interface chez les agences pilotes, il est apparu que l’ergonomie du site n’était pas à remettre en question, bien au contraire, elle semble réjouir un grand nombre d’utilisateur. Par contre, une question surgit systématiquement : quelle est la source des données. Tant que les utilisateurs ne comprennent pas la façon dont a été calculée la donnée, et comment la rattacher à la multitude d’indicateurs déjà en place dans le système d’information, ils n’utiliseront pas l’outil de prévision. En effet, pour que l’outil soit accepté, il est crucial que les données affichées soient fiables, homogènes et d’une qualité maîtrisée. Il ne suffit pas d’avoir accès à une prévision, l’interprétation et l’utilisation de la prévision dépendent de la qualité de l’information rapportent Forslund et Jonsson, 2007 (53).

Deux nouvelles pages ont été créées pour amener l’utilisateur à accepter la donnée présentée. Une page d’aide apporte la définition de la donnée, sa règle de calcul et ses sources. Le dénominateur commun de toutes les données du transporteur est la position. C’est pourquoi, une autre page donne, pour chacune des agences, et pour les 7

jours passés, l'ensemble des positions prises en compte dans les indicateurs fournis. Les informations liées à la position et son type sont affichées. En expédition l'outil recense les positions typées TDE, TEX, TEXTT, TLMHgr, TLHgr, TLHct, TLMHct. En distribution, il liste les positions marquées TDL, TLL, TPQ, TLMLocal. C'est une information qui est très appréciée par les utilisateurs, car inexistante ailleurs, dans le système d'information. L'information liée à la position est mieux maîtrisée par les agents du terrain que le typage de la position. Les noms des **expéditeurs**, des destinataires sont des informations liées à la position. A partir de ces informations, le chef de quai sait tout de suite si la position est bien passée à son quai, à quelle heure elle est passée, sa provenance et sa destination. A partir de là il acceptera plus facilement la donnée fournie puisqu'il pourra l'approximer mentalement.

## 4.8 Mode d'emploi du système de prévision « Horizons »

Les prévisions sont utilisées à des fins différentes. Nous l'avons mentionné dans la section 2.1.4 et section 2.2.1, les chefs de services cherchent à en déduire un planning de leurs effectifs, les chefs d'agences élaborent le budget et fixent des objectifs, les commerciaux relancent les clients. Suivant l'utilité que l'utilisateur a des prévisions il va les lire différemment. Il va préférer la prévision mensuelle, à moyen terme, au détriment de la prévision journalière, à court terme, ou vice versa. C'est pourquoi l'application « Horizons » propose d'afficher les prévisions aussi bien à 2 semaines, qu'à 5 semaines et 10 semaines. La version « Horizons » v2.0 propose des prévisions mensuelles à horizon de 24 mois. Nous rappelons à l'utilisateur que l'incertitude augmente avec l'horizon des prévisions.

La fiabilité des prévisions dépend aussi de la fiabilité des données d'entrée. Des erreurs de saisie sur le poids des positions peuvent accentuer les erreurs de prévision. Enfin, la qualité des prévisions dépend de la volatilité des données autour de leur moyenne. Dès lors, les prévisions chiffrées ne doivent pas être prises pour argent comptant. Elles sont un support pour être débattues et corrigées lors de réunions regroupant le comité de validations. Il a pour objectif de consolider les prévisions au niveau de l'agence. Comme le rappelle Fildes et al. (51), la prévision statistique, en milieu opérationnel, doit être corrigée et confirmée par des experts métiers.

La figure 4.20 montre le déroulement des opérations. C'est un ensemble d'étapes nécessaires à la fabrication d'une prévision juste et acceptée par l'ensemble des acteurs. L'historique des séries à prévoir est extrait des bases de données. Le moteur de calcul des prévisions s'appuie sur l'hypothèse que l'activité des agences de transport, ainsi que leur environnement, sont stables. Autrement dit, l'avenir à toutes les chances de se dérouler comme le passé, en prenant en compte une évolution progressive du phénomène

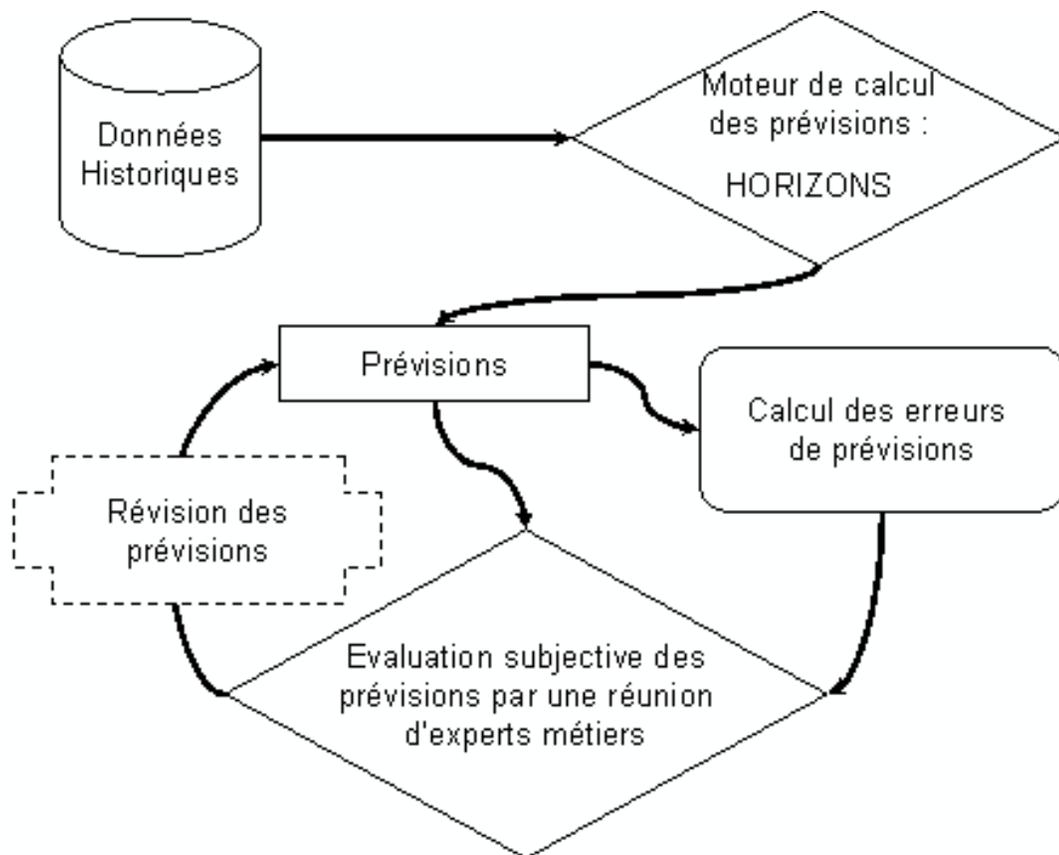


FIGURE 4.20 – Process d'utilisation du système de prévision « Horizons »

étudié. L'algorithme de prévision « *ne tient compte ni des modifications structurelles internes aux agences et au groupe, ni des changements conjoncturels du marché* » lié au transport, explique Bortolotti (17). G. Chevillon (29), ajoute que « *les méthodes de prévision ne sont malheureusement pas parfaites et n'ont pas vocation à devenir de purs exercices mathématiques sans contenu économique* ». Pour pallier à l'absence des données qualitatives, il est conseillé aux agences de transport de mettre à l'ordre du jour des réunions hebdomadaires, la prévision d'activité. D'habitude ces réunions hebdomadaires sont organisées pour traiter des problèmes passés, présents et à venir. Les responsables de services sont présents. Ils viennent du service commerce, du service de la direction, du service **quai** et du service de l'**exploitation**. Personne ne peut prétendre mieux connaître l'activité de l'agence de transport que chacun des responsables de services. Franco et Lignerolles, 2000 (54), précisent bien que « *les utilisateurs des prévisions doivent connaître le contenu informel disponible, maîtriser la valeur sémantique des données, la culture de l'entreprise et avoir une forte connaissance de l'environnement, notamment en terme d'historique d'actions et d'événements relatifs au contexte de l'entreprise. Ce recul et cette maîtrise du contexte sont des prérequis pour éviter au*

*mieux de perdre du temps sur certaines découvertes évidentes ou de tirer des fausses conclusions d'analyse de données bruitées* ». Pour faire court, les experts métiers utilisent leurs capacités de jugement pour valider la faisabilité des prévisions proposées. L'application « Horizons » leur fournit, en plus, un ensemble d'indicateurs leur permettant une prise de décision réfléchie. Ces indicateurs sont :

- la donnée réalisée en  $A - 1$ ,
- l'analyse du comportement du modèle sur la base des statistiques de fiabilité,
- la donnée prévisionnelle et son intervalle de confiance,
- des indicateurs de mesures permettant de juger de la pertinence des prévisions.

Pour les utiliser à bon escient, les experts métiers doivent être formés à la lecture de ces résultats. Par la suite, ces indicateurs croisés avec l'échange d'informations qualitatives en provenance des différents services de l'agence de transport, doivent permettre de s'approcher au plus près de la valeur réelle future. Passé leurs débats, les responsables de service valident ou ajustent la donnée prévisionnelle. Pour la valider, ils sélectionnent « prévision Horizons » dans la liste des motifs d'ajustement, sinon, ils choisissent un autre motif d'ajustement.

D'autre part, comme pour l'horizon de prévision, l'erreur de prévision n'est pas perçue de la même manière par l'ensemble des utilisateurs de prévision (voir section 2.2.11). D'abord pour le prévisionniste, une erreur de +5% est identique à une erreur de -5%. Elle est centrée en 0 et c'est ce qui l'importe. Par contre, l'exploitant accordera davantage d'importance à l'une ou à l'autre, suivant les moyens humains et matériels dont il dispose pour faire face à l'activité prévue. Une erreur positive (la prévision est supérieure au réalisé) coûte de l'argent à l'agence car elle aura planifié trop de personnel, par contre elle aura su répondre à l'activité. C'est essentiel pour son image de marque. Une erreur inférieure à zéro contraint l'agence à repousser l'activité du jour au lendemain, par faute de moyens. Elle aura planifié trop peu de main d'œuvre pour faire face à l'activité réelle. Ensuite, il est conseillé aux agences de transport de fixer leur tolérance, face à la moyenne des erreurs absolues, par rapport au réalisé (MAPE : Mean absolute percentage Error), en fonction de ce que cela représente en heures productives. Par exemple, si une erreur de prévision de 10% représente moins de 8 heures productives, elle est acceptable car elle n'a pas amené le chargé de planification à embaucher une personne de plus ou de moins pour la période considérée.

Il est également rappelé aux agences de transport, qu'une prévision peut être vue de différentes manières,

- est-ce un objectif à atteindre ?
- est-ce un chiffre qui devra être atteint quoi qu'il arrive ?
- est-ce le chiffre qui devrait être atteint si tout se passe sans imprévus ?
- est-ce une tendance ?

En fonction de la réponse, la communication des prévisions se fait différemment. Il est préconisé aux agences de transport de considérer la prévision fournie par l'outil « Horizons », avec son intervalle de confiance, comme un chiffre qui va se réaliser si tout se passe comme prévu. Néanmoins, les responsables de l'agence de transport peuvent retoucher les prévisions proposées par le système « Horizons » dans l'optique de les utiliser comme un objectif à atteindre.

Les responsables d'agence de transport ont l'habitude de parler du taux de productivité. Nous l'avons mentionné dans la section 1.1.2.2.3, c'est le ratio entre le poids des marchandises passées à quai et les heures productives sur le quai. Les responsables d'agence utilisent ce taux comme un indicateur de performance. Il leur permet de déterminer si les heures productives ont été utilisées correctement. Il faut savoir qu'une fois par semaine, les responsables de service achètent des heures productives en fonction de leurs besoins. Les besoins varient selon la charge de travail due aux fluctuations d'activité. Cette charge de travail comptée en heures productives est réalisée par des intérimaires. L'heure d'intérim coûte très cher à l'agence. C'est pourquoi les responsables d'agence souhaitent, par la maîtrise des prévisions d'activité, planifier, au plus juste, le nombre d'heures productives et donc le nombre d'intérimaires.

Comme pour les poids de marchandises, nous distinguons le taux de productivité en expédition de celui en distribution. Rappelons que, schématiquement, l'activité d'expédition correspond à l'activité du quai jour, et que l'activité de distribution correspond à l'activité du quai de nuit. Ainsi, en anticipant les heures productives, le système de prévision « Horizons » donne la possibilité de planifier l'équipe de jour et l'équipe de nuit. La prévision des heures productives est le résultat du ratio entre le poids prévu du jour  $J$  et la médiane des six derniers taux de productivité pour le jour de la semaine du jour  $J$ . C'est-à-dire que si le jour  $J$  prévu est un lundi, le taux de productivité prévu est la médiane du taux de productivité réalisé les six lundis précédents le jour  $J$ . Dans ce sens, la prévision des heures productives découle de la prévision des poids. Nous n'avons pas, ici, construit un modèle de prévision spécifique qui prenne en compte les différentes composantes de la série chronologique à prévoir. Néanmoins les résultats prévisionnels se rapprochent fortement des résultats réalisés. Les responsables d'agences surveillent de près les taux de productivité et aiment à les rappeler quand bon leur semble.

Après avoir été discutée et validée par les différentes parties, la prévision d'activité (poids, nombre de positions et taux de productivité) doit être publiée au niveau de l'agence de transport. Pour communiquer sur la prévision et motiver les équipes, le module « Horizons » permet d'afficher sur une même page, pour la semaine en cours et la semaine à venir, le graphique et les tableaux de bord, de la prévision et de l'activité réalisée. En imprimant cette page et en la diffusant dans l'agence de transport, les employés peuvent comparer le réalisé et la prévision des derniers jours. Ainsi dans un premier temps, les

employés comparent l'activité affichée et la charge de travail qu'ils ont réalisées. Dans un deuxième temps, ils visionnent la prévision des jours à venir et se font, ainsi, une idée précise de la charge de travail qui leur est demandée. D'autre part, les salariés sont satisfaits quand ils voient que l'activité réalisée a atteint ou dépassé la prévision et le sont moins dans le cas contraire. Cet affichage de l'activité passée et des prévisions a notamment pour but d'impliquer un peu plus les salariés à la bonne marche de l'agence.

Non seulement l'outil « Horizons » donne les prévisions de l'activité, mais il donne aussi la possibilité, a posteriori, d'analyser l'activité en naviguant dans le temps. L'utilisateur peut exporter les tableaux hebdomadaires dans un intervalle de temps donné et les analyser par le biais d'un tableur. La partie graphique permet également à l'utilisateur d'identifier rapidement les grosses erreurs de saisie sur le poids des positions. C'est le cas lorsque la courbe des poids forme un pic à l'image de la figure 4.21.

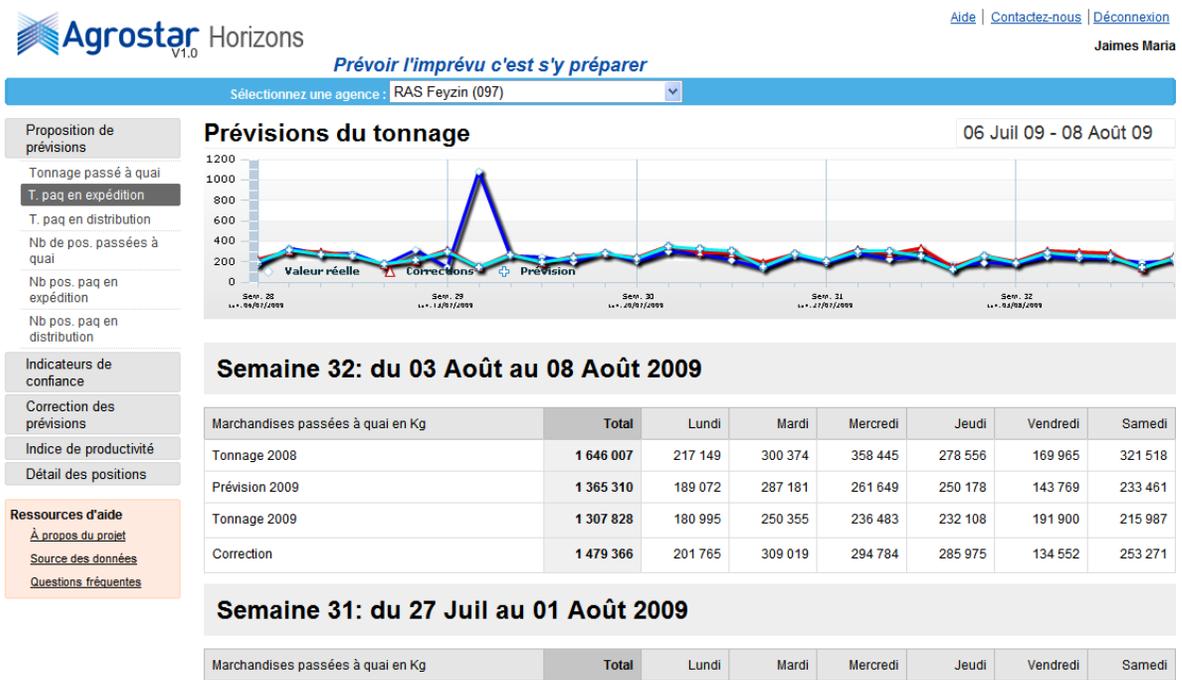


FIGURE 4.21 – Valeur extrême identifié rapidement grâce à l'interface web « Horizons »

## 4.9 Retour sur investissement

La section 2.1.4 donne des pistes sur la façon d'estimer le retour sur investissement, suite à la mise en place d'un système de prévision. Nous avons vu que le montant in-

vesti est souvent facilement identifiable. En revanche, l'argent gagné l'est moins. Nous voulons montrer que non seulement le système de prévision « Horizons » est un outil complet qui a été industrialisé au niveau d'un groupe d'envergure internationale, mais que, par ailleurs, le groupe retire de substantiels bénéfices à l'utiliser.

Depuis le 1<sup>er</sup> Janvier 2009, le système de prévision « Horizons » est en production dans les agences de transport TFE. L'année 2009 doit permettre de communiquer autour du produit, de former les futurs utilisateurs, de capitaliser l'expérience et d'adapter le produit suivant les usages qui en sont faits. La communication a pour but de convaincre les responsables d'agences de l'intérêt qu'il y a à utiliser l'application. Même si ce n'est pas la seule raison d'utiliser les prévisions, la communication insiste sur l'anticipation de l'activité pour améliorer la planification des ressources humaines sur le quai. Cette perspective, nous amène à une possible première évaluation du gain apporté par le système de prévision. Sans compter la convivialité de l'interface, l'accès rapide à une information juste, et mise à jour quotidienne, nous pouvons nous attendre raisonnablement à ce que la prévision d'activité face gagner 24 heures de production par mois. Cela équivaut en moyenne à moins d'une heure par jour (24 heures / 26 jours ouvrés par mois). Au niveau groupe, ce bénéfice monte à 1.680 heures mensuelles (70 agences × 24 heures). Sur la base d'un smic horaire, le gain atteint 14.817€ par mois (8, 82€ × 1680 heures) ou 177.811€ par an, sans compter les charges patronales.

Avant la mise en place du système de prévision « Horizons », les différentes régions du groupe (Bretagne Pays de la Loire, Sud-ouest, Nord-Ouest, Rhône-Alpes, Sud-est, Est) possédaient chacune leur propre système de prévision. Pour être maintenu à jour, les systèmes développés sur tableur avaient besoin au minimum d'une demi-journée par semaine du temps de travail d'un contrôleur de gestion. Sur une base annuelle de 28 k€, le gain pour le groupe monte à 16.800€ par an (25k€ / 10 temps × 6 régions), sans compter les charges.

Le gain se retrouve également dans le confort d'utilisation. La partie graphique de l'application permet, rapidement, de constater des anomalies dans les indicateurs affichés. Cette facilité d'usage autorise d'être très réactif et de corriger rapidement la source de l'erreur constatée visuellement.

L'affichage de l'activité passée et prévue (voir section 4.7.3), doit motiver les équipes. Le but recherché est un gain de productivité. Il semble néanmoins que l'affichage aux yeux de tous soit à utiliser avec modération. Se rendant compte de l'activité produite, l'employé aurait tendance à réclamer une hausse de salaire, nous a révélé un responsable d'équipe.

Un moyen d'estimer le retour sur investissement est de laisser la parole aux utilisateurs. Lydie Hoarau, responsable des ressources humaines chez TFE Bourge, utilise régulièrement l'application. Elle avance que les prévisions des indicateurs poids et nombre de positions passées à **quai**, sont pertinentes. Les prévisions des indicateurs en expédition

sont moins justes, car il y a moins d'activité. Cela engendre un historique des données volatiles. Lydie comprend la cause de l'erreur et corrige en conséquence les prévisions en expédition. Elle ajoute utiliser les prévisions d'activité pour estimer la productivité de la semaine suivante.

Julien Vidal, chef de **quai** chez TFE Lille, utilise l'outil « Horizons » régulièrement. A partir des prévisions journalières sur un horizon de deux mois, il affecte un effectif quai journalier. L'effectif est calculé en fonction de la prévision des poids et d'une productivité avisée. Il estime que les prévisions sont justes à 95%, en moyenne, sur la semaine. Cependant, il pense que 5% d'erreur c'est encore beaucoup, au vu des tonnages hebdomadaires que cela représente. 5% d'erreur équivalent à 2 ou 3 personnes à la semaine.

Thierry Jaffredou, contrôleur de gestion chez TFE Rennes exploite, les prévisions fournies par « Horizons » pour planifier les équipes du quai. Il dit retrouver des données cohérentes qu'il est capable de croiser avec des informations de sources différentes. Comme il est préconisé par la méthodologie, Thierry Jaffredou réunit les chefs de service une fois par semaine, pour croiser les informations qualitatives qu'ils ont du terrain, dans le but d'affiner les prévisions proposées par l'outil « Horizons ». T. Jaffredou est parvenu à convaincre le chef de quai de l'agence d'utiliser l'outil de prévision. D'autre part, il travaille en étroite collaboration avec le prévisionniste pour lui suggérer des évolutions de l'outil.

Il ressort des discussions avec les utilisateurs un réel engouement pour la prévision d'activité. Ils sont persuadés que c'est un moyen d'améliorer la productivité et, plus généralement, la logistique. Ils sont les premiers à en redemander. Ils sont nombreux à penser que le métier évolue. Avant l'activité se répartissait équitablement sur une journée de travail. L'expédition se faisait pendant la journée, tandis que la distribution se faisait la nuit. Aujourd'hui la logistique se recentre sur des plages horaires plus étroites. L'activité demande plus de flexibilité. Mais pour décider des plages horaires de travail, les décideurs ont besoin de prévision d'activité par créneaux horaire. Cette demande de prévision émanant du terrain est un signe positif qui nous semble important d'insérer dans les critères de retour sur investissement. C'est le signe que les prévisions sont devenues indispensables pour le pilotage de l'activité.

Nous avons vu qu'un critère du retour sur investissement est le besoin qu'ont les agences de transport en données prévisionnelles. Un moyen de calculer ce besoin est de compter le nombre de connexions mensuelles au site « Horizons ». Aujourd'hui, il est d'environ 350 connexions mensuelles. Bien entendu, un même utilisateur n'est compté qu'une fois par jour. Une trentaine de personnes se connectent plus d'une fois par semaine. Ce qui est le signe que ces personnes suivent avec intérêt les prévisions et l'évolution de leur activité réelle.

Nous venons de montrer que le retour sur investissement ne se calculait pas uniquement en termes de gain financier. Une partie du gain est du temps gagné, que nous pour-

rions chercher à convertir en argent, une autre partie est l'accroissement de la réactivité, une autre est la prise de conscience, par les managers, de l'importance des données prévisionnelles, etc. Le groupe STEF-TFE a pris conscience des enjeux et c'est pour-quoi il met les moyens pour inciter toutes les agences de transport du groupe à utiliser à bon escient le système de prévision « Horizons ». Le soutien des instances dirigeantes a également permis de sortir une version 2.0 du produit. Elle fournit des prévisions à moyen terme, c'est-à-dire des prévisions d'activité mensuelle à horizon de 24 mois. Ces prévisions ont pour vocation d'aider à fixer les budgets.

## 4.10 Industrialisation : Prévision à moyen terme

La version 2.0 du système de prévision « Horizons » propose des prévisions d'activité à moyen terme. Elle n'est en service que depuis Octobre 2009. Les indicateurs poids des marchandises et nombre de **positions** passées à quai mensuellement sont prévus à horizon fin 2010. Début 2010, l'horizon de prévision passera à fin 2011 et ainsi de suite. Ces prévisions mensuelles serviront, entre autres, à établir les budgets. Comme pour les prévisions à court terme, des statistiques sont fournies pour permettre à l'utilisateur de se faire sa propre opinion sur la confiance à apporter aux prévisions proposées. De manière générale, tout ce qui se trouvait dans « Horizons » V1.0 pour les prévisions journalières, se retrouve dupliqué dans « Horizons » V2.0 pour les données mensuelles.

« Horizons » V2.0 est la preuve qu'il est possible, en moins de 6 mois, de rédiger un modèle mathématique de prévision, de le programmer informatiquement, d'instaurer la communication entre applications et, enfin, de proposer par une interface graphique conviviale une information passée et prévisionnelle juste et mise à jour quotidiennement. C'est le travail effectué sur « Horizons » V1.0 qui a permis d'identifier, de définir les étapes et de normaliser le processus de mise en production d'un système de prévision. La version « Horizons » V2.0 est présentée dans le rapport de stage de M. Jaimes (68). M. Jaimes reprend la méthodologie, les processus de développement, les méthodes analytiques et les recommandations décrits dans cette thèse pour développer un système de prévision améliorées, qui anticipe mensuellement l'activité des agences de transport du groupe STEF-TFE. Ainsi, cette thèse aura aussi permis de décrire, d'une manière détaillée et pratique une méthode de mise place d'un système de prévision dans son complet. La méthode a fait ses preuves dans le domaine du transport, au sein du groupe STEF-TFE. Elle attend d'être appliquée chez les clients du groupe et notamment chez les industriels de l'agroalimentaire.

# Chapitre 5

## Simulation des flux de marchandises sur un quai de messagerie

La compréhension et la modélisation d'un **quai** de messagerie sont un domaine peu exploré. L'objectif de ce chapitre est de contribuer de manière significative à son exploration. Nous avons précédemment décrit les flux physiques et informatiques occasionnés par le transport d'une marchandise. Ces informations nous permettront de prescrire les entrées et sorties du modèle de quai de messagerie

Disposer d'un tel modèle implémenté dans un code de calcul efficace permettra d'estimer le temps de traitement de la marchandise sur un quai. Cela permettra également de prévoir les ressources matérielles et humaines nécessaires pour le passage à quai des marchandises. Enfin, cela ouvrira la possibilité d'optimiser les processus industriels en jeu sur les quais.

Nous construisons un modèle flexible permettant de simuler les flux de marchandises, aussi bien sur un petit quai que sur un quai de grande taille, à partir des données qui sont à notre disposition. Dans un premier temps, nous décomposons le quai en plusieurs compartiments. L'hypothèse est de considérer que la marchandise associée à une **position** donnée passe par chacun des compartiments du quai. Le passage d'un compartiment à un autre est décrit par un flux, lui-même modélisé par une fonction de transfert. Les fonctions de transfert interviennent dans un système d'équations différentielles ordinaires décrivant l'évolution de la masse de marchandises dans chacun des compartiments. Deux modèles simulant les flux sont développés. Le premier, appelé « modèle fin », modélise les flux à l'échelle de la position. Le deuxième modèle, appelé modèle « agrégé », modélise les flux à l'échelle d'un ensemble de positions.

Ces deux modèles sont implémentés dans l'environnement R, puis testés sur des cas

très simples. Leur capacité à se comporter conformément au fonctionnement réel d'un quai est exhibée. En revanche les tests mettent en évidence que pour être utilisé de façon opérationnelle il est indispensable de les implémenter avec une méthodologie de calcul scientifique, ce qui est hors de portée de cette thèse.

Enfin, en vue de leur implémentation, un modèle permettant d'interconnecter plusieurs quais est construit. Des exemples de formulations de questions industrielles en termes d'optimisation de problèmes induisant ces modèles sont décrits de manière relativement détaillée comme perspectives.

## 5.1 Modèle « fin »

Nous modélisons l'activité du quai par un modèle à cinq compartiments et 4 fonctions de transfert.

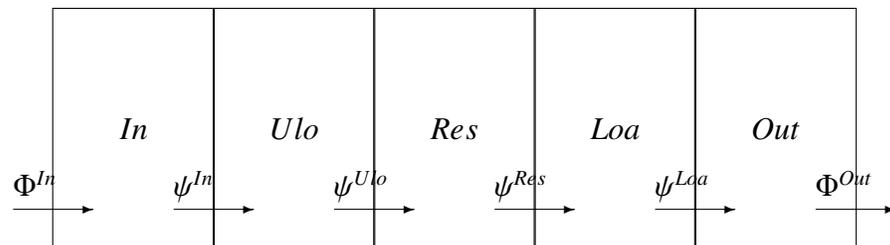


FIGURE 5.1 – Décomposition du quai en 5 compartiments

La figure 5.1 illustre la façon dont le quai est découpé en 5 aires. La première, appelée « **aire d'entrée** » (ou In), n'appartient pas physiquement au quai. Elle représente la place sur laquelle les chauffeurs garent leur camion, pour faire décharger la marchandise. Avant le déchargement, un certain nombre d'opérations sont nécessaires (bon de livraison, ...). Ensuite, les marchandises sont déchargées dans « **l'aire de déchargement** » (ou Unlo) pour être conditionnées. Si elles ne sont pas planifiées pour repartir sous les 24 heures, elles sont transférées dans une réserve ou dans une aire de stockage. Nous appelons ces deux endroits, « **aire de résidence** » (ou Res). Après leur attente dans l'aire de résidence, les marchandises passent dans « **l'aire de chargement** » (ou Loa). Cette aire permet aux manutentionnaires de bien ranger les palettes de marchandises devant les portes de sortie où sont stationnés les camions pour le chargement. La remorque du

camion dans laquelle se fait le chargement constitue l'« **aire de sortie** » (ou Out).

La vitesse de transfert de la marchandise d'une aire à une autre dépend de l'emballage de la marchandise et du nombre de personnes affectées à cette tâche. De même, d'autres aspects importants sont à considérer. Par exemple, le schéma de décomposition du quai en compartiments ne prend pas en compte toutes les situations. Les marchandises peuvent être directement transférées de l'aire de déchargement vers l'aire de chargement sans passer par l'aire de résidence. D'autre part, les grands quais, bénéficient de personnel dédié. Il existe les réceptionnistes, les préparateurs, les chargeurs. Alors que le personnel des petits quais est souvent polyvalent.

Le modèle développé cherche à prendre en compte toutes ces particularités. Il doit être valide sur une échelle de temps de quelques minutes. En effet, les applications pour lesquelles nous projetons de l'utiliser nécessitent des modèles à temps discrets utilisant des pas de temps variant entre une valeur  $\Delta t_{\min}$ , de 15 minutes environ et une valeur  $\Delta t_{\max}$  de l'ordre de 3 heures. A partir de ces considérations, nous construisons un modèle qui utilise des équations ordinaires différentielles.

Pour  $A \in \{In, Ulo, Res, Loa, Out\}$ ,  $p$  est une position parmi un ensemble  $Pos$  et  $t \in \mathbb{R}$ , nous définissons  $S^{A,(p)}(t)$  comme la masse de marchandises rattachées à la position  $p$  et présente dans l'aire  $A$ , à la date  $t$ . Pour être plus exact, c'est la moyenne de cette masse sur un intervalle de longueur  $\Delta t_{\min}$ , centré autour de  $t$ . Nous définissons également,  $\Phi^{In,(p)}(t)$  comme le flux de masse de marchandises associées à la position  $p$  et qui entre dans l'aire d'entrée, à la date  $t$ . (Concrètement, ce flux est induit par les camions qui se garent pour déchargement.) Sa définition précise consiste à dire que la masse de marchandises associées à la position  $p$  qui entre dans l'aire d'entrée entre la date  $t$  et  $t + \Delta t$  avec  $\Delta t > \Delta t_{\min}$  est

$$\int_t^{t+\Delta t} \Phi^{In,(p)}(s) ds. \quad (5.1)$$

De la même façon, nous définissons  $\Phi^{Out,(p)}(t)$  comme le flux de la masse des marchandises attachées à la position  $p$  qui quittent l'aire de sortie. Posons  $S^A(t) = \sum_{p \in Pos} S^{A,(p)}(t)$ ,

la masse totale de marchandises présentes sur l'aire  $A$  à la date  $t$ . Ainsi, par le biais de fonctions de transfert, il est possible de définir les flux de masse de marchandises circulant entre les aires. Définissons la fonction de transfert  $\psi^{In,(p)}(t)$  comme le flux de la masse de marchandises attachées à la position  $p$  et qui circule entre l'aire d'entrée et l'aire de déchargement, à la date  $t$ . De la même manière  $\psi^{Ulo,(p)}(t)$  est la fonction de transfert entre l'aire de résidence et l'aire de chargement, de même que  $\psi^{Loa,(p)}(t)$  est la fonction de transfert entre l'aire de chargement et l'aire de sortie. Il est raisonnable de poser :

$$\begin{cases} \psi^{In,(p)}(t) = \tilde{\psi}^{In}(\Pi(p, t), S^{Ulo}(t), n^{In,(p)}(t)) \chi(S^{In,(p)}(t)), \\ \text{où } \chi(S) = 1 \text{ si } S > 0 \text{ et } 0 \text{ si } S \leq 0. \end{cases} \quad (5.2)$$

L'équation 5.2 met en avant que  $\psi^{In,(p)}(t)$  est une fonction de la position. La façon dont  $\tilde{\psi}^{In}$  dépend de  $p$  doit prendre en compte l'emballage des marchandises de la position  $p$  et, certainement, d'autres critères relatifs à la position  $p$ . Certains emballages sont, tout simplement, plus rapide à déplacer que d'autres.  $\Pi(p, t)$  représente les informations liées à la position  $p$  à la date  $t$ .  $\tilde{\psi}^{In}$  est également une fonction de  $S^{Ulo}$  ce qui permet de prendre en compte que  $S^{Ulo}$  peut être surchargé.  $\tilde{\psi}^{In}$  dépend aussi du nombre d'employés (ou heures productives)  $n^{In,(p)}(t)$  dédié au transfert. La fonction  $\chi(S^{In,(p)}(t))$  interrompt le transfert quand il ne reste plus de marchandises liées à la position  $p$  dans l'aire d'entrée. De la même façon, nous posons :

$$\psi^{Ulo,(p)}(t) = \tilde{\psi}^{Ulo}(\Pi(p, t), S^{Res}(t), n^{Ulo}(t))\chi(S^{Ulo,(p)}(t)), \quad (5.3)$$

$$\psi^{Res,(p)}(t) = \tilde{\psi}^{Res}(\Pi(p, t), v, S^{Loa}(t), n^{Res}(t))\chi(S^{Res,(p)}(t)), \quad (5.4)$$

$$\psi^{Loa,(p)}(t) = \tilde{\psi}^{Loa}(\Pi(p, t), S^{Out}(t), n^{Loa}(t))\chi(S^{Loa,(p)}(t)). \quad (5.5)$$

Ayant défini ces quantités et fonctions de transfert, nous pouvons écrire les équations différentielles ordinaires modélisant le transfert de la masse de marchandises d'une position  $p$  à travers les compartiments du quai :

$$\frac{dS^{In,(p)}}{dt} = \Phi^{In,(p)} - \psi^{In,(p)}, \quad (5.6)$$

$$\frac{dS^{Ulo,(p)}}{dt} = \psi^{In,(p)} - \psi^{Ulo,(p)}, \quad (5.7)$$

$$\frac{dS^{Res,(p)}}{dt} = \psi^{Ulo,(p)} - \psi^{Res,(p)}, \quad (5.8)$$

$$\frac{dS^{Loa,(p)}}{dt} = \psi^{Res,(p)} - \psi^{Loa,(p)}, \quad (5.9)$$

$$\frac{dS^{Out,(p)}}{dt} = \psi^{Loa,(p)} - \Phi^{Out,(p)}. \quad (5.10)$$

La variation de la masse de marchandises présentes dans chacune des aires du quai est le résultat de la différence entre les flux d'entrées et de sorties de ces mêmes aires.

Dès que le modèle est formulé, il faut montrer qu'il est bien adapté pour des quais de grande taille. Pour ce faire, nous construisons des fonctions  $\tilde{\psi}^A$  et  $\Phi^A$  adaptées. Nous devons également démontrer que le modèle est assez flexible pour fonctionner avec des quais de taille réduite. C'est pourquoi nous présenterons les notions de chevauchement d'aires et d'heures productives adaptés  $n^A(t)$ . L'information liée à la position s'écrit :

$$\Pi(p, t) = (p, m(p), N_p(p, t), T_{In}(p), T_{Out}(p), Pub(p)). \quad (5.11)$$

$m(p)$  est la masse de marchandises liées à la position  $p$ .  $N_p(p, t)$  est le nombre d'emballages (palettes, rolls, bacs) liés à la position  $p$ . Le nombre d'emballages dépend de la date  $t$ . En effet les marchandises d'une position peuvent représenter par exemple 10

palettes lors de l'enlèvement et seulement 9 lors de la **livraison**. Entre deux, elles auront été gerbées, ou les colis auront été reconditionnés pour gagner de la place.  $T_{In}(p)$  est la date à laquelle le camion contenant la position  $p$  est prêt à être déchargé.  $T_{Out}(p)$  est la date à laquelle le camion contenant la position  $p$  est prêts à quitter le quai.  $Pub(p)$  est la liste des positions à décharger du camion avant d'accéder à la position  $p$ . Pour les grands quais choisir la fonction de transfert  $\psi^{In}$  suivante :

$$\tilde{\psi}^{In}(\Pi(p, t), S^{Ulo}(t), n^{In,(p)}(t)) = \frac{1}{\tau^{In}} \frac{m(p)}{N_p(p, t)} n^{In,(p)}(t) \chi(S_{\max}^{Ulo} - S^{Ulo}(t)), \quad (5.12)$$

semble pertinent.

Dans 5.12,  $\tau^{In}$  est le temps qu'il faut pour transférer un emballage de l'aire d'entrée vers l'aire de déchargement.  $S_{\max}^{Ulo}$  est la capacité maximale de l'aire de déchargement.  $\frac{m(p)}{N_p(p, t)}$  est la masse moyenne par emballage de la position  $p$ . L'équation 5.12 traduit le fait que la masse par emballage qui transfère de l'aire d'entrée vers l'aire de déchargement, est proportionnelle à la masse présente sur chaque emballage et à la force de travail affectée à la tâche de déchargement de la position  $p$ . La masse par emballage, qui transfère de l'aire d'entrée vers l'aire de déchargement, est aussi inversement proportionnelle au temps qu'il faut pour décharger un emballage. Le facteur  $\chi(S_{\max}^{Ulo} - S^{Ulo})$  permet de stopper le transfert quand  $S_{\max}^{Ulo}$  est atteint. Notez qu'il est possible d'affecter  $+\infty$  à  $S_{\max}^{Ulo}$  si l'aire de déchargement n'est jamais saturée. La force de travail,  $n^{In,(p)}(t)$ , allouée au déchargement de la position  $p$  peut être définie suivant un indice de priorité que nous souhaitons accorder à la position  $p$ , pondéré par ce que représente la masse de marchandises de la position  $p$  parmi l'ensemble de la masse à décharger, et en fonction aussi des indices de priorité, accordés aux autres positions. Ce qui peut être traduit par la formule suivante :

$$n^{In,(p)}(t) = \frac{\int_{t-\Delta t_{\min}}^t \gamma(p, s) S^{In,(p)}(s) ds}{\int_{t-\Delta t_{\min}}^t \sum_{p' \in Pos} \gamma(p', s) S^{In,(p')}(s) ds} n^{In}(t), \quad (5.13)$$

$\gamma(p, t) \in [0, +\infty]$  est l'indice de priorité. Un indice élevé permet d'augmenter la force de travail allouée au déchargement de la position  $p$ .  $\gamma(p, t)$  peut être défini de différentes manières. Une façon simple consiste à poser :

$$\gamma(p, t) = \gamma(p) = \frac{\tilde{\gamma}}{T_{Out}(p) - T_{In}(p)}, \quad (5.14)$$

pour tout  $t \in [T_{In}(p), T_{Out}(p)]$  et pour une constante donnée  $\tilde{\gamma}$ ,

Une deuxième proposition est de lier  $\gamma(p, t)$  à la longueur d'un intervalle de temps :

$$\gamma(p, t) = \min(\tilde{\gamma}_1 + \frac{\tilde{\gamma}_2}{T_{Out}(p) - t}; \tilde{\gamma}_3), \quad (5.15)$$

pour tout  $t \in [T_{In}(p), T_{Out}(p)]$  et pour des constantes  $\tilde{\gamma}_1, \tilde{\gamma}_2, \tilde{\gamma}_3$  fixées.

Cette procédure permet d'augmenter l'indice de priorité au fur et à mesure que nous nous approchons de la date prévue de départ du quai. Enfin,  $\gamma(p, t)$  peut être considéré comme une fonction à optimiser.

Dans tous les cas de figures l'expression 5.13 nous semble préférable à la suivante :

$$n^{In,(p)}(t) = \frac{\gamma(p, t)S^{In,(p)}(t)}{\sum_{p' \in Pos} \gamma(p', t)S^{In,(p')}(t)} n^{In}(t). \quad (5.16)$$

Cette dernière ne prend en effet pas en compte le temps nécessaire de la force de travail pour changer de poste ou simplement pour passer d'une aire à une autre. De plus, utiliser la fonction 5.16 accélère le début de chargement et ralentit sa fin. La fonction accorde sans discernement une forte force de travail quand la masse de marchandises est importante et une faible force de travail quand la masse de marchandises est faible. La fonction 5.15 permet d'attribuer une force de travail selon une masse moyenne, sur un intervalle de temps de longueur  $\Delta t_{\min}$ . Ceci a l'avantage de prendre en compte les pauses prises par les employés ou les changements de poste.

Avant de définir la fonction de transfert  $\tilde{\psi}^A$  pour  $A \in \{Ulo, Res, Loa\}$ , généralisons la notation.  $\tau^{Ulo}$  est pour le temps passé à transférer un emballage de l'aire de déchargement vers l'aire de résidence.  $\tau^{Res}$  est pour le temps passé à transférer un emballage de l'aire de résidence vers l'aire de chargement.  $\tau^{Loa}$  est pour le temps passé à transférer un emballage de l'aire de chargement vers l'aire de sortie. En désignant par  $S_{\max}^A$  la capacité maximale de l'aire  $A$ ,  $n^A(t)$  et  $n^{A,(p)}(t)$  sont liés par :

$$n^{A,(p)}(t) = \frac{\int_{t-\Delta t_{\min}}^t \gamma(p, s)S^{A,(p)}(s)ds}{\int_{t-\Delta t_{\min}}^t \sum_{p' \in Pos} \gamma(p', s)S^{A,(p')}(s)ds} n^A(t), \quad (5.17)$$

où  $n^A(t)$  est la force de travail allouée à l'aire  $A$  et  $n^{A,(p)}(t)$  la force de travail allouée à l'aire  $A$  pour exécuter une tâche sur la position  $p$ . Ainsi, nous pouvons écrire les égalités suivantes pour définir  $\tilde{\psi}^{Ulo}$  et  $\tilde{\psi}^{Loa}$

$$\tilde{\psi}^{Ulo}(\Pi(p, t), S^{Res}(t), n^{Ulo,(p)}(t)) = \frac{1}{\tau^{Ulo}} \frac{m(p)}{N_p(p, t)} n^{Ulo,(p)}(t) \chi(S_{\max}^{Res} - S^{Res}(t)), \quad (5.18)$$

$$\tilde{\psi}^{Loa}(\Pi(p, t), S^{Out}(t), n^{Loa,(p)}(t)) = \frac{1}{\tau^{Loa}} \frac{m(p)}{N_p(p, t)} n^{Loa,(p)}(t) \chi(S_{\max}^{Out} - S^{Out}(t)). \quad (5.19)$$

La fonction  $\tilde{\psi}^{Res,(p)}$  dépend d'une variable  $\nu$  qui représente la date à laquelle le transfert des marchandises de la position  $p$ , entre l'aire de résidence et l'aire de chargement commence. Cette variable permet ainsi de prendre en compte le fait que la marchandise peut rester un certain temps sur l'aire de résidence. La définition de  $\tilde{\psi}^{Res,(p)}$  est alors :

$$\left\{ \begin{array}{l} \tilde{\psi}^{Res,(p)}(\Pi(p, t), \nu(p), S^{Loa}(t), n^{Res,(p)}(t)) = \frac{1}{\tau^{Res}} \frac{m(p)}{N_p(p, t)} n^{Res,(p)}(t) \chi(S_{\max}^{Loa} - S^{Loa}(t)) \\ \quad \text{si } t \geq \nu(p) \\ = 0 \text{ sinon.} \end{array} \right. \quad (5.20)$$

Le calcul de  $n^{Res,(p)}$  nécessite une définition de  $\gamma(p, t)$ . Si avant d'arriver dans l'aire de résidence,  $\gamma(p, t)$  est défini par la fonction 5.14 il est nécessaire de la définir de la manière suivante :

$$\gamma(p, t) = \frac{\tilde{\gamma}}{T_{Out}(p) - \nu(p)} \text{ pour tout } t \in [\nu(p), T_{Out}(p)]. \quad (5.21)$$

D'autre part, si la définition 5.15 est choisie, la fonction précédente reste valide. Pour définir  $\nu(p)$  plusieurs choix peuvent être faits. Prenons le plus simple :

$$\nu(p) = T_{Out}(p) - \tilde{\nu}, \quad (5.22)$$

où la constante  $\tilde{\nu}$  est un temps estimé pour qu'une position  $p$  effectue le trajet entre l'aire de résidence et l'aire de chargement. Bien entendu, la constante  $\tilde{\nu}(\Pi(p, t))$  dépend de  $\Pi(p, t)$ . L'idée qui se cache sous cette dépendance est que plus il y a d'emballages associés à la position  $p$ , plus le temps  $\tilde{\nu}(\Pi(p, t))$  est long.  $\tilde{\nu}(\Pi(p, t))$  peut également être inconnu et, dans ce cas, être déterminé par une procédure d'optimisation.

Considérons  $\Phi^{In,(p)}$  et  $\Phi^{Out,(p)}$ . Désignons  $\delta_{[t=T]}$  la masse de Dirac à la date  $T$  et  $\tilde{\delta}^{\Delta t, T}(t)$  la régularisation suivante :

$$\left\{ \begin{array}{ll} \tilde{\delta}^{\Delta t, T}(t) = 0 & \text{si } t < T - \Delta t, \\ = \frac{t - T + \Delta t}{\Delta t^2} & \text{si } T - \Delta t \leq t < T, \\ = \frac{-t + T + \Delta t}{\Delta t^2} & \text{si } T \leq t < T + \Delta t, \\ = 0 & \text{si } t \geq T + \Delta t. \end{array} \right. \quad (5.23)$$

Nous posons

$$\Phi^{Out,(p)} = m(p)\delta_{[t=T_{Out}(p)]}. \quad (5.24)$$

Ce qui signifie que la masse totale associée à la position  $p$  est transférée en dehors de l'aire de sortie à la date  $T_{Out}(p)$  où

$$\Phi^{Out,(p)} = m(p)\delta^{\frac{\Delta t_{\min}}{4}, T_{Out}(p)}(t), \quad (5.25)$$

est la régularisation de 5.24.

Pour  $\Phi^{In,(p)}$  nous introduisons un temps  $\tilde{T}(p)$ . Il représente le temps passé à décharger les positions  $Pub(p)$  avant d'accéder à la position  $p$ , dans le camion. Nous définissons alors :

$$\Phi^{In,(p)} = m(p)\delta_{[t=T_{In}(p)+\tilde{T}(p)]}, \quad (5.26)$$

ou

$$\Phi^{In,(p)} = m(p)\delta^{\frac{\Delta t_{\min}}{4}, T_{In}(p)+\tilde{T}(p)}(t). \quad (5.27)$$

Dans ces définitions,  $\tilde{T}(p)$  peut être déclaré comme :

$$\tilde{T}(p) = \tau^{In} \sum_{p' \in Pub(p)} N_p(p', T_{In}(p)), \quad (5.28)$$

mais peut également faire l'objet d'un processus d'optimisation, que nous soulignerons plus tard.

Le modèle qui vient d'être présenté est adapté à des quais de grande taille. Pour des quais de taille plus modeste ou des quais qui ne fonctionnent pas exactement comme nous l'avons décrit ci-dessus, le modèle peut être adapté. Nous présentons maintenant des exemples de ces adaptations.

Premièrement, la force de travail peut ne pas être allouée à une aire mais à un ensemble d'aires. Dans ce contexte,  $\mathcal{A}$  représente un ensemble d'aires. Par exemple,  $\mathcal{A}$  peut représenter  $\{In, Ulo\}$  ou  $\{Res, Loa\}$  ou  $\{In, Ulo, Res, Loa\}$ . Alors pour un  $A$  donné dans  $\mathcal{A}$  nous définissons l'heure productive adaptée avec la formule suivante :

$$n^{A,(p)}(t) = \frac{\int_{t-\Delta t_{\min}}^t \gamma(p, s) S^{A,(p)}(s) ds}{\int_{t-\Delta t_{\min}}^t \sum_{p' \in Pos} \sum_{A' \in \mathcal{A}} \gamma(p', s) S^{A',(p')}(s) ds} n^{\mathcal{A}}(t), \quad (5.29)$$

où  $n^{\mathcal{A}}(t)$  est la force de travail allouée à l'ensemble  $\mathcal{A}$ , à la date  $t$ . La formule 5.29 remplace la formule 5.17. Il est possible d'utiliser la formule 5.17 pour un intervalle de temps donné et la formule 5.29 pour un autre intervalle de temps défini.

Ensuite un **quai** peut ne pas comporter d'aire de résidence. Alors nous utilisons ce que nous avons précédemment appelé le chevauchement d'aires. Le chevauchement d'aires consiste à remplacer l'équation 5.6 par :

$$S^{Res,(p)}(t) = S^{Ulo,(p)}(t) \text{ pour tout temps } t. \quad (5.30)$$

Il est possible de chevaucher plus de deux aires et que le chevauchement est compatible avec la distribution de la force de travail définie en 5.29 et 5.17

## 5.2 Modèle « agrégé »

Nous présentons dans cette section un modèle pour les flux de marchandises passant à quai à un niveau plus macroscopique. Typiquement le pas de temps minimum d'un modèle à temps discret déduit de ce modèle « agrégé » sera de l'ordre de  $4\Delta t_{\min} \simeq 1h$ . En conséquence, ce modèle nécessite moins de variables, moins de détails dans les données et s'avère plus facile à utiliser. Son principal défaut est que ce modèle n'est pas

aussi précis que celui développé entre les équations 5.1 et 5.28, en particulier il regroupe les positions pour en perdre le détail.

Commençons par définir la fonction  $\Phi^{In}(t)$  qui donne pour une date  $t$ , le flux de la masse de marchandises de l'aire d'entrée.  $\Phi^{Out}(t)$  est le flux de la masse de marchandises qui sort de l'aire de sorties à la date  $t$ . La masse de marchandises présentes sur une aire  $A$  est :

$$S^A(t) = \sum_{p \in Pos} S^{A,(p)}(t) \text{ pour } A\{In, Ulo, Res, Loa, Out\}. \quad (5.31)$$

La somme prend en compte les positions  $p \in Pos$  dont la date de passage à quai est comprise entre  $T_{In}(p)$  et  $T_{Out}(p)$  :  $T_{In}(p) \leq t \leq T_{Out}(p)$ .

Nous déduisons des équations 5.1 à 5.28 le modèle « agrégé ». Les équations 5.12, 5.18, 5.19 et 5.20 utilisent la quantité :

$$\frac{1}{\tau^A} \frac{m(p)}{N_p(p, t)} \text{ pour } A\{In, Ulo, Res, Loa, Out\}. \quad (5.32)$$

Celle-ci peut être moyennée à partir d'un historique d'observations afin d'obtenir les coefficients moyens suivants :

$$\omega^A \text{ pour } A\{In, Ulo, Res, Loa, Out\}. \quad (5.33)$$

Nous définissons ensuite la fonction de transfert  $\psi^{In}(t)$ . Elle représente le flux de masse de marchandises qui passent de l'aire d'entrée vers l'aire de déchargement à la date  $t$ .

$$\psi^{In}(t) = \bar{\psi}^{In}(S^{Ulo}(t), n^{In}(t))\chi(S^{In}(t)), \quad (5.34)$$

avec  $\bar{\psi}^{In}$  défini par

$$\bar{\psi}^{In}(S^{Ulo}, n^{In}) = \omega^{In} n^{In} \chi(S_{\max}^{Ulo} - S^{Ulo}). \quad (5.35)$$

De la même façon nous définissons

$$\psi^{Ulo}(t) = \bar{\psi}^{Ulo}(S^{Res}(t), n^{Ulo}(t))\chi(S^{Ulo}(t)), \quad (5.36)$$

$$\psi^{Loa}(t) = \bar{\psi}^{Loa}(S^{Out}(t), n^{Loa}(t))\chi(S^{Loa}(t)), \quad (5.37)$$

avec

$$\bar{\psi}^{Ulo}(S^{Res}, n^{Ulo}) = \omega^{Ulo} n^{Ulo} \chi(S_{\max}^{Res} - S^{Res}), \quad (5.38)$$

$$\bar{\psi}^{Loa}(S^{Out}, n^{Loa}) = \omega^{Loa} n^{Loa} \chi(S_{\max}^{Out} - S^{Out}). \quad (5.39)$$

Nous avons vu que la marchandise pouvait être stockée un certain temps (plus de 24h) sur l'aire de résidence. C'est pourquoi la fonction  $S_{Rem}^{Res}$  donne, à tout moment, la masse de marchandises qui reste sur l'aire de résidence. Ainsi

$$\bar{\psi}^{Res}(S^{Loa}, n^{Res}) = \omega^{Res} n^{Res} \chi(S_{\max}^{Loa} - S^{Loa}), \quad (5.40)$$

et  $\psi^{Res}(t)$  est définie par

$$\psi^{Res}(t) = \bar{\psi}^{Res}(S^{Loa}(t), n^{Res}(t))\chi(S^{Res}(t) - S_{Rem}^{Res}(t)). \quad (5.41)$$

Le système d'équations différentielles constituant le modèle « agrégé » s'écrit :

$$\frac{dS^{In}}{dt} = \Phi^{In} - \psi^{In}, \quad (5.42)$$

$$\frac{dS^{Ulo}}{dt} = \psi^{In} - \psi^{Ulo}, \quad (5.43)$$

$$\frac{dS^{Res}}{dt} = \psi^{Ulo} - \psi^{Res}, \quad (5.44)$$

$$\frac{dS^{Loa}}{dt} = \psi^{Res} - \psi^{Loa}, \quad (5.45)$$

$$\frac{dS^{Out}}{dt} = \psi^{Loa} - \Phi^{Out}. \quad (5.46)$$

Le modèle 5.31-5.42 est construit pour fonctionner sur des quais de grande taille. Cependant pour des quais plus petits ou dans lesquels le personnel n'est pas affecté à une aire donnée, il est possible de développer une procédure de recouvrement d'aires ou d'heures productives adaptées.

### 5.3 Modèle de l'organisation des quais en réseau

Dans l'intention de construire un modèle du réseau de transport, nous avons monté des connexions entre le modèle décrit par les formules 5.1-5.28 et le modèle développé par les formules 5.31-5.42. Suite à cela, considérons un réseau formé de  $n$  quais,  $\{h_1, \dots, h_n\} = \mathcal{H}$ . Parmi ces quais, certains appartiennent au transporteur TFE, ils forment le sous ensemble  $\mathcal{H}_{mod}$ , et d'autres sont la propriété des confères, ils forment le sous ensemble  $\mathcal{H}_{umm}$ . Ainsi  $\mathcal{H}_{mod} \cap \mathcal{H}_{umm} = \emptyset$ ,  $\mathcal{H}_{mod} \cup \mathcal{H}_{umm} = \mathcal{H}$ . Nous connaissons le fonctionnement des quais appartenant à  $\mathcal{H}_{mod}$  et nous l'avons modélisé, par contre nous ne savons rien des autres quais. Nous posons aussi  $K_{i,j}$  comme l'ensemble de voyages possibles entre un quai n°  $i$  et un quai n°  $j$ . Notons que s'il n'existe pas de voyage entre n°  $i$  et n°  $j$  alors  $K_{i,j} = \emptyset$ . Dans le cas contraire, il semble raisonnable d'écrire

$$K_{i,j} = \{k_{i,j}^1, k_{i,j}^2, \dots, k_{i,j}^l, \dots, k_{i,j}^{L_{i,j}}\}, \quad (5.47)$$

$k_{i,j}^l$  est le voyage nommé  $l$  entre le quai n°  $i$  et le quai n°  $j$ . Les voyages  $k_{i,j}^l$  durent un temps  $T_{i,j}^l$ . C'est le temps nécessaire à effectuer le voyage.  $L_{i,j}$  est le nombre de voyages possibles entre le quai n°  $i$  et le quai n°  $j$ .

La masse de marchandises rattachées à la position  $p$ , et en transfert dans l'aire d'entrée du quai  $i$ , est décrite par la fonction  $\Phi_i^{In,(p)}$  (voir formule 5.1). Dans l'optique de décrire la provenance de cette marchandise, nous introduisons  $R_i^{\lambda,(p)}(t)$ .  $R_i^{\lambda,(p)}(t)$  est la masse de marchandises rattachées à la position  $p$ , qui entre dans l'aire d'entrée du quai n°  $i$  suite à une ramasse chez l'expéditeur  $\lambda$ . De la même façon,  $B_i^{\lambda,(p)}(t)$  est la masse de marchandises rattachées à la position  $p$  directement amenée par l'expéditeur  $\lambda$  dans l'aire d'entrée du quai n°  $i$ . Enfin  $F_i^{j,(p)}(t)$  est la masse de marchandises rattachées à la position  $p$ , en provenance du quai n°  $j$  et qui entre dans l'aire d'entrée du quai n°  $i$ . Ainsi,

$$\Phi_i^{In,(p)}(t) = \sum_{\lambda \in \Lambda_i} R_i^{\lambda,(p)}(t) + \sum_{\lambda \in \Lambda_i} B_i^{\lambda,(p)}(t) + \sum_{j=1}^n F_i^{j,(p)}(t), \quad (5.48)$$

$\Lambda_i$  est l'ensemble des expéditeurs clients de l'agence de transport n°  $i$ . En sortie de quai, nous introduisons les quantités suivantes :  $D_i^{\alpha,(p)}(t)$  est la masse de marchandises attachées à la position  $p$ , qui quitte le quai n°  $i$  pour être livrée chez le destinataire  $\alpha$ .  $P_i^{\alpha,(p)}(t)$  est la masse de marchandises attachées à la position  $p$ , qui est prise par le destinataire  $\alpha$ , au quai n°  $i$ .  $E_i^{j,(p)}(t)$  est la masse de marchandises attachées à la position  $p$ , qui quitte le quai n°  $i$ , pour être déchargée au quai n°  $j$ . Ainsi,

$$\Phi_i^{Out,(p)}(t) = \sum_{\alpha \in \mathcal{A}_i} D_i^{\alpha,(p)}(t) + \sum_{\alpha \in \mathcal{A}_i} P_i^{\alpha,(p)}(t) + \sum_{j=1}^n E_i^{j,(p)}(t). \quad (5.49)$$

$\mathcal{A}_i$  est l'ensemble des confrères travaillant avec l'agence du quai n°  $i$ .

Remarquons que les marchandises qui quittent le quai n°  $i$  pour le quai n°  $j$ , à une date  $t$ , sont les mêmes marchandises qui arrivent au quai n°  $j$ , en provenance du quai n°  $i$ , à une date  $t + T_{i,j}^l$ . Ainsi nous connectons le modèle 5.1-5.28 du quai n°  $i$  avec celui du quai n°  $j$ .

$$F_i^{j,(p)}(t) = \sum_{l=1}^{L_{i,j}} a(k_{i,j,p}^l) E_j^{i,(p)}(t - T_{i,j}^l) \quad (5.50)$$

$a(k_{i,j,p}^l)$  est la proportion de la masse de marchandises attachées à la position  $p$ , transportée dans le voyage  $k_{i,j}^l$ , allant du quai n°  $i$  vers le quai n°  $j$ . En particulier,

$$\sum_{l=1}^{L_{i,j}} a(k_{i,j,p}^l) = 1 \text{ pour tout } p \in Pos. \quad (5.51)$$

Pour connecter les différentes instances des modèles 5.31-5.42 indexées par  $i = 1, \dots, n$  nous procédons de la manière suivante :

$$\begin{aligned} R_i^\lambda(t) &= \sum_{p \in Pos} R_i^{\lambda,(p)}(t) \quad , \quad D_i^\alpha(t) = \sum_{p \in Pos} D_i^{\alpha,(p)}(t), \\ B_i^\lambda(t) &= \sum_{p \in Pos} B_i^{\lambda,(p)}(t) \quad , \quad P_i^\alpha(t) = \sum_{p \in Pos} P_i^{\alpha,(p)}(t), \\ F_i^j(t) &= \sum_{p \in Pos} F_i^{j,(p)}(t) \quad , \quad E_i^j(t) = \sum_{p \in Pos} E_i^{j,(p)}(t), \end{aligned} \quad (5.52)$$

et

$$\begin{aligned}\Phi_i^{In}(t) &= \sum_{\lambda \in \Lambda_i} R_i^\lambda(t) + \sum_{\lambda \in \Lambda_i} B_i^\lambda(t) + \sum_{j=1}^n F_i^j(t), \\ \Phi_i^{Out}(t) &= \sum_{\alpha \in \mathcal{A}_i} D_i^\alpha(t) + \sum_{\alpha \in \mathcal{A}_i} P_i^\alpha(t) + \sum_{j=1}^n E_i^j(t),\end{aligned}\tag{5.53}$$

et

$$F_i^j(t) = \sum_{l=1}^{L_{i,j}} \bar{a}(k_{i,j}^l) E_j^i(t - T_{i,j}^l),\tag{5.54}$$

avec  $\bar{a}(k_{i,j}^l)$ , qui satisfait

$$\sum_{l=1}^{L_{i,j}} \bar{a}(k_{i,j}^l) = 1,\tag{5.55}$$

est la proportion de masse de marchandises entre le quai n°  $i$  et n°  $j$ , qui prend le voyage  $k_{i,j}^l$ .  $\bar{a}$  peut-être déduit après observation du terrain ou d'un ensemble d'enregistrements dans la base de données.

Il est également possible de moyenner les formules 5.50 et 5.54

$$\begin{aligned}\int_t^{\Delta t} F_i^{j,(p)}(s) ds &= \sum_{l=1}^{L_{i,j}} \bar{a}(k_{i,j}^l) \int_{t-T_{i,j}^l}^{t-T_{i,j}^l+\Delta t} E_j^{i,(p)}(s) ds, \\ \int_t^{\Delta t} F_i^j(s) ds &= \sum_{l=1}^{L_{i,j}} \bar{a}(k_{i,j}^l) \int_{t-T_{i,j}^l}^{t-T_{i,j}^l+\Delta t} E_j^i(s) ds.\end{aligned}\tag{5.56}$$

## 5.4 Tests de validation

Cette section tient à montrer la faisabilité de l'utilisation des modèles « fin » et « agrégé » dans le but d'optimiser des aspects de fonctionnement d'un quai de messagerie. Le premier aspect est la détermination de la quantité de main d'œuvre nécessaire pour faire face à l'activité tout en la répartissant subtilement dans les différentes aires du quai définies dans le modèle « agrégé ». Une seconde utilisation est faite pour déterminer la quantité de main d'œuvre par **position**, et ceci, en optimisant l'indice de priorité  $\gamma(p, t)$  de l'équation 5.14. Le troisième exemple d'utilisation amène une solution pour fixer  $\tilde{v}(p)$ , la date à laquelle la position  $p$  sort de l'aire de résidence pour l'aire de sortie.

### 5.4.1 Résultats obtenus par le modèle « fin »

Le modèle décrit dans la section 5.1 a été programmé dans l'environnement R. Afin de démontrer les performances du modèle, nous simulons l'activité d'un quai à partir d'un exemple simple. Nous faisons entrer deux positions à une heure d'intervalle. Elles restent à quai pendant dix heures. Nous voulons observer comment leur masse remplit les différentes zones du quai au cours de ces dix heures.

Pour ce faire, le modèle décrit ci-dessus, nécessite de fixer des paramètres pour décrire les positions, la main d'œuvre disponible sur le quai, et la taille du quai. Quatre mises en situation vont permettre de comprendre comment se comporte le modèle suivant les paramètres que nous faisons varier. La première mise en situation positionne les paramètres comme indiqué dans les tableaux suivants. Deux positions de 5 tonnes et 20 palettes chacune arrivent à quai, respectivement à 1h et 2h. Elles sont programmées pour un départ à 11h et 12h. Leurs paramètres pour déterminer l'indice de priorité via la formule 5.14 sont identiques. Les positions peuvent être déchargées en même temps. Les différentes aires du quai sont assez grandes pour entreposer la somme des masses des deux positions. La quantité de main d'œuvre est équitablement répartie sur chaque poste. Le temps de transfert d'un emballage (palette) d'une zone à une autre est estimé à une minute. Pour estimer le temps de transfert des positions de la zone de résidence vers la zone de sortie, nous exécutons le modèle une première fois pour visualiser le temps nécessaire au déchargement. Ce temps est alors affecté au paramètre  $\nu$ .

N° position : $p$	Masse en tonnes : $m(p)$	Nombre de palettes : $N_p$	Positions devant : $Pub(p)$	Date de déchargement : $T_{In}(p)$	Date de départ : $T_{Out}(p)$
0R4080329341	5	20	<i>Null</i>	1	10
0R4080328864	5	20	<i>Null</i>	2	11

TABLE 5.1 – Paramètres de position

La figure 5.2 montre le résultat obtenu par l'exécution du modèle avec les paramètres ci-dessus.

Le premier graphique de la figure 5.2 montre l'évolution de la somme des masses des marchandises contenues dans les deux positions, sur chacune des aires. Il montre également l'évolution de la masse totale entreposée dans le quai.

Le deuxième et troisième graphique de la figure 5.2, montrent l'évolution de la masse dans chacune des aires, pour chacune des positions.

La courbe « In » représente l'évolution de la masse de marchandises présentes dans

Temps de transfert de la position $p$ entre la zone de résidence et la zone de sortie : $\nu$	Paramètre 1 de l'indice de priorité : $\gamma_1$	Paramètre 2 de l'indice de priorité : $\gamma_2$	Paramètre 3 de l'indice de priorité : $\gamma_3$
8	1	1	5
8	1	1	5

TABLE 5.2 – Paramètres de position, suite

Poids max en zone d'entrée, en tonnes : $S_{\max}^{In}$	Poids max en zone de déchargement, en tonnes : $S_{\max}^{Ulo}$	Poids max en zone de résidence, en tonnes : $S_{\max}^{Res}$	Poids max en zone de chargement, en tonnes : $S_{\max}^{Loa}$	Poids max en zone de sortie, en tonnes : $S_{\max}^{Out}$
20	20	20	20	20

TABLE 5.3 – Paramètres de quai

l'aire d'entrée. La courbe « Ulo » représente l'évolution de la masse de marchandises contenues dans l'aire de déchargement. La courbe « Res » représente l'évolution de la masse de marchandises contenues dans l'aire de résidence. La courbe « Loa » représente l'évolution de la masse de marchandises contenues dans l'aire de chargement. La courbe « Out » représente l'évolution de la masse de marchandises contenues dans l'aire de sortie. Alors que les aires de déchargement, de résidence et de chargement sont des aires physiquement matérialisées sur un quai de messagerie, l'aire d'entrée, comme l'aire de sortie, représentent la somme des aires contenues dans les remorques qui se mettent à quai, pour être déchargées ou chargées. Rappelons qu'un camion contient les marchandises d'un ensemble de positions. Le camion et sa remorque se placent à quai pour être déchargés à une heure définie. Après avoir été manutentionnée et déplacée de l'aire de déchargement vers l'aire de résidence, puis de l'aire de résidence vers l'aire de chargement, la marchandise d'une même **position** est chargée dans un camion pour un départ à une heure planifiée. C'est la date de départ ou la date à laquelle la marchandise quitte l'aire de sortie.

La courbe « In » du premier graphique de la figure 5.2, montre que la date de début de déchargement de la position n° 0R4080329341 est à 1h00. Le deuxième graphique montre que son déchargement se poursuit jusqu'à, approximativement, 7h30. Ce même graphique montre également qu'il y a un ralentissement du transfert de la masse de la position n° 0R4080329341 de l'aire d'entrée vers l'aire de déchargement. C'est simplement la conséquence de l'entrée de la deuxième position sur le quai. L'évolution de la répartition de la masse sur le quai de la deuxième position (n° 0R4080328864) est vue

Temps pour transférer un emballage de la zone d'entrée vers la zone de déchargement en fractions d'heures : $\tau^{In}$	Temps pour transférer un emballage de la zone de déchargement, vers la zone de résidence en fractions d'heures : $\tau^{Ulo}$	Temps pour transférer un emballage de la zone de résidence vers la zone de chargement, en fractions d'heures : $\tau^{Res}$	Temps pour transférer un emballage de la zone de chargement vers la zone de sortie en fractions d'heures : $\tau^{Loa}$
1/60	1/60	1/60	1/60

TABLE 5.4 – Paramètres d'équipe de quai

Main d'œuvre disponible pour le déchargement : $n^{In}$	Main d'œuvre disponible pour le transfert vers la zone de résidence : $n^{Ulo}$	Main d'œuvre disponible pour le transfert vers la zone de chargement : $n^{Res}$	Main d'œuvre disponible pour le chargement : $n^{Loa}$
0, 1	0, 1	0, 1	0, 1

TABLE 5.5 – Paramètres d'équipe de quai, suite

sur le troisième graphique de la figure 5.2. La deuxième position se présente pour déchargement à 2h et quitte l'aire de sortie à 11h00.

Rappelons que la main d'œuvre disponible pour le déchargement est équivalente à la main d'œuvre disponible pour le transfert des marchandises de l'aire de déchargement vers l'aire de résidence. Ainsi, aussitôt que la marchandise est déchargée, elle est transférée en zone de résidence. C'est pourquoi, comme le montrent les trois graphiques de la figure 5.2, la zone de déchargement n'accumule pas de stock. Pendant que l'aire d'entrée se vide, les aires de résidence et de sortie se remplissent (voir courbe « Res » et « Out »). La zone de sortie se remplit à partir de 02h00, car nous avons fixé le paramètre  $\nu$  à 8 heures. La zone de chargement ne se remplit pas. La raison de cela est que les marchandises sorties de la zone de résidence sont transférées immédiatement dans la zone de sortie. La position n° 0R4080329341 part, comme convenu, à 10h00 et la position n° 0R4080328864 part à 11h00. Pendant que les marchandises passent en zone de sortie, les zones d'entrée et de résidence se vident. La courbe « Tot » permet de visualiser la masse de marchandises présentes à chaque instant sur le quai.

La deuxième mise en situation reprend les mêmes paramètres que la précédente à ceci près qu'elle réduit la main d'œuvre disponible pour le transfert des marchandises de l'aire de déchargement vers l'aire de résidence. Cette main d'œuvre passe de 0, 1 à 0, 08. La

$\Delta t_{\min}$
1/4

TABLE 5.6 – Paramètres du modèle

Figure 5.3 illustre le résultat obtenu.

Comme précédemment, la courbe « In » affiche deux pics correspondant à l'arrivée des positions dans l'aire d'entrée. La chute de la courbe entre 1h00 et 2h00 et ensuite entre 2h00 et 7h00 correspond au temps de déchargement des camions. L'équipe pour transférer la marchandise déchargée entre l'aire de déchargement et l'aire de résidence étant moins grande que l'équipe de déchargement, les marchandises s'accumulent progressivement sur l'aire de déchargement (voir courbe « Ulo »). C'est aussi pourquoi l'aire de résidence se charge moins que dans la situation précédente. Remarquons également que la courbe « Res » décline à partir de 3h00. C'est dû au fait que la marchandise transférée dans l'aire de résidence est aussitôt transférée dans l'aire de chargement pour être déplacée dans l'aire de sortie. La courbe « Out » possède une rupture de pente à 3h00. C'est l'heure à laquelle la deuxième position commence à être chargée dans les camions. La courbe « Out » atteint le palier de 10 tonnes à 9h00. C'est le signe que les marchandises des deux positions sont chargées et prêtes à quitter l'aire de sortie. A 10h00 la courbe « Out » connaît sa première chute, pour indiquer le départ de la position n° 0R4080329341. Après quoi, pendant 1 heure, la courbe stagne à 5 tonnes et chute à 11h00 pour atteindre 0 tonne. La position n° 0R4080328864 est partie.

De la même façon, lorsque nous diminuons la main d'œuvre disponible pour charger la marchandise (voir le paramètre  $n^{Loa}$ ), un stock se forme sur l'aire de chargement et ainsi la courbe « Loa » devient positive.

Une troisième mise en situation fixe les paramètres tels que nous l'avons fait pour la première simulation, considère que l'aire de résidence n'est pas assez grande pour accueillir plus d'une tonne de marchandises ( $S_{\max}^{Res} = 1$ ). Le résultat illustré par la figure 5.4, indique que cette contrainte a pour effet de stocker le surplus de marchandises dans l'aire de déchargement. La figure 5.4 montre parfaitement que l'aire de résidence (voir courbe « Res ») commence à se remplir à 1h00, avec l'arrivée de la première position. A 1h30, elle atteint sa capacité maximum fixée à une tonne. Au même moment, l'aire de déchargement commence à stocker de la marchandise. La rupture de pente de la courbe « Ulo » à 2h00 indique que la marchandise provient de la première et deuxième position. Comme attendu, l'aire de déchargement se vide avant l'aire de résidence.

La dernière mise en situation, considère que 6h30 pour décharger deux positions de 5 tonnes c'est long. 6h30 c'est le temps pendant lequel la courbe « In » reste positive, c'est-à-dire entre 1h00 et 7h30. La première idée qui vient à l'esprit pour diminuer le temps de

déchargement est d'augmenter la main d'œuvre dédiée au déchargement. Nous fixons tous les paramètres tels qu'ils étaient dans la première mise en situation et nous passons la valeur du paramètre  $n^{In}$  de 0, 1 à 0, 5.

La courbe « In » de la figure 5.5 indique que les positions ont été déchargées bien plus rapidement que dans la première mise en situation. Elles ont, toutes deux, été déchargées en moins d'une heure. Par contre, nous n'avons pas augmenté la main d'œuvre pour transférer les marchandises de l'aire de déchargement vers l'aire de résidence. C'est pourquoi nous constatons que la courbe « Ulo » prend de l'ampleur, c'est-à-dire que les marchandises résident plus longtemps sur l'aire de déchargement. Encore une fois, l'aire de déchargement se vide bien avant l'aire de résidence, ce qui indique que les marchandises sont constamment en mouvement. D'autre part, la somme des masses sur le quai est de 5 tonnes entre 1h00 et 2h00, et entre 10h00 et 11h00. Elle est de 10 tonnes entre 2h00 et 10h00. Cette courbe est un indicateur de la validité du modèle.

Ces 4 mises en situation montrent que, sur ces cas simples, le modèle se comporte conformément à un quai de messagerie. Elles montrent également que le modèle, tel qu'il a été programmé, permet de simuler beaucoup de mises en situation. Le modèle indique, par des valeurs aberrantes (négative par exemple) qu'une mise en situation souhaitée peut ne pas être réalisable. Par exemple, la figure 5.6 montre le résultat lorsque nous reprenons la première mise en situation et que nous voulons faire sortir la première position à 6h00. Le modèle nous indique, par la présence de valeurs négatives sur l'aire de sortie, que cet objectif n'est pas réalisable avec les moyens mis en œuvre. Il faudrait pour cela augmenter la capacité de main d'œuvre dédiée à cette position dans les différentes aires du quai. Nous pouvons suivre position par position, la répartition des poids sur le quai, il est alors aisé de cibler un dysfonctionnement.

Comme il vient d'être montré, le modèle programmé sur le logiciel R fonctionne. Par contre, R ne permet pas de le faire fonctionner avec un grand nombre de positions. Les exemples ci-dessus montrent la gestion de deux positions sur le quai. Pour ce faire, notre portable, équipé d'un processeur duo de 1,8 GHz et 2,5 Go de RAM a nécessité un peu moins de 5 minutes. Sachant qu'une agence de transport gère quotidiennement entre 1.000 et 2.000 positions, il faudrait à notre ordinateur, certes peu puissant, plus de 3 jours et demi pour simuler une mise en situation réelle. Nous pensons qu'une solution, pour augmenter significativement les performances d'exécution, est de programmer le modèle mathématique sur un langage compilé tel que C++. Une autre solution est d'agréger le modèle. Les résultats issus de simulations du modèle « agrégé » sont présentés dans la section suivante (section 5.4.2).

## 5.4.2 Résultats obtenus par le modèle « agrégé »

Comme le modèle « fin », le modèle « agrégé » a été développé dans l'environnement R. Le but n'est pas de le programmer pour en faire une application industrielle, mais de démontrer la crédibilité du modèle mathématique. Nous cherchons également à montrer que les temps de réponse du modèle « agrégé » sont plus rapides que ceux du modèle « fin ». Pour témoigner de la ressemblance des résultats obtenus par les deux modèles, nous exécutons le modèle « agrégé », avec des paramètres identiques à ceux des mises en situation exposées ci-dessus (section 5.4.1), et d'autres qui sont le résultat de l'agrégation des paramètres du modèle « fin ». Enfin, nous superposons les courbes obtenues par les deux modèles.

Le modèle « agrégé » nécessite moins de paramètres que le modèle « fin ». Il ne détaille pas, position par position, la répartition des marchandises sur le quai. Pour coller à la première mise en situation évoquée dans la section 5.4.1, le modèle « agrégé » nécessite les paramètres suivants.

N° position	Masse en tonnes : $m(p)$	Date de déchargement : $T_{In}(p)$	Date de départ : $T_{Out}(p)$	Temps de transfert de la position $p$ entre la zone de résidence et la zone de sortie : $\nu$
0R4080329341	5	1	10	8
0R4080328864	5	2	11	8

TABLE 5.7 – Paramètres de position

Poids max en zone d'entrée, en tonnes : $S_{\max}^{In}$	Poids max en zone de déchargement, en tonnes : $S_{\max}^{Ulo}$	Poids max en zone de résidence, en tonnes : $S_{\max}^{Res}$	Poids max en zone de chargement, en tonnes : $S_{\max}^{Loa}$	Poids max en zone de sortie, en tonnes : $S_{\max}^{Out}$
20	20	20	20	20

TABLE 5.8 – Paramètres de quai

La figure 5.7 superpose les résultats obtenus par les deux modèles étudiés. Elle permet

Coût humain horaire moyen pour transférer une unité de masse entre l'aire d'entrée et l'aire de déchargement : $\omega^{In}$	Coût humain horaire moyen pour transférer une unité de masse entre l'aire de déchargement et l'aire de résidence : $\omega^{Ulo}$	Coût humain horaire moyen pour transférer une unité de masse entre l'aire de résidence et l'aire de chargement : $\omega^{Res}$	Coût humain horaire moyen pour transférer une unité de masse entre l'aire de chargement et l'aire de sortie : $\omega^{Out}$
15	15	15	15

TABLE 5.9 – Paramètres de quai, suite

Main d'œuvre disponible pour le déchargement : $n^{In}$	Main d'œuvre disponible pour le transfert vers la zone de résidence : $n^{Ulo}$	Main d'œuvre disponible pour le transfert vers la zone de chargement : $n^{Res}$	Main d'œuvre disponible pour le chargement : $n^{Loa}$
0, 1	0, 1	0, 1	0, 1

TABLE 5.10 – Paramètres d'équipe de quai

$\Delta t_{\min}$
1/4

TABLE 5.11 – Paramètres du modèle

de constater que les deux modèles vident l'aire d'entrée à la même vitesse. Le transfert de la marchandise de l'aire de déchargement vers l'aire de résidence se déroule également de la même façon selon les deux modèles. Les deux modèles commencent le chargement à la même heure. Par contre, le modèle « fin » charge l'aire de sortie un peu plus lentement que le modèle « agrégé ». C'est pourquoi l'aire de résidence se remplit moins avec le modèle « agrégé ». Ce phénomène peut provenir de la fonction  $n^{A,(p)}(t)$  5.17 qui n'existe pas pour le modèle « agrégé ». Comme prévu, l'aire de sortie se vide à 10h00 puis à 11h00. Matériellement, c'est l'heure de départ des camions.

Sous les mêmes conditions, le temps d'exécution de la première mise en situation avec le modèle « fin » a été de 4 : 47 mn alors que l'exécution du modèle « agrégé » n'a nécessité que 1 : 37 mn, soit trois fois moins de temps.

## 5.5 Perspectives

Les tests ci-dessus démontrent la fiabilité de l'utilisation des modèles « fin » et « agrégé » pour décrire, comprendre et optimiser les processus industriels. Ils ont aussi mis en évidence les difficultés liées à cette utilisation. D'un point de vue informatique, si nous souhaitons monter en charge, c'est-à-dire, simuler l'évolution des masses de marchandises sur un quai de messagerie, avec plus de 1000 positions et plus de 700 tonnes de marchandises, il faut traduire le modèle à l'aide d'un langage de programmation compilé. D'autre part, il s'est avéré que le choix des valeurs à donner aux paramètres est délicat. Le modèle « fin » demande à ce que nous fixons 24 paramètres. Le déploiement industriel d'un modèle de ce type nécessite une étude précise pour fixer la valeur de ces paramètres pour chacun des quais du transporteur.

Parmi les perspectives il est intéressant de noter comment les questions industrielles peuvent se poser en termes d'optimisation d'une quantité liée au modèle « fin » ou « agrégé ». Les sections suivantes montrent comment utiliser les modèles « fin » et « agrégé » dans le but d'optimiser des aspects de fonctionnement d'un quai de messagerie. Le premier aspect est la détermination de la quantité de main d'œuvre nécessaire pour faire face à l'activité tout en la répartissant subtilement dans les différentes aires du quai définies dans le modèle « agrégé ». Une seconde utilisation est faite pour déterminer la quantité de main d'œuvre par position, et ceci, en optimisant l'indice de priorité  $\gamma(p, t)$  de l'équation 5.14. Le troisième exemple d'utilisation amène une solution pour fixer le temps qu'une position  $p$  met pour passer de l'aire de résidence à l'aire de sortie. Avec l'aide d'une expertise terrain, nous pourrions envisager de traduire en termes de coûts le passage à quai des marchandises. Il ne s'agirait plus d'optimiser les ressources et le temps, mais des coûts.

### 5.5.1 Calcul de la quantité de main d'œuvre

Nous utilisons le modèle 5.31-5.42. Les flux d'entrées et de sorties de quai sont respectivement  $\Phi^{In}(t)$  et  $\Phi^{Out}(t)$ . Les flux de marchandises qui transitent entre les différentes aires de quai sont  $\psi^{In}$ ,  $\psi^{Res}$  et  $\psi^{Loa}$ . Leurs définitions sont données par les équations 5.34-5.41. Nous considérons ici que les quantités de force de travail  $n^{In}(t)$ ,  $n^{Ulo}(t)$ ,  $n^{Res}(t)$  et  $n^{Loa}(t)$  sont inconnues. Nous cherchons à les déterminer. Nous savons qu'à une date  $t_i$ , les quantités de masses de marchandises  $S^{In}(t_i)$ ,  $S^{Res}(t_i)$ ,  $S^{Loa}(t_i)$  et  $S^{Out}(t_i)$  sont connues. Nous cherchons à définir la quantité de main d'œuvre nécessaire,  $n$ ,  $n^{In}$ ,  $n^{Res}$ ,  $n^{Loa}$  et  $n^{Out}$ , pour traiter l'activité dans l'intervalle de temps  $[t_i, t_f]$ .

Nous savons que,

$$n^{In}(t) + n^{Ulo}(t) + n^{Res}(t) + n^{Loa}(t) = n(t) \quad \forall t \in [t_i, t_f], \quad (5.57)$$

et la question ci-dessus, posée en termes d'optimisation, revient à minimiser

$$\int_{t_i}^{t_f} n(t) ds, \quad (5.58)$$

sous la contrainte,

$$\frac{dS^{Out}}{dt} = \psi^{Loa} - \psi^{Out} \geq 0. \quad (5.59)$$

D'autres contraintes peuvent être ajoutées. Par exemple pour imposer une stabilité des quantités de main d'œuvre allouées nous imposons,

$$\left| \frac{dn^A}{dt}(t) \right| \leq n_{\max}, \quad \forall t \in [t_i, t_f], \quad (5.60)$$

où

$$\frac{1}{n^A(t) + 1} \left| \frac{dn^A}{dt}(t) \right| \leq n_{\max}, \quad \forall t \in [t_i, t_f]. \quad (5.61)$$

Pour une constante  $n_{\max}$ . De plus, nous pouvons imposer à  $n$  d'être constant sur les intervalles  $[t_i, t_1], [t_1, t_2], \dots, [t_k, t_f]$ .

## 5.5.2 Allocation de la main d'œuvre à chaque position

Nous travaillons ici avec le modèle « fin » 5.6-5.11. Pour chacune des positions  $p$  qui passent à quai entre une date  $t_i$  et  $t_f$ , nous connaissons la masse  $S^{A,(p)}(t_i)$  sur les différentes aires  $A \in \{In, Ulo, Res, Loa, Out\}$ . Nous considérons que les quantités  $n^{In}(t)$ ,  $n^{Ulo}(t)$  et  $n^{Res}(t)$ ,  $n^{Loa}(t)$  sont données pour tout  $t \in [t_i, t_f]$ . Elles ont, par exemple, été calculées par la procédure décrite en 5.5.1. Nous considérons aussi comme connues les quantités  $\Phi^{Out,(p)}(t)$  relatives à chaque position  $p$ . Celles-ci peuvent être déterminées par les formules 5.24 ou 5.25. Pour ce qui concerne  $\Phi^{In,(p)}(t)$ , la quantité est donnée par la formule 5.26, dans laquelle  $\tilde{T}(p)$  est une inconnue à déterminer. Les flux de marchandises sur le quai sont donnés par les formules 5.2-5.6, 5.12, 5.13, 5.16-5.18 où  $\nu(p) = \tilde{\nu}(\Pi(p, t))$  est posé et  $n^{A,(p)}$  est obtenu par 5.13 et 5.17, pour  $A \in \{In, Ulo, Res, Loa, Out\}$ .  $\gamma(p, t)$  est une fonction inconnue à déterminer. Enfin, la question de l'allocation de la main d'œuvre aux positions consiste à chercher  $\tilde{T}(p)$  et  $\gamma(p, t)$  tel que

$$\sum_{p/S^{A,(p)}(t)>0} n^{A,(p)}(t) \leq n^A(t) \quad \text{pour } A \in \{In, Ulo, Res, Loa, Out\}, \quad (5.62)$$

$$\gamma(p, t) = 0 \text{ pour tout } t \in [T_{In}(p), T_{in}(p) + \tilde{T}(p)], \quad (5.63)$$

où

$$\tilde{T}(p) = \tau^{In} \sum_{p' \in Pub(p)} \frac{N_p(p', T_{In}(p))}{n^{In,(p')}(t)}, \quad (5.64)$$

et

$$\frac{dS^{Out,(p)}}{dt}(t) = \psi^{Loa,(p)}(t) - \Phi^{Out,(p)}(t) \geq 0 \text{ pour tout } t \in [t_i, t_f], \quad (5.65)$$

minimisent

$$\int_{t_i}^{t_f} \sum_{\substack{p/S^{A,(p)}(t) > 0 \\ A \in \{In, Ulo, Res, Loa, Out\}}} \left| \frac{dn^{A,(p)}}{dt}(t) \right| dt. \quad (5.66)$$

La contrainte 5.62 indique que la somme des quantités de force de travail par position  $n^{A,(p)}(t)$  ne peut pas dépasser la quantité de force de travail globale  $n^A(t)$ .  $\tilde{T}(p)$  est le temps nécessaire à décharger les marchandises du camion avant d'atteindre celles de la position  $p$ . La formule 5.64 définit le temps  $\tilde{T}(p)$ . La formule 5.63 indique que l'indice de priorité de la position  $p$  est nul tant que la marchandise n'est pas atteignable dans le camion. La contrainte 5.65 indique que le flux de sorties de quai d'une position  $p$ , à une date  $t$ , ne peut pas dépasser ce qui est chargé de cette même position dans les camions, à la date  $t$ .

La formule 5.66 permet de minimiser le temps nécessaire à la force de travail pour passer d'une position à une autre.

### 5.5.3 Optimiser le temps $\tilde{\nu}(p)$

Le dernier problème d'optimisation consiste à calculer la valeur de  $\tilde{\nu}(p)$ . C'est la date à laquelle la position  $p$  sort de l'aire de résidence pour l'aire de sortie. Nous considérons que le nombre de palettes  $N_p(p, t)$  de la position  $p$  reste constant entre  $\nu(p)$  et  $T_{out}(p)$ .  $\nu(p)$  est défini par 5.22 avec un paramètre  $\tilde{\nu}$  dépendant de  $p$  et fixé par :

$$\tilde{\nu}(p) = \tilde{\nu}(\Pi(p, t)) = aN_p(p, T_{out}(p)) + b, \quad (5.67)$$

pour deux constantes  $a$  et  $b$ . Le problème de minimisation pour déterminer  $a$  et  $b$  consiste à considérer le modèle 5.6-5.10 dans l'intervalle de temps  $[t_i, t_f]$  avec les valeurs connues  $S^{A,(p)}(t_i)$  pour chaque position  $p$  passée à quai entre  $t_i$  et  $t_f$  et  $A \in \{In, Res, Loa, Out\}$ . Nous connaissons également  $\Phi^{In,(p)}$  et  $\Phi^{Out,(p)}$  pour tout  $t \in [t_i, t_f]$ . Les flux internes au quai sont donnés par 5.18-5.20, 5.24-5.27. Les valeurs de  $a$  et  $b$  doivent minimiser

$$\sum_{p/T_{In}(p) > t_i \text{ et } T_{Out}(p) < t_f} |\tilde{\nu}(p)|, \quad (5.68)$$

sous les contraintes

$$a \geq 0, b \geq 0 \text{ et} \tag{5.69}$$

$$\frac{dS^{Out,(p)}}{dt}(t) = \psi^{Loa,(p)}(t) - \Phi^{Out,(p)}(t) \geq 0 \text{ pour tout } t \in [t_i, t_f]. \tag{5.70}$$

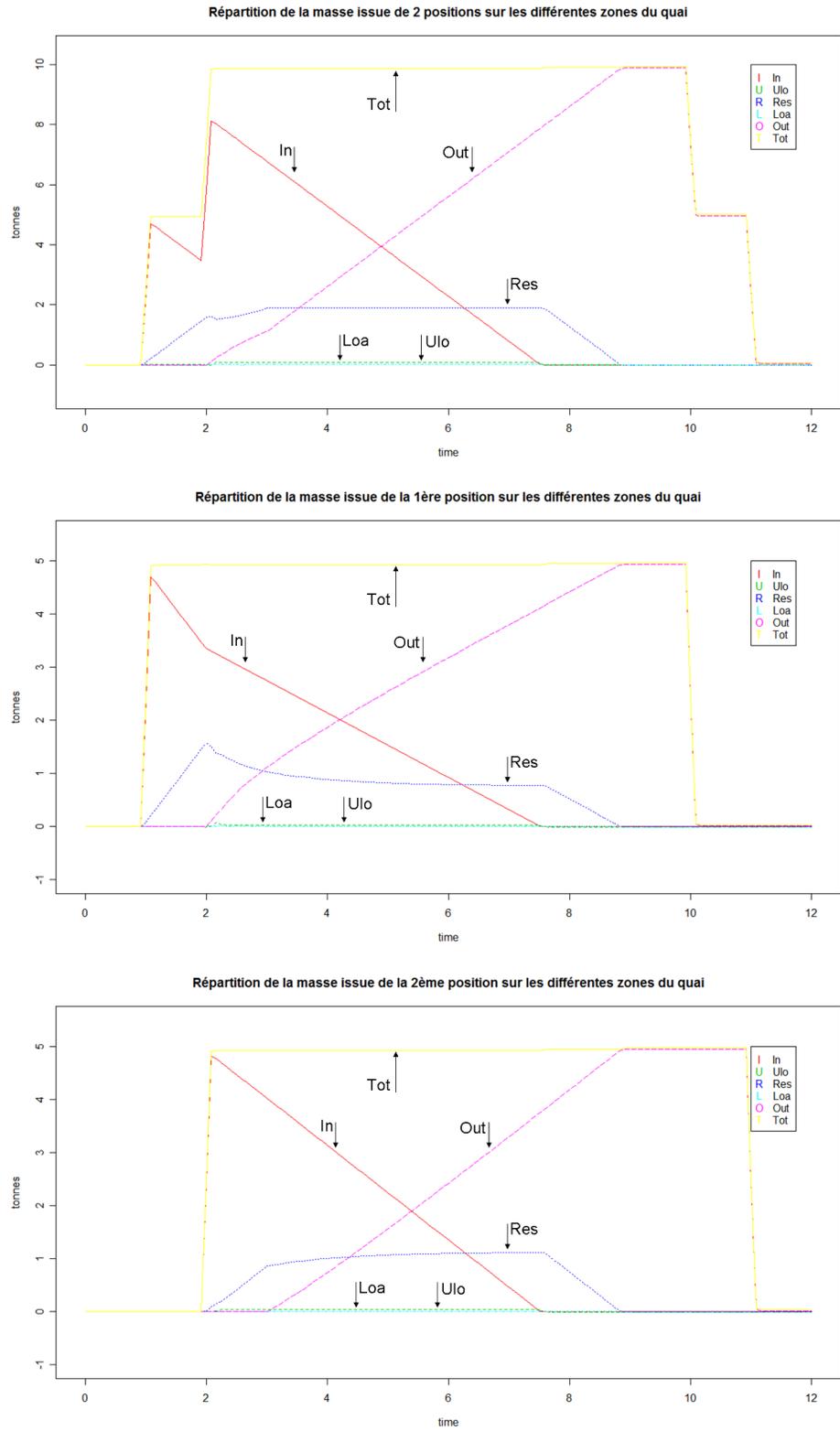


FIGURE 5.2 – Répartition de la masse des marchandises sur les différentes zones du quai, suivie du détail pour chaque position

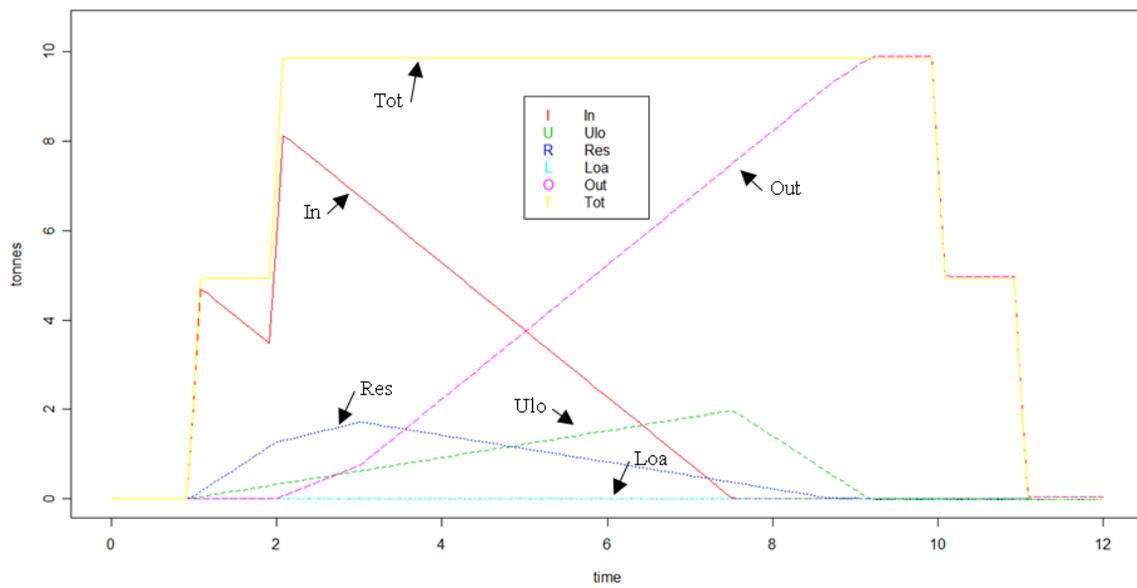


FIGURE 5.3 – Répartition de la masse des marchandises sur les différentes zones du quai, avec  $n^{Res} = 0,08$

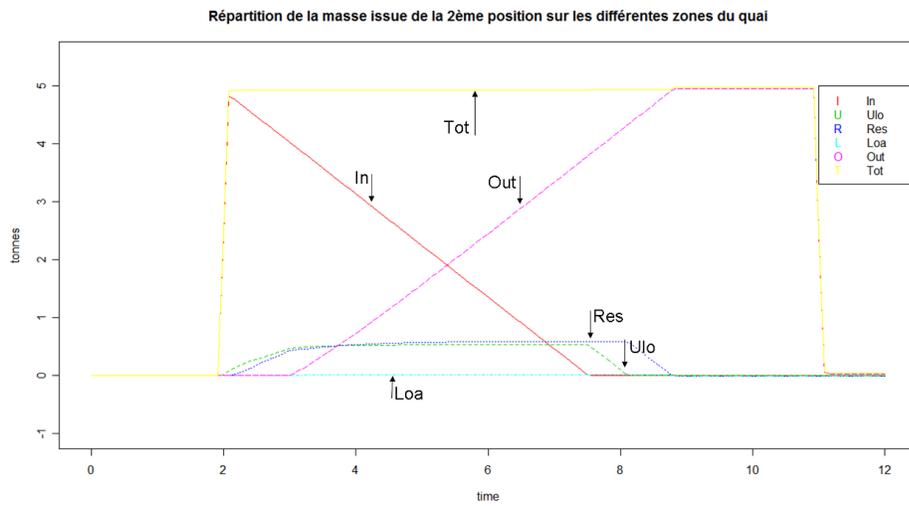
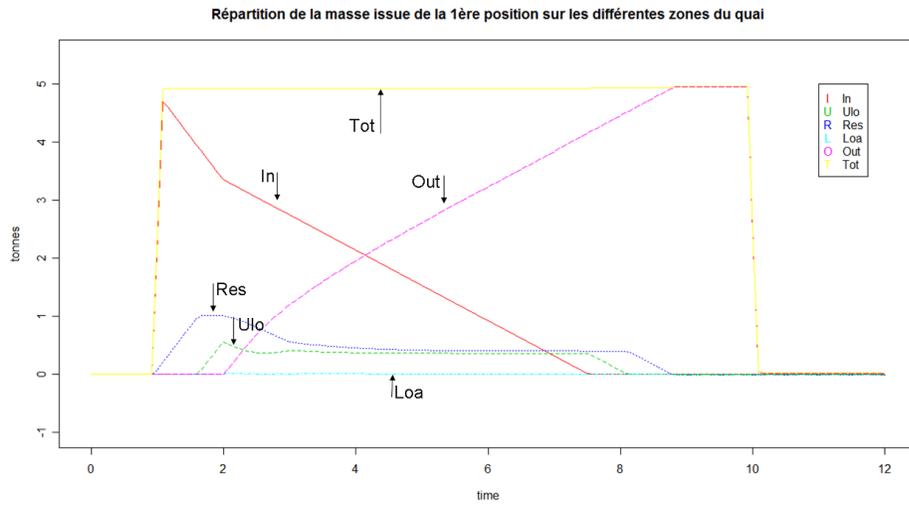
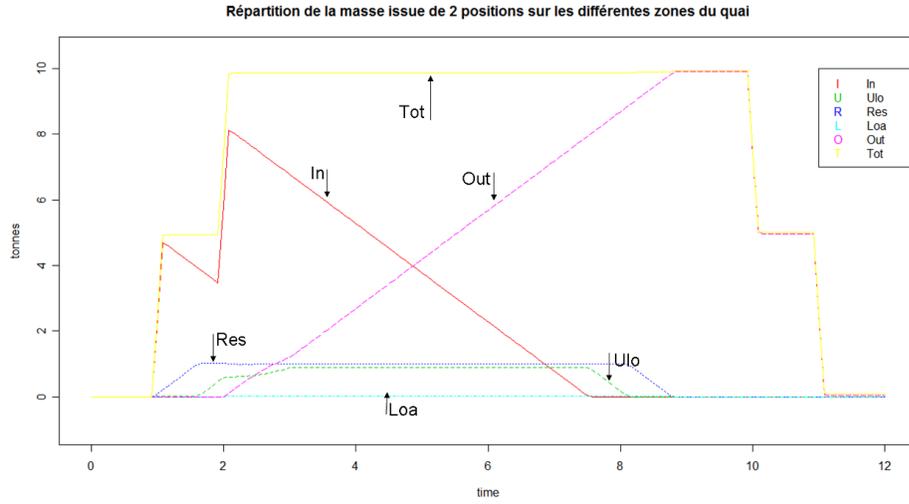


FIGURE 5.4 – Répartition de la masse des marchandises sur les différentes zones du quai, avec  $S_{\max}^{Res} = 1$

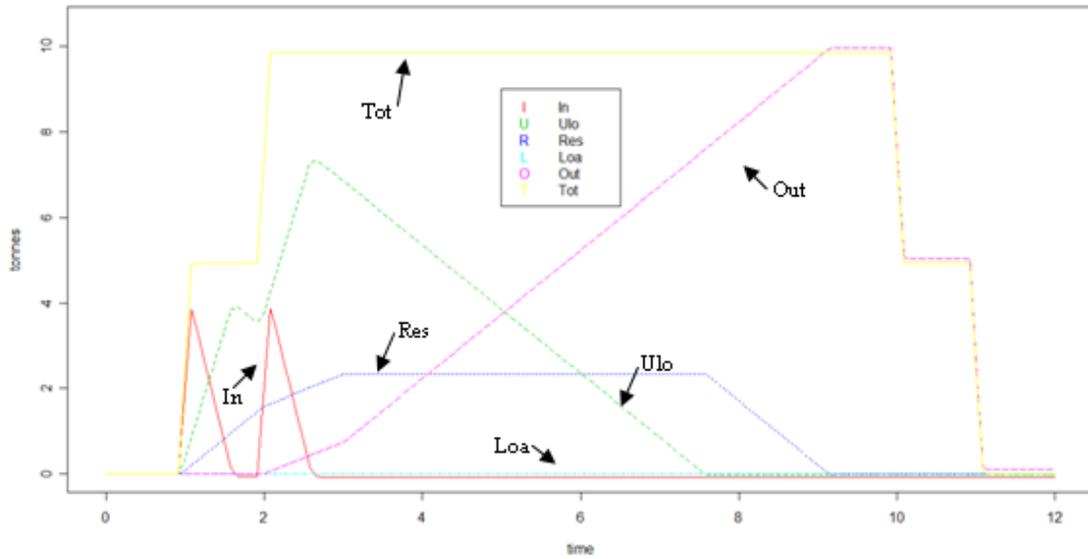


FIGURE 5.5 – Répartition de la masse des marchandises sur les différentes zones du quai, avec  $n^{ln} = 0,5$

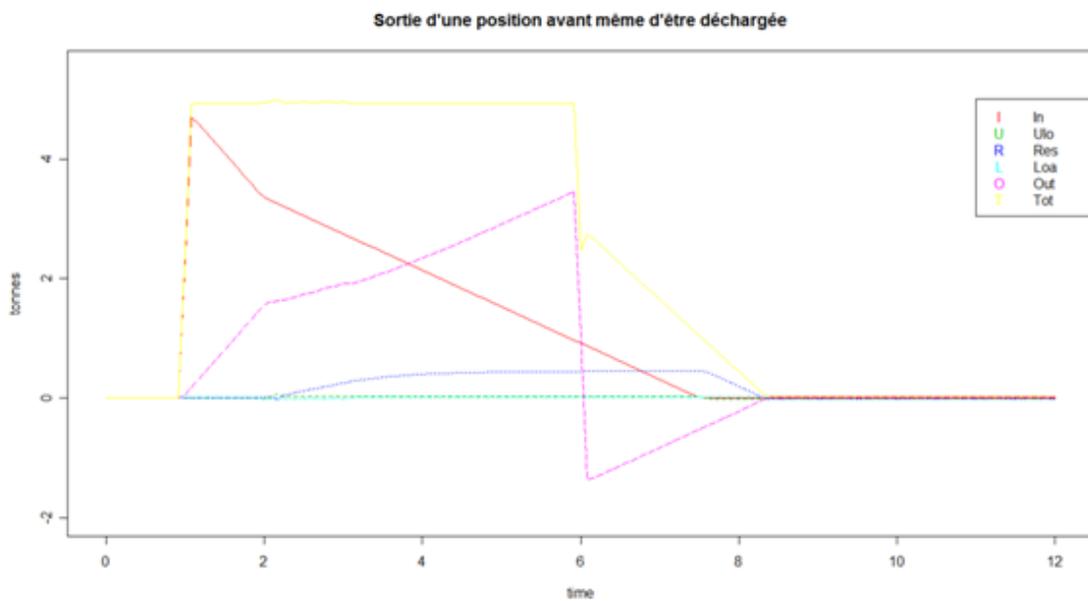


FIGURE 5.6 – Incohérence dans la répartition de la masse des marchandises sur les différentes zones du quai

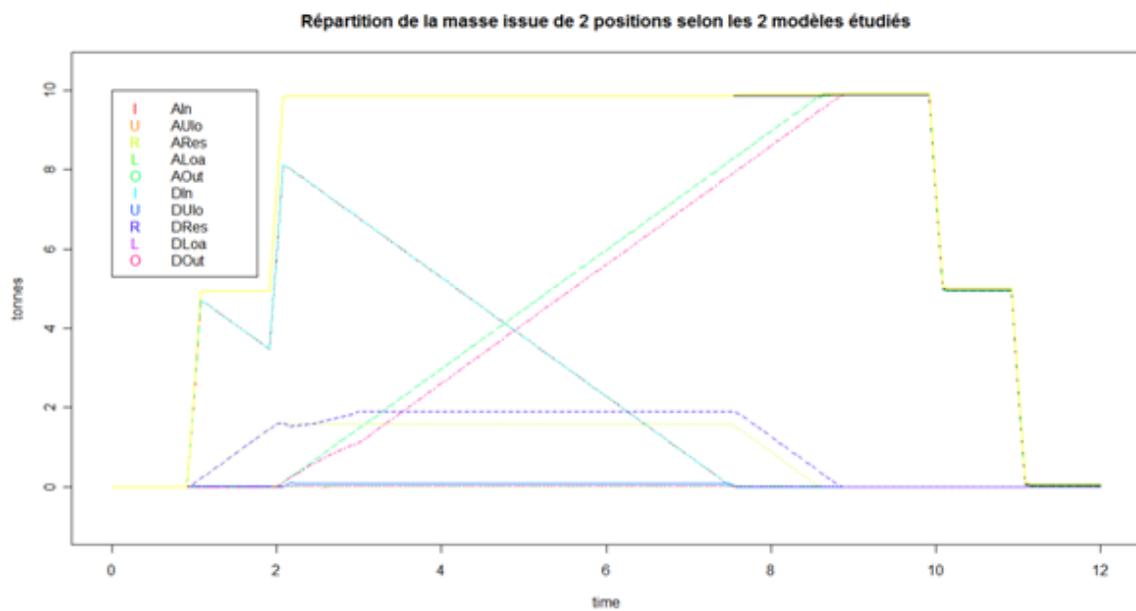


FIGURE 5.7 – *Superposition des résultats du modèle « fin » avec ceux du modèle « agrégé »*

# Conclusion

Rappelons que cette thèse a été réalisée dans un cadre industriel et en réponse à des besoins précis. A l'origine, il s'agissait de concevoir une solution permettant d'anticiper à très court terme le poids, le nombre de positions et l'équivalent en palettes au sol, des marchandises transportées par les agences d'un transporteur sous température dirigée. Nous avons montré dans le chapitre 1 que malgré des avis divergents, des souhaits difficilement réalisables et des définitions hétérogènes, le travail effectué a su synthétiser et recentrer le besoin des sites de transport du groupe STEF-TFE en information prévisionnelle. Pour fournir un système de prévision opérationnel, une documentation approfondie de l'existant dans le monde industriel et académique a été nécessaire. C'est l'objet du chapitre 2.

Le bilan de l'existant présenté dans le chapitre 3 a été un élément moteur à la définition précise des besoins de prévision, à la normalisation des indicateurs, à la rédaction d'un cahier des charges et à l'élaboration des recommandations.

Au de là du contexte précis dicté par le groupe STEF-TFE, nous nous sommes efforcés de généraliser la méthodologie produite. Cette méthodologie est documentée dans le chapitre 4. Elle est utilisable dans divers secteurs industriels et notamment ceux appartenant à la chaîne logistique. Bien entendu, le modèle mathématique de prévision sera à adapter, mais l'analyse de l'environnement, les choix dans l'organisation, la gestion de projet, le développement informatique, l'industrialisation, la formation du personnel et la communication sont exportables.

La volonté d'interconnecter les sites de transport pour améliorer les prévisions a conduit à étudier le fonctionnement des quais de messagerie. La compréhension et la maîtrise des flux de marchandises entre le déchargement et le chargement à quai, ont conduit à leurs modélisations. Elle a été documentée dans le chapitre 5. Notre travail a montré que cette modélisation permet de simuler le fonctionnement d'un quai et de poser certains problèmes industriels en termes de question d'optimisation utilisant le modèle. La modélisation qui a été faite du quai est inédite. Un article est en cours de formalisation pour être soumis dans une revue internationale.

Les perspectives d'étude sur le thème de la prévision sont nombreuses. A commencer par prévoir de nouveaux indicateurs, comme cela a déjà été fait avec la méthodologie proposée (68). Après un an de recueil des données qualitatives, lors des ajustements de prévision (voir 4.4.12), il est envisageable de les utiliser pour améliorer les prévisions. Rappelons qu'il s'agit d'expliquer quotidiennement le pourquoi des variations d'activité. En posant l'hypothèse que ces variations sont répétitives, le système de prévision pourra déclencher une alerte et ainsi avertir, par avance, les utilisateurs de leurs apparitions.

Un domaine qui dans cette thèse, pour des raisons de prescriptions initiales, n'a pas été assez largement investi est la recherche de facteurs explicatifs. Rappelons qu'il était essentiel d'aboutir à un modèle mathématique de prévision commun à toutes les agences de transport. Rappelons aussi que les agences ont une activité qui diffère selon leur position géographique et c'est pourquoi il semblait difficile de trouver des facteurs explicatifs communs. Cependant, il en existe certainement.

Nous avons mentionné qu'un point faible du modèle est le nombre trop important de paramètres. Une façon de réduire le nombre de paramètres est d'estimer la saisonnalité par une transformée de Fourier et de classifier des événements calendaires en fonction de leurs effets. Il est également envisagé d'utiliser la corrélation qui peut exister entre les séries temporelles des 70 agences pour améliorer la modélisation. Les corrélations proviennent du fait que les agences forment un réseau et s'expédient de la marchandise entre elles.

Par ailleurs, il reste à implémenter le modèle de l'organisation des quais en réseau (voir section 5.3). Ce modèle permettra d'investir dans la recherche des interactions entre agences de transport.

La modélisation des flux de marchandises traversant le quai en équations mathématiques a permis de poser les bases d'un sujet de recherche. Cette étude a révélé de formidables possibilités pour l'optimisation de l'organisation d'un quai et a donné des réponses à des problématiques industrielles. L'étude a aussi pointé les difficultés, notamment celle du grand nombre de paramètres à maîtriser. Les algorithmes de résolution d'équations différentielles ordinaires et d'optimisation sont gourmands en ressources informatiques. Le nombre important de positions à insérer dans le modèle ralentit aussi les traitements d'exécution. Nous avons en outre mentionné que le déploiement industriel d'un modèle de ce type nécessite une étude précise pour fixer la valeur de ces paramètres pour chacun des quais du transporteur.

Ces deux derniers points ne pourront être abordés de manière efficace qu'en utilisant une méthodologie de calcul scientifique.

# Bibliographie

- [1] Conception d'un quai de chargement. catalogue, Schiltz, 2005.
- [2] Comment agir sur la demande? Rapport technique, econokoi.org, [http://www.econokoi.org/IMG/article\\_PDF/marketing-comment-agir-sur-la-demande.pdf](http://www.econokoi.org/IMG/article_PDF/marketing-comment-agir-sur-la-demande.pdf), 2006.
- [3] Bovas. ABRAHAM et Johannes LEDOLTER : *Statistical Methods for Forecasting*. Wiley-Interscience, 1983.
- [4] Kenneth B. ACKERMAN : CPFR - How it could change the warehouse. *Warehousing Forum*, 15(10):1–2, 2000.
- [5] J.C. ANDRASKI : Supply chain collaboration. <http://www.foodlogistics.com/collaboration.html>, 1999.
- [6] G. ANSION : *Les méthodes de prévisions en économie*. Armand Colin, 1990.
- [7] J.Scott ARMSTRONG : *Principles of Forecasting : A Handbook for Researchers and Practitioners*. Springer, 2001.
- [8] J.Scott ARMSTRONG et Roderick J. BRODIE : Forecasting for Marketing. *International Thompson Business Press*, pages 92–119, 1999.
- [9] J.Scott ARMSTRONG et Fred COLLOPY : Integration of Statistical Methods and Judgment for Time Series Forecasting : Principles from Empirical Research. *International Journal of Forecasting*, 18:159–161, 1998.
- [10] Antoine ARTOUS : Massification, groupage et flux tendus : les nouvelles lois du métier. *L'officiel des Transporteurs*, 2355 supplément(Avril):29, 2006.
- [11] Voluntary Interindustry Commerce Standards ASSOCIATION : Collaborative planning, forecasting, and replenishment. <http://www.cpfr.org>, 1998.
- [12] Sofiane AYADI : Le Supply Chain Management : Vers une optimisation globale des flux. Working paper, Université Catholique de Lyon, 2005.

- [13] M. BARRATT et A. OLIVERIA : Exploring the experience of collaborative planning initiatives. *International Journal of Physical Distribution & Logistics Management*, 31:266–89, 2001.
- [14] Paul BEYNON-DAVIES : *Systèmes d'information - Conception et mise en oeuvre*. AF-NOR, 1989.
- [15] Craig H. BISHOP, Brian J. ETHERTON et Majumdar Sharanya J. : Adaptive sampling with the ensemble transform Kalman filter. Part I : Theoretical aspects. *Monthly weather review*, 129(3):420–436, 2001.
- [16] Barry W. BOEHM : A spiral model of software development. *Computer*, 21(5):61–72, 1988.
- [17] Renato BORTOLOTTI : Analyse du système de prévision des effectifs d'élèves dans l'enseignement secondaire postobligatoire. Note d'information, Service de la recherche en éducation, Genève, 2005.
- [18] Régis. BOURBONNAIS et Jean Claude USUNIER : *Prévision des ventes, théorie et pratique*. Economica, 2007.
- [19] George Box et Gwilym JENKINS : *Time series analysis : Forecasting and control*. San Francisco : Holden-Day Inc, 1970.
- [20] François BRANCHE : Transport de messagerie. *Revue technique de l'ingénieur*, dossier n°AG8151 V2, 2006.
- [21] J. Peter BROCKWELL et A. Davis RICHARD : *Introduction to Time Series and Forecasting*. Springer, 1996.
- [22] S.R. BRUBACHER et G. TUNNICLIFFE : Interpolating Time Series with Application to the Estimation Holiday Effects on Electricity Demand. *Applied Statistics*, 25(6):107–116, 1976.
- [23] S W BURGIN, S L GOLICIC et T M MCARTHY : Leveraging Collaborative Forecasting to Optimize Suplly Chain Performance. *Council of Logistics Management*, pages 243–251, 2000.
- [24] B. BURTSCHY et C. MENENDIAN : A propos de prévision à court terme de la production industrielle. *Revue De Statistique Appliquée*, 28(2):5–24, 1980.
- [25] Christoph Heinrich Diedrich Buys-BALLOT : Les changements périodiques de temperature. *Utrecht*, 1847.

- [26] Valentina CARBONE : *Le rôle des prestataires logistiques en europe - Intégration des chaînes et alliances logistiques*. Thèse, Ecole Nationale des Ponts et Chaussées, 2004.
- [27] Chris CHATFIELD : *Time-Series Forecasting*. Chapman & Hall/CRC, 2000.
- [28] Hongmei CHEN, Brani VIDAKOVIC et Dimitri MAVRIS : Multiscale Forecasting Method using ARMAX Models. Paper, Georgia Institute of Technologie.
- [29] Guillaume CHEVILLON : Analyse économétrique et compréhension des erreurs de prévision. *Revue de l'OFCE*, 95:327–356, 2005.
- [30] Guillaume CHEVILLON : Econométrie. Document de travail, OFCE & Univ of Oxford, 2005.
- [31] Geoffroy CLUZEL : Rentabilité d'un système d'information. Approche théorique. *Revue technique de l'ingénieur*, dossier n°AG5310, 2006.
- [32] CNRS : Conduite de projet - Systèmes d'information. <http://www.dsi.cnrs.fr/conduite-projet/>, 2009.
- [33] Gilbert COLLETAZ : Les Critères de sélection. Cours, Université d'Orléans, Master ESA, 2007.
- [34] D. Jonathan. CRYER et Kung-Sik CHAN : *Time Series Analysis with applications in R*. Springer, 2008.
- [35] J.J. DAUDIN, C. DUBY, S. ROBIN et P. TRÉCOURT : Analyse de Séries Chronologiques. Cours, INA-PG, Mathématiques, <http://www.agroparistech.fr/IMG/pdf/Polychro.pdf>, 1996.
- [36] J.-C. DAVID, M. et Michaud : *La prévision : approche empirique d'une méthode statistique*. Masson, 1989.
- [37] Wilfried DESPAGNE : Etude préliminaire à un modèle de prévision à court terme de l'activité d'un transporteur sous température dirigée. *Modulad*, 39:95–106, 2008.
- [38] Wilfried DESPAGNE : Etude préliminaire à un modèle de prévision à court terme de l'activité d'un transporteur sous température dirigée. In *MAJECSTIC, Session : Modélisation*, Marseille, France, 2008.
- [39] Wilfried DESPAGNE : Prévision à court terme de l'activité des 57 plates-formes du transporteur sous température dirigée TFE. In *EGC, Session : Fouille de données temporelles*, Sophia Antipolis , France, 2008.

- [40] Wilfried DESPAGNE : Short-term forecast for logistic activities. *In Information System & Data Analysis*, Vannes, France, 2008.
- [41] Wilfried DESPAGNE : A Forecasting Support System Dedicated to Temperature-Controlled Goods Hauling. *Springer, Computational Statistics*, Soumis:–, 2009.
- [42] Wilfried DESPAGNE : A Forecasting System Developed under R, Dedicated to Temperature-Controlled Goods Hauling. *In useR! Session : Focus Forecasting in Practice, book of abstracts p 50*, Rennes, France, 2009.
- [43] Wilfried DESPAGNE : A Forecasting System Developed under R, Dedicated to Temperature-Controlled Goods Hauling. *In DSC 2009, Session 7b, book of abstracts p 30*, Copenhagen, Denmark, 2009.
- [44] Hubert d'HONDT : ERP et Planification, convictions et retours d'expérience. [http://www.journaldunet.com/solutions/printer/031106\\_chro\\_dhondt.shtml](http://www.journaldunet.com/solutions/printer/031106_chro_dhondt.shtml), 2007.
- [45] D. DIEHN : Seven steps to build a successful collaborative forecasting process. *The Journal of Business Forecasting Methods & Systems*, 19:23,28–9, 2000/2001.
- [46] Danièle DUFOUR, Jean-Guy et Patier : Du transport de marchandises en ville à la logistique urbaine. *Revue technique de l'ingénieur*, dossier n°AG8210, 2006.
- [47] Jean-Marie DUFOUR : Histoire de l'analyse des séries chronologiques. [http://www.euroestech.net/resources/tsa\\_history.pdf](http://www.euroestech.net/resources/tsa_history.pdf), 2003.
- [48] Daniel EILSTEIN et AL. : Séries temporelles et modèles de régression. Rapport, Institut de veille sanitaire - Département santé environnement, [http://fulltext.bdsp.ehesp.fr/Invs/Rapports/2005/rapport\\_series\\_temporelles.pdf?8WW88-3X698-WJ944-J94WM-8XXG4](http://fulltext.bdsp.ehesp.fr/Invs/Rapports/2005/rapport_series_temporelles.pdf?8WW88-3X698-WJ944-J94WM-8XXG4), 2005.
- [49] Nathalie FABBE-COSTES : Système d'information logistique et transport. *Revue technique de l'ingénieur*, dossier n°AG8030 V2, 1999.
- [50] Laurent FERRARA et Dominique GUEGUAN : Analyse d'intervention et prévisions. Problématique et application à des données RATP. *Revue de Statistique Appliquée*, 48(2):55–72, 2000.
- [51] Robert FILDES, Paul GOODWIN et Michael LAWRENCE : The design feature of forecasting support systems and their effectiveness. *Decision Support Systems*, 42:351–361, 2006.
- [52] Jean-Paul FITOUSSI : Débat sur les prévisions. *Revue de l'OFCE*, 63:84–92, 1997.

- [53] Helena FORSLUND et Patrik JONSSON : The impact of information quality on supply chain performance. *International Journal of Operations & Production Management*, 27:90–107, 2007.
- [54] Jean-Michel FRANCO et Sandrine LIGNEROLLES : *Piloter l'entreprise grâce au data warehouse*. Eyrolles, 2000.
- [55] Placide FRESNAIS : Méthode d'accès et performances. *In Ecole Modulad - SFdS*, Port-aux-Rocs, Le Croisic, 1999.
- [56] André GAC : Logistique intégrée : le froid - Action du froid sur les denrées périssables. *Revue technique de l'ingénieur*, dossier n°A9065, 2001.
- [57] Bernadette GOVAERTS : Calcul Statistique sur Ordinateur. <http://www.stat.ucl.ac.be/cours/stat2430/documents/c5reglin.pdf>, 2007.
- [58] C.W.J. GRANGER : *Forecasting economic time series*. Academic Press, 1986.
- [59] Amara GRAPS : An Introduction to Wavelets. *IEEE Computational Sciences and Engineering*, 2(2):50–61, 1995.
- [60] Pierre-Julien GUAY : Introduction aux licences de logiciels libres GNU/GPL. Rapport technique, NTIC.org, <http://ntic.org>, [lePortaildesTIC](http://lePortaildesTIC), 2002. Direction du soutien aux établissements du ministère de l'Education du Québec.
- [61] Olaf HELMER et Norman DALKEY : An Experimental Application of the Delphi Method to the Use of Experts. *Management Science*, 9(3):458–467, 1963.
- [62] François HETMAN : *Le langage de la prévision*. Futuribles S.E.DE.I.S, 1969.
- [63] Charles C. HOLT : Forecasting seasonals and trends by exponentially weighted moving averages. *International Journal of Forecasting*, 20(1):5–10, 1957.
- [64] François-Bernard HUYGHE : Web 2.0 : influence, outils et réseaux. *Revue Internationale d'Intelligence Economique*, page Série Publications Numériques, 2009.
- [65] F. IBANEZ et M. ETIENNE : Le filtrage des séries chronologiques par l'analyse en composantes principales de processus. *Journal de Recherche Océanographique*, 16:27–33, 1992.
- [66] J.P. INDJEHAGOPIAN : Analyse des effets d'intervention sur les prévisions issues de modèles ARIMA. *Revue De Statistique Appliquée*, 29(4):5–30, 1981.
- [67] R. IRELAND et R. BRUCE : CPFR : only the beginning of collaboration. *Supply Chain Management Review*, pages 80–8, 2000.

- [68] Maria JAIMES : Développement d'un modèle de prévisions mensuelles qui anticipe l'activité des agences de transport de STEF-TFE. Stage, STEF-TFE, 2009.
- [69] Franck JEDRZEJEWSKI : *Introduction aux méthodes numériques*. Springer, 2005.
- [70] P.G.W. KEEN et M.S. SCOTT MORTON : *Decision Support Systems : An Organizational Perspective*. Addison Wesley, Reading, OrgPersp, 1978.
- [71] Sharon KOZICKI et P. A. TINSLEY : Vector rational error correction. Working Paper 98-03, Federal Reserve Bank of Kansas City, 1998.
- [72] Dominique LADIRAY : Econométrie, prévision et analyse des données. *Revue Modulad*, (35):6–24, 2006.
- [73] L. LAPIDE : New developments in business forecasting. *The Journal of Business Forecasting Methods & Systems*, 18:24–5, 1999.
- [74] Monique LE GUEN : La boîte à moustaches de TUKEY un outil pour initier à la Statistique. *SFDS*, 4:1–3, 2001.
- [75] Sébastien LECHEVALIER : Une introduction à l'économétrie des séries temporelle. *La revue des sciences économiques et sociales*, (13):45–51, 1998.
- [76] Matthieu LEMOINE et Florian PELGRIN : Introduction aux modèles espace-état et au filtre de Kalman. *Revue de l'OFCE*, 86:203–229, 2003.
- [77] Lon-Mu LIU, Siddhartha BHATTACHARYYA, Stanley L. SCLOVE et Rong CHEN : Data mining on time series : an illustration using fast-food restaurant franchise data. *Computational Statistics & Data Analysis*, 37:455–476, 2001.
- [78] R. LOUFOULLAH : Méthodes d'extrapolation à court terme des chroniques passées. *Revue de Statistique Appliquée*, 11(2):21–38, 1963.
- [79] L'e magazine cetelem de la DISTRIBUTION : Surgelé, les spécialistes tirent leur carte du jeu ! <http://www.dyn-com.com/dyncom104/marche.htm>, 2006.
- [80] Edward J. MARIEN : Demand Planning and Sales Forecasting : A Supply Chain Essential. *The Supply Chain Management Review*, pages 76–86, 2001.
- [81] S. MARKRIDAKIS, S.C. WHEELWRIGHT et R.J. HYNDMAN : *Forecasting Methods and Applications, Third Edition*. John Wiley & Sons, Inc., 1998.
- [82] 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.

- [83] Terea M. McCARTHY et Susan L. GOLICIC : Implementing collaborative forecasting to improve supply chain performance. *International Journal of Physical Distribution & Logistics Management*, 32:431–454, 2002.
- [84] John Tom MENTZER et Roger GOMES : Further Extensions of Adaptive Extended Exponential Smoothing and Comparison with the M-Competition. *Journal of the Academy of Marketing Science*, 22:372–382, 1994.
- [85] John Tom MENTZER et A.M. MOON : *Sales Forecasting Management - A demand management approach*. SAGE, 2005.
- [86] John Tom MENTZER et Jon SCHROETER : Integrating Logistics Forecasting Techniques, Systems, and Administration : The Multiple Forecasting System. *Journal of Business Logistics*, 15:205–225, 1994.
- [87] Séverine MEUNIER MARTINS : *Approche stratégique pour la prévision de données de gestion : Méthode ciblée intégrant des éléments subjectifs et mathématiques*. Thèse, Ecole Polytechnique Fédérale de Lausanne, 2003.
- [88] MINISTÈRE DE L'ÉCOLOGIE DE L'ÉNERGIE DU DÉVELOPPEMENT DURABLE ET DE L'AMÉNAGEMENT DU TERRITOIRE : Arrêté du 29 août 2008 portant agrément d'un organisme pour délivrer aux opérateurs les attestations de capacité ... *Journal officiel de la république française*, texte 2:sur 119, 2008.
- [89] Guy MÉLARD : *Méthodes de prévisions à court Terme*. Editions Ellipses, 1990.
- [90] Guy MÉLARD : Initiation à l'analyse des séries temporelles et à la prévision. *Modulad*, 35:82–129, 2006.
- [91] A.M. MOON, John Tom MENTZER et E.D THOMAS : Customer Demand Planning at Lucent Technologie. *Industrial Marketing Management*, 29:19–26, 2000.
- [92] A MORINEAU, A-E SAMMARTINO, M GETTLER-SUMMA et C PARDOUX : Analyse des données et modélisation des séries temporelles. *Revue de Statistique Appliquée*, 42(4):61–81, 1994.
- [93] Chantal MORLEY : *Gestion d'un projet système d'information ...* InterEditions, 1999.
- [94] J. MÉRAUD et A. TYMEN : Les variations saisonnières de l'activité économique. *Revue de Statistique Appliquée*, 42(4):15–68, 1960.
- [95] Akal MUSTAFA : Forecasting Turkey's tourism revenues by ARMAX model. *Energy Policy*, 25:565–580, 2004.
- [96] Seetharama L. NARASIMHAN, Dennis W. McLEAVY et Peter BILLINGTON : *Production Planning and Inventory Control*. Prentice-Hall, 1995.

- [97] OECD : Feature Article : Seasonal Adjustment. Rapport technique, OECD - Main Economic Indicators, <http://www.oecd.org/dataoecd/48/0/2073999.pdf>, 1999.
- [98] Laure PICHOT : *Stratégie de déploiement d'outils de pilotage de chaînes logistiques : Apport de la classification*. Thèse, Institut National des Sciences Appliquées de Lyon, 2006.
- [99] J.M. POGGI : Préviation non paramétrique de la consommation électrique. *Revue de Statistique Appliquée*, 42(4):83–98, 1994.
- [100] Gérard POUMEYROL : La logistique des produits à température dirigée. *Revue Générale du Froid*, 81:59–60, 1991.
- [101] Ricco RAKOTOMALALA : TANAGRA : un logiciel gratuit pour l'enseignement et la recherche. in *Actes de EGC'2005, RNTI-E-3*, 2:697–702, 2005.
- [102] S. RESSE : The human aspects of collaborative forecasting. *The Journal of Business Forecasting Methods & Systems*, 19:3–9, 2000/2001.
- [103] S.P. SCHNAARS : An evaluation of rules for selecting an extrapolation model on yearly sales forecasts. *Interfaces*, 16:100–1007, 2006.
- [104] Nicholas SEIERSEN : Systèmes d'information en logistique. *Revue technique de l'ingénieur*, dossier n°A9050, 1998.
- [105] Nicholas SEIERSEN : Systèmes d'information logistique. *Revue technique de l'ingénieur*, dossier n°AG5300, 2006.
- [106] SETRA/CSTR : Pourquoi faut-il désaisonnaliser ? Note d'information, Ministère de l'Ecologie, de l'Energie, du Développement Durable et de l'Aménagement du Territoire, 2001.
- [107] J. SMÂROS : Collaborative forecasting : a selection of practical approaches. *International Journal of Logistics*, 6:245–258, 2003.
- [108] Valentina STAN : Une étude comparative de logiciels de prévision automatique de séries chronologiques. *Modulad*, 33:108–128, 2005.
- [109] Tarkan TAN : Using Imperfect Advance Demand Information in Forecasting. *Journal of Management Mathematics*, 19:163–173, 2007.
- [110] Gerard J. TELLIS : The Price Elasticity of Selective Demand. *Journal of Marketing Research*, 25(4):331–341, 1988.

- [111] Charlotte TERROLLE : *Evolution 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.
- [112] J. Jordan THOMAS et R. Savioz MARCEL : Faut-il combiner les prévisions de modèles VAR ? Bulletin trimestriel, Banque Nationale Suisse, Zurich, 2003.
- [113] Thomas THRON, Gabor NAGHY et Niaz WASSAN : Evaluating alternative supply chain structures for perishable products. *The international journal of logistics Management*, 18:364–384, 2007.
- [114] Stéphane TUFFÉRY : *Data mining et statistique décisionnelle : l'intelligence des données*. Technip, Paris, 2007.
- [115] Michel VATÉ : *Statistique Chronologique et Prévisions*. Economica, 1993.
- [116] M. C. VIANO et A. PHILIPPE : *Cours de Séries Temporelles*. Cours, Université des Sciences et Technologies de Lille, 2004.
- [117] S VOLKAN et A. SERTAÇ : ARIMA forecasting of primary energy demand by fuel in Turkey. *Energy Policy*, 35:1701–1708, 2007.
- [118] William Wu-Shyong WEI : *Time series analysis : univariate and multivariate methods*. Addison-Wesley publ, 1994.
- [119] Handik WIDIARTA, S. VISWANATHAN et Rajesh PIPLANI : Forecasting aggregate demand : An analytical evaluation of top-down versus bottom-up forecasting in a production planning framework. *International Journal of Production Economics*, 118:87–94, 2009.
- [120] WIKIPÉDIA : Fêtes et jours fériés en France. [http://fr.wikipedia.org/wiki/F%C3%AAtes\\_et\\_jours\\_f%C3%A9ri%C3%A9s\\_en\\_France](http://fr.wikipedia.org/wiki/F%C3%AAtes_et_jours_f%C3%A9ri%C3%A9s_en_France), 2009.
- [121] N. WILSON : Game plan for a successful collaborative forecasting process. *The Journal of Business Forecasting Methods & Systems*, 20:3–6, 2001.
- [122] P. R. WINTERS : Forecasting sales by exponentially weighted moving averages. *Management Science*, 6:324–342, 1960.
- [123] Xiande ZHAO, Jinxing XIE et Janny LEUNG : The impact of forecasting model selection on the value of information sharing in a supply chain. *European Journal of Operational Research*, 142:321–344, 2002.
- [124] Eric ZIVOT, Jiahui WANG et Siem Jan KOOPMAN : State Space Modeling in Macroeconomics and Finance Using SsfPack for S+FinMetrics. Paper, University of Washington, <http://faculty.washington.edu/ezivot/statespacesurvey.pdf>, 2003.

# Glossaire

**agence** Centre TFE regroupant des activités commerciales et logistiques (transport et/ou prestation) servant de plate-forme de transit, de ramasse et de livraison pour la marchandise.. 2, 12, 19, 20, 22, 26–30, 35–37, 41, 42, 44, 63, 66, 69, 71, 97, 125, 211

**bordereau** Groupe de positions regroupées à l'initiative du transporteur. Nous distinguons plusieurs types de bordereaux : de groupage, de ramasses, de livraison, etc... (cf. définitions).. 101, 103, 106

**exploitation** D'une façon générale, les différents services d'exploitation ont en charge l'organisation des transports. Ces différentes exploitations sont identifiées au 2<sup>ème</sup> niveau de la structure de l'application GTI d'une agence par une codification issue d'une table.. 32, 103, 194

**expéditeur** Tiers qui expédie la marchandise à un destinataire. C'est souvent également le donneur d'ordre et le lieu de chargement.. 8, 21, 23, 29, 32, 35, 37, 100–102, 106, 193, 211

**expédition** Remise d'une marchandise par un expéditeur à un transporteur pour acheminement jusqu'au destinataire mentionné sur la lettre de voiture, quelque soit le mode utilisé (direct ou messagerie). Attention : ce terme ne s'applique pas au transporteur. Celui-ci ne fait pas de l'expédition mais de la ré-expédition de marchandises qu'il a enlevées ou qui lui sont confiées.. 6, 8, 20, 22, 27–29, 32, 33, 35–37, 42, 68, 96, 107, 110, 126

**groupage** Regroupement de marchandises destinées à plusieurs clients et remis à une même agence de dégroupage.. 2, 8, 17, 106

**lettre de voiture** Document d'accompagnement de la marchandise matérialisant le contrat de transport (récépissé, CMR, ...). 25, 28, 100, 106

**livraison** Transport et mise à disposition de la marchandise auprès du destinataire (reconnu sur la lettre de voiture) sur le lieu de déchargement (destinataire, intermédiaire, plate-forme, ...).. 7, 8, 12, 17, 22, 25, 27, 29, 42, 96, 126, 205

**lot complet** Terme usuel qui désigne le transport d'un lot en direct depuis le lieu de chargement jusqu'au lieu de déchargement et justifiant l'usage d'un véhicule exclusif.. 2, 8

**messagerie** Ensemble des opérations (ramasse, passage à quai, traction, passage à quai, livraison) permettant l'acheminement des marchandises à travers le réseau de plates-formes.. 2, 12, 17, 19, 96, 106

**ordre de transport** Demande de transport effectuée par un donneur d'ordre. Synonyme : position.. 36

**position** Traduction informatique de l'envoi d'un expéditeur à un destinataire (expéditeur, nombre de colis, poids, destinataire, localité du destinataire).. 25, 28, 30, 32, 33, 101, 126, 200, 201, 212, 214

**quai** Partie d'une plate-forme ou d'un entrepôt, ou d'une usine, servant à décharger, à trier et à recharger la marchandise. Quai est souvent utilisé dans le sens de plate-forme.. 8, 9, 22, 23, 28, 33, 44, 46, 68, 101, 125, 127, 128, 134, 138, 142, 159, 164, 165, 184–186, 190, 194, 196, 198, 199, 201, 208, 210

**ramasse** Opération de transport consistant à prendre en charge la marchandise chez un expéditeur pour l'acheminer jusqu'au quai de l'agence. Nous considérons pour une ramasse la quantité globale de marchandises à prendre en charge, et non le détail de chaque position constituant la ramasse. Une ramasse génère automatiquement, dès son affectation à un voyage, un point d'arrêt de type « Chargement » et un point d'arrêt de type « Déchargement » à l'agence.. 6, 8, 12, 20, 24, 25, 27–29, 32, 33, 42, 96

**régulière** Une série temporelle est dite régulière à temps discret lorsque les valeurs sont observées suivant un pas de temps régulier.. 135, 181, 182

**segment** Partie de parcours d'une position sur un voyage entre 1 point de chargement et un point de déchargement. Il existe 3 types de segment : - segment LD (livraison directe) = parcours depuis le lieu de chargement jusqu'au lieu de déchargement final, - segment EN (enlèvement) = parcours depuis le lieu de chargement jusqu'au lieu de passage à quai, - segment LI (Livraison) = parcours depuis le lieu de passage à quai jusqu'au lieu de déchargement final.. 33, 35, 68

**série temporelle** Une série temporelle ou encore chronique (TS : time series en anglais) est une suite d'observations ordonnées dans le temps.. 79, 82, 83, 108, 134

**tourné** Sous-ensemble des points à livrer ou à ramasser (habituellement par le même voyage) - à distinguer du voyage qui peut regrouper plusieurs tournées.. 20, 25, 27, 30, 32, 33, 36, 104, 126

**voyage** Le voyage est le parcours de la marchandise entre au minimum 2 points d'arrêts. Il peut également être constitué d'une succession de point d'arrêt. Un voyage d'Expédition comprend souvent un aller et un retour. Un voyage de Distribution (ou mixte Distribution/Ramasses) comprend souvent une « boucle » depuis le quai de l'agence jusqu'au retour au quai de l'agence. Un voyage d'Affrètement comprend souvent un aller simple (ou retour simple). Le voyage est à distinguer de la Mission du chauffeur. Le voyage peut comprendre une ou plusieurs tournées.. 22, 100, 101, 104, 106, 211, 212