

UNIVERSITÉ DU QUÉBEC À MONTRÉAL

ÉVALUATIONS DES RISQUES DANS LES CORRIDORS DE TRANSPORT
POUR L'AIDE HUMANITAIRE : LE CAS DU PROGRAMME
ALIMENTAIRE MONDIAL AU NIGER

MÉMOIRE
PRÉSENTÉ
COMME EXIGENCE PARTIELLE
DE LA MAÎTRISE EN TECHNOLOGIE DE L'INFORMATION

PAR
MARIE-PIER SÉGUIN

MAI 2019

UNIVERSITÉ DU QUÉBEC À MONTRÉAL
Service des bibliothèques

Avertissement

La diffusion de ce mémoire se fait dans le respect des droits de son auteur, qui a signé le formulaire *Autorisation de reproduire et de diffuser un travail de recherche de cycles supérieurs* (SDU-522 – Rév.07-2011). Cette autorisation stipule que «conformément à l'article 11 du Règlement no 8 des études de cycles supérieurs, [l'auteur] concède à l'Université du Québec à Montréal une licence non exclusive d'utilisation et de publication de la totalité ou d'une partie importante de [son] travail de recherche pour des fins pédagogiques et non commerciales. Plus précisément, [l'auteur] autorise l'Université du Québec à Montréal à reproduire, diffuser, prêter, distribuer ou vendre des copies de [son] travail de recherche à des fins non commerciales sur quelque support que ce soit, y compris l'Internet. Cette licence et cette autorisation n'entraînent pas une renonciation de [la] part [de l'auteur] à [ses] droits moraux ni à [ses] droits de propriété intellectuelle. Sauf entente contraire, [l'auteur] conserve la liberté de diffuser et de commercialiser ou non ce travail dont [il] possède un exemplaire.»

REMERCIEMENTS

Je tiens à remercier d'abord et avant tout mes trois directeurs de recherche : Walter Rei (ESG UQAM), Marie-Ève-Rancourt (HEC Montréal) et Claudio Contardo (ESG UQAM). Vous êtes absolument remarquables ! Vous avez su m'appuyer à merveille dans ce beau parcours pluridisciplinaire. Vous avez été présents, coordonnés et m'avez donné la chance d'apprendre de vous. J'en suis extrêmement reconnaissante. Une belle part de la réussite de ce projet vous revient !

J'aimerais remercier les gens qui partagent leurs connaissances en répondant à des questions techniques sur des sites comme StackExchange ou qui ont créé leur propre site web pour expliquer des concepts complexes. Une petite part du bon fonctionnement des modèles présentés dans ce travail vous revient !

Je veux aussi remercier les Fonds de recherche nature et technologie du Québec (FRQNT) pour avoir cru en mon projet et mes capacités de chercheure en devenir ! Sans l'appui que vous m'avez donné, je n'aurais jamais pu m'investir si pleinement dans la grande aventure de la maîtrise avec mémoire.

Je remercie aussi tous les membres de ma famille et mes amies. Vous avez toujours été là quand il le fallait, beau temps, mauvais temps ! Une part de la réussite de ce projet vous revient aussi !

Scott, thank you so much for being who you are. Without you it wouldn't have been so easy to push through all the difficulties I faced. A nice part of this achievement belongs to you, who holds me up on your shoulders.

TABLE DES MATIÈRES

LISTE DES FIGURES	vii
LISTE DES TABLEAUX	xi
LISTE DES ACRONYMES	xiii
RÉSUMÉ	xv
INTRODUCTION	1
CHAPITRE I CONTEXTE OPÉRATIONNEL DU PAM	7
1.1 Structure physique de la chaîne d'approvisionnement	8
1.2 Processus de distribution	9
1.3 Modèle d'optimisation	11
1.4 Sources d'incertitude et risques	14
CHAPITRE II PROBLÉMATIQUE, QUESTION DE RECHERCHE ET OBJECTIFS	17
2.1 Définitions des termes et des concepts	17
2.2 Problématique	19
2.3 Question de recherche	20
2.4 Objectifs	21
CHAPITRE III REVUE DE LITTÉRATURE	23
3.1 Identification des sources d'incertitude	24
3.2 Catégorisation et impacts des sources d'incertitude	25
3.2.1 La catégorisation des sources d'incertitude	26
3.2.2 L'impact des sources d'incertitude sur la chaîne d'approvision- nement	27
3.3 L'évaluation des risques	30
3.3.1 Les modèles prédictifs	31

3.3.2	Les types de simulation	34
3.3.3	L'application des méthodes en contexte d'aide humanitaire . .	37
CHAPITRE IV APPROCHE MÉTHODOLOGIQUE		39
4.1	Identifier et catégoriser les sources d'incertitude	41
4.2	Collecter et analyser les données	43
4.2.1	Les données objectives	44
4.2.2	Les données subjectives	47
4.3	Élaborer les modèles pour estimer les paramètres stochastiques	48
CHAPITRE V RÉSULTATS		59
5.1	Les sources d'incertitude	59
5.1.1	La catégorisation des sources d'incertitude	60
5.1.2	L'effet des sources d'incertitude sur les paramètres stochastiques	66
5.2	L'analyse des données	70
5.2.1	Les paramètres stochastiques à prédire	70
5.2.2	Les variables explicatives	84
5.3	Les modèles prédictifs	97
5.3.1	Comparaison des modèles et des erreurs de prédiction	97
5.3.2	Choix du modèle prédictif	105
CHAPITRE VI DISCUSSION		121
6.1	Génération de scénarios et simulation	122
6.2	Prédire pour évaluer les risques	132
CONCLUSION		139
BIBLIOGRAPHIE		143

LISTE DES FIGURES

Figure	Page
1.1 Routes pour l'approvisionnement au Niger (LCA, 2014)	8
1.2 Processus de la chaîne d'approvisionnement du PAM	10
1.3 Réseau d'approvisionnement du PAM au Niger (exemple pour l'op- timisation des coûts lors d'un don de moins de 2500 TM)	12
2.1 Application des termes et concepts	19
4.1 Schéma de la méthodologie quantitative	40
4.2 Exemple d'arbre de décision de base	53
5.1 Réseau avec sources d'incertitude affectant les délais	67
5.2 Réseau avec sources d'incertitude affectant les coûts	68
5.3 Réseau avec sources d'incertitude affectant les pertes	69
5.4 Histogramme des quantités transportées lors du transport maritime	72
5.5 Histogramme des quantités réparties dans les camions	73
5.6 Histogrammes des délais de transport maritime du continent d'ori- gine au port de Cotonou au Bénin	75
5.7 Histogrammes des délais de transport maritime du continent d'ori- gine au port de Lomé au Bénin	76
5.8 Histogrammes des délais de manutention aux ports	78
5.9 Histogrammes des délais de transport terrestre de Cotonou au Niger	79
5.10 Histogrammes des délais de transport terrestre de Lomé au Niger	80
5.11 Matrice de corrélation des paramètres stochastiques en sol africain	82
5.12 Boîtes à moustaches des délais de transport maritime en fonction du transporteur maritime	87

5.13 Nuage de points des délais de transport maritime en fonction de la distance nautique	87
5.14 Boîtes à moustaches des délais en sol nigérien (manutention et transport) en fonction du mois	95
5.15 Boîtes à moustaches des délais en sol nigérien (manutention et transport) en fonction du transitaire responsable	96
5.16 Exemple d'arbre de régression pour prédire le délai total en sol africain	104
5.17 Décroissance des erreurs (MSE) du modèle de <i>RandomForest</i> final, prédisant les délais de transport maritime, en fonction du nombre d'arbres.	110
5.18 Classifications des variables explicatives par ordre d'importance selon le modèle <i>RandomForest</i> final prédisant les délais de transport maritime.	110
5.19 Décroissance des erreurs (MSE) du modèle de <i>RandomForest</i> final, prédisant les délais en sol africain, en fonction du nombre d'arbres.	114
5.20 Classifications des variables explicatives par ordre d'importance selon le modèle de <i>RandomForest</i> final prédisant les délais en sol africain.	115
5.21 Classifications des variables explicatives par ordre d'importance selon le modèle <i>Boosting</i> final prédisant les délais en sol africain.	116
5.22 Décroissance des erreurs de validation croisée du modèle <i>Boosting</i> final, prédisant les délais en sol africain, en fonction du nombre d'arbres.	117
5.23 Graphiques de dépendance partielle entre une variable explicative et la prédiction du délai en sol africain.	118
5.24 Distribution des erreurs de prédictions (résidus) pour les trois modèles finaux en sol africain.	120
6.1 Schéma des activités du chapitre 6	122
6.2 Histogramme des prédictions du modèle <i>Combined</i> sur les scénarios de la simulation DeBase	132

6.3	Boîtes à moustaches comparant les prédictions du modèle <i>Combined</i> en fonction de l'intervalle de sélection des variables incertaines les plus importantes	133
6.4	Boîtes à moustaches comparant les prédictions du modèle <i>Combined</i> en fonction du transitaire sélectionné	135

LISTE DES TABLEAUX

Tableau	Page
5.1 Les catégories de sources d'incertitude	61
5.2 Analyse des pertes (en TM)	74
5.3 Statistiques descriptives des paramètres stochastiques	81
5.4 Les variables pour expliquer les délais de transport maritime . . .	85
5.5 Les variables pour expliquer les délais en sol africain (partie 1) . .	89
5.6 Les variables pour expliquer les délais en sol africain (partie 2) . .	91
5.7 Comparaison des modèles prédictifs pour les délais de transport maritime. Sur l'ensemble de validation (30%) de dix échantillons.	98
5.8 Comparaison des modèles prédictifs pour les délais de manutention au port. Sur l'ensemble de validation (30%) de dix échantillons. .	101
5.9 Comparaison des modèles prédictifs pour les délais de transport terrestre. Sur l'ensemble de validation (30%) de dix échantillons. .	101
5.10 Comparaison des modèles prédictifs pour les délais en sol africain (manutention et transport). Sur l'ensemble de validation (30%) de dix échantillons.	102
5.11 Comparaison des modèles d'arbres décisionnels multiple pour prédire les délais de transport maritime. Sur l'ensemble de test (15%) de dix échantillons.	108
5.12 Comparaison des modèles d'arbres décisionnels multiple pour prédire les délais en sol africain. Sur l'ensemble de test (15%) de dix échantillons.	112
6.1 Valeurs à utiliser pour la variable explicative AnnRevCarrierCo .	123
6.2 Variables explicatives les plus importantes pour l'estimation du délai en sol africain	124

6.3	Autres variables explicatives pour estimer le délai en sol africain .	126
6.4	Valeurs des variables explicatives connues dans l'exemple de simulation DeBase	127
6.5	Estimation des variables explicatives inconnues dans l'exemple de simulation DeBase	129
6.6	Échantillonnage des variables importantes des modèles de simulation	130
6.7	Niveau de risque des simulations	135

LISTE DES ACRONYMES

ASC	Approche de simulation combinée
COMPAS	<i>Commodity Movement Processing and Analysis system</i>
CP	<i>Complexity Parameter</i>
EDP	<i>Extended Delivery Point</i>
ETA	<i>Estimated time of arrival</i>
GBM	<i>Generalized Boosted Regression Models</i>
GRCA	Gestion des risques de la chaîne d'approvisionnement
HCR	Haut Commissariat des Nations Unies pour les réfugiés
LCA	<i>Logistics Capacity Assesment</i>
LTI	<i>Landside Transportation Instruction</i>
LESS	<i>Logistics Execution Support System</i>
MAE	<i>Mean Absolute Error</i>
MAPE	<i>Mean Absolute Percentage Error</i>
ME	<i>Mean Error</i>
MPE	<i>Mean Percentage Error</i>
MSE	<i>Mean Square Error</i>
ONU	Organisation des Nations Unies
PAM /	Programme alimentaire mondial /
WFP	<i>World Food Programme</i>
RFQ	<i>Request for quotation</i>
RMSE	<i>Root Mean Square Error</i>
TM	Tonne métrique
UNOCHA	<i>United Nations Office for the Coordination of Humanitarian Affairs</i>

RÉSUMÉ

Ce projet de recherche porte sur l'évaluation des risques dans la chaîne logistique de distribution des denrées alimentaires au Niger. Le Niger est un pays enclavé situé en Afrique de l'Ouest considéré comme l'un des moins développés et des plus pauvres au monde. Il est instable et tend à subir les effets des catastrophes à évolutions lentes : les sécheresses, les inondations et les mouvements de population causés par l'instabilité des pays voisins. Le Niger prend aussi plusieurs années à se relever de ses épisodes de famine. Dans le cadre de notre recherche, nous étudions les aspects logistiques du Programme alimentaire mondial (PAM) en ciblant les activités et les décisions qui doivent être prises entre la confirmation d'un don de denrées alimentaires et son arrivée aux entrepôts des grandes villes du Niger. Lorsqu'un don est confirmé à l'attention du Niger, les responsables de la logistique doivent choisir le(s) corridor(s) pour la distribution de la nourriture et déterminer qu'elle compagnie de transitaires parmi celles disponibles au(x) port(s) sélectionné(s) deviendra responsable de le récupérer à l'arrivée du navire, d'en faire le dédouanement et de le transporter par camion jusqu'au Niger. Notre projet vise à développer un modèle à la fois descriptif et prédictif des risques permettant d'évaluer les décisions prises en contexte d'incertitude.

Les phénomènes aléatoires affectant les corridors de transport ont été identifiés et catégorisés. Des données empiriques ont été récoltées afin de comprendre la variabilité des paramètres stochastiques affectant la performance des opérations logistiques. Des modèles prédictifs d'apprentissage statistique ont été développés, évalués et améliorés. Les meilleurs modèles ont été sélectionnés pour effectuer la prédiction du délai de transport maritime et du délai total en sol africain. Les modèles prédictifs basés sur les arbres de régressions multiples se sont montrés plus performants. L'intérêt d'inclure les sources d'incertitude de type sécuritaire, grâce à l'analyse des conflits régionaux, a été prouvée. Enfin, nous illustrons l'utilisation des solutions technologiques développées en générant des scénarios contextuels lorsque les informations nécessaires aux modèles sont incertaines au moment de la prise de décision. La technique de simulation de type Monte Carlo est utilisée pour évaluer les risques et les décisions prises. Les résultats des modèles développés pourraient servir à instancier des modèles prescriptifs utilisés au moment de la prise de décision.

Mots-clés : gestion des risques, logistique humanitaire, modèles prédictifs, simulation Monte Carlo.

ABSTRACT

This research project evaluates the risks affecting the logistical chain for food distribution to Niger. A West African landlocked country considered as one of the poorest and least developed countries of the world. Niger tends to suffer slow onset disasters : floods, droughts, famines, and displaced population caused by conflict and terrorism in its surrounding countries. We study the activities and decisions of the World Food Programme in Niger, from the moment a donation is confirmed until its arrival at the distribution centers. When a donation is confirmed for Niger, the logistics team must decide on which corridor the donations will travel and which forwarder company will be responsible for the handling at port, the custom clearance and the inland transportation. Our project aims at developing a descriptive and predictive model able to evaluate the risks and the decision taken under numerous uncertainty sources.

The random events affecting the transport corridors have been identified and categorized. Empirical data have been collected for a better understanding of the variability of the stochastic parameters that affect the performance of the logistic operations. Predictive models based on supervised statistical learning were iteratively developed, evaluated and improved. The best models were chosen to predict the maritime transportation time and the total inland operation time. The predictive models based on multiple regression trees had the best performance. The high importance of the security type of uncertainty source was proven with the inclusion of the of regional historical conflict data. We illustrate how generating contextual scenarios can facilitate the usage of the technological solutions during the decision making process when the needed information is incomplete. Therefore, Monte Carlo simulation technique is used to evaluate the risk and the decision making.

Key words : risks analysis, humanitarian logistics, predictive models, Monte Carlo simulation.

INTRODUCTION

L'aide humanitaire peut se faire dans plusieurs contextes, mais elle vise toujours à venir en aide et répondre aux besoins d'une population en difficulté. Elle peut s'activer en situation d'urgence lorsque, par exemple, une région est frappée par un tremblement de terre ou un tsunami. Elle peut être déployée en situation de crise ou de conflit ; une forme d'aide qui se distingue par l'importante implication des questions de sécurité et de politique au cœur de son élaboration. Selon l'évaluation du Bureau de la coordination des affaires humanitaires des Nations Unies (UNOCHA, 2017), les conflits représenteront la cause principale des besoins en aide humanitaire pour 2018. L'aide humanitaire peut aussi se faire sous forme de projet d'aide au développement ainsi elle peut, entre autre, contribuer à combattre ou prévenir la pauvreté ou encore les épidémies comme le choléra et le paludisme (UNOCHA, 2016). Cette dernière forme d'aide humanitaire diffère des opérations en situation de crise ou d'urgence, car elle est faite de manière régulière sur un horizon à plus long terme (Rancourt *et al.*, 2015; Holguín-Veras *et al.*, 2012; Altay et Green, 2006). L'Organisation des Nations Unies (ONU) est probablement l'acteur le plus important mondialement avec une présence et des programmes d'actions déployés dans de nombreux pays en difficultés. L'une de ses divisions, le Programme alimentaire mondial (PAM), se dédie à lutter contre la faim et l'insécurité alimentaire à travers le monde. L'ONU est considérée comme la plus grande organisation qui s'affaire à distribuer l'aide alimentaire aux communautés dans le besoin, notamment dans de nombreux pays d'Asie et d'Afrique. Les actions que le PAM entreprend sont de quatre types : les opérations d'urgence, les opérations de développement, les opérations spéciales et les opérations d'interventions prolon-

gées de secours et de redressements (WFP, 2017a). Selon un rapport sur le pays (WFP, 2017b), le PAM agit au Niger depuis 1968. L'équipe logistique du PAM est installée dans le pays à long terme. Lorsque le pays vit des situations catastrophiques, l'organisme doit aussi mener des opérations d'urgence. Ainsi, c'est la même équipe de logisticiens déployés au Niger qui s'assure d'acheminer les denrées alimentaires aux bénéficiaires et ce, quelque soit le type d'opération. Pour ce projet, nous nous concentrerons sur l'ensemble des opérations menées par le PAM au Niger.

Nous analysons le cas du Niger, un pays enclavé situé à l'ouest de l'Afrique considéré comme l'un des moins développés au monde. En effet, le Niger possède une population des plus pauvres et un indice de développement humain classé 188^e sur 188 pays répertoriés par l'ONU en 2015 (WFP, 2016). Il présente aussi un taux de fertilité des plus élevé ; sa population est actuellement de plus de 18 millions et elle augmente de 3,9% par année. Le PAM signale que 46,6% des enfants souffrent de malnutrition chronique. L'équipe du PAM travaille en sol nigérien pour appuyer le développement des communautés par des activités telles que les «*school meals*» ou les «*take-home rations*» (WFP, 2016). Le PAM au Niger doit aussi agir en situation d'urgence lorsqu'il y a des épisodes de famine ; la dernière a eu lieu en 2011 et les habitants peinent toujours à s'en relever. Ainsi, le Niger subit les effets de catastrophes à évolutions lentes. Les actions du PAM visent à aider les habitants à améliorer leurs mécanismes de survie pour subvenir à leurs besoins malgré tous les facteurs d'insécurité qui les entourent (inondations, sécheresses, activités terroristes, instabilité des pays voisins et épidémies). Une aide marquée est portée aux femmes enceintes, aux enfants de moins de cinq ans, aux milieux scolaires, ainsi qu'aux personnes en déplacement qui fuient les crises pour pallier les problèmes d'insécurité alimentaire.

Pour tout type d'aide humanitaire, la chaîne logistique débute généralement par

un don (monétaire ou en nature) en provenance de l'international. Dans le cas de l'aide alimentaire, les vivres doivent être transportés jusqu'à un port d'entrée, où ils peuvent être entreposés avant d'être transportés vers les centres de distributions pour ensuite être remis aux bénéficiaires. Jusqu'à ce jour, plusieurs chercheurs ont développé des outils d'aides à la décision pour des réseaux de distribution en contexte de réponse aux crises (Anaya-Arenas *et al.*, 2014). Par contre, des outils spécifiquement conçus pour des cas pratiques d'aide humanitaire sont nécessaires (Holguín-Veras *et al.*, 2012). Notre projet de recherche appliqué au cas du PAM compte y contribuer. Nous étudions les aspects logistiques, spécifiquement les risques, associés à l'aide alimentaire en ciblant les activités et les décisions à prendre entre le port d'entrée et les centres de distributions. La problématique décisionnelle considérée est complexe, elle est donc abordée par étape. Le mémoire vise à déterminer comment évaluer et mesurer les sources d'incertitude qui affectent les corridors de transport. L'étude cible le cas du PAM actuellement actif au Niger. Les sources d'incertitude sont nombreuses et leurs occurrences peuvent modifier le déroulement prévu des activités du PAM, ce qui affecte le processus décisionnel des dirigeants. Les événements aléatoires qui sont traités peuvent affecter les coûts associés au transport des denrées, les délais de livraison ou entraîner des pertes en termes de quantités de denrées. Notre étude vise à développer un modèle de type quantitatif, à la fois descriptif et prédictif des risques permettant d'évaluer les décisions prises en contexte d'incertitude grâce à des critères d'évaluation pour des solutions possibles. Autrement dit, les modèles permettent d'évaluer les décisions des responsables du PAM en fonction de leur niveau d'appétence aux risques. Des modèles prédictifs d'apprentissage-machine et la technique de simulation de type Monte Carlo, basé sur la génération de scénario, sont utilisés pour évaluer les risques qu'entraînent des délais de livraison prolongés.

Dans ce projet, l'élaboration de solutions technologiques a été faite en suivant les étapes classiques de la recherche opérationnelle (Ackoff et Sasini, 1968; Churchman *et al.*, 1957). Des modèles prédictifs et de simulation sont développés pour répondre à une problématique décisionnelle concrète. De manière itérative, les modèles sont testés, améliorés et rigoureusement évalués. Nous illustrons leur utilisation pratique en contexte d'incertitude. En développant ainsi des solutions pour un problème réel dans la phase de redressement suite à des catastrophes à évolutions lentes, nous répondons à l'appel de Galindo et Batta (2013) et Holguín-Veras *et al.* (2012). Nos modèles quantitatifs en gestion des risques, lorsque la prise de décision est difficile, répondent à l'appel de Tang et Musa (2011). De plus, peu d'études ont analysé à la fois les sources et les conséquences des risques (Heckmann *et al.*, 2015). En analysant l'impact des aspects temporels (p. ex., les saisonnalités) nos modèles d'apprentissage statistiques contribuent à combler un manque, identifié par Heckmann *et al.* (2015), au sein de la littérature scientifique. Le projet et les outils développés apportent aussi deux contributions pratiques pour l'organisation : 1) en permettant d'évaluer un plan d'approvisionnement un contexte d'incertitude, 2) en permettant de comparer et de confirmer ou d'infirmer les décisions qui sont généralement prises, en fonction de certains critères. Ainsi les dirigeants du PAM pourront se servir de notre modèle pour améliorer leur technique décisionnelle et leurs choix futurs.

La suite de ce mémoire se divise en sept chapitres. Le premier chapitre décrit le contexte opérationnel du PAM. Le deuxième présente la problématique étudiée, précise la question de recherche et les trois sous-objectifs visés. Le troisième chapitre comprend une revue de littérature portant sur les sources d'incertitude et leurs impacts sur la chaîne d'approvisionnement en contexte d'aide humanitaire. Notre revue de littérature présente aussi les outils et les techniques utilisées pour évaluer les risques. Le quatrième chapitre décrit l'approche méthodologique et le

cinquième les résultats. On analyse les sources d'incertitude, les données collectées et les modèles prédictifs. Dans la sixième chapitre nous présentons une preuve de concept en illustrant l'utilisation des modèles prédictifs pour contribuer à la prise de décision. Nous discutons de la génération de scénarios afin de répondre à notre question de recherche. Le dernier chapitre vient conclure le mémoire.

CHAPITRE I

CONTEXTE OPÉRATIONNEL DU PAM

L'approvisionnement en denrées alimentaires par le PAM au Niger a lieu dans un contexte difficile. L'instabilité au Niger est certainement particulière, comme le pays est traversé en majeure partie par le Sahel, il est caractérisé par un climat très chaud et aride qui cause de grandes sécheresses. Le Sahel est une zone de transition semi-désertique entre le désert du Sahara au nord et les régions tropicales et humides du centre de l'Afrique. Le territoire du Niger est très vaste, mais globalement il a deux saisons ; une saison sèche et une saison des pluies. Toutes deux menacent les récoltes presque chaque année, car l'une engendre la désertification et l'autre les inondations, ce qui entraîne des variations dans les besoins en aide alimentaire.

Dans ce chapitre, nous décrivons la structure physique de la chaîne du PAM et le processus de distribution en précisant ses acteurs. L'accent est mis sur les décisions prises par les logisticiens du PAM au Niger. Une représentation graphique du réseau logistique du PAM au Niger et son modèle d'optimisation sont présentés. Nous complétons le chapitre en abordant l'incertitude entourant les activités du PAM.

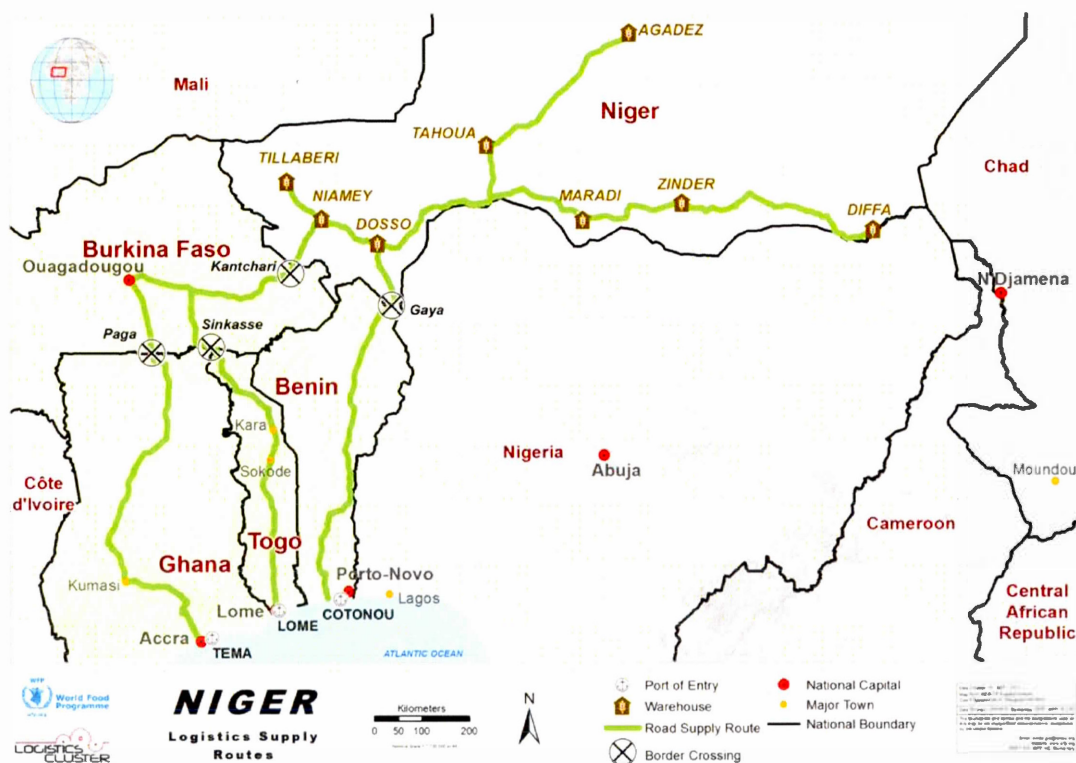


Figure 1.1 Routes pour l'approvisionnement au Niger (LCA, 2014)

1.1 Structure physique de la chaîne d'approvisionnement

La chaîne d'approvisionnement en denrées alimentaires est déclenchée par l'approbation d'un don destiné au Niger. Tel que présenté par la Figure 1.1, les denrées peuvent être acheminées à destination du port de Cotonou au Bénin, du port de Lomé au Togo ou du port de Tema au Ghana. À noter que le Niger est un pays enclavé ; les denrées doivent traverser un ou deux autres pays avant de franchir la frontière du Niger et donc traverser un ou deux postes douaniers selon le port d'arrivée. La distance de transport varie considérablement entre les corridors. Il est important de spécifier que, l'un des objectifs du PAM est d'être en mesure d'apporter de l'aide humanitaire jusqu'à ce que la population dans le besoin re-

prenne le contrôle et développe des mécanismes afin d'arriver à subvenir à ses besoins de façon autonome. C'est pourquoi le PAM contracte des compagnies locales (des transitaires) pour effectuer la manutention au port, le dédouanement et le transport jusqu'à ses entrepôts en sol nigérien. Le PAM possède des entrepôts dans neuf grandes villes au Niger, où les vivres sont acheminés pour ensuite être distribués aux bénéficiaires par le biais de différents programmes.

1.2 Processus de distribution

Le processus décisionnel décrit dans cette section représente une problématique du PAM au Niger et il s'insère dans le processus de la chaîne d'approvisionnement schématisé dans la Figure 1.2. Le processus à l'étude débute lorsqu'une donation est confirmée à l'attention du Niger par le bureau chef du PAM situé à Rome. Les responsables de la logistique au Niger doivent dès lors prendre une première décision; ils doivent sélectionner le(s) corridor(s) pour la distribution de la nourriture. Ils doivent sélectionner le port d'arrivée pour les denrées (Cotonou, Lomé ou Tema). Ils pourraient aussi décider de diviser le don entre deux ou trois de ces ports. En fonction de la quantité du don confirmé, ils doivent prendre une deuxième décision; ils doivent faire le choix d'une compagnie de transitaires (*forwarder*, en anglais) parmi celles disponibles au(x) port(s) sélectionné(s). Un contrat est attribué, chaque année, à un seul transitaire par port pour les dons de 2 500 tonnes métriques (TM) et moins; des appels d'offres doivent être faits pour les dons de tailles supérieures. Le transitaire sélectionné pour un don devient responsable de le récupérer à l'arrivée du navire, d'en assurer le dédouanement et de le transporter par camions jusqu'aux entrepôts du PAM au Niger (la destination finale des denrées pour le problème étudié). Ce sont les chauffeurs des camions et les responsables de la compagnie de transitaires qui font le choix de la route à prendre pour se rendre au Niger. La troisième forme de décision qui

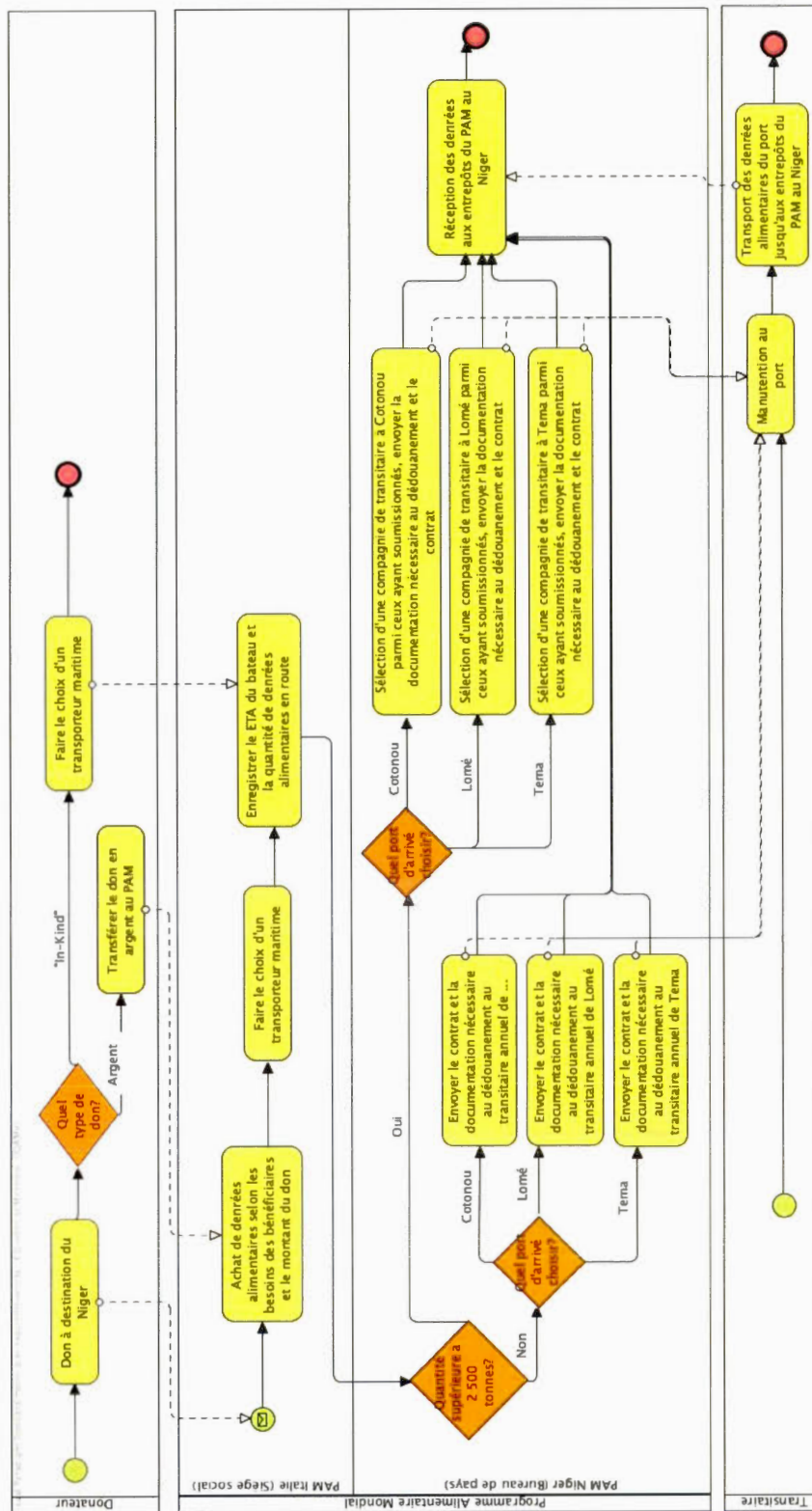


Figure 1.2 Processus de la chaîne d'approvisionnement du PAM

doit être prise est celle concernant le ratio de la quantité du don à transporter sur chacune des routes disponibles. Ces décisions sont d'ordre tactique et gagneraient à être réévaluées, car elles sont actuellement prises par les responsables du PAM au Niger sans support technologique. En plus de décrire la séquence des activités et des décisions du PAM au Niger, la Figure 1.2 décrit les activités faites par les autres acteurs du processus d'approvisionnement ; les donateurs et le siège social du PAM en Italie (en amont) ainsi que les transitaires avec qui le PAM basé au Niger transige.

Les denrées sont acheminées aux bénéficiaires selon un processus qui est considéré hors de la portée de ce projet. Ainsi, contrairement à la plupart des articles qui sont présentés dans notre revue de littérature, la demande des bénéficiaires ne sera pas considérée comme un paramètre aléatoire dans le cadre du présent projet, car les quantités requises à chaque entrepôt sont des paramètres connus et fixés à l'avance par l'équipe du PAM. Ces quantités à livrer aux entrepôts sont déterminées selon un processus complexe ciblant différentes catégories de bénéficiaires et prenant en compte l'apport de nombreux programmes d'interventions visant à renforcer la résilience des communautés. Malheureusement, le PAM au Niger reçoit rarement suffisamment de dons pour répondre complètement aux besoins de la population, ce sont donc les bénéficiaires et les communautés jugées les plus vulnérables qui sont desservies en priorité lorsqu'un don est confirmé. À titre d'exemple, en 2016, le PAM a reçu des dons ne finançant que 25% du budget estimé nécessaire pour répondre à la demande totale (WFP, 2016).

1.3 Modèle d'optimisation

Dans la discipline de la recherche opérationnelle, le problème à l'étude est défini comme un problème de synthèse de réseau. Ce problème inclut aussi la résolution

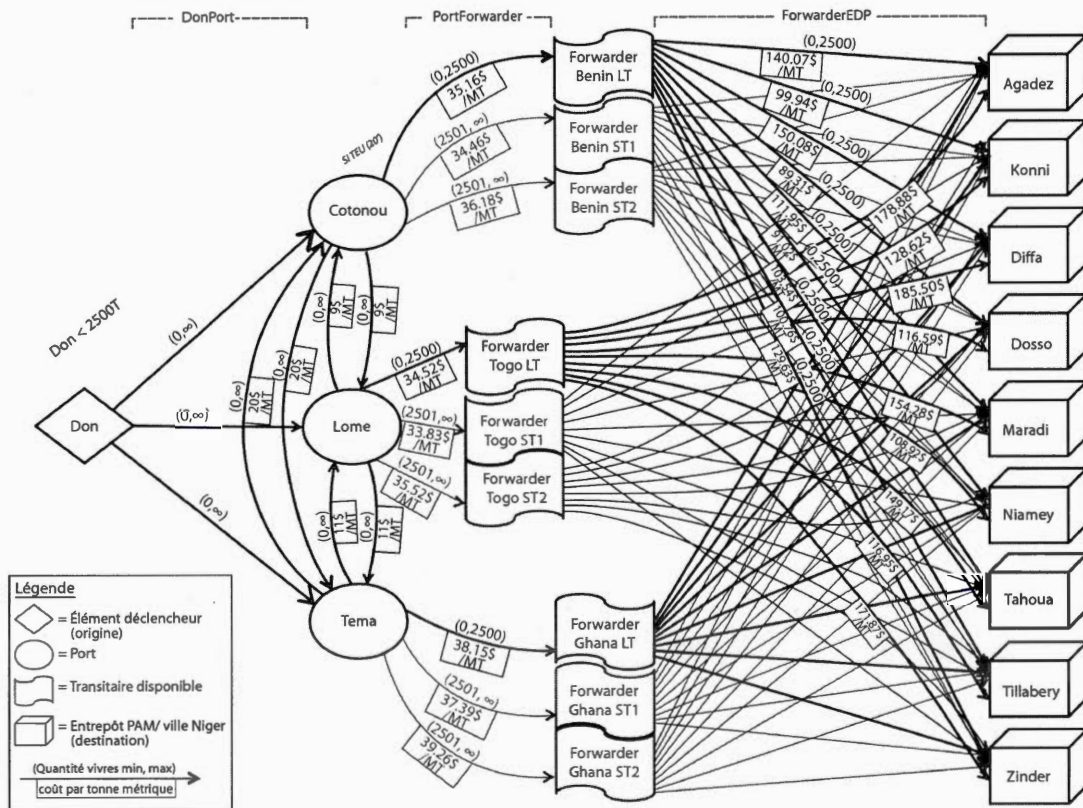


Figure 1.3 Réseau d'approvisionnement du PAM au Niger (exemple pour l'optimisation des coûts lors d'un don de moins de 2500 TM)

de type «flow dans un réseau» (*Network flow problem*, en anglais), voir l'exemple de la Figure 1.3. Le réseau développé est constitué de nœuds et d'arcs et il présente un ensemble de décisions devant être prises pour répondre au problème. Il comporte quatre types de nœud : le don, les ports, les transitaires et les entrepôts au Niger. Ces nœuds sont reliés par des arcs. Sur la partie supérieure des arcs, on retrouve les quantités minimales et maximales pouvant être acheminées et sous chacun de ces arcs, la valeur du paramètre associée au déplacement des vivres pour cette section du réseau (les coûts dans l'exemple de la Figure 1.3). Les arcs **DonPorts** représentent le transport maritime entre l'origine du don et son premier port d'arrivée en Afrique de l'ouest. Les arcs **Interports** donne la possibilité au don d'être séparé en deux destinations. Les arcs **PortForwarder** reflètent la manutention ; la transition des vivres entre le navire et les camions. Les arcs **ForwarderEDP** viennent simplifier la route que les camions empruntent et le ou les postes douaniers qu'ils doivent traverser jusqu'à leur destination finale, soit les entrepôts du PAM situés dans les grandes villes du Niger.

Un modèle d'optimisation a été développé pour traduire la problématique du PAM au Niger. Ce modèle, au stade de prototype, n'est pas encore en fonction. Les objectifs poursuivis, les variables de décision et les contraintes ont été définis. Le PAM poursuit un double objectif pour la distribution de l'aide alimentaire ; il cherche à minimiser à la fois les délais pour la réception des denrées aux entrepôts, mais aussi les coûts totaux engendrés par le transport en sol africain. Le niveau de priorité entre les deux objectifs varie en fonction de l'urgence des besoins. Le modèle a des variables de décision pour déterminer la quantité de vivres qui emprunte chacun des arcs du réseau (variables de flot). Aussi, des variables de décision de type binaires peuvent être associées aux nœuds centraux, soit ceux des ports et des transitaires. Ces variables sont binaires, car des décisions doivent être prises pour déterminer quel port et quel transitaire seront sélectionnés. Le

modèle a aussi des variables de décisions binaires pour permettre la sélection des arcs. Ce type de variables permet d'imposer au modèle, par exemple, l'utilisation d'un seul navire pour le transport maritime. Les limites minimales et maximales pour tous les arcs sont traduites sous forme de contraintes dans le modèle. La conservation du flux de denrées transportées à travers le réseau, la possibilité d'utiliser un arc uniquement si son noeud émetteur a été sélectionné et la non-négativité des variables sont aussi des contraintes essentielles à la modélisation du problème.

Tel qu'il l'est fait par les logisticiens du PAM, nous faisons l'hypothèse que les prix pour l'acheminement et la manutention des vivres sont les mêmes pour tous les types de denrées. Cette agrégation est raisonnable étant donné que les prix varient que très peu d'une denrée à une autre. Les transitaires font très peu de distinction entre les types de denrées transportées à l'intérieur des contrats de transport.

1.4 Sources d'incertitude et risques

À l'aide de logiciel adapté à la programmation mathématique, le modèle de programmation linéaire présenté à la section précédente peut être résolu de façon efficace dans sa version déterministe; sans tenir compte des facteurs d'incertitudes. Dans le cadre de ce projet, on se concentre sur le besoin d'intégrer un grand nombre de phénomènes aléatoires au raisonnement décisionnel. Ces phénomènes peuvent faire varier les coûts et les délais pour faire parvenir le don au Niger ou même engendrer des pertes de denrées.

La situation géographique et les conditions climatiques difficiles du Niger viennent complexifier la distribution de l'aide alimentaire. Selon le(s) corridor(s) sélectionné(s), les denrées traversent une ou des zone(s) de port(s) et un ou des poste(s)

douanier(s). Les niveaux de congestion des ports et des postes douaniers sont variables et difficilement prévisibles. Nathan Associates Inc. (2013) parlent d'inefficacité des opérations dans les ports de l'Afrique de l'Ouest et précisent que la congestion impacte fortement les délais et les coûts de la marchandise en transit, surtout aux ports de Cotonou et de Lomé. En sol africain, le transport se fait sur des routes pavées ou non pavées, qui sont parfois considérées impraticables en raison des conditions climatiques (LCA, 2014). Les pays limitrophes comportent aussi des risques qui touchent directement les activités du PAM et qui viennent s'ajouter à ceux présents en sol nigérien. Le Nigéria, situé à sa frontière sud-est est l'un des territoires où réside le groupe terroriste Boko Haram. La présence et les actions de ce groupe génèrent des mouvements de populations en sol nigérien et un risque particulièrement élevé d'incursion près de la ville de Diffa en bordure du lac Tchad (un lac partagé entre le Niger, le Nigeria et le Tchad). Il est notamment difficile pour le PAM d'avoir des installations sécuritaires dans cette ville. Les réfugiés en sol nigérien proviennent principalement du Nigéria et du Mali (UNOCHA, 2016). En outre, les transitaires contactés par le PAM pour acheminer les denrées à partir d'un port de l'Afrique de l'Ouest jusqu'aux entrepôts du Niger constituent un facteur qui apporte son important lot d'incertitude en ce qui a trait aux délais de livraison. Les différentes sources d'incertitude affectant les activités du PAM au Niger seront vues plus en profondeur dans les prochains chapitres de ce mémoire.

CHAPITRE II

PROBLÉMATIQUE, QUESTION DE RECHERCHE ET OBJECTIFS

Étant donné que notre projet de recherche se place au croisement de plusieurs disciplines de recherche, c'est-à-dire la recherche opérationnelle, la gestion de la chaîne d'approvisionnement, la gestion des risques et la logistique humanitaire, il est important de clarifier la terminologie liée à l'analyse des risques. Ce chapitre débute donc par la définition des termes et des concepts utilisés et se poursuit avec la présentation de la problématique, de la question de recherche et des objectifs du projet.

2.1 Définitions des termes et des concepts

Nous voulons souligner que l'utilisation des termes risque et incertitude ainsi que leur définition n'est pas universellement acceptée (Klibi *et al.*, 2010; Tang et Musa, 2011; Heckmann *et al.*, 2015). Jüttner *et al.* (2003) ont fait remarquer que le terme risque peut être utilisé pour référer à deux choses bien distinctes : la source du risque ou la conséquence du risque. Heckmann *et al.* (2015) ont par ailleurs comparé les études qui ont classifié les risques de la chaîne d'approvisionnement selon les sources des risques avec celles qui les ont classifiés selon leurs impacts. Au sein des articles qu'ils ont considérés, les techniques sont partagées ; le même nombre de recherches ont été menées sous les deux cas d'interprétation. Ces mêmes

auteurs précisent que très peu de recherches ont analysé à la fois les sources et les conséquences des risques. Dans le cadre de ce projet, nous avons choisi de nous en tenir aux définitions présentées dans la norme ISO 31000 (Lark, 2015, p. 16) sur le management du risque, où le risque est défini comme «l'effet de l'incertitude sur l'atteinte des objectifs» et l'incertitude comme «l'état, même partiel, de défaut d'information concernant la compréhension ou la connaissance d'un événement de ses conséquences ou de sa vraisemblance». Nous souhaitons préciser que le risque a toujours une certaine probabilité d'occurrence et un impact quantifiable sur les activités de l'organisation. Il est aussi important de spécifier que les sources d'incertitude peuvent évoluer et affecter positivement ou négativement les objectifs d'une organisation. Ainsi, pour la suite de ce mémoire, nous utiliserons les termes sources d'incertitude et risques (conséquences) pour bien distinguer les concepts. De plus, le terme paramètre sera utilisé pour parler des métriques non contrôlables présentes dans le contexte informationnel du problème étudié, ces paramètres peuvent être considérés certains ou incertains. Tandis que le terme variable sera utilisé pour parler des décisions à prendre (Govindan *et al.*, 2017).

Dans un contexte informationnel où un paramètre est connu, ou présumé connu, on dit qu'il est un paramètre déterministe (Heckmann *et al.*, 2015). Dans le cas où un paramètre est considéré comme étant incertain (variant de façon aléatoire), les chercheurs parlent plutôt de paramètres stochastiques (Govindan *et al.*, 2017). La Figure 2.1 présente des exemples pour clarifier les termes et les concepts qui sont abordés dans ce mémoire. Il est possible d'analyser les conséquences des risques ; l'effet qu'ont les sources d'incertitude sur les paramètres stochastiques considérés ou sur la performance d'un système dans son ensemble. La manifestation d'une source d'incertitude (p.ex., groupes terroristes actifs, inondations, etc.) peut entraîner de la variation dans la valeur d'un ou de plusieurs paramètres stochastiques (demande, coûts, délais, pertes, etc.). Dans le cadre de notre recherche, nous ana-

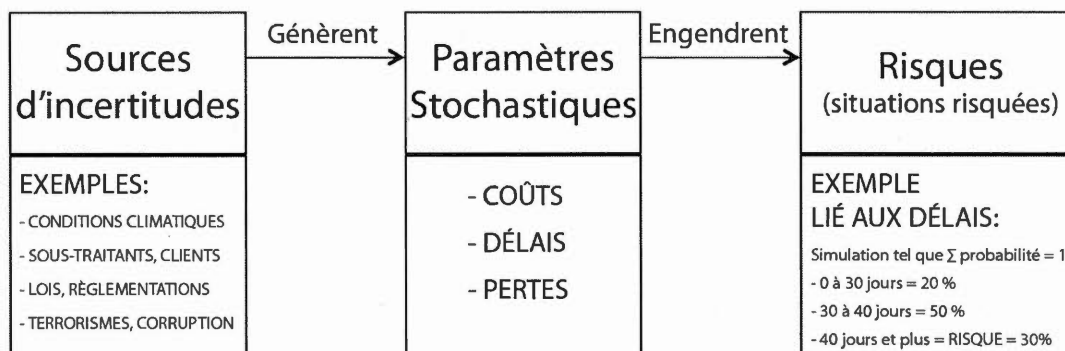


Figure 2.1 Application des termes et concepts

lysons trois paramètres stochastiques, soit les délais, les coûts et les pertes, car ils sont directement liés aux objectifs du PAM. Les variations dans les valeurs qu'ils prennent peuvent engendrer des risques pour l'organisation. La Figure 2.1 donne un exemple où les sources d'incertitude engendrent un impact sur les délais. Dans cet exemple, le risque, en termes de situation risquée, est calculé par la probabilité que le délai soit supérieur à 40 jours. La Figure montre bien le contexte dans lequel se présente le problème et la manière dont il est considéré par le présent projet de recherche.

2.2 Problématique

Les activités du PAM au Niger se déroulent dans un contexte marqué par la présence de nombreuses sources d'incertitude importantes. Étant donné leurs caractères aléatoires, il est très difficile pour le PAM d'en tenir compte lorsque vient le temps de choisir le corridor approprié pour le transport. L'organisation souhaite minimiser les coûts et les délais totaux associés au transport des denrées alimentaires, mais la valeur de ceux-ci, calculée en dollars et en jours, varie considérablement sur un même corridor. De plus, les logisticiens constatent parfois des pertes, en termes de quantité, entre l'envoi et la réception d'un don. Ces trois

paramètres stochastiques directement liés aux objectifs globaux du PAM sont difficiles à estimer. De plus, les différentes sources d'incertitude n'impactent pas les paramètres stochastiques de la même façon. La distribution de probabilité et le niveau d'impact d'une source d'incertitude pourraient varier selon le port d'arrivée, la section du corridor, la saison, etc. Il est donc bénéfique d'approfondir la notion de risque dans le réseau logistique qui a été développé pour le cas du PAM (voir Figure 1.3) en analysant les sources d'incertitude qui lui sont associées afin d'arriver à évaluer les risques pour l'organisation. Nous avons voulu comprendre l'incertitude associée à chaque arc du réseau et développer un modèle capable de générer des scénarios traduisant le contexte particulier du PAM au Niger. Pour ce faire, des informations contextuelles ont été nécessaires (historique des données et avis d'experts). Ces informations sont utilisées, comme intrant, pour des modèles permettant d'estimer la valeur des paramètres stochastiques.

Nous croyons qu'en fonction de l'appétence aux risques de l'organisation, le caractère aléatoire d'un ensemble de sources d'incertitude pourrait entraîner un changement de corridor pour l'approvisionnement des denrées dans le but d'atténuer les risques encourus.

2.3 Question de recherche

La complexité de la problématique qui a été présentée nous a menés à la question de recherche suivante :

Comment générer des scénarios valides permettant de quantifier les paramètres stochastiques afin d'évaluer les risques affectant les corridors de transport et contribuer à une meilleure prise de décision par les responsables de la logistique du PAM au Niger ?

2.4 Objectifs

Pour parvenir à évaluer les risques et appuyer la prise de décision du PAM, nous avons identifié les objectifs suivants : 1) identifier les sources d'incertitude et les catégoriser, 2) déterminer les modèles pour estimer les paramètres stochastiques, et 3) générer les scénarios.

Nous avons aussi identifié deux contributions pratiques pour le PAM. La première est de pouvoir évaluer un plan d'approvisionnement dans un contexte d'incertitude ; de comparer puis confirmer ou infirmer les décisions généralement prises en fonction de certains critères. Ainsi les dirigeants pourront se servir de notre modèle pour revoir leur technique décisionnelle et améliorer leurs choix futurs. La deuxième contribution pratique est qu'une fois validé, le modèle permettant de générer des scénarios pourra servir à instancier un modèle d'optimisation à la fois multi-objectif et stochastique qui pourra être utilisé par le PAM lors de la prise de décision.

CHAPITRE III

REVUE DE LITTÉRATURE

D'après Leiras *et al.* (2014), plusieurs chercheurs ont concentré leurs efforts sur les phases de préparation et de réponse aux crises ou aux désastres naturels. Par contre, d'après ces mêmes chercheurs ainsi que plusieurs autres (Altay et Green, 2006; Kovács et Spens, 2007; Overstreet *et al.*, 2011; Natarajarathinam *et al.*, 2009; Kunz et Reiner, 2012; Galindo et Batta, 2013), la phase de redressement suite aux catastrophes continue de recevoir que très peu d'attention. Il est aussi précisé que l'impact des catastrophes à évolution lente nécessite généralement des projets de développement à long terme (Leiras *et al.*, 2014), ce qui contribue à la difficulté de mener des études à terme sur ce sujet.

Dans ce chapitre, nous présentons une revue de la littérature associée à la problématique évoquée au chapitre précédent. Nous couvrons d'abord la littérature portant sur les méthodes d'identification des sources d'incertitude, les façons de les catégoriser et les impacts qu'elles peuvent avoir sur la chaîne d'approvisionnement en contexte d'aide humanitaire. Nous abordons la littérature ciblant les approches d'évaluation des risques : les modèles prédictifs et la simulation. Nous complétons par les différentes techniques utilisées pour simuler les sources d'incertitude et leurs cas d'applications en contexte d'aide humanitaire.

3.1 Identification des sources d'incertitude

À travers les années, un grand nombre d'auteurs ont élaboré des méthodes pour déterminer la typologique des sources d'incertitude d'une entreprise (Miller, 1992; Simons, 1999; Tchankova, 2002; Slywotzky et Drzik, 2005; Dumitriu, 2006), leur but était d'arriver à élaborer une technique pour tracer le profil de risque complet des entreprises. À la lecture de ces articles, on remarque qu'avec le temps les sources d'incertitude évoluent et se multiplient. Les auteurs mentionnés sont passés d'une perspective en silo à une approche holistique de la gestion des risques (Gordon *et al.*, 2009). Ces méthodes émergent de la gestion intégrée des risques d'entreprise (*enterprise risk management*, en anglais).

Dans leur article sur la création de valeur lors de la conception de réseaux logistiques, Klibi *et al.* (2010) ont proposé une façon structurée de faire l'inventaire des sources d'incertitude¹ basée sur trois grandes catégories : les actifs endogènes, les partenaires de la chaîne d'approvisionnement et les facteurs géographiques exogènes. Les auteurs soulignent l'importance de ne pas seulement considérer les facteurs aléatoires courants (taux de change, volatilité des prix et variation dans la demande) et présentent l'intérêt de tenir compte des événements extrêmes tels que les catastrophes naturelles et les actes terroristes. Ghadge *et al.* (2013) proposent une approche méthodologique pour modéliser les risques de la chaîne d'approvisionnement. Selon leur modèle, l'identification des sources d'incertitude se fait en deux étapes, soit la taxonomie des sources d'incertitude (regroupement par catégorie) et l'analyse de la tendance des risques (l'évolution à travers le temps et l'étendue de leur probabilité). Les chercheurs favorisent une approche systémique en utilisant la classification par attributs de Burtonshaw-Gunn (2009). Les auteurs considèrent six attributs (processus, organisation, lieu, données, application

1. Ces auteurs utilisent le terme source de vulnérabilité et non pas source d'incertitudes.

et technologie) comme des sphères de changement.

Des articles présentant des revues de la littérature (Tang et Musa, 2011; Singhal *et al.*, 2011) ont examiné les recherches dans le domaine de la gestion des risques de la chaîne d'approvisionnement (GRCA) et les ont classées selon les sources d'incertitude qu'elles ont considérées. Tang et Musa (2011) les classent selon les flux de matériaux, les flux monétaires et les flux d'informations et Singhal *et al.* (2011) selon leurs caractéristiques, soit des caractéristiques opérationnelles, de marché, de stratégie d'affaires, des produits et autres. Dans les deux cas, on remarque que peu d'articles qui ont analysé plus d'une catégorie de source d'incertitude à la fois. D'ailleurs, Rao et Goldsby (2009) affirment que très peu d'attention est accordée à l'identification des risques de la chaîne d'approvisionnement au sein de la littérature et Ghadge *et al.* (2013) que d'importantes recherches dans le domaine de la GRCA étudient les risques en dyade ou qu'un seul à la fois.

En contexte d'aide humanitaire, Coppola (2015) est le seul auteur que nous avons identifié qui propose des techniques d'identification des sources d'incertitude dans son livre sur la gestion des désastres où il se concentre sur celles ayant un potentiel de catastrophe. Il parle de sources géologiques, météorologiques, hydrologiques, biologiques, technologiques, politiques et sociales et les regroupe sous trois types : naturels, technologique et intentionnel. L'auteur précise que différentes méthodes peuvent être utilisées pour identifier les risques et il mentionne qu'elles devraient être combinées et qu'elles doivent être exhaustives pour être effectives.

3.2 Catégorisation et impacts des sources d'incertitude

Dans cette deuxième section, les études qui ont soit élaboré des catégories de sources d'incertitude ou analysé leurs impacts sur la chaîne d'approvisionnement en contexte d'aide humanitaire sont présentés.

3.2.1 La catégorisation des sources d'incertitude

Klibi *et al.* (2010) proposent de catégoriser les sources d'incertitude d'un réseau selon trois types qui ne peuvent être quantifiés de la même façon étant donné la qualité de l'information qui peut être récoltée à leur sujet. Les auteurs utilisent les termes *randomness*, *hazard* et *deep uncertainty* pour les distinguer. Dans le premier cas, beaucoup de données sont accessibles et les distributions de probabilités des paramètres aléatoires analysés sont connues. Les sources d'incertitude de cette catégorie sont récurrentes. Le deuxième type d'incertitude énoncée représente les crises et les événements extrêmes. Dans ces cas, les probabilités d'occurrences sont très faibles, mais les impacts sont majeurs pour l'organisation. Cette seconde catégorie doit être traitée différemment de la première et elle ne doit pas être négligée surtout lorsque l'on analyse une organisation sujette à des catastrophes. D'après ces auteurs, il n'y a aucun modèle de réseau de la chaîne d'approvisionnement qui prend en compte cette forme d'incertitude. Le troisième type d'incertitude qu'ils ont défini est celui où l'information est manquante, car elle est soit non disponible ou impossible à collecter. Il y a manque d'informations à la fois pour quantifier les probabilités d'occurrence et les impacts. Klibi et Martel (2012) donnent des exemples d'événements liés au terrorisme et à l'instabilité politique pour cette catégorie.

Tang (2006) a présenté deux grandes catégories qui ont par la suite été utilisées par plusieurs auteurs (Sawik, 2011; Sadghiani *et al.*, 2015; Baghalian *et al.*, 2013) : les risques opérationnels et les risques de ruptures. Ainsi, Sawik (2011), Govindan *et al.* (2017) et Kleindorfer et Saad (2005) regroupent les phénomènes aléatoires dont la probabilité d'occurrence est faible et dont l'impact sur la chaîne d'approvisionnement est très grand en une seule catégorie qu'ils appellent les risques de ruptures (catégorie appelée événement catastrophique par Knemeyer *et al.*, 2009).

Les ruptures dans la chaîne peuvent être causées par des phénomènes naturels ou être causées par l'homme. Les auteurs y incluent un grand nombre de sources d'incertitude tels que les inondations, les tremblements de terre, les ouragans, les attaques terroristes, les crises économiques, les grèves et même la défaillance des équipements et des systèmes dans le cas de Kleindorfer et Saad (2005). Dans leur revue de littérature, Govindan *et al.* (2017) présentent la notion d'environnements décisionnels incertains qu'ils catégorisent sous trois groupes : 1) l'environnement décisionnel où le gestionnaire connaît la distribution de probabilité des paramètres stochastiques, 2) l'environnement décisionnel où le gestionnaire n'a aucune information sur la distribution de probabilité des paramètres stochastiques et 3) l'environnement décisionnel dit *Fuzzy* où il y a soit ambiguïté dans les options de décision ou les limites du domaine d'intérêt sont vagues. Environ 73 % des études qu'ils ont recensées se retrouvent dans l'environnement informationnel du premier groupe et que 19 % ont traité les risques de ruptures.

3.2.2 L'impact des sources d'incertitude sur la chaîne d'approvisionnement

Peu d'études semblent avoir vraiment cherché à comprendre l'impact des sources d'incertitude sur la chaîne d'approvisionnement. Knemeyer *et al.* (2009) ont proposé un processus de planification proactif pour répondre aux risques de catastrophes. Ils mentionnent le besoin d'identifier les sources d'incertitude, leur probabilité d'occurrence et leur conséquence en se basant sur des données ou sur le jugement d'expert. Le but étant de pouvoir évaluer les pertes causées par les événements catastrophiques. Les auteurs calculent l'impact de ceux-ci sur les ressources affectées (ressources humaines et financières, produits, infrastructures et informations) et proposent des mesures pour atténuer la vulnérabilité des organisations. Ils précisent que les pertes peuvent excéder la valeur des bâtiments affectés.

Les études sont d'autant plus rares en contexte d'aide humanitaire. Nous avons identifié trois études qui ont eu comme objectif d'identifier et de quantifier les facteurs attribuables aux coûts élevés des transports dans les corridors d'approvisionnement africain soit Nathan Associates Inc. (2013), Rancourt *et al.* (2014) et Teravaninthorn et Raballand (2009). D'abord, Nathan Associates Inc. (2013) a publié un rapport pour la Banque mondiale portant sur cinq importants corridors logistiques de l'Afrique de l'Ouest, dont celui allant du port de Cotonou au Bénin jusqu'à Niamey, la capitale du Niger. Ils ont séparé les coûts en deux grandes catégories, soit les coûts financiers des services logistiques et les coûts cachés, ces derniers étant définis comme étant les coûts économiques attribuables aux délais et aux sources d'incertitude. Cette seconde catégorie est particulièrement importante en Afrique de l'Ouest. Selon leur étude et les cas qu'ils ont analysés, du corridor «Cotonou-Niamey», elle représente entre 14% et 50% des coûts totaux. Les résultats de leur étude démontrent que la congestion au port de Cotonou est de loin le problème principal de la chaîne logistique « Cotonou-Niamey » et la première source de coûts et de délais (Nathan Associates Inc., 2013, p. 79).

Pour leur part, Rancourt *et al.* (2014) ont analysé le cas du PAM basé en Éthiopie, un pays situé à l'est de l'Afrique où l'incertitudes est particulièrement élevée dans la région de la Somalie. Le PAM basé en Éthiopie apporte un soutien en aide alimentaire sur un horizon à long terme et les chercheurs ont analysé les contrats régissant les activités entre le PAM et ses transporteurs. Ils ont regroupé les données qu'ils ont recueillies en trois grandes catégories de variables indépendantes, soit celles représentant les coûts liés aux opérations des transporteurs (distance, qualité des routes, perception du risque, etc.), celles incluant les mesures en lien avec la structure du marché du transport et celles regroupant les facteurs socio-économiques. Avec ces variables, les chercheurs ont procédé à des

régressions linéaires multiples pour expliquer la variation de la variable dépendante : le coût des transports. Ils ont remarqué que les contrats dans les zones considérées risquées ont des tarifs plus élevés. Les compagnies de transports se protègent ainsi de certains risques. De plus, contrairement à ce qui est le cas en Amérique du Nord (Chainalytics, 2014; Rodrigue *et al.*, 2009), les résultats qu'ils ont observés démontrent que la distance n'explique qu'une petite partie de la variabilité des coûts de transport, soit 9% sur les corridors internationaux et 27% sur les corridors domestiques. Cette observation est très étonnante sachant que souvent le prix du transport augmente presque proportionnellement avec la distance. Dans le cas de l'Amérique du Nord, un prix fixe et un prix variable constant par kilomètre expliquent entre 80 et 85 pour cent de la variabilité des coûts (Chainalytics, 2014 cité par Rancourt *et al.*, 2014). C'est une constatation importante pour notre étude puisque la majorité des trois corridors que l'on étudie sont internationaux et leurs distances varient grandement. Les résultats de Rancourt *et al.* (2014) ont montré que ce sont l'état des routes (pavée ou non pavée) et l'intensité de la concurrence qui expliquent le mieux les coûts de transport; des variables ayant aussi été identifiées comme étant des facteurs clés par l'étude de la Banque mondiale (Teravaninthorn et Raballand, 2009).

Jahre *et al.* (2016) ont considéré un problème de conception de réseau dans le cas du Haut Commissariat des Nations unies pour les réfugiés (HCR), une division des Nations Unies. Ils ont développé un modèle permettant d'intégrer des facteurs contextuels (politiques et sécuritaires) lors de la sélection des entrepôts pour le prépositionnement des denrées en combinant les besoins et les difficultés en situation de réponse d'urgence avec ceux des opérations en continu. Ils ont voulu quantifier l'impact d'un réseau fusionné sur les coûts et les temps de distribution de la marchandise. L'effet des sources d'incertitude qu'ils ont identifiées a aussi été analysé sur les coûts et les délais. Conjointement avec les responsables du

HCR et en fonction des données disponibles, les auteurs ont décidé d'attribuer des poids (des pourcentages associés aux pays considérés) sur certaines sources d'incertitude à caractères politiques et de sécurités parmi les suivantes : niveau de sécurité du pays, risque de pillage, niveau de difficulté à trouver de la main-d'œuvre locale, qualité de la relation avec le gouvernement local et la possibilité de coopération avec les autres organismes humanitaires présents dans le pays. Selon les auteurs, d'autres chercheurs ont discuté des facteurs politiques et sécuritaires, mais personne ne les a quantifiés et modélisés avant eux.

3.3 L'évaluation des risques

Dans des cas où le système étudié est complexe, et qu'il y a de l'interaction entre les sources d'incertitude, la prise de décision et son évaluation deviennent donc un défi (Comes *et al.*, 2011; Bozorgi-Amiri *et al.*, 2013). De nombreuses méthodes sont utilisées pour quantifier les sources d'incertitude et permettre d'évaluer les risques. Même si des mesures telles que la variance et l'écart-type sont reconnus comme problématiques, car elles considèrent également la déviation positive par rapport à la valeur la plus probable, elles sont utilisées dans de nombreuses recherches pour mesurer le risque (Heckmann *et al.*, 2015). Pour mesurer le risque de détérioration (*downside-risk*) certains auteurs, comme Poojari *et al.* (2008) et Sawik (2011), utilisent la valeur à risque (*value at risk*) ou la valeur conditionnelle à risque (*conditional value at risk*) dans des modèles d'optimisation où la qualité des données recueillies le permet (Sarykalin *et al.*, 2008). Pour évaluer les dimensions stochastiques d'un problème décisionnel ou d'un système complexe, des modèles prédictifs ou des modèles de simulation peuvent aussi être développés. Les différentes méthodes utilisées et leurs domaines d'applications seront présentés dans les sections qui suivent. Les rares cas propres à l'aide humanitaire sont décrits.

3.3.1 Les modèles prédictifs

Le forage des données (*data mining*) est essentiel à la construction de modèles prédictifs. Plusieurs techniques pour explorer les données (p. ex., l'analyse en composantes principales) et bâtir des modèles (p. ex., régression linéaire et analyse discriminante) existent depuis longtemps en statistique (Shmueli *et al.*, 2017). Le forage de données peut être fait pour répondre à deux types de problèmes en apprentissage statistique. Les différents auteurs du domaine s'entendent pour parler d'apprentissage non supervisé (*unsupervised learning*) lorsque l'utilisateur veut apprendre des tendances (*patterns*) et des relations dans les données, sans spécifier d'intentions. Alors qu'en apprentissage supervisé (*supervised learning*), l'utilisateur spécifie la variable qu'il veut prédire (le paramètre stochastique étudié). Cette variable à prédire peut être soit qualitative (catégorique) ou quantitative (numérique). On parle de modèles de classification lorsque l'on veut prédire dans quelle catégorie la variable se trouve et de modèles de régression lorsque l'on veut prédire sa valeur numérique. Les livres de Hastie *et al.* (2009), de James *et al.* (2013) et de Shmueli *et al.* (2017) sont d'excellentes références sur le sujet.

Les domaines d'applications pour les modèles prédictifs sont vastes : l'aéronautique (Rebollo et Balakrishnan, 2014; Wang *et al.*, 2018), la médecine (Zupancic *et al.*, 2003; Hilbert *et al.*, 2014), l'éducation (Cortez et Silva, 2008; Miguéis *et al.*, 2018; Masci *et al.*, 2018), la finance (Salehi et Davoudi Pour, 2016; Fitzpatrick et Mues, 2016), etc. Selon le type de données collectées, il est possible que la relation entre les variables explicatives (les intrants du modèle) et la variable à prédire (l'ex-trant du modèle) soit linéaire. C'est le cas pour l'étude de Zupancic *et al.* (2003). Les auteurs affirment avoir développé un modèle de régression linéaire ayant un excellent pouvoir prédictif des coûts accessoires reliés aux naissances prématurés. Toujours dans le domaine de la médecine, Hilbert *et al.* (2014) ont créé un modèle

d'arbre de décision pour mieux gérer les risques de réadmission des patients dans les hôpitaux. L'arbre de décision leur a permis de créer des règles décisionnelles simples, compréhensibles et transparentes pour les cliniciens. Les modèles d'arbres de régression et de régressions linéaires permettent de facilement interpréter les résultats obtenus.

Lorsque la précision de la prédiction est plus importante que l'interprétabilité du modèle, les chercheurs peuvent opter pour des modèles d'apprentissages-machine (*machine learning*) plus complexes. Pour améliorer la performance d'un modèle unique, plusieurs modèles peuvent être combinés pour prédire. Le modèle de forêt d'arbres décisionnels (*random forest*) est l'un de ces modèles, où plusieurs arbres de décisions sont combinés. Entre autres, les chercheurs Cafri *et al.* (2018) utilisent le modèle de forêt aléatoire pour prédire les risques de mortalité ou de défaillance de la prothèse suite à une opération de remplacement total du genou. Ils précisent que la technique permet d'atténuer le surapprentissage (*overfitting*) des données et qu'elle est d'autant plus intéressante lorsque les données comportent des événements rares. D'autres chercheurs (McCluskey *et al.*, 2014; Schiltz *et al.*, 2018; Elith *et al.*, 2008) favorisent plutôt les arbres de régressions boostés (*boosted regression trees*). Cette technique permet aussi de combiner plusieurs arbres de décisions, mais les arbres sont bâtis séquentiellement. Les algorithmes de ce type apprennent lentement sur les données. Chaque arbre est bâti en utilisant les informations issues des arbres précédents. La façon de quantifier l'insuffisance du modèle (*lack of fit*) et le réglage des paramètres entre les itérations peuvent différer d'un algorithme à l'autre (Elith *et al.*, 2008). D'autres modèles complexes existent et sont utilisés pour prédire la valeur des paramètres stochastiques étudiés. Par exemple, Oneto *et al.* (2018) utilisent les réseaux de neurones (*neural networks*) pour prédire les retards des trains. Dans leur article portant sur la prédiction de la violence en milieu carcéral, Bačák et Kennedy (2018) incitent sur l'importance

de comparer plusieurs méthodes prédictives. L'objectif étant de trouver le modèle qui offre la meilleure performance prédictive dans le contexte étudié. Entre autres, Chen *et al.* (2018), Chen *et al.* (2017) et Rebollo et Balakrishnan (2014) ont comparé plusieurs modèles. Ils affirment tous que leur modèle de forêt d'arbres décisionnels surpasse les autres modèles prédictifs qu'ils ont considérés.

Pour évaluer la performance prédictive des modèles, les techniques et les métriques utilisées diffèrent. En général, l'objectif est de construire un modèle qui pourra bien prédire sur un nouvel échantillon. Varian (2014) indique que c'est une pratique conventionnelle en apprentissage-machine de diviser les données pour en faire différents ensembles (entraînement, test et validation). L'auteur indique que l'ensemble de validation et de test sont souvent combinés pour ne former qu'un seul ensemble. Wang *et al.* (2018), Smith *et al.* (2013), Zupancic *et al.* (2003) et Hilbert *et al.* (2014) ont séparé les données qu'ils ont utilisées en différents ensembles. Par contre, leurs choix pour la partition des données diffèrent. Wang *et al.* (2018) ont créé un ensemble d'entraînement qui représente 64% de leurs données, un ensemble de validation de 16% et un ensemble de test de 20%. Smith *et al.* (2013), Zupancic *et al.* (2003) et Hilbert *et al.* (2014) ont utilisé seulement que deux ensembles. Pour Smith *et al.* (2013), l'ensemble d'entraînement représente 90% des données et l'ensemble de test est de 10%. Zupancic *et al.* (2003) ont séparé également leurs données et Hilbert *et al.* (2014) ont utilisé leurs données datant de l'année 2010 pour l'entraînement et leurs données datant de 2011 comme ensemble de test. La méthode de validation croisée (*cross-validation*) peut aussi être utilisée pour valider la performance prédictive d'un modèle. Rebollo et Balakrishnan (2014) et Chen *et al.* (2017) l'ont fait. Ils ont partitionné leurs données en dix sous-ensemble d'entraînement et validé leurs résultats en dix étapes. Ainsi, toutes les données sont utilisées pour l'entraînement sans toutefois surapprendre. Cafri *et al.* (2018) ont pour leur part utilisé la méthode d'évaluation intégrée à l'algorithme de forêt

aléatoire. Comme chaque arbre est construit à partir d'un sous ensemble de données, les observations qui n'ont pas été utilisées dans la construction de l'arbre (*out of bag*) sont utilisées pour l'évaluation.

Pour ce qui est des métriques utilisées pour évaluer la performance des modèles, elles varient aussi d'un projet à l'autre. L'erreur moyenne au carré (*mean squared error*, MSE) ou la racine de l'erreur moyenne au carré (*root mean squared error*, RMSE) sont des mesures d'erreur répandues (Wang *et al.*, 2018; Baćak et Kennedy, 2018; Cafri *et al.*, 2018). Wang *et al.* (2018) considèrent aussi la moyenne des erreurs en valeur absolue (*mean absolute error*, MAE), une mesure qui pénalise moins les erreurs causées par les valeurs aberrantes. Smith *et al.* (2013) et Zupancic *et al.* (2003) utilisent plutôt le coefficient de corrélation R^2 pour évaluer la performance de leurs modèles.

3.3.2 Les types de simulation

Galindo et Batta (2013) font remarquer que dans plusieurs articles les techniques de simulation et de statistique sont combinées. Dans bien des cas, les statistiques sont utilisées pour modéliser l'incertitude et la simulation pour des fins de validation. Il existe quelques techniques pour simuler les événements aléatoires présents dans un système, en fonction de la problématique considérée, les chercheurs utilisent l'une ou l'autre des techniques suivantes : la simulation de type multi-agents (*agent based simulation*), la simulation à événements discrets (*discrete event simulation*) ou la simulation de type Monte Carlo.

La simulation de type multi-agents est utilisée pour modéliser les comportements des agents, des entités autonomes considérées intelligentes. Dépendamment des spécifications incluses par le concepteur du modèle, un agent peut réagir, raisonner, mémoriser et même prendre des initiatives selon ses connaissances et ses

expériences. Ce type de simulation est donc dynamique et adaptative (Chan *et al.*, 2010). La simulation à événements discrets est aussi dynamique, mais ce sont des événements distincts qui sont simulés et qui viennent modifier l'état du système à travers le temps. Or, ces deux premières méthodes de simulations s'appliquent bien pour les cas où le chercheur souhaite représenter les changements dans l'état d'un système à travers le temps. Par contre, la notion d'interaction diffère entre les deux techniques. Dans le cas de la simulation multi agents ce sont les agents qui interagissent tandis que dans celui de la simulation à événements discrets ce sont les processus qui peuvent interagir (Chan *et al.*, 2010).

L'approche de modélisation des risques proposée par Ghadge *et al.* (2013) indique que pour un niveau de probabilité donné, le modèle considère l'impact des sources d'incertitude comme étant soit de niveau élevé ou faible. Dans le cas où l'impact est élevé, l'exposition aux risques ne peut être atténuée et il y a propagation du risque sur le système. Sous cette base, leur étude présente et compare deux approches de modélisation qu'ils implémentent en parallèle. Une modélisation statistique prévue pour prédire le comportement des risques et une modélisation par la simulation à événements discrets pour évaluer leurs impacts pour une probabilité donnée; tous deux calculent l'impact total cumulé sur les coûts et les délais. Le modèle de simulation permet de démontrer que dans certains cas, un ensemble d'événements peut venir perturber un réseau en entraînant une augmentation des délais sans toutefois affecter les coûts.

La simulation de type Monte Carlo qui requiert la génération de scénarios représentatifs de l'état possible d'un système à un certain moment dans le temps. Elle est qualifiée de statique par certains (Sobolev *et al.*, 2011). Les scénarios peuvent être soit générés par le biais des distributions propres aux paramètres stochastiques ou par la combinaison d'informations objectives et subjectives lorsque la problématique le nécessite (Hansen *et al.*, 2016; Klibi et Martel, 2012; Knemeyer

et al., 2009; Poojari *et al.*, 2008; Banks, 2005). Schwartz (1992) explique que la planification de scénario nous sert à interrompre notre incrédulité envers les possibilités du futur, elle nous permet de penser qu'une de ces issues pourrait être possible et nous prépare donc pour l'imprévu. Dans leur article, Klibi et Martel (2012) proposent une approche de modélisation des risques associés à la chaîne d'approvisionnement. Leur méthodologie inclut l'identification des sources d'incertitude, leur évolution, leur niveau d'exposition et leur impact sur le réseau. Ils regroupent les sources de type hazards en multi-hazards et ce, par région. Ils estiment l'occurrence et les conséquences de ses événements sur le réseau étudié en utilisant des métriques appropriées et des fonctions de recouvrements. Selon eux et Banks (2005) la loi de Poisson peut être utilisée, dans les modèles pratiques, pour déterminer le nombre d'événements extrêmes qui surviennent dans une période donnée. Par ces méthodes, ils génèrent des scénarios qui se veulent représentatifs du futur sur un horizon donné.

Hansen *et al.* (2016) ont utilisé l'approche de simulation combinée (ASC). Cette approche combine la méthode d'analyse de scénarios avec celle de la simulation par ordinateur. Les auteurs ont appliqué L'ASC pour comparer les coûts des différentes stratégies d'impartition d'une entreprise de production. Les résultats de l'analyse des scénarios narratifs sont intégrés comme intrants pour la simulation numérique à événements discrets. L'accent est mis sur l'interaction entre les différentes parties prenantes et sur l'interaction entre les deux méthodes. Les résultats de la simulation offrent un retour d'information utile pour la création de nouveaux scénarios. Les chercheurs précisent que l'ASC augmente le niveau de confiance qu'ont les décideurs et les différentes parties prenantes envers le modèle, leur participation au développement les motive à utiliser le modèle.

Nous avons identifié qu'un seul article ayant utilisé à la fois les modèles prédictifs et la simulation. Sundaramoorthi *et al.* (2009) l'ont fait pour contribuer à une

meilleure prise de décision en milieu hospitalier. Ils ont développé des arbres de classification et de régression pour prédire le temps qu'une infirmière passe avec un patient. La simulation qu'ils ont développée sert à évaluer l'assignation des infirmières aux patients en utilisant les informations issues des modèles prédictifs. Les auteurs disent avoir utilisé une approche nouvelle. Ils indiquent que l'approche permet de réduire les états possibles de la simulation la rendant plus efficace.

3.3.3 L'application des méthodes en contexte d'aide humanitaire

Wei *et al.* (2002) ont utilisé les réseaux de neurones pour prédire les inondations catastrophiques, ils ont utilisé des données sur les zones inondables en Chine et prouvé les bienfaits de leur approche. Mis à part ces auteurs, nous n'avons identifié aucun autre article développant des modèles prédictifs pour l'aide humanitaire.

L'article d'Iakovou *et al.* (2014) propose l'utilisation de la simulation d'évènements discrets pour estimer les coûts engendrés par l'utilisation d'une seule source d'approvisionnement et cherche à prouver les bienfaits d'une deuxième source. Krejci (2015) a analysé l'interaction entre les systèmes de coordination des organisations d'aide dans les phases de préparation et de réponse aux désastres. Elle a développé une simulation dite hybride, en combinant la méthode de génération d'évènements discrets et celle basée sur les agents. Banomyong et Sopadang (2010) ont proposé un modèle logistique pour améliorer les opérations d'intervention lors de crises ou de désastres naturels par l'utilisation de la simulation Monte-Carlo. Ils ont utilisé le modèle élaboré par Banomyong *et al.* (2009), pour le cas du tsunami ayant frappé la côte sud-est de la Thaïlande en 2004, comme processus de base pour leur modèle. Par contre, comme leurs données ont été récoltées auprès des acteurs de la crise, plusieurs années après les évènements, leur simulation est basée sur une distribution triangulaire des délais de chaque activité, une simplification majeure

susceptible de mal représenter les réelles situations vécues.

Enfin, une équipe de chercheurs du Georgia Institute of Technology aux États-Unis (Alvarenga *et al.*, 2010) a mené un projet d'envergure en Afrique de l'Est. Leur projet était constitué de deux parties distinctes, mais intégrées : un modèle de simulation des délais dans les six ports d'entrées étudiés et un modèle d'optimisation des activités de transport routier. Le modèle de simulation développé servait à prédire le nombre de cargos allant être déchargés à chaque port chaque jour. L'objectif était d'arriver à prévoir les journées où les cargaisons attendues allaient quitter le port en direction de leur destination. Leur simulation utilisait des paramètres tels que les caractéristiques des cargaisons, la date prévue d'arriver, les caractéristiques des quais, le taux de service et les données historiques de déchargement à chaque port. Leur outil de simulation pouvait contribuer d'une part au choix de port pour une nouvelle cargaison et d'autre part, les résultats obtenus pouvaient servir à titre d'intrants pour le modèle d'optimisation des transports terrestres. À notre connaissance, personne n'a encore développé un modèle de simulation couvrant l'ensemble d'un réseau logistique en contexte d'aide humanitaire.

Le présent projet de recherche viendra contribuer à la littérature par l'étude d'un problème réel dans la phase de redressement suite à des catastrophes à évolutions lentes et ainsi répondre à l'appel de Galindo et Batta (2013) et Holguín-Veras *et al.* (2012) pour des études contextuelles et pratiques dans cette phase d'aide humanitaire. Puis, nous comptons répondre à la demande de Tang et Musa (2011) qui affirment qu'il y a un trou dans la littérature et un potentiel pour développer des modèles quantitatifs en gestion des risques où la prise de décision est difficile.

CHAPITRE IV

APPROCHE MÉTHODOLOGIQUE

Nos travaux de recherche suivent les grands principes de la recherche opérationnelle comme ils ont été définis par Churchman *et al.* (1957). Nous avons créé des modèles prédictifs et de simulation. Nous l'avons fait dans le but de répondre à une problématique organisationnelle concrète, soit la prise de décision des acteurs du PAM en contexte d'incertitude. Plusieurs modèles ont été itérativement créés, solutionnés, testés et améliorés. Les modèles, instanciés par des données réelles, ont été rigoureusement évalués. Nos solutions technologiques pourraient être implantées et maintenues à jour par l'organisation. Ces solutions et leurs applications pratiques dans un domaine peu étudié apportent une contribution scientifique significative.

La méthodologie proposée pour le mémoire se divise en trois étapes : 1) identifier et catégoriser les sources d'incertitude selon la grille méthodologique développée par Dumitriu (2006), 2) collecter et analyser les données quantitatives, et 3) élaborer les modèles pour estimer les paramètres stochastiques selon les techniques de l'apprentissage-machine. Ces trois étapes méthodologiques ont été prévues pour nous permettre d'atteindre nos deux premiers objectifs de recherche : 1) identifier les sources d'incertitude et les catégoriser et 2) déterminer les modèles pour estimer les paramètres stochastiques. L'étape un suit une approche méthodologique plus

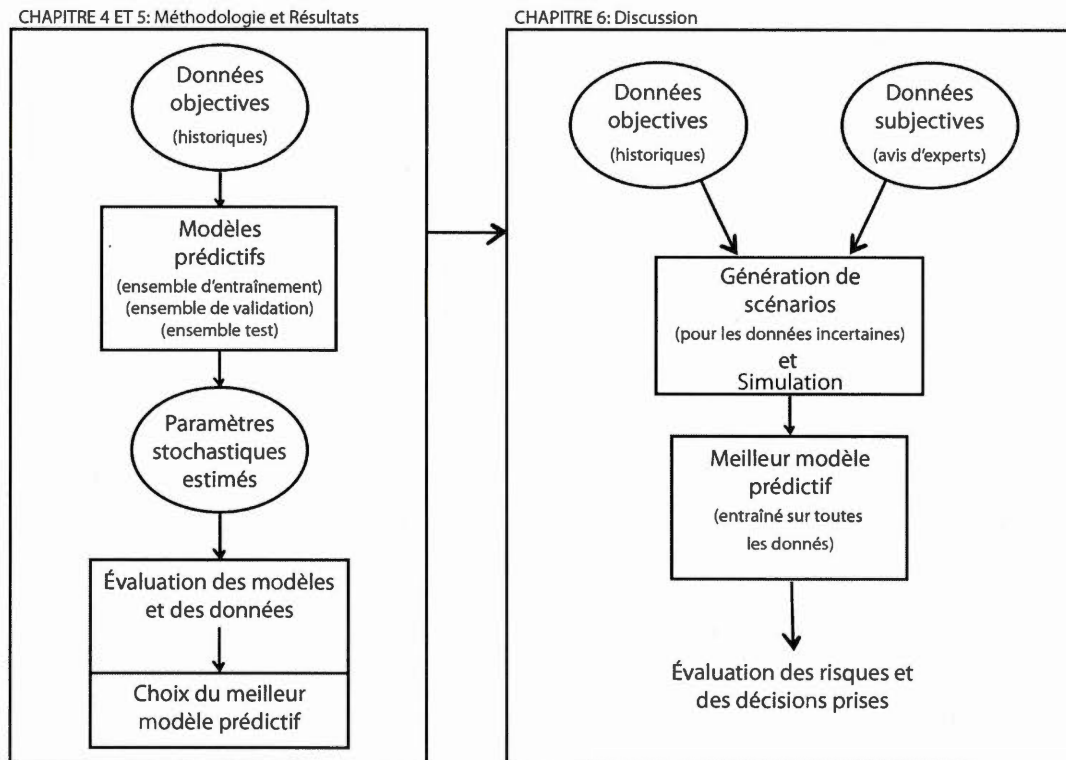


Figure 4.1 Schéma de la méthodologie quantitative

qualitative alors que les étapes deux et trois sont plus quantitatives, car on fait des analyses statistiques. Les sections suivantes du chapitre quatre présentent chacune de ces étapes et précisent les actions qui ont été faites, soit la méthodologie propre à chaque étape. Les résultats de ces étapes sont présentés dans le chapitre cinq. La génération de scénarios et l'évaluation des risques, qui permet de répondre à l'objectif trois (générer les scénarios) et à notre question de recherche, sont des étapes qui sont décrites dans la discussion (chapitre 6). La Figure 4.1 illustre les étapes méthodologiques quantitatives du projet.

4.1 Identifier et catégoriser les sources d'incertitude

La première étape de notre approche méthodologique nécessite l'identification des sources d'incertitude en lien avec les activités du PAM au Niger. Il a été important d'identifier toutes les sources d'incertitude qui sont susceptibles d'affecter la performance de la chaîne d'approvisionnement du PAM en entraînant des variations sur les coûts encourus, les délais de transport ou les pertes de denrées alimentaires (les trois paramètres stochastiques de notre étude). Nous nous sommes concentrés sur les sources d'incertitude qui peuvent être associées aux nœuds et aux arcs du réseau logistique illustrés à la Figure 1.3, car ce sont celles qui sont directement en lien avec la problématique. Les sources d'incertitude liées aux donateurs (p. ex., le risque lié à la réputation du PAM) ou celles liées aux bénéficiaires (p. ex., le risque de boycottage) n'ont donc pas été analysées puisqu'ils sont en amont ou en aval du réseau étudié. Pour parvenir à dresser le portrait des sources d'incertitude associées à la problématique, plusieurs techniques d'identification peuvent être valables tant qu'elles sont utilisées de façon exhaustive (Coppola, 2015).

Nous avons utilisé le modèle holistique d'identification des risques de Dumitriu (2006). Ce cadre méthodologique est prévu pour identifier la totalité des risques d'une entreprise. Chaque catégorie de risque doit représenter un ensemble de sources d'incertitude sans qu'il y ait de redondance. La grille méthodologique développée par Dumitriu (2006) contient 11 grandes catégories de risques avec leurs sources respectives, pour un total de 46 sources d'incertitude. La technique a été testée auprès de 500 entreprises des plus performantes financièrement aux États-Unis (*Fortune 500 companies*). Comme le modèle fût conçu pour être appliqué à tout type d'entreprise, nous l'avons jugé approprié pour un organisme à but non lucratif. Nous avons choisi de suivre cette démarche d'analyse pour identifier les sources d'incertitude, car elle permet d'être assidu pour faire un inventaire

complet. Nous avons examiné point par point toutes les sources d'incertitude qui pourraient engendrer des risques pour l'organisation. Le réseau logistique a permis de structurer notre application de la méthode d'identification des risques. C'est aussi grâce à des discussions avec les experts du milieu, à la lecture des rapports officiels du PAM (WFP, 2015; WFP, 2016) et à la lecture de la littérature scientifique que nous avons bien compris les difficultés rencontrées par le PAM ainsi que le contexte dans lequel leurs activités s'insèrent. Nous avons exploré la littérature en gestion des risques de la chaîne d'approvisionnement (Singhal *et al.*, 2011; Tang et Musa, 2011; Rao et Goldsby, 2009), celle en contexte d'aide humanitaire (Holguín-Veras *et al.*, 2012; Jahre *et al.*, 2016; Coppola, 2015) et celle qui a ciblé l'Afrique comme terrain d'étude (Nathan Associates Inc., 2013; Teravaninthorn et Raballand, 2009; Rancourt *et al.*, 2014).

Rappelons que Klibi *et al.* (2010) catégorisent (sous trois catégories) les sources d'incertitude d'un réseau en fonction de la qualité de l'information pouvant être récoltée pour les représenter. On ne peut pas traiter toutes les sources d'incertitude de la même façon, car pour certaines nous sommes en mesure d'extraire une distribution représentative de l'historique des événements, elles sont qualifiées de type *randomness* selon Klibi *et al.* (2010), mais pour d'autres ce n'est pas possible. Pour les sources d'incertitude pouvant entraîner des événements catastrophiques, l'information est souvent rare (*hazard*) ou difficilement récoltable (*deep uncertainty*). Ainsi, dans le cadre de ce projet, nous avons catégorisé les sources d'incertitude identifiées selon deux catégories : celles qui peuvent entraîner des événements aléatoires et celles susceptibles d'engendrer des événements catastrophiques. Nous définissons les événements catastrophiques comme ceux ayant une probabilité d'occurrence très faible combinée à un niveau d'impact très élevé. Une même source d'incertitude peut être susceptible d'entraîner des événements aléatoires et des événements catastrophiques. Banks (2005) l'illustre bien en indiquant

que l'accumulation d'eau engendre souvent des inondations mineures, mais que certaines inondations sont significativement plus grandes et capables de détruire des infrastructures et de faire des victimes.

Une fois que les sources d'incertitude ont été identifiées, il était important d'identifier quel(s) paramètre(s) stochastique(s) ces sources sont susceptibles d'affecter, et sur quelle section du réseau elles agissent. Les résultats liés à la démarche qui vient d'être proposée sont présentés à la première section du chapitre sur les résultats.

4.2 Collecter et analyser les données

La seconde étape de notre approche méthodologique consiste en une collecte de données. Il était essentiel de collecter d'abord des données pour représenter les paramètres stochastiques étudiés (les délais, les coûts et les pertes). Puis, d'autres données ont été collectées dans le but de prédire et de comprendre le comportement de ces paramètres stochastiques. Nous avons tenté d'obtenir des informations qui représentent chaque catégorie d'évènement (aléatoire et catastrophique). Nous avons recueilli un maximum de données objectives, des données historiques qui proviennent soit de l'interne (de l'historique des activités du PAM) ou de l'externe (données publiques). Nous avons fait un diagnostic de ces données et procédé à des analyses statistiques. Suite à ces analyses, nous sommes parfois retournés collecter des données supplémentaires. Des données subjectives ont ensuite été collectées afin d'avoir l'information manquante pour représenter les sources d'incertitude et les paramètres stochastiques associés à chaque arc du réseau. Étant sur les lieux et grâce à leur expérience, les gestionnaires du PAM possèdent des informations qui sont importantes. Ces données subjectives étaient d'autant plus nécessaires pour la formulation d'évènements catastrophiques.

4.2.1 Les données objectives

Nous avons recueilli des données historiques auprès des logisticiens du PAM et dans les bases de données disponibles publiquement. Il a d'abord été important de recueillir des données historiques pour représenter les paramètres stochastiques associés à chaque arc du réseau logistique (Figure 1.3). Dans le cas des délais et des pertes, des données sont enregistrées par les systèmes d'information du PAM chaque fois qu'un don leur est dédié. Les données collectées proviennent à la fois de l'ancien système d'information utilisé par le PAM, Commodity movement processing and analysis system (COMPAS), et du plus récent système, Logistics Execution Support System (LESS) (WFP, 2016). Vers la fin du premier trimestre de l'année 2016, le PAM a changé son système d'information logistique. Les logisticiens du PAM ont extrait les données disponibles de ces deux systèmes d'information et nous les ont transmises sous forme de fichiers Excel. Les logisticiens récoltent aussi des données de façon manuelle, c'est d'ailleurs le cas pour le transport maritime. Ces informations sont enregistrées dans des tableurs Excel. La structure des données varie grandement d'un fichier à l'autre. Les données provenant du système COMPAS avaient été analysées et prétraitées par les logisticiens du PAM. Nous avons majoritairement travaillé sur les tableurs présentant des données brutes. Nous voulions l'information décrivant les activités sur chaque corridor : son continent d'origine, le port d'arrivée, les compagnies de transport, le type de denrée, les quantités, ainsi que les dates où les activités importantes ont eu lieu. Nous avons ciblé les données susceptibles de nous aider à comprendre la variabilité des paramètres stochastiques. En créant notre base de données, un nettoyage important a dû être fait, soit par l'identification des erreurs et des manques, la corrections et la standardisation de l'écriture. Les données ont été traitées pour en faire ressortir l'information voulue. Par exemple, la date de départ du camion a été soustraite à la date d'arrivée à quai du navire pour obtenir le délai de ma-

nutrition. Les quantités de denrées sont aussi généralement enregistrées dans le système à chaque étape importante (chargement du navire, chargement du camion, et réception au Niger). Grâce à ces données, nous avons pu calculer les pertes. Les données historiques représentant les délais et les pertes sont essentielles, car elles sont utilisées comme intrants pour la modélisation des paramètres stochastiques. Les données ont été analysées grâce à R, un langage de programmation destiné aux statistiques et à la science des données (R Core Team, 2018) et à RStudio, un environnement de développement intégré pour programmer en R (RStudio Team, 2016). Nous avons décidé de procéder ainsi, car en plus de nous permettre de faire des analyses statistiques sur les données recueillies, R nous a aussi permis de développer des modèles prédictifs et de générer des scénarios pour un modèle de simulation.

Pour ce qu'il y est des coûts, c'est par l'historique des contrats signés entre le PAM et ses compagnies de transitaires qu'il est possible de récolter des données. Les contrats donnent de l'information sur la structure des coûts et on peut percevoir le niveau de variabilité avec plusieurs observations. Pour parvenir à analyser et estimer les coûts, un grand nombre de contrats sont nécessaires. Sahinyazan *et al.* (2017) se sont déplacés en Afrique et sont parvenus à collecter l'information issue de 213 contrats entre les transitaires et le PAM. Pour notre part, la cueillette de ces documents contenant de l'information jugée sensible était difficile. Les logisticiens ont été en mesure de nous faire parvenir seulement quelques exemples d'appel d'offres (appelé *Request for quotation*, RFQ, en anglais). Le RFQ est envoyé aux compagnies de transitaires lorsqu'un don est confirmé pour le Niger. Il donne des détails sur le don. Par exemple, le moment estimé de l'arrivée du navire (*estimated time of arrival*, ETA), le type ou les types de denrées et les quantités devant être transportées à chaque destination. Les compagnies de transitaires sont appelées à répondre au RFQ à l'intérieur d'un délai spécifique en précisant les coûts et les

délais prévus pour les différentes activités logistiques. Une fois les réponses des transitaires reçues, les responsables de la logistique du PAM se rencontrent pour faire le choix de la compagnie de transitaires retenue. Les différentes offres sont rassemblées et présentées sur un document appelé *Tender*. Nous avons reçu que deux documents de ce type. C'est suffisant pour bien comprendre le fonctionnement des contrats, mais ce n'est pas suffisant pour pouvoir concevoir un modèle capable d'estimer les coûts futurs.

Ensuite, nous avons recueilli des données pour représenter toutes les sources d'incertitude identifiées. Certaines informations pouvaient être déduites ou extraites des données du PAM (p. ex., la saison, les parties prenantes et les retards dans la documentation). Pour les autres, nous nous sommes retournés vers les bases de données publiques. Étant donné que le niveau de congestion des ports de l'Afrique de l'Ouest a été confirmé comme étant problématique par Nathan Associates Inc. (2013) ainsi que par les responsables du PAM au Niger, des données représentatives de la congestion dans ce secteur étaient, entre autres, nécessaires. Des données publiques abrégées peuvent être utilisées comme des proxies pour quantifier certaines sources d'incertitude. Klibi et Martel (2012) mentionnent plusieurs sites internet où l'on peut avoir accès à des indices variés (politiques, stabilité régionale, performance économique, vulnérabilité aux catastrophes naturelles, qualité des infrastructures publiques et autres). Holguín-Veras *et al.* (2012) affirment que l'on peut s'attendre à ce qu'un modèle d'aide à la décision basé sur des mesures proxys puisse fonctionner raisonnablement bien lorsqu'il s'applique à la phase de redressement à long terme suite aux catastrophes, car les opérations de cette phase prennent place dans un environnement assez stable. Un critère important pour notre collecte de données publiques était de trouver des données qui couvraient, au minimum, la période couverte par les données issues des systèmes du PAM (de l'année 2014 à 2016). De façon itérative, tout au long de notre collecte de données

publiques, une analyse des données a été faite avec le logiciel R. L'objectif était de comprendre les relations de dépendances et les corrélations, et d'utiliser les données pour entraîner des modèles prédictifs. Les données manquantes pour représenter les paramètres stochastiques et les sources d'incertitude qui les affectent ont été identifiées.

4.2.2 Les données subjectives

Il est fréquent que les données objectives ne suffisent pas pour traiter les sources d'incertitude, surtout lorsque la probabilité d'occurrence est très basse et les impacts sur la chaîne sont très grands (Klibi et Martel, 2012; Knemeyer *et al.*, 2009; Banks, 2005). L'historique des données seul ne permet donc pas toujours de bien représenter ces événements catastrophiques. Par ailleurs, Banks (2005) indique que les sources d'incertitude liées aux actes terroristes doivent être considérées de façon objective et subjective, car même s'ils arrivent sans prévenir comme la plupart des catastrophes naturelles, ils ne se produisent pas de façon aléatoire, ils ont été prévus. La collecte de données subjective nous a donc permis d'aller chercher l'expérience de l'utilisateur. L'utilisateur des modèles et spécialiste du terrain doit préciser ce à quoi il s'attend. S'il y a des cas où l'information sur une section du réseau est manquante, l'utilisateur doit identifier, au minimum, la valeur qu'il considère comme la plus probable pour ce paramètre. Nous obtiendrons donc une distribution statistique discrète ou une valeur unique pour représenter un paramètre sur une section du réseau donnée.

Ce sont d'autant plus les résultats des modèles prédictifs qui ont guidé les besoins en données subjectives. Les données subjectives sont combinées aux données objectives lors de la génération des scénarios nécessaire à la simulation. Comme Schwartz (1992) et Hansen *et al.* (2016) l'indiquent, l'implication des gestionnaires

lors de la génération des scénarios est bénéfique, car elle contribue à étendre leur compréhension par rapport à la problématique à laquelle ils sont confrontés. Enfin, les sources d'incertitude qui ne sont pas susceptibles d'entraîner des événements catastrophiques et qui sont bien représentées par l'historique des activités du PAM ne nécessitent pas de collecte de données subjectives.

4.3 Élaborer les modèles pour estimer les paramètres stochastiques

Pour estimer la valeur des paramètres stochastiques à l'étude, nous avons utilisé des modèles d'apprentissage statistique supervisés. Que ce soit pour estimer les coûts, les délais ou les pertes, ce sont des modèles de régression qui devraient être utilisés, car les variables à prédire sont numériques. Nous avons utilisé les données collectées pour représenter les paramètres stochastiques lors du transport maritime, de la manutention au port et du transport terrestre. Elles sont les variables à prédire. Nous utilisons le terme variables explicatives pour parler des données qui ont été récoltées pour expliquer et comprendre la variabilité des variables à prédire. Ces variables sont appelées prédicteurs (*predictors*) par la plupart des auteurs.

Certaines de nos variables explicatives étaient corrélées entre elles et nous avons fait le choix de les conserver. Nous avons développé des modèles dans le but de prédire, le mieux possible, la valeur prise par le paramètre stochastique dans le futur. Nous n'avons pas cherché à attribuer, avec précision, quel effet observé sur le paramètre stochastique est attribuable à quelle variable explicative. Les variables explicatives n'ont pas non plus été classées en ordre d'importance selon leur niveau de contribution au modèle. Mullainathan et Spiess (2017) expliquent que c'est une pratique omniprésente dans le domaine de l'apprentissage-machine comparativement aux estimations statistiques traditionnelles. Kutner *et al.* (2005)

précisent que la multicolinéarité n'empêche pas d'obtenir un bon modèle capable de prédire avec précision. La multicolinéarité affecte les coefficients et les *p-values* dans les modèles de régression, mais elle n'affecte pas nécessairement la capacité d'inférence et de prédiction.

Actuellement, les logisticiens du PAM comparent les délais prévus par les transitaires pour effectuer les opérations avec les délais historiques moyens sur les corridors. C'est donc avec le délai moyen qu'ils évaluent les offres des compagnies. Par conséquent, nous avons utilisé la moyenne historique, pour un paramètre stochastique sur une section d'un corridor comme valeur prédite. La moyenne, comme estimation d'un paramètre stochastique, nous a servi de référence pour l'évaluation. Nous avons comparé la performance prédictive de cette moyenne estimée pour évaluer la performance des autres modèles développés. Pour ce faire, nous avons choisi de partitionner nos données en différents ensembles. Pour comparer les modèles développés et analyser leurs performances prédictives, nous avons créé deux ensembles. L'échantillonnage est fait grâce au générateur de nombre pseudo-aléatoire du logiciel R. La fonction *sample()* a été utilisée. Le premier ensemble est celui d'entraînement, il comporte 70% des observations de la base de données. L'échantillonnage est fait à l'intérieur d'un vecteur de données; en pigeant parmi le nombre de lignes de la base de données. Les modèles prédictifs sont développés en utilisant les données de cet ensemble. Le second ensemble est celui prévu pour la validation. L'ensemble de validation est constitué des observations qui n'ont pas été sélectionnées pour l'ensemble d'entraînement, il comporte 30% des données. La sélection du meilleur modèle est faite à partir de cet ensemble. Nous avons procédé ainsi pour prédire la valeur des paramètres stochastiques et évaluer les résultats pour les opérations de transport maritime, de manutention au port et de transport terrestre.

Tel que James *et al.* (2013) l'indique, il n'y a pas de modèle prédictif dominant.

Un modèle peut bien fonctionner sur certaines données, mais un autre modèle pourrait avoir de meilleurs résultats prédictifs pour un problème semblable dans un contexte différent. C'est pourquoi nous avons décidé de comparer la performance prédictive de la régression linéaire, de l'arbre de décision et de la forêt aléatoire avec celle obtenue pour la moyenne estimée. Notre objectif n'était pas d'identifier le meilleur modèle pour prédire, mais d'identifier un bon modèle. Idéalement un modèle qui offre de la stabilité dans la qualité de ses prédictions. La modification de l'échantillon ne change pas de façon marquée les résultats prédictifs. Ce niveau de stabilité a été analysé sur dix échantillons tirés aléatoirement à partir des données collectées.

Nous avons développé des modèles de régression linéaires. Pour ce faire, la fonction *lm()* du logiciel R a été utilisée. Un modèle de régression linéaire multiple prend la forme suivante :

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p + \epsilon , \quad (4.1)$$

où y est la variable à prédire, dans notre étude celle-ci correspond au paramètre stochastique pour une section du réseau (transport maritime, la manutention au port ou le transport par camion). Cette variable est la variable dépendante. Toutes les variables explicatives collectées ont été utilisées (les variables indépendantes en statistique classique), soit de x_1 à x_p variables explicatives, p étant le nombre total de variables explicatives collectées. ϵ est le terme qui représente l'erreur (le bruit). Il capture les autres facteurs qui influencent la variation de la variable à prédire ; ce qui n'est pas capturé par le modèle. L'hypothèse que ce terme est aléatoire et de moyenne zéro est faite. Les constantes β_0 à β_p sont les seuls paramètres du modèle de régression linéaire multiple. Ils doivent être estimés grâce aux données. Ces paramètres, une fois estimés, prennent la forme $\hat{\beta}_0$ à $\hat{\beta}_p$ dans la formule utilisée

pour les prédictions :

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \dots + \hat{\beta}_p x_p . \quad (4.2)$$

Les paramètres sont estimés avec l'approche des moindres carrés ordinaire (*ordinary least squares*). Ils sont estimés en minimisant la somme des carrés des résidus (les erreurs de prédiction). Shmueli *et al.* (2017) expliquent que l'objectif du modélisateur de la régression linéaire peut prendre deux formes : 1) expliquer l'effet moyen des variables explicatives (statistiques classiques) ou 2) prédire la valeur de la variable à prédire sur de nouvelles données (exploration de données). Dans le premier cas, l'attention du modélisateur est axée sur les coefficients (β) dans le second, sur les prédictions ($\hat{y}$). Dans le cas présent, la priorité a été donnée à la qualité de la prédiction définie selon différentes mesures d'erreur, elles sont présentées à la fin de cette section. Le coefficient de détermination R^2 du modèle de régression linéaire est donné par la fonction $lm()$. Le R^2 prend toujours une valeur qui se situe entre 0 et 1, mais il n'est pas toujours défini par le même calcul. Dans la fonction de régression linéaire $lm()$, R calcule le coefficient de détermination à l'aide de la corrélation. Pour la régression linéaire multiple, il effectue le carré de la corrélation des valeurs prédites avec leurs vraies valeurs (James *et al.*, 2013). Lorsque le coefficient de détermination R^2 du modèle de régression linéaire était jugé trop bas, nous avons collecté de nouvelles variables explicatives disponibles publiquement afin de mieux comprendre la variabilité du paramètre stochastique étudié. Plusieurs itérations entre le développement de modèles de régression linéaire multiples et la collecte de données ont été faites sur cette base.

Pour prédire les paramètres stochastiques des opérations logistiques, nous avons aussi développé des arbres de régression (Breiman *et al.*, 1984). Dans ce modèle, les relations entre les variables explicatives et la variable prédictive peuvent être complexes et pas nécessairement linéaires. Mullainathan et Spiess (2017) expliquent que les méthodes basées sur les arbres permettent les interactions entre

les variables explicatives. Un arbre de régression de base est construit en deux étapes (James *et al.*, 2013). Premièrement, l'espace des prédicteurs (variables explicatives) est segmenté en différentes régions. L'ensemble des valeurs possibles pour $x_1, x_2, \dots, x_p$ est divisé en j sous-ensembles rectangulaires distincts, qui ne se chevauchent pas, $R_1, R_2, \dots, R_j$. Ainsi, les R_j sont formés d'une combinaison de plusieurs conditions de type $x_p \geq t_k$. Deuxièmement, les prédictions sont faites pour chaque sous-ensembles. La même valeur est prédite pour toutes les observations qui font partie du sous-ensembles R_j . C'est la moyenne des vraies valeurs des variables à prédire faisant partie du sous-ensembles R_j dans l'ensemble d'entraînement. Pour la conception du modèle, l'objectif est le même que pour la régression linéaire. Les sous-ensembles sont déterminées afin de minimiser la somme des carrés des résidus (des erreurs de prédiction).

Pour développer les modèles d'arbre de régression simple sur le logiciel R, nous avons installé le package `rpart` (Therneau et Atkinson, 2018) et utilisé la fonction `rpart()`. Un exemple simplifié d'arbre de régression, pour prédire le délai d'opération, est présenté à la Figure 4.2. Le package `rpart.plot` (Milborrow, 2018) et l'utilisation de la fonction `prp()` ont été nécessaires pour créer l'arbre de décision tel qu'illustré à la Figure 4.2. Il est impossible de considérer toutes les segmentations possibles pour l'espace des prédicteurs (James *et al.*, 2013). L'arbre est donc construit du haut vers le bas. Son interprétation est simple, car il présente des règles clairement définies. La première séparation de l'arbre se fait à partir du meilleur prédicteur avec un certain point de rupture. L'approche est connue sous le nom de séparation binaire récursive (*Recursive Binary Splitting*). Dans l'exemple de la Figure 4.2, la variable explicative la plus importante est celle du nombre de conflits total et son point de rupture est de 12 conflits. Si la condition est respectée, il faut suivre la branche de gauche de l'arbre et la branche de droite si elle ne l'est pas. Ainsi, lorsque le nombre de conflits total est plus petit que

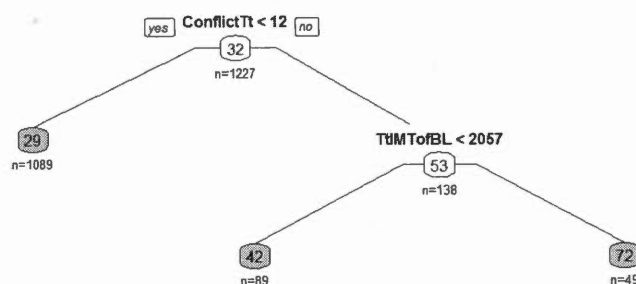


Figure 4.2 Exemple d'arbre de décision de base

12, le délai prédit est de 29 jours. Lorsqu'il est supérieur ou égal à 12, la variable explicative indiquant la quantité totale inscrite au certificat de connaissance (*bill of lading*) est considérée. Si cette quantité est inférieure à 2 057 TM, le délai prédit est de 42 jours. Si la quantité est supérieure ou égale à ce point de rupture, le délai est alors de 72 jours. La prédiction de 72 jours est basée sur la moyenne des délais observés pour les 49 cas qui ne respectent pas les deux conditions (en caractère gras dans la Figure 4.2). Chaque branche peut être séparée de nouveau, augmentant la complexité et la profondeur de l'arbre.

Certains paramètres peuvent être spécifiés à la conception du modèle d'arbre de régression. Avec la fonction *rpart()*, la profondeur de l'arbre peut être contrôlée avec le paramètre de complexité (*Complexity parameter*, CP). Il est difficile de déterminer le bon moment pour arrêter le développement de l'arbre, et donc la bonne valeur pour ce paramètre (Shmueli *et al.*, 2017). Par défaut, le CP prend la valeur de 0,01 dans la fonction *rpart()*. Plus le CP est petit, plus le niveau de détail de l'arbre est grand. De manière générale, l'erreur diminue plus le nombre de noeuds augmente, jusqu'au point où il peut y avoir surapprentissage (Shmueli *et al.*, 2017). Ainsi, l'arbre est bâti en minimisant la somme de deux objectifs. Le coût de l'arbre $C(T)$, basé sur les erreurs de prédiction, est minimisé et la

complexité de l'arbre $|T|$ l'est aussi. Le CP, identifié par α dans la formule qui suit, représente le poids associé à l'objectif de la complexité de l'arbre :

$$C\alpha(T) = C(T) + \alpha|T| . \quad (4.3)$$

Une constatation importante a été faite grâce à l'apprentissage machine : en faisant la moyenne de plusieurs petits modèles on obtient, très souvent, un modèle avec un meilleur pouvoir prédictif sur de nouvelles données (Varian, 2014). James *et al.* (2013) et Shmueli *et al.* (2017) mentionnent que la performance prédictive des arbres de décisions peut être grandement améliorée par les méthodes qui agrègent plusieurs arbres de décisions. Pour un modèle donné, l'idée étant que la variance associée aux prédictions fournies par un arbre unique est généralement élevée. Ce problème peut-être résolu, en partie, en effectuant la moyenne des prédictions obtenues via l'utilisation de plusieurs arbres pour le même modèle. Le modèle de forêt d'arbres décisionnels (*random forest*) en est un exemple. Ainsi, pour notre modèle de forêt d'arbres de régression (Breiman, 2001), nous avons choisi d'utiliser la fonction *randomForest()* issue du package *randomForest* (Liaw et Wiener, 2002). La méthode de forêt aléatoire est basée sur le *bagging* (pour *bootstrap aggregation*), originalement de Breiman (1996). Les arbres de régression sont bâtis à partir de différents échantillons aléatoires (avec remplacement) issus des données fournies pour la construction du modèle (l'ensemble d'entraînement). De plus, dans le modèle de forêt aléatoire, un sous ensemble aléatoire de variables est choisi parmi les p variables explicatives à chaque fois qu'une séparation est effectuée dans un arbre. La variance dans les prédictions du modèle est d'autant réduite, car les arbres sont décorrélés (Breiman, 2001). Dans la fonction *randomForest()*, par défaut le paramètre *ntree*, indiquant le nombre d'arbres dans la forêt, est de 500 arbres. Le paramètre *mtry*, indiquant le nombre de variables explicatives à considérer pour chaque séparation de l'arbre, est aussi un paramètre qui devrait être ajusté par l'utilisateur (Smith *et al.*, 2013; Ouallouche *et al.*, 2018).

Par défaut, il prend le quotient du nombre total de variable explicatives par trois. Similaire à la régression linéaire, la fonction retourne la proportion de la variation qui est expliquée par le modèle. Dans le cadre de cette fonction le R^2 est calculé ainsi :

$$1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} . \quad (4.4)$$

Dans ce cas ci, le R^2 indique la proportion de la variance, de la variable à prédire, qui est expliquée par le modèle. Le modèle de forêt aléatoire est développé à partir de n observations, et les y_i sont les vraies valeurs de la variable à prédire. La moyenne de cette variable est notée $\bar{y}$ et les valeurs prédites par le modèle sont notées $\hat{y}_i$. Nous avons analysé le R^2 sur les modèles développés, mais nous l'avons pas utilisé pour évaluer la performance prédictive de nos modèles. Nous avons fait ce choix, car le R^2 n'est pas toujours évalué de la même façon par les modèles. Par ailleurs, le R^2 ne permet pas d'évaluer la performance d'un modèle sur de nouvelles données.

Pour analyser la performance prédictive des modèles de régression développés, nous avons utilisé les mesures d'erreurs issues de la fonction *accuracy()* de R. Cette fonction est disponible avec l'installation du package *forecast* (Hyndman *et al.*, 2018) dans R. La fonction *accuracy()* calcule cinq métriques pour analyser les résultats prédictifs d'un modèle. L'erreur moyenne (*Mean Error*, ME) est calculée ainsi :

$$ME = \frac{1}{n} \sum_{i=1}^n e_i . \quad (4.5)$$

ME correspond à la moyenne des erreurs e_i de prédictions $(y_i - \hat{y}_i)$ sur les n observations de l'échantillon. L'erreur indique la différence entre la vraie valeur du paramètre et la valeur prédite. Cette erreur est en termes de jours lorsque les paramètres stochastiques sont les délais, de quantités pour les pertes et de dollars pour les coûts. Une erreur de prédiction peut être inférieure ou supérieure

à la valeur observée. Le résultat du calcul de ME indique dans quelle mesure, en moyenne, les prédictions sous-estiment ou surestiment les vraies valeurs. L'erreur absolue moyen (*Mean Absolute Error*, MAE) se calcule de façon similaire :

$$MAE = \frac{1}{n} \sum_{i=1}^n |e_i| . \quad (4.6)$$

Dans le calcul de MAE, la valeur absolue des erreurs sont additionnées. Le résultat obtenu par ce calcul indique l'erreur moyenne dans l'unité mesurée (en jours pour les délais, en quantités de denrées pour les pertes et en dollars pour les coûts). Il y a aussi des mesures d'erreurs relatives. Elles sont données en pourcentage. Le pourcentage d'erreur moyen (*Mean Percentage Error*, MPE) est calculé ainsi :

$$MPE = \frac{1}{n} \sum_{i=1}^n \frac{e_i}{y_i} . \quad (4.7)$$

Pour chaque observation, l'erreur de prédiction est divisée par la vraie valeur observée pour un paramètre stochastique. L'équation 4.7 permet d'évaluer, en pourcentage, les erreurs de prédiction sous-estiment ou surestiment les vraies valeurs des observations. Pour obtenir le pourcentage d'erreur moyen en valeur absolue (*Mean Absolute Percentage Error*, MAPE) le calcul est similaire :

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{e_i}{y_i} \right| . \quad (4.8)$$

Avec l'équation 4.8, on obtient la moyenne des erreurs relatives en pourcentage. Enfin, la cinquième mesure d'erreur donnée par la fonction *accuracy()* est celle de la racine carrée de la moyenne des erreurs mises au carré (*Root Mean Square Error*, RMSE). Le RMSE est calculé par l'équation suivante :

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n |e_i|^2} . \quad (4.9)$$

Comparativement au MAE, le RMSE pénalise plus grandement les erreurs aberrantes, car les erreurs sont d'abord élevées au carré. Le résultat du RMSE est

obtenu dans la même unité que le paramètre stochastique étudié. Dans le cadre de notre étude, nous avons comparé les résultats pour les MAE, MAPE et RMSE. Contrairement au ME et MPE, ces trois métriques donnent des informations qui peuvent être comparées pour évaluer les modèles. Plus leur valeur est petite, meilleur est le modèle. Il est peu utile d'analyser les résultats prédictifs sur l'ensemble d'entraînement, car l'objectif est de vérifier si le modèle est bon pour estimer la valeur d'un paramètre en présence de nouvelles données. Ce sont donc les mesures d'erreurs (MAE, RMSE et MAPE) calculées sur l'ensemble de validation qui ont été analysées. La performance prédictive de la moyenne estimée a été comparée à celle de la régression linéaire, de l'arbre de décision et de la forêt aléatoire. Le modèle prédictif qui présentait, en moyenne, les meilleurs résultats a été sélectionné.

Une fois le meilleur type de modèle sélectionné, nous l'avons amélioré. Différentes versions de ce modèle ont été comparées et les paramètres de ses modèles ont été ajustés. Pour ce faire, nous avons cette fois partitionné nos données en trois ensembles. L'ensemble d'entraînement comporte 70% des données, l'ensemble de validation comporte 15% et l'ensemble de test 15%. Le premier ensemble a servi à concevoir les modèles. Avec le second, nous avons comparé différents paramètres pour les modèles. Nous avons sélectionné les paramètres qui donnaient, en moyenne, les meilleurs résultats. Avec les paramètres choisis, nous avons comparé les résultats prédictifs des différentes versions du modèle sur l'ensemble de test pour dix échantillons différents. La version permettant, en moyenne, une meilleure performance prédictive et une certaine stabilité (avec un écart-type bas) a été sélectionnée. Ce modèle, a ensuite été entraîné sur la totalité des données collectées et a été utilisé pour la simulation. De plus amples détails sur les versions des modèles et leurs paramètres sont donnés dans le chapitre sur les résultats.

CHAPITRE V

RÉSULTATS

Dans ce chapitre, nous présentons les résultats obtenus suite à l'application de notre démarche méthodologique. D'abord, les sources d'incertitude pouvant affecter les opérations du PAM sont présentées. Nous abordons l'effet qu'elles peuvent avoir sur les paramètres stochastiques étudiés et la section du réseau logistique impliquée. Ensuite, nous présentons une analyse descriptive des données récoltées. Les données collectées pour représenter les paramètres stochastiques sont analysées et les corrélations sont énoncées. Les données récoltées pour comprendre la variabilité des paramètres stochastiques sont énumérées et des analyses statistiques sont présentées. Enfin, les résultats des modèles prédictifs sont comparés et les modèles finaux sont sélectionnés.

5.1 Les sources d'incertitude

Dans cette section, nous identifions, décrivons et donnons des exemples pour chaque source d'incertitude. Nous abordons aussi l'effet que peuvent avoir les sources d'incertitude sur les paramètres stochastiques étudiés. Sur trois graphiques (un pour les délais, un pour les coûts et un pour les pertes), les sources d'incertitude sont associées à la section du réseau logistique qu'elles peuvent affecter.

5.1.1 La catégorisation des sources d'incertitude

Pour parvenir à la catégorisation des sources d'incertitude présentées dans cette section, nous avons retenu uniquement les sources jugées capables d'avoir un impact sur les paramètres stochastiques à l'étude. Nous avons nommé les catégories afin que les sources puissent être intuitivement classifiées dans une et une seule catégorie. Les discussions avec les acteurs du PAM nous ont permis de bien comprendre l'ampleur du contexte d'incertitude dans lequel ils opèrent. Les experts ont mentionné qu'il est possible qu'un navire en route pour le Niger soit redirigé vers une autre destination. Cette situation arrive parfois lorsque le don est jugé plus nécessaire ailleurs. Par exemple, lors d'une catastrophe naturelle soudaine dans un autre pays. Cette source d'incertitude, susceptible d'affecter le délai de transport maritime, ne fait donc pas partie des sources d'incertitude catégorisées parce que si le don est redirigé vers un autre pays, il n'y a plus de décisions à prendre pour ce don dans le système étudié. Les sources d'incertitude qui sont présentées dans cette section n'ont pas été classifiées selon un ordre d'importance, elles ont toutes été jugées suffisamment importantes pour être incluses à l'étude. Les six catégories de sources d'incertitude répertoriées sont résumées dans le tableau 5.1 et décrites dans les prochains paragraphes.

L'environnement représente une source d'incertitude importante. Pour cette catégorie nommée **Environnementale**, ce sont principalement les conditions climatiques en l'Afrique de l'Ouest qui doivent être analysées ; les événements naturels qui pourraient survenir. Dans le cas du Niger, les saisons sont importantes. La saison des pluies s'échelonne typiquement du mois de juin au mois d'octobre inclusivement, le mois d'août étant le plus pluvieux en moyenne (The World Bank Group, 2018). Lorsque les pays de l'Afrique de l'Ouest reçoivent de grandes quantités de pluies durant une certaine période, on parle rapidement d'inondations.

Tableau 5.1 Les catégories de sources d'incertitude

Catégorie	Source d'incertitude	Description	Exemple
E	Environnementale	Causé par les conditions climatiques.	Inondation. Sécheresse.
I	Interne	Causé par le PAM. Généralement lors d'un retard ou d'un manque à la documentation requise.	Certificat de connaissance. Document de dédouanement.
L	Légale	Lorsqu'il y a procédure légale (d'autres arrivages pourraient être retardés)	Conflit avec une compagnie de transitaire. Manque à la procédure douanière.
P	Parties prenantes : transporteur maritime (tm), port (p), transitaire (t), douane (d)	Causé par les compagnies de transport ou les lieux à traverser.	Congestion (port ou douane). Défaillance des équipements.
R	Ressource	Causé par les variations du prix ou la disponibilité des ressources	Pétrole.
S	Sécuritaire	Causé par l'insécurité présente dans les pays traversés	Corruption. Soulèvement populaire. Terrorisme.

Elles surviennent presque chaque année et peuvent causer des dommages importants. On peut penser à des glissements de terrain, des pertes de maisons, de récoltes, de bétail, et même de vies humaines. En date du 27 août 2018, il y avait déjà 36 décès causés par les inondations cette année là (UNOCHA, 2018). Il est clair que les inondations peuvent entraîner des événements catastrophiques. Les routes de l'Afrique de l'Ouest qui sont considérées belles et pavées ne sont pas nécessairement aussi résistantes aux intempéries ni durables comme celles de la

plupart de pays développés. Des pluies prolongées et en grandes quantités peuvent rapidement les rendre impraticables durant quelques jours (un exemple d'évènement aléatoire). Les pluies abondantes peuvent clairement entraîner des délais sur les routes, mais lorsque la situation est extrême elles peuvent aussi engendrer des pertes de denrées et des coûts pour l'organisation. La situation est aussi difficile en situation de grandes chaleurs et de sécheresse prolongées. Les activités logistiques (manutention, transport et dédouanement) peuvent être ralenties, parce qu'il est plus difficile de travailler à plein régime lors de températures extrêmes. Tous ces éléments, entrant dans la catégorie **Environnementale**, contribuent grandement aux épisodes de famine que traversent les Nigériens. D'autres phénomènes naturels peuvent arriver et entrent aussi dans cette catégorie ; vents violents causant des vagues très hautes dans l'océan, tsunamis, tremblement de terre, etc.

Les logisticiens du PAM basé au Niger sont responsables d'accomplir certaines tâches cruciales à certains moments du processus d'approvisionnement. S'ils sont en retard dans l'accomplissement de leurs tâches, ils peuvent être responsables d'engendrer des retards ou des coûts supplémentaires. Nous avons catégorisé cette source d'incertitude comme étant de type **Interne**. Dans cette catégorie, ce qui nous a semblé le plus important c'est le besoin de faire parvenir la documentation nécessaire au bon moment. Nous avons identifié trois documents essentiels au processus d'approvisionnement ; le certificat de connaissance (*bill of lading* en anglais), les documents de dédouanement, et les instructions de transport (*Land-side Transportation Instruction, LTI*, en anglais). Ces documents doivent tous être transmis à temps et idéalement en avance. Le certificat de connaissance est d'abord transmis aux logisticiens du Niger par les acteurs du PAM en Italie (siège social). Les logisticiens au Niger peuvent recevoir ce document plusieurs mois avant l'arrivée du navire, mais c'est possiblement problématique s'ils le reçoivent avec très peu d'avance ou avec du retard. Ce document de connaissance est en-

suite transmis au transitaire qui est responsable du don lors de son arrivée au port. Comme le transitaire est responsable de la manutention, du dédouanement et du transport, ils doivent aussi recevoir les autres documents (LTI et dédouanements) dans un délai raisonnable (idéalement quelque temps avant l'arrivée du don). Si les documents sont transmis en retard, l'arrivée du don au Niger pourrait être retardée. Si le transitaire doit attendre des documents avant de procéder, elle pourrait demander un dédommagement et ainsi engendrer des coûts supplémentaires pour le PAM. Il en est de même s'il y a un manque à la documentation nécessaire au dédouanement. Les délais et les coûts peuvent, dans un tel cas, augmenter de façon considérable. Les dons subséquents peuvent même être retenus par les douaniers ; un exemple d'évènement catastrophique pour le PAM. Le transitaire peut parfois être responsable d'un manque à la procédure douanière, mais dans cette catégorie on ne considère que les erreurs commises par les acteurs du PAM.

L'incertitude peut aussi provenir d'autres entités. Nous en avons identifié quatre : transporteur maritime, transitaire, port et douane. Ces quatre entités sont quatre sous-catégories de la source d'incertitude nommée **Partie prenante**. Les compagnies de transport maritime peuvent être responsables de la variation dans les délais, les pertes et les coûts associés aux transports en mer. Certaines compagnies sont plus fiables que d'autres ; elles vont faire le transport dans un délai raisonnable et prévisible. Elles peuvent aussi être plus rigoureuses que d'autres sur les procédures de protection des denrées alimentaires et donc engendrer moins de pertes. Les compagnies de transport peuvent ainsi être responsables d'un niveau plus ou moins élevé d'incertitude pour le PAM. Les employés des compagnies peuvent aussi être la cause de l'incertitude (niveau de productivité, erreur humaine, action malveillante, ou autre). Les équipements ou les dispositifs de sécurité utilisés par les compagnies peuvent aussi faire défaut ou être mal entretenus, ce qui peut générer des variations dans les trois paramètres stochastiques étudiés.

L'incertitude s'explique de la même façon pour les transitaires. On doit ajouter la sous-traitance du transitaire comme source d'incertitude dans cette catégorie. Le transitaire pourrait contracter une autre compagnie pour effectuer les opérations logistiques dont il est responsable. Il y a aussi le sur chargement des camions ; une pratique courante chez les transitaires pour sauver des coûts (Nathan Associates Inc., 2013; Teravaninthorn et Raballand, 2009). L'état des routes peut venir aggraver l'état des camions même qu'en Afrique, il peut y avoir pénurie de camions (Nathan Associates Inc., 2013). Dans le cas précis de l'aide alimentaire pour le Niger, ce sont uniquement les transitaires qui sont responsables du transport par camions entre les ports et les entrepôts. Ils sont totalement responsables de la flotte de camion qu'ils utilisent. C'est pourquoi la source d'incertitude liée aux camions pour le transport terrestre fait partie de la sous-catégorie des transitaires.

Le port où arrive le don à destination du Niger peut être la cause de variations dans les délais et même causer des pertes de denrées. Si les installations du port sont insuffisantes pour répondre à la demande et que le nombre de navires augmente dans la zone d'ancrage, on va parler de congestion et nécessairement de délais supplémentaires. Les équipements ou les dispositifs de sécurité du port peuvent faire défaut ou être mal entretenus. Les employés du port peuvent, eux aussi, être la cause de l'incertitude (grève, erreur humaine, action malveillante, ou autre). Les explications sont les mêmes pour les postes douaniers qui doivent être traversés (congestion, équipement, employés, etc.). Les routes, pavées ou non pavées, auraient pu être considérées comme une source d'incertitude en soi. Étant donné que les routes sur les corridors étudiés (entre les ports et les entrepôts au Niger) ont toutes été identifiées comme des routes primaires et donc pavées (LCA, 2018), nous ne l'avons pas fait.

Pour mener à bien les activités d'une organisation, certaines ressources (autres que les ressources humaines) sont nécessaires. Les variations dans le prix et la

disponibilité de ces ressources peuvent être source d'incertitude. En Afrique de l'Ouest, c'est plutôt inhabituel, mais c'est possible qu'il n'y ait plus de pétrole disponible dans les réservoirs. On parle alors de pénurie de pétrole. C'est pourquoi nous avons identifié la catégorie **Ressource** comme source d'incertitude liée aux délais et aux coûts. Lorsqu'il y a pénurie de pétrole, les délais vont automatiquement augmenter, à moins que le transitaire ait fait des réserves. Lorsqu'il y a des hausses majeures dans le prix de cette ressource, le montant payé par le PAM pour le transport des denrées peut devenir inapproprié. Dans un tel cas, il n'est plus rentable pour le transitaire de respecter le contrat. Le transitaire va demander aux responsables du PAM de renégocier le prix du contrat à la hausse. La source de cette augmentation des coûts pour le PAM, c'est le prix de la ressource (et non le transitaire).

Si le PAM vient à être en conflit avec l'un des transitaires qu'il a contractés ou s'il y a eu un manque à la documentation douanière et que la situation ne s'est pas réglée, des procédures légales peuvent être entamées. C'est pourquoi nous avons identifié la source d'incertitude de type **Légale** comme une source pouvant entraîner des événements catastrophiques. Il y a peu de chance que ce genre de procédure ait lieu, mais l'impact est généralement très grand. Les coûts et les délais pour le transport du don pourraient devenir considérables, surtout si les cargaisons subséquentes sont retenues par la douane.

Le niveau de sécurité peut varier grandement d'un pays à un autre et d'une période à une autre en Afrique de l'Ouest. Un don destiné aux bénéficiaires du Niger doit traverser un ou deux pays avant d'entrer au Niger. Les pays régionaux (le Bénin, le Togo, le Ghana, le Burkina Faso ainsi que le Niger) sont source d'incertitude en ce qui a trait au niveau d'insécurité (ou de sécurité) qu'ils présentent. La catégorie de type **Sécuritaire** vient compléter la liste des sources d'incertitude identifiée pour le processus d'approvisionnement en aide alimentaire vers le Niger.

Bon nombre de facteurs d'incertitude entrent dans cette catégorie. Il est connu, et discuté par Nathan Associates Inc. (2013), qu'il y a de nombreux blocages sur les routes (*roadblocks* en anglais) des corridors de l'Afrique de l'Ouest. Les véhicules sont arrêtés pour des paiements informels et des pots-de-vin. La corruption est très répandue, elle ralentit les déplacements et contribue à augmenter les coûts des transitaires. Dans les contrats de transport, entre le PAM et le transitaire, les montants sont ajustés pour tenir compte de ces frais. Les ralentissements sur les routes peuvent aussi être causés par des soulèvements populaires. La population peut descendre dans les rues et protester pour différentes raisons. Des entités (autres que les ports ou les douanes) peuvent faire la grève et ralentir les véhicules sur les routes. De plus, lors de ces blocages et ralentissements, il peut y avoir des vols ou pillages de denrées alimentaires engendrant des pertes pour le PAM. De plus, l'instabilité des pays voisins du Niger apporte son lot d'incertitude. Par exemple, la guerre au Mali contribue aux mouvements de population et aux réfugiés. Les attaques terroristes, actes sporadiques ou activités des groupes régionaux, sont aussi source d'incertitude. Le groupe Boko Haram, basé au Nigéria, cause des mouvements de population entrant au Niger et de l'incertitude marquée dans la ville de Diffa où se trouvent des entrepôts du PAM. Les pirates du Nigéria se déplacent parfois vers le port de Cotonou au Bénin et suscitent de l'incertitude en mer. En fonction de la situation, les sources d'incertitude qui ont été énumérées et entrant dans la catégorie **Sécuritaire**, peuvent entraîner des événements aléatoires ou catastrophiques.

5.1.2 L'effet des sources d'incertitude sur les paramètres stochastiques

Nous avons vu que les sources d'incertitude peuvent affecter les paramètres stochastiques de délais, de coûts et de pertes pour différentes raisons. Nous précisons maintenant, pour chaque paramètre stochastique, quelle source d'incertitude

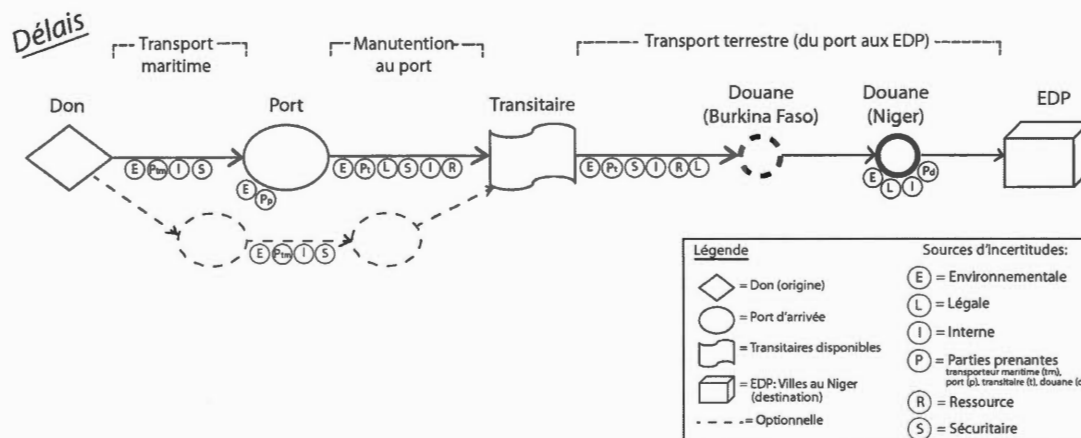


Figure 5.1 Réseau avec sources d'incertitude affectant les délais

identifiée affecte quelle section du réseau logistique. Les trois graphiques qui sont présentés offrent une version agrégée du réseau de la Figure 1.3. Les arcs ont été combinés pour montrer que les trois grandes étapes opérationnelles : le transport maritime, la manutention au port et le transport terrestre. À chaque étape opérationnelle, ce sont les mêmes sources d'incertitude qui sont susceptibles d'avoir un impact, sans égards à leur origine ou leur destination. La section en mer (transport maritime) inclut la période où le navire est ancré dans la zone du port. L'étape de la manutention au port débute lorsque le bateau est à quai et termine lorsque le transport terrestre débute (lors du chargement du camion).

La Figure 5.1 présente les sources d'incertitude qui peuvent avoir un effet sur les délais. On voit que les sources d'incertitude Environnementale, Interne, Sécuritaire et le transporteur maritime (comme Partie prenante) peuvent avoir un impact sur les délais lors du transport en mer. Si le don est séparé entre deux ports d'arrivées, ces mêmes sources d'incertitude peuvent avoir un effet lors du transport entre les ports. En mer, ce ne sont que les sources d'incertitude de type Sécuritaire qui sont susceptibles d'engendrer des conséquences catastrophiques (accident pétrolier et pirates). Dans la zone du port, les sources environ-

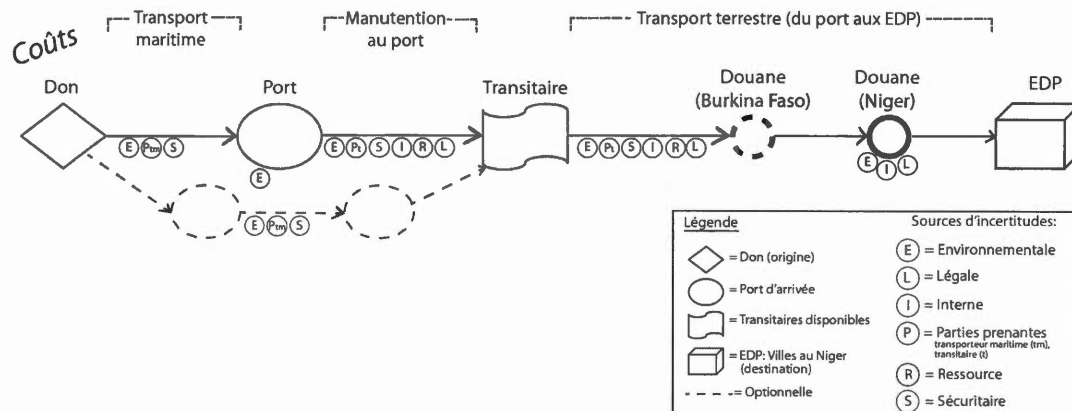


Figure 5.2 Réseau avec sources d'incertitude affectant les coûts

nementales et le port en soi peuvent avoir leurs effets sur les délais. Les effets sur les délais sont généralement aléatoires, mais pourraient être catastrophiques dans le pire des cas. Toutes les six catégories de sources d'incertitude que nous avons identifiées (le transitaire comme seule **Partie prenante**) peuvent avoir un effet sur les délais de manutention au port et de transport terrestre. Plusieurs exemples de sources d'incertitude affectant ces sections du réseau ont été donnés précédemment. Dans la plupart des cas, l'impact qu'ont les événements (aléatoires ou catastrophiques) sur les activités du PAM dépend du niveau des stocks dans les entrepôts au Niger ; de l'urgence des besoins en aide alimentaire à cette période.

La Figure 5.2 précise les sources d'incertitude qui sont susceptibles d'avoir un effet sur les coûts des activités logistiques du PAM. Ces coûts sont détaillés dans les contrats entre le PAM et les compagnies de transport. Lors d'événements jugés majeurs par les parties (route primaire inutilisable, grève au port, hausse marquée du prix du pétrole, etc.) les contrats sont renégociés à la hausse. Les coûts occasionnés par d'éventuelles procédures légales doivent aussi être considérés. On parle donc ici de coûts directs. Les coûts dus aux pertes de denrées ne sont pas inclus, car les effets sur les pertes sont analysés séparément. Sur la section du

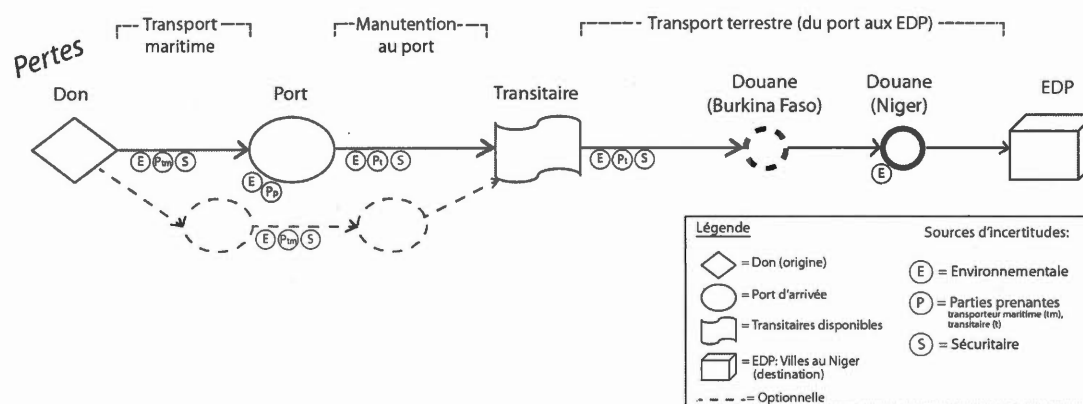


Figure 5.3 Réseau avec sources d'incertitude affectant les pertes

transport maritime, ce sont les mêmes sources d'incertitude qui peuvent affecter les coûts (les mêmes qu'indiqué pour les délais), à l'exception de la source de type **Interne**. Cette source pourrait engendrer de l'attente dans la zone d'ancrage, mais elle ne devrait pas engendrer des coûts sur cette section du réseau (seulement que pour la manutention et le transport terrestre). Le ou les port(s) ainsi que la ou les douane(s) visités durant le transport ne devraient pas non plus être directement responsables de variations dans les coûts (possiblement que sur les délais et les pertes).

Le réseau de la Figure 5.3 indique les sources d'incertitude pouvant affecter les pertes. Les pertes sont en termes de quantité de denrée perdue entre le port d'origine international et les entrepôts du PAM au Niger. Ainsi les pertes peuvent avoir lieu lors du transport maritime, lors de la manutention au port ou durant le transport terrestre. Les catégories de sources d'incertitude de types **Légale**, **Interne** et **Ressource** ne sont pas susceptibles d'affecter les pertes.

5.2 L'analyse des données

Une collecte de données a été effectuée. Nous avons d'abord collecté des données historiques auprès des logisticiens du PAM pour représenter les paramètres stochastiques de l'étude (les variables à prédire au sein des modèles prédictifs). Puis, des données ont été collectées dans le but de comprendre la variabilité des paramètres stochastiques. Ces variables explicatives proviennent soit de l'historique des opérations du PAM ou de bases de données disponibles en ligne. Dans cette section nous présentons et analysons les résultats de ces collectes. Dans un premier temps, les valeurs paramètres stochastiques à prédire (les coûts, les pertes et les délais associés aux activités logistiques du PAM) sont statistiquement analysés. Leurs corrélations, entre elles et entre les principales activités logistiques, sont précisées. Dans un second temps, les variables explicatives sont décrites en précisant si elles ont été collectées pour représenter une ou des sources d'incertitude. Comme nous avons collecté un très grand nombre de variables explicatives, nous présentons une analyse statistique et des représentations graphiques seulement pour un sous-ensemble de variables. Les modifications qui ont dû être faites sur les données sont précisées.

5.2.1 Les paramètres stochastiques à prédire

Les coûts associés à chaque arc du réseau logistique sont importants à analyser, car ils ont un impact direct sur les activités du PAM. Des coûts élevés peuvent avoir des impacts négatifs sur les quantités de denrées acheminées au Niger. Selon les logisticiens du PAM, les coûts varient beaucoup d'un contrat à l'autre. Les facteurs qui expliquent les variations de prix ne sont pas bien connus. Les coûts encourus par le PAM peuvent aussi varier après la signature du contrat. Comme nous l'avons mentionné dans la section précédente, les transitaires peuvent de-

mander de renégocier les prix des contrats à la hausse pour différentes raisons (augmentation du prix du pétrole, grève au port, inaccessibilité des routes, etc.). Pour le transport maritime, il nous a été impossible de récolter des données représentatives des coûts. Principalement, parce que les logisticiens du PAM basés au Niger n'en sont pas responsables. Cette section du transport est sous la responsabilité du siège social du PAM en Italie. Lors d'un don monétaire, ces derniers sélectionnent le fournisseur de denrée et la compagnie maritime responsable du transport en mer. Par contre, les frais de transport ainsi que la logistique sont parfois assumés par les donateurs. C'est généralement le cas pour les dons de produits alimentaires (*in kind*) provenant des États-Unis. Pour ce qui est des coûts reliés à la manutention au port et au transport terrestre, leur historique peut-être collecté via les contrats signés entre les transitaires et le PAM. Nous avons reçu que quelques exemples de ces contrats. On y retrouve les coûts pour le dédouanement au port, la manutention, les formalités douanières à destination, les escortes (au besoin), le transport routier, les frais de service (incluant les rapports de suivi journalier) et les autres frais (au besoin). Les transitaires peuvent préciser ces coûts par type de denrée et par destination. Le PAM sélectionne généralement le transitaire offrant le meilleur prix moyen pour toutes les destinations. Nous ne sommes pas en mesure de dire si les changements effectués (*a posteriori*) aux contrats sont répertoriés par les responsables. Aucune analyse statistique n'a été faite sur les coûts, étant donné un échantillon de données trop petit.

Nous avons reçu des données issues des systèmes d'information du PAM. Nous en avons extrait les informations importantes. Grâce à quatre fichiers Excel (Cotonou 2016, Cotonou 2017, Lomé 2016 et Lomé 2017) nous avons pu créer une base de données associée au transport maritime. Elle comporte 205 observations qui ont été collectées, manuellement, par les logisticiens du PAM. Le premier cas répertorié quitte son origine en décembre 2015, et le dernier arrive en Afrique

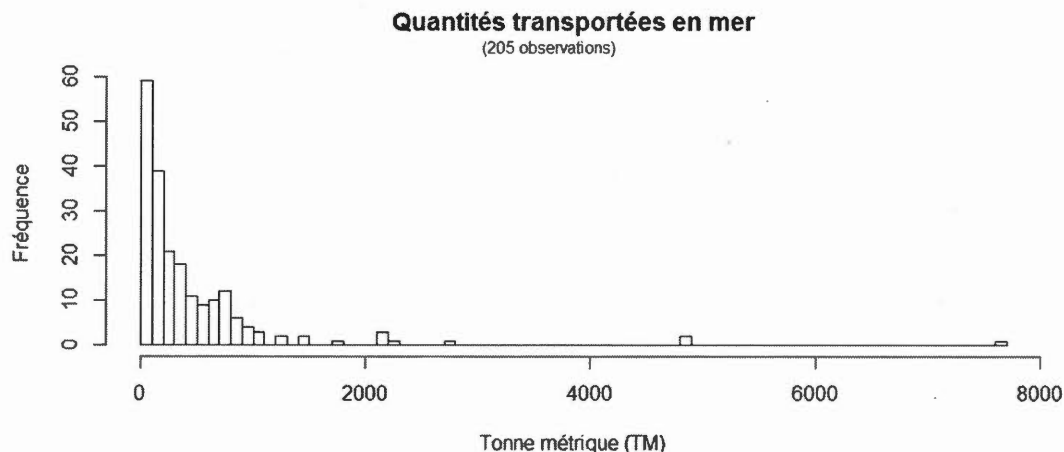


Figure 5.4 Histogramme des quantités transportées lors du transport maritime

de l'Ouest en juillet 2017. Chaque instance représente un don dédié à un certain projet du PAM (P. ex., les opérations d'interventions prolongées de secours et de redressements). La Figure 5.4 présente la distribution empirique des quantités de dons transportées en mer. Elles sont en tonne métrique (MT) et varient entre 1,65 MT et 7 694,85 MT. La médiane est de 217 MT alors que la moyenne est de 454 MT. La distribution des données est asymétrique, il y a un plus grand nombre de dons dont les quantités sont inférieures à la moyenne. La 5.4 montre clairement que les dons de 2 000 MT et plus sont assez rares. Pour calculer les pertes, il faut soustraire, des quantités répertoriées au début de l'activité logistique, les quantités enregistrées à la fin de l'activité (ou au début de l'activité suivante). Les données récoltées pour représenter le transport maritime ne couvrent pas la même période que celles récoltées pour la manutention et le transport terrestre. Il a été impossible pour nous d'associer les quantités reçues par navire avec celles distribuées dans les camions. Les données récoltées ne nous permettent pas de calculer les pertes lors du transport maritime.

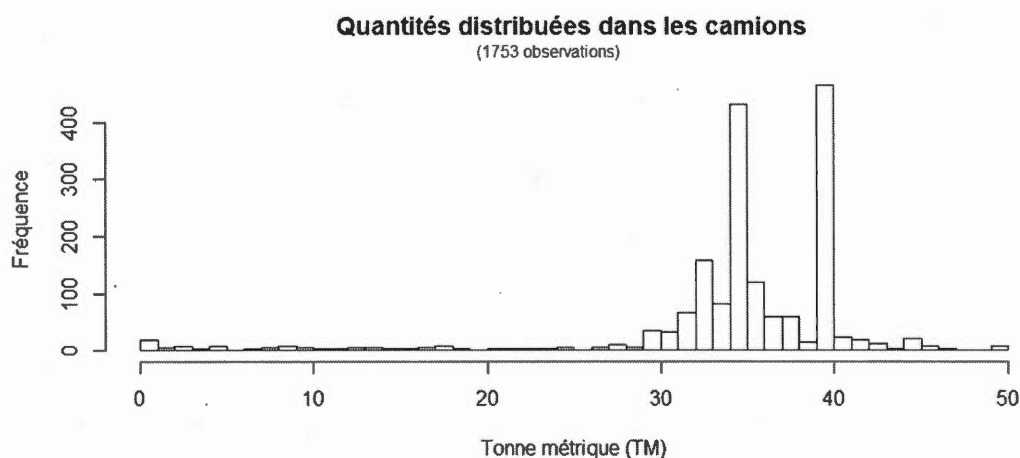


Figure 5.5 Histogramme des quantités réparties dans les camions

Les premières informations issues du système COMPAS datent du mois de mai 2014 et les dernières, d'avril 2016. La base de données qui a été conçue pour représenter la manutention au port et le transport terrestre comporte 1 753 observations. L'histogramme présenté à la Figure 5.5 illustre la distribution empirique des quantités qui ont été réparties (*dispatched*) dans les camions. Généralement, ces quantités se situent entre 29 MT et 45 MT. Étant donné que les cas observés pour le transport maritime ne sont pas les mêmes que ceux des opérations en sol nigérien, il n'a pas été possible de calculer les pertes survenues spécifiquement lors de la manutention. Nous avons été en mesure de soustraire les quantités reçues aux entrepôts du Niger, des quantités réparties dans les camions lors de la manutention. Ainsi, les pertes ayant eu lieu lors du transport par camion ont été calculées grâce aux informations collectées via les fichiers Excels du PAM. Le Tableau 5.2 indique la quantité de cas observés sous sept catégories. Il y a 97 cas parmi les 1 753 observations où la quantité, soit lors de la manutention ou lors de la réception, n'a pas été enregistrée. L'information concernant les pertes n'est donc pas disponible pour ces observations. Dans plus de 68% des cas, il n'y a eu

Tableau 5.2 Analyse des pertes (en TM)

	Aucune information	Négative < 0	Nulle = 0	Très faible] 0; 0,2]	Faible] 0,2; 1]	Moyenne] 1; 8]	Grande] 8; 40]
Qté d'obs	97	47	1198	347	49	7	8
Proportion	5,53%	2,68%	68,34%	19,79%	2,8%	0,4%	0,46%

aucune perte et près de 20% des cas présentent des pertes inférieures à 0,2 MT. Il est difficile d'expliquer pourquoi on constate 47 instances où les pertes calculées sont négatives. Dans ces cas, les quantités réceptionnées au Niger sont enregistrées comme étant plus grandes que celles qui ont quitté le port. C'est probablement attribuable à une mauvaise collecte de données. Le PAM n'a effectivement pas les ressources nécessaires pour faire un suivi rigoureux des quantités distribuées tout au long de sa chaîne logistique. Il y a d'ailleurs huit observations où le camion semble être arrivé au Niger vide ou presque vide. Ces informations n'ont pas été confirmées comme étant exactes par les logisticiens du PAM.

Comme il y avait un nombre considérable de données manquantes ou erronées pour représenter les pertes, nous avons fait le nettoyage des données reçues du PAM en ciblant les délais de transport. Pour le transport maritime, deux observations ont dû être éliminées, car elles présentaient un délai négatif. La date de leur arrivée au port de l'Afrique de l'Ouest était avant la date de leur départ à l'internationale. Nous avons jugé ces observations comme étant erronées. Une fois ces deux cas éliminés, nous avons 205 observations pour représenter le transport maritime. Les Figures 5.6 et 5.7 illustrent les délais de transport selon l'origine et la destination du don. Les délais de transport sont en jours. Les dons proviennent de quatre continents : l'Afrique, l'Amérique du Nord, l'Asie et l'Europe. La Figure 5.6 montre la distribution empirique, par origine, des 86 dons à destination du port de Cotonou au Bénin. Puis, la Figure 5.7 présente les 119 cas à destination du port de Lomé au Togo. Les dons proviennent majoritairement de l'Amérique

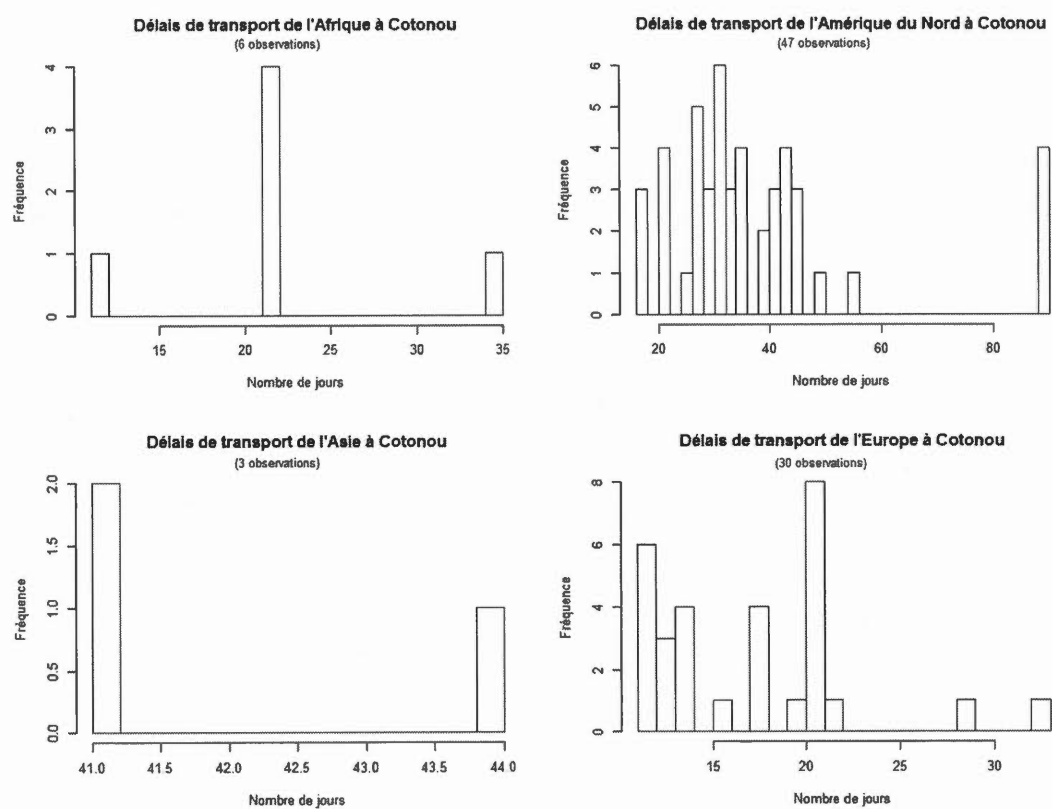


Figure 5.6 Histogrammes des délais de transport maritime du continent d'origine au port de Cotonou au Bénin

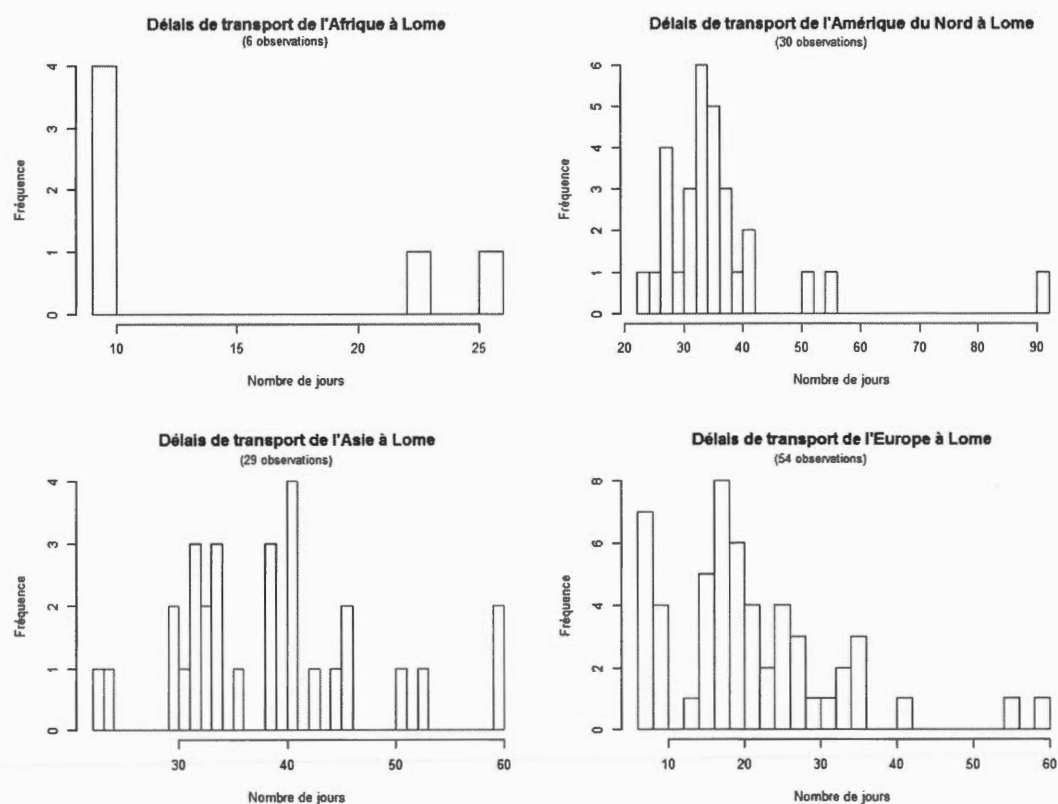


Figure 5.7 Histogrammes des délais de transport maritime du continent d'origine au port de Lomé au Bénin

du Nord et de l'Europe. Les délais pour le transport maritime varient beaucoup. Le cas le plus rapide est de sept jours et le plus long de 91 jours. La moyenne des cas observés est de 29.47 jours et la médiane de 28 jours.

Dans notre base de données présentant les opérations en sol nigérien (manutention et transport terrestre), il y avait aussi deux cas avec des délais opérationnels négatifs qui ont été retirés. Tous les autres cas ont été conservés même s'ils présentaient des valeurs très loin des valeurs jugées normales. Nous avons conservé toutes les valeurs aberrantes (*outliers*), car il n'était pas possible pour nous de confirmer qu'elles étaient erronées. Certaines d'entre elles représentent peut-être des cas qu'on peut qualifier de catastrophiques ou de très risqués pour le PAM. Avec des délais aussi longs, le PAM pourrait de ne pas être en mesure de répondre à temps à la demande des bénéficiaires les plus vulnérables.

Les histogrammes de la Figure 5.8 illustrent la distribution des délais pour effectuer la manutention au port de Cotonou et au port de Lomé. Parmi les données collectées, 1 346 cas ont été observés à Cotonou et 407 à Lomé. Il est très fréquent que la manutention en zone de port soit complétée à l'intérieur d'une journée. Dans ces cas, la valeur pour le délai de manutention est de zéro jour, car le début et la fin de l'opération ont eu lieu le même jour. Les opérations de manutention ont été effectuées le jour même, dans 1,56% des cas observés à Cotonou et dans 9,83% de ceux observés au port de Lomé. Au total, les opérations de manutentions répertoriées ont duré entre zéro et 112 jours. La valeur médiane des observations est de 9 jours alors que la moyenne est de 11,33 jours. La variabilité entre les données est donc très grande et plusieurs cas prennent plus d'un mois. Nous avons identifié 136 valeurs aberrantes parmi les 1 753 cas observés, ce sont tous des délais supérieurs aux délais considérés normaux. Pour les calculer, nous avons utilisé la méthode de Tukey (1977). On additionne à la valeur du troisième quartile (14 jours) le produit de l'écart interquartile (11 jours) avec 1,5. Dans le

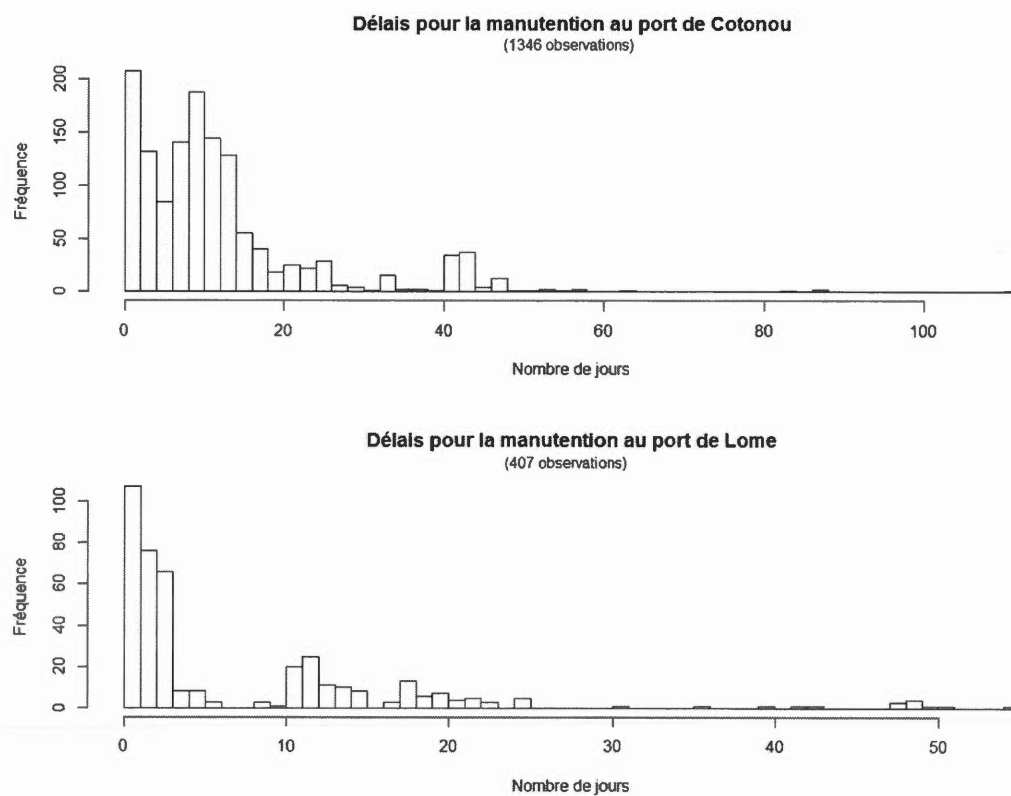


Figure 5.8 Histogrammes des délais de manutention aux ports

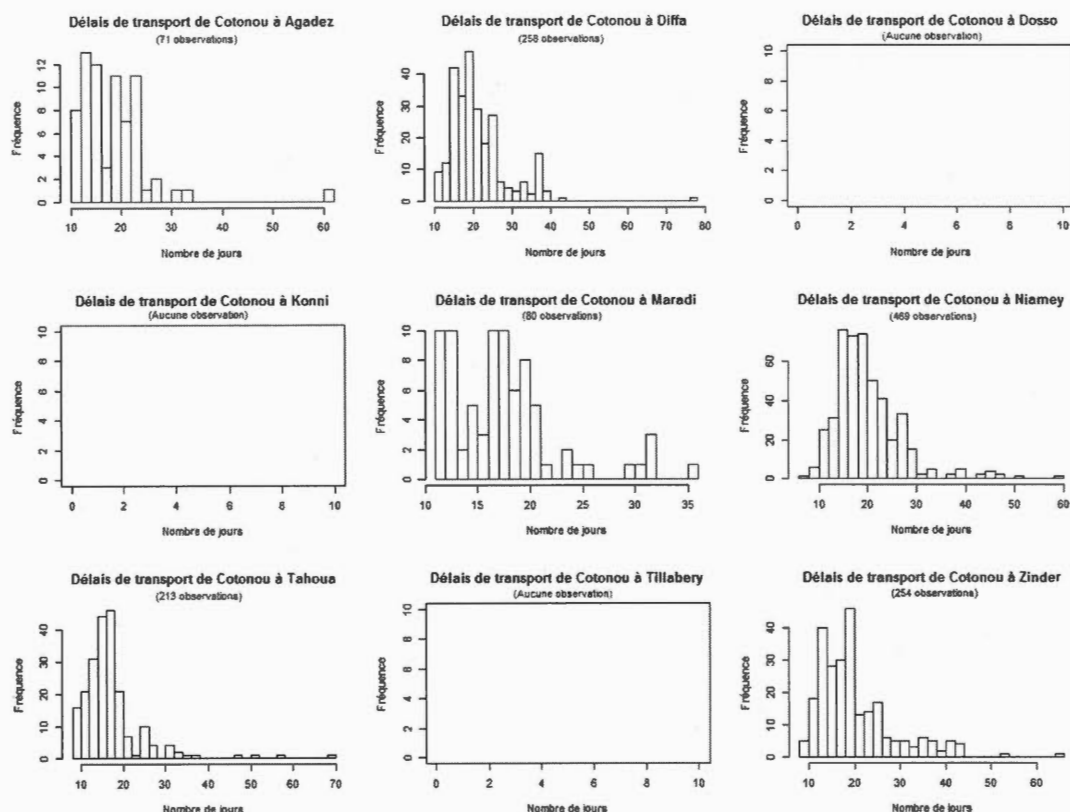


Figure 5.9 Histogrammes des délais de transport terrestre de Cotonou au Niger

cas de la manutention en zone de port le calcul est : $14 + (11 \times 1,5) = 30,5$ jours. Il y a donc 136 observations avec des valeurs supérieures à la borne supérieure de 30,5 jours pour le délai de manutention.

Il y a aussi beaucoup de variabilité dans les délais observés pour le transport par camion. La Figure 5.9 présente les cas répertoriés où l'origine est Cotonou. Il y a neuf histogrammes, un pour chaque destination au Niger (les villes où se trouvent les entrepôts du PAM). Il n'y a eu aucun cas observé où l'origine était le port de Cotonou au Bénin et la destination, les villes de Dosso, Konni ou Tillabery. Les logisticiens nous ont mentionné que les entrepôts de ces trois villes sont offerts, sans frais, par le gouvernement nigérien. Le PAM évite d'envoyer des denrées

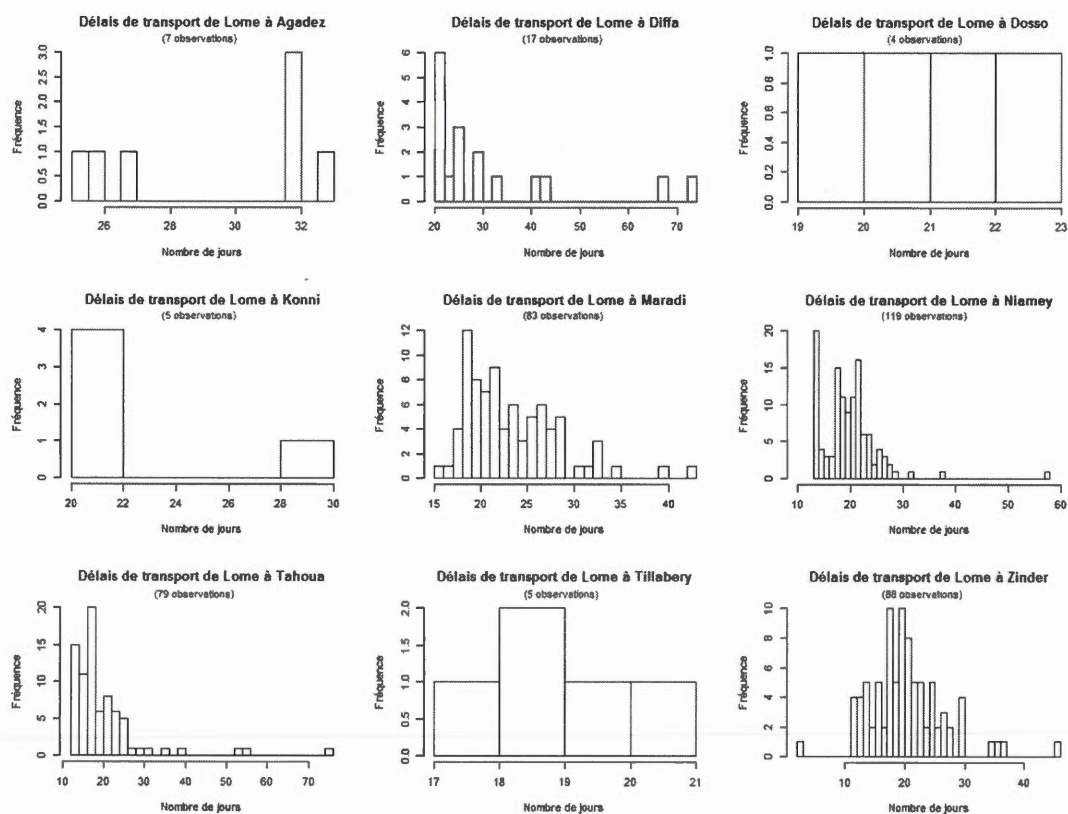


Figure 5.10 Histogrammes des délais de transport terrestre de Lomé au Niger

Tableau 5.3 Statistiques descriptives des paramètres stochastiques

Paramètres stochastiques	Période	Nombre total d'obs.	Min.	1er quartile	Médiane	Moyenne	3e quartile	Max.
LossOnTruck : Perte transport terrestre	MAI-14 à AVR-16	1 753 obs.	-44,8	0	0	-0,1914	0	39,9
DonPort : Délai transport maritime	DEC-15 à JUL-17	205 obs.	7	19	28	29,47	36	91
PortForwarder : Délai manutention	MAI-14 à AVR-16	1 753 obs.	0	3	9	11,33	14	112
ForwarderEDP : Délai transport terrestre	MAI-14 à AVR-16	1 753 obs.	2	16	19	20,43	23	78

directement en provenance de l'international dans ces entrepôts, les achats locaux ou régionaux sont priorisés. La Figure 5.10 montre les neuf histogrammes des délais de transport en provenance du port de Lomé au Togo. Il y a que très peu de cas à destination des trois villes mentionnées. Comme pour la manutention au port, on remarque beaucoup de variabilité dans les délais de transport terrestre et plusieurs observations indiquent des délais très longs. Le cas observé où le délai est le plus court est 2 jours, cette valeur est calculée comme étant une valeur aberrante (inférieure à la borne inférieure). Le second délai le plus court est de 7 jours et le délai observé le plus long est de 78 jours. Le délai médian est de 19 jours et le délai moyen de 20,43 jours. Toutes les valeurs supérieures à 33,5 jours sont des valeurs aberrantes. Il y a 96 observations de ce type parmi toutes les valeurs observées pour le transport terrestre.

Le Tableau 5.3 résume les données qui ont été récoltées pour représenter les paramètres stochastiques. Les paramètres stochastiques sont, dans le cadre des modèles prédictifs, les variables à prédire. Le Tableau 5.3 précise le nom de la variable à prédire, la période couverte par les données collectées, le nombre total d'observations

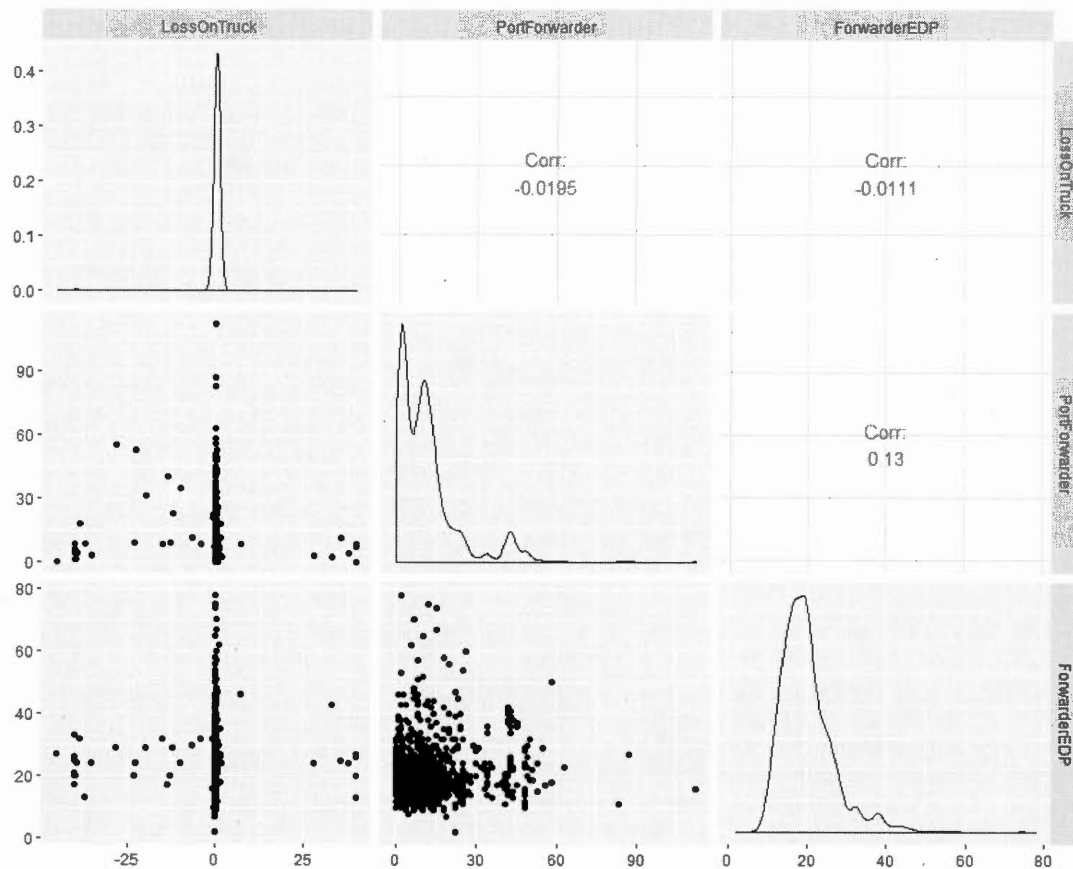


Figure 5.11 Matrice de corrélation des paramètres stochastiques en sol africain

et les statistiques descriptives des valeurs observées pour chaque variable (minimum, 1er quartile, médiane, moyenne, 3e quartile et maximum). Rappelons que les données représentatives du transport maritime ont été récoltées manuellement tandis que les données représentatives des opérations en sol africain proviennent du système d'information COMPAS. Ils ne couvrent pas les mêmes années et ils présentent des cas différents.

Une analyse des corrélations entre les paramètres stochastiques a été faite avec le logiciel R. La matrice de corrélation présentée à la Figure 5.11 a été produite grâce l'utilisation du package GGally (Schloerke *et al.*, 2018). Sur la diagonale

on peut voir la distribution des variables à prédire. La Figure 5.11 inclut aussi des graphiques en nuage de points qui représentent chaque couple de valeurs observées. La matrice indique qu'une très faible corrélation entre les pertes lors du transport par camion et les délais de manutention (-0,0195). Nous arrivons à la même conclusion pour la corrélation entre les pertes lors du transport par camion et le transport terrestre (-0,01110). De plus, nous jugeons que le coefficient de corrélation de 0,13, entre le délai de manutention et de transport terrestre, est aussi très bas. Les données récoltées ne nous permettent pas de tester les corrélations entre le transport maritime et les opérations en sol nigérien. Nous n'avons aucune raison de croire qu'elles seraient corrélées. Rappelons que ce ne sont pas les mêmes entités qui sont responsables des denrées et que l'environnement opérationnel y est très différent. C'est pourquoi nous considérons les paramètres stochastiques de notre étude comme étant indépendants.

Nous avons décidé de ne pas poursuivre notre analyse sur les pertes. Nous avons fait ce choix étant donné que la grande majorité des données collectées présentent des pertes nulles et parce que nous avons un nombre considérable de données manquantes et de données erronées pour représenter ce paramètre stochastique. Nous croyons qu'il serait possible de faire une analyse plus approfondie sur les pertes s'il y avait un système plus rigoureux en place pour faire le suivi des quantités. Pour ce qui est des coûts, il est certain que le PAM se doit de minimiser ses coûts logistiques. Par contre, le PAM à peu de choix, la compétition entre les transitaires est faible en Afrique de l'Ouest. Comme les opérations du PAM au Niger sont idéalement temporaires, le PAM ne peut pas non plus accomplir seul ses opérations logistiques. Puis, dans le contexte de marché dans lequel il opère, le PAM n'est pas en mesure d'implanter des systèmes de pénalités (lorsqu'un contrat n'est pas respecté). Enfin, nous ne sommes pas allés plus loin dans notre analyse des coûts principalement parce que les données récoltées ne nous le permettaient

pas.

Malgré tout, nous croyons que dans le contexte étudié, la meilleure évaluation à faire porte sur les délais. En évaluant les délais, on évalue la qualité des services achetés (pour accomplir les opérations logistiques) par le PAM. Les bénéficiaires les plus vulnérables ont des besoins nutritionnels urgents et les impacts des délais logistiques prolongés peuvent être très graves pour eux. C'est pourquoi nous avons tenté de comprendre et de prédire la variabilité dans les délais pour les activités logistiques (transport maritime, manutention et transport terrestre).

5.2.2 Les variables explicatives

Les variables explicatives, ont été récoltées dans le but de comprendre et expliquer la variabilité dans les valeurs des paramètres stochastiques. Nous avons collecté des données historiques pour expliquer les délais nécessaires à l'accomplissement des opérations logistiques et les sources d'incertitude qui les affectent.

Le Tableau 5.4 résume les variables explicatives qui ont été collectées pour mieux comprendre les délais de transport maritime. Six de ces variables proviennent des données reçues du PAM. La quantité (en TM) transportée, le continent d'origine du don, le mois au cours duquel le don a quitté son origine, le port d'arrivée des denrées, la compagnie maritime responsable du transport et le type de denrées transporté. Les variables **Port** et **CarrierGr** servent à représenter la source d'incertitude nommée **Partie prenante** causé par les entités port et transporteur maritime respectivement. Des modifications ont été faites dans les données pour les variables **CarrierGr** et **CommodityGr**. Par exemple, la catégorie de denrées nommée **MIXCSB** regroupe les données originalement nommées : **CSB**, **CSB++**, **CSB+**, **CSB-SCP** et **Super Cereal (CSB+) With Sugar**. La catégorie **MIXCSB** est l'une des onze catégories de denrée alimentaire (variable **CommodityGr**) réper-

Tableau 5.4 Les variables pour expliquer les délais de transport maritime

Variable	Type	Description	Source d'incertitude représentée
Données du PAM			
MTN	Numérique continue	Quantité, en tonne métrique, de denrées transportées.	-
Continent	Catégorique nominal (4)	Le continent d'origine du don.	-
DepartureMonth	Catégorique nominal (12)	Le mois pendant lequel le don quitte son origine (à l'international).	-
Port	Catégorique nominal (2)	Le port d'arrivée en Afrique de l'Ouest (Cotonou ou Lomé).	Pp
CarrierGr	Catégorique nominal (10)	La compagnie de transport maritime responsable.	Ptm
CommodityGr	Catégorique nominal (11)	Le type de denrée alimentaire transportée.	-
Données publiques			
NauticalMiles	Numérique continue	La distance nautique entre le port d'origine et le port d'arrivée.	-
ExpectedTime	Numérique continue	Le temps de transport estimé entre le port d'origine et le port d'arrivée.	Ptm et E
AnnRevCarrierCo	Numérique continue	Le revenu annuel (estimé) de la compagnie de transport maritime (en M\$ US).	Ptm

toriées dans cette base de données.

Trois variables ont été récoltées en ligne pour expliquer les délais de transport maritime. Ayant l'information sur le port d'origine du don (généralement à l'international) nous avons pu trouver, sur le site Ports.com (2018), la distance nautique entre le port d'origine et le port de destination. En estimant la vitesse du navire à 21 noeuds (vitesse moyenne d'un navire), nous avons aussi obtenu le temps prévu de transport. Pour la majorité des compagnies de transport maritime répertoriées, nous avons été en mesure de trouver leur revenu annuel sur leur site web respectif. Lorsque possible, nous avons utilisé les revenus de l'année 2015. Pour les plus petites compagnies qui ne précisaient pas leurs revenus en ligne, c'est grâce à des sites web qui les décrivaient que nous sommes parvenus à estimer leur revenu annuel. La variable *ExpectedTime* nous sert à représenter la source d'incertitude de type *Environnementale* ainsi qu'à évaluer le transporteur maritime pour la source d'incertitude *Partie prenante*. C'est possible en regardant l'écart entre la valeur observée pour le paramètre (le délai maritime) et le temps estimé (la valeur de la variable explicative). Aucune des données recueillies ne nous permet de représenter la source d'incertitude de type *Interne* ou celle de type *Sécuritaire*. Il pourrait être intéressant de le faire dans le cadre d'une étude future.

Le graphique de la Figure 5.12 montre, avec des boîtes à moustaches, la distribution des délais pour chaque compagnie de transport maritime. Un trait plus foncé indique la médiane de la distribution, alors qu'un losange noir indique sa moyenne. À l'extérieur des moustaches, un cercle indique une valeur aberrante à la manière de Tukey (1977). D'après les cas observés, c'est le *CARRIER6* qui fait, avec une moyenne de 20,82 jours, le transport maritime le plus rapidement. C'est aussi la compagnie pour lequel nous avons le plus grand nombre d'observations. Trois des dix compagnies (*CARRIER3*, *CARRIER5* et *CARRIER6*) ont effectué, à elles seules, 79,5% du transport maritime. La Figure 5.13 donne une idée de

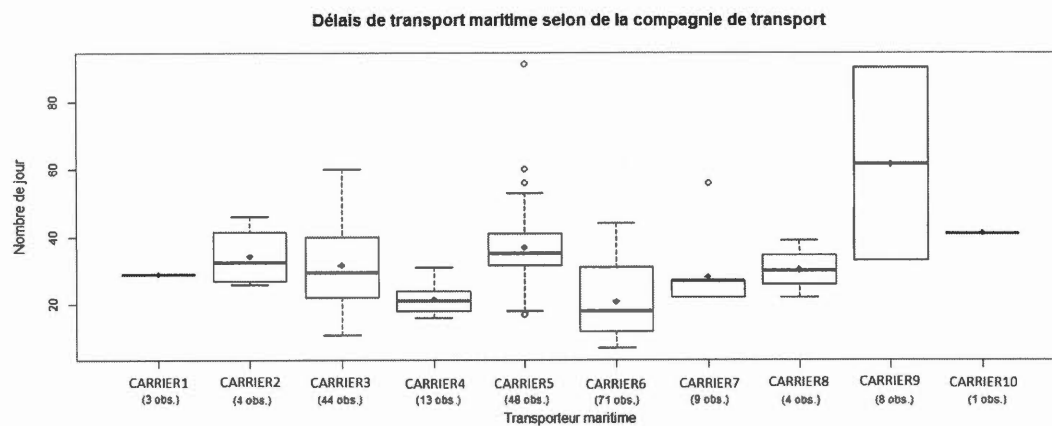


Figure 5.12 Boîtes à moustaches des délais de transport maritime en fonction du transporteur maritime

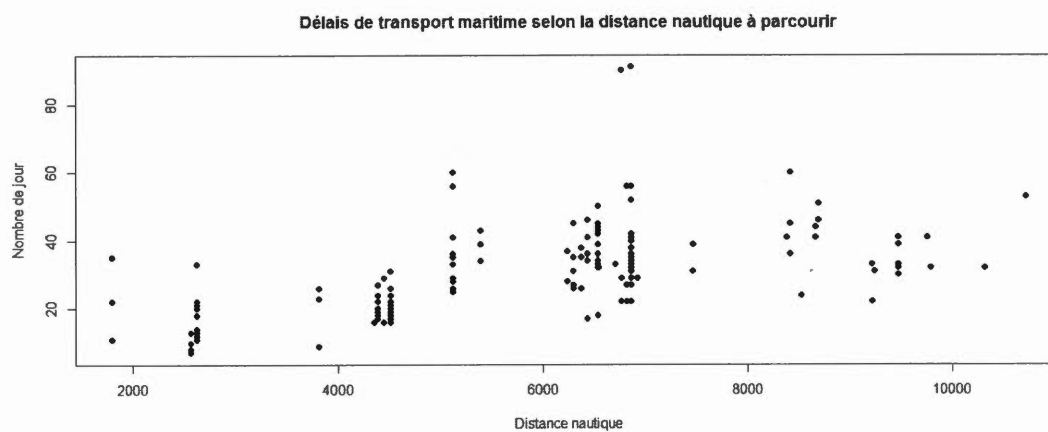


Figure 5.13 Nuage de points des délais de transport maritime en fonction de la distance nautique

la distribution en indiquant, avec un nuage de points, les couples d'observations pour la variable à prédire `DonPort` et la variable explicative `NauticalMiles`. Le graphique permet d'observer une certaine tendance ; lorsque la distance nautique augmente, le délai augmente aussi. Le coefficient de corrélation de 0,5932, entre le délai de transport maritime et la distance nautique, confirme cette corrélation positive. Il y a aussi corrélation positive entre le délai maritime et la quantité transportée (0,343) et entre le délai maritime et le revenu annuel des compagnies de transport maritime (0,2902).

Pour expliquer les délais en sol africain (manutention et transport par camion) nous avons récolté 69 variables provenant des données du PAM ou de bases de données publiques. Elles sont présentées en deux tableaux. Le Tableau 5.5 décrit les variables explicatives ayant été collectées parmi les fichiers reçus du PAM. Il y a le port d'arrivée du navire (où les camions sont chargés), la destination au Niger (où se trouvent les entrepôts du PAM), le transitaire mandaté pour le don, le type de denrées transportées et le mois du chargement des camions. Une variable a été créée pour représenter la source d'incertitude *Interne*. Le retard, en jour, dans l'envoi des documents (certificat de connaissance, dédouanement et LTI) au transitaire est calculé. Lorsque les documents sont envoyés en avance, cette variable prend la valeur de zéro. Il y a deux variables liées à la quantité de dons transportée : `DispatchMT` indique la quantité (en TM) qui a été chargée dans un camion, tandis que la variable `Tt1MTofBL` donne l'information sur la quantité totale du don. Cette dernière variable indique le total de tous les lots répartis dans les camions ayant le même certificat de connaissance. Les deux variables sont corrélées négativement avec le délai de manutention. Le coefficient de corrélation est de -0,19 avec la variable `DispatchMT` et -0,23 avec `Tt1MTofBL`. Plus la quantité est grande, dans les camions ou au total du don, plus le délai de manutention est court. Puis, deux variables ont été conçues pour mieux comprendre les saisons

Tableau 5.5 Les variables pour expliquer les délais en sol africain (partie 1)

Variable	Type	Description	Source d'incertitude représentée
Données du PAM			
LoadPort	Catégorique nominal (2)	Le port d'arrivée en Afrique de l'Ouest (Cotonou ou Lomé).	Pp
Destination	Catégorique nominal (9)	Villes où se trouve les entrepôts (EDP) au Niger.	-
Forwarder	Catégorique nominal (6)	Lee transitaire responsable.	Pt
Commodity	Catégorique nominal (8)	Le type de denrée transportée.	-
Month	Catégorique nominal (10)	Le mois où les camions sont chargés au port de l'Afrique de l'Ouest.	E
DelayDoc	Numérique discrète	Jour de retard pour l'envoi des documents au transitaire ; prends la valeur de 0, sinon.	I
DispatchMT	Numérique continue	Quantité, en tonne métrique, de denrées expédiées par l'instance.	-
TtlMTofBL	Numérique continue	Quantité totale, en tonne métrique, de denrées expédiées pour le <i>Bill of Lading</i> .	-
RainSeason	Catégorique nominal (2)	"oui" si les camions sont chargés au port entre le mois de juin et octobre inclusivement ; "non", sinon.	E
HarvestSeason	Catégorique nominal (2)	"oui" si les camions sont chargés au port entre septembre et décembre inclusivement ; "non", sinon.	E
NBofCustoms	Numérique discrète	Le nombre de postes douaniers à traverser. 2 si le don arrive au port du Bénin ; 3 si Togo.	Pd

et représenter la source d'incertitude *Environnementale*. Ces variables binaires indiquent si le don est traité au port durant la saison des pluies (entre juin et octobre) ou durant la saison des récoltes (entre septembre et décembre). Le nombre de frontières (et de postes douaniers) à traverser pour se rendre au Niger est identifié pour représenter la catégorie d'incertitude *Partie prenante douane*. La variable *NBofCustoms* prend la valeur de deux si le don arrive au port de Cotonou, et trois s'il arrive au port du Togo.

Les autres variables ont été créées grâce à des données historiques récoltées en ligne. Elles sont énumérées et décrites à l'intérieur du Tableau 5.6. D'abord, en utilisant les routes tracées par le Logistics Capacity Assesment (2014) nous avons calculé la distance, en kilomètre, séparant le port d'arrivée de la destination au Niger avec Google Maps (2018). Une analyse de la corrélation entre les délais de transport terrestre et la distance à parcourir, en kilomètre, a été faite. Avec un coefficient de corrélation de 0,131, la distance explique très peu le délai nécessaire pour transporter les denrées alimentaires entre le port d'arrivée en Afrique de l'Ouest et le Niger. Pour essayer de comprendre l'influence du prix du pétrole sur les délais de transport en Afrique et pour représenter la source d'incertitude de type *Ressource*, nous avons créé la variable *OilPrice*. Cette variable prend la valeur du baril de pétrole brut européen (U.S. Energy Information Administration, 2018), en dollar américain, pour le mois pendant lequel le camion a été chargé au port. Afin de comprendre si le niveau de pluie a un impact sur les délais, nous avons identifié la quantité de pluie étant tombée en moyenne durant le mois du chargement du camion. Le niveau de pluie est analysé avec cinq variables distinctes. Il y a la pluie moyenne (en millimètre) tombée au Niger, au port d'entrée (Cotonou ou Bénin selon le cas) et au Burkina Faso. Nous avons aussi une variable pour traduire le niveau de précipitation sur le corridor. *AvRainTtlCor* prend la valeur de la pluie moyenne au Niger additionné à la pluie moyenne au port d'entrée.

Tableau 5.6 Les variables pour expliquer les délais en sol africain (partie 2)

Variable ou groupe de variable	Type	Description	Source d'incertitude représentée
Données publiques			
DistanceKM	Numérique continue	La distance, en kilomètre, entre le port d'arrivée en Afrique de l'Ouest et les EDP.	-
OilPrice	Numérique continue	Prix du pétrole brut, en dollar par baril, le mois du chargement du don au port.	R
AvRain _i <i>i</i> = Niger, Port, BurFaso, TtlCor, Ttl2Mois	Numérique continue	Quantité de pluie moyenne, en millimètre, le mois du chargement du don au port. 5 variables.	E
TtlTrade _i <i>i</i> = Niger, PortEntry, BurFaso	Numérique continue	Valeur totale des échanges (import + export) le mois du chargement du don au port (\$ US). 3 variables.	Pp, Pt, Pd
CPI _i <i>i</i> = Niger, Benin, Togo, BurFaso, AvCorridor	Numérique continue	L'indice de perception de la corruption l'année du chargement du don au port. 5 variables.	S
INFORM _i <i>i</i> = Niger, Benin, Togo, BurFaso, AvCor	Numérique continue	L'indice du risque global l'année du chargement du don au port. 5 variables.	S
Conflict _i <i>i</i> = Niger, Benin, Togo, BurFaso, Ttl, Niger2M, Benin2M, Togo2M, BurFaso2M, Ttl2M	Numérique discrète	Nombre de conflits total (tous types) répertoriés durant le mois du chargement du don au port. Aussi pour deux mois (courant et précédent). 10 variables.	S
Fatal _i <i>i</i> = Niger, Benin, Togo, BurFaso, Ttl, Niger2M, Benin2M, Togo2M, BurFaso2M, Ttl2M	Numérique discrète	Nombre de décès (lié aux conflits de tous types) répertorié durant le mois du chargement du don au port. Aussi pour deux mois (courant et précédent). 10 variables.	S
Conflict _i <i>i</i> = NigerBattle, NigerRemViolence, NigerRiotsProtest, NigerStratDev, NigerViolenceToCivil, BattleTtl, ...	Numérique discrète	Nombre de conflits répertoriés durant le mois du chargement du don au port. Pour chaque type de conflit survenu au Niger et pour chaque type de conflit au total (sur le corridor). 10 variables.	S
Fatal _i <i>i</i> = NigerBattle, NigerRemViolence, NigerRiotsProtest, NigerStratDev, NigerViolenceToCivil, BattleTtl, ...	Numérique discrète	Nombre de décès répertoriés durant le mois du chargement du don au port. Pour chaque type de conflit survenu au Niger et pour chaque type de conflit au total (sur le corridor). 10 variables.	S

Puis, la variable `AvRainTtl2Mois` indique le total des précipitations du corridor pour le mois courant (`AvRainTtl1Cor`) et le mois précédent. Cette dernière variable a été créée dans le but de voir si l'accumulation de pluie (sur une période de deux mois) avait un impact sur les opérations. L'information donnée par ces cinq variables est issue des données collectées sur le site du The World Bank Group (2018). Les données mensuelles sur les précipitations étaient disponibles de 1901 à 2015. Pour ne pas avoir de données manquantes, nous avons dû estimer les valeurs de ces variables pour les quatre premiers mois de l'année 2016. Nous avons utilisé la moyenne historique des données disponibles (de 1901 à 2015) pour les quatre mois où les données étaient manquantes.

Il était important de collecter des données pour traduire la congestion (aux ports et aux postes douaniers). Nous avons décidé d'utiliser les niveaux d'échanges commerciaux (importations et exportations) dans les pays étudiés comme variables proxies. Nous avons trouvé des données mensuelles appropriées sur le site International Trade Center (2018). Ces données permettent aussi de représenter le niveau de demande pour les services des transitaires disponibles aux ports. Nous savons qu'il y a peu de compétition entre les transitaires. Lorsque la demande augmente, leur capacité à remplir les contrats signés, dans les délais prévus, peut devenir difficile. Ainsi, nous avons créé trois variables explicatives pour représenter les échanges commerciaux et donc les sources d'incertitude des **Parties prenantes** port, transitaire et douane. Nous avons fait le total mensuel des importations et des exportations (en dollars américains) du Niger, du port d'entrée et du Burkina Faso. Dans notre base de données nous utilisons la valeur totale des échanges le mois où le camion a été chargé au port. Les échanges commerciaux du Niger sont corrélés négativement avec les délais de manutention (-0,30). Quand il y a plus d'importation et d'exportation au Niger, les délais pour la manutention en zone de port tendent à être plus courts.

Pour représenter la source d'incertitude de type **Sécuritaire**, nous avons collecté de nombreuses variables. Les deux premiers groupes de variables sont les indices de la perception de la corruption (Transparency international, 2018) et les indices de la gestion des risques (Index for risk management, 2018). Les deux types d'indices sont disponibles sur une base annuelle. L'indice global du risque d'un pays est évalué selon trois critères : le niveau de danger et d'exposition (*Hazard and Exposure*, en anglais) ; le niveau de vulnérabilité (*Vulnerability*) ; et le manque de capacité d'adaptation (*Lack of Coping Capacity*). Chaque catégorie se divise en sous-catégorie et ensuite en sous-composantes. Pour obtenir l'indice global d'un pays, 53 points sont évalués au total. L'indice de la perception de la corruption est d'ailleurs inclus à l'analyse de l'indice de la gestion du risque. Dans les deux cas, nous avons utilisé leur valeur pour l'année pendant lequel le don a été traité au port. Nous avons créé cinq variables par indice. Il y en a une pour le Niger, le Bénin, le Togo, le Burkina Faso et une qui calcule la moyenne des indices des pays visités le long du corridor. Pour cette dernière variable, on fait la moyenne de l'indice du Bénin et du Niger lorsque le don arrive à Cotonou. Lorsqu'il arrive au port de Lomé, ce sont les indices du Togo, du Burkina Faso et du Niger qui sont utilisés.

Toutes les autres variables explicatives ont été créées grâce aux données récoltées sur le site du Humanitarian Data Exchange (2018). Des bases de données sur les conflits armés y sont disponibles pour les pays en voie de développement. Les données sont collectées, en temps réel, par des experts internationaux. Les événements, violents ou non violents, sont répertoriés en spécifiant, entre autres, la date, le lieu, les groupes impliqués, le type d'événement et le nombre de décès survenues (s'il y a lieu). Les événements et les mortalités qu'ils ont causées sont répertoriés qu'une seule fois.

À partir de ces bases de données, nous avons calculé le nombre de conflits total

ayant eu lieu, par mois, pour chaque pays étudié. Nous avons créé cinq variables `ConflitNiger`, `ConflitBenin`, `ConflitTogo`, `ConflitBurFaso` et `ConflitTtl`. La dernière variable calcule le total des conflits répertoriés dans les pays présents sur le corridor. Puis, cinq variables présentent le total des conflits pour deux mois. Pour les calculer, nous avons utilisé le nombre de conflits ayant eu lieu le mois du chargement des camions et nous les avons additionnés à ceux ayant eu lieu le mois précédent. Avec la même logique, nous avons créé dix variables pour représenter le nombre total de mortalités liées aux événements du mois. Nous avons aussi voulu savoir si certains types d'événements étaient susceptibles d'avoir un impact plus grand sur les délais. Ainsi nous avons discrétisé les variables par type d'évènement. Nous avons créé cinq nouvelles variables pour le Niger (`NigerBattle`, `NigerRemViolence`, `NigerRiotsProtest`, `NigerViolenceToCivil` et `NigerStratDev`). Elles représentent les affrontements, les violences perpétrées à distance, les protestations et émeutes, les violences envers les civils et les événements de développement stratégique. Dans cette dernière catégorie, on retrouve, entre autres, les grèves, vols et incursions. Cinq variables représentant les types d'événements ont aussi été créées pour calculer le nombre total d'événements ayant eu lieu dans les pays du corridor. Enfin, avec la même logique, dix variables ont été créées pour représenter le nombre total de mortalités liées aux cinq types d'événements. Il y a cinq variables de ce type pour le Niger et cinq pour le total du corridor.

Les deux graphiques illustrés à la Figure 5.14 montrent les distributions des délais, pour la manutention et le transport par camion, selon le mois de chargement des camions. Aucun don, parmi les données historiques récoltées, n'a été traité en zone de port durant les mois de novembre et de décembre. On remarque qu'en moyenne, les délais pour effectuer les opérations de manutention et de transport par camion sont plus longs en avril. D'après l'historique des données récoltées,

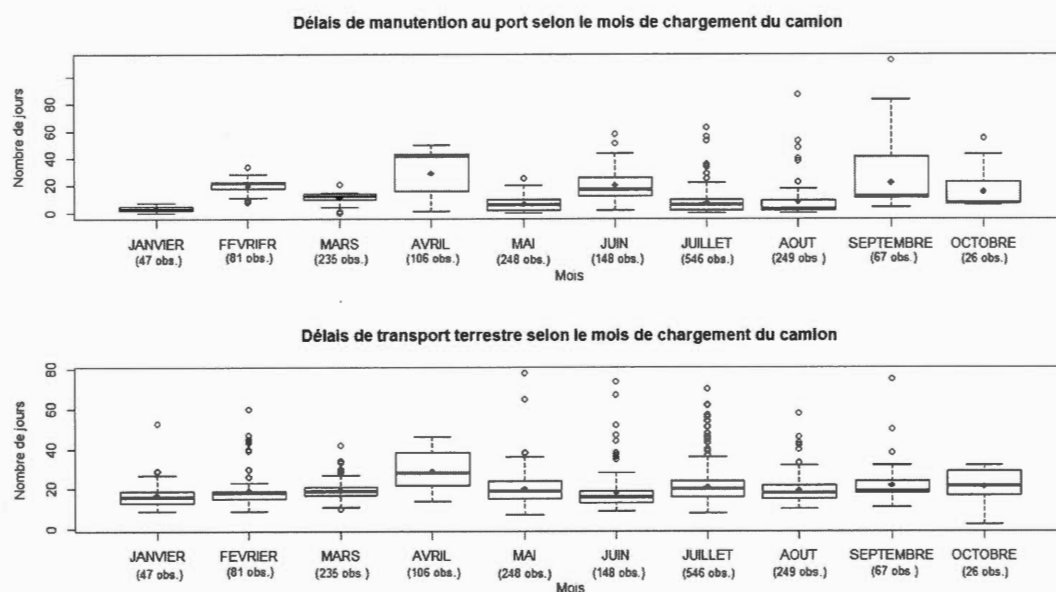


Figure 5.14 Boîtes à moustaches des délais en sol nigérien (manutention et transport) en fonction du mois

c'est en juillet que le PAM a effectué le plus grand nombre de chargements au port. Durant la période des récoltes (de septembre à décembre) beaucoup moins de dons sont acheminés vers le Niger.

La Figure 5.15 montre les distributions empiriques des délais, pour la manutention et pour le transport terrestre, de chaque transitaire répertorié. Pour les délais de manutention, on peut voir qu'historiquement certaines compagnies ont été plus fiables que d'autres. Par exemple, l'étendue de la distribution des délais de la compagnie Bénin3 est moins grande que celle de Bénin2 ou de Bénin4. La compagnie Bénin4 a été responsable d'un plus grand nombre de dons et la distribution de ses délais historiques (de manutention et de transport) comporte bon nombre de valeurs anormalement élevées. Au niveau des délais de transport terrestre, il est plus difficile d'affirmer qu'une compagnie est plus performante ou fiable qu'une

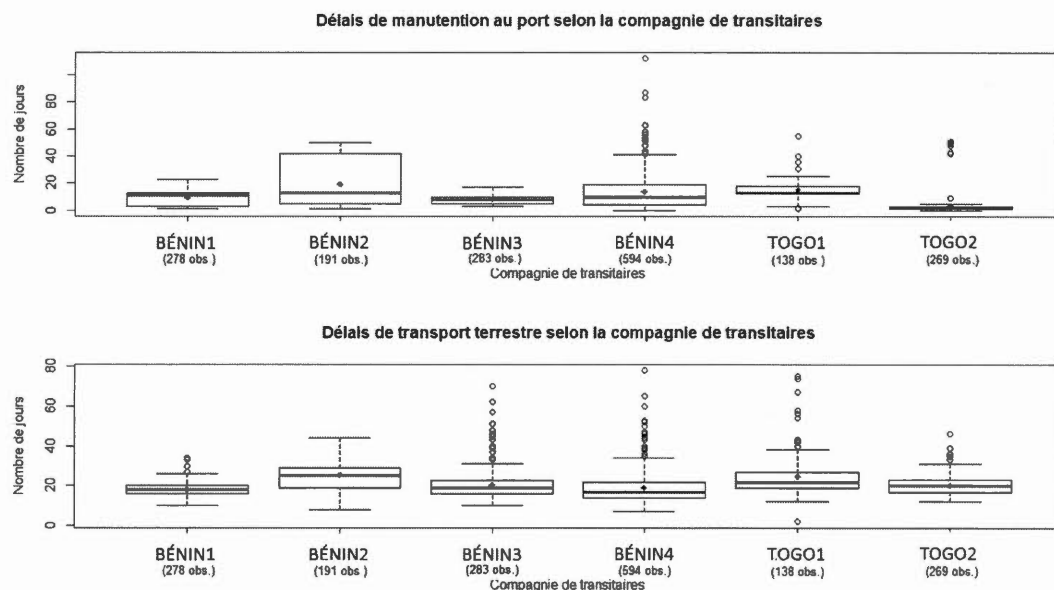


Figure 5.15 Boîtes à moustaches des délais en sol nigérien (manutention et transport) en fonction du transitaire responsable

autre.

Plusieurs données publiques ont été collectées comme des mesures proxies pour représenter les sources d'incertitude identifiées. Mise à part la source d'incertitude **Légale**, les données récoltées nous permettent d'analyser l'impact des cinq autres sources d'incertitude identifiées (**Environnementale**, **Interne**, **Parties prenantes**, **Ressource** et **Sécuritaire**) sur les délais en sol africain. Pour la source **Légale**, nous souhaitons avoir accès à l'historique des litiges avec les transitaires ainsi qu'à l'historique des pénalités liées aux cargaisons retenues aux ports ou aux douanes. Nous l'avons cru possible, car Jahre *et al.* (2016) ont eu accès à l'historique des réclamations d'assurance du HCR. Malheureusement, les responsables du PAM n'ont pas été en mesure de nous transmettre ce type de données. Nous croyons qu'il serait valable d'analyser les sources d'incertitude **Légale** dans

le cadre d'un autre projet de recherche.

5.3 Les modèles prédictifs

Dans le cas où les données le permettraient, des modèles prédictifs pourraient être développés pour estimer les coûts et les pertes reliés aux activités logistiques. Dans le cadre de cette recherche, des modèles prédictifs ont été élaborés pour estimer les délais de transport maritime, les délais de manutention au port, et les délais de transport par camion. D'abord, les résultats et les erreurs de prédiction des modèles de régression linéaire, d'arbre de décision et de forêt aléatoire sont comparés avec la moyenne estimée. Ensuite, le modèle ayant le plus grand potentiel prédictif est amélioré. Nous expliquons le choix des paramètres du modèle et analysons ses différentes versions.

5.3.1 Comparaison des modèles et des erreurs de prédiction

Pour prédire les délais de transport maritime (la variable `DonPort`) nous avons utilisé toutes les variables explicatives présentées au Tableau 5.4. Nous avons partitionné les 205 observations de notre base de données en un ensemble d'entraînement (de 70%) et un ensemble de validation (de 30%). Nous l'avons fait dix fois afin de comparer les résultats de dix échantillons tirés aléatoirement en assumant que chaque observation avait la même probabilité d'être pigée. Les erreurs de prédiction MAE, MAPE et RMSE ont été calculées, sur les dix échantillons, pour les quatre modèles prédictifs (Moyenne estimée, Régression linéaire, Arbre de décision et Forêt aléatoire). Le minimum, le maximum, la moyenne et l'écart-type des trois mesures d'erreur sont présentés dans le Tableau 5.7.

Pour calculer les résultats de la moyenne estimée, nous avons fait la moyenne de la variable `DonPort` pour les observations incluses à l'ensemble d'entraînement.

Tableau 5.7 Comparaison des modèles prédictifs pour les délais de transport maritime. Sur l'ensemble de validation (30%) de dix échantillons.

Mesure d'erreur		Moyenne Estimée	Regression linéaire *	Arbre de décision	Forêt aléatoire
RMSE	Minimum :	12,09	9,22	8,31	7,61
	Maximum :	18,30	12,44	13,16	11,68
	Moyenne :	14,80	11,08	10,97	9,16
	Écart-type :	1,82	1,20	1,65	1,31
MAE	Minimum :	9,60	6,41	5,59	5,35
	Maximum :	13,29	9,25	9,02	7,21
	Moyenne :	10,96	7,72	7,43	6,22
	Écart-type :	1,02	0,66	1,03	0,62
MAPE	Minimum :	41,03	25,26	20,76	18,24
	Maximum :	72,35	37,62	32,37	26,70
	Moyenne :	54,22	30,68	27,43	22,86
	Écart-type :	10,03	4,26	3,64	2,94
Précision : * Noter qu'il n'y a aucune prédiction possible pour le 3 ^e , 4 ^e , 5 ^e et 6 ^e échantillon. Les résultats de la régression linéaire sont donc basés sur six échantillons seulement.					

Par exemple, dans le premier échantillon, le modèle prédit 30,35664 jours pour effectuer le transport maritime. Ce même délai est prédit pour les 62 observations de l'ensemble de validation. Rappelons que le RMSE, calculé par l'équation 4.9, pénalise plus fortement les plus grandes erreurs de prédiction que le MAE, calculé par l'équation 4.6. Les résultats pour ces deux calculs sont donnés en jours. Tandis que les résultats de MAPE sont donnés en pourcentage. Lorsque la moyenne est utilisée comme estimée, l'erreur de RMSE est de 14,8 jours en moyenne, celle de MAE est de 10,96 jours en moyenne et le MAPE moyen est de 54,22% (relativement aux vraies valeurs).

En faisant l'estimation du délai maritime avec la régression linéaire, les résultats sont améliorés. Le RMSE est amélioré de près de six jours et le MAE de plus de trois jours. Alors que le MAPE est amélioré de 23,54%. Le coefficient de détermi-

nation est calculé sur l'ensemble d'entraînement. Sur les dix échantillons, le R^2 est de 0,68 en moyenne et le R_{adj}^2 de 0,56. Le R_{adj}^2 est ajusté selon le nombre de variables explicatives utilisées dans le modèle. La régression linéaire ne permet pas de faire une prédiction, sur l'ensemble de validation, lorsque cet ensemble comporte une nouvelle catégorie pour une variable catégorique. Par exemple, dans notre 3^e échantillon, il y avait une compagnie de transport maritime (variable `CarrierGr`) dans l'ensemble de validation qui n'était pas présente dans l'ensemble d'entraînement. Comme le modèle n'a pas appris avec cette catégorie, il n'est pas en mesure de faire une prédiction en sa présence. Parmi les dix échantillons, il y en a quatre pour lesquels le modèle n'a pas pu faire de prédiction. C'est, un point faible du modèle de régression linéaire. Pour chaque échantillon, un arbre de décision a été conçu à partir des données de l'ensemble d'entraînement. Dans les dix cas, la première séparation a été faite à partir de la variable `NauticalMiles`. Dans tous les cas, la condition était `NauticalMiles < 4 816`. Si la condition était respectée, la prédiction était de 17 ou 18 jours, dans le cas contraire elle se situait entre 37 et 39 jours. La performance prédictive du modèle d'arbre de décision est tout juste supérieure à celle de la régression linéaire. C'est le modèle de forêt aléatoire qui présente les meilleurs résultats pour les trois métriques. Pour l'estimation du délai de transport maritime, l'erreur moyenne se situe entre 6,22 jours (MAE) et 9,16 jours (RMSE). Par rapport aux vraies valeurs, l'erreur commise est de 22,86% en moyenne. C'est une amélioration de plus de 30% par rapport au modèle de référence. En moyenne, le pourcentage de variation expliqué par le modèle de forêt aléatoire est de 57,16% (calculé par l'équation 4.4).

Les modèles ont été utilisés avec leurs paramètres par défaut. Pour l'arbre de décision, conçu avec la fonction `rpart()` de R, le CP est de 0,01. Pour la forêt aléatoire, conçu avec la fonction `randomForest()`, le nombre d'arbres est déterminé à 500 et le nombre de variables explicatives considérées à chaque séparation est

de $p/3$, soit de trois variables explicatives étant donné le total de neuf ($p = 9$). Nous croyons que les résultats des trois modèles développés (régression linéaire, arbre de décision et forêt aléatoire), pour prédire les délais de transport maritime, pourraient être améliorés. La date estimée de l'arrivée du navire (le ETA) pourrait être ajoutée comme une variable explicative. Il n'a pas été possible pour nous de l'inclure, car nous n'avons pas cette information. Elle figure, entre autres, sur les contrats signés entre le PAM et le transitaire.

Pour estimer le délai de manutention en zone de port, nous avons aussi utilisé dix échantillons. Les 1 753 observations ont été partitionnées sous deux ensembles (entraînement avec 70% des données et validation avec 30%) et les erreurs de prédiction ont été calculées par les trois métriques (RMSE, MAE et MAPE). Comme il y a plusieurs observations pour lesquelles le délai de manutention prend la valeur de zéro jour, le MAPE ne peut pas être calculé (impossible de diviser par zéro). Les modèles prédictifs ont été développés de la même manière. Le paramètre du CP est de 0,01 dans l'arbre de décision. La forêt aléatoire compte 500 arbres et le nombre de variables explicatives utilisées pour effectuer chaque séparation dans ces arbres est de 23 ($p/3$ où $p = 69$). Les résultats des RMSE et MAE sont présentés à l'intérieur du Tableau 5.8. Les résultats de la régression linéaire sont à nouveau très similaires à ceux de l'arbre de décision. Sur l'ensemble d'entraînement, le R^2 est de 0,81, en moyenne et le R^2_{adj} de 0,80. La régression linéaire n'a pas pu faire de prédiction sur deux des dix échantillons. C'est la variable **Commodity** qui a posé problème avec une nouvelle catégorie dans l'ensemble de validation. En moyenne, le modèle de forêt aléatoire domine. Les erreurs de prédictions faites par le MAE sont d'environ deux jours et celles par le RMSE de moins de quatre jours et demi. Le modèle de forêt aléatoire améliore, en moyenne, les estimations de plus de six jours et demi par rapport au modèle de référence. En moyenne, le R^2 calculé par la fonction *randomForest()* est de 87%.

Tableau 5.8 Comparaison des modèles prédictifs pour les délais de manutention au port. Sur l'ensemble de validation (30%) de dix échantillons.

Mesure d'erreur		Moyenne Estimée	Regression linéaire *	Arbre de décision	Forêt aléatoire
RMSE	Minimum :	11,21	5,27	5,55	3,67
	Maximum :	13,23	8,00	7,57	5,65
	Moyenne :	12,04	5,94	6,27	4,45
	Écart-type :	0,62	0,91	0,58	0,67
MAE	Minimum :	7,60	3,42	3,20	1,87
	Maximum :	8,53	3,93	3,68	2,20
	Moyenne :	8,08	3,58	3,51	2,01
	Écart-type :	0,34	0,17	0,16	0,11
Précision :		* Noter qu'il n'y a aucune prédiction possible pour le 4 ^e et 10 ^e échantillon. Les résultats de la régression linéaire sont donc basés sur huit échantillons seulement.			

Tableau 5.9 Comparaison des modèles prédictifs pour les délais de transport terrestre. Sur l'ensemble de validation (30%) de dix échantillons.

Mesure d'erreur		Moyenne Estimée	Regression linéaire *	Arbre de décision	Forêt aléatoire
RMSE	Minimum	6,82	6,14	6,05	5,86
	Maximum	7,89	9,71	7,39	7,22
	Moyenne	7,51	7,15	6,90	6,58
	Écart-type	0,40	1,11	0,43	0,44
MAE	Minimum	4,96	4,29	4,30	3,91
	Maximum	5,39	4,99	4,78	4,56
	Moyenne	5,23	4,52	4,63	4,19
	Écart-type	0,16	0,22	0,15	0,23
MAPE	Minimum	25,61	21,72	22,68	19,97
	Maximum	29,69	25,68	25,75	24,42
	Moyenne	27,87	23,82	24,30	21,89
	Écart-type	1,30	1,33	1,03	1,69
Précision :		* Noter qu'il n'y a aucune prédiction possible pour le 4 ^e et 10 ^e échantillon. Les résultats de la régression linéaire sont donc basés sur huit échantillons seulement.			

Tableau 5.10 Comparaison des modèles prédictifs pour les délais en sol africain (manutention et transport). Sur l'ensemble de validation (30%) de dix échantillons.

Mesure d'erreur		Moyenne Estimée	Regression linéaire *	Arbre de décision	Forêt aléatoire
RMSE	Minimum :	13,62	8,19	8,83	7,15
	Maximum :	16,06	9,26	10,84	8,85
	Moyenne :	14,98	8,84	9,85	8,03
	Écart-type :	0,74	0,37	0,68	0,57
MAE	Minimum :	9,87	5,69	6,27	4,78
	Maximum :	11,32	6,11	7,29	5,55
	Moyenne :	10,63	5,96	6,70	5,06
	Écart-type :	0,52	0,15	0,29	0,25
MAPE	Minimum :	33,33	18,79	21,67	15,90
	Maximum :	38,68	19,99	24,95	18,34
	Moyenne :	36,73	19,47	22,49	16,78
	Écart-type :	1,79	0,52	0,97	0,73
Précision : * Noter qu'il n'y a aucune prédiction possible pour le 4 ^e et 10 ^e échantillon. Les résultats de la régression linéaire sont donc basés sur huit échantillons seulement.					

Nous avons obtenu de moins bons résultats pour prédire les délais de transport par camion. D'ailleurs les analyses de corrélation qui ont été faites, entre les délais de transport terrestre et les variables explicatives, étaient peu concluantes. La plupart des variables explicatives étaient corrélées au délai de manutention, mais pas aux délais de transport par camion. Le Tableau 5.9 présente les résultats des modèles pour prédire les délais de transport terrestre. Comparativement à une estimation basée sur la moyenne historique, bâtir un modèle de forêt aléatoire améliore les erreurs RMSE et MAE d'un jour seulement. En pourcentage, le meilleur modèle améliore les erreurs de six points seulement. Les mauvais résultats des modèles sont d'autant plus prouvés par les coefficients de détermination. Le R^2 prend la valeur de 0,27, en moyenne, selon le modèle de régression linéaire et de 0,2447 d'après le modèle de forêt aléatoire.

Les opérations de manutention au port et de transport par camion sont faites de façon successive. Ainsi, pour estimer le délai total en Afrique, l'erreur de prédiction du délai de manutention au port doit être additionnée à l'erreur de prédiction du délai de transport. En additionnant la moyenne des RMSE pour les deux meilleures modèles (forêt aléatoire pour manutention et transport) on obtient une erreur de 11 jours. Nous avons décidé de comparer les résultats obtenus par les deux modèles (désagrégés) avec les résultats pour un modèle prédisant le délai total en sol africain. Ainsi, nous avons créé une nouvelle variable à prédire. Cette variable nommée **PortEDP** représente la somme, pour chaque observation, du délai de manutention avec le délai de transport terrestre. Les valeurs des variables **PortForwarder** et **ForwarderEDP** sont additionnées.

Les résultats du modèle prédictif agrégé des délais en sol africain sont inscrits au Tableau 5.10. Le meilleur modèle, parmi les modèles considérés, est toujours celui de la forêt aléatoire. Avec environ huit jours comme erreur de RMSE, nous croyons que ce modèle agrégé a un meilleur pouvoir prédictif que la version désagrégée qui a été présentée. Dans cette version agrégée, l'erreur calculée par le MAE est plus de 50% moindre qu'en estimant avec la moyenne et le MAPE moyen est amélioré de près de 20%. La Figure 5.16 illustre l'arbre de régression conçu à partir des 1 227 observations de l'ensemble d'entraînement du 1er échantillon. 1 089 observations respectent la première condition; le nombre de conflits total ayant eu lieu le mois où le camion a été chargé au port est de moins de 12. D'autres règles, basées sur des points de rupture, doivent être considérées avant d'arriver à des prédictions. Les prédictions, en jours, sont indiquées à l'intérieur des noeuds grisés. Dans tous les cas, la prédiction est basée sur un minimum de 7 observations. Plusieurs variables explicatives ne sont pas utilisées par cet arbre. Un modèle de forêt aléatoire contient, par défaut, 500 arbres de décision similaire à celui de la Figure 5.16. Le pourcentage moyen de la variation de la variable à

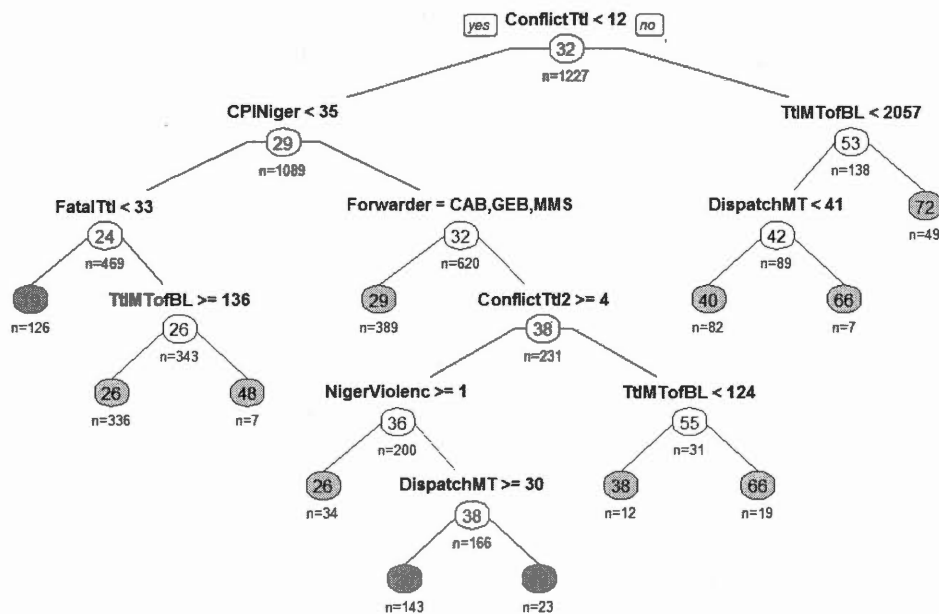


Figure 5.16 Exemple d'arbre de régression pour prédire le délai total en sol africain

prédire (**DonPort**) expliqué par le modèle de forêt aléatoire est de 71,38%. C'est un résultat que nous considérons satisfaisant étant donné la difficulté de prédire les délais de transport en Afrique de l'Ouest. Un avantage additionnel du modèle agrégé est qu'il permet d'obtenir des résultats prédictifs pour le MAPE, car au minimum, le délai total en sol africain est de dix jours. Pour toutes ces raisons, nous avons décidé de poursuivre nos analyses prédictives, avec le modèle agrégé, du délai total en sol africain.

À la fois pour estimer le délai de transport maritime et le délai total en sol africain, c'est le modèle de forêt aléatoire qui a offert, basé sur dix échantillons, les meilleurs résultats. Dans les deux cas, c'est le modèle ayant le plus grand potentiel prédictif pour le contexte étudié. Nous avons tenté d'améliorer ces modèles, nous

présentons les différentes versions développées et les ajustements qui ont été faits dans la prochaine section. Noter qu'il n'aurait pas été valable de tenter d'agréger le modèle maritime avec celui en sol africain. D'une part, car ce ne sont pas les mêmes parties qui sont responsables des denrées pour ces opérations et d'autre part, parce que les données collectées ne couvrent pas les mêmes dons.

5.3.2 Choix du modèle prédictif

Nous avons décidé de tester différentes versions des modèles d'arbres décisionnels multiples. Nous nous sommes basés sur le livre de James *et al.* (2013) et avons comparé les modèles de *bagging*, *random forest* et *boosting*. Pour ce faire, les données collectées ont été séparées en trois ensembles. Les modèles ont été développés à partir de l'ensemble d'entraînement comportant 70% des données. Le choix des paramètres a été effectué à partir des résultats prédictifs, testé sur l'ensemble de validation, comportant 15% des données. Le dernier ensemble, celui de test avec 15% des données, n'a été utilisé qu'à la toute fin, pour le choix du modèle final. Chaque ensemble a été construit de manière aléatoire et nous avons fait dix nouveaux échantillonnages pour analyser les résultats.

Le modèle appelé *Bagging* est une version simplifiée du modèle de forêt aléatoire présenté précédemment. Comme pour le modèle de forêt aléatoire, les arbres sont développés à la manière de Breiman (1996). Pour chaque arbre, un sous-ensemble, sélectionné aléatoirement, est utilisé parmi les données fournies au modèle. L'arbre est construit à partir de ce sous-ensemble. Pour ce modèle, nous avons aussi utilisé la fonction *randomForest()*. Par contre, le paramètre *mtry* prend la valeur de neuf dans le cas où l'on prédit le délai de transport maritime et de 69 dans celui en sol africain, soit le nombre total de variables explicatives pour chaque modèle. Ainsi, dans le cas du modèle que l'on appelle *Bagging*, lorsqu'une séparation est effectuée

à l'intérieur d'un arbre, toutes les variables explicatives sont considérées pour cette séparation. C'est la seule différence par rapport au modèle de forêt aléatoire. Dans les deux cas, le nombre d'arbres doit être défini par le paramètre *ntree*. Pour nos modèles que nous appelons *RandomForest*, nous avons testé différentes valeurs pour le paramètre *mtry*.

Le modèle appelé *Boosting* comporte lui aussi plusieurs arbres de régression qui sont combinés pour la prédiction, mais il est conçu autrement. Il s'appuie sur l'idée qu'il est plus facile d'utiliser plusieurs règles très générales que de se baser sur une règle de prédiction très précise (Schapire, 2003). Dans ce modèle, les arbres sont construits séquentiellement à partir d'une version modifiée des données originales. L'algorithme commence à partir des données originales. Suite au premier arbre bâti, il utilise les erreurs de prédictions (les résidus) pour bâtir l'arbre suivant. Les résidus sont mis à jour entre les arbres (diminuant d'une itération à l'autre). Ainsi, contrairement aux deux autres modèles, la construction de chaque arbre dépend fortement des arbres qui ont déjà été construits. En général, le modèle apprend lentement et les arbres sont peu profonds (James *et al.*, 2013). Pour nos modèles *Boosting* nous avons décidé d'utiliser le package *gbm* (Greenwell *et al.*, 2018) de R. Le *Generalized Boosted Regression Models* (GBM), est l'implémentation d'une extension de l'algorithme *AdaBoost* développé par Freund et Schapire (1999) et du *gradient boosting machine* développé par Friedman (2001). Nous avons utilisé la fonction *gbm()* pour concevoir nos modèles. D'après James *et al.* (2013) et Schiltz *et al.* (2018), les valeurs de trois paramètres doivent être sélectionnées pour le modèle *Boosting*. Le nombre d'arbres doit être spécifié avec le paramètre *n.trees*. Le paramètre *shrinkage* indiquant le rythme d'apprentissage du modèle. Il est typiquement de 0,01 ou de 0,001. Lorsque la valeur de ce paramètre est très petite, le nombre d'arbres pour avoir un modèle performant pourrait devoir être très grand. Puis, le paramètre *interaction.depth* qui contrôle le niveau de

complexité. Il indique le nombre de séparations de l'arbre et donc le nombre d'interactions entre les variables explicatives pour chaque arbre.

En utilisant les données des ensembles de validation, nous avons comparé les résultats prédictifs pour les paramètres importants des modèles d'arbres de régression multiple. Pour déterminer le nombre d'arbres des modèles *Bagging* et *RandomForest*, utilisant la fonction *randomForest()*, nous avons considéré les résultats de l'étude menée par Probst et Boulesteix (2018). Les auteurs ont analysé l'intérêt d'ajuster le paramètre du nombre d'arbre dans les modèles de forêt aléatoire. Ils indiquent que l'amélioration du modèle se fait avec les 250 premiers arbres, après quoi l'amélioration est très faible. Nous avons fait le choix d'utiliser un grand nombre d'arbres sans que le délai de résolution soit affecté. Nous avons donné la valeur de 1 000 au paramètre *ntree* dans tous les modèles *Bagging* et *RandomForest*. Nous n'avons pas tenté d'optimiser les autres paramètres des modèles. Certains performaient très bien sur un ensemble, mais moins bien sur d'autres. Nous avons décidé de faire nos choix pour les paramètres en priorisant la stabilité des résultats. Nous avons testé les paramètres sur dix ensembles et avons choisi les paramètres qui offraient de bons résultats dans la majeure partie des cas.

Dans le cas du transport maritime, pour le modèle de *RandomForest*, nous avons comparé cinq valeurs (1, 2, 3, 4 et 5) pour le paramètre *mtry*. Nous avons choisi d'utiliser *mtry* = 4 (par défaut il était de trois). Après vérifications, un nombre d'arbres augmenté à 1250 n'augmentaient pas la performance prédictive des deux premiers modèles (*Bagging* et *RandomForest*). Dans le modèle *Boosting*, Elith *et al.* (2008) précisent que le paramètre indiquant le rythme de l'apprentissage (*shrinkage*) et celui de la complexité (*interaction.depth*) doivent être déterminés avant de sélectionner le nombre d'arbres. Nous avons fait nos premières analyses, pour les paramètres, en utilisant un grand nombre d'arbres, soit 5 000 arbres. Avec *n.trees* = 5000, nous avons utilisé la validation croisée pour identifier les meilleurs

Tableau 5.11 Comparaison des modèles d'arbres décisionnels multiple pour prédire les délais de transport maritime. Sur l'ensemble de test (15%) de dix échantillons.

Mesure d'erreur		Moyenne Estimée	<i>Bagging</i> 1000 arbres, mtry= 9	<i>RandomForest</i> 1000 arbres, mtry= 4	<i>Boosting</i> 2500 arbres, <i>Shrinkage</i> = 0.01, <i>Depth</i> = 2
RMSE	Minimum :	11,94	6,53	6,86	8,21
	Maximum :	23,66	13,62	14,91	13,98
	Moyenne :	16,39	10,03	9,99	10,55
	Écart-type :	3,57	2,25	2,46	1,99
MAE	Minimum :	9,71	4,72	5,11	6,14
	Maximum :	16,79	8,73	9,29	9,82
	Moyenne :	11,62	6,85	6,82	7,57
	Écart-type :	2,25	1,23	1,28	1,08
MAPE	Minimum :	29,25	17,19	18,47	23,03
	Maximum :	80,14	29,54	29,04	33,13
	Moyenne :	50,86	22,56	22,30	29,94
	Écart-type :	17,40	3,64	3,14	3,40

paramètres pour l'apprentissage et la complexité. Six valeurs (0,001, 0,05, 0,04, 0,03, 0,02 et 0,01) ont été testées pour le rythme de l'apprentissage et cinq (1, 2, 3, 4, 5) pour la complexité. Nous avons déterminé que la meilleure valeur pour le paramètre *shrinkage* était de 0,01. Sur dix échantillons, nous avons testé les valeurs 2, 3 et 4 pour le paramètre *interaction.depth*. La valeur de deux était meilleure au niveau des RMSE et MAE, mais légèrement moins bonne que trois au niveau du MAPE. Nous avons décidé d'utiliser *interaction.depth* = 2, car il est préférable de restreindre le niveau de complexité des arbres afin d'éviter le surapprentissage. Ensuite, nous avons comparé les résultats prédictifs du modèle avec 5 000 arbres avec ceux de 2 500 et 10 000 arbres sur dix échantillons. C'est le modèle avec 2 500 arbres qui offrait, en moyenne, la meilleure performance prédictive.

Après avoir sélectionné les paramètres des modèles, nous avons comparé leurs résultats prédictifs sur l'ensemble de test des dix échantillons. Les statistiques des-

criptives des erreurs de prédictions sont inscrites au Tableau 5.11. En moyenne, les résultats des modèles *Bagging* et *RandomForest* sont très similaires, et ils surpassent tous les deux le modèle *Boosting* pour les trois métriques. Nous avons choisi le modèle *RandomForest*, parce qu'en moyenne, les erreurs calculées par le RMSE, le MAE et le MAPE sont inférieures à celles du modèle *Bagging*. Notez que lorsqu'elles ont été calculées sur l'ensemble de validation, les erreurs étaient encore plus petites, soit 7,97, en moyenne pour le RMSE, 5,61 pour le MAE et 21,68% pour le MAPE.

Une fois le modèle *RandomForest* choisi, avec 1 000 arbres et 4 variables explicatives considérées à chaque séparation d'un arbre, nous l'avons refait en utilisant la totalité des observations disponibles (soit $n = 205$). Ce faisant, le R^2 calculé par le modèle de *RandomForest* augmente de plus de 10%. Avec le modèle final pour prédire le délai de transport maritime, il est de 67,85%. Rappelons que le modèle *RandomForest* a une méthode d'évaluation des erreurs prédictive intégrée à l'algorithme. Les observations qui n'ont pas été utilisées pour la construction de l'arbre sont utilisées pour calculer le MSE. Le MSE calculé par le modèle final est de 77.83 jours², donc le RMSE est de 8,82 jours. Nous croyons que le modèle pourrait être amélioré si les données collectées représentaient un plus grand nombre de dons. La Figure 5.17 illustre la diminution des erreurs lorsque le nombre d'arbres augmente. On remarque qu'à partir de 500 arbres, le modèle ne s'améliore plus beaucoup, les erreurs sont stables. Parmi les 1 000 arbres, c'est le 798^e qui minimise les MSE.

Le modèle *RandomForest* permet aussi de classer le pouvoir prédictif des variables explicatives. La Figure 5.18 illustre les résultats issus de la fonction *VarImpPlot()*. Le graphique de gauche indique l'augmentation des MSE lorsque la variable explicative est permutée (remplacée aléatoirement). L'importance de la variable est mesurée par la diminution moyenne de la performance, résultant de la permutation. Ce niveau d'importance est calculé sur le nombre d'arbres total

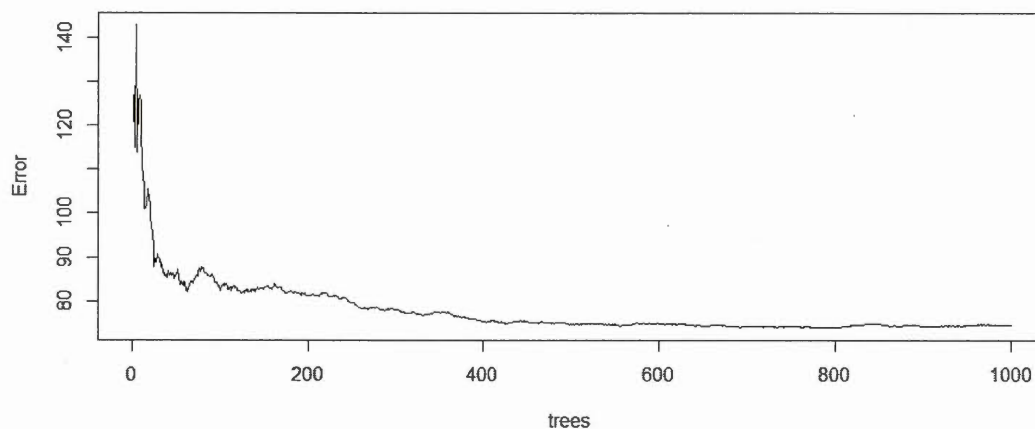


Figure 5.17 Décroissance des erreurs (MSE) du modèle de *RandomForest* final, prédisant les délais de transport maritime, en fonction du nombre d'arbres.

Ordres d'importance des variables

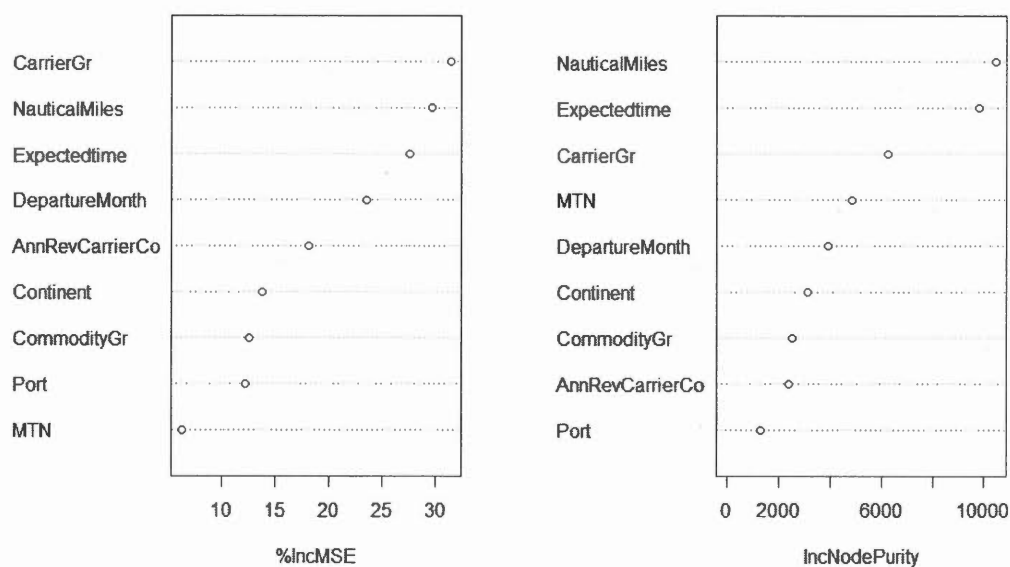


Figure 5.18 Classifications des variables explicatives par ordre d'importance selon le modèle *RandomForest* final prédisant les délais de transport maritime.

(Hastie *et al.*, 2009). Le graphique de droite indique à quel point la variable réduit l'impureté des noeuds. Pour chaque variable explicative, la diminution de la somme des carrés des résidus est calculée lorsqu'une séparation est effectuée à partir de cette variable dans un arbre. Cette diminution totale (pour tous les arbres de la forêt) représente la contribution de la variable explicative. Pour les deux graphiques, un chiffre plus grand indique une plus grande contribution de la variable explicative. Selon le modèle final prédisant les délais de transport maritime, les trois variables les plus importantes sont : *CarrierGr*, *NauticalMiles* et *Expectedtime*. Ce qui nous indique que la distance et la compagnie responsable du transport sont d'importants prédicteurs du délai de transport maritime. Les deux variables les moins importantes, bien qu'utilisées par le modèle, sont : *CommodityGr* et *Port*. Le type de denrées transportées et le port de destination en Afrique de l'Ouest n'influencent que très peu le délai de transport.

Nous avons suivi une démarche très similaire pour développer les différentes versions des modèles prédictifs (avec arbres décisionnels multiples) du délai total en sol africain. Le nombre d'arbres dans la forêt du modèle *Bagging* et *RandomForest* est de 1 000. Nous avons comparé les résultats prédictifs, sur l'ensemble de validation de dix échantillons, pour quatre valeurs (22, 23, 24 et 25) attribuées au paramètre *mtry* dans le modèle *RandomForest*. 23 était la valeur que le paramètre prenait par défaut. Nous avons décidé d'utiliser 24, car il offrait, en moyenne, de meilleurs résultats (25 était presque équivalent). Pour le modèle *Boosting*, nous avons, dans ce cas-ci, débuté nos analyses avec 10 000 arbres. Nous avons utilisé la validation croisée pour voir quelles valeurs pour les paramètres avaient le meilleur potentiel sur nos données. Quatre valeurs (0,01, 0,02, 0,03 et 0,05) ont été considérées pour le rythme d'apprentissage et trois (2, 3, 4) pour la complexité. Nous avons remarqué que lorsque le paramètre *shrinkage* est plus grand (p. ex., 0,03 ou 0,05) les erreurs de validations croisées, calculées par le modèle, remontent plus

Tableau 5.12 Comparaison des modèles d'arbres décisionnels multiple pour prédire les délais en sol africain. Sur l'ensemble de test (15%) de dix échantillons.

Mesure d'erreur		Moyenne Estimée	<i>Bagging</i> 1000 arbres, mtry= 69	<i>RandomForest</i> 1000 arbres, mtry= 24	<i>Boosting</i> 10000 arbres, <i>Shrinkage</i> = 0.01, <i>Depth</i> = 3	<i>Combined</i> (<i>RandomForest</i> & <i>Boosting</i>)
RMSE	Minimum :	13,31	6,88	6,72	6,43	6,43
	Maximum :	16,15	8,60	8,44	8,36	8,33
	Moyenne :	14,45	7,64	7,50	7,37	7,31
	Écart-type :	0,89	0,62	0,62	0,60	0,62
MAE	Minimum :	9,23	4,29	4,21	4,38	4,23
	Maximum :	11,49	5,36	5,31	5,42	5,30
	Moyenne :	10,43	4,86	4,83	4,87	4,76
	Écart-type :	0,68	0,35	0,37	0,34	0,35
MAPE	Minimum :	32,35	14,71	14,35	14,84	14,43
	Maximum :	39,91	17,61	17,55	17,82	17,44
	Moyenne :	36,69	16,37	16,16	16,40	16,02
	Écart-type :	2,26	1,02	1,10	1,06	1,04

rapidement à partir d'un certain nombre d'arbres. Elles sont plus stables avec des valeurs plus petites. Nous avons donc testé, sur dix échantillons, la valeur 0,01 et 0,02 pour le paramètre *shrinkage* et les valeurs 3 et 4 pour le paramètre *interaction.depth*. La combinaison *shrinkage* = 0,01 et *interaction.depth* = 3 a été sélectionnée, mais les trois autres options offraient des résultats prédictifs, sur les ensembles de validation, très similaires. Ensuite, nous avons comparé les résultats pour *n.trees* = 10000 avec ceux de trois autres quantités (5 000, 15 000 et 20 000) pour ce paramètre. Nous avons décidé d'utiliser 10 000 arbres, bien que le modèle de 15 000 arbres offrait des résultats très similaires.

Nous avons comparé les erreurs de prédiction, des modèles de *Bagging*, *RandomForest* et *Boosting*, sur l'ensemble de test de nos dix échantillons. Les résultats sont présentés à l'intérieur du Tableau 5.12. En moyenne, le modèle *RandomForest* révèle des erreurs de prédiction plus faibles au niveau du MAE et MAPE, soit

4,83 jours et 16,16% respectivement. Par contre, c'est le modèle *Boosting* qui a les erreurs les plus basses au niveau des RMSE, avec 7,37 jours. Pour faire le choix de notre modèle final, nous ne voulions pas prioriser une métrique plus qu'une autre. Nous voulions sélectionner le modèle qui offre, en moyenne, les meilleurs résultats prédictifs pour les trois métriques (RMSE, MAE et MAPE). Nous avons fait des analyses isolées sur nos prédictions. Nous avons comparé le délai réel avec le délai prédit par les modèles (*RandomForest* et *Boosting*). Nous avons remarqué que dans certains cas, l'un des modèles prédisait une valeur supérieure au vrai délai observé, alors que l'autre prédisait une valeur inférieure à ce délai. C'est possible, car les deux modèles utilisent une procédure bien différente pour bâtir leurs arbres. Entre autres, *RandomForest* développe tous les arbres de la forêt en parallèle alors que *Boosting* les développe de manière séquentielle. Baćak et Kennedy (2018) mentionnent que les meilleures estimations sont parfois produites par une combinaison pondérée de différents modèles prédictifs. Nous avons donc décidé de calculer les résultats prédictifs qu'offrirait un modèle prédisant la moyenne entre les deux prédictions. Ce modèle prédictif, combinant la prédiction faite par le modèle *RandomForest* et celle faite par *Boosting*, est appelé *Combined*. Les résultats de ce modèle sont présentés à la dernière colonne du Tableau 5.12. Ce modèle est plus robuste que les autres, car ses erreurs de prédiction, calculées par les trois métriques, sont inférieures à celles de tous les autres modèles. Il possède plus de stabilité dans la qualité de ses prédictions. C'est pourquoi nous avons sélectionné le modèle *combined* comme modèle final.

D'autres combinaisons de modèles plus sophistiqués auraient pu être testées. L'amélioration marginale, entre nos modèles d'arbres décisionnels multiples et le modèle *Combined* sélectionné, ne nous permet pas de penser que l'amélioration apportée par un modèle plus sophistiqué serait substantielle. Il n'a pas été jugé nécessaire d'analyser d'autres combinaisons de modèles prédictifs dans le cadre de

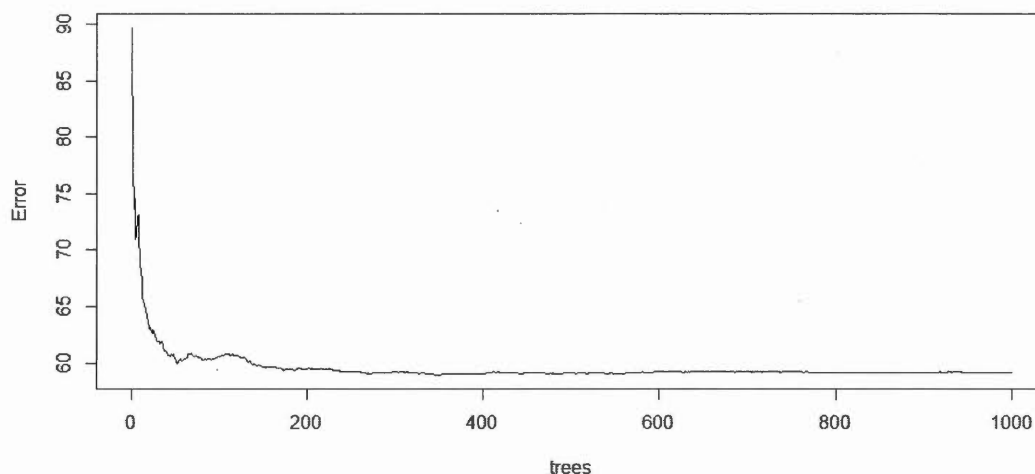


Figure 5.19 Décroissance des erreurs (MSE) du modèle de *RandomForest* final, prédisant les délais en sol africain, en fonction du nombre d'arbres.

ce projet, mais nous croyons qu'il y a un potentiel de recherche sur ce sujet.

Étant donné que le modèle *combined* nécessite l'élaboration des deux autres meilleurs modèles (*RandomForest* et *Boosting*), nous les avons refaits en utilisant la totalité des données collectées (soit $n = 1753$). Le R^2 calculé par le modèle final *RandomForest* est de 73,28%. Le MSE est de 59,6 jours², donc le RMSE de 7,72 jours. La Figure 5.19 montre l'évolution de l'erreur de type MSE plus le nombre d'arbres de la forêt augmente. Jusqu'à 300 arbres, le modèle s'améliore. Il n'y a que très peu d'amélioration par la suite. Les MSE sont minimisées lorsqu'il y a 351 arbres dans le modèle. Les 69 variables explicatives sont utilisées par le modèle. Le niveau d'importance des variables explicatives est illustré à la Figure 5.20. Les deux graphiques présentent les 30 plus importants prédicteurs. Dans le cas du graphique de gauche (indiquant l'augmentation de l'erreur lorsque la variable est permutée), il y a quatre variables explicatives qui se démarquent

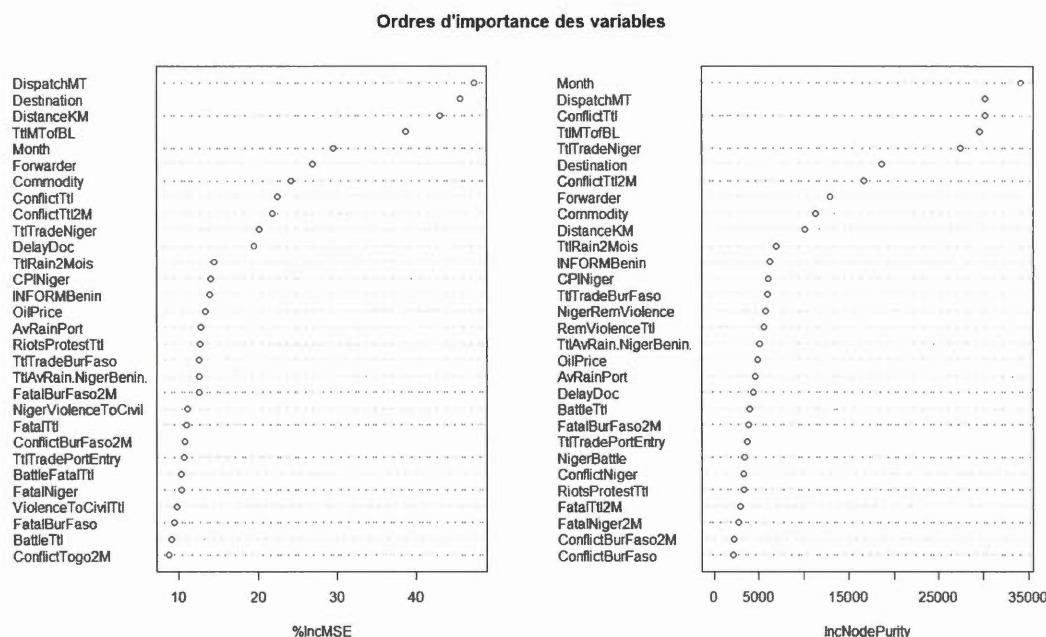


Figure 5.20 Classifications des variables explicatives par ordre d'importance selon le modèle de *RandomForest* final prédisant les délais en sol africain.

par leur écart vis-à-vis les autres. Ce sont les variables *DispatchMT*, *Destination*, *DistanceKM* et *TtlMTofBL* qui sont les plus importantes. Dans le cas du graphique de droite (basé sur la contribution de la variable aux noeuds des arbres) il y en a cinq : *Month*, *DispatchMT*, *ConflictTtl*, *TtlMTofBL* et *TtlTradeNiger*.

Ces classifications de l'importance des variables concordent avec celle effectuée par le modèle final *Boosting*. Tel qu'illustré à la Figure 5.21, le modèle *Boosting* identifie les cinq variables suivantes comme étant les plus importantes : *Month*, *DispatchMT*, *TtlMTofBL*, *ConflictTtl* et *Destination*. Nous avons donc, au total, sept variables explicatives considérées plus importantes par les modèles. Deux de ces variables, *ConflictTtl* et *TtlTradeNiger*, ont été créées pour représenter des sources d'incertitude (Sécuritaire et Partie prenante). Le modèle *Boosting* n'utilise pas les 69 variables explicatives qu'on lui a fournies, il en utilise 61.

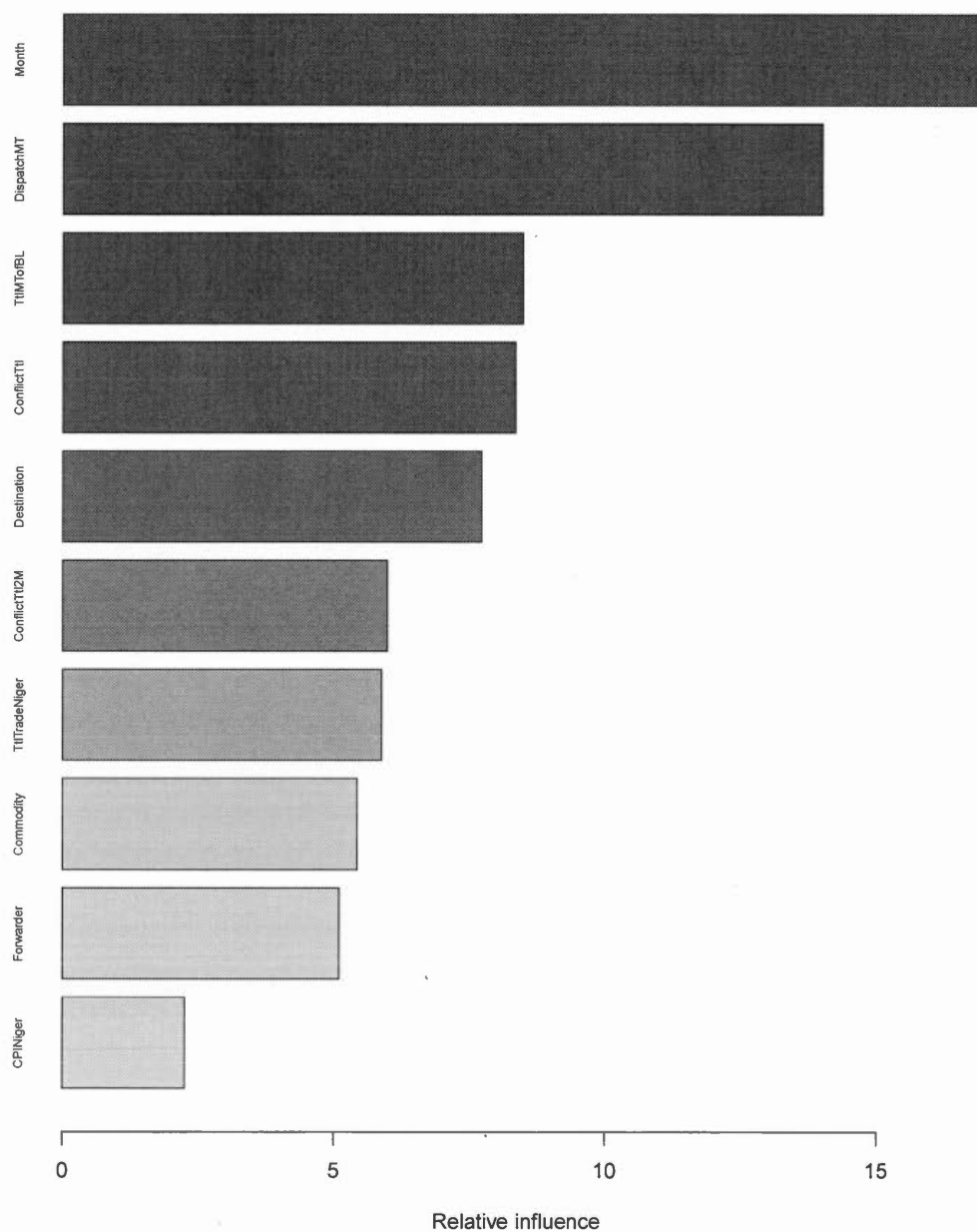


Figure 5.21 Classifications des variables explicatives par ordre d'importance selon le modèle *Boosting* final prédisant les délais en sol africain.

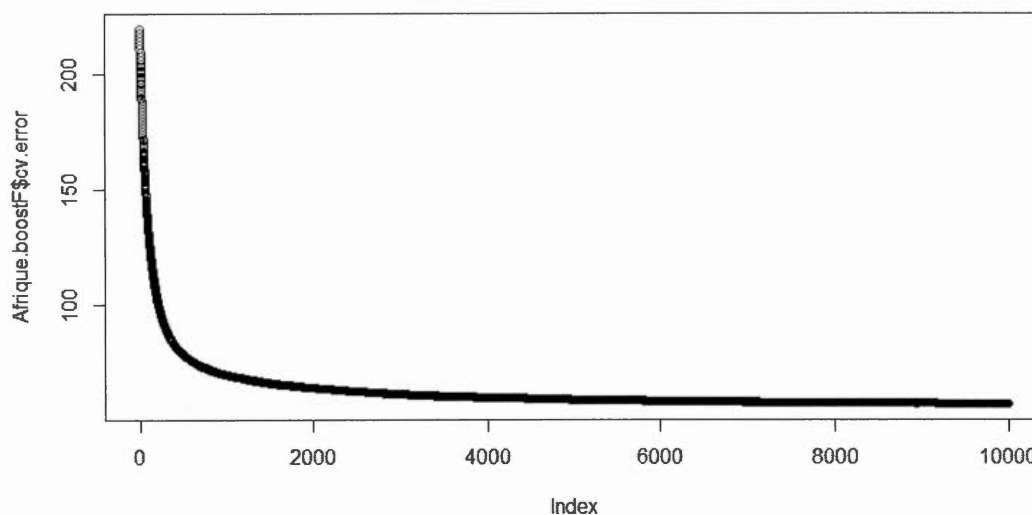


Figure 5.22 Décroissance des erreurs de validation croisée du modèle *Boosting* final, prédisant les délais en sol africain, en fonction du nombre d'arbres.

Il y a, entre autres, quatre variables (parmi les cinq) de type $INFORM_i$ qui ne sont pas utilisées par le modèle. Pour l'indice du risque, il n'y a que la variable $INFORM_{AvCor}$ qui est conservée. Elle est classée au 28^e rang par le modèle *Boosting*. Les erreurs de validation croisées sont minimisées lorsque le 10 000^e arbre est créé. La Figure 5.22 montre l'évolution des erreurs de validation croisées, calculées par le MSE, lorsque le nombre d'arbres augmente. On remarque que le modèle s'améliore toujours. Il est possible de surapprendre avec le un modèle de type *Boosting*, en fonction des données et des autres paramètres sélectionnés, les erreurs de validation croisées augmentent lorsqu'un nombre d'arbres trop élevé est sélectionné. Dans le cas de notre modèle final, nous avons fait le test et c'est autour de 15 000 arbres que le modèle cesse de s'améliorer. Dans modèle *Boosting* final, comprenant 10 000 arbres, le MSE est de 55,73 jours, et le RMSE de 7,47 jours.

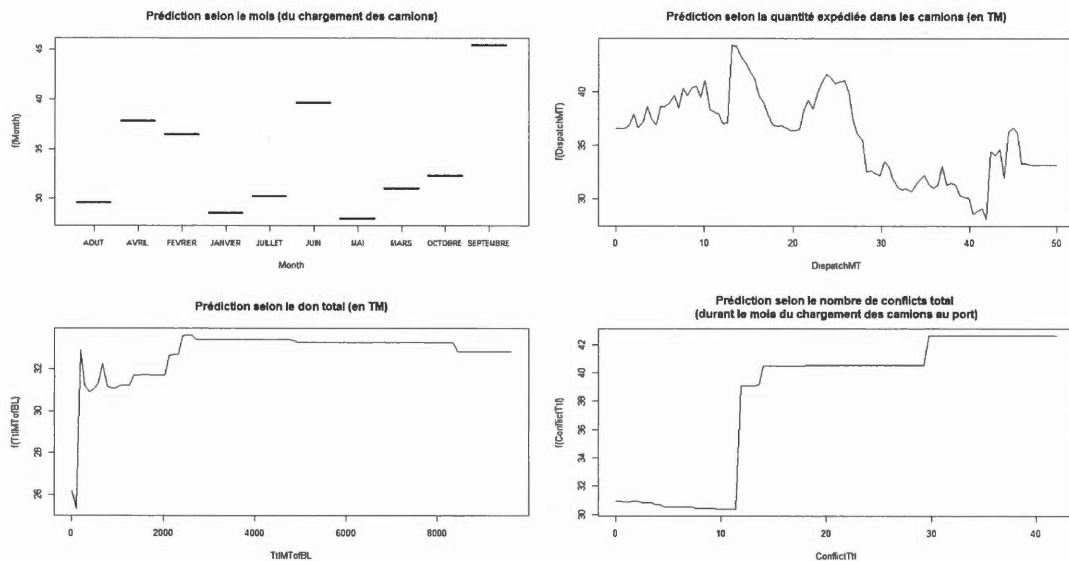


Figure 5.23 Graphiques de dépendance partielle entre une variable explicative et la prédiction du délai en sol africain.

Le modèle *Boosting* utilisé permet aussi de visualiser comment la prédiction change lorsqu'une variable explicative est modifiée (sans que les autres variables explicatives le soient). Les graphiques illustrés par la Figure 5.23 montrent la relation de dépendance qui existe entre les quatre plus importantes variables (faire référence à la Figure 5.21) et la prédiction faite pour le délai en sol africain. D'après les données collectées, le mois où, en moyenne, le délai observé en sol africain est le plus long est le mois d'avril (58 jours) et le plus rapide est janvier (20,2 jours). Pourtant le premier graphique indique que le modèle prédit, en moyenne, un délai plus long pour le mois de septembre et un délai plus rapide au mois de mai. La prédiction varie de plus de 15 jours en fonction du mois où les camions sont chargés au port. Le graphique de dépendance partielle entre la variable *DispatchMT* et la prédiction pour la variable *PortEDP* montre à quel point la prédiction varie selon la quantité transportée. Rappelons que la Figure 5.5 montrait que les quantités transportées par camion sont généralement entre 29 et 45 TM. Dans 91%

des cas, la variable *DispatchMT* prend une valeur dans l'intervalle [29, 45], c'est donc la section du graphique la plus importante à analyser. Dans cette section, il y a une tendance indiquant que plus le camion est rempli, plus le délai prédit est court. Par contre, la variation dans la valeur du délai prédit est faible. Entre 29 et 41 MT le délai diminue d'environ cinq jours, puis il remonte lorsque la variable *DispatchMT* est de plus 42 MT. Le troisième graphique nous semble un peu moins marquant, car la quantité inscrite sur le certificat de connaissance est rarement très basse. Elle est inférieure à 250 MT que dans 10% des cas. Dans le reste du graphique, le délai prédit ne varie que d'environ deux jours. Le graphique de la variable *ConflictTtl* démontre un élément très intéressant, le délai augmente d'environ dix jours lorsque le nombre total de conflits est supérieur à 12. C'est le même point de rupture qui débute l'arbre de régression présenté à la Figure 5.16. Nous avons fait le test, et si l'arbre de régression unique avait été utilisé comme modèle final, bâti sur toutes les données, il aurait débuté par la règle de séparation $\text{ConflictTtl} < 12$. Le modèle *Boosting* est pourtant construit de manière bien différente. Ces observations prouvent l'apport prédictif de cette variable initialement créée pour représenter la source d'incertitude *Sécuritaire*.

Enfin, pour les deux modèles nécessaires à la construction du modèle final ainsi que pour le modèle final *Combined*, nous avons calculé les résidus des modèles. Pour chaque observation, la valeur prédite par le modèle a été soustraite de la vraie valeur du délai en sol africain. Ainsi, à la Figure 5.24 nous pouvons observer la distribution des résidus pour le modèle *RandomForest*, *Boosting* et *Combined*. Dans les trois cas, la moyenne des résidus est de zéro (avec 0,05, 0,009 et 0,03 respectivement). Il y a quelques erreurs extrêmes (positives), elles sont dues au fait que nous avons conservé les observations aux valeurs aberrantes ; les cas possiblement catastrophiques. Ainsi, nos plus grandes erreurs de prédictions sont faites lorsque nous prédisons une valeur inférieure à la vraie valeur.

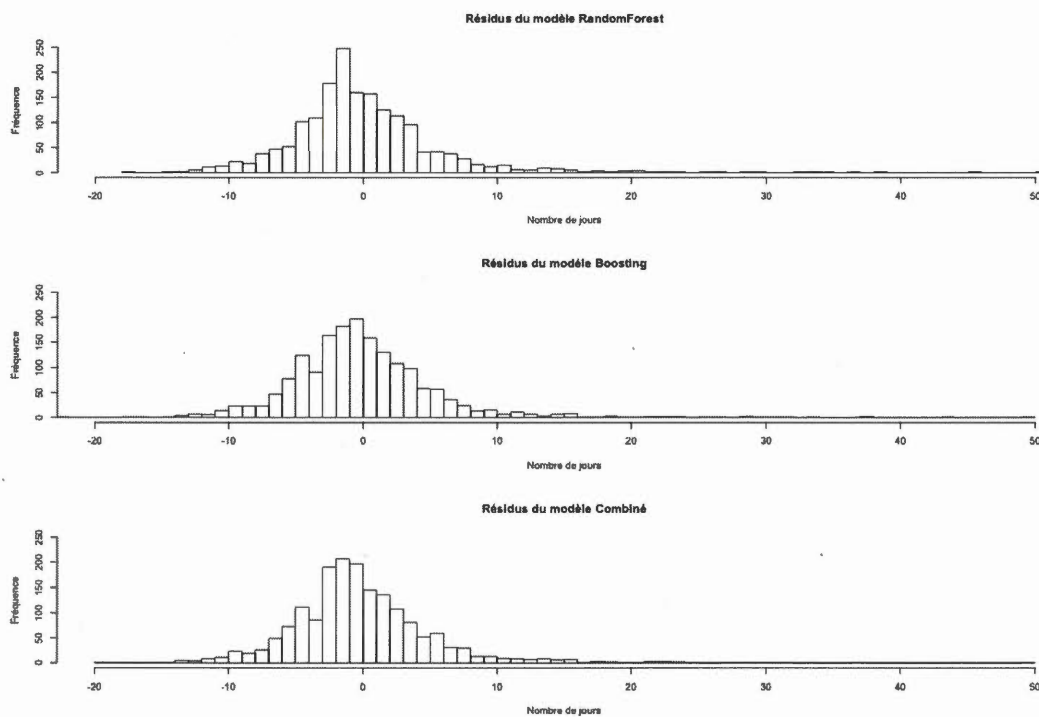


Figure 5.24 Distribution des erreurs de prédictions (résidus) pour les trois modèles finaux en sol africain.

CHAPITRE VI

DISCUSSION

En identifiant et catégorisant les sources d'incertitude, notre premier objectif a été atteint. En déterminant le modèle pour estimer le délai de transport maritime (*RandomForest*), et celui pour estimer le délai en sol africain (*Combined*), notre deuxième objectif a été atteint. Dans ce chapitre, nous présentons une preuve de concept en illustrant l'utilisation des modèles prédictifs pour contribuer à la prise de décision. En plus d'atteindre notre troisième objectif (générer les scénarios), nous répondons à notre question de recherche.

La Figure 6.1 présente les activités qui ont été menées pour la discussion. D'abord, nous identifions les variables explicatives dont les valeurs sont connues lors de la prise de décision et celles qui ne le sont pas. Les variables explicatives inconnues doivent être estimées par l'utilisateur ; par les experts du milieu. En prenant l'exemple d'un don plus récent qui ne fait pas partie des données utilisées pour le développement des modèles, nous illustrons l'utilisation du modèle prédisant les délais en sol africain. Des scénarios ont été générés pour représenter les valeurs possibles des variables explicatives. Nous présentons les simulations développées et expliquons comment nous avons généré 1 000 scénarios pour chacune. Nous utilisons ensuite le modèle prédictif pour estimer le délai logistique de chaque scénario. Les prédictions des délais de chaque simulation sont analysées et comparées.

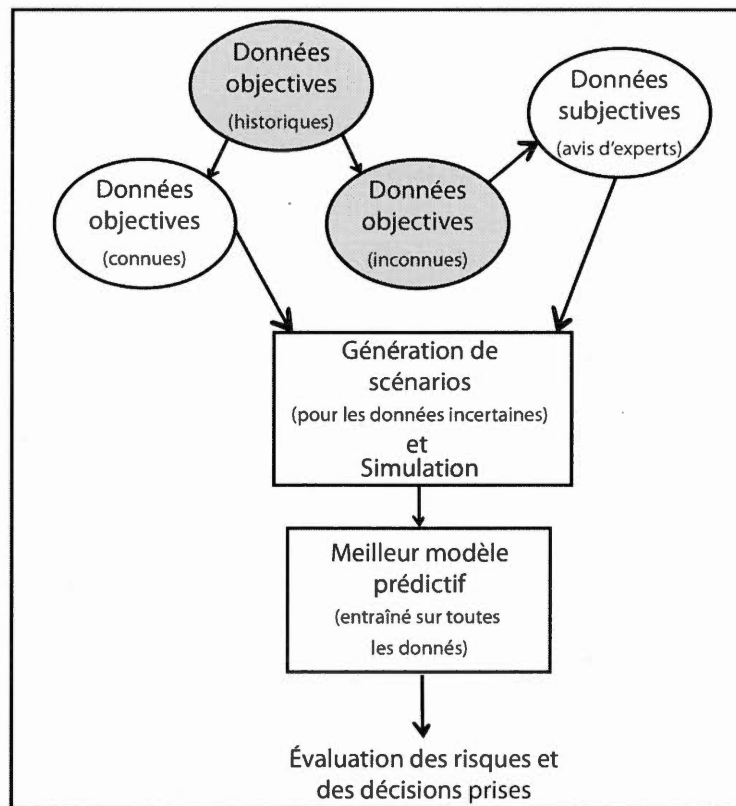


Figure 6.1 Schéma des activités du chapitre 6

Enfin, nous démontrons comment évaluer les risques en contexte d'incertitude. Les résultats prédictifs des différentes options décisionnelles sont aussi comparés afin d'évaluer les décisions prises.

6.1 Génération de scénarios et simulation

Pour quantifier le délai de transport maritime, la génération de scénario est très simple. Lors de l'utilisation du modèle prédictif *RandomForest*, les valeurs des neuf variables explicatives (faire référence au Tableau 5.4) sont facilement accessibles à l'utilisateur. Les responsables du PAM en Italie donnent plusieurs informations importantes pour ce modèle aux logisticiens basés au Niger. Ils obtiennent ainsi les va-

Tableau 6.1 Valeurs à utiliser pour la variable explicative AnnRevCarrierCo

Compagnie maritime	Revenu annuel estimé (en M\$ US)	Compagnie maritime	Revenu annuel estimé (en M\$ US)
CARRIER1	0,57	CARRIER2	2 554
CARRIER3	17 700	CARRIER4	3 269
CARRIER5	40 308	CARRIER6	2 258
CARRIER7	499	CARRIER8	2 300
CARRIER9	55,5	CARRIER10	3 318

leurs des variables MTN, Continent, CarrierGr, CommodityGr et DepartureMonth. Les logisticiens obtiennent aussi l'information sur le port d'origine du don par le PAM en Italie. Le port doit être inscrit sur le site Ports.com (2018) pour obtenir les valeurs des variables NauticalMiles et ExpectedTime. Ces deux dernières variables explicatives nécessitent aussi l'information sur le port d'arrivée. Les logisticiens vont pouvoir comparer les résultats obtenus par le modèle prédictif pour leurs différentes options de port d'arrivées (généralement Cotonou ou Lomé). La variable Port indique ce choix. Pour ce qui est de la variable AnnRevCarrierCo, si la compagnie de transport maritime est identifiée à l'intérieur du Tableau 6.1, les logisticiens inscrivent le revenu qui a été estimé. Sinon, ils doivent identifier (ou estimer) le revenu annuel de la nouvelle compagnie de transport maritime, en million de dollars américains. Lors de l'utilisation du modèle, il y aura donc qu'une seule valeur (catégorique ou numérique) pour représenter chaque variable explicative du modèle. Toutes les valeurs des variables sont connues, ou présumées connues. Ainsi, dans tous les scénarios, les valeurs des variables explicatives estimant le délai de transport maritime sont les mêmes. Le modèle prédictif peut donc être utilisé tel quel pour quantifier le paramètre stochastique.

La situation est différente pour quantifier le délai en sol africain. Au moment de l'utilisation du modèle (de la prise de décision), les informations pour représenter les variables explicatives ne sont pas toujours disponibles. Il y a des variables dont

Tableau 6.2 Variables explicatives les plus importantes pour l'estimation du délai en sol africain

Variable	Caractéristique	Provenance de l'information
DispatchMT	Inconnue (Numérique)	Valeurs historiques
Destination	Connue (Catégorique)	Prédéterminée par le PAM Niger
DistanceKM	Présumée connue (Numérique)	Google Maps (2018)
Tt1MTofBL	Connue (Numérique)	PAM Italie
Month	Inconnue (Catégorique)	Avis d'experts
ConflictTt1	Inconnue (Numérique)	Valeurs historiques et avis d'experts
Tt1TradeNiger	Inconnue (Numérique)	Valeurs historiques et avis d'experts

les valeurs sont connues (ou présumées connues), mais d'autres dont les valeurs sont inconnues. Comme les valeurs de ces dernières sont incertaines, elles doivent être estimées. Le modèle *Combined* sélectionné pour prédire le délai en sol africain nécessite deux modèles d'arbres décisionnels multiples, soit *RandomForest* et *Boosting*. Ces deux modèles nous ont indiqué qu'il y a sept variables explicatives qui se démarquent au niveau de leur importance pour la prédiction. C'est pourquoi une attention particulière a été portée sur ces variables. Tel qu'indiqué par le Tableau 6.2, trois d'entre elles sont connues (ou présumée connue) au moment de l'utilisation du modèle alors que quatre sont inconnues. La destination des denrées est prédéterminée par les logisticiens (variable *Destination*). La distance à parcourir a été calculée pour chaque couple d'origine et destination (variable *DistanceKM*). La quantité totale du don est inscrite sur le certificat de connaissance transmis par le siège social du PAM (variable *Tt1MTofBL*). Les valeurs de ces trois variables connues sont constantes pour tous les scénarios d'une simulation. Pour les variables incertaines, c'est en utilisant l'historique de leurs valeurs ou grâce à l'avis des experts du milieu qu'elles ont été estimées. Dans les modèles de simulations, différents scénarios ont été testés pour les variables *DispatchMT*, *Month*, *ConflictTt1* et *Tt1TradeNiger*. Nous croyons que les variables *Month*,

`ConflictTtl` et `TtlTradeNiger` peuvent être anticipées par les experts du milieu. Le ETA du navire aide à estimer le mois où les camions sont chargés au port (variable `Month`). C'est l'expérience des logisticiens et leurs anticipations des conflits et des échanges commerciaux dans les pays de l'Afrique de l'Ouest qui aide à estimer les variables `ConflictTtl` et `TtlTradeNiger`.

Les 62 autres variables explicatives utilisées par le modèle prédictif final sont énumérées dans le Tableau 6.3. Les variables `LoadPort` et `Forwarder` représentent les décisions à prendre par les responsables du PAM au Niger. Les résultats du modèle prédictif devront être comparés pour les différentes options décisionnelles. Cinq variables sont connues (ou présumées connues). L'information concernant le type de denrée (variable `Commodity`) est donnée par le siège social du PAM en Italie. Le nombre de douanes à traverser est connu, car il est déterminé en fonction de la variable `LoadPort`. L'information concernant les saisons (variables `RainSeason` et `HarvestSeason`) est présumée connue, car elle est déterminée en fonction du mois où les camions sont chargés au port. Les valeurs des indices `INFORM` sont connues lors de l'utilisation du modèle, car à la fin du mois de novembre 2018 l'information pour l'année 2019 était déjà disponible publiquement (groupe de variables `INFORMi`). La variable `DelayDoc` est inconnue lorsqu'un don est confirmé pour le Niger, mais les logisticiens ont un certain contrôle sur cette variable. Ils peuvent faire parvenir les documents au transitaire en avance et donc déterminer que cette variable prend la valeur de zéro pour tous les scénarios. Toutes les autres variables sont inconnues au moment de l'utilisation du modèle, elles doivent être estimées par l'utilisateur. Le Tableau 6.3 précise aussi quelles variables doivent être estimées en tenant compte du port d'arrivée des denrées.

Pour illustrer l'utilisation du modèle prédisant les délais en sol africain, nous prenons l'exemple d'un don qui ne fait pas partie des données qui ont été collectées. Ce don n'a donc pas été utilisé pour le développement des modèles. Nous avons les

Tableau 6.3 Autres variables explicatives pour estimer le délai en sol africain

Variable ou groupe de variable	Caractéristique	Variable ou groupe de variable	Caractéristique
LoadPort	À sélectionner (Catégorique)	Forwarder	À sélectionner (Catégorique)
Commodity	Connue (Catégorique)	NBofCustoms	Connue (Numérique); associée à la variable LoadPort
RainSeason	Présumée connue (Catégorique); associée à la variable Month	HarvestSeason	Présumée connue (Catégorique); associée à la variable Month
DelayDoc	Inconnue (Numérique)	OilPrice	Inconnue (Numérique)
AvRain: i = Niger, Port, BurFaso, TtlCor, Ttl2Mois	Inconnue (Numérique); AvRainPort, AvRainCor et AvRainTtl2Mois associées à la variable LoadPort	TtlTrade: i = PortEntry, BurFaso	Inconnue (Numérique); TtlTradePortEntry associée à la variable LoadPort
CPI: i = Niger, Benin, Togo, BurFaso, AvCorridor	Inconnue (Numérique)	INFORM: i = Niger, Benin, Togo, BurFaso, AvCor	Connue (Numérique) INFORMAvCor est associée à la variable LoadPort
Conflict: i = Niger, Benin, Togo, BurFaso, Niger2M, Benin2M, Togo2M, BurFaso2M, Ttl2M	Inconnue (Numérique) ConflictTtl2M est associée à la variable LoadPort	Fatal: i = Niger, Benin, Togo, BurFaso, Ttl, Niger2M, Benin2M, Togo2M, BurFaso2M, Ttl2M	Inconnue (Numérique) FatalTtl et FatalTtl2M sont associées à la variable LoadPort
Conflict: i = NigerBattle, NigerRemViolence, NigerRiotsProtest, NigerStratDev, NigerViolenceToCivil, BattleTtl, ...	Inconnue (Numérique) Les cinq variables pour le total du corridor (exemple ConflictBattleTtl) sont associées à la variable LoadPort	Fatal: i = NigerBattle, NigerRemViolence, NigerRiotsProtest, NigerStratDev, NigerViolenceToCivil, BattleTtl, ...	Inconnue (Numérique) Les cinq variables pour le total du corridor (exemple FatalBattleTtl) sont associées à la variable LoadPort

Tableau 6.4 Valeurs des variables explicatives connues dans l'exemple de simulation DeBase

Variable	Valeur variable connue	Variable	Valeur variable connue
Destination	DIFFA	DistanceKM	2128
Tt1MTofBL	12840	LoadPort	COTONOU
Commodity	CERRIC	NBofCustoms	2
INFORMNiger	7.3	INFORMBenin	4.4
INFORMTogo	4.1	INFORMBurFaso	5.4
INFORMAvCor	5.85	DelayDoc	0

informations le concernant, nous avons le RFQ et le *Tender* qui lui sont associés ainsi que les délais réels en sol africain pour quatre observations. Dans un cas le délai a été de 33 jours, dans deux cas il a été de 50 jours, et dans un cas de 51 jours. Dans cet exemple, le navire était prévu pour arriver le 8 avril 2017, selon le ETA inscrit sur les documents. Il est plutôt arrivé le 22 mai 2017, soit 44 jours plus tard. Le choix du transitaire a été fait le 14 mars 2017. C'est donc avant ce moment que notre modèle prédictif devrait être utilisé. Dans cet exemple, c'est plus de deux mois avant le chargement des camions au port, mais ce délai varie selon le don.

Les informations connues pour ce don lors de la prise de décision se trouvent dans le Tableau 6.4. La première simulation que nous avons développée est nommée simulation DeBase. Nous avons pris l'exemple du riz (*Commodity* = CERRIC) à destination de la ville de Diffa au Niger. Le port qui a été sélectionné est Cotonou donc la distance calculée est de 2 128 KM et le nombre de postes douaniers à traverser est deux. Il y a 7 710 MT de riz dans ce don, mais le certificat de connaissance inclut quatre types de denrées, à destination de six villes, et totalise 12 840 MT. En 2017, la valeur de l'indice INFORM pour les pays de l'Afrique de l'Ouest analysés se situait entre 4,1 et 7,3. Le Niger était le pays avec le plus haut niveau de risque. Nous faisons l'hypothèse que les documents sont envoyés

au transitaire à temps ($\text{DelayDoc} = 0$). Les variables énumérées dans le Tableau 6.4 ont pris les mêmes valeurs pour tous les scénarios de toutes les simulations dont le port d'arrivée était Cotonou.

Le Tableau 6.5 présente les choix qui ont été faits pour estimer les variables inconnues dans la simulation DeBase. Le transitaire qui a réellement effectué les opérations logistiques pour ce don est BÉNIN1, nous avons donc utilisé ce choix pour la simulation DeBase. Sachant que le navire devait arriver au port de Cotonou vers le 8 avril nous avons décidé de répartir les probabilités pour le mois du chargement des camions en trois : un tiers des probabilités pour le mois d'avril, un tiers pour le mois de mai et un tiers pour le mois de juin. Cette estimation doit être faite par l'utilisateur, au meilleur de ses connaissances. Comme le mois de juin fait partie de la saison des pluies, la variable `RainSaison` est positive pour le tiers des cas. Les mois estimés ne font pas partie de la saison des récoltes, donc `HarvestSeason` = non pour tous les scénarios. Le prix du baril de pétrole brut a été estimé en faisant la moyenne du prix moyen des douze mois de l'année précédente (année 2016). Les indices de corruption (groupe de variables CPI_i) ont été estimés en utilisant la valeur connue la plus récente, soit l'indice de l'année précédente pour tous les scénarios.

Les estimations des autres variables inconnues sont issues de deux bases de données distinctes. La première est celle qui a été conçue pour développer les modèles prédictifs, elle est dénotée par BD1. Pour estimer le niveau des échanges commerciaux des variables `TtlTradeBurFaso` et `TtlTradePortEntry`, nous avons utilisé la moyenne des valeurs observées pour ces variables dans BD1. Ces variables ont donc pris la même valeur pour tous les scénarios. Les variables indiquant le niveau de pluie (groupe AvRain_i) ont été estimées en tenant compte de leur corrélation avec les mois. Nous avons utilisé la moyenne de leurs valeurs historiques, dans BD1, lorsque le mois du chargement du camion était le même que celui du scéna-

Tableau 6.5 Estimation des variables explicatives inconnues dans l'exemple de simulation DeBase

Variable ou groupe de variable	Estimation variable inconnue	Variable ou groupe de variable	Estimation variable inconnue
Forwarder	BÉNIN1	Month	33.33% = AVRIL, 33.33% = MAI et 33.33% = JUIN
RainSeason	33.33% = oui et 66.66% = non	HarvestSeason	100% = non
OilPrice	Moyenne des 12 mois de l'année 2016	AvRain _i _i = Niger, Port, BurFaso, TtlCor, Ttl2Mois	Moyenne des cas historiques du PAM au Niger (BD1) pour de mois de l'observation
TtlTrade _i _i = PortEntry, BurFaso	Moyenne des cas historiques du PAM au Niger (BD1)	CPI _i _i = Niger, Benin, Togo, BurFaso, AvCorridor	Valeur la plus récente (indice de 2016)
Conflict _i _i = Niger, Benin, Togo, BurFaso, Niger2M, Benin2M, Togo2M, BurFaso2M, Ttl2M	Moyenne des cas historiques du PAM au Niger (BD1) pour de mois de l'observation	Fatal _i _i = Niger, Benin, Togo, BurFaso, Ttl, Niger2M, Benin2M, Togo2M, BurFaso2M, Ttl2M	Moyenne des cas historiques du PAM au Niger (BD1) pour de mois de l'observation
Conflict _i _i = NigerBattle, NigerRemViolence, NigerRiotsProtest, NigerStratDev, NigerViolenceToCivil, BattleTtl, ...	Moyenne des cas historiques du PAM au Niger (BD1)	Fatal _i _i = NigerBattle, NigerRemViolence, NigerRiotsProtest, NigerStratDev, NigerViolenceToCivil, BattleTtl, ...	Moyenne des cas historiques du PAM au Niger (BD1)

rio de la simulation. Parmi les variables qui répertorient les conflits et les décès totaux dans les villes de l'Afrique de l'Ouest (ConflictNiger, FatalNiger, etc.), plusieurs étaient corrélées négativement avec les niveaux de précipitation. C'est pourquoi nous avons estimé les 20 variables de ce type de la même manière que les variables représentant les précipitations (moyenne historique pour le mois dans BD1 associée au mois du scénario). Pour les variables de conflits et de décès désagrégées par type de conflits, elles ont été estimées en utilisant leur moyenne dans BD1, cette valeur a été utilisée dans tous les scénarios.

Les trois variables inconnues les plus importantes (DispatchMT, ConflictTtl et

Tableau 6.6 Échantillonnage des variables importantes des modèles de simulation

Simulation	Provenance et intervalle de sélection de la variable		
	DispatchMT	ConflictTtl	TtlTradeNiger
DeBase	BD1 [0 , ∞]	BD2 [0 , ∞]	BD2 [0 , ∞]
MT2050	BD1 [20 , 50]	BD2 [0 , ∞]	BD2 [0 , ∞]
MT3040	BD1 [30 , 40]	BD2 [0 , ∞]	BD2 [0 , ∞]
MT40	Uniforme [40]	BD2 [0 , ∞]	BD2 [0 , ∞]
Conflict042	BD1 [0 , ∞]	BD1 [0 , ∞]	BD2 [0 , ∞]
Conflict316	BD1 [0 , ∞]	BD1 [3 , 16]	BD2 [0 , ∞]
Conflict310	BD1 [0 , ∞]	BD1 [3 , 10]	BD2 [0 , ∞]
Trade1655	BD1 [0 , ∞]	BD2 [0 , ∞]	BD1 [160 000, 550 000]
Trade2040	BD1 [0 , ∞]	BD2 [0 , ∞]	BD1 [200 000, 400 000]
Trade3032	BD1 [0 , ∞]	BD2 [0 , ∞]	BD1 [300 000, 320 000]

TtlTradeNiger) ont été analysées par différentes simulations. Toutes les simulations qui ont été conçues comportent 1 000 scénarios. Le Tableau 6.6 présente, pour chacune de ces simulations, la provenance et l'intervalle de sélection des variables importantes. Tous les échantillonnages ont été faits avec remplacement en utilisant la fonction *sample()* de R. Dans la simulation DeBase, la variable DispatchMT est échantillonnée aléatoirement parmi les observations historiques de BD1. Afin de mieux représenter les valeurs possibles des conflits totaux et des échanges commerciaux du Niger, nous avons utilisé notre seconde base de données dénotée BD2. Cette base de données comporte uniquement que des données collectées publiquement. Toutes les variables qui ont été collectées en ligne pour représenter les sources d'incertitude font partie de cette base. La BD2 contient une valeur historique par mois pour toutes les années disponibles. Ainsi, dans la simulation DeBase, ConflictTtl a été échantillonnée aléatoirement parmi les données historiques mensuelles comprises entre décembre 1997 et décembre 2017. L'échantillonnage aléatoire de TtlTradeNiger a été fait dans les données historiques mensuelles comprises entre de janvier 2007 et septembre 2017.

Dans les autres simulations présentées par le Tableau 6.6, nous avons changé que la provenance et l'intervalle de sélection d'une des variables importantes. Nous avons bâti trois simulations pour analyser la variable *DispatchMT*. Dans tous les cas son échantillonnage provient de la BD1 et l'échantillonnage des variables *ConflictTtl* et *TtlTradeNiger* reste inchangé par rapport à la simulation De-Base. Dans la simulation nommée MT2050 la sélection aléatoire est faite seulement parmi les valeurs comprises entre 20 MT et 50 MT. Les observations qui représentent des camions très peu remplis sont donc exclues. Cet intervalle est réduit entre 30 MT et 50 MT pour la simulation MT3050. Dans la simulation MT40 la variable *DispatchMT* prend la valeur de 40 MT pour tous les scénarios. La variable *ConflictTtl* est aussi analysée grâce à trois simulations dont les valeurs sont sélectionnées dans la BD1. Dans la simulation DeBase, où les valeurs de la variable provenaient de BD2, *ConflictTtl* prenait des valeurs allant jusqu'à 47 conflits. En modifiant la base de données pour BD1 et en utilisant toutes les observations disponibles, la variable prend une valeur maximale de 42 conflits dans la simulation Conflict042. L'intervalle a été réduit entre 3 et 16 conflits pour la simulation Conflict316. Nous avons choisi cet intervalle, car les valeurs prises par la variable *ConflictTtl* durant les mois de l'année 2017 sont comprises dans cet intervalle. La simulation Conflict310 comprend les vraies valeurs de la variable pour les mois d'avril, mai et juin, soit entre 3 et 10 conflits. Les échanges commerciaux du Niger (la somme des importations et des exportations mensuelle) ont été analysés de manière similaire. La simulation nommée Trade1655 indique les échanges commerciaux dont la valeur est comprise entre 160 000 et 550 000 dollars américains. Cet écart est réduit entre 200 000 et 400 000 dollars pour la simulation Trade2040 et entre 300 000 et 320 000 dollars pour la simulation Trade3032. L'intervalle de sélection de la dernière simulation comprend les vraies valeurs de la variable pour les mois simulés.

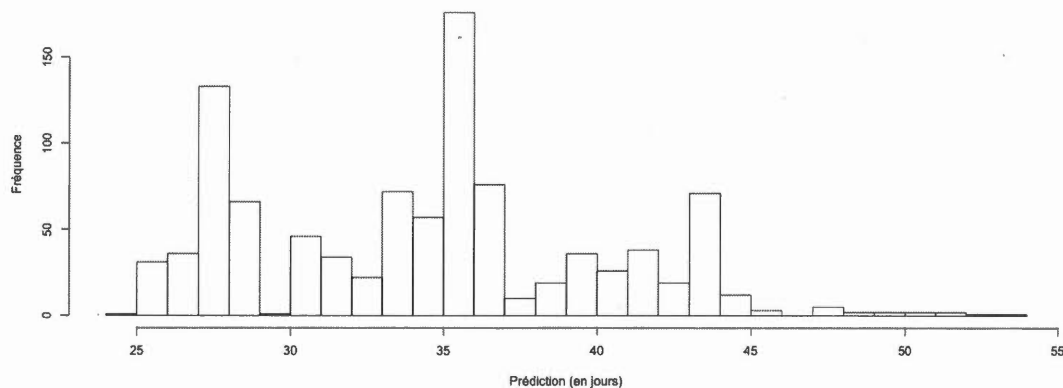


Figure 6.2 Histogramme des prédictions du modèle *Combined* sur les scénarios de la simulation DeBase

6.2 Prédire pour évaluer les risques

Les 1 000 scénarios d'une simulation sont utilisés pour prédire le délai total en sol africain. Le modèle prédictif *Combined*, entraîné sur toutes les données de la BD1, est utilisé pour prédire chaque scénario d'une simulation. La Figure 6.2 présente la distribution des prédictions pour la simulation DeBase. Les délais prédits sont compris entre 24 jours et 53 jours et ne suivent pas une loi de distribution de probabilité facilement identifiable. La valeur la plus souvent prédite (le mode) est de 36 jours. Sachant que le délai moyen en sol africain de toutes les observations de BD1 est de 31,77 jours, nous croyons qu'un délai de 40 jours ou plus est un délai risqué. Dans ce contexte, le risque est calculé par la probabilité que le délai soit supérieur ou égal à 40 jours. Avec 184 scénarios considérés risqués, parmi les 1 000 scénarios de la simulation DeBase, le risque est de 18,4%.

La Figure 6.3 présente la distribution des prédictions faites sur les scénarios des dix simulations développées pour analyser les variables importantes. Pour la ma-

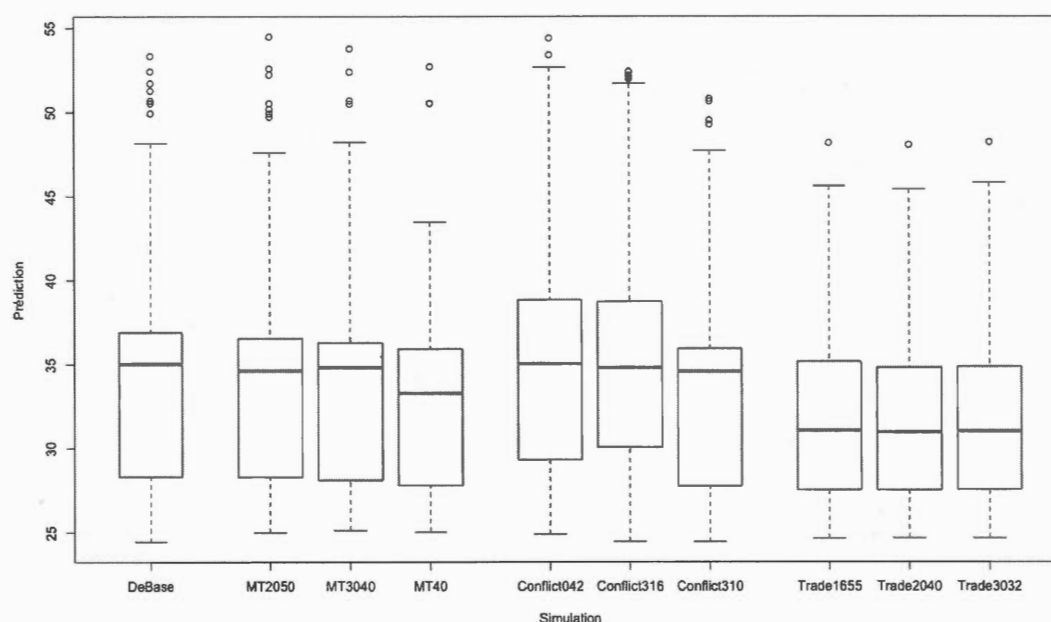


Figure 6.3 Boîtes à moustaches comparant les prédictions du modèle *Combined* en fonction de l'intervalle de sélection des variables incertaines les plus importantes

majorité des simulations, lorsque l'étendue de sélection d'une variable est réduite, l'étendue de la distribution de la prédiction l'est aussi. Dans le cas de la variable *DispatchMT*, il y a peu de différence dans la prédiction quand l'intervalle d'échantillonnage est réduit pour la simulation MT2050 ou MT3040. Par contre, lorsque la valeur de la variable est de 40 MT pour tous les scénarios, le délai médian est réduit de 1,75 jour et l'écart interquartile diminue considérablement. Cette observation concorde avec l'information illustrée par le deuxième graphique de dépendance partielle, issue du modèle *Boosting*, présenté à la Figure 5.23. Le délai est plus bas lorsque la variable *DispatchMT* prend une valeur entre 40 et 42 MT. Les expériences de simulation effectuées sur la variable *ConflictTtl* présentent aussi des observations intéressantes. Lorsque l'on modifie la base de données, l'étendue de la distribution des prédictions augmente. Plus on diminue l'intervalle de sélection de

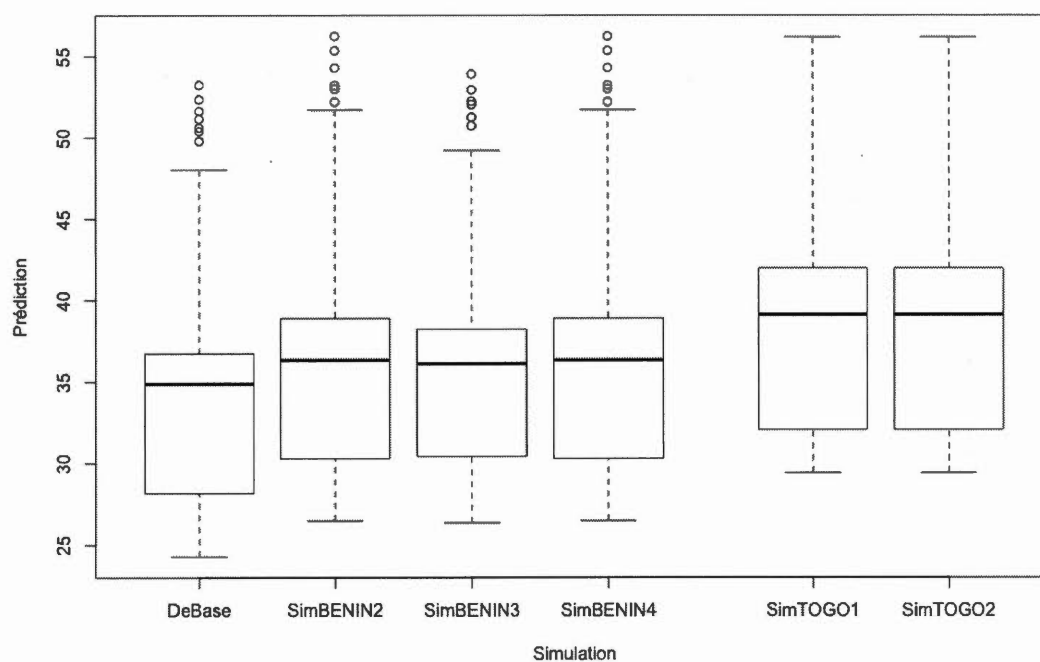
cette variable, plus l'étendue et la médiane de la distribution diminuent. L'impact des changements (modification de la base de données et réduction de l'intervalle de sélection) est d'autant plus marqué pour la variable **TtlTradeNiger**, car le délai médian est réduit de quatre jours. Nous sommes parvenus à comprendre cet impact grâce à une analyse des scénarios risqués de la simulation DeBase. Dans 75% des scénarios risqué, la variable **TtlTradeNiger** prend une valeur inférieure à 160 000\$ et dans 90%, elle est inférieure à 190 000\$. Ainsi lorsque les activités commerciales du pays sont basses, les délais sont plus longs. Cette observation est contre-intuitive sachant que nous avons collecté ces données afin de traduire la congestion au port et aux douanes, et le niveau de demande des transitaires. Dans BD2 les échanges commerciaux sont très bas pour les années 2007 à 2010 tandis que dans la BD1, couvrant les dons reçus entre mai 2014 et avril 2016, la valeur minimale de la variable **TtlTradeNiger** est 161 697\$. Changer l'intervalle de sélection de la variable, en échantillonnant dans BD1, n'engendre presque pas de changement sur les prédictions.

Le Tableau 6.7 indique le risque des différentes simulations, calculé par la probabilité que le délai soit de 40 jours ou plus. En réduisant l'intervalle de la variable **DispatchMT**, le risque diminue légèrement. En modifiant la provenance des données de la variable **ConflictTtl**, le risque augmente, mais lorsque l'intervalle de sélection est inférieur à 12 conflits il est plus bas. Le risque est diminué de façon plus importante, soit d'au moins 12,5%, lorsque la provenance de la variable **TtlTradeNiger** est modifiée et que l'intervalle de sélection est supérieur à 160 000\$.

Les logisticiens du PAM doivent décider du port d'entrée des denrées en Afrique de l'Ouest et du transitaire responsable des opérations logistiques en sol africain. Les données historiques de BD1 comportent six transitaires, quatre sont basés à Cotonou au Bénin et deux à Lomé au Togo. Dans la simulation DeBase, c'est

Tableau 6.7 Niveau de risque des simulations

Simulation	Risque
DeBase	18.4%
MT[20,50]	16.9%
MT[30,40]	15.5%
MT40	15.9%
Conflict[0,42]	21.7%
Conflict[3,16]	20.9%
Conflict[3,10]	16.3%
Trade[16,55]	5.9%
Trade[20,40]	3.3%
Trade[30,32]	3.5%

Figure 6.4 Boîtes à moustaches comparant les prédictions du modèle *Combined* en fonction du transitaire sélectionné

le transitaire nommé BÉNIN1 qui était responsable du don. Nous avons créé cinq nouvelles simulations afin de comparer les prédictions du modèle *Combined* lorsque la variable *Forwarder* est modifiée. La Figure 6.4 montre la distribution des prédictions en fonction du transitaire responsable. Pour les trois nouvelles simulations ayant un transitaire basé au Bénin, il y a uniquement que le transitaire (variable *Forwarder*) qui est modifié. Lorsque le transitaire est basé au Togo, 21 variables doivent être modifiées par rapport à la simulation DeBase. Le port, le transitaire, le nombre de douanes et toutes les variables qui présentent une valeur moyenne ou totale pour le corridor doivent être ajustés. Parmi les données historiques de BD1, il n'y a aucune observation dont le port d'arrivée est Lomé et le mois de chargement des camions, avril ou mai. Ainsi, pour les variables *AvRainPort*, *TtlRain2M*, *ConflictTtl2M*, *FatalTtl* et *FatalTtl2M*, nous avons utilisé la moyenne des valeurs observées pour ces variables dans BD2. Comme pour la simulation DeBase, la moyenne est calculée sur les valeurs du même mois que la simulation. De plus, pour une comparaison plus juste, nous avons utilisé que les observations de BD2 qui couvrent la même période que BD1, soit de mai 2014 à avril 2016. Dans le contexte des scénarios qui ont été générés, les résultats comparatifs illustrés par la Figure 6.4 indiquent que le transitaire BÉNIN1 était le meilleur choix. Les délais prédits pour la simulation DeBase sont plus courts que pour les autres simulations. Le modèle prédictif ne fait pas de distinction entre l'option du transitaire BÉNIN2 ou BÉNIN4. Pour les mêmes 241 scénarios des simulations SimBENIN2 et SimBENIN4 le délai prédit est d'au moins 40 jours, indiquant un risque de 24,1%. Les prédictions pour la simulation SimBENIN3 sont très similaires avec un risque calculé à 23,3%. Aucune différence n'est faite par le modèle entre les deux transitaires du Togo. Le délai moyen est de 38,16 jours et le risque est de 41,8% dans les deux cas.

Nous avons démontré comment générer des scénarios pour évaluer les risques grâce

à l'utilisation des modèles prédictifs développés. Les experts du milieu et leurs anticipations des valeurs prises par les données incertaines jouent un rôle clé dans la génération de ces scénarios. Sheffi (2005) disait que la résilience de la chaîne logistique se bâtit au travers de sa flexibilité, le cas présenté en est un bon exemple. Les modèles développés peuvent être utilisés à nouveau, pour la prise de décision, lorsque de l'information plus à jour parvient aux logisticiens. Avec les résultats des modèles mis à jour, l'utilisateur pourrait décider de modifier le choix du port d'arrivée pour mitiger les risques alors que le navire est en route.

CONCLUSION

Dans ce mémoire, nous avons étudié une problématique décisionnelle rencontrée par le PAM, la division de L'ONU dédiée à lutter contre la faim dans le monde. Nous avons analysé les activités logistiques du PAM basé au Niger, un pays enclavé de l'Afrique de l'Ouest subissant les effets de catastrophes à évolutions lentes. Nous avons ciblé les activités et les décisions à prendre entre le port d'entrée et les centres de distributions.

Nous avons d'abord identifié et catégorisé les sources d'incertitude pouvant affecter les coûts, les délais et les pertes. Nous avons collecté et analysé des données empiriques provenant des systèmes d'information du PAM et de bases de données publiques. La majorité des sources d'incertitude identifiées ont été représentées, soit les sources Environnementale, Interne, Ressources, Sécuritaire et les Parties prenantes (transporteur maritime, port, transitaire et douane). La catégorie Légale ne l'a pas été faute de données historiques sur les litiges avec les transitaires, les ports et les douanes. Cette source d'incertitude pourrait être considérée dans le cadre de recherches futures. De plus, la source de type Technologique n'a pas été considérée. Avec l'évolution technologique que connaît l'Afrique et les récents changements dans les systèmes d'information du PAM, nous croyons qu'elle pourrait aussi faire l'objet d'une nouvelle étude.

Des modèles d'apprentissage statistiques supervisés pour estimer les délais de transport maritime et les délais des opérations logistiques en sol africain (manutention et transport) ont été développés. Dans les deux cas, les modèles prédictifs basés sur les arbres de régressions multiples ont offert une meilleure performance

prédictive. Cette performance a été évaluée par trois mesures d'erreurs, soit le MAE, le MAPE et le RMSE. Parmi les 69 variables explicatives collectées pour estimer le délai total en sol africain, le modèle *Combined* (utilisant les prédictions des modèles *RandomForest* et *Boosting*) a identifié sept prédicteurs importants. Les conflits, sur les corridors de transport, ont prouvé leur impact sur les délais. Quand le nombre de conflits total répertorié au cours du mois du chargement du camion est supérieur à 12 conflits, le PAM devrait prévoir des délais considérablement plus longs. Nous avons ainsi démontré l'importance d'inclure des données représentant les sources d'incertitude de type *Sécuritaire* à l'analyse. Les niveaux d'échanges commerciaux totaux (importations et exportations) ont aussi prouvé leur impact sur les délais. De manière contre-intuitive, car les données ont été collectées pour représenter la congestion, lorsque les niveaux d'échanges commerciaux du Niger sont bas les délais prédits par les modèles sont plus longs.

Enfin, une preuve de concept a été faite pour illustrer l'utilisation des modèles dans un contexte où certaines informations sont manquantes. Nous avons démontré comment générer des scénarios pour évaluer les risques. Nous avons calculé le risque par la probabilité que le délai soit supérieur à une certaine valeur jugée risquée. En générant des scénarios pour simuler les différentes options décisionnelles, nous avons pu confirmer les décisions prises dans le cadre d'un exemple réel. Les modèles développés pourraient mieux performer si un plus grand nombre de variables étaient connues lors de leur utilisation. Pour ce faire, nous recommandons au PAM de collecter des données correspondant à la date où le don est confirmé. Les données collectées publiquement, qui sont utilisées dans les modèles, pourraient être associées à cette date plutôt qu'à la date du chargement des camions au port. Ainsi, la valeur de ses variables explicatives serait plus facile à identifier par l'utilisateur lors de la prise de décision.

Nos solutions technologiques pratiques apportent une contribution scientifique

significative dans un domaine peu étudié. Elles contribuent à combler les manques identifiés par Galindo et Batta (2013), Holguín-Veras *et al.* (2012), Tang et Musa (2011) et Heckmann *et al.* (2015) dans la littérature portant sur l'aide humanitaire et la gestion des risques de la chaîne d'approvisionnement. Des recherches futures sont recommandées pour analyser les coûts et les pertes associés aux corridors de transport. Une étude portant sur les activités du PAM dans les autres pays de la région (p. ex., le Mali) serait intéressante et pourrait permettre d'analyser le port de Tema au Ghana. Pour améliorer les modèles qui ont été développés, il serait valable d'obtenir des données associées au jour de la confirmation du don. Pour la section maritime du corridor, le ETA de chaque navire serait important.

BIBLIOGRAPHIE

- Ackoff, R. L. et Sasinieni, M. W. (1968). *Fundamentals of operations research*. New York : John Wiley & Sons.
- Altay, N. et Green, W. G. (2006). Or/ms research in disaster operations management. *European journal of operational research*, 175(1), 475–493.
- Alvarenga, R., Ergun, Ö., Li, J., Mata, F., Shekhani, N. et Slaton, D. (2010). *Optimizing food aid delivery in Africa*. [Document non publié]. Atlanta : H. Milton Stewart School of Industrial and Systems Engineering, Georgia Institute of Technology.
- Anaya-Arenas, A. M., Renaud, J. et Ruiz, A. (2014). Relief distribution networks : a systematic review. *Annals of Operations Research*, 223(1), 53–79.
- Bačák, V. et Kennedy, E. H. (2018). Principled machine learning using the super learner : an application to predicting prison violence. *Sociological Methods & Research*, 1–24.
- Baghalian, A., Rezapour, S. et Farahani, R. Z. (2013). Robust supply chain network design with service level against disruptions and demand uncertainties : A real-life case. *European Journal of Operational Research*, 227(1), 199–215.
- Banks, E. (2005). *Catastrophic risk : analysis and management*. Chichester : John Wiley & Sons.
- Banomyong, R., Beresford, A. et Pettit, S. (2009). Logistics relief response model : the case of thailand's tsunami affected area. *International Journal of Services Technology and Management*, 12(4), 414–429.
- Banomyong, R. et Sopadang, A. (2010). Using monte carlo simulation to refine emergency logistics response models : a case study. *International Journal of Physical Distribution & Logistics Management*, 40(8/9), 709–721.
- Bozorgi-Amiri, A., Jabalameli, M. et Al-e Hashem, S. M. (2013). A multi-objective robust stochastic programming model for disaster relief logistics under uncertainty. *OR spectrum*, 35(4), 905–933.

- Breiman, L. (1996). Bagging predictors. *Machine learning*, 24(2), 123–140.
- Breiman, L. (2001). Random forests. *Machine learning*, 45(1), 5–32.
- Breiman, L., Friedman, J. H., Olshen, R. A. et Stone, C. J. (1984). *Classification and regression trees*. Monterey : Wadsworth International Group.
- Burtonshaw-Gunn, S. (2009). *The essential management toolbox : Tools, models and notes for Managers and Consultants*. Chichester : John Wiley & Sons.
- Cafri, G., Li, L., Paxton, E. W. et Fan, J. (2018). Predicting risk for adverse health events using random forest. *Journal of Applied Statistics*, 45(12), 2279–2294.
- Chainalytics (2014). State of truckload transportation. Shipper semi-annual release. Freight Market Intelligence Consortium.
- Chan, W. K. V., Son, Y.-J. et Macal, C. M. (2010). Agent-based simulation tutorial-simulation of emergent behavior and differences between agent-based simulation and discrete-event simulation. In *Proceedings of the 2010 Winter Simulation Conference, Baltimore, MD, USA* (p. 135–150). IEEE.
- Chen, W., Xie, X., Wang, J., Pradhan, B., Hong, H., Bui, D. T., Duan, Z. et Ma, J. (2017). A comparative study of logistic model tree, random forest, and classification and regression tree models for spatial prediction of landslide susceptibility. *Catena*, 151, 147–160.
- Chen, W., Zhang, S., Li, R. et Shahabi, H. (2018). Performance evaluation of the gis-based data mining techniques of best-first decision tree, random forest, and naïve bayes tree for landslide susceptibility modeling. *Science of the Total Environment*, 644, 1006–1018.
- Churchman, C. W., Ackoff, R. L. et Arnoff, E. L. (1957). *Introduction to operations research*. New York : John Wiley & Sons.
- Comes, T., Hiete, M., Wijngaards, N. et Schultmann, F. (2011). Decision maps : A framework for multi-criteria decision support under severe uncertainty. *Decision Support Systems*, 52(1), 108–118.
- Coppola, D. P. (2015). *Introduction to international disaster management* (3e éd.). Boston : Butterworth-Heinemann.
- Cortez, P. et Silva, A. M. G. (2008). Using data mining to predict secondary school student performance. In A. C. Brito et J. feliz Teixeira (dir.), *Proceedings of the 2008 Future Business Technology Conference at the*

- University of Porto, Porto, Portugal, April 9-11, 2008* (p. 5–12). Ostend, Belgium : EUROSIS.
- Dumitriu, C. (2006). How to build a company risk profile a methodology based upon the risks disclosed by fortune 500 companies. *Risk Analysis V : Simulation and Hazard Mitigation.*, 1, 129–142.
- Elith, J., Leathwick, J. R. et Hastie, T. (2008). A working guide to boosted regression trees. *Journal of Animal Ecology*, 77(4), 802–813.
- Fitzpatrick, T. et Mues, C. (2016). An empirical comparison of classification algorithms for mortgage default prediction : evidence from a distressed mortgage market. *European Journal of Operational Research*, 249(2), 427–439.
- Freund, Y. et Schapire, R. (1999). A short introduction to boosting. *Journal of Japanese Society for Artificial Intelligence*, 14(5), 771–780.
- Friedman, J. H. (2001). Greedy function approximation : a gradient boosting machine. *Annals of statistics*, 29(5), 1189–1232.
- Galindo, G. et Batta, R. (2013). Review of recent developments in OR/MS research in disaster operations management. *European Journal of Operational Research*, 230(2), 201–211.
- Ghadge, A., Dani, S., Chester, M. et Kalawsky, R. (2013). A systems approach for modelling supply chain risks. *Supply Chain Management : An International Journal*, 18(5), 523–538.
- Google Maps. (2018). *Google Maps*. Récupéré de <https://www.google.ca/maps/>
- Gordon, L. A., Loeb, M. P. et Tseng, C.-Y. (2009). Enterprise risk management and firm performance : A contingency perspective. *Journal of Accounting and Public Policy*, 28(4), 301–327.
- Govindan, K., Fattahi, M. et Keyvanshokoo, E. (2017). Supply chain network design under uncertainty : A comprehensive review and future research directions. *European Journal of Operational Research*, 263(1), 108–141.
- Greenwell, B., Boehmke, B., Cunningham, J. et Developers, G. (2018). *gbm : Generalized Boosted Regression Models*. R package version 2.1.4
- Hansen, M. S., Rasmussen, L. B. et Jacobsen, P. (2016). Interactive foresight simulation. *Technological Forecasting and Social Change*, 103, 214–227.

- Hastie, T., Tibshirani, R. et Friedman, J. (2009). *The elements of statistical learning : data mining, inference, and prediction* (2e éd.). New York, NY : Springer.
- Heckmann, I., Comes, T. et Nickel, S. (2015). A critical review on supply chain risk-definition, measure and modeling. *Omega*, 52, 119–132.
- Hilbert, J. P., Zasadil, S., Keyser, D. J. et Peele, P. B. (2014). Using decision trees to manage hospital readmission risk for acute myocardial infarction, heart failure, and pneumonia. *Applied health economics and health policy*, 12(6), 573–585.
- Holguín-Veras, J., Jaller, M., Van Wassenhove, L. N., Pérez, N. et Wachtendorf, T. (2012). On the unique features of post-disaster humanitarian logistics. *Journal of Operations Management*, 30(7), 494–506.
- Humanitarian Data Exchange. (2018). *Armed Conflict Location and Event Data Project (ACLED)*. Récupéré de https://data.humdata.org/organization/acled?sort=metadata_modified+desc
- Hyndman, R., Athanasopoulos, G., Bergmeir, C., Caceres, G., Chhay, L., O'Hara-Wild, M., Petropoulos, F., Razbash, S., Wang, E. et Yasmeeen, F. (2018). *forecast : Forecasting functions for time series and linear models*. R package version 8.4
- Iakovou, E., Vlachos, D., Keramydas, C. et Partsch, D. (2014). Dual sourcing for mitigating humanitarian supply chain disruptions. *Journal of Humanitarian Logistics and Supply Chain Management*, 4(2), 245–264.
- Index for risk management. (2018). *INFORM country risk profiles for 191 countries*. Récupéré de <http://www.inform-index.org/Countries/Country-profiles/iso3/NER>
- International Trade Center. (2018). *International trade statistics 2001-2018*. Récupéré de <http://www.intracen.org/itc/market-info-tools/trade-statistics/>
- Jahre, M., Kembro, J., Rezvanian, T., Ergun, O., Håpnes, S. J. et Berling, P. (2016). Integrating supply chains for emergencies and ongoing operations in unhcr. *Journal of Operations Management*, 45, 57–72.
- James, G., Witten, D., Hastie, T. et Tibshirani, R. (2013). *An introduction to statistical learning : with application in R*. New York : Springer.
- Jüttner, U., Peck, H. et Christopher, M. (2003). Supply chain risk management : outlining an agenda for future research. *International Journal of Logistics : Research and Applications*, 6(4), 197–210.

- Kleindorfer, P. R. et Saad, G. H. (2005). Managing disruption risks in supply chains. *Production and operations management*, 14(1), 53–68.
- Klibi, W. et Martel, A. (2012). Scenario-based supply chain network risk modeling. *European Journal of Operational Research*, 223(3), 644–658.
- Klibi, W., Martel, A. et Guitouni, A. (2010). The design of robust value-creating supply chain networks : a critical review. *European Journal of Operational Research*, 203(2), 283–293.
- Knemeyer, A. M., Zinn, W. et Eroglu, C. (2009). Proactive planning for catastrophic events in supply chains. *Journal of operations management*, 27(2), 141–153.
- Kovács, G. et Spens, K. M. (2007). Humanitarian logistics in disaster relief operations. *International Journal of Physical Distribution & Logistics Management*, 37(2), 99–114.
- Krejci, C. C. (2015). Hybrid simulation modeling for humanitarian relief chain coordination. *Journal of Humanitarian Logistics and Supply Chain Management*, 5(3), 325–347.
- Kunz, N. et Reiner, G. (2012). A meta-analysis of humanitarian logistics research. *Journal of Humanitarian Logistics and Supply Chain Management*, 2(2), 116–147.
- Kutner, M. H., Nachtsheim, C. J., Neter, J., Li, W. et al. (2005). *Applied linear statistical models* (5e éd.). New York : McGraw-Hill Irwin.
- Lark, J. (2015). *ISO 31000 - Management du risque - Guide pratique pour les PME*. ISO, ITC et ONUDI, Suisse.
- Leiras, A., de Brito Jr, I., Queiroz Peres, E., Rejane Bertazzo, T. et Tsugunobu Yoshida Yoshizaki, H. (2014). Literature review of humanitarian logistics research : trends and challenges. *Journal of Humanitarian Logistics and Supply Chain Management*, 4(1), 95–130.
- Liaw, A. et Wiener, M. (2002). Classification and regression by randomforest. *R News*, 2(3), 18–22. Récupéré de <https://CRAN.R-project.org/doc/Rnews/>
- Logistics Capacity Assesment. (2014). *Niger road network*. Récupéré de <http://dlca.logcluster.org/display/public/DLCA/2.3+Niger+Road+Network;jsessionid=9B94FC5382C26E3F2AD9105B0BDC16C5>
- Logistics Capacity Assesment. (2018). *Niger road network*. Récupéré de <https://dlca.logcluster.org/display/public/DLCA/2.3+Niger+Road+Network>

- Masci, C., Johnes, G. et Agasisti, T. (2018). Student and school performance across countries : A machine learning approach. *European Journal of Operational Research*, 269(3), 1072–1085.
- McCluskey, W. J., Daud, D. Z. et Kamarudin, N. (2014). Boosted regression trees : An application for the mass appraisal of residential property in malaysia. *Journal of Financial Management of Property and Construction*, 19(2), 152–167.
- Miguéis, V. L., Freitas, A., Garcia, P. J. et Silva, A. (2018). Early segmentation of students according to their academic performance : A predictive modelling approach. *Decision Support Systems*, 115, 36–51.
- Milborrow, S. (2018). *rpart.plot : Plot 'rpart' Models : An Enhanced Version of 'plot.rpart'*. R package version 3.0.5
- Miller, K. D. (1992). A framework for integrated risk management in international business. *Journal of international business studies*, 23(2), 311–331.
- Mullainathan, S. et Spiess, J. (2017). Machine learning : an applied econometric approach. *Journal of Economic Perspectives*, 31(2), 87–106.
- Natarajarathinam, M., Capar, I. et Narayanan, A. (2009). Managing supply chains in times of crisis : a review of literature and insights. *International Journal of Physical Distribution & Logistics Management*, 39(7), 535–573.
- Nathan Associates Inc. (2013). *Logistics Cost Study of Transport Corridors in Central and West Africa*. [Rapport final]. Africa Transport Unit : The World Bank.
- Oneto, L., Fumeo, E., Clerico, G., Canepa, R., Papa, F., Dambra, C., Mazzino, N. et Anguita, D. (2018). Train delay prediction systems : A big data analytics perspective. *Big data research*, 11, 54–64.
- Ouallouche, F., Lazri, M. et Ameer, S. (2018). Improvement of rainfall estimation from msg data using random forests classification and regression. *Atmospheric Research*, 211, 62–72.
- Overstreet, R. E., Hall, D., Hanna, J. B. et Kelly Rainer Jr, R. (2011). Research in humanitarian logistics. *Journal of Humanitarian Logistics and Supply Chain Management*, 1(2), 114–131.
- Poojari, C. A., Lucas, C. et Mitra, G. (2008). Robust solutions and risk measures for a supply chain planning problem under uncertainty. *Journal of the Operational Research Society*, 59(1), 2–12.

- Ports.com. (2018). *Sea route and distance*. Récupéré de <http://ports.com/sea-route/>
- Probst, P. et Boulesteix, A.-L. (2018). To tune or not to tune the number of trees in random forest. *Journal of Machine Learning Research*, 18(181), 1–18.
- R Core Team (2018). *R : A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria
- Rancourt, M.-È., Bellavance, F. et Goentzel, J. (2014). Market analysis and transportation procurement for food aid in ethiopia. *Socio-Economic Planning Sciences*, 48(3), 198–219.
- Rancourt, M.-È., Cordeau, J.-F., Laporte, G. et Watkins, B. (2015). Tactical network planning for food aid distribution in kenya. *Computers & Operations Research*, 56, 68–83.
- Rao, S. et Goldsby, T. J. (2009). Supply chain risks : a review and typology. *The International Journal of Logistics Management*, 20(1), 97–123.
- Rebollo, J. J. et Balakrishnan, H. (2014). Characterization and prediction of air traffic delays. *Transportation research part C : Emerging technologies*, 44, 231–241.
- Rodrigue, J.-P., Comtois, C. et Slack, B. (2009). *The geography of transport systems* (2e éd.). New York : Routledge.
- RStudio Team (2016). *RStudio : Integrated Development Environment for R*. RStudio, Inc., Boston, MA
- Sadghiani, N. S., Torabi, S. et Sahebjamnia, N. (2015). Retail supply chain network design under operational and disruption risks. *Transportation Research Part E : Logistics and Transportation Review*, 75, 95–114.
- Sahinyazan, F. G., Verter, V. et Rancourt, M.-È. (2017). *Price Flexible Transportation Procurement Contracts for Food Aid Delivery in Developing Countries*, volume CIRRELT-2017-65. [Rapport technique]. Montreal : Centre interuniversitaire de recherche sur les réseaux d'entreprise, la logistique et le transport (CIRRELT).
- Salehi, M. et Davoudi Pour, M. (2016). Bankruptcy prediction of listed companies on the tehran stock exchange. *International Journal of Law and Management*, 58(5), 545–561.

- Sarykalin, S., Serraino, G. et Uryasev, S. (2008). Value-at-risk vs. conditional value-at-risk in risk management and optimization. In *State-of-the-Art Decision-Making Tools in the Information-Intensive Age* (p.270–294). Informs.
- Sawik, T. (2011). Selection of supply portfolio under disruption risks. *Omega*, 39(2), 194–208.
- Schapire, R. E. (2003). *The boosting approach to machine learning : An overview*, Dans C. C. B. M. David D.Denison, Mark H.Hansen et B. Yu (dir.). *Nonlinear estimation and classification. Lecture Notes in Statistics*, volume 171, (p. 149–171). New York : Springer.
- Schiltz, F., Masci, C., Agasisti, T. et Horn, D. (2018). Using regression tree ensembles to model interaction effects : a graphical approach. *Applied Economics*, 50(58), 6341–6354.
- Schloerke, B., Crowley, J., Cook, D., Briatte, F., Marbach, M., Thoen, E., Elberg, A. et Larmarange, J. (2018). *GGally : Extension to 'ggplot2'*. R package version 1.4.0
- Schwartz, P. (1992). The art of the long view. *Director*, 45(10), 78.
- Sheffi, Y. (2005). Manage risk through resilience. *CEO Magazine*, 214, 28–29.
- Shmueli, G., Bruce, P. C., Yahav, I., Patel, N. R. et Lichtendahl Jr, K. C. (2017). *Data mining for business analytics : concepts, techniques, and applications in R*. New Jersey : John Wiley & Sons.
- Simons, R. L. (1999). *A note on identifying strategic risk*. Harvard Business School.
- Singhal, P., Agarwal, G. et Mittal, M. L. (2011). Supply chain risk management : review, classification and future research directions. *International Journal of Business Science & Applied Management*, 6(3), 15–42.
- Slywotzky, A. J. et Drzik, J. (2005). Countering the biggest risk of all. *Harvard business review*, 83(4), 78–88.
- Smith, P. F., Ganesh, S. et Liu, P. (2013). A comparison of random forest regression and multiple linear regression for prediction in neuroscience. *Journal of neuroscience methods*, 220(1), 85–91.
- Sobolev, B. G., Sanchez, V. et Vasilakis, C. (2011). Systematic review of the use of computer simulation modeling of patient flow in surgical care. *Journal of medical systems*, 35(1), 1–16.

- Sundaramoorthi, D., Chen, V. C., Rosenberger, J. M., Kim, S. B. et Buckley-Behan, D. F. (2009). A data-integrated simulation model to evaluate nurse-patient assignments. *Health care management science*, 12(3), 252-268.
- Tang, C. S. (2006). Perspectives in supply chain risk management. *International Journal of production economics*, 103(2), 451-488.
- Tang, O. et Musa, S. N. (2011). Identifying risk issues and research advancements in supply chain risk management. *International journal of production economics*, 133(1), 25-34.
- Tchankova, L. (2002). Risk identification – basic stage in risk management. *Environmental Management and Health*, 13(3), 290-297.
- Teravaninthorn, S. et Raballand, G. (2009). *Transport Prices and Costs in Africa : A Review of the International Corridors*. Washington, DC : The World Bank.
- The World Bank Group. (2018). *Climate change knowledge portal*. Récupéré de <http://sdwebx.worldbank.org/climateportal/>
- Therneau, T. et Atkinson, B. (2018). *rpart : Recursive Partitioning and Regression Trees*. R package version 4.1-13
- Transparency international. (2018). *Corruption Perception Index 2017*. Récupéré de https://www.transparency.org/news/feature/corruption_perceptions_index_2017
- Tukey, J. W. (1977). *Exploratory data analysis*. Massachusetts : Addison-Wesley.
- United Nations Office for the Coordination of Humanitarian Affairs (2016). Plan de réponse humanitaire. janvier - décembre 2017. Récupéré de <https://www.humanitarianresponse.info/operations/niger>
- United Nations Office for the Coordination of Humanitarian Affairs (2017). Aperçu de la situation humanitaire mondiale 2018. Récupéré de http://www.unocha.org/sites/unocha/files/dms/GHO_2018_FR-Abridged.pdf
- United Nations Office for the Coordination of Humanitarian Affairs. (2018). *Niger : Situation des inondations*. Récupéré de https://reliefweb.int/sites/reliefweb.int/files/resources/snasphot_inondation_27082018.pdf
- U.S. Energy Information Administration. (2018). *Petroleum and other liquids*. Récupéré de <https://www.eia.gov/dnav/pet/hist/LeafHandler.ashx?n=PET&s=RB RTE&f=M>

- Varian, H. R. (2014). Big data : New tricks for econometrics. *Journal of Economic Perspectives*, 28(2), 3–28.
- Wang, Z., Liang, M. et Delahaye, D. (2018). A hybrid machine learning model for short-term estimated time of arrival prediction in terminal manoeuvring area. *Transportation Research Part C : Emerging Technologies*, 95, 280–294.
- Wei, Y., Xu, W., Fan, Y. et Tasi, H.-T. (2002). Artificial neural network based predictive method for flood disaster. *Computers & industrial engineering*, 42(2-4), 383–390.
- World Food Programme (2015). Politique en matière de gestion globale des risques. Rome : Session annuelle du Conseil d'administration. Récupéré de <http://documents.wfp.org/stellent/groups/public/documents/eb/wfpdoc063948.pdf>
- World Food Programme (2016). Fighting hunger worldwide : Enhancing the resilience of chronically vulnerable populations in niger. Standard project report. (PRRO 200583). Republic of (NE) : World Food Program. Récupéré de https://docs.wfp.org/api/documents/7c2a8dc48c4b4126bd30c53d3aabbf15b/download/?_ga=2.141317426.1632739240.1494097104-49524839.1494097091
- World Food Programme. (2017a). *Operation*. Récupéré de <http://www.wfp.org/operations>
- World Food Programme. (2017b). *WFP Niger country brief*. Récupéré de <http://www1.wfp.org/countries/niger>
- Zupancic, J. A., Richardson, D. K., O'Brien, B. J., Schmidt, B. et Weinstein, M. C. (2003). Daily cost prediction model in neonatal intensive care. *International journal of technology assessment in health care*, 19(2), 330–338.