[Page de garde]

HEC MONTRÉAL

Prédiction de la demande et répartition des ressources pour un service préhospitalier d'urgence

par

Lotfi Chikh

Sciences de la gestion
(Option Gestion des opérations)

Mémoire présenté en vue de l'obtention
du grade de maîtrise ès sciences en gestion
(M. Sc.)

Décembre 2019

Résumé

Le but principal de cette étude est de développer une approche capable de prévoir le nombre d'appels par période de la journée pour un service préhospitalier d'urgence et de déterminer le nombre d'ambulances nécessaires pour répondre à cette demande. Elle vise également à déterminer comment gérer efficacement les ressources pour atteindre les citoyens dans le besoin dans les meilleurs délais. Quatre modèles de prévision basés sur les arbres de régression et des fonctions de lissage exponentiel ont été proposés pour prévoir les appels reçus. Ces prévisions par zones et par heures de la journée serviront à déterminer le nombre d'appels probables en utilisant la distribution de Poisson. Par la suite, un modèle d'optimisation a été développé pour déterminer les emplacements optimaux des ambulances selon plusieurs critères en termes de couverture. Enfin, une simulation a été utilisée pour valider les résultats obtenus pour différents scénarios en considérant des objectifs d'ordre économique soit, par exemple le nombre d' ambulances utilisées, ou de performances, par exemple le temps de réponse moyen. Les résultats ont permis de montrer qu'avec une stratégie de partage de ressources, en planifiant un niveau de service de 95%, il est possible d'obtenir un temps de réponse moyen de 7,85 minutes avec un pourcentage d'appels répondus en bas du seuil standard de 10 minutes de 61%. D'une autre manière et avec la même stratégie, il est possible aussi d'atteindre un temps de réponse moyen de 6,97 minutes et 70% d'appels répondus en bas de 10 minutes, en augmentant le niveau de couverture à 99%. Les résultats ont permis aussi de montrer qu'il est possible de réduire le nombre de ressources jusqu'à un certain pourcentage pour trouver un équilibre entre le nombre de ressources utilisées et le service offert à la population.

L'approche développée dans le cadre de ce travail pourrait être utilisée dans d'autres contextes et d'autres régions pour aider le gestionnaire à prendre les décisions concernant la gestion des ressources selon leurs objectifs.

Table des matières

Liste des tableaux et des figures

Liste des tableaux

Liste des figures

Avant-propos

Un service préhospitalier d'urgence (SPU) représente un maillon essentiel dans le processus d'intervention auprès des personnes dans le besoin, car, lorsque les premiers secours sont donnés à temps, on peut sauver des vies humaines ou minimiser l'impact de l'incident. Dans ce contexte, il est essentiel d'avoir une bonne planification pour faire face à une demande souvent imprévisible et changeante à la fois. Il faut donc se prémunir de stratégies de gestion et d'outils afin de mener à bien les interventions et offrir un service de qualité.

Notre recherche porte sur le développement d'un outil de décision pour l'analyse de la demande et l'allocation des ressources afin d'obtenir un bon arbitrage entre le service à la population et les différents coûts associés à la gestion d'un tel service. Pour ce faire, plusieurs scénarios de couvertures seront développés puis analysés par simulation afin de déterminer la stratégie à favoriser dans différents contextes.

Remerciements

Je tiens à remercier tous ceux qui ont contribué de près ou de loin à la réalisation de ce travail. J'aimerais surtout remercier Yossiri Adulyasak et Valérie Bélanger pour leur patience, leurs conseils et leur suivi constant. Je tiens également à remercier mes parents, ma femme et tous les membres de ma famille et belle famille qui m'ont toujours encouragé et qui ont eu confiance en mes capacités tout au long de ce processus.

Introduction

Le service préhospitalier d'urgence (SPU) est le premier intervenant auprès des citoyens en cas d'urgence ou d'incidents. Il assure les premiers secours ainsi que le transport vers un hôpital à la suite d'un appel d'urgence. Le caractère social et économique de ce secteur fait en sorte que le service doit intervenir rapidement auprès des appelants. Il est donc important de bien étudier la demande future et localiser efficacement les ambulances afin de répondre aux besoins des patients dans de meilleurs délais.

Les demandes pour les SPUs arrivent généralement par téléphone à partir d'un numéro d'urgence (911 pour l'Amérique du Nord) vers un centre de traitement qui évalue l'urgence de l'appel. Des pompiers et une ambulance avec une équipe paramédicale est envoyée par la suite vers les lieux de l'incident pour donner les premiers soins, évaluer la situation sur place, puis déterminer si le patient doit être transporté vers un hôpital. Si c'est le cas, l'équipe paramédicale assure le transport vers l'hôpital de destination et reste au chevet du patient jusqu'à ce que ce dernier soit pris en charge par l'hôpital. Par la suite, si aucun appel est en attente, l'équipe est renvoyée vers un poste d'attente. La figure 1 démontre le processus de réponse des appels pour le SPU tel que présenté dans Reuter-Oppermann, van den Berg et Vile (2017) .

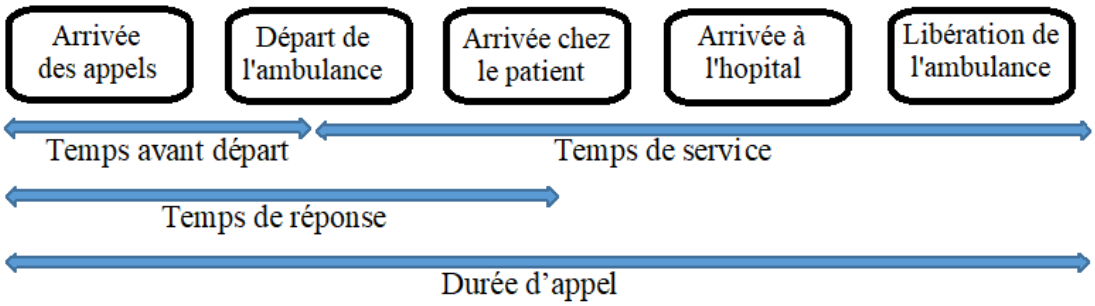

Figure 1 Processus de réponse (adapté de Reuter-Oppermann, van den Berg et Vile (2017))

Le SPU doit respecter des standards de performance fixés par les opérateurs des services ambulanciers, ou bien par le gouvernement (van Barneveld, Bhulai et van der Mei, 2016), comme la satisfaction de la clientèle, le taux d'appels répondus, le temps de réponse, pour

ne nommer que quelques exemples. La mesure la plus importante pour le SPU est le temps de réponse (Chen *et al.*, 2015; Peleg et Pliskin, 2004). Le temps de réponse est défini comme le temps écoulé entre la réception de l'appel au centre de traitement et l'arrivée de l'ambulance sur les lieux de l'incident. Les SPUs doivent donc s'assurer de répondre aux appels dans les délais adéquats. Cela amène toutefois certains coûts, notamment les frais liés aux contrats de service, les frais engagés pour l'achat ou pour la location des équipements, de même que les dépenses administratives et opérationnelles.

La planification des SPUs passe par trois niveaux. Le niveau stratégique détermine la localisation des garages et des centres opérationnels qui gèrent les appels provenant du centre 9-1-1. Il détermine aussi le nombre d'ambulances nécessaires et s'intéresse à la main-d'œuvre nécessaire. Le niveau tactique s'occupe des horaires de travail, de la formation d'équipes et de l'affectation des ambulances, alors que le niveau opérationnel s'intéresse aux règles de gestion, incluant l'affectation des appels aux équipes pour la répartition, les politiques de redéploiement et le choix du centre hospitalier (Bélanger, Ruiz et Soriano, 2019; van Barneveld, Bhulai et van der Mei, 2016). En plus des enjeux complexes au niveau opérationnel, les gestionnaires des SPUs font face aussi à des enjeux sociaux et économiques. Ils doivent donc trouver un équilibre entre la gestion de la performance et la gestion de coût (Setzler, Saydam et Park, 2009). Vile *et al.* (2012), ont cité qu'en 2007, les dépenses en heures supplémentaires des ambulanciers ont dépassé les 3 000 000 £ (welsh Ambulance Services NHS Trust). Ils rajoutent que pour faire face à une demande accrue tout en maintenant les coûts à un niveau acceptable, il serait nécessaire de faire des prévisions exactes. Dans ce contexte, il serait essentiel de bien étudier et analyser la demande, car c'est à partir de là que l'ensemble des décisions pourront être prises. L'incertitude quant au nombre d'appels reçus et leur localisation complexifie toutefois le processus de planification. Des méthodes de prévision peuvent aider à faire face à cette incertitude.

Pour déployer de façon optimale le personnel et les véhicules nécessaires pour les interventions, le défi d'un gestionnaire serait donc d'arriver à égaler la demande réelle à la demande estimée. À partir des prévisions, les gestionnaires doivent déterminer les besoins en ressources puis les répartir de manière à offrir un service adéquat à la

population. Il est donc nécessaire de bien anticiper la demande afin de prendre les bonnes décisions et parvenir à gérer efficacement les ressources. Cela nous amène à notre question de recherche :

- De quelle manière un service ambulancier d'urgence peut prévoir la demande future et comment ces prévisions peuvent être utilisées pour optimiser la répartition des ressources de façon à minimiser les coûts d'exploitation tout en gardant un niveau de qualité et de couverture adéquat?

Malgré les différentes recherches menées sur la prévision de la demande pour les SPUs, il n'existe pas une meilleure façon de faire, car la meilleure méthode dépend du cas étudié et des critères d'intérêt (Tyralis et Papacharalampous, 2017). Par contre, pour faire une bonne prévision, il serait nécessaire de chercher la précision, la stabilité et l'efficacité (Lin Lin *et al.*, 2017b). Dans le cadre de ce mémoire, nous nous intéressons d'abord à effectuer des prévisions de la demande à l'aide de différents modèles. Les résultats du meilleur modèle serviront à calculer le nombre d'appels probables. Par la suite, nous utiliserons un modèle d'optimisation (*set covering*) pour déterminer la localisation des postes d'attente et le nombre d'ambulances à affecter à chaque poste d'attente de façon à couvrir tout le territoire dans des délais raisonnables. Une simulation sera utilisée pour valider différents scénarios en termes de localisation et d'affectation des ressources, et ce, en présence d'incertitude. La méthodologie développée sera validée grâce aux données du SPU du comté de Marin, en Californie (États-Unis). Ces données sont disponibles et publiques pour une période allant du 1er mars 2013 au 30 juin 2017[1]. D'après le recensement de 2015, la population de ce comté avait atteint 261 221 habitants pour une superficie totale de 2 145 km², incluant, 1 336 km² de terre et 799 km² d'eau. Les habitants sont surtout regroupés sur le côté est du comté. Le reste du territoire contient de nombreuses localités qui dépendent de l'agriculture et du tourisme.

[1] https://data.marincounty.org/Public-Health/Emergency-Medical-Service-Incidents/swth-izpe (consulté le 15 novembre 2017)

Le présent mémoire inclut six chapitres. Le premier chapitre présente une revue de la littérature sur les modèles utilisés pour la prévision de la demande, la localisation des véhicules et la simulation dans le contexte d'un SPU. Dans le deuxième chapitre, nous exposons la méthodologie utilisée dans le cadre de notre étude ainsi que les outils choisis. Le troisième chapitre porte sur l'analyse des données du SPU du comté du Marin, Californie et l'analyse des variables utilisées. Dans le quatrième chapitre, les différents modèles et techniques de choix de variables utilisées pour faire des prévisions sur la demande sont présentés. Le cinquième chapitre est consacré à la localisation et l'affectation des ressources, alors que le sixième et dernier chapitre est consacré à la simulation du SPU selon différents scénarios. Enfin, une conclusion est donnée pour résumer les objectifs de l'étude, ainsi que les résultats obtenus avec la simulation pour des fins de comparaisons et validations.

Chapitre 1
Revue de la littérature

La revue de la littérature portera sur les études antérieures et les recherches effectuées dans le domaine de la prévision de la demande, la localisation des ressources et la simulation. Dans un premier temps, nous présentons les méthodes de prévisions utilisées pour prévoir la demande dans divers contextes, et plus particulièrement dans le contexte d'un SPU. Par la suite, nous présentons les principales recherches portant sur la localisation des ambulances et l'affectation des ressources ainsi que les modèles utilisés dans ce contexte. Enfin, nous présentons quelques études sur la simulation des SPUs avec différents types de données.

1.1 Prévision de la demande

Prévoir les arrivées des appels pour un SPU n'est pas une tâche facile. Certains SPUs utilisent les données historiques afin de déterminer les taux d'arrivées des appels pour chaque période pour des fins de planification et de budgétisation (Trudeau *et al.*, 1989) ou pour la localisation des postes d'attente (Chen *et al.*, 2015). Ces techniques quantitatives décomposent l'historique des appels en séries de données chronologiques temporelles. Dans un contexte de prévision de la demande d'un SPU, le choix de variables constitue un critère d'entrée important qui peut avoir un effet direct sur la performance du modèle (Gunes et Szechtman, 2005), car plus les variables ont un pouvoir prévisionnel élevé, moins les modèles génèrent des erreurs de prévisions. Dans certains cas, ces variables peuvent être de nature climatique, démographiques ou sociales (Kamenetzky, Shuman et Wolfe, 1982). Généralement, ces variables sont utilisées pour expliquer un phénomène, par exemple : le taux élevé d'appels par des personnes âgées plus qu'autre type de raison (McConnel et Wilson, 1998), ou bien pour déterminer la demande future en général. Les variables (temps) sont aussi utilisées dans le même contexte pour prévoir la demande grâce à son évolution dans une série chronologique. Parfois, un mélange de variables peut augmenter le pouvoir prévisionnel des modèles, comme dans le cas de McCarthy et al. (2008), qui ont fusionné des données climatiques aux données temporelles pour étudier l'arrivée des patients aux urgences.

Trois méthodes sont principalement utilisées dans le contexte des SPUs : des méthodes utilisant des variables exogènes ou causales (Kvålseth et Deems, 1979), des méthodes utilisant des variables endogènes ou extrapolatives (Channouf *et al.*, 2007) et une combinaison des deux (Babu et Reddy, 2014).

1.1.1 Méthodes exogènes ou causales

Les méthodes exogènes ou causales ne dépendent pas seulement du comportement ou de l'historique de la variable dépendante, c'est-à-dire la variable qu'on prédit, mais aussi du comportement des variables prédictives, c'est-à-dire les variables qu'on utilise pour prédire la variable dépendante. McConnel et Wilson (1998) ont utilisé des variables sociodémographiques pour prédire le nombre d'appels et le type d'urgence médicale à traiter, par exemple les cas d'arrêt cardiaque. Avec les méthodes causales, il est important de bien choisir les variables qui ont un pouvoir prédictif et fournissent des prévisions précises (Tyralis et Papacharalampous, 2017).

1.1.1.1 Les modèles de régression linéaire

Les modèles de régression linéaire sont des modèles simples qui utilisent la méthode causale pour déterminer la relation linéaire qui existe entre les variables prédictives et les variables dépendantes. Les travaux de Aldrich, Hisserich et Lave (1971) montrent comment les variables socio-économiques permettent de prédire le nombre de demandes d'interventions d'urgence ainsi que les types de demandes à l'aide de la régression des moindres carrés. Micheletti *et al.* (2010) ont utilisé des variables exogènes et un modèle de régression linéaire pour estimer le nombre moyen d'appels par jour pour les services préhospitaliers d'urgence de la ville de Milan (Italie). Ils ont utilisé des variables climatiques en plus d'autres variables reliées à des incidences d'épidémies et de pureté de l'air. Leur but était de fournir des résultats pour appuyer les décisions du centre de traitement des appels quant à sa planification à moyen et à long terme. Kvålseth et Deems (1979) ont utilisé des modèles similaires pour étudier la demande non urgente des services d'urgences médicales de la ville d'Atlanta, en Georgia (États-Unis). En plus des variables à caractères démographiques et socio-économiques, ils ont pris en compte d'autres variables comme des variables qui décrivent la cause de l'appel. L'objectif de leur étude

était de développer des modèles linéaires multi variés pour la demande globale et pour chaque catégorie d'urgences médicales.

Durant les années 60 et 70, les chercheurs se sont d'abord intéressés aux méthodes de régression linéaire pour prédire les arrivées d'appels pour les SPUs, puis, à travers le temps, d'autres méthodes ont vu le jour pour donner naissance à des modèles et algorithmes plus développés (Bontempi, Taieb et Le Borgne, 2012). Ce sont les modèles non linéaires construits à partir d'algorithmes et utilisant des combinaisons très complexes, mais qui sont faciles à utiliser et à interpréter. Cela ne veut pas dire qu'ils sont meilleurs que les précédents. En effet, des études ont démontré que certains modèles, bien que moins complexes, peuvent être plus performants que d'autres selon le contexte d'étude (Gunes et Szechtman, 2005; Setzler, Saydam et Park, 2009).

1.1.1.2 Modèles d'apprentissage automatique

Des méthodes non linéaires ont aussi été utilisées dans le traitement et l'analyse de données de séries chronologiques. Les plus connues sont les méthodes d'ensembles appelés aussi classificateurs ensemblistes. Ils utilisent des combinaisons de prévisions individuelles afin de prévoir ou classifier de nouvelles données. Les modèles les plus utilisés pour les SPUs sont les réseaux de neurones artificiels (RNA) (Hinton, Osindero et Teh, 2006) et les forêts aléatoires (FA) (Breiman, 2001). Quelques chercheurs se sont servi des réseaux de neurones pour étudier la demande des SPUs. Ils ont ensuite comparé cette méthode à d'autres techniques souvent utilisées dans le même contexte comme les méthodes d'auto-régression linéaires. Denton (1995) nous propose une comparaison entre les méthodes basées sur la régression et ceux basées sur les réseaux de neurones pour démontrer que l'élimination des ambiguïtés dans les variables prédictives permet aux réseaux de neurones d'apprendre avec succès sur les modèles d'apprentissages. Cela dit, dans des conditions idéales, les réseaux de neurones offrent de meilleures prévisions que ceux des modèles de régression. Les méthodes utilisant les RNA peuvent donc être une alternative intéressante aux méthodes de régression. Setzler, Saydam et Park (2009) ont utilisé un RNA pour la prévision de la demande des services ambulanciers, puis l'ont comparé à la méthode de moyenne mobile. Dans ce contexte, des zones de 2x2 et de 4x4

miles ont été utilisées avec des périodes d'une heure et trois heures. Ils ont conclu que le niveau de précision dépendait des niveaux de granularité temporelle et spatiale, mais ne dépendait pas des modèles utilisés.

Bien que les forêts aléatoires (FA) soient moins présentes dans les études qui ne sont intéressées aux SPUs, ils ont été souvent utilisés dans le contexte de séries chronologiques et de sélection de variables. Ils peuvent donc être une bonne alternative aux méthodes linéaires. Les FA sont aussi des techniques d'apprentissage automatique qui sont beaucoup utilisées de nos jours pour les problèmes de régression et de classification. Tyralis et Papacharalampous (2017) ont utilisé les forêts aléatoires afin de sélectionner des variables prédictives et tester la performance de leur modèle en utilisant des données de séries temporelles courtes. Ils ont montré que les FA peuvent être un outil fiable pour la prévision des séries chronologiques et que leurs performances seraient reliées à la sélection de variables. Comme le choix de variables est important (Genuer, Poggi et Tuleau-Malot, 2010), la sélection du nombre d'arbres dans les forêts aléatoires peut avoir un impact sur le niveau de performance du modèle. Contrairement au fait de croire que l'ajout d'arbres réduit le taux d'erreurs, le nombre limité d'arbres peut donner de meilleurs résultats de prévision (Probst et Boulesteix, 2017).

1.1.2 Méthodes endogènes ou extrapolatives

À la différence de la méthode causale qui utilise la corrélation entre les variables prédictives et la variable dépendante, les méthodes endogènes ou extrapolatives utilisent l'historique de la variable dépendante dans une série chronologique. Elles sont utilisées lorsque la série contient assez d'informations sur sa structure pour prédire les prochains évènements. Les méthodes endogènes ou extrapolatives incluent des méthodes comme la moyenne mobile, le lissage exponentiel ou la méthode de Holt-Winter. Matteson et al. (2011) précisent que parmi les recherches qui se sont penchées sur les modèles de séries chronologiques, la majorité ont utilisé des modèles linéaires gaussiens. Ils ajoutent que ces méthodes sont moins précises lorsque le taux d'arrivée est peu élevé. Cela met donc en évidence qu'il est important d'étudier et de développer des techniques propres aux contextes où le taux d'arrivée est relativement bas, comme c'est le cas pour le comté étudié dans le cadre de ce projet.

1.1.2.1 Modèles autorégressifs

Il existe plusieurs types de modèles appartenant à la famille des modèles autorégressifs. Trudeau et al. (1989) ont utilisé le modèle ARIMA (*Auto Regressive Integrated Moving Average*) de Box et Jenkins (1976) pour prédire les demandes quotidiennes et horaires des SPUs de la ville de Montréal pour ensuite les utiliser pour la planification et l'exploitation de leurs ressources. Indépendamment des tendances et des saisonnalités, la moyenne mobile et le lissage exponentiel sont les méthodes les plus utilisées lorsque les prévisions sont faites à court terme. Lorsqu'il s'agit du modèle de lissage exponentiel, deux cas de figure se présentent. En plus du paramètre du lissage pour le niveau d'erreur, le modèle de Holt est utilisé lorsque l'historique de la demande révèle des tendances, tandis que le modèle de Winters est utilisé lorsque la demande présente une saisonnalité. Baker et Fitzpatrick (1986) se sont servis du modèle de lissage exponentiel de Winters pour déterminer la demande des SPUs dans quatre comtés de la Caroline du Sud. Ils ont formulé un programme à objectifs multiples en combinant les appels urgents et les appels de routine. Après avoir déterminé les paramètres optimaux du modèle de lissage exponentiel, les résultats ont été comparés aux résultats d'un modèle de régression linéaire multiple et aux résultats d'un modèle de lissage exponentiel à objectif unique. L'approche à objectifs multiples a pu donner une bonne estimation de la demande globale avec un taux d'erreur quadratique moyenne (MSE) bas. Pour traiter des données saisonnières et non saisonnières, Hyndman et Khandakar (2007) ont présenté deux algorithmes du langage R pour la prévision automatique des séries chronologiques : un algorithme de lissage exponentiel et un algorithme par étapes avec les modèles ARIMA. Les techniques d'analyse non paramétrique comme l'analyse du spectre du singulier, ou *Singular Spectre Analysis* (SSA) en anglais sont aussi utilisées pour l'analyse des séries chronologiques. En plus d'être flexibles dans leurs structures, ces techniques permettent d'avoir des résultats supérieurs dans le cas de prévisions à long terme comparativement aux méthodes conventionnelles des séries temporelles et des résultats semblables à celles-ci pour des prévisions à court terme (Vile *et al.*, 2012).

1.1.3 Les méthodes hybrides

Étant toujours à la quête de meilleurs résultats de prédiction, les chercheurs ont envisagé la combinaison de deux ou plusieurs modèles de prévision pour voir si une telle alliance peut donner un meilleur résultat que lorsque les modèles sont utilisés individuellement. À ce jour, les méthodes hybrides ont été utilisées dans plusieurs contextes, qui peuvent inspirer le développement de méthodes pour les SPUs. En combinant un modèle de séries chronologiques ARIMA à un autre modèle de régression, Bell et Hillmer (1983) ont modélisé les effets de variation du jour de négoce et des vacances de Pâques par un modèle régressif. Ils ont pu constater que cette combinaison donne un meilleur résultat lorsque le modèle ARIMA est utilisé seul. Dans leur étude, Zou et Yang (2004) montrent comment un algorithme appelé AFTER combine des ordres différents du modèle ARIMA en exécutant plusieurs réglages séquentiels pour accentuer sa performance prédictive. La combinaison appropriée de modèles linéaires et non linéaires peut donc augmenter la robustesse de la prévision. Babu et Reddy (2014) ont conçu et utilisé un modèle hybride à partir du modèle linéaire ARIMA et un modèle non linéaire RNA dans la prévision de séries chronologiques. Les tests de comparaison ont démontré que le modèle hybride permet d'obtenir de meilleures performances que lorsque les modèles sont utilisés individuellement. Plusieurs modèles de prévision ont été utilisés avec le système d'information géographique (SIG) à savoir, une moyenne mobile, un réseau neuronal artificiel, une régression sinusoïde et une régression du vecteur de support (*support vector regression*). Dans ce contexte, différents découpages de zones et de périodes ont été pris en compte. Ils ont conclu que les intervalles de temps de 3 h et des zones de 3 km² permettaient d'avoir une meilleure précision sur l'estimation de la demande dans plusieurs districts de la ville de New Taipei.

Les études antérieures ont démontré qu'indépendamment de la nature du modèle, certains d'entre eux peuvent être plus performants que d'autres selon le contexte d'étude, les critères d'entrées, comme le choix de variables, et le niveau de granularité temporelle et spatiale. D'autre part, il a été démontré qu'une prévision qui provient d'une combinaison de modèles peut donner un meilleur résultat qu'une prévision individuelle et qu'un modèle hybride peut être plus performant qu'un modèle simple. Le tableau 1.1 résume les

travaux que nous avons présentés sur les méthodes et les modèles utilisés pour la prévision de la demande.

Tableau 1.1 Synthèse des modèles de prévision de la demande dans la littérature

Modèles de prévision						
Auteurs	**Variables**		**Type de modèle**			**Modèles utilisés**
	Causales	**Extrapolatives**	**Linéaires**	**Apprentissage automatique**	**Hybrides**	
Aldrich et al., 1971	X		X			Régression des moindres carrés
Kvålseth et Deems, 1979	X	X	X			Régression linéaire
Kamenetzky, Shuman et Wolfe, 1982	X		X			Régression linéaire
Bell et Hillmer, 1983		X	X		X	ARIMA+autres modèles de régression
JR Baker et Fitzpatrick, 1986		X	X			Lissage de Wintters,POGP
Trudeau et al., 1989		X	X			ARIMA, modèle économétrique
McConnel, et al., 1998	X		X			Régression linéaire
Zou et Yang, 2004		X			X	Algorithme AFTER
Channouf et al., 2007		X	X			ARIMA, H-W
Hyndman et Khandakar, 2007		X	X			Lissage exponentiel, ARIMA
McCarthy et al., 2008	X		X			Régression linéaire
Setzler, Saydam et Park, 2009	X			X		RNA
Micheletti et al., 2010	X		X			Régression linéaire

Vile et al., 2012			X			SSA(analyse du spectre singulier)
Babu et Reddy, 2014	X	X	X	X	X	RNA, ARIMA
Chen et al., 2015	X	X	X	X	X	RNA, moyenne mobile, rég.sunsoiide, rég.de support de vecteurs

1.2 Localisation des ambulances

La localisation des ambulances constitue un élément essentiel à considérer dans la gestion du SPU. En effet, la localisation des ambulances peut avoir un impact sur la qualité du service et le temps de réponse. Beaucoup de modèles ont été utilisés pour déterminer la localisation optimale des ambulances (Tavakoli et Lightner, 2004) incluant des modèles déterministes à couverture simple et des modèles probabilistes et stochastiques (Bélanger, Ruiz et Soriano, 2012). Dans ce qui suit, nous présentons les modèles déterministes à couverture simple et multiple puisqu'ils sont en lien avec la méthode choisie dans le cadre de ce mémoire.

1.2.1 Modèles déterministes à couverture simple

L'un des premiers modèles utilisés pour la localisation des ambulances est celui de Toregas et al. (1971). Ce dernier, nommé le problème de localisation avec couverture totale (PLCT) ou *location set covering problem* (LSCP) en anglais, vise à assurer la couverture de toutes les zones à l'intérieur d'une distance ou d'un délai prédéterminé, en minimisant le nombre de véhicules utilisés. D'autres modèles ont vu le jour par la suite, comme le modèle de Church et ReVelle (1974), qui est une extension du modèle précèdent et qui traite le problème de localisation avec couverture maximale (PLCM) ou *maximal covering location problème* (MCLP) en anglais. Ce modèle vise à maximiser le nombre d'appels couverts, toujours à l'intérieur d'un temps prédéfini en utilisant un nombre donné de véhicules. Repede et Bernardo (1994) y rajoutent la variation temporelle pour contrer le changement dans la demande dans le temps. Villegas, Castañeda et Blandón (2012) combinent le problème de couverture maximale attendue (MEXCLP) avec (PLCT) pour

déterminer d'abord le nombre d'ambulances pouvant intervenir aux incidents de la route, puis, les postes d'attente des ambulances.

1.2.2 Les modèles déterministes à couverture multiple

Avec les modèles déterministes à couverture simple, les ambulances sont considérées comme disponibles pour répondre aux appels lorsque ces derniers surviennent, ce qui ne reflète pas toujours la réalité. En effet, il peut arriver qu'un appel arrive pendant que l'ambulance est occupée à servir un autre appel. C'est à partir de cet aspect que les modèles de couverture totale ont été améliorés pour prendre en compte le risque qu'un véhicule ne soit pas disponible. Daskin et Stern (1981) ont été les premiers à introduire les modèles de couverture multiple qui traitent le problème de couverture avec objectifs hiérarchiques (PCOH) ou *hierarchical objective set covering problème* (HOSC) en anglais. Ce dernier vise d'abord à minimiser le nombre de véhicules utilisés, et par la suite, à maximiser la couverture multiple des zones. Un autre modèle de couverture multiple, appelé modèle avec double standard (MDS) ou *double standard model* (DSM) en anglais, proposé par Gendreau, Laporte et Semet (1997) vise à assurer un temps d'intervention pour une proportion de la population et un autre temps pour l'ensemble de la population dans le même modèle. Ce modèle a été par la suite modifié par Doerner et al. (2005). Dans cette étude, ils ajoutent un troisième élément qui limite le nombre d'habitants affectés à un véhicule. Ce modèle a été utilisé pour localiser les ambulances en Autriche. Storbeck (1982) a abordé le problème de localisation avec couverture maximale multiple (PLMM) ou *maximal-multiple location covering problème* (MMLCP) en anglais par programmation par objectifs de manière à minimiser la population non couverte et maximiser la population couverte.

Il existe aussi d'autres types de modèles appelés modèles probabilistes, qui considèrent les sources d'incertitudes en relation avec les SPUs. Dans le cas de notre étude, les modèles probabilistes n'ont pas été utilisés puisque nous utilisons un modèle combiné qui traite d'abord le problème d'optimisation déterministe et par la suite, les ressources d'incertitude sont ajoutées dans la simulation. Le tableau 1.2 présente une synthèse sur les modèles déterministes de localisation des ambulances.

Tableau 1.2 Synthèse des modèles de localisation des ambulances.

Modèles de localisation des ambulances			
Auteurs	**Modèles**		**Problème de localisation traité et modèles utilisés**
	Modèle à couverture multiple	**Modèle à couverture simple**	
Toregas et al. (1971)	X		Problème de localisation avec couverture totale (PLCT)
Church et ReVelle (1974)	X		Problème de localisation avec couverture maximale (PLCM)
Daskin et Stern, (1981)		X	Problème de couverture avec objectifs hiérarchiques (PCOH)
Storbeck (1982)		X	Problème de localisation avec couverture maximale multiple (PLMM)
Repede et Bernardo (1994)	X		Problème de localisation avec couverture espérée maximale avec variation de vitesse (TIMEXCLP)
Gendreau, Laporte et Semet (1997)		X	Modèle avec double standard (MDS)
Doerner et al. (2005)		X	Modèle avec double standard (MDS) + extension
Villegas, Castañeda et Blandón (2012)	X		Combinaison de (PLCEM) avec (PLCT)

1.3 Simulation des services préhospitaliers d'urgence

La simulation est un outil intéressant pour les SPUs. Elle permet d'évaluer les solutions suggérées à travers des scénarios dans un contexte plus réaliste que la programmation mathématique. L'un des premiers à avoir utilisé la simulation pour améliorer le service préhospitalier d'urgence de New York, États-Unis est Savas (1969). Il a comparé les différentes options de localisation de véhicules en prenant en considération le temps de réponse et le taux d'occupation des ambulances. Swoveland et al. (1973) ont utilisé la simulation pour la localisation des ambulances de la ville de Vancouver, Canada en priorisant les appels pour estimer le temps de réponse et le taux d'utilisation des ambulances. D'autre part, Lubicz et Mielczarek (1987) ont utilisé un modèle de simulation pour analyser le nombre de véhicules à utiliser pour répondre à une demande croissante. Afin de développer un modèle de simulation d'événements discrets pour décrire le fonctionnement du centre de traitement médical de Shanghai, Chine, Zhu, McKnew et Lee (1992) ont évalué le taux d'arrivée des appels d'urgence à différentes périodes de la journée. Pour un autre mode de transport, Gunes et Szechtman (2005) ont utilisé la simulation pour une société d'ambulances aérienne opérant à partir de deux bases situées au nord de la Californie, États-Unis. L'une des stations opère seulement cinq jours par semaine, alors que l'autre opère les sept jours sur sept. Le but de la simulation est d'évaluer l'impact opérationnel de l'ajout de journées de travail pour la première station. Finalement, Pinto, Silva et Young (2015) ont proposé des méthodes pour construire des modèles de simulation génériques utilisés pour la ville brésilienne Belo Horizonte. Ces modèles visaient à déterminer le nombre d'ambulances par station puis à analyser les scénarios d'augmentation de la demande. Aboueljinane, Sahin et Jemai (2013) présentent une multitude d'études sur les modèles de simulation des SPUs. Le tableau 1.3 résume les travaux que nous avons présentés sur la simulation.

Tableau 1.3 Synthèse des modèles de simulation pour le SPU

Modèles de Simulation		
Auteurs	**Objectif de la simulation**	**Données (ville ou lieu)**
Savas (1969)	Évaluer le temps de réponse et le taux d'occupation des ambulances	New York, États-Unis
Swoveland et al. (1973)	Priorisation des appels	Vancouver, Canada
Lubicz et Mielczarek (1987)	Analyser la variation du nombre de véhicules à utiliser pour une demande croissante	Pologne
Zhu, McKnew et Lee (1992)	Évaluer le taux d'arrivée des appels d'urgence à différentes périodes de la journée	Shanghai, Chine
Gunes et Szechtman (2005)	Évaluer l'impact opérationnel de l'ajout de journées de travail	Nord de la Californie, États-Unis
Silva et Young (2015)	Analyser les scénarios d'augmentation de la demande	Belo Horizonte, Brésil

L'objectif de notre étude est d'utiliser certains de ces modèles pour, d'abord prévoir le nombre d'appels futurs pour le SPU dans l'espace et dans le temps, puis déterminer le nombre d'ambulances et leur localisation optimale pour desservir toute la population. Par la suite, les solutions obtenues seront validées à l'aide d'une simulation. Dans le chapitre suivant, nous allons voir en détail la méthodologie utilisée dans notre étude.

Chapitre 2
Méthodologie

Ce chapitre portera sur l'approche utilisée pour l'étude, ainsi que le choix des outils et des méthodes de recherche inspirés de la revue de littérature. L'étude a été menée à partir des données du SPU du comté du Marin, en Californie. Elle comporte trois phases principales qui sont présentées à la figure 2.1.

Tout d'abord, l'étude commence par la réalisation de prévisions pour la demande des SPUs à partir de différents modèles. Les résultats des prévisions sont évalués et comparés pour déterminer le modèle le plus précis. Par la suite, des prévisions par heure et par zone sont effectuées pour déterminer le nombre d'appels probables à partir d'une distribution de Poisson utilisée afin de déterminer le nombre d'appels correspondant au 95^{ième} et au 99^{ième} percentile. Ce nombre d'appels sera ensuite utilisé dans le modèle d'optimisation, plutôt que la valeur moyenne, ce qui permettra de déterminer les ressources nécessaires pour couvrir 95% et 99% des appels et ainsi assurer un bon niveau de service. En fonction du nombre d'appels par unité spatio-temporelle, le nombre d'ambulances nécessaires pour répondre aux appels probables est fixé.

La deuxième phase est consacrée à la définition des emplacements optimaux des postes d'attente à l'aide d'un modèle d'optimisation visant à assurer la couverture de toutes les zones à l'intérieur d'un délai cible. Un système d'information géographique (SIG) a été utilisé pour calculer les distances entre les zones.

Enfin, une simulation est réalisée afin d'analyser les scénarios proposés, c'est-à-dire le nombre d'ambulances choisies et leur localisation, dans un contexte plus réaliste. Néanmoins, avant de commencer, nous allons présenter les phases de collecte, de nettoyage et de transformation des données.

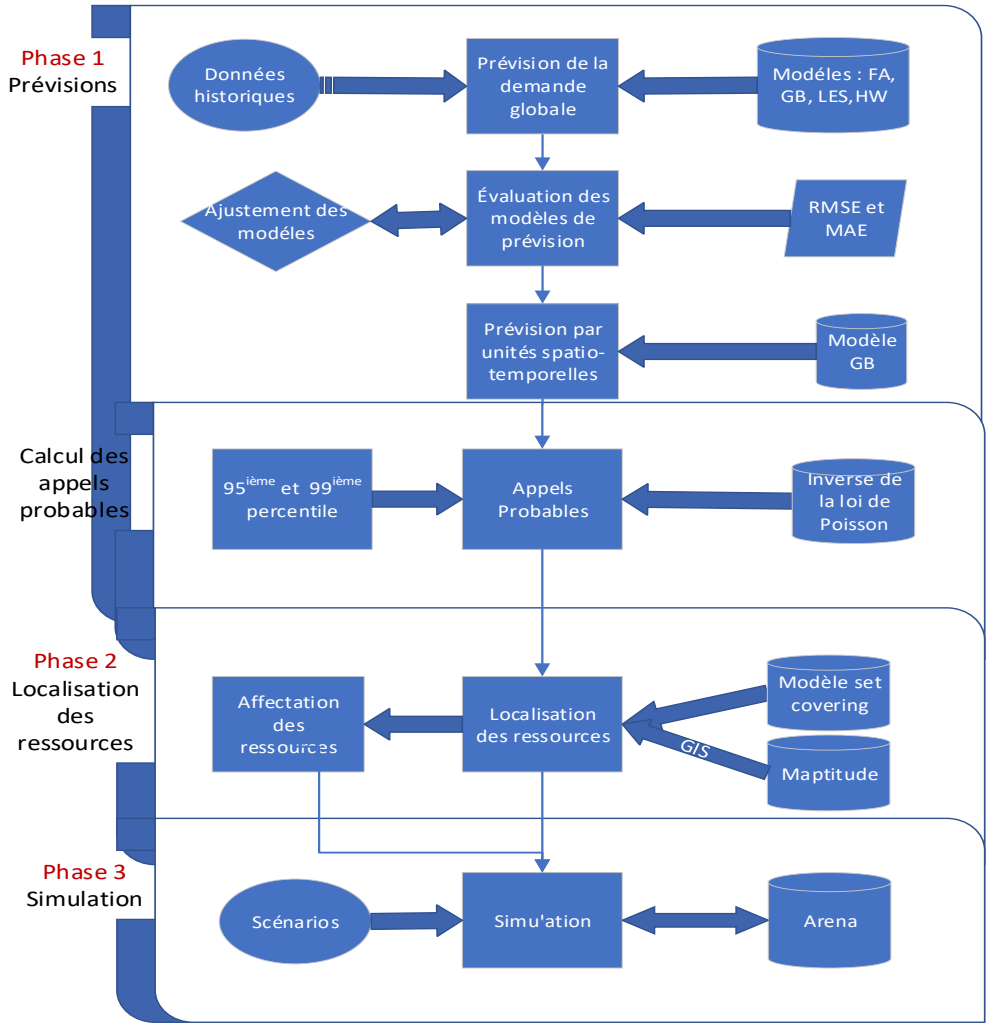

Figure 2.1 Étapes du processus de recherche

2.1 Collecte de données

Des données ont été collectées à partir du site public des données dénominalisées ouvertes du comté du Marin, Californie (voir page 17). Ces données incluent 71 983 enregistrements incluant l'ensemble des appels reçus par le SPU pour la période du 1 mars 2013 au 30 juin 2017. Chaque enregistrement décrit l'heure associée à chaque évènement important depuis la réception de l'appel jusqu'au transport du patient à l'hôpital dans le cas échéant. Il donne aussi une description de l'état du patient et des protocoles utilisés par les techniciens paramédics. Une géo localisation, un code postal ainsi qu'une identification unique sont aussi attribués à chaque incident pour un total de 27 variables. Ces variables sont présentées au tableau 2.1.

Tableau 2.1 Variables enregistrées par le service d'urgence du comté du Marin

Variables	Variables (suite)
Numéro d'incident	Détail de blessure
Temps de réception de l'appel	Endroit de la blessure
Temps d'affectation du véhicule	Date de la blessure
Temps du départ du véhicule vers la scène	Protocole utilisé par le personnel
Temp d'arrivée du véhicule à la scène	Âge du patient
Temps d'arrivée du véhicule chez le patient	Comté du patient
Temps que le véhicule ait quitté la scène	Genre du patient
Temps d'arrivée du véhicule à l'hôpital	Disposition
Adresse de l'incident	Endroit où le patient est transporté
Cité de l'incident	Géolocalisation de l'incident
Code postal de l'incident	Latitude de l'incident
Comté de l'incident	Longitude de l'incident
Premières Impressions	Identification unique
Première blessure	

2.2 Choix des outils

Pour situer les appels géographiquement, diviser les zones de demande et calculer les délais de déplacement entre elles, nous avons utilisé MAPTITUDE. Ce dernier est un logiciel qui utilise le système d'information géographique (SIG) contenant des instruments cartographiques permettant la visualisation et l'analyse de données. Le logiciel R version 3.4.3 a été utilisé pour traiter et nettoyer les données, développer les modèles de prévision, transformer et créer les variables.

Pour localiser les ressources, nous avons utilisé un modèle de couverture (*set covering model*) (Toregas *et al.*, 1971) implémenté avec EXCEL pour déterminer le nombre de ressources nécessaires pour couvrir tout le territoire. Finalement, le simulateur ARENA a servi pour représenter le processus, simuler et analyser différents scénarios développés. Ces logiciels sont implantés dans un système 64-bit avec un Processor i5-4300M CPU @ 2.60GHz 2.59 GHz et une carte mémoire de 4.00 GB.

2.3 Nettoyage des données

Nous avons effectué un nettoyage de données en éliminant les doublons, par exemple deux enregistrements pour le même appel. Nous avons éliminé les valeurs aberrantes comme les codes postaux hors périmètre à l'aide du logiciel R et d'un système d'information géographique (SIG) pour arriver à 71 665 observations. Nous avons aussi supprimé les enregistrements avec des codes postaux erronés, puis séparé les années (5 années), les mois (12 mois), les semaines (52 semaines), les jours (1-7) et les heures d'appels (0-23) pour faciliter l'agrégation et l'analyse des appels par période de temps. Nous avons transformé les semaines de telle sorte que le lundi soit le premier jour de la semaine au lieu du dimanche. Ces données nous permettront d'analyser les tendances et les saisonnalités de la demande.

2.4 Création et ajout de variables

Puisque nous faisons face à une variation de la demande dans une série chronologique, les variables (temps) utilisées ont été obtenues à partir de la fonction *tk_augment_timeseries_signature* du paquet *time kit*. Cette fonction ajoute les caractéristiques de la série temporelle et fragmente la variable date et temps en variable de différence, d'année, de mois, de jours, d'heure, de minute, y compris le jour de l'année, du mois ou de la semaine[2]. Elle génère, entre autres, des variables temps pour le matin, le soir, la fin de semaine, et ce pour un total de 28 fragments. L'avantage de la fonction *tk_augment_timeseries_signature*, est qu'elle produit automatiquement et rapidement les variables en allant chercher minutieusement les informations qui pourraient être pertinentes pour le modèle. Cette méthode permet d'économiser beaucoup de temps. Cependant, il est important d'évaluer la pertinence de chaque variable ainsi que son niveau d'importance. Nous verrons plus loin le choix de variables et la technique employée pour tester chacune d'elles afin d'évaluer leur puissance afin de sélectionner celles qui ont un plus grand pouvoir prédictif.

[2] https://www.rdocumentation.org/packages/timekit/versions/0.3.1/topics/tk_augment_timeseries

2.5 Division des données et création de zones

Afin d'effectuer les prévisions et de tester la performance des modèles, les données ont été divisées en deux parties. La partie «entrainement» contient les enregistrements de mars 2013 à décembre 2015, pour un total de 34 mois. La partie «test», servant à évaluer la performance des modèles, contient les enregistrements de 18 mois allant de janvier 2016 à juin 2017. Les codes postaux ont été utilisés pour décomposer le comté en 29 zones numérotées aléatoirement (voir Figure 2.2). Pour le besoin de notre étude, nous avons considéré 28 zones seulement puisque la 29^{ième} zone est inhabitée et comprend seulement 2 appels d'urgences pour toute la période étudiée. Ainsi, seuls les appels venant de la zone géographique avec un code postal allant de 94900 à 95000 sont considérés. Les autres appels à l'extérieur de ces zones ont été éliminés. De cette manière, seulement 2 appels ont été éliminés.

Figure 2.2 Découpage de zones

2.6 Modèles de prévisions

Après avoir choisi, créé les variables, et formé les zones, nous avons fait des prévisions sur la demande en utilisant à la fois des méthodes extrapolatives qui utilise l'historique

des appels dans le passé à partir d'une série chronologique, et des méthodes causales qui utilisent la corrélation entre les variables prédictives et l'arrivée des appels. Plus particulièrement, nous avons utilisé les forêts aléatoires (FA), le *Gradient Boosting* (GB), le lissage exponentiel simple (LES) et le lissage exponentiel *Holt-Winters* (H-W). Notre choix a été inspiré par les différents modèles de la revue de littérature. Ces modèles sont présentés plus en détail à la section 2.8. Nous avons aussi testé et appliqué différents paramètres, puis choisi ceux qui donnaient le meilleur résultat en termes de performance. Nous avons aussi passé en revue les différentes techniques de sélection des variables ainsi que leur capacité à améliorer la performance des modèles. Ces techniques sont présentées à la sous-section 2.7. Enfin, nous avons choisi le modèle en fonction du taux d'erreur le plus bas, mesuré par l'erreur absolue moyenne (*mean absolute error* - MAE), la racine de l'erreur moyenne quadratique (*root mean square error* -RMSE) et le pourcentage de l'erreur absolue moyenne MAE(%).

Dans ce qui suit, nous utiliserons la notation suivante :

y_t : Nombre d'appels au temps t

$\hat{y}_{t+1}$: Prédiction de y au temps $(t+1)$

n : Nombre de périodes

x_{it} : Valeur d'entrée de la variable i au temps t.

X : Ensemble des variables prédictives

Nous utilisons également x_i pour représenter le vecteur des valeurs d'entrée pour la variable i sur toutes les périodes de temps. Cette notation est utilisée dans les modèles sans régresseurs temporels (modèles non-chronologiques).

2.6.1 Choix de variables

Avant de construire le modèle de prévision, il est impératif de passer par l'étape de sélection des variables pour rendre le modèle plus précis et moins complexe. Pour cela, plusieurs méthodes existent pour des cas de régression, de classification, et de

regroupement des données. Ces méthodes sont utilisées pour identifier et supprimer les variables qui ne contribuent pas à la prévision ou qui la diminuent. Dans certains modèles, le choix de variables se fait manuellement, alors que dans d'autres, comme les modèles ensemblistes, la sélection de variables est automatique à partir d'algorithmes. L'objectif de ce processus est de réduire la taille des données pour permettre l'analyse et l'interprétation des résultats efficacement. La méthode de choix de variables se fait soit à travers la transformation ou bien la sélection de variables qui, à son tour, se fait en fonction de l'objectif, des ressources disponibles et du niveau d'optimisation souhaité (Jovic, Brkić et Bogunovic, 2015). Il existe trois méthodes de sélection de variables, soit la méthode Filtre (*Filter*), la méthode Symbioses (*Wrapper*) et la méthode intégrée (*Embedded*).

2.6.1.1 La méthode Filtre

La méthode Filtre sélectionne les variables en fonction des scores attribués aux variables à partir de mesures statistiques pour déterminer leur corrélation avec la variable dépendante. Le choix de variables se fait manuellement. Il est donc indépendant de tout algorithme d'apprentissage (Guerif et Bennani, 2007).

2.6.1.2 La méthode Symbiose

La méthode Symbiose consiste à évaluer les sous-ensembles de variables et détecter les interactions entre elles à partir de plusieurs tests possibles, comme le test-t ou le test-F (Yang, Chuang et Yang, 2010). Cette méthode utilise la régression par étape et le choix de variables s'effectue automatiquement à partir de trois approches : en choisissant les variables (*forward selection*), en les éliminant les variables (*backward elimination*) ou en utilisant une combinaison des deux approches précédentes (*stepwise selection*) jusqu'à ce que l'ajout ou l'élimination de variables supplémentaire n'améliore plus la performance du modèle (Derksen et Keselman, 1992). La méthode *stepwise* est plus rapide que les autres méthodes de sélection et peut gérer une grande quantité de variables, mais peut-être moins performant lorsque la multicolinéarité entre les variables est élevée (Chong et Jun, 2005).

2.6.1.3 La méthode intégrée

La méthode intégrée utilise un algorithme d'apprentissage qui permet de faire simultanément la sélection et la classification de variables. La méthode intégrée la plus connue est *LASSO* (Han, Li et Shi, 2017). Nous avons choisi différents modèles appartenant aux deux types de méthodes de prévisions (causales et extrapolative). Ces modèles ont été utilisés de pair avec les techniques de sélection de variables discutées précédemment.

2.6.2 Les forêts aléatoires (FA) ou forêts d'arbres décisionnels

Les forêts aléatoires (Breiman, 2001) sont formées de plusieurs arbres décisionnels entrainés sur des données différentes et où l'algorithme effectue un apprentissage sur ces prédicteurs et utilise des échantillons *Bootstrap* où chaque arbre est construit selon un ensemble de données différent. Le tableau 2.2 démontre les étapes générales de l'algorithme des forêts aléatoires (Lin *et al.*, 2017a).

Tableau 2.2 Étapes de l'algorithme des forêts aléatoires

Étape	
1	Ensemble de données d'apprentissage
2	Sélection du nombre d'arbres à construire (*n-trees*)
3	Pour chaque arbre, faire
4	| Générer un échantillon Bootstrap
5	| Construire un arbre de régression.
6	| Pour chaque coupure faire
7	| | Sélection de x_i variables au hasard parmi toutes les variables X
8	| | Choix des meilleures variables / point de coupure entre les variables choisies
9	| | Diviser le nœud en deux.
10	| fin
11	| Utiliser des critères d'arrêt pour déterminer quand un arbre est complet.
12	Fin

Les forêts aléatoires sont des méthodes de classification et de régression très puissantes et font partie des techniques d'apprentissage automatiques. Ils sont simples à utiliser et donnent souvent de bons résultats lorsqu'il s'agit de prévisions de grandes dimensions. La force de ce modèle est la combinaison des résultats de plusieurs prédicteurs au lieu de n'en considérer qu'un seul.

Exemple d'un arbre

La figure 2.3 présente un exemple d'un arbre de décision dans lequel le choix au premier nœud serait si oui ou non, la variable am.pm est plus petite qu'un nombre R. Si oui, une autre décision se prend au niveau de la variable heure (*hour*) pour choisir encore si oui ou non la valeur de l'heure est inférieure à U. Les valeurs A, B et C représentent un nombre ou un pourcentage et aussi le poids de la variable sélectionnée.

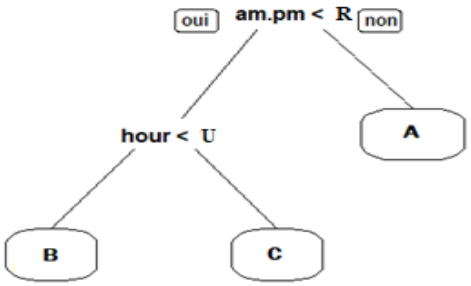

Figure 2.3 Schéma d'un arbre de décision

2.6.3 Le Gradient Boosting (GB)

L'idée principale du GB est d'ajouter à chaque itération un nouvel arbre entrainé en respectant l'erreur de l'ensemble des arbres appelés modèle d'apprentissage de base (base-learning model) (Schapire, 2003). L'opération est séquentielle et s'arrête lorsque l'algorithme juge que l'ajout d'un nouvel arbre n'améliorera plus la performance du modèle. Un arbre se forme à partir d'informations sur les arbres précédentes. Le modèle final est un modèle d'ajout par étapes de b arbres individuelles. L'algorithme de base pour les arbres de régression boostés peut être représenté comme suit :

$$f(X) = \sum_{b=1}^{B} f^{b}(X), \qquad (1)$$

où B représente le nombre d'arbres de régression boostés.

2.6.4 Le lissage exponentiel simple (LES)

Le lissage exponentiel simple est une méthode de prévision des séries chronologiques sans tendances et sans saisonnalités. Elle consiste à donner un poids plus grand aux observations les plus récentes. Le poids de chaque observation est désigné par un paramètre de lissage α. La période $t+1$ représente le moment où la prévision a été effectuée pour la valeur *y*.

$$\hat{y}_{t+1} = \alpha y_t + (1-\alpha)\hat{y}_t \quad (1)$$

Le coefficient de lissage appelé α est compris entre 0 et 1 et le coefficient $(\alpha - 1)$ s'applique à la prédiction précédente.

2.6.5 Le lissage de Holt-Winters (HW)

Comme nous l'avons déjà mentionné dans la section précédente, le lissage exponentiel simple est utilisé lorsque les séries chronologiques ne présentent pas de tendance ou de saisonnalité. Cela dit, lorsque ces dernières sont enregistrées à travers le temps, le modèle de lissage exponentiel triple de Holt-Winters (Holt, 1957; Winters, 1960) est utilisé pour capturer les tendances et la saisonnalité. Trois paramètres de lissage α, β et γ sont utilisés respectivement pour le niveau (*L*), la tendance (*T*) et saisonnalité (*S*) au moment *t*.

$$L_t = \alpha\left(y_t - S_{t-p}\right) + (1-\alpha)(L_{t-1} + T_{t-1}) \quad (1)$$

$$T_t = \beta(L_t - L_{t-1}) + (1-\beta)T_{t-1} \quad (2)$$

$$S_t = \gamma(y_t - L_{t-1} - T_{t-1}) + (1-\gamma)S_{t-p} \quad (3)$$

La fréquence de la saisonnalité est indiquée par *p,* et le nombre de périodes de prévision *k* est donné en combinant L_t, b_t et S_t au temps *t*. La longueur du cycle saisonnier pour les trois paramètres de lissage est comprise entre 0 et 1. Cette méthode peut être développée selon la nature de la saisonnalité. Si cette dernière est assez stable à travers le temps, la structure choisie est alors additive, sinon, elle est multiplicative. La saisonnalité que nous avons détecté dans nos données parait stable durant la période étudiée. Nous allons choisir alors la méthode additive.

La formule utilisée est donc la suivante :

$$\hat{y}_{t+k} = L_t + kT_t + S_{t+k-p(h+1)}, \quad (4)$$

où h est la partie entière de $(k-1)/p$, ce qui assure que les indices saisonniers utilisés pour la prévision proviennent du dernier ensemble des données observées.

2.7 Évaluation des modèles de prévision

Cette partie sera consacrée à l'estimation du niveau de précision de chaque modèle. Pour cela, certaines métriques seront utilisées pour évaluer les écarts d'erreur entre la prévision et la réalité. Chaque modèle sera évalué pour différentes techniques de choix de variables puis comparé aux autres modèles pour choisir le modèle qui minimise l'erreur de prévision.

2.7.1 Erreur absolue moyenne

L'erreur absolue moyenne ou *Mean Absolute Error* (MAE) en anglais est la moyenne des différences absolues entre la valeur réelle et une valeur estimée. Elle est indifférente face à la direction des erreurs.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_t - \hat{y}_t|$$

Le pourcentage de l'erreur absolue moyenne MAE(%) est aussi utilisé pour comparer les résultats des modèles.

2.7.2 Racine de l'erreur moyenne quadratique

La racine de l'erreur moyenne quadratique, ou *Root Mean Square Error* (RMSE) en anglais, est une mesure très populaire puisque les erreurs sont mises au carré avant le calcul de leurs moyennes. Ainsi, contrairement à la MAE, les grandes erreurs reçoivent un poids plus élevé.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_t - \hat{y}_t)^2}$$

2.8 Détermination du nombre d'appels probables

Nous avons utilisé l'inverse de la loi de Poisson pour calculer le nombre d'appels probables selon des pourcentages d'appels à couvrir. Les moyennes de prévisions ont été transformées pour calculer le nombre d'appels probable par heure pour couvrir 95% et 99% des appels (voir la figure 2.4). Autrement dit, à partir des moyennes trouvées à partir des modèles de prévision, nous avons calculé le nombre d'appels correspondant au 95e et au 99e percentile. Ceci permettra de déterminer le nombre de ressources nécessaires pour répondre aux appels probables avec un bon niveau de service.

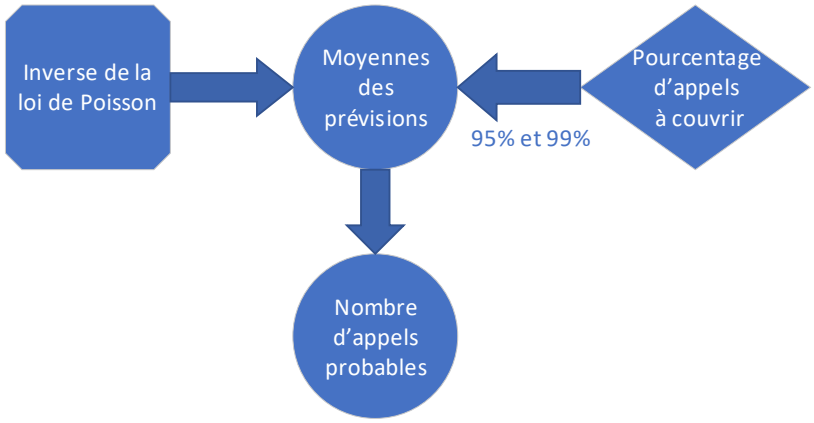

Figure 2.4 Transformation des moyennes de prévisions en nombre d'appels probables

2.8.1 Loi de Poisson

La loi de Poisson est une loi de probabilité qui s'applique aux évènements rares. Elle est construite avec un seul paramètre (θ) qui est à la fois la moyenne et la variance. Nous avons choisi d'utiliser cette distribution à cause du nombre bas des appels par unités spatio-temporelles. Pour un nombre moyen d'appels θ dans un intervalle de temps, la probabilité d'avoir précisément d appels est :

$$P(y = d) = \frac{\theta^d}{d_!}\, e^{-\theta} \quad , \tag{1}$$

où d est un nombre d'appels ($d = 0, 1, 2, \ldots$).

L'inverse de la loi de Poisson nous permet d'avoir le nombre d'appels probables correspondant au 95e et 95e percentile. Lorsque le nombre d'appels est calculé en fonction

des périodes et des zones, une répartition des ressources efficace est alors nécessaire pour couvrir le territoire.

2.9 Localisation de ressources

Le but des prévisions est de trouver le nombre d'appels probables à couvrir, ce qui va permettre ensuite de déterminer le nombre d'ambulances nécessaires pour y arriver. Il sera ensuite nécessaire de déterminer les emplacements optimaux des ambulances. Puisque nous disposons d'un nombre de zones peu élevé et aussi d'un nombre d'appels relativement faible, nous avons donc choisi d'utiliser un modèle de localisation avec couverture totale, tel que proposé dans Toregas *et al.* (1971). L'objectif de ce modèle est de minimiser le nombre de postes d'attente à utiliser parmi un ensemble d'emplacements possibles *J* de manière à couvrir toutes les zones de demande *I* à l'intérieur d'un temps de déplacement prédéfini *S.* Lorsque la zone $\hat{g}$ est couverte par l'emplacement *j à* l'intérieur de *S,* le paramètre $a_{\hat{g}j} = 1$, et si *j* est choisi comme poste d'attente, la variable $v_j = 1$.

La fonction objectif du modèle est :

$$Min \sum_{\hat{g} \epsilon J} v_j \tag{1}$$

Sous les contraintes :

$$\sum_{j \epsilon J} a_{\hat{g}j} \, v_{\hat{g}} \geq 1 \; \forall \, \hat{g} \in I, \tag{2}$$

$$v_j \in \{0, 1\} \; \forall \; j \in J. \tag{3}$$

Le nombre de ressources nécessaires pendant la journée est déterminée à partir du nombre d'appels probables. Nous avons donc calculé le nombre d'ambulances nécessaires par heure et par poste d'attente en fonction des appels prévus.

Lorsque les emplacements sont déterminés, il faut valider si ce choix reflète la réalité et permet d'atteindre les mesures de performance ciblées, d'où la nécessité de faire des simulations.

2.10 Simulations de la demande

Beaucoup d'études ont utilisé la simulation pour étudier les systèmes de gestion des SPUs (Aboueljinane, Sahin et Jemai, 2013). Le but de la simulation est de reproduire un système à partir de modèles qui présentent les mêmes comportements et caractéristiques que le système étudié pour mieux comprendre son comportement (Kelton, 2002). Dans notre cas, simuler le système d'intervention des SPUs nous permet d'apprendre sur les performances ainsi que l'utilisation des ressources. Ceci nous permet aussi de comparer les approches que nous avons proposées et de prendre des décisions concernant la gestion des ressources.

Nous avons donc décidé de développer un simulateur qui nous permettrait de reproduire facilement les opérations que nous voulons simuler avec les paramètres désirés et d'évaluer les résultats obtenus. Nous avons donc choisi d'utiliser le logiciel ARENA pour implanter le modèle de simulation. L'utilisation d'un tel logiciel ne nécessite aucune programmation en langage standard tout en permettant de simuler des systèmes très complexes. Nous avons donc créé un système d'intervention pour les SPUs à partir du nombre de ressources déterminé à partir du nombre d'appels probables par heure et par zone et des emplacements optimaux de ces ressources.

Chapitre 3
Analyse de données

Dans cette partie, nous présentons une analyse des données collectées. Dans un premier temps, nous analysons les tendances et les saisonnalités de la demande selon les mois, les heures, les jours, les semaines et les zones. Nous allons ensuite présenter une analyse des variables que nous avons créées et allons utiliser dans les modèles de prévision.

3.1 Agrégation des appels par période

À partir de l'historique des appels pour 52 mois, nous avons agrégé les appels par période dans le but de déterminer la fréquence des appels par heure telle qu'illustrée à la figure 3.1. Ces valeurs ont permis de déterminer la distribution du nombre d'appels par heure. Dans 48% du temps, aucun appel n'est reçu pour l'ensemble du territoire. Ceci montre le faible nombre d'appels par période et complexifie les prévisions, surtout lorsque l'on agrège aussi le nombre d'appels par zone.

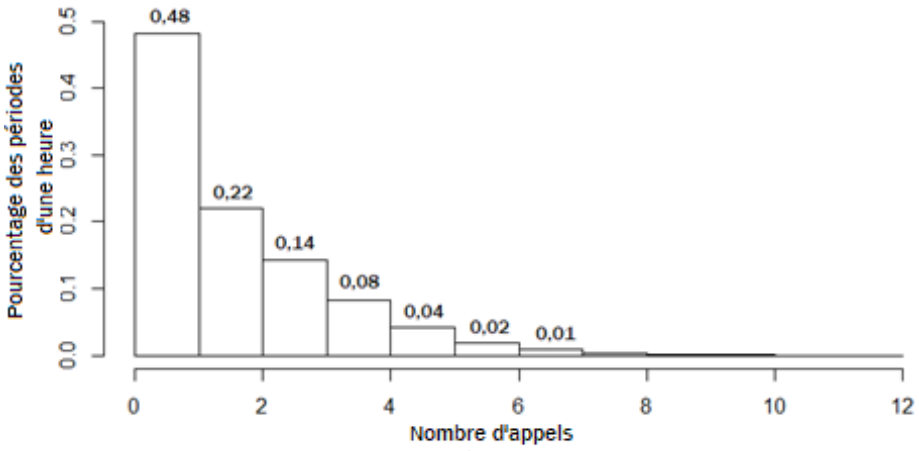

Figure 3.1 Distribution du nombre d'appels par période d'une heure sur l'ensemble du territoire

3.2 Analyse de tendance

Nous avons utilisé l'ensemble des données pour analyser la demande et avoir une vue d'ensemble de son évolution à travers les années et son comportement dans le temps. Cela permet de détecter des saisonnalités que nous pourrons utiliser pour la répartition des ressources par exemple le niveau de la demande à différentes périodes de la journée.

3.2.1 Tendance mensuelle

La figure 3.2 présente le nombre d'appels par mois pour la période analysée, soit mars 2013 à juin 2017. Les niveaux d'appels marquent une légère baisse vers la fin 2014 et une hausse du niveau d'appels entre 2015 et 2017. Cela est peut-être dû à la croissance démographique (McConnel et Wilson, 1998).

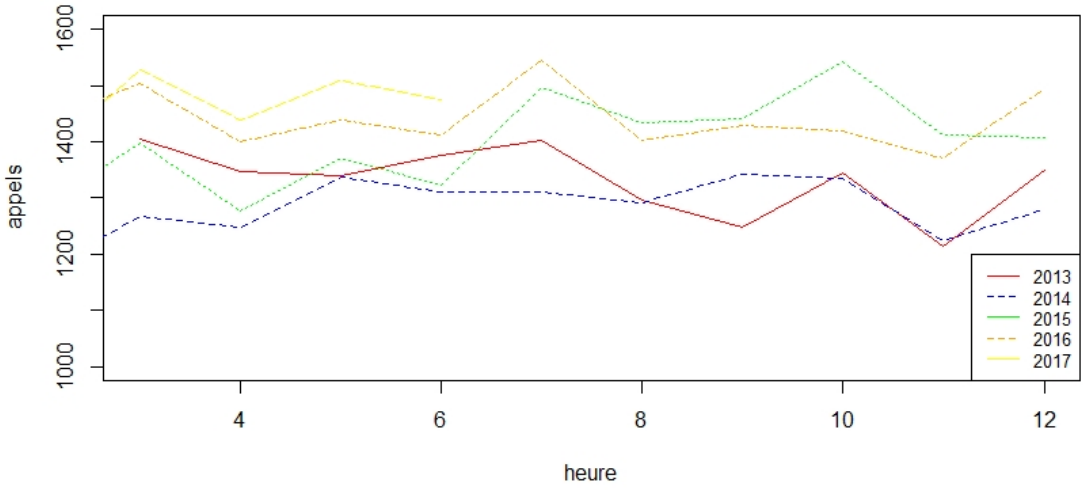

Figure 3.2 Nombre d'appels par mois

3.2.2 Tendance hebdomadaire

La figure 3.3 présente le nombre d'appels par semaine pour les années disponibles. On constate que le niveau d'appels est presque identique d'une année à une autre, à part quelques pics comme pour la semaine 41 de l'année 2015, qui peut être due à des causes climatiques (McCarthy *et al.*, 2008), à des épidémies (Kane *et al.*, 2014) ou autres.

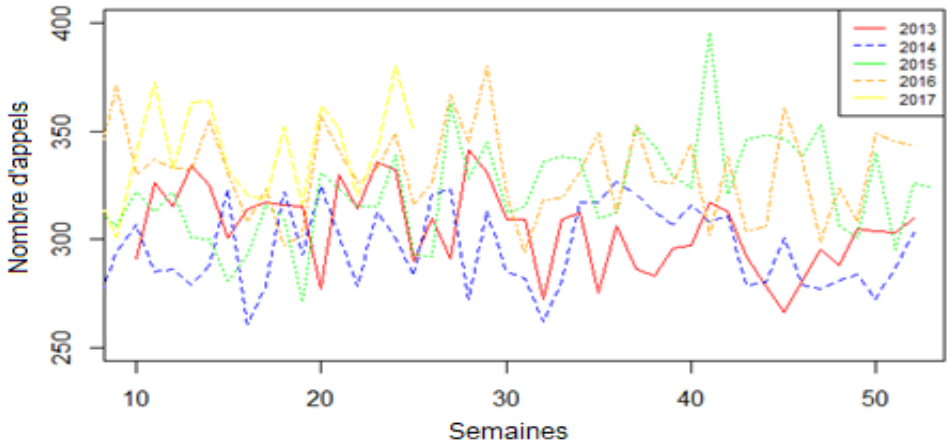

Figure 3.3 Nombre d'appels par semaine et par année

3.2.3 Tendance journalière et horaire

La représentation journalière du nombre d'appels pour les années disponibles est présentée à la figure 3.4 . Il est possible de constater que la demande augmente légèrement les vendredis et les samedis. Ceci est généralement le cas pour toutes les années.

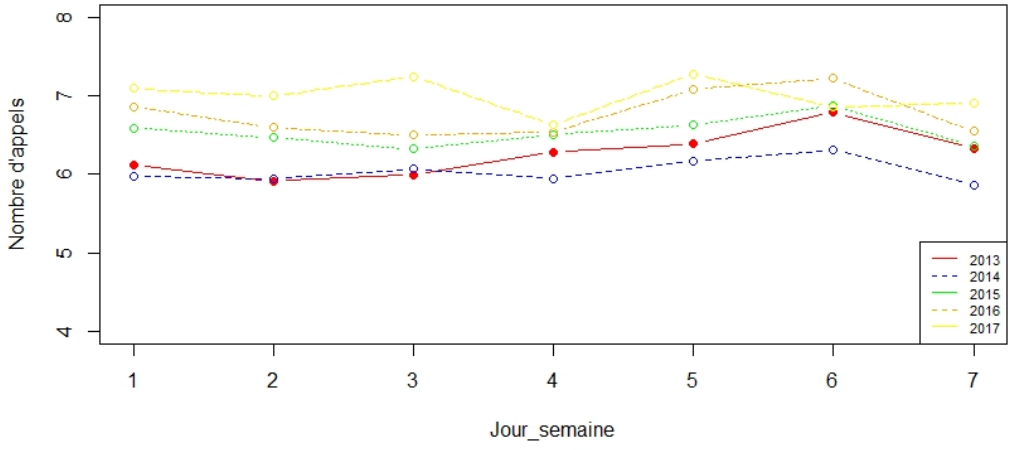

Figure 3.4 Nombre d'appels par jour de la semaine

Comme le montre la figure 3.5, les heures de la journée peuvent être divisées en 4 phases, le matin (6h-12h), après-midi (13h-18h), soir (19h-0h) et la nuit (1h-5h). Durant la

journée, la demande augmente graduellement à partir de 6h et atteint son maximum vers midi. Cependant, la demande reste plus ou moins stable jusqu'à 18h pour rechuter graduellement pendant la nuit. Cette saisonnalité peut informer les SPUs sur le changement de la demande durant la journée pour gérer les ressources efficacement.

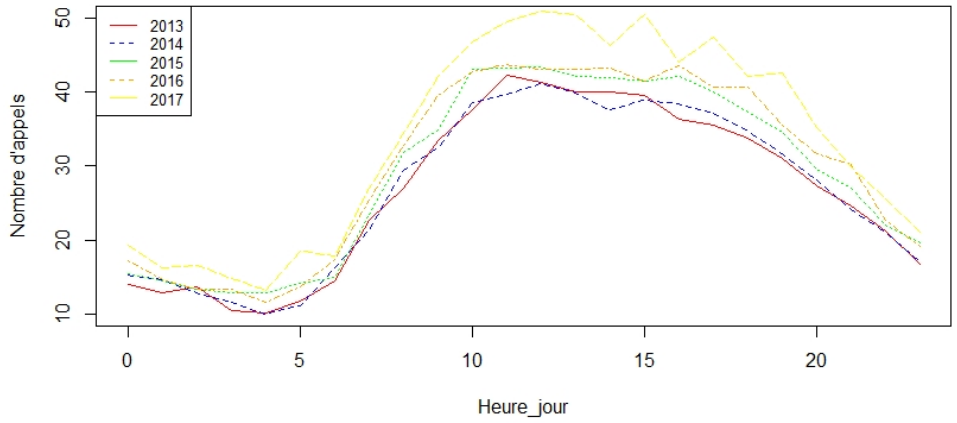

Figure 3.5 Nombre d'appels par heure de la journée

3.3 Présentation et analyse des variables

Dans le tableau 3.1, nous présentons les variables (temps) que nous avons créées pour effectuer les prévisions. Nous avons pu garder 29 fragments numériques après avoir éliminé des variables redondantes comme la variable factorielle (*month.lb*) qui réplique les mêmes valeurs de la variable numérique (*month*) représentant les mois de l'année. La description de chaque variable numérique est présentée au tableau 3.1.

Tableau 3.1 Présentation des nouvelles variables temps

Variables	Description
index.num	Index numérique (1362114-1498878)
Year	Année (2013-2017)
Half	La demi-composante de l'index (1-2)
Quarter	Composante trimestrielle de l'index (1-4)
Month.xts	Mois de l'année (0-11)
Hour	Heures de la journée (0-23)
hour12	Heures par demi-journée (0-11)
am.pm	Matin et après-midi (1-2)
wday.xts	Jours de la semaine (0-6)
Mday	Jours du mois (1-31)
Qday	Jours du quart de l'année (1-92)
Yday	Jours de l'année (1-366)
Mweek	Semaines du mois (1-5)
Week	Semaines de l'année commençant le dimanche (1-53)
week.iso	Semaines de l'année commençant le lundi (1-53)
week2	Fréquence bihebdomadaire (0-1)
week3	Fréquence trihebdomadaire (0-2)
week4	Fréquence quadrihebdomadaire (0-3)
mday7	Division des jours du mois par sept (1-5)
appel 1 à appel 10	Appels historiques décalés de 1 à 10 période (0, $+\infty$)

Les données contiennent un index numérique pour désigner chaque évènement. À part les variables désignant les années, les mois, les semaines, les jours et les heures de la journée, nous avons des variables pour des jours et des semaines commençant le dimanche ou le lundi, des fréquences différentes pour les semaines et une division des jours du mois par sept. Les variables appel1 à appel10 sont des variables de fréquences différentes. Par exemple, appel1 est décalé d'une heure, appel2 est décalé de deux heures et ainsi de suite jusqu'à appel10, qui est décalée de 10 périodes. Ces variables ont été créées pour trouver des corrélations potentielles entre elles et la variable réponse, et seront utilisées avec les modèles causaux.

3.4 Test de significativité

L'analyse de variance ou ANOVA est effectuée pour savoir s'il existe une relation entre les appels et les différentes variables de temps que nous avons générés. La tableau 3.2 représente cette relation par la somme des carrés des écarts par rapport à la moyenne calculée et les *p-value* pour un seuil de confiance de 95%. Le tableau suivant montre les résultats du test de variance pour les variables qui ont une forte corrélation avec les appels.

Tableau 3.2 Analyse de variance

	Df	Sum.Sq	Mean.Sq	F value	Pr(>F)
hour	1	573	573,4	2095,62	< 2e-16***
hour12	1	17	17,3	63,062	2.03e-15***
am.pm	1	1177	1176,7	4300,81	< 2e-16 ***
appel1	1	55	55	200,98	< 2e-16 ***
appel3	1	6	6,2	22,535	2.07e-06***

Au seuil de 95%, les *p-value* dans la colonne (Pr (>F)) confirment que le nombre d'appels dépend linéairement des variables heures de la journées (*hour*), demi-journées (*hour12*), du jour et du soir (*am.pm*), des appels historiques décalés d'une période (*appel1*) et des appels historiques décalés de trois périodes (*appel3*). Cela dit, il est possible qu'avec d'autres combinaisons et interactions entre variables, la force explicative de certaines variables augmente. Il est possible aussi que l'effet de certaines variables change en présence d'autres variables à cause du niveau de corrélation qui existe entre elles. C'est pour cette raison qu'il serait essentiel de tester la corrélation entre les variables car, il se peut qu'elle influence le résultat de notre modèle de prévision.

3.5 Analyse de colinéarité

Le corrélogramme présenté à la figure 3.6 permet de visualiser la corrélation entre les variables à partir du coefficient de corrélation *r* compris entre (1 et -1). Nous pouvons donc évaluer la dépendance entre plusieurs variables à partir des couleurs et formes des cercles de la matrice. Nous nous intéresserons plus à voir la relation entre les variables temps que nous avons obtenu et les variables (appel) décalées que nous avons créées. La couleur bleue signifie que la corrélation est positive, la couleur rouge indique que la

corrélation est négative, alors que la couleur blanche signifie que la corrélation est nulle. Les coefficients de corrélation sont proportionnels aux tailles des cercles et à l'intensité de leur couleur. Plus le cercle est grand et de couleur foncée, plus la corrélation est forte entre les variables. Il n'existe aucune corrélation entre les variables décalées (appel1 à appel10) et les autres variables de temps. Lorsqu'il y a une corrélation forte entre deux variables, leur pouvoir prédictif diminue. Dans ce cas, il faut choisir celle qui a une plus grande corrélation avec la variable dépendante. Il est donc nécessaire de connaitre cette corrélation avant d'effectuer les prévisions.

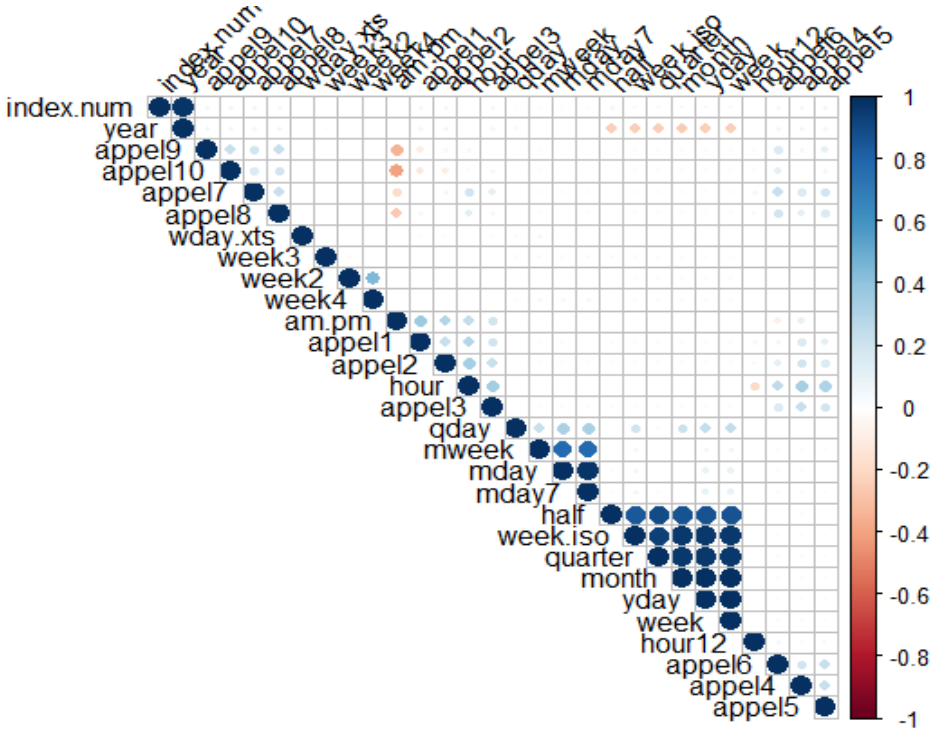

Figure 3.6 Matrice de corrélation des variables

3.6 Distribution des appels par zone

La figure 3.7 présente le nombre d'appels annuel par zone. Tel qu'illustré, la zone 12 reçoit le plus grand nombre d'appels, suivie par la zone 16, 14 et 7 pour les différentes années de l'étude. La nature géographique du comté du Marin et la concentration disparate de sa population font en sorte qu'il y a un grand déséquilibre en termes de demande entre les différentes zones.

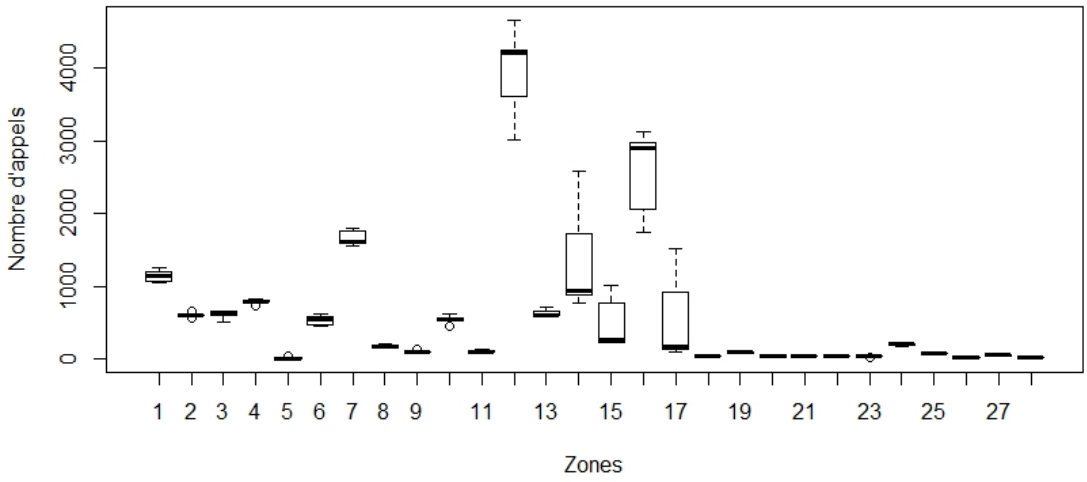

Figure 3.7 Nombre d'appels annuel par zone

Comme la figure 3.8 le démontre, la concentration des appels est plus élevée dans la zone du centre-ville située au sud-est du comté du Marin, un peu plus dispersée à l'ouest et presque nulle du côté nord-est. La zone en jaune sur la figure représente la concentration des appels. La concentration des appels sur ce côté de la carte est due au grand nombre de terres agricoles que le compté comporte sur le côté ouest.

Figure 3.8 Concentration de la demande en jaune sur l'ensemble du territoire

3.7 Conclusion

En général, la détermination géographique des niveaux de concentration de demande aide les gestionnaires à déterminer les emplacements optimaux et à répartir efficacement les ressources pour desservir la population dans les plus brefs délais. L'information que peut contenir chaque parcelle spatio-temporelle diffère selon le niveau d'agrégation. Le niveau de précision des prévisions dépendent alors de l'information que peut contenir dans ces parcelles (Setzler, Saydam et Park, 2009). À partir de l'analyse de données, nous constatons que la demande fluctue pendant la journée créant ainsi une saisonnalité identique pour tous les jours de la semaine. La demande est très élevée pour certaines zones et presque nulle pour d'autres.

Dans le chapitre suivant, nous allons d'abord voir comment ajuster les paramètres d'entrée des modèles de prévisions, puis comparer les erreurs de prévisions pour choisir les prévisions du modèle le plus performant.

Chapitre 4
Prévision de la demande

Dans ce chapitre, nous allons utiliser les différents modèles de prévision que nous avons présentés dans le chapitre 2 pour prévoir le nombre d'appels par zone et par heure pour le SPU du comté du Marin, en Californie. Pour cela, nous avons utilisé les variables présentées au tableau 3.1. Pour chaque modèle, nous avons appliqué différents jeux de paramètres, puis choisis ceux qui donnaient le meilleur résultat en termes de performance. Nous avons appliqué les techniques de sélection de variables aux modèles pour enfin les évaluer et les comparer sur la base du MAE, du RMSE et du MAE(%) qui est le pourcentage de l'erreur absolue moyenne par rapport à la moyenne des appels reçus. Ainsi, nous considérons les modèles basés sur les arbres de régression (qui traitent la non-linéarité), ainsi que les modèles de séries chronologiques comme modèles prédictifs. À partir des moyennes de prévisions, nous avons utilisé la distribution de Poisson pour calculer le nombre d'appels probables au niveau de couverture de 95% et 99%.

4.1 Ajustement des modèles

Pour évaluer la performance des modèles, il est nécessaire d'effectuer les prévisions sur l'ensemble des données *entraînement* puis de comparer les résultats aux données *test* à l'aide des mesures MAE, RMSE et MAE(%).

4.1.1 Le modèle des forêts aléatoires (FA)

La méthode de recherche *Grid* de type *full cartesian* du paquet R permet de construire une grille avec des combinaisons d'hyperparamètres et chercher parmi différents modèles de prévision celui qui est le plus performant. Les paramètres finaux retenus sont : le nombre d'arbres à chaque itération (*n-tree*)= 300, le nombre d'échantillons sur lesquels nous faisons un apprentissage (*mtry*) = 20, et le taux d'échantillonnage (*sample_rate*) est de 0.55. Les 20 échantillons sont formés pour faire la validation croisée et choisir automatiquement les variables. Ils aident aussi à améliorer l'apprentissage du modèle. Le modèle retenu présente un MAE = 1.13536, un RMSE = 1.49613 et un MAE(%) = 56,86%.

4.1.2 Le modèle Gradient Boosting GB

La méthode de recherche (*grid*) utilisée avec les forêts aléatoires est aussi utilisée avec le GB. Les paramètres sélectionnés par l'algorithme de recherche sont : colonne du taux d'échantillonnage par arbre (*col_sample_rate*) = 0.8, le taux pour baisser la fonction de perte du GB (*learn_rate*) = 0.05, la profondeur de chaque arbre (*max_depth*) = 3 et le minimum d'observations autorisées (*min_rows*) est de 10 observations. Parmi les différents modèles formés, un modèle est choisi avec 10 variables prédictives. La figure 4.1 démontre le choix et le poids de variables que GB a sélectionnés pour construire son modèle le plus performant. Les résultats ont pu fournir un MAE égal à 1.13371, un RMSE= 1.48681 et un MAE(%) = 56,78%.

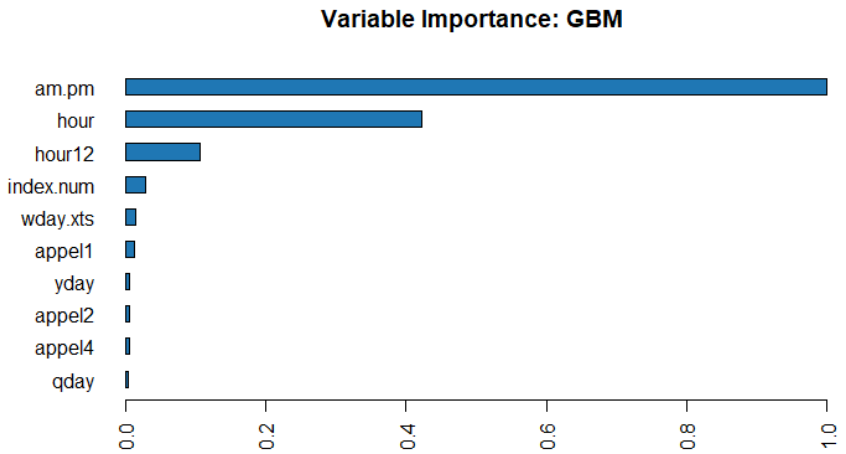

Figure 4.1 Variable d'importances (GB)

4.1.3 Le modèle lissage exponentiel simple (LES)

D'abord, nous avons transformé les appels en format séries temporelles, puis nous avons cherché le α optimal entre (0.01 et 0.99) par des séquences de 0.01. La figure 4.2 représente le RMSE en fonction du paramètre α. La valeur de α RMSE est de 0.76.

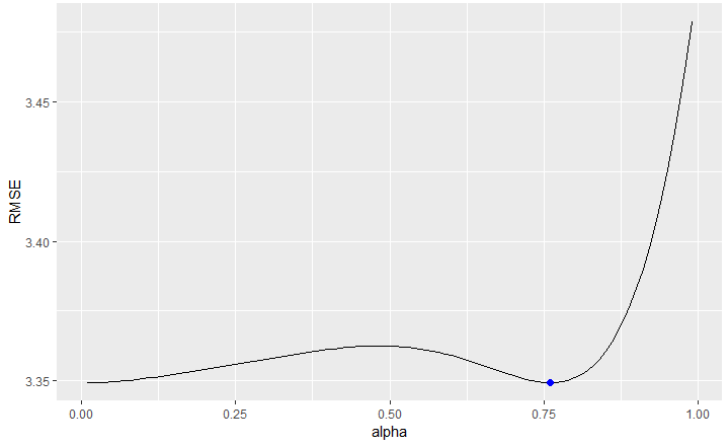

Figure 4.2 Choix de paramètre α optimal (LES)

En effectuant l'évaluation de la performance du modèle, nous constatons à partir du résultat dans le tableau 4.1 que les taux d'erreurs pour l'ensemble de test sont plus bas que celui de l'apprentissage, ce qui signifie que notre modèle n'est pas surévalué.

Tableau 4.1 Résultats de prévisions avec le lissage exponentiel simple

Données	RMSE	MAE	MAE(%)
Apprentissage	2.965952	2.273243	-
Test	2.109726	1.563341	78,29%

4.1.4 Le modèle Holt-Winters (HW)

Pour appliquer cette méthode, nous utilisons la fonction (*ets*) du paquet R pour calculer l'erreur, la tendance et la saisonnalité. Pour cela, nous avons cherché le paramètre α qui nous donne le plus bas taux d'erreur. Ce paramètre est égal à 8e-04.

En regardant les résidus dans la figure 4.3 nous constatons que les erreurs progressent au même niveau dans le temps ce qui confirme que la méthode additive est la plus appropriée pour nos données. Cependant, si la progression n'était pas du même niveau dans le temps, cela signifie qu'il existe une saisonnalité et donc la méthode appropriée serait multiplicative.

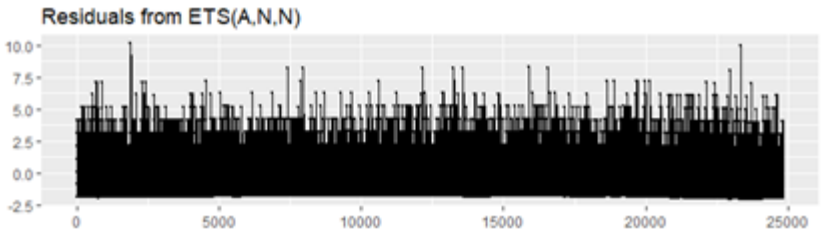

Figure 4.3 Distribution des résidus (HW)

4.1.4.1 Autocorrélation et choix du modèle

Les graphiques d'autocorrélation ACF constituent un outil visuel utile pour déterminer si une série est stationnaire. Lorsque la série temporelle est stationnaire, le coefficient d'autocorrélation dans PACF tend vers zéro après deux ou trois périodes retardées (Gamberini *et al.*, 2010). Si la série est en corrélation avec les périodes décalées, il existe généralement des composantes de tendance ou de saisonnalité. Dans le cas de notre étude, nous pouvons conclure que ces composantes peuvent être présentes (voir figure 4.4).

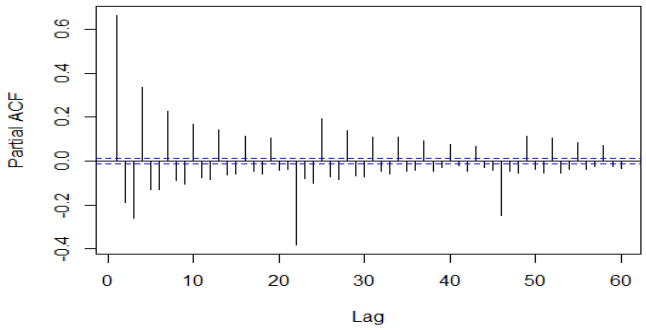

Figure 4.4 Diagrammes d'autocorrélation partielle

4.1.4.2 Ajustement du modèle

La fonction ARIMA () du paquet R permet de spécifier l'ordre du modèle HW en générant automatiquement un ensemble de valeurs optimales (*f, d, q*) pour l'erreur, la tendance et la saisonnalité qui optimisent les critères d'ajustement du modèle. Le modèle choisi est donc ARIMA(3,1,1). Les résultats ont permis d'avoir un MAE = 1.15560, RMSE = 1.51619 et un MAE(%) = 57,87%.

4.2 Comparaison des erreurs de prévision

Le tableau 4.2 contient les MAE, RMSE ainsi que les MAE(%) pour chaque modèle étudié. Ces erreurs reflètent l'écart de la moyenne de prévision par rapport à la moyenne réelle des appels pour le comté du Marin pour la période test de 18 mois (de janvier 2016 à juin 2017). Le choix de variables se fait automatiquement avec la méthode symbiose à partir des algorithmes des Forêts aléatoires et *Gradient Boosting*.

Tableau 4.2 Comparaison des modèles de prévision

Méthodes	**Méthode de Sélection de variables**	**MAE**	**RMSE**	**MAE(%)**
FA	Symbiose	1.13536	1.49613	56,86%
GB	Symbiose	1.13371	1.48681	56,78%
LES	-	1.56334	2.10972	78,29%
HW	-	1.15560	1.51619	57,87%

En examinant les valeurs des RMSE et des MAE, le modèle GB semble performer mieux que les autres modèles suivis de RF. Cela confirme que le choix automatique des variables par les algorithmes donne un meilleur résultat. Les différences entre les valeurs de RMSE et MAE ne sont pas très grandes. Cela confirme que les erreurs individuelles pour les différentes heures de prévision ont presque le même poids, car si elles étaient trop éloignées, les valeurs de RMSE prendraient un poids plus grand. Maintenant que les moyennes de prévision sont calculées, nous allons les utiliser pour déterminer le nombre d'ambulances nécessaire pour couvrir toutes les zones. Nous avons converti les moyennes de prévisions à l'aide de l'inverse de la loi de Poisson en nombre d'appels probables pour deux niveaux de service, soit couvrir 95% et 99% des appels.

4.3 Calcul du nombre d'appels probables

À partir des résultats de prévisions par zone et par heure, nous avons calculé le nombre d'appels probable pour 672 parcelles spatio-temporelles (24h x 28 zones) à l'aide de la distribution de Poisson. Les valeurs trouvées sont ensuite utilisées pour déterminer le

nombre d'ambulances nécessaires par zone et par période de la journée. Nous avons donc choisi d'utiliser les moyennes de prévisions de deux types de modèles : méthode d'ensembles et méthode autorégressive. Nous avons choisi les résultats du lissage de *HW* et le *Gradient Boosting* pour comparer le nombre d'appels probable calculé pour chaque modèle avec la distribution de Poisson.

4.3.1 Calcul du nombre d'appels probables des prévisions

Pour calculer le nombre d'appels probables par zone et par période d'une heure et en utilisant la distribution de Poisson, nous avons déterminé deux niveaux de service pour couvrir 95% et 99% d'appels. D'abord, nous avons calculé la moyenne de chaque parcelle, puis utilisé l'inverse de la loi de Poisson pour trouver les k appels probables. Les tableaux 4.2 et 4.3 montrent les valeurs d'appels probables correspondant au 95^e et le 99^e percentile. Les tableaux 4.3 et 4.4 montrent les valeurs des appels probables pour chaque unité spatio-temporelle à 95% et 99% respectivement.

Tableau 4.3 Nombre appels probables au niveau 95%

Nombre d'appels prédits par zone et par heure de la journée avec une couverture de 95% (Distribution de Poisson)																													
	z1	z2	z3	z4	z5	z6	z7	z8	z9	z10	z11	z12	z13	z14	z15	z16	z17	z18	z19	z20	z21	z22	z23	z24	z25	z26	z27	z28	TTL
h0	1	0	0	1	0	0	1	0	0	0	0	1	0	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	8
h1	0	0	0	0	0	0	1	0	0	0	0	1	0	1	0	1	1	0	0	0	0	0	0	0	0	0	0	0	5
h2	1	0	0	0	0	0	1	0	0	0	0	1	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	5
h3	1	0	0	0	0	0	1	0	0	0	0	1	0	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	7
h4	0	0	0	0	0	0	1	0	0	0	0	1	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	4
h5	0	0	0	1	0	0	1	0	0	0	0	1	0	1	0	1	1	0	0	0	0	0	0	0	0	0	0	0	6
h6	1	0	0	0	0	0	1	0	0	0	0	1	0	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	7
h7	1	1	1	1	0	1	1	0	0	0	0	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	12
h8	1	1	1	1	0	1	1	0	0	1	0	2	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	14
h9	1	1	1	1	0	1	1	0	0	1	0	2	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	14
h10	1	1	1	1	0	1	1	0	0	1	0	2	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	14
h11	1	1	1	1	0	1	1	0	0	1	0	2	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	14
h12	1	1	1	1	0	1	1	0	0	1	0	2	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	14
h13	1	1	1	1	0	1	1	0	0	1	0	2	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	14
h14	1	1	1	1	0	1	1	0	0	1	0	2	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	14
h15	1	1	1	1	0	1	1	0	0	1	0	2	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	14
h16	1	1	1	1	0	1	1	0	0	1	0	2	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	14
h17	1	1	1	1	0	1	1	0	0	1	0	2	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	14
h18	1	1	1	1	0	1	1	0	0	1	0	2	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	14
h19	1	1	1	1	0	1	1	0	0	1	0	2	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	14
h20	1	1	1	1	0	1	1	0	0	1	0	2	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	14
h21	1	1	1	1	0	1	1	0	0	1	0	2	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	14
h22	1	1	1	1	0	0	1	0	0	1	0	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	12
h23	1	1	0	1	0	0	1	0	0	0	0	1	0	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	9

Tableau 4.4 Nombre appels probables au niveau 99%

Nombre d'appels prédits par zone et par heure de la journée avec une couverture de 95% (Distribution de Poisson)																												
	z1	z2	z3	z4	z5	z6	z7	z8	z9	z10	z11	z12	z13	z14	z15	z16	z17	z18	z19	z20	z21	z22	z23	z24	z25	z26	z27	z28	TTL
h0	1	1	1	1	0	1	1	0	0	1	1	2	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	15
h1	1	1	1	1	0	1	1	0	0	1	1	2	1	1	1	1	1	0	0	0	0	0	0	1	1	0	1	0	18
h2	1	1	1	1	0	1	1	0	0	1	0	2	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	14
h3	1	1	1	1	0	1	1	0	0	1	0	2	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	14
h4	1	1	1	1	0	1	1	0	0	1	0	2	1	1	1	1	1	0	0	0	0	0	0	1	0	0	0	0	15
h5	1	1	1	1	0	1	1	0	0	1	0	2	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	14
h6	1	1	1	1	0	1	1	1	0	1	1	2	1	1	1	1	1	0	1	0	0	0	0	1	0	0	1	0	19
h7	1	1	1	1	0	1	2	1	0	1	0	2	1	2	1	2	1	0	1	0	0	0	1	1	1	0	0	0	22
h8	1	1	1	1	0	1	2	1	1	1	1	3	1	2	1	2	1	0	1	0	0	0	0	1	0	0	1	0	24
h9	2	1	1	1	0	1	2	1	1	1	1	3	1	2	1	2	2	1	0	0	0	0	0	1	1	0	1	0	27
h10	2	1	1	1	0	1	2	1	1	1	1	3	1	2	1	2	2	0	1	0	0	1	0	1	0	1	0	0	27
h11	2	1	1	2	0	1	2	1	1	1	1	3	1	2	1	2	2	0	1	1	0	0	0	1	1	1	0	0	29
h12	2	1	1	2	0	1	2	1	1	1	1	3	1	2	2	2	2	1	1	0	0	1	0	1	1	0	1	0	31
h13	2	1	1	1	0	1	2	1	1	1	1	3	1	2	1	2	2	0	1	0	0	0	0	1	1	0	0	0	26
h14	2	1	1	1	0	1	2	1	1	1	1	3	1	2	1	2	2	0	1	0	1	0	0	1	1	0	1	0	28
h15	2	1	1	1	0	1	2	1	1	1	1	3	1	2	1	2	2	1	1	0	1	0	1	1	1	0	1	0	30
h16	2	1	1	1	0	1	2	1	1	1	1	3	1	2	1	2	2	0	1	0	1	0	0	1	1	0	0	1	28
h17	2	1	1	1	0	1	2	1	1	1	1	3	1	2	1	2	2	0	1	0	0	0	0	1	1	0	0	0	26
h18	2	1	1	1	0	1	2	1	1	1	0	3	1	2	2	2	2	1	1	0	0	1	0	1	1	0	0	0	28
h19	2	1	1	1	0	1	2	1	1	1	1	3	1	2	1	2	2	0	1	0	0	0	0	1	0	0	0	0	25
h20	2	1	1	1	0	1	2	1	0	1	1	2	1	2	1	2	1	0	0	0	0	0	0	1	1	0	0	0	22
h21	1	1	1	1	0	1	2	1	0	1	1	2	1	2	1	2	1	0	0	0	1	0	0	1	1	0	0	0	22
h22	1	1	1	1	0	1	2	0	1	1	1	2	1	2	1	2	1	0	1	0	0	1	0	0	0	0	0	0	21
h23	1	1	1	0	0	1	1	0	0	1	0	2	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	13

4.3.2 Comparaison du nombre d'appels probables des différents modèles

Nous avons produit et comparé les nombres d'appels probables prédits par les deux modèles de prévision selon les deux niveaux de service. Les figures 4.5 et 4.6 montrent le nombre d'appels probables par heure de la journée, et ce, pour les deux modèles.

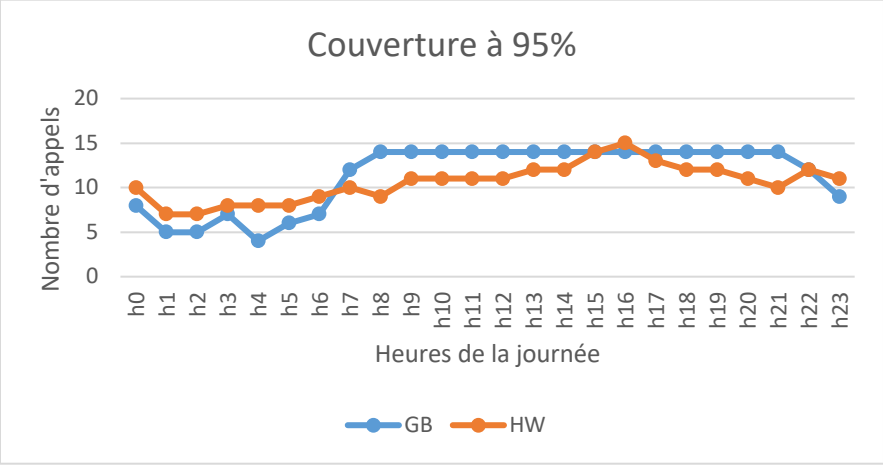

Figure 4.5 Nombre d'appels probables par heure au niveau 95%

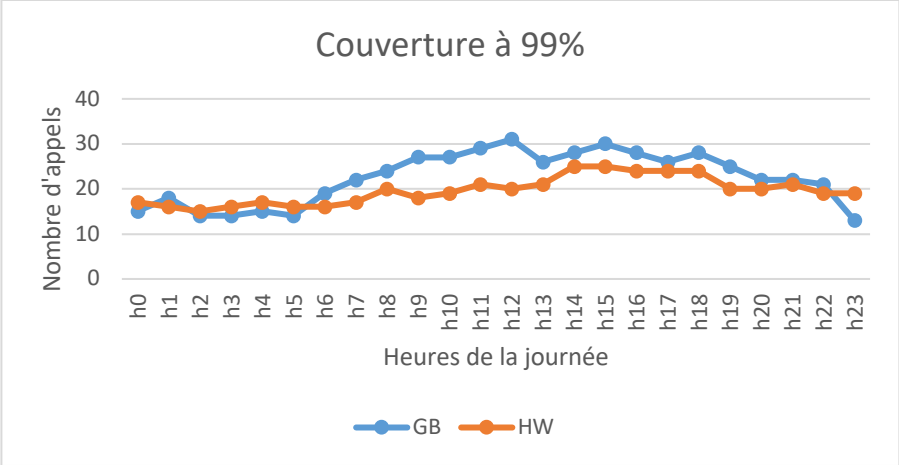

Figure 4.6 Nombre d'appels probables par heure au niveau 99%

Les courbes se rapprochent dans certaines périodes et s'éloignent légèrement dans d'autres. Cependant, La courbe du modèle GB estime plus d'appels pour la période où il y a un plus grand nombre d'appels (7h – 21h) pour les deux niveaux 95% et 99%. Cela est bien, car vaut mieux, dans le contexte d'un SPU, en avoir plus d'appels à considérer pour le calcul du nombre d'ambulances que d'en avoir moins et se trouver en face d'une pénurie. La courbe du modèle HW, quant à elle, ne suit pas les fluctuations de la courbe GB.

4.4 Conclusion

La comparaison des erreurs de prévisions des différents modèles a permis de choisir le modèle GB pour sa performance à prévoir les appels futurs avec le plus bas taux d'erreur. Les moyennes de prévisions par unité spatio-temporelle ont servi pour calculer le nombre d'appels probables à couvrir à partir de la distribution de Poisson. Puisque le niveau de service est très important pour les SPUs, nous avons calculé le nombre d'appels au niveau 95% et 99%. Cela va nous permettre de déterminer le nombre de ressources nécessaires pour couvrir tous les appels. Le nombre total d'appels par heure pour toutes les zones donne des valeurs différentes selon le modèle de prévision. L'analyse graphique des deux modèles a permis de visualiser l'éloignement et le rapprochement des courbes en sous-estimant ou bien sur estimant le nombre d'appels par période. Le modèle HW performe moins que le modèle GB par rapport à l'estimation du nombre d'appels probables.

Maintenant que nous avons transformé les moyennes de prévision du modèle GB en nombre d'appels, nous allons chercher l'emplacement optimal pour les ambulances et le nombre d'ambulances à localiser à chaque emplacement de façon à assurer à la fois un bon service tout en minimisant le nombre de ressources utilisées.

Chapitre 5
Répartition des ressources

Dans ce chapitre, nous allons déterminer les emplacements optimaux des ambulances pour couvrir toutes les zones pendant la journée ainsi que le nombre d'ambulances à affecter à chaque emplacement. Le but principal est de réduire au minimum le nombre de postes d'attente à utiliser et atteindre les zones dans un délai maximal prédéterminé pour répondre aux appels efficacement.

5.1 Localisation des ambulances

La densité de la population du comté du Marin en 2017 est de 193 habitant/km²[3]. Ce chiffre est bas comparativement à d'autres endroits comme la ville de Montréal qui compte 4 662 habitant/km². En effet, la majorité des terres sont agricoles avec des petites agglomérations de population éparpillées sur le côté ouest du comté. Trois hôpitaux peuvent accueillir les patients, et les services médicaux d'urgence sont fournis par six prestataires de services d'incendie et trois sociétés privées. Dans le cadre de ce mémoire, nous considérons une gestion centralisée de ces prestataires. La répartition des ambulances est une tâche délicate pour les gestionnaires d'un SPU, car certaines ambulances pourraient être immobilisées à des endroits ou périodes où le niveau d'appel est très bas. Pour couvrir toutes les 28 zones, nous avons choisi d'utiliser le modèle de couverture totale (*set covering*) de Toregas *et al.* (1971). Ce modèle vise à minimiser le nombre d'ambulances afin de couvrir tout le territoire selon un temps de déplacement prédéterminé (Reuter-Oppermann, van den Berg et Vile, 2017). Dans ce modèle, nous présumons qu'une seule ambulance est localisée à chaque emplacement. Dans notre cas, nous fixerons, dans un deuxième temps, le nombre d'ambulances à localiser à chacun de ces emplacements. Le temps cible (temps de réponse moyen) est souvent fixé par les opérateurs des services ambulanciers, ou bien par le gouvernement. Par exemple, aux Pays-Bas, ce temps ne doit pas dépasser 15 minutes dans 95% des cas les plus urgents

[3] https://www.opendatanetwork.com/entity/0500000US06041/Marin_County_CA/geographic.population.density?year=2017

(van Barneveld, Bhulai et van der Mei, 2016). La loi de 1973 aux États-Unis stipule que ce temps doit être dans un délai de 10 minutes dans les zones urbaines et 30 minutes dans les zones rurales (Iannoni, Morabito et Saydam, 2009). Lorsque nous avons appelé le centre de traitement des appels du comté du Marin, en Californie, ils nous ont mentionné qu'ils visaient un temps de réponse moyen de 10 minutes, mais qu'il était difficile d'atteindre ce chiffre vu la nature géographique de la région considérée comme semi-urbaine et l'éloignement de certaines zones. Pour utiliser le modèle de couverture, nous avons créé deux limites au niveau du temps de réponse : un délai de 10 minutes (T1) et un délai de 15 minutes (T15). Comme la nature et la localisation des appels diffèrent, la durée totale de l'appel qui est le temps entre la réception de l'appel et la libération de l'ambulance diffère aussi. Afin d'estimer la durée totale et le temps de réponse, il a été nécessaire de calculer le temps de déplacement entre les zones à l'aide du gestionnaire d'information géographique de Maptitude. La figure 5.1 représente un exemple de calcul de distance entre deux zones selon une vitesse moyenne-rapide.

Figure 5.1 Exemple de calcul de temps de conduite entre deux zones

Cela dit, puisque, sur l'exemple présenté à la figure 5.1, le temps de déplacement entre le poste d'attente 3 et la zone 16 est de 13 minutes (voir annexe 1), cette dernière serait couverte dans les délais de 15 minutes, mais pas dans le cas où le délai de 10 minutes est pris en compte.

5.2 Détermination des postes d'attente

Dans le but d'économiser les coûts d'inexploitation des ressources tout en respectant un temps de réponse convenable, nous avons donc résolu le modèle d'optimisation (*set covering*) (Toregas *et al.*, 1971) présenté à la section 5.1. Le modèle a été implémenté avec Excel. La détermination des emplacements des ambulances aide d'abord à couvrir le plus grand nombre d'incidents (Rojas-Trejos, González-Velasco et López-Ramírez, 2017) et aide les gestionnaires à mieux gérer la répartition des ambulances. Dans les deux cas, nous avons calculé le nombre minimal d'emplacements nécessaire afin de couvrir toutes les zones de demande à l'intérieur du délai fixé. Les figures 5.2 et 5.3 montrent les résultats obtenus et illustrent pour chaque poste d'attente choisi les zones couvertes par ce même poste en considérant respectivement T10 et T15. Nous avons observé que la solution optimale avec T10 comporte sept postes d'attente tandis que la solution avec T15 comporte six postes d'attente. Puisqu'un des postes d'attente ne sert qu'une seule zone (zone9) dans le cas où le délai a été fixé à 10 minutes, nous avons attribué cette zone au poste d'attente 19 en faisant une exception puisque le temps de déplacement entre ces deux zones est de 17 minutes au lieu qu'il soit en bas ou égal à 10 minutes. Le nombre de postes d'attente minimal devient alors six au lieu de sept postes d'attentes. Nous arrivons donc à avoir le même nombre de postes d'attente pour les deux délais cibles avec des localisations différentes.

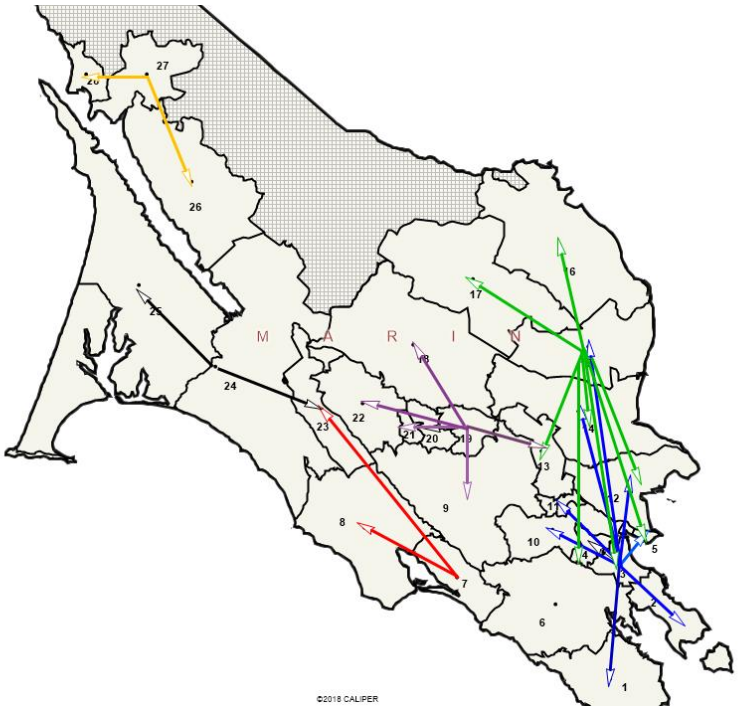

Figure 5.2 Détermination des postes d'attentes selon T10

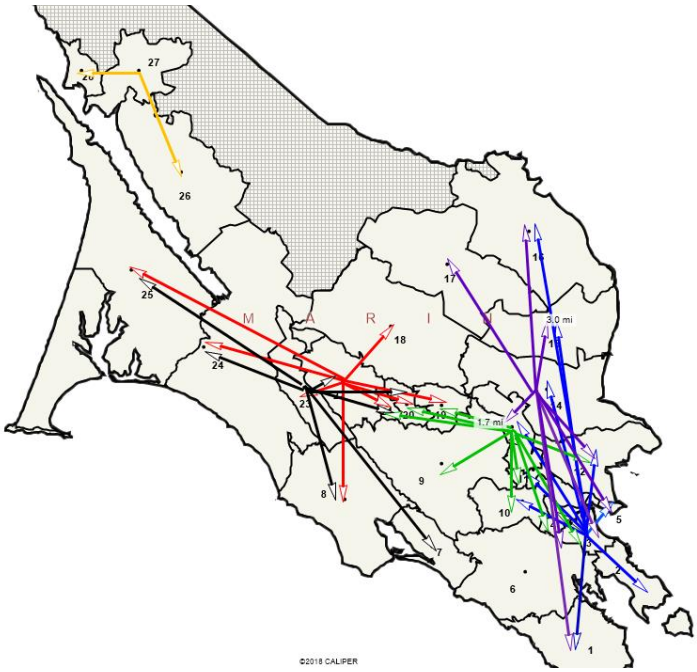

Figure 5.3 Détermination des postes d'attentes selon T15

Les figures 5.4 et 5.5 permettent d'avoir une meilleure visualisation des zones couvertes à partir des postes d'attente selon respectivement T10 et T15. Sur ces schémas, les losanges représentent les postes d'attentes et les cercles regroupent les zones atteignables à partir de ces postes d'attente. Certaines zones bénéficient d'une double couverture comme c'est le cas pour les zones 4, 5, 12 et 14 qui peuvent être couvertes à la fois par les postes 3 et 15 à l'intérieur d'un temps de déplacement de 10 minutes (voir la figure 5.4). Alors que d'autres zones bénéficient d'une triple couverture, comme c'est le cas pour les zones 4 et 12 qui sont atteignables à partir des postes d'attentes 3, 13 et 14 selon un temps de déplacement de 15 minutes (voir figure 5.5). Cela veut dire que si un poste d'attente est à court de ressources, il peut utiliser les ressources d'un autre poste d'attente en priorisant l'ambulance disponible la plus proche.

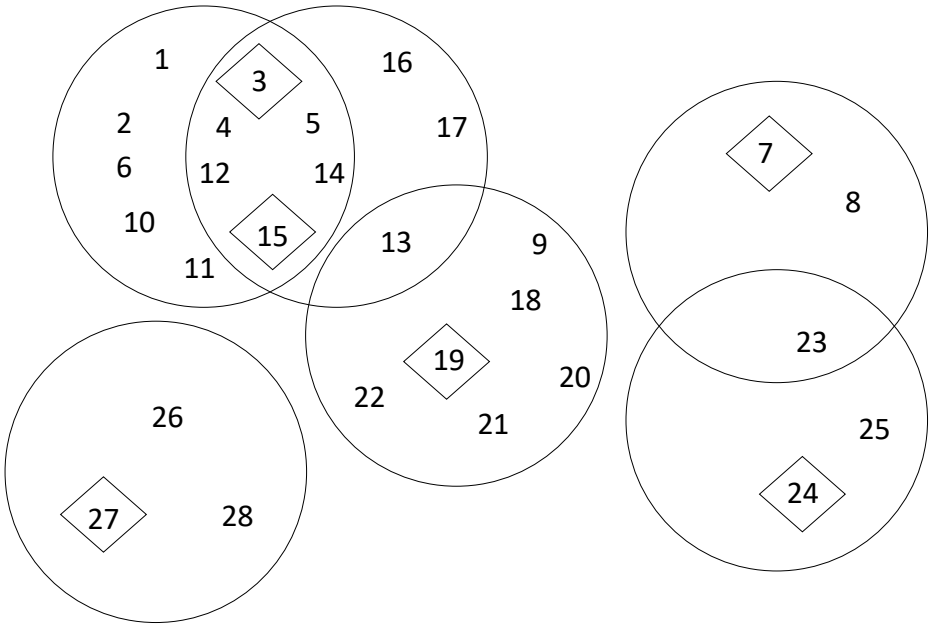

Figure 5.4 Couverture de toutes les zones par les postes d'attente en losanges selon T10

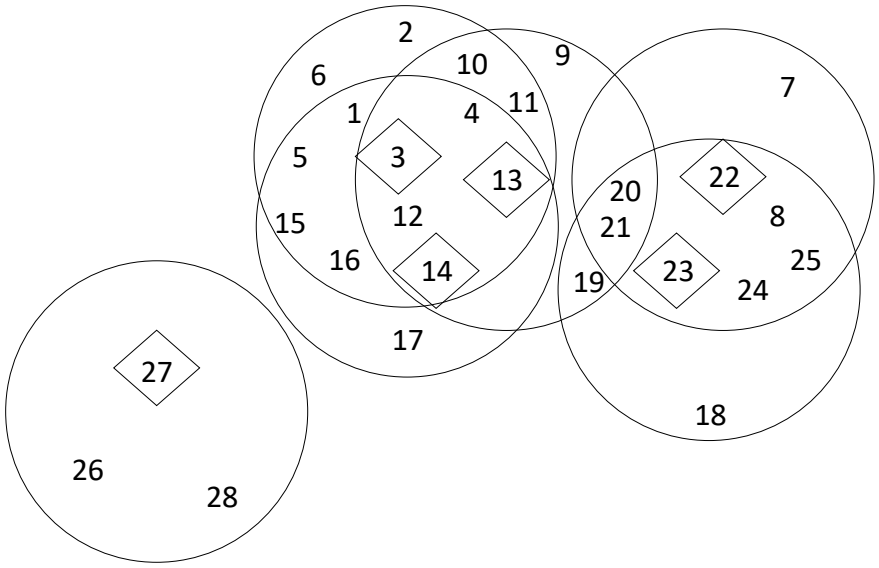

Figure 5.5 Couverture de toutes les zones par les postes d'attente en losanges selon T15

Maintenant que les postes d'attentes sont déterminés, il est nécessaire de déterminer le nombre de véhicules à affecter à chaque poste d'attente afin de répondre à l'ensemble des demandes dans les délais impartis.

5.3 Détermination du nombre d'ambulances

Comme nous l'avons déjà évoqué dans le chapitre précédent, le but des prévisions était de transformer les moyennes des prévisions en nombre d'appels probables à partir de la distribution de poisson selon deux niveaux de services, soit couvrir 95% ou bien 99% des appels. Le nombre d'ambulances est déterminé selon le niveau de couverture par zone et par heure de la journée. Par exemple, pour la période 0 heure du matin, si notre objectif est de couvrir 95% des appels de cette période, nous devrons donc couvrir 8 appels. En présumant qu'une ambulance peut répondre à un appel par heure, nous aurons besoin de 8 ambulances disponibles sur cette période d'une heure. Si notre objectif est de couvrir 99% des appels, nous utiliserons plutôt 15 ambulances (voir tableau 4.5 et 4.6 du chapitre précédent).

5.4 Conclusion

Dans ce chapitre, nous avons utilisé un modèle d'optimisation pour choisir les emplacements optimaux pour la localisation des postes d'attente des ambulances. Le

choix est déterminé selon deux temps de déplacement entre les zones. Les choix permettent de réduire le nombre de ressources à utiliser et permettent en même temps d'avoir une couverture totale de toutes les zones. À partir de là, nous pouvons affecter le nombre nécessaire de ressources aux ces postes d'attentes durant les différentes heures de la journée suivant le nombre d'appels probables.

Maintenant que nous avons déterminé le nombre de ressources nécessaires et les emplacements optimaux, il est important de valider ce choix avec la simulation. Pour cela, nous allons créer dans le prochain chapitre plusieurs scénarios pour évaluer les différentes situations possibles d'emplacement et de nombre d'ambulances.

Chapitre 6
Simulation de la demande

Pour valider les décisions qui ont été prises dans un contexte plus réaliste et déterminer la meilleure stratégie qui permettrait une utilisation plus efficiente des ressources, nous avons comparé deux approches de gestion : une gestion centralisée où chaque poste d'attente peut desservir l'ensemble de zones de demande et une gestion décentralisée où chaque poste d'attente et les ambulances qui y sont affectées sont assignées à un secteur donné, et ne peuvent être utilisées par d'autres secteurs.

6.1 Développement du modèle de simulation

Pour effectuer la simulation, nous avons utilisé le simulateur ARENA présenté au chapitre 2. Le principe est de générer des entités, correspondant aux appels qui entrent dans le système et se déplacent à travers différents modules reliés entre eux qui permettent de représenter l'ensemble des tâches réalisées, par exemple le déplacement de l'ambulance sur le lieu de l'incident, ou le transport du patient vers l'hôpital. La figure 6.1 représente l'outil de simulation que nous avons utilisé dans notre étude.

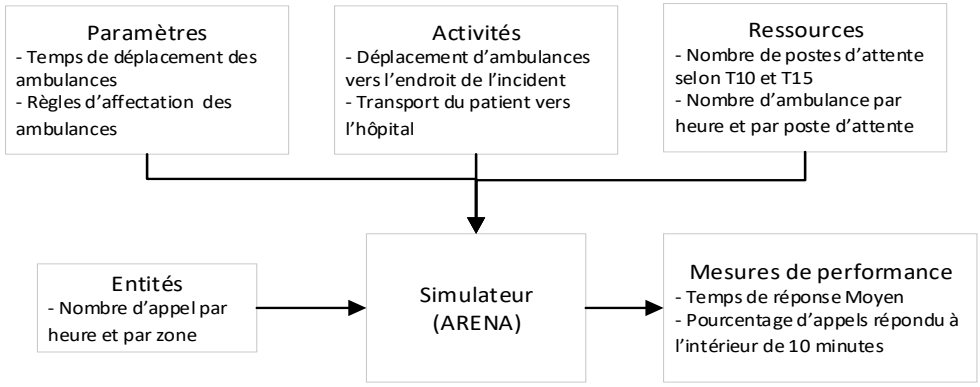

Figure 6.1 Modèle de simulation du système d'intervention des SPUs

Ces modules permettent d'effectuer les activités du processus pour un plus haut niveau d'authenticité. D'autres modules sont ajoutés pour enregistrer l'état d'évolution du système au fil du temps et mesurer sa performance. Les données sont programmées pour déterminer le nombre d'appels entrants par heure et par zone, ainsi que le nombre de ressources nécessaires par heure et par poste d'attente. Le principe est que, pour chaque appel entrant, nous assignons l'ambulance disponible la plus proche. Lorsque l'ambulance arrive sur le lieu de l'incident, elle doit transporter le patient vers l'hôpital le plus proche. Ce temps est constant et calculé à partir de la distance entre la zone desservie et l'hôpital le plus proche. Par la suite, l'ambulance quitte l'hôpital pour aller servir un autre appel ou bien retourner au poste d'attente. La figure 6.2 illustre une partie du processus de traitement d'un appel pour la zone 1. Pour chaque zone, un processus semblable est réalisé.

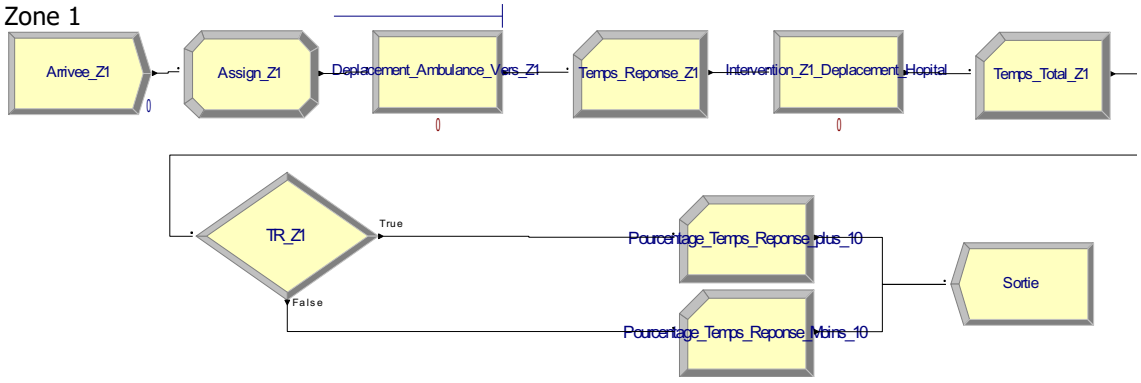

Figure 6.2 Processus d'intervention pour la zone 1

Le processus commence avec le module *Create* qui désigne les arrivées des appels selon les données incluses dans le module *Schedule* qui comporte le nombre d'appels moyen par heure et par zone pour la période de 24 heures. Par la suite le module *Assign* permet d'enregistrer le temps d'arrivée des appels. L'activité de déplacement de l'ambulance vers le lieu de l'incident est déterminée par le module *Process*, qui désigne selon un ordre de préférence, une ambulance à affecter à l'appel reçu et conséquemment, le temps de déplacement pour atteindre l'appel qui varie selon le poste d'attente de l'ambulance choisie. Un autre module semblable est utilisé pour déterminer le déplacement de l'ambulance du lieu de l'incident vers l'hôpital le plus proche. Dans ce cas, le temps de déplacement est constant. Les deux modules *process* sont reliés à des modules *Record*

pour calculer le temps de réponse moyen et le temps total moyen. Ces temps nous permettront d'évaluer la performance du système. Nous avons aussi voulu connaitre le nombre d'appels traités selon des temps prédéterminés pour l'ensemble du système. Pour cela, nous avons rajouté un module *Separate* qui permet de dupliquer le processus. Par la suite, nous avons relié chaque flux dupliqué à un module *Decide* pour calculer le nombre d'appels traités à l'intérieur d'un temps de réponse de 10 minutes. Finalement, le module *Dispose* est ajouté à la fin de chaque flux pour arrêter le processus et libérer les ressources. Lorsque les ressources ne sont pas disponibles, les appels vont attendre jusqu'à ce que la prochaine ressource soit disponible. Comme le nombre d'appels entrant varie par zone et par heure pour 24 heures, nous avons aussi fixé le nombre de ressources nécessaires pour chaque période grâce au module de données *Schedule*. Nous avons simulé des scénarios dans lesquels nous avons utilisé deux types d'approches : l'approche centralisée qui permet le partage des ressources de tous les postes d'attente et l'approche décentralisée qui permet l'utilisation des ressources d'un poste d'attente uniquement par les zones couvertes par ce poste. Dans la simulation et avec l'approche centralisée, nous assignons les ambulances aux appels selon un ordre de priorité déterminé en fonction de la distance entre le poste d'attente et le lieu de l'incident. Le nombre d'ambulances par poste d'attente diffère selon le niveau de service que nous voulons donner et la localisation des postes d'attente diffère selon le délai de déplacement T10 et T15.

6.2 Description des approches

6.2.1 Approche centralisée

L'approche centralisée consiste à utiliser en priorité les ressources du poste d'attente le plus proche pour desservir une zone en particulier. Si les ressources à ce poste d'attente ne sont pas disponibles, les ressources des autres postes d'attente seront utilisées pour répondre aux appels placés dans ces zones. Lorsque plusieurs ressources sont disponibles, la ressource la plus proche sera assignée à l'appel. Le but principal de cette approche est d'optimiser l'utilisation des ressources en cas de besoin. Les Figures 6.1 et 6.2 démontrent les solutions optimales des postes d'attentes pour atteindre les zones selon les deux temps de déplacement maximaux T10 et T15. Dans chaque figure, les zones en vert sont les zones couvertes par le poste d'attente de cette figure.

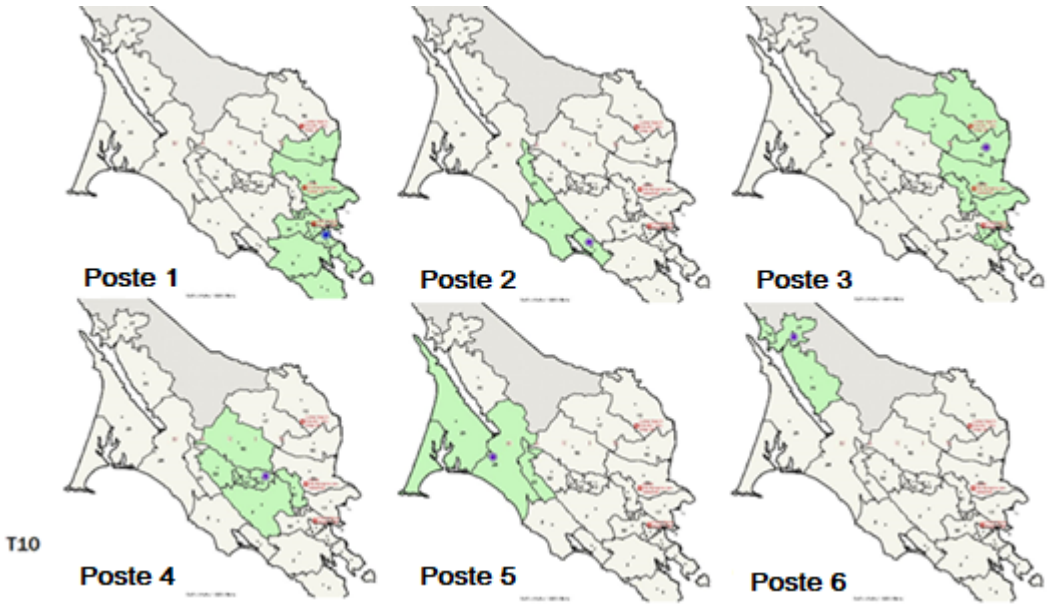

Figure 6.3 Couverture totale des zones selon T10

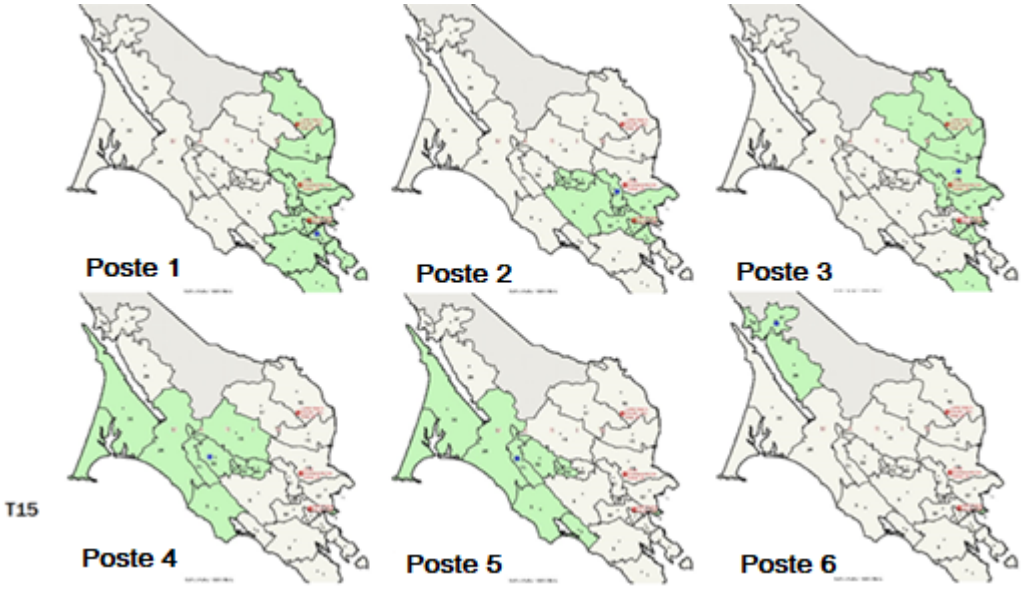

Figure 6.4 Couverture totale des zones selon T15

6.2.2 Approche décentralisée

L'approche décentralisée considère que chaque secteur est indépendant et gère sa propre flotte. Aucune ambulance ne peut intervenir dans une zone à l'extérieur de son secteur. Dans ce cas, les ressources sont partagées seulement à la verticale, c'est-à-dire, par les zones appartenant au même secteur. Un secteur est formé lorsque les zones d'un secteur sont atteignables à partir du poste d'attente à l'intérieur d'un temps prescrit. En revanche si une zone bénéficie d'une double ou triple couverture, elle fera partie du secteur avec le poste d'attente le plus proche de cette zone.

La Figure 6.3 illustre le découpage des secteurs selon les deux approches pour le temps de déplacement T10 et T15, chaque secteur étant représenté par une couleur différente.

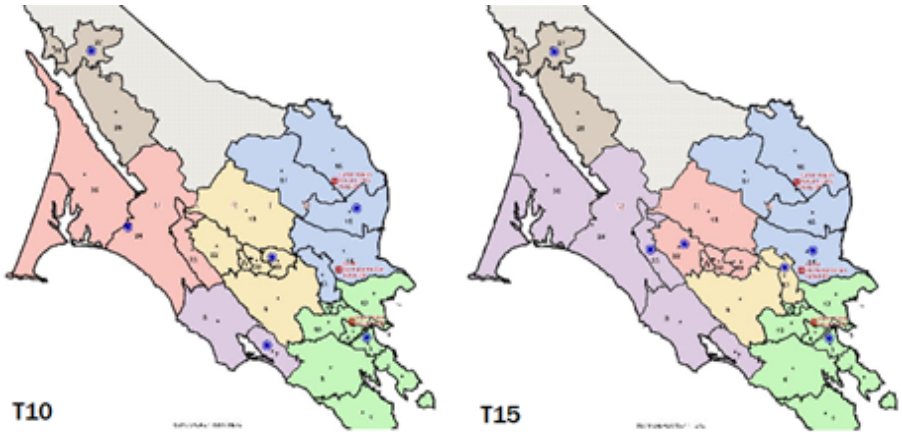

Figure 6.5 Couverture totale des zones selon T10 et T15

6.3 Scénarios de simulation

Nous avons donc créé et comparé huit scénarios. Le tableau 6.1 illustre les différents scénarios développés, incluant les deux approches, les temps T10 et T15 ainsi que les deux niveaux de couverture d'appels.

Tableau 6.1 Caractéristiques des différents scénarios étudiés

Scénarios	Délai cible	Pourcentage d'appels répondus	Approche de gestion
S1	10	95%	Centralisée
S2	10	99%	Centralisée
S3	15	95%	Centralisée
S4	15	99%	Centralisée
S5	10	95%	Décentralisée
S6	10	99%	Décentralisée
S7	15	95%	Décentralisée
S8	15	99%	Décentralisée

Pour les scénarios, nous avons effectué 100 itérations de 24 heures en considérant une période de transition d'une heure. Pour évaluer les résultats de simulation, nous avons choisi d'utiliser les temps de réponse moyens et le pourcentage du nombre d'appels répondus à l'intérieur d'un temps limite de 10 minutes pour le temps de réponse comme mesures de performance et le nombre total d'heures travaillées comme mesure associée à l'utilisation des ressources et les coûts qui en découlent.

6.4 Résultats de la simulation

Nous avons résumé les résultats des huit scénarios au tableau 6.2. En comparant ces résultats, il est clair que l'approche centralisée présentée dans les quatre premiers scénarios permet d'avoir de meilleures performances au niveau du temps de réponse moyen, car, avec le partage de ressources entre les postes d'attente, le système est plus flexible et permet d'atteindre de meilleurs résultats.

Au niveau des temps de déplacement, les scénarios avec un temps T10 présentent de meilleures performances que le temps T15 parce que les délais de déplacement entre les postes d'attentes et les zones qui lui sont assignées sont plus courts. Avec les 2 niveaux de service, nous obtenons de meilleures performances avec le niveau de couverture de 99% puisque nous utilisons plus de ressources que pour couvrir 95% des appels.

Le scénario S6 de l'approche décentralisée présente d'aussi bons résultats que le scénario S2. À ce niveau, le nombre de ressources est si élevé qu'il n'y presque pas de différence

entre les deux approches au niveau du temps de réponse moyen, et aussi au niveau du nombre d'appels répondu à l'intérieur de dix minutes qui représente 70% du nombre d'appels traités pour S2, et 69% pour S6 (voir tableau 6.2).

Tableau 6.2 Résultats des scénarios

Scénario	Appel traité moyen	Nombre maximal d'ambulances par heure	Nombre total d'heures travaillées	Taux d'utilisation d'ambulances	Temps de réponse moyen (minutes)	Pourcentage du temps de réponse < 10 minutes
S1	312	14	271	37%	7,85	61%
S2	312	31	538	12%	6,97	70%
S3	312	14	271	26%	8,32	62%
S4	312	31	538	13%	7,58	64%
S5	310	14	271	43%	11,04	60%
S6	312	31	538	23%	6,96	69%
S7	299	14	271	42%	35,64	41%
S8	312	31	538	24%	8,44	45%

Afin d'évaluer l'impact de la variation du nombre de ressources sur la performance du système, nous avons augmenté le nombre d'heures travaillées de 10% pour les huit scénarios, puis réduit le nombre de ressources pour les quatre meilleurs scénarios de la stratégie centralisée seulement. Nous avons évalué le temps de réponse moyen ainsi que le pourcentage du nombre d'appels répondus en bas de 10 minutes. Les résultats sont présentés au tableau 6.3 et pour le tableau 6.4 pour l'augmentation et la réduction du nombre de ressources respectivement.

L'augmentation du nombre de ressources a permis d'améliorer les résultats légèrement pour les scénarios qui prennent en compte l'approche centralisée (S1 et S3). En revanche, avec l'approche décentralisée, l'augmentation du nombre de ressources augmente considérablement le niveau de performance du système (S6 et S8). Ceci confirme que l'approche décentralisée donne d'aussi bons résultats que l'approche centralisée, mais

requiert un plus grand nombre de ressources pour atteindre des performances équivalentes.

Tableau 6.3 Résultats après augmentation du nombre de ressources de 10%

Scénario	Nombre total d'heures travaillées	Taux d'utilisation d'ambulances	Temps de réponse moyen (minutes)	Temps Total moyen (minutes)	% Temps de réponse < 10 minutes
S1	298	34%	7,74	18,05	62%
S2	592	21%	6,97	17,27	70%
S3	298	24%	8,15	15,69	63%
S4	592	13%	7,57	15,11	64%
S5	298	41%	11,02	21,21	62%
S6	592	17%	6,77	17,08	71%
S7	298	41%	35,62	43,16	41%
S8	592	17%	7,59	15,13	64%

Cependant, lorsque l'on réduit le nombre de ressources, les scénarios qui sont moins affectés par la variation du temps de réponse sont S1 et S2 pour répondre respectivement à 95% et 99% des appels. Lorsque l'on réduit les ressources de 10%, le temps de réponse moyen augmente de 0,51 minutes avec S1, et de 0,01 minutes avec S2. En revanche, avec S3 et S4 ce temps augmente de 1,05 et 0,04 minutes respectivement. Avec une réduction de 20% des ressources, le temps de réponse moyen augmente de 1,67 minutes pour S1 pour atteindre 10,03 minutes, et augmente de 0,04 minutes pour S2. Cependant, ce temps augmente de 4,45 minutes pour S3 et de 0,17 minutes pour S4. Cela dit, nous pouvons arriver à avoir un temps de réponse moyen de dix minutes avec un niveau de service de 95% et une réduction de ressources de 20%. D'une autre manière, nous pouvons réduire jusqu'à 30% des ressources pour un niveau de service de 99% et avoir un temps de réponse moyen aussi bas que 7,5 minutes (voir tableau 6.4). Cela dit, le niveau de service augmente avec l'augmentation du nombre de ressources.

Tableau 6.4 Résultats de réduction de ressources pour les scénarios de l'approche centralisée

Scénario	Pourcentage de réduction du nombre de ressources	Temps de réponse moyen (minutes)	Taux d'utilisation des ressources	Pourcentage de Temps de réponse < 10 minutes
S1	10%	8,36	41%	59%
	20%	10,03	48%	46%
	30%	19,59	60%	20%
	40%	50,37	73%	17%
	50%	145,09	80%	10%
S2	10%	6,98	23%	70%
	20%	7,02	26%	70%
	30%	7,50	31%	69%
	40%	12,03	45%	45%
	50%	71,91	72%	12%
S3	10%	9,37	32%	59%
	20%	13,82	43%	35%
	30%	23,78	54%	20%
	40%	18,59	58%	24%
	50%	50,75	72%	9%
S4	10%	7,62	15%	64%
	20%	7,79	18%	64%
	30%	8,01	21%	63%
	40%	13,15	31%	52%
	50%	18,77	55%	15%

Si nous planifions les ressources en considérant un niveau de service de 95%, le temps de réponse moyen que nous obtenons est de 7,85 minutes avec 61% d'appels traités à l'intérieur de 10 minutes. Cependant, Si nous planifions les ressources en considérant un niveau de service de 99%, le temps de réponse moyen obtenu est 6,97 avec 70 % d'appels traités à l'intérieur de 10 minutes. La différence entre les deux objectifs est le nombre de ressources planifiées pour répondre aux appels.

L'objectif des SPUs du Marin est d'avoir un temps de réponse moyen égal ou inférieur à 10 minutes. Pour une utilisation minimale des ressources, il est possible d'avoir un temps de réponse moyen égal à dix minutes avec une réduction de 20% du nombre de ressources avec S1 (voir figure 6.6), et un temps de réponse moyen de 7,50 minutes avec une réduction de 30% du nombre de ressources avec S2 (voir figure 6.7). À ce moment-là, le nombre d'appels répondu à l'intérieur de 10 minutes est 46% pour S1 et 69% pour S2. Pour atteindre ces résultats, il faut planifier un nombre maximal de 14 ambulances pour un niveau de service de 95% et 31 ambulances pour un niveau de service de 99%. Il faut aussi planifier un nombre d'heures travaillées de 271 heures pour un niveau de service de 95% et 538 heures pour un niveau de service de 99%.

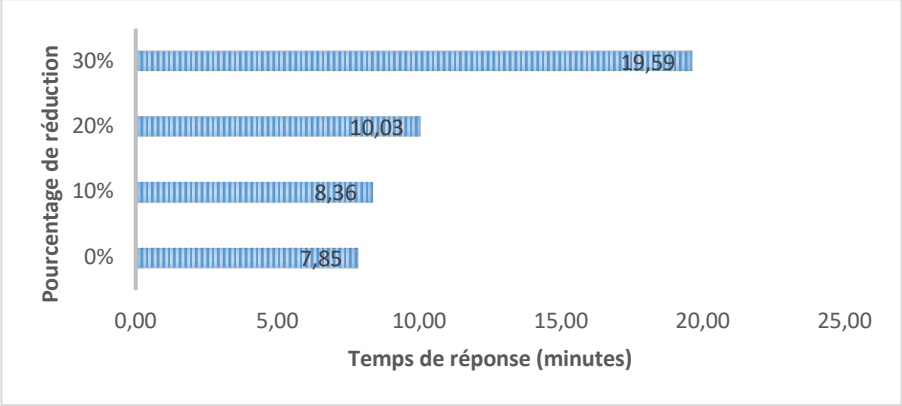

Figure 6.6 Réduction du nombre de ressources avec une couverture d'appels de 95%

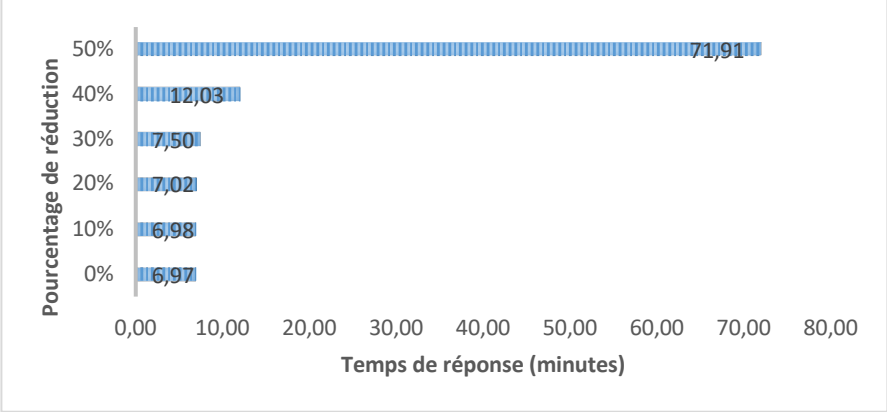

Figure 6.7 Réduction du nombre de ressources avec une couverture d'appels de 99%

6.5 Conclusion

La stratégie de partage de ressources entre les postes d'attente permet d'avoir de meilleurs résultats au niveau de temps de réponse moyen. Cela confirme que même si certains postes sont plus éloignés des zones de demande, avoir recours aux ressources d'autres postes permet d'avoir une meilleure performance que dans le cas où les ressources ne sont pas partagées entre les différents postes d'attente. Cependant, si le nombre de ressources est très grand, l'approche décentralisée peut donner de bonnes performances avec le temps de déplacement T10. En revanche, notre objectif est d'utiliser le minimum de ressources tout en gardant un niveau de service adéquat. Avec la réduction du nombre d'ambulances, et conséquemment, d'heures travaillées, nous avons pu avoir de bonnes performances jusqu'à une certaine limite. Les meilleurs scénarios sont ceux que nous avons effectués avec le temps de déplacement T10 au niveau de couverture de 95% et 99%. Dépendamment du nombre d'appels que nous voulons couvrir, le scénario 1 et le scénario 2 présentent les meilleurs résultats.

Conclusion

D'un point de vue économique et social, l'objectif de cette étude est de trouver un outil capable de réduire les coûts d'exploitation des ressources des SPUs tout en offrant un service adéquat à la population. Plusieurs choix de stratégies sont possibles désormais à partir des différentes solutions de couverture et assignations des ambulances aux postes d'attentes potentiels. Le choix optimal des postes d'attente permet à certaines zones d'avoir une double ou bien triple couverture. Cela dit, le partage de ressources entre les postes d'attente permet d'avoir de meilleures performances au niveau du temps de réponse. Le temps de déplacement entre les zones permet d'avoir différents choix de poste d'attente et à un impact sur le temps de réponse. D'un autre côté, le nombre d'ambulances par poste d'attente est déterminé à partir du nombre d'appels probables par heure et par zones couvertes par ce poste selon le niveau de service voulu. Si l'objectif est de répondre à 99% des appels, le nombre total d'heures travaillées est presque deux fois plus que le nombre total d'heures travaillées pour répondre à 95% des appels. Il faut donc choisir entre niveaux de service et les ressources utilisées. En revanche, il est possible d'améliorer ce choix en réduisant les heures travaillées pour les deux niveaux de service. Les résultats ont montré que plusieurs solutions sont possibles pour trouver un équilibre entre le temps réponse, le nombre de ressources et le niveau de service souhaité. La meilleure stratégie sera donc de combiner plusieurs solutions pour répondre à l'objectif des SPUs et trouver la combinaison clé entre économie et niveau de service.

Bibliographies

Aboueljinane, Lina, Evren Sahin et Zied Jemai (2013). « A review on simulation models applied to emergency medical service operations », *Computers & Industrial Engineering*, vol. 66, no 4, p. 734-750.

Aldrich, Carole A, John C Hisserich et Lester B Lave (1971). « An analysis of the demand for emergency ambulance service in an urban area », *American Journal of Public Health*, vol. 61, no 6, p. 1156-1169.

Babu, C Narendra et B Eswara Reddy (2014). « A moving-average filter based hybrid arima–ann model for forecasting time series data », *Applied Soft Computing*, vol. 23, p. 27-38.

Baker, JR et KE Fitzpatrick (1986). « Determination of an optimal forecast model for ambulance demand using goal programming », *Journal of the Operational Research Society*, vol. 37, no 11, p. 1047-1059.

Bélanger, A. Ruiz et P. Soriano (2019). « Recent optimization models and trends in location, relocation, and dispatching of emergency medical vehicles », *European Journal of Operational Research*, vol. 272, no 1, p. 1-23.

Bélanger, Angel Ruiz et Patrick Soriano (2012). « Déploiement et redéploiement des véhicules ambulanciers dans la gestion d'un service préhospitalier d'urgence », *INFOR: Information Systems and Operational Research*, vol. 50, no 1, p. 1-30.

Bell, William R et Steven C Hillmer (1983). « Modeling time series with calendar variation », *Journal of the American statistical Association*, vol. 78, no 383, p. 526-534.

Box, George EP et Gwilym M Jenkins (1976). « Time series analysis: Forecasting and control san francisco », *Calif: Holden-Day*.

Breiman, Leo (2001). « Random forests », *Machine learning*, vol. 45, no 1, p. 5-32.

Channouf, Nabil, Pierre L'Ecuyer, Armann Ingolfsson et Athanassios N Avramidis (2007). « The application of forecasting techniques to modeling emergency medical system calls in calgary, alberta », *Health care management science*, vol. 10, no 1, p. 25-45.

Chen, Albert Y, Tsung-Yu Lu, Matthew Huei-Ming Ma et Wei-Zen Sun (2015). « Demand forecast using data analytics for the preallocation of ambulances », *IEEE journal of biomedical and health informatics*, vol. 20, no 4, p. 1178-1187.

Chong, Il-Gyo et Chi-Hyuck Jun (2005). « Performance of some variable selection methods when multicollinearity is present », *Chemometrics and intelligent laboratory systems*, vol. 78, no 1-2, p. 103-112.

Church, Richard et Charles ReVelle (1974). « The maximal covering location problem », *Papers in regional science*, vol. 32, no 1, p. 101-118.

Daskin et Edmund H Stern (1981). « A hierarchical objective set covering model for emergency medical service vehicle deployment », *Transportation Science*, vol. 15, no 2, p. 137-152.

Denton, James W (1995). « How good are neural networks for causal forecasting? », *The Journal of Business Forecasting*, vol. 14, no 2, p. 17.

Derksen, Shelley et Harvey J Keselman (1992). « Backward, forward and stepwise automated subset selection algorithms: Frequency of obtaining authentic and noise variables », *British Journal of Mathematical and Statistical Psychology*, vol. 45, no 2, p. 265-282.

Doerner, Karl F, Walter J Gutjahr, Richard F Hartl, Michaela Karall et Marc Reimann (2005). « Heuristic solution of an extended double-coverage ambulance location problem for austria », *Central European Journal of Operations Research*, vol. 13, no 4, p. 325-340.

Gamberini, Rita, Francesco Lolli, Bianca Rimini et Fabio Sgarbossa (2010). « Forecasting of sporadic demand patterns with seasonality and trend components: An empirical comparison between holt-winters and (s) arima methods », *Mathematical Problems in Engineering*, vol. 2010.

Gendreau, Michel, Gilbert Laporte et Frédéric Semet (1997). « Solving an ambulance location model by tabu search », *Location science*, vol. 5, no 2, p. 75-88.

Genuer, Robin, Jean-Michel Poggi et Christine Tuleau-Malot (2010). « Variable selection using random forests », *Pattern Recognition Letters*, vol. 31, no 14, p. 2225-2236.

Guerif, Sebastien et Younès Bennani (2007). « Sélection de variables en apprentissage numérique non supervisé ».

Gunes, Ersan et Roberto Szechtman (2005). « A simulation model of a helicopter ambulance service », communication présentée au *Proceedings of the 37th conference on Winter simulation*, Orlando, Florida.

Han, Kai, Chao Li et Xin Shi (2017). « Autoencoder feature selector », *arXiv preprint arXiv:1710.08310*.

Hinton, Geoffrey E, Simon Osindero et Yee-Whye Teh (2006). « A fast learning algorithm for deep belief nets », *Neural computation*, vol. 18, no 7, p. 1527-1554.

Holt, CC (1957). « Forecasting seasonals and trends by exponentially weighted moving averages, office of naval research », *Research memorandum*, vol. 52.

Hyndman, Rob J et Yeasmin Khandakar (2007). *Automatic time series for forecasting: The forecast package for r*, Monash University, Department of Econometrics and Business Statistics ….

Iannoni, Ana Paula, Reinaldo Morabito et Cem Saydam (2009). « An optimization approach for ambulance location and the districting of the response segments on highways », *European Journal of Operational Research*, vol. 195, no 2, p. 528-542.

Jovic, Alan, Karla Brkić et N. Bogunovic (2015). *A review of feature selection methods with applications*, 1200-1205 p.

Kamenetzky, Ricardo D, Larry J Shuman et Harvey Wolfe (1982). « Estimating need and demand for prehospital care », *Operations research*, vol. 30, no 6, p. 1148-1167.

Kane, Michael J, Natalie Price, Matthew Scotch et Peter Rabinowitz (2014). « Comparison of arima and random forest time series models for prediction of avian influenza h5n1 outbreaks », *BMC bioinformatics*, vol. 15, no 1, p. 276.

Kelton, W David (2002). *Simulation with arena*, McGraw-hill.

Kvålseth, Tarald O et John M Deems (1979). « Statistical models of the demand for emergency medical services in an urban area », *American Journal of Public Health*, vol. 69, no 3, p. 250-255.

Lin, Ziming Wu, Longxin Lin, Angzhan Wen et Jin Li (2017a). « An ensemble random forest algorithm for insurance big data analysis », *Ieee Access*, vol. 5, p. 16568-16575.

Lin, Lin, Fang Wang, Xiaolong Xie et Shisheng Zhong (2017b). « Random forests-based extreme learning machine ensemble for multi-regime time series prediction », *Expert Systems with Applications*, vol. 83, p. 164-176.

Lubicz, Marek et Bożena Mielczarek (1987). « Simulation modelling of emergency medical services », *European Journal of Operational Research*, vol. 29, no 2, p. 178-185.

Matteson, David S, Mathew W McLean, Dawn B Woodard et Shane G Henderson (2011). « Forecasting emergency medical service call arrival rates », *The Annals of Applied Statistics*, vol. 5, no 2B, p. 1379-1406.

McCarthy, Melissa L, Scott L Zeger, Ru Ding, Dominik Aronsky, Nathan R Hoot et Gabor D Kelen (2008). « The challenge of predicting demand for emergency department services », *Academic Emergency Medicine*, vol. 15, no 4, p. 337-346.

McConnel, Charles E et Rosemary W Wilson (1998). « The demand for prehospital emergency services in an aging society », *Social Science & Medicine*, vol. 46, no 8, p. 1027-1031.

Micheletti, Alessandra, Daniela Morale, Daniele Rapati et Paola Nolli (2010). *A stochastic model for simulation and forecasting of emergencies in the area of milano*, 1-6 p.

Peleg, Kobi et Joseph S Pliskin (2004). « A geographic information system simulation model of ems: Reducing ambulance response time », *The American journal of emergency medicine*, vol. 22, no 3, p. 164-170.

Pinto, LR, PMS Silva et TP Young (2015). « A generic method to develop simulation models for ambulance systems », *Simulation Modelling Practice and Theory*, vol. 51, p. 170-183.

Probst, Philipp et Anne-Laure Boulesteix (2017). « To tune or not to tune the number of trees in random forest », *Journal of Machine Learning Research*, vol. 18, p. 181:181-181:118.

Repede, John F et John J Bernardo (1994). « Developing and validating a decision support system for locating emergency medical vehicles in louisville, kentucky », *European journal of operational research*, vol. 75, no 3, p. 567-581.

Reuter-Oppermann, Melanie, Pieter L van den Berg et Julie L Vile (2017). « Logistics for emergency medical service systems », *Health Systems*, vol. 6, no 3, p. 187-208.

Rojas-Trejos, Carlos Alberto, Julián González-Velasco et Mario Alberto López-Ramírez (2017). « Optimization model for the location of prehospital care ambulances in the city of cali colombia », *International Journal of Combinatorial Optimization Problems and Informatics*, vol. 8, no 3, p. 64-70.

Savas, ES (1969). « Simulation and cost-effectiveness analysis of new york's emergency ambulance service », *Management science*, vol. 15, no 12, p. B-608-B-627.

Schapire, Robert E (2003). « The boosting approach to machine learning: An overview », dans *Nonlinear estimation and classification*, Springer, p. 149-171.

Setzler, Hubert, Cem Saydam et Sungjune Park (2009). « Ems call volume predictions: A comparative study », *Computers & Operations Research*, vol. 36, no 6, p. 1843-1851.

Storbeck, James E (1982). « Slack, natural slack, and location covering », *Socio-Economic Planning Sciences*, vol. 16, no 3, p. 99-105.
Swoveland, C, D Uyeno, I Vertinsky et R Vickson (1973). « Ambulance location: A probabilistic enumeration approach », *Management Science*, vol. 20, no 4-part-ii, p. 686-698.
Tavakoli, Asad et Constance Lightner (2004). « Implementing a mathematical model for locating ems vehicles in fayetteville, nc », *Computers & Operations Research*, vol. 31, no 9, p. 1549-1563.
Toregas, Constantine, Ralph Swain, Charles ReVelle et Lawrence Bergman (1971). « The location of emergency service facilities », *Operations research*, vol. 19, no 6, p. 1363-1373.
Trudeau, Pierre, Jean-Marc Rousseau, Jacques A Ferland et Jean Choquette (1989). « An operations research approach for the planning and operation of an ambulance service », *INFOR: Information Systems and Operational Research*, vol. 27, no 1, p. 95-113.
Tyralis, Hristos et Georgia Papacharalampous (2017). « Variable selection in time series forecasting using random forests », *Algorithms*, vol. 10, no 4, p. 114.
van Barneveld, TC, Sandjai Bhulai et Robert D van der Mei (2016). « The effect of ambulance relocations on the performance of ambulance service providers », *European Journal of Operational Research*, vol. 252, no 1, p. 257-269.
Vile, Julie Leanne, JW Gillard, Paul R Harper et Vincent A Knight (2012). « Predicting ambulance demand using singular spectrum analysis », *Journal of the Operational Research Society*, vol. 63, no 11, p. 1556-1565.
Villegas, Juan G, Carolina Castañeda et Katherine A Blandón (2012). « Mejoramiento de la localización de ambulancias de atención prehospitalaria en medellín (colombia) con modelos de optimización », *CLAIO/SBPO2012*, vol. 123, p. 12.
Winters, Peter R (1960). « Forecasting sales by exponentially weighted moving averages », *Management science*, vol. 6, no 3, p. 324-342.
Yang, Cheng-Huei, Li-Yeh Chuang et C Hong Yang (2010). « Ig-ga: A hybrid filter/wrapper method for feature selection of microarray data », *Journal of Medical and Biological Engineering*, vol. 30, no 1, p. 23-28.
Zhu, Zhiwei, Mark McKnew et Jim Lee (1992). *Effects of time-varied arrival rates: An investigation in emergency ambulance service systems*, 1180-1186 p.
Zou, Hui et Yuhong Yang (2004). « Combining time series models for forecasting », *International journal of Forecasting*, vol. 20, no 1, p. 69-84.

Annexes

Annexe 1 : Tableau 4.1 distance entre les zones avec une vitesse de conduite moyen-rapide

	z1	z2	z3	z4	z5	z6	z7	z8	z9	z10	z11	z12	z13	z14	z15	z16	z17	z18	z19	z20	z21	z22	z23	z24	z25	z26	z27	z28
z1	0	16	10	12	14	16	29	36	31	20	20	17	22	14	17	19	27	31	27	27	29	33	40	44	47	46	41	47
z2	15	0	10	12	15	17	32	38	31	19	20	17	22	17	17	19	27	32	26	27	28	33	40	44	46	46	41	47
z3	10	10	0	5	6	13	27	34	23	13	13	9	14	10	10	12	20	24	20	19	22	26	32	36	41	40	34	40
z4	13	13	5	0	8	15	30	36	20	9	10	10	13	10	10	14	20	25	19	18	20	24	32	36	39	40	35	41
z5	13	13	6	8	0	16	30	37	25	15	15	10	16	10	10	14	20	25	20	20	23	27	34	38	41	40	35	41
z6	16	17	10	14	17	0	15	20	29	20	23	19	24	19	19	21	29	34	29	29	31	34	27	36	40	48	43	49
z7	29	31	26	28	31	15	0	6	30	36	36	33	36	33	33	35	33	28	28	26	25	19	12	20	25	32	39	45
z8	35	38	32	34	37	20	6	0	28	39	36	39	30	33	34	31	27	23	23	20	19	14	7	16	19	27	34	39
z9	31	30	24	20	25	30	29	28	0	24	20	23	14	24	24	27	34	24	17	17	19	20	30	34	38	44	45	54
z10	20	19	10	8	14	23	36	40	24	0	13	17	15	17	17	20	27	27	20	20	22	26	33	38	41	46	41	47
z11	19	19	10	7	13	20	36	36	20	10	0	16	12	16	14	19	24	22	16	15	16	20	29	35	36	44	41	44
z12	15	16	9	10	10	19	32	39	23	16	15	0	15	10	10	12	20	23	19	19	20	26	32	36	39	39	34	39
z13	22	20	14	10	15	25	35	30	15	15	10	13	0	14	10	18	25	17	10	10	12	17	24	27	31	37	38	45
z14	17	17	10	10	10	20	33	33	25	18	18	9	16	0	5	8	15	17	19	17	19	23	26	30	33	34	30	37
z15	17	18	10	14	13	20	35	32	27	20	18	10	18	6	0	4	10	20	22	20	22	24	26	32	32	30	26	32
z16	20	19	13	16	15	22	36	30	30	20	20	15	19	8	6	0	12	22	24	20	24	23	24	28	32	29	22	29
z17	27	27	20	20	20	30	31	26	34	28	28	19	27	15	10	9	0	17	20	18	20	18	19	23	26	24	25	30
z18	20	31	25	25	25	34	28	23	24	27	24	23	18	17	19	20	18	0	10	9	10	14	16	20	23	27	28	34
z19	25	27	20	16	20	30	27	20	17	20	17	19	10	19	20	23	22	10	0	4	5	9	16	20	24	31	32	44
z20	26	26	20	18	20	29	26	20	16	19	16	19	10	17	19	22	19	9	3	0	3	7	14	18	22	28	30	35
z21	28	29	23	19	23	32	24	20	19	20	18	20	10	19	22	24	20	10	5	3	0	7	13	18	20	29	32	37
z22	33	32	26	23	27	34	19	13	20	26	20	25	16	23	26	23	19	15	9	7	6	0	7	12	15	23	30	35
z23	39	39	33	30	34	27	10	7	29	32	29	32	23	26	27	24	20	16	16	14	13	7	0	8	12	20	27	33
z24	44	44	38	35	38	35	20	15	34	37	34	36	28	30	31	28	24	20	20	18	17	10	9	0	10	20	27	33
z25	47	47	41	38	42	39	24	19	37	40	37	39	31	33	34	31	28	23	24	22	20	15	12	10	0	20	31	36
z26	48	48	42	43	42	46	31	29	43	46	42	40	36	36	33	30	26	26	29	27	27	22	19	19	23	0	10	17
z27	43	43	37	38	37	46	37	32	45	44	44	35	39	31	27	25	25	28	32	30	31	28	26	27	29	10	0	6
z28	48	48	42	44	43	51	43	38	55	50	49	41	48	37	33	31	31	34	37	35	37	34	31	32	35	16	6	0

Annexe 2 :Tableau des temps de déplacements entre les zones les trois hôpitaux du comté

	Hopital1	Hopital2	Hopital3	Minimum
zone1	13	16	20	13
zone2	13	16	20	13
zone3	6	9	13	6
zone4	3	13	15	3
zone5	7	10	13	7
zone6	16	19	23	16
zone7	29	33	36	29
zone8	36	35	33	33
zone9	20	25	28	20
zone10	8	17	21	8
zone11	6	15	18	6

zone12	11	9	12	9
zone13	11	16	19	11
zone14	10	5	7	5
zone15	10	6	4	4
zone16	15	10	6	6
zone17	20	16	11	11
zone18	24	20	23	20
zone19	16	21	25	16
zone20	18	20	23	18
zone21	19	23	26	19
zone22	23	26	25	23
zone23	29	29	26	26
zone24	34	32	30	30
zone25	37	35	33	33
zone26	40	35	31	31
zone27	35	30	25	25
zone28	40	37	32	32