



HAL
open science

Méthodologie d'évaluation du risque aérien sur un site sensible

Pierre Masselot, Pascal Lezaud, Karim Mehadhebi

► **To cite this version:**

Pierre Masselot, Pascal Lezaud, Karim Mehadhebi. Méthodologie d'évaluation du risque aérien sur un site sensible. 2013. hal-00874397

HAL Id: hal-00874397

<https://enac.hal.science/hal-00874397>

Submitted on 21 Nov 2013

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

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

RAPPORT DE RECHERCHE

Mai 2013



Méthodologie d'évaluation du risque aérien sur un site sensible



N° : RR/ENAC/2013/01

La référence aéronautique

www.enac.fr →

École Nationale de l'Aviation Civile

Auteurs :

Pierre Masselot*, Pascal Lezaud*, Karim Mehadhebi**

*ENAC, MAIAA, F-31055 Toulouse, France

** DSNA-DTI



Résumé : Ce rapport présente une méthodologie probabiliste d'estimation de risque de chute d'un aéronef sur une installation sensible (site nucléaire, installation chimique, par exemple) en se fondant sur des données caractérisant les flux de trafic aérien au voisinage du site sensible considéré. Seule l'aviation commerciale a été prise en compte.

Une première approche, s'appuyant sur les données radar, est basée sur un processus de Poisson de droites. La deuxième approche qui s'appuie sur les données plan de vol est basée sur une analyse statistique des écarts entre la trajectoire réelle et une trajectoire idéalisée par des tronçons rectilignes.

Méthodologie d'évaluation du risque aérien sur un site sensible

P. Masselot, P. Lezaud, K. Mehadhebi

27 mai 2013

1 Introduction

L'objectif de ce document est de présenter une méthodologie probabiliste d'estimation de risque de chute d'un aéronef sur une installation sensible (site nucléaire, installation chimique, par exemple) en se fondant sur des données caractérisant les flux de trafic aérien au voisinage du site sensible considéré. En outre, cette méthodologie devra être conforme à la réglementation applicable pour les études de sécurité dans le domaine de la navigation aérienne telle que précisée dans la circulaire européenne CE2096.

Le point de vue adopté dans ce rapport consiste à définir « la chute d'un aéronef sur une installation sensible » comme accident ou événement catastrophique. Un tel choix permet d'utiliser la terminologie introduite dans la circulaire CE2096, en particulier le terme de danger ou d'événement redouté devra être compris comme étant « toute condition, événement ou circonstance qui pourrait provoquer la chute d'un aéronef sur un site sensible ». La méthodologie adoptée est fondée sur l'hypothèse qu'un accident est le résultat d'une succession d'événements conduisant le système d'un état considéré comme « sûr » à des états de plus en plus « dégradés » jusqu'à l'occurrence de l'accident. En quelque sorte, pour chaque scénario considéré, le système suit une trajectoire dans son espace d'état aboutissant à l'accident. Un danger sera un état dégradé particulier, choisi sur une des trajectoires probables. Les états de la trajectoire en amont de ce danger seront les causes du danger, tandis que les états en aval seront les effets du danger ; l'effet final étant l'accident (cf. figure 1).

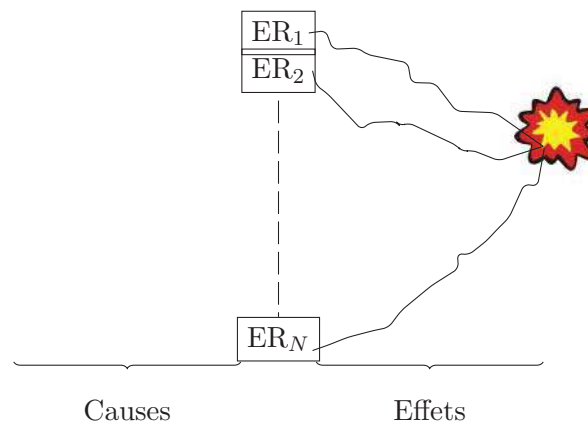


FIGURE 1 – Illustration *en tamis* des dangers

Ainsi, estimer le risque de chute nécessitera une première étape pour identifier les événements redoutés et en modéliser les causes ainsi que les effets, puis une seconde pour estimer les probabilités d'occurrence de ces événements. Enfin l'étape finale, devra permettre d'estimer le risque associé à l'accident par l'étude probabiliste des effets de chaque événement redouté.

Ce document se découpe en une première partie qui précisera la méthode mise en œuvre conformément à la circulaire CE2096, ainsi que la métrique utilisée pour exprimer le risque de chute d'un aéronef. Par exemple dans le domaine du trafic aérien, le risque d'accident aérien s'exprime en nombre d'accident par heure de vol ; il dépend du type de trafic et de la phase de vol (décollage – atterrissage, croisière).

Dans une seconde partie, nous détaillerons les données que nous utiliserons pour notre étude, à savoir des données radar et des données plan de vol. Ceci nous permettra prendre en compte la complexité du trafic aérien au voisinage de chaque site sensible. Intuitivement, pour chaque site sensible, l'exposition au risque de chute d'un avion dépend de la densité de ce trafic au voisinage du site et de sa complexité (un site situé au voisinage d'un aéroport est plus exposé au risque de chute d'un aéronef qu'un site situé en dehors de tout trafic).

Dans une troisième partie, nous proposerons des modèles probabilistes pour évaluer le risque de chute d'un aéronef en vol commercial sur un site sensible, sachant l'événement redouté « perte de contrôle ». Ces modèles devront prendre en compte le taux d'accident d'un avion sur sa trajectoire, la modélisation probabiliste du point de chute, et la complexité des trajectoires au voisinage du site considéré (soit par une analyse détaillée au cas par cas, soit par une modèle statistique). Le risque lié à l'aviation générale n'est pas étudié dans ce rapport. La difficulté inhérente à ce type d'aviation réside dans l'identification des événements redoutés et dans l'évaluation des probabilités associées à partir des bases de données disponibles (ENAC, BEA, etc.).

Enfin en annexe, nous ferons quelques rappels en probabilité et introduirons les processus ponctuels de droites dans le plan. Ces processus nous serviront à décrire de façon statistique les trajectoires des aéronefs au voisinage d'un site sensible.

2 Démarche méthodologique

Définir une méthodologie d'analyse du risque est le point de départ de toute étude de sécurité. Elle définit les concepts et conditionne l'enchaînement des modèles proposés. Comme explicité dans l'introduction, la méthodologie mise en œuvre dans cette étude s'appuie sur la circulaire CE2096. Cette circulaire précise en particulier les notions de risque et de danger (ou événement redouté).

2.1 Approche adoptée conformément à la CE2096

Le terme de risque est fréquemment utilisé dans le langage courant avec différentes interprétations : un accident ou danger, une probabilité, une fréquence. Une succession d'événements est normalement qualifiée comme un risque si elle implique potentiellement un accident. Le degré du risque dépendra des conséquences possibles de cette succession, soit en termes de pertes humaines soit en termes économiques. Par conséquent pour appréhender un risque, il convient d'étudier sa fréquence d'apparition et sa sévérité. Ainsi, la circulaire CE2096 définit le risque comme étant « la combinaison de la probabilité la plus élevée ou la fréquence d'un événement aux conséquences dommageables provoqués par un danger et de la gravité de ces conséquences ».

Pour la navigation aérienne, la traduction probabiliste du risque est une fréquence maximale d'accident par heure de vol à ne pas dépasser. Pour notre étude, l'accident sera la chute d'un aéronef sur un site sensible, il convient alors de définir une métrique pour caractériser ce risque.

La représentation du risque fait appel à un processus stochastique d'occurrences de l'accident. Ce processus est considéré markovien, i.e. la probabilité d'apparition d'un accident ne dépend que de l'état présent du système. En d'autres mots, la connaissance du passé ne modifie pas la probabilité d'intervention d'un accident conditionnée à l'état présent. Par exemple, le fait qu'un accident ne soit pas arrivé depuis longtemps ne signifie pas qu'un accident est imminent. Cette propriété traduit bien le caractère imprévisible d'un accident qui peut survenir subitement, sans que l'on ne le voit venir.

Cette hypothèse conduit par exemple à supposer que le temps entre deux événements successifs est décrit par une variable aléatoire d'intensité un taux d'occurrence par heure.

L'estimation d'un risque repose sur la définition, la modélisation et l'identification des dangers. Un danger est un état ayant une cause, un état qui le précède et un effet, qui est un état le succédant (cf. figure 1). Chaque état peut apparaître selon une probabilité conditionnée aux états en amont dans la trajectoire. On voit ainsi que l'on peut définir le risque comme étant une trajectoire, i.e. une suite d'états (d'événements). La base de l'objectif de sécurité est donc l'identification des événements redoutés, qui correspondent à des événements directement préjudiciables quelque soit le contexte (par exemple une panne ou un endormissement lorsque l'on se trouve en voiture). Ces événements redoutés tamisent le risque en étant les événements que l'architecture doit permettre d'éviter.

Lors d'une étude de risque, le danger est l'événement *central* de la trajectoire. Il est précédé par des événements précurseurs (des causes), qui sont des événements non dangereux par nature, mais qui participent à la chaîne qui a des conséquences malheureuses. Ce principe est illustré dans la figure 2. Afin d'être exhaustif, il faut également anticiper les conséquences possibles de l'événement redouté. Ainsi, un endormissement au volant peut provoquer une simple sortie de route ou une collision mortelle. Finalement, une fois les dangers identifiés, on modélise les trajectoires dans lesquelles ils interviennent par une modélisation en *nœud papillon*.

Rappelons que dans notre étude le terme de danger ou d'événement redouté est relatif au risque ou accident défini comme étant la « chute d'un aéronef sur une installation sensible ».

2.2 Diagramme en nœud papillon d'un événement redouté

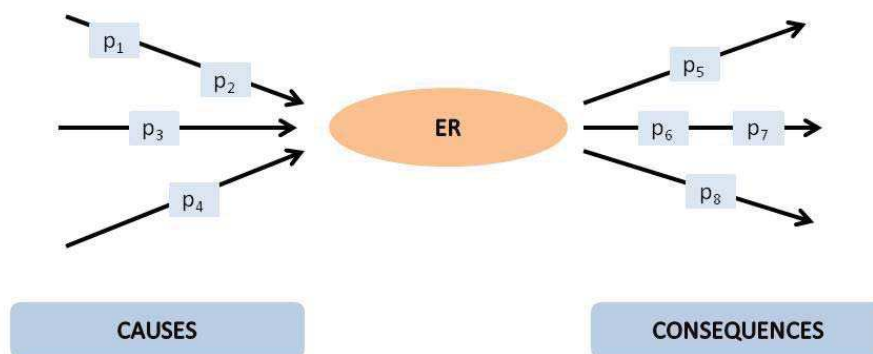


FIGURE 2 – Nœud papillon : chaque rectangle est un état avec une certaine probabilité P_i de se produire conditionnellement à l'état en amont.

La méthode du nœud papillon a pour but de visualiser concrètement des scénarios pouvant aboutir à un accident. Il permet la modélisation des trajectoires d'événements qui constituent le risque. Une

modélisation en nœud papillon est centrée sur un événement redouté qui est l'événement que l'on veut éviter et qui peut avoir des conséquences indésirables. Cette méthode permet en fait de modéliser individuellement chaque événement redouté, tel que défini dans la section précédente. Comme le montre la figure 2, cet état se trouve au centre du nœud papillon. A gauche du nœud se trouvent les événements précurseurs (ou causes) et à droite se trouvent les conséquences (ou effets). Il s'agit donc de modéliser les enchaînements d'événements pouvant amener à cet événement redouté, puis les conséquences possibles pouvant découler de cet événement central.

Cette méthode illustre bien la principale difficulté de l'estimation de risques : déterminer les événements redoutés. En d'autres termes, pour un scénario de risque donné, quel est l'état déclencheur de la catastrophe ? En effet, si pour un scénario de risque donné l'on situe l'événement redouté trop en amont dans la trajectoire, les conséquences possibles sont bien trop nombreuses pour pouvoir être estimées et a contrario, s'il est placé trop en aval, il est extrêmement difficile de reconstituer les branches de l'arbre de défaillances. De ce fait, à partir de l'état considéré comme l'événement redouté, les effets doivent être suffisamment simples à trouver. Comme dit plus haut, les événements redoutés doivent tamiser le risque en mettant en valeur ses composantes principales.

Une difficulté supplémentaire réside dans la présence possible et probable de causes communes, i.e. des causes pouvant altérer plusieurs états en même temps. Ainsi, une cause commune peut déclencher plusieurs branches à la fois. Nous pouvons citer comme exemple le séisme du Japon qui, en plus d'affaiblir la centrale de Fukushima, a contribué à déclencher le tsunami qui lui fut fatal. Ces causes communes ont le mérite de mettre en évidence deux limitations majeures de la modélisation en nœud papillon : (1) les branches ne sont pas forcément indépendantes les unes des autres. En effet, l'altération d'un état d'une branche peut entraîner ou aller de pair avec l'altération d'un état d'une autre branche. (2) les branches ne sont pas forcément disjointes, c'est-à-dire qu'un événement peut participer à plusieurs branches.

C'est pourquoi, en pratique, la démarche d'identification des événements redoutés doit être complétée par une analyse des modes communs de défaillance. Différentes approches sont possibles pour mener cette analyse, que nous ne détaillerons pas ici car cela sort du périmètre de notre étude.

2.3 Application de la méthodologie

L'évaluation des scénarios pouvant conduire à la chute d'un aéronef sur un site sensible est bien évidemment liée au type d'aviation considérée :

- l'aviation commerciale ;
- l'aviation militaire ;
- l'aviation générale.

En raison du manque de données et de connaissances concernant le fonctionnement de l'aviation militaire, ce type d'aviation ne sera pas développé ici. L'aviation commerciale est caractérisée par l'usage des règles de vol aux instruments (IFR), alors que l'aviation générale (ou de loisir) par celle des règles de vol à vue (VFR). Ces spécificités fondamentales impliquent une méthodologie différente entre les deux types d'aviation civile.

Nous allons commencer par identifier, pour chacun des deux cas nous intéressant, le niveau auquel nous situons les événements redoutés.

2.3.1 Le cas de l'aviation commerciale : un seul ER

En aviation commerciale, nous pouvons nous ramener à l'identification d'un seul ER : la perte de contrôle d'un avion qui conduit celui-ci à s'écraser au sol, suite à une trajectoire difficilement prévisible. Cet ER est suffisant car il est exhaustif (tous les scénarios de risque passent par lui). En effet, les aéronefs commerciaux ayant accès à tous les indicateurs nécessaires, une chute passe obligatoirement

par une perte de contrôle de l'avion. De plus, l'effet de cet ER ainsi que sa fréquence d'occurrence sont calculables, comme nous allons le montrer par la suite.

Dans le cas de l'aviation commerciale, nous n'avons donc qu'un nœud papillon à modéliser dont le centre est la perte de contrôle de l'aéronef. L'ER étant unique, la modélisation de la cause n'est pas nécessaire pour calculer la fréquence d'occurrence de cet ER. En effet, cette fréquence peut-être calculée à partir des comptes-rendus d'accidents réalisés par divers organismes (par ex. BEA), en ne retenant que les accidents dus à une perte de contrôle de l'avion. Quand à l'effet, il s'agit de la chute d'un aéronef sur un site sensible. En d'autres termes, il va s'agir de modéliser la cinématique de chute afin de déterminer la probabilité qu'un aéronef en train de chuter, termine sa trajectoire sur le site en question.

Afin d'évaluer le risque qu'un avion chute sur une installation, il faut d'abord évaluer le risque que celui-ci perde le contrôle à proximité de l'installation. Ceci nécessite donc la modélisation des trajectoires d'avion afin d'y associer une probabilité de chute. Pour ceci, nous considérons que sa route suit une ligne droite entre l'aéroport de départ et l'aéroport d'arrivée, avec une déviation latérale calculée par rapport à la route théorique. Cette déviation sera considérée aléatoire, ce qui a pour effet de rendre le processus stochastique. L'accident aérien sera ensuite modélisé comme un processus ponctuel suivant la trajectoire de l'avion auquel on associe une fréquence de chute. Les statistiques montrant que les accidents surviennent principalement en phase d'approche, la fréquence associée sera différente entre les phases d'approche et de vol en-route. Au sein de la portion en-route, le processus sera par contre considéré comme homogène.

2.3.2 Le cas de l'aviation générale : plusieurs ERs

Dans l'aviation générale, à la différence de l'aviation commerciale, les aéronefs ne suivent pas de trajectoires prédéfinies. Leur trajet est donc beaucoup plus aléatoire, et la prévision de trajectoire devient plus incertaine. Par conséquent, l'approche utilisée pour l'aviation commerciale et décrite précédemment n'est plus adaptée. En effet, l'évaluation du risque lié à l'aviation commerciale s'obtient par modélisation de la chute d'un aéronef le long de sa trajectoire nominale ; il suffit donc de considérer un échantillon de trafic nominal. Or considérer seulement les trajectoires « nominales » pour l'aviation générale présente l'inconvénient de masquer des scénarios pertinents d'un point de vue opérationnel, par exemple une entrée involontaire dans une zone interdite, une perte des conditions de vol à vue (VMC). De plus, l'aviation générale se basant beaucoup plus sur les décisions du pilote, une chute peut résulter d'autres facteurs qu'une perte de contrôle de l'aéronef (p. ex. une mauvaise appréciation de distance, réaction trop tardive devant un événement imprévu...).

Ainsi, il est nécessaire d'identifier toutes les causes directes de chute ou de collision avec une installation (endormissement, panne moteur...). Ceci nécessite d'évaluer le risque de plusieurs événements redoutés, donc de modéliser plus de nœuds papillons. Dans chaque cas, il s'agira de déterminer ce qui peut amener l'événement redouté à se produire (p.ex. contrôles peu fréquents pour la panne). Les effets sont également modélisés afin de connaître les effets probables de chaque événement redouté (p. ex. l'avion ira tout droit si l'ER produit est un endormissement, une spirale si c'est une panne moteur...).

La modélisation du risque de chute sur un site sensible pour l'aviation générale est donc plus complexe et nécessite l'introduction de plusieurs ERs. L'analyse des rapports d'accidents récoltés au BEA s'avère indispensable dans ce contexte.

2.4 La métrique utilisée

Dans l'aviation civile, la métrique utilisée pour le risque est le nombre moyen d'accident par heure de vol. Le choix de cette métrique résulte d'une gestion du risque du type régulateur/opérateur, le régulateur spécifiant un niveau de risque acceptable et l'opérateur apportant la preuve que le niveau de risque réel est bien inférieur à ce niveau de risque.

Dans le cas du risque de chute d'un aéronef sur un site sensible en particulier, le risque prend la forme d'un nombre de chutes par année et par mètre carré (*chutes/an/m²*). L'étude que nous cherchons à mener à bien n'est pas nouvelle dans la mesure où elle a déjà été réalisée dans le cas des sites nucléaires, en se basant sur les statistiques d'accidents passés. Le risque λ_a est alors évalué par le produit

$$\lambda_a = M_a \times X_{ar} \times \frac{1}{S_F}$$

où M_a représente le nombre annuel de vols selon la catégorie considérée (aviation générale, commerciale ou militaire), X_{ar} est le taux de chute régional pour la zone considérée (en route, approche, aéroport) et S_F est la superficie de la France.

Précisons que la référence temporelle utilisée est l'année, puisque nous supposerons que le site considéré est soumis au risque aérien de façon continue. Ainsi, le cadre de référence n'est pas le même que celui adopté dans l'aviation civile, puisque dans ce cas le risque n'est présent que pendant la durée d'un vol. Il faudra donc passer d'une référence en « nombre d'accidents par heure de vol » en « nombre de chute par heure ». Le taux λ_a défini ci-dessus permet d'obtenir la probabilité P qu'en N années il n'y ait pas de chute observée sur un site de superficie S par la formule

$$P = \exp(-\lambda_a \times N \times S).$$

Pour terminer cette partie, notons que la méthodologie que nous proposons dans cette étude utilisera des données sur les trajectoires des aéronefs au voisinage du site considéré, ceci afin de prendre en compte l'exposition au risque aérien du site en fonction de la complexité du trafic qui le survole.

3 Description des données statistiques disponibles

La possession de multitudes de données aériennes pour la statistique est un apport très important de l'ENAC. En effet, depuis l'avènement de l'informatique, les données concernant tous les vols se produisant partout dans le monde ont commencé à pouvoir s'accumuler sans problèmes. Ainsi, pour l'application pratique de la méthodologie, trois types de données devront être utilisés. Les données d'accidents récoltées notamment par le BEA seront d'abord présentées, puis la description de chaque vol par le biais des plots radar et des plans de vol.

3.1 Données d'accidents et de trafic

Comme explicité dans la section a, les données d'accident sont indispensables pour l'étude de risque. En effet, en recensant aussi bien les crashes que les collisions bénignes lors de tours de piste, ces données sont la clé permettant d'estimer la fréquence d'occurrence de tous les événements redoutés, que ce soit pour l'aviation générale ou commerciale. Ces recensements d'accidents sont notamment effectués par le BEA, qui publie régulièrement des rapports d'enquêtes sur les accidents survenus grâce à une déclaration des acteurs ou observateurs de l'accident (pilotes, contrôleurs, etc.). Ces données d'accident sont cependant très difficiles à exploiter à cause de la forme qu'elles prennent. En effet, comme nous venons de le dire, les résultats d'enquête sont sous forme de rapport, ce qui ne constitue pas des données exploitables directement par les statisticiens. Le calcul de l'estimation des fréquences d'occurrence des accidents, nécessite donc la transformation des données sous une forme plus exploitable (base de données oracle par exemple), ce qui n'est pas le sujet du travail ici. Par conséquent, lors de la description de la méthodologie, nous considérerons toujours que nous disposons d'une valeur d'intensité, même si celle-ci reste à calculer en pratique.

3.2 Données radar

Plus exploitables sont les données décrivant chaque vol, et notamment les données radar (aussi appelées « plots » radar). En effet, en vol, chaque avion dispose d'une signature (appelée « transpondeur ») que des radars disposés en différents endroits en France permettent de suivre donc de connaître leur position réelle. La position est relevée toutes les quatre secondes permettant ainsi le suivi d'une trajectoire de manière très précise. Il existe également des données radar sous un format agrégé, i.e. on garde une trace toutes les 4 minutes. Ce format est appelé format IMAGE et est le plus accessible lorsqu'il s'agit de récupérer des données. En termes plus pratiques, un relevé de position correspond à une donnée contenant un certains nombre de variable telles que l'identifiant du vol, sa position (en nautiques par rapport à une origine placée proche de Poitiers), sa vitesse (en nœuds) ou encore son altitude (en pieds). Ces données seront celles qui nous seront le plus utile, comme nous le verrons par la suite.

3.3 Données plans de vol

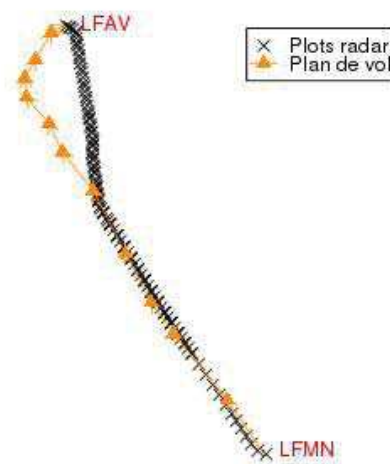


FIGURE 3 – Plots radar et balises constituant le plan de vol, d'un trajet allant de Nice à Valenciennes. On voit que l'avion ne suit pas forcément exactement sa trajectoire prévue.

Pour finir, allant de pair avec les données radar, sont disponible les données plan de vol. Ces données correspondent à une route « théorique » que doit emprunter l'avion et qui est déposée avant le vol. Un plan de vol correspond simplement en une succession de balises fixes au-dessus desquelles l'avion doit passer durant son vol. La route tracée par ses balises peut cependant être assez complexe et faire réaliser à l'avion des petits détours indésirables. Ainsi, en réalité l'avion ne passera pas nécessairement par toutes les balises et privilégiera une trajectoire différente selon la météo ou les conflits potentiels. Ceci est illustré par la figure 3 : plots radar et balises constituant le plan de vol, d'un trajet allant de Nice à Valenciennes. On voit que l'avion ne suit pas forcément exactement sa trajectoire prévue. qui

montre le plan de vol (flèches orange) ainsi que les plots radar (croix noires) d'un vol entre Nice et Valencienne. Enfin, notez que le dépôt d'un plan de vol n'est obligatoire que pour les vols commerciaux. Ces données sont donc inexistantes pour l'aviation générale.

4 Mise en œuvre pratique de l'approche pour l'aviation commerciale

Nous ne considérerons dans la suite que le cas de l'aviation commerciale. L'estimation du risque lié à l'aviation de loisir est plus complexe que pour l'aviation commerciale. En effet, les causes de chutes se révèlent bien trop nombreuses dans ce cas et un trop grand nombre de facteurs peut intervenir. De plus, avec le nombre de causes possibles, l'accident peut avoir des conséquences très variables et la chute de l'avion paraît plus difficile à modéliser.

Grâce aux compte rendus d'accident ou à l'expérience de pilotes et contrôleurs, on peut évaluer la fréquence d'occurrence de chacun des ER. Cette fréquence sera pondérée par la densité d'aérodromes situés à proximité (les vols loisirs sont souvent des vols locaux). De plus, comme expliqué dans la section 2.3.2, on modélise la trajectoire des aviatisions volant à vue par un couple position/vitesse. En disposant des traces radar et en observant le comportement des aéronefs aux abords du site, on peut imaginer les conséquences de chacun des ER.

Pour l'aviation commerciale, le fait de n'avoir qu'un ER permet de simplifier l'évaluation du risque pour les vols commerciaux. En effet, il s'agit de modéliser uniquement le nœud papillon associé à cet ER, tel qu'expliqué dans la section 2.2. Ainsi, en pratique, l'estimation du risque se décompose en trois parties. Deux parties concernent la partie droite du nœud papillon, i.e. les causes, à savoir l'évaluation du risque selon la phase de vol (approche/en-route), et la modélisation des trajectoires d'avion qui nous donne la proportion d'avions survolant le site considéré. Enfin, les effets sont étudiés par la modélisation de la cinématique de chute, qui donne une aire probable de chute. Il est à noter que les méthodes présentées dans cette section ne concernent que les avions commerciaux en phase de croisière ou en début de phase d'approche (cette notion est définie ci-après, dans la section 4.1). En effet, les deux méthodes se basent sur une modélisation rectiligne des trajectoires, les rendant ainsi inaptes à modéliser les aéronefs VFR de l'aviation générale, où les avions IFR qui peuvent se trouver en hippodrome par exemple.

4.1 Uniformité du processus

Nous avons affirmé plus haut que la probabilité de chute est différente selon la phase de vol. Dans le cas IFR, nous pouvons identifier plusieurs phases de vol :

- Roulage : l'avion roule afin de se placer sur la piste de décollage ;
- Décollage / Montée initiale : l'avion quitte la piste et débute sa phase ascendante tout en se plaçant dans la direction qui lui convient ;
- Montée : les becs/volets sont rentrés et l'avion monte jusqu'à son niveau de croisière ;
- Croisière (en-route) : l'avion vole à niveau constant (c'est généralement la phase la plus longue d'un vol) ;
- Descente : l'avion commence à descendre en s'approchant de l'aéroport de destination. A la fin de cette phase, si le contrôle le demande, l'avion peut être placé en hippodrome ;
- Approche finale / atterrissage : l'avion se met en position puis atterrit.

En sus de ces phases, nous définissons comme phase d'approche, les phases de montée et de descente, que nous considérerons de la même manière que la phase de croisière (elles peuvent durer longtemps), mais pour lesquelles nous utiliserons un taux de chute différent. En effet, comme le montre la figure 4, les statistiques indiquent que le nombre d'accidents est bien plus élevé lors des phases d'approche que de la phase de croisière. Ceci peut s'expliquer par le caractère évolutif du vol, ce qui implique une charge de travail plus importante que lors du vol croisière. Le risque d'erreur du pilote est donc plus

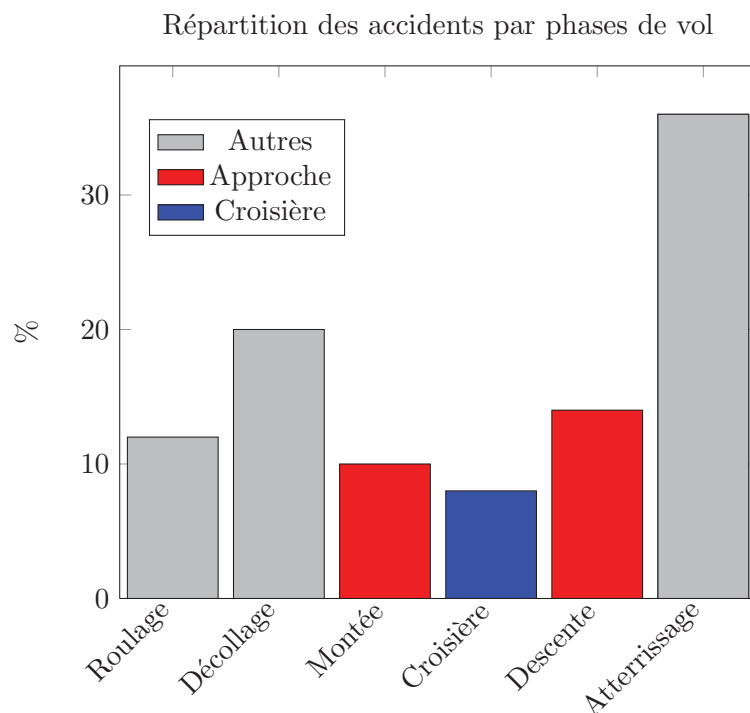


FIGURE 4 – Répartition des accidents par phase de vol (cf. [1]).

important. Ainsi, dans les méthodes présentées plus loin, le taux de chutes par heure de vol nécessaire au calcul de risque sera différent pour les phases d’approche ou pour le vol en-route, le risque associé à ces derniers étant plus important (cf. figure 4). Notez que par la suite, les autres phases que approche et croisière ne seront pas prises en compte car, d’une part ce sont des phases très spécifiques pendant lesquels l’avion ne suit plus une trajectoire rectiligne, et d’autre part, il n’est pas censé y avoir de bâtiments dans les zones proches des aéroports.

Enfin, dans le cas des avions commerciaux, l’altitude peut être considérée comme un critère pour définir la différence entre approche et croisière. En effet, les avions voyagent en croisière à une altitude supérieure au niveau 200 (20000 pieds) ; ils ne sont donc en dessous de ce plancher que lorsqu’ils sont en phase d’approche. De ce fait, il n’y aura pas de discrimination particulière à l’altitude dans les calculs de risque car cette différence est déjà prise en compte dans le découpage approche/croisière. Ainsi, les avions volant plus haut pouvant potentiellement s’écraser loin du point de départ de la chute, si l’on considère que les avions volant près du plancher ont le même champs de chute qu’un avion volant très haut, un plus grand nombre d’avions que nécessaire sera pris en compte dans le calcul du risque. Ceci aura pour effet de légèrement surévaluer le risque, ce qui est une démarche classique de l’estimation du risque en cas d’incertitude.

4.2 Cinématique de la trajectoire de chute

Une fois que l’état « l’aéronef commence à chuter » s’est déclenché, il faut ensuite modéliser la partie droite du nœud papillon, i.e. les conséquences de la chute. En l’occurrence, nous voulons connaître le risque qu’a l’avion, à partir du moment où il commence sa chute, de tomber dans une zone définie. Pour ceci, il est nécessaire de connaître la zone de chute probable de l’avion, connaissant sa position lors de la perte de contrôle. Cette zone peut-être calculée grâce à l’outil des Minimum Volume Set (MVS) qui seront présentés ci-après.

A partir de cette zone, nous pouvons définir une zone de risque potentiel autour d'un site. Lorsqu'un avion pénètre dans cette zone, il existe un risque que celui-ci chute sur le site concerné. Il est démontré dans la deuxième sous section que cette zone correspond à une extension du site par un cercle.

4.2.1 Zone de chute probable : les Minimum Volume Set

Soit un avion se déplaçant selon l'axe des abscisses et amorçant sa chute à l'origine, nous définissons sa zone de chute probable \mathcal{K}_α ($0 < \alpha < 1$) comme

$$\mathbb{P}(\text{chute} \in \mathcal{K}_\alpha) = 1 - \alpha.$$

Nous supposons que \mathcal{K}_α est un convexe ; il sera toujours possible de le choisir de la sorte. Comme expliqué dans la section 4.1, \mathcal{K}_α ne dépendra pas de l'altitude et sera calculé globalement. Intuitivement, on peut imaginer \mathcal{K}_α comme un « pinceau » qui se promène sur la route de l'avion.

Pour un avion amorçant sa chute en un point $M \in \mathbb{R}^2$ et suivant une trajectoire faisant un angle θ avec l'axe des abscisses, la zone de chute associée sera $\mathcal{R}_\theta \mathcal{K}_\alpha \oplus M$ où \mathcal{R}_θ désigne la rotation d'angle θ de centre l'origine et,

$$\mathcal{R}_\theta \mathcal{K}_\alpha \oplus M = \{M + \varphi, \varphi \in \mathcal{R}_\theta \mathcal{K}_\alpha\}$$

est le translaté de $\mathcal{R}_\theta \mathcal{K}_\alpha$ par le vecteur \overrightarrow{OM} (cf. figure 5).

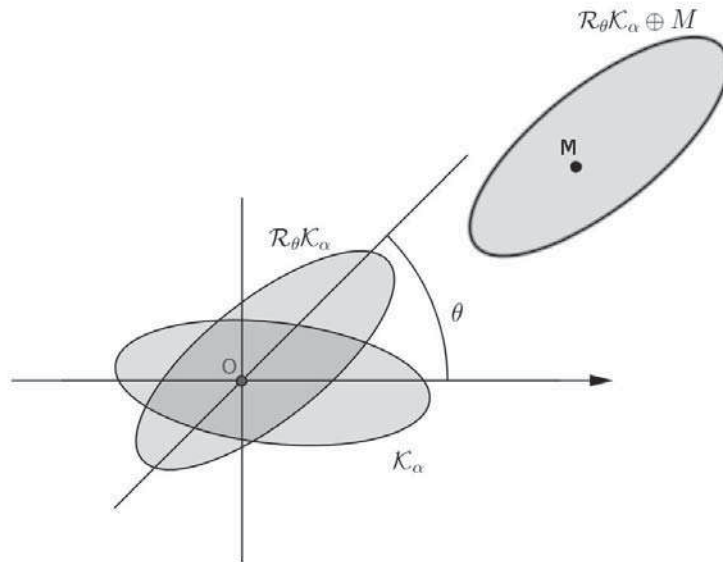


FIGURE 5 – Illustration du convexe $\mathcal{R}_\theta \mathcal{K}_\alpha \oplus M$. La zone de chute est translatée jusqu'au point M et une rotation d'angle θ est effectuée.

Pour donner un exemple de convexe \mathcal{K}_α , nous pouvons citer une étude de l'Interstate Aviation Committee (communication privée). Cette étude réalisée sur 5 cas répertoriés situe les points de chute à l'intérieur d'un cercle d'un rayon de 12NM, centré sur la position du début de la trajectoire de chute. En fait, il y a 4 points d'impact dans un cône orienté dans la direction initiale du vol et un seul cas orienté à 180° . Clairement, le choix d'un cercle est très conservateur. Dans un premier temps, il faudrait

enrichir ces données en explorant les bases de données d'accident. Ensuite, le cercle ne semblant pas être la forme géométrique la plus adaptée, nous aimerions définir une région de confiance plus précise. Pour ceci, nous pouvons utiliser l'outil des « Minimum Volume Set » (MVS). Si l'on considère une densité de probabilité, un MVS de degré α (compris entre 0 et 1) est défini comme étant le plus petit volume (selon une mesure quelconque), représentant une probabilité supérieure à α . Intuitivement, il s'agit du lieu dans lequel la probabilité est la plus « concentrée ». Par exemple, la figure 6 montre un MVS d'ordre $\alpha = 0,7$ d'un mélange gaussien. La partie en magenta représente le plus petit volume possible dans lequel on puisse trouver 70% des observations. L'estimation d'un MVS peut-être calculée, par exemple, sur une estimation de la densité de probabilité qu'est l'histogramme [5].

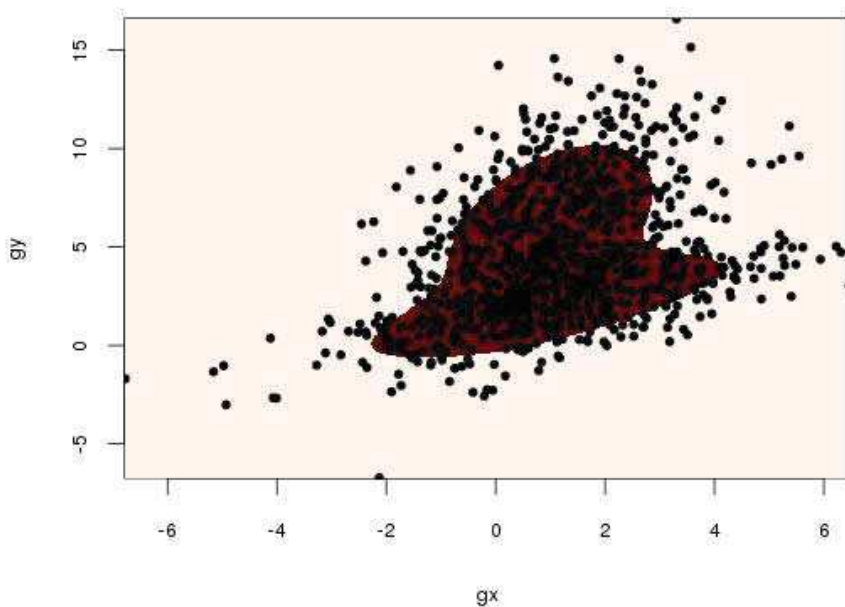


FIGURE 6 – MVS à $\alpha = 0.7$ d'un mélange gaussien. 500 observations tirées de la densité considérée sont également reportées.

En pratique, il faut bien évidemment disposer de données. Et pour obtenir un résultat suffisamment précis, il est important de disposer d'un plus grand nombre d'exemples d'accidents que 5. Cette estimation doit donc s'accompagner de démarche auprès du BEA pour obtenir ce qui nous est nécessaire. Signalons que l'accident de l'AF447 (Rio-Paris) vérifie les conclusions de cette étude.

4.2.2 Zone de risque potentiel

La zone de risque potentiel \mathcal{C} associée à un site spécifique, matérialisé par une surface Δ , est définie comme l'ensemble des points M du plan tels que si un avion suit une trajectoire orientée d'angle θ et amorce une chute en M , alors l'intersection entre sa zone de chute $\mathcal{R}_\theta \mathcal{K}_\alpha \oplus M$ et le site Δ est non vide, soit :

$$\mathcal{C} = \{M \in \mathbb{R}^2 ; \exists \theta : \mathcal{R}_\theta \mathcal{K}_\alpha \oplus M \cap \Delta \neq \emptyset\}.$$

De manière analogue, nous pouvons dire que M appartient à \mathcal{C} si et seulement si il existe un θ tel que l'ensemble $\mathcal{R}_\theta \mathcal{K}_\alpha \oplus M$ intersecte le site Δ . Ainsi, il est aisé de voir que si au moins un θ vérifie

la condition qui fait que M appartient à \mathcal{C} , alors l'union sur tous les θ possibles vérifie cette propriété (cf. figure 7).

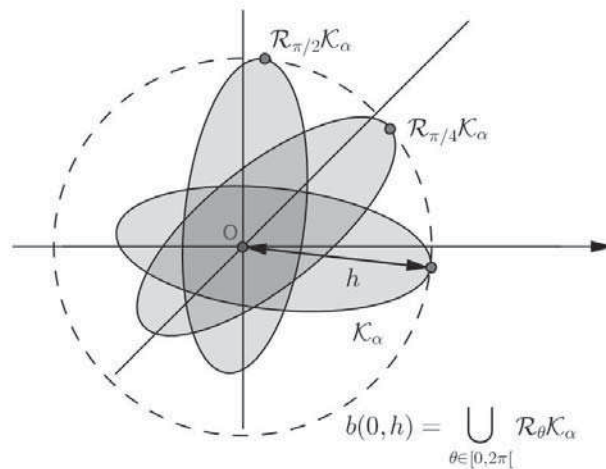


FIGURE 7 – L'union de toutes les rotations de \mathcal{K}_α forment un cercle.

Nous pouvons donc reformuler la zone de risque potentiel de la façon suivante :

$$\mathcal{C} = \{M \in \mathbb{R}^2 : (\bigcup_{\theta \in [0, 2\pi[} \mathcal{R}_\theta \mathcal{K}_\alpha) \oplus M \cap \Delta \neq \emptyset\}.$$

L'ensemble

$$b(O, h) = \bigcup_{\theta \in [0, 2\pi[} \mathcal{R}_\theta \mathcal{K}_\alpha$$

n'est autre que la boule centrée en l'origine O et de rayon $h = \max\{\|k\| : k \in \mathcal{K}_\alpha\}$. Il est alors aisé de voir que nous pouvons exprimer \mathcal{C} comme la somme (de Minkowski)

$$\mathcal{C} = \Delta \oplus b(O, h).$$

De manière imagée, nous pouvons voir la zone \mathcal{C} comme obtenue par élargissement de la frontière du site d'une longueur h .

Cette zone de risque potentiel est utilisée dans les deux méthodologies présentées ci-après. En effet, ces deux méthodologies ont pour but de calculer le nombre de chutes survenant dans ce volume, mais en adoptant deux approches différentes. Il est à savoir que la grande difficulté de l'estimation du risque est due au fait que l'on travaille sur des probabilité infinitésimales, qui nécessiteraient donc un grand nombre de données pour être estimées. Cette masse de données étant impossible à traiter il est donc important de contourner ce problème.

Ainsi, la première approche est basée sur une modélisation statistique des trajectoires d'avion par rapport :

- à un segment de référence qui lie l’aéroport de départ à l’aéroport d’arrivée ;
- à une ligne brisée qui lie les balises constituant le plan de vol.

Ceci permet d’inférer la densité du trafic au dessus de la zone \mathcal{C} , et donc de confronter cette densité à un risque de chute plus global. Dans cette méthode, une loi agrégée des écarts à la trajectoire « théorique » de référence est estimée, afin de représenter chaque trajectoire par une ligne droite (brisée) « perturbée » par cette loi de déviation. Cette méthode permet ainsi de prendre en compte le trafic dans sa globalité ainsi que les cas extrêmes possibles, mais il peut lui être reproché d’utiliser une représentation trop « simpliste » de la complexité du trafic. Cette approche présente également l’avantage de n’avoir besoin que des données plan de vol pour être appliquée, i.e. des aéroports liés par chaque vol et des balises qui constituent les plans de vol.

La deuxième approche, en revanche, est basée sur les traces radar récupérées directement au voisinage d’un site, en extrapolant le risque induit par la présence d’un aéronef dans ce volume. Dans cette méthode, le trafic observé est considéré représentatif et chaque trajectoire observée est considérée selon sa longueur dans la zone de risque potentiel. Cette deuxième méthodologie a l’avantage d’être plus proche des données disponibles mais dépend de l’échantillon de trajectoires sélectionné pour l’estimation du risque. La représentativité de cette méthodologie est de plus très liée à la taille de la zone de risque potentiel \mathcal{C} . Le fait de mal choisir la taille d’échantillon peut ainsi biaiser l’estimation du risque.

5 Première approche : modélisation des trajectoires d’avions

5.1 Modélisation des déviations

La méthodologie présentée ici est basée sur une modélisation statistique des trajectoires suivies par les avions. Cette méthodologie vise à utiliser plutôt les données plan de vol des avions (données COURAGE, cf. section 3.3), données relativement facile à avoir. L’idée centrale de cette modélisation est de considérer qu’un avion suit une trajectoire prévue avec des déviations latérales aléatoires. Cette trajectoire prévue peut être de deux types :

- dans la majorité des cas la trajectoire ressemblera à une ligne brisée. En effet, dans le cas des avions commerciaux, un avion doit suivre un plan de vol qui est constitué de plusieurs balises que l’avion doit survoler. Entre deux balises, la trajectoire sera rectiligne. Un avantage non négligeable de ce découpage du vol est de pouvoir définir une loi des déviations pour les phases d’approche ;
- dans d’autres cas cependant, le contrôleur peut choisir d’assigner une route directe à l’avion. Sa trajectoire se rapprochera donc fortement d’une ligne droite entre l’aéroport de départ et l’aéroport d’arrivée. Ce cas se présente souvent pour les vols de nuit ou à des moments de faible trafic.

La figure 3 illustre bien cette dualité en montrant une trajectoire qui suit bien son plan de vol puis coupe les dernières balises pour aller directement à sa destination. Ainsi, cette méthodologie se base sur l’idée d’ajuster une loi de probabilité de densité f_Y à la déviation latérale aléatoire, pour chacun des deux cas de figure. Il sera ainsi possible de calculer la probabilité qu’un aéronef se trouve dans une aire donnée (qui sera l’aire de risque potentiel \mathcal{C} définie en section 4.2.2) au moment de sa chute.

Une déviation latérale peut être définie comme la longueur du segment MM' où M' représente le projeté orthogonal de M sur le segment qui lie les balises B_1 et B_2 , les deux balises qui entourent le point M . Ainsi, nous pouvons dire que la déviation est la coordonnée d’un aéronef sur un axe orthogonal au segment B_1B_2 et centré sur ce même segment. A noter que dans le cas d’une trajectoire directe, les points B_1 et B_2 sont les aéroports de départ et de destination.

Afin de modéliser une « déviation nominale », la loi gaussienne est souvent choisie. Ceci est dû au fait qu’une déviation est en fait une somme d’un grand nombre de petites déviations aléatoires, qui

dépendent de facteurs différents. Or le théorème central limite nous dit qu'une somme de variables aléatoires indépendantes tend vers une gaussienne, justifiant ce choix. Cependant des comportements non nominaux pouvant conduire à de grands écarts, il convient en analyse de risque de les modéliser par une loi de type extrême, à déterminer empiriquement. Ainsi, nous supposons qu'il existe un seuil x_{sep} séparant le comportement nominal du comportement non nominal, d'où l'expression suivante de la densité

$$f_Y(x) = \begin{cases} f_1(x) & |x| < x_{\text{sep}} \\ f_2(x) & |x| \geq x_{\text{sep}} \end{cases}$$

où f_1, f_2 sont les densités pour respectivement le comportement nominal (a priori une gaussienne) et le comportement non nominal (une loi extrême).

5.1.1 Estimation empirique de la loi des déviations

L'estimation empirique de f_Y utilise les données IMAGE de plots radar (cf. section 3). Ces données sont une agrégation des traces radar récoltées dans le sens où elles donnent la position d'un aéronef toutes les 3 ou 4 minutes, au lieu de toutes les 4 secondes pour les traces radar brutes. Ces données sont disponibles pour tous les avions survolant le territoire français, même si l'avion ne fait que le survoler le (pour un vol allant d'Italie en Grande-Bretagne par exemple), depuis le 1^{er} janvier 2011 à aujourd'hui. De plus, ces données présentent l'avantage de contenir l'aéroport de départ et l'aéroport de destination pour tous les vols enregistrés, même si ceux-ci ne sont pas français.

Pour estimer f_Y , un échantillon de plus de 5000 vols IFR sélectionnés entre le 1^{er} et le 5 mai 2012 a été tiré de la base de données IMAGE. La méthode de calcul des déviations latérales est simple. Pour un plot radar, sa déviation est la distance signée à son projeté orthogonal. Pour comprendre le calcul, il faut savoir qu'en aviation civile, on travaille avec le cap d'une trajectoire qui correspond à son angle en rotation horaire par rapport à l'axe nord-sud. Ainsi, un avion volant plein nord aura un cap de 0 et un avion volant plein est aura un cap de $\pi/2$. Cette subtilité a finalement pour simple effet d'inverser les sinus et cosinus dans les calculs. La déviation latérale peut donc se calculer la façon suivante :

$$\text{offset}_p = (x_p - x_b) \cos \theta - (y_p - y_b) \sin \theta$$

où (x_p, y_p) sont les coordonnées du plot, (x_b, y_b) les coordonnées de la balise de fin du tronçon (de l'aéroport de destination dans le cas des trajets directs) et θ est le cap du tronçon.

Pour chaque vol dont on dispose, on calcule les déviations par rapport au plan de vol ainsi que par rapport à la droite qui lie l'aéroport de départ et l'aéroport d'arrivée. On considérera l'une ou l'autre des déviations, selon celle qui minimise la déviation maximum (en valeur absolue). La figure 8 montre la proportion de chaque type de déviations et nous permet de constater que 9 avions sur 10 sont fidèles à leur plan de vol initial. Notez que, dans le cas où l'avion suit les balises de son plan de vol, les déviations calculées par rapport au premier tronçon (entre l'aéroport de départ et la première balise) et au dernier tronçon (entre la dernière balise et l'aéroport de destination) vont enrichir un échantillon annexe permettant d'estimer la loi des déviations en phase d'approche. Le cas des déviations par rapport à la ligne brisée du plan de vol ne comprend donc que les déviations en phase de croisière, i.e. étant à plus d'une balise des aéroports.

La figure 9 montre la répartition des heures de départ des vols pour chacun des deux types. Si la forme des histogrammes est globalement identique, avec les modes au même moment de la journée, on peut tout de même constater qu'il y a une plus grande proportion de vols nocturne parmi les avions qui réalisent un trajet direct. Ceci est dû au fait que la possibilité de couper des balises est plus souvent accordée lorsque le trafic est faible, ce qui est notamment le cas la nuit.

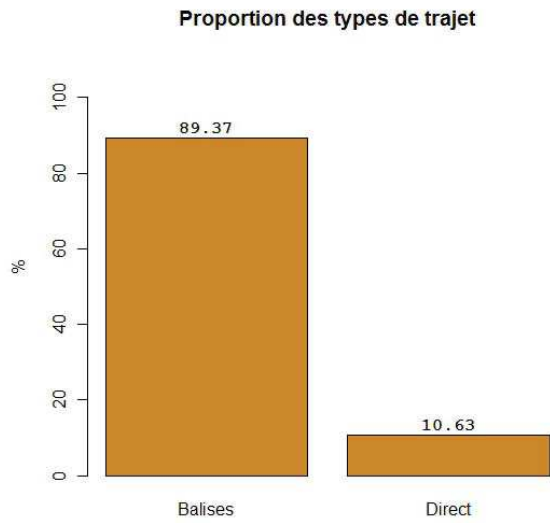


FIGURE 8 – Répartition des types de trajet sur l'ensemble des vols qui constituent le jeu de données.

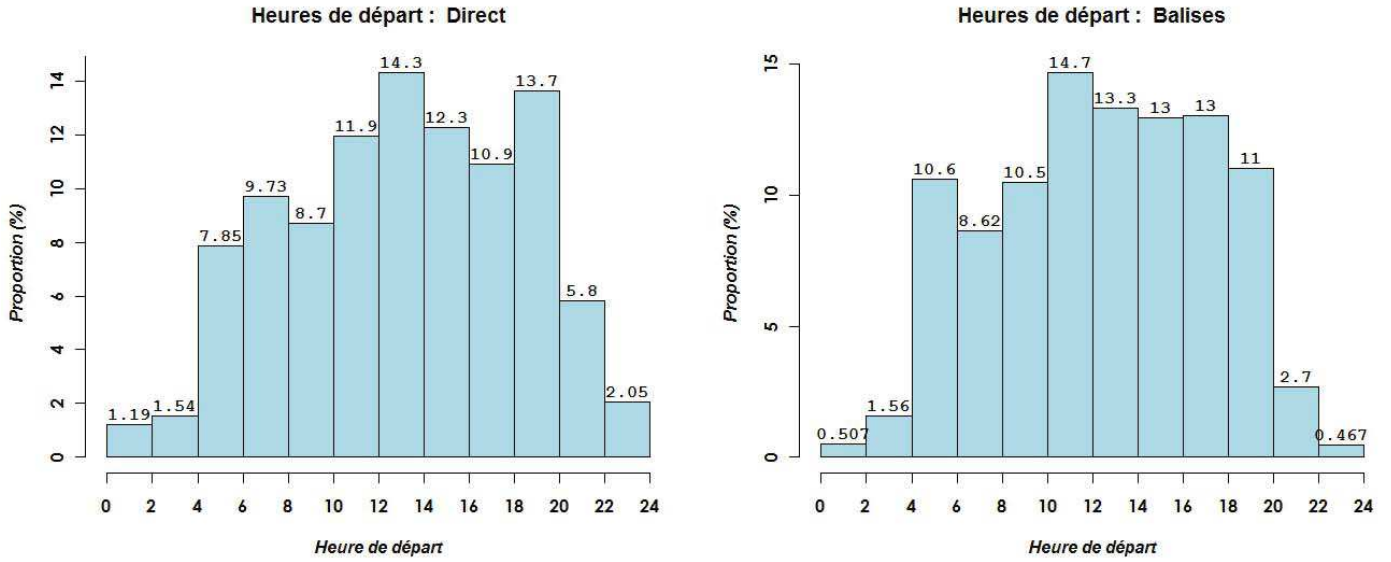


FIGURE 9 – Histogramme des heures de départ pour les deux types de vol. Les barres de l'historgramme ont des amplitudes de 2h.

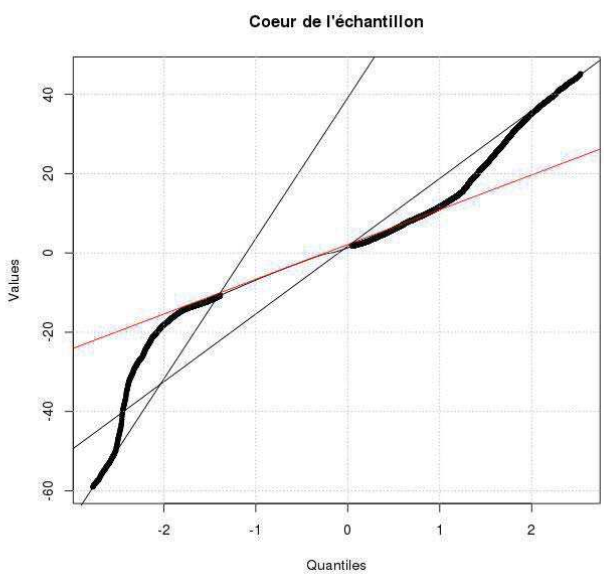
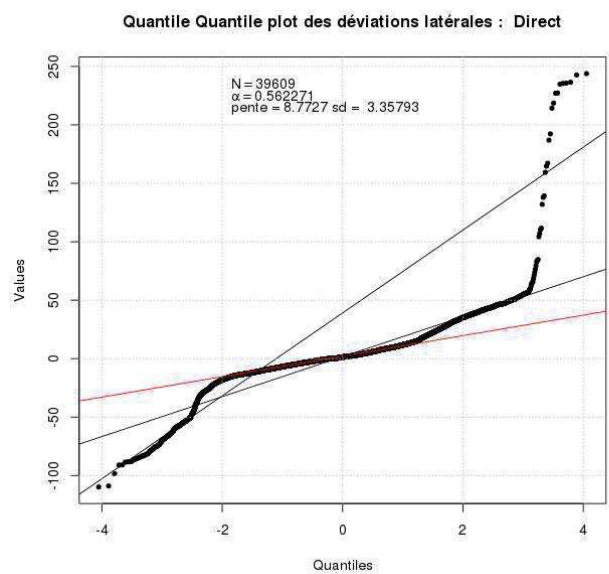
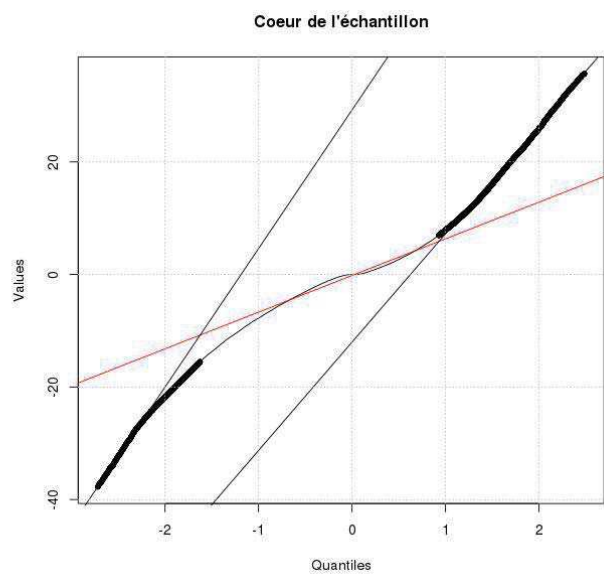
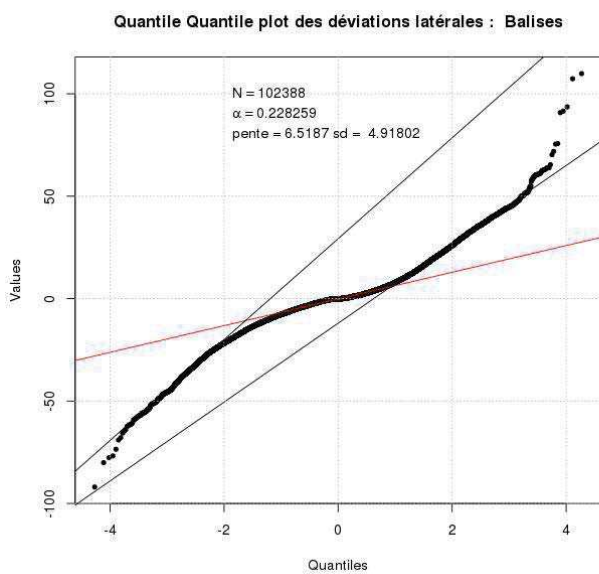
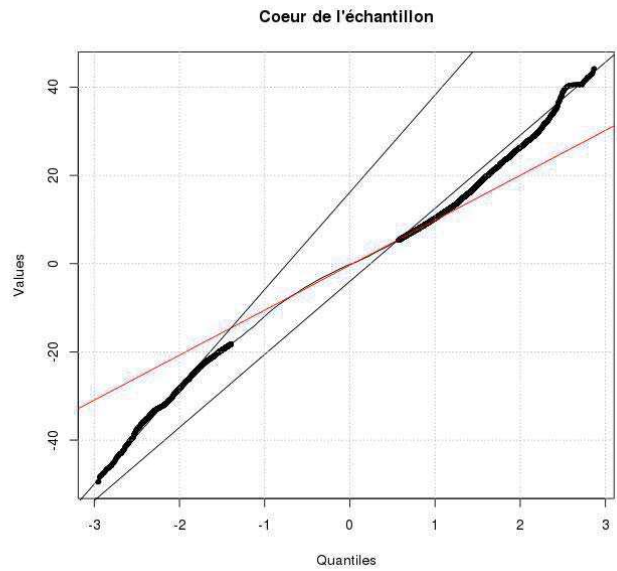
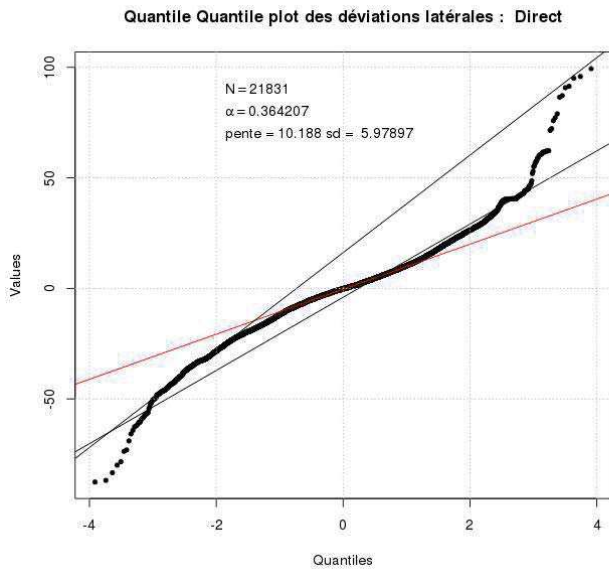


FIGURE 10 – Diagramme quantile-quantile par rapport à la loi gaussienne des déviations aux trajectoires directes (haut), plan de vol (milieu) et phase d'approche (bas). Le graphique de droite est un zoom sur le cœur de l'échantillon.

La première analyse que l'on peut faire des résultats est de vérifier la normalité des faibles déviations (que l'on appellera noyau de l'échantillon). Pour ceci, on peut tracer le diagramme quantile-quantile des déviations en fonction d'une loi normale centrée réduite (de moyenne nulle et de variance égale à 1). Ce diagramme est tracé en figure 10, respectivement pour les trajectoires directes (haut), pour les trajectoires qui suivent le plan de vol (milieu) et pour les phases d'approche (bas). En effet, si X est une variable de moyenne μ et de variance σ^2 , alors on sait que l'on peut l'écrire $X = \mu + \sigma Z$ où Z est justement une variable gaussienne centrée réduite. On s'attend donc en l'occurrence à ce que le cœur du diagramme suive une droite de coefficient directeur σ (μ sera nul car les déviations sont centrées).

L'analyse de la figures 10 indique que le noyau de l'échantillon ne s'ajuste pas bien à la loi gaussienne. En effet, nous voyons bien que, même au centre, la ligne ne suit pas bien la droite théorique. De plus, on peut voir que l'écart-type estimé sur le cœur de l'échantillon ne correspond pas à la pente de la droite théorique, ce qui devrait pourtant être le cas si les déviations étaient gaussiennes.

Ce phénomène est confirmé par le diagramme des probabilités de dépassement de chaque type de déviation, illustré en figure 11, respectivement pour les trajectoires directes (en haut à gauche), pour les trajectoires qui suivent le plan de vol (en haut à droite) et pour les phases d'approche (en bas). Ce diagramme trace la valeur $\mathbb{P}(|\text{déviations}| > x)$ en fonction de x , en échelle logarithmique (par le logarithme décimal plus précisément) afin de pouvoir également ajuster précisément les queues de distribution. Comme la distribution des déviations est censée être symétrique et centrée, cela revient à tracer $2(1 - F_Y(x))$ où $F_Y(x)$ est la fonction de répartition empirique des déviations. En effet, le diagramme montre un noyau d'échantillon (les points d'abscisse inférieurs à) rectiligne en échelle logarithmique. Or si celui-ci était gaussien, les points devraient suivre la ligne en pointillés bleus. Le noyau étant rectiligne en échelle logarithmique, cela signifie qu'il suit une loi exponentielle. En effet, si Y suit une loi exponentielle de paramètre λ , alors

$$\log_{10}(1 - F_Y(x)) = \log_{10}(e^{-\lambda x}) = -\frac{\lambda}{\ln 10}x,$$

où l'on voit qu'il s'agit bien d'une droite de coefficient directeur $a = -\lambda/\ln 10$. Ainsi, si l'on ajuste le noyau de l'échantillon par régression linéaire, on peut trouver le paramètre de la loi exponentielle $\lambda = -a \ln 10$.

Ces ajustements étant faits sur une probabilité en valeur absolue et comme nous sommes en réalité dans un cas où la variable est définie sur \mathbb{R} , on peut considérer que le noyau des déviations suit une loi de Laplace centrée. En effet, on sait que pour une variable aléatoire X qui suit une loi de Laplace de paramètres de position $\mu = 0$ et de paramètre d'échelle b , sa valeur absolue $|X|$ suit une loi exponentielle de paramètre $1/b$. On peut donc considérer que le noyau de nos déviations (les « petites déviations ») est distribué selon une loi de Laplace de paramètres μ et $b = 1/\lambda$ (cf. annexe 6.2.2). On peut confirmer ceci par l'analyse de la figure 12 pour les trajectoires directe (gauche), pour les trajectoires suivant le plan de vol (droite) et la figure 13 pour les phases d'approche. En effet, on constate que le noyau de l'échantillon suit plutôt bien la ligne rouge dans les trois cas et que la pente de cette ligne est proche de l'estimation du paramètre d'échelle b , ce qui est attendu dans le cas où l'échantillon est laplacien (b remplit le même rôle pour une Laplace que l'écart-type σ pour une gaussienne, voir l'annexe 6.2.2).

Nous venons donc d'établir que l'intuition a priori d'une déviation gaussienne était erronée et que la distribution considérée semble plutôt être une Laplace. Ceci s'explique par le fait que nous sommes finalement dans un cas où le théorème central limite ne s'applique plus car les déviations ne sont pas uniquement aléatoires et peuvent résulter de choix du contrôleur et du pilote. La loi de Laplace traduit plutôt le fait que lorsqu'un avion commence à dévier, sa probabilité de poursuivre sa déviation ne change pas, c'est-à-dire que le fait d'avoir déjà dévié de n nautiques n'a pas d'influence sur le fait de dévier de m nautiques supplémentaires. Cette intuition est sous-jacente dans le fait que le diagramme en échelle logarithmique soit rectiligne et la probabilité de continuer à dévier est représentée par la pente de la droite.

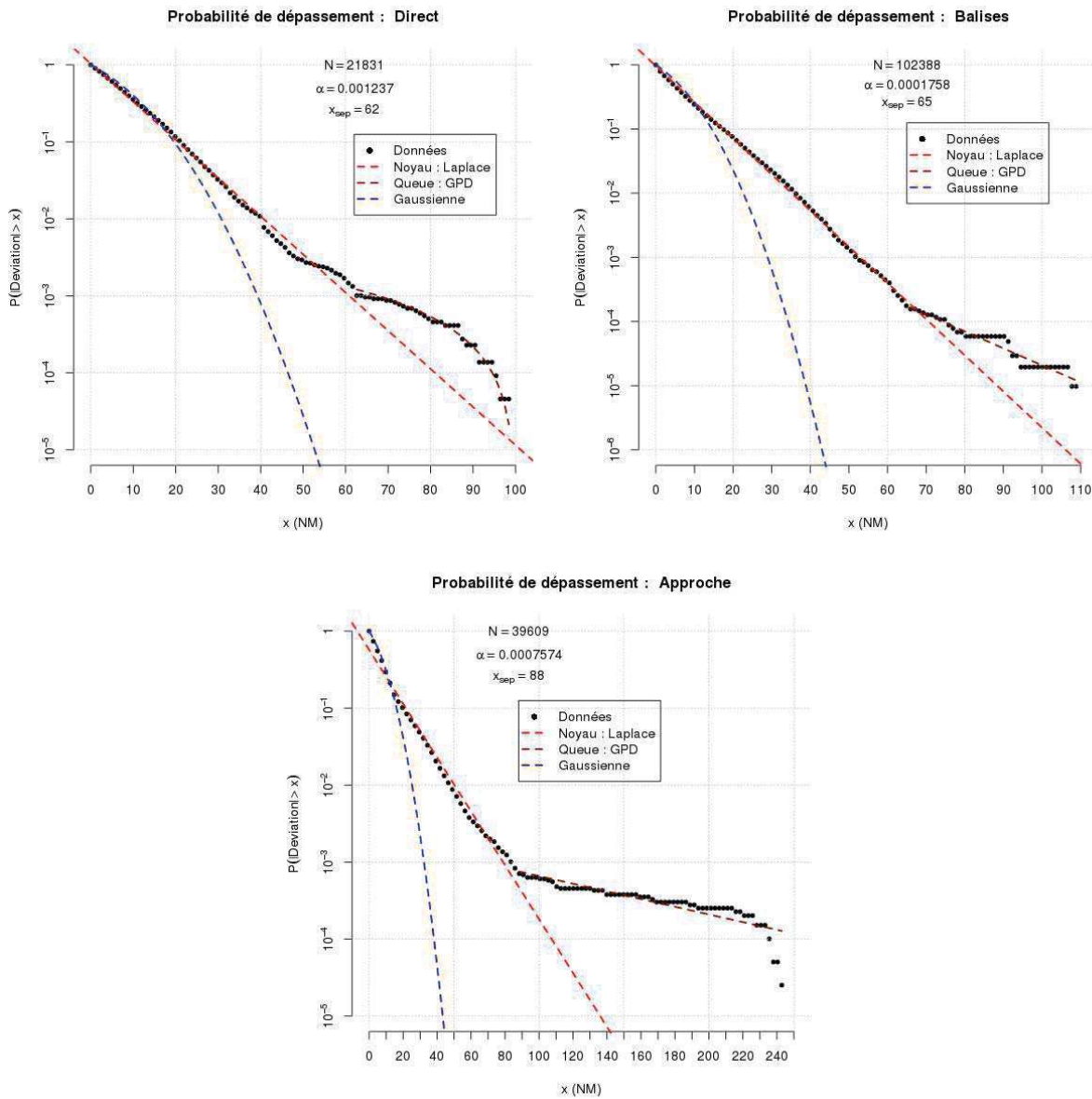


FIGURE 11 – Diagramme des probabilités de dépassement en échelle logarithmique pour les déviations par rapport à la trajectoire directe (en haut à gauche), par rapport à la ligne brisée formée par le plan de vol (en haut à droite) et en phase d’approche (en bas). Sur ce graphique est tracé $2(1 - F(x))$ où $F(x)$ est la fonction de répartition de la loi ou des données considérée.

Cependant ceci n’est vrai que jusqu’au point où l’on passe dans la queue de l’échantillon (le point x_{sep} dans les tracés de la figure 11). A partir de là, on peut considérer que le système est dans un état « dégradé » que l’on modélisera par une loi de Pareto généralisée (GPD) de fonction de répartition $G_{\xi, \sigma}(x)$ [3]. En effet, cette loi est traditionnellement utilisée pour modéliser les valeurs qui excèdent un seuil extrême (cf. annexe 6.2.3). L’intérêt d’utiliser cette loi pour modéliser les extrêmes est bien mis en évidence sur le tracé en haut à droite de la figure 11, où l’on voit que la queue est incurvée vers le bas, ce qui correspond, comme expliqué en annexe 6.2.3, à un paramètre de forme négatif.

Les estimations du paramètre de forme ainsi que les bornes d’un intervalle de confiance à 95% de la valeur du paramètre sont reporté dans le tableau 1. L’estimation des paramètres est faite par

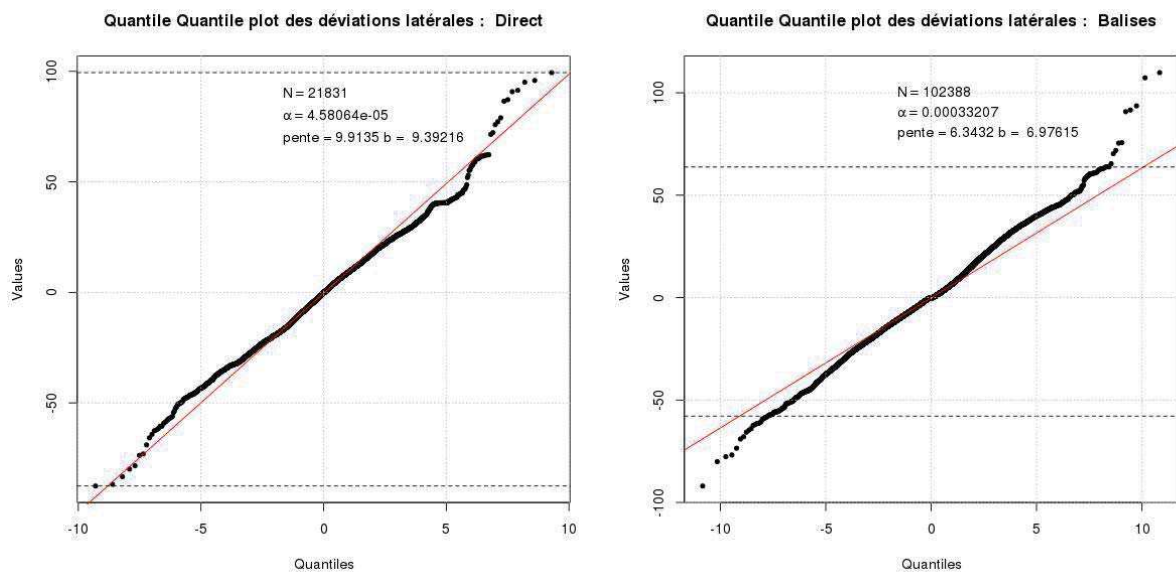


FIGURE 12 – Diagramme quantile-quantile par rapport à une loi de Laplace des déviations à la trajectoire directe (à gauche) et des déviations au plan de vol (à droite).

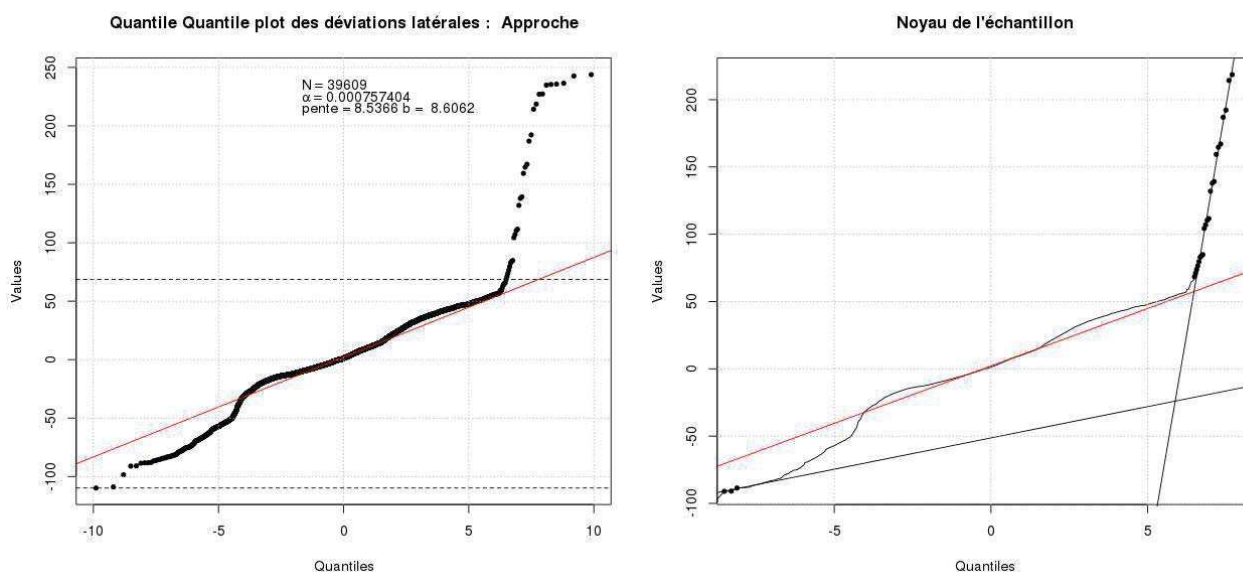


FIGURE 13 – Diagramme quantile-quantile par rapport à une loi de Laplace des déviations en phase d'approche.

maximum de vraisemblance et l'intervalle de confiance par bootstrap [2]. On remarque notamment que dans le cas des trajectoires plan de vol et de la phase d'approche, l'intervalle de confiance inclut la valeur zéro, nous empêchant de rejeter l'hypothèse de nullité du paramètre de forme. Dans ces deux cas nous considérerons donc $\xi = 0$. Le diagramme en échelle logarithmique des tracés en haut à droite et en bas de la figure 11 indiquent d'ailleurs que l'on peut considérer les queues comme une droite. Notons que lorsque ξ est nul, la queue de distribution peut également être modélisée par une loi de

	Borne inférieure	Estimation	Borne supérieure
Direct	-1,581	-0,7626	-0,165
Balises	-1,84	-0,3803	0,1548
Approche	-1,953	-1,234	0,5569

TABLE 1 – Estimation et intervalle de confiance à 95% de la valeur du paramètre de forme ξ de la GPD.

Laplace. En effet, on voit immédiatement que pour une variable aléatoire X suivant une loi de Laplace de paramètres $\mu = 0$ et b ,

$$\mathbb{P}(|X| > x) = 2\mathbb{P}(X > x) = \exp(-x/b) = \mathbb{P}(Z > x)$$

où Z suit une GPD de paramètres $\xi = 0$ et $\sigma = b$. La première égalité résulte du fait que la loi de Laplace est symétrique et centrée en 0.

Nous avons donc ajusté la loi de la déviation latérale Y , conditionnellement à sa valeur absolue. Maintenant, considérons que la probabilité que la valeur absolue de Y soit supérieure au seuil x_{sep} soit égale à α . Comme nous sommes dans un cas centré que nous considérons comme symétrique, nous avons

$$\mathbb{P}(Y \leq -x_{\text{sep}}) = \alpha/2, \quad \mathbb{P}(x_{\text{sep}} < Y \leq x_{\text{sep}}) = 1 - \alpha, \quad \mathbb{P}(Y > x_{\text{sep}}) = \alpha/2.$$

En pratique, nous estimons α par $\hat{\alpha} = k/N$ où k est le nombre de données dépassant x_{sep} en valeur absolue et N est le nombre total de points dont nous disposons. Nous pouvons donc finalement déduire la fonction de répartition de la loi de déviation latérale Y qui est

$$F_Y(x) = \begin{cases} \frac{\alpha}{2} [1 - G_{\xi, \sigma}(-x - x_{\text{sep}})], & x \leq -x_{\text{sep}} \\ \frac{\alpha}{2} + (1 - \frac{\alpha}{2}) L_{0, \lambda^{-1}}(x), & -x_{\text{sep}} < x \leq x_{\text{sep}} \\ 1 - \frac{\alpha}{2} [1 - G_{\xi, \sigma}(x - x_{\text{sep}})], & x > x_{\text{sep}} \end{cases}$$

où $G_{\xi, \sigma}$ est la fonction de répartition d'une GPD de paramètres ξ et σ , et $L_{0, \lambda^{-1}}$ est la fonction de répartition d'une loi de Laplace de paramètres $\mu = 0$ et $b = 1/\lambda$. Les valeurs des paramètres de la fonction de répartition sont reportées dans le tableau 2 pour chacun des trois cas.

	N	α	x_{sep}	λ	σ	ξ
Direct	21831	0,001237	62	8,765	29,1	-0,7626
Balises	102388	0,0001758	65	7,707	16,38	0
Approche	39608	0,0007574	88	12,42	87,7	0

TABLE 2 – Récapitulatif des coefficients de chacun des trois cas.

5.1.2 Calcul de la probabilité de chute sur un site

Grâce à l'utilisation d'une loi de probabilité pour modéliser la déviation, il est possible de calculer la probabilité que l'avion se trouve dans une aire quelconque. Pour commencer, soit X la variable aléatoire de l'avancement d'un aéronef sur son parcours entre un point B_1 (qui correspond à $X = 0$) et un point B_2 (qui correspond à $X = d$ avec d la distance entre B_1 et B_2), et Y la variable aléatoire de sa déviation latérale mesurée au point x . Pour le calcul de probabilité, nous faisons l'hypothèse que la vitesse de l'aéronef est constante sur l'axe B_1B_2 et qu'une chute peut survenir à n'importe quel moment, indépendamment de X . Ainsi, conditionnellement au fait qu'une chute survienne, nous considérons que la position de l'aéronef au moment de la chute suit une loi uniforme entre B_1 et

B_2 , soit de densité $f_X(x) = 1/d$ et de fonction de répartition $F_X(x) = x/d$, pour $x \in [0, d]$. La densité de probabilité de la déviation Y et sa fonction de répartition, lorsque $X = x$, seront notées respectivement $f_{Y|X=x}(y)$ et $F_{Y|X=x}(y)$. La dernière hypothèse que nous ferons est que la loi de la déviation ne dépend pas de l'avancement dans le parcours, i.e. que Y est identique pour toutes les valeurs de X , conformément à ce qui a été expliqué auparavant. Nous pouvons donc écrire la fonction de répartition de Y simplement $F_Y(y)$ et avons donc que la probabilité du couple peut s'écrire comme le produit des fonctions de répartition de chaque variable, i.e. :

$$\mathbb{P}(Y \leq y, X \leq x) = \mathbb{P}(Y \leq y)\mathbb{P}(X \leq x) = F_Y(y)F_X(x).$$

Conditionnellement à X , la probabilité de se trouver dans \mathcal{C} est simple à calculer, elle sera simplement

$$\mathbb{P}(y_1(x) > Y \leq y_2(x) | X = x) = F_Y(y_2(x)) - F_Y(y_1(x)),$$

où $y_1(x)$ et $y_2(x)$ représentent les bornes de \mathcal{C} pour la coordonnée d'avancement $X = x$. A partir de ces hypothèses, la probabilité devient aisée à calculer. En effet, considérons qu'une chute survienne, la probabilité que l'avion A se trouve dans l'aire \mathcal{C} est

$$\mathbb{P}(A \in \mathcal{C} | \text{chute}) = \int_{x_1}^{x_2} f_X(x) \int_{y_1(x)}^{y_2(x)} f_Y(y) dy = \frac{1}{d} \int_{x_1}^{x_2} [F_Y(y_2(x)) - F_Y(y_1(x))] dx.$$

où x_1 et x_2 sont les bornes de \mathcal{C} sur X .

Dans notre cas nous nous intéressons en fait au cas où l'avion est dans une certaine aire et chute, soit l'événement $A \in \mathcal{C} \cap \text{chute}$. Selon la formule des probabilités conditionnelles, nous savons que

$$\mathbb{P}(A \in \mathcal{C} \cap \text{chute}) = \mathbb{P}(A \in \mathcal{C} | \text{chute})\mathbb{P}(\text{chute}).$$

Il ne reste donc plus qu'à calculer le terme $\mathbb{P}(\text{chute})$.

Rappelons que nous modélisons l'apparition de chutes par un processus de Poisson (cf. annexe 6.3) d'intensité λ , où λ est un taux de chutes par unité de distance parcourue (en pratique, λ est un taux de chute par heure de vol, mais il suffit de diviser ce taux par la vitesse de l'avion pour l'obtenir en unité de distance). Le fait de vouloir connaître la probabilité que l'avion A chute durant un vol de distance d , revient à regarder la probabilité que le nombre d'occurrence durant la distance d noté N_d , est supérieur ou égal à 1. Le processus étant poissonien, cette probabilité suit une loi de Poisson d'intensité λd [4]. Nous avons donc

$$\mathbb{P}(\text{chute}) = \mathbb{P}(N_d \geq 1) = 1 - \mathbb{P}(N_d = 0) = 1 - e^{-\lambda d} \approx \lambda d,$$

car λ est très petit dans notre cas. Notez que cette probabilité revient à calculer la probabilité $\mathbb{P}(T \leq d)$ où T suit une loi exponentielle de paramètre λ ; T est en fait la variable aléatoire du temps de première occurrence.

Finalement, nous arrivons à la probabilité p que A chute sur \mathcal{C} alors qu'il se trouve entre B_1 et B_2

$$p = \mathbb{P}(A \in \mathcal{C} \cap \text{chute}) = \lambda \int_{x_1}^{x_2} [F_Y(y_2(x)) - F_Y(y_1(x))] dx.$$

Dans le cas où A suit une trajectoire directe entre ses aéroports de départ et destination, le calcul est celui qui vient d'être expliqué, avec d la longueur du trajet. Par contre, dans le cas où l'avion suit son plan de vol, l'avion peut représenter un danger sur plusieurs tronçons. Ainsi, si le trajet comporte I tronçons, que l'on note p_i la probabilité de chute sur \mathcal{C} lorsque l'avion se trouve sur le i -ème tronçon, d_i la longueur de ce même tronçon et d la longueur totale du parcours, nous calculons la probabilité totale

$$p = \frac{1}{d} \sum_{i=1}^I d_i p_i,$$

car nous considérons le déplacement de l'avion uniforme sur X pour tous les tronçons, en les considérant indépendants deux à deux. On peut s'attendre à ce qu'une majorité des p_i soient nuls, dans ce cas il n'ajouteront rien à la probabilité finale.

5.1.3 Calcul du risque

La section précédente indique le calcul de la probabilité qu'un avion soit dans une certaine aire lorsqu'il entame une chute. Seulement, conformément à ce qui a été dit dans la section 2.4, la métrique finale doit être un nombre de chutes par année et par mètre carré. Une année voit passer un grand nombre d'aéronefs, il faut donc prendre en compte cette densité de trafic en plus de transformer notre probabilité en un taux.

Si l'on considère l'ensemble \mathcal{A} d'avions représentant tous les vols écoulés durant un intervalle de temps T , dont chacun vole indépendamment des autres, chacun d'eux à une probabilité p_i de chuter dans \mathcal{C} , qui dépend de la longueur de son parcours et de l'éloignement de sa trajectoire théorique par rapport à la zone. Ainsi, à chaque vol, on peut associer une variable B_i prenant les valeurs 1 si l'avion chute dans \mathcal{C} , et 0 sinon. La variable B_i suit donc une loi de Bernoulli de paramètre p_i . A partir de là, nous pouvons définir une variable $\Lambda = \sum_i B_i$ qui correspondra au nombre de chute survenues parmi l'ensemble I , représentant donc le risque recherché. Finalement, en faisant l'hypothèse que tous les B_i sont indépendants deux à deux, nous calculons la valeur

$$\mathbb{E}(\Lambda) = \mathbb{E}\left(\sum_i B_i\right) = \sum_i \mathbb{E}(B_i) = \sum_i p_i$$

qui nous donne un nombre moyen de chute par intervalle de temps T . Ainsi $\mathbb{E}(\Lambda)/T$ correspondra à un taux de chute par heure et en divisant par la surface de \mathbb{C} , (sous l'hypothèse que les dimensions de \mathcal{C} sont relativement petite par rapport à la variance de la loi de l'écart, on pourra considérer une chute uniformément répartie sur le domaine \mathcal{C}), nous obtenons un taux de chutes par heure et par m².

5.1.4 Algorithme pratique

L'estimation du risque repose sur le calcul des probabilités en fonction des avions survolant la France. Seulement, en pratique, un grand nombre d'avions survolant la France chaque jour, le calcul sera réalisé sur un sous ensemble d'avions passant suffisamment près de \mathcal{C} dans un intervalle de temps donné (par exemple une journée). L'algorithme pratique est finalement simple, et se décrit pour un ensemble \mathcal{A} d'avions de la façon suivante :

Algorithm 1 Estimation du risque pour un ensemble \mathcal{A} d'avions : première approche

- 1: Détection, parmi \mathcal{A} , des n avions dont la trajectoire passe suffisamment proche de \mathcal{C}
 - 2: **for** $i = 1, \dots, n$ **do**
 - 3: Estimation de p_i , tel qu'en section 5.1.2
 - 4: **end for**
 - 5: Estimation du risque $\hat{\Lambda}_1 = \sum_{i=1}^n p_i$
 - 6: **return** $\hat{\Lambda}_1$
-

La principale difficulté de cet algorithme réside dans le calcul de l'intégrale intervenant dans l'expression de chacune des probabilités p_i , mais en pratique elle peut être résolue par des méthodes numériques.

Cet algorithme demandera un grand nombre de trajectoires afin d'estimer précisément le risque. La deuxième approche, présentée dans la section suivante, n'utilise que les avions qui survolent \mathcal{C} pour estimer le risque en se basant sur la théorie des processus de Poisson de droites.

5.2 Deuxième approche : approche probabiliste du trafic

La méthode présentée dans cette section vise à modéliser chaque site indépendamment, en prenant en compte uniquement les avions passant au voisinage du site sensible. Contrairement à la méthodologie précédente, celle-ci nécessite l'utilisation des données radar (données IMAGE, cf. section 3.2), qui sont plus précises mais aussi plus difficiles à obtenir. Pour appliquer cette méthode on se place ainsi dans le référentiel du site, en considérant celui-ci comme l'origine $(0,0)$ du repère. L'idée qui se trouve derrière cette méthode est justement ce passage au référentiel lié au site où, en partant d'un taux de chute horaire λ associé à un aéronef, on veut parvenir à un taux de chute horaire (ou annuel par exemple) sur le site en question. Il s'agira donc pour ceci de déterminer le passage aux abords du site, en passant par une utilisation des processus de Poisson (cf. annexe 6.3). Notez que la méthode présentée est valable pour les avions commerciaux en phase en-route ou en début de phase d'approche. En effet, l'utilisation de processus de droites nécessite de ne prendre en compte que des avions ayant une trajectoire rectiligne (ou très proche). Cette méthode ne représente en fait qu'une « brique » dans le calcul du risque attribué aux avions commerciaux et doit être complété par un risque du aux aéronefs en phase autre que l'approche ou la croisière. En effet, le site étudié peut se trouver très proche de l'hippodrome d'attente d'un aéroport ou être survolé par des avions en phase de montée initiale ou en phase de descente relativement à un aéroport proche du site. Une prise en compte des aéroports à proximité du site doit donc impérativement venir compléter la méthodologie présentée ici.

5.2.1 Risque induit par un aéronef

Il est intuitif que le risque dépend du passage à l'intérieur de la zone de risque potentiel \mathcal{C} (cf. section 4.2). Nous allons, pour commencer, déterminer le risque créé par la présence d'un seul aéronef. L'extension à plusieurs aéronefs est triviale dans la mesure où il suffira de sommer le risque créé par chaque avion.

Soit λ le nombre de perte de contrôle par heure de vol estimé a priori à partir des statistiques d'accidents, constant quelque soit l'aéronef. Pour un aéronef a , on note v_a sa vitesse en nœuds (mile nautique par heure), alors le rapport λ/v_a donne le nombre de perte de contrôle par mile nautique inhérent à a . Grâce à ce ratio, la connaissance de la distance parcourue par a dans la zone \mathcal{C} permet de déduire le risque induit par cet avion. Ainsi, si cette distance est de longueur L_a , alors le risque induit sera simplement $\lambda_a = \lambda L_a/v_a$. Par ailleurs, soit \mathcal{A} l'ensemble des avions survolant \mathcal{C} durant un intervalle de temps T (e.g. une journée), le risque global par unité de temps T est

$$\Lambda = \lambda \sum_{a \in \mathcal{A}} \frac{L_a}{v_a}.$$

Dans un souci de simplification du modèle, nous ferons l'hypothèse que la dispersion des vitesses des avions dans \mathcal{A} sont relativement faibles et n'ont que très peu d'influence sur le risque. Aussi, nous utiliserons, en guise de vitesse, la moyenne v des vitesses en croisière des aéronefs de \mathcal{A} . Ceci nous permet d'écrire finalement le risque sous la forme suivante :

$$\Lambda = \frac{\lambda}{v} \sum_{a \in \mathcal{A}} L_a. \quad (1)$$

Ainsi, au facteur λ/v qui ne dépend pas des caractéristiques de la zone, le risque global ne dépend que de la longueur totale des trajectoires survolant \mathcal{C} .

5.2.2 Incertitude associée au risque

Rappelons que nous cherchons à quantifier le risque pour un site selon le temps que passent les avions dans la zone de risque potentiel \mathcal{C} , avec l'expression (1), où \mathcal{A} représente l'ensemble des aéronefs

traversant \mathcal{C} , durant l'unité de temps donnée.

Seulement, en pratique, la trajectoire exacte des avions ne peut être prévue, ce qui implique une incertitude des paramètres déterminant la trajectoire (les valeurs p et θ , cf. figure 16 en annexe 6.3) et par extension sur L . De plus, on ne peut non plus connaître le nombre exact $N = \#\mathcal{A}$ d'avions survolant \mathcal{C} durant l'unité de temps choisie, ce nombre pouvant varier de façon drastique. Ainsi, nous nous intéressons au nombre

$$\mathbb{E}(\Lambda) = \frac{\lambda}{v} \mathbb{E}\left(\sum_{a \in \mathcal{A}} L_a\right).$$

Si l'on modélise l'ensemble des trajectoires du voisinage du site par un processus de Poisson de droites (cf. annexe 6.3) d'intensité λ_d , alors la valeur $\mathbb{E}(\Lambda)$ sera proportionnelle à l'aire $\nu_2(\mathcal{C})$ de la zone de risque potentiel \mathcal{C} . En effet, d'après l'identité (4) (cf. section 6.3) nous pouvons écrire

$$\mathbb{E}\left(\sum_{a \in \mathcal{A}} L_a\right) = \lambda_d \nu_2(\mathcal{C}).$$

Ainsi, le risque global attendu sur la zone \mathcal{C} s'exprime par

$$\mathbb{E}(\Lambda) = \frac{\lambda}{v} \lambda_d \nu_2(\mathcal{C}). \quad (2)$$

Notons que pour le processus de droite, nous considérons les droites formées par les trajectoires durant l'unité de temps T choisie. Ainsi la quantité $\lambda_d \nu_2(\mathcal{C})$ correspond à l'espérance de la longueur totale de l'intersection du processus de droites avec la zone de risque \mathcal{C} durant le temps T . Enfin, notons qu'en divisant $\mathbb{E}(\Lambda)$ par l'aire $\nu_2(\mathcal{C})$ de la partie considérée, nous obtenons l'intensité du processus par m^2 , ce qui permet de définir un risque indépendamment de la taille d'une zone. Mathématiquement, cela revient à considérer la quantité

$$\frac{\mathbb{E}(\Lambda)}{\nu_2(\mathcal{C})} = \lambda_d \left(\frac{\lambda}{v}\right),$$

qui exprime la valeur du risque en nombre de chutes/ m^2 durant l'unité de temps T .

5.2.3 Mise en œuvre pratique

En pratique, l'estimation des différents coefficients nécessite l'utilisation de deux types de données (cf. section 3) : les statistiques d'accident et les données radar des vols. Notons que les valeurs respectives de la vitesse moyenne v et du taux de chute horaire λ sont supposées connues. En effet, ces informations peuvent être aisément récupérées auprès d'organismes en lien avec l'aviation civile (DGAC, BEA, etc.). Cependant, pour plus de précision, nous considérerons deux valeurs de λ différentes ; une pour les avions en phase de croisière et une pour les avions en phase d'approche (hippodromes et phase d'atterrissage exclus, voir section 4.1).

Ainsi, on partitionne l'ensemble d'avion \mathcal{A} en deux ensembles disjoints $\mathcal{A}_{\text{approche}}$ et $\mathcal{A}_{\text{croisière}}$ de taille respective N_{approche} et $N_{\text{croisière}}$ ($N = N_{\text{approche}} + N_{\text{croisière}}$). Soit $\lambda_{\text{approche}}$ et $\lambda_{\text{croisière}}$ les taux de chute respectivement en approche et en croisière, alors nous prendrons pour valeur de λ la moyenne pondérée :

$$\lambda = \left(\frac{N_{\text{approche}}}{N}\right) \lambda_{\text{approche}} + \left(\frac{N_{\text{croisière}}}{N}\right) \lambda_{\text{croisière}}.$$

L'estimation du risque lié aux avions commerciaux en phase autre que les phases arrivée et départ d'un aéroport voisin du site, comportera deux principales étapes. Dans un premier temps, les statistiques d'accident devraient permettre de définir le convexe \mathcal{K}_α représentant la zone de chute probable d'un avion. Cette détermination doit être réalisée sur les avions pour lesquels cette méthodologie est valable, i.e. les avions commerciaux en phase en-route, lesquels ont des trajectoires très rectilignes. L'utilisation des MVS (cf. section 4.2) semble pertinente pour cela, tout en veillant à obtenir une zone

convexe (ce qui n'est pas forcément le cas en théorie pour les MVS), en définissant, par exemple, des lignes de niveau. Nous choisirons conformément à la section 4.2 pour \mathcal{K}_α , un cercle de rayon 12NM. Par suite, il est donc possible de déterminer la zone de risque potentiel \mathcal{C} , comme décrit dans la section 4.2.2.

Le choix de \mathcal{C} permet par la suite de sélectionner les plots radar au voisinage du site, pour un laps de temps donné. Ainsi, à partir des plots radar de la zone, il est possible de calculer une estimation \hat{L} de $\mathbb{E}(\sum_{a \in \mathcal{A}} L_a)$ en sommant les longueurs de chaque trajectoire dans \mathcal{C} . Cela revient simplement calculée à sommer la distance qui sépare deux plots consécutifs de chaque trajectoire comme illustré dans la figure 14.

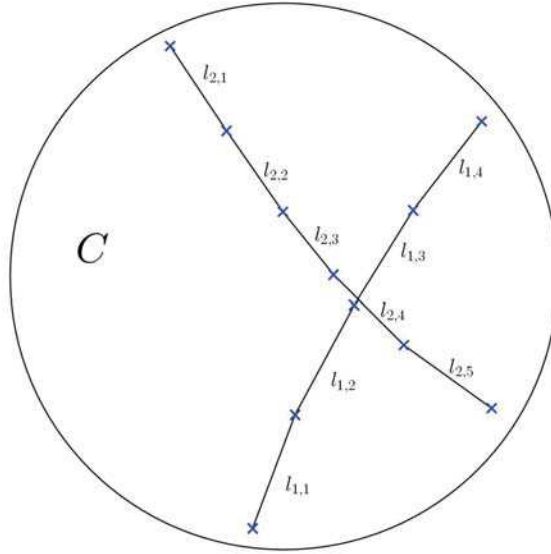


FIGURE 14 – Illustration de la méthode de calcul des longueurs de segment. Ici, la première trajectoire aura une longueur $L_1 = l_{1,1} + l_{1,2} + l_{1,3} + l_{1,4}$ et la deuxième $L_2 = l_{2,1} + l_{2,2} + l_{2,3} + l_{2,4} + l_{2,5}$, puis au final $\hat{L} = L_1 + L_2$.

En résumé, la méthodologie pratique est constitué des étapes suivantes :

Algorithm 2 Estimation du risque pour un ensemble \mathcal{A} d'avions : deuxième approche

Require: Estimation (ou récupération) des valeurs de v , $\lambda_{\text{approche}}$ et $\lambda_{\text{croisière}}$. Détermination de \mathcal{K}_α .

- 1: Calcul de λ ;
 - 2: détermination de \mathcal{C} ;
 - 3: calcul de $\hat{L} = \sum_{a \in \mathcal{A}} L_a$;
 - 4: calcul du risque global $\hat{\Lambda}_2 = \frac{\lambda}{v} \frac{\hat{L}}{\text{aire}(\mathcal{C})}$
 - 5: **return** $\hat{\Lambda}_2$
-

6 Annexes

6.1 Fonction d'intensité ou risque instantané

Étant donnée une variable aléatoire réelle positive T (par exemple la durée de vie d'un composant) de fonction de répartition $F(t) = \mathbb{P}(T \leq t)$, il est courant d'introduire la fonction de survie, notée $R(t)$ ou $\bar{F}(t)$, définie par

$$R(t) = \mathbb{P}(T > t) = 1 - F(t).$$

La fonction de survie étant décroissante et à valeurs dans l'intervalle $[0, 1]$, elle peut s'exprimer sous la forme $R(t) = \exp(-H(t))$, la fonction $H(t)$ étant appelée intensité cumulée ou risque cumulé. La fonction d'intensité ou de risque instantané $h(t)$ est alors définie par

$$h(t) = \frac{dH(t)}{dt} = -\frac{d \ln R(t)}{dt},$$

ce qui permet d'écrire

$$R(t) = \exp\left(-\int_0^t h(u)du\right).$$

Le risque instantané caractérise la probabilité d'une défaillance dans un futur immédiat, sachant un fonctionnement correct jusqu'à l'instant t . En effet, nous avons la relation suivante :

$$h(t) = \lim_{s \rightarrow 0} \frac{\mathbb{P}(T \leq t + s | T > t)}{s}.$$

Lorsque la loi de T admet une densité $f(t) = dF(t)/dt$, le risque instantané se calcule également par l'identité suivante :

$$h(t) = \frac{f(t)}{R(t)}.$$

6.2 Lois de probabilités utilisées en analyse de risque

6.2.1 Loi exponentielle

La loi exponentielle est très utilisée dans les études du risque en raison de sa propriété d'absence de mémoire. En effet, une variable aléatoire T positive suit une loi exponentielle de paramètre λ si sa densité de probabilité est

$$f(t) = \lambda e^{-\lambda t}, \quad t \geq 0.$$

Il est facile de vérifier que l'intensité $h(t)$ d'une loi exponentielle est constante et égale à λ , ce qui conduit à l'approximation

$$\mathbb{P}(t < T \leq t + dt | T > t) \approx \lambda dt,$$

qui caractérise l'absence de mémoire de cette loi. Une autre façon de formaliser ce résultat est de remarquer que pour la loi exponentielle, nous avons, pour tout $t \geq 0$ et $s \geq 0$, la relation

$$\mathbb{P}(T > t + s | T > t) = \mathbb{P}(T > s),$$

qui traduit que la probabilité qu'un phénomène se produise entre les instants t et $t + s$ s'il ne s'est pas produit avant t est la même que la probabilité qu'il se produise entre les instants 0 et s . La loi exponentielle, de par sa caractéristique de perte de mémoire, intervient nécessairement dans les processus markoviens. Ceci découle que lorsque ce type de processus atteint un certain état à l'instant t , il déclenche une horloge aléatoire de loi exponentielle, dont l'intensité dépend de l'état présent, qui lui indique le temps passé dans cet état avant d'effectuer une transition vers un autre état choisi avec une certaine probabilité.

Notons que l'intensité est un nombre d'incident par heure, et lorsque cette intensité est très faible, nous avons l'approximation suivante :

$$\mathbb{P}(T \leq t) = 1 - e^{-\lambda t} \approx \lambda t. \tag{3}$$

Attention cependant à l'utilisation abusive de cette approximation lorsque l'on superpose plusieurs phénomènes indépendants. Par exemple, supposons qu'une défaillance résulte de l'occurrence de n

phénomènes indépendants d'intensités respectives $\lambda_1, \lambda_2, \dots, \lambda_n$. Alors il n'y aura aucune défaillance sur l'intervalle $[0, t]$, si aucun des n phénomènes ne s'est produit dans cet intervalle de temps, d'où

$$\mathbb{P}(\text{aucune défaillance dans } [0, t]) = \prod_i \mathbb{P}(T_i > t) = e^{-\lambda t}, \quad \text{où } \lambda = \sum_i \lambda_i.$$

Une utilisation abusive de l'approximation (3) aurait conduit directement à l'approximation du 1^o ordre

$$\mathbb{P}(\text{aucune défaillance dans } [0, t]) \approx \prod_i (1 - \lambda_i t) \approx 1 - \lambda t.$$

Or rien ne garantit que λ soit petit, et donc que la probabilité ainsi obtenue soit correcte, voir même inférieure à 1.

6.2.2 Loi de Laplace

La loi de Laplace étant très utilisée dans ce rapport, il convient de la définir plus précisément. Cette loi est classiquement utilisée pour modéliser des jeux de données ayant des queues de distribution assez lourdes. Elle est souvent surnommée « double exponentielle », car sa densité peut-être vue comme l'association de deux exponentielles. Plus formellement, une variable aléatoire X suit une loi de Laplace de paramètre de position μ et de paramètre d'échelle $b > 0$, si elle a pour densité

$$f_X(x) = \frac{1}{2b} \exp\left(-\frac{|x - \mu|}{b}\right).$$

L'interprétation de ces deux paramètres est similaire aux paramètres de la loi de Gauss. En effet, à l'instar de cette dernière, le paramètre de position μ , représente la moyenne, la médiane, mais aussi le mode de la variable. Le paramètre d'échelle, bien que lié à la variance de X , n'est pas tout à fait égal à celle-ci, puisque $\sigma^2 = 2b^2$. Le paramètre b correspond plutôt à un écart absolu moyen à la moyenne, son estimateur du maximum de vraisemblance est

$$\hat{b} = \frac{1}{n} \sum_{i=1}^n |x_i - \mu|.$$

La fonction de répartition, obtenue par intégration de la densité, s'écrit

$$F_X(x) = \begin{cases} \frac{1}{2} \exp\left(-\frac{\mu-x}{b}\right) & \text{si } x < \mu \\ 1 - \frac{1}{2} \exp\left(-\frac{x-\mu}{b}\right) & \text{si } x \geq \mu \end{cases}$$

Une propriété que la loi de Laplace partage avec la loi gaussienne est le fait que X peut s'écrire comme la somme $X = \mu + bZ$, où Z est une variable aléatoire de loi de Laplace de paramètres $\mu = 0$ et $b = 1$. Cette propriété justifie l'utilisation de diagrammes quantile-quantile dans la partie 5.1.1.

6.2.3 Lois de Pareto généralisées

Les lois de Pareto généralisées (« Generalized Pareto Distribution » en anglais, abrégé GPD), sont utilisées en statistiques extrêmes pour modéliser les excès par rapport à un seuil relativement élevé. Supposons ce seuil égale à u , les GPD sont utilisées pour caractériser les probabilités du type $\mathbb{P}(X > u + x | X > u)$, avec $x \geq 0$; c'est à dire la probabilité qu'un phénomène dépasse la valeur $x + u$ sachant qu'elle a déjà dépassé le seuil u . Pour un seuil u suffisamment élevé, un théorème limite montre que la variable $X - u$ suit une GPD.

Les GPD sont caractérisées par un paramètre de forme ξ et les paramètres de position μ et d'échelle σ . La fonction de répartition est donnée par

$$G_{(\xi,\mu,\sigma)}(x) = \begin{cases} 1 - \left(1 + \frac{\xi(x-\mu)}{\sigma}\right)^{-1/\xi} & \text{si } \xi \neq 0, \\ 1 - \exp\left(-\frac{x-\mu}{\sigma}\right) & \text{si } \xi = 0. \end{cases}$$

pour $x \geq \mu$ quand $\xi \geq 0$, et $\mu \leq x \leq \mu - \sigma/\xi$ quand $\xi < 0$, où $\mu \in \mathbb{R}$, $\sigma > 0$ et $\xi \in \mathbb{R}$.

Nous utilisons seulement le cas $\mu = 0$ dans la section 5.1.1, ce qui implique que lorsque $\xi \geq 0$, il n'y a pas de limite supérieure sur la valeur de x , alors que pour $\xi > 0$, la limite supérieure du support de la loi est $\sigma/|\xi|$. La figure 15 montre le diagramme des probabilités de dépassement en échelle logarithmique de la fonction GPD pour différents paramètres de forme ξ . Cela permet de mettre en évidence le rôle de ξ car l'on voit qu'en échelle logarithmique, un ξ nul correspond à une ligne droite, un ξ positif donne une courbe convexe, alors qu'un ξ négatif donne une courbe concave.

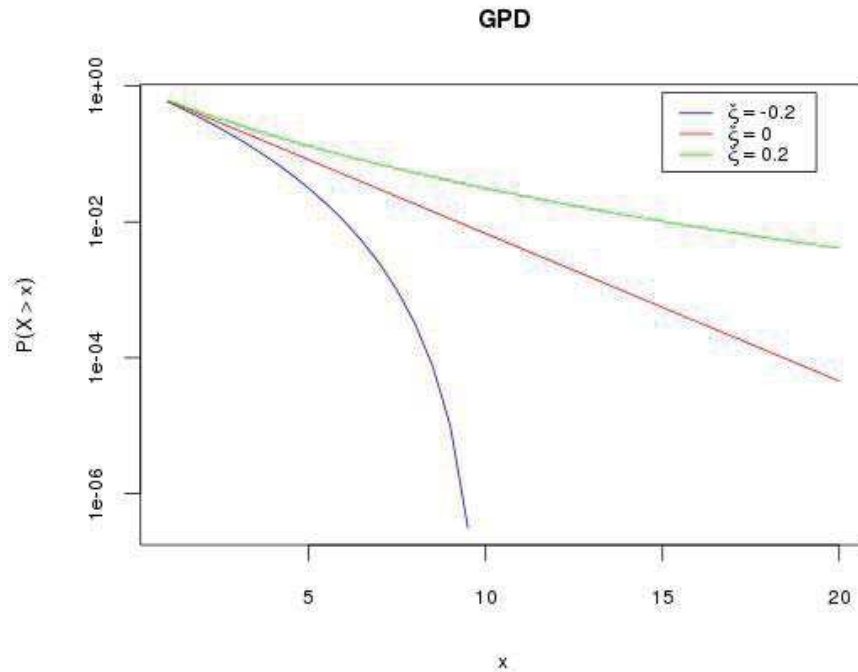


FIGURE 15 – Diagramme des probabilités de dépassement d'une GPD pour différentes valeurs de ξ avec σ fixé à 2.

Les moyenne, médiane et variance sont respectivement

$$\mu + \frac{\sigma}{1 - \xi} \quad (\xi < 1), \quad \mu + \frac{\sigma(2^\xi - 1)}{\xi}, \quad \frac{\sigma^2}{(1 - \xi)^2(1 - 2\xi)} \quad (\xi < 1/2).$$

6.3 Processus de Poisson de droites

L'étude des occurrences d'événements aléatoires peut être abordée d'une autre façon ; au lieu de considérer la durée de bon fonctionnement T , nous pouvons compter le nombre d'occurrence sur un intervalle de temps $[0, t]$. Nous construisons ainsi un processus $N(t)$ qui vaut 0 tant que l'événement n'a pas eu lieu et 1 dès l'apparition de celui-ci. Cette approche est très utile lorsque l'on considère des

successions d'événements dans le temps ; dans ce cas nous observons un processus ponctuel sur l'axe des temps et $N(t)$ est le processus de comptage associé.

Un processus ponctuel particulier est le processus de Poisson qui est caractérisé par le fait que $N(t)$ est distribué comme une loi de Poisson de moyenne $H(t)$, où $H(t)$ est l'intensité cumulée associée à la variable T . Lorsque $H(t) = \lambda t$, le processus de Poisson est dit homogène en temps, du fait que l'intensité instantanée h est constante.

Les processus ponctuels permettent aussi de modéliser la distribution de points dans le plan, ou dans d'autres espaces plus grands, mais aussi la distribution aléatoires d'objets géométriques. Comme exemple important, citons la modélisation de droites orientées dans le plan.

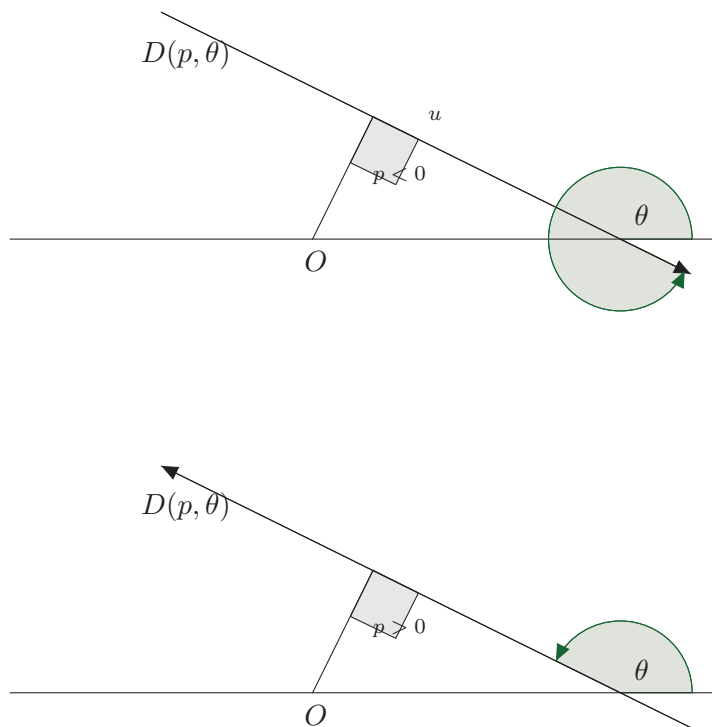


FIGURE 16 – Paramétrage d'une droite orientée dans le plan. Les deux dessins illustrent les cas d'une distance p respectivement négative et positive.

Pour décrire une droite orientée du plan, nous utiliserons les paramètres p et θ tels que décrit dans la figure 16. Plus précisément, étant donné un point origine O et une direction de référence, une droite orientée D est caractérisée par l'angle θ entre la direction de référence et la droite elle-même et par la distance perpendiculaire signée p de D au point O ; le signe étant positif si le point O est situé à gauche de D . Les paramètres auront pour domaine de définition $\theta \in]0, 2\pi]$ et p réel.

Un tel paramétrage permet de décrire le processus aléatoire de droites par un processus de Poisson sur l'espace $]0, 2\pi] \times \mathbb{R}$, ou ce qui est équivalent par une mesure d'intensité $\mu(dp, d\theta)$ définie par : si I et J sont des parties respectivement de \mathbb{R} et $]0, 2\pi]$, alors le nombre moyen de droites de paramètres

p, θ appartenant à $I \times J$ est donné par

$$\mu(I \times J) = \int_{I \times J} \mu(dp, d\theta).$$

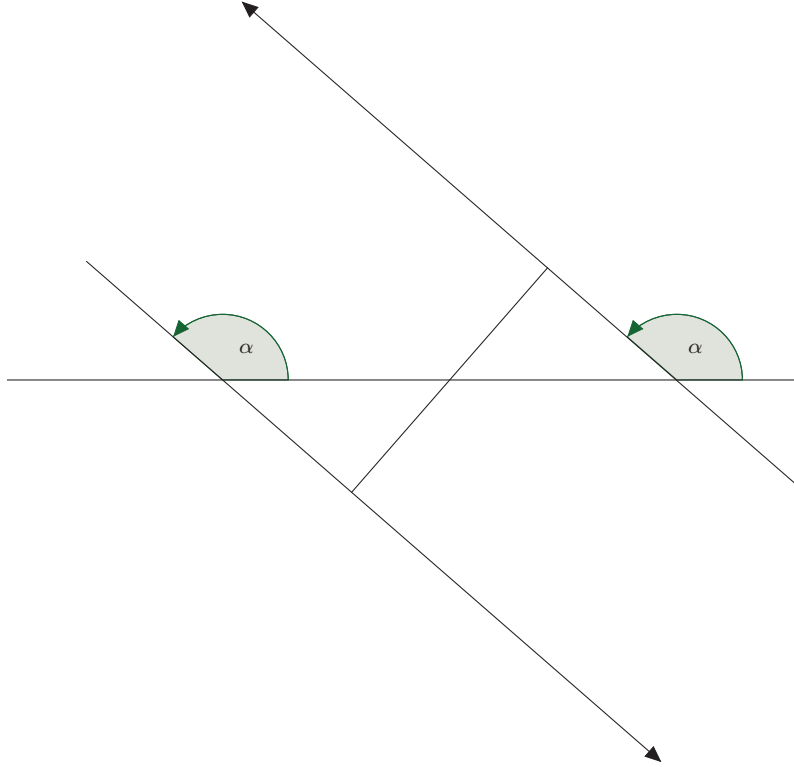


FIGURE 17 – Processus de droites de même angle α ; illustration pour une même valeur absolue de la distance p .

Si le processus est invariant par translation, c'est-à-dire que ces propriétés ne dépendent pas du choix de l'origine, alors μ est nécessairement de la forme

$$\mu(dp, d\theta) = L_A dp R(d\theta),$$

où L_A est une constante exprimant l'intensité moyenne du processus et $R(d\theta)$ est une densité de probabilité sur $]0, 2\pi]$, dénommée rose des directions du processus. Cette distribution peut s'interpréter comme la distribution des directions privilégiées des droites du processus. Si la distribution des droites est isotrope, donc indépendante du choix de l'axe de référence, alors $R(\theta) = d\theta/2\pi$. Comme contre-exemple, citons le cas d'une mesure de Dirac sur le cercle, $R(d\theta) = \delta_\alpha(\theta)$, c'est-à-dire que si I est une partie de $]0, 2\pi]$ contenant la valeur α , alors $\delta_\alpha(I)$ vaut 1, et 0 sinon. Cet exemple traduit un processus de droites de paramètre θ la seule valeur α ; pour la même valeur absolue $|p|$, nous obtenons deux droites parallèles d'orientation différente séparées d'une distance $2p$ (cf. figure 17).

Une interprétation de L_A s'obtient de la façon suivante. Pour toute partie B du plan, nous pouvons mesurer, en moyenne, la longueur totale des intersections avec B des droites du processus; cela définit une mesure μ^* par

$$\mu^*(B) = \mathbb{E}\left(\sum_{\ell} \nu_1(\ell \cap B)\right),$$

où ℓ sont les droites du processus et nu_1 désigne la mesure de longueur. Si le processus de droites est invariant par translation, la mesure μ^* est une mesure dans le plan invariante par translation, par conséquent elle est proportionnelle à la mesure d'aire (mesure de Lebesgue du plan), notée nu_2 , d'où $\mu^*(B) = c\nu_2(B)$. Choisissons pour B la boule $b(0, 1)$ centrée en l'origine et de rayon 1, et appliquons la relation précédente. Tout d'abord $\mu^*(b(0, 1)) = c\pi$, alors que

$$\begin{aligned}\mu^*(b(0, 1)) &= \mathbb{E}\left(\sum_{\ell} \nu_1(\ell \cap b(0, 1))\right) \\ &= \mathbb{E}\left(\sum_{\ell: |p| \leq 1} 2\sqrt{1-p^2}\right) \\ &= 2 \int_{|p| \leq 1} \sqrt{1-p^2} \mu(dp, d\theta) \\ &= 2L_A \int_{-1}^1 \sqrt{1-p^2} dp = L_A \pi\end{aligned}$$

Le passage à la deuxième ligne découle d'une propriété classique du cercle sur la puissance d'un point par rapport à un cercle, à savoir : soit M est un point quelconque et C un cercle de centre O et de rayon R , alors pour toute droite passant par M et coupant le cercle en A et B (cf. figure 18), nous avons l'identité $MA \times MB = |OM^2 - R^2|$. Il suffit d'appliquer cette formule pour $R = 1$ et M le point sur la droite à distance p de l'origine.

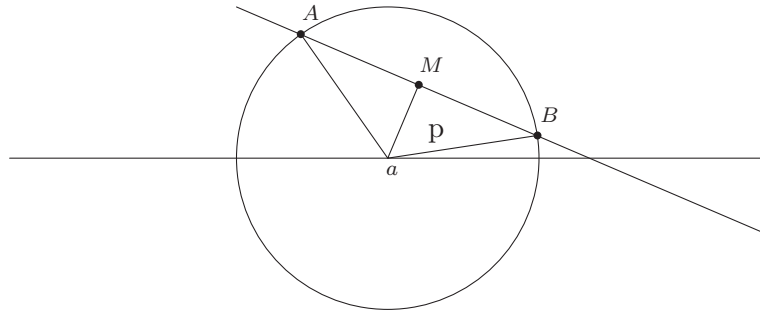


FIGURE 18 – Puissance d'un point par rapport à un cercle.

Il en découle que l'intensité L_A de la mesure aléatoire de longueur μ^* , dans le cas invariant par translation, est égale à la constante c et vérifie la relation suivante, valable pour toute partie (mesurable) B du plan :

$$L_A \nu_2(B) = \mathbb{E}\left(\sum_{\ell} \nu_1(\ell \cap B)\right). \quad (4)$$

La figure 19 illustre la simulation d'un processus de droites non orientées dans un cercle de rayon R , invariant par translation et isotrope. Pour cela, il suffit de tirer θ uniformément dans $[0, \pi]$ et p uniformément dans $[-R, R]$.

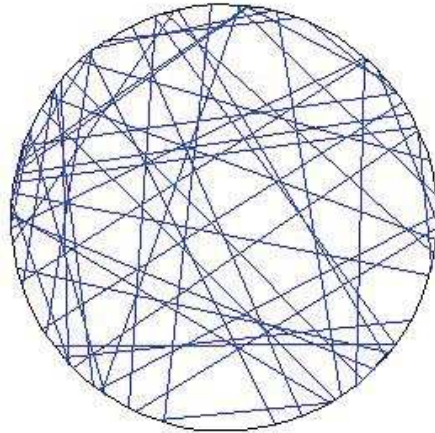
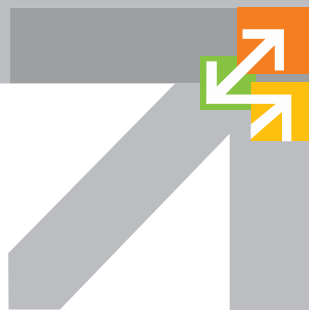


FIGURE 19 – Simulation d'un processus de droites non orientées dans un cercle.

Références

- [1] Freissinet, S. *Étude statistique des accidents aériens*. Consulté le 11 juillet 2012, sur 1001 Crash : <http://www.1001crash.com/index-page-statistique-lg-1-numpage-3.html>
- [2] Hastie, T., Tibshirani, R., & Friedman, J. *The Elements of Statistical Learning*. Springer, 2008.
- [3] International Civil Aviation Organisation. (2009). *A Unified Framework for Collision Risk Modelling in Support of the Manual on Airspace Planning Methodology for the Determination of Separation Minima* Doc 9689, 2009.
- [4] Kendall, W., Stoyan, D., & Mecke, J. (1987). *Stochastic geometry and its applications*. Wiley, 1987.
- [5] Scott, C., & Nowak, R. *Learning Minimum Volume Sets*. J. Machine Learning Res., 2006.



École Nationale de l'Aviation

7, avenue Édouard Belin

BP 54005

31055 Toulouse cedex 4

Tél. +33 (0)5 62 17 40 00

Fax +33 (0)5 62 17 40 23



La référence aéronautique

www.enac.fr →