Contributions à la simulation des évènements rares dans les systèmes complexes

Jérôme Morio

Université de Rennes I

Ecole doctorale MATISSE

Contributions à la simulation des évènements rares dans les systèmes complexes

Dossier de Synthèse

présenté et soutenu publiquement le 9 décembre 2013

pour l'obtention d'une

Habilitation à Diriger des Recherches de l'Université de Rennes I

(Spécialité génie informatique, automatique et traitement du signal)

par

Jérôme MORIO, Ingénieur de Recherche ONERA

Rapporteurs : Anne-Laure Fougères (Pr., Université Claude Bernard Lyon I)
Pierre-Etienne Labeau (Pr., Université Libre de Bruxelles)
Igor Nikiforov (Pr., Université de Technologie de Troyes)

Examinateurs : Josselin Garnier (Pr., Université Paris VII - Diderot)
Bertrand Iooss (Chercheur senior, EDF R&D)
François Le Gland (DR., INRIA Rennes)

Office National d'Etudes et de Recherches Aérospatiales

Mis en page avec la classe thloria.

Table des matières

Remerciements

Cette synthèse conclut 6 années de recherches sur l'analyse de systèmes complexes et plus particulièrement l'estimation de probabilité d'évènements rares. Il est donc temps de remercier les personnes sans lesquelles ces travaux n'auraient pu voir le jour.

Mes remerciements vont en premier lieu à ceux qui ont accepté d'être les rapporteurs de ce dossier d'Habilitation à Diriger des Recherches, Anne-Laure Fougères, Pierre-Etienne Labeau et Igor Nikiforov. Leurs remarques constructives ainsi que leurs propositions concernant les perspectives potentielles de ces travaux furent une source de motivation appréciable. La recherche sur les méthodes de simulation d'évènements rares a encore de beaux jours devant elle.

Ensuite, déterminer quelle méthode de simulation s'avère la mieux adaptée à un problème précis demeure une question ouverte qui implique une certaine part de subjectivité. C'est pourquoi je souhaite également remercier Bertrand Iooss qui nous a fait partager son regard sur l'applicabilité des méthodes de simulation d'évènements rares. Je suis certain que nous aurons l'occasion de collaborer sur ce sujet dans le cadre du GDR MASCOT-NUM ou ailleurs.
J'exprime par ailleurs toute ma gratitude à Josselin Garnier qui m'a fait l'honneur de présider ce jury de HDR.

Qui plus est, rédiger un mémoire d'habilitation est une décision complexe à prendre. Philippe Bidaud, directeur scientifique de la branche Traitement de l'Information et Systèmes de l'Onera, engendra cet élan initial qui me permit d'affronter mes craintes. Il me fournit de précieux conseils en vue de l'aboutissement de ce manuscrit ainsi que de sa soutenance.
De plus, le soutien et l'accompagnement de François Le Gland lors des différentes étapes de l'obtention de cette HDR à l'Université de Rennes I ont également été déterminants. Mes remerciements à François dépassent bien sûr le cadre de cette HDR. Nos collaborations, lors de l'encadrement de doctorants, ont toujours été très enrichissantes et complémentaires. J'ai également une pensée pour les différents directeurs de thèse avec lesquels j'ai collaboré, que sont Daniel Vanderpooten et Pierre Del Moral.

Par ailleurs, les travaux de cette HDR s'appuient très largement sur les résultats obtenus grâce au travail de recherches des doctorants que j'ai co-encadrés ou que je co-encadre actuellement, à savoir Rudy Pastel, Dalal Madakat, Damien Jacquemart et Christelle Vergé. Travailler avec eux au jour le jour est une source d'enrichissement mutuel, de remise en question et d'inspiration inestimable.

D'autre part, cette habilitation doit beaucoup à Mathieu Balesdent, ingénieur de recherche à l'Onera. Son arrivée dans mon unité, il y a deux ans, a donné un second souffle, une vision neuve à ces travaux sur la simulation d'évènements rares. J'assisterai avec plaisir et à n'en pas douter, d'ici quelques années, à sa soutenance de HDR.

Je remercie également ma hiérarchie et plus particulièrement Thérèse Donath, directrice du Département Conception et évaluation des Performances des Systèmes (DCPS), pour m'avoir embauché, il y a 6 ans, poussé à travailler sur la thématique des évènements rares et entendu mes envies de renouvellement. Cette habilitation raisonne donc aussi comme un nouveau départ, vers le DCPS de Toulouse, à partir de janvier 2014. Je suis certain que de futures collaborations, notamment au sein de ma nouvelle unité, verront le jour. J'ai aussi une pensée pour les collègues de mon unité actuelle, à Palaiseau, et leur souhaite beaucoup de réussite à venir.

Du reste, trouver des enseignements intéressants n'est pas toujours aisé pour les non académiques. C'est pourquoi, j'exprime tous mes remerciements à Vanessa Vitse et Laurent Dalmas pour leur confiance. Enseigner à l'Institut d'Optique a été un grand plaisir que je dois à François Goudail. J'espère que nous aurons à l'avenir l'occasion de nous recroiser.

Enfin, cette habilitation n'aurait pas pu voir le jour sans le soutien de ma famille qui a su m'encourager dans mes études, acceptant avec enthousiasme chacun de mes choix. Qui plus est, l'appui de ma femme, Anne, fut inestimable. Toujours à l'écoute afin de m'aider à relativiser mes difficultés et mes doutes, elle s'est impliquée pleinement et m'a aidé dans chaque étape de cette soutenance de HDR. Avec tout mon amour, je la remercie pour tout.

Motivations

En tant qu'ingénieur de recherche au centre ONERA de Palaiseau, mon travail de recherche s'est axé sur l'évaluation de performances de systèmes, et notamment l'estimation d'évènements rares depuis l'obtention de ma thèse de doctorat. Au contact notamment de la direction scientifique de l'ONERA, mais également par motivation personnelle, il m'est apparu nécessaire, dans un objectif de synthèse sur ces années de recherches, de regrouper mes travaux sur l'estimation d'évènements rares dans un dossier visant à l'obtention d'une Habilitation à Diriger des Recherches.
Bien que basé en région parisienne, mon choix pour déposer ce dossier s'est porté vers l'Université de Rennes 1, en raison de mon travail de collaboration régulier avec François Le Gland, Directeur de recherche à l'INRIA de Rennes, notamment par le biais du co-encadrement de thèse que j'ai effectué auprès de deux doctorants inscrits à l'Université de Rennes 1 et dont il est le directeur de thèse. C'est en raison d'une collaboration enrichissante d'un point de vue scientifique mais aussi humain, qu'il s'est porté garant de cette demande d'habilitation.
Ce dossier de synthèse retrace de façon organisée et concise, l'ensemble des recherches que j'ai menées à l'ONERA depuis l'obtention d'un doctorat en physique en 2007, à l'Université Paul Cézanne d'Aix-Marseille. Bien que mon travail de thèse fut également développé en partenariat avec l'ONERA, sur le thème de l'analyse d'images polarimétriques et interférométriques radar à l'aide de techniques statistiques et issues de la théorie de l'information, j'ai opéré un virage assez radical quant au sujet de mes recherches. En effet, la thématique de cette synthèse ne se situe pas dans la continuité de ma thèse. Cependant, l'expérience acquise sur les techniques statistiques durant ces trois années de thèse fut un atout pour mon travail de recherche sur les évènements rares.
Les trois principales parties qui composent ce dossier de synthèse présentent, dans un premier temps, le contexte encadrant ces recherches et précisent les différents thèmes scientifiques abordés au cours de celles-ci, puis détaillent mes activités de publications, de communications, d'enseignement, d'administration de la recherche et enfin d'encadrement. Dans une seconde partie, ce mémoire aborde certains aboutissements de mes recherches, ici certaines contributions scientifiques que j'ai proposées, développant les méthodes d'estimation d'évènements rares dans les systèmes complexes. Enfin, un choix représentatif de cinq publications dont je suis co-auteur est proposé en conclusion de ce mémoire.

Première partie

Présentation générale et résumé des activités de recherche

Chapitre 1

Présentation générale

Jérôme Morio, Ingénieur de recherche à l'ONERA (Office National d'Etudes et de Recherches Aérospatiales)
- Tél : 01 80 38 66 54
- Email : jerome.morio@onera.fr
- Adresse professionnelle : ONERA, DCPS/SSD, Chemin de la Hunière et des Joncherettes, 91123 Palaiseau Cedex
- Page personnelle : http ://www.onera.fr/staff/jerome-morio/

1.1 Historique

A la suite d'une formation en physique à l'Ecole Centrale de Marseille (Ecole Nationale Supérieure de Physique de Marseille), j'ai débuté ma première expérience de recherche scientifique en 2004 lors d'un stage de DEA de quatre mois sous la direction de François Goudail et Philippe Réfrégier à l'Institut Fresnel (Marseille) sur le thème de la caractérisation de l'information polarimétrique dans les images de Mueller. J'ai ensuite poursuivi cette étude dans une thèse également dirigée par François Goudail et Philippe Réfrégier, cette fois-ci en collaboration avec l'ONERA, sur l'analyse d'images polarimétriques et interférométriques SAR (Synthetic Aperture Radar) à l'aide de techniques statistiques et issues de la théorie de l'information. De plus, en parallèle, j'ai enseigné en licence les notions de systèmes temps réels à l'IUT de Salon de Provence. J'ai soutenu ma thèse en physique en novembre 2007 après avoir été engagé en CDI à l'ONERA en septembre de cette même année. Mon travail consiste désormais à analyser les performances de systèmes complexes, également au moyen de techniques statistiques. Ces systèmes sont cependant moins ancrés dans le domaine de la physique que ne l'étaient ceux étudiés durant ma thèse.

1.2 Curriculum Vitæ

- **2007/....** Ingénieur de recherche à l'ONERA, centre de Palaiseau, Département Conception et évaluation des Performances des Systèmes (DCPS), Unité Systèmes Spatiaux et de Défense (SSD).
- **2004-2007** Doctorat en Physique, spécialité optique, électromagnétisme et image, de l'Université Paul Cézanne d'Aix-Marseille, effectué à l'Institut

Fresnel, Marseille et à l'ONERA, Centre de Salon de Provence, Département ElectroMagnétisme et Radar (DEMR), sur l'analyse d'images polarimétriques et interférométriques SAR à l'aide de techniques statistiques et issues de la théorie de l'information.

- **2003-2004** DEA Optique Image Signal, option Image à l'Université Paul Cézanne d'Aix-Marseille. Mention Très Bien - Classement : 1er sur 30.
- **2001-2004** Diplôme d'ingénieur de l'Ecole Nationale Supérieure de Physique de Marseille (ENSPM) / Centrale Marseille.

1.3 Bilan rapide de l'activité de recherche

Le bilan de mes activités de recherche est résumé ci-après.

Publications et communications

- 24 publications dans des revues à comité de lecture dont 23 référencées sur Thomson Reuters Web of Knowledge, 2 articles soumis en cours de révision et 4 en cours de relecture.
- 10 communications à congrès avec comité de sélection.

Encadrement

- 4 co-encadrements à 50% de thèse. 1 thèse soutenue, 1 thèse en cours de finalisation et 2 thèses en cours.
- 5 encadrements à 100% de stagiaire de niveau Master 2.

Collaborations académiques

- Principales collaborations académiques : F. Le Gland (DR INRIA), D. Vanderpooten (Pr Univ. Paris Dauphine), P. Del Moral (DR INRIA).
- Membre du GDR Mascot-Num (Méthodes d'Analyses Stochastiques pour les COdes et Traitements NUMériques).

Rayonnement scientifique

- Activité de relecture pour les revues suivantes (6-7 articles par an) : Entropy, The Proceedings of the IEEE, Safety Science, Journal of Risk and Reliability, Optics Express, Aerospace Science and Technology, IET Control Theory & Applications, International Journal of Remote Sensing, Applied Optics, IEEE Transactions on Geoscience and Remote Sensing.
- Organisation, avec Fabrice Poirion (ONERA), d'une Journée Scientifique le 15/11/2012 sur les méthodes d'évènements rares au centre ONERA de Palaiseau (5 intervenants extérieurs, 90 participants dont 70 externes à l'ONERA). (http ://sites.onera.fr/MODNAT/node/9)

1.4 Discipline de recherche

1.4.1 Thématiques de recherche

Mon travail de recherche à l'ONERA consiste à analyser les performances de systèmes complexes liés à l'aéronautique et au domaine spatial, dans un cadre civil ou de défense nationale. Ces systèmes multidimensionnels et incertains peuvent être statiques ou dynamiques selon la dépendance temporelle de leurs sorties. Les techniques statistiques et de simulation trouvent donc une place prépondérante dans mes travaux, qui par ailleurs s'inscrivent dans le cadre de la

section 61 du CNU, génie informatique, automatique et traitement du signal. Les paragraphes suivants résument les thématiques de recherche que j'ai abordées dans un cadre de collaboration académique et contractuelle. De plus, les avancées scientifiques obtenues sont évoquées dans leur contexte et leurs perspectives éventuelles sont discutées.

Introduction

L'objectif de la thématique de recherche développée ci-après est d'analyser les performances de systèmes complexes statiques de type entrées-sorties (également appelés "boîtes noires") [36]. Ce modèle très générique permet de représenter de nombreuses applications dans les domaines de l'aérospatial et de la défense [64]. Il est tout à fait possible de simuler les performances d'un système aérospatial, en représentant par exemple la trajectoire d'un lanceur, à l'aide de ce type de modèle. Les entrées aléatoires sont notamment les caractéristiques aérodynamiques d'un lanceur et de ses équipements. La sortie correspond à la distance au plus près de la position prévue de retombée.

De manière générale, je supposerai par la suite qu'un modèle boîte noire est défini par des entrées aléatoires caractérisées par une loi de probabilité, par des sorties supposées également aléatoires, et par une fonction entrées-sorties déterministe. Analyser les performances d'un système stochastique de type boîte noire revient donc à caractériser les lois de probabilité de ses entrées et de ses sorties, à estimer un critère de performance (une probabilité ou un quantile par exemple) sur les sorties ou bien à analyser la sensibilité des sorties vis-à-vis des entrées. De manière générale, les méthodes Monte-Carlo, comme le montrent les références [72] et [121], permettent d'évaluer les performances d'un système boîte noire mais s'avèrent coûteuses en temps de calcul, en nombre de simulations et limitées en terme de vitesse de convergence, pour l'analyse de systèmes complexes. C'est pourquoi, bien que simple à modéliser, l'analyse de performances de systèmes entrées-sorties peut donc être difficile à mettre en place.

Définir des modèles de substitution de la boîte noire s'avère donc pertinent afin de réduire le temps de calcul des simulations. L'analyse de la sensibilité d'un code de calcul par rapport à ses entrées permet également de simplifier le modèle en réduisant le nombre d'entrées aléatoires à générer. Caractériser la loi de probabilités de la sortie peut permettre d'estimer à moindre coût des critères de performance sur celle-ci. Enfin, il peut s'avérer nécessaire d'accélérer la convergence des estimateurs Monte-Carlo, par exemple pour l'estimation d'évènements rares. Mes apports dans ces thèmes de recherche sont développés dans les paragraphes suivants.

Modèle de substitution

La simulation d'un système boîte noire avec des entrées multidimensionnelles peut s'avérer coûteuse en temps de calcul. Le budget de simulation est limité et ne permet pas d'appliquer directement sur ce modèle des techniques d'évaluation de performances génériques. Néanmoins, il est possible de définir un modèle de substitution au modèle boîte noire complexe, à l'aide d'échantillons entrées-sorties de celui-ci. L'estimation des sorties d'un modèle de substitution à partir d'échantillons des entrées est en règle générale peu coûteuse en temps de calcul.

Différentes structures de modèles ont été proposées : le krigeage, les réseaux de neurones, les splines, entre autres. L'apprentissage des poids d'un réseau de neurones est souvent problématique car le critère d'optimisation des poids de celui-ci est multidimensionnel et multimodal. Néanmoins, les réseaux de neurones sont très efficaces lorsque la base de données d'apprentissage est importante. J'ai ainsi comparé dans un rapport technique différents algorithmes d'apprentissage afin de régler les poids d'un réseau de neurones. Une application de ces algorithmes à un cas industriel de missile développé par MBDA-Missile Systems ayant 52 entrées indépendantes est également réalisée dans un rapport technique confidentiel. Le réseau de neurones se substitue avec succès au système complexe mais peut dépendre fortement de son apprentissage. J'ai ainsi pu mettre en place des méthodes d'évaluation de performances impossibles à appliquer directement sur un cas industriel. Le krigeage est également un autre type de modèle de substitution défini par des processus gaussiens. Cet algorithme est particulièrement adapté aux travaux statistiques car sa variance de prédiction peut être déterminée analytiquement en chaque point. Lorsque la base de données d'apprentissage est peu importante, le krigeage est une méthode efficace pour estimer un modèle de substitution. Le nombre d'appels au modèle s'avère limité avec un risque d'erreur de modélisation faible. Pour des bases de données d'apprentissage plus complètes, le calcul de la surface de réponse est complexe, du fait de l'inversion d'une matrice carrée qui a pour nombre de lignes le nombre d'échantillons de la base d'apprentissage et pose donc des problèmes de temps de calcul. Le couplage entre krigeage et algorithmes d'optimisation a été proposé dans la littérature. J'ai évalué l'efficacité de ce couplage pour le positionnement d'un capteur, afin de maximiser sa probabilité de détection, dans l'article [98]. Le temps de calcul nécessaire au positionnement du capteur a été fortement diminué, tout en conservant une forte probabilité de déterminer sa position optimale.

Bilan : 1 publication dans une revue à comité de lecture.

Caractérisation d'une loi de probabilité à base de mesures physiques

La caractérisation de la loi de sortie d'un code de calcul peut également être un atout dans le cadre de l'analyse de performances, afin, par exemple, d'estimer une probabilité de fausse alarme (PFA) ou de détection (PD). A cet effet, des estimateurs statistiques basés sur la théorie de l'information peuvent être notamment utilisés.

La distance de Bhattacharrya est un indicateur très pertinent de la distance entre deux lois de probabilité comme je le présente dans l'article [104]. Il existe ainsi une bijection entre l'aire sous la courbe PD/PFA estimée par simulation et la distance de Bhattacharrya. Cette approche est également valable dans le cas où les sorties du code de calcul à évaluer sont multidimensionnelles. L'entropie de Shannon correspond à l'indice de variabilité d'une distribution statistique. Celle-ci est différente de la variance et peut être considérée comme un indice de la dispersion d'une variable aléatoire. Des propriétés de l'entropie de Shannon ont ainsi été démontrées en optique afin de caractériser les différences sources de variabilité de la lumière, dans la publication [119]. J'ai également proposée une approche conjointe de la distance de Bhattacharrya et de l'entropie de Shannon dans l'article [105], afin de caractériser des distributions statistiques. Le cadre d'application développé dans cet article en collaboration avec l'Institut Fresnel

dans le cadre de ma thèse concerne les images polarimétriques et interférométriques SAR dont les réalisations peuvent être interprétées comme les sorties d'un code de calcul multidimensionnel.
Lorsque peu de réalisations de la sortie d'un code de calcul sont disponibles, il est difficile de caractériser la loi de celle-ci et d'évaluer les performances du système. J'ai proposé une approche basée sur les techniques d'estimation d'erreurs de détection par moindres carrés dans l'article [99], pour caractériser des objets spatiaux à l'aide de mesures radar. Les entrées du code de calcul sont des mesures radar bruitées et les sorties sont des paramètres orbitaux des objets. Chaque objet recherché est défini par des caractéristiques orbitales puis comparé à la sortie du code de calcul. Le nombre de mesures radar étant limité et celles-ci devant être traitées en temps réel, il faut être capable d'associer de façon efficace les échantillons de mesure à une classe d'objets, en minimisant la probabilité de fausse alarme. A l'aide de l'estimation proposée, 10 mesures radar sont alors nécessaires pour évaluer de manière fiable les caractéristiques d'un objet spatial ce qui représente un gain important par rapport aux méthodes existantes.

Bilan : 2 publications (hors thèse) dans des revues à comité de lecture.

Analyse de sensibilité

L'analyse de sensibilité consiste à déterminer les entrées du code de calcul les plus influentes sur sa sortie. Il existe principalement deux approches d'analyse de sensibilité, locale et globale :

- l'analyse de sensibilité globale étudie la variance des sorties d'un modèle et plus précisément, détermine la part de variance de la sortie, issue de l'incertitude de chaque entrée, notamment à l'aide de l'estimation d'indices de Sobol. Ces indices caractérisent la manière dont les différentes entrées influencent la sortie en supposant l'indépendance de celles-ci.
- l'analyse de sensibilité locale procède par évaluation des dérivées partielles de la sortie du code de calcul par rapport à ses différentes entrées. La moyenne et l'écart type de ces dérivées partielles pour plusieurs valeurs de l'espace d'entrée définissent l'influence d'une entrée sur la sortie.

J'ai comparé les résultats obtenus entre méthode d'analyse de sensibilité locale et globale dans l'article [92]. Il s'avère notamment que l'analyse de sensibilité locale n'est pas robuste aux effets non linéaires et aux interactions des entrées sur la sortie, contrairement à l'analyse de sensibilité globale. Ceci est dû au fait que les sensibilités locales ne prennent en compte que le gradient (différentielle) qui est linéaire. En revanche, les méthodes d'analyse de sensibilité globale sont souvent coûteuses en nombre de simulations ($\approx 500\times$ la dimension des entrées). Ce n'est pas le cas des méthodes locales qui mettent en évidence les variables les plus influentes à faible coût de simulation ($\approx 20\times$ la dimension des entrées). L'utilisation conjointe de méthode d'analyse de sensibilité globale et de modèle de substitution est une manière de résoudre ce problème.

Les paramètres des lois des entrées du code de calcul jouent également un rôle prépondérant sur la loi de sortie. En général, ceux-ci sont estimés à partir d'observations ou par connaissance *a priori* et sont ensuite supposés constants durant la simulation. Ces hypothèses peuvent par conséquent influencer l'évaluation d'un critère de sortie du code de calcul, tel que la probabilité de dépassement d'un seuil. J'ai ainsi proposé une méthode d'échantillonnage couplée à une analyse de

sensibilité globale dans l'article [93], afin d'évaluer l'influence de ces paramètres sur l'estimation d'un critère de sortie. Une estimation de l'impact d'une erreur commise sur un paramètre de loi des entrées y est également détaillée. Un classement des paramètres de loi en fonction de leur influence sur le critère de sortie peut être obtenu pour un code de calcul donné.

Bilan : 2 publications dans des revues à comité de lecture.

Estimation d'évènements rares sur la sortie d'un code de calcul

Afin d' estimer une probabilité sur la sortie d'un code de calcul de l'ordre 10^{-6} avec 10% d'erreur relative, environ 10^8 simulations Monte-Carlo sont théoriquement nécessaires, ce qui n'est pas réalisable sur des systèmes complexes, où le temps de calcul d'une simulation peut être trop important et souvent rédhibitoire. De plus, le caractère cyclique des générateurs de nombres aléatoires impliquerait également une corrélation entre les échantillons générés, ce qui pourrait biaiser l'estimation. Il est donc important de mettre en place des techniques d'estimation efficaces à faible coût de simulation. On dénombre essentiellement cinq techniques d'estimation d'évènements rares pertinentes : la théorie des valeurs extrêmes, l'échantillonnage préférentiel, les méthodes multi-niveaux, les méthodes de type FORM/SORM (First/Second Order Reliability Methods), enfin l'enrichissement adapté de métamodèles. J'ai proposé dans l'article [56] un état de l'art général et une comparaison de ces méthodes appliquées à un cas d'estimation de zones de retombées d'un engin aérospatial. Les domaines d'utilisation des différentes méthodes ont ainsi pu être déterminés.
Les apports méthodologiques dans le cadre de l'estimation d'évènements rares se sont principalement concentrés, jusqu'à présent, sur les méthodes d'échantillonnage préférentiel et les méthodes multi-niveaux. L'échantillonnage préférentiel utilise une loi de tirage auxiliaire des échantillons permettant de générer des évènements rares plus efficacement. Le choix de cette loi auxiliaire a un fort impact sur l'estimation d'une probabilité car celle-ci peut diminuer très fortement la variance de l'estimation, mais également, en cas de mauvais choix de loi, dégrader les performances d'estimation que l'on aurait obtenues avec des simulations Monte-Carlo. Il n'est possible de définir une loi auxiliaire optimale de tirage des échantillons que par le biais de la connaissance *a priori* de la probabilité à estimer, comme je le décris dans l'article [91]. Cette loi peut néanmoins être approchée itérativement de manière paramétrique en optimisant les variables d'une famille donnée de lois. Il est également possible de proposer une version non paramétrique de l'algorithme, basée sur l'optimisation itérative d'un estimateur à noyaux de densité, comme je le détaille dans l'article [94]. La dépendance à l'initialisation de cet algorithme peut parfois être problématique. Une extension de celui-ci, avec une densité initiale égale à la densité des tirages de type Monte-Carlo a été proposée dans l'article [95] pour l'estimation de quantile.
Les méthodes multi-niveaux consistent en l'estimation de plusieurs probabilités conditionnelles. La probabilité cible est ensuite estimée en utilisant le théorème de Bayes. J'ai proposé un cadre générique d'application de cette méthode dans l'article [102] en collaboration avec l'INRIA Rennes (DR François Le Gland). Ces méthodes sont généralement plus coûteuses en simulations que les méthodes d'échantillonnage préférentiel mais résistent beaucoup mieux au fléau de la dimension des entrées. Des méthodes multi-niveaux ont ainsi été développées dans

l'article [103] sur un système entrées-sorties de dimension 50, où l'échantillonnage préférentiel est inefficace. Une version multidimensionnelle et multi-lois a également été détaillée dans la publication [97], afin d'estimer des probabilités de collision entre aéronefs, dans un espace aérien non contrôlé, à partir d'une base de données réelles de trajectoires d'avions de tourisme. Ces résultats détaillés dans la référence [97] sont également une avancée dans le cadre de la caractérisation des probabilités de perte de séparation entre aéronefs et également de l'insertion des drones dans le trafic aérien. Par ailleurs, il faut également considérer les difficultés liées au réglage du paramétrage des méthodes multi-niveaux. Dans les cas simplifiés, au moins 5 paramètres différents doivent être réglés de façon à minimiser la variance de la probabilité estimée. Une méthode d'optimisation de ces paramètres a été développée, dans cette optique, à l'aide du krigeage et de l'algorithme d'enrichissement du krigeage EGO (Efficient Global Optimisation). Un article sur cette approche [96] a été publié et un second article est en cours de révision.
Dans le cas de sorties multidimensionnelles, la notion de quantile est plus complexe à définir comme je le précise dans l'article [100]. Une approche possible est l'estimation d'ensembles de densités à volume minimal pour un niveau de quantile donné. Pour des quantiles extrêmes, les méthodes Monte-Carlo sont, comme expliqué précédemment, inefficaces. L'article [101] a permis de décrire un algorithme d'échantillonnage préférentiel non paramétrique qui permet d'estimer des quantiles multidimensionnels extrêmes à moindre coût de simulation. Les méthodes multi-niveaux sont également développées, pour l'estimation de quantiles multidimensionnels extrêmes. Un article est en cours de révision sur cette problématique.

Bilan : 11 publications dans des revues à comité de lecture.

1.4.2 Perspectives de recherche à court et moyen termes

Influence des paramètres du modèle boîte noire sur l'estimation d'évènements rares

Pour caractériser l'influence de paramètres d'un modèle boîte noire (paramètres des densités d'entrées, par exemple) sur une probabilité en sortie, il est également intéressant de calculer les lois des phénomènes aléatoires à la défaillance. Autrement dit, cela revient à calculer les probabilités conditionnelles des variables aléatoires du système complexe sachant que ce dernier est dans un régime de défaillance. Le calcul de ces lois permet ainsi de caractériser tous les aléas et leurs corrélations qui conduisent justement à la défaillance. Estimer la loi des paramètres d'un modèle aléatoire, sachant que ce dernier évolue dans un état critique est une piste de recherche intéressante.
Cette thématique de recherche est actuellement traitée dans le cadre de la thèse de Christelle Vergé débutée en Octobre 2012 à la suite de son stage en collaboration avec l'INRIA Bordeaux (DR Pierre Del Moral) à l'aide de techniques particulaires. Une publication dans une revue à comité de lecture a été soumise sur ce sujet.

Influence de l'erreur d'un métamodèle

Dans de nombreuses applications, il est souvent difficile de propager des incertitudes directement dans des modèles complexes. 1000 calculs entrées-sorties du modèle sont, dans un cas général, nécessaires pour mettre en place les techniques de base d'analyse des incertitudes, telles que l'analyse de sensibilité, l'estimation d'évènements rares, ce qui peut être problématique en termes de coût/temps de simulation. Pour s'affranchir de ces limites, il peut être pertinent de développer un modèle de substitution, qui imite le comportement du modèle complexe à moindre coût de calcul, et ensuite d'effectuer la propagation d'incertitudes sur celui-ci. Néanmoins, le modèle de substitution ne peut être exactement conforme au modèle réel et introduit donc des erreurs d'estimation. Il est par conséquent nécessaire d'analyser l'influence d'un métamodèle sur la propagation d'incertitudes. Quelle est la part de l'erreur due au métamodèle, dans l'erreur globale et le biais d'estimation ? Dans quelle mesure utiliser les propriétés du métamodèle en vue de caractériser l'influence de celui-ci sur la quantité d'intérêt estimée ? Une première approche mêlant krigeage et importance sampling (paramétrique et non paramétrique) vient d'être accepté dans une revue à comité de lecture [5] pour l'estimation d'une probabilité d'évènements rare (voir article 5 de la Partie III contenant une sélection de publications). Un projet mené par l'ONERA a été proposé à l'ANR (Agence Nationale de la Recherche) sur un sujet qui englobe cette problématique.

Estimation d'évènements rares dans les systèmes dynamiques

Les méthodes d'estimation d'évènements rares peuvent également être utilisées dans un cadre dynamique. Le modèle de propagation des incertitudes n'est plus un code de calcul entrées-sorties mais une chaîne de Markov. Par exemple, les incertitudes sur la position d'un aéronef le long de son plan de vol peuvent être modélisées dans un espace aérien non contrôlé par une chaîne de Markov comme nous l'avons proposé dans l'article [57]. Celles-ci sont notamment dues aux pilotes et aux conditions atmosphériques, et sont caractérisées par une erreur gaussienne dont la variance augmente avec le temps. Pour estimer une probabilité de collision entre avions dans ce cadre dynamique, les techniques de simulation Monte-Carlo ne sont pas efficaces. Néanmoins, les techniques d'estimation d'évènements rares issues des études statiques ne sont pas implémentables de manière évidente et doivent donc être adaptées. Quel politique de seuillage choisir avec les méthodes multi-niveaux ? Peut-on définir des méthodes multi-niveaux pondérées efficaces ? Comment faire le lien entre approche stochastique et déterministe (équations aux dérivées partielles) ? La thèse de Damien Jacquemart, qui a débuté en octobre 2011 sur bourse DGA-ONERA en collaboration avec l'INRIA Rennes (DR François Le Gland), commence à fournir des réponses à certaines de ces différentes questions. Une publication d'état de l'art sur ce thème a été soumise. Un article sur le réglage de méthodes particulaires dans un cadre markovien a été publié [96]. Une communication mêlant technique de splitting et sampling (splitting pondéré) vient d'être acceptée dans la Winter Simulation Conference 2013. Deux autres articles sont en cours de relecture.

Optimisation stochastique

La thèse de Dalal Madakat traite de l'optimisation du retrait de débris spatiaux à l'aide d'une navette spatiale en collaboration avec l'Université Paris Dauphine (Pr. Daniel Vanderpooten). De manière théorique, il s'agit en fait de résoudre un problème de voyageur de commerce bicritère dépendant du temps. Les deux critères d'évaluation de la performance de retrait des débris sont en effet la masse de carburant consommé et la durée de la mission. Des algorithmes de résolution de type *Branch and Bound* ont été mis en place. Ils permettent de déterminer de manière exacte la frontière de Pareto des solutions possibles comme nous le développons dans l'article [84]. Dans le cas où le nombre de débris est important ou bien lorsque la durée maximum de la mission est longue, le temps de calcul des algorithmes Branch and Bound est élevé et impose des hypothèses simplificatrices. Une approche innovante serait de considérer les techniques multi-niveaux afin de les adapter au domaine de l'optimisation. De premières études ont été effectuées et semblent obtenir des résultats satisfaisants [23]. La comparaison des résultats obtenus entre les algorithmes Branch and Bound et les algorithmes multi-niveaux pourrait s'avérer pertinente à traiter. Comment adapter le cadre d'application des méthodes multi-niveaux à un problème d'optimisation multi-critère et dépendant du temps ? Quelles hypothèses statistiques faut-il définir pour implémenter ces méthodes ?

1.5 Activités d'enseignement

En tant que vacataire aux IUT de Velizy, de Salon de Provence et à l'Institut d'Optique Graduate School, j'ai assuré des cours relatifs à l'analyse des systèmes temps réels, aux mathématiques et à la reconnaissance statistique des formes. J'ai également participé à la rédaction du document nécessaire au cours de reconnaissance des formes et de systèmes temps réels.

- **2012/2013**
  - Reconnaissance statistique des formes en M2 (Institut d'Optique Graduate School) : 2 h de CM et 6 h de TD.
- **2011/2012**
  - Mathématiques en L2 (IUT de Vélizy) : 30 h de CM et 8 h de TD.
  - Reconnaissance statistique des formes en M2 (Institut d'Optique Graduate School) : 2 h de CM et 8 h de TD.
- **2010/2011**
  - Mathématiques en L2 (IUT de Vélizy) : 45 heures de TD.
  - Reconnaissance statistique des formes en M2 (Institut d'Optique Graduate School) : 8 h de TD et 2 h de CM.
- **2009/2010**
  - Mathématiques en L2 (IUT de Vélizy) : 45 heures de TD.
  - Reconnaissance statistique des formes en M2 (Institut d'Optique Graduate School) : 8 h de TD et 2 h de CM.
- **2008/2009**
  - Mathématiques en L2 (IUT de Vélizy) : 45 heures de TD.
  - Traitement du signal numérique en L2 (IUT de Vélizy) : 20 heures de TD.

- **2006/2007**
  - Analyse de systèmes temps réels en L3 (IUT de Salon de Provence) : 90 heures de CM-TD.
- **2005/2006**
  - Analyse de systèmes temps réels en L3 (IUT de Salon de Provence) : 90 heures de CM-TD.

1.6 Activités administratives et rayonnement scientifique

Les activités administratives qu'il m'a été donné d'entreprendre s'intègrent à la fois dans un cadre contractuel, mais impliquent également une prise de responsabilité scientifique au sein de l'ONERA, ainsi qu'une participation à la vie scientifique. Ces activités peuvent être résumées en cinq points.

- Participation à 8 contrats sous financement DGA (Direction Générale de l'Armement), FUI (Fonds Unique Interministériel) et UE (Union Européenne). Rédaction des 21 rapports techniques correspondants.
- Participation en tant que membre nommé au Conseil Scientifique de la Branche Traitement de l'Information et Système (TIS) de l'ONERA pour le DCPS depuis 2009. La branche TIS emploie 250 personnes de l'ONERA dont 100 dans mon département, le DCPS. Ce Conseil Scientifique compte 16 membres nommés par les différents départements, dont le DCPS et se réunit 2 fois par an.
- Gestion, en tant que responsable, de la feuille de route AMOSYS (Analyse et MOdélisation de SYStèmes) au DCPS depuis 2011. Je participe à l'orientation des activités de recherche sur l'analyse des systèmes complexes.
- Activité de relecture pour les revues suivantes (6-7 articles par an) : Entropy, The Proceedings of the IEEE, Safety Science, Journal of Risk and Reliability, Optics Express, Aerospace Science and Technology, IET Control Theory & Applications, International Journal of Remote Sensing, Applied Optics, IEEE Transactions on Geoscience and Remote Sensing.
- Organisation, avec Fabrice Poirion (ONERA), d'une Journée Scientifique le 15/11/2012 sur les méthodes d'évènements rares au centre ONERA de Palaiseau (5 intervenants extérieurs, 90 participants dont 70 externes à l'ONERA). (http ://sites.onera.fr/MODNAT/node/9)
- Organisation, avec F. Barbaresco (Thales Air System), d'un groupe de travail ONERA-Thales-UPMC (Université Pierre-et-Marie-Curie, Paris VI) sur l'estimation d'évènements rares, depuis février 2013. Une collaboration Thales-ONERA-Université Catholique de Louvain est en cours de mise en place pour l'analyse des cas extrêmes de turbulence aérodynamique après le passage d'un avion.
- Participation à un groupe de travail ONERA-INRIA-ENSTA sur le contrôle optimal de drones pour l'évitement de collisions.
- Coordinateur d'un article à 11 auteurs pour la revue Aerospace Lab sur l'analyse de performances des systèmes complexes.

Chapitre 2

Résumé des activités de recherche

2.1 Publications dans des revues à comité de lecture

Le descriptif, situé ci-dessous, des 24 articles acceptés dans des revues à comité de lecture et auxquels j'ai participé est un aperçu de mes activités de recherche.

- J. Morio and F. Goudail, Influence of diattenuator, retarder and polarizer in polar decomposition of Mueller matrices, *Optics Letters*, 2004, Vol. 29, Issue 19, pp. 2234-2236.
- P. Réfrégier and J. Morio, Shannon entropy of partially polarized and partially coherent light with Gaussian fluctuations, *Journal of the Optical Society of America A*, 2006, Vol. 23, Issue 12, pp. 3036-3044.
- J. Morio, F. Goudail, X. Dupuis, P. Dubois-Fernandez and P. Réfrégier, Polarimetric and interferometric SAR image partition into statistically homogeneous regions based on the minimization of the stochastic complexity, *IEEE Transactions on Geoscience and Remote Sensing*, 2007, Vol. 45, Issue 11, pp. 3599-3609.
- J. Morio, P. Réfrégier, P. Dubois-Fernandez, F. Goudail and X. Dupuis, Information theory based approach for contrast analysis in polarimetric and/or interferometric SAR images, *IEEE Transactions on Geoscience and Remote Sensing*, 2008, Vol. 46, Issue 8, pp. 2185-2196.
- J. Morio and F. Muller, Onboard sensor orientation analysis with kriging method, *Acta Astronautica*, 2009, Vol. 64, Issues 2-3, pp. 374-381.
- J. Morio, P. Réfrégier, P. Dubois-Fernandez, F. Goudail and X. Dupuis, A characterization of Shannon entropy and Bhattacharyya measure of contrast in PolInSAR images, *The Proceedings of the IEEE*, 2009, Vol. 97, Issue 6, pp. 1097-1108.
- J. Morio and F. Muller, Radar measurements analysis for spatial object classification, *International Journal of Intelligent Defence Support Systems*, 2009, Vol. 2, n°2, pp. 91-104.
- J. Morio and R. Pastel, Sampling technique for launcher impact safety zone estimation, *Acta Astronautica*, 2010, Vol. 66, Issues 5-6, pp. 736-741.
- J. Morio, Importance sampling : how to approach the optimal density ?, *European Journal of Physics*, 2010, Vol. 31, n°2, pp. 41- 48.
- J. Morio, R. Pastel and F. Le Gland, An overview of importance splitting

for rare event simulation, *European Journal of Physics*, 2010, Vol. 31, n°5, pp. 1295-1303.

- J. Morio and F. Muller, Spatial object classification by radar measurements, *Aerospace Science and Technology*, 2010, Vol. 14, Issue 4, pp. 259-265.
- J. Morio, Nonparametric Adaptive Importance Sampling for the probability estimation of a launcher impact position, *Reliability Engineering & System Safety*, 2011, Vol. 96, Issue 1, pp. 178-183.
- J. Morio, R. Pastel et F. Le Gland, Estimation de probabilités et de quantiles rares pour la caractérisation d'une zone de retombée d'un engin, *Journal de la Société Française de Statistique*, 2011, Vol. 152, n°4, pp. 1-29.
- J. Morio, Influence of input PDF parameters of a model on a failure probability estimation, *Simulation Modelling Practice and Theory*, 2011, Vol. 19, Issue 10, pp. 2244-2255.
- J. Morio, Global and local sensitivity analysis methods for a physical system, *European Journal of Physics*, 2011, Vol. 32, n°6, pp 1577-1583.
- J. Morio and R. Pastel, Plug-in estimation of d-dimensional density minimum volume set of a rare event in a complex system, *Proceedings of the IMechE, Part O, Journal of Risk and Reliability*, 2012, Vol. 226, n°3, pp. 347-345.
- J. Morio, R. Pastel and F. Le Gland, Missile target accuracy estimation with importance splitting, *Aerospace Science and Technology*, 2013, Vol. 25, Issue 1, pp. 40-44.
- J. Morio, T. Lang and C. Le Tallec, Estimating separation distance loss probability between aircraft in uncontrolled airspace in simulation, *Safety Science*, 2012, Vol. 50, Issue 4, pp. 995-1004.
- J. Morio, Extreme quantile estimation with non parametric adaptive importance sampling, *Simulation Modelling Practice and Theory*, 2012, Vol. 27, pp. 76-89.
- D. Jacquemart and J. Morio, Conflict probability estimation between aircraft with dynamic Importance Splitting, *Safety Science*, 2013, Vol. 51, n°1, pp. 94-100.
- D. Madakat, J. Morio and D. Vanderpooten, Biobjective planning of an active debris removal mission, *Acta Astronautica*, 2013, Vol. 84, pp. 182-188.
- R. Pastel, J. Morio and F. Le Gland, Improvement of satellite conflict prediction reliability through use of the adaptive splitting technique, à paraître dans la revue *Proceedings of the IMechE, Part G : Journal of Aerospace Engineering*.
- J. Morio, D. Jacquemart, M. Balesdent and J. Marzat, Optimisation of interacting particle system for rare event estimation, *Computational Statistics and Data Analysis*, 2013, Vol. 66, pp. 117-128.
- M. Balesdent, J. Morio and J. Marzat, Kriging-based adaptive importance sampling algorithms for rare event estimation, *Structural Safety*, 2013, Vol. 44, pp. 1-10.

2.2 Publications en révision ou en cours de relecture

Les publications suivantes sont en cours de révision ou de relecture dans des revues à comité de lecture.

- R. Pastel, J. Morio and F. Le Gland, Extreme quantile and density level set estimation via the adaptive splitting technique, en cours de révision dans la revue *Journal of Computational Science*.
- M. Balesdent, J. Morio and J. Marzat, Recommendations for the tuning of rare event probability estimators, en cours de révision dans la revue *Reliability Engineering & System Safety*.
- J. Morio, M. Balesdent, D. Jacquemart and C. Vergé, A survey of rare event estimation methods for static input-output models, en cours de relecture dans la revue *Information Sciences*.
- D. Jacquemart, J. Morio and F. Le Gland, A comparison of rare event simulation methods for time-continuous Markov process, en cours de relecture dans la revue *Statistics and Computing*.
- D. Jacquemart and J. Morio, Adaptive interacting particle system for rare events, en cours de relecture dans la revue *Methodology and computing in Applied Probability*.
- C. Vergé, J. Morio and P. Del Moral, Interacting island particle algorithm for safety analysis, en cours de relecture dans la revue *Structural Safety*.

2.3 Autres publications dans des revues

J'ai également participé à un article et été responsable de rédaction d'un second dans la revue à comité de lecture externe, éditée par l'ONERA, *Aerospace Lab*.

- S. Defoort, M. Balesdent, P. Klotz, P. Schmollgruber, J. Morio, J. Hermetz, C. Blondeau, G.Carrier, N. Bérend, Multidisciplinary Aerospace System Design : Principles, Issues and Onera Experience, *Aerospace Lab*, N°4, May 2012.
- J. Morio, H. Piet-Lahanier, F. Poirion, J. Marzat, C. Seren, S. Bertrand, Q. Brucy, R. Kervarc, S. Merit, P. Ducourtieux, R. Pastel, An Overview of Probabilistic Performance Analysis Methods for Large Scale and Time-Dependent Systems, *Aerospace Lab*, N°4, May 2012.

2.4 Liste des communications avec comité de sélection

Les communications suivantes ont été effectuées dans des colloques avec comité de sélection.

- J. Morio, P. Réfrégier, F.Goudail, P Dubois-Fernandez, and X. Dupuis, Influence of polarimetry and interferometry in the partition of PolInSAR images, EUSAR 2006, Dresden 16-18/05/2006.
- J. Morio, P. Réfrégier, P Dubois-Fernandez, F.Goudail and X. Dupuis, Application of information theory measures to polarimetric and interferometric SAR images, 5th International Conference on Physics in Signal & Image Processing, PSIP 2007, Mulhouse 30/01/2007-02/02/2007.
- R. Pastel, J. Morio and H. Piet-Lahanier, Rare event probability estimation through a kriging algorithm, 7th International Workshop on Rare Event Simulation, RESIM 2008, Rennes 24-26/09/2008.
- R. Pastel, J. Morio and F. Le Gland, Estimating satellite versus debris collision probabilities via the adaptive splitting technique, RESIM, Cambridge, 21-22/06/2010.
- R. Pastel, J. Morio and F. Le Gland, Representing spatial distribution via the Adaptive Importance Splitting technique and Isoquantile Curves, EMS 2010, Piraeus, 17-22/08/2010.
- R. Pastel, J. Morio and F. Le Gland, Estimating Satellite Versus Debris Collision Probabilities via the Adaptive Splitting Technique, 3th International Conference on Computer modeling and simulation, ICCMS 2011, Mumbai 7-9/01/2011.
- R. Pastel, J. Morio and F. Le Gland, Satellites collision probabilities : switching from numerical integration to the adaptive splitting technique, 4th European Conference for Aerospace Sciences, EUCASS 2011, Saint Petersburg, 4-8/07/2011.
- D. Madakat, J. Morio and D. Vanderpooten, Planification biobjective d'une mission de retrait de débris spatiaux, 13ème congrès de la ROADEF, Angers, 11-13/04/2012.
- D. Madakat, J. Morio and D. Vanderpooten, An ELECTRE TRI method for space debris sorting, 54th CORS/MOPGP'2012, Niagara Falls, 11-13/06/2012.
- D. Jacquemart, F. Le Gland and J. Morio, A combined importance splitting and sampling algorithm for rare event estimation, accepté à la Winter Simulation Conference 2013.

2.5 Conférences sur invitations et séminaires

Les présentations suivantes ont été effectuées dans des séminaires, parfois sur invitation.

- J. Morio, X. Dupuis, P Dubois-Fernandez, F. Goudail and P. Réfrégier, Polarimetric parameter estimation and classification using an automatic partition algorithm, 6th DLR-CNES Workshop Information Extraction and Scene Understanding for Meter Resolution Images, Oberpfaffenhofen, 15-17/11/2005.
- J. Morio, P. Réfrégier, F. Goudail, P. Dubois-Fernandez et X. Dupuis, Analyse de l'entropie de Shannon dans les images polarimétriques et interférométriques radar, Journée Analyse d'Images Multivariées, Paris, 23/11/2006.
- J. Morio, Estimation d'évènements rares par importance sampling non pa-

ramétriques, JSO évènements rares, Châtillon, 8/12/2009 (Invité).
- J. Morio, Application de méthodes d'évènements rares pour l'estimation de zones de retombées lanceurs, Journée scientifique "ARF Stochastique 2010", Châtillon, 11/02/2010 (Invité).
- R. Pastel, J. Morio and F. Le Gland, From satellite versus debris collision probabilities to the adaptive splitting technique, Rare Events Simulation Workshop, Bordeaux, octobre 2010.
- J. Morio, Importance sampling et splitting, Formation CSDL (Complex System Design Lab), Clamart, 12-13/02/2012 (Invité).
- J. Morio et M. Balesdent, Quelques voies d'améliorations en importance sampling et splitting, Journée scientifique évènements rares, Palaiseau, 15/11/2012 (Invité).

2.6 Liste des activités d'encadrement d'étudiants en master 2

L'encadrement à 100% de cinq étudiants en stage de master 2 a été une expérience enrichissante, que je détaille ci-dessous.

- Rudy Pastel, *Application du krigeage pour l'estimation de densité de probabilité et la simulation d'évènements rares*, 3ème année ENSTA, 2008.
- Alizée Bertrand, *Analyse de la population des objets spatiaux par des graphes*, 3ème année ENSHEEIT, 2009.
- Naushad Remtoula, *Caractérisation statistique du risque de collision dans le trafic aérien*, 3ème année ENSIMAG, 2010.
- Damien-Barthélémy Jacquemart, *Estimation de la probabilité de collision d'avions par méthodes d'évènements rares dynamiques : développement de méthodes d'importance splitting*, Master 2 probabilités et statistiques de l'Université d'Aix-Marseille I, 2011.
- Christelle Vergé, *Estimation du risque de collision entre deux objets spatiaux par méthodes d'événements rares*, Master 2 Modélisation Statistique et Stochastique de l'université de Bordeaux I, 2012.

2.7 Liste des activités d'encadrement d'étudiants en thèse

Ma participation à l'encadrement des 4 thèses suivantes est un point primordial de mon travail de recherche.

- Encadrement (50%) de la thèse ONERA de Rudy Pastel (2008-2011), effectuée à l'ONERA, dont le sujet est l'*Estimation de probabilités d'évènements rares et de quantiles extrêmes. Applications dans le domaine aérospatial.* Cette thèse réalisée en collaboration avec François le Gland (Directeur de thèse) de l'INRIA Rennes, a été soutenue le 14/02/2012 à l'Université de Rennes 1 dans le cadre de l'école doctorale MATISSE. Rudy Pastel a reçu

le Prix des Doctorants de la branche TIS de l'ONERA en 2011 et est actuellement en post-doctorat à BASF Ludwigshafen en Allemagne.

- Encadrement (50%) de la thèse ONERA de Dalal Madakat (2009-2012), effectuée à l'ONERA, ayant pour sujet l'*Optimisation du retrait de débris spatiaux*, en collaboration avec Daniel Vanderpooten (Directeur de thèse) de l'Université Paris Dauphine dans le cadre de l'école doctorale EDDIMO. La présoutenance de cette thèse est prévue pour juin 2013. Dalal Madakat est actuellement ATER à l'Université Paris Dauphine.
- Encadrement (50%) de la thèse ONERA/DGA de Damien-Barthélémy Jacquemart (2011-2014), effectuée à l'ONERA, ayant pour sujet l'*Estimation de la probabilité de collision d'avions par méthodes d'évènements rares dynamiques*, en collaboration avec François le Gland (Directeur de thèse) de l'INRIA Rennes dans le cadre de l'école doctorale MATISSE.
- Encadrement (50%) de la thèse ONERA/CNES de Christelle Vergé (2012-2015), effectuée à l'ONERA, ayant pour sujet l'*Estimation du risque de collision entre deux objets spatiaux par méthodes d'événements rares* en collaboration avec Pierre Del Moral (Directeur de thèse) de l'INRIA Bordeaux dans le cadre de l'école doctorale de l'Ecole Polytechnique.

2.8 Liste des activités effectuées en relation avec le milieu industriel

2.8.1 Participation à des contrats de recherche

Mes activités sur la caractérisation des systèmes complexes sont essentiellement développées dans le cadre de contrats de recherche pour le domaine de la défense et de l'aérospatiale.

- DGA Expertise de la Défense Aérienne Elargie (150 keuros).
- DGA EXpertise NATO Air Command and Control System (200 keuros).
- DGA Plan d'Etude Amont Rénovation à mi-vie ASMPA (200 keuros).
- DGA Etude NAVigation GUIdage 2 (75 keuros).
- DGA EXpertise DROne 2 (75 keuros).
- FUI Complex System Design Lab (125 keuros).
- UE HYPOW Protection of Critical Infrastructures against High Power Microwave Threats (60 keuros).
- UE P^{2}ROTECT Prediction, Protection & Reduction of OrbiTal Exposure to Collision Threats (150 keuros).

2.8.2 Liste des rapports techniques internes

21 rapports techniques internes provenant des études contractuelles listées précédemment ont également été rédigés. Ces rapports techniques sont en très grande majorité non accessibles pour des raisons de confidentialité interne et de protection du secret de la Défense Nationale.

Deuxième partie

Contributions de recherche

Contexte de la recherche

Les activités de recherche que je mène au sein de l'ONERA ont pour principale caractéristique d'être particulièrement variés. En effet, elles intègrent différents domaines, telles que les statistiques, l'optimisation ou encore la recherche opérationnelle, mais elles demeurent néanmoins en lien avec les domaines de l'aéronautique, l'aérospatiale ou la défense. Ce mémoire de synthèse est centré sur la thématique principale de mes activités de recherches, à savoir la caractérisation des évènements rares dans les systèmes complexes. Ces travaux sont le fruit d'une conjoncture académique et industrielle. En effet, ils bénéficient, d'une part, d'une collaboration universitaire avec François Le Gland (DR INRIA Rennes) et Pierre Del Moral (DR INRIA Bordeaux) dans le cadre du co-encadrement de 3 thèses. D'autre part, ces travaux demeurent en lien avec les problématiques des industriels et des acteurs étatiques dans un cadre contractuel, tels que MBDA-Missile Systems, la DGA, Thales Air Systems ou encore le CNES (Centre National d'Etudes Spatiales). Ce manuscrit a donc pour but de synthétiser l'ensemble des contributions que j'ai pu apporter à ce thème de recherche qu'est la caractérisation des évènements rares dans les systèmes complexes, qu'elles soient méthodologiques ou plus ancrées dans des projets industriels.

L'évaluation de performances de systèmes complexes et des risques qui y sont associés nécessite des techniques d'estimation de plus en plus performantes. En effet, dans le secteur spatial, celui de la défense ou de la finance, comme dans de nombreux autres, on tend à caractériser et par conséquent à probabiliser les risques inhérents à l'application de ces systèmes complexes. C'est pourquoi les domaines d'application concernant la sûreté de fonctionnement sont très étendus. Les études sur l'estimation d'évènements rares permettent, par exemple, de répondre à des questions telles que : « Quelle est la probabilité qu'un missile retombe à plus de Z mètres de sa cible ? Comment déterminer la zone de retombée à 99% d'un étage de lanceur spatial ? Quelle est la probabilité de collision entre deux avions dans l'espace aérien ? Quelle est la probabilité de collision entre un débris spatial et un satellite ? ». Mais pour répondre à ces interrogations, les méthodes d'estimation et d'intégration les plus couramment utilisées, comme les simulations Monte-Carlo ou Quasi Monte-Carlo, entre autres, n'aboutissent pas à des résultats suffisamment précis. En effet, le nombre d'échantillons disponible est souvent insuffisant pour calculer de manière précise des probabilités aussi faibles que celles que l'on souhaite estimer. Il est donc nécessaire de développer des techniques adaptées à l'estimation de telles probabilités.
Les différentes contributions issues de mes recherches sur l'estimation des évènements rares ainsi que leurs perspectives de recherche associées sont développées dans les chapitres suivants.

Chapitre 1

Comparaison et synthèse des méthodes d'estimation d'évènements rares

1.1 Choix de modélisation

Préciser le cadre de la modélisation détaillée dans la suite du manuscrit est en premier lieu indispensable afin de développer un contexte précis et efficace de la recherche sur l'estimation d'évènements rares.
Considérons une variable aléatoire $\mathbf{X}$ de dimension d avec une densité de probabilité (PDF) h_0, une fonction positive scalaire $\phi : \mathbb{R}^d \to \mathbb{R}$ et enfin un seuil S scalaire. Les différentes composantes de $\mathbf{X}$ seront notées par la suite $\mathbf{X} = (X^1, X^2, ..., X^d)$. La fonction ϕ est supposée déterministe et indépendante du temps. Celle-ci représente par exemple un code de calcul numérique entrées-sorties. Ce type de modèle est très général et est notamment utilisé dans de nombreuses applications, comme dans les références [64], [103], [6], [50], [71] et [141]. La sortie de la fonction ϕ, définie par $Y = \phi(\mathbf{X})$, est supposée aléatoire, de PDF inconnue g et de dimension 1. Notons que le cas des processus stochastiques n'entre pas dans le cadre de ce modèle et n'est donc pas traité dans ce manuscrit. Ce sujet est néanmoins actuellement analysé dans le cadre d'une thèse ONERA/DGA que je co-encadre en collaboration avec l'INRIA Rennes (Voir, par exemple, article 2 de la Partie III contenant une sélection de publications). Mes recherches sont essentiellement focalisées sur l'estimation de deux quantités d'intérêt de la sortie Y :

- la probabilité $\mathbb{P} = P(\phi(\mathbf{X}) > S)$ de dépassement d'un seuil S par la sortie Y.
- le α-quantile q_α de la variable Y défini par $\alpha = \int_{-\infty}^{q_\alpha} g(y)dy$

Différentes techniques d'estimation de quantile et de probabilité ont été développées lorsque le budget de simulations N disponible est peu important par rapport à la probabilité à estimer, c'est à dire lorsque $\mathbb{P} < \frac{1}{N}$. Mes collaborateurs et moi-même avons rédigé un état de l'art des techniques de simulation d'évènements rares, qui est en cours de révision, afin de tenter de répondre à la question suivante : "Quelle est la méthode d'estimation d'évènements rares la mieux adaptée au problème traité ?". Nous avons également publié des comparaisons de résul-

tats de simulations dans l'article [56], plus particulièrement sur les techniques de simulations d'évènements rares. Les sections suivantes proposent un bref bilan de ce panorama des méthodes d'estimation de probabilités faibles. Enfin, une synthèse originale sous forme de questions-réponses et permettant une sélection des méthodes d'évènements rares adaptés à un modèle ϕ donné est proposé à la fin de ce chapitre.

1.2 Présentation des méthodes d'estimation d'évènements rares

Les méthodes d'estimation d'évènements rares peuvent tout d'abord être classées en 4 grandes catégories citées ci-dessous et développées dans les sections suivantes :

- les méthodes de simulation d'évènements rares,
- les techniques statistiques,
- les techniques d'estimation basées sur des approximations paramétriques issues des études de fiabilité,
- les approches fondées sur la caractérisation d'un modèle de substitution de la fonction ϕ.

1.2.1 Techniques de simulation d'évènements rares

Méthodes Monte-Carlo

Une manière simple d'estimer $\mathbb{P}$ est d'utiliser les méthodes Monte-Carlo comme notamment dans les références [134], [90], [135], [122], [109], et [43]. Le principe de celles-ci est de générer N échantillons indépendants et identiquement distribués (i.i.d.) $\mathbf{X}_1, ..., \mathbf{X}_N$ selon la PDF h_0 et d'évaluer les échantillons, en sortie de la fonction ϕ, $\phi(\mathbf{X}_1), ..., \phi(\mathbf{X}_N)$. La probabilité $\mathbb{P}$ est alors estimée par :

$$\hat{\mathbb{P}}^{CMC} = \frac{1}{N} \sum_{i=1}^{N} \mathbf{1}_{\phi(\mathbf{X}_i)>S} \tag{1.1}$$

où $\mathbf{1}_{\phi(\mathbf{X}_i)>S}$ prend pour valeur 1 si $\phi(\mathbf{X}_i) > S$ et 0 si ce n'est pas le cas. La convergence de l'estimateur est assurée par la loi des grands nombres. Un indicateur possible de l'efficacité de l'estimation par méthode Monte-Carlo est donnée par son erreur relative, définie comme le rapport entre l'écart-type et la moyenne de l'estimation. L'erreur relative de l'estimation Monte-Carlo s'exprime en fonction de $\mathbb{P}$ par :

$$\frac{\sigma_{\hat{\mathbb{P}}^{CMC}}}{\mathbb{P}} = \frac{1}{\sqrt{N}} \frac{\sqrt{\mathbb{P} - \mathbb{P}^2}}{\mathbb{P}} \tag{1.2}$$

Lorsque l'évènement est rare, la probabilité $\mathbb{P}$ tend vers 0 et on observe par conséquent :

$$\lim_{\mathbb{P}\to 0} \frac{\sigma_{\hat{\mathbb{P}}^{CMC}}}{\mathbb{P}} = \lim_{\mathbb{P}\to 0} \frac{1}{\sqrt{N\mathbb{P}}} = +\infty \tag{1.3}$$

Pour estimer une probabilité $\mathbb{P}$ de l'ordre de 10^{-4} avec une erreur relative de 10%, 10^6 échantillons sont requis, ce qui est souvent rédhibitoire en temps de calcul, dans la pratique, malgré la simplicité de mise en place de la technique.

Importance sampling

Une alternative simple aux méthodes Monte-Carlo est l'échantillonnage d'importance ou importance sampling (IS) comme expliqué dans les articles [111], [72], [120], [40], [10], [13], [51], [17] et [88]. L'idée directrice de l'importance sampling est de générer les échantillons $\mathbf{X}_1, ..., \mathbf{X}_N$ avec une PDF auxiliaire h capable de générer des échantillons tels que $\phi(\mathbf{X}) > S$. Le changement de densité de tirage des échantillons est pris en compte dans l'estimateur de probabilité en ajoutant un poids. L'estimateur IS de probabilité $\widehat{\mathbb{P}}^{IS}$ est donné par l'équation suivante :

$$\widehat{\mathbb{P}}^{IS} = \frac{1}{N}\sum_{i=1}^{N} \mathbf{1}_{\phi(\mathbf{X}_i)>S} \frac{h_0(\mathbf{X}_i)}{h(\mathbf{X}_i)} \tag{1.4}$$

Le terme $\widehat{\mathbb{P}}^{IS}$ est un estimateur non biaisé de la probabilité $\mathbb{P}$. Sa variance s'exprime alors de la manière suivante

$$Var\left(\widehat{\mathbb{P}}^{IS}\right) = \frac{Var\left(\mathbf{1}_{\phi(\mathbf{X})>S} w(\mathbf{X})\right)}{N} \tag{1.5}$$

avec $w(\mathbf{X}) = \frac{h_0(\mathbf{X})}{h(\mathbf{X})}$. La variance de $\widehat{\mathbb{P}^{IS}}$ dépend fortement du choix de $h(.)$. Si h est bien choisie, l'estimateur d'importance sampling peut avoir une variance très faible contrairement aux méthodes Monte-Carlo et inversement. L'objectif de l'importance sampling est de minimiser la variance de l'estimation et il est, par conséquent, possible de définir une densité auxiliaire optimale qui minimise $Var\left(\widehat{\mathbb{P}}^{IS}\right)$. Comme les variances sont des quantités positives, la densité auxiliaire optimale h_{opt} est déterminée en annulant la variance dans l'équation 1.5. La densité h_{opt} s'exprime alors, comme décrit dans l'article [20], de la manière suivante :

$$h_{opt} = \frac{\mathbf{1}_{\phi(\mathbf{X})>S} h_0}{\mathbb{P}} \tag{1.6}$$

La densité auxiliaire optimale h_{opt}, en raison du fait qu'elle dépende de la probabilité recherchée $\mathbb{P}$, n'est donc pas utilisable en pratique. Néanmoins, h_{opt} peut être intéressante pour déterminer une PDF auxiliaire d'échantillonnage efficace. Des changements de densités de type changement d'échelle, translation ou à base d'exponential twisting, comme décrites dans les articles [26] et [59], sont par ailleurs très efficaces dans certains cas particuliers de fonction ϕ. Dans le cas général, une procédure itérative est nécessaire pour déterminer une densité auxiliaire efficace. Celle-ci peut être appliquée sur les paramètres d'une famille de densités comme dans le cas de la méthode de cross-entropy dans les références [125] et [29], ou bien sur des estimateurs à noyaux de densité dans le cadre de la méthode NAIS (Non-parametric Adaptive Importance Sampling) dans les articles [143], [108] et [107]. Les méthodes paramétriques résistent mieux à l'augmentation de la dimension d des entrées que les méthodes NAIS. Cependant, le choix de la famille de densités doit être adapté aux caractéristiques de la densité h_{opt} qui n'est pas connue à l'avance.

Importance splitting

La simulation multi-niveaux, décrite dans les articles [63], [124], [31], [3], [4], [15], [14] et [25], également appelée importance splitting, subset simulation,

sequential Monte Carlo ou subset sampling, est une méthode d'estimation d'espérances mathématiques tout particulièrement utile dans le cadre de simulations d'évènements rares et des densités de probabilités conditionnelles. Plutôt que de s'attaquer directement au problème complexe à résoudre, on le réexprime en une suite de problèmes plus simples, de manière à ce que l'enchaînement de leurs solutions permette de résoudre le problème initial. Définissons l'ensemble $\mathbf{A} = \{\mathbf{X} \in \mathbb{R}^d | \phi(\mathbf{X}) > S\}$. L'objectif est donc d'estimer la probabilité $\mathbb{P} = P(\mathbf{X} \in \mathbf{A})$. Le principe de l'importance splitting est de déterminer des sur-ensembles de $\mathbf{A}$ et ensuite d'estimer $P(\mathbf{X} \in \mathbf{A})$ à l'aide de probabilités conditionnelles. Soient $\mathbf{A}_0 = \mathbb{R}^d \supset \mathbf{A}_1 \supset ... \supset \mathbf{A}_{n-1} \supset \mathbf{A}_n = \mathbf{A}$, une suite décroissante de sous-ensembles de $\mathbb{R}^d$, ayant pour plus petit élément $\mathbf{A} = \mathbf{A}_n$. La probabilité recherchée $\mathbb{P}$ peut s'écrire de la manière suivante :

$$P(\mathbf{X} \in \mathbf{A}) = \prod_{k=1}^{n} P(\mathbf{X} \in \mathbf{A}_k | \mathbf{X} \in \mathbf{A}_{k-1}) \tag{1.7}$$

où $P(\mathbf{X} \in \mathbf{A}_k | \mathbf{X} \in \mathbf{A}_{k-1})$ est la probabilité que $\mathbf{X} \in \mathbf{A}_k$ sachant $\mathbf{X} \in \mathbf{A}_{k-1}$. Un choix optimal des ensembles $\mathbf{A}_k$, $k = 0, ..., n$ est obtenu quand $P(\mathbf{X} \in \mathbf{A}_k | X \in \mathbf{A}_{k-1}) = p$, où p est une constante, c'est-à-dire quand toutes les probabilités conditionnelles sont égales. La variance de $\mathbb{P}$ est en effet minimale dans cette configuration, comme le décrivent les articles [73] et [21]. Si la variance de l'estimation de chaque probabilité conditionnelle est faible, alors l'importance splitting sera plus performant que les méthodes Monte-Carlo. Il est complexe de définir à l'avance les ensembles $\mathbf{A}_k$ pour une fonction ϕ quelconque. Une version adaptative du splitting a été proposée dans l'article [25] et permet de s'affranchir de cette difficulté. Les méthodes de splitting sont très adaptées pour des évènements de probabilité inférieure à 10^{-4} et sont toujours efficaces lorsque la dimension d des entrées est importante. Néanmoins, le budget de simulation nécessaire pour l'application de l'algorithme est assez important, au minimum 10000 échantillons, et le réglage des paramètres correspondant est très sensible. De plus, le splitting est essentiellement adapté aux cas où la densité de $\mathbf{X}$ est gaussienne. Si ce n'est pas le cas, une transformation des entrées (transformation de Rosenblatt définie dans l'article [123], de Nataf définie quant à elle dans l'article [106],...) peut s'avérer nécessaire.

Autres algorithmes de simulation

Dans des cas simples de fonction ϕ et pour des familles usuelles de densité de $\mathbf{X}$, les techniques basées sur l'utilisation de variables antithétiques, comme c'est le cas dans les références [51] et [128], ou de variables de contrôles, dans les reférences [89] et [47], permettent également de réduire, en simulation, la variance de l'estimateur de probabilité. Leur potentielle utilisation dans un cadre général est *a priori* limitée.

1.2.2 Techniques statistiques d'estimation d'évènements rares

Les techniques statistiques permettent d'estimer ou de borner la probabilité $\mathbb{P}$ pour un jeu fixé d'échantillons $\phi(\mathbf{X}_1), ..., \phi(\mathbf{X}_N)$. Les approches statistiques principalement utilisées sont la théorie des valeurs extrêmes et la théorie des grandes déviations.

Théorie des valeurs extrêmes

Soient $\phi(\mathbf{X}_1), ..., \phi(\mathbf{X}_N)$ une série d'échantillons i.i.d. de la variable Y, G_Y la fonction de répartition de Y et $M_N = \max(\phi(\mathbf{X}_1), ..., \phi(\mathbf{X}_N))$ le maximum des observations. L'évolution de la loi des maxima peut s'obtenir de la manière suivante :

$$\begin{aligned} P(M_N \leq y) &= P(\max(\phi(\mathbf{X}_1), ..., \phi(\mathbf{X}_N)) \leq y) \\ &= P(\phi(\mathbf{X}_1) \leq y, \phi(\mathbf{X}_2) \leq Y, ..., \phi(\mathbf{X}_N) \leq y) = (G_Y(y))^N \end{aligned}$$

Dans le cas d'un échantillon i.i.d, la loi des maxima peut être déterminé dès lors que G_Y est connue. Le résultat fondamental de la théorie des valeurs extrêmes est que, sous des conditions de régularité très générales, la loi des valeurs extrêmes appartient à une famille composée de trois lois de probabilités, comme cela est démontré dans les articles [42], [75], [76], [28] et [39]. S'il existe deux suites de nombres réels (a_N, b_N), avec $a_N > 0$ et une fonction de distribution H_0 non dégénérée telles que :

$$P\left(\frac{M_N - b_N}{a_N} \leq t\right) = G^N(a_N t + b_N) \to H_0(t), \qquad N \to +\infty \tag{1.8}$$

pour tout $t \in \mathbb{R}$, alors H_0 ne peut être qu'une distribution de type Fréchet, Gumbel ou Weibull négative. Ces trois distributions peuvent en fait se regrouper en une seule famille paramétrique : la distribution généralisée des valeurs extrêmes (Generalized Extreme Value = GEV en anglais). La distribution GEV se définit de la manière suivante :

$$H_0(t) = \exp\left(-\left[1 + \varepsilon\frac{t-\mu}{\sigma}\right]^{-\frac{1}{\varepsilon}}\right)$$

où $(\mu, \sigma, \varepsilon)$, avec $\sigma > 0$, sont respectivement les paramètres de localisation, d'échelle et de forme de la distribution GEV. Une version multivariée de la théorie des valeures extrêmes a également été définie [44].
On caractérise ainsi la densité de probabilité du maximum d'un échantillon issu d'une variable aléatoire. Néanmoins, pour un jeu de données réelles, la détermination de la GEV correspondante n'est pas évidente. En effet, il faut séparer en blocs aléatoires ou blocs maxima, les données d'entrées puis déterminer les valeurs maxima de chacun des blocs. La densité GEV est alors estimée à partir de l'échantillon de ces valeurs maxima.
Cependant, le problème que nous étudions a trait à l'estimation d'une probabilité de dépassement de seuil, comme le présentent également les articles [27], [87], [116] et [24]. Dans ce cadre, l'approche POT (Peaks Over Threshold) est la plus couramment utilisée dans les applications proposées que l'on peut observer dans l'ensemble des ouvrages ayant pour thème les valeurs extrêmes. Selon cette approche, considérons $\mu_{end} = \sup(y \in \mathbb{R} | G_Y(y) < 1) \leq \infty$, la borne supérieure du support de G_Y. La densité de la variable Y, conditionnée à dépasser un seuil μ, approche une densité de probabilité de Pareto généralisée (ou GPD, pour *Generalized Pareto Distribution*) quand μ tend vers μ_{end}. En effet, si Y vérifie l'équation (1.8) développée ci-dessus, alors on peut en déduire :

$$P(Y \leq t | Y > \mu) \to H(t), \qquad \mu \to \mu_{end}$$

$$H(t) = 1 - \left(1 + \varepsilon \frac{t-\mu}{\sigma}\right)^{-\frac{1}{\varepsilon}}$$

où $(\mu, \sigma, \varepsilon)$ sont respectivement les paramètres de localisation, d'échelle et de forme de la distribution GPD avec $\sigma > 0$. Si l'on connaît $H(t)$ et $P(Y > \mu)$ pour la série $\phi(\mathbf{X}_1), ..., \phi(\mathbf{X}_N)$, on est alors capable d'estimer les quantiles et les probabilités de la variable aléatoire Y au delà du seuil μ. A partir des données, il faut donc estimer les différents paramètres de la GPD $(\mu, \sigma, \varepsilon)$ et différentes approches sont alors disponibles.
La théorie des valeurs extrêmes peut être appliquée sans rééchantillonnage et pour de faibles valeurs de N, contrairement aux techniques de simulation. Néanmoins, l'estimation des paramètres $(\mu, \sigma, \varepsilon)$ de la GPD est souvent complexe et parfois peu précise. Lorsqu'un rééchantillonnage est possible, ces méthodes s'avèrent souvent moins efficaces que les techniques de simulations.

Théorie des grandes déviations

La théorie des grandes déviations caractérise le comportement asymptotique de suites de lois de probabilités, comme cela est décrit dans les références [33] et [34]. D'un point de vue plus détaillé, elle permet d'analyser la manière dont une suite de lois de probabilité dévie de son comportement habituel défini par la loi des grands nombres. La théorie des grandes déviations est notamment utilisée afin d'évaluer la convergence de certains algorithmes d'estimation, comme cela est présenté dans les articles [37], [126], [48], [30] et [32].
Soient $D_N = J(\phi(\mathbf{X}_1), ..., \phi(\mathbf{X}_N))$ une suite de variables aléatoires indexées par N avec J une fonction scalaire continue ayant pour espérance D, et la suite associée $V_N = D_N - D$. V_N satisfait le principe des grandes déviations avec un taux continue I si la limite suivante existe :

$$\lim_{N\to\infty} \frac{1}{N} \log[P(\mid V_N \mid > \gamma)] = -I(\gamma) \tag{1.9}$$

L'existence de cette limite implique pour de fortes valeurs de N que :

$$P(\mid V_N \mid > \gamma) \approx \exp\left(-NI(\gamma)\right) \tag{1.10}$$

La probabilité $P(\mid V_N \mid > \gamma)$ décroît exponentiellement à mesure que N augmente, à un taux qui dépend de γ. Cette approximation est un résultat fondamental de la théorie des grandes déviations. Si cette limite n'existe pas, alors $P(\mid V_N \mid > \gamma)$ a un comportement singulier ou décroît à une vitesse plus rapide qu'exponentielle. Si la limite est égale à 0, alors la probabilité $P(\mid V_N \mid > \gamma)$ décroît avec N plus lentement que le terme $\exp(-Na)$ avec $a > 0$. Le calcul du taux I est souvent complexe mais peut être obtenu par le biais du théorème de Gärtner-Ellis, tel que dans l'article [138]. Dans de nombreuses situations, I correspond à une divergence de Kullback-Leibler.
La théorie des grandes déviations ne peut pas être appliquée directement à l'estimation de la probabilité $\mathbb{P}$ en pratique car la densité de Y est inconnue *a priori*. Lorsque V_N est une somme de N variables aléatoires identiques, la théorie des grandes déviations peut néanmoins être utile pour estimer des bornes de variation de la probabilité $P(\mid V_N \mid > \gamma)$. La caractérisation de la déviation d'un estimateur Monte-Carlo est un exemple direct de ce cas particulier.

1.2.3 Techniques d'estimation basées sur des approximations paramétriques issues des études de fiabilité

FORM/SORM

FORM/SORM (First/Second Order Reliability Methods), présentés dans les articles [142], [12], [85] et [74], sont des algorithmes approximant la région limite de défaillance où $\phi(\mathbf{X}) > S$ dans l'espace des entrées $\mathbf{X}$, de manière paramétrique. Ces méthodes sont souvent utilisées, notamment dans le domaine de la fiabilité des matériaux. Pour appliquer cette procédure, la PDF de $\mathbf{X}$ doit être gaussienne, centrée et réduite. Si ce n'est pas le cas, une transformation des entrées (transformation de Rosenblatt, de Nataf...) s'avère nécessaire. Nous supposerons par la suite que $\mathbf{X}$ est une densité gaussienne, centrée et réduite. Les méthodes FORM/SORM imposent, en premier lieu, de déterminer le point de défaillance le plus probable dans l'espace des entrées. Cela correspond au point de la surface limite de défaillance $S - \phi(\mathbf{X}) = 0$ dans l'espace des entrées, ayant une distance minimale avec l'origine. Ce point nommé β est donc défini par :

$$\beta = \underset{x}{\operatorname{argmin}} \mid\mid \mathbf{X} \mid\mid \tag{1.11}$$

sous la contrainte $S - \phi(\mathbf{X}) = 0$ et où $||\ .\ ||$ correspond à la norme euclidienne. Plusieurs algorithmes d'estimation de β ont notamment été proposés dans les références [52], [114], [118] et [41].
Il faut enfin approcher la surface $S - \phi(\mathbf{X}) = 0$ au point de défaillance le plus probable β à l'aide d'une droite dans le cas de FORM ou d'un paraboloïde dans le cas de SORM, tel que décrit dans l'artcicle [11]. La probabilité est alors estimée dans le cas de FORM par :

$$\hat{\mathbb{P}}^{FORM} = \Omega(- \mid\mid \beta \mid\mid) \tag{1.12}$$

où Ω est la fonction de répartition d'une PDF gaussienne, centrée et réduite.
Les méthodes FORM/SORM requièrent un budget N de simulations peu important pour obtenir des résultats satisfaisants. Néanmoins, les différentes hypothèses de FORM/SORM rendent celles-ci peu robustes aux non-linéarités de la fonction ϕ ainsi que dans les cas où les entrées de la fonction ϕ ont une forte dimension. De plus, les caractéristiques de h_{opt} peuvent également influencer la précision de l'estimation donnée par FORM/SORM.

Line sampling

L'idée sous-jacente de l'échantillonnage de lignes ou line sampling (LS), décrite dans les articles [69], [70] et [131], est d'utiliser des lignes plutôt qu'un échantillonnage aléatoire pour estimer les entrées $\mathbf{X}$ qui aboutissent à $\phi(\mathbf{X}) > S$. De manière équivalente à FORM/SORM, pour appliquer cette procédure, la PDF de $\mathbf{X}$ doit être gaussienne, centrée et réduite. Si ce n'est pas le cas, une transformation des entrées (transformation de Rosenblatt, de Nataf...) s'avère nécessaire. Nous supposerons par la suite que $\mathbf{X}$ est une densité gaussienne, centrée et réduite. L'ensemble $\mathbf{A} = \{\mathbf{X} \in \mathbb{R}^d | \phi(\mathbf{X}) > S\}$ peut s'exprimer de la manière suivante :

$$\mathbf{A} = \{\mathbf{X} \in \mathbb{R}^d | X^1 \in \mathbf{A}_1(\mathbf{X}^{-1})\}. \tag{1.13}$$

où l'ensemble $\mathbf{A}_1(\mathbf{X}^{-1})$ est défini sur $\mathbb{R}$ et dépend de $\mathbf{X}^{-1} = (X^2, X^3, ..., X^d)$. Des ensembles similaires à $\mathbf{A}_1$ peuvent être définis par rapport à n'importe quelle direction dans l'espace d'entrée, quelque soit l'ensemble $\mathbf{A}$. La probabilité $\mathbb{P}$ peut alors être décrite par les équations suivantes :

$$\begin{aligned}\mathbb{P} &= \int_{\mathbb{R}^d} \mathbf{1}_{\phi(\mathbf{X})>S} h_0(\mathbf{X}) d\mathbf{X} \\ &= \int_{\mathbb{R}^d} \mathbf{1}_{\mathbf{X}\in\mathbf{A}} h_0(\mathbf{X}) d\mathbf{X} \\ &= \int_{\mathbb{R}^{d-1}} \int_{\mathbb{R}} \mathbf{1}_{X^1\in\mathbf{A}_1} h_0(\mathbf{X}) dX^1 d\mathbf{X}^{-1}\end{aligned}$$

A l'aide des hypothèses de gaussianité sur $\mathbf{X}$, la probabilité $\mathbb{P}$ est défini par :

$$\mathbb{P} = \mathbb{E}\left(P(X^1 \in \mathbf{A}_1)\right) \tag{1.14}$$

où $\mathbb{E}$ est l'espérance mathématique par rapport à la variable $\mathbf{X}^{-1}$. La probabilité recherchée $\mathbb{P}$ s'exprime donc comme l'espérance d'une variable aléatoire $P(X^1 \in \mathbf{A}_1)$ par rapport à la variable $\mathbf{X}^{-1}$. Cette espérance se traduit en pratique par l'estimation Monte-Carlo suivante :

$$\hat{\mathbb{P}}^{LS} = \frac{1}{N_C} \sum_{i=1}^{N_C} (P(X^1 \in \mathbf{A}_1(\mathbf{X}_i^{-1}))) \tag{1.15}$$

où $(\mathbf{X}_i^{-1})$ sont des échantillons de la variable aléatoire $\mathbf{X}^{-1}$. La probabilité $P(X^1 \in \mathbf{A}_1(\mathbf{X}_i^{-1}))$ peut alors être approchée de manière paramétrique dans le cas de régions de défaillance simples pour estimer $\hat{\mathbb{P}}^{LS}$. Cette méthode ne requiert pas nécessairement un budget N de simulations important pour obtenir des résultats valables. Les caractéristiques de h_{opt} influencent tout de même la précision de l'estimation donnée par la méthode LS.

1.2.4 Approches fondées sur la caractérisation d'un méta-modèle de la fonction ϕ

Pour des fonctions ϕ coûteuses en temps de calcul ou en ressources par exemple, il peut être intéressant d'utiliser un modèle réduit approchant la fonction ϕ afin de réduire le nombre d'appels à ϕ. Un grand nombre de méthodes de réduction de modèles ont été proposées et comparées dans l'article [137]. Les modèles réduits classiques à base de polynômes, splines et réseaux de neurones ont notamment été associés à la technique de simulation FORM, comme c'est le cas dans les références [19], [49] et [132]. Les polynômes de chaos et les méthodes Monte-Carlo ont également été utilisés pour estimer des évènements rares, tel que décrit dans l'article [81]. L'association de machines à vecteurs de support et de la méthode d'importance splitting a enfin été réalisée dans les références [7] et [16]. D'une situation à une autre, il est évidemment compliqué de déterminer quel couple de méthodes modélisation réduite et échantillonnage est adapté au problème considéré.
La modélisation réduite par krigeage, décrite dans les articles [86], [129] et [127],

présente certains avantages dans le cadre de l'estimation d'évènements rares. En effet, ce type de modèle réduit basé sur des processus gaussiens permet d'estimer la variance de l'erreur du modèle prédit par le krigeage. Un estimateur de la confiance dans le modèle réduit peut être défini en tout point de l'espace des entrées. Cet indicateur peut être utilisé directement pour raffiner le modèle, c'est-à-dire pour effectuer de nouvelles simulations où l'erreur du modèle réduit est la plus importante. Le krigeage a été associé avec les méthodes Monte-Carlo dans l'article [38], importance sampling dans les articles [58], [132] et [35] ou importance splitting dans les références [139], [83] et [9]. La manière de raffiner le modèle de krigeage est un point important de la méthode et différentes stratégies ont été proposées dans les références [35], [115] et [8] afin d'exploiter la description probabiliste donnée par le krigeage et minimiser le nombre d'appels à la fonction ϕ. Une comparaison numérique de différentes méthodes d'estimation d'évènements rares associées au krigeage est proposée dans l'article [82].
L'utilisation de métamodèle peut très fortement réduire le temps de calcul nécessaire pour obtenir une estimation de probabilités satisfaisante. Néanmoins, il est souvent difficile d'estimer le biais introduit par l'utilisation d'un modèle réduit pour l'estimation d'une probabilité. De plus, dans le cas où la fonction ϕ est fortement non linéaire, les modèles réduits sont inadaptés.

1.3 Synthèse

L'étude et la pratique de ces différents algorithmes nous ont permis d'établir une stratégie pour appliquer et développer la méthode d'estimation d'évènements rares la plus adaptée au cas considéré. Cette synthèse est présentée sous forme de questions auxquelles il est impératif de répondre avant de commencer les estimations de manière pratique.

1. Est-il possible d'utiliser la fonction ϕ pour obtenir de nouveaux échantillons de sa sortie ? Si le rééchantillonnage n'est pas possible, un unique jeu d'échantillons $\phi(\mathbf{X}_1), ..., \phi(\mathbf{X}_N)$ est alors disponible. Dans ce cas, la théorie de valeurs extrêmes et l'utilisation de métamodèle sont les seuls algorithmes applicables. Si le rééchantillonnage est possible, la théorie de valeurs extrêmes n'est souvent pas la méthode la plus efficace.
2. La densité de Y ou la fonction ϕ sont-elles analytiquement connues ? Si c'est le cas, il peut être intéressant d'observer les résultats obtenus par la théorie des grandes déviations, via de simples changements de mesure en importance sampling ou bien par le biais des techniques de variables de contrôle ou variables antithétiques. Si ces méthodes ne sont pas efficaces, il faut alors se tourner vers des algorithmes plus généraux et plus complexes à implémenter.
3. La région des entrées qui aboutit à $\phi(\mathbf{X}) > S$ est-elle approximativement connue ? Si oui, alors les méthodes FORM/SORM et NAIS peuvent être efficaces pour l'estimation d'événements rares.
4. La région des entrées qui aboutit à $\phi(\mathbf{X}) > S$ est-elle multimodale ? Si c'est le cas ou si la réponse à cette question n'est pas connue, l'utilisation de l'importance sampling optimisé par cross-entropy, des méthodes FORM/SORM ou line sampling peut être risquée.

5. Quelle est la dimension du problème ? Si $d > 10$, la méthode importance splitting et la méthode importance sampling optimisée par cross-entropy, résistent souvent mieux aux fortes dimensions. Pour des faibles dimensions, toutes les techniques sont adaptées.
6. Quel est le budget N de simulations disponible ? Si $N > 1000$, les méthodes importance sampling et importance splitting sont applicables. Si $N < 1000$, les méthodes FORM/SORM, line sampling et théorie des valeurs extrêmes sont alors recommandées. Les méthodes importance sampling et importance splitting peuvent néanmoins être appliquées quand $N < 1000$ mais, dans ce cas précis, un modèle réduit doit leur être associé.
7. La fonction ϕ est-elle fortement non-linéaire ? Si c'est le cas, les méthodes FORM/SORM, line sampling et les modèles réduits sont à utiliser avec précaution car ils peuvent engendrer un biais d'estimation. La méthode importance splitting résiste en revanche efficacement aux non-linéarités du problème.

Répondre à ces questions permet donc de mieux cibler en pratique les algorithmes à appliquer et à privilégier. De plus, il est important de préciser qu'une connaissance au moins partielle de la fonction ϕ est nécessaire afin de pouvoir utiliser les algorithmes d'estimation d'événements rares précédemment présentés.

1.4 Perspectives de recherche

Les recherches les plus actives actuellement sur les techniques d'estimation d'évènements rares sont tournées essentiellement vers les techniques d'importance splitting, l'utilisation de métamodèles et les lois des valeurs extrêmes. En effet, le réglage des techniques d'importance splitting et l'influence de celui-ci sur l'estimation de la probabilité est une problématique encore ouverte à la recherche. De même, la technique de raffinage d'un métamodèle pour l'estimation d'évènements rares et l'analyse de l'erreur introduite par celui-ci sont également des questions encore peu balisées par la communauté scientifique. En revanche, l'importance sampling ou FORM/SORM sont des techniques qui sont globalement bien maîtrisées. Les innovations dans ces domaines interviennent donc essentiellement sur la résolution de problèmes particulièrement précis.
Ce panorama des différentes méthodes d'estimation d'événements rares a suscité de ma part certaines interrogations auxquelles certaines avancées proposées dans la suite de ce manuscrit vont tenter d'apporter des réponses.

Chapitre 2

Contributions au développement de la technique de non-parametric adaptive importance sampling (NAIS)

2.1 Introduction

L'algorithme NAIS a été proposé dans la référence [143] et ensuite dans l'article [108] afin d'estimer des évènements rares. L'objectif de l'algorithme NAIS est d'utiliser un estimateur à noyaux de densité pour approcher la densité optimale h_{opt} de tirage définie dans l'équation 1.6. Cet algorithme a donc l'avantage d'éviter un choix paramétrique de densité qui peut devenir problématique lorsque la fonction ϕ est inconnue. En effet, l'optimisation paramétrique de la densité d'importance peut être très efficace mais uniquement si le modèle de densité à optimiser a une forme proche de la densité optimale h_{opt}. La multimodalité de h_{opt} est notamment une source de problèmes dans le cas où l'optimisation est paramétrique. L'algorithme NAIS s'affranchit donc de ces difficultés, même si la forme du noyau de densité a malgré tout une influence sur l'applicabilité de cette technique. Par exemple, si $\mathbf{X}$ suit une loi à queue lourde et que les noyaux de l'estimateur de densité sont gaussiens, la méthode NAIS ne sera alors pas efficace. Les estimateurs à noyaux de densité sont également peu performants pour approcher une densité lorsque la dimension augmente (d$>$10). Il en est de même pour l'algorithme NAIS qui n'est généralement applicable que dans des cas où la dimension de $\mathbf{X}$ est inférieure à 10.
En utilisant simplement les algorithmes NAIS proposés dans les articles de référence [143] et [108], il est à noter qu'il n'est pas possible d'obtenir de résultats valables d'estimation de probabilités pour une fonction ϕ quelconque. En effet, ces algorithmes NAIS supposent que l'utilisateur est déjà capable de générer des évènements tels que $\phi(\mathbf{X}) > S$. Cela est en fait rarement le cas dans un cadre réaliste. Les articles [94] et [95] de nos recherches ont donc consistés à améliorer l'approche NAIS de façon à ce qu'elle soit utilisable, de manière générale, pour l'estimation d'évènements rares et cela, sans connaissance *a priori* sur les régions de ϕ qui aboutissent à $\phi(\mathbf{X}) > S$. La section suivante décrit donc la démarche proposée et publiée dans les références [94] et [95] (voir article 1 de la Partie III

contenant une sélection de publications) et une extension de l'algorithme NAIS pour l'estimation de lignes de niveaux de densité publiée dans l'article [101].

2.2 Amélioration proposée de l'algorithme NAIS

L'approche proposée s'inspire des méthodes paramétriques telles que l'optimisation par cross-entropy. La démarche consiste donc à approcher la densité auxiliaire optimale h_{opt} de tirage pour un seuil $S_i < S$ afin qu'il soit possible de générer des échantillons de $\mathbf{X}$ tels que $\phi(\mathbf{X}) > S_i$. Cette nouvelle densité, notée h_i, nous permet de générer de nouveaux échantillons de $\mathbf{X}$ tels que $\phi(\mathbf{X}) > S_i$ et donc d'approcher la densité h_{i+1}, densité auxiliaire optimale de tirage pour le seuil $S_{i+1} > S_i$. En itérant ce processus jusqu'à ce que le seuil courant dépasse S, l'algorithme aboutit à l'estimation de la densité auxiliaire optimale de tirage pour le seuil S et à l'estimation de $\mathbb{P}$. La définition des seuils S_i s'effectue à l'aide de ρ-quantiles des échantillons courants. Les détails de l'algorithme sont décrits dans les étapes suivantes. Pour des raisons de simplicité, l'algorithme est présenté pour un estimateur à noyaux gaussiens mais d'autres types de noyaux peuvent être considérés.

1. Soit $k = 1$ et $\rho \in [0, 1]$
2. Générer les échantillons $\mathbf{X}_1^{(k)}, ..., \mathbf{X}_N^{(k)}$ selon la PDF h_{k-1} et appliquer la fonction ϕ pour déterminer $Y_1^{(k)} = \phi(\mathbf{X}_1^{(k)}), ..., Y_N^{(k)} = \phi(\mathbf{X}_N^{(k)})$
3. Estimer le ρ-quantile empirique $q_k = min(S, Y_{\lfloor \rho N \rfloor}^{(k)})$
4. Estimer $I_k = \frac{1}{kN} \sum_{j=1}^{k} \sum_{i=1}^{N} \mathbf{1}_{\phi(\mathbf{X}_i^{(j)}) \geq q_k} \frac{h_0(\mathbf{X}_i^{(j)})}{h_{j-1}(\mathbf{X}_i^{(j)})}$
5. Mettre à jour l'estimateur à noyaux de densité par :

$$h_k(\mathbf{X}) = \frac{1}{kNI_k \det(B_k)} \sum_{j=1}^{k} \sum_{i=1}^{N} w_j(\mathbf{X}_i^{(j)}) K_d \left(B_k^{-1} \left(\mathbf{X} - \mathbf{X}_i^{(j)} \right) \right) \tag{2.1}$$

   où K_d est une PDF gaussienne de dimension d avec une moyenne nulle et de matrice de covariance identité. Le terme $B_k = diag(b_k^1, ..., b_k^d)$ est une matrice diagonale dont les coefficients peuvent être optimisés par le critère AMISE (Asymptotic Mean Integrated Square Error) tel que décrit dans les références [134] et [46]. Les poids w_j sont définis par $w_j = \mathbf{1}_{\phi \geq q_k} \frac{h_0}{h_{j-1}}$.
6. Si $q_k < S$, $k \leftarrow k + 1$, retour à l'étape 2 de l'algorithme.
7. Estimer la probabilité recherchée $\widehat{\mathbb{P}}^{NAIS} = \frac{1}{N} \sum_{i=1}^{N} \mathbf{1}_{\phi(\mathbf{X}_i^{(k)}) > S} \frac{h_0(\mathbf{X}_i^{(k)})}{h_{k-1}(\mathbf{X}_i^{(k)})}$

L'approche que nous avons proposée permet donc d'étendre l'applicabilité de l'algorithme décrit dans les articles [143] et [108]. La valeur ρ est fixée par l'utilisateur et influence la vitesse de convergence de l'algorithme. Poser $\rho = 0.9$ est souvent un choix pertinent tel que cela est présenté dans l'article [95], où une version de l'algorithme adaptée à l'estimation de quantile est également proposée. La figure 2.1 présente l'estimation d'un $(1-10^{-5})$-quantile dans le cas simple où $\mathbf{X}$ est une densité gaussienne, centrée et réduite, de dimension 1. De nombreux résultats de simulations sont analysés dans la référence [95], ainsi que des cas de simulations applicatifs plus complexes et des comparaisons avec d'autres méthodes d'estimation.

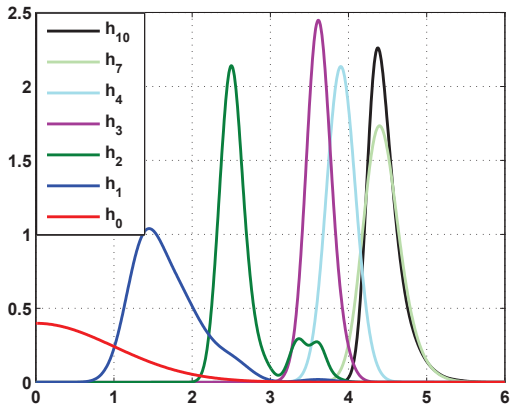


FIGURE 2.1 – Densité de tirage h_i de l'algorithme NAIS pour l'estimation d'un $(1-10^{-5})$-quantile d'une densité gaussienne, centrée et réduite, de dimension 1.

2.3 Cas des lignes de niveaux de densité

L'algorithme NAIS que nous avons proposé dans la section précédente a également été adapté au cas de l'estimation de lignes de niveaux de densité (Minimum Volume Set en anglais) dans le cadre de l'article [101]. En effet, lorsque la sortie Y de la fonction ϕ est multidimensionnelle, il est intéressant de caractériser celle-ci par des lignes de niveaux. Ce cas de figure est notamment présent dans un code de calcul de retombée d'étages de lanceurs spatiaux développé à l'ONERA. La position de retombée est alors définie par 2 coordonnées. La problématique de la caractérisation des lignes de niveau pour les évènements rares est en fait assez peu traitée de manière générale. Les paragraphes suivants résument l'approche que nous avons mise en place.

2.3.1 Définition

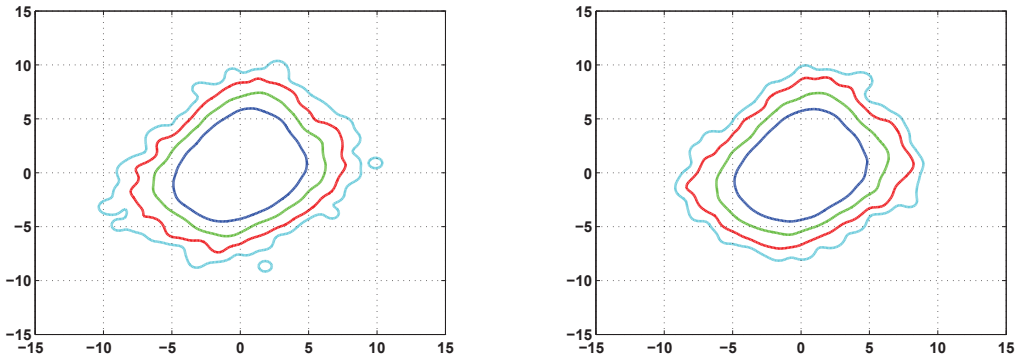
La ligne de niveau t définie par $\mathbf{L}(t)$ issue d'une variable multidimensionnelle $Y \in \mathbb{R}^r$, de PDF g, est donnée par l'équation suivante :

$$\mathbf{L}(t) = \{y \in \mathbb{R}^r : g(y) \geq t\} \tag{2.2}$$

pour $t \geq 0$. La fonction de distribution des lignes de niveaux de g est caractérisée par une application ϵ avec les probabilités telles que :

$$\begin{aligned} \epsilon : [0, \sup g] &\rightarrow [0,1] \\ t &\rightarrow \int_{\mathbf{L}(t)} g(y)dy = P\left(y \in \mathbf{L}(t)\right) = \alpha \end{aligned}$$

Ainsi, la ligne de niveau $\mathbf{L}(t)$ correspondant au niveau t de la densité g est la région de surface minimale de probabilité α. Certaines conditions de régularité sont tout de même nécessaires et décrites dans les articles [110] et [117]. En pratique, l'objectif est d'estimer, pour une valeur de probabilité α, la ligne de niveau $\mathbf{L}(t)$ associée. L'utilisation de l'algorithme NAIS devient donc intéressante lorsque α prend des valeurs proches de 1.



a) b)

FIGURE 2.2 – Estimation de lignes de niveau pour un code de calcul de retombée à l'aide de 10^4 simulations NAIS a) et de 10^6 simulations Monte Carlo b), pour différentes valeurs de probabilités α ($\alpha = 0.95$ en bleu, $\alpha = 0.99$ en vert, $\alpha = 0.999$ en rouge et $\alpha = 0.99999$ en turquoise).

2.3.2 Principe

Un estimateur $\hat{t}$ dit "plug-in" de t, pour estimer une ligne de niveau $\mathbf{L}(t)$ de probabilité α, est défini par le α-quantile de la variable aléatoire $g(\phi(\mathbf{X}))$. La densité g est inconnue mais elle peut être approchée par un estimateur à noyaux de densité. Lorsque la probabilité α n'est pas rare, les méthodes Monte-Carlo sont efficaces. Il suffit alors de générer des échantillons $\phi(\mathbf{X}_i)$, $i = 1, ..., N$, et d'estimer la densité à noyaux $\hat{g}$ de ces échantillons. L'estimateur $\hat{t}$ de la ligne de niveau $\mathbf{L}(t)$ de probabilité α est alors défini par le α-quantile des échantillons $\hat{g}(\phi(\mathbf{X}_i))$, $i = 1, ..., N$. Cet estimateur n'est pas fiable lorsque la probabilité α est faible. Nous avons proposé dans l'article [101] d'adapter la procédure décrite à la section 2.2 afin d'obtenir un estimateur efficace de $\mathbf{L}(t)$ avec un nombre limité d'échantillons. Un exemple d'application est donné dans la figure 2.2 pour un code de calcul de retombée d'étages de lanceurs spatiaux.

2.4 Perspectives de recherche

L'amélioration proposée permet d'utiliser l'algorithme NAIS de manière efficace. Néanmoins, cette approche peut être perfectionnée. En effet, le calcul actuel de la matrice B_k de l'estimateur à noyaux de densité, à l'aide du critère AMISE, n'est pas optimal. Une expression de la variance optimale dans chaque dimension est donnée dans l'article [143], mais n'est pas calculable en pratique. Des approches d'optimisation paramétrique pourraient donc améliorer l'estimation de la variance mais il faudrait vérifier si le gain en matière de précision d'estimation, pour la probabilité recherchée est suffisant par rapport à l'augmentation de la complexité de calcul.
L'algorithme NAIS est essentiellement applicable à des cas de faibles dimensions ($d < 10$). Une récente approche, décrite dans la référence [108], mêlant NAIS et méthode de quasi Monte-Carlo, permet l'échantillonnage uniquement dans un sous-espace de $\mathbb{R}^d$ et améliore donc le traitement des cas à fortes dimensions.

Cette approche pourrait être améliorée en la couplant à une analyse de sensibilité. En effet, les variables les plus influentes, de manière globale, sur la sortie du code de calcul ne le sont pas toujours dans la zone de fonctionnement de l'évènement rare d'intérêt.
Dans le cas où la fonction ϕ est très coûteuse en temps de calcul, l'utilisation des techniques de métamodèle couplées avec l'estimation par importance sampling peut s'avérer très intéressante. Un article où je suis coauteur vient d'être accepté sur cette thématique dans la revue à comité de lecture Structural Safety [5] (voir article 5 de la Partie III contenant une sélection de publications). Le nombre d'échantillons de la fonction ϕ réellement évalués est alors fortement diminué d'un facteur allant de 4 à 20, sans dégradation de la qualité d'estimation. Cet article propose également plusieurs innovations. En effet, le couplage importance sampling et krigeage a été réalisé pour l'algorithme NAIS mais également pour l'algorithme d'importance sampling optimisé par cross-entropy, pour des probabilités inférieures à 10^{-3}. De plus, à l'aide des propriétés statistiques du krigeage, un intervalle de confiance sur la probabilité estimée a été déterminé. Enfin, la caractérisation de la part d'erreur due au métamodèle, présente dans l'erreur inhérente à l'estimation d'une probabilité, a également été approchée. Ces résultats prometteurs doivent être confirmés sur des cas complexes à forte dimension.

Chapitre 3

Contributions au développement de la méthode d'importance splitting

3.1 Introduction

Les techniques de splitting ont été développées dans les années 50 en physique et décrites dans l'article [63], et connaissent un véritable essor à l'heure actuelle, comme le montrent les références [31], [3], [4], [15], [77], [78] et [14]. Celles-ci sont très efficaces pour estimer des probabilités d'évènements rares sur des fonctions ϕ non linéaires et à forte dimension, ce qui est un atout en pratique par rapport aux méthodes d'importance sampling. Néanmoins, lorsque l'on applique la méthode d'importance splitting, le nombre de simulations nécessaires pour obtenir une estimation fiable peut s'avérer rédhibitoire. C'est pourquoi le couplage des techniques de splitting avec des modèles réduits est une approche qui est de plus en plus étudiée.
Le splitting peut également s'avérer complexe à régler dans des cas réalistes. La mise en place d'une approche à base de métamodèle afin de déterminer des réglages efficaces du paramétrage du splitting est actuellement en cours de révision (révision mineure) dans une revue à comité de lecture. Cet algorithme est présenté à la suite de cette section. Il consiste en l'optimisation des paramètres du splitting (niveau de quantile, taille de l'échantillon,...) afin de minimiser la variance de l'estimateur de probabilité, à l'aide de krigeage. Ces résultats permettent de proposer une recommandation de réglage adapté à un cas général de simulation.
Dans le cadre de trois collaborations, en interne, avec des organismes étatiques ou des entreprises, le splitting a enfin permis de répondre aux questions de sécurité et de fiabilité qui s'étaient présentées. Je dresserai donc un bilan de ces activités mélant à la fois recherche et application qui ont abouti aux trois publications suivantes : [97], [113] et [103].

3.2 Réglage de l'estimateur d'importance splitting

3.2.1 Implémentation

Le principe du splitting a été décrit rapidement dans la section 1.2.1. Dans le cas où $\mathbf{X}$ suit une densité de probabilité gaussienne, l'algorithme de splitting adaptatif présenté dans la référence [25] peut se résumer de la manière suivante afin d'estimer $P(\phi(X) > S) = P(X \in \mathbf{A})$ avec un budget de simulations maximum N_{max} :

1) Soient $k = 0$, $j = 0$ et $N_{used} = 0$ le budget de simulations utilisées. Définir les valeurs du nombre d'échantillons N, du quantile β, de la force d'agitation ζ et du nombre d'applications du noyau N_{app}.
2) Générer N échantillons $\mathbf{X}_1^{(0)}, ..., \mathbf{X}_N^{(0)}$ de h_0, puis les échantillons correspondants $\phi(\mathbf{X}_1^{(0)}), ..., \phi(\mathbf{X}_N^{(0)})$. Le paramètre N_{used} est mis à jour de la manière suivante $N_{used} = N_{used} + N$.
3) Estimer le β-quantile $q^{(0)}$ de $\phi(\mathbf{X}_1^{(0)}), ..., \phi(\mathbf{X}_N^{(0)})$.
4) Tant que $q^{(k)} < S$ et $N_{used} < N_{max}$:
   a) Remplacer les simulations, telles que $\{\mathbf{X}_i^{(k)} \mid \phi(\mathbf{X}_i^{(k)}) < q^{(k)}\}$ par des échantillons uniformément choisis avec replacement parmi $\{\mathbf{X}_i^{(k)} \mid \phi(\mathbf{X}_i^{(k)}) > q^{(k)}\}$, puis nommer ces nouveaux échantillons $\mathbf{Z}_1^0, ..., \mathbf{Z}_N^0$
   b) Pour $j = 1$ à N_{app}, estimer les échantillons $\mathbf{Z}_1^j, ..., \mathbf{Z}_N^j$ avec :

$$\mathbf{Z}^j = \begin{cases} \frac{\mathbf{Z}^{j-1} + \zeta\mathcal{N}(0_d, I_d)}{\sqrt{1+\zeta^2}} & \text{si } \phi(\frac{\mathbf{Z}^{j-1} + \zeta\mathcal{N}(0_d, I_d)}{\sqrt{1+\zeta^2}}) > q^{(k)} \\ \mathbf{Z}^{j-1} & \text{sinon} \end{cases}$$

   c) Poser $\mathbf{X}_l^{(k+1)} = \mathbf{Z}_l^{N_{app}}$ $\forall l = 1, ..., N$ et estimer le β-quantile $q^{(k+1)}$ de $\phi(\mathbf{X}_1^{(k+1)}), ..., \phi(\mathbf{X}_N^{(k+1)})$
   d) Poser $k = k + 1$ et mettre à jour la valeur N_{used} avec $N_{used} = N_{used} + N \times N_{app}$
5) Estimer la probabilité recherchée avec :

$$\hat{\mathbb{P}}^{AST} = (1-\beta)^k \times \frac{1}{N_{max} - N_{used}} \sum_{i=1}^{N_{max}-N_{used}} \mathbf{1}_{\phi(X_i^{(k)})>S}$$

Si le budget de simulations est atteint avant que $q^{(k)} < S$, on considère que le splitting échoue dans l'estimation de la probabilité.

3.2.2 Impact des choix de paramétrage dans l'efficacité de l'importance splitting

Comme nous l'avons vu dans la section précédente, plusieurs paramètres doivent être réglés par l'utilisateur pour un résultat de simulation pertinent :

- *le nombre d'échantillons N.*
  Les quantiles $q^{(k)}$ sont en effet mieux estimés si, à chaque itération, le nombre d'échantillons N est important, mais une trop grande valeur de N réduit le nombre d'étapes intermédiaires possibles et augmente le budget de simulations nécessaire.

- *le quantile β.*
  Si la valeur de β est trop faible, la convergence du splitting peut ne pas être atteinte avant l'épuisement du budget de simulations. A l'inverse, si la valeur de β est trop forte, les quantiles $q^{(k)}$ seront mal estimés pour un nombre N d'échantillons donné.
- *la force d'agitation du noyau ζ.*
  Si le nombre de transitions refusées est important, cela signifie que le paramètre ζ a une valeur trop forte et inversement.
- *Le nombre d'application du noyau N_{app}.*
  Si N_{app} est trop faible, les échantillons sont dépendants et l'estimation par splitting peut être biaisé. Dans le cas contraire, de nouveaux échantillons sont générés, même si les échantillons sont déjà indépendants, ce qui n'améliore pas l'estimation.

Le compromis entre ces différents paramètres est complexe à trouver même dans des cas simples.

3.2.3 Approche proposée

Considérons la simulation numérique d'un cas test représentatif sur lequel on cherche à régler les techniques de splitting adaptatif. Une fonction boîte noire peut ainsi être définie. Les paramètres $\alpha = [N, \beta, \zeta, N_{app}]$ sont les entrées de celle-ci et la sortie correspondante est un indice de performance scalaire s. Dans le cas du splitting, s est l'erreur relative de la probabilité d'estimation. Les paramètres α sont supposés appartenir à un ensemble borné $\mathbb{F}$. Le problème de réglage optimal peut être formalisé de la manière suivante :

$$\widehat{\alpha} = \arg\min_{\alpha \in \mathbb{F}} s(\alpha). \tag{3.1}$$

Ce problème est complexe car la seule information disponible est la valeur de la performance obtenue en certaines valeurs de α. La méthodologie proposée utilise le modèle de krigeage afin d'approximer la fonction inconnue reliant les paramètres α et le critère de performance $s(\cdot)$. La procédure EGO (Efficient Global Optimization) décrite dans la référence [62], est ensuite employée pour explorer l'espace des paramètres qui pourrait aboutir à un meilleur critère de performances, c'est-à-dire à une plus faible erreur relative et donc à un meilleur réglage de α.
La stratégie est donc la suivante. Soit m un nombre de valeurs de paramètres α, pour lesquelles l'erreur relative du splitting est connue. Cet ensemble est noté $A_m = \{\alpha_1, ..., \alpha_m\}$. Les erreurs relatives correspondantes sont assemblées dans le vecteur de dimension m, $\mathbf{s}_m = [s(\alpha_1), ..., s(\alpha_m)]^{\mathrm{T}}$. En se basant sur cette connaissance initiale, le krigeage, défini dans la référence [86], permet de construire un modèle approché $\widehat{S}$ de la fonction boîte noire, reliant α et s, en la modélisant par un processus gaussien $S(\cdot)$. Un des avantages du krigeage est sa capacité à estimer l'écart-type $\widehat{\sigma}$ de l'erreur de prédiction, c'est-à-dire le niveau de confiance lié à $\widehat{S}$. Les algorithmes d'optimisation basés sur le krigeage permettent d'aboutir à un compromis entre recherche locale (près de l'optimum) et recherche globale (position dans l'espace des paramètres où l'incertitude du modèle réduit est forte), comme cela est proposé dans l'article [60]. Une de ces stratégies est la procé-

dure EGO, proposée dans la référence [62], que l'on peut décrire de la manière suivante :

1. Echantillonner m valeurs de paramètres α, pour calculer A_m et $\mathbf{s}_m$.
2. Générer le modèle de krigeage associé A_m et $\mathbf{s}_m$.
3. Déterminer $s^m_{\min} = \min_{i=1,\dots,m} \{s(\alpha_i)\}$.
4. Estimer $\widehat{\alpha}_{m+1} = \arg\max_{\alpha \in \mathbb{F}} EI(\alpha, s^m_{\min}, \widehat{S}, \widehat{\sigma})$.
5. Ajouter $\widehat{\alpha}_{m+1}$ à A_m et $s\left(\alpha_{m+1}\right)$ à $\mathbf{s}_m$.
6. Si $m > m_{\max}$, retourner $s^m_{\min}$ comme étant la solution optimale.
   Sinon, retour à l'étape 2 avec $m \leftarrow m+1$.

Le principe général de cet algorithme itératif est de remplacer le problème d'optimisation difficile à résoudre directement (3.1), par une optimisation répétée avec une fonction plus simple nommée *Expected Improvement* (EI, à l'étape 4), dont l'expression est donnée par :

$$\begin{aligned} EI(\alpha, s^m_{\min}, \widehat{S}, \widehat{\sigma}) &= \left(s^m_{\min} - \widehat{S}(\alpha)\right) \Omega\left(\frac{s^m_{\min} - \widehat{S}(\alpha)}{\widehat{\sigma}(\alpha)}\right) \\ &+ \widehat{\sigma}(\alpha)\omega\left(\frac{s^m_{\min} - \widehat{S}(\alpha)}{\widehat{\sigma}(\alpha)}\right), \end{aligned} \tag{3.2}$$

où ω et Ω sont respectivement la densité de probabilité et la fonction de répartition d'une loi normale. Cette fonction est peu coûteuse à évaluer car elle implique la prédiction du krigeage et sa variance. Elle peut donc plus facilement être optimisée à chaque étape de la procédure par un algorithme auxiliaire. Dans nos tests, l'algorithme DIRECT, proposé dans la référence [61], a été utilisé. Grâce au critère EI, plusieurs positions candidates de l'espace des paramètres sont ainsi explorées et l'optimum des performances est alors déterminé pour le cas d'application considéré. Le critère d'arrêt de la procédure EGO est $m_{\max}$ et correspond au budget alloué pour l'évaluation de la boîte noire reliant paramètres et critère de performances. D'autres critères d'arrêt auraient pu être utilisés comme des seuils sur les valeurs successives du maximum de l'EI, à l'étape 4, comme cela est le cas dans l'article [130].

3.2.4 Résultats

Nous avons testé notre approche sur des cas tests et également des codes de simulation. Le budget global de simulations a été fixé à 500000 échantillons pour des probabilités estimées variant entre 10^{-4} et 10^{-7}. Les réglages qui permettent d'obtenir des estimateurs efficaces de probabilité avec le splitting sont précisés dans le tableau 3.1. Ces résultats recoupent, par exemple, les préconisations avancées dans l'article [25], sur le paramètre de quantile. Les paramétrages suivants peuvent ainsi servir de base de départ afin de bénéficier d'une utilisation satisfaisante du splitting, pour des fonctions ϕ très diverses. On peut néanmoins ensuite améliorer ce paramétrage, au cas par cas, pour diminuer l'erreur relative de manière optimale, bien que cela ne soit pas toujours possible dans des cas industriels. Un article sur cette thématique est actuellement relu après une première révision dans la revue *Reliability Engineering & System Safety*.

Paramètres	N	N_{app}	α	β
Recommandations	[3500,11000]	[3,4]	[0.35,0.45]	[0.60,0.85]

TABLE 3.1 – Recommandations de réglage du splitting.

3.3 Application et développement des techniques de splitting pour l'aérospatiale et la défense

La technique de splitting nous a permis de répondre à différentes questions majeures en terme de sécurité et de fiabilité, posées par des organismes étatiques ou des entreprises. Ces études ont abouti aux 3 articles [103], [97] (voir article 4 de la Partie III contenant une sélection de publications) et [113], dont deux sont rapidement résumés dans les sections suivantes.

3.3.1 Caractérisation de la collision débris-satellite

Le 10 février 2009, un satellite actif commercial Iridium et un satellite russe inactif sont entrés en collision bien que leur configuration ne semblait pas propice à un tel évènement, comme cela est expliqué dans l'article [65]. L'impact a notamment été à l'origine de la propagation de nombreux débris dont la plupart sont capables et risquent d'endommager d'autres satellites en orbite. En collaboration avec le CNES, nous nous sommes demandés comment améliorer la précision de l'estimation de la probabilité de collision entre un débris et un satellite. Cela a en pratique deux intérêts majeurs. En effet, la diminution de l'incertitude sur la probabilité permet de réduire le risque de collision, mais également d'allonger la durée de vie d'un satellite, en annulant, par exemple, des manoeuvres d'évitement de débris qui ne seraient pas nécessaires. L'approche que nous avons proposée est basée sur le splitting.
L'incertitude sur la position initiale d'un objet spatial est donnée par une PDF gaussienne à 6 dimensions (3 dimensions pour la position et 3 dimensions pour la vitesse). La moyenne de la position initiale est décrite dans les "two-lines elements" fournis par le NORAD (North American Aerospace Defense Command). La trajectoire de l'objet est ensuite propagée, sur une certaine durée, à l'aide de l'outil de propagation de trajectoire d'objets spatiaux SGP4 (Simplified General Perturbations), développé également par le NORAD. Ce code est déterministe et la position d'un objet à un instant donné ne dépend donc que de l'incertitude sur la position initiale. En appliquant cette procédure à deux objets spatiaux, on peut facilement estimer la probabilité que leur distance relative soit inférieure à un certain seuil.
Les techniques de type importance sampling paramétrique ou non paramétrique ne sont pas efficaces pour estimer une probabilité de collision débris-satellite avec cette modélisation. En effet, pour la densité d'importance, différentes familles de PDF ont été testées et optimisées par cross-entropy, mais sans résultat. De même, l'algorithme NAIS n'est pas efficace, du fait de la dimensionnalité du problème ($d > 10$). La technique de splitting est donc parfaitement adaptée, comme le montre les résultats suivants. Le cas de collision choisi est présenté dans la figure 3.1, dans laquelle deux zones de conflit ont une influence sur la probabilité d'un

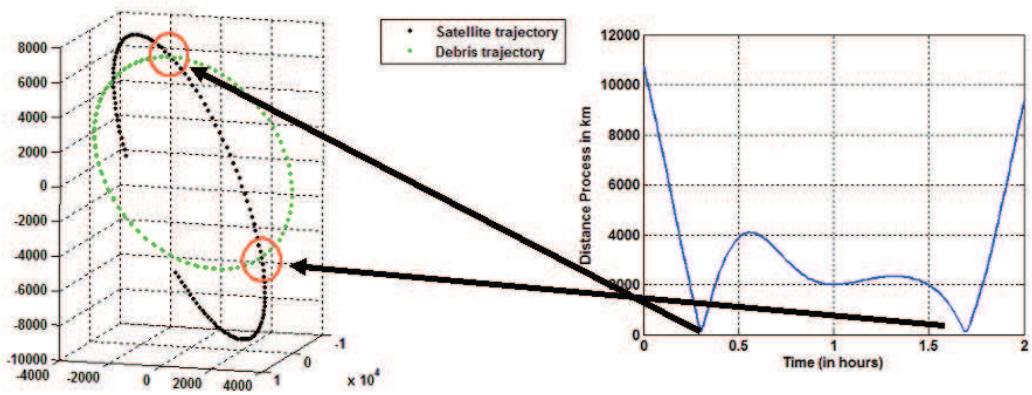


FIGURE 3.1 – Trajectoire du débris et du satellite, ainsi que leur distance relative.

Méthodes	Nombre de simulations	Estimation de la probabilité	Erreur relative
Monte-Carlo	100000	$8.3\ 10^{-4}$	0.10
Monte-Carlo	10000	$8.1\ 10^{-4}$	0.42
Splitting	10000	$8.5\ 10^{-4}$	0.20

TABLE 3.2 – Résultat d'estimation de probabilité de collision.

tel cas.

La distance de collision est fixée à 10 mètres. La table 3.2 présente l'estimation de la probabilité de collision obtenue pour la technique d'importance splitting et, en guise de référence, pour les méthodes Monte-Carlo. La technique de splitting est très efficace bien que la fonction ϕ soit en fait particulièrement non linéaire. Cela explique pourquoi les autres méthodes d'estimation d'évènements rares testées ne sont pas pertinentes. Cette étude pourrait avoir un impact sur les opérationnels du CNES qui surveillent les satellites gérés par cet organisme. L'implémentation de techniques particulaires pour estimer une probabilité de collision, est en cours de réflexion dans les cas où le satellite et le débris ont une faible vitesse relative.

3.3.2 Estimation de la probabilité de collision en espace aérien non contrôlé

Maintenir une distance de séparation minimum entre deux avions dans l'espace aérien est essentiel dans le management du trafic aérien. Cette exigence est généralement garantie dans l'espace contrôlé, par les contrôleurs aériens qui imposent aux avions de voler à certains niveaux d'altitude ou sur certains caps. Les positions des avions sont connues grâce à leur transpondeur et aux différents radars terrestres. Les statistiques de collision ou de perte de séparation sont par conséquent bien évaluées. L'analyse de celles-ci permet notamment de faire évoluer les règles en vigueur, comme le montrent les articles [18] et [80]. De plus, la modélisation du trafic aérien en espace contrôlé a été longuement étudiée et simulée notamment, dans les articles [66], [133], [140] et [55]. Ce n'est pas le cas de l'espace aérien non contrôlé qui constitue l'essentiel de l'espace aérien, à l'ex-

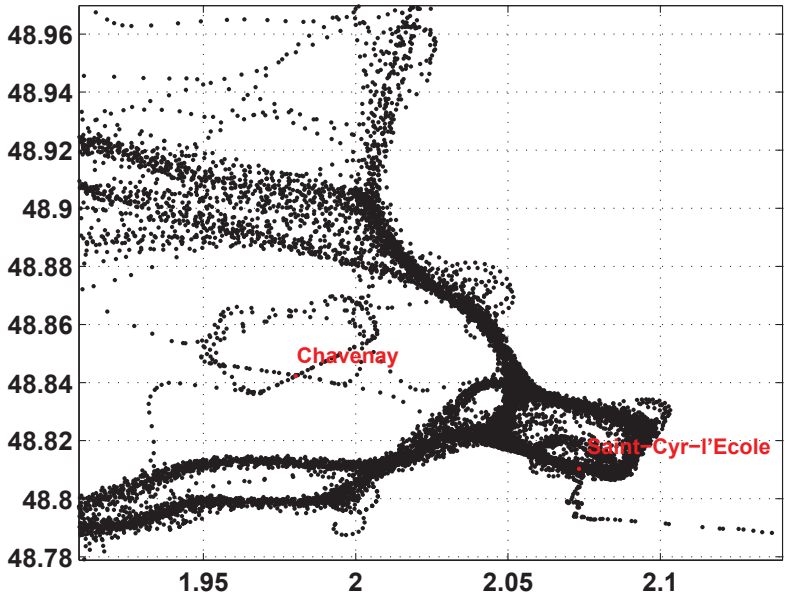


FIGURE 3.2 – Base de données de trajectoires d'aéronefs en longitude et en latitude au voisinage de Saint-Cyr-l'Ecole.

ception du voisinage des aéroports. En effet, il est difficile d'évaluer les risques de collision et de perte de séparation, car le nombre d'avions, leur position et leur direction ne sont pas connus.
Un fort besoin d'estimation de ces risques a abouti à une collaboration avec la DGA et Claude Le Tallec, un des spécialistes français des drones (voir article 4 de la Partie III contenant une sélection de publications). En effet, l'espace aérien non contrôlé est envisagé pour intégrer les vols de drones, comme il est précisé dans les références [1], [112], [2], [68] et [67]. Nous avons donc travaillé en espace aérien non contrôlé, sur une base de données de 170 trajectoires autour de l'aérodrome de Saint-Cyr-L'Ecole dans les Yvelines, décrite dans la figure 3.2. A l'aide de cette approche, une modélisation du problème équivalent à un modèle boîte noire a été développée. Une approche de splitting général, a été mise en place afin d'estimer une carte des risques autour de cet aérodrome. L'enjeu méthodologique était la forte dimensionnalité du problème ($d>50$) et la non gaussianité de ces entrées, dépassant le cadre de la simple application du splitting. Un exemple de résultat obtenu dans le cadre de cette étude est explicité dans la figure 3.3 sous forme de carte de probabilité de collision.
Cette étude a montré que les risques de collision dans l'espace aérien non contrôlé sont en fait très faibles, ce qui étaye les quelques statistiques disponibles sur le sujet. Les plus grands risques de collision se situent aux alentours des aérodromes et aéroports, mais diminuent fortement dès que l'on s'éloigne de ces régions. Ces résultats ont donc permis d'affiner la position de la DGA sur les stratégies et la réglementation de vol de drones dans l'espace aérien.

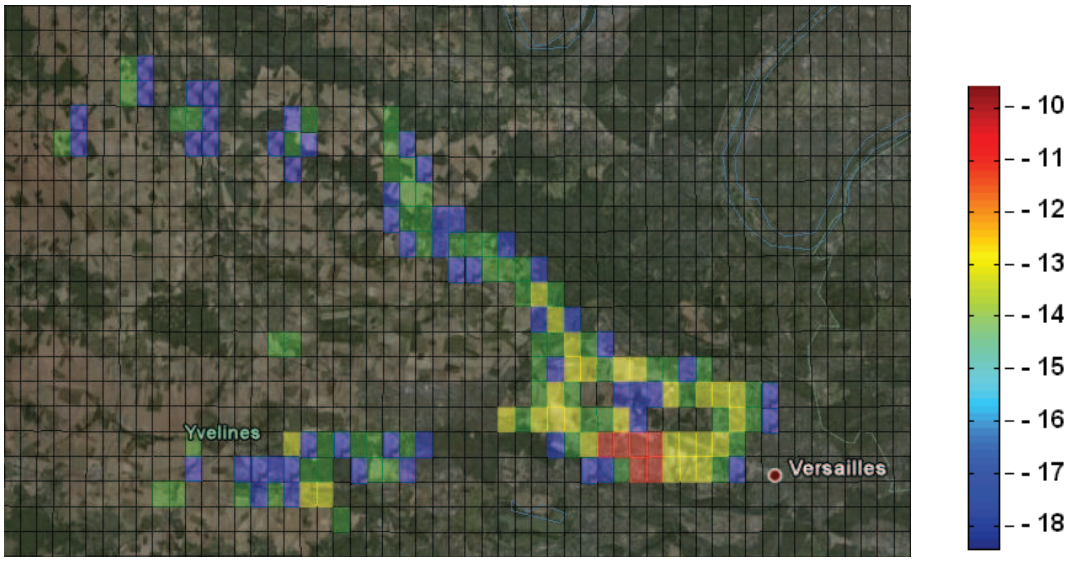


FIGURE 3.3 – Carte de probabilité de collision estimée par splitting dans la région de Saint-Cyr-l'Ecole. L'échelle de couleur sur la probabilité est logarithmique.

3.4 Perspectives de recherche

Le nombre de simulations nécessaires pour utiliser les techniques de splitting sont au minimum de l'ordre de 10^4 échantillons. Cela peut être trop lourd à mettre en place sur des codes complexes et notamment des codes de calcul industriels. L'utilisation conjointe de splitting et de modèle réduit est une perspective directe de ces travaux. Des algorithmes de ce type ont néanmoins déjà été proposés récemment dans la référence [83], bien que l'erreur d'estimation due au métamodèle ne soit pas quantifiée.
La politique d'agitation des échantillons est prépondérante dans le splitting et peut consommer beaucoup de simulations. Il est important de caractériser ce qu'est une bonne politique d'agitation des échantillons et de déterminer l'apport que celle-ci peut avoir sur la réduction de variance de l'estimation. Ces questions sont complexes et nécessitent des développements dans la méthodologie et dans la simulation. En effet, d'une part, le réglage de la force d'agitation ζ du noyau markovien et du nombre d'application N_{app} de celui-ci, est complexe mais d'autre part, la compréhension de l'effet de ces deux paramètres sur la variance me semble particulièrement importante pour améliorer le splitting. Des approches par plan d'expérience sur ces deux paramètres pourraient être une source potentielle d'idées pour la compréhension des enjeux liés à la politique d'agitation.
L'estimation de lignes de niveau à l'aide de différentes méthodes d'évènements rares peut également être intéressante. Un article dans une revue à comité de lecture est actuellement en cours de révision et propose l'estimation de lignes de niveau par la technique d'importance splitting, en complément des méthodes d'importance sampling que nous avons développées en section 2.3. Les techniques de splitting ont l'intérêt d'être, dans ce cadre, moins coûteuse en temps de calcul que les méthodes NAIS.
La recherche menée à l'ONERA sur le splitting concerne actuellement essentielle-

ment le cas de processus stochastiques et non le cadre statique. Ce modèle n'a pas été développé dans ce dossier mais pose tout de même des problèmes méthodologiques et applicatifs très actuels. Une thèse est actuellement en cours, en 2ème année, en collaboration avec l'INRIA de Rennes (DR INRIA François Le Gland). Une publication [57] a été réalisée en 2013 et un article, dans cette thématique, a également été accepté dans la revue *Computational Data and Statistical Analysis* [96] (voir article 2 de la Partie III contenant une sélection de publications) basé sur le réglage de l'algorithme développé par P. Del Moral et J. Garnier dans les articles [32] et [45]. Une communication sur le splitting pondéré dans un cadre dynamique a de plus été accepté dans la Winter Simulation Conference 2013. Un article d'état de l'art des méthodes d'estimation d'évènements rares dans un cadre dynamique et un article sur l'optimisation de système particulaire ont également été soumis.

Chapitre 4

Analyse de sensibilité et évènements rares

4.1 Introduction

L'analyse de la sensibilité de la sortie d'un code de calcul par rapport à ses différentes entrées est une problématique majeure de la compréhension d'une fonction boîte noire. Des analyses de sensibilité locale (autour d'un point de fonctionnement) ou globale (basée sur la variance) peuvent être ainsi effectuées afin de déterminer les entrées les plus influentes sur la sortie de la fonction ϕ [54]. Néanmoins, il n'existe pas de technique qui permette d'analyser la sensibilité d'une probabilité d'un dépassement de seuil sur la sortie, par rapport aux entrées du code boîte noire. Cela est d'autant plus important dans le cadre des évènements rares qu'une entrée peut être particulièrement influente sur le comportement global de la sortie, mais être peu influente sur la probabilité de l'évènement rare, et inversement. De plus, de nombreuses techniques qui résistent peu à la dimension, telles que la méthode NAIS, peuvent être couplées à des analyses de sensibilité locale ou globale, afin de réduire la dimension de l'espace à échantillonner. Ce choix est-il toujours judicieux ? Cette question est en fait peu traitée dans la littérature scientifique [79]. Nous avons donc proposé une approche, décrite dans l'article [93] (voir article 3 de la Partie III contenant une sélection de publications), basée sur les techniques d'importance sampling, pour répondre à ce type de questions. Les sections suivantes de ce chapitre décrivent donc cette démarche.

4.2 Particularités de la modélisation

Considérons le modèle boîte noire décrit dans la section 1.1. Supposons dorénavant que la densité h_0 de $\mathbf{X}$ est incertaine. De manière plus précise, le modèle paramétrique de h_0 est connu, mais les paramètres $\mathbf{\Theta} = (\Theta_1, ..., \Theta_m)$ de h_0 sont aléatoires et de densité $f_{\mathbf{\Theta}}$. Par exemple, si h_0 est une PDF gaussienne, alors $\mathbf{\Theta}$ décrit la moyenne et la matrice de covariance de h_0. Dans de nombreux cas de modélisation industrielle, il n'est pas toujours possible d'avoir une idée précise de la PDF de toutes les entrées de la fonction ϕ. La plupart du temps, pour des raisons de simplicité, les paramètres des PDF sont fixés à leur valeur la plus

vraisemblable. Cela peut néanmoins être risqué si certaines composantes de $\mathbf{\Theta}$ ont une forte influence sur la probabilité cible $\mathbb{P}_\theta = P(\phi(\mathbf{X}) > S|\mathbf{\Theta} = \theta)$ car un biais peut apparaître dans l'estimation. L'approche proposée ici permet donc de prendre en compte un large éventail de PDF pour les entrées de ϕ.
L'objectif est de caractériser les influences des paramètres $\mathbf{\Theta}$ sur la probabilité $\mathbb{P}_\theta$, à l'aide de techniques d'importance sampling et d'analyse de sensibilité.

4.3 Description de la méthode d'analyse

Les techniques d'importance sampling ou d'importance splitting permettent de déterminer des PDF d'échantillonnage efficaces, afin d'estimer la probabilité $\mathbb{P}_\theta$, pour une valeur θ fixée. Soient $< \theta >$ la valeur la plus vraisemblable des paramètres aléatoires $\mathbf{\Theta}$ et $\mathbf{X}_1, ..., \mathbf{X}_N$ un jeu de N échantillons de PDF $h_{<\theta>}$ efficace pour estimer $\mathbb{P}_{<\theta>}$. Un estimateur de $\mathbb{P}_{<\theta>}$ est donné par :

$$\widehat{\mathbb{P}}_{<\theta>} = \frac{1}{N}\sum_{i=1}^{N} \mathbf{1}_{\phi(\mathbf{X}_i)>S} \frac{h_{0,\mathbf{\Theta}=<\theta>}(\mathbf{X}_i)}{h_{<\theta>}(\mathbf{X}_i)} \tag{4.1}$$

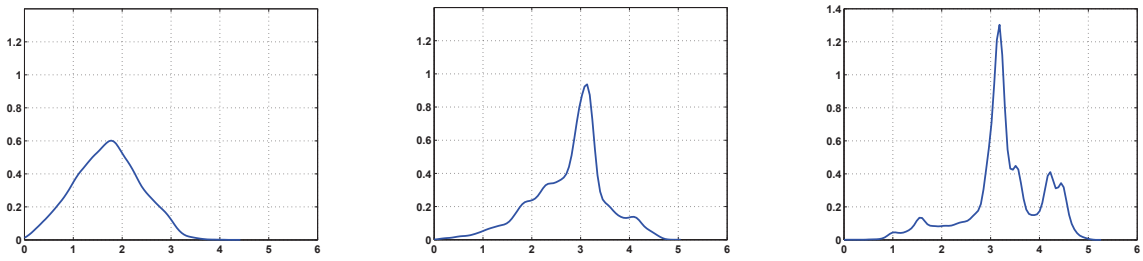
Les échantillons $\mathbf{X}_1, ..., \mathbf{X}_N$ qui aboutissent à $\phi(\mathbf{X}) > S$, appartiennent toujours à la région de défaillance, quelque soit la valeur θ prise par $\mathbf{\Theta}$, car la région de défaillance est indépendante de $\mathbf{\Theta}$ et ne dépend que des propriétés de la fonction ϕ. Ces échantillons peuvent donc être utilisés afin d'estimer $\mathbb{P}_\theta$ pour différentes valeurs de θ de la manière suivante :

$$\widehat{\mathbb{P}}_{\theta} = \frac{1}{N}\sum_{i=1}^{N} \mathbf{1}_{\phi(\mathbf{X}_i)>S} \frac{h_{0,\mathbf{\Theta}=\theta}(\mathbf{X}_i)}{h_{<\theta>}(\mathbf{X}_i)} \tag{4.2}$$

Il est donc possible d'estimer $\mathbb{P}_\theta$ quelque soit la valeur prise par $\mathbf{\Theta}$ à l'aide de $\widehat{\mathbb{P}}_{<\theta>}$ sans avoir à réechantillonner. Cependant, cet estimateur a ses limites. En effet, si $\widehat{\mathbb{P}}_{<\theta>}$ et $\widehat{\mathbb{P}}_\theta$ ont des valeurs très différentes, alors l'estimateur $\widehat{\mathbb{P}}_\theta$, sera biaisé. Néanmoins, en pratique si $10^{-2} < \frac{\mathbb{P}_\theta}{\widehat{\mathbb{P}}_{<\theta>}} < 10^2$ pour $N = 10000$, alors l'estimateur $\widehat{\mathbb{P}}_\theta$ sera efficace et c'est dans ce cas de figure que nous nous placerons par la suite. Si ce n'était pas le cas, augmenter N ou déterminer un nouveau jeu d'échantillons $\mathbf{X'}_1, ..., \mathbf{X'}_N$ adapté à $\mathbf{\Theta} = \theta$ sont des alternatives possibles pour valider cette hypothèse.
En reformulant l'équation 4.2, l'expression suivante se déduit :

$$\widehat{\mathbb{P}}_\theta = \Gamma(\mathbf{\Theta}) = \Gamma(\Theta_1, ..., \Theta_m) \tag{4.3}$$

où Γ est une fonction scalaire boîte noire $\Gamma : \mathbb{R}^m \to \mathbb{R}$. Les paramètres des PDF d'entrées de ϕ sont les entrées de celle-ci et la probabilité de dépassement de seuil correspond à sa sortie. Il est donc possible d'appliquer sur Γ des méthodes d'analyse de sensibilité globale pour déterminer les paramètres Θ_i, $i = 1, ..., m$ les plus influents sur la probabilité. L'estimation d'indices de Sobol, décrite dans les références [53] , [136] et [135], a été la méthode choisie dans l'article [93], afin d'évaluer cette influence.



$\Theta_i = 0.1\ \forall i$ a) $\Theta_i = 0.55\ \forall i$ b) $\Theta_i = 1\ \forall i$ c)

FIGURE 4.1 – PDF de Y estimée avec 10000 échantillons Monte-Carlo pour différentes valeurs de Θ_i

Indice de Sobol Total	valeur	écart-type
S_{T_1}	0.22	0.01
S_{T_2}	0.19	0.01
S_{T_3}	0.09	0.01
S_{T_4}	0.09	0.02
S_{T_5}	0.38	0.02
S_{T_6}	0.14	0.01
S_{T_7}	0.11	0.01

TABLE 4.1 – Moyenne et écart-type des indices de Sobol totaux estimés sur 100 répétitions avec 50000 simulations Monte-Carlo.

4.4 Résultats de simulation

Pour appliquer la méthode décrite à la section précédente, un code de simulation d'un missile supersonique à stato-réacteur a été privilégié. Ce code possède 7 entrées supposées indépendantes, issues de la centrale inertielle du missile, et une sortie, la distance de sa position de retombée par rapport à la cible. Ces 7 entrées ont une moyenne nulle et un écart-type Θ_i, $i = 1, ..., 7$. Les paramètres Θ_i sont aléatoires et de PDF uniforme sur l'intervalle $[0.1, 1]$. La sortie Y du code de calcul est par conséquent incertaine. La figure 4.1 propose en effet des estimations de la PDF de Y à l'aide d'un estimateur à noyaux de densité pour différentes valeurs de Θ_i. L'écart entre ces PDF montre que les paramètres Θ_i ont une influence non négligeable sur la densité de Y. La probabilité d'intérêt dans ce cadre est fixée à $\mathbb{P} = P(Y > 4.5)$. Nous avons donc appliqué la stratégie développée à la section précédente afin de déterminer quels sont les paramètres Θ_i les plus influents sur $\mathbb{P}$, en estimant des indices de Sobol totaux. Ces résultats sont résumés dans le tableau 4.1 où S_{T_i} correspond à l'indice de Sobol total estimé pour le paramètre Θ_i. Certains paramètres, tels que Θ_5 ou Θ_2 sont beaucoup plus influents que d'autres sur l'estimation de $\mathbb{P}$ et doivent être fixés avec précaution lors de l'utilisation de la fonction ϕ. Une erreur sur leur valeur pourrait impliquer un fort biais dans l'estimation de $\mathbb{P}$ lors des simulations.

4.5 Perspectives de recherche

L'approche proposée est efficace mais peut être améliorée et notamment la robustesse de l'estimateur de probabilité de dépassement $\widehat{\mathbb{P}}_\theta$. Avec une procédure plus adaptée et un raffinage des échantillons $\mathbf{X}_i$, il est sans doute possible d'obtenir une estimation de $\mathbb{P}_\theta$ plus fiable. Une approche à base de métamodèle est actuellement en cours d'évaluation. De plus, en pratique, il n'est pas toujours possible de modéliser la loi des paramètres $\mathbf{\Theta}$ car il y a peu d'études effectuées sur ces paramètres. Il serait donc intéressant d'envisager ce type d'analyse avec une approche qui permettrait l'abandon d'hypothèses sur la loi des paramètres $\mathbf{\Theta}$.
Une autre alternative visant à caractériser l'influence de paramètres sur une probabilité est de calculer les lois des phénomènes aléatoires menant à la défaillance $\phi(\mathbf{X}) > S$. Cela revient, par conséquent, à calculer les probabilités conditionnelles des variables aléatoires du système complexe, sachant que ce dernier est dans un régime de défaillance. Le calcul de ces lois permet ainsi de caractériser tous les aléas et leur corrélation qui conduisent justement à une telle défaillance. Cette thématique de recherche est actuellement traitée dans le cadre d'une thèse, en fin de première année, en collaboration avec l'INRIA de Bordeaux (DR INRIA Pierre Del Moral), à l'aide de technique statistique basée sur des îlots de particules. Cette méthode appelée notamment SMC^2 (SMC^2=Sequential Monte Carlo sur les îlots et Sequential Monte Carlo sur les particules de chaque îlot) mise en place dans le cadre d'application de filtrage dans l'article [22] a été développée de manière originale pour caractériser les lois conditionnelles de $\mathbf{\Theta}$ sachant l'évènement rare $Y > S$ dans le cadre de cette thèse. Une publication dans la revue à comité de lecture Structural Safety a été soumise sur ce sujet. L'analyse théorique de la convergence de cet algorithme est actuellement en cours d'étude.

Troisième partie

Sélection de publications

Chapitre 1

Extreme quantile estimation with non parametric adaptive importance sampling, J. Morio, *Simulation Modelling Practice and Theory*, 2012, Vol. 27, pp. 76-89

Simulation Modelling Practice and Theory 27 (2012) 76–89

Contents lists available at SciVerse ScienceDirect

Simulation Modelling Practice and Theory

journal homepage: www.elsevier.com/locate/simpat

Extreme quantile estimation with nonparametric adaptive importance sampling

Jérôme Morio *

ONERA, The French Aerospace Lab, BP 80100, 91123 Palaiseau Cedex, France

ARTICLE INFO

Article history:
Received 20 July 2011
Received in revised form 9 May 2012
Accepted 18 May 2012
Available online 20 June 2012

Keywords:
Statistics
Rare event
Quantile estimation
Importance sampling

ABSTRACT

In this article, we propose a nonparametric adaptive importance sampling (NAIS) algorithm to estimate rare event quantile. Indeed, Importance Sampling (IS) is a well-known adapted random simulation technique. It consists in generating random weighted samples from an auxiliary distribution rather than the distribution of interest. The optimization of this auxiliary distribution is often very difficult in practice. First, we review how to define the optimal auxiliary density of IS for quantile estimation in the general case and then propose a nonparametric method based on Gaussian kernel density estimator to approach the optimal auxiliary density that does not assume an initial PDF guess. This method is then finally applied to theoretical cases to demonstrate the efficiency of the proposed NAIS algorithm and on the quantile estimation of the temperature of a forest fire detection simulator.

1. Introduction

Estimating rare event probability and quantile with a good accuracy is an important source of interest in reliability and safety. Different methods have been investigated for this purpose like splitting or sampling. Importance splitting [1–4] also called subset simulation [5–10] is a valuable algorithm to estimate both rare events and quantiles. It does not suffer much of the curse of dimensionality but the algorithm parameters are not always easy to set. Importance sampling [11–15] is also a well-known technique that consists in generating random weighted samples from an auxiliary distribution rather than the distribution of interest. It is easy to implement but the main issue of this algorithm is the determination of a valuable auxiliary distribution. Rare event quantiles can also be estimated via controlled sampling and regression meta-models [16]. Line sampling is also widely employed in structural reliability problems for estimating very small probabilities with a reduced number of simulations [17–21]. Extreme value theory [22,23] finally deals with the extreme deviations from probability distributions and is often used for assessing risk in finance or meteorology domain. Well-known strategies in computer experiment community that consists in using a surrogate model can also help for rare event estimation process [24–26].

In this article, we focus on importance sampling algorithm that is a widely used variance reduction simulation technique. The main idea of IS is to generate weighted samples from an auxiliary probability density function (PDF) rather than the distribution of interest. The IS quantile and probability estimation variance can be greatly reduced with respect to a general Monte Carlo estimate. A major difficulty of IS is the selection of an efficient auxiliary distribution function from which pseudorandom numbers are generated since a wrong choice of auxiliary distribution leads to worse estimations than Monte Carlo simulations. Parametric sampling distributions, if available at all, are often inadequate for high-dimensional integrals over irregular regions. One possible remedy is to use a nonparametric importance sampling to estimate the unknown optimal sampling function. This kind of approach is notably proposed in [27–31] for expectation estimation but not for quantile

* Tel.: +33 1 80 38 66 54.
E-mail address: jerome.morio@onera.fr

1569-190X/$ - see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.simpat.2012.05.008

estimation. In this article, we generalize this method for quantile estimation and also propose an improvement of the usual NAIS technique since the initial auxiliary PDF guess is no more necessary in this article NAIS version. When the sampling dimension is too large, one can also use the nonparametric partial importance sampling algorithm which applies nonparametric importance sampling to a certain subspace [29].

In this article, we review the fundamentals of IS technique and define the optimal IS auxiliary density for quantile estimation. We then propose an innovative nonparametric adaptive importance sampling algorithm for quantile estimation without any hypothesis on the initial PDF contrary to the algorithms presented in [27,28,30]. The performances of the proposed NAIS algorithm for rare quantile estimation are then applied to different test cases.

2. Importance sampling

2.1. Monte Carlo method

In this section, we review the fundamentals of IS technique for quantile estimation. We consider a black box input–output system ϕ with $\phi : \mathbb{R}^d \to \mathbb{R}$ a continuous scalar function. This system has a d-dimensional random input X with a PDF $f : \mathbb{R}^d \to \mathbb{R}$ and a 1-dimensional output $Y = \phi(X)$. We assume in this article that Y is also a random variable with probability density g. In this article, we focus on the general case where we want to estimate the α-quantile q_α of the variable Y. We propose to define the quantile q_α with the following equation:

$$\alpha = \int_{q_\alpha}^{+\infty} g(y)dy \tag{1}$$

This definition is a transposition of the quantile definition for continuous random variables to the complementary cumulative distribution function (CCDF) of Y instead of cumulative distribution function (CDF). One relevant way to estimate q_α is to use Monte-Carlo techniques [32,33]. One determines an estimator of the CCDF $G(y)$ (also called survival function or reliability function) of Y which is defined in the following way:

$$G(y) = \int_{\mathbb{R}} \mathbf{1}_{\{z \geqslant y\}} g(z)dz = \int_{\mathbb{R}^d} \mathbf{1}_{\{\phi(x) \geqslant y\}} f(x)dx \tag{2}$$

Then, one generates in practice a set of N independent and identically distributed samples of X, $(X_1, \ldots, X_N)$ and applies the function ϕ to these samples to determine a set of samples $(Y_1, \ldots, Y_N)$ that follows the PDF g of Y. The Monte Carlo estimator of the CCDF $\widehat{G}^{MC}$ of Y is given by:

$$\widehat{G}^{MC}(y) = \frac{1}{N}\sum_{i=1}^{N} \mathbf{1}_{\{Y_i \geqslant y\}} \tag{3}$$

The Monte Carlo estimator of the α-quantile of the variable Y is thus:

$$\hat{q}_\alpha^{MC} = sup\{y, \widehat{G}^{MC}(y) > \alpha\} \tag{4}$$

This estimator can be refined with interpolation and smoothing methods [34].

In statistics, the accuracy of an estimator is expressed in term of asymptotic variance in the case of unbiased estimators that satisfies a central limit theorem. The study of the estimate variance alone is not sufficient to assess the performance of the estimate since one has also to consider its bias. In the following of the article, we will consequently present the theoretical value of the target quantiles if available or a valuable estimation based on Monte Carlo method over a high number of simulations, the mean estimated target quantiles and their variance for the different methods.

It can be shown that the variance of $\widehat{G}^{MC}(y)$ depends on N and on the theoretical CCDF $G(y)$. The variance of $\widehat{G}^{MC}(y)$ relatively to the PDF f is given by:

$$Var_f(\widehat{G}^{MC}(y)) = \frac{Var(\mathbf{1}_{\{Y_i \geqslant y\}})}{N} \tag{5}$$

The random variables $\mathbf{1}_{\{Y_i \geqslant y\}}$ are independent and identically distributed and follow a Bernoulli distribution with parameter $p = G(y)$. The variance of this Bernoulli distribution is given by $p(1-p)$. The variance of $\widehat{G}^{MC}(y)$ is thus given by

$$Var_f(\widehat{G}^{MC}(y)) = \frac{G(y)(1-G(y))}{N} \tag{6}$$

We can find in [35,36] that $\hat{q}_\alpha^{MC}$ has the following asymptotic expansion:

$$\hat{q}_\alpha^{MC} = q_\alpha + \frac{\alpha(1-\alpha)}{\sqrt{N}g^2(q_\alpha)}\xi + O(1/N) \tag{7}$$

where ξ follows a standard normal distribution. This formulation gives the limit of the quantile estimate, its asymptotic normality and its asymptotic variance. The quantile variance is then larger when a low quantile is estimated since $g(q_\alpha) \to 0$. Monte-Carlo techniques are clearly not adapted to estimate such low expectations.

2.2. Basics of importance sampling

The objective of IS is thus to reduce the estimation variance of $\widehat{G}^{MC}(Y)$ without increasing N. The idea is to generate the samples $X_1,\ldots,X_N$ from an auxiliary PDF h and then estimate G in the following way:

$$\widehat{G}^{IS}(y) = \frac{1}{N}\sum_{i=1}^{N}\mathbf{1}_{\{Y_i \geqslant y\}}\frac{f(X_i)}{h(X_i)} = \frac{1}{N}\sum_{i=1}^{N}\mathbf{1}_{\{\phi(X_i) \geqslant y\}}\frac{f(X_i)}{h(X_i)} \tag{8}$$

The auxiliary PDF h must be chosen with care. The PDF h must be positive when f is positive, and its tail must be sufficiently heavy with respect to the tail of f in order to have an estimate with finite variance.

The IS estimator of the α-quantile of the variable Y is then defined by:

$$\hat{q}_{\alpha}^{IS} = sup\{y, \widehat{G}^{IS}(y) > \alpha\} \tag{9}$$

The term $\widehat{G}^{IS}$ is an unbiased estimator of G since:

$$E_h(\widehat{G}^{IS}(y)) = \frac{1}{N}\sum_{i=1}^{N}\int_{\mathbb{R}^d}\mathbf{1}_{\{\phi(x_i) \geqslant y\}}\frac{f(x_i)}{h(x_i)}h(x_i)dx_i = \int_{\mathbb{R}^d}\mathbf{1}_{\{\phi(x) \geqslant y\}}f(x)dx = G(y) \tag{10}$$

with E_h the expected value operator relatively to the PDF h. The variance of $\widehat{G}^{IS}$ is estimated by:

$$Var_h(\widehat{G}^{IS}(y)) = \frac{Var_h(\mathbf{1}_{\{\phi(X) \geqslant y\}}w(X))}{N} \tag{11}$$

with $w(X) = \frac{f(X)}{h(X)}$. One can then obtain

$$Var_h(\widehat{G}^{IS}(y)) = \frac{1}{N}\left(\int_{\mathbb{R}^d}\mathbf{1}^2_{\{\phi(x) \geqslant y\}}w(x)^2h(x)dx - G(y)^2\right) \tag{12}$$

$$= \frac{1}{N}(E_h(\mathbf{1}_{\{\phi(X) \geqslant y\}}w(X)^2) - G(y)^2) \tag{13}$$

The quantile estimates follow a normal distribution because of the central limit theorem. The variance of this normal distribution is given when $N \to \infty$ by

$$\frac{1}{g^2(q_\alpha)}\left(\int_{\mathbb{R}^d}\frac{\mathbf{1}_{\{\phi(x) \geqslant q_\alpha\}}f^2(x)}{h(x)}dx - \alpha^2\right)$$

(see [16]). The variance of IS estimate depends strongly on the choice of h and also depends on N. If h is well adapted to the considered case, then the IS estimate variance can be very low. Conversely, if h is not well-chosen to the estimation of $G(y)$ and q_α, the IS estimate results can have a higher variance than Monte Carlo estimate. The IS estimate with the lower variance is obtained with the IS optimal auxiliary density defined in the following section.

2.3. IS optimal auxiliary density

The IS optimal auxiliary density h_{opt} is the auxiliary PDF that minimizes the variance $Var_h(\widehat{G}^{IS}(y))$. Since variances are non-negative quantities, the lower possible bound of a variance is thus zero. IS optimal auxiliary density h_{opt} can be determined by cancelling the variance in Eq. (13)

$$\frac{1}{N}(E_{h_{opt}}(\mathbf{1}_{\{\phi(X) \geqslant y\}}w(X)^2) - G(y)^2) = 0 \tag{14}$$

We have thus the following expression:

$$E_{h_{opt}}\left(\mathbf{1}_{\{\phi(X) \geqslant y\}}\frac{f(X)^2}{h_{opt}(X)^2}\right) = G(y)^2 \tag{15}$$

and consequently

$$E_{h_{opt}}\left(\mathbf{1}_{\{\phi(X) \geqslant y\}}\frac{f(X)^2}{h_{opt}(X)^2G(y)^2}\right) = 1 \tag{16}$$

With h_{opt} defined with the following equation,

$$h_{opt}(x) = \frac{\mathbf{1}_{\{\phi(x) \geqslant y\}}f(x)}{G(y)} \tag{17}$$

the relation Eq. (16) is verified. Consequently, the PDF h_{opt} implies a zero variance estimate and as such the optimal choice of density. The auxiliary PDF h_{opt} depends unfortunately on $G(y)$ which is the unknown quantity that we try to estimate. The PDF h_{opt} is thus unusable in practice.

However, this approach is difficult to implement directly for quantile estimation because the variance of quantile depends on $g^2(q_\alpha)$. The derivative of g at q_α is not explicit and annealing the quantile variance is thus not useful theoretically. Nevertheless, using Eq. (17), a simple and practically feasible alternative, is to choose the following optimal IS auxiliary density to estimate the α-quantile q_α:

$$h_{opt}(x) = \frac{\mathbf{1}_{\{\phi(x) \geqslant q_\alpha\}} f(x)}{\alpha} \tag{18}$$

This auxiliary density is often suggested in different articles [15,16] without further theoretical justification but based on the similarity with Eq. (17). In the following, our objective is thus to determine an IS auxiliary density that approaches the optimal density h_{opt} to estimate the quantile q_α and also allows an easy generation of random numbers. Different parametric methods exist but are not in the scope of this article. We propose a new nonparametric algorithm using a mixture of Gaussian kernels K_d [37] to determine an estimated IS optimal density $\hat{h}_{opt}$ for quantile estimation in the following section.

3. Nonparametric adaptive importance sampling (NAIS) for quantile estimation

In this section we present a new nonparametric adaptive importance sampling algorithm to estimate the α-quantile q_α of Y. The objective of the proposed NAIS technique is to approximate IS optimal auxiliary density with Gaussian kernel function. The principle is to apply IS algorithm from an initial auxiliary PDF g_0 that is not optimal. The simplest case is of course to choose $g_0 = f$ when one has no idea of an efficient initial distribution. This is an important advance on NAIS since the main articles [27,38,28,29,31] on NAIS assume that the distribution g_0 is able to generate at least some rare events. It is not obviously the case in practice. The proposed NAIS algorithm is thus the following:

(1) Set $k = 0$ and $g_0 = f$, and define parameters N_k, γ and k_{max}
(2) Generate N_k random numbers $X_1^{(k)}, \ldots, X_{N_k}^{(k)}$ from PDF g_k.
(3) Estimate the CCDF G by

$$\widehat{G}_k^{NAIS}(y) = \frac{1}{m_{k+1}} \sum_{j=0}^{k} \sum_{i=1}^{N_j} \mathbf{1}_{\{\phi(x_i^{(j)}) \geqslant y\}} \frac{f(x_i^{(j)})}{g_j(x_i^{(j)})} \tag{19}$$

where $m_{k+1} = \sum_{j=0}^{k} N_k$. Taking in account all the previous samples enables to obtain a more efficient and stable kernel density estimate. The IS estimator of the α-quantile of the variable Y is then obtained with:

$$\hat{q}_{\alpha,k}^{NAIS} = sup\left\{y, \widehat{G}_k^{NAIS}(y) > \alpha\right\} \tag{20}$$

In practice, $\hat{q}_{\alpha,k}^{NAIS}$ is easily performed considering the samples $\phi\left(X_1^{(k)}\right), \ldots, \phi\left(X_{N_k}^{(k)}\right)$ with a weight respectively equal to $\frac{f\left(X_1^{(k)}\right)}{g_k\left(X_1^{(k)}\right)}, \ldots, \frac{f\left(X_{N_k}^{(k)}\right)}{g_k\left(X_{N_k}^{(k)}\right)}$

(4) Estimate the parameter γ by $\gamma = min\left(\hat{q}_{\alpha,k}^{NAIS}, \gamma_\rho^k\right)$ where the term γ_ρ^k is the ρ-quantile of the unweighted samples $\phi\left(X_1^{(k)}\right), \ldots, \phi\left(X_{N_k}^{(k)}\right)$. We then set

$$w_k(x) = \frac{\mathbf{1}_{\{\phi(x) > \gamma\}} f(x)}{g_k(x)} \tag{21}$$

to approximate the optimal IS density defined in Eq. (18).
How to select the samples that will be used to estimate the new sampling PDF at the following step? In this article, we use the quantile γ_ρ^k to determine these samples. The samples that are potentially interesting to determine q_α are located above q_α. Unfortunately, at the start of algorithm, very few samples are above q_α, and thus it is not possible to determine a valuable sampling PDF with them. For that purpose, the definition of γ_ρ^k increases the number of samples that will be taken into account to estimate the sampling PDF at the following step. This approach is inspired from Cross-entropy optimization of importance sampling technique defined in [39–41].

(5) Define the PDF g_{k+1} with

$$g_{k+1}(x) = \frac{1}{m_{k+1} \widehat{G}_k^{NAIS}(\gamma) \det(B_{k+1})} \sum_{j=0}^{k} \sum_{i=1}^{N_j} w_j\left(x_i^{(j)}\right) K_d\left(B_{k+1}^{-1}\left(x - x_i^{(j)}\right)\right) \tag{22}$$

where K_d is standard d-dimensional Gaussian function with zero mean and a diagonal covariance matrix $B_{k+1} = diag\left(b_{k+1}^1, \ldots, b_{k+1}^d\right)$. During the convergence, the PDF g_{k+1} has to become close to the optimal sampling density de-

fined in Eq. (18). The PDF g_{k+1} is thus defined by similarity with this density. Indeed, by definition of the weight w in Eq. (21), the term

$$\frac{1}{m_{k+1}\det(B_{k+1})}\sum_{j=0}^{k}\sum_{i=1}^{N_j} w_j\left(x_i^{(j)}\right)K_d\left(B_{k+1}^{-1}\left(x-x_i^{(j)}\right)\right)$$

is a weighted kernel approximation of function $\mathbf{1}_{\{\phi(x)>\gamma\}}f(x)$. The variable $\widehat{G}^{NAIS}(\gamma)$ enables g_{k+1} to be a PDF and is an estimation of probability $P(\phi(X)>\gamma)$. Equivalent definitions of NAIS PDF can be found in [27–29] for probability estimation.
The adapted bandwidth in each dimension is optimized according to the AMISE (asymptotic Mean integrated Square Error) criterion [32,42]. As noticed in [43], when estimating a probability density function, the standard kernel estimator is efficient for densities that are not far from Gaussian in shape, but it can perform very poorly when the shape is far from Gaussian, especially, near the PDF tail. In that case, more efficient kernels have been described in [44].

(6) If $k < k_{max}$, return to the stage (2) of the algorithm and set $k = k + 1$.
(7) Estimate the CCDF G with the samples at iteration k_{max}

$$\widehat{G}^{NAIS}(y)=\frac{1}{m_{k+1}}\sum_{j=0}^{k_{max}}\sum_{i=1}^{N_j}\mathbf{1}_{\{\phi(x_i^{(j)})\geq y\}}\frac{f\left(x_i^{(j)}\right)}{g_j\left(x_i^{(j)}\right)} \tag{23}$$

Another alternative to estimate the CCDF G is also given by

$$\widehat{G}^{NAIS}(y)=\frac{1}{N_{k_{max}}}\sum_{i=1}^{N_{k_{max}}}\mathbf{1}_{\{\phi(x_i^{(k_{max})})\geq y\}}\frac{f\left(x_i^{(k_{max})}\right)}{g_{k_{max}}\left(x_i^{(k_{max})}\right)} \tag{24}$$

Both CCDF estimators are efficient in the proposed algorithm. The IS estimator of the α-quantile of the variable Y is then obtained with:

$$\hat{q}_\alpha^{NAIS}=sup\{y,\widehat{G}^{NAIS}(y)>\alpha\} \tag{25}$$

In the first steps of the algorithm, it is probable that $\gamma_\rho^k < \hat{q}_{\alpha,k}^{NAIS}$ if $\hat{q}_{\alpha,k}^{NAIS}$ is an extreme quantile. At iteration k, The samples $X_i^{(j)}$, $j = 1,\ldots,k$ that have a strictly positive weight in the definition of PDF g_{k+1} are above γ_ρ^k after application of ϕ. Consequently, the PDF g_{k+1} with function ϕ generates samples $\phi\left(X_1^{(k+1)}\right),\ldots,\phi\left(X_{N_{k+1}}^{(k+1)}\right)$ at iteration $k+1$ that are above γ_ρ^k. By definition, the ρ-quantile γ_ρ^{k+1} of $\phi\left(X_1^{(k+1)}\right),\ldots,\phi\left(X_{N_{k+1}}^{(k+1)}\right)$ is then greater than γ_ρ^k. The parameter γ increases consequently during NAIS convergence.

When γ_ρ^k becomes greater than $\hat{q}_{\alpha,k}^{NAIS}$, that is, when the NAIS convergence has been reached, then the samples $X_i^{(j)}$ $j = 1,\ldots,k$ that have a strictly positive weight in the sampling PDF g_{k+1} are above $\hat{q}_{\alpha,k}^{NAIS}$ after application of ϕ. The PDF g_{k+1} with function ϕ generates samples $\phi\left(X_1^{(k+1)}\right),\ldots,\phi\left(X_{N_{k+1}}^{(k+1)}\right)$ at iteration $k+1$ that are above $\hat{q}_{\alpha,k}^{NAIS}$. By definition, the ρ-quantile γ_ρ^{k+1} of samples $\phi\left(X_1^{(k+1)}\right),\ldots,\phi\left(X_{N_{k+1}}^{(k+1)}\right)$ is then greater than $\hat{q}_{\alpha,k+1}^{NAIS}$. The parameter γ is then equal to $\hat{q}_{\alpha,k+1}^{NAIS}$.

The value of ρ is set by the user and influences the convergence speed of the algorithm. A discussion about the value of ρ is given in Section 4.1.1.

Convergence of NAIS has notably been studied in [28,27]. NAIS algorithm has always been used to estimate probability. In the proposed version of this algorithm, we show that it is possible to estimate a quantile with accuracy. Moreover, the function g_0 is initially equal to the density f. In the algorithm proposed by [27], the pdf g_0 is set by the user in order to generate rare events, that is, to generate samples x_i so that $\phi(x_i) > q_\alpha$. If no sample or only very few samples are above the threshold, the NAIS algorithm cannot be used or furnish wrong estimation of quantiles and probabilities. In realistic situations, it is nearly impossible to find a valuable initial PDF and consequently the NAIS algorithm is not efficient. It is no more the case with the proposed version since one estimates the quantile q_α step by step. One does not try to directly generate samples that are above q_α. Indeed we iteratively generate samples that are above γ. It is an easier task than before since one has $\gamma \leqslant q_\alpha$. At each iteration, the value of γ increases until it reaches q_α. This principle enables to estimate quantile when one has not any knowledge of an efficient initial PDF for X.

The termination criterion of the proposed algorithm is a maximum number of iterations k_{max} because of simulation purposes. Indeed, one wants to simulate with a constant number of total simulations whatever the value of k_{max}. One can also choose an alternative termination criterion based on the convergence evaluation of the NAIS algorithm. In the previous section, the asymptotic properties of the classical IS estimates have been exposed and one would have expected that these properties would have been used to build a termination criterion. In the adaptive context, the asymptotic behavior of the estimate is much more involved and as such cannot be used to build a termination criterion. However, one can propose for instance to stop the NAIS algorithm once a reliable and stable estimate of the quantile of interest is obtained. If $\hat{q}_{\alpha,k-1}^{NAIS}$ and $\hat{q}_{\alpha,k}^{NAIS}$ are respectively the NAIS q_α estimate at iteration $k-1$ and k, a possible termination criterion is:

$$\left|\frac{\hat{q}_{\alpha,k-1}^{NAIS} - \hat{q}_{\alpha,k}^{NAIS}}{\hat{q}_{\alpha,k}^{NAIS}}\right| < \beta \tag{26}$$

where β is predefined by the user. If the successive quantile estimates do not change, the algorithm is stopped. This termination criterion enables to reduce the required total number of simulations for a studied case. Since the total number of simulations varies then from a trial to another, it can be difficult to compare the NAIS results with other different methods.

4. Application on simple test cases

In the previous section, we have presented the principle of the NAIS algorithm. In the following, we propose to apply NAIS algorithm to simple cases in order to evaluate the performances of the algorithm and the influence of the parameters ρ and N_k.

4.1. One-dimensional Gaussian PDF

First, we consider a very simple case where the random variable X follows a one-dimensional Gaussian PDF f with zero mean and a variance equal to 1, $N(0,1)$. The function ϕ is the identity function so that $Y = X$. We try to estimate with NAIS different quantile values of Y. The NAIS method is applied with the following parameters:

- $g_0 = f$ is a Gaussian PDF with zero mean and a variance equal to 1.
- The sampling sizes are set so that $N_0 = N_1 = \ldots = N_9 = 1000$.
- k_{max} is set to 9 in this toy example so that the total number of simulations is equal to 10,000. Nevertheless, depending on the value of N_i and ρ, the NAIS convergence can be reached before the iteration k_{max}.
- ρ is set to 0.1.

The Table 1 gives different mean α-quantile estimations of Y with NAIS algorithm. Each quantile estimation has been repeated 100 times and then an average has been computed. When it deals about rare event quantile estimation, one also defines the relative deviation of an estimate by the ratio between its standard deviation and its estimated mean or the theoretical quantile value if the true quantile is known. The theoretical values are also presented in the Table 1. All the chosen quantiles are well estimated with NAIS in mean when they are compared to the theory. The $\hat{q}_{\alpha}^{NAIS}$ relative deviation is also low even for the 10^{-7}-quantile. The Fig. 1 shows the evolution of the PDF g_k obtained with the NAIS algorithm for the estimation of 10^{-5}-quantile. From the PDF g_0 which is not adapted to the estimation of rare quantile, the NAIS algorithm enables us to update the sampling PDF at each iteration in order to approximate the optimal auxiliary density h_{opt}. The PDF g_{10} is indeed a valuable approximation of h_{opt}. The Figs. 2 and 3 show the different sample histograms generated throughout the different NAIS iterations. The intermediate values of γ_{ρ}^{k} and $\hat{q}_{\alpha,k}^{NAIS}$ are also given.

To reduce the computation time, one should be interested to not recycle the samples of previous iterations in the NAIS process. This approach is not efficient and raises some issues during the NAIS convergence. In Fig. 4, we present a convergence problem that can occur when sample recycling of previous NAIS iterations is not performed. The diversity of the sample population decreases during the NAIS convergence and does not estimate a valuable quantile. The 10^{-5}-quantile is estimated in that case to 3.87 with a 12.9% relative deviation.

Nonparametric importance sampling always comes with increased computational burden that is why we also provide runtime execution. On a standard PC, NAIS algorithm takes about 6 s to deliver one quantile estimation with 10,000 simulations whatever the value of α whereas it takes 0.005 s to generate 10,000 Monte Carlo simulations and estimate a quantile. Consequently, if the function ϕ does not require a large execution runtime and it is the case when ϕ is the identity function, generating a lot of Monte Carlo simulations can be more efficient than NAIS.

4.1.1. Influence of the ρ parameter

In this section, we propose to analyze the influence of parameter ρ on NAIS quantile estimation in the case presented in the previous section. In Table 2, we vary the value ρ and compare the 10^{-5}-quantile estimate for $\rho = \{0.01, 0.05, 0.1, 0.2, 0.3, 0.5, 0.7, 0.9\}$. In Table 3, we present the NAIS quantile estimation at different iterations for two values of ρ. If the value ρ is too large, the NAIS convergence is not reached and an underestimated quantile value is determined. It is confirmed in Fig. 5 where we estimate experimentally the number of iterations needed to reach convergence in function of ρ. We assume in that case that convergence is reached with the criterion given in Eq. (26) with $\beta = 0.05$. Low values ρ are able to estimate valuable quantile with a low relative deviation of the estimation in this simple case but can lead to instability when one considers more difficult cases. A trade-off between stability and performances has to be done to choose an efficient

Table 1
Estimation of different α-quantiles with 10,000 NAIS simulations for a one-dimensional Gaussian PDF.

α	0.1	0.05	0.01	10^{-3}	10^{-5}	10^{-7}
Theoretical quantile value	1.28	1.64	2.33	3.09	4.26	5.20
$\hat{q}_{\alpha}^{NAIS}$	1.28	1.64	2.33	3.09	4.27	5.20
NAIS relative deviation	0.5%	0.3%	0.2%	0.1%	0.09%	0.09%

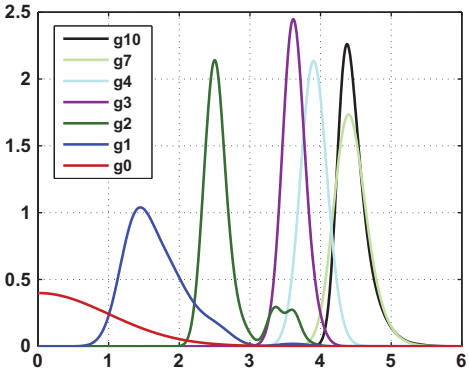


Fig. 1. The estimated optimal IS density g_i with a mixture of Gaussian kernels for $i = 0,\ldots,10$ adapted to the estimation of the 10^{-5}-quantile for a one-dimensional Gaussian PDF.

value ρ. From our experience, a valuable ρ value is $\rho = 0.1$ since it does not require a high number of simulations but also reaches convergence with a good confidence. Knowing the value of α and ρ, an efficient rule of thumb to determine the number of necessary iterations k_{max} is given by:

$$\rho^{k_{max}+1} > \alpha \tag{27}$$

It is due to the iterative use of the ρ-quantile in the expression of γ in the NAIS algorithm. It gives a minimum iteration number where we can hope that the convergence has been reached.

4.1.2. Influence of the N_k parameter

In this section, we propose to analyze the influence of the number of samples N_k on NAIS quantile estimation proposed in Section 4.1 while keeping the total number of simulations constant and equal to 10,000. In that case, the value of k_{max} is then varied depending on the value of N_k and is thus equal to $\frac{10{,}000}{N_k} - 1$. In Table 4, we vary the value N_k and compare the 10^{-5}-quantile estimate for $N_k = \{100, 200, 500, 1000, 2000, 5000\}$. If the value N_k is too large, there are not enough iterations of the NAIS algorithm to reach convergence and quantile is underestimated. If a too low value N_k is considered, valuable quantile estimate can be determined. In the same way as the parameter ρ, too few samples at each iteration can lead to instability. From our experience, relevant values of ρ and N_k for the algorithm efficiency are $\rho = 0.9$ and $N_k = 1000$. If one considers a high number of samples N_k, then the parameter ρ can be slightly increase to reach convergence so that at least 100 samples $X_i^{(k)}$, $i = 1,\ldots,N_k$ be over the new threshold γ at iteration $k+1$. Conversely, if one considers a low number of samples N_k, then the parameter ρ should be decrease to reach convergence.

4.2. Weibull distribution

In this section, we apply NAIS algorithm to estimate quantiles of a Weibull distribution with scale parameter $\lambda = 1$ and shape parameter $k = 2$. The kernel density estimator is still based on normal distributions and not adapted to the Weibull distribution. The NAIS method is applied with the following parameters:

- $g_0 = f$ is a Weibull PDF with scale parameter $\lambda = 1$ and shape parameter $k = 2$.
- The sampling sizes are set so that $N_0 = N_1 = \ldots = N_9 = 1000$.
- k_{max} is set to 9 so that the total number of simulations is equal to 10,000.
- ρ is set to 0.1.

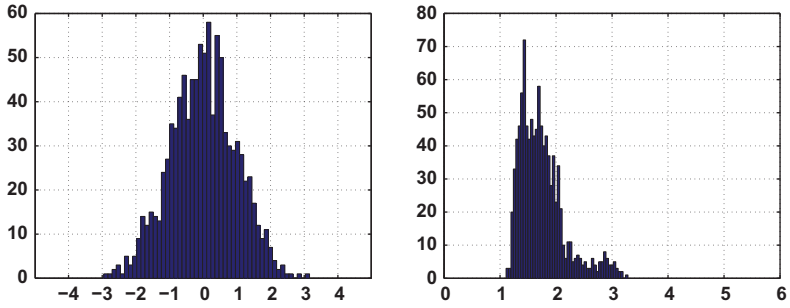
The Table 5 presents the results obtained with NAIS algorithm. Even if the density kernel shape is not adapted to Weibull distribution, the NAIS algorithm gives some interesting results.

The density kernel in auxiliary sampling PDF tail must be sufficiently heavy with respect to the tail of f in order to obtain valuable results. If it is not the case, the NAIS algorithm is not efficient as any parametric importance sampling methods are with an unadapted auxiliary density. The standard kernel estimator is efficient for densities that are not far from Gaussian in shape, but it can perform very poorly when the shape is far from Gaussian, especially, near the boundaries. In that case, more efficient kernels have been described in [44] and can be applied when Gaussian kernels are not valuable.

4.3. Multidimensional cases

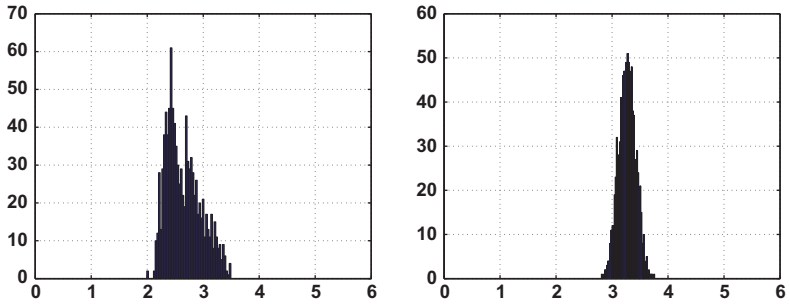
4.3.1. Test case with multidimensional function

In the following section, we consider the case where $X = (X_1, X_2, \ldots, X_5)$ follows a multidimensional Gaussian PDF with mean $(0,0,0,0,0)$ and a covariance matrix equal to I_5, the identity matrix of $\mathbb{R}^5$. The function ϕ is given by the following function called Ackley function:



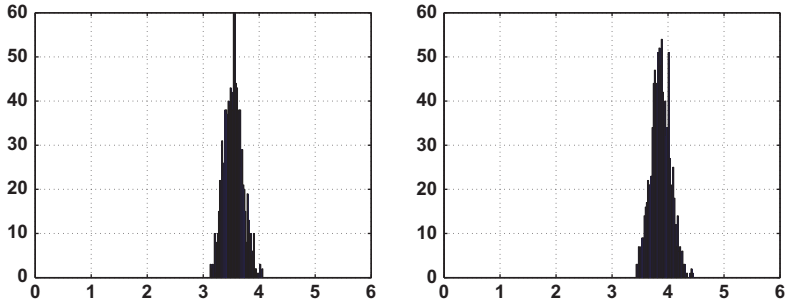
samples $X_1^{(0)},..,X_{1000}^{(0)}$ generated with g_0 samples $X_1^{(1)},...,X_{1000}^{(1)}$ generated with g_1

$\gamma_\rho^0 = 1.26$ and $\hat{q}_{\alpha,0}^{NAIS} = 3.05$ $\gamma_\rho^1 = 2.26$ and $\hat{q}_{\alpha,1}^{NAIS} = 3.28$



samples $X_1^{(2)},..,X_{1000}^{(2)}$ generated with g_2 samples $X_1^{(3)},..,X_{1000}^{(3)}$ generated with g_3

$\gamma_\rho^2 = 3.11$ and $\hat{q}_{\alpha,2}^{NAIS} = 3.50$ $\gamma_\rho^3 = 3.47$ and $\hat{q}_{\alpha,3}^{NAIS} = 3.76$



samples $X_1^{(4)},..,X_{1000}^{(4)}$ generated with g_4 samples $X_1^{(5)},..,X_{1000}^{(5)}$ generated with g_5

$\gamma_\rho^4 = 3.75$ and $\hat{q}_{\alpha,4}^{NAIS} = 4.02$ $\gamma_\rho^5 = 4.08$ and $\hat{q}_{\alpha,5}^{NAIS} = 4.20$

Fig. 2. Sample histograms and intermediate parameter values with NAIS algorithm adapted to the estimation of the 10^{-5}-quantile for a one-dimensional Gaussian PDF.

$$Y = \phi(X) = -20\exp\left(-0.2\sqrt{\frac{1}{d}\sum_{i=1}^{d}X_i^2}\right) - \exp\left(\frac{1}{d}\sum_{i=1}^{d}\cos(2\pi X_i)\right) + 20 + e \tag{28}$$

This function is nonlinear because of the cosinus function which also implies several local maxima and one global maximum on (0,0,0,0,0). The effective dimension is 5 but the interactions between inputs are very low. The local maxima imply disjoint failure modes that are difficult to cope with in parametric modelisation. We then estimate different rare quantiles of the random variable Y with NAIS algorithm. We have applied NAIS algorithm with the following parameters:

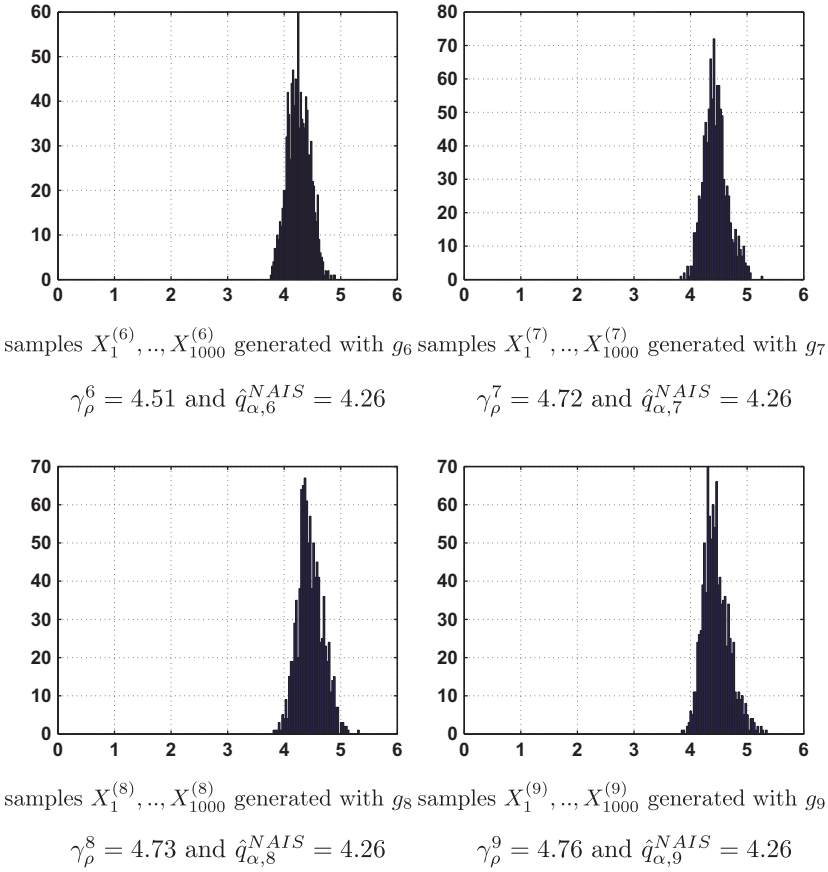


Fig. 3. Sample histograms and intermediate parameter values with NAIS algorithm adapted to the estimation of the 10^{-5}-quantile for a one-dimensional Gaussian PDF.

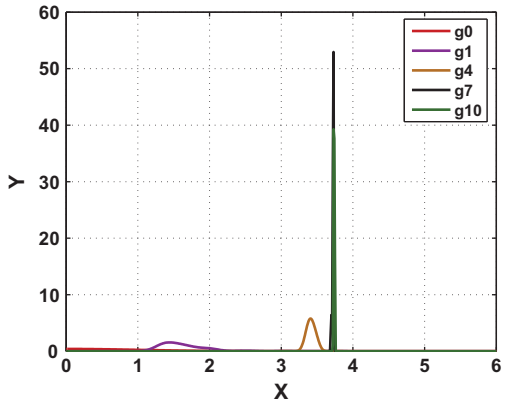


Fig. 4. Convergence problem when recycling of samples is not performed.

Table 2
Estimation of 10^{-5}-quantiles with 10,000 NAIS simulations with different values of ρ for a one-dimensional Gaussian PDF.

ρ	0.01	0.05	0.1	0.2	0.3	0.5	0.7	0.9
$\hat{q}_\alpha^{NAIS}$	4.26	4.27	4.27	4.27	4.16	3.16	1.94	0.31
Relative deviation	0.5%	0.2%	0.1%	0.1%	2.3%	2.9%	2.3%	11.5%

Table 3
Evolution of the quantile estimate during the iteration process for two values of ρ.

Iteration number	0	1	2	3	4	5	6	7	8	9
γ for $\rho = 0.1$	1.29 (3.5%)	2.34 (3.6%)	3.02 (5.1%)	3.39 (8.3%)	3.71 (6.7%)	4.01 (2.2%)	4.23 (1.7%)	4.26 (0.3%)	4.27 (0.1%)	4.27 (0.1%)
γ for $\rho = 0.5$	0 (1893%)	0.67 (8.9%)	1.17 (4.7%)	1.57 (3.8%)	1.90 (2.9%)	2.19 (2.1%)	2.48 (2.4%)	2.72 (2.6%)	2.95 (2.6%)	3.16 (2.8%)

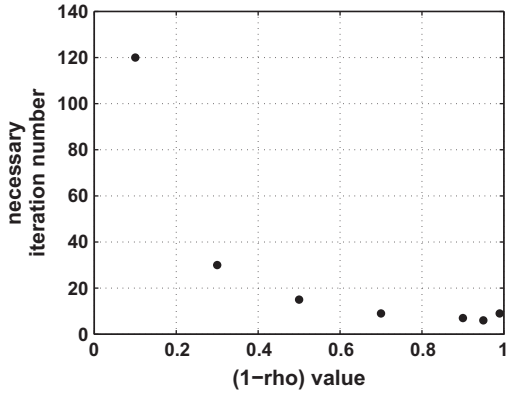


Fig. 5. Evaluation of the necessary number of iterations to reach NAIS convergence in function of the value of ρ.

Table 4
Estimation of 10^{-5}-quantiles with 10,000 NAIS simulations with different values of N_k for a one-dimensional Gaussian PDF.

N_k	100	200	500	1000	3000	5000
$\hat{q}_\alpha^{NAIS}$	4.27	4.27	4.27	4.26	3.07	2.37
Relative deviation	0.5%	0.1%	0.1%	0.1%	3.8%	66%

Table 5
Estimation of different α-quantiles with 10,000 NAIS simulations for a Weibull PDF with $\lambda = 1$ and $k = 2$.

α	0.1	0.05	0.01	10^{-3}	10^{-5}	10^{-7}
Theoretical quantile value	1.52	1.73	2.14	2.62	3.39	4.01
$\hat{q}_\alpha^{NAIS}$	1.52	1.73	2.14	2.62	3.39	4.00
NAIS relative deviation	0.2%	0.2%	0.2%	0.1%	0.1%	0.9%

- the initial PDF $g_0 = f$.
- k_{max} is set to 9.
- The sampling sizes are set so that $N_0 = N_1 = \ldots = N_9 = 1000$.
- ρ is set to 0.1.

The Table 6 gives the α-quantile estimations on 100 repetitions with NAIS algorithm. Even in a multidimensional case, the NAIS relative deviation estimation is low and the different estimations are not biased. A significant Monte Carlo simulation has been performed in order to evaluate NAIS convergence. Gaussian kernel enables us to well approximate the IS optimal auxiliary density. On a standard PC, NAIS algorithm takes about 32 s to estimate a quantile with 10,000 simulations of the function ϕ. As previously stated, the computation burden of nonparametric algorithm can be an issue when the function ϕ is simple to simulate. In Tables 7 and 8, we show the 10^{-5}-quantile estimation for different values of ρ and N_k parameters. In Table 9, we also present the evolution of NAIS quantile estimate during the process for two value of ρ. In Fig. 6, we finally present the required number of iterations to reach convergence. Same conclusions as in Sections 4.1.2 and 4.1.1 can be drawn.

4.3.2. Effect of dimension on NAIS estimation

In this section, we propose to study the influence of the problem dimension on the NAIS estimation. Let us consider the simple case where the d-dimensional input $X = (X_1, X_2, \ldots, X_d)$ follows a d-dimensional Gaussian PDF with zero mean and a

Table 6
Estimation of different α-quantiles with NAIS and Monte Carlo simulations for a multidimensional case.

α	$\hat{q}_\alpha^{MC}$ with $N = 10^6$	MC relative deviation with $N = 10^6$ (%)	$\hat{q}_\alpha^{NAIS}$ with $N = 10^4$	NAIS relative deviation with $N = 10^4$ (%)
0.1	6.49	0.03	6.49	0.7
0.05	6.89	0.03	6.90	0.5
0.01	7.64	0.05	7.63	0.5
10^{-3}	8.46	0.1	8.44	0.5
10^{-5}	9.70	0.6	9.67	0.5
10^{-7}	10.26	3.2	10.6	0.8

Table 7
Estimation of 10^{-5}-quantiles with 10,000 NAIS simulations with different values of ρ for a multidimensional case.

ρ	0.01	0.05	0.1	0.2	0.3	0.5	0.7	0.9
$\hat{q}_\alpha^{NAIS}$	9.66	9.69	9.68	9.66	9.66	9.67	6.66	4.52
Relative deviation	0.5%	0.5%	0.6%	0.5%	0.7%	0.6%	1.6%	1.5%

Table 8
Estimation of 10^{-5}-quantiles with 10,000 NAIS simulations with different values of N_k for a multidimensional case.

N_k	100	200	500	1000	3000	5000
$\hat{q}_\alpha^{NAIS}$	9.47	9.64	9.69	9.65	9.67	9.67
Relative deviation	3%	2%	1%	0.4%	0.2%	1%

Table 9
Evolution of the quantile estimate during the iteration process for two values of ρ.

Iteration number	0	1	2	3	4	5	6	7	8	9
$\hat{q}_{\alpha,k}^{NAIS}$ for $\rho = 0.1$	6.48 (1.1%)	9.16 (3.6%)	9.65 (0.7%)	9.66 (0.7%)	9.66 (0.8%)	9.66 (0.6%)	9.66 (0.6%)	9.66 (0.6%)	9.67 (0.7%)	9.66 (0.6%)
$\hat{q}_{\alpha,k}^{NAIS}$ for $\rho = 0.5$	5.08 (0.8%)	6.19 (1.0%)	6.95 (0.9%)	7.37 (1.2%)	7.95 (1.3%)	8.52 (1.3%)	9.13 (1.3%)	9.64 (1.7%)	9.66 (0.6%)	9.67 (0.7%)

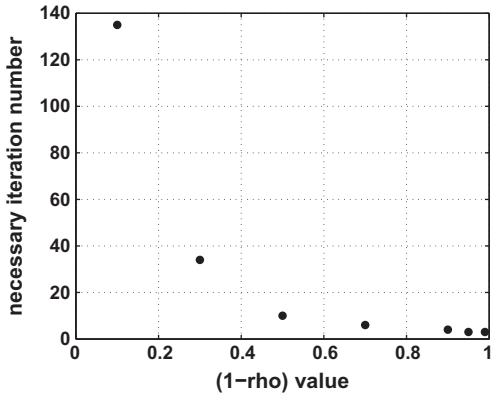


Fig. 6. Necessary number of iterations to reach NAIS convergence in function of the value of ρ.

covariance matrix equal to $\frac{1}{d} I_d$. The output Y is simply given by $Y = \sum_{i=1}^{d} X_i$. We then estimate the 10^{-5}-quantile of Y using NAIS algorithm for different values d with 10,000 simulations. This quantile is theoretically constant whatever the value of d and equal to 4.26. The results given by NAIS are presented in Table 10. One of the NAIS limits is the impact of problem dimension on the algorithm performances. When $d > 9$, applying NAIS with 10,000 simulations is not efficient because of the difficulty to obtain a valuable nonparametric sampling density with few samples. This conclusion is inherent to NAIS algorithms [29]. In that case, a higher number of samples has to be considered to obtain valuable estimations. A recent approach proposed in [29] based on sensitivity analysis of the inputs is a solution to this problem.

Table 10
Estimation of 10^{-5}-quantiles with 10,000 NAIS simulations for different values d.

d	2	3	4	5	6	7	8	9	10	11	15
$\hat{q}_\alpha^{NAIS}$	4.26	4.27	4.27	4.26	4.26	4.26	4.26	4.27	4.05	3.5	3.2
Relative deviation	0.2%	0.2%	0.2%	0.2%	0.3%	0.5%	0.6%	5.1%	9.8%	12.5%	18.7%

5. Application on onboard sensor orientation analysis for fire forest detection simulator

A simulation platform ϕ has been developed at ONERA that enables the estimation of forest fire temperature T from space vehicles [45] with infrared images. In this article, we focus on the case where three satellites with a geocentric orbit that load onboard an infrared sensor try to detect a forest fire start. The infrared sensor is assumed to be able to point different regions of the Earth from its orbit. We call, in the following, position, the region of the Earth that is pointed by the sensor. The ability for a sensor to detect a forest fire will depend on the distance between the position pointed by the sensor and the forest fire, the point angle of the sensor, the weather conditions (clouds, wind, etc.). For that purpose, we model with 7 independent random inputs X this uncertainty in the simulation platform. These 7 random inputs are given by:

- Distance between the position pointed by the sensor and the forest fire (correlated Gaussian PDF in 2 dimensions).
- Point angle of the sensor (correlated Gaussian PDF in 2 dimensions).
- Weather conditions: temperature, radiance, and absorption (independent Gaussian PDF in 3 dimensions).

The forest fire risk is then evaluated with $|T - T_0| = \phi(X)$ where T is the temperature estimated by the simulator and T_0 the real temperature for a given scenario. If T is strongly greater than T_0, there is a high risk of forest fire. If T is significantly lower than T_0, it means that the sensors are not able to estimate a valuable temperature. Both cases are of interest to analyze the performances of the simulation and the risks on the populations. Nevertheless, most of the time when $|T - T_0| > S$, it is nearly always due to $T > T_0$.

The initial PDF of each parameter X is either Gaussian or exponential PDF. The output of the simulation platform $|T - T_0|$ is assumed to be a random variable with PDF g. In Fig. 7, we present the density estimation of g with a mixture of Gaussian kernels on 10,000 samples of $|T - T_0|$. We also present in Table 11 the total Sobol indices [46,47] estimated on a surrogate model of function ϕ based on a neural network. Sobol indices enable to perform global sensitivity analysis of complex numerical models by calculating variance-based importance measures of the input variables. All the 7 inputs have a significant influence on the output. The effective dimension of the problem is thus also 7 because the Sobol indices'cross-terms are non-negligible.

In Table 12, we present the mean quantile estimation on 10 repetitions obtained for 10,000 NAIS simulations with $\rho = 0.1$ and $N_k = 1000$. The same kind of results is presented in Table 12 for 10,000 Monte Carlo simulations. We also propose to compare these results with two different parametric adaptive importance sampling algorithms based on cross-entropy (CE) [39–41] optimization and on Stochastic Approximation (AS) and stationary points [15]. We propose to use a Gaussian parametric sampling PDF $N(\mu, \Sigma)$ where μ is the mean vector and Σ is the covariance matrix in both cases. As stated in [15,39], one has to determine the parameter μ^* so that:

- μ^* minimizes the Kullback–Leibler distance between $N(\mu, \Sigma)$ and h_{opt} in the case of cross-entropy optimization. One has thus:

$$\mu_{CE}^* = \underset{\mu}{\operatorname{argmin}}(K(h_{opt}, N(\mu, \Sigma)) \tag{29}$$

  where K is the Kullback–Leibler distance.
- μ^* minimizes the variance of the weighted CCDF estimate. One has thus:

$$\mu_{AS}^* = \underset{\mu}{\operatorname{argmin}}(Var_{N(\mu,\Sigma)}(\widehat{G}^{IS}(y))) \tag{30}$$

In Table 12, we also present the results obtained for 10,000 parametric IS simulations optimized with CE and AS algorithm. With the same number of samples, NAIS algorithm seems to give valuable results when compared to Monte Carlo or parametric importance sampling. Monte Carlo method underestimates rare quantile because not enough samples are generated. Parametric importance sampling is not very efficient since it is difficult to find a good parametric model of input auxiliary importance sampling PDF. CE and AS give similar results as NAIS with a higher number of samples. To obtain these results, the parametric models considered in CE and AS have to well-fit the importance sampling optimal auxiliary PDF which is not known *a priori* and implies a greater number of samples than NAIS when the chosen parametric models is not exactly adapted.

Runtime execution for a quantile estimation is in fact nearly the same whatever the chosen methods. Indeed, 11 s is necessary to generate a value of $|T - T_0|$ from the simulation platform ϕ for a set of inputs. To generate the 10,000 input–output

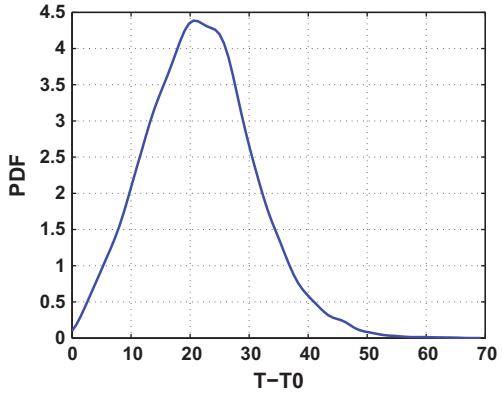


Fig. 7. Kernel density estimation of the output $|T - T_0|$ over 10,000 samples.

Table 11
Mean of first and total Sobol indice estimations over 100 retrials with 50,000 Monte Carlo simulations.

Input	1	2	3	4	5	6	7
First Sobol indices	0.10	0.08	0.07	0.09	0.03	0.08	0.05
Total Sobol indices	0.25	0.21	0.24	0.33	0.21	0.22	0.30

Table 12
Estimation of different α-quantiles with MC, NAIS, CE and AS in the case of fire detection simulator. The computation time required for all the methods is about 30 h.

α	$\hat{q}_\alpha^{MC}$ with $N = 10^4$	MC relative deviation with $N = 10^4$ (%)	$\hat{q}_\alpha^{NAIS}$ with $N = 10^4$	NAIS relative deviation with $N = 10^4$ (%)	$\hat{q}_\alpha^{CE}$ with $N = 10^4$	CE relative deviation with $N = 10^4$ (%)	$\hat{q}_\alpha^{AS}$ with $N = 10^4$	AS relative deviation with $N = 10^4$ (%)
0.1	33.3	0.4	33.2	0.8	33.3	0.7	33.3	0.6
0.05	37.5	0.5	37.4	0.4	37.3	0.7	37.4	0.8
0.01	46.4	0.7	46.6	0.6	46.6	0.7	46.6	0.6
10^{-3}	56.5	1.2	57.2	0.3	57.4	2.2	57.3	2.3
10^{-5}	62.1	4.2	66.4	0.7	67.3	6.2	67.1	5.6
10^{-7}	62.1	4.2	80.7	1.2	79.3	8.1	79.7	7.6

simulations, it requires consequently about 30 h. The sampling method runtime execution is consequently negligible (5 s for parametric methods CE and AS and 35 s for NAIS) and each method requires similarly the same computation time.

6. Conclusion

In this article, we have described a NAIS algorithm for rare quantile estimation. Its purpose is to iteratively approach the IS optimal auxiliary density with a mixture of Gaussian kernels. In this paper, we generalize NAIS algorithm to quantile estimation and also propose an approach that does not rely on an initial guess of g_0. First, we reviewed the principle of IS and then described the process of this algorithm. We finally applied this new NAIS algorithm to different one-dimensional and multidimensional test cases. The NAIS results are compared with Monte Carlo simulations and showed great improvement since the nonparametric density fitted efficiently the optimal auxiliary density.

A limit of NAIS algorithm is the impact of the curse of dimensionality on runtime execution when the input X have a too high dimension. Nevertheless, a recent approach proposed in [29] can be a solution to this problem. NAIS final sampling density can also be used to determine a valuable parametric density family which would be more efficient in a high dimension.

A potential perspective to this work is the improvement of NAIS bandwidth estimation. Indeed, it is currently estimated with the AMISE criterion which is not an optimal method for adaptive importance sampling. Considering a cross-entropy optimization of NAIS bandwidth could improve the general performances of the algorithm.

Acknowledgements

The author thanks Rudy Pastel, Mathieu Balesdent and François Le Gland for fruitful discussions and the anonymous reviewers for their valuable remarks. The author thanks for partial funding the Ministére de l'Economie, de l'Industrie et de l'Emploi (Minister for the Economy, Industry and Employment) in the frame of CSDL (Complex System Design Lab) project of French competitiveness cluster Systematic – Paris Région.

References

[1] P. L'Ecuyer, V. Demers, B. Tuffin, Splitting for rare event simulation, in: Proceeding of the 2006 Winter Simulation Conference, 2006, pp. 137–148.
[2] F. Cerou, A. Guyader, Adaptive particle techniques and rare event estimation, in: Proceedings of ESAIM Conference, vol. 19, 2007, pp. 65–72.
[3] J. Morio, R. Pastel, F. Le Gland, An overview of importance splitting for rare event simulation, European Journal of Physics 31 (2010) 1295–1303.
[4] F. Cerou, P. Del Moral, T. Furon, A. Guyader, Rare event simulation for a static distribution, INRIA report 6792, 2009, pp. 1–18.
[5] S.K. Au, J.L. Beck, Estimation of small failure probabilities in high dimensions by subset simulations, Probabilistic Engineering Mechanics 16 (4) (2001) 263–277.
[6] S.K. Au, J.L. Beck, Subset simulation and its application to seismic risk based on dynamic analysis, Journal of Engineering Mechanics 129 (8) (2001) 1–17.
[7] S.K. Au, Reliability-based design sensitivity by efficient simulation, Computers and Structures 83 (2005) 1048–1061.
[8] S.K. Au, Z.H. Wang, S.M. Lo, Compartment fire risk analysis by advanced Monte Carlo simulation, Probabilistic Engineering Mechanics 29 (9) (2007) 2381–2390.
[9] L. Katafygiotis, S.H. Cheung, A two-stage subset simulation-based approach for calculating the reliability of inelastic structural systems subjected to gaussian random excitations, Computers Methods in Applied Mechanics and Engineering 194 (12–16) (2005) 1581–1595.
[10] L. Katafygiotis, S.H. Cheung, Application of spherical subset simulation method and auxiliary domain method on a benchmark reliability study, Structural Safety 29 (2007) 194–207.
[11] M. Denny, Introduction to importance sampling in rare-event simulations, European Journal of Physics (2001) 403–411.
[12] P.H. Borcherds, Importance sampling: an illustrative introduction, European Journal of Physics (2000) 405–411.
[13] T. Booth, J. Hendricks, Importance estimation in forward Monte-Carlo calculations, Nuclear Technology/Fusion 9 (1984) 90–100.
[14] I. Beichl, F. Sullivan, The importance of importance sampling, Computing in Science and Engineering Approximating the Permanent via Importance Sampling with Application to the Dimer Covering Problem (1999) 71–73.
[15] D. Egloff, M. Leippold, Quantile estimation with adaptive importance sampling, The Annals of Statistics 38 (2010) 1244–1278.
[16] C. Cannamela, J. Garnier, B. Iooss, Controlled stratification for quantile estimation, Annals of Applied Stats 2 (4) (2008) 1554–1580.
[17] P.S. Koutsourelakis, H.J. Pradlwarter, G.I. Schueller, Reliability of sructures in high dimensions, Part I: Algorithms and application, Probabilistic Engineering Mechanics 19 (2004) 409–417.
[18] H.J. Pradlwarter, M.F. Pellissetti, C.A. Schenk, G.I. Schueller, A. Kreis, S. Fransen, A. Calvi, M. Klein, Realistic and efficient reliability estimation for aerospace structures, Computer Methods in Applied Mechanics and Engineering 194 (2005) 1597–1617.
[19] H.J. Pradlwarter, G.I. Schueller, P.S. Koutsourelakis, D.C. Charmpis, Application of line sampling simulation method to reliability benchmark problems, Structural Safety 29 (2007) 208–221.
[20] G.I. Schueller, H.J. Pradlwarter, P.S. Koutsourelakis, A critical appraisal of reliability estimation procedures for high dimensions, Probabilistic Engineering Mechanics 19 (2004) 463–474.
[21] G.I. Schueller, H.J. Pradlwarter, Benchmark study on reliability estimation in higher dimensions of structural systems – an overview, Structural Safety 29 (3) (2007) 167–182.
[22] J. Pickands, Statistical inference using extreme order statistics, Annals of Statistics 3 (1975) 119–131.
[23] S.G. Coles, An Introduction to Statistical Modeling of Extreme Values, Springer, New York, 2001.
[24] J. Bect, D. Ginsbourger, L. Li, V. Picheny, E. Vazquez, Sequential design of computer experiments for the estimation of a probability of failure, Statistics and Computing 22 (3) (2012) 773–793.
[25] V. Picheny, D. Ginsbourger, O. Roustant, R.T. Haftka, N.H. Kim, Adaptive designs of experiments for accurate approximation of target regions, Journal of Mechanical Design 132 (7) (2010) 1–9.
[26] P. Ranjan, D. Bingham, G. Michailidis, Sequential experiment design for contour estimation from complex computer codes, Technometrics 50 (4) (2008) 527–541.
[27] P. Zhang, Nonparametric importance sampling, Journal of the American Statistical Association 91 (434) (1996) 1245–1253.
[28] J.C. Neddermeyer, Computationally efficient nonparametric importance sampling, Journal of the American Statistical Association 104 (2009) 788–802.
[29] J.C. Neddermeyer, Non-parametric partial importance sampling for financial derivative pricing, Quantitative Finance (2010) 1469–7696.
[30] J. Morio, How to approach the importance sampling density, European Journal of Physics 31 (2010) 41–48.
[31] J. Morio, Nonparametric adaptive importance sampling of the probability estimation of launcher impact position, Reliability Engineering and System Safety 96 (1) (2011) 178–183.
[32] B. Silverman, Density estimation for statistics and data analysis, in: Monographs on Statistics and Applied Probability, Chapman and Hall, London, 1986.
[33] C. Robert, G. Casella, Monte Carlo Statistical Methods, Springer, New York, 2005.
[34] T. Dielman, C. Lowry, R. Pfaffenberger, A comparison of quantile estimators, Communications in Statistics – Simulation and Computation 23 (1994) 355–371.
[35] H.A. David, Order Statistics, Wiley, New York, 1981.
[36] B.C. Arnold, N. Balakrishnan, H.N. Nagaraja, A First Course in Order Statistics, Wiley Interscience, New York, 1992.
[37] S. Barber, All of statistics: a concise course in statistical inference, Journal of the Royal Statistical Society Series A 168 (1) (2005) 261.
[38] Y.B. Kim, D.S. Roh, M.Y. Lee, Nonparametric adaptive importance sampling for rare event simulation, in: Proceedings of the Winter Simulation Conference, Orlando, USA, vol. 1, 2000, pp. 767–772.
[39] T.H. de Mello, R.Y. Rubinstein, Rare Event Estimation for Static Models via Cross-Entropy and Importance Sampling, John Wiley, New York, 2002.
[40] R.Y. Rubinstein, Optimization of computer simulation models with rare events, European Journal of Operations Research 99 (1997) 89–112.
[41] R. Rubinstein, The cross-entropy method for combinatorial and continuous optimization, Methodology and Computing in Applied Probability 2 (1999) 127–129.
[42] I.K. Glad, N.L. Hjort, N.G. Ushakov, Mean-squared error of kernel estimators for finite values of the sample size, Journal of Mathematical Sciences 146 (2007) 5977–5983.
[43] M.P. Wand, J.S. Marron, D. Ruppert, Transformations in density estimation, Journal of the American Statistical Association 86 (1991) 343–361.
[44] A. Charpentier, A. Oulidi, Beta kernel quantile estimators of heavy-tailed loss distributions, Statistics and Computing 20 (1) (2009) 35–55.
[45] J. Morio, F. Muller, Onboard sensor orientation analysis with Kriging method, Acta Astronautica 2–3 (2009) 374–381.
[46] A. Saltelli, S. Tarantola, K.-S. Chan, A quantitative model-independent method for global sensitivity analysis of model output, Technometrics 41 (1999) 39–56.
[47] A. Saltelli, Making best use of model evaluation to compute sensitivity indices, Computer Physics Communication 145 (2002) 280–297.

Chapitre 2

Optimisation of interacting particle system for rare event estimation, J. Morio, D. Jacquemart, M. Balesdent and J. Marzat,

Computational Statistics and Data Analysis, 2013, Vol. 66, pp. 117-128

Computational Statistics and Data Analysis 66 (2013) 117–128

Contents lists available at SciVerse ScienceDirect

Computational Statistics and Data Analysis

journal homepage: www.elsevier.com/locate/csda

Optimisation of interacting particle systems for rare event estimation

Jérôme Morio [a,*], Damien Jacquemart [a,b], Mathieu Balesdent [a], Julien Marzat [a]

[a] Onera - The French Aerospace Lab, BP 80100, 91123 Palaiseau Cedex, France

[b] INRIA Rennes, ASPI Applications of Interacting Particle Systems to Statistics Campus de Beaulieu, 35042 Rennes, France

ARTICLE INFO

Article history:
Received 29 October 2012
Received in revised form 19 March 2013
Accepted 25 March 2013
Available online 1 April 2013

Keywords:
Rare event
Probability
Interacting particle system
Optimisation
Hyperparameter
Surrogate model
Input–output model
Kriging

ABSTRACT

The interacting particle system (IPS) is a recent probabilistic model proposed to estimate rare event probabilities for Markov chains. The principle of IPS is to apply alternatively selection and mutation stages to a set of initial particles in order to estimate probabilities or quantiles more accurately than with usual estimation techniques. The practical issue of IPS is the tuning of a parameter in the selection stage. Kriging-based optimisation strategy with a low simulation cost is thus proposed in order to minimise the probability estimate relative error. The efficiency of the proposed strategy is demonstrated on different test cases.

1. Introduction

Rare event estimation has become a large area of research in the reliability engineering and system safety domain. Importance sampling and importance splitting are the most well-known rare event simulation techniques for time-dependent complex systems. Importance sampling (Glynn and Iglehart, 1989; Juneja and Shahabuddin, 2001) consists in changing the initial probability distribution so that the rare event appears more frequently. The determination of an efficient auxiliary sampling density is far from obvious and often requires an optimisation stage and also some *a priori* knowledge of the complex system. For a general Markov process with continuous state space, there does not seem to be a method that determines an efficient measure of importance sampling, because of the very broad variety of instances that can be involved. It is thus difficult to apply importance sampling in realistic cases for Markov processes with continuous state space.

The idea of importance splitting (also called subset simulation, subset sampling or sequential Monte Carlo) is to express the sought probability as a product of conditional probabilities that can be estimated within a reasonable computation time. These probability estimates are obtained by applying several stages of selection/mutation on an interacting particle population. The particles most likely to reach the rare event are duplicated (mutation stage) and the others are killed (selection stage). Particle filter algorithms also consider the same principles: the particles that suit best the observations are duplicated and the others are killed. This idea has been first proposed in a physical context in 1951 (Kahn and Harris, 1951), and numerous variants have been worked out since. Many particle models such as sampling branching models, excursion space exploration, branching particle systems, or related multi-level techniques have been proposed based on importance

* Corresponding author. Tel.: +33 1 80 38 66 54.
E-mail addresses: jerome.morio@onera.fr, jerome_morio@yahoo.fr (J. Morio).

0167-9473/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.csda.2013.03.025

sampling and splitting approaches. A detailed review of these algorithms is available in Del Moral and Lezaud (2006) and Del Moral (2004).

Interacting particle system (IPS) (Del Moral and Garnier, 2006) is considered here. It is a relatively new algorithm that improves the estimation of rare event probabilities based on evolutionary principles and importance sampling method. IPS is indeed equivalent to an importance splitting algorithm where the interacting particles have a resampling weight. The selection/mutation stages in IPS are nevertheless applied at different times of the Markov process evolution, unlike importance splitting where selection/mutation stages are carried out on the whole Markov trajectories. IPS can thus be interesting when compared to importance splitting if simulating the whole Markov trajectories is time-consuming. Moreover, an efficient IPS sampling weight can be determined through the definition of a selection function, whereas it is not the case for importance sampling. The theoretical efficiency of this approach has been demonstrated notably in Del Moral and Garnier (2006), Carmona et al. (2009) and Carmona and Crépey (2010) for the estimation of a rare event on a Markov chain at a finite time. The first application of interacting particle systems in the context of rare events arose in optical fibre communication (Garnier and Del Moral, 2006). Nevertheless, the practical applicability of IPS highly depends on the tuning of a hyperparameter which strongly influences the IPS convergence. Unfortunately, an initial guess of its value is not accessible for realistic applications such as in industrial cases.

The main existing tools for tuning hyperparameters are cross-validation and its variants (k-fold cross-validation, leave-one-out cross-validation, generalised cross-validation; see Golub et al., 1979). Cross-validation may be used to assess the performance level for a given value of the hyperparameter vector and then an optimisation procedure may be called upon to find the best tuning for these hyperparameters. In Kohavi and John (1995) and Hutter et al. (2007), such approaches have been exploited *via* a discretisation of the hyperparameter space. Bayesian networks have also been advocated in Pavón et al. (2008), by considering previous simulation runs as prior knowledge. Various other techniques have been employed for tuning purpose, such as Monte Carlo simulations (Gold and Sollich, 2003), neural networks (Korniyenko et al., 2006) or evolutionary algorithms (Powell, 2002; Lin et al., 2008). In particular, the use of genetic algorithms should be mentioned with applications in control design (Varsek et al., 1993) or reliability assessment (Gen and Yun, 2006). Nevertheless, these approaches reveal to be computationally intensive, since they require a large exploration of the hyperparameter space. This may be prohibitive when the evaluation of the method performance is achieved *via* costly computer simulations, and even more when the number of those simulations is limited.

Kriging-based global optimisation (Santner et al., 2003) is employed here to provide optimal values of the hyperparameters at a limited computational cost. For example, these tools have been successfully applied to the optimal tuning of fault detection strategies in Marzat et al. (2010). The present paper reports the application of this strategy to the tuning of IPS algorithm, which does not seem to have been addressed in the literature so far. The principle of IPS is first presented and the impact of its tuning parameter on probability estimation is described on a simple case. Section 3 is devoted to the presentation of the automatic tuning methodology. Finally, the proposed optimisation strategy is applied in Section 4 on a realistic case of aircraft conflict probability estimation with very promising results.

2. Interacting particle system

2.1. General problem of rare event estimation

Rare event estimation can be characterised by a threshold exceedance of a real Markovian functional at deterministic time n. Consider a Markov chain $(X_n)_{n\in\mathbb{N}}$ with state space $E \subset \mathbb{R}^d$. It is assumed that the initial law $\mathbb{P}(X_0 \in dx)$ and the transitions kernels $\mathbb{P}(X_j \in dx', X_{j-1} = x)$ for $j = 1, \ldots, n$, have a density with respect to the Lebesgue measure, denoted by $\pi_0(x_0)$ and $\pi_j(x_j|x_{j-1})$. Hence, the density of the random trajectory $(X_0, \ldots, X_n)$ of the chain is given by

$$Q_n(x_0, x_1, \ldots, x_n) = \pi_0(x_0) \prod_{j=1}^{n} \pi_j(x_j|x_{j-1}). \tag{1}$$

The probability of interest is denoted by $\mathbb{P}_n(A)$ and defined by

$$\mathbb{P}_n(A) = \mathbb{P}\{V(X_n) \in A\}, \tag{2}$$

where V is a function from E to $\mathbb{R}$. Over the last few years, different algorithms (Bucklew, 2004; Cérou et al., 2006) have been proposed to estimate the probability $\mathbb{P}_n(A)$ when this probability is rare ($<10^{-4}$). Emphasis is put here on the IPS algorithm (Del Moral and Garnier, 2006; Carmona et al., 2009; Carmona and Crépey, 2010), which is one of the most efficient to solve this kind of problem. The following section reviews its principle.

2.2. Principle

For a discrete-time process, importance sampling algorithms (Bucklew, 2004) can help sampling rare events and improve the probability estimation in terms of variance. They consist in generating random weighted samples from an auxiliary distribution rather than the distributions of interest π_j. Nevertheless, in most cases, the studied system is sufficiently

complicated so that it is impossible for the user to determine properly a valuable auxiliary sampling distribution. Importance splitting (Cérou et al., 2006, 2005; Jasra et al., 2008; Carvalho and Lopes, 2007; Voutilainen and Kaipio, 2005) is another rare event estimation method. Its principle is to express the sought probability as a product of conditional probabilities that can be estimated within a reasonable computation time. However, it is not very adapted to the estimation of rare event in finite time. From past experience, IPS and splitting performances are generally similar.

The idea of IPS is to select and favour, at each iteration time $j = 1, \ldots, n$, the Markov chains that are more likely to reach the rare set A. The IPS algorithm consists in a set of N paths $(X_j^{(i)})_{1\leq j\leq n},\ i = 1, 2, \ldots, N$. The initial generation is a set of N samples $X_0^{(1)}, \ldots, X_0^{(N)}$ independently generated from the initial distribution of the chain π_0. The trajectories are updated from j to $j + 1$ to advantage the ones that can potentially reach the rare event A. The IPS algorithm is performed in two steps. First, the selection stage consists in choosing with replacement N paths according to an empirical weighted measure built with a function G_j that depends on $X_{0:j}^{(1)}, \ldots, X_{0:j}^{(N)}$ where $X_{0:j}^{(i)}$ is the vector $(X_0^{(i)}, X_1^{(i)}, \ldots, X_j^{(i)})$. The function G_j has to be positive and favours the paths that are more likely to reach the rare set. Secondly, the mutation stage consists in applying the Markov transition kernel π_{j+1} to the trajectory evolution. The complete algorithm is presented in the following:

(i) Set $j = 0$. Define the sample size N and the positive weight functions G_j. Initialise weights to $W_0^{(1)} = \cdots = W_0^{(N)} = 1$.
(ii) Sample independently $X_0^{(1)}, \ldots, X_0^{(N)}$ from law π_0. A set of N particles $(X_0^{(i)}, W_0^{(i)}),\ i = 1, \ldots, N$, can be formed.
(iii) Estimate the normalising constant

$$\eta_j = \frac{1}{N}\sum_{i=1}^{N} G_j(X_{0:j}^{(i)}).$$

(iv) Choose independently N paths according to the empirical distribution

$$\mu_j(d\tilde{X}, d\tilde{W}) = \frac{1}{N\eta_j}\sum_{i=1}^{N} G_j(X_{0:j}^{(i)})\delta_{\left(X_j^{(i)}, W_j^{(i)}\right)}(d\tilde{X}, d\tilde{W}).$$

The selected particles are denoted by $(\tilde{X}_j^{(i)}, \tilde{W}_j^{(i)})$ for $i = 1, \ldots, N$.
(v) For $i = 1, \ldots, N$, the particle $(\tilde{X}_j^{(i)}, \tilde{W}_j^{(i)})$ is transformed into $(X_{j+1}^{(i)}, W_{j+1}^{(i)})$ in the following way:

$$\tilde{X}_j^{(i)} \underset{\pi_{j+1}}{\rightarrow} X_{j+1}^{(i)},$$

where the mutation are performed independently, and

$$W_{j+1}^{(i)} = \tilde{W}_j^{(i)} \times \left(G_j(\tilde{X}_{0:j})\right)^{-1}.$$

(vi) If $j < n$, then go back to stage (iii).
(vii) Estimate $\mathbb{P}_n(A) \approx \widehat{p}_{\text{IPS}}$ with

$$\widehat{p}_{\text{IPS}} = \prod_{j=1}^{n-1}\eta_j \times \left(\frac{1}{N}\sum_{i=1}^{N}\mathbb{1}_A(X_n^{(i)})\left((W_n^{(i)})^{-1}\right)\right).$$

The IPS probability estimate is unbiased as demonstrated in Del Moral and Garnier (2006). The original paper Del Moral and Garnier (2006) also proposes to use one of the two following weighted functions G_j, written here with respect to the present paper problem formulation:

$$G_j^{\beta}(X_{0:j}) = \exp(\beta V(X_j)) \quad \text{for some } \beta > 0, \tag{3}$$

and

$$G_j^{\alpha}(X_{0:j}) = \exp(\alpha(V(X_j) - V(X_{j-1}))) \quad \text{for some } \alpha > 0. \tag{4}$$

The function G_j^{α} often gives better results (Del Moral and Garnier, 2006) in practice and the function G_j^{β} will not be considered in the following. Note that using these functions does not require to store the whole trajectories $X_{0:j}^{(i)}$ in memory but only $X_j^{(i)}$ or $X_{j-1}^{(i)}$ at each step.

IPS optimal tuning is nevertheless needed when one uses the functions G_j^{β} or G_j^{α}. In general, for two different probability estimations on the same Markov chain, the performance of IPS algorithm is different for the same values of α (or β). So, although asymptotic variance expression can be computed (Del Moral and Garnier, 2006), optimal values of these parameters depend on the unknown probability to estimate.

2.3. Application and tuning problem

Let us consider the following toy case. The Markov chain considered is the Gaussian random walk $X_{j+1} = X_j + W_{j+1}$ where the $W_{j=1,\ldots,n}$ are i.i.d. (independent and identically distributed) Gaussian random variables with zero mean and variance 1

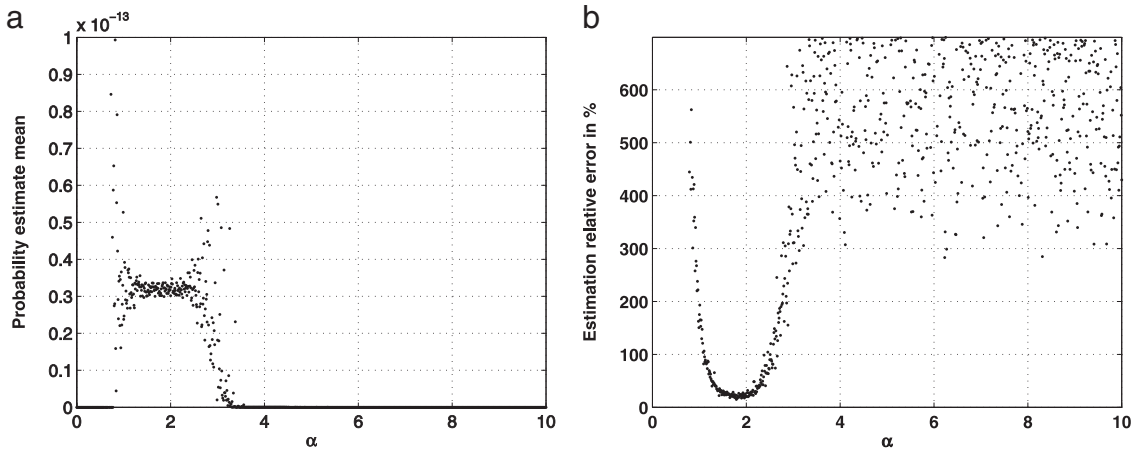


Fig. 1. IPS probability estimate mean and relative error for different values of α and $\mathbb{P}\{X_n > 30\}$.

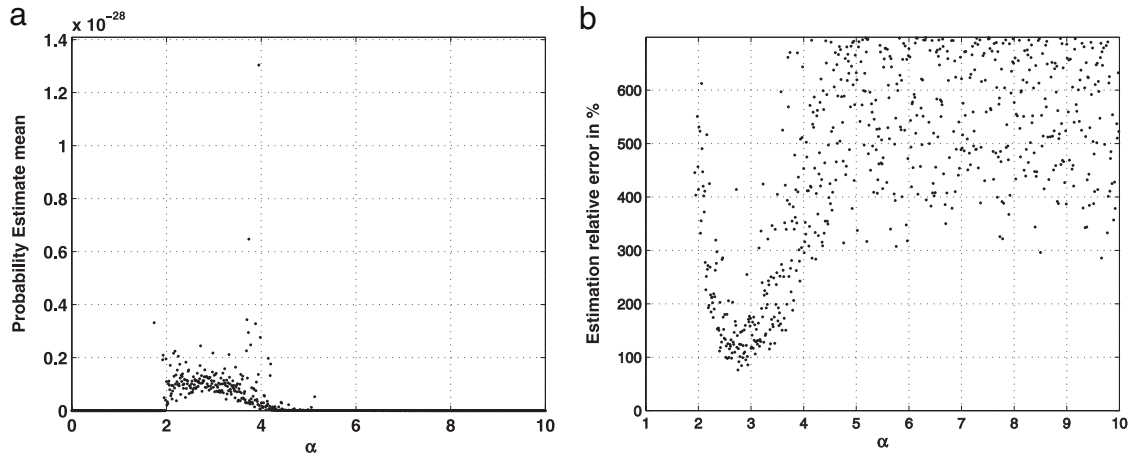


Fig. 2. IPS probability estimate mean and relative error for different values of α and $\mathbb{P}\{X_n > 45\}$.

and with $X_0 = 0$. One assumes in the following that $n = 16$ and one wants to estimate $\mathbb{P}\{X_n > 30\}$. The IPS algorithm with weight function given in Eq. (4) is applied for different values of α and with $N = 2 \cdot 10^4$ particles. These estimations have been repeated 50 times to determine IPS estimate mean and relative error where the IPS relative error is defined by the ratio between IPS estimate standard deviation and IPS estimate mean. Fig. 1(a) shows the IPS probability estimate mean, and the relative error is given in Fig. 1(b). The theoretical probability is $3.18 \cdot 10^{-14}$. The IPS algorithm converges to the theoretical probability even if there is a high variability of the IPS results depending on the value of α. Moreover, the choice of α depends on the sought probability. Indeed, the same experiments have been performed to estimate $\mathbb{P}\{X_n > 45\}$. As shown in Fig. 2, the optimal parameter α is equal to 2.75, which means the tuning of α is only valuable for a given probability estimation. A wrong choice for this value could thus lead to erroneous results. In very simple cases, it is possible to determine theoretically the optimal value of α as shown in Del Moral and Garnier (2006) such as the Gaussian random walk. Nevertheless, in realistic situations, it is not possible to determine the optimal value α for the IPS algorithm and it is difficult to use IPS efficiently since it can lead to a strong misestimation. In industrial applications, it is not possible to sample as many Markov chains as in this simple case to tune the parameter α. It is thus necessary to use an efficient optimisation algorithm with a low simulation cost. In the following, one proposes an optimisation methodology to determine valuable values of the IPS hyperparameter α at a relatively low computation cost.

3. Optimal determination of hyperparameters

The method proposed to determine an optimal IPS hyperparameter is based on design and analysis of computer experiments (DACE) methods. The objectives of DACE are as follows:

- to reduce the uncertainty in the experimental area, Gramacy and Lee (2009).
- to find extrema, Jones et al. (1998).

- to find contours, Ranjan et al. (2008).
- to find boundaries that are contours with large gradients, Banerjee and Gelfand (2006).

In this article, simulation is used to estimate the most efficient hyperparameter α of the IPS system being simulated, that is, the hyperparameter α that gives the lowest IPS relative error. A recent and efficient DACE method is considered for minimising the IPS relative error.

Consider the numerical simulation of a representative test case on which IPS to be tuned is applied. One can define a black-box function where the hyperparameter value α is an input that should return as an output a scalar performance index s. In the case of IPS, s is the IPS probability relative error. IPS probability relative error is defined by the ratio between IPS estimate standard deviation and IPS estimate mean. s is computed by repetitive estimations of the probability given by the IPS model. The hyperparameter α is only assumed to belong to a known bounded set $\mathbb{F}$. The tuning problem could then be formalised as the search for the optimal hyperparameter such that

$$\widehat{\alpha} = \arg\min_{\alpha\in\mathbb{F}} s(\alpha). \tag{5}$$

This is a difficult problem, since the only available information is the performance value at sampled locations of α. The proposed tuning methodology uses a Kriging model to approximate the unknown mapping from the hyperparameter space $\mathbb{F}$ to the performance criterion $s(\cdot)$. The so-called Efficient Global Optimization (EGO) algorithm (see Jones et al., 1998) is then employed to explore areas of the hyperparameter space that might lead to best performance and finally find an optimal tuning. The proposed strategy is now detailed.

Consider that a small number, m, of possible α tunings have already been experimented, resulting in the set $A_m = \{\alpha_1, \ldots, \alpha_m\}$. The corresponding performance indices are gathered in the m-dimensional vector $\mathbf{s}_m = [s(\alpha_1), \ldots, s(\alpha_m)]^{\mathrm{T}}$. Based on this initial knowledge, Kriging (Matheron, 1963) makes it possible to build a surrogate model $\widehat{S}$ of the black-box function between α and s, by modelling it as a Gaussian process $S(\cdot)$ with mean function mean $(\cdot)$ and covariance function cov $(\cdot, \cdot)$ (Lefebvre et al., 1996). One of the advantage of Kriging is the ability to estimate the variance $\widehat{\sigma}^2$ of the prediction error, that is the confidence level on the surrogate model $\widehat{S}$. Kriging-based optimisation algorithms achieve a trade-off between local search (near the best known optimum) and global search (locations where the uncertainty on the surrogate is strong) (Jones, 2001). One of these strategies is the EGO procedure (Jones et al., 1998), which proceeds as follows:

(i) Sample m hyperparameter values α to compute A_m and $\mathbf{s}_m$.
(ii) Fit a Kriging model on A_m and $\mathbf{s}_m$.
(iii) Find $s^m_{\min} = \min_{i=1,\ldots,m}\{s(\alpha_i)\}$.
(iv) Find $\widehat{\alpha}_{m+1} = \arg\max_{\alpha\in\mathbb{F}} EI(\alpha, s^m_{\min}, \widehat{S}, \widehat{\sigma})$.
(v) Append $\widehat{\alpha}_{m+1}$ to A_m and $s\,(\alpha_{m+1})$ to $\mathbf{s}_m$.
(vi) If $m > m_{\max}$, return $s^m_{\min}$ as the minimum relative error found.
Else, go to Step (ii) with $m \leftarrow m + 1$.

The working principle of this iterative algorithm is to replace the initial intractable optimisation problem (5) by the repeated optimisation of a much lighter function called the *Expected Improvement* (EI at Step (iv)), whose closed-form expression is defined by Schonlau (1997)

$$\mathrm{EI}(\alpha, s^m_{\min}, \widehat{S}, \widehat{\sigma}) = \int_{-\infty}^{s^m_{\min}} \left[s^m_{\min} - \widehat{S}\right] h\left[\widehat{S}\right] d\widehat{S}, \tag{6}$$

where $h\left[\widehat{S}\right]$ is the distribution of $\widehat{S}$. EI assumes that this distribution is a Gaussian distribution with the estimated mean $\widehat{S}$ and a standard deviation equal to the estimated predictor standard deviation $\widehat{\sigma}$. The previous equation then becomes

$$\mathrm{EI}(\alpha, s^m_{\min}, \widehat{S}, \widehat{\sigma}) = \left(s^m_{\min} - \widehat{S}(\alpha)\right) \Psi\left(\frac{s^m_{\min} - \widehat{S}(\alpha)}{\widehat{\sigma}(\alpha)}\right) + \widehat{\sigma}(\alpha)\psi\left(\frac{s^m_{\min} - \widehat{S}(\alpha)}{\widehat{\sigma}(\alpha)}\right), \tag{7}$$

in which ψ and Ψ are respectively the probability density and cumulative distribution functions of the normal distribution. This function is computationally light, since it only involves the Kriging linear prediction and standard deviation. It could thus be optimised at each step *via* an auxiliary algorithm to be chosen (in our experiments, DIRECT Jones et al., 1993 was used). Thanks to the guidance of this sampling criterion, several optimal candidate points are explored iteratively, which provides in the end a set of appropriate hyperparameters α for the application considered. The stopping condition of EGO is $m_{\max}$, which is the budget allotted for the black-box evaluations (it could also be expressed in terms of computation time). Note that other stopping criteria could also be considered, such as a thresholds on the successive maximum values of EI obtained at Step (iv) (Schonlau, 1997).

Computation cost in simulation time or in IPS generated trajectories are highly correlated since the Kriging optimisation computation time can be neglected when compared to IPS simulation one. Considering time or IPS generated trajectories for computation cost is equivalent. The computational cost C in IPS generated trajectories of the proposed Kriging-based IPS

Table 1
General results of hyperparameter α optimisation on a toy case.

	$\alpha_{\min}$	$s^m_{\min}$ (%)
Average result	1.93	21.3
Worst result	2.01	23.5
Best result	1.92	18.8

can be derived in the following way. Indeed, there is $m_{\max}$ Kriging evaluations of s with the proposed strategy. Moreover, each estimation of the relative error s requires n_{rep} repetitive IPS estimations with N particles. The computational cost in IPS generated trajectories is then simply given by the following equation:

$$C = m_{\max} \times n_{\text{rep}} \times N. \tag{8}$$

In the following, two IPS runs have thus the same computation cost if the same number of IPS particle trajectories are generated. The parameter n_{rep} has been set to 50 for the different simulation cases but it can be decreased for computational reasons in realistic complex situations. Moreover, the proposed set $\mathbb{F}$ is defined as $[0,\ 10]$ in the simulation cases. This interval is well-adapted to general situations, although low probabilities ($\mathbb{P}_n(A) < 10^{-20}$) or small Markov chain length ($n < 5$) may require an expansion of the set $\mathbb{F}$.

4. Application of the proposed strategy

In this section, we apply the methodology proposed in the previous section to a toy case and a realistic situation of aircraft conflict.

4.1. Toy case

The toy case described in Section 2.3 is considered with the optimisation strategy described in Section 3. The number m of initial values of α where the IPS relative error is estimated is set to 5 in order to decrease the computation cost. These values are determined using random Latin Hypercube Sampling (LHS) (McKay et al., 1979). The number $m_{\max}$ of authorised α samples is 15. The relative error s and the probability mean estimate are estimated with $n_{\text{rep}} = 50$ repetitive estimations of the probability of interest. An example of IPS relative error optimisation is given in Fig. 3. In this case, the algorithm finds an IPS minimum relative error equal to 22% obtained for $\alpha_{\min} = 1.96$ where $s(\alpha_{\min}) = s^m_{\min}$. This tuning value is very efficient and a low computation effort has been required to determine this approximated optimum. It is thus an improvement for the users of IPS algorithm since an initial guess of α is not obvious. We repeated this optimisation 20 times with different initial LHS and estimated the average, worst and best performances of the optimisation strategy in Table 1. In the worst case of these 20 repetitions, the IPS relative error obtained with the proposed tuning is lower than 25%.

It is also interesting to compare the results given by single runs of IPS without using the proposed Kriging based approach but with a large number of particles to keep the computational cost constant. In these IPS single run simulations, the number of particles is set to $m_{\max} \times N = 3 \cdot 10^5$. The number of repeated simulations to estimate the probability and its relative error is kept to $n_{\text{rep}} = 50$. The corresponding probability and relative error estimate are given in Fig. 4. If the initial choice of α is included in the interval $[1.15;\ 2.45]$, that is, if the chosen value of α is not too far from its optimum value, then it is better in terms of relative error to consider IPS simulations with a high number of particles than the proposed Kriging-based IPS optimisation. For other values of α, the algorithm proposed in this article is more efficient.

4.2. Aircraft trajectory model

A model for computing conflict probability estimation between two aircraft is presented in this section. It is based on realistic measurements (Paielli and Field, 1998) of 9500 aircraft flying at 29 000 ft over a period of 20 min following a straight line, with a constant nominal speed of 10 nmi/min (nmi = nautical miles) (Jacquemart and Morio, 2013).

An aircraft trajectory is modelled by a 2-dimensional stochastic process in discrete time defined by

$$X_{j+1} = X_j + vh + \sqrt{h}\sigma_j^h W_n,$$

where X_j is the position of the aircraft at time j, v is a two-dimensional speed vector, h is the discretisation time step, σ_j^h is a squared correlation matrix and W_j is a standard two-dimensional Gaussian random variable. The initial position error is given by $X_0 \sim \mathcal{N}(0,\ \Lambda_0)$, for a given covariance matrix Λ_0.

The cross-correlation between the along-track and cross-track errors is small enough to be considered as null. The vertical position error can be negligible (Paielli and Erzberger, 1997) when compared to along-track and cross-track position errors, which is why the trajectory model considered is two-dimensional. Let us define $X_j = (X_j^a, X_j^c)$ the aircraft position at time j with X_j^a the along-track aircraft position and X_j^c the cross-track aircraft position. If it is assumed that the aircraft has a

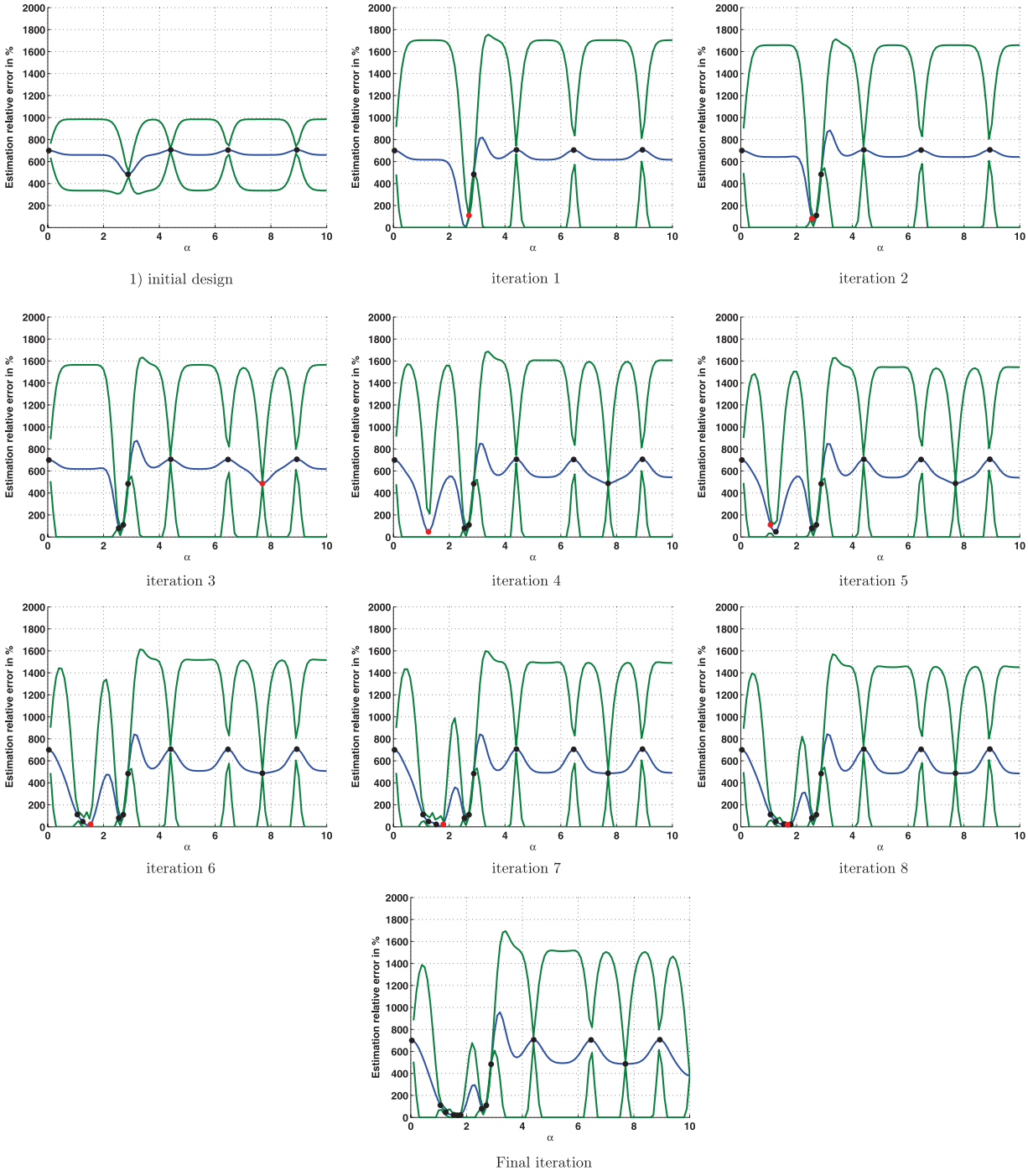


Fig. 3. Evolution of the IPS relative error during the optimisation of hyperparameter α with the EGO procedure. (Blue curves: $\widehat{S}$; green curves: $\hat{S} \pm \widehat{\sigma}$; black points: the current design; red points: new proposed samples.) (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

constant speed v in a coordinate system where the abscissa axis is parallel to aircraft trajectory, and taking into account the independence between the along-track and cross-track errors, one has

$$\begin{cases} X^a_{j+1} = X^a_j + vh + \sqrt{h} g^a(jh) W^a_j \\ X^c_{j+1} = X^c_j + \sqrt{h} g^c(jh) W^c_j, \end{cases} \tag{9}$$

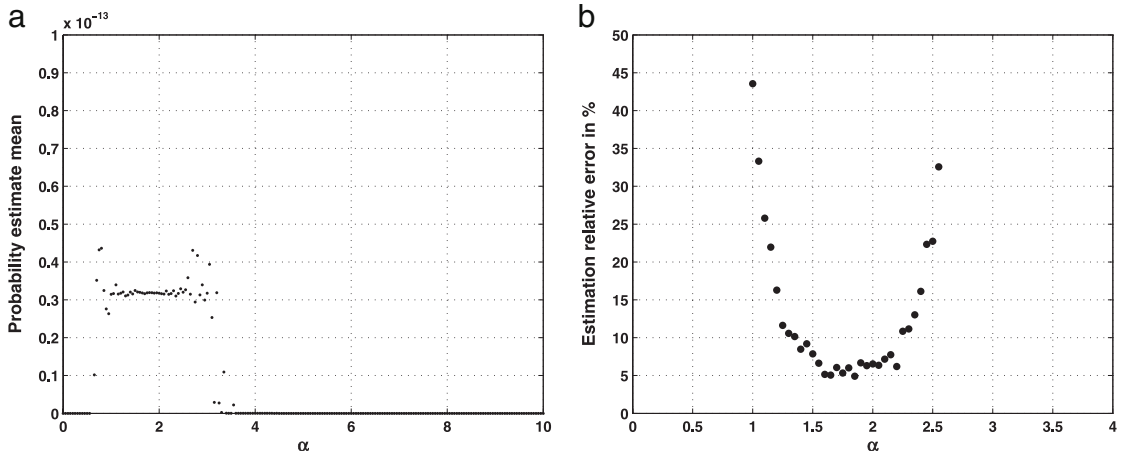


Fig. 4. IPS probability estimate mean and relative error for different values of α, $\mathbb{P}\{X_n > 30\}$ and 300 000 IPS particles.

where h is the discretisation step and W_j^a et W_j^c are two independent standard one-dimensional Gaussian random variables, with initial conditions

$$\begin{cases} X_0^a \sim \mathcal{N}(0, \nu_{0,a}^2) \\ X_0^c \sim \mathcal{N}(0, \nu_{0,c}^2). \end{cases} \tag{10}$$

The initial conditions are experimentally set to Jacquemart and Morio (2013)

$$\nu_{0,a} = \nu_{0,c} = \nu_0 = 0.333. \tag{11}$$

The functions g^a and g^c are defined as

$$g^a(x) = \sqrt{2(r_a^2 x + r_a \nu_0)}, \tag{12}$$

and

$$g^c(x) = \sqrt{2(2p^2x^3 + 3pqx^2 + (2p\nu_0 + q^2)x + 2q\nu_0 x)} \tag{13}$$

with $x \in \mathbb{R}^+$ and $r_a = 0.233$, $p = 0.0068$ and $q = 0.328$ (Jacquemart and Morio, 2013).

Now, if the speed vector has an angle θ with an horizontal axis, the aircraft position is obtained by applying a rotation matrix of angle θ. One has then

$$X_{j+1} = X_j + R_\theta \begin{pmatrix} vh \\ 0 \end{pmatrix} + \sqrt{h} R_\theta \begin{pmatrix} g^a(jh) & 0 \\ 0 & g^c(jh) \end{pmatrix} W_j, \tag{14}$$

with $R_\theta = \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix}$. The term X_0 is defined by

$$X_0 = R_\theta \begin{pmatrix} \nu_{0,a} & 0 \\ 0 & \nu_{0,c} \end{pmatrix} W_0, \tag{15}$$

where W_0 is the 2-dimensional standard Gaussian random variable.

4.2.1. Conflict probability between aircraft

If one considers two aircraft indexed by 1, 2, the two following stochastic processes define the aircraft trajectories:

$$\begin{cases} X_{j+1}^1 = X_j^1 + R_{\theta 1} \begin{pmatrix} v_1 h \\ 0 \end{pmatrix} + \sqrt{h} R_{\theta 1} \begin{pmatrix} g^a(jh) & 0 \\ 0 & g^c(jh) \end{pmatrix} W_j^1 \\ X_{j+1}^2 = X_j^2 + R_{\theta 2} \begin{pmatrix} v_2 h \\ 0 \end{pmatrix} + \sqrt{h} R_{\theta 2} \begin{pmatrix} g^a(jh) & 0 \\ 0 & g^c(jh) \end{pmatrix} W_j^2. \end{cases} \tag{16}$$

The process to be studied is the distance between the two aircraft defined by

$$D_j = \|X_j^1 - X_j^2\|, \quad j \in \{0, \ldots, n\}. \tag{17}$$

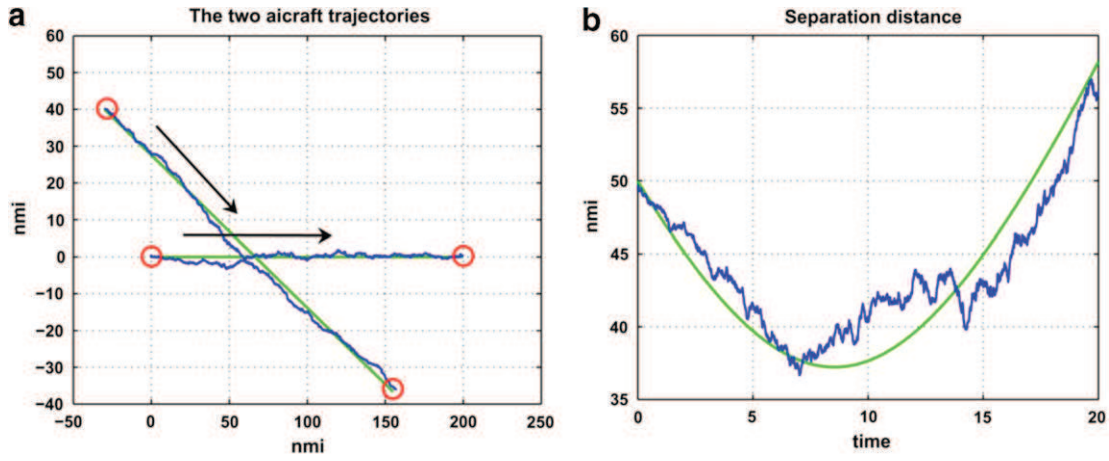


Fig. 5. (a) Top view of a sample of the two aircraft trajectory. (b) Associated separation distance.

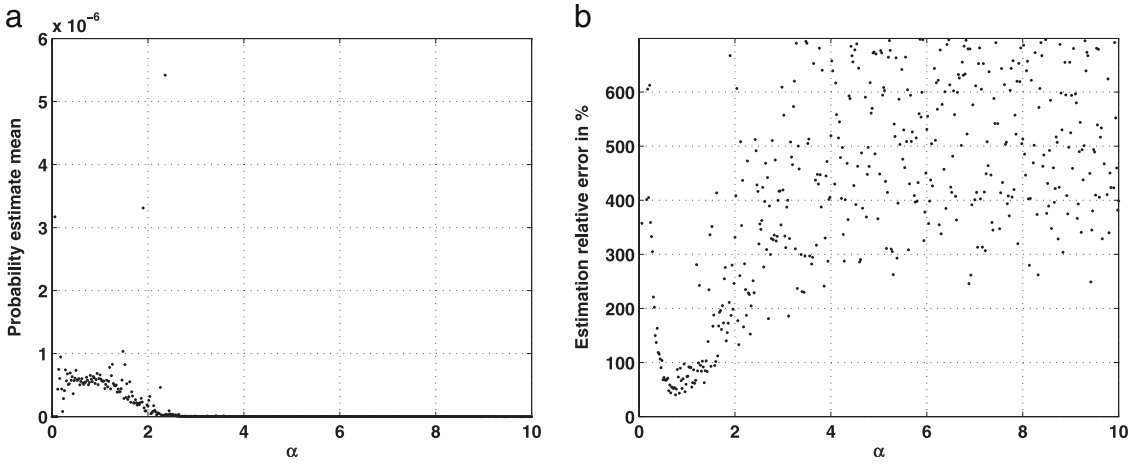


Fig. 6. IPS probability estimate mean and relative error for different values of α.

where n is set in the proposed case to $n = 200$ which is equivalent to a period of 20 min. The probability of interest in this application is

$$\mathbb{P}(\exists j \in \{0, \ldots, n\}, D_j < \delta), \tag{18}$$

where δ is a specified distance threshold. For instance, setting δ as the sum of the two semi-wingspans of the aircraft corresponds to the probability of collision. When one talks about conflict probability, δ is set to 8 *nmi* in this article. Fig. 5 shows a sample of a couple of trajectories with a selected flight plan. The problem defined in Eq. (18) is not exactly the same as the one described in Section 2.1 but can be modified to fit in this formalism. The idea is to introduce the minimum process M_j of the Markov chain D_j before time j,

$$M_j = \min_{l=0,\ldots,j} D_l.$$

The chain M_j is not Markovian. It is thus necessary to apply IPS algorithm on the process S_j

$$S_j = (X_j^1, X_j^2, M_j), \tag{19}$$

which is Markovian and then to estimate the probability $\mathbb{P}(M_n < \delta)$ since

$$\mathbb{P}(M_n < \delta) = \mathbb{P}(\exists j \in \{0, \ldots, n\}, D_j < \delta).$$

The IPS algorithm with weight function given in Eq. (4) is applied for different values of α and with $N = 2 \cdot 10^4$ particles to estimate the probability $\mathbb{P}(M_n < \delta)$. These estimations have been repeated 50 times to determine IPS estimate mean and relative error *s*. Fig. 6(a) shows the IPS probability estimate mean and IPS relative error is given in Fig. 6(b). The probability estimated with a $5 \cdot 10^6$ Monte Carlo simulations is equal to $0.6 \cdot 10^{-6}$, to be used as a reference value. The IPS algorithm converges to this probability even if there is a high variability of the IPS results depending on the value of α. A wrong choice

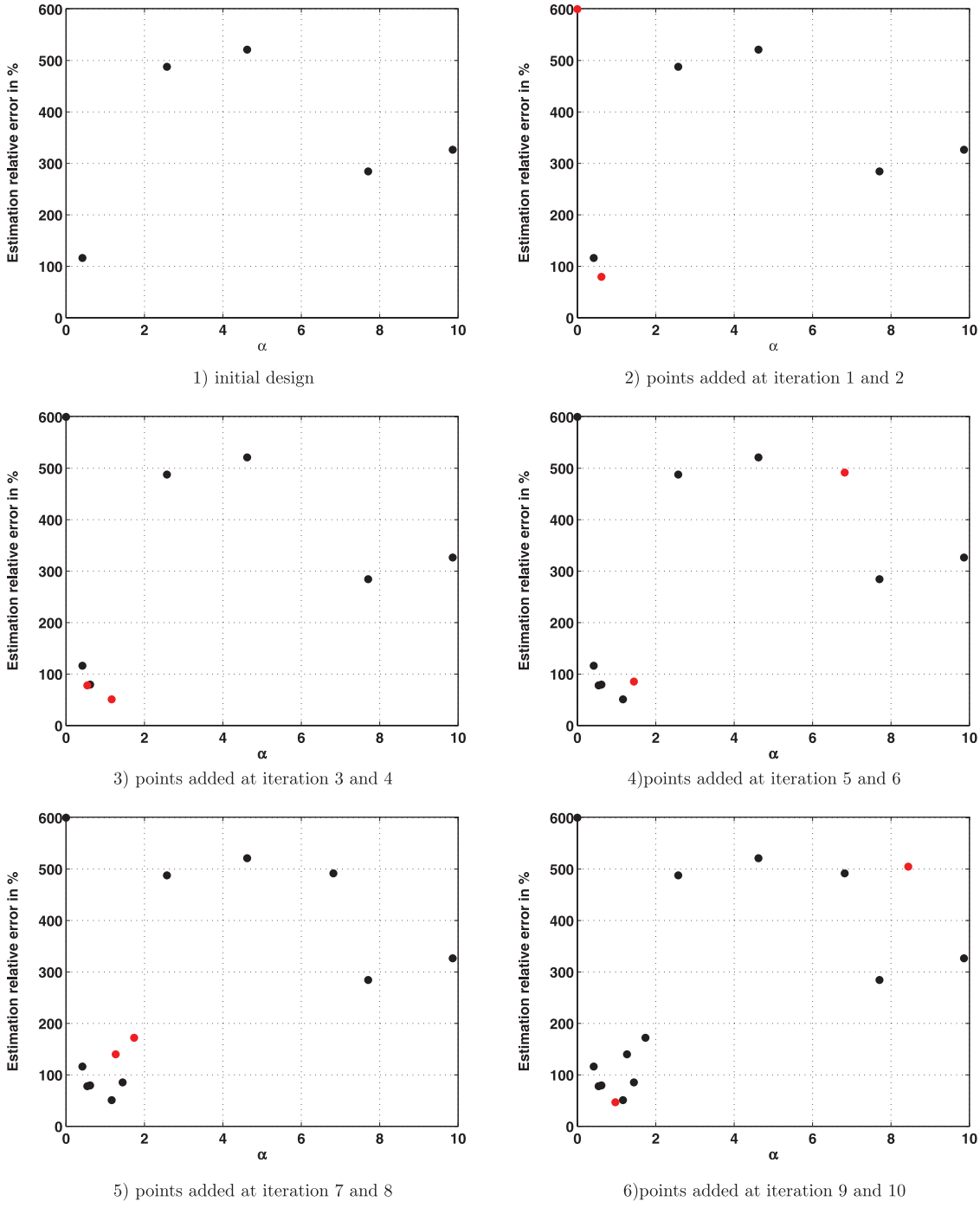


Fig. 7. Evolution of the IPS relative error during the optimisation of hyperparameter α for a realistic case of aircraft conflict with the EGO procedure. (Black points: the current design; red points: new proposed samples.) (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

of this value could lead to erroneous results and moreover only a very small interval of α values gives valuable results. In this realistic case of aircraft conflict, it is impossible to determine theoretically an optimal value of α for the IPS algorithm which may reduce its applicability. If one uses the α optimisation strategy described in Section 3, one can obtain the results given in Fig. 7. The number m of initial values of α, where the IPS relative error is estimated, is 5 and they are determined using random LHS. The number $m_{\max}$ of authorised α samples is still equal to 15. In this case, the algorithm finds an IPS minimum relative error equal to 51% obtained for $\alpha_{\min} = 0.91$. This tuning value is thus very efficient. We repeated this optimisation 20 times with different initial LHS and estimated the average, worst and best performance of the optimisation strategy in

Table 2
General results of hyperparameter α optimisation on a realistic aircraft conflict.

	$\alpha_{\min}$	$s^m_{\min}$ (%)
Average result	0.89	59
Worst result	1.02	70
Best result	0.93	47

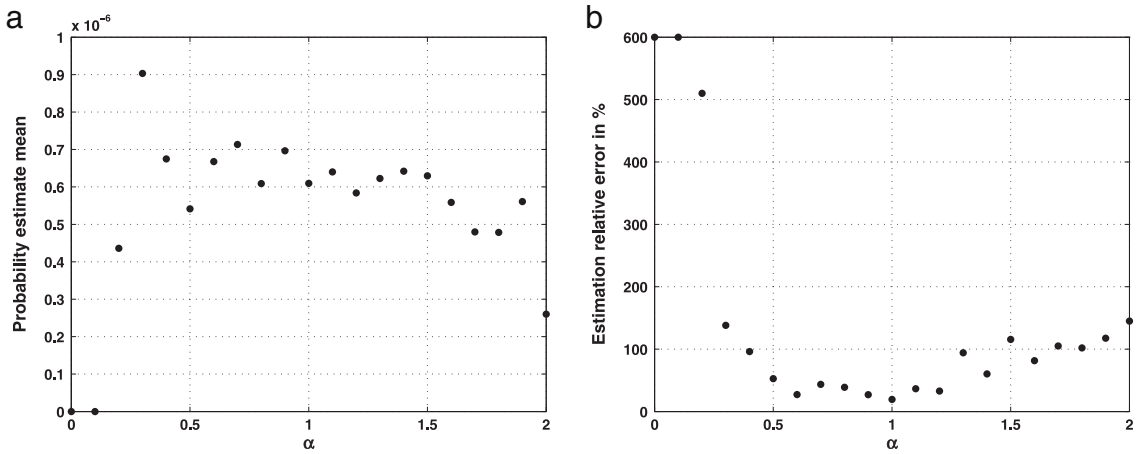


Fig. 8. IPS probability estimate mean and relative error for different values of α and 300 000 IPS particles.

Table 2. In the worst case of these 20 repetitions, the IPS relative error obtained with the proposed tuning is lower than 70%. It is the best result one can obtain with the IPS method for such a low probability estimate and this limited number N of trajectory simulations. The only way to reduce the estimate relative error is no more the improvement of the α tuning but the increase of N.

The results given by single runs of IPS without using the proposed Kriging-based approach but with a large number of particles can be compared. The computational cost is also kept constant. In these IPS single run simulations, the number of particles is set to $m_{\max} \times N = 3 \cdot 10^5$. The number of repeated simulations to estimate the probability and its relative error is kept equal to $n_{\text{rep}} = 50$. The corresponding probability and relative error estimate are given in Fig. 8. If the initial choice of α is included in the interval [0.5; 1.3], then it is better in terms of relative error to consider IPS simulations with a high number of particles than the proposed Kriging-based IPS optimisation. For other values of α, the algorithm proposed in this article is more efficient.

5. Conclusion

In this article, a Kriging-based strategy for estimating the optimal IPS hyperparameter has been proposed. This tuning is an issue since it is very complicated to determine *a priori* efficient values for α. This is no more the case with the proposed method that enables a practical use of IPS since it does not require a high simulation cost. This algorithm has been applied successfully to the Gaussian random walk and to a realistic case of aircraft conflict estimation. A possible perspective to this work is the ability to modify the value of α during the IPS process in order to improve the estimation accuracy. For that purpose, there exist some alternative adaptive IPS models based on adaptive resampling procedures (Del Moral et al., 2012; Wolters, 2012). The idea is to use a sequence of weighted approximations up to the first time the product of sampling weights induce some degenerate estimate with respect to some criteria such as weight variance or entropy. It would be thus very interesting to connect the Kriging-IPS algorithm with these more classical resampling procedures.

Another perspective of this approach is to take into account the simulation error. The number of repetitions to estimate s has been set to 50 in this article. Thus, the estimation of relative error is relatively accurate and taking simulation error into account will not change the results of this article. Nevertheless, if one wants to reduce the number of repetitions to estimate s for computation time reason, it becomes necessary to consider the simulation error in the Kriging model. Mean and standard deviation of the relative error are still available since the relative error is estimated with Monte-Carlo estimations. The use of the Kriging model with noisy simulations has been reviewed in Picheny et al. (in press).

Acknowledgements

The work of Damien Jacquemart is financially supported by DGA (Direction Générale de l'Armement) and Onera, The French Aerospace Lab. The authors thank the two anonymous reviewers for their valuable remarks.

References

Banerjee, S., Gelfand, A.E., 2006. Bayesian wombling: curvilinear gradient assessment under spatial process models. Journal of the American Statistical Association 1487–1501.

Bucklew, J.A., 2004. Introduction to Rare Event Simulation. Springer.

Carmona, R., Crépey, S., 2010. Particle methods for the estimation of credit portfolio loss distributions. International Journal of Theoretical and Applied Finance 13, 577–602.

Carmona, R., Fouque, J.P., Vestal, D., 2009. Interacting particle systems for the computation of rare credit portfolio losses. Finance and Stochastics 13, 613–633.

Carvalho, C.M., Lopes, H.F., 2007. Simulation-based sequential analysis of Markov switching stochastic volatility models. Computational Statistics & Data Analysis 51, 4526–4542.

Cérou, F., Del Moral, P., LeGland, F., Lezaud, P., 2005. Limit theormes for the multilevel splitting algorithm in the simulation of rare events. In: 2005 Winter Simulation Conference. IEEE Press, pp. 682–691.

Cérou, F., Del Moral, P., LeGland, F., Lezaud, P., 2006. Genetic genealogical models in rare event analysis. ALEA, Latin American Journal of Probability and Mathematical Statistics 1, 181–203. Paper 01-08.

Del Moral, P., 2004. Feynman–Kac formulae. In: Genealogical and Interacting Particle Systems with Applications. Springer Series: Probability and Applications. New York, USA.

Del Moral, P., Doucet, A., Jasra, A., 2012. On adaptive resampling procedures for sequential Monte Carlo methods. Bernoulli 18, 252–278.

Del Moral, P., Garnier, J., 2006. Genealogical particle analysis of rare events. The Annals of Applied Probability 15, 2496–2534.

Del Moral, P., Lezaud, P., 2006. Branching and interacting particle interpretations of rare event probabilities. In: Blom, H., Lygeros, J. (Eds.), Stochastic Hybrid Systems. In: Lecture Notes in Control and Information Science, vol. 337. Springer, Berlin, Heidelberg, pp. 277–323.

Garnier, J., Del Moral, P., 2006. Simulations of rare events in fiber optics by interacting particle systems. Optics Communications 267, 205–214.

Gen, M., Yun, Y., 2006. Soft computing approach for reliability optimization: state-of-the-art survey. Reliability Engineering & System Safety 91, 1008–1026.

Glynn, P., Iglehart, D., 1989. Importance sampling for stochastic simulations. Management Science 35, 1367–1392.

Gold, C., Sollich, P., 2003. Model selection for support vector machine classification. Neurocomputing 55, 221–249.

Golub, G.H., Heath, M., Wahba, G., 1979. Generalized cross-validation as a method for choosing a good ridge parameter. Technometrics 21, 215–223.

Gramacy, R.B., Lee, H.K.H., 2009. Adaptive design and analysis of supercomputer experiments. Technometrics 51, 130–145.

Hutter, F., Hoos, H.H., Stutzle, T., 2007. Automatic algorithm configuration based on local search. In: Proceedings of the National Conference on Artificial Intelligence, pp. 1152–1160.

Jacquemart, D., Morio, J., 2013. Conflict probability estimation between aircraft with dynamic importance splitting. Safety Science 51, 94–100.

Jasra, A., Doucet, A., Stephens, D.A., Holmes, C.C., 2008. Interacting sequential Monte Carlo samplers for trans-dimensional simulation. Computational Statistics & Data Analysis 52, 1765–1791.

Jones, D.R., 2001. A taxonomy of global optimization methods based on response surfaces. Journal of Global Optimization 21, 345–383.

Jones, D.R., Perttunen, C.D., Stuckman, B.E., 1993. Lipschitzian optimization without the Lipschitz constant. Journal of Optimization Theory and Applications 79, 157–181.

Jones, D.R., Schonlau, M.J., Welch, W.J., 1998. Efficient global optimization of expensive black-box functions. Journal of Global Optimization 13, 455–492.

Juneja, S., Shahabuddin, P., 2001. Splitting-based importance-sampling algorithm for fast simulation of Markov reliability models with general repair-policies. IEEE Transactions on Reliability 50, 235–245.

Kahn, H., Harris, E., 1951. Estimation of particle transmission by random sampling. National Bureau of Standards Applied Mathematical Series 12, 27–30.

Kohavi, R., John, G.H., 1995. Automatic parameter selection by minimizing estimated error. In: Proceedings of the Twelfth International Conference on Machine Learning, pp. 304–312.

Korniyenko, O.V., Sharawi, M.S., Aloi, D.N., 2006. Neural network based approach for tuning Kalman filter. In: Proceedings of the IEEE International Conference on Electro/Information Technology. Lincoln, USA.

Lefebvre, J., Roussel, H., Walter, E., Lecointe, D., Tabbara, W., 1996. Prediction from wrong models: the Kriging approach. IEEE Antennas and Propagation Magazine 38, 35–45.

Lin, S.W., Ying, K.C., Chen, S.C., Lee, Z.J., 2008. Particle swarm optimization for parameter determination and feature selection of support vector machines. Expert Systems with Applications 35, 1817–1824.

Marzat, J., Walter, E., Piet-Lahanier, H., Damongeot, F., 2010. Automatic tuning via Kriging-based optimization of methods for fault detection and isolation. In: Proceedings of the IEEE Conference on Control and Fault-Tolerant Systems, Nice, France, pp. 505–510.

Matheron, G., 1963. Principles of geostatistics. Economic Geology 58, 1246–1266.

McKay, M.D., Beckman, R.J., Conover, W., 1979. A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. Technometrics 21, 239–245.

Paielli, R., Erzberger, H., 1997. Conflict probability estimation for free flight. Journal of Guidance, Control and Dynamics 20, 588–596.

Paielli, R., Field, M., 1998. Empirical test of conflict probability estimation. In: 2nd USA–Europe Air Traffic Management Research and Development Seminar ATM-1998.

Pavón, R., Díaz, F., Luzón, V., 2008. A model for parameter setting based on Bayesian networks. Engineering Applications of Artificial Intelligence 21, 14–25.

Picheny, V., Wagner, T., Ginsbourger, D., 2013. A benchmark of Kriging-based infill criteria for noisy optimization. In: Structural and Multidisciplinary Optimization (in press).

Powell, T.D., 2002. Automated tuning of an extended Kalman filter using the downhill simplex algorithm. Journal of Guidance, Control and Dynamics 25, 901–908.

Ranjan, P., Bingham, D., Michailidis, G., 2008. Sequential experiment design for contour estimation from complex computer codes. Technometrics 50, 527–541.

Santner, T.J., Williams, B.J., Notz, W., 2003. The Design and Analysis of Computer Experiments. Springer-Verlag, Berlin, Heidelberg.

Schonlau, M., 1997. Computer experiments and global optimization. Ph.D. Thesis, University of Waterloo, Canada.

Varsek, A., Urbancic, T., Filipic, B., 1993. Genetic algorithms in controller design and tuning. IEEE Transactions on Systems, Man and Cybernetics 23, 1330–1339.

Voutilainen, A., Kaipio, J.P., 2005. Sequential Monte Carlo estimation of aerosol size distributions. Computational Statistics & Data Analysis 48, 887–908.

Wolters, M.A., 2012. A particle swarm algorithm with broad applicability in shape-constrained estimation. Computational Statistics & Data Analysis 56, 2965–2975.

Chapitre 3

Influence of input PDF parameters of a model on a failure probability estimation, J. Morio, *Simulation Modelling Practice and Theory*, 2011, Vol. 19, Issue 10, pp. 2244-2255

Simulation Modelling Practice and Theory 19 (2011) 2244–2255

Contents lists available at SciVerse ScienceDirect

Simulation Modelling Practice and Theory

journal homepage: www.elsevier.com/locate/simpat

Influence of input PDF parameters of a model on a failure probability estimation

Jérôme Morio *

Onera - The French Aerospace Lab, BP 80100, FR-91123 Palaiseau Cedex, France

ARTICLE INFO

Article history:
Received 14 February 2011
Received in revised form 5 August 2011
Accepted 6 August 2011
Available online 9 September 2011

Keywords:
Sensitivity analysis
Rare event simulation
Monte Carlo methods
Sobol indices
Importance sampling
Importance splitting

ABSTRACT

Critical probability estimation is of major interest in safety and reliability applications. In this article, we focus on a black-box model with multidimensional random input X and one random output Y. We consider the estimation of probability P that Y exceeds a threshold S. We assume that the random input X follows a multidimensional parametric density with parameters δ and thus the probability P will depend on the values of δ. In this paper, we analyze the sensitivity of the critical probability P to the model parameters δ. We propose a methodology that estimates Sobol indices with low computation cost. This strategy enables us to determine which statistical parameters have a great influence on the value of the probability and require a valuable determination. The last part of this article applies the proposed technique on a realistic case of missile collateral damage estimation.

1. Introduction

Simulation and computing take a larger part in the dimensioning of real systems. In this article, we focus on rare failure probability estimation of a black-box model. This estimation is performed with Monte Carlo simulations or with more adapted techniques like importance sampling or importance splitting. These algorithms assume that the input PDF (probability density function) parameters (such as the mean and the variance in the case of Gaussian input density) are perfectly known and well-determined. It is of course not always the case in realistic situations. In this article, we propose to analyze the influence of the uncertainty of input density parameters on a failure probability estimation. It is an original and important issue in reliability and safety since it is often of major interest to determine which input density parameters have to be well estimated in order to not bias the probability estimation. Similar approaches have been proposed in the case of local sensitivity method with importance splitting [1–3], extreme value theory [4,5], fault tree [6,7] or probability bounding [8,9].

For that purpose, we propose a method to estimate Sobol indices of the input PDF parameters that characterizes their influence on the rare event probability. Even if rare event estimation and global sensitivity method are generally well-known, their combined use in this article is not very developed in the scientific literature [10,11]. The main difficulty of sensitivity analysis methods is the high required number of samples generated with the simulation code in order to obtain an accurate estimation. The approach proposed in this article requires some samples to estimate the failure probability for a set of input density parameters. Then, the sensitivity analysis does not need the generation of new samples which is another innovative aspect of this article.

In this article, we describe the problem statement and propose a three stage methodology to analyze the sensitivity of a rare probability estimation to the uncertainty of input density parameters. Sobol indices of each input density parameter are

* Tel.: +33 1 80 38 66 54.
E-mail address: jerome.morio@onera.fr

1569-190X/$ - see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.simpat.2011.08.003

then determined to rank their different influences. The last part of the article concerns the application of this approach on missile collateral damage probability estimation.

2. Problem statement

In this section, we propose to analyze the general context of the proposed study. We consider a black-box system ϕ modeled by a continuous scalar function $\phi : \mathbb{R}^d \to \mathbb{R}$. The input of this function is a d-dimensional random variable $X = (X^{(1)}, X^{(2)}, \ldots, X^{(d)})$ that follows a probability density function f and the output is defined by the variable Y such as $Y = \phi(X)$. Each random variable $X^{(i)}$ follows a PDF $f_i(\delta_i)$ with δ_i, a set of PDF parameters. The PDF parameter δ_i is for instance the variance of a Gaussian PDF. Without any lack of generality, we will suppose to simplify the notation in the following that each input PDF only depends on one parameter.

In this paper, we focus on the estimation of a failure probability $P(\phi(X) > S) = P$ with S a threshold. A simple way to estimate this probability is to consider Monte Carlo methods [12–16]. For that purpose, one generates independent and identically distributed samples $X_1, \ldots, X_N$ from the PDF f and then estimates the probability with

$$P^{MC} = \frac{1}{N}\sum_{i=1}^{N} \mathbf{1}_{\phi(X_i)>S} \tag{1}$$

where $\mathbf{1}_{\phi(X_i)>S}$ is equal to 1 if $\phi(X_i) > S$ and 0 otherwise. When S has a high value, the probability is low and cannot be accurately estimated with Monte Carlo. The estimate relative deviation is indeed too important in this case. Alternative methods such as importance sampling (IS) method [17–25] or importance splitting (ISp) [26–29] can be used.

In many situations, the PDF model and its parameters are uncertain but are assumed constant to simplify the problem. This situation is of course not realistic. In this paper, we assume that the PDF of parameter δ_i is defined by g_i. The PDF g_i can be for instance a uniform random variable on the interval $[\langle\delta_i\rangle - c_i, \langle\delta_i\rangle + c_i]$ where $\langle\delta_i\rangle$ is the mean parameter of δ_i and where c_i is experimentally determined or also a Gaussian PDF with mean parameter $\langle\delta_i\rangle$ and a variance parameter v_i. This uncertainty has an influence on the probability estimation and consequently the probability P depends on the value of $\delta_1, \delta_2, \ldots, \delta_d$.

Nevertheless, in order to estimate Sobol indices of $\delta_1, \delta_2, \ldots, \delta_d$, all these parameters have to be independent. It is indeed a sine qua none condition for Sobol indice estimation. This is of course a limit of the proposed approach since it is not always true for every kind of densities. As in every application of Sobol indices, it is thus necessary to be careful that this independence assumption is valid, otherwise Sobol indice estimations are not valuable. Contrary to the usual application of Sobol indices, the $X^{(i)}$ components are not required to be independent but have to be described by PDF with independent parameters.

In this paper, we propose a methodology to study the influence of $(\delta_1, \delta_2, \ldots, \delta_d)$ on the probability P with a low computation cost.

3. Sensitivity analysis of model parameters on a failure probability

It consists in analyzing the influence of the input PDF parameters δ_i on the value of the probability P. The proposed strategy is the following:

- Determine an efficient sampling density h to estimate P for a fixed value of $(\delta_1, \delta_2, \ldots, \delta_d)$. One can notably choose that $(\delta_1, \delta_2, \ldots, \delta_d) = (\langle\delta_1\rangle, \langle\delta_2\rangle, \ldots, \langle\delta_d\rangle)$. h_{opt} is theoretically defined with:
$$h_{opt} = \frac{\mathbf{1}_{\phi(X)>S} f(X)}{P(\langle\delta_1\rangle, \langle\delta_2\rangle, \ldots, \langle\delta_d\rangle)} \tag{2}$$
The PDF h_{opt} cannot be derived since it depends on the unknown probability P. Nevertheless rare event methods enable to generate samples according to efficient sampling PDF h that approximates the PDF h_{opt}. The more h is close to h_{opt}, the more efficient is the rare event algorithm. At the end of this stage, we can obtain a set of N samples $X_1, \ldots, X_N$ generated from h and estimate the probability $\widehat{P}(\langle\delta_1\rangle, \langle\delta_2\rangle, \ldots, \langle\delta_d\rangle)$.
- Estimate the probability depending on the value of the model parameters δ_i with importance sampling thanks to the following equation:
$$\widehat{P}(\delta_1, \ldots, \delta_d) = \frac{1}{N}\sum_{i=1}^{N} \mathbf{1}_{\phi(X_i)>S} \frac{f_{\delta_1,\ldots,\delta_d}(X_i)}{h(X_i)} \tag{3}$$
where the term $f_{\delta_1,\ldots,\delta_d}$ means that f depends on $\delta_1, \ldots, \delta_d$. One has thus determined a direct link between probability estimate $\widehat{P}$ and model parameters $\delta_1, \ldots, \delta_d$.
- Make a sensitivity analysis with, for instance, Sobol indices to determine which density parameters are the most influential on the value of $\widehat{P}$. If the variation of a model parameter δ_i gives a large variation of the value $\widehat{P}$, then $\widehat{P}$ is very sensitive to δ_i. A ranking of the different input density parameters can then be obtained and one can determine which input density parameters have to be accurately estimated to have confidence in the model.

These stages are analyzed in detail in the following of the article.

4. Determining an efficient sampling PDF to estimate *P*

The value of $(\delta_1, \delta_2, \ldots, \delta_d)$ is firstly set to $(\langle\delta_1\rangle, \langle\delta_2\rangle, \ldots, \langle\delta_d\rangle)$. Then, when one wants to estimate a probability P, a very simple method is to use Monte Carlo simulations using Eq. (1). Knowing the true probability P that $(\phi(X) > S)$, the relative deviation $\frac{\sigma_{P^{MC}}}{P^{MC}}$ is defined by

$$\frac{\sigma_{P^{MC}}}{P^{MC}} = \frac{1}{\sqrt{N}} \frac{\sqrt{P - P^2}}{P} \tag{4}$$

The Monte Carlo estimate relative deviation tends to $+\infty$ when one focuses on rare event estimation, that is when $P \to 0$ for a given sample-size N. Monte Carlo techniques are consequently not adapted to such estimation. The optimal sampling density h_{opt} to estimate P is theoretically defined by Eq. (2). Only one sample generated with this density h_{opt} is sufficient to obtain a zero-variance estimator of P but since it depends on the unknown probability P, it cannot be used directly. Nevertheless, if one determines a density that approaches the optimal density h_{opt} and also allows an easy generation of random numbers, the probability P is then efficiently estimated. There are two main methods to estimate rare event probability: importance sampling (IS) and importance splitting (ISp). Both methods are described in the following.

4.1. Importance sampling

Let us define $w(X) = \frac{f(X)}{h(X)}$. The idea of IS is to generate samples $X_1, \ldots, X_N$ from an auxiliary PDF h and then estimate P in the following way:

$$P^{IS} = \frac{1}{N} \sum_{i=1}^{N} \mathbf{1}_{\phi(X_i)>S} w(X_i) \tag{5}$$

The variance of P^{IS} is estimated by:

$$Var(P^{IS}) = \frac{Var(\mathbf{1}_{\phi(X)>S} w(X))}{N} \tag{6}$$

The auxiliary PDF h must be chosen with care. The PDF h must be positive when f is positive because of the condition of absolute continuity, and its tail must be sufficiently heavy with respect to the tail of f in order to have an estimate with finite variance. The variance of IS estimate depends notably on the choice of h. If h is well-chosen, then the IS estimate variance can be very low. Conversely, if h is not adapted to the estimation, the IS estimate results can have a higher variance than Monte-Carlo estimate. A lot of methods are available to estimate a valuable sampling PDF h and are notably described in different references [17–24,30–33]. Non parametric importance sampling, cross-entropy optimization, sequential Monte Carlo samplers, exponential tilting have notably been investigated. They consist most of the time in an optimization algorithm. The PDF h can be notably estimated within a parametric PDF family or derived with nonparametric kernel density estimator. Efficient IS often involves adaptive and recursive estimation. Since it is not the objective of this article to compare the different importance sampling optimization algorithm, one will only consider cross-entropy and nonparametric importance sampling in order to determine the PDF h.

Once a valuable sampling PDF h to estimate the probability P has been determined, it is then possible to generate N samples $X_1, \ldots, X_N$ that follow the density h and estimate the probability $\widehat{P}(\langle\delta_1\rangle, \langle\delta_2\rangle, \ldots, \langle\delta_d\rangle)$ with Eq. (3).

4.2. Importance splitting

Importance splitting is an alternative to importance sampling that is often well adapted to very rare events. In importance splitting, the underlying distribution is not modified. There is no change of the probability laws driving the model but an acceleration of the rate of occurrence of the rare event is performed by duplicating the promising samples.

Let us define

$$\mathbf{A} = \{x \in \mathbb{R}^d | \phi(x) > S\} \tag{7}$$

One has consequently $P(X \in \mathbf{A}) = P(\phi(X) > S) = P$. The principle of ISp is to iteratively estimate supersets of the set A and then to estimate $P(X \in \mathbf{A})$ with conditional probabilities.

Let us define $\mathbf{A}_0 = \mathbb{R}^d \supset \mathbf{A}_1 \supset \cdots \supset \mathbf{A}_{n-1} \supset \mathbf{A}_n = \mathbf{A}$, a decreasing sequence of $\mathbb{R}^d$ subsets with smallest element $\mathbf{A} = \mathbf{A}_n$. They can be defined with

$$\mathbf{A}_k = \{x \in \mathbb{R}^d | \phi(x) \geqslant S_k\}$$

for $k = 1, \ldots, n$ with an increasing sequence of thresholds

$$S_1 \leqslant \cdots \leqslant S_{n-1} \leqslant S_n = S$$

The probability $P(X \in \mathbf{A})$ can be then rewritten in the following way:

$$P(X \in \mathbf{A}) = \prod_{k=1}^{n} P(X \in \mathbf{A}_k | X \in \mathbf{A}_{k-1})$$

where $P(X \in \mathbf{A}_k | X \in \mathbf{A}_{k-1})$ is the probability that $X \in \mathbf{A}_k$ knowing that $X \in \mathbf{A}_{k-1}$. An optimal choice of the sequence $\mathbf{A}_k$, $k = 0 \dots n$ is given when $P(X \in \mathbf{A}_k | X \in \mathbf{A}_{k-1}) = p$, where p is a constant, that is when all the conditional probabilities are equal. The variance of $P(X \in \mathbf{A})$ is indeed minimized in this configuration as shown in [34,35]. Consequently, if each $P(X \in \mathbf{A}_k | X \in \mathbf{A}_{k-1})$ is well estimated, then the probability $P(X \in \mathbf{A})$ is estimated more accurately with ISp than with a direct estimation by Monte Carlo simulations. Defining the A_k sequence and generating samples with conditional densities are still a domain of research. At the end of importance splitting, one has obtained a set of N samples $X_1, \dots, X_N$ that follow a PDF h.

We propose in the following to analyse the different stages of the algorithm ISp to estimate $P(\phi(X) > S) = P(X \in \mathbf{A})$.

(1) Set $k = 0$.
(2) Generate N_0 samples $X_1^0, \dots, X_{N_0}^0$ from f.
(3) Estimate S_1 as the α-quantile of samples $\phi(X_1^0), \dots, \phi(X_{N_0}^0)$.
(4) While $S_{k+1} < S$, do
  (a) Set $k = k + 1$.
  (b) Generate N_k samples $X_1^k, \dots, X_{N_k}^k$ from f^k, the density of X restricted to the set $\mathbf{A}_k = \{x \in \mathbb{R}^d | \phi(x) \geqslant S_k\}$. Unfortunately, generating directly independent samples from the f^k conditional densities is in most cases impossible as they are usually unknown. A f-reversible Markovian kernel can be used to generate new samples with density f^k [36,26].
  (c) Estimate S_{k+1} as the α-quantile of samples $\phi(X_1^k), \dots, \phi(X_{N_k}^k)$.
(5) Estimate the probability with

$$P^{ISp} = (1-\alpha)^k \times \frac{1}{N_k} \sum_{i=1}^{N_k} \mathbf{1}_{\phi(X_i^k) > S}$$

The sampling PDF h can be then defined as a sum of weighted Dirac measures over the samples X_i^k. It is then possible to generate N samples $X_1, \dots, X_N$ that follow the density h and estimate the probability $\widehat{P}(\langle \delta_1 \rangle, \langle \delta_2 \rangle, \dots, \langle \delta_d \rangle)$ with Eq. (3). Kernel density estimator or histogram over the samples X_i can also be used but are not required.

5. From probability estimation to sensitivity analysis

5.1. Principle

Let us define $\delta = (\delta_1, \delta_2, \dots, \delta_d)$ and $\langle \delta \rangle = (\langle \delta_1 \rangle, \langle \delta_2 \rangle, \dots, \langle \delta_d \rangle)$. With importance sampling or splitting, one can thus determine a set of N samples $X_1, \dots, X_N$ that follows a PDF h to estimate $P(\langle \delta \rangle)$ with:

$$\widehat{P}(\langle \delta \rangle) = \frac{1}{N} \sum_{i=1}^{N} \mathbf{1}_{\phi(X_i) > S} w(X_i) \tag{8}$$

Since one has used a valuable rare event technique to determine h, h is efficient to estimate $P(\langle \delta \rangle)$ with a reasonable variance but is it still the case to estimate $P(\delta)$ for every combination of δ? If h is an efficient PDF to estimate $P(\langle \delta \rangle)$, it means that most of the samples $X_1, \dots, X_N$ fall into the set A defined in Eq. (7). The set A is completely independent from the values of δ since it only depends on the properties of the function ϕ. Consequently, the samples $X_1, \dots, X_N$ generated from h can still be used to estimate $P(\delta)$. The value of $P(\delta)$ can be estimated with:

$$\widehat{P}(\delta) = \frac{1}{N} \sum_{i=1}^{N} \mathbf{1}_{\phi(X_i) > S} \frac{f_\delta(X_i)}{h(X_i)} \tag{9}$$

The validity of this probability estimation depends on the variability of the parameters $\delta_1, \delta_2, \dots, \delta_d$ and on their impact for P. It is thus necessary to verify with caution if $\widehat{P}$ approaches well P with Eq. (9). The following examples gives some numerical examples of how to use Eq. (9) to estimate probabilities.

5.2. Example on simple cases

In this section, we propose to analyse the efficiency of the methodology proposed at the previous sections in order to estimate a probability. We consider the case where X is a simple Gaussian PDF with zero mean and a standard deviation equal to σ and try to estimate the probability $P(X > 5)$. We firstly set σ to 1 and apply a rare event algorithm to determine a valuable

sampling PDF h to estimate $P(X > 5)$. For that purpose, we propose to use Non parametric Adaptive Importance Sampling (NAIS) algorithm proposed in [21,37] and cross-entropy (CE) optimization [38,22,39] with 10,000 samples. Other importance sampling optimizations could be used (see Section 4.1) and may give better results than NAIS or CE depending on the considered case. The objective of this article is not to compare importance sampling methods but the choice of NAIS and CE is reasonable considering that they are among the most well-known IS optimization. Then, we generate 10,000 samples from the PDF h obtained with NAIS and with CE and then estimate the probability $P(X > 5)$. To evaluate the influence of σ on the probability, we also estimate $P(X > 5)$ with Eq. (9) for different values of σ without regenerating any samples.

In Fig. 1, we present the PDF h obtained with NAIS and the optimal sampling PDF for different values of σ. The optimal sampling PDF can be plotted since we know the value of theoretical probability $P(X > 5)$. In Table 1, we present the estimation of $P(X > 5)$ for different values of σ with the samples generated with PDF h. These probabilities are compared to the theoretical ones $P^{Th}(X > 5)$. One can notice that the probabilities are well estimated with a low relative deviation if $\sigma < 3$, otherwise the estimated probability is not correct. Indeed, NAIS and CE sampling PDF have been set with $\sigma = 1$; if $\sigma > 3$, the distribution tail is too heavy when compared to NAIS and CE sampling PDF. NAIS and CE sampling PDF are thus not efficient in that case with only 10,000 samples. To improve the estimation, more samples have to be drawn or NAIS and CE sampling PDF have to be updated to better perform this estimation. NAIS algorithm takes about 7 s on a standard PC to deliver one estimation with 10,000 samples, CE takes about 2 s, whereas it takes 0.005 second to generate 10,000 Monte Carlo simulations.

We also consider the case where $X = (X_1, X_2, \ldots, X_5)$ follows a multidimensional Gaussian PDF with mean 0 and a covariance matrix equal to αI_5 with I_5 the identity matrix of $\mathbb{R}^5$. The function ϕ is given by the following function called Ackley function:

$$Y = \phi(X) = -20\exp\left(-0.2\sqrt{\frac{1}{d}\sum_{i=1}^{d} X_i^2}\right) - \exp\left(\frac{1}{d}\sum_{i=1}^{d}\cos(2\pi X_i)\right) + 20 + e \tag{10}$$

We then applied the same methodology as previously described for IS to estimate $P(\phi(X) > 9.5)$. Indeed, we set α to 1 and apply NAIS algorithm and CE optimization to determine a valuable sampling PDF h with 10,000 samples. We finally generate 10,000 samples from the PDF h obtained with NAIS and with CE and then estimate the probability $P(\phi(X) > 9.5)$. In Table 2, we present the estimation of $P(\phi(X) > 9.5)$ for different values of α with the samples generated with PDF h. These probabilities are compared with the ones obtained with 10^6 Monte Carlo simulations $P^{MC}(\phi(X) > 9.5)$. If $\alpha > 2$, NAIS and CE sampling PDF are not efficient with only 10,000 samples. NAIS and CE sampling PDF have been determined for $\alpha = 1$. When parameter α varies too much, the NAIS and CE sampling PDF are not adapted. To improve the estimation, more samples have to be drawn or NAIS and CE sampling PDF have to be updated to better perform this estimation. NAIS algorithm takes about 40 s on a standard PC to deliver one estimation with 10,000 samples, CE takes about 12 s, whereas it takes 0.05 s to generate 10,000 Monte Carlo simulations.

The proposed technique estimates probabilities at low cost when some input PDF parameters vary. We show on simple examples that it can be very efficient as soon as the probability does not vary too much for a set of input parameters. It is hard to determine a priori if the probability estimate will be efficient or not. Nevertheless, one can notice that if estimated probability $\widehat{P}(\delta_1, \delta_2, \ldots, \delta_d)$ stays in the interval $[10^{-3}\widehat{P}(\langle\delta_1\rangle, \langle\delta_2\rangle, \ldots, \langle\delta_d\rangle), 10^{3}\widehat{P}(\langle\delta_1\rangle, \langle\delta_2\rangle, \ldots, \langle\delta_d\rangle)]$, the probability is very efficiently estimated. If it is no more the case, the probability estimates can then be biased. In realistic situations, it is consequently necessary to check how the probability to be estimated varies with different sets of input PDF parameters through random simulations.

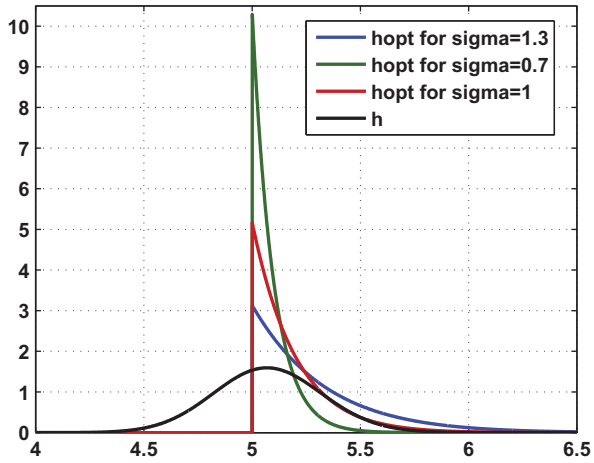


Fig. 1. Optimal sampling density for different values of standard deviation σ and NAIS sampling PDF h in black.

Table 1
Mean estimation of $P(X > 5)$ with NAIS and CE algorithm for different values of the standard deviation σ of Gaussian PDF X over 100 retrials.

σ	$P^{Th}(X > 5)$	$\widehat{P}^{NAIS}(X > 5)$ (relative deviation)	$\widehat{P}^{CE}(X > 5)$ (relative deviation)
0.7	4.6×10^{-13}	4.6×10^{-13} (1.5%)	4.7×10^{-13} (15.2%)
0.8	2.1×10^{-10}	2.1×10^{-10} (1.2%)	2.1×10^{-10} (11.3%)
0.9	1.4×10^{-8}	1.4×10^{-8} (1.0%)	1.3×10^{-8} (10.7%)
1	2.9×10^{-7}	2.9×10^{-7} (0.8%)	2.9×10^{-7} (7.8%)
1.1	2.7×10^{-6}	2.7×10^{-6} (0.7%)	2.7×10^{-6} (13.5%)
1.2	1.5×10^{-5}	1.5×10^{-5} (0.8%)	1.5×10^{-5} (15.2%)
1.3	6×10^{-5}	6.0×10^{-5} (1.0%)	6.0×10^{-5} (17.0%)
1.5	4.3×10^{-4}	4.2×10^{-4} (2%)	4.4×10^{-4} (22.1%)
2	6.2×10^{-3}	5.9×10^{-3} (5.3%)	6.3×10^{-3} (27.8%)
3	4.7×10^{-2}	4.2×10^{-2} (10%)	4.6×10^{-2} (29.3%)
5	0.16	9.4×10^{-2} (15%)	1.5×10^{-1} (30.8%)
10	0.31	0.15 (16%)	0.27 (30.1%)

Table 2
Estimation of $P(\phi(X) > 9.5)$ with NAIS and CE algorithm for different values of the covariance matrix $\alpha\Sigma$ of multidimensional Gaussian PDF X.

α	$P^{MC}(\phi(X) > 9.5)$ (relative deviation)	$P^{NAIS}(X > 9.5)$ (relative deviation)	$P^{CE}(X > 9.5)$ (relative deviation)
0.7	1.5×10^{-7} (244%)	7.2×10^{-8} (2.5%)	7.6×10^{-8} (13.2%)
0.8	1.1×10^{-6} (102%)	7.1×10^{-7} (2.1%)	7.3×10^{-7} (12.5%)
0.9	6.0×10^{-6} (35%)	4.2×10^{-6} (1.2%)	4.1×10^{-6} (8.9%)
1	2.1×10^{-5} (32%)	1.7×10^{-5} (0.8%)	1.6×10^{-5} (7.8%)
1.1	6.5×10^{-5} (9.4%)	5.5×10^{-5} (0.8%)	5.5×10^{-5} (8.6%)
1.2	1.7×10^{-4} (6%)	1.5×10^{-4} (1.0%)	1.4×10^{-4} (8.9%)
1.3	3.9×10^{-4} (4.5%)	3.2×10^{-4} (1.3%)	3.3×10^{-4} (10.2%)
1.5	1.4×10^{-3} (2.7%)	1.2×10^{-3} (1.7%)	1.2×10^{-3} (17.2%)
2	1.0×10^{-2} (0.9%)	9.8×10^{-2} (3.3%)	1.2×10^{-2} (25.3%)
3	7.0×10^{-2} (0.3%)	7.5×10^{-2} (6.8%)	7.2×10^{-2} (22.8%)
5	0.29 (0.1%)	0.20 (10.5%)	0.25 (27.5%)
10	0.68 (0.004%)	0.50 (15.2%)	0.61 (30.2%)

Instead of $\langle\delta\rangle$, one can choose a value δ_{min} that makes the probability $P(\delta)$ less rare so that Monte Carlo estimation techniques could be used. This would ease the estimation of $P(\delta)$ but, in the same way as before, it is difficult to ensure that the probability to be estimated will not be biased for other values of δ, that is, to ensure that $\widehat{P}(\delta_1, \delta_2, \ldots, \delta_d)$ stays in the interval $[10^{-3}\widehat{P}(\delta_{min}), 10^{3}\widehat{P}(\delta_{min})]$. In real case, one often knows the mean value of δ and its variation band. It is thus more logical to focus on the estimation of $P(\langle\delta\rangle)$ than on the estimation of $P(\delta_{min})$.

5.3. Functional interpretation

Since our objective is to study the influence of $\delta_1, \ldots, \delta_d$ on $\widehat{P}$, one can rewrite Eq. (3) in the following way

$$\widehat{P} = \Gamma(\delta_1, \ldots, \delta_d) \tag{11}$$

where Γ is a continuous scalar function $\Gamma : \mathbb{R}^d \to \mathbb{R}$. The d variables $\delta_1, \ldots, \delta_d$ are the inputs of this function Γ. Each input δ_i follows a PDF g_i.

6. Sensitivity analysis of failure probability $\widehat{P}$

The role of sensitivity analysis is to determine the influence of the parameters $(\delta_1, \ldots, \delta_d)$ on the probability estimate $\widehat{P}$. Consequently, we focus on the estimation of Sobol indices to determine this influence [40–43,14,44,45]. Indeed, in the general case of a nonlinear model Γ, one can estimate the influence of the model inputs by using a decomposition of Γ in elementary functions:

$$\widehat{P} = \Gamma(\delta_1, \ldots \delta_d) = \Gamma_0 + \sum_{i=1}^{d} \Gamma_i(\delta_i) + \sum_{1 \leqslant i < j \leqslant d} \Gamma_{ij}(\delta_i, \delta_j) + \ldots + \Gamma_{1\ldots d}(\delta_1, \ldots, \delta_d) \tag{12}$$

where Γ is assumed to be integrable, Γ_0 is a constant and

$$\int_{I_{i_k}} \Gamma_{i_1,\ldots i_s}(\delta_{i_1}, \ldots, \delta_{i_s}) d\delta_{i_k} = 0$$

is true $\forall k = 1 \ldots s$ and for $\{i_1, \ldots, i_s\} \subseteq \{1, \ldots, d\}$ where I_{i_k} is the support of density g_{i_k}. If the inputs δ_i are random and independent, one can obtain the well-known ANOVA decomposition:

$$Var(\widehat{P}) = V = \sum_{i=1}^{d} V_i + \sum_{1 \leqslant i < j \leqslant d} V_{ij} + \ldots + V_{1\ldots d}$$

with

$$V_i = Var(E(\widehat{P}|\delta_i))$$
$$V_{ij} = Var(E(\widehat{P}|\delta_i, \delta_j)) - V_i - V_j$$
$$V_{ijk} = Var(E(\widehat{P}|\delta_i, \delta_j, \delta_k)) - V_i - V_j - V_k - V_{ij} - V_{ik} - V_{jk}$$
$$\ldots$$

with E the mathematical expectation. One can then define the Sobol sensitivity indices at first order S_i for the variable δ_i with

$$S_i = \frac{V_i}{V} = \frac{Var(E(\widehat{P}|\delta_i))}{Var(\widehat{P})}$$

Sensitivity indices at second order can also be derived relatively to the variables δ_i and δ_j:

$$S_{ij} = \frac{V_{ij}}{V}$$

The interpretation of the sensitivity indices is easy since they vary between 0 and 1 and their sum is equal to 1. If S_i is close to 1, then the variable δ_i has a great influence on $\widehat{P}$. When the number of variables d increases, the number of Sobol indices increases exponentially and thus, the estimation of all these indices is impossible. For that purpose, total sensitivity indices S_{T_i} are then introduced for each variable δ_i:

$$S_{T_i} = \sum_{k \# i} S_k$$

where $\#i$ represents all the sets of indices that contain i. For instance, if $d = 3$, one has then:

$$S_{T_1} = S_1 + S_{12} + S_{13} + S_{132}$$

The Sobol indices are then often estimated with Monte Carlo methods [41,43]. The main difficulty of Sobol indice estimation is often the too high number of generated samples to obtain an accurate estimation (convergence rate in $\sqrt{N}$ where N is the sample size). It is notably the case when the simulation code is time consuming. 10,000 samples are often necessary for the estimation of one Sobol indice with a relative deviation of about 10%. The use of deterministic quasi Monte Carlo sequence (for instance LPτ Sobol sequences) can reduce the number of required samples to estimate Sobol indices [46]. Another method of Sobol indice estimation is to consider FAST algorithm [47,48] based on Fourier transform which is relatively more accurate than Monte Carlo and can be applied to total sensitivity indice estimation.r-LHS sampling [49] has also been proposed to estimate first order Sobol indices.

In the proposed application, with a set of variables $\delta_1, \ldots, \delta_d$, it is very simple to estimate the probability $\widehat{P}$ and thus, the number of required samples is not an issue and Monte Carlo methods can be used. Once the Sobol indices have been determined, it is then very simple to determine which input PDF parameters $\delta_1, \ldots, \delta_d$ are the most influential on the probability estimation.

7. Application

7.1. Missile simulation

We consider in this article a missile that is 6.55 m long and weighs 1100 kg. It is a supersonic stand-off missile powered by a liquid-fuel ramjet. It flies at Mach 2 to Mach 3, with a range between 100 km and 350 km depending on flight profile. This missile is modeled with a continuous black-box computer code ϕ with seven independent inputs X and one output $D = \phi(X)$ where D is the target distance. It takes about 2 s on a standard PC to compute the value D knowing X. The inputs of this black-box simulation code are the following:

- Plane inertial measurement unit in position and speed (four inputs).
- Missile inertial measurement unit in position and speed (three inputs).

These seven inputs $X^{(i)}$ follow a Gaussian PDF with 0 mean and a standard deviation equal to δ_i. The PDF parameter δ_i is assumed to be random and follow a uniform PDF on the interval $[0.1, 1]$.

The output of the simulation code D is the target distance and is consequently a random variable. The Fig. 2 shows the target distance PDF estimated with kernel density estimator for three different sets of input PDF parameters $(\delta_1, \ldots, \delta_d)$. The differences between these 3 PDF shows that the input PDF parameters $(\delta_1, \ldots, \delta_d)$ have a nonnegligible influence on the target distance D.

7.2. Sobol indice estimation of a failure probability

In this subsection, we focus on the probability estimation $P(D > 4.5) = P(\phi(X) > 4.5)$ and try to evaluate the influence of the input PDF parameter uncertainties on this probability. We firstly set all the input PDF parameters δ_i to $<\delta_i> = 0.55$. For this set of parameters, one can determine an efficient sampling PDF h to estimate the probability $P(D > 4.5)$. This PDF has been obtained with NAIS. The Fig. 3 presents the density of D with Gaussian kernel estimator when the samples $X_1,\ldots,X_N$ ($N = 10{,}000$) are generated with h. Then using Eq. (8), one estimates the probability that $\widehat{P}(D > 4.5) = 0.0106$ with $\langle\delta_i\rangle = 0.55$ $\forall i = 1 \ldots 7$. One determines with Eq. (9) the value of $\widehat{P}(D > 4.5)$ for every set of δ_i values. For instance, if $\delta_i = 0.1$ $\forall i$, one has $\widehat{P}(D > 4.5) = 0.0052$ and if $\delta_i = 1$ $\forall i$ one gets that $\widehat{P}(D > 4.5) = 0.0133$. The value of $\widehat{P}(D > 4.5)$ depends of course on the value of δ_i. NAIS algorithm takes about 6 h on a standard PC to deliver one estimation with 10,000 samples; whatever the chosen sampling method, the computation time is relatively the same.

In Fig. 4, one shows 100 estimations of $P(D > 4.5)$ obtained with 10^3 simulations and NAIS algorithm for different random values of input PDF parameters without using Eq. (9) but directly the simulation code ϕ. The probability stays in the interval proposed at the end of Section 5.2 and consequently, we can assume that the probabilities given with Eq. (9) are not biased.

In the following, we propose to rank the influence of parameters δ_i on the value of $\widehat{P}(D > 4.5)$ with Sobol indices as proposed in Section 6. The Tables 3 and 4 present the mean estimations of first order and total Sobol indices over 100 retrials with 50,000 Monte Carlo simulations. The standard deviation of the estimations is also provided. The 7 input PDF parameters do not contribute equally to the probability variability. The parameters δ_5 and δ_1 have the highest Sobol indices in this case. Consequently, it is necessary to well estimate their values since a misestimation will cause a high variation of the probability P. The other input PDF parameters are less influential on the probability at first order.

The same study has been done for the probability $P(D > 7)$. One determines a valuable sampling PDF to estimate $P(D > 7)$ with NAIS when $<\delta_i> = 0.55$ $\forall i = 1 \ldots 7$. Then, using the same methodology as previously described, one estimates that $\widehat{P}(D > 7) = 9.410^{-7}$ with $<\delta_i> = 0.55$ $\forall i = 1 \ldots 7$. Thanks to Eq. (9), if $\delta_i = 0.1$ $\forall i$, one has $\widehat{P}(D > 7) = 1.310^{-8}$ and if $\delta_i = 1$ $\forall i$ one gets that $\widehat{P}(D > 7) = 7.810^{-6}$. The value of $\widehat{P}(D > 7)$ depends of course on δ_i. In Fig. 5, one shows 100 NAIS estimations of $P(D > 7)$ obtained with 10^4 simulations for different random values of input PDF parameters without using Eq. (9) but directly the simulation code ϕ. The probability stays in the interval proposed at the end of Section 5.2 and consequently, the probabilities given with Eq. (9) will probably not be biased.

In the following, we then rank the influence of parameters δ_i on the value of $\widehat{P}(D > 7)$ with Sobol indices. The Tables 5 and 6 present the mean estimations of first order and total Sobol indices over 100 retrials with 50,000 Monte Carlo simulations. Parameters δ_6 and δ_5 are the most influent input PDF parameters on $P(D > 7)$. The most influent input PDF parameters depend on the estimated probability. For instance, $P(D > 4.5)$ is sensitive to δ_1 but it is not case for $P(D > 7)$.

In this section, we have shown that input PDF parameters are clearly influential on the probability estimate. Moreover a rank of their degree of influence is provided through Sobol indice estimation.

7.3. Simulation with other realistic PDF

To evaluate the influence of the PDF model on the probability estimation, one has modified the PDF of the inputs in the previous example described in Section 7.1. We now assume that the different input parameters follow a Gumbel distribution. These seven inputs $X^{(i)}$ follow a Gumbel PDF with location parameter:

$$(\mu_1 = 0.22, \mu_2 = 0.18, \mu_3 = -0.16, \mu_4 = 0.15, \mu_5 = 0.02, \mu_6 = -0.07, \mu_7 = 0.05)$$

and a scale parameter equal to δ_i. The PDF parameter δ_i is assumed to be random and follow a uniform PDF on the interval [0.5,2]. The considered distribution is asymmetric and has a heavier tail than Gaussian distribution.

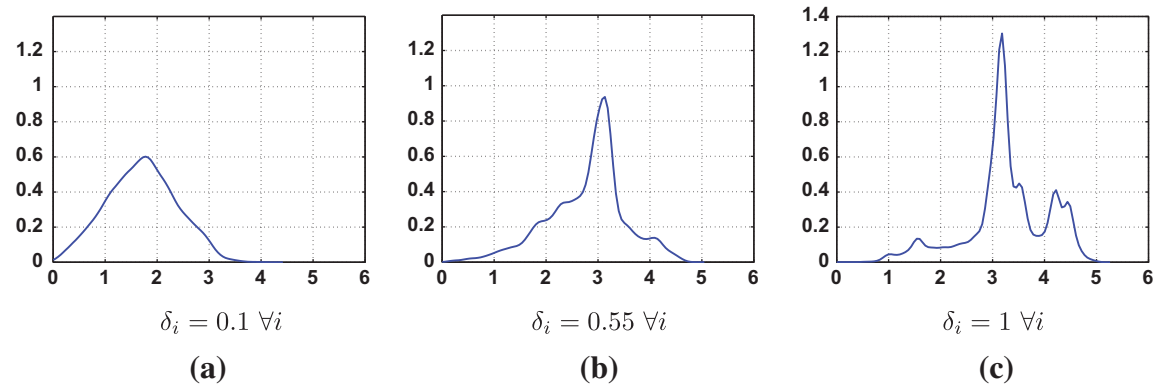


Fig. 2. Probability density function of target distance estimated with 10,000 Monte Carlo simulations for different sets of input PDF parameters.

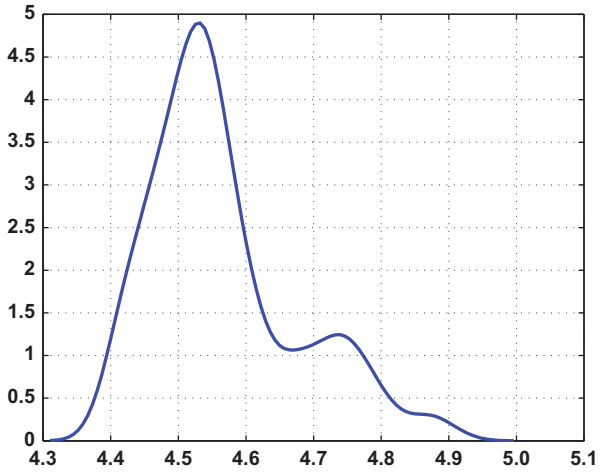


Fig. 3. Probability density function of target distance estimated with 10,000 samples generated from h.

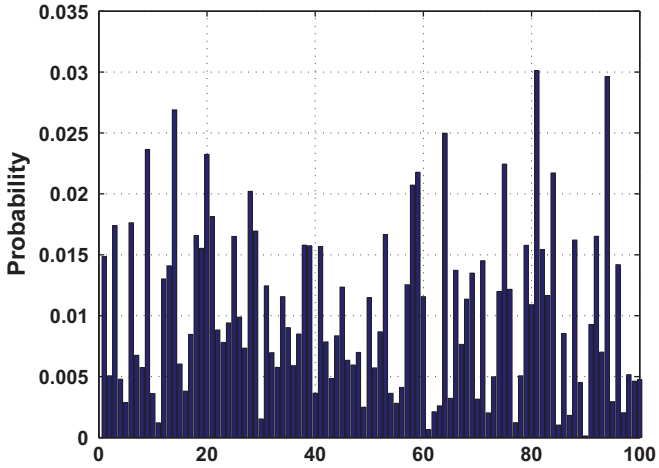


Fig. 4. 100 estimations of $P(D > 4.5)$ for different random sets of input PDF parameters.

Table 3
Mean of first order Sobol indice estimations over 100 retrials with 50,000 Monte Carlo simulations.

Sobol indices	Value	Standard deviation
S_1	0.15	0.01
S_2	0.09	0.008
S_3	0.018	0.01
S_4	0.023	0.002
S_5	0.31	0.01
S_6	0.06	0.004
S_7	0.07	0.003

We estimate the probability $P(D > 7)$. In the same way we previously described, one determines a valuable sampling PDF to estimate $P(D > 7)$ with NAIS when $\langle \delta_i \rangle = 1.25\ \forall i = 1 \ldots 7$. Then, using the same methodology as previously described, one estimates that $\widehat{P}(D > 7) = 6.710^{-5}$ with $\langle \delta_i \rangle = 1.25\ \forall i = 1 \ldots 7$. Thanks to Eq. (9), if $\delta_i = 0.5\ \forall i$, one has $\widehat{P}(D > 7) = 8.2.10^{-7}$ and if $\delta_i = 2\ \forall i$ one gets that $\widehat{P}(D > 7) = 2.110^{-4}$. The value of $\widehat{P}(D > 7)$ depends of course on δ_i. In Fig. 6, one shows 100 NAIS estimations of $P(D > 7)$ obtained with 10^4 simulations for different random values of input PDF parameters without using Eq. (9) but directly the simulation code ϕ. The probability stays in the interval proposed at the end of Section 5.2.

In the following, we then rank the influence of parameters δ_i on the value of $\widehat{P}(D > 7)$ with Sobol indices. The Tables 7 and 8 present the mean estimations of first order and total Sobol indices over 100 retrials with 50,000 Monte Carlo simulations.

Table 4
Mean of Total Sobol indice estimations over 100 retrials with 50,000 Monte Carlo simulations.

Sobol indices	Value	Standard deviation
S_{T_1}	0.22	0.01
S_{T_2}	0.19	0.01
S_{T_3}	0.09	0.01
S_{T_4}	0.09	0.02
S_{T_5}	0.38	0.02
S_{T_6}	0.14	0.01
S_{T_7}	0.11	0.01

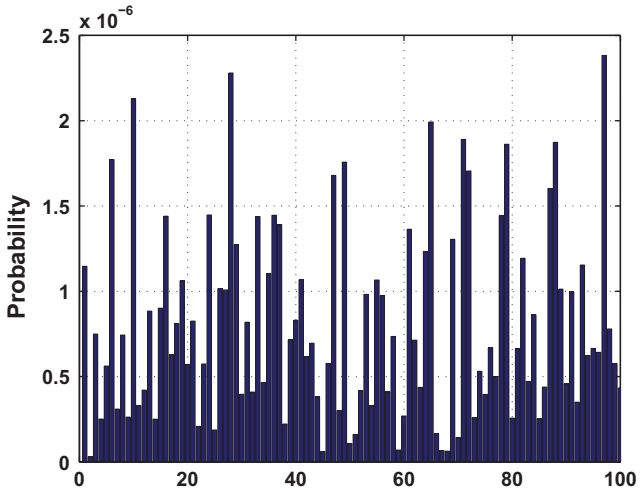


Fig. 5. 100 estimations of $P(D > 7)$ for different random set of input PDF parameters.

Table 5
Mean of first order Sobol indice estimations over 100 retrials with 50,000 Monte Carlo simulations.

Sobol indices	Value	Standard deviation
S_1	0.06	0.007
S_2	0.05	0.008
S_3	0.03	0.006
S_4	0.023	0.003
S_5	0.18	0.008
S_6	0.30	0.01
S_7	0.04	0.005

Table 6
Mean of total Sobol indice estimations over 100 retrials with 50,000 Monte Carlo simulations.

Sobol indices	Value	Standard deviation
S_{T_1}	0.22	0.01
S_{T_2}	0.13	0.01
S_{T_3}	0.07	0.01
S_{T_4}	0.05	0.01
S_{T_5}	0.38	0.02
S_{T_6}	0.42	0.02
S_{T_7}	0.13	0.01

Parameters δ_6 and δ_5 are the most influent input PDF parameters on $P(D > 7)$. These results are not too far from those obtained with Gaussian distributions. Indeed, the scale parameters δ_i of Gumbel distribution are directly linked to the

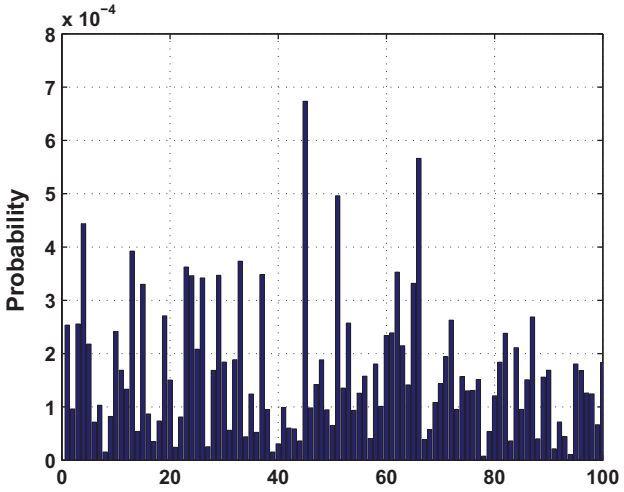


Fig. 6. 100 estimations of $P(D > 7)$ for different random set of input PDF parameters.

Table 7
Mean of first order Sobol indice estimations over 100 retrials with 50,000 Monte Carlo simulations.

Sobol indices	Value	Standard deviation
S_1	0.08	0.01
S_2	0.12	0.01
S_3	0.015	0.005
S_4	0.020	0.001
S_5	0.22	0.01
S_6	0.24	0.005
S_7	0.12	0.005

Table 8
Mean of total Sobol indice estimations over 100 retrials with 50,000 Monte Carlo simulations.

Sobol indices	Value	Standard deviation
S_{T_1}	0.12	0.01
S_{T_2}	0.22	0.02
S_{T_3}	0.12	0.02
S_{T_4}	0.13	0.01
S_{T_5}	0.32	0.03
S_{T_6}	0.35	0.01
S_{T_7}	0.09	0.01

distribution variance and thus, similar behavior on sensitivity analysis of a rare event are determined when the scale parameter, in the case of Gumbel distribution, or the variance parameter, in the case of Gaussian distribution, are random.

The efficiency of some rare event estimation methods depend on the tail behavior of the target distribution. In this article, we show that if a valuable sampling PDF is determined, the proposed methodolgy can be applied.

8. Conclusion

In this article, we have proposed an original methodology to analyze the influence of the input PDF parameter variability on a rare failure probability. The proposed method based on the estimation of a standard probability and on Sobol indices has been described in the case of a general problem where the model is a black-box system. The last part of this article concerns the application of this algorithm on the estimation of missile collateral damage probability. We firstly show that it is important to estimate input PDF parameters since they influence strongly the value of the output model probability. Moreover, a rank of the influence of the PDF parameters on the probability estimate can be obtained with Sobol indices.

References

[1] S. Song, Z. Lu, H. Qiao, Subset simulation for structural reliability sensitivity analysis, Reliability Engineering & System Safety 94 (2) (2009) 658–665.
[2] A.A. Taflanidis, J.L. Beck, Stochastic subset optimization for reliability optimization and sensitivity analysis in system design, Computers & Structures 87 (5–6) (2009) 318–331.
[3] E. Zio, N. Pedroni, Sensitivity analysis of the model of a nuclear passive system by means of subset simulation, Procedia – Social and Behavioral Sciences 2 (6) (2010) 7778–7779.
[4] G. Chan, H. Yang, Sensitivity analysis on ruin probabilities with heavy-tailed claims, Statistical Methodology 2 (1) (2005) 59–63.
[5] R.W. Katz, Extreme value theory for precipitation: sensitivity analysis for climate change, Advances in Water Resources 23 (2) (1999) 133–139.
[6] B. Fischhoff, P. Slovic, S. Lichtenstein, Fault trees: sensitivity of estimated failure probabilities to problem representation, Journal of Experimental Psychology: Human Perception and Performance 4 (2) (1978) 330–344.
[7] P.V. Suresh, A.K. Babar, V.V. Raj, Uncertainty in fault tree analysis: a fuzzy approach, Fuzzy Sets and Systems 83 (2) (1996) 135–141.
[8] S. Ferson, W.T. Tucker, Sensitivity analysis using probability bounding, Reliability Engineering & System Safety 91 (10–11) (2006) 1435–1442.
[9] M. Oberguggenberger, J. King, B. Schmelzer, Classical and imprecise probability methods for sensitivity analysis in engineering: a case study, International Journal of Approximate Reasoning 50 (4) (2009) 680–693.
[10] S.K. Au, J.L. Beck, Subset simulation and its application to seismic risk based on dynamic analysis, Journal of Engineering Mechanics 129 (8) (2001) 1–17.
[11] M. Munoz Zuniga, J. Garnier, E. Remy, E. de Rocquigny, Adaptive directional stratification for controlled estimation of the probability of a rare event, Reliability Engineering & System Safety, in Press. doi:10.1016/j.ress.2011.06.016.
[12] B. Silverman, Density Estimation for Statistics and Data Analysis, Chapman and Hall, London, 1986.
[13] G.A. Mikhailov, Parametric Estimates by the Monte Carlo Method, VSP, Utrecht (NED), 1999.
[14] I.M. Sobol, A Primer for the Monte Carlo Method, CRC Press, Boca Raton, Fl., 1994.
[15] C.P. Robert, G. Casella, Monte Carlo Statistical Methods, Springer, New York, 2005.
[16] H. Niederreiter, J. Spanier, Monte Carlo and Quasi-Monte Carlo Methods, Springer, 2000.
[17] I. Beichl, F. Sullivan, The importance of importance sampling, Computing in Science and Engineering 1 (1999) 71–73.
[18] P.H. Borcherds, Importance sampling: an illustrative introduction, European Journal of Physics (2000) 405–411.
[19] M. Denny, Introduction to importance sampling in rare-event simulations, European Journal of Physics (2001) 403–411.
[20] Z.I. Botev, D. Kroese, T. Taimre, Generalized cross-entropy methods with applications to rare-event simulation and optimization, Simulation 11 (2007) 785–806.
[21] P. Zhang, Nonparametric importance sampling, Journal of the American Statistical Association 91 (434) (1996) 1245–1253.
[22] R. Rubinstein, Optimization of computer simulation models with rare events, European Journal of Operations Research 99 (1997) 89–112.
[23] P. Glynn, Importance Sampling for Monte Carlo Estimation of Quantiles, Publishing House of Saint Petersburg University, 1996.
[24] J. Morio, How to approach the importance sampling density, European Journal of Physics 31 (2010) 41–48.
[25] J. Morio, R. Pastel, F. Le Gland, Estimation of rare event probabilities and quantiles applied to an aerospace vehicle fallen safety zone analysis, Journal de la SociTtT Française de Statistiques 3 (3) (2011) 1–29.
[26] F. Cerou, P. Del Moral, T. Furon, A. Guyader, Rare Event Simulation for a Static Distribution, RESIM, vol. 7, 2008, pp. 107–115.
[27] F. Cerou, A. Guyader, Adaptive Particle Techniques and Rare Event Estimation, ESAIM, vol. 19 2007, pp. 65–72.
[28] P. L'Ecuyer, V. Demers, B. Tuffin, Splitting for rare event simulation, in: Proceeding of the 2006 Winter Simulation Conference, 2006, pp. 137–148.
[29] P. Glasserman, P. Heidelberger, P. Shahabuddin, T. Zajic, Splitting for rare event simulation: analysis of simple cases, in: Proceeding of the 1996 Winter Simulation Conference, 1996, pp. 302–308.
[30] H.E. Daniels, Saddlepoint approximations in statistic, Ann. Math. Statis 5 (1954) 631–650.
[31] R. Lugannani, S. Rice, Saddlepoint approximation for the distribution of the sum of independent random variables, Advances in Applied Probability 12 (1980) 475–490.
[32] P.D. Moral, A. Doucet, A. Jasra, Sequential monte carlo samplers, Journal of the Royal Statistical Society: Series B 68 (3) (2006) 411–436.
[33] A. Doucet, S. Godsill, C. Andrieu, On sequential monte carlo sampling methods for bayesian filtering, Statistics and Computing 10 (3) (2000) 197–208.
[34] A. Lagnoux, Rare event simulation, Probability in the Engineering and Informational Science 20 (2006) 45–66.
[35] F. Cerou, P. Del Moral, F. Le Gland, P. Lezaud, Genetic genealogical models in rare event analysis, INRIA report 1797, 2006, pp. 1–30.
[36] L. Tierney, Markov chains for exploring posterior distributions, Annals of Statistics 22 (1994) 1701–1762.
[37] J. Morio, Non parametric adaptive importance sampling of the probability estimation of launcher impact position, Reliability Engineering & System Safety 96 (1) (2011) 178–183.
[38] T.H. de Mello, R. Rubinstein, Rare Event Estimation for Static Models via Cross-Entropy and Importance Sampling, John Wiley, New York, 2002.
[39] R. Rubinstein, The cross-entropy method for combinatorial and continuous optimization, Methodology and Computing in Applied Probability 2 (1999) 127–129.
[40] W. Hoeffding, A class of statistics with asymptotically normal distributions, Annals of Mathematical Statistics 19 (1948) 293–325.
[41] I. Sobol, S. Kuchereko, Sensitivity estimates for nonlinear mathematical models, Mathematical Modelling and Computational Experiments 1 (1993) 407–414.
[42] T. Homma, A. Saltelli, Importance measures in global sensitivity analysis of nonlinear models, Reliability Engineering & System Safety 52 (1996) 1–17.
[43] A. Saltelli, Making best use of model evaluation to compute sensitivity indices, Computer Physics Communication 145 (2002) 280–297.
[44] A. Saltelli, S. Tarantola, K.-S. Chan, A quantitative model-independent method for global sensitivity analysis of model output, Technometrics 41 (1999) 39–56.
[45] J.W. Hall, Uncertainty-based sensitivity indices for imprecise probability distributions, Reliability Engineering & System Safety 91 (10–11) (2006) 1443–1451.
[46] A. Saltelli, K. Chan, E. Scott, Global Sensitivity Analysis – The Primer, Wiley, 2000.
[47] H. Cukier, R. Levine, K. Schuler, Nonlinear sensitivity analysis of multiparameter model systems, Journal of Computational Physics 26 (1978) 1–42.
[48] A. Saltelli, S. Tarantola, K. Chan, A quantitative model-independent method for global sensitivity analysis of model output, Technometrics 41 (1) (1999) 39–56.
[49] M. McKay, R. Beckman, W. Conover, A comparison of three methods for selecting values of input variables in the analysis of output from a computer code, Technometrics 21 (2) (1979) 239–245.

Chapitre 4

Estimating separation distance loss probability between aircraft in uncontrolled airspace in simulation, J. Morio, T. Lang and C. Le Tallec, *Safety Science*, 2012, Vol. 50, Issue 4, pp. 995-1004

Safety Science 50 (2012) 995–1004

Contents lists available at SciVerse ScienceDirect

Safety Science

journal homepage: www.elsevier.com/locate/ssci

Estimating separation distance loss probability between aircraft in uncontrolled airspace in simulation

Jérôme Morio *, Thibault Lang, Claude Le Tallec

ONERA – The French Aerospace Lab, BP 80100, F-91123 Palaiseau, France

ARTICLE INFO

Article history:
Received 23 May 2011
Received in revised form 18 July 2011
Accepted 9 December 2011
Available online 20 January 2012

Keywords:
Uncontrolled airspace
Collision probability
Statistics
Rare event
Quantile estimation
Importance splitting

ABSTRACT

In this article, collision probability between aircraft in uncontrolled airspace is estimated. For that purpose, a large database of aircraft trajectories in the vicinity of Saint-Cyr-l'Ecole airfield (France) is considered and maps of probability collision from simulated aircraft are then estimated. Since the collision between aircraft is a rare event, we applied an importance splitting estimation technique rather than crude Monte Carlo simulations to reduce the variance of the probability estimation. In this study, we demonstrate the high local variability of collision probability in uncontrolled airspace and conclude on the difficulty to set general probability requirements.

1. Introduction

Maintaining a specific minimum separation distance between two aircraft to avoid collisions is mandatory in air traffic management (ATM). This safety rule is generally guaranteed by the air traffic control (ATC) that demands aircraft to fly at set levels or level bands, on defined routes or in certain directions. The aircraft positions are also well-known thanks to transponder and radar. Collision or separation loss statistics are consequently easily evaluated. The collection and analysis of data on hazardous air traffic management incidents have also been an important task to determine issues and improve ATM (Brooker, 2005; Leva et al., 2009). ATM modelling and simulation in controlled airspace (Kirkland et al., 2004; Shangwen and Ming-hua, 2010; Wang et al., 2010; Irvine, 2001) have been widely discussed in the literature. Based on these results, regulations have been set. On the contrary, when one considers uncontrolled airspace, it is difficult to evaluate the collision or separation loss risk with confidence. Indeed, the aircraft number, position and routes are neither known nor recorded. There is thus currently a specific need to estimate the probability of collision or separation loss in uncontrolled airspace.

For that purpose, we have obtained a 170 trajectory database of aircraft in uncontrolled airspace (class G airspace) over Saint-Cyr-l'Ecole, France, airfield. In this article, our objective is thus to estimate collision or separation loss probability with this trajectory database over the region of Saint-Cyr-l'Ecole in simulation. It will provide an overview of the mean collision risk in the general case and of its local variability. This type of study could be interesting for regulatory purposes about the integration of unmanned aircraft into uncontrolled airspace (Allouche, 2000; Ostwald and Hershey, 2007; Asmat et al., 2006; Kochenderfer et al., 2008b; Kochenderfer et al., 2008a). The effect of their integration on aircraft safety is a hard question to answer since the current safety conditions in uncontrolled airspace is not well characterized.

This paper presents the general context of airspace class and then details the aeronautical issue near Saint-Cyr-l'Ecole airfield. It describes the trajectory database that will be used in simulation and how collision probabilities are derived. As Monte Carlo (Mikhailov, 1999; Sobol, 1994; Robert and Casella, 2005) simulations are not accurate enough to estimate rare event probabilities with an affordable simulation, using importance splitting algorithm (Cerou et al., 2008; Cerou and Guyader, 2007; L'Ecuyer et al., 2006; Glasserman et al., 1996; Morio et al., 2010) is suggested and based on the recursive estimation of conditional probabilities. The final section of this article is dedicated to collision probability analysis over Saint-Cyr-l'Ecole.

2. Context

This section describes airspace class and flight rules that are followed by the aircraft in the trajectory database. The flight situation in the vicinity of Saint-Cyr-l'Ecole airfield is then presented.

* Corresponding author. Tel.: +33 1 80 38 66 54.
E-mail addresses: jerome.morio@onera.fr (J. Morio), thibault.lang@onera.fr (T. Lang), Claude.Le_Tallec@onera.fr (C.L. Tallec).

0925-7535/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.ssci.2011.12.014

2.1. Flight rules and airspace class

2.1.1. Flight rules

Two different flight rule sets coexist currently in airspace (Thom and Godwin, 2007): the visual flight rules (VFR) and the instrumental flight rules (IFR). VFR are a set of regulations which allow a pilot to operate an aircraft in weather conditions generally clear enough to allow the pilot to maintain ground in sight and see where the aircraft is going. The weather conditions are supposed to be sufficient to sense and avoid potential collisions with other aircraft. IFR permit an aircraft to operate in instrument meteorological conditions (IMC), which have much lower weather minimums than VFR. In this article, all the aircraft trajectories that are considered in the database follow the VFR rules. VFR flights are operated in the case of visual weather conditions (VMC). VMC are characterized by an horizontal visibility above 5–8 km depending on the flight height and down to 1.5 km in uncontrolled airspace at altitude below 3000 feet. The distance to the clouds in VMC conditions is equal or greater than 1500 m in horizontal plan and 300 m vertically. Flight is also permitted in uncontrolled airspace below 3000 feet just outside the clouds.

2.1.2. Airspace class

There are two different types of airspace: controlled and uncontrolled airspace. In controlled airspace, ATC has the authority to control air traffic, the level of which varies with the different airspace classes. Controlled airspace is established mainly for two different reasons:

- high traffic density areas (for instance, near airfields)
- IFR traffic under ATC guidance

Controlled airspace (class A–E airspace) usually exists in the immediate vicinity of major airfields, where aircraft are carrying out procedures for departures, approaches and transit routes.

In uncontrolled airspace (class F, G airspace), ATC service is unnecessary or cannot be provided for practical reasons. ATC does not exercise any executive authority in uncontrolled airspace, but may provide basic information services to aircraft in radio contact. Flight in uncontrolled airspace will typically be under VFR in the studied case. Aircraft operating under IFR should not expect separation from other traffic. In most countries, it is common to provide uncontrolled airspace in areas where significant air transport or military activity is not expected. Each national aviation authority determines how it uses the airspace classifications in its airspace design. Indeed, Fig. 1 presents the different kinds of airspace class in France.

All the aircraft in the trajectory database operate in a class G airspace. More precisely, a class G airspace is an uncontrolled airspace where the ATC clearance is not required, the separation is not provided, and traffic information is provided if possible. In France, class G airspaces are either located below flight level 115 and above flight level 660.

2.2. Saint-Cyr-l'Ecole airfield

Saint-Cyr-l'Ecole airfield is a French airfield located at 21 km southwest from Paris, France in the territory of Saint-Cyr-l'Ecole town (Yvelines). International Civil Aviation Organization (ICAO) code of this airfield is LFPZ. Its geographic coordinates are 48°48′37″ North, 2°04′24″ East. Its elevation above mean sea level is 113 m and the airfield area is 80 ha. The airfield has two runways in grass of direction 11*L*/29*R* and 11*R*/29*L* and with respective dimensions $890 \times 100\ m^2$ and $867 \times 60\ m^2$. Fig. 2a and b shows Saint-Cyr-l'Ecole location on a France map and an aerial photography of Saint-Cyr-l'Ecole airfield. Visual Approach Charts (VAC) maps for visual landing and visual approach at Saint-Cyr-l'Ecole airfield are provided in Figs. 3 and 4.

Fig. 5 presents the public airfields (plane icons) in Saint-Cyr-l'Ecole airfield surroundings. The different ICAO codes correspond to Toussus-le-Noble (LFPN), Velizy-Villacoublay (LFPV), Chavenay-Villepreux (LFPX) and Beynes-Thivernal (LFPF). This airfield network is located at less than 5 flight minutes away from Saint-Cyr-l'Ecole airfield with significant traffic from one airfield to another. Green tags with ICAO code LFPZSNOR, LFPZSER, LFPZSNL, and LFPZWES describe the entry and the exit points of Saint-Cyr-l'Ecole airfield. The entry point of the airfield is located at 1100 feet and the exit points at 1500 feet. One can also notice that a VOR (VHF Omnidirectional Radio range) station is located near Toussus-le-Noble airfield and implies a locally higher traffic density.

The following section focuses more precisely on the description of the aircraft database and details how the probability estimates are computed by simulation.

3. Flight simulation and collision probability estimation

In this section, we propose to present the aircraft trajectory database that will be used extensively in this article and then we present a specific statistical technique to estimate collision probabilities.

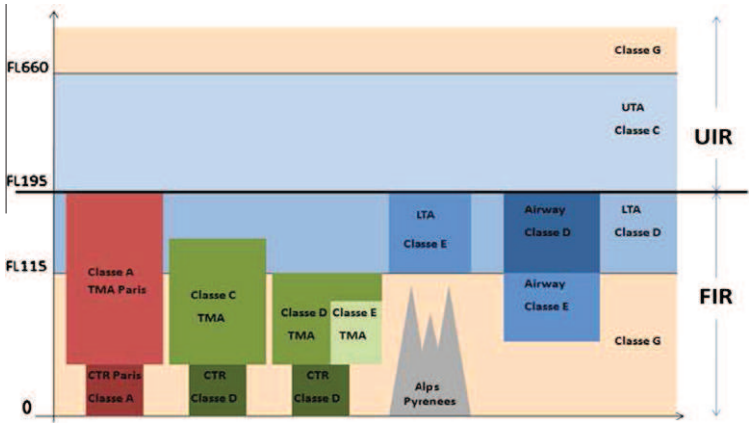


Fig. 1. French airspace description (Flight information region (FIR), Upper flight information region (UIR), TMA (Terminal Maneuvring Area), LTA (Lower Traffic Area), UTA (Upper Traffic Area), CTR (Contol Traffic Region) and FL (Flight Level)).

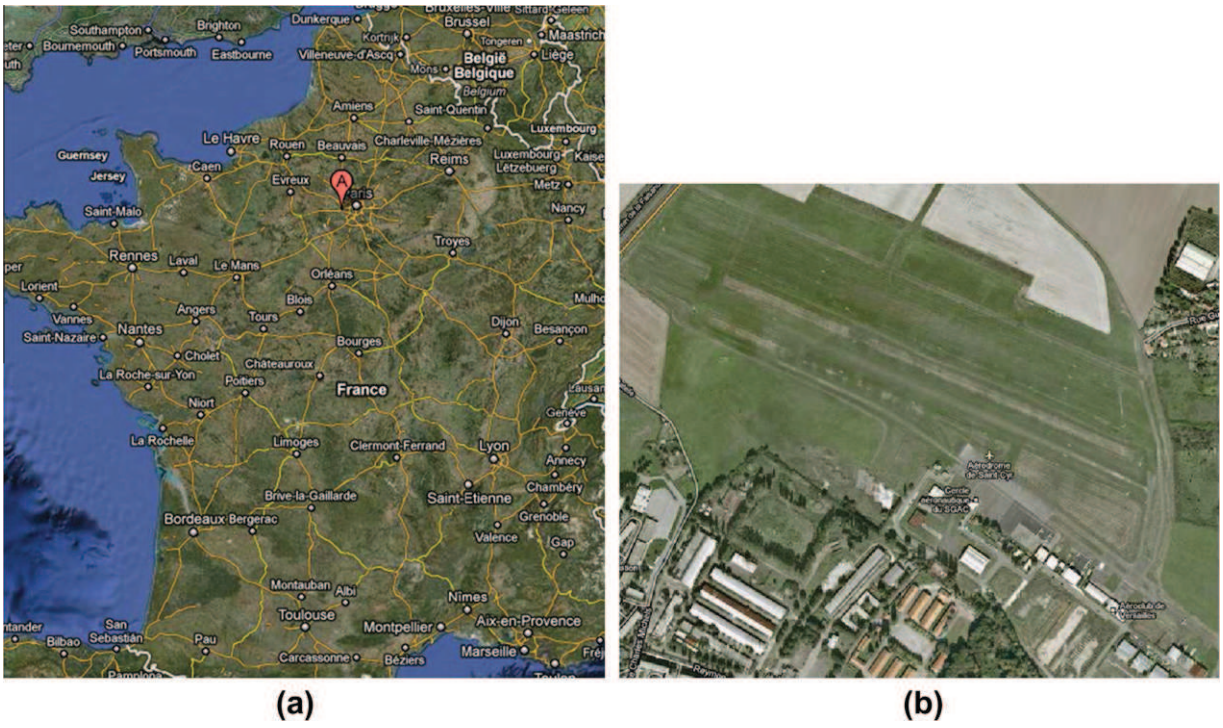


Fig. 2. Location of Saint-Cyr-l'Ecole (a) and aerial photography of Saint-Cyr-l'Ecole airfield (b). Credits given to Google Maps and Geoportail.fr

3.1. Trajectory database

In a class G airspace, aircraft number, position and routes are not known and collected. It is quite impossible to estimate collision probabilities with confidence in a class G airspace. For that purpose, 170 aircraft flight trajectories over Saint-Cyr-l'Ecole airfield in class G airspace were collected via GPS during year 2010. The aircraft that have been used to obtain this database were Robin DR-400 and Robin DR-221. During the 169 h of flight, aircraft positions are updated every 11 s. It is consequently an important database that can give an overview of collision probabilities over Saint-Cyr-l'Ecole area. Fig. 6 presents the 170 trajectory positions of the database. We focus more precisely on Saint-Cyr-l'Ecole country in Fig. 7 to estimate collision probabilities. The trajectory database is adapted to this estimation since all the trajectories take off or land at Saint-Cyr-l'Ecole airfield. Thus, the density of trajectories in this country is very high and one may think that reliable probability estimates can be found. A 3D view of the trajectories is proposed in Fig. 8. A mean trajectory lasts about 1 h, takes off and lands at Saint-Cyr-l'Ecole airfield, has a mean height of 458 m. All the trajectories follow the rules applicable in a class G airspace and follow the rules of visual landing and take off applied at Saint-Cyr-l'Ecole airfield.

3.2. Collision probability estimation

The objective of this article is to estimate collision probabilities in a class G airspace. This is obtained by determining the probability that the separation between aircraft be lower than S meters.

For that purpose, we propose to simulate days of flight traffic around Saint-Cyr-l'Ecole airfield. The number of flight movements (takeoff or landing) around Saint-Cyr-l'Ecole airfield is set to 500 per day in agreement with the statistics available for this airfield. The number of takeoffs and landings are assumed to be equal. Thus, the variable N of flights is set to 250.

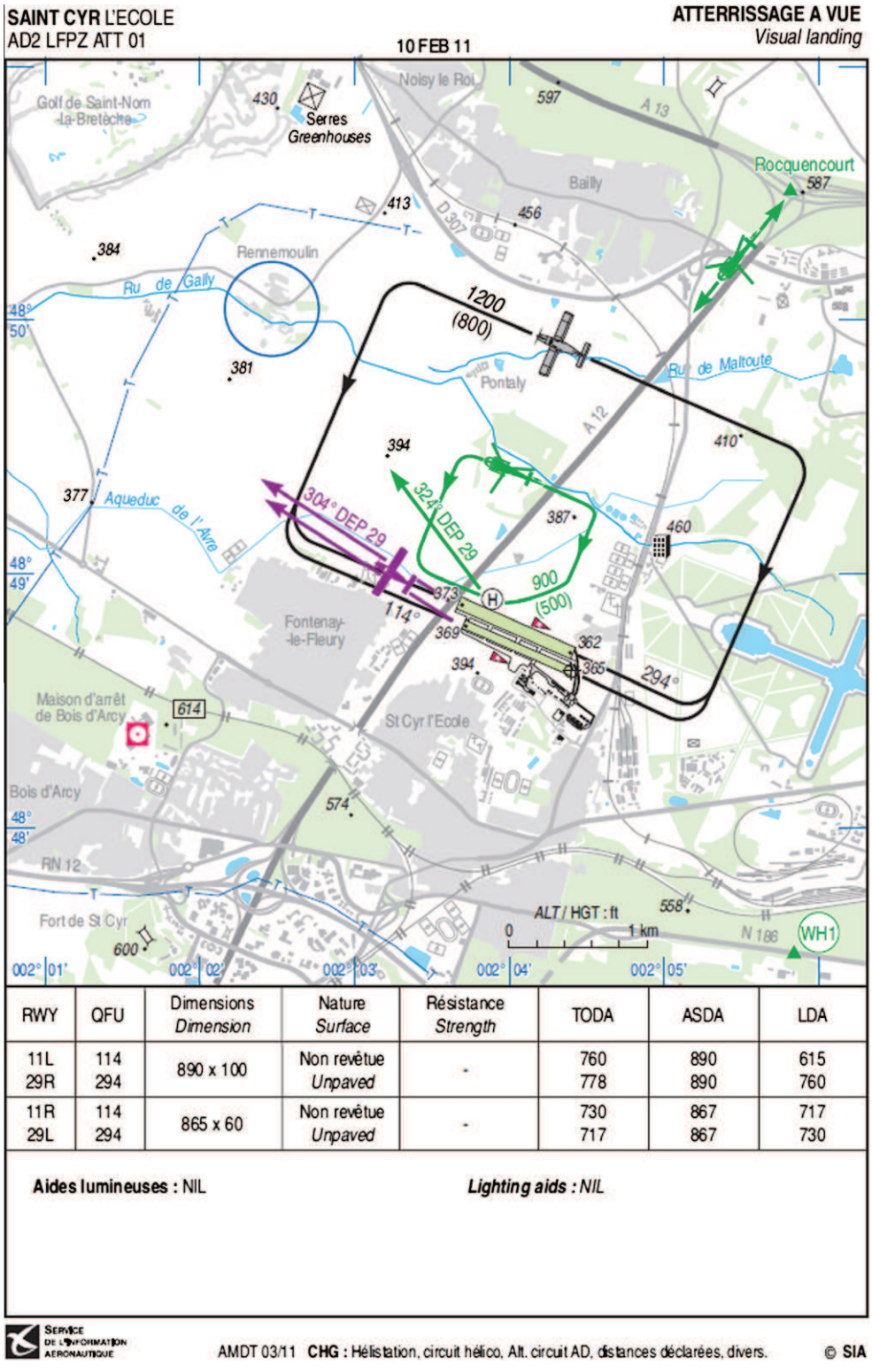
The sampling density f represents the density of aircraft takeoffs during a day function of day hours. It is illustrated in Fig. 9. There are two peaks during a day (one at 10 am and one at 3 pm) and the density of aircraft takeoffs decreases at midday. A sample histogram of this PDF is given in Fig. 10 to evaluate the number of movements at the different moments of a day. It is a realistic model of the traffic density at Saint-Cyr-l'Ecole airfield.

3.3. Monte Carlo methods

Monte Carlo methods are classical tools for probability estimation. Monte Carlo simulation of safety relevant air traffic scenarios have already been used to evaluate for systemic accident (Stroeve et al., 2009). Monte Carlo methods can be applied in the following way to estimate the collision probabilities around Saint-Cyr-l'Ecole airfield:

(1) Generate N aircraft simulations over Saint-Cyr-l'Ecole. An aircraft simulation i is defined by a random departure time T_i, $i = 1,\ldots,N$ that follows the PDF f and a trajectory. The trajectory is chosen with a uniform law in the 170 real trajectory database. Since the simulated aircraft follows exactly the real chosen trajectory with the same speed and the same position, the position of each aircraft is exactly known during all the simulation process.
(2) Estimate the minimum distance d^{ij} between aircraft i and j during the simulation process. This distance is a function of the departure time of aircraft i and j. It can be expressed as $d^{ij} = \phi(T_i, T_j)$ with $\phi : \mathbb{R}^2 \to \mathbb{R}$ a continuous function.
(3) Estimate, during the simulation time, the number C of conflicts between aircraft, that is the number of times the distance d^{ij} for all i, and with $j > i$ is lower than S. Let us define H as the number of simulated flight hours. The probability that the separation between aircraft be lower than S per flight hour is given by the ratio $P^{MC} = \frac{C}{H}$.

Unfortunately, this estimator is not efficient because the target probability is rare. Indeed, the relative deviation of the estimator P^{MC} is given by the ratio $\frac{\sigma_{P^{MC}}}{P^{MC}}$ with $\sigma_{P^{MC}}$, the standard deviation of P^{MC}. Knowing the true probability P, one has



RWY	QFU	Dimensions *Dimension*	Nature *Surface*	Résistance *Strength*	TODA	ASDA	LDA
11L 29R	114 294	890 x 100	Non revêtue *Unpaved*	-	760 778	890 890	615 760
11R 29L	114 294	865 x 60	Non revêtue *Unpaved*	-	730 717	867 867	717 730

Aides lumineuses : NIL **Lighting aids :** NIL

Fig. 3. VAC map of visual landing at Saint-Cyr-l'Ecole airfield.

$$\frac{\sigma_{P^{MC}}}{P^{MC}} = \frac{1}{\sqrt{N}} \frac{\sqrt{P - P^2}}{P} \tag{1}$$

Considering rare event probability estimation, that is when P takes low values, one has

$$\lim_{P \to 0} \frac{\sigma_{P^{MC}}}{P^{MC}} = \lim_{P \to 0} \frac{1}{\sqrt{NP}} = +\infty \tag{2}$$

The relative deviation of Monte Carlo estimation is very important. This leads to the conclusion that Monte Carlo methods are not adapted to rare event probability estimation and thus not suited to collision probability estimation. To overcome these difficulties, a recent advanced method to estimate rare event probabilities is derived and applied to the case of collision probability.

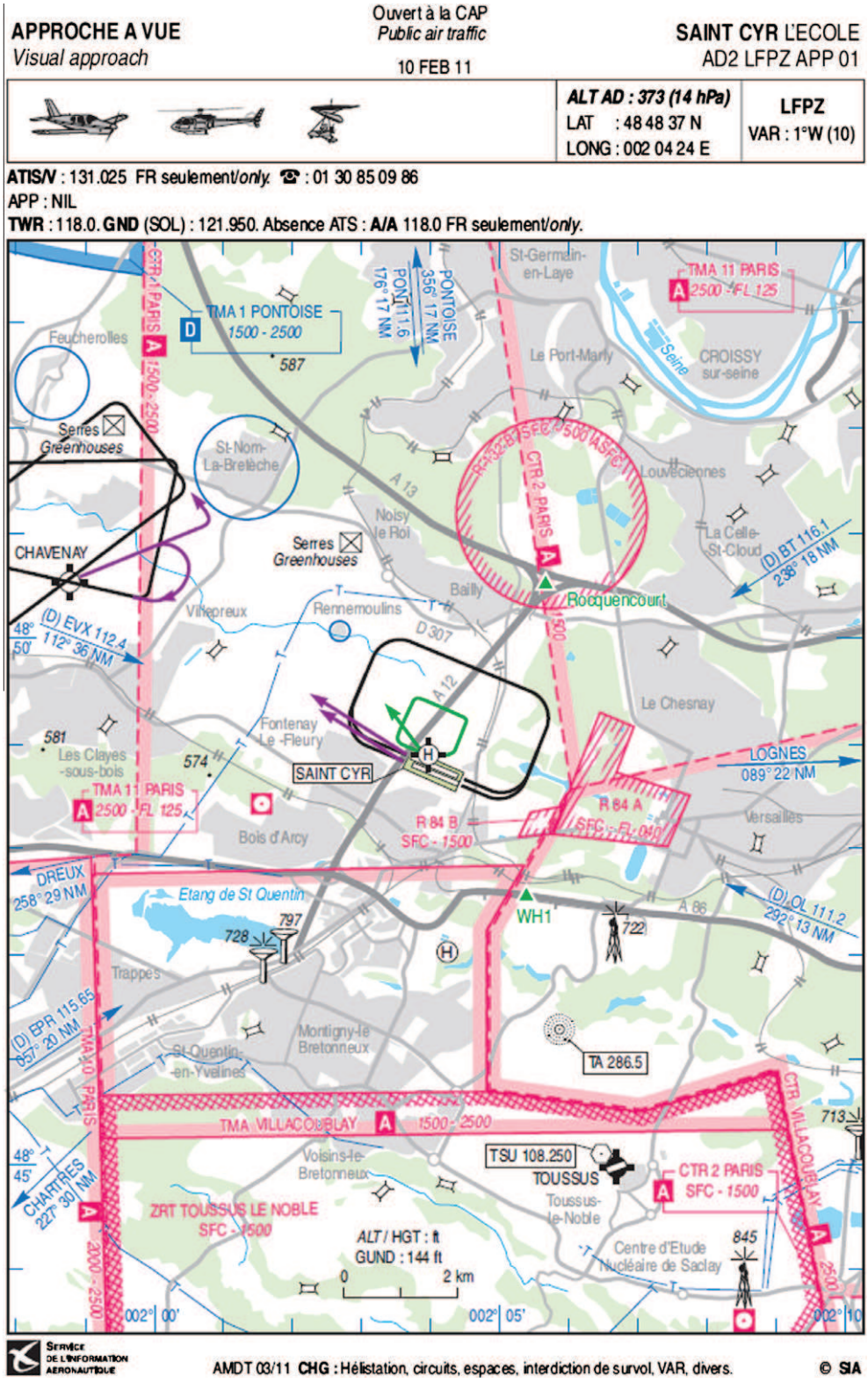


Fig. 4. VAC map of visual approach at Saint-Cyr-l'Ecole airfield.

3.4. Importance splitting for collision probability estimation

We consider a very efficient but not well-known algorithm called importance splitting (ISp). ISp is an alternative to Monte Carlo methods. We firstly propose to present the principle of importance splitting and then to implement it to the case of collision probability estimation.

3.4.1. Principle

Let us define the two-dimensional departure time random vector $\widetilde{T} = (T_i, T_j)$ of aircraft i and j. If one considers the set $\mathbf{A}^{ij} = \{\widetilde{T} \in \mathbb{R}^2 | d^{ij} < S\}$, the density on the set $\mathbf{A}^{ij}$ represents the probability that the distance between aircraft i and j be lower than S. The objective of this article is to determine the probability on the

Fig. 5. Other airfields in the surroundings of Saint-Cyr-l'Ecole airfield.

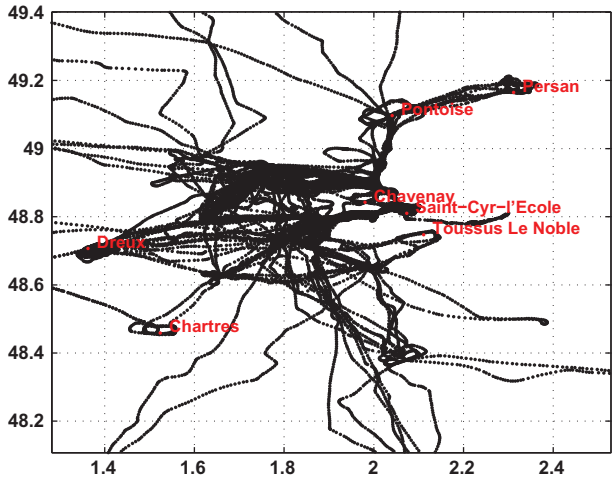


Fig. 6. The 170 aircraft trajectories of database plotted in longitude and latitude.

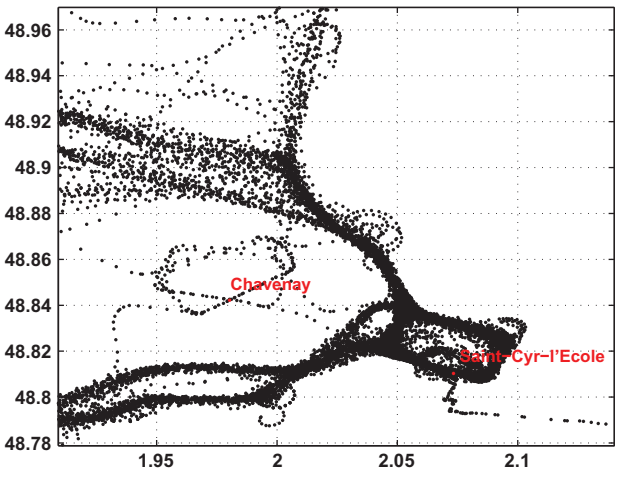


Fig. 7. Trajectory database positions in longitude and latitude in the vicinity of Saint-Cyr-l'Ecole.

set $A = \bigcup_{i=1\ldots N,\ j>i} \mathbf{A}^{ij}$. We propose to estimate independently the probability on the set $\mathbf{A}^{ij}$ and determine the probability on the set A without taking in account the potential correlation between the sets $\mathbf{A}^{ij}$. This assumption is realistic since the aircraft follow their routes in the simulation in an independent manner. The

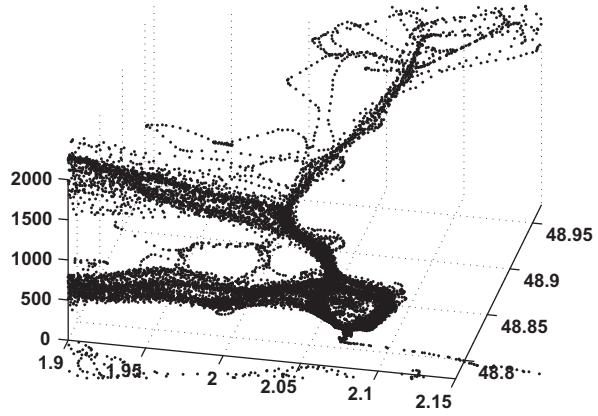


Fig. 8. Trajectory database positions in longitude, latitude and height (in meters) in the vicinity of Saint-Cyr-l'Ecole.

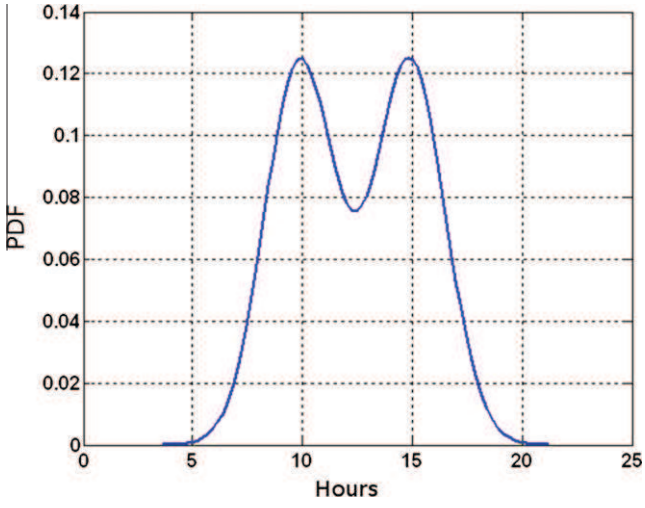


Fig. 9. Density of aircraft takeoffs during a day at Saint-Cyr-l'Ecole airfield.

principle of ISp is to estimate iteratively supersets of the set $\mathbf{A}^{ij}$ and then to estimate $P(\widetilde{T} \in \mathbf{A}^{ij})$ with conditional probabilities.

Let us define $\mathbf{A}_0^{ij} = \mathbb{R}^2 \supset \mathbf{A}_1^{ij} \supset \cdots \supset \mathbf{A}_{n-1}^{ij} \supset \mathbf{A}_n^{ij} = \mathbf{A}^{ij}$, a decreasing sequence of $\mathbb{R}^2$ subsets with smallest element $\mathbf{A}^{ij} = \mathbf{A}_n^{ij}$. The probability $P(\widetilde{T} \in \mathbf{A}^{ij})$ can be then rewritten in the following way :

$$P(\widetilde{T} \in \mathbf{A}^{ij}) = \prod_{k=1}^{n} P\left(\widetilde{T} \in \mathbf{A}_k^{ij} \middle| \widetilde{T} \in \mathbf{A}_{k-1}^{ij}\right)$$

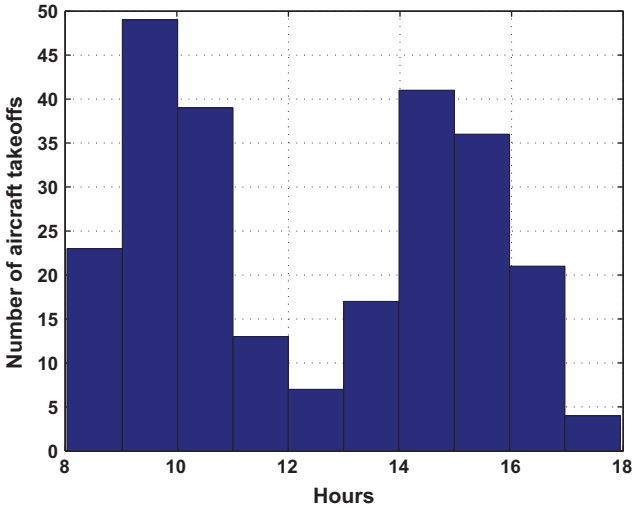


Fig. 10. Histogram of aircraft takeoffs during a day at Saint-Cyr-l'Ecole airfield.

where $P\left(\widetilde{T} \in \mathbf{A}_k^{ij} \middle| \widetilde{T} \in \mathbf{A}_{k-1}^{ij}\right)$ is the probability that $\widetilde{T} \in \mathbf{A}_k^{ij}$ knowing that $\widetilde{T} \in \mathbf{A}_{k-1}^{ij}$. An optimal choice of the sequence $\mathbf{A}_k^{ij}, k = 0 \ldots n$ is given when $P\left(\widetilde{T} \in \mathbf{A}_k^{ij} \middle| \widetilde{T} \in \mathbf{A}_{k-1}^{ij}\right) = p$, where p is a constant, that is when all the conditional probabilities are equal. The variance of $P\left(\widetilde{T} \in \mathbf{A}^{ij}\right)$ is indeed minimized in this configuration as shown in (Lagnoux, 2006; Cerou et al., 2006). Consequently, if each $P\left(\widetilde{T} \in \mathbf{A}_k^{ij} \middle| \widetilde{T} \in \mathbf{A}_{k-1}^{ij}\right)$ is well estimated, then the probability $P(\widetilde{T} \in \mathbf{A}^{ij})$ is estimated more accurately with ISp than with a direct estimation by Monte Carlo.

The subset $\mathbf{A}_k^{ij}$ sequence is easily evaluated in the following way. Indeed, it can be defined with $\mathbf{A}_k^{ij} = \{\widetilde{T} \in \mathbb{R}^2 | d^{ij} < S_k\}$ for $k = 1 \ldots n$ with $S = S_n < S_{n-1} < \ldots < S_k < \ldots < S_1$. Determining the sequence $\mathbf{A}_k^{ij}$ is equivalent to choose some values for S_k, with $k = 1 \ldots n$. One has presented in this subsection how to estimate the probability that the distance between aircraft i and j be lower than S. In the following, we propose to explain how this algorithm is implemented and how the target probability is estimated.

3.4.2. Implementation

The different stages of the algorithm ISp to estimate $P(d^{ij} < S)$ are analyzed here.

(1) Set $k = 0$, $f_0 = f \times f$ and $N_0 = N$.
(2) Generate N_k aircraft simulations over Saint-Cyr-l'Ecole with departure times $\widetilde{T}_1^{(k)}, \ldots, \widetilde{T}_{N_k}^{(k)}$ generated from f_k, the conditional density of $\widetilde{T}$ when $\widetilde{T}$ is restricted to the set $\mathbf{A}_k^{ij}$. Since f_k is not always easily sampled from in practice, one can use the Metropolis–Hastings algorithm to generate new departure times (Tierney, 1994; Roberts et al., 1997).
(3) Estimate the α-quantile $q_\alpha^{(k)}$ of the samples $\phi\left(\widetilde{T}_1^{(k)}\right), \ldots, \phi\left(\widetilde{T}_{N_k}^{(k)}\right)$ and set $S_{k+1} = q_\alpha^{(k)}$.
(4) Determine the subset $\mathbf{A}_{k+1}^{ij}$ with $\mathbf{A}_{k+1}^{ij} = \{\widetilde{T} \in \mathbb{R}^2 | d^{ij} < S_{k+1}\}$ and the conditional density f_{k+1}. By definition of quantile, one has $P\left(\widetilde{T} \in \mathbf{A}_{k+1}^{ij} \middle| \widetilde{T} \in \mathbf{A}_k^{ij}\right) = cste = 1 - \alpha$.
(5) If $S_{k+1} > S$, set $k = k + 1$ and go back to stage (2) of the algorithm. Otherwise, set $k = k + 1$ and estimate the probability that the distance between aircraft i and j be lower than S with

$$P^{ij} = (1-\alpha)^k \times \frac{1}{N_k} \sum_{i=1}^{N_k} \mathbf{1}_{\phi\left(\widetilde{T}_i^{(k)}\right) < S}$$

where $\mathbf{1}_{\phi\left(\widetilde{T}_i^{(k)}\right) < S}$ is equal to 1 if $\phi\left(\widetilde{T}_i^{(k)}\right) < S$ and 0 otherwise. The collision probability per flight hour is given by the ratio

$$P^{ISp} = \sum_{i=1}^{N} \sum_{j=i+1}^{N} \frac{P^{ij}}{H} \tag{3}$$

In the algorithm, the parameters N_k are set by the user and in most cases so that $N_k = N$ whatever the value of k. The variance of the estimate has been shown to be lower than the Monte Carlo estimate variance for rare events and its convergence has been deeply analyzed in (Cerou et al., 2008; Cerou and Guyader, 2007; L'Ecuyer et al., 2006). In the following section, we propose to apply this algorithm to estimate accurate collision probabilities between aircraft in class G airspace over Saint-Cyr-l'Ecole airfield.

4. Results

In this section, we propose to analyze the collision or separation loss probability results. Firstly, we focus on separation loss probability estimates by flight hour in the simulation. Then, one derives spatial collision probability map in order to characterize the spatial variability of the probability estimate. We finally make the synthesis of these results by raising some limits that has to be taken into account in order to improve probability estimates.

4.1. Overall results

We have applied Monte Carlo simulations and importance splitting algorithm described in the previous sections in order to evaluate the probability that the distance between two aircraft be lower than S meters by hour of flight. Table 1 presents the estimation results obtained for $H = 100{,}000$ h of simulated flight with

Table 1
Collision and separation loss probability estimates with Monte Carlo and importance splitting simulations for about 100,000 h of simulated flight ($N = 250$).

Distance S (m)	P^{MC} (relative error)	P^{ISp} (relative error)
10	0 (?%)	3.210^{-10} (92%)
50	0 (?%)	5.710^{-9} (53%)
100	1.510^{-8} (122%)	3.810^{-8} (38%)
200	7.510^{-7} (78%)	7.810^{-7} (33%)
500	1.410^{-4} (42%)	1.510^{-4} (28%)

Table 2
Collision and separation loss probability estimates with Monte Carlo and importance splitting simulations with a low number of flight movements ($N = 100$).

Distance S (m)	P^{MC} (relative error)	P^{ISp} (relative error)
10	0 (?%)	4.210^{-11} (97%)
50	0 (?%)	4.210^{-10} (51%)
100	7.310^{-8} (143%)	9.210^{-8} (42%)
200	1.710^{-8} (87%)	3.610^{-8} (38%)
500	6.310^{-5} (54%)	6.210^{-5} (25%)

Table 3
Collision and separation loss probability estimates with Monte Carlo and importance splitting simulations with a high number of flight movements ($N = 1000$).

Distance S (m)	P^{MC} (relative error)	P^{ISp} (relative error)
10	4.510^{-8} (152%)	4.210^{-8} (65%)
50	6.210^{-7} (77%)	4.210^{-7} (42%)
100	9.610^{-5} (54%)	9.210^{-5} (29%)
200	3.810^{-4} (37%)	3.610^{-4} (23%)
500	6.110^{-2} (10%)	6.210^{-2} (12%)

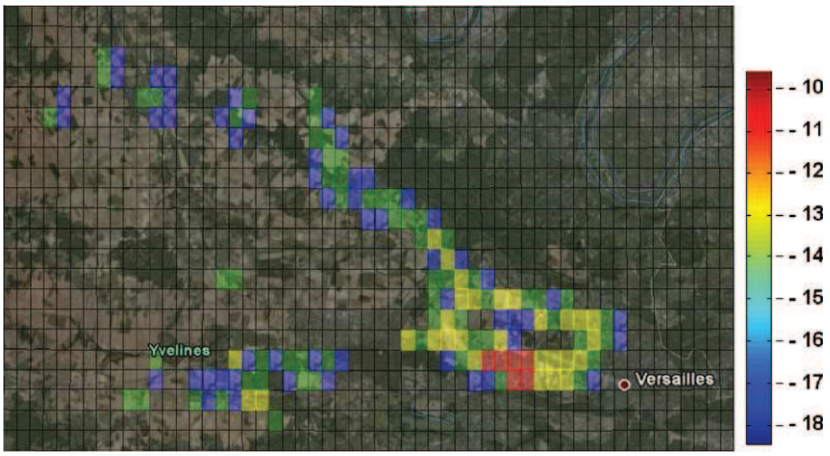


Fig. 11. Map of collision probability estimates with importance splitting over Saint-Cyr-l'Ecole region. The color scale depends on the logarithm of the probability. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Monte Carlo and importance splitting simulations. These estimates have been obtained over 100 retrials that also allow us to compute the estimation relative error.

The estimated probabilities are very low which explains why Monte Carlo simulations are not adapted to such estimations. Monte Carlo methods are notably not able to generate randomly events where the distance between two aircraft be lower than 10 m. On the contrary, importance splitting can evaluate rare event with a valuable relative deviation.

ESSAR4 assumes that maximum tolerable probability of ATM directly contributing to a severe accident (mostly collisions) of a commercial air transport aircraft of 1.55×10^{-8} accidents per controlled flight hour (Eurocontrol, 2001). It seems that in the class G airspace over Saint-Cyr-l'Ecole the collision probability is lower than in controlled airspace because the air traffic is less dense. When the threshold S increases, the probability is of course higher. Nevertheless even with the higher threshold $S = 500$ m, the probability is lower than 10^{-3} event per hour of flight.

In fact, these probabilities are directly linked to traffic densities. Indeed, we show in Table 2, that if we consider a lower number of flight movements at Saint-Cyr-l'Ecole airfield $N = 100$ for the same $H = 100{,}000$ h of simulated flight, the separation loss probabilities decrease non-linearly. On the contrary, as shows in Table 3, if the number of movements increases to reach $N = 1000$, the separation loss probabilities strongly increase. The airfield surroundings are saturated by aircraft and consequently the mean distance between each aircraft decreases and involves a high separation loss probability. The knowledge of the air traffic density over a region is thus very important to obtain valuable probability results.

The simulation results give an order of the collision and separation loss probabilities one can expect in a class G airspace. These levels could be used to set minimum safety requirements in class G airspace.

4.2. Maps of collision probabilities

In this section, we propose to evaluate the separation loss probability in function of the aircraft position. For that purpose, we use the simulation results obtained at the previous section and keep for each separation loss, their exact spatial position. It is then not difficult to estimate the separation loss probability per flight hour and per surface unity. We present in Fig. 11, a map of the mean probability over 100 retrials that the distance between two aircraft be lower than 50 m by hour of flight obtained with importance splitting.

The most dangerous sectors over Saint-Cyr-l'Ecole are in the vicinity of the field and the airfield pattern. It is no wonder since there is a high traffic density in this sector and that consequently increases the risk of separation loss. When no color has been set in map, it means that no separation loss has been encountered during the simulation in this area. As we can notice, there are many sectors with a very low risk of separation loss. One can conclude that there is a high variability of separation loss probability in function of the considered region. This fact has to be taken into account to draw general conclusions for separation loss probability in class G airspace.

In Fig. 12a–c, we present the separation loss maps for different flight levels. At low flight level, the separation loss probability is high only near Saint-Cyr-l'Ecole airfield and this risk is low in the other areas. At higher level flight, the separation loss probability is high in the different directions required to land and to take off from Saint-Cyr-l'Ecole airfield.

4.3. Limits of the simulation results

The simulation results obtained in the previous sections give an overview of the separation loss probabilities in the vicinity of Saint-Cyr-l'Ecole. Nevertheless, one has to qualify these estimations since, as in every simulation modelling (Brooker, 2006), several assumptions have been made:

- The pilot in the loop is not taken in account. Indeed, some separation losses between aircraft in class G airspace will be avoid by their pilots (Graham and Orr, 1970). VFR conditions are assumed in the simulation. Thus, there is a non-negligible probability that a pilot can see that other aircraft is at a low distance from his own aircraft and decides to change his direction to avoid a possible conflict. The probability estimates found in this article could be then decreased by this factor.
- The potential air traffic involved by other airfields is not considered. The number of aircraft in flight over Saint-Cyr-l'Ecole is maybe slightly greater than the one we consider.
- we have not introduced uncertainties in the trajectory database. Indeed, a simulated aircraft follow exactly the position and speed defined in the trajectory database. It could be interesting to develop a valuable model to add uncertainties on the aircraft positions and speed to increase the variety of the aircraft trajectories. In the same time, it will also reduce the impact of the trajectory database and of its genericity for probability estimation.

It is difficult to evaluate quantitatively the impact of these assumptions on the separation loss probability estimation in class G airspace.

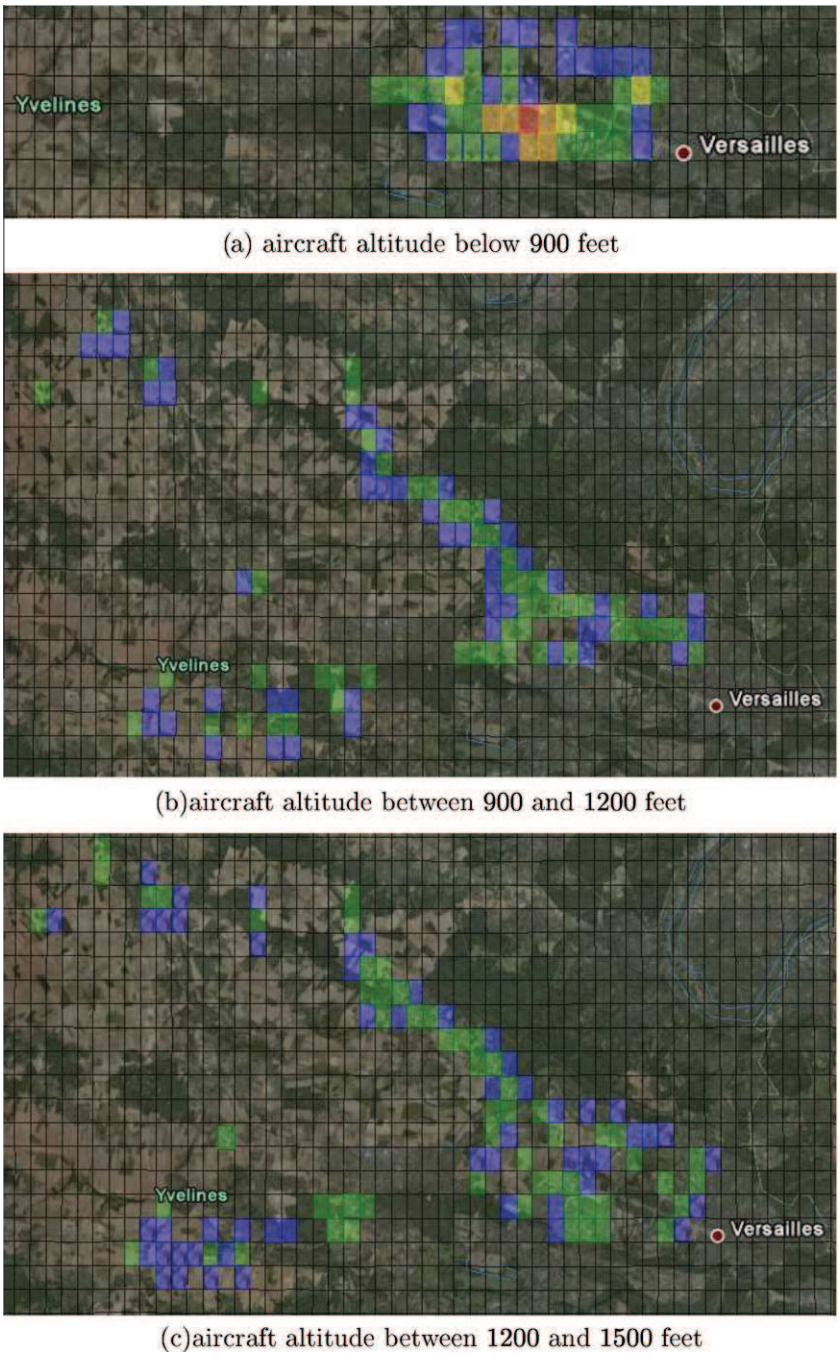


Fig. 12. Map of collision probability estimates with importance splitting in the vicinity of Saint-Cyr-l'Ecole. (Same color scale as in Fig. 11.) (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

5. Conclusion

In this article, we proposed to estimate the separation loss probability estimation in uncontrolled class G airspace. For that purpose, we have built a trajectory database over Saint-Cyr-l'Ecole and computed a simulation to estimate these probabilities. It enables us to estimate separation loss probabilities in class G airspace near Saint-Cyr-l'Ecole and then to analyze their spatial variability.

What are the practical implications of this research? As statistics of collision or separation loss are hard to compute by experience in uncontrolled airspace, these simulation results give a first idea of the safety probability one can expect in this type of airspace. Collision or separation loss risks in uncontrolled airspace are not very significant in the considered case and it seems that it is not mandatory to increase the safety requirements in uncontrolled airspace. An interesting area of research is also the integration of UAV (Unmanned Aerial Vehicle) in uncontrolled airspace. The obtained results show that UAV could be integrated safely in some sectors of uncontrolled airspace. Indeed the risk of collision between an UAV and an aircraft would be relatively low in uncontrolled airspace unless near airfields where the risk can be quite higher than in other sectors.

Because of the high spatial variability of the estimated probabilities, it would be interesting to make a same study in other regions where the air traffic is potentially significant. Tourist regions, (the Loire valley, France) or regions where air control monitoring is provided (the Mediterranean area, South of Sicily, Italy) can be of interest in that case.

Acknowledgements

This work is supported financially by DGA (Direction Générale de l'Armement). The authors are thankful to J. Cornu (Onera), A. Delloye (DGA), S. Mérit (Onera) R. Pastel (Onera), H. Piet-Lahanier (Onera), N. Tessaud (DGA) and A. Vergne for fruitful discussions and the two anonymous reviewers for their valuable remarks.

References

Allouche, M., 2000. The integration of UAVs in airspace. Air Space Europe 2, 101–104.
Asmat, J., Rhodes, B., Umansky, J., Villavicencio, C., Yunas, A., Donohue, G., Lacher, A., 2006. UAS safety: unmanned aerial collision avoidance system (UCAS). In: Systems and Information Engineering Design Symposium 2006, IEEE, pp. 43–49.
Brooker, P., 2005. Reducing mid-air collision risk in controlled airspace: lessons from hazardous incidents. Safety Science 43, 715–738.
Brooker, P., 2006. Air traffic management accident risk. Part 1: The limits of realistic modelling. Safety Science 44, 419–450.
Cerou, F., Del Moral, P., Furon, T., Guyader, A., 2008. Rare event simulation for a static distribution, RESIM. pp. 107–115.
Cerou, F., Del Moral, P., Le Gland, F., Lezaud, P., 2006. Genetic genealogical models in rare event analysis. INRIA Report 1797, 1–30.
Cerou, F., Guyader, A., 2007. Adaptive particle techniques and rare event estimation. ESAIM. pp. 65–72.
Eurocontrol, 2001. Risk assessment and mitigation in ATM. Eurocontrol Safely Regulatory Requirement. ESARR4 1.0.
Glasserman, P., Heidelberger, P., Shahabuddin, P., Zajic, T., 1996. Splitting for rare event simulation: analysis of simple cases. In: Proceeding of the 1996 Winter Simulation Conference, pp. 302–308.
Graham, W., Orr, R., 1970. Separation of air traffic by visual means: an estimate of the effectiveness of the see-and-avoid doctrine. Proceedings of the IEEE 58, 337–361.
Irvine, R., 2001. A simplified approach to conflict probability estimation. In: Digital Avionics Systems, 2001. DASC. The 20th Conference, vol. 2, pp. 7F5/1–7F5/12.
Kirkland, I.D.L., Caves, R.E., Humphreys, I.M., Pitfield, D.E., 2004. An improved methodology for assessing risk in aircraft operations at airports, applied to runway overruns. Safety Science 42, 891–905.
Kochenderfer, M., Espindle, L., Griffith, J., Kuchar, J., 2008a. A Bayesian approach to aircraft encounter modeling. In: AIAA Guidance, Navigation and Control Conference and Exhibit, 2008, GNCC, pp. 1–21.
Kochenderfer, M., Espindle, L., Griffith, J., Kuchar, J., 2008b. Encounter modeling for sense and avoid development. In: Integrated Communications, Navigation and Surveillance Conference, 2008. ICNS 2008, pp. 1–10.
Lagnoux, A., 2006. Rare event simulation. Probability in the Engineering and Informational science 20, 45–66.
L'Ecuyer, P., Demers, V., Tuffin, B., 2006. Splitting for rare event simulation. In: Proceeding of the 2006 Winter Simulation Conference, pp. 137–148.
Leva, M.C., Ambroggi, M.D., Grippa, D., Garis, R.D., Trucco, P., StrSter, O., 2009. Quantitative analysis of ATM safety issues using retrospective accident data: the dynamic risk modelling project. Safety Science 47, 250–264.
Mikhailov, G.A., 1999. Parametric Estimates by the Monte Carlo Method. VSP, Utrecht (NED).
Morio, J., Pastel, R., Le Gland, F., 2010. An overview of importance splitting for rare event simulation. European Journal of Physics 31, 1295–1303.
Ostwald, P., Hershey, W., 2007. Helping Global Hawk fly with the rest of US. In: Integrated Communications, Navigation and Surveillance Conference, 2007. ICNS '07, pp. 1–11.
Robert, C.P., Casella, G., 2005. Monte Carlo Statistical Methods. Springer, New York.
Roberts, G., Gelman, A., Gilks, W., 1997. Weak convergence and optimal scaling of random walk Metropolis algorithms. The Annals of Applied Probability 7, 110–120.
Shang-wen, Y., Ming-hua, H., 2010. Estimation of air traffic longitudinal conflict probability based on the reaction time of controllers. Safety Science 48, 926–930.
Sobol, I.M., 1994. A Primer for the Monte Carlo Method. CRC Press, Boca Raton, FL.
Stroeve, S.H., Blom, H.A., Bakker, G.B., 2009. Systemic accident risk assessment in air traffic by monte carlo simulation. Safety Science 47, 238–249.
Thom, T., Godwin, P., 2007. The Air Pilot's Manual: Aviation Law And Meteorology. Air Pilot Publisher Ltd., Cranfield, England.
Tierney, L., 1994. Markov chains for exploring posterior distributions. Annals of Statistics 22, 1701–1762.
Wang, W., Jiang, X., Xia, S., Cao, Q., 2010. Incident tree model and incident tree analysis method for quantified risk assessment: an in-depth accident study in traffic operation. Safety Science 48, 1248–1262.

Chapitre 5

Kriging-based adaptive importance sampling algorithms for rare event estimation, *Structural Safety*, M. Balesdent, J. Morio and J. Marzat, 2013, Vol. 44, pp. 1-10

Structural Safety 44 (2013) 1–10

Contents lists available at SciVerse ScienceDirect

Structural Safety

journal homepage: www.elsevier.com/locate/strusafe

Kriging-based adaptive Importance Sampling algorithms for rare event estimation

Mathieu Balesdent*, Jérôme Morio, Julien Marzat

Onera – The French Aerospace Lab, F-91123 Palaiseau, France

ARTICLE INFO

Article history:
Received 29 May 2012
Received in revised form 25 April 2013
Accepted 25 April 2013

Keywords:
Surrogate model
Importance Sampling
Rare event estimation
Input–output function
Kriging

ABSTRACT

Very efficient sampling algorithms have been proposed to estimate rare event probabilities, such as Importance Sampling or Importance Splitting. Even if the number of samples required to apply these techniques is relatively low compared to Monte-Carlo simulations of same efficiency, it is often difficult to implement them on time-consuming simulation codes. A joint use of sampling techniques and surrogate models may thus be of use. In this article, we develop a Kriging-based adaptive Importance Sampling approach for rare event probability estimation. The novelty resides in the use of adaptive Importance Sampling and consequently the ability to estimate very rare event probabilities (lower than 10^{-3}) that have not been considered in previous work on similar subjects. The statistical properties of Kriging also make it possible to compute a confidence measure for the resulting estimation. Results on both analytical and engineering test cases show the efficiency of the approach in terms of accuracy and low number of samples.

1. Introduction

This paper focuses on the joint use of adaptive sampling techniques and surrogate models to estimate probability of rare events for costly *input–output* functions $\phi(.)$ that can represent, for instance, a simulation code. We assume that an important computation time is required to simulate output samples of $\phi(.)$ so that the number of calls should be minimum. We propose here a strategy to avoid the evaluation of the function ϕ by replacing it by a surrogate model ($\phi(.) \rightarrow \hat{\phi}(.)$). The problem under study is to detect whether the output of ϕ exceeds a given threshold S. In this case, in order to efficiently build the surrogate of the expensive function with a minimal number of evaluations, the surrogate model $\hat{\phi}(.)$ does not need to be the exact representation of $\phi(.)$ [1,2]. Indeed, if $\mathbf{1}_{\phi(.)>S}$ and $\mathbf{1}_{\hat{\phi}(.)>S}$ take the same values over the input space, the estimated probability will be the same with $\phi(.)$ or $\hat{\phi}(.)$.

Importance Sampling is a well-known adapted random simulation technique [3]. It consists in generating random weighted samples from an auxiliary distribution rather than the distribution of interest. The optimization of this auxiliary distribution is often difficult in practice, yet it can be made adaptively. This procedure often requires more than 10,000 evaluations of the function $\phi(.)$ to be efficient. The aim of the paper is thus to present a methodology combining an Importance Sampling probability estimator with a surrogate model for estimating very rare event probabilities (say 10^{-6} to 10^{-10}). To this end, we propose to develop a Kriging-based methodology derived from [1,2,4] to adaptive Importance Sampling techniques. We select two adaptive Importance Sampling algorithms: Cross-Entropy optimization [5,6] and an improved version of non parametric adaptive Importance Sampling [7,8].

2. Rare event estimation

2.1. Context

Consider a d-dimensional random vector $\mathbf{X}$ with a probability density function (pdf) h_0 and the need to estimate the probability that $P(\phi(\mathbf{X}) > S)$ with ϕ, a continuous "input–output" scalar function $\phi : \mathbb{R}^d \rightarrow \mathbb{R}$ and S a threshold. The function ϕ simply represents the input–output system in a considered application domain. We assume here that $\phi(\mathbf{X})$ is a random variable. A simple way to estimate this probability is to consider Monte Carlo methods [9–13]. For that purpose, one generates independent and identically distributed samples $\mathbf{X}_1$,...., $\mathbf{X}_N$ from the pdf h_0 and then estimates the probability with

$$\hat{P}^{\mathrm{MC}} = \frac{1}{N}\sum_{i=1}^{N} \mathbf{1}_{\phi(\mathbf{X}_i)>S}, \tag{1}$$

where $\mathbf{1}_{\phi(\mathbf{X}_i)>S}$ is equal to 1 if $\phi(\mathbf{X}_i) > S$ and 0 otherwise. The relative deviation of the Monte Carlo estimator $\hat{P}^{\mathrm{MC}}$ is given by the ratio $\frac{\sigma_{\hat{P}^{\mathrm{MC}}}}{\hat{P}^{\mathrm{MC}}}$ with $\sigma_{\hat{P}^{\mathrm{MC}}}$, the standard deviation of $\hat{P}^{\mathrm{MC}}$. Knowing the true probability P that $(\phi(\mathbf{X}) > S)$, one has

$$\frac{\sigma_{\hat{P}^{\mathrm{MC}}}}{\hat{P}^{\mathrm{MC}}} = \frac{1}{\sqrt{N}}\frac{\sqrt{P - P^2}}{P}. \tag{2}$$

* Corresponding author. Tel.: +33 180386608.
E-mail address: mathieu.balesdent@onera.fr (M. Balesdent).

0167-4730/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.strusafe.2013.04.001

Considering the case of rare event probability estimation, for which P takes low values, one has

$$\lim_{P\to 0}\frac{\sigma_{\hat{P}^{\mathrm{MC}}}}{\hat{P}^{\mathrm{MC}}}=\lim_{P\to 0}\frac{1}{\sqrt{NP}}=+\infty. \tag{3}$$

Relative deviation of the Monte Carlo estimation is very high and thus, one can conclude that Monte Carlo methods are not well suited to rare event probability estimation. Different alternatives to Monte Carlo can be considered, such as Importance Sampling [3,7,14–19], Importance Splitting [20–26] or Extreme Value Theory [27–32]. An adaptive version of Importance Sampling techniques is considered here.

2.2. Adaptive Importance Sampling

2.2.1. Principle of Importance Sampling

The objective of Importance Sampling (IS) is to reduce the variance of the Monte Carlo estimator $\hat{P}^{\mathrm{MC}}$ without increasing the number of samples N [3,14,19]. The main idea is to generate the samples $\mathbf{X}_1,...,\mathbf{X}_N$ with an auxiliary density h and to introduce a weight in the probability estimate. The IS probability estimate $\hat{P}^{\mathrm{IS}}$ is then given by

$$\hat{P}^{\mathrm{IS}}=\frac{1}{N}\sum_{i=1}^{N}\mathbf{1}_{\phi(\mathbf{X}_i)>S}\frac{h_0(\mathbf{X}_i)}{h(\mathbf{X}_i)}. \tag{4}$$

The term $\hat{P}^{\mathrm{IS}}$ is an unbiased estimate of the probability P. Its variance is given by the following equation,

$$Var\left(\hat{P}^{\mathrm{IS}}\right)=\frac{Var\left(\mathbf{1}_{\phi(\mathbf{X})>S}w(\mathbf{X})\right)}{N}, \tag{5}$$

with $w(\mathbf{X})=\frac{h_0(x)}{h(x)}$. The variance of $\hat{P}^{\mathrm{IS}}$ strongly depends on the choice of h. If h is well-chosen, the Importance Sampling estimate has then a much smaller variance than the Monte-Carlo estimate. The objective of Importance Sampling is to decrease the estimation variance and one can thus define an optimal IS auxiliary density that minimizes $Var(\hat{P}^{\mathrm{IS}})$. Since variances are non-negative quantities, the optimal auxiliary density h_{opt} is determined by canceling the variance in Eq. (5). It is well-known that h_{opt} is then defined as [33]

$$h_{\mathrm{opt}}=\frac{\mathbf{1}_{\phi(x)>S}h_0}{P}. \tag{6}$$

The optimal auxiliary density h_{opt} depends unfortunately on the probability P that one tries to estimate and is not exploitable in practice. Nevertheless, h_{opt} can be useful to determine an efficient sampling pdf. Indeed, a valuable sampling auxiliary pdf h will be close to the pdf h_{opt} according to a given criterion. Optimizing the auxiliary sampling pdf is then necessary. For this purpose, two algorithms are described in the following sections.

2.2.2. Cross-Entropy optimization of Importance Sampling auxiliary density

Let us define $h(\cdot,\boldsymbol{\lambda})$, a family of densities indexed by a parameter vector $\boldsymbol{\lambda}\in\Delta$ where Δ is the multidimensional space of pdf parameters. The parameter vector $\boldsymbol{\lambda}$ is, for instance, the mean and the variance in the case of Gaussian densities. The objective of Importance Sampling with Cross-Entropy (CE) is to determine the parameter vector $\boldsymbol{\lambda}_{\mathrm{opt}}$ that minimizes the Kullback–Leibler divergence between $h(\cdot,\boldsymbol{\lambda}_{\mathrm{opt}})$ and h_{opt} [5,6]. The value of $\boldsymbol{\lambda}_{\mathrm{opt}}$ is thus obtained as follows

$$\lambda_{\mathrm{opt}}=\underset{\lambda}{\mathrm{argmin}}\left\{D\left(h_{\mathrm{opt}},h(\cdot,\lambda)\right)\right\}, \tag{7}$$

where the notation $x_{\mathrm{opt}}=\underset{x}{\mathrm{argmin}}\{f(x)\}$ stands for the value of x that minimizes $f(x)$ and D is the Kullback–Leibler divergence defined between two densities p and q by

$$D(q,p)=\int_{\mathbb{R}^d}q(x)\ln(q(x))\,dx-\int_{\mathbb{R}^d}q(x)\ln(p(x))\,dx. \tag{8}$$

For example, $h(\cdot,\boldsymbol{\lambda})$ can be chosen in the parameterized Gaussian density family, that gives (if $\phi:\mathbb{R}\to\mathbb{R}$):

$$h(x,\lambda)=\frac{1}{\sqrt{2\pi}\lambda_1}\exp\left(-\frac{(x-\lambda_2)^2}{2\lambda_1^2}\right), \tag{9}$$

with $\boldsymbol{\lambda}=[\lambda_1,\lambda_2]$. Determining $\boldsymbol{\lambda}_{\mathrm{opt}}$ with Eq. (7) is not obvious since it depends on the unknown pdf h_{opt}. In fact, it can be shown [5] that Eq. (7) is equivalent to

$$\lambda_{\mathrm{opt}}=\underset{\lambda}{\mathrm{argmax}}\left\{\mathbb{E}\left[\mathbf{1}_{\phi(\mathbf{X})>S}\ln(h(\mathbf{X},\lambda))\right]\right\}, \tag{10}$$

where $\mathbb{E}$ defines the expectation operator. In fact, one does not focus directly on Eq. (10) but proceeds iteratively to estimate $\boldsymbol{\lambda}_{\mathrm{opt}}$ with an iterative sequence of thresholds,

$$q_0<q_1<q_2<\cdots<q_k<\cdots\le S, \tag{11}$$

chosen adaptively using quantile definition.

At iteration k, the value $\boldsymbol{\lambda}_{k-1}$ is available and one maximizes in practice

$$\lambda_k=\underset{\lambda}{\mathrm{argmax}}\frac{1}{N}\sum_{i=1}^{N}\mathbf{1}_{\phi(\mathbf{X}_i)>\gamma_k}\frac{h_0(\mathbf{X}_i)}{h(\mathbf{X}_i,\lambda_{k-1})}\ln(h(\mathbf{X}_i,\lambda)). \tag{12}$$

where the samples $\mathbf{X}_1,...,\mathbf{X}_N$ are generated with $h(\cdot,\boldsymbol{\lambda}_{k-1})$. The probability $\hat{P}^{\mathrm{CE}}$ is then estimated with Importance Sampling at the last iteration. The Cross-Entropy optimization algorithm for Importance Sampling density is

(1) $k=1$, consider h_0 and set $\rho\in[0,1]$;
(2) Generate the population $\mathbf{X}_1,\cdots,\mathbf{X}_N$ according to the pdf $h(\cdot,\boldsymbol{\lambda}_{k-1})$ (or $h_0(\cdot)$ if $k=1$) and apply the function ϕ in order to have $Y_1=\phi(\mathbf{X}_1),\ldots,Y_N=\phi(\mathbf{X}_N)$;
(3) Compute the empirical ρ-quantile $q_k=\min(S,Y_{\lfloor\rho N\rfloor})$, where $\lfloor a\rfloor$ denotes the largest integer that is smaller than or equal to a;
(4) Optimize the parameters of the auxiliary pdf family as

$$\lambda_k=\underset{\lambda}{\mathrm{argmax}}\left\{\frac{1}{N}\sum_{i=1}^{N}\left[1_{\phi(\mathbf{X}_i)>q_k}\frac{h_0(\mathbf{X}_i)}{h(\mathbf{X}_i,\lambda_{k-1})}\ln[h(\mathbf{X}_i,\lambda)]\right]\right\};$$

(5) If $q_k<S$, $k\leftarrow k+1$, go to Step 2;
(6) Estimate the probability $\hat{P}^{\mathrm{CE}}(\phi(\mathbf{X}>S))=\frac{1}{N}\sum_{i=1}^{N}1_{\phi(\mathbf{X}_i)>S}\frac{h_0(\mathbf{X}_i)}{h(\mathbf{X}_i,\lambda_{k-1})}$.

In Step 1, ρ is a parameter which has to be set carefully. Typical value of ρ according to [34] can be chosen in [0.9, 0.99]. Choosing a high value of ρ reduces the number of CE iterations but weakens the quality of the final auxiliary pdf. Contrariwise, a relatively low value of ρ allows to better estimate $h(\cdot,\lambda_k)$ but implies a more important number of iterations. For more details about CE mechanisms, one can refer to [5].

2.2.3. Non parametric adaptive Importance Sampling

The objective of Non parametric adaptive Importance Sampling (NAIS) technique [7,8] is to approximate the IS optimal auxiliary density given in Eq. (6) with a kernel function (for example a Gaussian kernel function). NAIS does not require the choice of a pdf family and is thus more flexible than a parametric model. Its iterative principle is relatively similar to CE optimization and can be described by the following steps.

(1) $k = 1$ and set $\rho \in [0,1]$;
(2) Generate the population $\mathbf{X}_1^{(k)}, \ldots, \mathbf{X}_N^{(k)}$ according to the pdf h_{k-1}, apply the function ϕ in order to have $Y_1^{(k)} = \phi(\mathbf{X}_1^{(k)}), \ldots, Y_N^{(k)} = \phi(\mathbf{X}_N^{(k)})$;
(3) Compute the empirical ρ-quantile $q_k = \min(S, Y_{\lfloor \rho N \rfloor}^{(k)})$;
(4) Estimate $I_k = \frac{1}{kN} \sum_{j=1}^{k} \sum_{i=1}^{N} \mathbf{1}_{\phi(\mathbf{X}_i) \geq q_k} \frac{h_0(\mathbf{X}_i)}{h_{j-1}(\mathbf{X}_i)}$;
(5) Update the Gaussian kernel sampling pdf with

$$h_k(\mathbf{X}) = \frac{1}{kNI_k det(B_k)} \sum_{j=1}^{k} \sum_{i=1}^{N} w_j \left(\mathbf{X}_i^{(j)}\right) K_d\left(B_k^{-1} \times \left(\mathbf{X} - \mathbf{X}_i^{(j)}\right)\right), \quad (13)$$

where $K_d(\mathbf{X})$ is, for example, the standard d-dimensional Gaussian function with zero mean and covariance identity matrix, B_k is a diagonal covariance matrix $B_k = \mathrm{diag}(b_k^{1\,2}, \ldots, b_k^{d\,2})$ and $w_j = \mathbf{1}_{\phi(\mathbf{X}_i^{(j)}) \geq q_k} \frac{h_0(\mathbf{X}_i^{(j)})}{h_{j-1}(\mathbf{X}_i^{(j)})}$. For example, if $\phi : \mathbb{R}^d \to \mathbb{R}$, $K_d(B_k^{-1} \times (\mathbf{X} - \mathbf{X}_i^{(j)})) = \prod_{l=1}^{d} \frac{1}{\sqrt{2\pi} b_k^l} \exp(-\frac{(X_l - X_{li}^{(j)})^2}{2 b_k^{l\,2}})$.

The coefficients of matrix B_k are optimized according to the Asymptotic Mean Integrated Square Error (AMISE) criterion [9,35]. A valuable B_k matrix minimizes the mean integrated square error between the kernel density of the samples and the true density of the samples. Even if the true density of the samples is not available, it is possible to derive an asymptotic estimate of this error when $N \to \infty$.

(6) If $q_k < S$, $k \leftarrow k + 1$, go to Step 2;
(7) Estimate the probability $\hat{P}^{\mathrm{NAIS}}(\phi(\mathbf{X} > S)) = \frac{1}{N} \sum_{i=1}^{N} 1_{\phi(\mathbf{X}_i^{(k)}) > S} \frac{h_0(\mathbf{X}_i^{(k)})}{h_{k-1}(\mathbf{X}_i^{(k)})}$.

The use of Importance Sampling optimization with NAIS or CE often requires about 10,000 calls to ϕ. These algorithms are thus not implementable for very time-consuming simulation codes. We will show that the use of a surrogate model can greatly reduce this computational burden.

2.2.4. Metamodel-based rare event probability estimators

Being able to build an efficient surrogate model, which allows to reduce the number of calls to the expensive input–output function while keeping a good accuracy, is a key point in rare event probability estimation. A great number of methods have been proposed and compared in recent years. A survey of the different metamodel methods can be found in [36]. In this section, we present the main surrogate models that have already been combined with Importance Sampling or Monte Carlo estimators. Classical deterministic surrogate models such as polynomials, splines have been tested and compared to Neural Networks and First Order Reliability Method (FORM) [37–39]. Polynomial Chaos has been associated with Monte Carlo sampling to estimate failure probabilities [40]. Support Vector Machines have also been employed to estimate domains of failure [41] and coupled to a rare event estimator such as subset sampling [42].

Kriging method [43–45] presents some advantages in rare event probability estimation. Indeed, this surrogate model is based on a Gaussian process, which allows to estimate the variance of the prediction error and consequently to define a confidence domain of the surrogate model. This indicator can be directly used to refine the model, i.e., to choose new points where the real function should be evaluated so as to improve the accuracy of the model. Kriging has been extensively used with classical Monte Carlo estimator [4], Importance Sampling method [1,2,39] or Subset Simulation [46–48]. We will detail the Kriging theory in the next section. How to update the Kriging model is a key point and different strategies have been proposed [1,49,50] to exploit its probabilistic properties to evaluate a minimal number of points on the real expensive input–output function. A numerical comparison of different Kriging-based methods to estimate a probability of failure can be found in [51].

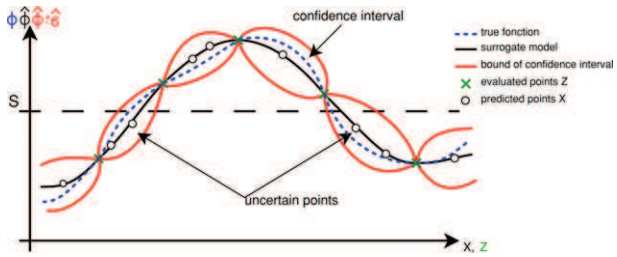


Fig. 1. Kriging model and corresponding confidence interval.

3. Adaptation of Kriging-based surrogate model to adaptive Importance Sampling

3.1. Kriging theory

Kriging [43,44,52] is a statistical surrogate model that can be used to approximate the input–output function ϕ on its input space $\mathbb{X} \subset \mathbb{R}^d$. It only requires a small initial Design Of Experiments (DOE) with p samples, $\mathcal{Z} = \{\mathbf{Z}_1, \ldots, \mathbf{Z}_p\}$, $\mathcal{Z} \in \mathbb{X}^p$, for which the function values have been computed and stored into $\boldsymbol{\phi}_p = [\phi(\mathbf{Z}_1), ..., \phi(\mathbf{Z}_p)]^{\mathrm{T}}$. The Kriging model consists of a Gaussian process $\Phi(\cdot)$ that is expressed, for any input vector $\mathbf{X} \in \mathbb{X}$, as

$$\Phi(\mathbf{X}) = m(\mathbf{X}) + \zeta(\mathbf{X}), \quad (14)$$

where the mean function $m(\cdot)$ is an optional regression model estimated from available data (polynomial in $\mathbf{X}$, for example) and $\zeta(\cdot)$ is a zero-mean Gaussian process with covariance function $\mathrm{cov}(\cdot, \cdot)$. Since the actual covariance function is unknown in practice, it can be modeled as

$$\mathrm{cov}(\zeta(\mathbf{Z}_i), \zeta(\mathbf{Z}_j)) = \sigma_\zeta^2 \mathrm{Corr}(\mathbf{Z}_i, \mathbf{Z}_j), \quad (15)$$

where σ_ζ^2 is the process variance and $\mathrm{Corr}(\cdot, \cdot)$ is a parametric correlation function. A classical choice for this correlation function is

$$\mathrm{Corr}(\mathbf{Z}_i, \mathbf{Z}_j) = \exp\left(-\sum_{l=1}^{d} \theta_l |Z_{il} - Z_{jl}|^{p_l}\right), \quad (16)$$

where the parameters $0 < p_l \leq 2$ reflect the smoothness of the interpolation (2 is the smoothest), while the θ_l are scale factors that can be estimated, e.g., by maximum likelihood [44].

Kriging provides an optimal unbiased linear predictor at any $\mathbf{X} \in \mathbb{X}$ as

$$\hat{\phi}(\mathbf{X}, \mathcal{Z}) = m(\mathbf{X}) + \mathbf{r}(\mathbf{X}, \mathcal{Z})^{\mathrm{T}} \mathbf{R}^{-1}(\mathcal{Z})(\boldsymbol{\phi}_p(\mathcal{Z}) - \mathbf{m}_p(\mathcal{Z})), \quad (17)$$

where

$$\begin{cases} \mathbf{R}_{|ij}(\mathcal{Z}) = \mathrm{Corr}(\mathbf{Z}_i, \mathbf{Z}_j) \\ \mathbf{r}(\mathbf{X}, \mathcal{Z}) = [\mathrm{Corr}(\mathbf{X}, \mathbf{Z}_1), \ldots, \mathrm{Corr}(\mathbf{X}, \mathbf{Z}_p)]^{\mathrm{T}} \\ \mathbf{m}_p(\mathcal{Z}) = [m(\mathbf{Z}_1), \ldots, m(\mathbf{Z}_p)]^{\mathrm{T}} \end{cases}. \quad (18)$$

Moreover, using Gaussian processes makes it possible to compute confidence intervals for the prediction (17) through the variance

$$\hat{\sigma}^2(\mathbf{X}, \mathcal{Z}) = \sigma_\zeta^2 \left(1 - \mathbf{r}(\mathbf{X}, \mathcal{Z})^{\mathrm{T}} \mathbf{R}^{-1}(\mathcal{Z}) \mathbf{r}(\mathbf{X}, \mathcal{Z})\right). \quad (19)$$

This ability to quantify the prediction uncertainty will be further exploited in Section 3.3 to find lower and upper bounds for the probability estimate. An illustration of Kriging interpolation and corresponding confidence interval is given in Fig. 1.

3.2. Adaptation of Kriging surrogate model to adaptive Importance Sampling techniques

3.2.1. Methodology

At the lth iteration, let us denote $\mathcal{Z} = \{\mathbf{Z}_1, ..., \mathbf{Z}_p\}$ the current DOE used to build the Kriging surrogate model, $\mathbb{X}^l$ the population generated by the auxiliary pdf, $\tilde{\mathbb{X}}^l$ the population of uncertain points such that

$$\tilde{\mathbb{X}}^l = \left\{\mathbf{X}_i \in \mathbb{X}^l | \mathbb{E}\left(\hat{\phi}\left(\mathbf{X}_i, \mathcal{Z}\right)\right) - k_l\, \hat{\sigma}\left(\mathbf{X}_i, \mathcal{Z}\right) < q_l < \mathbb{E}\left(\hat{\phi}\left(\mathbf{X}_i, \mathcal{Z}\right)\right) + k_l\, \hat{\sigma}\left(\mathbf{X}_i, \mathcal{Z}\right)\right\}, \quad (20)$$

with k_l a parameter determining the confidence level of the surrogate model (e.g., $k_l = 1.96$ corresponds to a confidence level of 95%). $\tilde{\mathbb{X}}^l$ are supposed to be the points that can be potentially misclassified (i.e., their estimations are above the threshold while the true values $\phi(\tilde{\mathbb{X}}^l)$ are below or *vice versa*). Consequently, in order to accurately compute the auxiliary density (intermediate iterations) or the estimation of the probability (final iteration), one has to check whether $\mathbb{E}(\phi(\tilde{\mathbb{X}}^l))$ exceeds the current threshold q_l (or S for the final iteration). This is achieved by adding new sampled points in the current DOE in order to improve the surrogate model. The strategy used to choose these points is detailed in the next section.

Since we use an adaptive Importance Sampling method, we have to define a metamodel not only for the threshold S but also for each of the intermediary quantiles. Defining a metamodel with using k_S at each iteration of the Importance Sampling is not a good strategy because it requires to compute on ϕ many points for defining very accurately intermediary thresholds which do not intervene directly in the probability calculation. In order to adapt the accuracy of the metamodel to the adaptive versions of Importance Sampling, we have chosen to define for each iteration, k as a function of the current ρ-quantile q and S the threshold that defines the probability to estimate: $P(\phi(\mathbb{X}) > S)$, so that at the lth iteration of the adaptive Importance Sampling method, we have:

$$k_l = k_S \frac{q_l}{S}. \quad (21)$$

This heuristics allows us to reduce at the minimum the number of points to evaluate on ϕ at the intermediary iterations and to have an accuracy of $k_S = 1.96$ (95%) for the probability estimation at the final iteration of the algorithm.

3.2.2. Sampling strategy

As in [1,49,50], we use the intrinsic properties of Kriging to define a refinement strategy that requires a minimum number of evaluation points to build the surrogate model.

The new point appended to the DOE at the kth iteration should be the one that best characterizes the uncertain population. Let $AN(\mathcal{Z}, \mathbf{z}, \tilde{\mathbb{X}}^k)$ be the criterion quantifying the improvement of the global uncertainty of $\tilde{\mathbb{X}}^k$ by appending $\mathbf{z}$ to the current DOE $\mathcal{Z}$. It holds

$$\mathbf{z}^* = \underset{\mathbf{z}\in\mathbb{R}^d}{\operatorname{argmax}}\left(AN\left(\mathcal{Z}, \mathbf{z}, \tilde{\mathbb{X}}^k\right)\right) = \underset{\mathbf{z}\in\mathbb{R}^d}{\operatorname{argmax}}\left\{\sum_{i=1}^{\mathrm{Card}\left(\tilde{\mathbb{X}}^k\right)}\left[\hat{\sigma}\left(\tilde{\mathbf{X}}_i, \mathcal{Z}\right) - \hat{\sigma}\left(\tilde{\mathbf{X}}_i, \mathcal{Z}, \mathbf{z}\right)\right]\right\}, \quad (22)$$

with

- $\hat{\sigma}\left(\tilde{\mathbf{X}}, \mathcal{Z}\right)$, the standard deviation of the prediction in $\tilde{\mathbf{X}}$ corresponding to the Kriging model built from the current DOE $\mathcal{Z}$,
- $\hat{\sigma}\left(\tilde{\mathbf{X}}, \mathcal{Z}, \mathbf{z}\right)$ the standard deviation of the prediction in $\tilde{\mathbf{X}}$ corresponding to the Kriging model built from the extended DOE $\{\mathcal{Z}, \mathbf{z}\}$,
- $\mathrm{Card}(\tilde{\mathbb{X}}^k)$, the number of points present in the uncertain population $\tilde{\mathbb{X}}^k$.

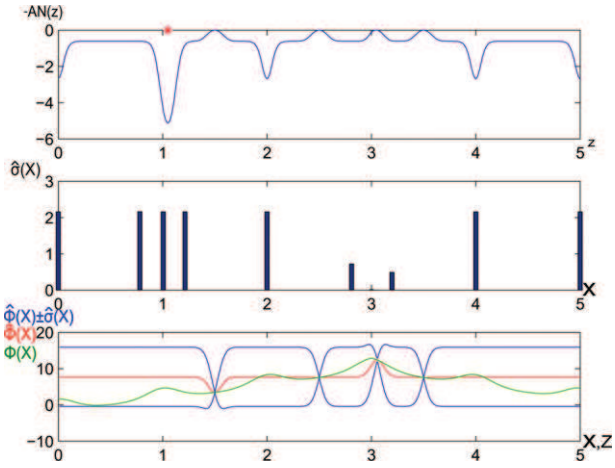


Fig. 2. Top: criterion *AN* (opposite of the), center: relative standard deviation of each point, and bottom: surrogate model.

3.2.2.1. Determination of $\hat{\sigma}\left(\tilde{\mathbf{X}}, \mathcal{Z}\right)$ and $\hat{\sigma}\left(\tilde{\mathbf{X}}, \mathcal{Z}, \mathbf{z}\right)$

From Section 3.1, we have

$$\mathbf{r}(\mathbf{X}, \mathcal{Z}) = [\mathrm{Corr}(\mathbf{X}, \mathbf{Z}_1), \ldots, \mathrm{Corr}(\mathbf{X}, \mathbf{Z}_p)]^{\mathrm{T}}, \quad (23)$$

$$\mathbf{R}(\mathcal{Z}) = \begin{bmatrix} \mathrm{Corr}(\mathbf{Z}_1, \mathbf{Z}_1) & \cdots & \mathrm{Corr}(\mathbf{Z}_p, \mathbf{Z}_1) \\ \vdots & \ddots & \vdots \\ \mathrm{Corr}(\mathbf{Z}_p, \mathbf{Z}_1) & \cdots & \mathrm{Corr}(\mathbf{Z}_p, \mathbf{Z}_p) \end{bmatrix}, \quad (24)$$

and we also recall that the prediction variance (19) is

$$\hat{\sigma}\left(\tilde{\mathbf{X}}_i, \mathcal{Z}\right)^2 = \hat{\sigma}_\zeta^2\left[1 - \mathbf{r}\left(\tilde{\mathbf{X}}_i, \mathcal{Z}\right)^{\mathrm{T}} \mathbf{R}^{-1}(\mathcal{Z})\, \mathbf{r}\left(\tilde{\mathbf{X}}_i, \mathcal{Z}\right)\right], \quad (25)$$

with the process variance that can be estimated by maximum-likelihood as

$$\hat{\sigma}_\zeta^2 = \frac{(\boldsymbol{\phi}_p(\mathcal{Z}) - \mathbf{m}_p(\mathcal{Z}))^{\mathrm{T}} \mathbf{R}^{-1}(\mathcal{Z})(\boldsymbol{\phi}_p(\mathcal{Z}) - \mathbf{m}_p(\mathcal{Z}))}{\mathrm{Card}(\mathcal{Z})}. \quad (26)$$

Similarly, $\hat{\sigma}\left(\tilde{\mathbf{X}}_i, \mathcal{Z}, \mathbf{z}\right)$ can be expressed as follows

$$\hat{\sigma}\left(\tilde{\mathbf{X}}_i, \mathcal{Z}, \mathbf{z}\right)^2 = \hat{\sigma}_{\zeta'}^2\left[1 - \mathbf{r}'\left(\tilde{\mathbf{X}}_i, \mathcal{Z}, \mathbf{z}\right)^{\mathrm{T}} \mathbf{R}'^{-1}(\mathcal{Z}, \mathbf{z})\, \mathbf{r}'\left(\tilde{\mathbf{X}}_i, \mathcal{Z}, \mathbf{z}\right)\right], \quad (27)$$

with

$$\mathbf{r}'(\mathbf{X}, \mathcal{Z}, \mathbf{z}) = \left[\mathbf{r}(\mathbf{X}, \mathcal{Z})^{\mathrm{T}}\ \mathrm{Corr}(\mathbf{X}, \mathbf{z})\right]^{\mathrm{T}}, \quad (28)$$

$$\mathbf{R}'(\mathcal{Z}, \mathbf{z}) = \begin{bmatrix} \mathrm{Corr}(\mathbf{Z}_1, \mathbf{Z}_1) & \cdots & \mathrm{Corr}(\mathbf{Z}_p, \mathbf{Z}_1) & \mathrm{Corr}(\mathbf{Z}_p, \mathbf{z}) \\ \vdots & \ddots & \vdots & \vdots \\ \mathrm{Corr}(\mathbf{Z}_p, \mathbf{Z}_1) & \cdots & \mathrm{Corr}(\mathbf{Z}_p, \mathbf{Z}_p) & \mathrm{Corr}(\mathbf{Z}_p, \mathbf{z}) \\ \mathrm{Corr}(\mathbf{z}, \mathbf{Z}_1) & \cdots & \cdots & \mathrm{Corr}(\mathbf{z}, \mathbf{z}) \end{bmatrix}. \quad (29)$$

Since $\phi(\mathbf{z})$ is unknown, we propose to approximate $\hat{\sigma}_{\zeta'}$ by $\hat{\sigma}_\zeta$, which is the estimated process standard deviation for the DOE $\mathcal{Z}$. The error induced by this approximation is expected to be small when $\mathrm{Card}(\mathcal{Z})$ is large enough. This standard deviation may also be assumed to be known and kept constant [53].

The criterion *AN* allows us to determine the points with the biggest influence on the total standard deviation of the uncertain set. CMA-ES [54] is used to optimize *AN*, since it may be non linear and present local optima (Fig. 2). This auxiliary optimization strategy has been exploited in [2,49] in a similar context. In order to get the Kriging parameters for the DOE $\mathcal{Z}$, we use the DACE toolbox [55].

The sampling process is included in the initial adaptive Importance Sampling algorithms (Step 2 of the CE and adaptive NAIS algorithms):

2-1 Build a Kriging model from the current DOE $\mathcal{Z}$,

Table 1
Results obtained for $l = 6$.

	$\mathbb{E}(P)$	$\sigma(P)/\mathbb{E}(P)(\%)$	N1	N LHS	N2	$\mathbb{E}(\Gamma)$
CE with Kriging	4.4×10^{-3}	24	600	20	36	1.1×10^{-4}
NAIS with Kriging	4.4×10^{-3}	21	748	20	122	2.1×10^{-5}
AK-MCS	4.4×10^{-3}	?	10^6		126	
MC REF	4.4×10^{-3}		10^6			
Equivalent MC – 122 points	3.9×10^{-3}	149	122		122	
CE without Kriging	4.4×10^{-3}	24	601		601	
NAIS without Kriging	4.3×10^{-3}	19	750		750	

Table 2
Results obtained for $l = 7$.

	$\mathbb{E}(P)$	$\sigma(P)/\mathbb{E}(P)(\%)$	N1	N LHS	N2	$\mathbb{E}(\Gamma)$
CE with Kriging	2.2×10^{-3}	28.2	1200	20	48	3.0×10^{-5}
NAIS with Kriging	2.1×10^{-3}	17	893	20	127	1.2×10^{-5}
AK-MCS	2.2×10^{-3}	?	10^6		96	
MC REF	2.2×10^{-3}		10^6		10^6	
Equivalent MC – 127 points	2.8×10^{-3}	174	127		127	
CE without Kriging	2.2×10^{-3}	24	1200		1200	
NAIS without Kriging	2.2×10^{-3}	18	887		887	

Table 3
Results obtained for the Rastrigin function.

	$\mathbb{E}(P)$	$\sigma(P)/\mathbb{E}(P)(\%)$	N1	N LHS	N2	$\mathbb{E}(\Gamma)$
CE with Kriging	7.4×10^{-2}	21	903	20	188	5.4×10^{-4}
NAIS with Kriging	7.4×10^{-2}	26	784	20	200	6.3×10^{-4}
AK-MCS	7.4×10^{-2}	?	6×10^4		416	
MC REF	7.3×10^{-2}		10^6		10^6	
Equivalent MC – 200 points	7.3×10^{-2}	27	200		200	
CE without Kriging	7.4×10^{-2}	19	903		903	
NAIS without Kriging	7.3×10^{-2}	24	792		792	

2-2 Predict $\hat{\phi}(\mathbf{X}_1, \mathcal{Z}), \ldots, \hat{\phi}(\mathbf{X}_N, \mathcal{Z})$ using the surrogate model, compute the confidence domain and determine $\tilde{\mathbb{X}}$,
2-3 While $\mathrm{Card}(\tilde{\mathbb{X}}) \neq 0$
  2-3-1 Optimize the criterion *AN* with using CMA-ES to find $\mathbf{z}^*$,
  2-3-2 Evaluate $\phi(\mathbf{z}^*)$ and add it to the current DOE, $N \leftarrow N + 1$
  2-3-3 Reestimate the parameters of the Kriging model from the new DOE,
  2-3-4 Predict $\hat{\phi}(\mathbf{X}_1, \mathcal{Z}), \ldots, \hat{\phi}(\mathbf{X}_N, \mathcal{Z})$ again, compute the confidence domain and compute $\tilde{\mathbb{X}}$.

An illustration of the *AN* criterion is displayed in Fig. 2 for a given uncertain set of points. The top figure represents the opposite of *AN*, which is used during the optimization, and the star indicates the location of the optimal point. In the central figure, the different bars represent the standard deviation of the estimated points. Finally, in the bottom figure, the green curve stands for the true (but unknown) function, the red curve stands for the Kriged function and the blue curve stands for the confidence interval. We can see here that this mono-dimensional example has a current DOE of four points: {1.5, 2.5, 3.1, 3.5} and nine points have been generated using the current pdf: {0, 0.8, 1, 1.2, 2, 2.8, 3.2, 4, 5}. The criterion favors sampling in areas with a high density of uncertain points. In the bottom figure, the ϕ function is extrapolated below $x = 1.5$ and above $x = 3.5$. The "flat" behavior of the kriged estimate in these domains is due to the chosen regression model m as defined in Eq. (14) which has been taken as zero order polynomial in this case.

3.3. Estimation of relative error induced by the use of the surrogate model

Once the Importance Sampling algorithms have converged, all the points stay at a distance from the threshold equal to at least $k\sigma$. Consequently, we can bound the error generated by the use of the surrogate model for the probability estimation to $1 - F^{-1}(k)$, where F is the cumulative distribution function of the reduced centered normal law (which pdf is defined by $p_{\mathcal{N}}(x) = \frac{1}{\sqrt{2\pi}} \exp(-\frac{x^2}{2})$). More precisely, let d_j be the distance between the estimation $\hat{\phi}(\mathbf{X}_j)$ and the threshold S. We define:

$$\alpha_j = \frac{d_j}{\hat{\sigma}(\mathbf{X}_j, \mathcal{Z})}, \tag{30}$$

the ratio between the standard deviation of the predicted point $\mathbf{X}_j$ and its distance from the threshold S. Let $P_j = 1 - F^{-1}(\alpha_j)$ the probability of misclassification, i.e., the probability that the prediction is above the threshold while the true value is underneath (and *vice versa*). Uncertainty induced by the Kriging model on the estimated probability is

$$\hat{P}_{\mathrm{Krig}} = \frac{1}{M} \sum_{w=1}^{M} \left(\frac{1}{N} \sum_{j=1}^{N} \Delta(\mathbf{X}_j) \frac{h_0(\mathbf{X}_j)}{h_{k-1}(\mathbf{X}_j)} \right), \tag{31}$$

with $\Delta(\mathbf{X}_j)$ a coefficient that takes the value 1 with a probability P_j and 0 with probability $1 - P_j$ if $\hat{\phi}(\mathbf{X}_j) > S$, and M the number of runs used to compute the Monte Carlo estimate $\hat{P}_{\mathrm{Krig}}$. Similarly, it takes the

value 0 with probability P_j and 1 with probability $1-P_j$ if $\hat{\phi}(\mathbf{X}_j) < S$. Its expression is

$$\Delta(\mathbf{X}_j) = \overline{XOR\left(\mathcal{B}(P_j, 1-P_j), 1_{\hat{\phi}(\mathbf{X}_j)>S}\right)}, \tag{32}$$

where $\mathcal{B}(P_j, 1-P_j)$ is the Bernoulli distribution indexed on P_j. Finally, the relative part of the IS estimator error due to the Kriging model is

$$\Gamma = \sigma\left(\hat{P}_{\text{Krig}}\right), \tag{33}$$

with σ the standard deviation of the estimator defined in Eq. (31). This estimator allows us to define a 95% confidence interval for the Kriged probability estimation using

$$[P_{\text{IS}} - 1.96\Gamma,\ P_{\text{IS}} + 1.96\Gamma]. \tag{34}$$

This interval, defined with Γ, has to be distinguished with the total confidence interval of the IS probability estimate. Indeed, Γ defines only the IS estimator error due to the Kriging model and does not take into account the inherent error of the IS estimator.

The probability and the confidence interval on the IS estimator error due to the Kriging model for 10 estimations of the IS algorithm is given for the aerospatial test case in Fig. 8.

4. Results

Before applying the Kriging-based adaptations of the Cross-Entropy and adaptive NAIS algorithms to an aerospace industrial case, we propose to test the new algorithms on several academic cases proposed in [4,39,56,57]. These cases have been proposed in order to evaluate the efficiency of different surrogate-based techniques coupled with classical Monte Carlo and non adaptive Importance Sampling methods. In particular, we compare the proposed algorithms to AK-MCS proposed in [4]. Note that the comparison with AK-MCS in terms of standard deviation of the estimator was not possible due to the lack of the standard deviation value in [4].

4.1. Benchmark of analytical test cases

The different test cases of the benchmark have been transformed with respect to their original formulations in order to apply the proposed adaptive methods. To this end, the analytical problems have been modified for threshold-exceeding probability estimation. For all the results presented here, $N1$ stands for the total number of points used by estimators, N_{LHS} is the number of points used in the initial sampling by LHS (Latin Hypercube Sampling), and $N2$ is the number of points that are actually evaluated using the true function ϕ_i. A hundred estimations have been generated for each of the cases in order to compute the standard deviation of the estimators. An expensive Monte Carlo estimation has been achieved in order to find the reference probability. Furthermore, for comparison sake, a second Monte Carlo, with a simulation budget equivalent to the maximal number of points evaluated by Kriging CE or ANAIS, has also been performed. All the random variables present in the test cases are considered as uncorrelated.

4.1.1. Four-branch system

The first test case consists in a four-branch system. The input space is $\mathbb{X} = \mathbb{R}^2$. The input variables are distributed according to a normal distribution $\mathcal{N}(0_2, I_2)$.

$$S_1 = 10, \tag{35}$$

$$\mathbf{X} \sim \mathcal{N}(0_2, I_2), \tag{36}$$

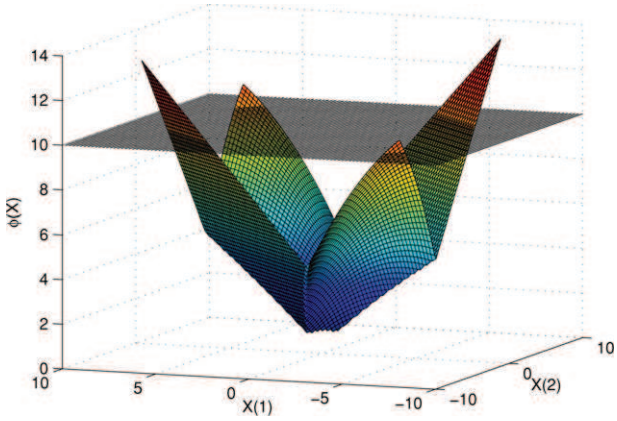


Fig. 3. Four-branch system.

$$\phi_1(X_1, X_2) = \min\begin{cases} 7 - 0.1(X_1 - X_2)^2 + \dfrac{X_1 + X_2}{\sqrt{2}} \\ 7 - 0.1(X_1 - X_2)^2 - \dfrac{X_1 + X_2}{\sqrt{2}} \\ 10 - (X_1 - X_2) - \dfrac{1}{\sqrt{2}} \\ 10 - (X_2 - X_1) - \dfrac{1}{\sqrt{2}} \end{cases}. \tag{37}$$

In this formulation, l is a parameter that can be equal to 6 or 7. The function ϕ_1 and the threshold S_1 are given in Fig. 3. The results obtained for this test case are given in Tables 1 and 2. An example of sampling function evolution for NAIS algorithm is described in Figs. 4 and 5. These figures show the adaptation of the sampling distribution in order to generate samples in the threshold exceedance area. The number of samples per iteration is equal to 200 for CE and 250 for NAIS. Concerning CE, the average number of required iterations is equal to 3 for $l=6$ and 6 for $l=7$. Concerning NAIS, the average number of required iterations is equal to 3 for $l=6$ and 3.5 for $l=7$.

For these two test cases, we can see that the adaptive Cross-Entropy method coupled with Kriging metamodel allows to divide by 2 and 3 the computational cost compared to AK-MCS. Moreover, the value of $\mathbb{E}(\Gamma)$ is very low compared to the probability to estimate, that indicates a good confidence in the estimation. This is confirmed by comparing the probability found by CE and NAIS to the true probability found by MC with 10^6 simulations. The coefficient of variation of CE is relatively higher than NAIS, that can be explained by the choice of parametrized pdf family (here two-dimensional normal density function) that is less appropriate to multimodal regions (Fig. 5). Nevertheless, this coefficient is still relatively low.

4.1.2. Rastrigin function

The second test case is the Rastrigin function, which is often used to test global optimization algorithms, due to its highly multimodal characteristics. The input space is $\mathbb{X} = \mathbb{R}^2$ and the input variables are distributed according to a normal distribution $\mathcal{N}(0_2, I_2)$. The function ϕ_2 and the threshold S_2 are displayed in Fig. 6. The results obtained for this test case are given in Table 3.

$$S_2 = 20, \tag{38}$$

$$\mathbf{X} \sim \mathcal{N}(0_2, I_2), \tag{39}$$

$$\phi_2(X_1, X_2) = \sum_{i=1}^{2}\left(X_i^2 - 5\cos(2\pi X_i)\right). \tag{40}$$

For this test case again, we can see that both adaptive CE and NAIS coupled with Kriging metamodel allow to divide by 2 the computational cost compared to AK-MCS. The values of $\mathbb{E}(\Gamma)$ are higher than

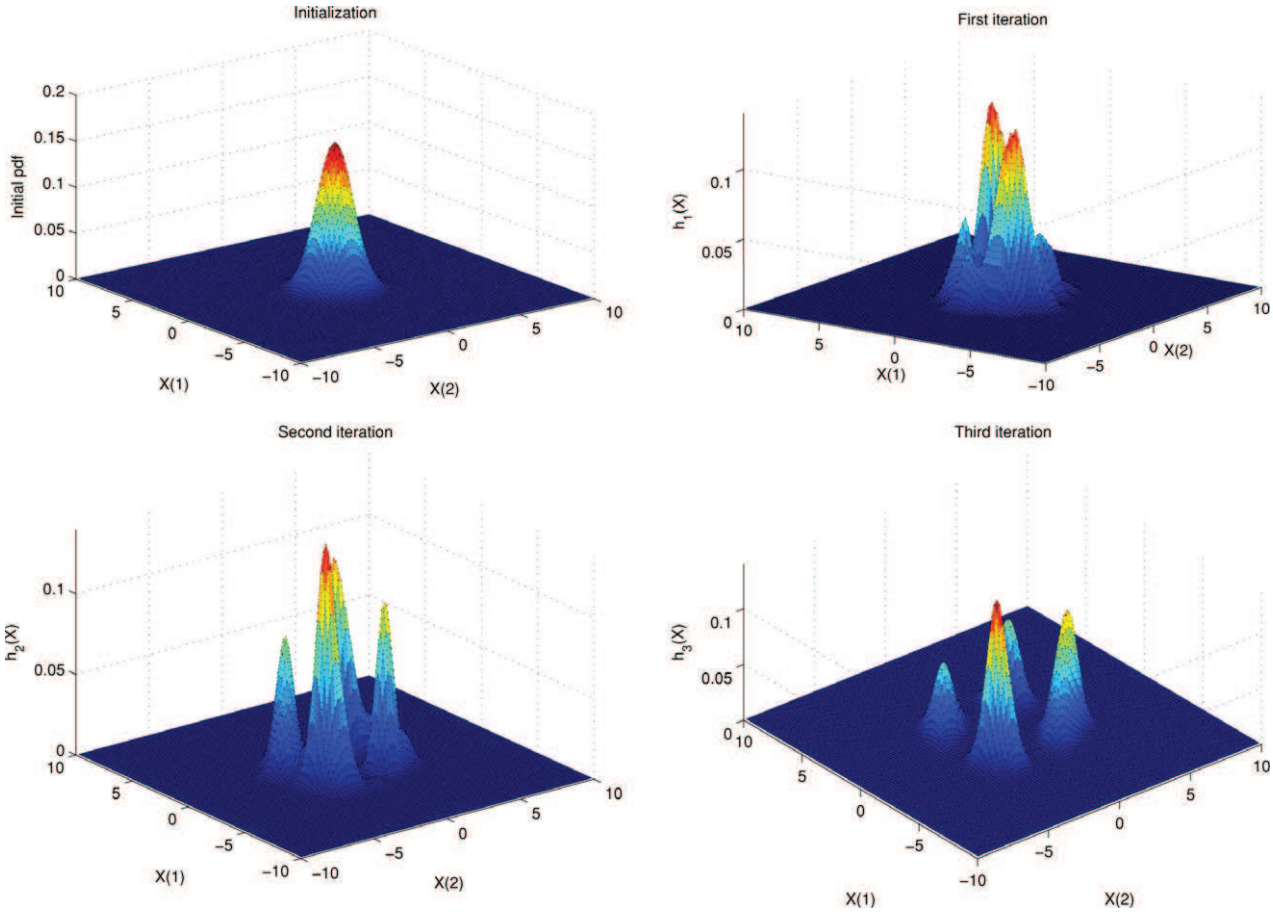


Fig. 4. Sampling function evolution (NAIS).

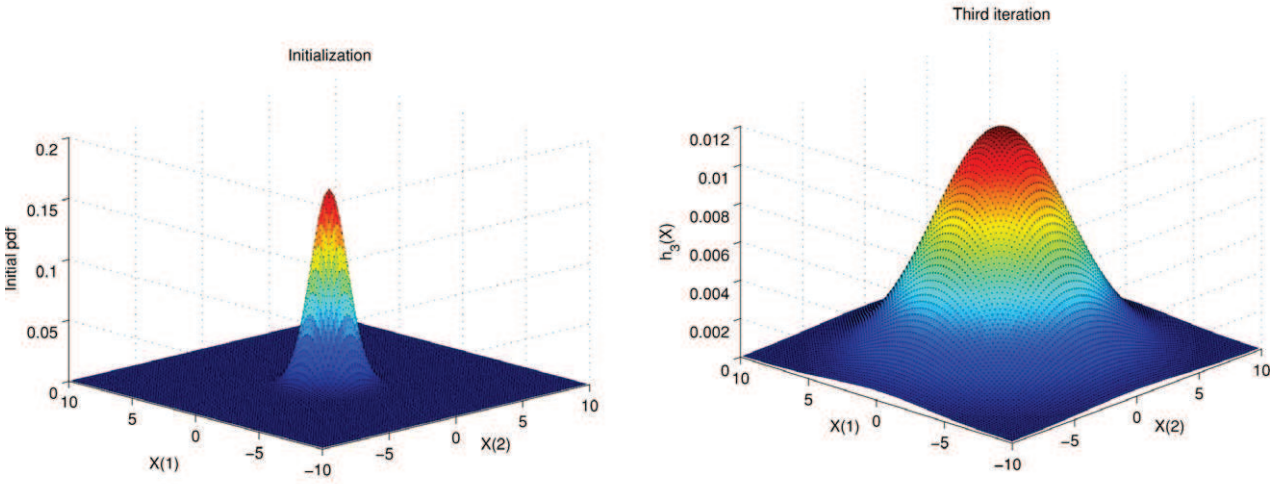


Fig. 5. Sampling function evolution (CE).

in the first case, but the probability found by CE and NAIS is again very close to the true probability found by MC and is equal to the one found by AK-MCS. Concerning this test case, the average number of required iterations is equal to 3 for CE and 2 for NAIS. The number of samples per iteration is equal to 300 for CE and 400 for NAIS.

4.1.3. Nonlinear oscillator

4.1.3.1. Description The last academic test case is a one-degree-of-freedom nonlinear oscillator. This example has been used in [38,58,59]. The input space is $\mathbb{X} = \mathbb{R}^6$ and the different variables are distributed according to 6 Gaussian distributions with different means and standard deviations:

$$S_3 = 10\,, \tag{41}$$

$$\mathbf{X} = [c_1, c_2, r, m, t_1, F_1]^{\mathrm{T}}\ [\mathcal{N}(1, 0.1),\ \mathcal{N}(0.1, 0.01),\ \mathcal{N}(1, 0.05),\ \mathcal{N}(0.5, 0.05),\ \mathcal{N}(1, 0.2),\ \mathcal{N}(1, 0.2)]^{\mathrm{T}}\,, \tag{42}$$

Table 4
Results obtained for the nonlinear oscillator function.

	$\mathbb{E}(P)$	$\sigma(P)/\mathbb{E}(P)(\%)$	N1	N LHS	N2	$\mathbb{E}(\Gamma)$
CE with Kriging	2.8×10^{-2}	21	1200	60	55	1.6×10^{-4}
NAIS with Kriging	2.9×10^{-2}	30	339	60	92	2.5×10^{-4}
AK-MCS	2.8×10^{-2}	?	7×10^4		58	
MC REF	2.8×10^{-2}		10^6		10^6	
Equivalent MC – 92 points	2.7×10^{-2}	65	92		92	
CE without Kriging	2.8×10^{-2}	21	1200		1200	
NAIS without Kriging	2.9×10^{-2}	29	302		302	

Table 5
Results obtained for the booster fallout zone estimation with $S = 650$ m.

$S = 650$ m	$\mathbb{E}(P)$	$\sigma(P)/\mathbb{E}(P)(\%)$	N1	N LHS	N2	$\mathbb{E}(\Gamma)$
CE with Kriging	1.9×10^{-5}	18	6000	40	192	1.8×10^{-7}
NAIS with Kriging	2.0×10^{-5}	14	5080	40	466	1.0×10^{-7}
MC REF	1.9×10^{-5}		1e8		1e8	
Equivalent MC – 466 points	Not accessible	Not defined	466		466	
CE without Kriging	1.9×10^{-5}	17	6000		6000	
NAIS without Kriging	1.9×10^{-5}	14	5100		5100	

Table 6
Results obtained for the booster fallout zone estimation with $S = 720$ m.

$S = 720$ m	$\mathbb{E}(P)$	$\sigma(P)/\mathbb{E}(P)(\%)$	N1	N LHS	N2	$\mathbb{E}(\Gamma)$
CE with Kriging	2.1×10^{-7}	28	7500	40	248	1.7×10^{-9}
NAIS with Kriging	2.1×10^{-7}	13	9000	40	500	1×10^{-9}
MC REF	2.1×10^{-7}		1e9		1e9	
Equivalent MC – 500 points	Not accessible	Not defined	500		500	
CE without Kriging	2.1×10^{-7}	27	7500		7500	
NAIS without Kriging	2.1×10^{-7}	14	8980		8980	

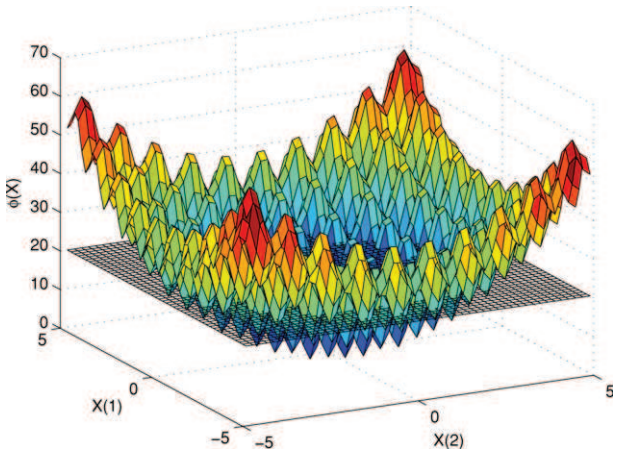


Fig. 6. Rastrigin function.

$$\phi_2(c_1, c_2, m, r, t_1, F_1) = 10 - 3r + \left| \frac{2F_1}{c_1 + c_2} \sin\left(\sqrt{\frac{c_1 + c_2}{m}} \frac{t_1}{2} \right) \right|. \quad (43)$$

The results obtained for this test case are given in Table 4.

For this test case, the results of adaptive CE coupled with Kriging metamodel are very close to the ones obtained with AK-MCS. Moreover, the average number of required iterations (respectively number of samples per iteration) is equal to 4 for CE (respectively 300) and 2 for NAIS (respectively 150).

4.1.4. Synthesis

The adaptive version of Importance Sampling globally allows to estimate the probability of rare event with less points than AK-MCS. Unfortunately, the performed comparisons with AK-MCS have been performed with incomplete information because the coefficient of variation of AK-MCS is not provided is the reference paper [4].

For each of the test cases, the empiric mean of the estimated probability is very close to the "true" probability found by extensive Monte Carlo. The value of the estimator $\mathbb{E}(\Gamma)$ is always much lower than the probability, which provides a good level of confidence to the obtained results. This comparison points out the advantages of Kriging-based adaptive Importance Sampling methods to estimate rare event probability with respect to AK-MCS, traditional IS and MCS method.

Simulations with using only CE and NAIS (without Kriging) have also been performed. It can therefore be seen that the proposed approach requires significantly less sampled points than the standard algorithms while maintaining the same performance level. In the next section, we analyze the efficiency of the proposed strategies on a real-world aerospace application case.

4.2. Estimation of launch vehicle booster fallout zone

4.2.1. Presentation of the case study

Spatial launch vehicle fall-back safety zone estimation is a very important problem in space application since the consequence of a mistake can be dramatic for the populations. We consider in this article a solid rocket booster that is the first stage of a launch vehicle. Its mass is about 35,000 kg and the launch point is at 112 km height with a slope of 15°. At the end of its mission, the rocket booster falls into the sea at some distance of a predicted position. Similar models

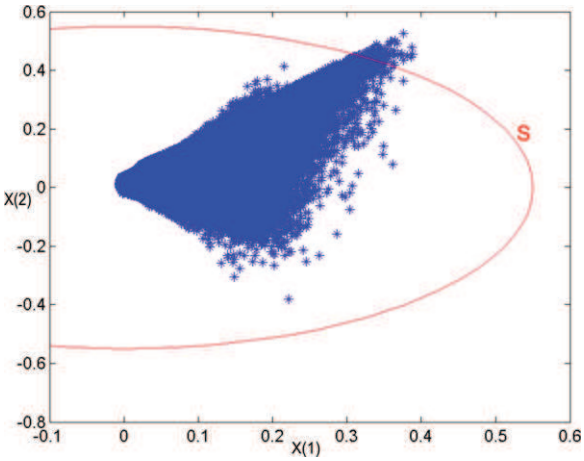


Fig. 7. Booster fallout zone and threshold.

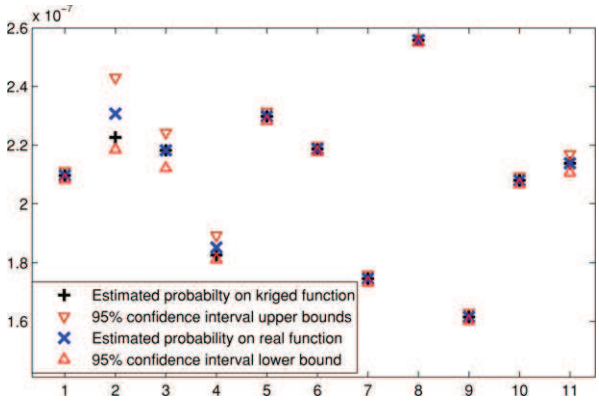


Fig. 8. Estimated probability on Kriging model and real function, and confidence interval of approximation error due to the use of Kriging model.

have already been analyzed in [60].

The launch vehicle stage fall-back is thus modeled as an input–output function ϕ with a vector X of four Gaussian inputs and one output $Y = \phi(X)$, representing the distance between the estimated launch stage fall-back position and the predicted one. In this case study, we aim at estimating the probability that the distance to the predicted impact position exceeds 0.65 km: $P_1(\phi(X) > 0.65)$ and 0.72 km: $P_2(\phi(X) > 0.72)$. Several components of X can thus influence its impact position:

- Meteorological conditions (two inputs). The wind variations during the fall-back can influence the impact position.
- Launch vehicle mass (one input). The mass of the different parts of the launch vehicle is also slightly random during the fall-back.
- The slope angle between the vertical axis and the velocity vector (one input) (Fig. 7).

4.2.2. Results

The results obtained for the two series estimations are given in Tables 5 and 6. A hundred estimations have been generated for each of the cases in order to compute the standard deviation of the estimators.

For each case, the estimated probabilities are very close to the true ones, obtained via Monte Carlo with a very long computation time. The Kriging-based versions of the algorithms allow to divide the computation cost of CE by 26 for the two cases and the computation cost of NAIS by 10 for the first case and 16 for the second one. This way, the Kriging-based adaptive versions of CE and NAIS allow to greatly improve the efficiency of these algorithms and make it possible to apply this algorithm with very restricted simulation budget (a few hundred samples). The equivalent MC estimations with analog bugdets as Kriging-based versions of CE and ANAIS have not converged because the simulation budget is too restricted to sample a point above the thresholds with the initial pdf.

The number of samples per iteration is equal to 1500 for CE and 1000 for NAIS. Concerning CE, the average number of required iterations is equal to 4 for $S = 650$ m and 5 for $S = 720$ m. Concerning NAIS, the average number of required iterations is equal to 5 for $S = 650$ m and 9 for $S = 720$ m.

Fig. 8 shows an example of the estimator behavior for 11 estimations of the second test case probability using adaptive CE method. For nine of the 11 estimations, the probability found on the Kriging-based model is equal to the probability found while evaluating the real function. Furthermore, for the two other cases, the estimated probability is always located in the 95% confidence interval defined by $\mathbb{E}(\Gamma)$. This illustrates the consistency of the proposed confidence interval of approximation error due to the use of Kriging model.

5. Conclusions

A new method combining Kriging surrogate models with adaptive Importance Sampling algorithms has been proposed for estimating the probability of rare events. The proposed scheme requires very few samples, making this method suitable for time-consuming simulation codes. Moreover, the statistical properties of Kriging have permitted the definition of a confidence interval for the probability estimate, which is a noticeable feature. The performance of the Kriging-based adaptive Importance Sampling approach has been illustrated and compared on analytical test-cases from the literature. A realistic aeronautical example has also been presented to show the feasibility and accuracy of the method. These promising results should now be confirmed on complex, high-dimensional problems that cannot be addressed previously due to computational cost.

Acknowledgments

The authors are thankful to the anonymous reviewers for their valuable remarks, and to Dr. Rudy Pastel (Onera), Dr. Hélène Piet-Lahanier (Onera) and Dr. François Le Gland (INRIA) for fruitful discussions. The authors thank for partial funding from the Ministère de l'Economie, de l'Industrie et de l'Emploi (Minister for the Economy, Industry and Employment) in the frame of CSDL (Complex System Design Lab) project of French competitiveness cluster System@tic – Paris Région.

References

[1] Dubourg V, Deheeger E, Sudret B. Metamodel-based importance sampling for the simulation of rare events. In: Faber M, Kohler J, Nishilima K, editors. Proceedings of the 11th International Conference of Statistics and Probability in Civil Engineering (ICASP2011), Zurich, Switzerland.
[2] Janusevskis J, Le Riche R. Simultaneous Kriging-based estimation and optimization of mean response. J Global Optim 2013;55:313–36.
[3] Beichl I, Sullivan F. The importance of importance sampling. Comput Sci Eng 1999;1:71–3.
[4] Echard B, Gayton N, Lemaire M. AK-MCS: an active learning reliability method combining Kriging and Monte Carlo Simulation. Struct Saf 2011;33:145–54.
[5] Rubinstein R, Kroese D. The Cross-Entropy method: a unified approach to combinatorial optimization, Monte-Carlo simulation and machine learning (information science and statistics). Springer; 2004.
[6] Homem-de Mello T. A study on the cross-entropy method for rare event probability estimation. INFORMS J Comput 2007;19:381–94.
[7] Zhang P. Nonparametric importance sampling. J Am Stat Assoc 1996;91:1245–53.
[8] Morio J. Extreme quantile estimation with nonparametric adaptive importance sampling. Simul Model Pract Theory 2012;27:76–89.

[9] Silverman B. Density estimation for statistics and data analysis. London: Chapman and Hall; 1986.
[10] Mikhailov GA. Parametric estimates by the Monte Carlo method. Utrecht (NED): VSP; 1999.
[11] Sobol IM. A primer for the Monte Carlo method. Boca Raton, Fl: CRC Press; 1994.
[12] Robert CP, Casella G. Monte Carlo statistical methods. New York: Springer; 2005.
[13] Niederreiter H, Spanier J. Monte Carlo and quasi-Monte Carlo methods. Springer; 2000.
[14] Borcherds PH. Importance sampling: an illustrative introduction. Eur J Phys 2000:405–11.
[15] Denny M. Introduction to importance sampling in rare-event simulations. Eur J Phys 2001:403–11.
[16] Botev ZI, Kroese D, Taimre T. Generalized cross-entropy methods with applications to rare-event simulation and optimization. Simulation 2007;11:785–806.
[17] Rubinstein R. Optimization of computer simulation models with rare events. Eur J Oper Res 1997;99:89–112.
[18] Glynn P. Importance sampling for Monte Carlo estimation of quantiles. Publishing House of Saint Petersburg University; 1996, pp. 180–5.
[19] Morio J. How to approach the importance sampling density. Eur J Phys 2010;31:41–8.
[20] Glasserman P, Heidelberger P, Shahabuddin P, Zajic T. Splitting for rare event simulation: analysis of simple cases. In: Proceeding of the 1996 Winter Simulation Conference. p. 302–08.
[21] L'Écuyer P, Demers V, Tuffin B. Splitting for rare event simulation. In: Proceeding of the 2006 Winter Simulation Conference. p. 137–48.
[22] L'Écuyer P, Le Gland F, Lezaud P, B Tuffin. Splitting methods. In: G Rubino, B Tuffin (Eds.), Monte Carlo methods for rare event analysis. Chichester: John Wiley & Sons; 2009, pp. 39–61.
[23] Cérou F, Guyader A. Adaptive particle techniques and rare event estimation. ESAIM 2007;19:65–72.
[24] Cérou F, Del Moral P, Furon T, Guyader A. Rare event simulation for a static distribution. RESIM 2008;7:107–15.
[25] Cérou F, Del Moral P, Furon T, Guyader A. Sequential Monte Carlo for rare event estimation. Stat Comput 2011:1–14.
[26] Morio J, Pastel R, Le Gland F. An overview of importance splitting for rare event simulation. Eur J Phys 2010;31:1295–303.
[27] Gilli M, Këllezi E. An application of extreme value theory for measuring risk. Comput Econ 2006;27:207–28.
[28] Piera-Martinez M, Vazquez E, Walter E, Fleury G, Kielbasa R. Estimation of extreme values, with application to uncertain systems. In: Proceedings of the 14th IFAC Symposium on System Identification. 2006.
[29] Davison A, Smith R. Models for exceedances over high thresholds (with discussion). J Roy Stat Soc 1990;52:393–442.
[30] McNeil A, Saladin T. The peaks over threshold method for estimating high quantiles of loss distributions. In: Proceedings of the 2nd International ASTIN Colloquium. 1997.
[31] Pickands J. Statistical inference using extreme order statistics. Ann Stat 1975;3:119–31.
[32] Coles S. An introduction to statistical modeling of extreme values. New York: Springer; 2001.
[33] Bucklew JA. Introduction to rare event simulation. Springer; 2004.
[34] Boer PTD, Kroese D, Mannor S, Rubinstein R. A tutorial on the cross-entropy method. Ann Oper Res 2002;134.
[35] Glad IK, Hjort NL, Ushakov NG. Mean-squared error of kernel estimators for finite values of the sample size. J Math Sci 2007;146:5977–83.
[36] Sudret B. Meta-models for structural reliability and uncertainty quantification. In: Asian-Pacific symposium on structural reliability and its applications Singapore: Singapore; 2012.
[37] Bucher C, Bourgund U. A fast and efficientresponse surface approach for structural reliability problems. Struct Saf 1990;7:57–66.
[38] Gomes H, Awruch A. Comparison of response surface and neural network with other methods for structural reliability analysis. Struct Saf 2004;26:49–67.
[39] Schueremans L, Van Gemert D. Use of Kriging as meta-model in simulation procedures for structural reliability. In: 9th international conference on structural safety and reliability, Rome; 2005. p. 2483–90.
[40] Li J, Xiu D. Evaluation of failure probability via surrogate models. J Comput Phys 2010;229:8966–80.
[41] Basudhar A, Missoum S, Sanchez A. Limit state function identification using support vector machines for discontinuous responses and disjoint failure domains. Probabilist Eng Mech 2008;23:1–11.
[42] Bourinet J-M, Deheeger F, Lemaire M. Assessing small failure probabilities by combined subset simulation and support vector machines. Struct Saf 2011;33:343–53.
[43] Matheron G. Principles of geostatistics. Econ Geol 1963;58:1246.
[44] Sasena M. Flexibility and efficiency enhancements for constrained global design optimization with Kriging approximation. Ph.D. thesis, University of Michigan; 2002.
[45] Santner TJ, Williams BJ, Notz W. The design and analysis of computer experiments. Berlin, Heidelberg: Springer-Verlag; 2003.
[46] Vazquez E, Bect J. A sequential Bayesian algorithm to estimate a probability of failure. In: Proceedings of the 15th IFAC symposium on system identification, Saint-Malo, France, July 6–8. p. 546–50.
[47] Li L, Bect J, Vazquez E. Bayesian subset simulation: a Kriging-based subset simulation algorithm for the estimation of small probabilities of failure. In: Proceedings of PSAM 11 and ESREL 2012, 25–29 June 2012, Helsinki, Finland.
[48] Bect J, Ginsbourger D, Li L, Picheny V, Vazquez E. Sequential design of computer experiments for the estimation of a probability of failure. Stat Comput 2012;22:773–93.
[49] Picheny V. Improving accuracy and compensating for uncertainty in surrogate modeling. Ph.D. thesis, University of Florida; 2009.
[50] Baudoui V, Klotz P, Hiriart-Urruty J-B, Jan S, Morel F. Local Uncertainty Processing (LOUP) method for multidisciplinary robust design optimization. Struct Multidiscip Optim 2012:1–16.
[51] Li L, Bect J, Vazquez E. A numerical comparison of two sequential Kriging-based algorithms to estimate a probability of failure. In: Uncertainty in Computer Model Conference UK: Sheffield; 2010, July, 12–14.
[52] Kleijnen JPC. Kriging metamodeling in simulation: a review. Eur J Oper Res 2009;192:707–16.
[53] Cornford D, Csató L, Opper M. Sequential, bayesian geostatistics: a principled method for large data sets. Geogr Anal 2005;37:183–99.
[54] Hansen N. The CMA evolution strategy: a comparing review. In: JA Lozano, P Larranaga, I Inza, E Bengoetxea (Eds.), Towards a new evolutionary computation Studies infuzziness and soft computating; 2006, vol. 192, pp. 75–102.
[55] Lophaven S, Nielsen H, Songdergaard J. DACE a MATLAB Kriging toolbox, Technical Report IMM-TR-2002–12. Technical University of Denmark; 2002.
[56] Au S, Ching J, Beck J. Application of subset simulation methods to reliability benchmark problems. Struct Saf 2007;29:183–93.
[57] Schueremans L, Van Gemert D. Benefit of splines and neural networks in simulation based structural reliability analysis. Struct Saf 2005;27:246–61.
[58] Rajashekhar M, Ellingwood B. Comparison of response surface and neural network with other methods for structural reliability analysis. Struct Saf 1993;12:205–20.
[59] Gayton N, Bourinet J, Lemaire M. CQ2RS: a new statistical approach to the response surface method forreliability analysis. Struct Saf 2003;25:99–121.
[60] Morio J. Non-parametric adaptive importance sampling for the probability estimation of a launcher impact position. Reliab Eng Syst Saf 2011;96:178–83.

Bibliographie

[1] M. Allouche. The integration of UAVs in airspace. *Air and Space Europe*, 2(1) :101–104, 2000.

[2] J. Asmat, B. Rhodes, J. Umansky, C. Villavicencio, A. Yunas, G. Donohue, and A. Lacher. UAS safety : Unmanned aerial collision avoidance system (UCAS). In *Systems and Information Engineering Design Symposium, 2006 IEEE*, pages 43–49, april 2006.

[3] S. K. Au. Reliability-based design sensitivity by efficient simulation. *Computers and Structures*, 83 :1048–1061, 2005.

[4] S. K. Au and J. L. Beck. Estimation of small failure probabilities in high dimensions by subset simulations. *Probabilistic Engineering Mechanics*, 16(4) :263–277, 2001.

[5] M. Balesdent, J. Morio, and J. Marzat. Kriging-based adaptive importance sampling algorithms for rare event estimation. *Structural Safety*, 13 :1–10, 2013.

[6] I. Banerjee, S. Pal, and S. Maiti. Computationally efficient black-box modeling for feasibility analysis. *Computers & Chemical Engineering*, 34(9) :1515–1521, 2010.

[7] A. Basudhar, S. Missoum, and A. Sanchez. Limit state function identification using Support Vector Machines for discontinuous responses and disjoint failure domains. *Probabilistic Engineering Mechanics*, 23 :1–11, 2008.

[8] V. Baudoui, P. Klotz, J.-B. Hiriart-Urruty, S. Jan, and F. Morel. LOcal Uncertainty Processing (LOUP) method for multidisciplinary robust design optimization. *Structural and Multidisciplinary Optimization*, pages 1–16, 2012.

[9] J. Bect, D. Ginsbourger, L. Li, V. Picheny, and E. Vazquez. Sequential design of computer experiments for the estimation of a probability of failure. *Statistics and Computing*, 22 :773–793, 2012.

[10] I. Beichl and F. Sullivan. The importance of importance sampling. *Computing in science and engineering*, pages 71–73, Mars/Avril 1999.

[11] P. Bjerager. On computation methods for structural reliability analysis. *Structural Safety*, 9(2) :79–96, 1990.

[12] R. Bjerager. *Methods for structural reliability computation*, pages 89–136. Springer Verlag, New York, 1991.

[13] P. H. Borcherds. Importance sampling : An illustrative introduction. *European Journal of Physics*, pages 405–411, April 2000.

[14] Z. I. Botev and D. P. Kroese. An efficient algorithm for rare-event probabilityn estimation, combinatorial optimization, and counting. *Methodol. Comput. Appl. Probab.*, 10(4) :471–505, 2012.

[15] Z. I. Botev and D. P. Kroese. Efficient Monte-Carlo simulation via the generalized splitting method. *Statistics and Computing*, 22(1) :1–16, 2012.

[16] J.-M. Bourinet, F. Deheeger, and M. Lemaire. Assessing small failure probabilities by combined subset simulation and support vector machines. *Structural Safety*, 33 :343–353, 2011.

[17] M. Broniatowski and V. Caron. Small variance estimators for rare event probabilities. *ACM Trans. Model. Comput. Simul.*, 23(1) :7 :1–7 :23, January 2013.

[18] P. Brooker. Reducing mid-air collision risk in controlled airspace : Lessons from hazardous incidents. *Safety Science*, 43(9) :715 – 738, 2005.

[19] C. Bucher and U. Bourgund. A fast and efficient response surface approach for structural reliability problems. *Structural Safety*, 7 :57–66, 1990.

[20] J. A. Bucklew. *Introduction to Rare Event Simulation*. Springer, 2004.

[21] F. Cérou, P. Del Moral, F. Le Gland, and P. Lezaud. Genetic genealogical models in rare event analysis. *ALEA, Latin American Journal of Probability and Mathematical Statistics*, 1 :181–203, 2006. Paper 01–08.

[22] N. Chopin, P. E. Jacob, and O. Papaspiliopoulos. SMC^2 : an efficient algorithm for sequential analysis of state space models. *Journal of the Royal Statistical Society : Series B (Statistical Methodology)*, 75(3) :397–426, 2013.

[23] M. Chouchane, S. Paris, F. Le Gland, and M. Ouladsine. Splitting method for spatio-temporal sensors deployment in underwater systems. In Jin-Kao Hao and Martin Middendorf, editors, *Evolutionary Computation in Combinatorial Optimization*, volume 7245 of *Lecture Notes in Computer Science*, pages 243–254. Springer Berlin Heidelberg, 2012.

[24] S. G. Coles. *An introduction to statistical modeling of extreme values*. Springer, New York, 2001.

[25] F. Cérou, P. Del Moral, T. Furon, and A. Guyader. Sequential Monte-Carlo for rare event estimation. *Statistics and Computing*, 22 :795–808, 2012.

[26] H. E. Daniels. Saddlepoint approximations in statistics. *Ann. Math. Statist*, 25(4) :631–650, 1954.

[27] A. C. Davison and R. L. Smith. Models for exceedances over high thresholds (with discussion). *Journal of the Royal Statistical Society*, 52 :393–442, 1990.

[28] L. de Haan and J. Pickands. Stationary min-stable stochastic processes. *Probability Theory and Related Fields*, 7 :477–492, 1986.

[29] T. Homem de Mello and R. Y. Rubinstein. *Rare event estimation for static models via cross-entropy and importance sampling*. John Wiley, New York, 2002.

[30] T. Dean and P. Dupuis. Splitting for rare event simulation : a large deviation approach to design and analysis. *Stochastic processes and their applications*, 119(2) :562–587, 2009.

[31] P. Del Moral. *Feynman-Kac Formulae, Genealogical and Interacting Particle Systems with Applications. Probability and its Applications.* Springer, New York, 2004.

[32] P. Del Moral and J. Garnier. Genealogical particle analysis of rare events. *The Annals of Applied Probability*, 15 :2496–2534, 2005.

[33] A. Dembo and O. Zeitouni. *Large deviations techniques and applications*. Springer-Verlag, New York, 1998.

[34] F. den Hollander. *Large Deviations*. American Mathematical Society, New York, 2008.

[35] V. Dubourg, E. Deheeger, and B. Sudret. Metamodel-based importance sampling for the simumation of rare events. In *Faber, M. J. Kohler and K. Nishilima (Eds.), Proceedings of the 11th International Conference of Statistics and Probability in Civil Engineering (ICASP2011), Zurich, Switzerland*, 2011.

[36] W. Duch, R. Setiono, and J. M. Zurada. Computational intelligence methods for rule-based data understanding. *Proceedings of the IEEE*, 92(5) :771–805, 2004.

[37] P. Dupuis and H. Wang. Importance sampling, large deviations, and differential games. *Stochastics and Stochastics Reports*, 76 :481–508, 2004.

[38] B. Echard, N. Gayton, and M. Lemaire. AK-MCS : An active learning reliability method combining Kriging and Monte-Carlo Simulation. *Structural Safety*, 33 :145–154, 2011.

[39] P. Embrechts, C. Kluppelberg, and T. Mikosch. *Modelling Extremal Events for Insurance and Finance*. Springer Verlag, Berlin, 1997.

[40] S. Engelund and R. Rackwitz. A benchmark study on importance sampling techniques in structural reliability. *Structural Safety*, 12(4) :255–276, 1993.

[41] O. Dietlevsen et H.O Madsen. *Structural Reliability Methods*. John Wiley and Sons, New York, 1996.

[42] R. A. Fisher and L. H. C. Tippett. On the estimation of the frequency distributions of the largest or smallest member of a sample. Proceedings of the Cambridge Philosophical Society, 1928.

[43] G. S. Fishman. *Monte-Carlo : Concepts, Algorithms, and Applications*. Springer, New York, 1996.

[44] A.-L. Fougères. Multivariate extremes. In B. Finkenstädt and H. Rootzén, editors, *Extreme Values in Finance, Telecommunications, and the Environment, Monographs on Statistics and Applied Probability 99*, pages 373–388. Chapman and Hall/CRC, Boca, Raton, 2004.

[45] J. Garnier and P. Del Moral. Simulations of rare events in fiber optics by interacting particle systems. *Optics Communications*, 267(1) :205–214, 2006.

[46] I. K. Glad, N. L. Hjort, and N. G. Ushakov. Mean-squared error of kernel estimators for finite values of the sample size. *Journal of Mathematical Sciences*, 146 :5977–5983, 2007.

[47] P. Glasserman. *Monte-Carlo Methods in Financial Engineering*. Springer, New York, 2003.

[48] P. Glasserman and Y. Wang. Counter examples in importance sampling for large deviations probabilities. *Ann. Appl. Prob.*, 7 :731–746, 1997.

[49] H. M. Gomes and A. M. Awruch. Comparison of response surface and neural network with other methods for structural reliability analysis. *Structural Safety*, 26 :49–67, 2004.

[50] A. Grancharova, J. Kocijan, and T. A. Johansen. Explicit output-feedback nonlinear predictive control based on black-box models. *Engineering Applications of Artificial Intelligence*, 24(2) :388–397, 2011.

[51] J. M. Hammersley and D. C. Handscomb. *Monte-Carlo methods*. Methuen, London, 1964.

[52] A. M. Hasofer and N. C. Lind. An exact and invariant first-order reliability format. *Journal of Engineering Mechanics*, 100 :111–121, 1974.

[53] W. Hoeffding. A class of statistics with asymptotically normal distributions. *Annals of Mathematical Statistics*, 19 :293–325, 1948.

[54] B. Iooss. Revue sur l'analyse de sensibilité globale de modèles numériques. *Journal de la Société Française de Statistique*, 152(1) :1–23, 2011.

[55] R. Irvine. A simplified approach to conflict probability estimation. In *Digital Avionics Systems, 2001. DASC. The 20th Conference*, volume 2, pages 7F5/1 –7F5/12, oct 2001.

[56] R. Pastel J. Morio and F. Le Gland. Estimation de probabilités et de quantiles rares pour la caractérisation d'une zone de retombée d'un engin. *Journal de la Société Française de Statistique*, 152(4) :1–29, 2011.

[57] D. Jacquemart and J. Morio. Conflict probability estimation between aircraft with dynamic importance splitting. *Safety Science*, 51(1) :94–100, 2013.

[58] J. Janusevskis and R. Le Riche. Simultaneous Kriging-based estimation and optimization of mean response. *Journal of Global Optimization*, 55 :313–336, 2012.

[59] J. L. Jensen. *Saddlepoint Approximations*. Oxford University Press, USA, 1995.

[60] D. R. Jones. A taxonomy of global optimization methods based on response surfaces. *Journal of Global Optimization*, 21(4) :345–383, 2001.

[61] D. R. Jones, C. D. Perttunen, and B. E. Stuckman. Lipschitzian optimization without the Lipschitz constant. *Journal of Optimization Theory and Applications*, 79(1) :157–181, 1993.

[62] D. R. Jones, M. J. Schonlau, and W. J. Welch. Efficient global optimization of expensive black-box functions. *Journal of Global Optimization*, 13(4) :455–492, 1998.

[63] H. Kahn and T. E. Harris. Estimation of particle transmission by random sampling. *Appl. Math. Ser.*, 12 :27–30, 1951.

[64] A. J. Keane and P. B. Nair. *Computational Approaches for Aerospace Design*. John Wiley & Sons, Ltd, 2005.

[65] T. S. Kelso. Analysis of the Iridium 33-Cosmos 2251 collision. Rapport interne, 2009.

[66] I. D. L. Kirkland, R. E. Caves, I. M. Humphreys, and D. E. Pitfield. An improved methodology for assessing risk in aircraft operations at airports, applied to runway overruns. *Safety Science*, 42(10) :891–905, 2004.

[67] M. J. Kochenderfer, L. P. Espindle, J. D. Griffith, and J. K. Kuchar. A Bayesian approach to aircraft encounter modeling. In *AIAA Guidance, Navigation and Control Conference and Exhibit, 2008 GNCC*, pages 1 –21, august 2008.

[68] M. J. Kochenderfer, L. P. Espindle, J. D. Griffith, and J. K. Kuchar. Encounter modeling for sense and avoid development. In *Integrated Communications, Navigation and Surveillance Conference, 2008. ICNS 2008*, pages 1–10, may 2008.

[69] P. S. Koutsourelakis, H. J. Pradlwarter, and G. I. Schueller. Reliability of sructures in high dimensions, part I algorithms and application. *Probabilistic Engineering Mechanics*, 19(4) :409–417, 2004.

[70] P.S. Koutsourelakis. Reliability of structures in high dimensions, part II theoretical validation. *Probabilistic Engineering Mechanics*, 19(4) :419–423, 2004.

[71] V. Kreinovich and S. A. Ferson. A new Cauchy-based black-box technique for uncertainty in risk analysis. *Reliability Engineering & System Safety*, 85(1-3) :267–279, 2004.

[72] D. P. Kroese and R. Y. Rubinstein. Monte-Carlo methods. *Wiley Interdisciplinary Reviews : Computational Statistics*, 4(1) :48–58, 2012.

[73] A. Lagnoux. Rare event simulation. *Probability in the Engineering and Informational science*, 20 :45–66, 2006.

[74] T. Lassen and N. Recho. *Fatigue Life Analyses of Welded Structures*. ISTE Wiley, New York, 2006.

[75] M. R. Leadbetter, G. Lindgren, and H. Rootzén. *Extremes and related properties of random sequences and series*. Springer Verlag, New York, USA, 1983.

[76] M. R. Leadbetter and H. Rootzén. Extremal theory for stochastic processes. *Annals of Probability*, 16 :431–478, 1988.

[77] P. L'Écuyer, V. Demers, and B. Tuffin. Rare events, splitting, and quasi–Monte Carlo. *ACM Transactions on Modeling and Computer Simulation*, 17(2 (Special issue honoring Perwez Shahabuddin)), April 2007. Article 9.

[78] P. L'Écuyer, F. Le Gland, P. Lezaud, and B. Tuffin. Splitting methods. In Gerardo Rubino and Bruno Tuffin, editors, *Monte Carlo Methods for Rare Event Analysis*, chapter 3, pages 39–61. John Wiley & Sons, Chichester, 2009.

[79] P. Lemaître, E. Sergienko, A. Arnaud, N. Bousquet, F. Gamboa, and B. Iooss. Density modification based reliability sensitivity analysis. submitted, 2013.

[80] M. C. Leva, M. De Ambroggi, D. Grippa, R. De Garis, P. Trucco, and O. Sträter. Quantitative analysis of ATM safety issues using retrospective accident data : The dynamic risk modelling project. *Safety Science*, 47(2) :250 – 264, 2009.

[81] J. Li and D. Xiu. Evaluation of failure probability via surrogate models. *Journal of Computational Physics*, 229 :8966–8980, 2010.

[82] L. Li, J. Bect, and E. Vazquez. A numerical comparison of two sequential Kriging-based algorithms to estimate a probability of failure. In *Uncertainty in Computer Model Conference, Sheffield, UK, July, 12-14,2010*.

[83] L. Li, J. Bect, and E. Vazquez. Bayesian Subset Simulation : a Kriging-based subset simulation algorithm for the estimation of small probabilities of failure. In *Proceedings of PSAM 11 and ESREL 2012, 25-29 June 2012, Helsinki, Finland*, 2012.

[84] D. Madakat, J. Morio, and D. Vanderpooten. Biobjective planning of an active debris removal mission. *Acta Astronautica*, 84 :182 – 188, 2013.

[85] H. O. Madsen, S. Krenk, and N. C. Lind. *Methods of structural safety*. Springer-Verlag, Englewood Cliffs, 1986.

[86] G. Matheron. Principles of geostatistics. *Economic Geology*, 58(8) :1246–1266, 1963.

[87] A. McNeil and T. Saladin. The peaks over threshold method for estimating high quantiles of loss distributions. Proceedings of the 28th International ASTIN Colloquium,Cairns, 1997.

[88] R. E. Melchers. Importance sampling in structural systems. *Structural Safety*, 6(1) :3–10, 1989.

[89] S. P. Meyn. *Control Techniques for Complex Networks*. Cambridge University Press, 2007.

[90] G. A. Mikhailov. *Parametric Estimates by the Monte Carlo Method*. VSP, Utrecht (NED), 1999.

[91] J. Morio. How to approach the importance sampling density ? *Eur. J. Phys*, 31 :41–48, 2010.

[92] J. Morio. Global and local sensitivity analysis methods for a physical system. *Eur. J. Phys.*, 32 :1577–1583, 2011.

[93] J. Morio. Influence of input pdf parameters of a model on a failure probability estimation. *Simulation Modelling Practice and Theory*, 19(10) :2244–2255, 2011.

[94] J. Morio. Non parametric adaptive importance sampling of the probability estimation of launcher impact position. *Reliability engineering and system safety*, 96(1) :178–183, 2011.

[95] J. Morio. Extreme quantile estimation with nonparametric adaptive importance sampling. *Simulation Modelling Practice and Theory*, 27 :76–89, 2012.

[96] J. Morio, D. Jacquemart, M. Balesdent, and J. Marzat. Kriging-based adaptive importance sampling algorithms for rare event estimation. *Computational Statistics & Data Analysis*, 66 :117–128, 2013.

[97] J. Morio, T. Lang, and C. Le Tallec. Estimating separation distance loss probability between aircraft in uncontrolled airspace in simulation. *Safety Science*, 50(4) :995–1004, 2012.

[98] J. Morio and F. Muller. Onboard sensor orientation analysis with Kriging method. *Acta Astronautica*, 2-3 :374–381, jan-feb 2009.

[99] J. Morio and F. Muller. Spatial object classification by radar measurements. *Aerospace Science and Technology*, 14(4) :259–265, 2010.

[100] J. Morio and R. Pastel. Sampling technique for launcher impact safety zone estimation. *Acta Astronautica*, 66(5-6) :736–741, 2010.

[101] J. Morio and R. Pastel. Plug-in estimation of d-dimensional density minimum volume set of a rare event in a complex system. *Proceedings of the IMechE, Part O, Journal of Risk and Reliability*, 226(3) :337–345, 2012.

[102] J. Morio, R. Pastel, and F. Le Gland. An overview of importance splitting for rare event simulation. *European Journal of Physics*, 31 :1295–1303, 2010.

[103] J. Morio, R. Pastel, and F. Le Gland. Missile target accuracy estimation with importance splitting. *Aerospace Science and Technology*, 25(1) :40–44, 2013.

[104] J. Morio, P. Réfrégier, P. Dubois-Fernandez, F. Goudail, and X. Dupuis. Information theory based approach for contrast analysis in polarimetric and/or interferometric sar images. *IEEE Trans. on Geoscience and Remote Sensing*, 48(8) :2185–2196, 2008.

[105] J. Morio, P. Réfrégier, P. Dubois-Fernandez, F. Goudail, and X. Dupuis. A characterization of shannon entropy and bhattacharyya measure of contrast in polinsar images. *The Proceedings of the IEEE*, 97(6) :1097–1108, 2009.

[106] A. Nataf. Distribution des distributions dont les marges sont données. *Comptes rendus de l'Académie des Sciences*, 225 :42–43, 1962.

[107] J. C. Neddermeyer. Computationally efficient nonparametric importance sampling. *Journal of the American Statistical Association*, 104 :788–802, 2009.

[108] J. C. Neddermeyer. Non-parametric partial importance sampling for financial derivative pricing. *Quantitative Finance*, pages 1469–7696, 2010.

[109] H. Niederreiter and J. Spanier. *Monte Carlo and Quasi-Monte Carlo Methods*. Springer, 2000.

[110] J. Nuñez Garcia, Z. Kutalik, K.-H. Cho, and O. Wolkenhauer. Level sets and minimum volume sets of probability density functions. *International Journal of Approximate Reasoning*, 34(1) :25–47, 2003.

[111] G. C. Orsak and B. Aazhang. A class of optimum importance sampling strategies. *Information Sciences*, 84(1-2) :139–160, 1995.

[112] P. Ostwald and W. Hershey. Helping Global Hawk fly with the rest of US. In *Integrated Communications, Navigation and Surveillance Conference, 2007. ICNS '07*, pages 1–11, 30 2007-may 3 2007.

[113] R. Pastel, J. Morio, and F. Le Gland. Improvement of satellite conflict prediction reliability through use of the adaptive splitting technique. *Proceedings of the IMechE, Part G : Journal of Aerospace Engineering*, to appear, 2013.

[114] L. Pei-Ling and A. Der Kiureghian. Optimization algorithms for structural reliability. *Structural Safety*, 9(3) :161–177, 1991.

[115] V. Picheny. *Improving accuracy and compensating for uncertainty in surrogate modeling.* Thèse de doctorat, University of Florida, 2009.

[116] J. Pickands. Statistical inference using extreme order statistics. *Annals of Statistics*, 3 :119–131, 1975.

[117] W. Polonik. Minimum volume sets and generalized quantile processes. *Stochastic Processes and their Applications*, 69(1) :1–24, 1997.

[118] R. Rackwitz and B. Flessler. Structural reliability under combined random load sequences. *Computers and Structures*, 9(5) :489–494, 1978.

[119] Ph. Réfrégier and J. Morio. Shannon entropy of partially polarized and partially coherent light with gaussian fluctuations. *J. Opt. Soc. Am. A*, 23(12) :3036–3044, 2006.

[120] J.-F. Richard and W. Zhang. Efficient high-dimensional importance sampling. *Journal of Econometrics*, 141(2) :1385–1411, 2007.

[121] B. D. Ripley. *Stochastic Simulation.* Wiley & Sons, New York, USA, 1987.

[122] C. Robert and G. Casella. *Monte Carlo Statistical Methods.* Springer, New York, 2005.

[123] M. Rosenblatt. Remarks on a multivariate transformation. *Annals of Mathematical Statistics*, 23 :470–472, 1952.

[124] M. N. Rosenbluth and A. W. Rosenbluth. Monte-Carlo calculation of the average extension of molecular chains. *J. Chem. Phys*, 356 :356–359, 1955.

[125] R. Rubinstein and D. Kroese. *The Cross-Entropy Method : A Unified Approach to Combinatorial Optimization, Monte-Carlo Simulation and Machine Learning (Information Science and Statistics).* Springer, 2004.

[126] J. S. Sadowsky. Evaluation of large deviation probabilities via importance sampling. In *Signals, Systems and Computers, 1994. 1994 Conference Record of the Twenty-Eighth Asilomar Conference on*, volume 1, pages 30–34, 1994.

[127] T. J. Santner, B. J. Williams, and N. I. Notz. *The design and analysis of computer experiments.* 2003.

[128] P. K. Sarkar and M. A. Prasad. Variance reduction in Monte-Carlo radiation transport using antithetic variates. *Annals of Nuclear Energy*, 19(5) :253–265, 1992.

[129] M. K. Sasena. *Flexibility and Efficienvy enhancements for constrained global design optimization with Kriging approximation.* Thèse de doctorat, University of Michigan, 2002.

[130] M. Schonlau. *Computer Experiments and Global Optimization.* PhD Thesis, University of Waterloo, Canada, 1997.

[131] G. I. Schueller, H. J. Pradlwarter, and P. S. Koutsourelakis. A critical appraisal of reliability estimation procedures for high dimensions. *Probabilistic Engineering Mechanics*, 19 :463–474, 2004.

[132] L. Schueremans and D. Van Gemert. Use of Kriging as Meta-model in simulation procedures for structural reliability. In *9th International conference on structural safety and reliability, Rome*, pages 2483–2490, 2005.

[133] Y. Shang-Wen and H. Ming-Hua. Estimation of air traffic longitudinal conflict probability based on the reaction time of controllers. *Safety Science*, 48(7) :926–930, 2010.

[134] B. W. Silverman. Density estimation for statistcs and data analysis. In *Monographs on Statistics and Applied Applied Probability*. London : Chapman and Hall, 1986.

[135] I. M. Sobol. *A Primer for the Monte Carlo Method*. CRC Press, Boca Raton, Fl., 1994.

[136] I. M. Sobol and S. Kucherenko. Sensitivity estimates for non linear mathematical models. *Mathematical Modelling and Computationnal Experiments*, 1 :407–414, 1993.

[137] B. Sudret. Meta-models for structural reliability and uncertainty quantification. In *Asian-Pacific Symposium on Structural Reliability and its Applications, Singapore, Singapore 2012*, 2012.

[138] H. Touchette. The large deviation approach to statistical mechanics. *Physics Reports*, 478(1-3) :1–69, 2009.

[139] E. Vazquez and J. Bect. A Sequential Bayesian algorithm to estimate a probability of failure. In *15th IFAC, Symposium on System Identification (SYSID'09), 5 pages, Saint-Malo, France, July 6-8*, 2009.

[140] W. Wang, X. Jiang, S. Xia, and Q. Cao. Incident tree model and incident tree analysis method for quantified risk assessment : An in-depth accident study in traffic operation. *Safety Science*, 48(10) :1248–1262, 2010.

[141] K. Worden, C. X. Wong, U. Parlitz, A. Hornstein, D. Engster, T. Tjahjowidodo, F. Al-Bender, D. D. Rizos, and S.D. Fassois. Identification of pre-sliding and sliding friction dynamics : Grey box and black-box models. *Mechanical Systems and Signal Processing*, 21(1) :514–534, 2007.

[142] Z. Yan-Gang and O. Tetsuro. A general procedure for first/second-order reliabilitymethod (FORM/SORM). *Structural Safety*, 21(2) :95–112, 1999.

[143] P. Zhang. Nonparametric importance sampling. *Journal of the American Statistical Association*, 91(434) :1245–1253, September 1996.