THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par l'Université Toulouse 3 - Paul Sabatier

Présentée et soutenue par

Ambre DIET

Le 7 avril 2021

Une approche de bout en bout du tolérancement statistique sous contraintes industrielles :Contribution au jumeau virtuel industriel.

Ecole doctorale : **EDMITT - Ecole Doctorale Mathématiques, Informatique et Télécommunications de Toulouse**

Spécialité : **Mathématiques et Applications**

Unité de recherche :
IMT : Institut de Mathématiques de Toulouse

Thèse dirigée par
Nicolas COUELLAN

Jury
M. Jean Yves DANTAN, Rapporteur
M. Jean-Marc BARDET, Rapporteur
M. Xavier GENDRE, Examinateur
Mme Mathilde MOUGEOT, Examinatrice
M. Jean Philippe NAVARRO, Examinateur
M. Nicolas COUELLAN, Directeur de thèse

INSTITUT DE MATHÉMATIQUES DE TOULOUSE

Abstract

Université Paul Sabatier
IMT

Doctor of Philosophy

An end-to-end approach of statistical tolerancing under industrial constraints: Contribution to the design/industrial virtual twin

by Ambre DIET

In the manufacturing process of a product, various assembly steps are necessary. Several types of requirements have to be met at each level and involve considerations about dimensional uncertainties on the parts to be assembled. Tolerancing is the activity in charge of the management of these uncertainties and takes place both in the product development phase and in the series production phase. In the context of the aeronautics industry, in particular with regards to tolerancing on aerostructures, specificities have to be taken into account in the development of adequate methods and tools. Prior to production, one of the main issues of tolerancing amounts to allocate tolerance limits suited to a given acceptable scrap rate. The aim is to allow the actors concerned with tolerance intervals to agree on a consistent and robust tolerance value. A statistical methodology based on a Chernov bound approach applied to a sum of uniform distributions is proposed. In the production phase, the availability of measurement data allows to refine the statistical tolerancing approach. The linear model often considered can be corrected to serve new approaches. A methodology to manage acceptance criteria on tolerance values is proposed, basing the decision support on risk concepts pertinently defined for industrial actors. Within the framework of the revision of tolerance sharing in an assembly, an optimization problem is formulated with appropriate industrial costs in order to propose the optimal tolerance re-sharing in a stack chain. Finally, the proposed methodologies are implemented in tools allowing industrial processing and end-to-end management of tolerances from elementary parts to final product assembly, thus contributing to the elaboration of the product virtual twin.

INSTITUT DE MATHÉMATIQUES DE TOULOUSE

Résumé

Université Paul Sabatier
IMT

Doctor of Philosophy

An end-to-end approach of statistical tolerancing under industrial constraints: Contribution to the design/industrial virtual twin

by Ambre DIET

Dans le processus de fabrication d'un produit, diverses étapes d'assemblage sont nécessaires. Plusieurs types d'exigences sont à respecter à chaque niveau et ils impliquent de considérer les incertitudes de dimensions sur les pièces à assembler. Le tolérancement est l'activité en charge de la gestion de ces incertitudes et intervient à la fois en phase de développement du produit et en phase série. Dans le contexte de l'industrie aéronautique, en particulier en ce qui concerne le tolérancement sur les aérostructures, des spécificités sont à prendre en compte pour l'élaboration de méthodes et outils adéquats. Avant la mise en production, une des problématiques principales du tolérancement est l'allocation de limites de tolérance adaptées à un certain taux acceptable de rebut. Le but est de permettre aux acteurs concernés par les intervalles de tolérance de s'accorder sur une valeur de tolérance cohérente et robuste. Une méthodologie statistique basée sur une approche type borne de Chernov appliquée à une somme de distributions uniformes est proposée. En phase de production, la disponibilité de données de mesure permet de raffiner la démarche du tolérancement statistique. Le modèle linéaire considéré peut être corrigé à la faveur de nouvelles approches. Une méthodologie de gestion des critères d'acceptation sur les valeurs de tolérance est également proposée, en basant l'outil d'aide à la décision sur des notions de risques définies en adéquation avec les acteurs industriels. Dans le cadre de la révision du partage de tolérances dans un assemblage, un problème d'optimisation est formulé avec des coûts industriels appropriés afin de proposer le re-partage optimal de tolérances dans une chaîne de côte. Enfin, les méthodologies proposées sont implémentées dans les outils permettant le traitement industriel et la gestion de bout en bout des tolérances depuis les pièces élémentaires jusqu'à l'assemblage final du produit, contribuant ainsi à l'élaboration du jumeau virtuel du produit.

Contents

List of Figures

List of Tables

List of Abbreviations

ACTI	**A**ssembly **C**ri**T**ical **I**tem
AKC	**A**ssembly **K**ey **C**haracteristic
API	**A**pplication **P**rogramming **I**nterface
ASCR	**A**irbus **S**afety **C**oefficient **R**esult
CAT	**C**omputer **A**ided **T**olerancing
CP	**C**apability indicator of a **P**rocess measuring the dispersion
CPK	**C**apability indicator of a **P**rocess measuring the offset
CTI	**C**ri**T**ical **I**tem
DIVA	**D**ata **I**ntegrator for **V**ariation **A**nalysis
DTN	**D**esign **T**echnical **N**ote
KC	**K**ey **C**haracteristic
LR	**L**ong **R**ange
MAP	**M**ise **A**u **P**oint
ME	**M**anufaturing **E**ngineering
MSN	**M**anufaturing **S**erial **N**umber
OOT	**O**ut **O**f **T**olerance
PFD	**P**ackage **F**rontier **D**rawing
PKC	**P**rimary **K**ey **C**haracteristic
PCTI	**P**rimary **C**ri**T**ical **I**tem
RSS	**R**ootsquare **S**um of **S**quares
SA	**S**ingle **A**isle
STR	**S**pecific **T**olerance **R**equirement
WC	**W**orst **C**ase

Chapter 1

Introduction

1.1 Contexte

Dans le processus de fabrication d'un avion, de nombreux domaines de compétence sont concernés. Parmi eux, le tolérancement est une activité peu connue du grand public. Dans le département d'ingénierie structure avion, l'équipe tolérancement gère les incertitudes liées à la géométrie et aux dimensions des différentes pièces à assembler lors de toutes les étapes dans la fabrication d'un avion, et plus généralement d'un produit. La nécessité de gérer les tolérances provient des exigences liées à la production, aux performances ou à l'esthétique du produit. Dans le cas spécifique de la structure aéronautique, les incertitudes dimensionnelles sont en effet fortement reliées aux exigences imposées par la sécurité, la qualité ou encore la performance (aérodynamique par exemple) de l'avion ainsi qu'aux exigences d'assemblage (capabilités industrielles, outillages, . . .).

D'un point de vue plus général, on réduit souvent le tolérancement à son activité finale qui consiste à rassembler des spécifications dimensionnelles dans les documents contractuels de définition de l'avion. Dans ce contexte, le tolérancement, aussi appelé GD&T (pour *Geometric Dimensioning and Tolerancing*), peut pourtant être abordé de différentes façons. On distingue souvent le tolérancement géométrique (GPS pour *Geometrical Product Specification*) du tolérancement dimensionnel. Le premier se réfère aux limitations en termes de forme et de position (l'inclinaison, le parallélisme, la localisation, le battement, . . .). Le tolérancement dimensionnel fait quant à lui référence aux limites de dimension ou de taille des pièces d'un assemblage. Ces deux aspects sont complémentaires afin de communiquer au mieux sur les attentes concernant la géométrie et les dimensions des pièces d'un assemblage. Le langage du tolérancement est régi par un système de normes internationales réactualisées régulièrement. Les notions, les symboles, et le vocabulaire nécessaires pour la compréhension et l'interprétation des tolérances y sont définis.

Ce travail de thèse porte sur la gestion des tolérances quelle que soit leur nature. Les défauts de forme ou d'orientation ne sont pas pris en compte en détail dans les méthodologies développées. Toutes les problématiques sont ramenées à l'effet d'une tolérance quelconque sur les dimensions dans un assemblage.

1.2 Notions élémentaires, terminologie et vocabulaire

Cette section définit précisément les notions requises du tolérancement et nécessaires pour la suite de ce manuscrit.

1.2.1 Qu'est ce qu'un assemblage ?

Un ensemble est défini par un certain nombre de pièces ou de sous-ensembles assemblés pour remplir une fonction spécifique. Une caractéristique clé est une exigence pour une dimension de pièce spécifique. Une même pièce peut avoir plusieurs caractéristiques, c'est-à-dire différentes dimensions d'intérêt. Par exemple, la hauteur et la longueur sont deux caractéristiques d'une même pièce. Il existe différentes classes de criticité pour les caractéristiques :

- PKC (Primary Key Characteristic),
  caractéristique finale dont dépend la performance de l'avion.
- AKC (Assembly Key Characteristic),
  caractéristique d'assemblage impliquée dans une chaîne de côte avec pour sortie un PKC.

- STR (specific Tolerance Requirement),
  exigence spécifique non liée à un PKC.
- MKC (Manufacturing Key Characteristic),
  caractéristique élémentaire pour la dimension d'une pièce.

Chaque assemblage implique une exigence globale et un certain nombre de caractéristiques. Par exemple, une exigence PKC dépend de la valeur de plusieurs AKC.

Pour Airbus, la criticité est définie selon deux axes: l'importance telle que définie dans la norme européenne dédiée à la qualité en aéronautique EN9100[1] (permet de choisir entre KC, CTI, ...) et le niveau de la tolérance dans la chaîne de côte (si c'est une exigence finale, il s'agit de PKC ou PCTI autrement ce sera AKC, ACTI, MKC, ...). Ces deux axes de classification vont induire un traitement différent des tolérances lors des phases de design et de production (niveau ciblé, controle qualité, ...).

1.2.2 Vocabulaire lié au tolérancement dans un assemblage

Comme déjà abordé précédemment, il y a dans un assemblage deux concepts principaux : en entrée de l'assemblage, les contributeurs sont représentés par des caractéristiques de dimensions à assembler entre elles. En sortie, la caractéristique associée à l'exigence de niveau supérieur est en quelque sorte le résultat de l'assemblage. Le lien entre les entrées et les sorties est idéalisé par ce qu'on appelle le modèle de tolérance. Ce concept est illustré dans un assemblages de trois contributeurs dans la Figure 1.1.

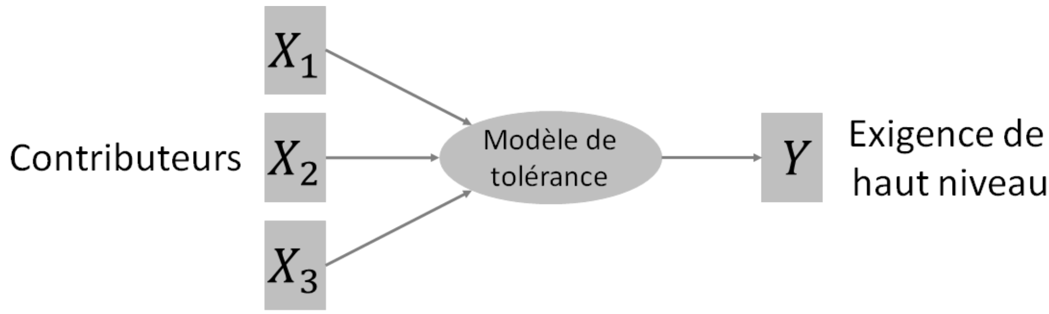

FIGURE 1.1: Modélisation générale d'un assemblage

Dans le cadre du tolérancement, on définit la chaîne de côte d'un assemblage associée à une exigence comme le catalogue des contributeurs et leurs limites de tolérance. Ainsi, chaque assemblage dans la fabrication d'un produit est associé à une chaîne de côte théorique qui servira de brique de base pour les méthodologies d'amélioration de la démarche du tolérancement.

1.2.3 Niveau d'assemblage

Il existe plusieurs étapes d'assemblage lors de la fabrication d'un avion, avec un effet de cascade : le résultat d'un assemblage devient une pièce à assembler au niveau suivant, et ainsi de suite (Figure 1.2). Par conséquent, la variation potentielle d'une caractéristique a un effet cascade aux différents niveaux.

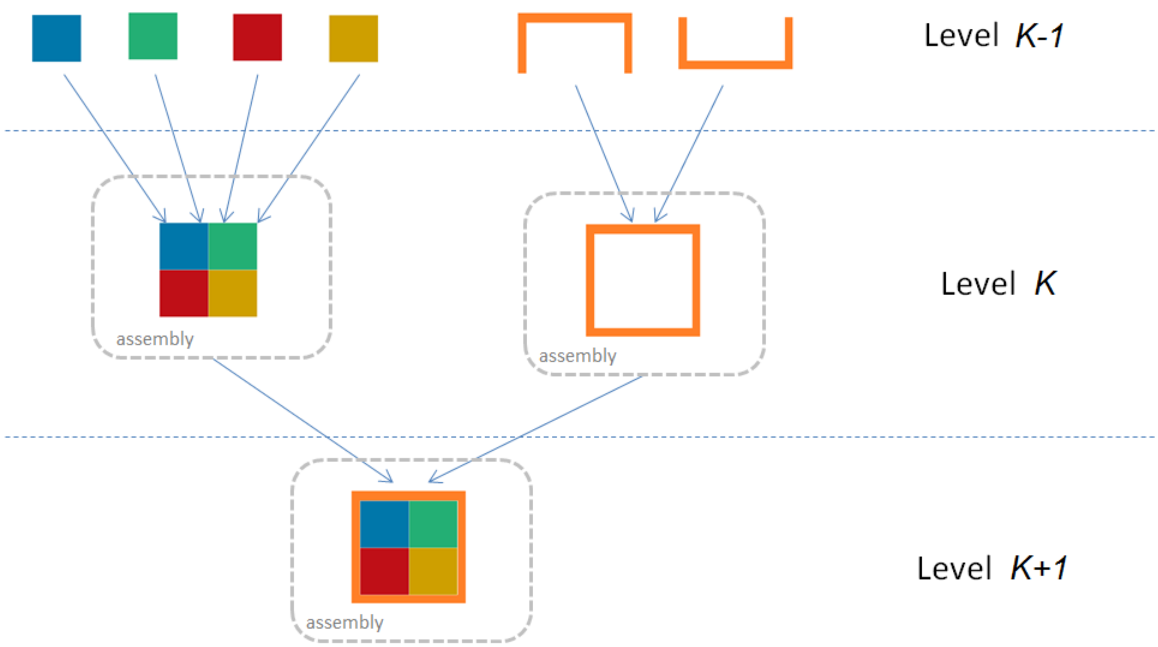

FIGURE 1.2: Exemple de différents niveaux d'assemblage

En général, pour un avion, il y a 5 étapes principales d'assemblage : Le niveau des pièces élémentaires, le niveau de l'assemblage, le niveau du lot de travail, le niveau de la section et le niveau de l'avion (Figure 1.3).

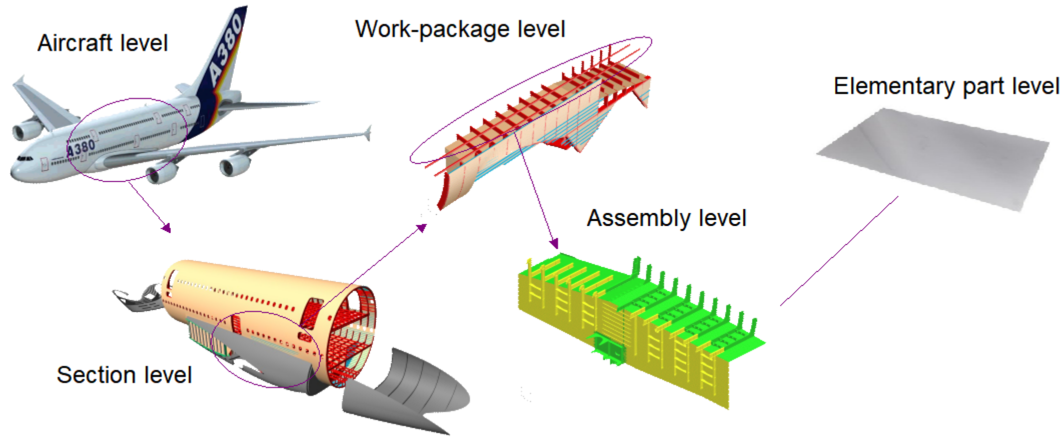

FIGURE 1.3: Niveaux d'assemblage d'un avion

1.2.4 Source de variabilité dans un assemblage

Dans l'Industrie, un produit est conçu et fabriqué pour des spécifications de dimension bien précises. La dimension théorique d'une caractéristique est appelée valeur nominale. En réalité, le produit ne peut théoriquement pas être produit aux dimensions exactes spécifiées sur le plan. Il existe différentes sources de variation de dimension. Pendant la fabrication, il s'agit par exemple de l'outil ou de la machine utilisée, de l'opérateur, des conditions environnementales (hydrométrie, température, . . .) ou encore des variables dans les procédés. Les conditions de transport et d'assemblage peuvent aussi influer, notamment si les points de support sont différents d'un poste de travail à un autre.

Une autre source d'incertitude est le moyen de mesure et sa précision, qui perturbe la perception du produit. Dans un processus de contrôle, le moyen de mesure et sa précision sont supposés être adaptés à la caractéristique mesurée. Dans la suite du manuscrit, il s'agit de la variabilité globale perçue qui est considérée, mélangeant ainsi les effets extérieurs et la précision de mesure. Cette variabilité est considérée au travers d'un intervalle de tolérance autour de la valeur nominale d'une caractéristique.

1.3 Motivations

Dans le contexte particulier de l'industrie aéronautique, le but de ces travaux est d'améliorer la démarche globale du tolérancement statistique dans l'entreprise en améliorant notablement la prise en compte des retours de production des usines.

1.3.1 Les problématiques du tolérancement étudiées

Afin de vérifier que les exigences en matière de structure sont conformes aux spécifications, les dimensions de chaque partie de l'avion doivent être gérées correctement. La fabrication d'un avion comporte plusieurs étapes, chacune correspondant à un niveau d'assemblage. Depuis les pièces élémentaires jusqu'à l'avion complet, chaque écart par rapport à la dimension nominale doit être pris en compte afin d'analyser l'impact sur les exigences globales et de définir un intervalle de tolérance raisonnable. La cascade des tolérances est fonctionnelle et correspond à un juste besoin industriel.

De nombreux acteurs sont impliqués dans les spécifications de tolérance : usines, stress, fournisseurs, transporteurs, ... L'équipe de tolérancement est chargée de coordonner toutes ces activités pour spécifier des valeurs de tolérance cohérentes, en connaissant les différentes contraintes rencontrées tout au long du processus de fabrication. Si on considère un assemblage précis à un certain stade du processus de fabrication, plusieurs problématiques liées au tolérancement apparaissent et sont à traiter.

Il s'agit d'abord de spécifier les exigences de haut niveau de l'assemblage. Cela signifie qu'il faut identifier quelles seront les dimensions qui devront être prises en compte dans l'étape suivante du processus de fabrication. Par exemple, si nous considérons l'assemblage de plusieurs pièces constituant un cadre de porte, le jeu entre le cadre et la porte est une des exigences de haut niveau. On suppose que l'étape suivante est l'installation de la porte dans son cadre. Une illustration simple est présentée dans la Figure 1.4. L'assemblage du cadre de porte doit être conforme aux exigences fonctionnelles pour l'étape suivante du processus de fabrication. Cela signifie que la valeur de la hauteur du cadre doit se situer dans un intervalle précis de valeurs afin que le jeu soit suffisant. Ainsi, la porte pourra s'insérer dans le cadre, sous réserve que la porte ne présente pas de défauts de dimensions. Remarquons que pour le fabricant du cadre, la hauteur du cadre devient l'exigence de haut niveau par l'effet cascade.

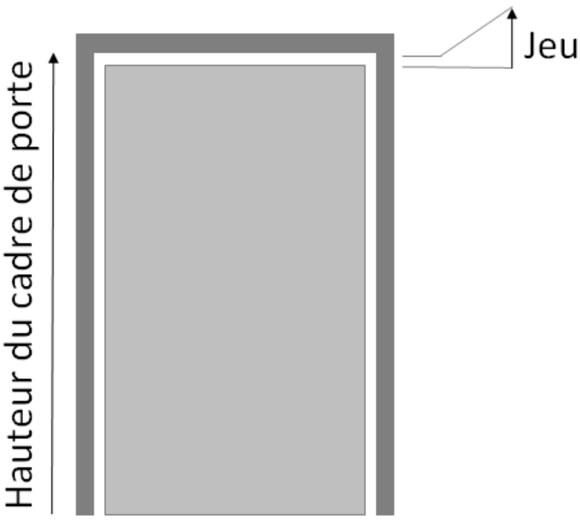

Figure 1.4: Exemple simplifié de l'installation d'une porte dans un cadre de porte

Une fois que les exigences d'assemblage sont identifiées, les spécialistes du tolérancement analysent l'incertitude sur les dimensions des pièces impliquées dans l'assemblage. L'approche *top down* consiste à cascader les exigences de haut niveau sur les tolérances de dimension de ces pièces. L'objectif est de définir un intervalle de tolérance approprié permettant de respecter les exigences de haut niveau. Considérons notre exemple précédent : même s'il existe un intervalle de valeurs acceptables, la hauteur du cadre ne doit être ni trop grande ni trop courte car l'exigence fonctionnelle ne serait pas satisfaite. Le respect de l'exigence dépend de la variabilité des dimensions des pièces impliquées dans l'assemblage du cadre. Quelle doit être la tolérance sur les dimensions impliquées dans l'assemblage afin que les exigences de haut niveau de cette étape soient respectées ?

Supposons maintenant que les intervalles de tolérance pour les dimensions impliquées dans l'assemblage soient connues. L'objectif est de prévoir le résultat sur l'incertitude des exigences de haut niveau. Il s'agit de connaître l'intervalle dans lequel va évoluer la dimension associée à l'exigence fonctionnelle d'un assemblage, connaissant les limites de tolérances des dimensions des contributeurs de l'assemblage. Ensuite, nous pouvons comparer la plage de valeurs obtenue à l'objectif de l'exigence de niveau supérieur et évaluer la tenue ou non de cette exigence. En suivant l'exemple de la porte, cela signifie qu'il faut estimer l'intervalle de valeurs de la hauteur du cadre en connaissant la dimension des pièces composant le cadre de porte. Si la plage de valeurs de la hauteur du cadre dépasse l'exigence cible, la porte ne pourra pas forcément être montée dans le cadre. L'installation de la porte dans l'étape suivante du processus de fabrication ne sera pas possible : l'exigence de niveau supérieur n'est pas satisfaite si on considère les intervalles de tolérance initialement définis pour les dimensions qui contribuent à l'assemblage. Il faudra alors envisager un changement de nominal, de tolérance, d'outillage ou de procédé par exemple. Cette méthode de travail est connue sous le nom d'approche *bottom up*.

Pour un assemblage, on associe généralement une chaîne de côte. Cette chaîne de côte rassemble les informations des contributeurs de l'assemblage et y associe une valeur nominale ainsi qu'un intervalle de tolérance. Que ce soit dans une approche descendante *top down* ou ascendante *bottom up*, la méthode permettant de relier les dimensions d'un contributeur à l'exigence de niveau supérieur est basée sur un modèle de tolérance. Plusieurs modèles sont envisageables et seront détaillés plus loin. Le plus souvent, les paramètres de ce modèle de tolérance sont obtenus à partir de l'analyse 3D des assemblages, où sont modélisés les contributeurs et la caractéristique de l'exigence.

1.3.2 Le processus de tolérancement en phase série

Les tolérances sont généralement agrées entre les différents acteurs pendant la phase de développement. La décision est basée sur un modèle de tolérance et sur des projections concernant les procédés d'assemblage, les capabilités, la criticité, les outillages ou encore les retours d'expérience. Une fois en production, les mesures recueillies sur les dimensions des différentes caractéristiques des assemblages sont alors censées refléter les systèmes des tolérances agréées. Cependant, il arrive que ce ne soit pas le cas et il s'agit de comprendre ce qu'il se passe. Pour cela et grâce à la disponibilité des données de mesure, il devient possible de vérifier les hypothèses choisies pendant la phase de développement.

Tout d'abord il est possible de revoir les modèles de tolérance et d'en vérifier les paramètres. En effet, ce modèle de tolérance doit être cohérent avec les mesures de dimension des contributeurs et la caractéristique d'exigence d'un assemblage.

De plus, la dispersion observée sur une caractéristique de dimension peut facilement être confrontée aux limites de tolérance précédemment définies. En cas de différence majeure entre hypothèse et réalité, il est alors possible de revoir le partage de tolérance au sein d'un assemblage. En effet, un partage de tolérance est une façon de traduire une exigence en une répartition d'incertitude sur les différents contributeurs de l'assemblage. Si par exemple les mesures de la dimension d'un contributeur sont moins dispersées, alors l'incertitude sur ce contributeur est moindre et pourrait profiter à un autre contributeur plus incertain que prévu. Ce rééquilibrage de l'incertitude est envisageable quand la connaissance de la réalité de l'assemblage est suffisante, tant du côté des contributeurs (les entrées) que de l'exigence de l'assemblage (la sortie).

1.4 État de l'art

Dans cette section, un état de l'art général du tolérancement est proposé. Dans chaque chapitre, une revue bibliographique plus précise orientée sur la problématique et les méthodologies du chapitre est proposée.

1.4.1 Fondements du tolerancement

Les activités de tolérancement ont émergé dés l'essor de l'industrie au début du 20$^{\text{ème}}$ siècle. En effet, la géométrie et les dimensions d'un élément sont rarement parfaites, et la notion d'interchangeabilité des pièces dans un assemblage requiert des contraintes sur ces variations par rapport au plan théorique. Ces considérations ont poussé les ingénieurs à travailler sur les spécifications géométriques et dimensionnelles afin de limiter les coûts d'assemblage.

Dés 1938, des travaux formalisent le domaine des méthodes géométriques pour le dimensionnement et le tolérancement. En particulier, le livre de Parker [2] est souvent considéré comme la référence de base sur le tolérancement, parue en 1956. Les premières normes apparaissent pour uniformiser l'approche du tolérancement géométrique. Le principe de Taylor est énoncé. Il est aujourd'hui appelé exigence d'enveloppe et signifie que les dimensions maximales d'une pièce ne doivent pas être supérieures aux limites de taille maximales (idem pour les dimensions minimales qui ne doivent pas être inférieures aux limites de taille minimales). La surface d'une caractéristique dimensionnelle isolée ne doit pas excéder l'enveloppe imaginaire de forme idéale à la dimension au *maximum de matière*, un concept introduit par Chevrolet ([3]) et normé quelques années plus tard ([4]).

Le domaine du tolérancement est depuis en constante évolution. Dans [5], les auteurs proposent un cadre général pour le tolérancement adapté aux nouveaux challenges du domaine. On y retrouve les nombreuses problématiques liées à la gestion d'incertitudes depuis la phase design jusqu'au produit final. Celles-ci sont détaillées et reliées aux travaux déjà existants, depuis les débuts du tolérancement jusqu'à 2018.

1.4.2 La place du tolérancement statistique

La chaîne de côte est l'outil principal du tolérancement. Il s'agit d'un ensemble de maillons correspondant chacun à une caractéristique contribuant à l'assemblage. Pour chacun, un certain intervalle de tolérance est associé et correspond à la variabilité dimensionnelle de la caractéristique par rapport à sa valeur nominale.

En ce qui concerne le tolérancement dimensionnel, l'essor des différentes industries à partir des années 50 a mené à différentes approches proposées par plusieurs ingénieurs. En particulier, l'aspect statistique du tolérancement permet de traiter plus efficacement et plus économiquement la gestion de tolérances dimensionnelles. Les statistiques interviennent lorsqu'on suppose une densité de probabilité associée à une caractéristique de dimension, tel qu'illustré dans la Figure 1.5 avec des lois normales.

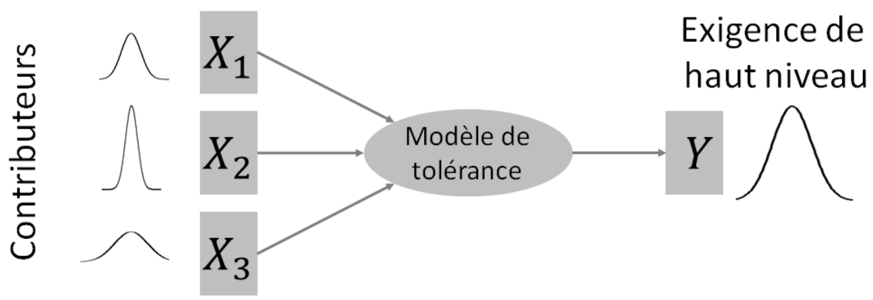

FIGURE 1.5: Modélisation d'un assemblage et lois de distribution

Pour une chaîne de côte, l'analyse statistique de tolérance est une des problématiques les plus abordées dans les travaux industriels. Il s'agit de déterminer la variation de sortie d'un assemblage en connaissant la variation des entrées. Historiquement, il existe deux méthodes principales détaillées dans [6] : le pire cas et l'approche dite statistique.

L'approche pire cas consiste à sommer les variabilités de chacun des maillons de la chaîne de côte pour estimer l'intervalle de variabilité de la caractéristique de sortie de l'assemblage. Cela garantit que les limites de l'intervalle associé à la caractéristique de sortie ne seront jamais dépassées, à condition que les tolérances des caractéristiques d'entrée que représentent les maillons soient respectées, ce qui n'est pas nécessairement le cas dans l'industrie.

La racine carrée de la somme des carrés (RSS) donne un résultat statistique qui repose sur l'hypothèse que les caractéristiques contribuant à l'assemblage sont produites suivant une distribution Gaussienne parfaite sur une plage de 6σ où σ représente l'écart type de la caractéristique tolérancée. La caractéristique de sortie sera ainsi elle aussi comprise dans un intervalle à 6σ. C'est souvent un résultat plus proche de la réalité mais qui nécessite la validation de l'hypothèse Gaussienne pour être cohérent. En phase de design, le niveau d'information disponible sur les distributions statistiques des contributeurs peut être limité (il n'y a pas de phase de pré-série pour la production des avions). Il s'avère alors risqué de prendre une forme de distribution Gaussienne pour définir les tolérances.

Au delà de ces deux approches basiques, de nombreux travaux sur l'analyse statistique des tolérances existent et reposent sur des simulations de Monte Carlo pour différents types d'assemblages ([7, 8, 9, 10, 11, 12, 13, 14]).

Par extension, les travaux de cette thèse utilisent les statistique non seulement pour l'analyse pure de tolérance (Chapitre 2), mais aussi à des fins plus spécifiques dans les Chapitre 3, 4 et 5.

1.4.3 Normes

Avec la mondialisation des industries, un système de normes internationales relatives à la modélisation du dessin technique et notamment au tolérancement a été mis en place. Ces normes sont régulièrement mises à jour et internationalement reconnues.

Il s'agit des normes ISO, acronyme de l' International Organization for Standardization. Dans le domaine, la norme ISO 14638:2015 est la norme GPS de base. Elle conceptualise la spécification géométrique des produits (ISO GPS) et structure les liens entre les normes ISO actuelles et futures pour répondre aux exigences du système ISO GPS. Les principales normes sont à titre d'exemple les normes ISO 8015 ([4]), ISO 5459 ([15]), ISO 1101 ([16]), ISO 14405 ([17]) ou encore ISO 17450 ([18]). Il existe aussi un organisme de normalisation américain appelé ASME pour American Society of Mechanical Engineers qui participe à l'élaboration des normes et publie des documentations. Le standard ASME Y14.5-2009 [19] concerne le tolérancement. Quelques nuances distinguent l'ISO et l'ASME, tel que des différences sur les symboles utilisés ou l'utilisation du principe de l'enveloppe selon lequel les tolérances dimensionnelles limitent également les caractéristiques géométriques si la cible est un élément dimensionnel. En effet, l'ASME permet l'usage d'indication de tolérance dimensionnelle pour une tolérance géométrique tandis que la norme ISO sépare strictement la tolérance dimensionnelle et géométrique.

Une analyse récente sur les normes ISO et leur impact sur la digitalisation de la production industrielle fortement liée à l'élaboration du jumeau virtuel d'un produit est détaillé dans [20].

Le travail présenté ici n'aborde pas directement ces systèmes de normalisation mais ceux-ci permettent la définition et l'harmonisation des notions, symboles et vocabulaire utilisés.

1.4.4 Processus du tolérancement

Comme indiqué dans [21], le processus de tolérancement repose sur plusieurs aspects principaux : la représentation de la tolérance, la spécification de la tolérance, la vérification et l'analyse de la tolérance, et enfin la synthèse de la tolérance. La représentation signifie l'adéquation entre les exigences fonctionnelles et la tolérance visée dans la sortie d'un assemblage représentée par une chaîne de côte. Par exemple dans un contexte aéronautique, la performance aérodynamique doit être assurée grâce à une définition correcte de la tolérance de sortie pour l'étape d'assemblage final. La spécification donne des informations sur la manière de relier une tolérance de sortie avec les contributeurs impliqués dans la chaîne de côte. Cet aspect est souvent appelé l'approche *top down*. Inversement, la vérification et l'analyse des tolérances concernent la concordance entre les tolérances des pièces ou des processus et les exigences fonctionnelles. Il s'agit d'une approche ascendante utilisant un modèle de tolérance. Enfin, la synthèse concerne la boucle de correction effectuée grâce aux informations supplémentaires rendues disponibles sur le produit. Dans [21], un modèle global basé sur les normes ISO et des considérations mécaniques est décrit afin de proposer un processus de tolérance complet et cohérent.

Dans l'entreprise Airbus, le processus du tolérancement s'articule tel que décrit dans la Figure 1.6. Le travail de cette thèse adresse principalement l'étape de vérification et d'analyse de tolérance pour le chapitre 2 et l'étape de synthèse pour les chapitres 3, 4 et 5.

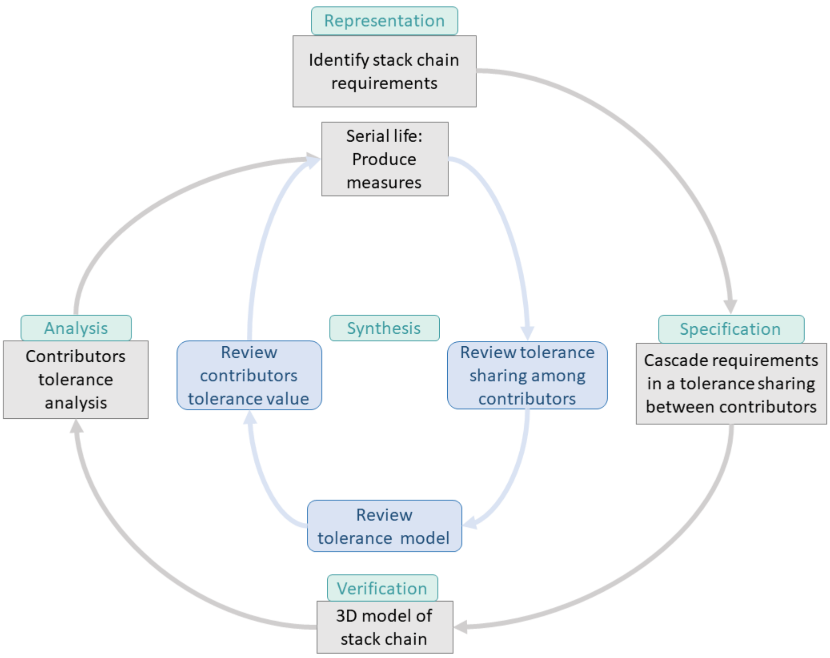

FIGURE 1.6: Airbus tolerancing process.

1.4.5 Outils pour le tolérancement

Il existe dans le commerce plusieurs outils pour le tolérancement. Dans [22], les auteurs font un point sur l'état actuel des modèles de représentation, de manipulation et d'analyse des données de dimensionnement et de tolérancement des principaux systèmes de Tolérancement Assisté par Ordinateur (TAO en français, CAT en anglais pour *Computer Aided Tolerancing*), désormais disponibles dans le commerce. Il existe notamment CETOL 6 Sigma de Sigmetrix ([23]), CATIA 3D FTD de Dassault Systèmes, VSA-GDT et VSA-3D de Engineering Animation Inc. ([24]), e-TolMate de Tecnomatrix ([25]), CAT-3DCS ([26]) ou MECAmaster ([27]).

L'approche linéaire du modèle de tolérance requiert le calcul des coefficients d'influence. Ce calcul de coefficients est notamment implémenté par MECAmaster (pp 93-104 [28]). Ce dernier logiciel est un de ceux utilisés par l'équipe tolérancement Airbus. Les résultats fournis par le logiciel sont ceux utilisés dans les modèles linéaires des travaux présentés par la suite.

Les logiciels tels que e-TolMate de Tecnomatrix et CAT-3DCS de DCS s'affranchissent du modèle de tolérance approximé linéairement. Leur fonctionnement repose sur des simulations (de type méthodes Monte-Carlo). Le logiciel 3DCS est aussi utilisé par les équipes Airbus, toujours dans l'optique d'estimer les coefficients d'influence dans le cadre d'un modèle de tolérance linéaire.

1.4.6 Approche qualité et métrologie

Les domaines de la qualité et de la métrologie sont étroitement liés au tolérancement. En effet la garantie du respect des tolérances spécifiées passe par la vérification des données de mesures et des analyses de qualité. Symétriquement, le contrôle statistique des processus ([29]) repose sur des indicateurs dépendant des limites de tolérance allouées aux caractéristiques mesurées.

Un exemple concret du lien entre le contrôle des processus et le tolérancement est la signification d'un résultat de type RSS pour l'analyse d'une tolérance en termes d'indicateurs bien connus en qualité : C_p et C_{pk}, définis dans le chapitre 4.3. Un résultat de type RSS signifie qu'il faut atteindre un C_p égal à 1 et un C_{pk} égal à 1 pour tous les contributeurs afin de garantir des C_p et C_{pk} de 1 également pour l'exigence finale. Ces indicateurs fréquemment utilisés en contrôle des processus et en qualité sont détaillés dans [30].

Cependant, et plus particulièrement dans l'industrie aéronautique où aucune présérie de contrôle n'est réalisée, les indicateurs couramment considérés pour surveiller les capacités de processus ne sont pas disponibles pendant la phase de conception car les mesures ne sont pas encore produites et la normalité des distributions ne peut pas être vérifiée. C'est ce constat qui motive l'analyse décrite dans le chapitre 2.

1.5 Spécificités du tolérancement dans l'aéronautique

L'industrie aéronautique présente quelques spécificités dans l'approche du tolérancement, distincte des autres industries telles que l'automobile ou encore l'électroménager.

1.5.1 Particularités des assemblages d'aérostructure

Par rapport à d'autres industries, les cadences de production Airbus sont faibles (Figure 1.7), de quelques centaines à quelques milliers d'avions à comparer aux millions de véhicules produits par un constructeur automobile. Cela modifie la façon d'aborder l'échantillonnage traditionnel pour le contrôle de la qualité, puisque peu d'observations sont disponibles. Tous les individus sont donc systématiquement contrôlés par la mesure. Même avec un jeu de données exhaustif, les mesures ne permettent pas forcément de refléter un taux de rebut souvent très bas. En effet, il est difficile de constater sur les données de mesure un taux de 0.1% lorsque seulement 1000 pièces sont produites est mesurées.

	A300/A310	A220/A320	A330/A340/A350	A380	Total
Total orders	816	16224	3125	251	**20416**
Total deliveries	816	9707	2273	243	**13039**
Aircraft in fleet	291	9188	2054	240	**11773**

FIGURE 1.7: Livraisons et commandes Airbus par programme au 31 octobre 2020.

De plus, le nombre de pièces à assembler dans un avion est bien supérieur aux nombres de pièces dans l'assemblage d'une voiture ou d'un appareil électroménager.

Les assemblages d'aérostructure sont généralement de grandes dimensions, de l'ordre de quelques mètres tandis que les déviations dimensionnelles considérées sont souvent exprimées en millimètres. Une autre particularité des aérostructures réside dans la façon de procéder pour un assemblage final. Chez Airbus, les différentes parties de l'avion sont produites en Angleterre, en Espagne, en Allemagne ou en France et sont assemblées dans une ligne d'assemblage finale comme celle de Toulouse. Cela implique le transport de pièces et des systèmes industriels discontinus, engendrant une variabilité supplémentaire. Enfin, une dernière spécificité sont les coûts souvent élevés liés aux assemblages. Cela ne permet pas la production de pré-séries pour la calibration des tolérances par l'analyse des mesures et des procédés. De même, les éléments coûteux d'un assemblage sont généralement retravaillés si les dimensions ne sont pas conformes. Dans certaines industries, il est acceptable d'écarter les pièces qui ne répondent pas aux spécifications. Avec les aérostructures, une pièce hors tolérance engendre généralement des coûts supplémentaires pour des réparations.

1.5.2 Contraintes industrielles

Les contraintes industrielles résident en premier lieu dans les outils de gestion des tolérances. Pour un avion, de nombreuses tolérances sont enregistrées, réparties sur plusieurs programmes, scénarios d'assemblage et niveaux d'assemblage. Les outils pour le tolérancement et le traitement des données de mesure doivent être suffisamment robustes, performants et rapides pour permettre une utilisation par tous les acteurs impliqués dans la gestion des tolérances.

Les approches proposées doivent aussi être compatibles avec les processus industriels appliqués dans l'entreprise. De même, il convient de respecter la culture industrielle déjà en place dans les équipes.

1.5.3 Le concept de jumeau virtuel chez Airbus

Dans un contexte général, un jumeau virtuel (*digital twin* en anglais) est une représentation numérique d'un objet, d'un processus ou d'un système, utilisée à diverses fins ([31, 32, 33, 34]). Dans le contexte de la structure aéronautique chez Airbus, il s'agit d'une maquette numérique qui permet de représenter la géométrie de l'avion. Cette représentation hybride implique plusieurs composants à assembler, chacun avec leur dimension et leur variabilité. La thèse récente ([35]) de J.L. Gregorio détaille précisément le concept de jumeau numérique dans le cadre de la maîtrise géométrique des structures aérodynamiques lors de leur processus d'assemblage.

Dans les travaux de ce manuscrit, nous ne traitons pas directement la géométrie des composants mais bien leur variabilité. Différentes approches sont présentées et contribuent à représenter les tolérances, leur risque associé et leur partage au sein des différents assemblages matérialisés dans le jumeau virtuel. Un tel jumeau numérique requiert une actualisation fréquente au fur et à mesure de la disponibilité de nouvelles informations. Dans le cadre des travaux de ce manuscrit, il s'agit de traiter les systèmes de tolérance en temps réels. Par exemple, les modèles de tolérance peuvent être corrigés à chaque nouvelle donnée de mesure. De même, les risques de hors tolérances, l'ajustement des paramètres de réglages ou le repartage de tolérance doivent pouvoir être mis à jour régulièrement.

1.6 Aperçu de la thèse

Ce travail de thèse s'articule autour des différents points d'amélioration du processus de tolérancement dans un contexte industriel spécifique. Ce processus ainsi que les contributions industrielles associées aux différents chapitres sont illustrés dans la Figure 1.8.

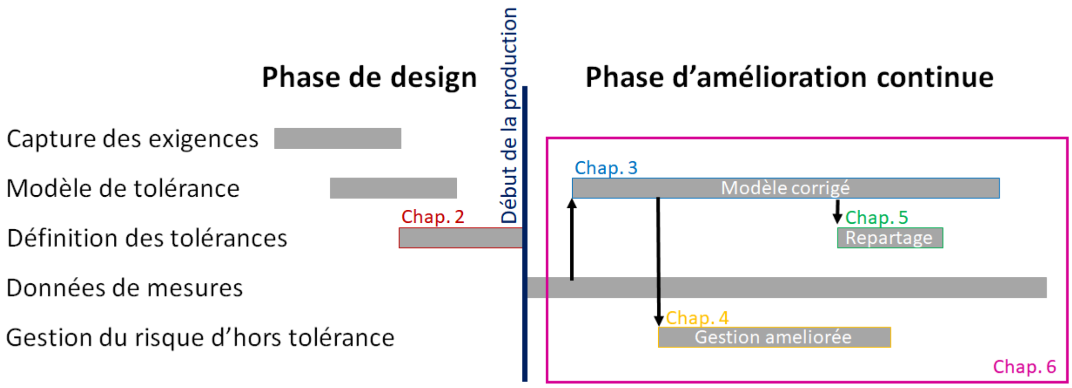

FIGURE 1.8: Schéma des étapes du processus tolérancement Airbus

Ce premier chapitre présente le contexte, les motivations, un état de l'art général autour du sujet ainsi que la terminologie et la définition des notions pré-requises pour la suite.

Le Chapitre 2 concerne la phase de développement et l'approche *bottom up*. En connaissant les limites de tolérance des contributeurs, le but est de donner un intervalle de tolérance sur la caractéristique d'exigence de niveau supérieur de l'assemblage en assurant un certain taux de rebut. Pour cela, un modèle de tolérance linéaire est considéré et les méthodes standards couramment utilisées sont présentées. Un cadre statistique est proposé pour dérouler une méthode de type borne de Chernov afin d'assurer un taux de rebut prédéfini lorsque la connaissance des contributeurs se limite aux deux bornes de tolérance. Des applications théoriques et industrielles sont détaillées.

Dans la suite d'un processus industriel standard, la disponibilité des mesures permet de corriger le modèle de tolérance initialement utilisé. C'est le sujet du Chapitre 3. Différentes alternatives concernant le modèle de tolérance sont discutées et l'utilisation du modèle linéaire est justifiée. Plusieurs méthodologies de correction du modèle sont détaillées et appliquées sur des exemples.

Le Chapitre 4 concerne la gestion des risques liés aux tolérances pendant la production du produit. Cela repose principalement sur des calculs de probabilités d'être hors tolérance, elles-mêmes basées sur des estimations de densité des variables aléatoires associées aux caractéristiques d'un assemblage. Une limite de tolérance sur une caractéristique implique nécessairement un risque figé de hors tolérance. La définition de critère d'acceptation sur les tolérances permet d'assurer plus de flexibilité pour une gestion par le risque lorsque des données de mesure sont disponibles. Des indicateurs concernant l'information relative à la connaissance de l'incertitude amenée par des données de mesure ainsi que des notions de risques industriellement pertinents sont définis puis calculés sur des exemples.

L'intérêt de la disponibilité des mesures est aussi exploité dans le Chapitre 5 dans lequel la problématique du re-partage de tolérance est traitée. Un état de l'art général des techniques d'optimisation au service du tolérancement est proposé. Une méthodologie adaptée au processus Airbus et à ses contraintes industrielles est proposée afin de répondre au besoin spécifique. Les critères de minimisation sont définis et justifiés, puis une méthode de résolution pour un problème d'optimisation mixte et non linéaire est détaillée. Des exemples permettent l'illustration des techniques proposées.

Le Chapitre 6 traite de l'implémentation des méthodes étudiées et de leurs intégrations dans les outils industriels. En particulier, un outil développé au cours d'un travail collaboratif entre les équipes internes Airbus et ce travail de thèse est présenté. Plusieurs cas d'études sont proposés pour illustrer les méthodologies de ce manuscrit.

Enfin, un dernier Chapitre 7 résume les travaux et aborde les perspectives.

1.7 Contributions

Le Chapitre 2 fait l'objet d'une publication ([36]) dans le journal *The International Journal of Advanced Manufacturing Technology, 111(11), 3571-3581*, sous le titre *A Chernov bound for robust tolerance design and application.* Une réfléxion plus générale sur l'approche de bout en bout du tolérancement ([37]) a été présenté à la 16ème conférence sur le *Computer Aided Tolerancing* (CAT) en du 15 au 17 juin 2020. Les méthodologies développées dans les Chapitres 2 et 4 font l'objet de demandes de brevet, référencées respectivement sous les numéros de dépôt 1912668 et 2008847 à l'Office des brevets français (INPI).

Bibliography

[1] Quality Management Systems. Requirements for Aviation, Space and Defence Organizations:. Technical report, BSI British Standards.

[2] Stanley Brampton Parker. *Drawings and Dimensions*. Pitman, 1956.

[3] General Motors Corporation Chevrolet Motor Division Engineering Department. *Chevrolet Draftsman's Handbook*. General Motors Sales Corporation, 1940.

[4] ISO Standard. *ISO 8015: fundamental tolerancing principle*. Geneva: International Organization for Standardization, 1985.

[5] Edward Morse, Jean-Yves Dantan, Nabil Anwer, Rikard Söderberg, Giovanni Moroni, Ahmed Qureshi, Xiangqian Jiang, and Luc Mathieu. Tolerancing: Managing uncertainty from conceptual design to final product. *CIRP Annals*, 67(2):695–717, 2018.

[6] K. W. Chase and W. H. Greenwood. Design Issues in Mechanical Tolerance Analysis. *Manufacturing Review, ASME*, vol 1(no 1):pp. 50–59, March 1988.

[7] Swami D Nigam and Joshua U Turner. Review of statistical approaches to tolerance analysis. *Computer-Aided Design*, 27(1):6–15, 1995. Publisher: Elsevier.

[8] Paul Varghese, Robert N Braswell, Ben Wang, and Chuck Zhang. Statistical tolerance analysis using FRPDF and numerical convolution. *Computer-Aided Design*, 28(9):723–732, 1996. Publisher: Elsevier.

[9] Victor J Skowronski and Joshua U Turner. Using Monte-Carlo variance reduction in statistical tolerance synthesis. *Computer-Aided Design*, 29(1):63–69, 1997. Publisher: Elsevier.

[10] JP Kharoufeh and M Jeya Chandra. Statistical tolerance analysis for non-normal or correlated normal component characteristics. *International Journal of Production Research*, 40(2):337–352, 2002. Publisher: Taylor & Francis.

[11] Ahmed Jawad Qureshi, Jean-Yves Dantan, Jérome Bruyère, and Régis Bigot. Tolerance analysis based on quantified constraint satisfaction problems. *Product Lifecycle Management: Geometric Variations*, pages 125–144, 2010. Publisher: Wiley Online Library.

[12] Jean-Yves Dantan, Nicolas Gayton, Ahmed Jawad Qureshi, Maurice Lemaire, and Alain Etienne. Tolerance analysis approach based on the classification of uncertainty (aleatory/epistemic). *Procedia CIRP*, 10:287–293, 2013. Publisher: Elsevier.

[13] Christopher Rausch, Mohammad Nahangi, Carl Haas, and Wanhua Liang. Monte Carlo simulation for tolerance analysis in prefabrication and offsite construction. *Automation in Construction*, 103:300–314, July 2019.

[14] Heping Peng and Zhuoqun Peng. An iterative method of statistical tolerancing based on the unified Jacobian–Torsor model and Monte Carlo simulation. *Journal of Computational Design and Engineering*, 7(2):165–176, April 2020.

[15] Geometrical Product Specifications. *ISO 5459 Geometrical Tolerancing-Datums and datum systems.* ISO, 2011.

[16] DIN EN ISO 1101. Geometrical Product Specifications (GPS)–Geometrical tolerancing–Tolerances of form, orientation, location and run-out. 2006.

[17] ISO 14405-1: 2010. *Geometrical product specifications (GPS)—dimensional tolerancing—part 1: linear sizes.* International Organization for Standardization Geneva, 2010.

[18] ISO 17450-1. *Geometrical product specifications (GPS)—General concepts—Part 1: Model for geometrical specification and verification.* International Organization for Standardization, Geneva, Switzerland, 2011.

[19] James D Meadows. *Geometric dimensioning and tolerancing: applications, analysis & measurement (per ASME Y14. 5-2009).* American Society of Mechanical Engineers, 2009.

[20] Edward P. Morse, Craig M. Shakarji, and Vijay Srinivasan. A Brief Analysis of Recent ISO Tolerancing Standards and Their Potential Impact on Digitization of Manufacturing. *Procedia CIRP*, 75:11–18, 2018.

[21] Jean-Yves Dantan, Alex Ballu, and Luc Mathieu. Geometrical product specifications — model for product life cycle. *Computer-Aided Design*, 40(4):493–501, April 2008.

[22] Umberto Prisco and Giuseppe Giorleo. Overview of current CAT systems. *Integrated Computer-Aided Engineering*, 9(4):373–387, November 2002.

[23] Anindita Upasana Sarma and K Yogesh. Three-Dimensional Tolerance Stack Utilizing CETOL. Technical report, SAE Technical Paper, 2020.

[24] Donald M. Wisniewski and Praveen Gomer. Tolerance Analysis Using VSA-3D® for Engine Applications. In Hoda A. ElMaraghy, editor, *Geometric Design Tolerancing: Theories, Standards and Applications*, pages 453–464. Springer US, Boston, MA, 1998.

[25] Francesco Chiesi and Lapo Governi. Tolerance analysis with eM-TolMate. *J. Comput. Inf. Sci. Eng.*, 3(1):100–105, 2003.

[26] https://www.3dcs.com/.

[27] Paul Clozel and Pierre-Alain Rance. MECAmaster: A Tool for Assembly Simulation from Early Design, Industrial Approach. In François Villeneuve and Luc Mathieu, editors, *Geometric Tolerancing of Products*, pages 241–273. John Wiley & Sons, Inc., Hoboken, NJ, USA, February 2013.

[28] Pierre Bourdet and Luc Mathieu. *Geometric Product Specification and Verification.* Springer Netherlands, Dordrecht, 2010.

[29] Douglas C Montgomery. *Introduction to statistical quality control.* John Wiley & Sons, 2020.

[30] Donald J. Wheeler and David Smith Chambers. *Understanding Statistical Process Control*. SPC Press. 1992.

[31] Werner Kritzinger, Matthias Karner, Georg Traar, Jan Henjes, and Wilfried Sihn. Digital Twin in manufacturing: A categorical literature review and classification. *IFAC-PapersOnLine*, 51(11):1016–1022, 2018.

[32] Guodong Shao and Moneer Helu. Framework for a digital twin in manufacturing: Scope and requirements. *Manufacturing Letters*, 24:105–107, April 2020.

[33] Tsega Y. Melesse, Valentina Di Pasquale, and Stefano Riemma. Digital Twin Models in Industrial Operations: A Systematic Literature Review. *Procedia Manufacturing*, 42:267–272, 2020.

[34] Yuqian Lu, Chao Liu, Kevin I-Kai Wang, Huiyue Huang, and Xun Xu. Digital Twin-driven smart manufacturing: Connotation, reference model, applications and research issues. *Robotics and Computer-Integrated Manufacturing*, 61:101837, February 2020.

[35] Jean-Loup Gregorio. *Contribution à la définition d'un jumeau numérique pour la maîtrise de la qualité géométrique des structures aéronautiques lors de leurs processus d'assemblage*. Theses, Université Paris-Saclay, March 2020. Issue: 2020UPASN008.

[36] Ambre Diet, Nicolas Couellan, Xavier Gendre, and Julien Martin. A Chernov bound for robust tolerance design and application. *The International Journal of Advanced Manufacturing Technology*, 111(11-12):3571–3581, December 2020.

[37] Ambre Diet, Nicolas Couellan, Xavier Gendre, Julien Martin, and Jean-Philippe Navarro. A statistical approach for tolerancing from design stage to measurements analysis. *Procedia CIRP*, 92:33–38, 2020.

Chapter 2

Design phase: Tolerance model and allocation

This chapter is the subject of a publication ([1]) in the journal *The International Journal of Advanced Manufacturing Technology, 111(11), 3571-3581*, under the title *A Chernov bound for robust tolerance design and application.*

2.1 Introduction

Dimensions may have some deviation from the designed value without significant impact on the quality and functional requirements of the final product. Tolerance intervals are defined according to engineering knowledge and scientific analysis in order to determine these acceptable variations. A deviation out of the determined tolerance bounds is considered non-compliant and imply an action such as an investigation or a modification in the process or the design. The perfect balance between functional requirements and process capability has to be found so that the specified tolerance interval is the most accurate possible. If the tolerance is too tight, the process might not have the capability to manufacture it and either there will be many rejected items or some costly improvement will be needed to produce compliant items. Otherwise, a too wide tolerance will lead to non-conformity with functional requirements of the final product and may lower the product performance. As there are often several steps in a manufacturing process, the propagation of uncertainty has also to be taken into account to specify the tolerance interval of following assembly steps.

In this chapter, the focus is on the tolerances allocation during the design phase in which tolerancing activity does not only aim at anticipating the margins of uncertainty but also help in predicting their effects on the various assembly steps. These involve different physical characteristics of parts, such as part length, hole position, pin, . . . , called features. All tolerancing issues and notations are detailed in the engineering drawing and related documentation practices [2] and [3].

In our case, there are no available dimensions measurements because the focus is on tolerance allocation in the design phase of a product prior production. Considering one specific assembly stage, one of the main concern is to assess the variability of an output feature of the assembly knowing the tolerance range of the input features. In Figure 2.1 which is a simple example in two dimensions, input features are the lengths of different parts and the output feature is the total length of interest in this assembly.

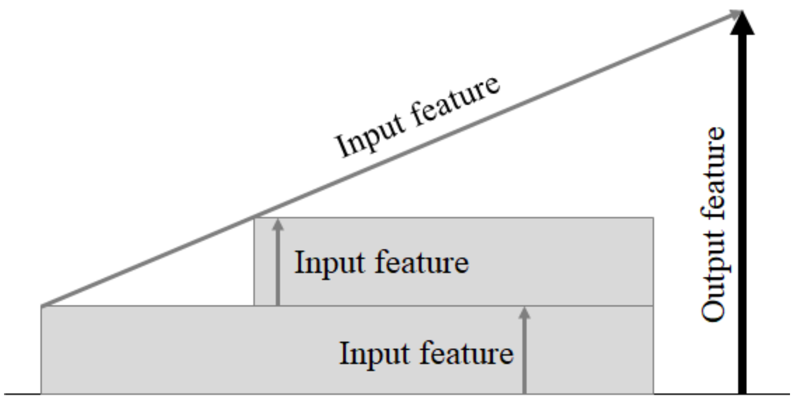

FIGURE 2.1: An assembly example: inputs and output features identification.

In the design phase, both inputs and output tolerance intervals are assumed to be centered around the nominal dimension. To determine the variation around this nominal dimension, there are two main methods detailed in [4]: Worst Case and statistical approaches. These methods propose different ways to define an output tolerance range based on inputs tolerances. Some other approaches based on sampling, fuzzy arithmetic or analytical procedures are reviewed in [5].

Worst Case approach (see [6]) is to consider all assembly parts delivered at their worst acceptable value (assembly output tolerance equals the sum of the input tolerances). Statistical approach, also called RSS approach (square Root of the Sum of

Squares), gives a result assuming all input features are normally distributed (assembly output tolerance equals the square root of the sum of squares of input features tolerances). Statistical result gives a much tighter tolerance range result than the worst case approach, but it does not hedge against the case where input are not reasonably close to their nominal value. To find a balance between this two approaches, Bender [7] proposed to multiply the statistical result by an empiric coefficient of 1.5 to obtain an inflated statistical result which is supposed to give a result tighter than the worst case but more conservative than RSS. However, this technique does not apply if number of assembly inputs is low as it gives a wider result than the worst case approach. Several other statistical methodologies have been studied to obtain the best trade-off between worst case and statistical approaches. For instance, Skowronski and Turner [8] proposed a method relying on Monte Carlo techniques. Choi *et al.* [9] studied an approach based on Taguchi's method requiring the definition a quadratic loss function. The tolerance allocation problem is formulated as a minimization of the sum of machining cost and quality loss. Manufacturing cost considerations for tolerance allocation is beyond the scope of this chapter. Pillet *et al.* [10] proposed to consider weighted inertial tolerancing. Inertial tolerancing works with mean square deviation (inertia) of the output feature as limit instead of considering a tolerance interval. Then, they applied a weighting system based on the number of assembly inputs to obtain a reasonable tolerance result. An other approach has been studied in [11] by taking an interest in the meaning of the conformity. Instead of limiting the assembly output variability, they propose a formal definition of statistical conformity that does not apply individually to a part but to a part population.

Note that tolerance intervals are highly related to the assemblies processes capabilities. Even if suppliers process capability indicators should be monitored as detailed in [12], the normality of features distributions can not always be verified.

One of the objective of the tolerancing is to assess the same confidence in a tolerance interval whatever the distribution of inputs are, as long as these inputs are delivered within the claimed tolerance range. Indeed, suppliers of parts receive a nominal value and two dimension limits. They are also required to follow a target distribution, however checking this compliance is difficult in practice. At the design stage, it is impossible to characterize the features distributions from measurement data. Uniform distribution is a better option to hedge against less favorable distributions of suppliers values.

Knowing lower and upper limits, the less informative distribution is the uniform distribution. It means results obtained with this assumption still stands for alternative distribution provided that distribution support is finite. If the support is not finite, as is the case for Gaussian distribution, uniform assumption is still a good candidate because this is a conservative approach.

A mathematical tool is proposed to define an accurate assembly output tolerance range considering uniform input features and taking into account the stack chain inputs structure. Indeed, result on output tolerance is highly dependant on how balanced is a stack chain. A balanced stack chain means that all contributors have the same impact on the output. Conversely, the predominance of one contributor in the assembly leads to an unbalanced stack chain. The aim is to present an analytical result that links stack chain inputs structure and output tolerance range. This kind of outcomes could be obtained from Monte Carlo simulation but such a procedure is not analytical and does not provide information on the link between inputs and output tolerances.

The chapter is organized as follows: statistical framework is introduced in Section 2.2, main results are presented in Section 2.3: First part is devoted to traditional approach on deviations and following parts detail improvement on the upper bound accuracy and balance term introduction. In Section 2.4, a simulation study is carried out in order to represent and compare our results. Finally, an example on airframe assembly with real inputs data is performed.

2.2 Statistical framework

Consider a set of input features $X_1, \ldots, X_n \in \mathbb{R}$ and an output feature $Y \in \mathbb{R}$. All input features are assumed to be independent random variables for the reason that assembly parts are supposed to be separately produced. The main interest here is in the variability of the output Y and especially in a way to define a tolerance range for this feature.

From an engineering perspective, input features $X_1, \ldots, X_n \in \mathbb{R}$ shall correspond to the stack chain contributors of an assembly such as parts dimensions, while the output feature Y relates to the top level requirement.

Each input feature is assumed to be centered around a nominal dimension and has its own variability characterized by its tolerance range $[-v_i, v_i], \forall i \in \{1, \ldots, n\}$ where $v_1, \ldots, v_n > 0$ are the tolerance bounds. This variability reflects the uncertainty linked to the process (temperature, control plan, ground motion, delivery types, ...).

In order to discuss about the feature Y, assembly step must be modeled to represent the link between inputs and output of the assembly. for isoconstrained mechanisms, a common approach in tolerancing is the linear coefficient model (see [13]). If the variations are supposed to be small around the nominal dimension, the linear approach is appropriate. More elaborated models than linear one could be considered, such as studied in [14] where assembly geometry is taken into account. In this chapter, the framework is to work with a linear model on centered tolerances.

Based on the knowledge of inputs tolerances and influence coefficients on the output, output result is seen as a linear combination of all inputs weighted by known influence coefficients (previously determined with a 3D CAD tool and only linked to the assembly geometry). Let denote $\alpha_1, \ldots, \alpha_n$ the coefficients for a linear tolerance model and input dimension features $Z_1, \ldots, Z_n$, then

$$Y = \sum_{i=1}^{n} \alpha_i Z_i.$$

For ease of notations, the weighted features denoted by $\alpha_1 Z_1, \ldots, \alpha_n Z_n$ are directly treated. These input features are denoted already multiplied by their respective influence coefficients by $X_1, \ldots, X_n$. In this formalism,

$$Y = \sum_{i=1}^{n} X_i.$$

For a given confidence level ρ, the aim is to determine the associated tolerance interval $[-t, t]$ for the output feature Y, verifying

$$\mathbb{P}(|Y| \geqslant t) \leqslant \rho.$$

The tolerance interval is determined based on the distribution of the output feature which depends on input features distributions.

A popular practice is to consider all features as Gaussian which leads to a Gaussian output feature. By applying the commonly used 6σ methodology, the confidence level is $\rho = 0.0027$. In Gaussian framework, the result is $t = 3\sigma_Y$ with σ_Y the standard deviation of the feature Y.

Within the Gaussian framework, the 6σ methodology gives the standard deviation of each input feature: $v_i/3, \forall i \in 1, \ldots, n$. As input features are assumed independent, the standard deviation of the output feature in the Gaussian case is $\frac{1}{3}\sqrt{\sum_{i=1}^{n} v_i^2}$. Again, the 6σ methodology leads to the interval $[-T_{RSS}, T_{RSS}]$ for the output feature tolerance, where $T_{RSS} = \sqrt{\sum_{i=1}^{n} v_i^2}$. This tolerance interval is commonly called the statistical result or RSS (Root Sum of Squared) result by the tolerancing community. However, as tolerance allocation is considered in the design phase, Gaussian assumption can not be verified from measurement data on features. Only tolerances bounds of input features $v_1, \ldots, v_n$ are available.

Input features are considered as uniform random variables, since it is the least informative available distribution given our knowledge about inputs. The purpose is to characterize the deviation of the sum of uniform independent random variables.

Killmann and Von Collani [15] studied the distribution of the sum of uniform features. Their idea was to explicitly calculate density of the sum but such a closed form is numerically intractable and therefore not suited to our context.

Note that our objective is to focus on the quantile of the distribution of Y ensuring a given probability ρ to be out of tolerance. This probability value is fixed in our framework and the tolerancing problem uniformly according to ρ is not planned to be addressed here. This is the point of view of the field of *optimal transport* as developed in [16] but controlling distribution tails leads to poor results in practice for reasonable values of ρ.

In the uniform case, input features standard deviations are now $v_i/\sqrt{3}, \forall i \in 1,\dots,n$, and the standard deviation of the output feature is $\frac{1}{\sqrt{3}}\sqrt{\sum_{i=1}^n v_i^2}$. In this case, the 6σ methodology is applied with standard deviations of uniform distributions and the output feature tolerance interval would be $[-\sqrt{3}\times T_{RSS}, \sqrt{3}\times T_{RSS}]$.

The coefficient $\sqrt{3}$ is an accurate coverage factor on the statistical result if the $v_1,\dots,v_n$ are all equal but it does not address the case where they are unbalanced. Yet, if one of the feature predominates over others, for a same confidence level the output tolerance interval should be tighter, as shown in the Figure 2.2.

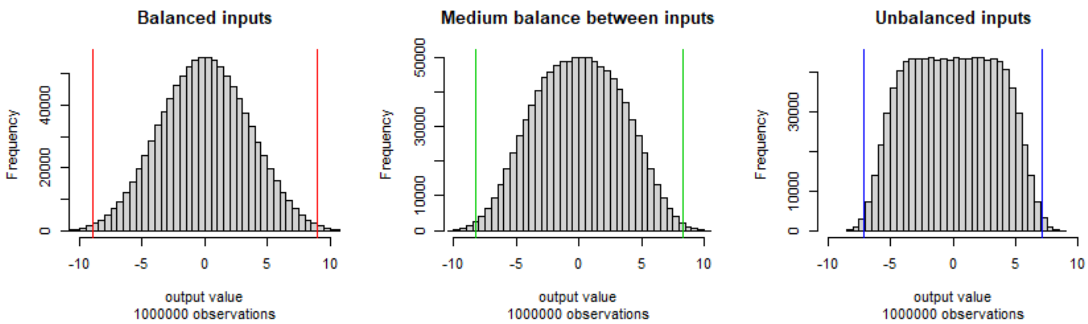

FIGURE 2.2: Output feature distribution for several balance ratio of inputs.

The aim of our approach is to introduce a shape coefficient in order to correct the RSS interval result assuming the input distributions are uniforms. This balance indicator aims to determine how inputs contribution to the output variation is distributed. Indeed, this coefficient will depend on how unbalanced input features are. It also depends on the selected confidence level ρ. The value $\sqrt{3}$ for this shape coefficient means that $v_1 = v_2 = \cdots = v_n$. The more input features are unbalanced, the lower the form coefficient value is.

Next, the focus will be on the role of this coefficient and its impact on the probability ρ.

2.3 Main results

If Gaussian independent input features are considered with a tolerance interval $[-v_i, v_i]$, $\forall i \in \{1, \ldots, n\}$, the associated standard deviation from the 6σ methodology is $v_i/3$, $\forall i \in \{1, \ldots, n\}$. If the features are denoted N_i and $N_i \sim \mathcal{N}(0, v_i/3)$, then $\forall i \in 1, \ldots, n$, the standard Gaussian deviation inequality gives

$$\mathbb{P}\left(\left|\sum_{i=1}^{n} N_i\right| \geqslant t\right) \leqslant 2\exp\left(-\frac{t^2}{2\sum_{i=1}^{n}\left(\frac{v_i}{3}\right)^2}\right)$$

and then

$$\mathbb{P}\left(\left|\sum_{i=1}^{n} N_i\right| \geqslant \frac{1}{3}\sqrt{2log\left(\frac{2}{\rho}\right)\sum_{i=1}^{n} v_i^2}\right) \leqslant \rho$$

that is equivalent to

$$\mathbb{P}\left(\left|\sum_{i=1}^{n} N_i\right| \geqslant l_\rho \times T_{RSS}\right) \leqslant \rho$$

with

$$l_\rho = \frac{1}{3}\sqrt{2log\left(\frac{2}{\rho}\right)}.$$

For fixed ρ and independent uniform input features $U_i \sim \mathcal{U}([-v_i, v_i])$, $\forall i \in 1, \ldots, n$, our aim is to determine f such that

$$\mathbb{P}\left(\left|\sum_{i=1}^{n} U_i\right| \geqslant f \times l_\rho \times T_{RSS}\right) \leqslant \rho. \tag{2.1}$$

This f coefficient aims to correct the RSS value obtained in the 6σ Gaussian case, together with the fixed value l_ρ which manages the confidence level ρ.

2.3.1 Hoeffding approach for the deviation of a sum of bounded random variables

Traditional approaches based on deviations are related to the Hoeffding inequality which provides an upper bound on the probability that the sum of bounded independent random variables deviates more than a certain amount.

As detailed in [17] and [18], this inequality applied to the sum of uniform independent random variables $Y = \sum_{i=1}^{n} X_i$ gives a non asymptotic upper bound for the probability of deviation. This result is summarized in the Proposition 1.

Proposition 1. *Let $v_1, \ldots, v_n > 0$, if $X_1, \ldots, X_n$ are independent random variables such that*

$$\forall i \in \{1, \ldots, n\},\ |X_i| \leqslant v_i,\ a.s.$$

then,

$$\forall t > 0, \mathbb{P}\left(\left|\sum_{i=1}^{n} X_i\right| \geqslant t\right) \leqslant 2\exp\left(-\frac{t^2}{2\sum_{i=1}^{n} v_i^2}\right).$$

Proof. See Section 2.6 in [18]. □

Setting $t = \sqrt{2\log\left(\frac{2}{\rho}\right)\sum_{i=1}^{n} v_i^2}$ leads to

$$\mathbb{P}\left(\sum_{i=1}^{n} X_i \geqslant \sqrt{2\log\left(\frac{2}{\rho}\right)\sum_{i=1}^{n} v_i^2}\right) \leqslant \rho,$$

and with $f = 3$ in order to match the expression (2.1). Now, it becomes

$$\mathbb{P}\left(\sum_{i=1}^{n} X_i \geqslant 3 \times l_\rho \times T_{RSS}\right) \leqslant \rho.$$

Hoeffding approach only takes into account the fact that random variables are bounded. However, here the information that features are uniform random variables is also available. This information will be used to find a tighter upper bound for the deviation of a sum of uniform random variables. As a result, a lower value for the coefficient f is obtained.

2.3.2 Chernov approach to improve the bound for a sum of uniform random variables

As it has just been mentioned, the Hoeffding approach is solely based on the support of the distribution involved in the deviation inequality. To improve such an upper bound, this distribution has to be considered more carefully. To this end, the well-known Cramér-Chernov bounding method is available. Such an approach is based on the following inequality derived from Markov's inequality and valid for any real random variable W and any real $\lambda > 0$,

$$\forall t > 0,\ \mathbb{P}\left(W - \mathbb{E}[W] \geqslant t\right) \leqslant e^{-\lambda t}\mathbb{E}\left[e^{\lambda(W-\mathbb{E}[W])}\right].$$

Thus, the method consists in optimizing this upper bound with respect to $\lambda > 0$ in order to exhibit a sharper deviation inequality.

The following proposition is stated and proved:

Proposition 2. *Let $v_1, \ldots, v_n > 0$, if $X_1, \ldots, X_n$ are independent random variables such that*

$$\forall i \in \{1, \ldots, n\},\ X_i \sim \mathcal{U}([-v_i, v_i]),$$

then,

$$\forall t > 0, \mathbb{P}\left(\left|\sum_{i=1}^{n} X_i\right| \geqslant t\right) \leqslant 2 \inf_{\lambda>0} \left\{\exp\left(\phi(\lambda, t)\right)\right\}$$

where the function ϕ is defined for any $\lambda, t > 0$ by

$$\phi(\lambda, t) = \sum_{i=1}^{n} \log\left(\frac{e^{\lambda v_i} - e^{-\lambda v_i}}{2\lambda v_i}\right) - \lambda t. \tag{2.2}$$

Proof. Let $\lambda > 0$, applying Markov inequality to the positive random variable $\exp\left(\lambda \sum_{i=1}^{n} X_i\right)$ gives the following upper bound on the probability that the sum of uniform independent random variables deviates more than $t > 0$

$$\mathbb{P}\left(\sum_{i=1}^{n} X_i \geqslant t\right) \leqslant \frac{\mathbb{E}[e^{\lambda \sum_{i=1}^{n} X_i}]}{e^{\lambda t}}.$$

Thus, using the symmetry of the uniform distribution, for any $t > 0$,

$$\mathbb{P}\left(\left|\sum_{i=1}^{n} X_i\right| \geqslant t\right) \leqslant 2\exp\left(\phi(\lambda, t)\right).$$

The upper bound is valid for any value of $\lambda > 0$ and the announced result follows by taking the infimum according to λ.

□

The optimization of ϕ function with respect to λ is highly related to the input features balance. Next, a way to characterize this dependency is detailed.

2.3.3 Dependency on the features balance

The aim of this section is to provide details on how to determine the value of λ which is obtained from a function minimization in the upper bound previously presented.

A concept of balance between input features is introduced. This balance represents the discrepancy between the uniform distributions parameters: if all uniform random variables have the same parameters, it means a perfect balance between tolerance bounds. Otherwise, one of the random variables within the sum may have a much larger support set than others and it leads to imbalance between tolerance bounds. The upper bound result from Proposition (2) is taken. The idea is to bound from above this result by introducing a specific term that identifies the influence of the balance within $v_1, \ldots, v_n$.

The focus is on the sum of logarithms in the function ϕ given in (2.2) that can be rewritten as

$$\sum_{i=1}^{n} \log\left(\frac{e^{\lambda v_i} - e^{-\lambda v_i}}{2\lambda v_i}\right) \tag{2.3}$$

$$= \lambda \sum_{i=1}^{n} v_i + \sum_{i=1}^{n} \log\left(\frac{1 - e^{-2\lambda v_i}}{2\lambda v_i}\right)$$

$$= n \log\left(\frac{1 - e^{-2\lambda \bar{v}}}{2\lambda \bar{v}}\right) + S_\lambda \tag{2.4}$$

with S_λ defined as follows

$$S_\lambda = \sum_{i=1}^{n} \left(\log\left(\frac{1 - e^{-2\lambda v_i}}{2\lambda v_i}\right) - \log\left(\frac{1 - e^{-2\lambda \bar{v}}}{2\lambda \bar{v}}\right)\right).$$

The term S_λ quantifies the imbalance between uniform distributions parameters. In the next two propositions, results are proposed about the upper bound on the probability that the sum of uniform independent random variables deviates from its expected value.

Proposition 3. *Let* $v_1, \dots, v_n > 0$ *and the mean* $\bar{v}$ *be defined as*

$$\bar{v} = \frac{1}{n} \sum_{i=1}^{n} v_i.$$

If $X_1, \dots, X_n$ *are independent random variables such that* $\forall i \in \{1, \dots, n\}$, $X_i \sim \mathcal{U}([-v_i, v_i])$ *then,*

$$\forall t > 0,\ \mathbb{P}\left(\left| \sum_{i=1}^{n} X_i \right| \geqslant t \right) \leqslant 2 \exp\left(\psi(\lambda_0, t)\right).$$

where for any $\lambda, t > 0$

$$\psi(\lambda, t) = -\lambda t + \lambda n \bar{v} + n \log\left(\frac{1 - e^{-2\lambda \bar{v}}}{2\lambda \bar{v}} \right) + \lambda \sum_{i=1}^{n} |v_i - \bar{v}|$$

and λ_0 *is such that*

$$\frac{\partial \psi(\lambda_0, t)}{\partial \lambda} = 0$$

For a set of tolerance bounds $v_1, \dots, v_n > 0$ and a fixed probability ρ, t verifies $\psi(\lambda_0, t) = \rho$. The value of interest t is obtained by inversion with respect to t of the function ψ. With this expression, the balance within $v_1, \dots, v_n$ appears via $\sum_{i=1}^{n} |v_i - \bar{v}|$. Indeed, this term is large for unbalanced values $v_1, \dots, v_n$ and small otherwise. Next, Propostion 3 is proven.

Proof. Let function h be defined as

$$\forall x > 0, \quad h(x) = \log\left(\frac{1 - e^{-x}}{x} \right).$$

This function is $\frac{1}{2}$-Lipschitz continuous. The proof is detailed below.

Proof. The definition of Lipschitz continuity of a function $f : \mathbb{R} \to \mathbb{R}$ is recalled. Let $L > 0$, if the function is such that

$$\forall x, y \in \mathbb{R},\ |f(x) - f(y)| \leqslant L\,|x - y|\,.$$

Then, f is said to be Lipschitz continuous with constant L. In this appendix, the previously claimed Lipschitz continuity of function h defined by

$$\forall x > 0, \quad h(x) = \log\left(\frac{1 - e^{-x}}{x}\right)$$

is proven with constant $L = 1/2$.

The first and second derivative functions of h are easily obtained as

$$\begin{aligned} \forall x > 0, \quad h'(x) &= \frac{e^{-x}}{1 - e^{-x}} - \frac{1}{x}, \\ h''(x) &= \frac{1 + (1 + x^2)e^{-x}}{x^2(1 - e^{-x})^2}. \end{aligned}$$

Since $h''(x) \geqslant 0$ for any $x > 0$, then h' is an non decreasing function. Moreover, $h'(x)$ tends to $-1/2$ when $x \to 0^+$ and to 0 when $x \to +\infty$. Thus, the conclusion is that $|h'(x)| \leqslant \frac{1}{2}$ and finally that h is Lipschitz continuous with $L = 1/2$ by a straightforward integration argument. □

The Lipschitz property for the function h is verified, therefore

$$\forall x, y > 0, \quad |h(x) - h(y)| \leqslant \frac{1}{2}\,|x - y|\,. \tag{2.5}$$

This inequality is applied for $x = 2\lambda v_i$ $\forall i \in \{1, \ldots, n\}$ and for $y = 2\lambda\bar{v}$ and sum the terms to obtain

$$S_\lambda \leqslant \lambda \sum_{i=1}^{n} |v_i - \bar{v}|\,.$$

The announced result follows from this upper bound on S_λ in the equation (2.4). □

In the previous proposition, the balance ratio of the v_i was quantified through the absolute values $|v_i - \bar{v}|$. It is natural to consider also the variance to this end and this is the purpose of the next proposition.

Proposition 4. *Let $v_1, \ldots, v_n > 0$, the mean $\bar{v}$ and the variance $\overline{\sigma}_v^2$ be defined as*

$$\bar{v} = \frac{1}{n}\sum_{i=1}^{n} v_i \quad \text{and} \quad \overline{\sigma}_v^2 = \frac{1}{n}\sum_{i=1}^{n}(v_i - \bar{v})^2.$$

If $X_1, \ldots, X_n$ are independent random variables such that $\forall i \in \{1, \ldots, n\}$, $X_i \sim \mathcal{U}([-v_i, v_i])$ then,

$$\forall t > 0,\ \mathbb{P}\left(\left|\sum_{i=1}^{n} X_i\right| \geqslant t\right) \leqslant 2\exp\left(\tilde{\psi}(\lambda_0, t)\right).$$

where for any $\lambda, t > 0$

$$\tilde{\psi}(\lambda, t) = -\lambda t + \lambda n\bar{v} + n\log\left(\frac{1 - e^{-2\lambda\bar{v}}}{2\lambda\bar{v}}\right) + \frac{n\lambda^2\overline{\sigma}_v^2}{2}$$

and λ_0 is such that

$$\frac{\partial\tilde{\psi}(\lambda_0, t)}{\partial\lambda} = 0$$

As in Proposition 3, tolerance bounds $v_1, \ldots, v_n > 0$ and a fixed probability ρ lead to a value t obtained by inversion with respect to t of the function $\tilde{\psi}$. The proof of Proposition 4 is as follows.

Proof. The Lipschitz continuity of h ensures the following inequality (see for example Lemma 1.2.3 in [19] for a proof of this result):

$$\forall x, y > 0, \quad |h(x) - h(y)| \leqslant \frac{L}{2}\|x - y\|^2.$$

This result applied to $x = 2\lambda v_i$ $\forall i \in \{1, \ldots, n\}$ and for $y = 2\lambda\bar{v}$ gives

$$\forall\lambda, v_1, \ldots, v_n > 0, \quad |h(2\lambda v_i) - h(2\lambda\bar{v})| \leqslant \frac{\lambda^2}{2}\|v_i - \bar{v}\|^2$$

and finally, since $S_\lambda = \sum_{i=1}^{n}(h(2\lambda v_i) - h(2\lambda\bar{v}))$,

$$S_\lambda \leqslant \frac{n\lambda^2\overline{\sigma}_v^2}{2}.$$

The announced result follows from this upper bound on S_λ in the equation (2.4). □

2.4 Applications

The first part of this section will describe how our upper bound behaves on different stack chains obtained from simulations. The second part will focus on a practical study on an industrial example of tolerance definition within an aircraft assembly.

2.4.1 Simulations

Tolerance design on an assembly example

First step is to simulate stack chains. Stack chains represented by features X_1, X_2, X_3, X_4, X_5 are randomly generated with a number of inputs $n = 5$. Their tolerance intervals values v_1, v_2, v_3, v_4, v_5 are also randomly generated between 1 and 5.

The aim is to assess on the output tolerance variability via tolerance intervals to be defined, and accepting an out-of-tolerance rate ρ. A stack chain is generated with tolerance inputs intervals bounds and traditional output results as following:

Table 2.1: Example of a stack chain characterization.

X_1	X_2	X_3	X_4	X_5	RSS	WC
$v_1 = 5$	$v_2 = 4$	$v_3 = 3$	$v_4 = 2$	$v_5 = 1$	7.4	15

The first approach is based on Monte Carlo methods. $N = 10^5$ observations are generated from $n = 5$ uniform distributions and are summed. The probability to be out of a given tolerance interval can therefore be asymptotically estimated and considered as a near theoretical result. The two following methods give an output interval bound according to the two approaches proposed in this chapter. The provided upper bound depends on the selected confidence level ρ. This level is the probability for the output feature to be out of the designed output tolerance interval. The higher the confidence level, the wider the tolerance interval is. Indeed, if more values to be out of tolerance are allowed, the output tolerance interval should be broader. Figure 2.3 illustrates the results obtained from Monte Carlo draws and from the methods detailed in this chapter.

Among the three methods, Figure 2.3 shows that the Lipschitz and Quadratic approaches give a looser upper bound than the Chernov method.

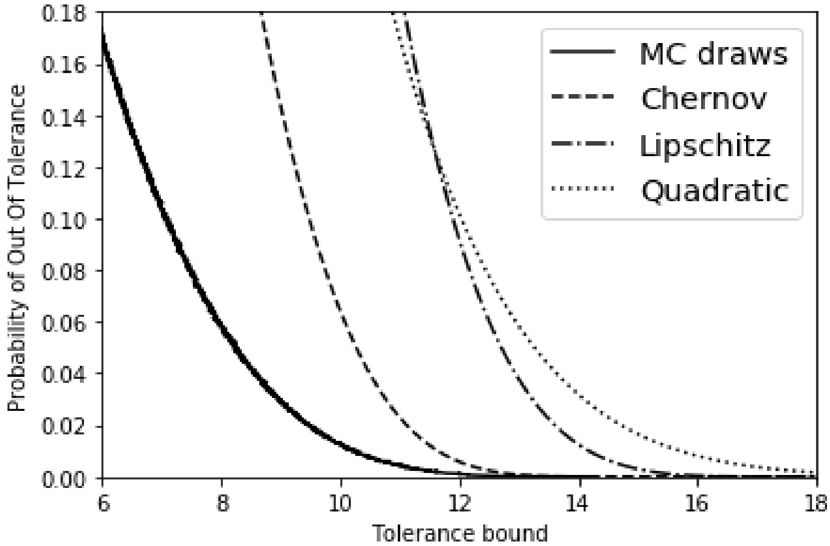

FIGURE 2.3: Example of the behavior of the probability to be out of tolerance with respect to the value of the output tolerance bound for the discussed approaches.

The benefits of the methods proposed in this chapter is that they do not require Monte Carlo draws, nor asymptotic estimation of the probability to be out of tolerance. Indeed, the provided bounds offer theoretical non asymptotic guarantees and eliminate any risk of rare events that Monte Carlo methods would not generate. Moreover, for large assemblies, the number of Monte Carlo draws needed to obtain a sufficiently sharp result would grow with the number of input features in the assembly. The discussed formula are closed and directly usable in practice and cheaper to compute than Monte Carlo simulations.

Influence of the assembly geometry

In order to represent the balance within a stack chain, previous sections introduced the following term

$$S_\lambda = \sum_{i=1}^{n} \left(\log\left(\frac{1 - e^{-2\lambda v_i}}{2\lambda v_i} \right) - \log\left(\frac{1 - e^{-2\lambda \bar{v}}}{2\lambda \bar{v}} \right) \right).$$

In particular, taking arbitrarily parameter $\lambda = 1$ leads to

$$S_1 = \sum_{i=1}^{n} \left(\log \left(\frac{1 - e^{-2v_i}}{2v_i} \right) - \log \left(\frac{1 - e^{-2\bar{v}}}{2\bar{v}} \right) \right).$$

This quantity can be used as an indicator of the balance of the stack chain. Indeed, the more balanced the stack chain is, the lower the value is and vice versa.

As mentioned in the previous part, one of our main issue is to take into account the traditional RSS result and the balance of the stack chain. This explains why hereafter the choice is to display the coefficient f with respect to some balance indicator such as S_1 or other dispersal measures within input features.

First, in Figure 2.4, the results are showed for the coefficient f obtained from a Monte Carlo simulation of uniform distributions with 2×10^5 drawn observations. Next, the coefficient f is displayed according to the Chernov methodology as detailed in Proposition 2. Finally, it shows that bounding by Lipschitz and Quadratic approaches directly depend on some balance factor.

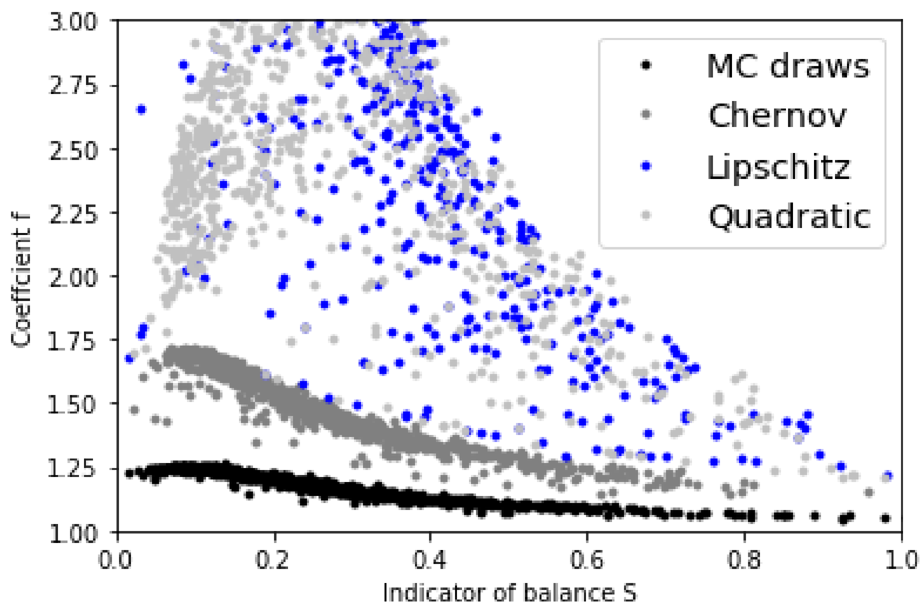

FIGURE 2.4: Link between the coefficient f and the balance factor S_1 with parameter $\rho = 0.05$.

An almost linear behavior of the result with respect to the balance factor S_1 is observed for the Monte Carlo approach and for the Chernov methodology. This factor S_1 seems to be a relevant indicator to characterize f coefficient. As expected and due to the upper bounds defined in these methods, both Lipschitz and Quadratic approaches give results much more conservative. Still the Quadratic approach is more accurate for small values of S_1. This is explained by the fact that the result with Quadratic approach in Proposition 4 takes into account the variance of input feature bounds. For a small S_1, input features are balanced and variance is a more regular control quantity for the structure of the stack chain than the sum of absolute deviations around the mean $\bar{v}$ introduced in Proposition 3.

2.4.2 Case study

In this part, the focus is on industrial practices at Airbus. First, the example of an assembly from an aircraft is taken and results from the methodology proposed in this chapter are showed. Then, the common process of tolerance definition at Airbus is detailed and explanation about how it is related to the approaches presented in the chapter are provided. Finally, all stack chains are represented in a real aeronautical product perimeter according to the balance factor.

An assembly example

The assembly in Figure 2.5 is related to a generic frame misalignment for an Airbus aircraft.

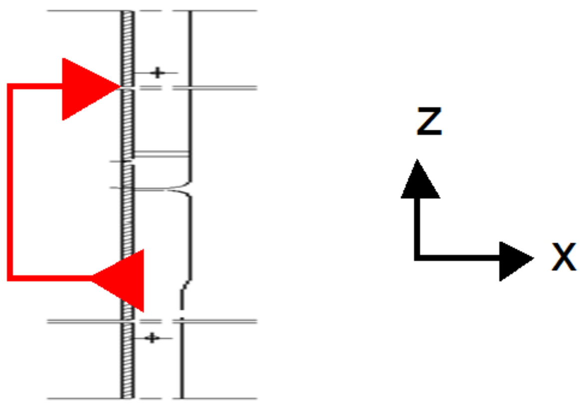

FIGURE 2.5: Example of vertical frame misalignment with respect to the last rigid point.

Table 2.2 gives the stack chain data of this requirement. Tolerance bounds value have been modified.

Name of the contributor	**Tolerance interval**
Frame 1	± 1
Frame 2	± 0.5
Process tolerance	± 0.25
Process tolerance	± 0.23
Process tolerance	± 0.2
Process tolerance	± 0.2
Process tolerance	± 0.15
Process tolerance	± 0.13
Process tolerance	± 0.1
Process tolerance	± 0.09

TABLE 2.2: Stack chain: frame misalignment - last rigid point.

It involves 10 input features in the assembly and tolerance data are scaled and unit free. Table 2.3 provides traditional tolerancing worst case and RSS results. The application of the different methods proposed in this chapter gives the results depicted in Figure 2.6.

Worst Case result	RSS result
± 2.85	± 1.23

TABLE 2.3: Result: frame misalignement - last rigid point.

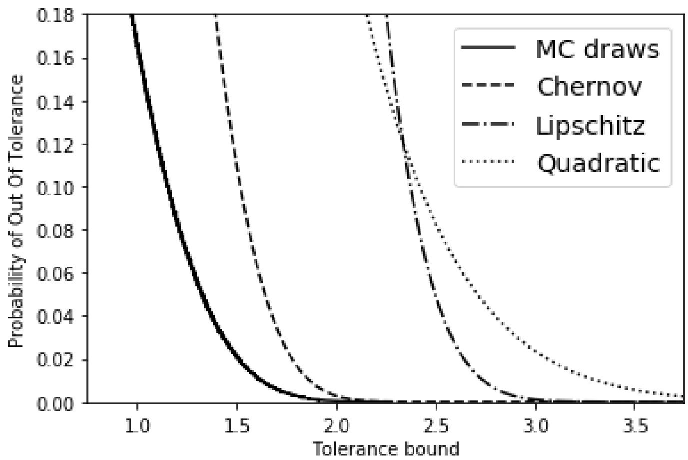

FIGURE 2.6: Real case study of a bound value according to the confidence level.

Results for the real case study are very similar to simulated data conclusion. Indeed, Lipschitz and Quadratic analytical approaches still do not give a sharp bound result compared to the Monte Carlo and Chernov approaches. In practice, the more accurate method in order to design an output tolerance should be the Chernov one. Note that the values for the probability of Out Of Tolerance probability considered are generally the lower percentages on the y-axis as the aim in practice is to limit the scrap rate and in Airbus practices, the reference value is often 0.27%. This reference value is displayed in the Figure 2.7 which focuses on the percentage below 1% for Out Of Tolerance probability. The limits defined by RSS and WC methods applied to the assembly example are also displayed.

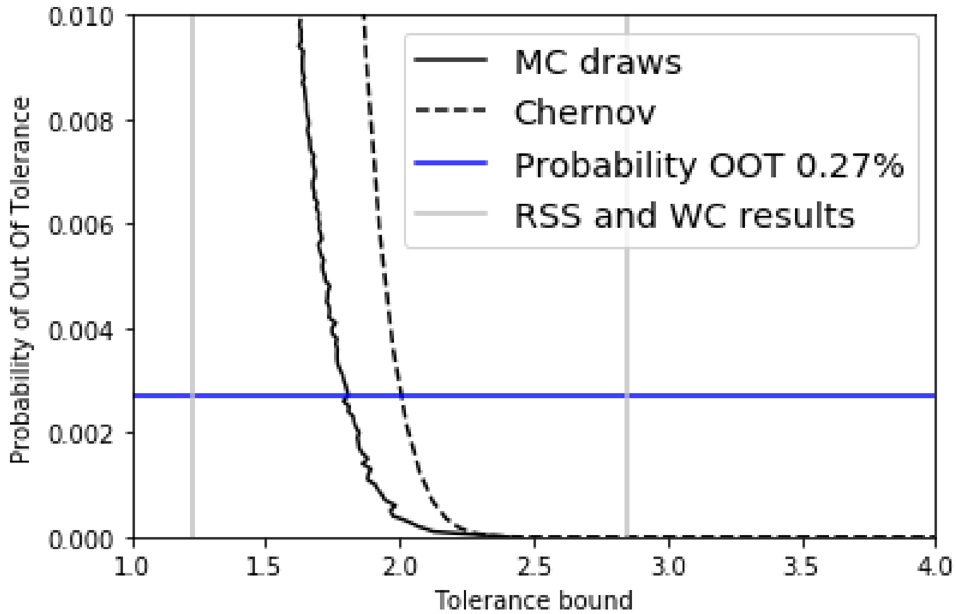

FIGURE 2.7: Real case study of a bound value according to the confidence level zoomed on relevant values for Airbus.

Industrial practices: Airbus

Among the methods used by Airbus to define a tolerance in the design phase, one of them is an approximation of Monte Carlo simulation data under uniform assumption for contributors distribution and a disproportion parameter. As for the Gaussian case, quantile at 0.27% are observed on Monte Carlo simulations and a linear regression with respect to the factor parameter is carried out to obtain the result. For a set of tolerance bound $v_1, \dots, v_n > 0$ for an input features balance ratio D, this rule gives an output feature tolerance interval $[-T_{Airbus}, T_{Airbus}]$ defined as:

$$T_{Airbus} = \beta \times (-0.56D + 1.04) \times T_{RSS} \tag{2.6}$$

with T_{RSS} as defined in previous parts,

$$\forall v_1, \ldots, v_n > 0, \quad D = \frac{\max_i(v_i) - \bar{v}}{\sum_{i=1}^{n} v_i}$$

and $\beta > 0$ a coefficient. In practice and without claim of universality, Airbus industrial approach takes $\beta = 1.6$ as a relevant value for business and in order to ensure continuity with former practices.

This D factor measures how far from the mean is the main contributor of the stack chain and has the advantage of being understandable. This quantity is highly correlated to the term S_1 previously introduced as shown on the Figure 2.8:

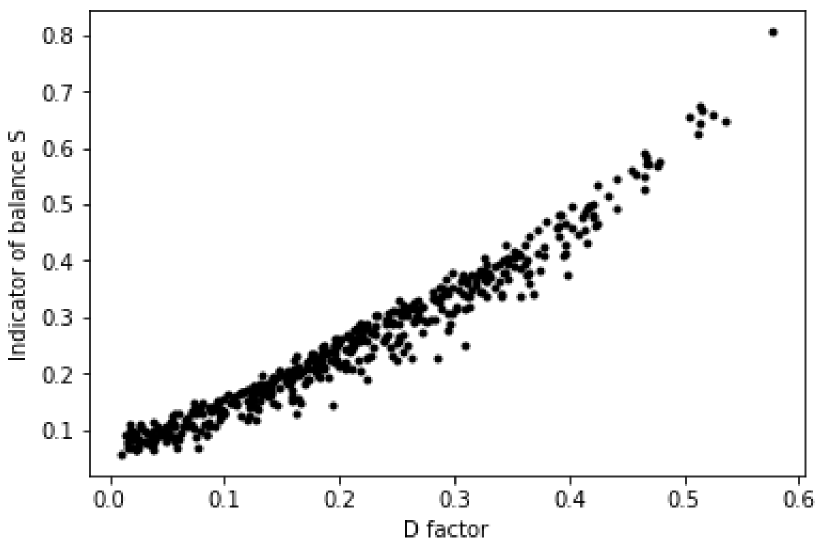

FIGURE 2.8: Correlation between the term S_1 and the balance factor D.

With this definition, a high D still implies an unbalanced stack chain. Conversely, a small value of B means a balanced between stack chain inputs. Figure 2.9 shows a few examples of this factor with the respect to the stack chain structure.

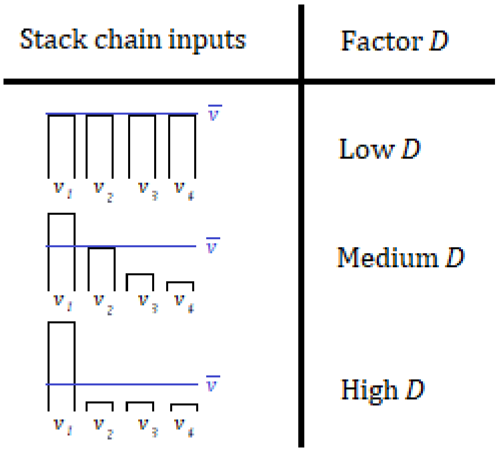

FIGURE 2.9: Stack chain structure and balance factor D.

In practice, 70% of the stack chains have a D value between 0.10 and 0.36. The Figure 2.10 details the repartition of the coefficient value for Airbus stack chains.

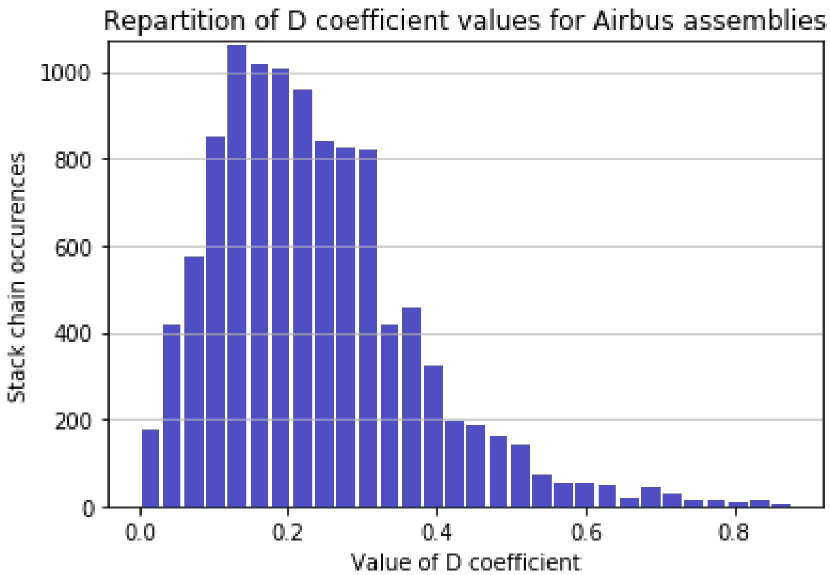

FIGURE 2.10: Repartion of D values for Airbus assemblies.

In the industrial example of the frame misalignment. The probability to be out of tolerance for the output feature is set at $\rho = 0.0027$, which corresponds to the acceptable 0.27% out of the interval from the 6σ methodology. Table 2.4 summarizes the tolerance interval obtained for the frame misalignment. Three results are displayed: the Monte Carlo approach with 200000 drawn observations for each input feature, the Chernov approach proposed in this chapter with $\rho = 0.0027$ and the industrial practice presented in (2.6).

Method	Monte Carlo $\rho = 0.27\%$	Chernov $\rho = 0.27\%$	Industrial practice
Tolerance interval	± 3.56mm	± 4.01mm	± 3.53mm

TABLE 2.4: Tolerance interval results according to the different approaches.

For a level $\rho = 0.27\%$, the result from the industrial rule is very close to the value observed on Monte Carlo simulations. The Chernov method gives a more conservative result but ensures a precise probability ρ to be out of the interval for the output feature.

Performance of the different approaches on industrial cases

Focusing on a real sample of aeronautical assemblies, all stack chains have been analyzed in order to obtain the value of the f coefficient times l_ρ for $\rho = 0.27\%$, according to the different methodologies. The Airbus rule that can be used for decision helping is also displayed. This is an approximation for the selected confidence level.

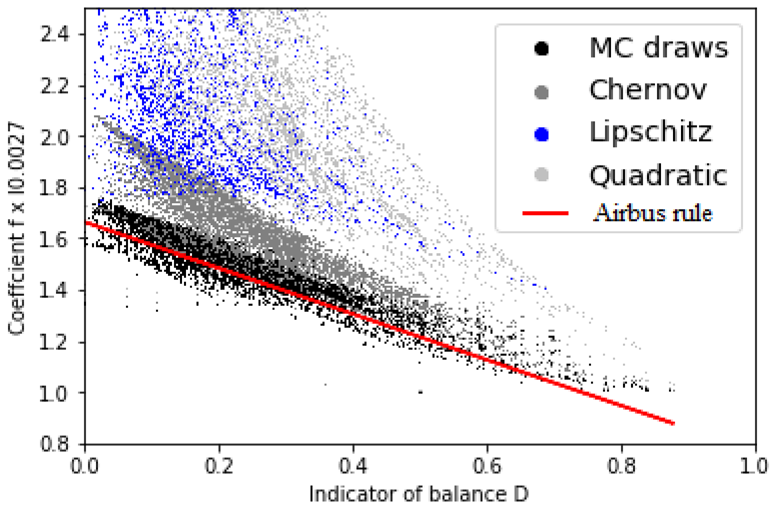

FIGURE 2.11: Link between the term introduced and the balance factor D.

The same trend that for simulated data is retrieved: Better result for Chernov approach and linearity in D. For Monte Carlo simulations, a similar linear behaviour is observed with respect to the balance factor D. The airbus rule seems to be a good approximation for evaluate the coefficient f times l_ρ with a confidence level $\rho = 0.27\%$.

2.5 Conclusion

Robust approaches are proposed for tolerance definition in the design phase allowing the management of confidence level. From known input tolerance intervals of an assembly and for a selected confidence level, the output tolerance interval can be determined. The result will be robust against poor or unknown industrial capabilities because uniform distributions on tolerance intervals are assumed for input features.

The Chernov method is particularly accurate and gives an output tolerance interval result close to the reality, tight enough to be industrially relevant, and ensures also the selected probability as confidence level. A balance factor is also provided. This factor is strongly related to how tight an interval should be according to the disproportion of the stack chain. An almost linear behavior of the result is obtained from the Chernov methodology with respect to this balance factor.

Future directions of this work would be to consider more adversarial distributions for input features. For instance, bimodal distributions or truncated distributions could be studied in order to hedge against industrial practices with the induced bias of machine or thrust effect. Additionally, extension to this work in more general nonlinear settings for the model output Y as a function of the input features could be thought of. Indeed, if the model is nonlinear but the variance of each input component is finite, [20] suggests that one could also control the model output as proposed in our study.

Bibliography

[1] Ambre Diet, Nicolas Couellan, Xavier Gendre, and Julien Martin. A Chernov bound for robust tolerance design and application. *The International Journal of Advanced Manufacturing Technology*, 111(11-12):3571–3581, December 2020.

[2] 9780791822234: Dimensioning and Tolerancing ANSI Y14.5M-1994 - AbeBooks - American Society Of Mechanical Engineers (ASME): 0791822230. Library Catalog: www.abebooks.fr.

[3] ISO 1101:2017. Library Catalog: standards.iteh.ai.

[4] Kenneth W. Chase and William Henry Greenwood. Design issues in mechanical tolerance analysis. *Manufacturing Review, ASME*, 1:50–59, 1988.

[5] Björn Heling, Thomas Oberleiter, Benjamin Schleich, Kai Willner, and Sandro Wartzack. On the Selection of Sensitivity Analysis Methods in the Context of Tolerance Management. *Journal of Verification, Validation and Uncertainty Quantification*, 4(1):011001, March 2019.

[6] W. H. Greenwood and K. W. Chase. Worst Case Tolerance Analysis with Nonlinear Problems. *Journal of Engineering for Industry*, 110(3):232–235, August 1988.

[7] A. Bender. Benderizing tolerances-a simple practical probability method of handling tolerances for limit-stack-ups. *Graphic Science*, 17, 1962.

[8] Victor J Skowronski and Joshua U Turner. Using Monte-Carlo variance reduction in statistical tolerance synthesis. *Computer-Aided Design*, 29(1):63–69, January 1997.

[9] Hoo-Gon R. Choi, Man-Hee Park, and Erik Salisbury. Optimal Tolerance Allocation With Loss Functions. *Journal of Manufacturing Science and Engineering*, 122(3):529–535, August 2000.

[10] Maurice Pillet, Daniel Duret, and Alain Sergent. Weighted inertial tolerancing. March 2003.

[11] Laurent Leblond and Maurice Pillet. Conformity and statistical tolerancing. *International Journal of Metrology and Quality Engineering*, 9:1, 2018.

[12] Paul J. Drake, editor. *Dimensioning and tolerancing handbook*. McGraw Hill, New York, 1999.

[13] S. Charles Liu and S. Jack Hu. Variation Simulation for Deformable Sheet Metal Assemblies Using Finite Element Methods. *Journal of Manufacturing Science and Engineering*, 119(3):368–374, August 1997.

[14] Yue Wang, Lei Li, Nathan W. Hartman, and John W. Sutherland. Allocation of assembly tolerances to minimize costs. *CIRP Annals*, 68(1):13–16, 2019.

[15] Frank Killmann and Elart von Collani. A Note on the Convolution of the Uniform and Related Distributions and Their Use in Quality Control. *Economic Quality Control*, 16(1), January 2001.

[16] Cédric Villani. *Topics in optimal transportation*. Number v. 58 in Graduate studies in mathematics. American Mathematical Society, Providence, RI, 2003.

[17] Wassily Hoeffding. Probability Inequalities for Sums of Bounded Random Variables. *Journal of the American Statistical Association*, 58(301):13–30, March 1963.

[18] Stéphane Boucheron, Gábor Lugosi, and Pascal Massart. *Concentration inequalities: a nonasymptotic theory of independence*. Oxford University Press, Oxford New York, NY, paperback edition, 2016. OCLC: 947132616.

[19] Yurii Nesterov. *Introductory Lectures on Convex Optimization: a Basic Course*. Springer, New York, NY, 2013. OCLC: 1066186560.

[20] Antoine Marchina. *Concentration inequalities for functions of independent random variables*. PhD thesis, Université Paris-Saclay, 2017.

Chapter 3

Tolerance model re-adjustment from in-service data

In this chapter, the use of linear tolerance model is discussed and approaches for model enhancement are proposed. Data availability, model uncertainties and integration effect are considered. Finally, implementations of proposed methodologies are detailed in examples.

3.1 About linear tolerance model and alternatives

The tolerance model represents the link between the inputs of the assembly, the contributors and the output (the top level requirement). An illustration of this notion for a number p of contributors is illustrated in Figure 3.1.

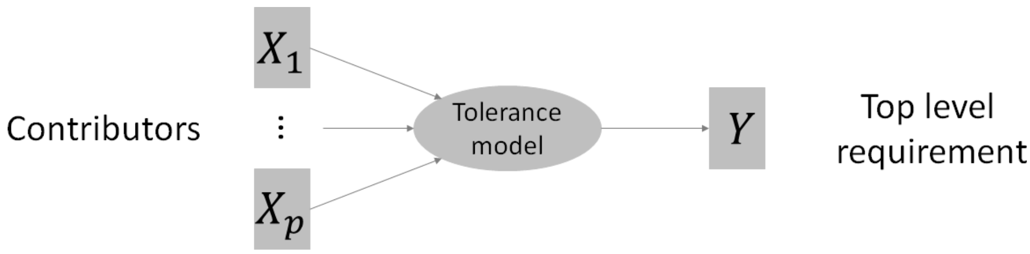

FIGURE 3.1: General context: assembly tolerance model

The linear model is widely used due to its easiness. In the literature, several other tolerancing models are available and each one has benefits and limitations. These models often involve considerations of geometrical parameters that make the storage and use of the tolerance model in the industrial context more complex than a linear approach. The vectorial tolerance approach presented in [1] needs to determine the contact surfaces between the components during the assembly process and contact parameters needs to be determined in order to discuss about the assembly output tolerance. The Technologically and Topologically Related Surfaces (TTRS, [2]) also relies on geometrical surfaces identifications and associations. The Small displacement torsor method ([3]) is a basic concept used in several improved approaches. Under the assumption of rigid body, rotations and translations vectors are considered to define a torsor that will be used to solve a general problem fitting a geometrical surface model to a set of points. The Tolerance-Map (T-Map) model is based on a hypothetical Euclidean point-space hypothesis. All variations possibilities in size and shape of this space are explored in order to define the variation range of features through a point to point analysis. Details about this approach are available in [4] and [5]. The Direct Linearization Method (DLM) detailed in [6] is a generalization of vectorial tolerancing approaches. The approach is based on a first order Taylor expansion and kinematic constraint equations. A comparison with Monte Carlo approach is discussed in [7].

More recently, the skin model and its discrete version skin model shape modelling ([8], [9]) have been considered. Such a model takes into account the physical interface between a part and its environment in order to consider surfaces to which the tolerances apply. A better approach to take into account surface defects and best represent reality is brought by skin model. However, this is more complex and difficult to manipulate and to store such a type of model. It is also less suitable for the propagation of the model in various industrial tools, which is why the linear model has been chosen as the basis for tolerance studies.

In [10], the interest and the impact for tolerance analysis of the linearization of the model to link features in contact with each other is studied. Linearization strategies are detailed and discussed for non-linear equations from the behavior model constraints,taking into account geometrical considerations.

In addition to its interest for tolerance analysis approaches detailed in [10], the linear model also presents advantages for the industrial use. First, this is an historical choice in the early stages of tolerancing, suited for the one dimension approach. The linear model has been generalized to 3D assemblies as an easy to understand model, facilitating communication between the different functions involved in tolerance management. It is also convenient for the use of the tolerance model in complex tools or software because linear approach is a basic component. The model parameters are easily logged into internal tools and allow tolerancing specialists to perform tolerancing studies and give all information for decision on dimension tolerances whatever the considered assembly. Moreover, the linear model presents an additive structure that is relevant for the decomposition between observed and non observed contributors in an assembly once tolerance models are in production and measurement data are considered. Finally, the use of a linear model for tolerancing is also acceptable if the small displacement hypothesis is valid as described in Section 3.2.

The linear approach for tolerance model is the one used in the sequel of this chapter. Other models treating non linear effects might be more effective from a certain perspective, but both industrial needs and management costs justify the consideration of the linear approach. To overcome the limitations of a tolerance model, strategies on different aspects where inconsistencies with reality are identified are considered. Solutions are proposed in Section 3.3 to correct the model in order to improve fitting to real data.

3.2 Linear approach framework - Initial tolerance model

3.2.1 Mathematical framework

Each assembly has an associated stack chain, composed of p contributors. These contributors are represented by features handled as random variables and assumptions on their distribution are made in order to define contributor's tolerance. The random variables associated to contributors are assumed to be independent. The assembly output is the top level requirement feature. This is the random variable associated to the dimension of interest that needs to be controlled in order to meet functional requirements of the final product. Each contributor impacts the top level requirement feature in its own way (Figure 3.2). A simple way to represent the relation between assembly inputs and output is through a first order model.

Assuming a linear link between contributors and top level requirement and thanks to 3D simulation, it is possible to compute coefficients of each contributor within a linear model. Let denote:

- $\alpha_1, \ldots, \alpha_p \in \mathbb{R}$, the influence coefficients,
- $X_1, \ldots, X_p$, representing contributors by real random variables,
- Y, the top level requirement as a real random variable.

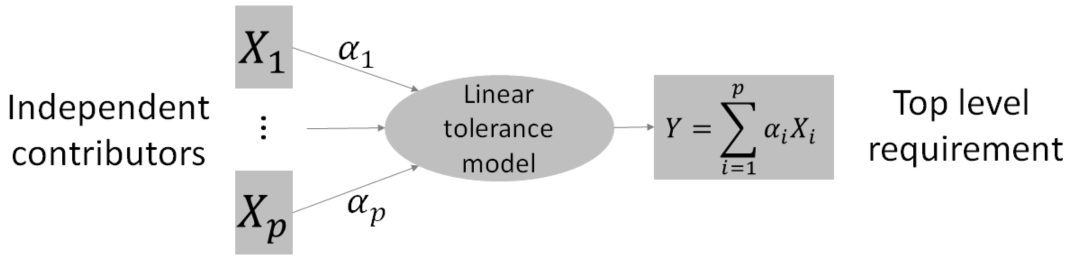

FIGURE 3.2: Linear tolerance model

Thus, the linear model leads to the relation

$$Y = \sum_{j=1}^{p} \alpha_j X_j. \tag{3.1}$$

This linear approach is justified through the Taylor expansion of order one. The relation function between assembly inputs and output is given by a smooth function f. Let us denote the small displacement vector $h \in \mathbb{R}^p$ and the sequence of random variables $X = (X_1, \ldots, X_p)$.

The Taylor expansion of f applied to X has the form

$$\forall h \in \mathbb{R}^p, f(X+h) = f(X) + \nabla f(X)^T h + o\left(||h||\right)$$

The first order term is the linear term. The second and above order terms represent the non linear effects not taken into account in a linear tolerance model and not explained by the first order expansion. We can choose to neglect these terms with respect to the first order linear term if we assume the small displacement hypothesis. Note that in the case of the linear model, the gradient is $\nabla f(X) = (\alpha_1, \ldots, \alpha_p)$.

This approach is still valid in the case with random perturbations instead of a fixed small displacement h. We can consider a random displacement H as a centered random variable in $\mathbb{R}^p$. The notation $o\left(||h||\right)$ becomes represented by $o_p(H)$,

$$f(X+H) = f(X) + \nabla f(X)^T H + o_p(H)$$

where $o_p(H)$ notation is detailed in [11], Section 2.2 ($o_p(1)$ denoting a sequence that converges to 0 in probability).

The term $o_p(H)$ represents residual effects that are not considered in the linear tolerance model. The linear coefficient are thus related to the derivatives values of the function linking assembly inputs contributors to the output top level requirement.

Through this notation and under the assumption that variations are reasonable, a linear representation is given by the first derivatives of the function f in order to obtain the linear coefficients $\alpha_1, \ldots, \alpha_p$ from Equation (3.1).

At the scale of aerostructures, it is relevant to consider these geometric variations as small displacements. Indeed, in the case of an aircraft, it is often a question of assembly involving components with dimension order of a meter, whereas the tolerances are generally expressed in millimeters.

3.2.2 Linear tolerance model on a simple assembly example

A simple example of a 2 dimensions assembly is displayed in Figure 3.3. The blue arrow represents the contributor feature and the green arrow represents the top level requirement feature, which is the total height of the assembly. We assume that the large grey part and the black stem have perfect known dimensions. The contributor is positioned according to d_2 and d_3 as described on the figure. Its nominal value is d_1.

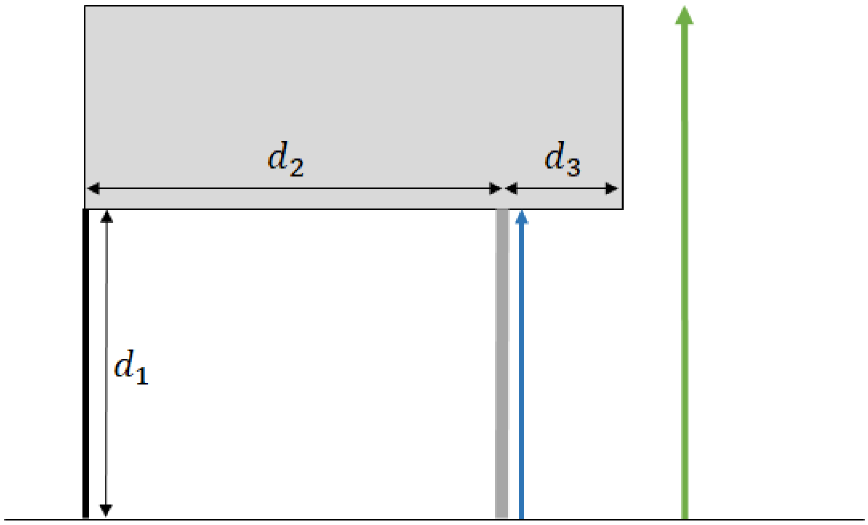

Figure 3.3: Simple assembly example

The deviation of the contributor nominal dimensions is represented by h in Figure 3.4. This variation in dimension leads to the formation of an angle θ represented in yellow on the figure. The impact on the higher level requirement is represented by the dimension Δ. This is the dimension that needs to be estimated.

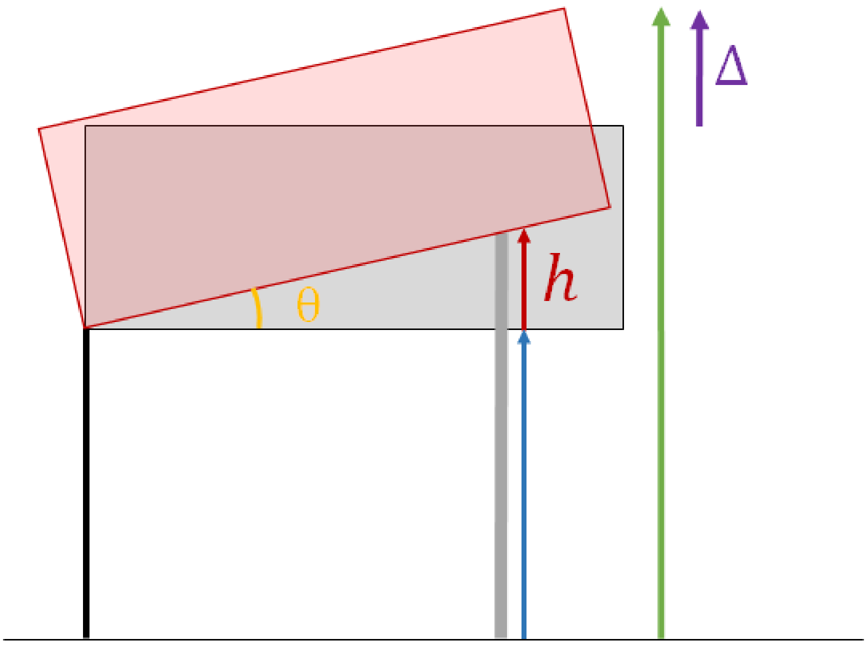

Figure 3.4: Simple assembly example with a variation

Within the framework of this simple assembly, it is possible to directly calculate the dimension Δ via the trigonometry formulas on the rotation of angle θ.

The first step is to calculate the angle θ.

$$\theta = \arctan\left(\frac{h}{d_2}\right)$$

Then, the dimension Δ is expressed in terms of θ, d_2 and d_3.

$$\Delta = (d_2 + d_3)\sin(\theta)$$

Finally, we can now express the deviation of the top level requirement Δ according to the deviation h of the contributor.

$$\Delta = (d_2 + d_3)\sin\left(\arctan\left(\frac{h}{d_2}\right)\right) \tag{3.2}$$

The trigonometry formulas allow to rewrite Δ as

$$\Delta = (d_2 + d_3)\left(\frac{\frac{h}{d_2}}{\sqrt{\left(\frac{h}{d_2}\right)^2 + 1}}\right).$$

The Taylor expansion to order 1 of this expression in the neighborhood of 0 gives

$$\Delta = (d_2 + d_3)\left(\frac{h}{d_2}\right) + O\left(h^2\right) = \left(1 + \frac{d_3}{d_2}\right)h + O\left(h^2\right). \tag{3.3}$$

The graph in the Figure 3.5 shows the results for Δ with the exact evaluation of (3.2) and the linear approximation proposed in (3.3). The parameter for this application are $d_2 = 30$ and $d_3 = 70$. As expected, the linear approximation is very accurate for small values of h, which represents small variations in the dimension of the contributor involved in this assembly. This examples shows that if the deviation from the nominal value of the contributor is small enough, the linear tolerance model is a relevant model for tolerance management.

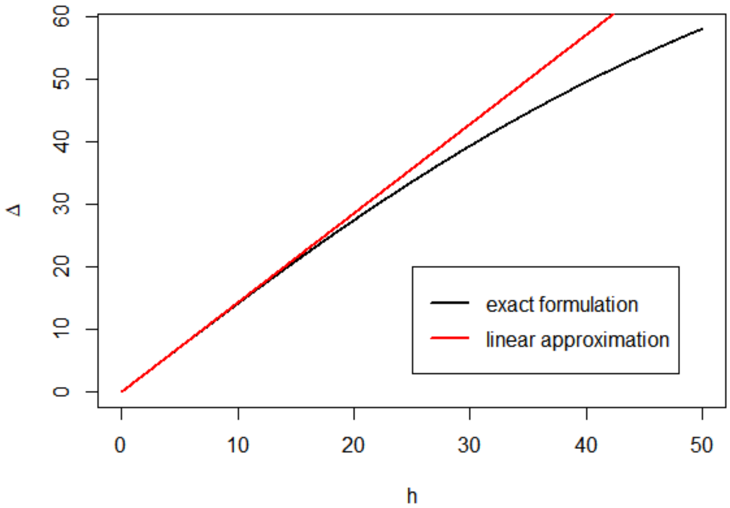

FIGURE 3.5: Exact Δ evaluation and linear approximation

As previously discussed, in an aeronautical context, the variations are often in millimeters while the size of an aerostructure is of the magnitude of tens of meters. Note that if d_2 and d_3 are expressed in meters, it would mean to focus on values of h below 0.01 on the Figure 3.5.

3.2.3 Data availability

Once a stack chain is used in production, some feedback measures become available. Unfortunately, it is common that not all contributors lead to available observations. Only a number m of contributors among the p contributors in the stack chain are measured, with $m \leq p$. We might also have observations for the top level requirement of the stack chain feature Y.

Industrial context

In the industrial context, there is a processing of measures in order to streamline measurement data. The defined methodology is in line with international standards. The main objective of the harmonization was and is to ensure consistent indicators used throughout design, preparation, execution and quality control process with respect to process performance. This implies validating the measurement data and then calculating process performance indicators or descriptive statistics. For instance, it is planned to verify the sample size, eliminate outliers, and perform multimodality and normality test. For each observed characteristic, a fit test is done to find the most appropriate distribution. Several distributions are considered and the associated parameters are estimated. Traditionally, process are designed to follow normal distributions but the fit test allows to verify or consider other common distributions. In the Airbus process, distributions available for testing are: Normal, Uniform, Triangular, Cauchy, Exponential, Lognormal, Gamma, Weibull, Minimum Extreme, Beta, Rayleigh and Folded Normal. The steps for the Airbus process performance measurement detailed in the internal document technical report [12] is summarized below. For a data set, carry out the following steps:

1. Statistical Representative Population:

   Decide on the number of observations required, depending on the type of analysis desired.

2. Outliers identification:

   With the common box-plot technique, identify all outliers from the data

3. Multimodality test:

   Provide if the points in the data set are unimodal or multimodal. The implementation of this test relies on Gaussian Mixture Models (GMM) and the Bayesian Information Criterion (BIC).

4. Normality test:

   Provide if the points in the data set comply with a normal distribution, through the Shapiro-Wilk test.

5. Goodness of fit:

   Identify the best distribution to fit the data through the Maximum Likelihood Estimation (MLE).

6. Descriptive statistics:

   Provide figures for business use.

7. Process Performance:

   Measure common performance indicators.

This process ensures the consistency and harmony of the data collected by Airbus teams in the various assembly plants. This processing on measurement data is a prerequisite for Airbus applications of the methodologies relying on measurement data analysis detailed in the sequel.

Empirical density estimation: benefits and disadvantages

Assuming to be able to store more than the distribution parameters, we can estimate observed features densities with kernel density estimators ([13]). In this case, we store all measurement data. Such a solution makes it possible to treat any type of distribution by freeing oneself from the parametric estimation of distributions with parameters. This is particularly interesting when the measurements do not verify the classical Gaussian hypothesis. For example, it is thus possible to take into account bimodal distributions sometimes encountered when two tolerance limits are required. In such a case, a normal distribution would tend to smooth out the information by fitting a centered distribution with a large variance, whereas it is in fact a distribution with two modes that represents a specific way of production.

The kernel estimation method is a way to fit a density on measurement data without considering a parametric model such as the maximum likelihood estimation (MLE) method or the method of moments. From an implementation point of view, the method of moments presents an interest which is to allow a readjustment of the estimated parameters of a distribution at each aggregation of measures in a data set. This works with storage limited to the parameters of distributions from a collection, but will not work with datasets classified as bimodal and which require the storage of all observations. It is precisely in the latter case that kernel density estimation is relevant.

The method of moments is a common way to estimate population parameters that can be expressed as expectations under a given distribution. The theoretical moments of the considered random variable are expressed as functions of the parameters to be estimated and then set equal to the sample empirical moments. The solution of this equation leads to the distribution parameters estimation. Most of the distributions used within an industrial context only needs the first and second order moments to estimate the few parameters that characterize them. The implementation advantage of this method is that empirical moments can be updated at each measurement. In this way, the computing cost is lower. Indeed, calculation done with a new measurement data only relates to the new observation and not all the sample population. Empirical moments for a sample of $n+1$ observations can be decomposed as a function of the empirical moment for the sample of n and the $n+1$-th observation. The following example shows the result for first and second order empirical moments for the mean and variance parameters.

For a sample with n observations associated to the random variable X_n, the empirical first order moment is the computed mean denoted as $\overline{X}_n$. When a new measure is available, the empirical mean is $\overline{X}_{n+1}$.

$$\begin{aligned}
\overline{X}_{n+1} &= \frac{1}{n+1}\sum_{k=1}^{n+1} X_k \\
&= \frac{1}{n+1}\left(\sum_{k=1}^{n} X_k + X_{n+1}\right) \\
&= \frac{1}{n+1}\left(n\overline{X}_n + X_{n+1}\right) \\
\overline{X}_{n+1} &= \overline{X}_n + \frac{1}{n+1}\left(X_{n+1} - \overline{X}_n\right)
\end{aligned}$$

A similar type of decomposition can be established for the second order empirical moment. The empirical variance is denoted as S_n^2.

$$\begin{aligned}
S_{n+1}^2 &= \frac{1}{n+1}\sum_{k=1}^{n+1}\left(X_k - \overline{X}_{n+1}\right)^2 \\
&= \frac{1}{n+1}\sum_{k=1}^{n+1}\left(\left(X_k - \overline{X_n}\right) + \left(\overline{X_n} - \overline{X}_{n+1}\right)\right)^2 \\
&= \frac{1}{n+1}\left[\sum_{k=1}^{n+1}\left(X_k - \overline{X_n}\right)^2\right] - \left(\overline{X}_{n+1} - \overline{X_n}\right)^2 \\
&= \frac{1}{n+1}\left[\sum_{k=1}^{n}\left(X_k - \overline{X_n}\right)^2\right] + \frac{1}{n+1}\left(X_{n+1} - \overline{X_n}\right)^2 - \left(\overline{X}_{n+1} - \overline{X_n}\right)^2 \\
S_{n+1}^2 &= \frac{n}{n+1}S_n^2 + \frac{1}{n+1}\left(X_{n+1} - \overline{X_n}\right)^2 - \frac{1}{(n+1)^2}\left(X_{n+1} - \overline{X_n}\right)^2
\end{aligned}$$

At the end, the empirical variance S_{n+1}^2 is expressed as a function of the empirical mean and variance of the sample with size n and the $n+1$-th observation, respectively $\overline{X}_n$, S_n^2 and X_{n+1}.

The main disadvantage of this method is that it will not be relevant if the distribution is not parametric. That is the interest of the empirical density estimation. However, the empirical density estimation requires to store all observations since the density is expressed for the contributor feature j as:

$$\forall x \in \mathbb{R}, \quad \hat{f}_{\tilde{X}_j^E}(x) = \frac{1}{qb} \sum_{i=1}^{q} K\left(\frac{x - x_j^i}{b}\right)$$

where K is a non-negative function such that $\int_{\mathbb{R}} K(x)dx = 1$ known as kernel and $b > 0$ is a bandwidth parameter. For a Gaussian kernel, the function K is

$$\forall x \in \mathbb{R}, \quad K(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}.$$

3.2.4 Measurement data and tolerance model link

In a general framework, the knowledge of the amount of information in the m measured contributors allows to re-adjust the tolerance model. Knowing the regression coefficients and denoting $\tilde{X}_1, \ldots, \tilde{X}_m$ the measured contributors, we can introduce the partial residuals e defined by:

$$\tilde{Y} = \sum_{j=1}^{m} \alpha_j \tilde{X}_j + \sum_{j=m+1}^{p} \alpha_j X_j + e.$$

The two sum terms are respectively representative of measured and unmeasured contributors. The goal is to retrieve as much information as possible about e, the error introduced when measurement data is available.

3.3 Linear model correction

In the sequel of this chapter, we will consider the linear model to represent the relationship between inputs and output of an assembly. Such a model of tolerances system involves several limitations. Indeed, the tolerance model does not perfectly reflects the reality encountered in plants and final assembly lines. The correction step proposed in this part aims to tackle this issue through different outlooks for tolerance model enhancement. This does not guarantee the explanation of the discrepancy between reality and the theoretical model. The goal is simply to recalibrate the model thanks to the available data in order to obtain a more adequate result, in the same way as in [14].

3.3.1 Influence coefficient sign correction

We focus on the signs of the coefficients and not on their value because generally, the availability of data in an industrial context does not allow this information to be retrieved. Indeed, not all contributors are measured, excluding any multivariate regression. Moreover, production rates can lead to the analysis of dataset with few observations. Studying the coefficient signs gives a more robust result with respect to these constraints.

Influence coefficient sign is a well known issue for centered tolerances management. Indeed, sign errors might happen during 3D experiment, values registration, or tolerance change discussion. For instance, a designer can imagine a feature in one direction, and when an operator measures, he will take the measurement in the other direction. This leads to a sign inversion. When tolerances are centered, influence coefficient signs do not have a significant impact on the theoretical model. However, when measurement data are included to the tolerance study, the sign of coefficients becomes crucial to perform relevant analysis on an individual or off-centered tolerances population.

To face this issue, there is a need for a tool to automatically correct influence coefficient signs. This is made possible by the availability of measurement data. These data allow to verify the concordance between what is observed and the model coefficient sign.

Several methods were considered. The first one relies on correlation study, while another is based on a regression approach for the binary variable of positive or negative sign.

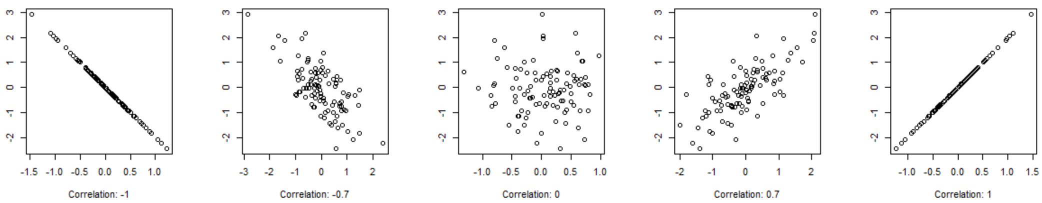

FIGURE 3.6: Pearson correlation coefficients examples

Correlation analysis

The correlation analysis allows to quantify the link between a contributor and its top level requirement. The correlation measures the dependency between two variables and we take the most common approach for correlation which is the Pearson product-moment correlation coefficient. This dependency quantifier measures the linear correlation between two variables.

The definition of the Pearson correlation coefficient $\rho_{X,Y}$ between two random variables X and Y with expected values μ_X and μ_Y and standard deviations σ_X and σ_Y is

$$\rho_{X,Y} = \mathrm{corr}(X,Y) = \frac{\mathrm{cov}(X,Y)}{\sigma_X \sigma_Y} = \frac{\mathrm{E}[(X-\mu_X)(Y-\mu_Y)]}{\sigma_X \sigma_Y} \tag{3.4}$$

The coefficient value is between -1 for a total negative correlation and 1 for a total positive correlation, as illustrated in Figure 3.6. In our case, what is interesting in this value is the sign. It allows to assess about the sign of the contributor influence coefficient. If the correlation between a contributor and its top level requirement is positive, then the associated influence coefficient should be positive, and conversely.

The correlation is used because covariance is scale dependant. With the correlation, significance thresholds could be defined according to the importance of a contributor in a stack chain. For instance in the Gaussian case with 6σ hypothesis, the notion of threshold correlation is defined and can be considered to decide about the significance of the correlation value. Let focus on the independent contributors $X_1, \ldots, X_p$, with respectively the tolerance bounds $\pm v_1, \ldots, \pm v_p$. The linear model gives $Y = \sum_{i=1}^{p} X_i$.

The correlation definition for the j-th contributor gives

$$\begin{aligned}\operatorname{corr}(X_j, Y) &= \frac{\operatorname{cov}(X_j, Y)}{\sigma_{X_j}\sigma_Y} = \frac{\operatorname{cov}(X_j, \sum_{i=1}^{p} X_i)}{\sigma_{X_j}\sigma_Y} = \frac{\operatorname{cov}(X_j, X_j)}{\sigma_{X_j}\sigma_Y} \\ &= \frac{\operatorname{var}(X_j)}{\sigma_{X_j}\sigma_Y} = \frac{\sigma_{X_j}^2}{\sigma_{X_j}\sigma_Y} = \frac{\sigma_{X_j}}{\sigma_Y} = \frac{v_j/3}{\sigma_Y}.\end{aligned}$$

We can express σ_Y as a function of the RSS result denoted by T_{RSS}.

$$\begin{aligned}\sigma_Y &= \sqrt{\operatorname{var}(Y)} = \sqrt{\sum_{i=1}^{p} \sigma_{X_i}^2} = \sqrt{\sum_{i=1}^{p} \left(\frac{v_j}{3}\right)^2} \\ &= \frac{1}{3}\sqrt{\sum_{i=1}^{p} v_j^2} = \frac{T_{RSS}}{3}\end{aligned}$$

The threshold correlation for the j-th contributor in the Gaussian case under 6σ hypothesis is therefore given by

$$\operatorname{corr}(X_j, Y) = \frac{v_j/3}{T_{RSS}/3} = \frac{v_j}{T_{RSS}}.$$

This correlation threshold help with- the limitation when a contributor has a weak impact in the stack chain. The correlation between a weak contributor and its top level requirement might not be significant. Indeed, if the impact of the contributor is negligible, then the correlation which is supposed to assess about the linear link between contributor and top level requirement will be close to zero. Note that from an industrial perspective, a sign error for a contributor with a low importance in the stack chain will not significantly disturb the output of the assembly and therefore does not represent a real threat.

Regression on signs

Another approach for influence coefficient signs check is based on a regression according to the binary variables of coefficient signs. We assume that all contributors have available feedback measurement data.

The linear model assumption gives

$$Y = \sum_{j=1}^{p} \alpha_j X_j$$

where $\alpha_1, \ldots, \alpha_p$ are influence coefficients whose sign requires validation. We assume absolute values of these coefficients $(|\alpha_1|, \ldots, |\alpha_p|)$ are known.

We define for any $j \in \{1, \ldots, p\}$ the corrected influence coefficients thanks to a binary variable $s_j \in \{-1, 1\}$ such that the corrected linear coefficient in the tolerance model is $s_j |\alpha_j|$.

The top level requirement is now represented by the expression

$$Y = \sum_{j=1}^{p} s_j |\alpha_j| X_j.$$

A linear regression on the binary variable s allows to obtain the sign of influence coefficient consistent with measurement data. The linear least squares method gives the optimal $\hat{s} = (\hat{s}_1, \ldots, \hat{s}_p) \in \{-1, 1\}^p$ values by minimizing the sum of squared residuals where absolute values $|\alpha_1|, \ldots, |\alpha_p|$ are given. The associated optimization problem is

$$(\hat{s}_1, \ldots, \hat{s}_p) = \underset{s \in \{-1,1\}^p}{\operatorname{argmin}} \sum_{k=1}^{n} \left(y^k - \sum_{j=1}^{p} s_j |\alpha_j| x_j^k \right)^2 \tag{3.5}$$

where $(y^1, \ldots, y^n)$ and $(x_j^1, \ldots, x_j^n)$ are n observations of Y and X_j for any $j \in \{1, \ldots, p\}$ respectively, and n is the number of observations.

To solve the optimization problem stated in (3.5), several solvers are available, based on simulated annealing ([15]) or branch and bound techniques ([16]). An example based on an exhaustive search with a simple application is detailed in 3.4.2.

3.3.2 Offset consideration - Integration effect

In the case where we have partial measurement data (Figure 3.7), it is possible to have a first information on the e error previously defined. This is an offset representative of a systematic bias.

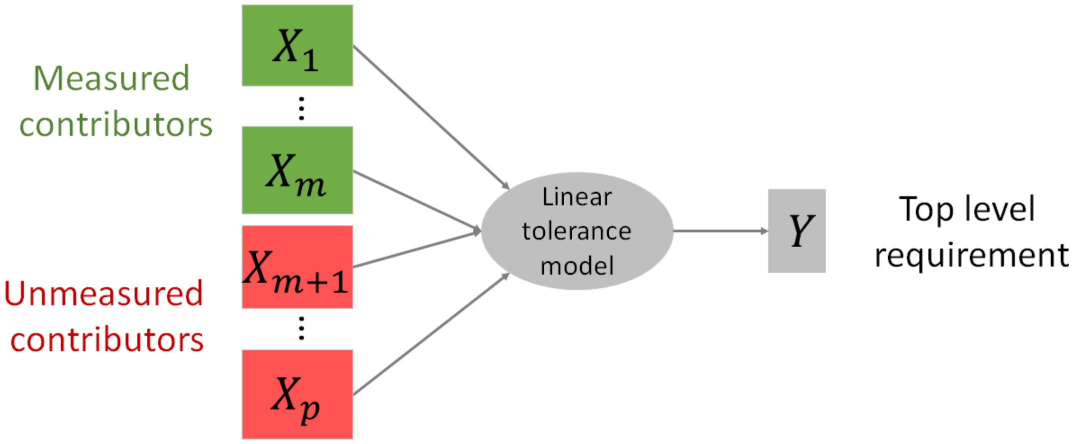

FIGURE 3.7: Measurement data availability on the linear tolerance model

If we assume all non-measured features to be centered, we have an information about the mean of e:

$$\mu_e = \mu_Y - \sum_{j=1}^{m} \alpha_j \mu_j$$

where μ_Y is the mean of observations of the top level requirement Y and $\mu_j = \frac{1}{n}\sum_{i=1}^{n} \tilde{x}_i^j$ is the mean of the contributor $\tilde{X}_j$.

This μ_e actually represents integration effects, for instance gravity effects that happen during the assembly or shift errors due to unmeasured contributors. These effects are not captured by the 3D model used to estimate regression coefficients as they are using a rigid-body approach. This step is a way to loop back to the model by knowing partial measurement data, and to identify systematic biases that can be found in assemblies.

3.4 Application on an example

3.4.1 Data

The simulated example in this section represents an assembly with three contributors, and the stack chain associated is detailed in Table 3.1, including tolerance, influence, feedback availability, measured mean and standard deviation information. These information needs to be corrected thanks to the methodology presented in this section.

Contributor	Tol. interval	Inf.	Sign	Feedback	Mean	St. Dev.
X_1	± 3	$\alpha_1 = 0.5$	$+$	Yes	$\mu_1 = 0.4$	$\sigma_1 = 0.9$
X_2	± 2	$\alpha_2 = 1$	$+$	No	*NA*	*NA*
X_3	± 1	$\alpha_3 = 1$	$+$	Yes	$\mu_3 = 0.3$	$\sigma_3 = 0.4$

TABLE 3.1: Top level requirements considered in this example

Data observations are simulated and noise is added in order to represent the uncertainty met in reality. The top level requirement of this stack chain has a targeted tolerance range $\pm X$, and measurement data are available. The mean for this simulated example is $\mu_Y = -0.4$. In this example, there are $n = 1000$ observations simulated using the correct information about the tolerance model parameters, that need to be retrieved through the correction process. As stated in Table 3.1, only X_1 and X_3 are observed. Figure 3.8 represents the histograms of simulated observations for these contributors. In this simulation, the aim is to mimic measurement and to consider both deviation and measurement accuracy.

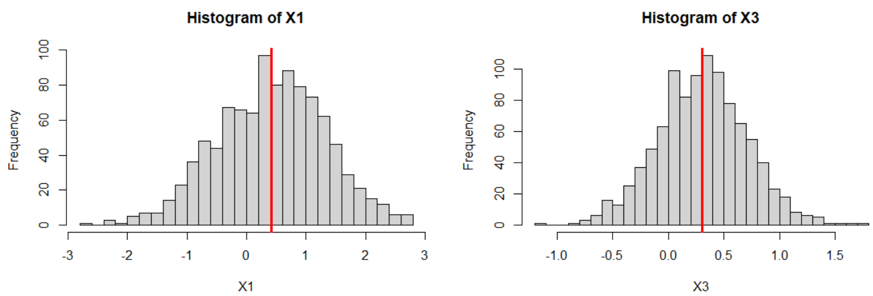

FIGURE 3.8: Histogram for observations of X_1 and X_3.

For the top level requirement measurement simulation, data are generated through the linear model and a Gaussian uncentered noise with a variance 0.5 is added to simulate external effects on an assembly (transport, tooling, . . .) and to reflect the given $\mu_Y = -0.4$. The simulated observations are summarized in an histogram displayed in Figure 3.9.

$$Y = \sum_{j=1}^{p} s_i \left|\alpha_j\right| X_j + \epsilon_Y$$

where $\epsilon_Y \sim \mathcal{N}(-0.3, 0.5)$.

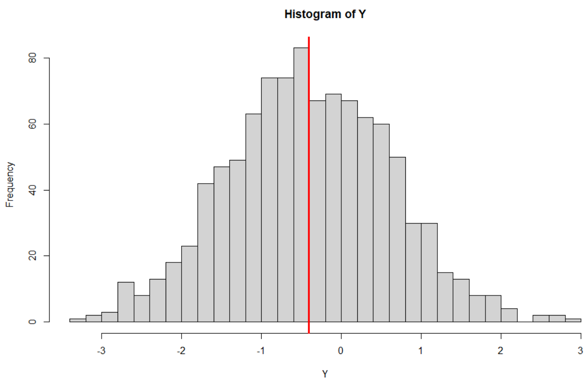

FIGURE 3.9: Histogram for observations of Y.

3.4.2 Computation

Influence coefficient sign correction - Correlation analysis

Figure 3.10 represents the observations of Y with respect to X_1 and X_3. Pearson correlations are given in Table 3.2 through a correlation matrix. The signs of this coefficients provide the correct influence coefficient sign for the first and last contributor.

Cor.	X_1	X_3	Y
X_1	1	0.04	0.44
X_3	0.04	1	-0.37
Y	0.44	-0.37	1

TABLE 3.2: Correlation matrix between observed random variables.

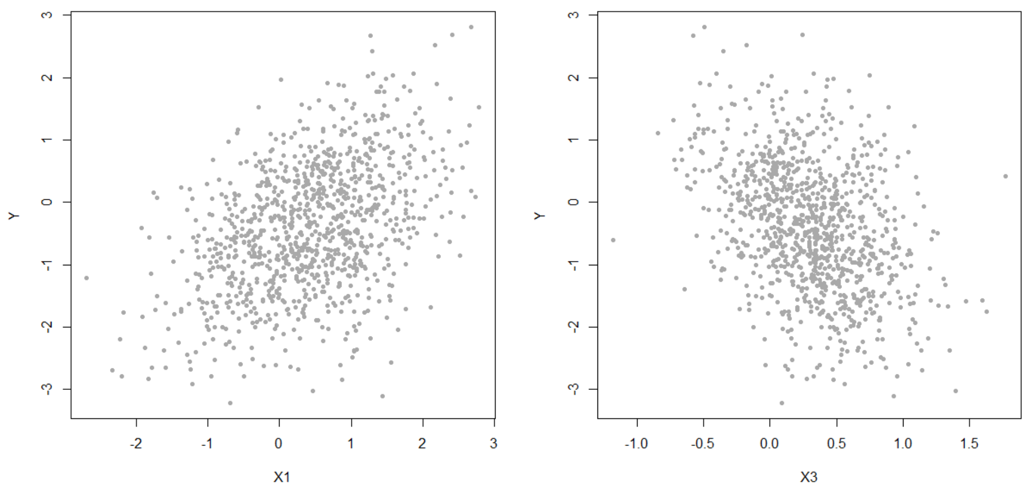

FIGURE 3.10: Scatterplot of Y observations with respect to X_1 and X_3.

Influence coefficient sign correction - Regression on signs

In this simple case, the minimization problem as stated in (3.5) is easy to solve. Indeed, there are only two coefficients signs to correct, s_1 and s_3, and it is therefore possible to quickly calculate for each combination of signs the value to be minimized,

$$\min_{(s_1,s_3)\in\{-1,1\}^2} \sum_{k=1}^{n} \left(y^k - \left(s_1 \left|\alpha_1\right| x_1^k + s_3 \left|\alpha_3\right| x_3^k \right) \right)^2$$

Table 3.3 gives the values of this quantity to be minimized for each sign combination.

	$s_1 = -1$	$s_3 = -1$
$s_1 = 1$	2205.75	1371.53
$s_3 = 1$	2878.68	2613.70

TABLE 3.3: Results of exhaustive search for minimization.

The minimum value is reached when $s_1 = 1$ and $s_3 = -1$. The corrected sign is positive for X_1 and negative for X_3, which is the same result than the correlation analysis.

Integration stack - Offset consideration

Thanks to the available data, the computation on means in order to detect an offset to be considered as an integration stack gives

$$\mu_e = \mu_Y - \sum_{j=1}^{m} \alpha_j \mu_j$$

$$\mu_e = -0.4 - (0.5 \times 0.4 + (-1) \times 0.3) = -0.3.$$

This offset μ_e can be integrated to the stack chain as a recurrent effect observed on the output of the assembly.

3.4.3 Result

As a result of the model correction step, the stack chain is now as detailed in Table 3.4. The influence coefficient signs are corrected and an offset is considered.

Contributor	Tol. interval	Corrected influence	Sign
X_1	± 3	0.5	$+$
X_2	± 2	1	$+$
X_3	± 1	-1	-
Integration stack	offset: $\mu_e = -0.3$	1	$+$

TABLE 3.4: Top level requirements considered in this example

Figure 3.11 gives intermediate results after the signs correction and the offset integration. The last blue histogram is the corrected histogram for Y observations, to be compared with the histogram in Figure 3.9 which represents the simulated reality. The distributions of error between this real Y observations and the estimation from the model at each correction step are displayed in the grey histograms. This error is the difference between real observations of Y and estimated one. For the model with sign correction and offset consideration, the error for each observation k is $y^k - (\sum_{i=1}^{p} \hat{s}_i |\alpha_i| x_i^k + \mu_e)$.

We can see that the distribution of errors between the actual data and the estimation of the empirical model is rather dispersed. The first step of sign correction allows to tighten it. The stage of the consideration of the offset allows to re-center it.

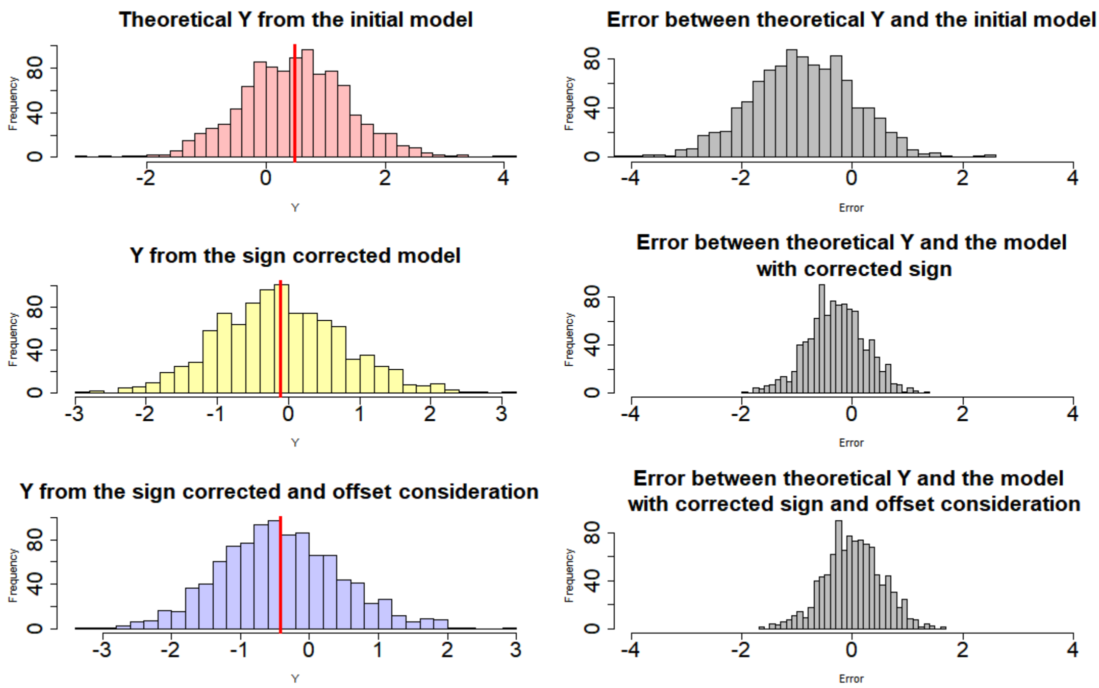

Figure 3.11: Intermediate results on Y after each corrections step and associated error distributions.

3.5 Conclusion

This chapter focused on the tolerance model to link the inputs to the output of an assembly. It is a basic element in the different ways to improve the tolerance approach discussed. There are, however, limitations to this approach to model correction. For instance, one of the challenges is to decide how many observed contributors allow to propose a consistent model correction. It would also be relevant to consider the number of measures needed.

The choice of the simple linear model is motivated by industrial constraints. To overcome the limitations of the model, various methods are considered and illustrated on a simulated example. The two main axes of correction are the addition of an integration stack and the validation of the linear influence coefficients. The goal is to propose a tolerance model in accordance with the measurements observed in reality in the plants where both production and assembly take place.

Bibliography

[1] A. Wirtz. Vectorial Tolerancing A Basic Element For Quality Control. *Proc. of 3rd CIRP Seminars on Computer Aided Tolerancing*, pages 115–128, 1993.

[2] A. Desrochers and A. Clément. A dimensioning and tolerancing assistance model for CAD/CAM systems. *The International Journal of Advanced Manufacturing Technology*, 9(6):352–361, November 1994.

[3] P. Bourdet. The concept of small displacement torsor in metrology. *Ser Adv Math Appl Sci*, 40:110–122, 1996.

[4] Gaurav Ameta. *Statistical tolerance analysis and allocation for assemblies using Tolerance-Maps*. ProQuest, 2006.

[5] Joseph K. Davidson and Jami J Shah. Using Tolerance-Maps to represent material condition on both a feature and a datum. *The 8th international CIRP seminar on computer aided tolerancing*, pages 92–101, 2003.

[6] *Advanced Tolerancing Techniques.*

[7] Jinsong Gao, Kenneth W Chase, and Spencer P Magleby. DE-Vol. 82, Volume 1, ASME 1995 Proceedings of the ASME Design Engineering Technical Conferences Boston, MA, Sept. 17-20, 1995, pp.353-360. page 10.

[8] Xingyu Yan and Alex Ballu. Tolerance analysis using skin model shapes and linear complementarity conditions. *Journal of Manufacturing Systems*, 48:140–156, July 2018.

[9] Benjamin Schleich, Nabil Anwer, Luc Mathieu, and Sandro Wartzack. Skin Model Shapes: A new paradigm shift for geometric variations modelling in mechanical engineering. *Computer-Aided Design*, 50:1–15, May 2014.

[10] A. Dumas, J.-Y. Dantan, and N. Gayton. Impact of a behavior model linearization strategy on the tolerance analysis of over-constrained mechanisms. *Computer-Aided Design*, 62:152–163, May 2015.

[11] Aad W. van der Vaart. *Asymptotic statistics.* Cambridge series in statistical and probabilistic mathematics. Cambridge Univ. Press, Cambridge, 1. paperback ed., 8. printing edition, 2007. OCLC: 838749444.

[12] Airbus. Product and Process Performance through Measurement Feedback – Calculation Methodology for Continuous Data, 2019.

[13] Bernard W Silverman. *Density estimation for statistics and data analysis*, volume 26. CRC press, 1986.

[14] Rubén Ibáñez, Emmanuelle Abisset-Chavanne, David González, Jean-Louis Duval, Elias Cueto, and Francisco Chinesta. Hybrid constitutive modeling: data-driven learning of corrections to plasticity models. *International Journal of Material Forming*, 12(4):717–725, July 2019.

[15] Scott Kirkpatrick, C Daniel Gelatt, and Mario P Vecchi. Optimization by simulated annealing. *science*, 220(4598):671–680, 1983. Publisher: American association for the advancement of science.

[16] AH Land and AG Doig. An automatic method of solving discrete programming problems. Econometrica. v28. 1960.

Chapter 4

Out-of-tolerance risk management in production

During the design phase of a product, tolerances on manufacturing assembly are defined. When the product is in production, measurement data on manufacturing processes become available. This allows to consider and calibrate new limits called acceptance criteria on tolerances based on what is observed. We propose a smart statistical approach to define such criteria using feedback measurement data to refine the initial tolerance model. The method allows to compute the risk to be out of tolerance on top level requirements and to calibrate acceptance criteria in order to meet these requirements reflecting the exact industrial need. A tool has been developed to illustrate this approach on aerostructures use-cases.

4.1 Introduction

Standard guidelines lead to define tolerances with a certain percentage of risk to be above requirement targets. The actual risk taken during definition can vary due to negotiations: this risk can be greater or lower according to what the assembly responsible agreed to accept. When all the tolerances are set and validated, the risk status is frozen. Similarly to the management and verification method stated in [1], the aim is also to offer a way to consider this risk to take industrial decisions. While the authors of [1] focus on tool risk and mean shift, we propose to consider available measurement data of assembly contributors and enhanced tolerance model.

Once tolerances specifications are in service, production gives feedback measurements on some features. These measures give additional information on the tolerance stack chain and especially on assumptions made during tolerance design phase.

A way to consider information made available in production is to provide acceptance criteria on tolerances. These acceptance criteria represent flexible values to be considered as reasonable limits instead of the traditional tolerance value. The advantage of such indicators is to fit to the available data and thus be more accurate according the observations. Unlike tolerances, acceptance criteria are valid only for a limited period of time during which the production behaves in the same way. However, considering acceptance criteria implies to observe and assess about the feedback data. It makes acceptance criteria less universal limits than what tolerance could be. A process control must be set up to ensure the relevance of acceptance criteria.

In the approach proposed in this chapter, the notion of risk for an assembly has to be precisely defined. Indeed, this is the basic component to quantify the conformity of an assembly and such a measure must reflect what is important for the industrial stakeholders. By focusing on a single assembly contributor on which acceptance criteria should be defined, one should not only look at the risk of being out of tolerance for the output of the assembly by a certain value for that contributor, but also at the risk for that contributor to be measured at that same value. In the manner of a probability/impact analysis often used in the general field of risk management, the definition of acceptance criteria on the tolerances of an assembly contributor must be based on two distinct notions of risk to be jointly considered.

This risk monitoring step relies on the tolerance model previously discussed in Chapter 3. The first step detailed in Section 3.3 is to re-adjust the tolerance model taken into account for tolerance definition thanks to production feedback. Then, for each stack chain, a standard computation on contributors distribution functions gives the risk to be above top level requirement targets. The first part of this chapter focuses on hypothesis assumed in this step and detail the estimation of the output top level requirement feature distribution. Then, an indicator of the amount of information in variability is proposed in order to better understand and identify differences between as-design model and enhanced model through feedback measurement data. This also contributes to the following part which is about different notions and definition of risk and its use for out-of tolerance monitoring. Finally, the last part focuses on a concrete example for application of the out-of-tolerance risk monitoring methodology proposed here to illustrate its benefits.

4.2 Output feature distribution

When the product is in design stage, tolerances both on contributors and top level requirement are set. The risk to be out of specification at the top level feature of an assembly is therefore fixed in accordance with the contributors tolerances and the selected value as top level target which should not be exceeded. During production phase, distributions of the contributors features are essential to value the risk at the assembly level. Knowing the stack chain of the assembly and eventually some contributors feedback data, the risk to be out of tolerance at the top level requirement can then be re-evaluated.

There are various ways to consider input distributions depending on the availability of measurement data. If we do not have feedback, the conservative approach is to assume a uniform distribution between the tolerance bounds as it is the less informative probability distribution. If we have feedback on the contributor, we can either use fitted Gaussian or empirical distribution as discussed in Chapter 3. In each case detailed in hereafter, the output feature distribution is obtained from the convolution product of inputs distributions and we discuss the numerical implementation later on.

4.2.1 Estimated distribution

During the design stage within the Airbus tolerancing process, contributors features are usually assumed to be uniform distributions, as detailed in Chapter 2. Therefore, the targeted tolerance values on the top level feature Y are defined on an output distribution which is a sum of p uniform features. We assumed the independence between contributors random variables and considering a linear model, the density of this output distribution in the design phase is then given by the convolution product

$$f_Y = f_{\alpha_1 X_1^{\mathcal{U}}} * \cdots * f_{\alpha_p X_p^{\mathcal{U}}}$$

where $\forall j \in 1, \ldots, p$, $X_j^{\mathcal{U}} \sim \mathcal{U}(-v_j, v_j)$ and $v_j > 0$.

In the case where we have measurement data available for m measured features, we can first assume normality. Such an assumption allows to store only the two parameters of the normal distribution μ_j and σ_j for any $j \in \{1, \ldots, m\}$. This might be more convenient for the methodology implementation in internal tools.

Only the Gaussian case is discussed here for the development of associated formulas, but the approach remains valid with any other type of known distribution with parameters that can be estimated through the method of moments. This would require to apply the step of fitting the distributions on the measures as detailed in Chapter 3 and to store both the type of distribution and the associated parameters.

At this point, the convolution product on densities concerns the m measured features considered as normal distribution, the $p - m$ non measured features remaining considered as uniform distributions and the model error ϵ whose distribution is estimated through the model correction from measurement data. Now the density of the top level feature can be re-evaluated as

$$f_{\alpha_1 \tilde{X}_1^{\mathcal{N}}} * \cdots * f_{\alpha_m \tilde{X}_m^{\mathcal{N}}} * f_{\alpha_{m+1} X_{m+1}^{\mathcal{U}}} * \cdots * f_{\alpha_p X_p^{\mathcal{U}}} * f_\epsilon$$

which is equivalent to

$$f_{\alpha_1 \tilde{X}_1^{\mathcal{N}} + \cdots + \alpha_m \tilde{X}_m^{\mathcal{N}}} * f_{\alpha_{m+1} X_{m+1}^{\mathcal{U}}} * \cdots * f_{\alpha_p X_p^{\mathcal{U}}} * f_\epsilon$$

where for any $j \in \{m+1, \ldots, p\}$, each unmeasured contributor $X_j^{\mathcal{U}} \sim \mathcal{U}(-v_j, v_j)$ and for any $j \in \{1, \ldots, m\}$, each measured contributor $\tilde{X}_j^{\mathcal{N}} \sim \mathcal{N}(\mu_j, \sigma_j^2)$ thus the sum of independent variables $\alpha_1 \tilde{X}_1^{\mathcal{N}} + \cdots + \alpha_m \tilde{X}_m^{\mathcal{N}} \sim \mathcal{N}(\sum_{j=1}^m \mu_j, \sum_{j=1}^m \sigma_j^2)$.

An alternative to parametric models is to consider empirical distributions of measured contributors as described in Section 3.2.3. If we convolute the kernel density estimators with initial uniform distributions of non-observed features, we obtain the following result as density estimation for the top level feature

$$\hat{f}_{\alpha_1 \tilde{X}_1^E} * \cdots * \hat{f}_{\alpha_m \tilde{X}_m^E} * f_{\alpha_{m+1} X_{m+1}^{\mathcal{U}}} * \cdots * f_{\alpha_p X_p^{\mathcal{U}}} * f_e.$$

Convolution product implementation

To perform the convolution products of inputs distributions and then obtain an estimation on the output distribution, there exists a numerical method based on the approximation of Poisson summation formula detailed in Section 4.2.2. This is a convenient method for efficiency in tools implementation of the approach. Indeed, Monte Carlo simulations could also be considered for output level requirement estimation instead of a convolution product. However, calculation cost is far more expensive, especially if we deal with rare events estimation, such as the probability to be out of tolerance that we expect as low as possible.

Applications of Monte Carlo method in tolerance analysis are discussed in [2]. A thought on the use of this method is proposed in [3], highlighting both advantages and drawbacks of Monte Carlo approach and convolution product in applications. For industrial needs and internal tools implementation, the approximation of Poisson summation formula is used in order to perform the convolution products.

4.2.2 Approximation of Poisson summation formula

As mentioned in [4], [5] and used by Lebrun in [6], a continuous random variable which is the sum of the values taken by the density of Y on an entire network is equal to the sum of the values taken by its characteristic function on another wisely selected network.

Let us focus on the framework given by normal distributions for measured contributors and uniform distributions for others as described above. A similar approach can be applied to other situations.

Let Y be a continuous random variable with density f_Y and characteristic function ϕ_Y such as, for any $u \in \mathbb{R}$, $\int_{\mathbb{R}} |\phi_Y(u)|\, du < \infty$ where $\phi_Y(u) = \mathbb{E}\left[e^{iuY}\right]$.

For any $r > 0$ and $y \in \mathbb{R}$, we have

$$\sum_{k\in\mathbb{Z}} f_Y\left(y + \frac{2k\pi}{r}\right) = \frac{r}{2\pi} \sum_{l\in\mathbb{Z}} \phi_Y(rl) e^{-irly}$$

and when integrating this sum of density, we obtain, for any $r > 0$ and $z, t \in \mathbb{R}$,

$$\sum_{k\in\mathbb{Z}} \left\{ F_Y\left(y + \frac{2k\pi}{r}\right) + F_Y\left(y - t + \frac{2k\pi}{r}\right) \right\} = \frac{rt}{2\pi} + \frac{r}{2\pi} \sum_{l\in\mathbb{Z}} \phi_Y(rl) \frac{e^{-irly} - e^{-irl(y-t)}}{irl}.$$

Using Fourier decomposition, we obtain, for any $r > 0$ and $z \in \mathbb{R}$,

$$f_Y(y) = \frac{r}{2\pi} \sum_{l\in\mathbb{Z}} \phi_Y\left(rl\right) e^{-irly} - \sum_{k\in\mathbf{Z}^*} f_Y\left(y + \frac{2k\pi}{r}\right).$$

Such an approximation allows to consider all kind of distributions for contributors and implement the convolution product to estimate the distribution of output feature and to compute the risk associated to the top level requirement target.

This implementation of the convolution product is interesting in terms of computing time. Another way to proceed would be with Monte Carlo simulations, as detailed above, and the number of simulations required would require much more computing time. The complexity of the approach with the approximation of Poisson summation formula is lower and therefore more interesting for the implementation in industrial tools.

4.3 Amount of information

The global purpose of this section is to introduce indicators for an a priori consideration of the relevance and performance of a risk analysis on a stack chain in the light of available feedback.

Several indicators already exist in the field of quality or statistical process control, such as the well known C_p and C_{pk} ([7]). These common quality indicators can be related to a contributor or a top level requirement and require tolerance limits and measurement data. These indicators can be computed at different stages of the manufacturing process. The C_p measures the dispersion with the standard deviation σ computed on the measurement data set. This characterizes the capacity of the process to fit between the upper limit v_u and lower limit v_l. The C_{pk} quantifies the process shifting through the mean μ with respect to the limits v_l and v_u.

$$C_p = \frac{v_u - v_l}{6\sigma}$$
$$C_{pk} = \frac{\min(\mu - v_l, v_u - \mu)}{3\sigma}$$

The aim of this section is to compare the as-designed variability and the measured variability either for contributors or top level requirement. It means that different hypothesis are made for contributors as detailed in the previous section and the different indicators presented in the sequel are built to quantify both variability and comparison between design and measurement reality. Thanks to this comparison, the relevance of acceptance criteria definition can be discussed.

4.3.1 Contributor individual indicator

In the design phase, we can quantify the amount of information about the feature variability for each contributor as the variance defined for the uniform distribution. Therefore, for all $j \in \{1, \dots, p\}$, we define the amount of information about the contributor variability as $\frac{v_j^2}{3}$ where v_j is the tolerance bound of the contributor and parameter of the associated uniform distribution.

For each observed contributor j, we propose to define the ratio R_j between the designed variability and the measured variability as

$$R_j = 3\frac{\sigma_j^2}{v_j^2} \tag{4.1}$$

where σ_j^2 is the variance of the j-th contributor.

Such an indicator gives an information on the margin in terms of variability for the contributor in this assembly. If R_j is equal to 1, it means there is no margin in terms of variability for this contributor. The lower is R_j, the higher the variability margin is. Note that, if the R_j value is above 1, it means that the observed variability is worse than the designed one for the considered contributor.

4.3.2 Top level indicators

In a stack chain, each contributor represents a part of the information observed at the output of the assembly on the top level feature that represents a final requirement on a manufacturing step. This information is apportioned according to certain assumptions made during the design phase, and the availability of measurements therefore allows the partitioning among the contributors of this information to be refined.

As contributors are assumed to be independent and $Y = \sum_{j=1}^{p} \alpha_j X_j$, we can assess about the global variability. In design phase, the global amount of information about expected variability for the top level feature is given by

$$I_v = \frac{1}{3} \sum_{j=1}^{p} v_j^2. \tag{4.2}$$

When feedback is available, the global amount of information about the re-valuated variability for the top level feature is the sum of two terms depending on observed and non-observed variability of the measured or not contributor features,

$$I'_v = \sum_{j=1}^{m} \sigma_j^2 + \sum_{j=m+1}^{p} \frac{v_j^2}{3}. \tag{4.3}$$

4.3.3 Comparison indicators

There are many ways to represent the sharing of information between contributors and to quantify the role each one plays in the assembly. A naive approach would be to simply consider whether a contributor is observed or not. If all contributors are measured, then it would be assumed that 100% of the assembly information is accessible. Conversely, if no contributors are measured, then the information of the assembly and its partitioning remains unknown. When only part of the contributors are measured, a naive indicator denoted by R_{naive} could be a ratio between the number of observed contributors and the total number of contributors in the assembly.

$$R_{\text{naive}} = \frac{m}{p}. \tag{4.4}$$

However, this approach does not allow to consider the relative importance of the contributors compared to each other. Indeed, it is frequent that some contributors of the stack chain are predominant while some have a very small impact on the output characteristic of the assembly. One way to overcome this limitation is to reason in terms of contributor variance. This allows a precise look at the partitioning of information between contributors, with or without the availability of measurement data.

We can define an indicator R which gives the global variability margin for the assembly as the ratio between the top level feature designed and re-valuated variability

$$R = \frac{\sum_{j=1}^{m} \sigma_j^2 + \sum_{j=m+1}^{p} \frac{v_j^2}{3}}{\sum_{j=1}^{p} \frac{v_j^2}{3}} = \frac{I_v'}{I_v}. \tag{4.5}$$

Before performing acceptance criteria, this last global indicator R gives the margin in terms of variability available for an assembly. If the ratio is above 1, it means acceptance criteria definition might not be very relevant for a tight targeted out-of-tolerance rate since measures performed on this assembly are already more dispersed than expected. Conversely, if the indicator value is below 1, it means acceptance criteria are practicable and profitable for one or several stack chain contributors. The lower the indicator, the more significant acceptance criteria will be.

From an industrial perspective, a similar indicator is relevant. This is the ratio between the variability explained only by the measured contributors and the re-valuated global variability for the top level requirement. We then define Q as

$$Q = \frac{\sum_{j=1}^{m} \sigma_j^2}{\sum_{j=1}^{m} \sigma_j^2 + \sum_{j=m+1}^{p} \frac{v_j^2}{3}} = 1 - \frac{I_v}{I_v'}. \tag{4.6}$$

Type	Name	Definition	Parameters
Quality / SPC	C_p	$\frac{v_u - v_l}{6\sigma}$	v_u upper limit v_l lower limit σ st. dev.
	C_{pk}	$\frac{min(\mu - v_l, v_u - \mu)}{3\sigma}$	v_u upper limit v_l lower limit μ mean σ st. dev.
Contributor individual indicator	R_j	$3\frac{\sigma_j^2}{v_j^2}$	v_j σ_j
Top level indicator	I_v	$\frac{1}{3}\sum_{j=1}^{p} v_j^2$	v_j p
	I'_v	$\sum_{j=1}^{m} \sigma_j^2 + \sum_{j=m+1}^{p} \frac{v_j^2}{3}$	v_j σ_j m p
Comparison indicator	R_{naive}	$\frac{m}{p}$	m p
	R	$\frac{\sum_{j=1}^{m} \sigma_j^2 + \sum_{j=m+1}^{p} \frac{v_j^2}{3}}{\sum_{j=1}^{p} \frac{v_j^2}{3}}$	v_j σ_j m p
	Q	$\frac{\sum_{j=1}^{m} \sigma_j^2}{\sum_{j=1}^{m} \sigma_j^2 + \sum_{j=m+1}^{p} \frac{v_j^2}{3}}$	v_j σ_j m p

TABLE 4.1: Indicators about amount of information in variability summary

This indicator Q belongs to $[0, 1]$. A value equal to 0 means that $I_v = I'_v$. In such a case, no contributor is measured and $m = 0$. Only non-observed contributors explain the assembly variability. Conversely, a value of 1 means that $m = p$ so the second term of the denominator nullifies itself. All contributors are observed and explain the assembly variability.

4.3.4 Indicators summary

The Table 4.1 summarizes all indicators detailed in this section, and Table 4.2 presents their use and the level to which the indicator applies, either at contributor level $X_1, \ldots, X_p$ or at the top level requirement Y.

Type	Name	Level	Use
Quality SPC	C_p	Y, X_j	Measure capabilities process (dispersion) Evaluate compliance between requirement (tolerance limits) and process (measurement data).
	C_{pk}	Y, X_j	Measure capabilities process (off center) Evaluate compliance between requirement (tolerance limits) and process (measurement data).
Contributor individual indicator	R_j	X_j	Appreciate the difference between the as-designed variability (through uniform assumption) and the reality (normal distribution fitted on measurement data).
Top level indicator	I_v	Y	Evaluate the global amount of information about expected variability at the top level requirement level during design phase.
	I'_v	Y	Evaluate the global amount of information about re-valuated variability at the top level requirement level considering measurement data.
Comparison indicator	R_{naive}	Y	Evaluate the proportion of information brought by the availability of measurement data. This does not take into account the relative importance of the contributors.
	R	Y	Evaluate the proportion of information about the global variability brought by the availability of measurement.
	Q	Y	Evaluate the proportion of global variability explained thanks to measurement data.

TABLE 4.2: Indicators about amount of information in variability use

4.4 Risk evaluation

Generally speaking, the risk for an assembly is defined as the out of tolerance percentage for the top level requirement of a stack chain. If the target value for the top level requirement is T_y, the risk is $P(Y > T_y)$. A common practice for tolerance design is to target a 0.27% out-of-tolerance rate for the top level requirement. This number comes from the Gaussian assumption for contributors and the widely used 6σ rule which leads to this rate value for the output top level requirement. During the serial phase, conformity definition and decision rules related to the out-of-tolerance rate depends on several parameters. More details are available in [8].

Tolerances defined on input features and assumptions about their distributions have a significant impact on the risk. Indeed, the distribution of the top level feature Y is completely dependent on the contributor features hypothesis as detailed in previous parts. The top level distribution can be estimated through the tolerance model and distribution hypothesis stated as following:

- (U): Uniform distribution assumption for all contributors
- (N): Normal distribution assumption for measured contributors and normal assumption for non-observed features
- (E): Empirically estimated distribution for measured contributors and normal assumption for non-observed features

A subtlety is introduced here by the observation of a feature to be added to the definition of risk. When an observation for a contributor is available, it means we know the value of one of the input features. If the focus is on the k-th observed feature, the measured value is x_k. The risk knowing the value of one of the input features and considering normal distributions for measured contributors is:

$$P_N(Y > T_y \mid \tilde{X_k} = x_k).$$

The distribution of $\tilde{X_k}$ can be estimated whether with Gaussian density or kernel density estimation as detailed in the previous part. In the Gaussian case for feedback, the risk to be out of tolerance for Y and to have a value equal to x_k for the k-th feature at the same time is

$$P_N(Y > T_y \mid \tilde{X_k} = x_k) f_{\tilde{X_k}}(x_k).$$

In the case of kernel density estimations, the risk is

$$P_E(Y > T_y \mid \tilde{X_k} = x_k) \hat{f}_{\tilde{X_k}}(x_k).$$

In this context, the so-called acceptance criteria would be for example $\pm x_k$. It is relevant to look at the risk that the contributor X_k is beyond these acceptance criteria. We can then jointly consider the risk of being out of tolerance at the output of the assembly. Finally, in the normal case and for symmetric acceptance criteria, the weighted risk to be considered industrially is

$$\int_{|t| > x_k} P_N(Y > T_y \mid \tilde{X_k} = t) f_{\tilde{X_k}}(t) dt.$$

Industrial use

In an industrial context, the process for using this notion of risk is as follows:

- Choose a contributor X_k in the stack chain on which to define acceptance criteria.
- Focus on a hypothetical measure x_k for this contributor.
- Check the probability τ to be out of tolerance at the top level. This depends on Y distribution, and thus on the way to fit distribution on other stack chain contributors measurement data. In the Gaussian framework, it means considering the probability $P_N(Y > T_y \mid \tilde{X_k} = x_k)$. For empirical density estimations, it means considering the probability $P_E(Y > T_y \mid \tilde{X_k} = x_k)$.
- Consider the distribution of the contributor X_k if measurement data are available. It allows to look at the probability to encounter a value x_k, represented by $f_{\tilde{X_k}}(x_k)$. In the Gaussian framework, the risk to be out of tolerance for Y and to have a value equal to x_k is $P_N(Y > T_y \mid \tilde{X_k} = x_k)\hat{f}_{\tilde{X_k}}(x_k)$. Same kind of result applies for the empirical framework.
- Decide whether the value x_k for the contributor X_k is acceptable, knowing the risk τ and eventually the probability that the contributor exceeds this value x_k.

The Airbus process is detailed in the dedicated section 4.5 about a simulated industrial example.

4.5 Example of a simulated industrial case

The following application example focuses on one contributor involved in three different assemblies. This example is similar to what one might encounter in an industrial context. We assume a reasonable quality in processes, allowing to work with normal distributions for assembly features if measurement data are available. Let us focus on the contributor 1, which is linked to three different stack chain.

4.5.1 Data

The contributors may or may not have feedback measurement, as well as the other contributors of the stack chains. This information and the targeted tolerance interval for the top level requirements are summarized in Table 4.3. Details for contributors and eventual feedback data are available in Table 4.4.

Top level requirement	Targeted tol. interval	Feedback
Top level req. 1	±4.5mm	Yes
Top level req. 2	±4.2mm	Yes
Top level req. 3	±4.0mm	Yes

TABLE 4.3: Top level requirements considered in this example

Top level req.	Contributor	Initial tolerance	Influence	Mean	Standard dev.
Top level req. 1	Contributor 1	±2mm	-1	1.46	0.97
	Contributor 2	±2mm	+1	-0.29	1.42
	Contributor 3	±0.5mm	+1	0.09	0.11
	Contributor 4	±0.4mm	-1	/	/
	Contributor 5	±0.4mm	-1	/	/
Top level req. 2	Contributor 1	±2mm	-1	1.46	0.97
	Contributor 6	±1mm	+1	-0.09	0.51
	Contributor 3	±0.5mm	+1	0.09	0.11
	Contributor 4	±0.4mm	-1	/	/
	Contributor 5	±0.4mm	-1	/	/
Top level req. 3	Contributor 1	±2mm	-1	1.46	0.97
	Contributor 7	±1mm	+1	/	/
	Contributor 3	±0.5mm	+1	0.09	0.11
	Contributor 4	±0.4mm	-1	/	/
	Contributor 5	±0.4mm	-1	/	/

TABLE 4.4: Stack chain information considered in this example

4.5.2 Indicator - Amount of information in variability

In this example with three different stack chains, we can compute the amount of information in variability ratio R presented in (4.5). This allows to characterize the global margin for the stack chain assembly in terms of variability and the performance potential of acceptance criteria methodology.

Stack chains	Ratio results R
Top level req. 1	1.07
Top level req. 2	0.71
Top level req. 3	0.75

TABLE 4.5: Indication about the amount of information in variability for stack chains - global ratio results.

The first stack chain has a ratio above 1, meaning that feedback observation shows that the variability within this stack chain is higher than expected. Indeed, the contributor 2 has a very large variance. If we focus on this top level ratio, we can define the individual ratios R_j between the designed variability and the measured variability for each contributor with feedback of the stack chain as stated in (4.1).

Stack chains	Ratio results R_j
Contributor 1	0.71
Contributor 2	1.51
Contributor 3	0.15

TABLE 4.6: Indication about the individual amount of information in variability for contributors - individual ratio results.

Table 4.6 shows that within the first stack chain, this is mainly the second contributor which presents a higher variability than expected. Contributor 2 is responsible for the global ratio above one for this stack chain, while other contributors with feedback have lower variablities than expected.

The two other stack chains in Table 4.5 have a lower ratio below one, meaning that variability is less than expected and acceptance criteria might perform well.

The top level requirement 3 has a slightly higher ratio result than the top level requirement 2 because even if stack chains are very similar, there are only two out of five contributors observed while there are three measured for the top level requirement 2.

4.5.3 Risk computation

As the tolerance we are interested in is involved in three top level requirements, we will perform three acceptance criteria analysis in order to evaluate the risk for this tolerance in each of the three stack chains. Once computations are done, the most restrictive risk has to be taken into account.

As stated in Section 4.2.1, we treat differently the feature with feedback measurement data and those without. If the contributor is not measured, then we assume a uniform distribution between the tolerance bounds. This is the case for the contributors 4, 5 and 7. Therefore, they are assumed to follow uniform distributions $\mathcal{U}(-0.4, 0.4)$, $\mathcal{U}(-0.4, 0.4)$ and $\mathcal{U}(-1, 1)$ respectively. Otherwise, if the contributor is measured and have observations available, a normal distribution can be fitted on observed data. In a first approach, we chose to assume the normality of measured contributors features, as industrial processes are supposed to meet requirements linked to indicators ensuring normality.In this example, the contributors 2, 3 and 6 follow normal distributions $\mathcal{N}(-0.29, 1.42)$, $\mathcal{N}(-0.09, 0.51)$ and $\mathcal{N}(-0.09, 0.11)$ respectively. Let us focus on the first top level requirement 1 to start the computation:

- First step is to compute the risk according the value $v \in \mathbb{R}$ of the contributor 1 we are interested in. To this end, we have to perform a convolution product. As we are focusing on a given value v of Contributor 1, we do not need its distribution in this first step as it is only an offset to consider in the linear tolerance model.

- We now consider all possible values in a range between -5mm and 5mm for this contributor. We perform the convolution product of the contributors distributions that are involved in the stack chain of the top level requirement 1. This convolution product associated to the each hypothetical value selected for contributor 1 gives the estimation of the top level requirement 1 feature distribution based on feedback available.

- This distribution can now be used to estimate the risk to be out of tolerance knowing the each possible values for Contributor 1. A way to represent the result from this approach is to plot the risk value depending on the value considered for Contributor 1, in a range between -5mm and 5mm in this example.

This methodology is applied to Contributor 1 in a same way for the two other stack chains of top level requirement 2 and top level requirement 3. Only the convolution product of the contributors distributions that are involved in the stack chain differs since the stack chain are different for each top level requirement. The results are presented in the plot presented in Figure 4.1. Each curve represents the risk associated to a top level requirement.

4.5.4 Results

The results of this analysis is presented in Figure 4.1. The X-axis represents the value that Contributor 1 could take. The associated curve gives the risk to be out of tolerance if Contributor 1 has the value considered on the X-axis, also called the impact risk. There are 3 different curves representing the 3 top level requirement in which Contributor 1 is involved. The Y-axis stands for the risk to be out of tolerance at a top level requirement level in percentage. The dotted line is the density of the distribution fitted on Contributor 1 measurement data. It helps the user to quantify the probability for Contributor 1 to have a specific value (from the X-axis).

The final acceptance criteria can be monitored both with the risk associated to the curves displayed in Figure 4.1 and the value of the density of the distribution for the considered contributor. Indeed, a high out of tolerance risk for a value is acceptable if the probability to encounter this value is very low. The guidelines for threshold percentage values rely on the engineering judgment of users.

4.5.5 Airbus process for the definition of acceptance criteria

Let us take an arbitrary value of 10% as the threshold for an acceptable out-of-tolerance rate at requirement level. This limits the probability $P_N(|Y| > T_y \mid \tilde{X}_1 = x_1)$. This value depends on the considered assembly and should be decided accordingly by industrial stakeholders. This first threshold stands for the the impact risk.

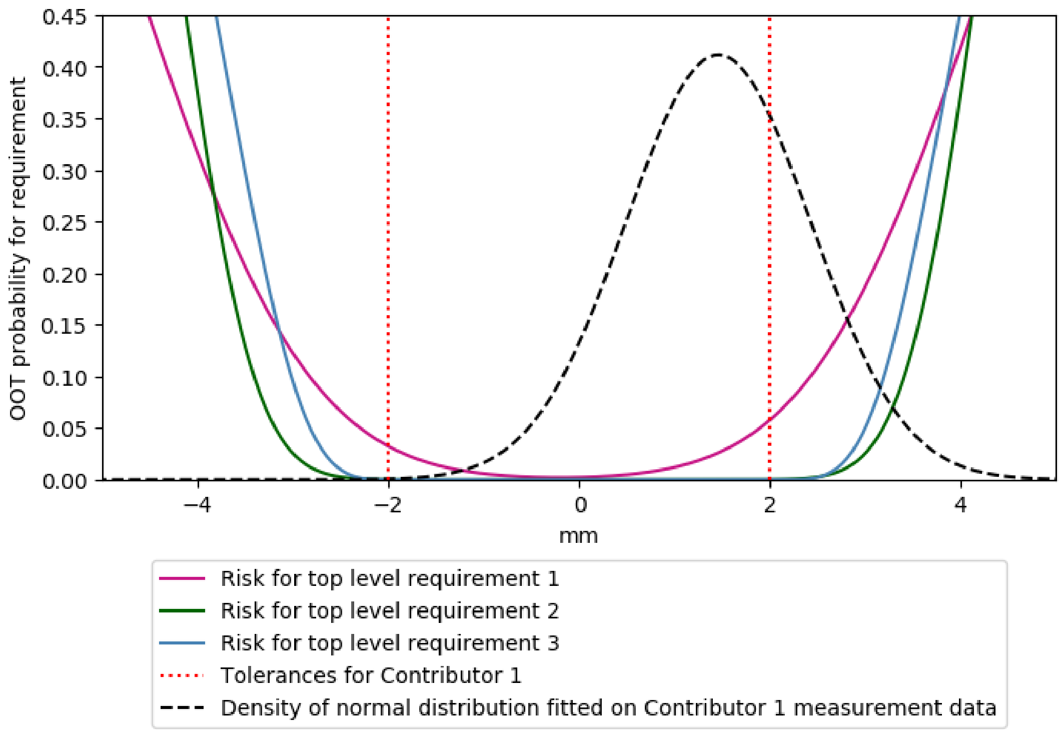

FIGURE 4.1: Example of results for acceptance criteria analysis

In our example, we can see that the top level requirement 1 gives the most restrictive result. Indeed, it is indeed this requirement that presents the highest risk, since its curve is globally above the other two. This is a logical result since the top level requirement 1 contains Contributor 2, whose measurements are more dispersed than expected. This is therefore an assembly for which the risk is higher because one contributor is already quite bad within the stack chain.

For a threshold of 10%, we can see on the graph that the associated acceptance criteria values for Contributor 1 would be -2.86 for the negative limit and 2.43 for the positive limit. Initial designed tolerances were ± 2, acceptance criteria for a 10% risk are extended to $-2.8|2.2$ and we have

$$P_N(|Y| > T_y \mid \tilde{X_1} = -2.86) = P_N(|Y| > T_y \mid \tilde{X_1} = 2.43) = 10\%.$$

It is also necessary to verify the probability of encountering such values for Contributor 1. A second threshold should be defined to limit the global risk taking into account both the impact risk and the probability to exceed a certain value.

Let us take an arbitrary value 3% for this second threshold to limit

$$\int_{|t|>x_1} P_N(|Y| > T_y \mid \tilde{X}_1 = t)\ f_{\tilde{X}_1}(t)dt.$$

The plot associated to this quantity is displayed on the Figure 4.2. This is the orange curve, which represents the weighted risk. This is the impact risk of out of tolerance rate at requirement level weighted by the probability to encounter a value that exceeds of such a criteria. The scale of this orange curve has been modified in order to visualize this quantity and does not have unit. We can see that only the positive side presents a certain risk. The weighted risk is the area under the orange curve beyond the positive criteria, 2.43. Indeed, the negative acceptance criteria is in the negative tail of the Gaussian density and weighted risk value is very close to 0.

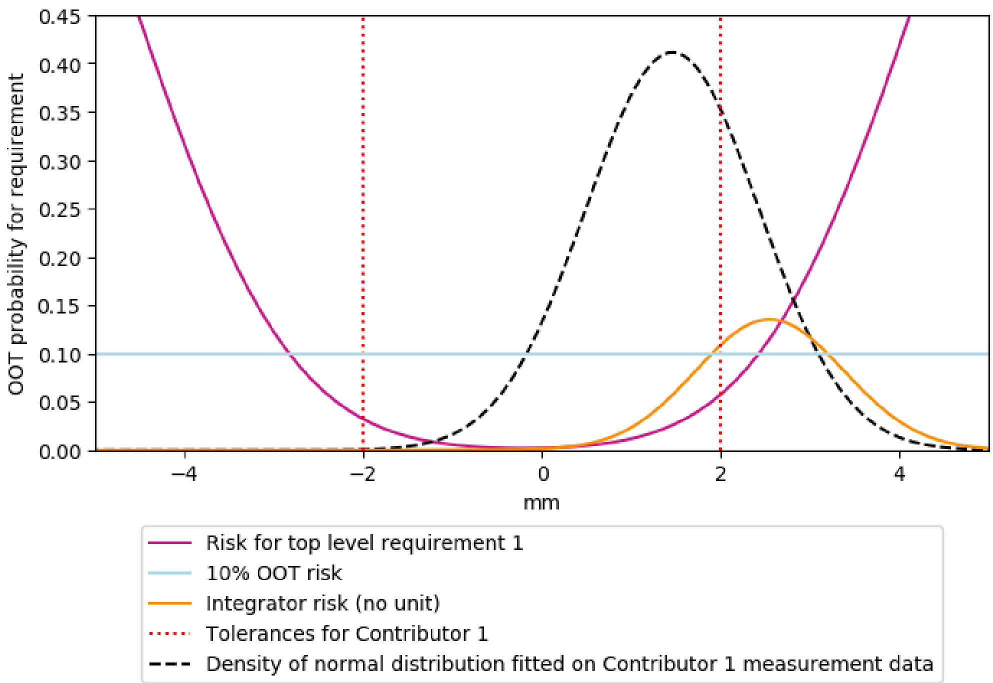

FIGURE 4.2: Advanced results for acceptance criteria analysis

In the example, the weighted risk value is the sum of the two following terms,

$$\int_{t<-2.86} P_N(|Y| > T_y \mid \tilde{X}_1 = t)\ f_{\tilde{X}_1}(t)dt = 0$$

and

$$\int_{t>2.43} P_N(|Y| > T_y \mid \tilde{X}_1 = t)\ f_{\tilde{X}_1}(t)dt = 0.0242.$$

The sum gives a weighted risk of 2.42% therefore these criteria are valid for the considered threshold of 3%.

4.6 Example of bimodal distributions fitted on data

This example focuses on the case of an assembly with one contributor with an associated feature classified as bimodal. We consider a simple simulated assembly with three contributors associated to random variables X_1, X_2 and X_3. These contributors all have a linear influence coefficient of 1, so the random variable associated to the top level requirement of the assembly is $Y = X_1 + X_2 + X_3$. If the contributors are measured, it allows the estimation of the random variable distributions. If the contributors are not measured, the assumption of uniform distribution is made for the associated random variables.

The distribution fitted on measurement data of the contributor X_1 through kernel density estimation presents two modes. The contributor X_2 does not have available measurement data. The acceptance criteria to be defined in this example concern another contributor X_3 in the assembly. Table 4.7 details the stack chains contributors of this example.

Contributors	Tolerance	Measurement data information
X_1	± 1	measured and bimodal or normal distribution
X_2	± 1	non measured, uniform distribution assumption
X_3	± 2	measured and acceptance criteria to be defined

Table 4.7: Stack chain contributors for a bimodal example

The histogram of observations on the contributor X_1 is displayed in Figure 4.3. The blue line represents the empirical density of the distribution fitted on the simulated data with Gaussian kernel. The bandwidth parameter is automatically adjusted and its value is 0.2. The red line is the density of a normal distribution with parameters mean and standard deviation computed on X_1 observations.

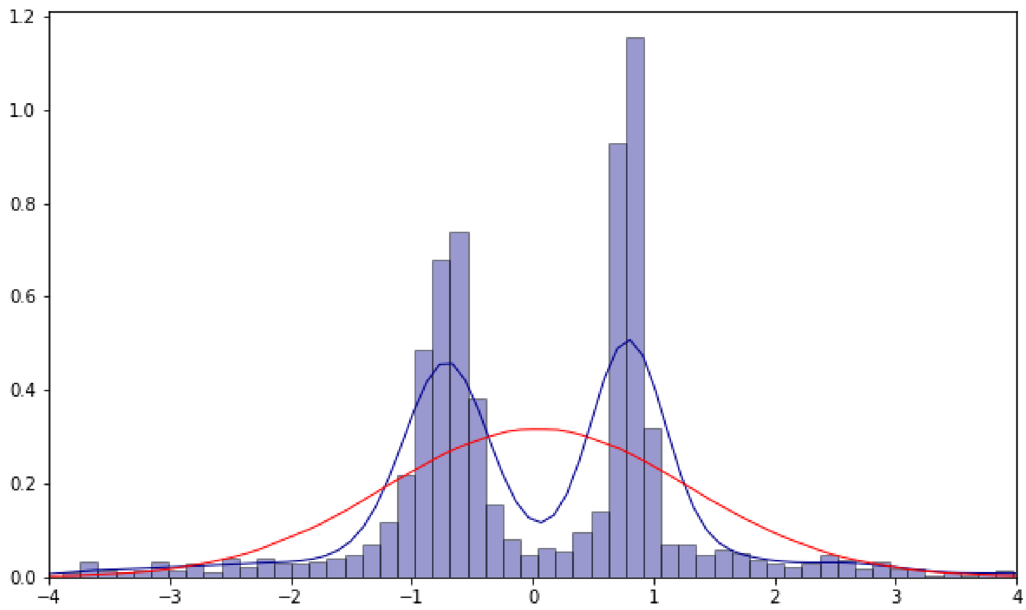

FIGURE 4.3: Example of a empirical compared to normal distribution for X_1

The estimation of the top level requirement distribution is possible through a convolution product of X_1 and X_2. X_3 is assumed to value zero for this illustration. The convolution product is between a bimodal and then a normal distribution for X_1 and a uniform distribution for X_2.

The Open TURNS package ([9]) is used for implementation. The densities in Figure 4.4 are respectively associated to the top level requirement density computed with the empirical density and the normal density fitted on X_1 measurement data. Note that the X-axis have different orders of magnitude.

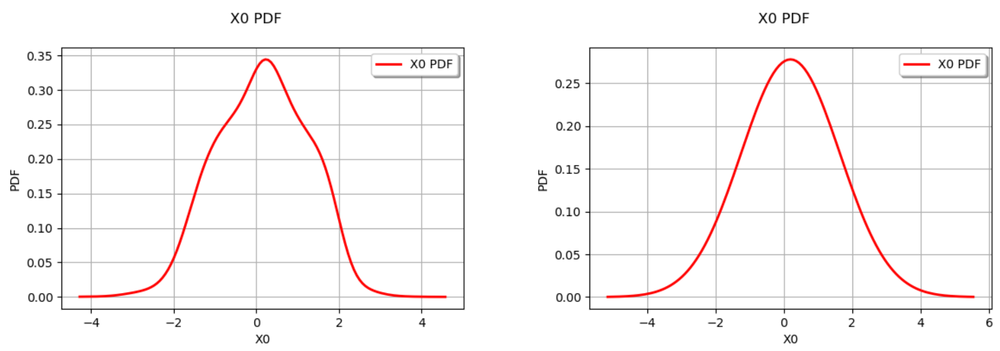

FIGURE 4.4: Top level requirement densities (empirical and normal case considered for the contributor X_1)

The plot representing risk to be out of tolerance according the values possible for the contributor X_3 is displayed in Figure 4.5. The risk computation is performed by the analysis of the convolution of the empirical or normal density of X_1, the uniform distribution of X_2 and the possible value selected for X_3 represented in the X-axis. The blue line is associated to the empirical density estimation of X_1 and the red line is the risk computed with the normal distribution fitted on X_1 measurement data. In the sequel, let us focus on the blue line as the risk curve for X_3 values and used for decision on acceptance criteria.

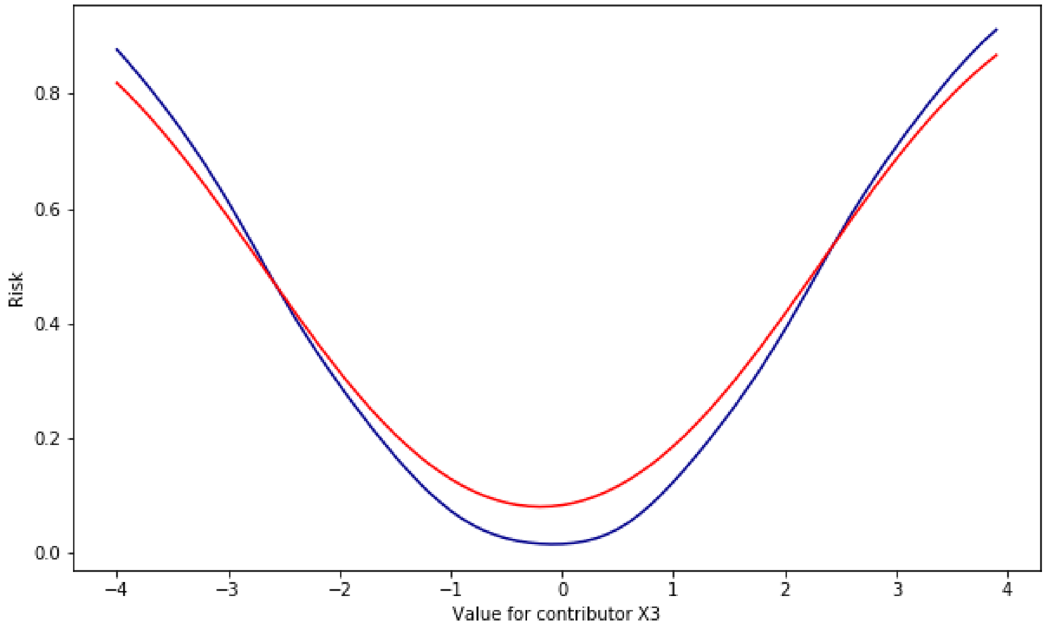

FIGURE 4.5: Result for the risk in the example - comparison between empirical and normal distribution for X_1

At this stage, we have not yet discussed the distribution of the contributor on which acceptance criteria are to be defined. It is the consideration of this distribution that allows us to analyze the probability that the risk presented in Figure 4.5 will occur. Here too, bimodality brings a subtlety. The histogram in Figure 4.6 represents simulated data for the third contributor on which critera need to be defined. As the distribution of the contributor X_3 has all its load at the extremities (see the green line in Figure 4.6), the probability of encountering a values around -2 or 1.3 and thus a high risk at the top level requirement is much higher than for the normal distribution.

Conversely, if the distribution fitted for X_3 is a centered normal as the yellow line in Figure 4.6, the load is concentrated on the center where risk is lower for top level requirement. This may lead to a wrong decision. Indeed, one could consider that a value of -2 for X_3 is unlikely to happen according the normal distribution, while the probability to encounter this value is higher according to the measurement data.

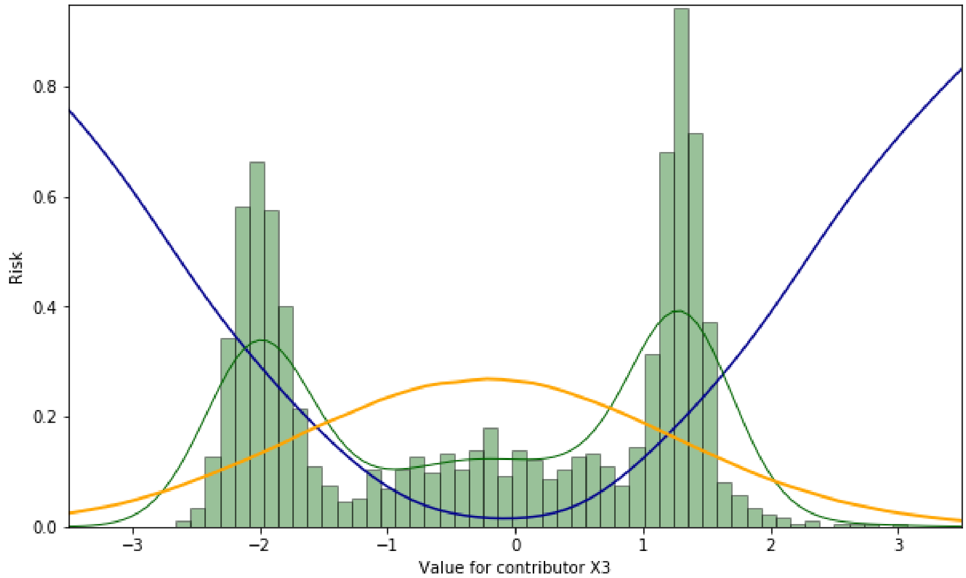

FIGURE 4.6: Risk result and consideration of X_3 distribution

Conclusion

To conclude, this chapter deals with the management of out-of-tolerance risk, in particular during the production phase. Based on a tolerance model and distribution hypotheses, the proposed approach can be adapted to any assembly, whatever the level in the aircraft structure. The approximation of the Poisson summation formula and its implementation in common programming tools allow the convolution product of the assumed laws of the contributors to estimate the output characteristic of the assembly. The interest of the approach is quantifiable by an indicator about the variability of a stack chain, defined according to the different feedback on the contributors.

The benefits of the method are related to the costs of non-conformities, which can be avoided thanks to the proposed approach. A global management tool has been developed to make it easier for users responsible for maintaining tolerances, particularly in the various Airbus plants. This tool will be detailed in Chapter 6.

Bibliography

[1] Alain Van Hoecke. Tool Risk Setting in Statistical Tolerancing and its Management in Verification, in Order to Optimize Customer's and Supplier's Risks. *Procedia CIRP*, 43:250–255, 2016.

[2] Huiwen Yan, Xiangji Wu, and Jiangxin Yang. Application of Monte Carlo Method in Tolerance Analysis. *Procedia CIRP*, 27:281–285, 2015.

[3] Bob J.J. Alting von Geusau. Monte Carlo versus numerical convolutions: Some critical remarks. *European Journal of Operational Research*, 15(2):259–269, February 1984.

[4] Joseph Abate and Ward Whitt. The Fourier-series method for inverting transforms of probability distributions. *Queueing Systems*, 10(1-2):5–87, January 1992.

[5] William Feller. *An introduction to probability theory and its applications. Vol. 2: ...* Wiley series in probability and mathematical statistics. Wiley, S.l.., 2. ed., [nachdr.] edition, 2009. OCLC: 608467098.

[6] R. Lebrun. Efficient time/space algorithm to compute rectangular probabilities of multinomial, multivariate hypergeometric and multivariate Pólya distributions. *Statistics and Computing*, 23(5):615–623, September 2013.

[7] Paul J. Drake, editor. *Dimensioning and tolerancing handbook.* McGraw Hill, New York, 1999.

[8] Geometrical Product Specifications. Inspection by Measurement of Workpieces and Measuring Equipment—Part 1: Decision Rules for Proving Conformity or Nonconformity With Specifications. *ISO Standard*, pages 14253–1, 2013.

[9] Michaël Baudin, Anne Dutfoy, Bertrand Iooss, and Anne-Laure Popelin. Open TURNS: An industrial software for uncertainty quantification in simulation. *arXiv preprint arXiv:1501.05242*, 2015.

Chapter 5

Smart tolerance redefinition

5.1 Context and state of the art

Mathematical optimization techniques have been considered for tolerance allocation, verification and variation management during product life cycle. A vast literature considers the ways to use optimization for tolerance management and tackle different specific problems. A recent and very complete map of the current state of the art about optimization techniques for tolerance related issues is detailed in [1], reviewing close to 300 papers. The aim of this first section is to introduce elementary tools to summarize the background of optimization for tolerancing.

A mathematical optimization problem consists in finding the best candidate available in a set S of allowed values to minimize or maximize an objective function f, also called cost. Optimization problems may be modeled in this general framework

$$\min_{x \in S} f(x). \tag{5.1}$$

In order to determine the optimal set of tolerances through the minimization of an objective function, a large part of researches about tolerance cost optimization take into account manufacturing cost. This type of cost is related to a part production ([2, 3, 4, 5, 6, 7, 8]). These manufacturing costs might be non-linear, for instance if they are related to density functions when considering default probabilities.

Another cost widely studied is the quality cost, often associated in the objective function to a quality loss function ([9, 10, 11, 7, 12, 13, 14, 15]). This quality loss function reflects the quality appreciation of customers on a dimension within (or not) a specified tolerance range. If a dimension is outside the tolerance bounds, satisfaction decrease drastically while a perfect dimension leads to the customer satisfaction. The Taguchi's method is a common approach ([16, 17]) but the function can also be defined by any arbitrary polynomial function with fixed degree. In between tolerance range, loss function provides a model for the appreciation on product quality, which is represented by the *loss*, as displayed in Figure 5.1 in the common approach of Taguchi. Generally, this is a quadratic curve given considering the square of the difference between the actual product dimension and the specified one, weighted by a coefficient as parameter (quadratic loss function). If it is a product population studied instead of only one individual, then variances and average quantities should be considered. More details are available in [18].

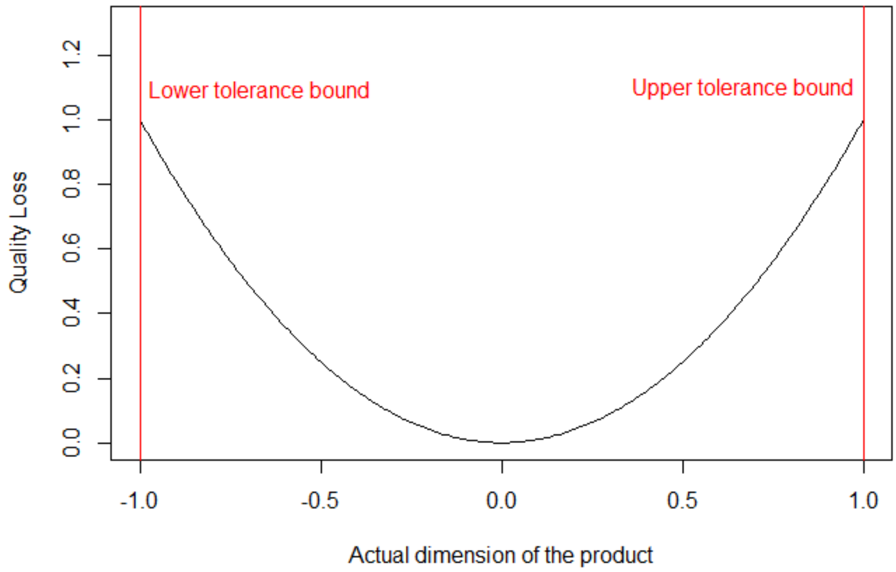

FIGURE 5.1: Quality loss function plot with arbitrary parameters.

Beyond different costs considerations, functional requirements might be considered through constraints into the optimization problem ([2, 3, 5, 6]). Such constraints are taken into account in the set S which only allows values satisfying the functional requirements. The different alternatives for assembly techniques, machines or processes could also be considered in the objective function of the optimization problem ([19, 5, 20, 21, 22, 23, 15]).

This leads to discrete optimization problem, where the set of allowed values S in 5.1 is a discrete set. A common and widely used tool to solve such an optimization problem is the class of genetic (or meta-heuristic or heuristic search) algorithms available both for constrained and unconstrained optimization problems. A review of the use of such algorithms in optimization is available in [24]. Genetic algorithms are used for tolerancing in [20, 25, 26, 23, 27, 8]. An other tool to solve such problem is exhaustive research method ([3, 28]) but such an approach is commonly unfeasible in practice.

Several other tolerancing problems can be considered in optimization problem and associated cost. In [22] and [29], the approaches take into account the interrelation of stack chains in the objective function associated to the optimization problem. In design phase, the verification cost (measurement tools, conformity assessment, ...) are considered in [30] in order to define best tolerances considering inspection cost. Another parameter which is the product degradation and time value of money is taken into account in [11] for the the minimization. Planning and event sequence for an assembly is considered in [31] and [32]. Assembly technique for a joint is considered in [33]. In [34] the notion of assembly step is taken into account by the system, meaning where the tolerance is logged in the production process. In [35], time variant deviation such as deformation, thermal expansion or parts mobility are specifically addressed in an optimization algorithm based on particle swarm. Finally, recycling cost is also a candidate for optimization variables as considered in [36]. More generally, [37] proposes the minimization both of economic and ecological costs, illustrated on automotive industry and materiel choice.

Our work is driven by industrial needs in tolerancing encountered at Airbus and has to suit available data flow. The approach considers cost related to non-quality and process to change a tolerance value. The proposed approach is a step that comes well beyond the development and tolerance allocation phase. The hypothesis is that a tolerance sharing is already in place and we consider a re-sharing. We focus on the goal to reduce money spent when non conformities occur without spending too much for tolerances change in drawings. This approach should be implemented in the internal tools on all stack chains with available measurement at all assembly levels for all programs. It needs to provide proper results in a reasonable amount of time.

5.2 Motivations

We propose to refine the tolerancing approach by identifying assemblies with capability disparities between contributors. Indeed, if a contributor is better produced than expected when tolerance has been designed, another contributor can benefit from this positive margin to enlarge its tolerances. Review a tolerance sharing within the stack is an easy way to reduce cost related to poor capabilities without changing assembly process or tooling

Several criteria have to be taken into account in order to identify the best opportunities for tolerance sharing. For instance, the number of drawings impacted by an assembly contributor or a top level requirement gives an information on how difficult it would be to initiate a change of design. The number of non conformity logged for a tolerance feature also gives an information on how valuable would be the change of design. Moreover, capabilities indicators such as Cp and Cpk (see [38]) can be used if feedback is available in order to consider the feasibility of a tolerance optimization.

Once opportunities for tolerance sharing are identified as detailed in Section 5.3, an optimization approach helps to find the best re-sharing solution. Criteria for the optimization objective function needs to be defined upstream and a relevant choice about costs to be minimized is discussed in Section 5.4. Then, the choice of a methodology and the implementation of optimization solver through different approaches are detailed in Section 5.5. Finally, a concrete application of the proposed methodology is presented in Section 5.6.

5.3 Tolerance re-sharing opportunities identification

A first crucial step amounts to clarify what is a *good* tolerance sharing. Indeed, this approach allows to harmonize and align objectives between different functions before considering new tolerance sharing within stack chains. This step also helps to define a concrete use case that appears to be relevant for demonstration of the added value in an industrial context.

Several factors are considered in the Airbus industrial context and allow to define a global indicator that can be used to discuss a tolerance sharing and assess its relevance. Among these various factors, we identified fields of criteria:

- **Model accuracy** - This refers to factors related to the accuracy of the tolerance model and eventually to the enhancement of this model.
- **Quality performance** - This includes indicators within the scope of quality control on measurement data.
- **Cost of implementation** - This concerns the cost represented by procedures to be implemented in order to modify a tolerance sharing.

Through these fields, performance of each criteria can be managed and support decision process. The table 5.1 summarizes factors identified as relevant to assess a tolerance sharing within a stack chain.

Cluster	**Criteria**
Model accuracy and feedback quality	Identified bias during model correction
	Ratio Q from Chapter 4, Equation (4.6)
	Percentage of measured contributors in the stack chain
	Measures availability on output feature of the assembly
Quality performance	Cp and Cpk of assembly contributors
	Cp and Cpk of output feature of the assembly
	Out of tolerance (scrap rate) at contributors level
	Out of tolerance (scrap rate) at top requirement level
Cost of implementation	Number of drawings impacted by the stack chain
	Number of contributors in the assembly stack chain

TABLE 5.1: Examples of criteria for tolerance re-sharing opportunities

Industrial stakeholders can define a score based on criteria displayed in Table 5.1 in order to measure the quality of a tolerance sharing. The aggregation of the various factors can be managed through a mean after normalization of the different variables. A weighting system could complete this score evaluation according to industrial reasons because some criteria are more significant than others.

The following section describes the modeling of the optimization problem considered in our work.

5.4 Problem statement

In a first approach, let us focus on an assembly with p contributors, represented by random variables $X_1, \dots, X_p$. We might have feedback measurement data on these contributors (or not). For each contributor, the definition of tolerance will induce costs of out of tolerance if the support of their associated distribution exceeds the tolerance interval. The tighter the contributor tolerance, the more expensive is the price. If measurement data are available for a contributor, these out of tolerance costs represent a reflection of the capacity to produce complying with the tolerance required.

If a contributor is observed, we are able to assess the out of tolerance rate that we expect for this contributor according to its tolerance bounds. This is a criteria that will be taken into account in the stack chain optimization. A perfect stack chain should ensure the consistency between the tolerance bounds defined for contributors and the tolerated interval for the top level requirement. Tolerance model used by tolerancing specialists allows to have a prediction of the top level requirement distribution when we have observations for contributor input features. Again, whatever the target bound value for this top level feature, we assume to be able to assess about the non conformity rate at the top level assembly output. The article [39] gives some methods to estimate such a scrap in order to perform tolerance sharing optimization.

In production phase, a change of tolerance design involves costs related to this modification. However, no matter how different is the new design from the old stack chain, the cost remains the same. The criteria which is relevant is a constant cost if a contributor tolerance interval is modified within a stack chain.

The problem to be solved is similar to multi-criteria optimization. Indeed, it involves three objective functions that are to be minimized and the result is set of solutions that define the best tradeoff between competing objectives. The different approaches and algorithms to be considered for this type of problem are detailed in [40]. We have chosen the so-called Weighted Sum Method which consists in scalarize a set of objectives into a single objective by adding each objective pre-multiplied by a user supplied weight. The following section describes in detail the modeling of the optimization problem considered in our work.

5.4.1 Variables and parameters definition

The quantities to be considered in our optimization approach in a stack chain are:

- **Parameters**: the targeted tolerance range for the assembly requirement, the contributors initial tolerances and their observed variability range according to feedback measurement data.

  We denote by v_y the tolerance bound targeted for the top level requirement of the stack chain. Hereafter, we consider p contributors in a stack chain representing an assembly. The contributors initial tolerances are treated as parameters in the optimization problem and denoted by $v = (v_1, ..., v_p) \in \mathbb{R}^p$.

- **Variables**: the proposed tolerances for each contributor and the fact to modify or not the initial as-designed tolerance of each contributor.

Let us define $(x, t) = ((x_1, \ldots, x_p), (t_1, \ldots, t_p)) \in \mathbb{R}^p \times \{0, 1\}^p$ already stated. The variables $(x_1, \ldots, x_p)$ are the proposed tolerance bounds for contributors and $(t_1, \ldots, t_p)$ are binary indicators to assess about the change or not of contributors tolerance bounds. The binary indicator is equal to 1 if the initial tolerance is changed for the contributor, and 0 otherwise.

5.4.2 Objective function

Let us introduce $c_1 : \mathbb{R}^p \times \{0, 1\}^p \longrightarrow \mathbb{R}$, $c_2 : \mathbb{R}^p \times \{0, 1\}^p \longrightarrow \mathbb{R}$ and $c_3 : \mathbb{R}^p \times \{0, 1\}^p \longrightarrow \mathbb{R}$ the cost functions associated to non-quality for contributors, non-quality for top level feature and a contributor tolerance change respectively.

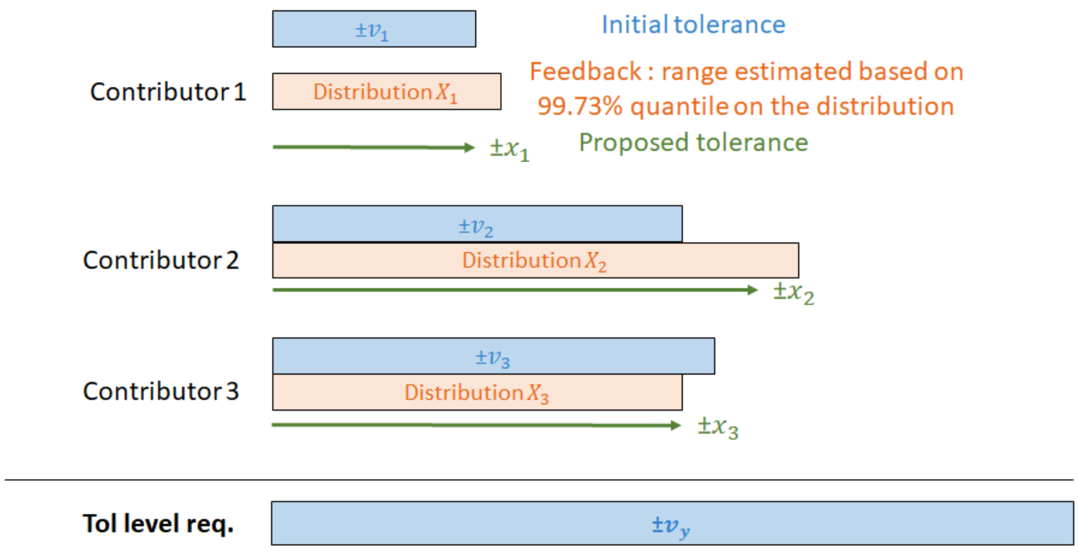

FIGURE 5.2: Stack chain example with 3 contributors.

The non-quality is expressed as the standardized sum of the probabilities to be out of designed tolerance interval for each contributor,

$$c_1(x,t) = \frac{1}{p}\sum_{i=1}^{p}\left[t_i\mathbb{P}\left(|X_i| > x_i\right) + (1-t_i)\mathbb{P}\left(|X_i| > v_i\right)\right],$$

where $X_1, \ldots, X_p$ are the random features associated to each contributor. Their distributions can be estimated through the measurement analysis and density estimation. For the i-th contributor, the tolerance limit taken into account is either x_i if the tolerance is modified, or the initial tolerance v_i otherwise.

This cost relates to the out of tolerance rate of all the contributors. The purpose of the term $\frac{1}{p}$ is to normalize this cost, which is then homogeneous to a probability whatever the number of contributors in the stack chain. This implies that each contributor has the same importance in the optimization of its non quality. A refined approach could be envisaged, so that the most important contributors in the chain are preponderant. For this, one could for example replace the standardization term $\frac{1}{p}$ by a set of differentiated terms for each contributor $\left\{\frac{x_1}{\sum_{i=1}^{p} x_i}, \ldots, \frac{x_p}{\sum_{i=1}^{p} x_i}\right\}$ or $\left\{\frac{x_1^2}{\sum_{i=1}^{p} x_i^2}, \ldots, \frac{x_p^2}{\sum_{i=1}^{p} x_i^2}\right\}$.

The probability for the top level feature to be out of its targeted interval is highly dependent of tolerance bounds applied to contributors via the linear tolerance model.

$$c_2(x,t) = \mathbb{P}\left(|Y(x,t)| > v_y\right),$$

where v_y is the target interval for top level feature $Y(x,t)$ which distribution depends now on the pair (x,t). Distributions densities can be estimated based on available information about assembly contributors. In the particular hypothesis of a Gaussian framework with independent features, $Y(x,t) \sim \mathcal{N}\left(0, \sqrt{\sum_{i=1}^{p}\left(\alpha_i\left(t_i x_i + (1-t_i)v_i\right)\right)^2}\right)$. Otherwise, any distributions could be considered using different estimation methods as presented in Chapter 4.

The cost of change is represented as a unit cost,

$$c_3(x,t) = \sum_{i=1}^{p} t_i.$$

The optimization problem can then be formulated as the following non-linear mixed integer programming (MINLP)

$$\min_{(x,t)\in\mathbb{R}^p\times\{0,1\}^p} \left\{\lambda_1 c_1(x,t) + \lambda_2 c_2(x,t) + \lambda_3 c_3(x,t)\right\}$$

where $\lambda_1, \lambda_2, \lambda_3 > 0$ are setting parameters to be defined according to engineering judgment. From an industrial point of view, these coefficients make it possible to adjust the relative importance of the costs in relation to each other. Often, one can consider a close order of magnitude for λ_1 and λ_2 because both represent a probability. The λ_3 represents a threshold from which it is beneficial to change the tolerance.

If $\lambda_1 = \lambda_2 = 0$, it means only the third cost is considered in the optimization problem. The proposed solution will be to not change any tolerance since any change would increase the third cost, the only one considered with this setting.
If $\lambda_1 = \lambda_3 = 0$, only the second cost is active, meaning that the solution is to reduce as much as possible the tolerance intervals of the contributors in order to reduce the rate of out-of-tolerance of the output to 0%. All the tolerances can be changed without any restriction since the third cost is inactive.

Conversely, if $\lambda_2 = \lambda_3 = 0$, the solution is to increase the contributors tolerance ranges so that the sum of out-of-tolerance rates of inputs is minimized. The rate of out-of-tolerance at the output will be very high since its cost will not be taken into account.

5.5 Methodology

This section presents different methods to solve the optimization problem previously stated.

5.5.1 Non-linear solver for continuous variables

In a first approach, we checked the validity of costs selected for the objective function by simplifying the non-linear mixed integer problem in order to quickly solve it as a continuous non-convex problem through Newton or Quasi-Newton's methods. As our optimization problem involves a discrete variable, the first step was to modify these binary indicators and cost function.

Relaxation of the problem

The previously presented optimization problem involves the variables $(t_1, \dots, t_p) \in \{0,1\}^p$. In order to use common solvers, we relaxed these binary indicators to continuous variables $\left(t'_1, \dots, t'_p\right) \in [0,1]^p$.

As we still want these variables to represent a tolerance change, the cost function of change $c_3(x,t) = \sum_{i=1}^p t_i$ has been changed for a continuous function $c'_3 : \mathbb{R}^p \times [0,1]^p \longrightarrow \mathbb{R}$

$$c_3(x,t') = \sum_{i=1}^{p} t'_i.$$

We now need to control the values of $\left(t'_1, \dots, t'_p\right)$ between 0 and 1 through an approximation function to represent how expensive is a tolerance change.
An arbitrary choice for any $i \in 1, \dots, p$,

$$t'_i = 1 - \exp\left(\frac{-\left(x_i - v_i\right)^2}{\omega}\right)$$

where ω is a running parameter related to the relaxation accuracy. This function is zero if initial tolerance is unchanged, and quickly grows to one if initial tolerance is changed. The main difference with the previously defined discrete function is that a slight modification of the initial tolerance is now allowed without increasing the cost of change. This inaccuracy adjusted by the ω parameter is acceptable from industrial perspective if we deal with large roundings for tolerance management. Figure 5.3 shows the value t'_k of an example contributor with initial tolerance value $v_k = 1$. Several values for ω parameter are displayed:$\omega = 10^{-3}$, $\omega = 10^{-4}$, $\omega = 10^{-5}$ and $\omega = 10^{-6}$ from smooth to more constrained.

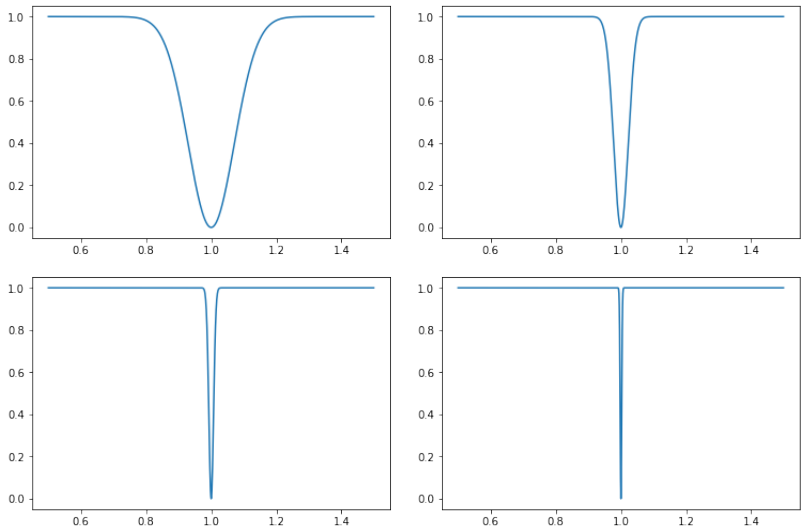

FIGURE 5.3: Illustrations of accuraccy ajusted by $\omega = 10^{-3}$, $\omega = 10^{-4}$, $\omega = 10^{-5}$ and $\omega = 10^{-6}$ for t'_k with an example $v_k = 1$.

Vector t' is entirely dependent on variables x and parameter v so we could drop it in the problem formulation. Finally, the relaxation form of the optimization problem is given

$$\min_{x \in \mathbb{R}^p} \{\lambda_1 c_1(x) + \lambda_2 c_2(x) + \lambda_3 c_3(x)\}$$

where $\lambda_1, \lambda_2, \lambda_3 > 0$ are setting parameters to be defined according to engineering judgment about importance of each cost. This expression is a minimization of a scalar function without constraints. Many options to solve it are available and we consider the Sequential Least Squares Programming (SLSQP) based on Newton's method, the derivative free Nelder-Mead algorithm (simplex) and the Limited memory-BFGS-B algorithm of the quasi-Newton's family.

Starting point specificity and limitations

The selected algorithm will converge to local minima and do not ensure the global optimal solution is reached. With the problem simplification and the introduction of the continuous cost of change, the objective function will become 'bumpy' and many local minima are generated in the space of solutions. The solution obtained is often unsatisfactory. These approaches are also highly dependent on the starting point, which leads to distinct local minima. To address these shortcomings, we implemented a multi-start strategy to test various starting point in order to increase the chance to reach a good local solution. We propose an experimental design based on combinations between initial tolerance value and observed tolerance bound for each contributor. This observed tolerance bound can be estimated as 99.73% quantile of the contributor feature distribution, as it is commonly assumed with the 6σ rule in the Gaussian framework. This strategy allows to cover a large part of the solution space. Among all the solutions given by the different starting points, taking the minimum gives a solution which is the better option to the optimization problem.

The limitations of this approach are the number of optimization to be performed in order to increase the chance to reach a good local optimal solution. Indeed, our experimental design for a multi-start strategy leads to run 2^p optimization loops and required a significant amount of time to reach the solution.

5.5.2 Greedy algorithm

A greedy algorithm is an iterative algorithm to approximately solve an optimization problem. It follows a defined heuristic to find the locally optimal solution at each stage. It might not lead to a global optimum. In some cases, a well-defined heuristic can provide a good approximation of the global solution. The advantage of this approach is that it runs and give a solution in a reasonable amount of time.

Problem solving and heuristics specificity

We remind the MINLP problem that needs to be solved

$$\min_{(x,t)\in\mathbb{R}^p\times\{0,1\}^p} \{\lambda_1 c_1(x,t) + \lambda_2 c_2(x,t) + \lambda_3 c_3(x,t)\}$$

where $\lambda_1, \lambda_2, \lambda_3 > 0$ are setting parameters to be defined according to engineering judgment about importance of each cost.

In the implementation of a greedy algorithm in our context, the iteration relies on the number of tolerance changed within the stack chain. Indeed, we consider differently the binary variables to assess about a tolerance change, as the algorithm is divided in several stages. Each stage represents several objective function evaluations and the one giving the best option is selected to go to the next step. In such a tree structure, there are as many stages as there are contributors in the stack chain.

The initial step consists in allowing the change of all the p contributors tolerances. Then, next stages allows to change only $p-1$ tolerances, and so on. At each step, the objective function is computed for each combination of tolerance change among the number of contributors allowed to change. The cost c_3 depends on the step of the greedy algorithm and therefore the number of tolerance modifications. While there is a result lower than the previous evaluation of the objective function, the tree traversal continues. It means that the decrease in the cost c_3 compensates for the increase in costs $c_1 + c_2$.

The decision on which contributor tolerance should be modified is driven by result obtained at each stage for objective function. At each step, the contributor tolerance which becomes fixed is the one with the lowest evaluation of the objective function and that is the greedy aspect of the algorithm. From a step to the next one, the cost of change (represented by c_3) is reduced as a tolerance contributor is fixed at each stage. The global cost might increase or decrease depending on other costs c_1 and c_2 related to inputs and output features quality (scrap rate). This is a trade off between costs c_1 and c_2 on one side and c_3 on the other side.

Algorithm 1: Greedy algorithm for tolerance resharing optimization

Maximum number of iteration: p ;

Initialization of iteration: $i = 0$;

Initialization of available changes of tolerances: all tolerances ;

Initialization for continuous solver: feedback tolerance ranges;

Result: Optimized set of the stack chain tolerances

while *objective function for* i *lower than for* $i - 1$ *AND* i *is lower than* p **do**

  List every combination of tolerance changes $p - i$ tolerance changes;

  forall *possible combination in this iteration* i **do**

    Quicksort of the objective function evaluations;

    Stock the contributor index to be fixed for the next iteration;

  end

  Evaluate objective function ;

  Obtain the associated solution: optimized set of the tolerances;

  Define the index for the tolerance to be frozen;

  Next iteration: $i = i + 1$;

end

In Figure 5.4, we give an example of the tree structure with a stack chain involving three contributors. The red dots means that the contributor is not allowed to change its tolerance while a green dots means the change is possible. In this example, Step 1 consists of three evaluations for the different combination red/green. The lowest value for the objective function is obtained with the first combination: green, green, red. This implies this value is also lower than the reference evaluation of Step 0. Then in Step 2, two values for the objective function are evaluated. Again, if the lowest value of these two evaluations is lower than Step 1 evaluation, the tree traversal continues and it comes to Step 3. In this simple case, it means that the best solution according our greedy algorithm is to not change anything in the stack chain tolerances. In more complex cases, we expect that the descent stops before reaching the tolerance sharing *status quo*.

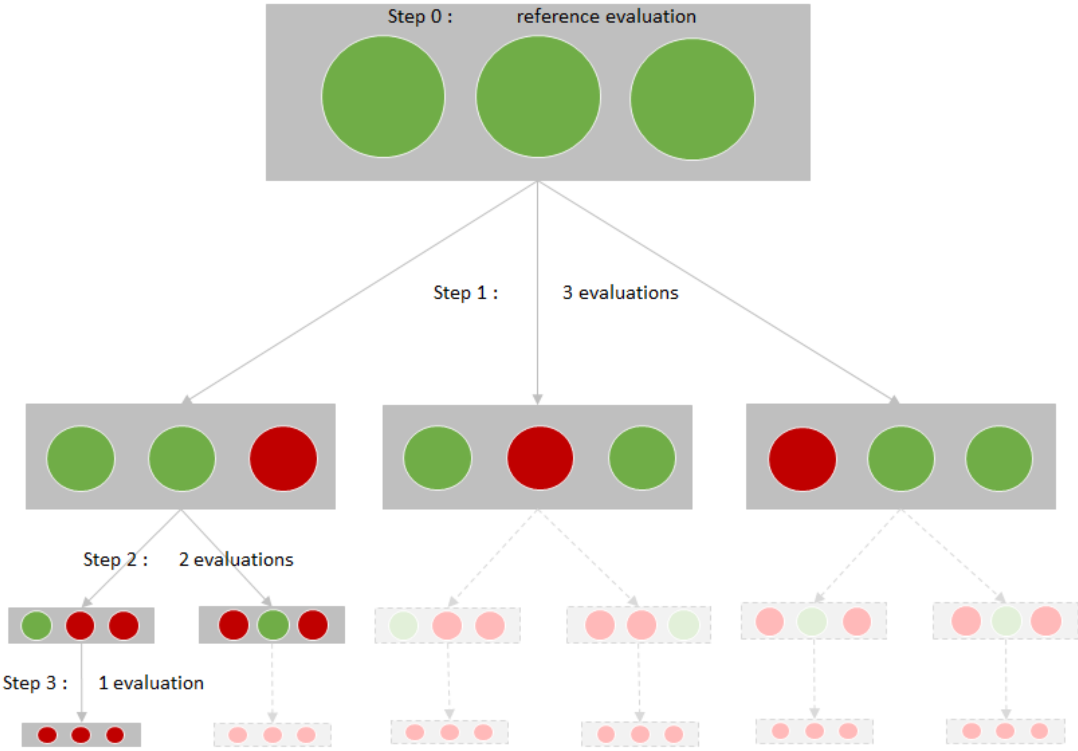

FIGURE 5.4: Outline of the greedy algorithm on a simple example.

Special case of global optimality for a greedy algorithm

A greedy algorithm does not guarantee an optimal global solution. In some cases this approach allows to reach a global optimum, but this is not the general case. Several well-known optimization problems illustrate the failure of a greedy algorithm to find a global optimal solution. For instance, this might be the case for some problems of change making (even if usually a greedy algorithm gives a global solution to this class of problem). The goal is to minimize the number of coins when making change. A greedy approach consists in systematically choosing the coin of higher value that is not higher than the remaining quantity to be returned ([41]). Let's take the example of a coin return of 6 with the denominations of the coins were 1, 3 and 4. The greedy algorithm will give the solution 4,1,1 for the change. However, a better solution to the problem is to render two coins of 3, and this is the global optimal solution. In the majority of cases, it is difficult to show that a greedy algorithm allows to obtain an optimal global solution. The point is to show, if a choice that seems the best at the moment is made, we should never reconsider an earlier choice. This is expressed by two properties that need to be verified to ensure that a greedy algorithm achieves an optimal global solution.

- **Property of the first greedy choice -** There is always an optimal solution which contains a first greedy choice.
- **Optimal sub-structure property -** Every optimal solution contains an optimal substructure. The substructure represents the resulting sub-problem when a choice on the problem is made.

In the specific case of tolerance sharing, these two properties are not trivial to demonstrate. The greedy algorithm as described above is based on the evaluation of contributor combinations to be modified in the stack chain. Therefore, the same solution may appear in several branches of the tree structure of the algorithm.

Initially, the cost of change is maximum as all contributors are allowed to change. The following step will freeze the tolerance which is the least beneficial when modified. And so on at each step until the benefit on the cost of change related to the freezing of a tolerance does not outweigh the gain in terms of the other two costs. From the stack chains studied, it seems that the sequence of operations to freeze tolerances at each stage does not matter. Intuitively, this means that the best solution can be reached by following different branches. In the industrial context and in all the cases studied, the greedy algorithm led to appropriate optimal solutions and these results are satisfactory.

5.5.3 Branch and Bound

Tailored for integer programming problems, branch and bound principle is based on state space search process. Solutions set are divided into subsets treated as branches, and forming a tree. A solution is searched thanks to the tree exploration, relying on successive estimation of the lower and upper bounds of branches of the search space.

Background

The branch and bound principle was first presented in [42] for discrete linear programming. Later on, binary problems have been discussed by [19] and [43]. In the following years, many works related to branch and bound techniques have been published, either for discrete, mixed, linear and non-linear optimization problem.

Branch and bound algorithm aims to find optimal solutions of various optimization problems, especially in discrete and combinatorial optimization. The principle of this general algorithm consists in a systematic enumeration of all candidate solutions in the space state. The candidates are then linked and clustered through a tree structure. This is the first step of the algorithm, called branching step. The second step is to drop large cluster of poor candidates (i.e. branches), in order to isolate the global optimal solution. In general, this is possible by using upper and lower estimated bounds of the objective function through relaxation of discrete variables. The confrontation of these evaluations for different branches allows to drop some of the tree branches where solution candidates are not enough competitive. This is the bounding step of the branch and bound technique. The details of the algorithm are available in [44] and [45].

Problem solving

Let us remind the initial MINLP problem that refers to optimization problems with continuous and discrete variables and nonlinear functions as objective functions:

$$\min_{(x,t)\in\mathbb{R}^p\times\{0,1\}^p} \{\lambda_1 c_1(x,t) + \lambda_2 c_2(x,t) + \lambda_3 c_3(x,t)\}$$

where $\lambda_1, \lambda_2, \lambda_3 > 0$ are setting parameters already stated.

To ensure that changing a tolerance impact the associated cost, the following constraint between x and t value is implemented, for any $i \in \{1, \ldots, p\}$,

$$(1 - t) \times (x_i - v_i) = 0.$$

We also implement an advanced constraints system in order to properly manage and differentiate the increase or decrease for a contributor's tolerance. To this end, we decompose the binary variables t_i into t_i^+ and t_i^- in order to discriminate between an increase and a decrease of tolerance for each contributor respectively. It may be useful for user to distinguish increase and decrease because it may lead to different costs, production systems or business constraints.

As a tolerance can only remain identical, increase or decrease its initial value, constraints system ensures this consistency. Obviously, a tolerance can not increase and decrease at the same time. The only possible configurations related to the binary variables for the i-th contributor is

- $t_i^+ = 0$ and $t_i^- = 0$, if the initial tolerance is not modified ($x_i = v_i$).
- $t_i^+ = 1$ and $t_i^- = 0$, if the proposed tolerance is greater than the initial ($x_i > v_i$).
- $t_i^+ = 0$ and $t_i^- = 1$, if the proposed tolerance is lower than the initial ($x_i < v_i$).

The constraints modelling is, for any $i \in \{1, \ldots, p\}$,

$$\begin{cases} t_i^+ \times t_i^- = 0 & \text{because either } t_i^+ = 1 \text{ or } t_i^- = 1, \\ t_i^+ \left(1 - t_i^-\right) \times (x_i - v_i) \geq 0 & \text{to manage } x_i \geq v_i, \\ t_i^- \left(1 - t_i^+\right) \times (x_i - v_i) \leq 0 & \text{to manage } x_i \leq v_i, \\ \left(1 - t_i^+\right)\left(1 - t_i^+\right) \times (x_i - v_i) = 0 & \text{to manage } t_i^+ = t_i^- = 0. \end{cases} \tag{5.2}$$

This constraints system ensures the relevance of the solution for a tolerance sharing.

Implementation

The optimization problem is formalized through a script using AMPL language (A Mathematical Programming Language). AMPL supports different solvers, either open source or commercial software. For instance, available solvers are CBC ([46]), CPLEX ([47]), FortMP ([48]), Gurobi ([49]), MINOS ([50]), IPOPT ([51]), SNOPT ([52]), KNITRO ([53]) or LGO ([54]).

For our experimental study, the branch and bound approach has been implemented through the Knitro solver ([53]). This solver particularly addresses large scale nonlinear mathematical optimization problems. It allows to solve MINLP through Nonlinear

Branch and Bound, Mixed-Integer Sequential Quadratic Programming (MISQP) and Quesada Grossman algorithm. The method used is the hybrid Quesada-Grossman (HQG) method based on the algorithm described in [55].

The objective function involves non linear functions which require specific library. Indeed, costs related to input and output non quality are based on distribution functions and survival functions. To this end, we manipulate functions defined in GSL, the GNU Scientific Library, a collection of numerical routines for scientific computing written in C.

5.6 Application

Let us introduce examples to illustrate methodologies presented in this chapter. The first part of this section presents a simple assembly in order to visualize the optimization structure problem. The second part focuses a more complex assembly closer from a real industrial case and associated results and comments.

5.6.1 A simple example

The first example involves only two contributors in order to illustrate the optimization problem through 3D plots.

Assembly description and data

The first example involves only two contributors in order to illustrate the optimization problem. Table 5.2 summarizes information about this assembly, involving initial tolerance value, feedback measurement and target for the top level requirement. For the sake of readability, influence coefficients in the linear model are all assumed to be equal to 1.

	Initial tolerance value	Observed tolerance interval
Contributor 1 (X_1)	± 1	± 3
Contributor 2 (X_2)	± 2	± 0.5
Top level requirement (Y)	± 2.3	

TABLE 5.2: Simple example - Stack chain information

Objective function visualization and result

The first cost about inputs quality is given by

$$\begin{aligned} c_1(x_1, x_2, t_1, t_2) = \frac{1}{2}(&t_1\mathbb{P}\left(|X_1| > x_1\right) + (1-t_1)\mathbb{P}\left(|X_1| > v_1\right) \\ &+t_2\mathbb{P}\left(|X_2| > x_2\right) + (1-t_2)\mathbb{P}\left(|X_2| > v_2\right)) \end{aligned} \tag{5.3}$$

where $x_1, x_2 \in \mathbb{R}$, $t_1, t_2 \in \{0, 1\}$ and finally $v_1 = 1$ and $v_2 = 2$ as stated in Table 5.2.

The second cost about inputs quality is

$$c_2(x_1, x_2, t_1, t_2) = \mathbb{P}\left(|Y(x_1, x_2, t_1, t_2)| > v_y\right) \tag{5.4}$$

where $x_1, x_2 \in \mathbb{R}$, $t_1, t_2 \in \{0, 1\}$ and $v_y = 2.3$ as stated in Table 5.2.

The third and last cost about inputs quality is defined as

$$c_3(x_1, x_2, t_1, t_2) = \mathbb{1}_{\{x_1 \neq v_1\}} + \mathbb{1}_{\{x_2 \neq v_2\}} = t_1 + t_2. \tag{5.5}$$

The total cost is the sum of these 3 functions, weighted by parameters $\lambda_1, \lambda_2, \lambda_3$ associated to each cost. Here, we arbitrary consider $\lambda_1 = 20$, $\lambda_2 = 20$ and $\lambda_3 = 20$ for illustration purpose. Note that first cost is the standardized sum of two probabilities, second cost is one probability expression while third cost is unitary. We place ourselves in the context of the continuous relaxation of the third cost to represent the objective function. With Gaussian hypothesis on contributors and top level requirement distributions and using $c'_3(x, t') = t'_1 + t'_2$ as detailed in 5.5.1, we can visualize the total cost function in Figure 5.5. Distributions are supposed to be centered normal distributions with parameters obtained for the 6σ rule: $X_1 \sim \mathcal{N}(0, v_1/3)$, $X_2 \sim \mathcal{N}(0, v_2/3)$ and $Y \sim \mathcal{N}(0, \sqrt{x_1^2 + x_2^2}/3)$. From the plot, we see that the cost function is not convex.

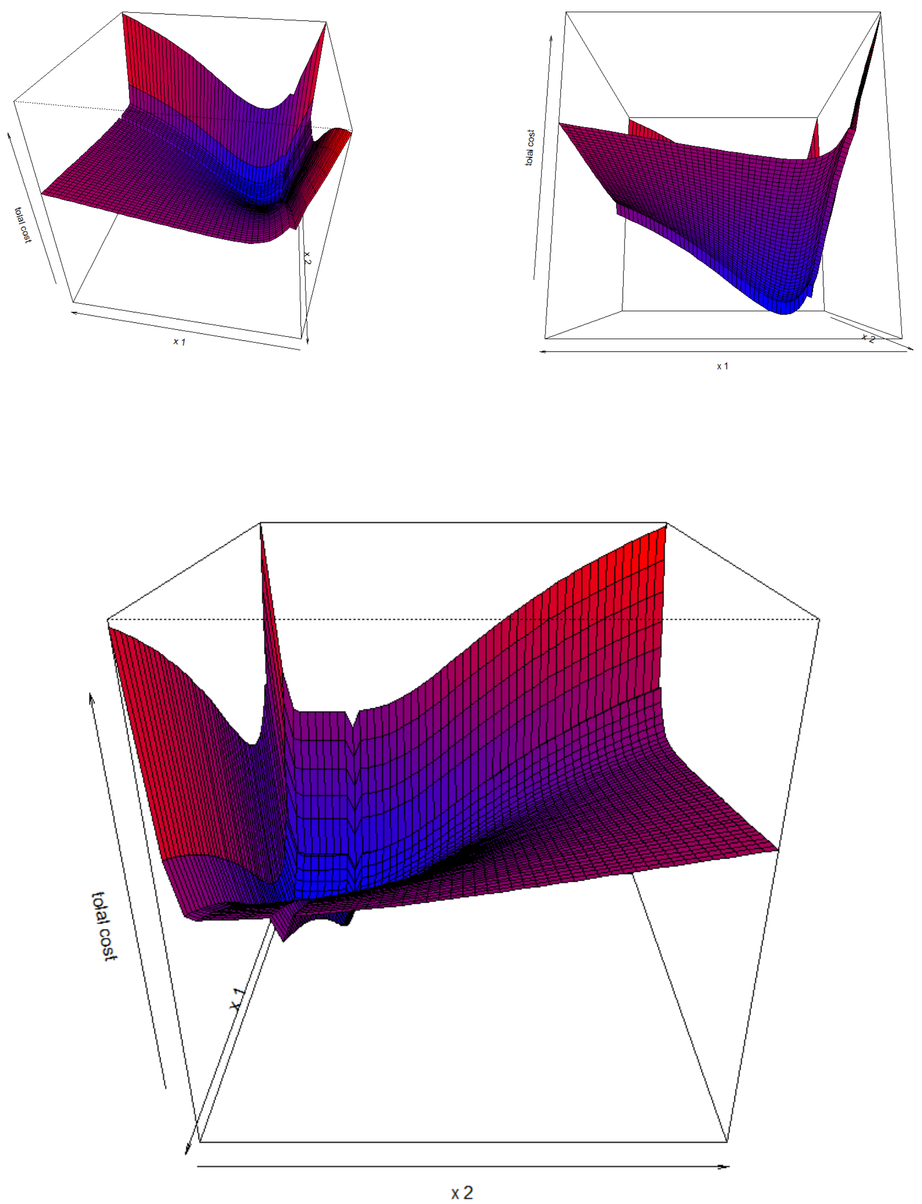

FIGURE 5.5: Total cost function according to costs of Contributor 1 and Contributor 2 when c_3 is replaced by c'_3.

While the surface seems smooth enough, we clearly distinguish the impact of the third cost c'_3 approximated to be continuous which draw two furrows around initial values of the two stacks. When dimension increases, such local minima increase as well. The strategy to alternate starting points between initial and observed tolerance value helps to reach the global minimum of the problem.

According to our hypotheses, the best solution of the optimization problem is $x_1 = 2.45$ and $x_2 = 0.49$. This global minimum is obtained with the Nelder-Mead method when initialization is the set of observed tolerance ranges. The starting point at the initial tolerance values leads to the local minimum $x_1 = 1$ and $x_2 = 2$. The greedy algorithm and the branch and bound lead to the same result for this simple example. The next part addresses a more complex assembly and optimization problem to be discussed, considering different ways to solve.

5.6.2 A more advanced example

Assembly description and data

Table 5.3 summarizes the information about the stack chain considered in this example. There are $p = 5$ contributors, each with an influence equal to 1 ($\alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = \alpha_5 = 1$). The tolerance limitation for top level requirement in this stack chain is equal to a range ± 11.5, which is the result provided by the ASCR approach detailed in Chapter 2.

	Initial tolerance value	Observed tolerance interval
Contributor 1 (X_1)	± 1	± 1.1
Contributor 2 (X_2)	± 2	± 2.8
Contributor 3 (X_3)	± 3	± 2
Contributor 4 (X_4)	± 4	± 4.4
Contributor 5 (X_5)	± 5	± 4.9
Top level requirement (Y)	± 11.5	NA

TABLE 5.3: A more advanced example - Stack chain information

Results, comments and discussion

The three approaches respectively described in Sections 5.5.1, 5.5.2 and 5.5.3 are applied. The calibration coefficient selected for this example are $\lambda_1 = 10$, $\lambda_2 = 1$ and $\lambda_3 = 0.01$.

The non-linear solver for continuous variables (Sequential Least Squares Programming) is used to solve the relaxed optimization problem, with a constant $\omega = 0.001$. The computation time is long enough as $2^5 = 32$ optimization loops are required by the experiment design for this approach as explained in 5.5.1. The greedy algorithm required three steps since two tolerances are changed. The branch and bound approach is implemented through AMPL language and KNITRO solver.

The result is represented in the graphs below. The three approaches lead to a very similar result. The green bar represents the resulting optimal range, while the blue bar corresponds to the tolerances initially defined and the orange bar to the dispersion intervals observed on the measurement data.

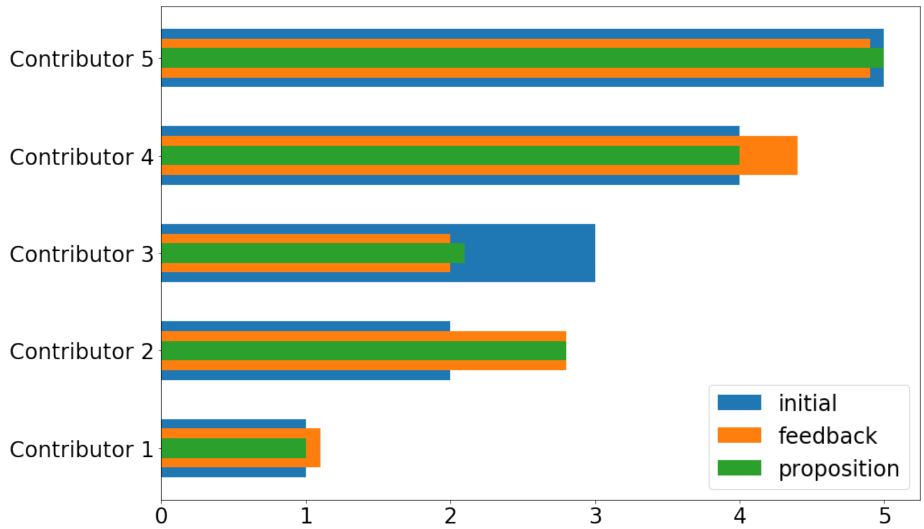

FIGURE 5.6: Result: identical for the three approaches

This tolerance resharing is relevant from an industrial point of view, since only the tolerances far from of their feedback are modified. the tolerance on Contributor 2 is extended while the tolerance on Contributor 3 is reduced. This new balance is economically interesting since it reduces the out of tolerances probability for the contributors while maintaining a reasonable risk on the output top level requirement. Moreover, only two tolerances are modified. The ASCR value for the new resharing is ±11.4 (against ±11.5 for the initial tolerance sharing). The difference between the three approaches is the number of calls to a continuous solver in the resolution.

The approach with a nonlinear solver solves a relaxed problem 32 times using Sequential Least Squares Programming (SLSQP). The greedy algorithm requires 15 resolutions by a continuous solver (SLSQP) in the different steps. According to Knitro's output, the branch and bound approach processes only one node and two subproblems, which correspond to two calls to a continuous solver. The disadvantage of Knitro is the implementation complexity in internal tools, that is why in Chapter 6 the greedy algorithm is used.

5.7 A variant of the optimization approach for adjustments (rigging)

It is common for a stack chain to be made up of so-called rigging contributors that are adjustment links. These rigging contributors are a source of variation in the stack chain if they are not specifically set to a certain value. These contributors are adjustable and the value to be assigned to them has to be defined in order to minimize the non-quality at the output of the assembly. The same rigging can be involved in several assemblies and therefore several stack chains, as illustrated in Figure 5.7 where the rigging contributor $X1$ is involved in three top level requirements.

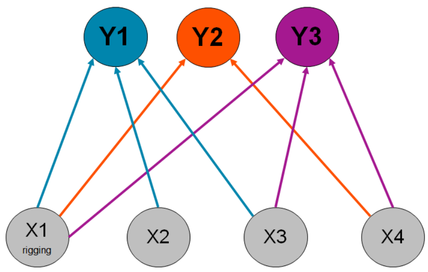

FIGURE 5.7: Rigging contributors interrelations between stack chains

In order to find the optimal value for this rigging, it is therefore necessary to globally evaluate the impact on the top level requirements of this adjustment contributor in each stack chain.

Thus, to find the optimal value of a rigging, the criterion to be minimized is similar to the cost c_2 which relates to the non-quality at the output of the assembly. However, it is now necessary to consider the interconnection of several stack chains in which the same rigging contributor can be involved.

Note that we are no longer working on tolerance intervals but on a value to be assigned to a contributor with an allocated tolerance range.

The non-quality of the assembly output is now computed from the distribution estimated for the random variable associated to the top level requirement. This distribution is obtained from the convolution product of the densities associated to the contributors. As already detailed in previous parts, the measured contributors have a fitted distribution on observations and the non-measured contributors are assumed to be uniform.

The novelty here is to consider a new type of contributor that is rigging. A rigging contributor take a precise value and might have some specific bounds. Usually, these bounds are the tolerance limits of the rigging contributor since these often represent physical setting limits. We denote the rigging contributors in a stack as x_{rig}.

To formalize the approach, the function to be minimized is written as

$$\sum_{k=1}^{K} \mathbb{P}\left(|Y_k(x_{\text{rig}})| > v_y\right)$$

where $Y_1, \ldots, Y_K$ are the feature associated to stack chains in which rigging contributors are involved. These distributions are estimated from measured and non-measured contributors as previously detailed, but in addition rigging contributors (one or more) need to be considered. This the newly introduced variable of the criterion function to be minimized, denoted x_{rig}.

The optimal rigging are the solution of the following minimization problem:

$$\min_{x_{\text{rig}} \in \mathbb{R}^r} \sum_{k=1}^{K} \mathbb{P}\left(|Y_k(x_{\text{rig}})| > v_y\right).$$

This methodology for rigging optimization in interrelated stack chains is implemented within an internal tool detailed in Chapter 6. Note that if we try to limit the number of rigging operation to perform, the problem becomes similar to the one stated in 5.4 and same solving techniques than in 5.5 could be applied.

Bibliography

[1] Martin Hallmann, Benjamin Schleich, and Sandro Wartzack. From tolerance allocation to tolerance-cost optimization: a comprehensive literature review. *The International Journal of Advanced Manufacturing Technology*, 107(11-12):4859–4912, April 2020.

[2] David H. Evans. Optimum Tolerance Assignment to Yield Minimum Manufacturing Cost. *Bell System Technical Journal*, 37(2):461–484, March 1958.

[3] F. H. Speckhart. Calculation of Tolerance Based on a Minimum Cost Approach. *Journal of Engineering for Industry*, 94(2):447–453, May 1972.

[4] M. F. Spotts. Allocation of Tolerances to Minimize Cost of Assembly. *Journal of Engineering for Industry*, 95(3):762–764, August 1973.

[5] W.-J. Lee and T. C. Woo. Optimum Selection of Discrete Tolerances. *Journal of Mechanisms, Transmissions, and Automation in Design*, 111(2):243–251, June 1989.

[6] Woo-Jong Lee, Tony C. Woo, and Shuo-Yan Chou. Tolerance synthesis for nonlinear systems based on nonlinear programming. *IIE Transactions*, 25(1):51–61, January 1993.

[7] J Lööf, T Hermansson, and R Söderberg. An Efficient Solution to the Discrete Least-Cost Tolerance Allocation Problem with General Loss Functions. In Joseph K. Davidson, editor, *Models for Computer Aided Tolerancing in Design and Manufacturing*, pages 115–124. Springer Netherlands, Dordrecht, 2007.

[8] K. Geetha, D. Ravindran, M. Siva Kumar, and M. N. Islam. Multi-objective optimization for optimum tolerance synthesis with process and machine selection

using a genetic algorithm. *The International Journal of Advanced Manufacturing Technology*, 67(9-12):2439–2457, August 2013.

[9] Bor-Wen Cheng and Saeed Maghsoodloo. Optimization of mechanical assembly tolerances by incorporating Taguchi's quality loss function. *Journal of Manufacturing Systems*, 14(4):264–276, January 1995.

[10] Hoo-Gon R. Choi, Man-Hee Park, and Erik Salisbury. Optimal Tolerance Allocation With Loss Functions. *Journal of Manufacturing Science and Engineering*, 122(3):529–535, August 2000.

[11] C.-Y. Chou and C.-L. Chang. Minimum-Loss Assembly Tolerance Allocation by Considering Product Degradation and Time Value of Money. *The International Journal of Advanced Manufacturing Technology*, 17(2):139–146, January 2001.

[12] Chung-Ho Chen and Hui-Sung Kao. The determination of optimum process mean and screening limits based on quality loss function. *Expert Systems with Applications*, 36(3):7332–7335, April 2009.

[13] Yi Zong and Jian Mao. Tolerance Optimization Design Based on the Manufacturing-costs of Assembly Quality. *Procedia CIRP*, 27:324–329, 2015.

[14] Yue Wang, Lei Li, Nathan W. Hartman, and John W. Sutherland. Allocation of assembly tolerances to minimize costs. *CIRP Annals*, 68(1):13–16, 2019.

[15] Tao Jing, Xitian Tian, Xiang Liu, Hao Hu, Min Zhang, and Bo Li. A multiple alternative processes-based cost-tolerance optimal model for aircraft assembly. *The International Journal of Advanced Manufacturing Technology*, 107(1-2):667–677, March 2020.

[16] Chin-Nung Liao and Hsing-Pei Kao. Supplier selection model using Taguchi loss function, analytical hierarchy process and multi-choice goal programming. *Computers & Industrial Engineering*, 58(4):571–577, May 2010.

[17] Wei-Ning Pi and Chinyao Low. Supplier evaluation and selection using Taguchi loss functions. *The International Journal of Advanced Manufacturing Technology*, 26(1-2):155–160, July 2005.

[18] W. Edwards Deming. *The New Economics for Industry, Government, Education.* MIT Press, October 2018.

[19] Egon Balas, Fred Glover, and Stanley Zionts. An Additive Algorithm for Solving Linear Programs with Zero-One Variables. *Operations Research*, 13(4):517–549, 1965. Publisher: INFORMS.

[20] Kenneth W. Chase, William Henry Greenwood, Bruce G. Loosli, and Loren F. Hauglund. Least Cost Tolerance Allocation for Mechanical Assemblies with Automated Process Selection, 1990. Library Catalog: www.semanticscholar.org.

[21] Nagraj Balakrishnan. A multiple-choice Knapsack model for tolerance allocation in mechanical assemblies. *IIE Transactions*, 25(4):13–14, July 1993.

[22] Pradeep K. Singh, Satish C. Jain, and Pramod K. Jain. Advanced optimal tolerance design of mechanical assemblies with interrelated dimension chains and process precision limits. *Computers in Industry*, 56(2):179–194, February 2005.

[23] P K Sing, S C Jain, and P K Jain. Comparative study of genetic algorithm and simulated annealing for optimal tolerance design formulated with discrete and continuous variables. *Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture*, 219(10):735–758, October 2005.

[24] Sergio Nesmachnow. An overview of metaheuristics: accurate and efficient methods for optimisation. *International Journal of Metaheuristics*, 3(4):320, 2014.

[25] J. Lee and G.E. Johnson. Optimal tolerance allotment using a genetic algorithm and truncated Monte Carlo simulation. *Computer-Aided Design*, 25(9):601–611, September 1993.

[26] Chae-Bogk Kim, P. Simin Pulat, Bobbie L. Foote, and Dong Hoon Lee. Least cost tolerance allocation and bicriteria extension. *International Journal of Computer Integrated Manufacturing*, 12(5):418–426, January 1999.

[27] L. Governi, R. Furferi, and Y. Volpe. A Genetic Algorithms-based Procedure for Automatic Tolerance Allocation Integrated in a Commercial Variation Analysis Software. *Journal of Artificial Intelligence*, 5(3):99–112, March 2012.

[28] P K Singh, S C Jain, and P K Jain. A genetic algorithm based solution to optimum tolerance synthesis of mechanical assemblies with alternate manufacturing processes—benchmarking with the exhaustive search method using the Lagrange multiplier. *Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture*, 218(7):765–778, July 2004.

[29] B. Heling, A. Aschenbrenner, M.S.J. Walter, and S. Wartzack. On Connected Tolerances in Statistical Tolerance-Cost-Optimization of Assemblies with Interrelated Dimension Chains. *Procedia CIRP*, 43:262–267, 2016.

[30] Giovanni Moroni, Stefano Petrò, and Tullio Tolio. Early cost estimation for tolerance verification. *CIRP Annals*, 60(1):195–198, 2011.

[31] K. Geetha, D. Ravindran, M. Siva Kumar, and M.N. Islam. Concurrent tolerance allocation and scheduling for complex assemblies. *Robotics and Computer-Integrated Manufacturing*, 35:84–95, October 2015.

[32] Z. Dong and W. Hu. Optimal process sequence identification and optimal process tolerance assignment in computer-aided process planning. *Computers in Industry*, 17(1):19–32, October 1991.

[33] Loïc Andolfatto, François Thiébaut, Claire Lartigue, and Marc Douilly. Quality- and cost-driven assembly technique selection and geometrical tolerance allocation for mechanical structure assembly. *Journal of Manufacturing Systems*, 33(1):103–115, January 2014.

[34] J.R. He. Tolerancing for manufacturing via cost minimization. *International Journal of Machine Tools and Manufacture*, 31(4):455–470, January 1991.

[35] M.S.J. Walter, T.C. Spruegel, and S. Wartzack. Least Cost Tolerance Allocation for Systems with Time-variant Deviations. *Procedia CIRP*, 27:1–9, 2015.

[36] Alain Etienne, Shirin Mirdamadi, Mehrdad Mohammadi, Roozbeh Babaeizadeh Malmiry, Jean-François Antoine, Ali Siadat, Jean-Yves Dantan, Reza Tavakkoli, and Patrick Martin. Cost engineering for variation management during the product and process development. *International Journal on Interactive Design and Manufacturing (IJIDeM)*, 11(2):289–300, May 2017.

[37] Steven Hoffenson, Andreas Dagman, and Rikard Söderberg. Tolerance Specification Optimization for Economic and Ecological Sustainability. In Michael Abramovici and Rainer Stark, editors, *Smart Product Engineering*, pages 865–874. Springer Berlin Heidelberg, Berlin, Heidelberg, 2013. Series Title: Lecture Notes in Production Engineering.

[38] Donald J. Wheeler and David Smith Chambers. *Understanding Statistical Process Control*. SPC Press, 1992. Google-Books-ID: XvMJAQAAMAAJ.

[39] Martin Hallmann, Benjamin Schleich, Björn Heling, Alexander Aschenbrenner, and Sandro Wartzack. Comparison of different methods for scrap rate estimation in sampling-based tolerance-cost-optimization. *Procedia CIRP*, 75:51–56, 2018.

[40] Kalyanmoy Deb. *Multi-objective optimization using evolutionary algorithms*. Wiley-Interscience series in systems and optimization. John Wiley & Sons, Chichester ; New York, 1st ed edition, 2001.

[41] Xuan Cai. Canonical Coin Systems for CHANGE-MAKING Problems. In *2009 Ninth International Conference on Hybrid Intelligent Systems*, pages 499–504, Shenyang, China, 2009. IEEE.

[42] A. H. Land and A. G. Doig. An Automatic Method of Solving Discrete Programming Problems. *Econometrica*, 28(3):497, July 1960.

[43] R. J. Dakin. A tree-search algorithm for mixed integer programming problems. *The Computer Journal*, 8(3):250–255, March 1965.

[44] Irene Charon, Olivier Hudry, and Anne Germa. *Méthodes d'optimisation combinatoire*. 1997.

[45] Michel Gondran and Michel Minoux. *Graphes et algorithmes*. 2009.

[46] John Forrest and Robin Lougee-Heimer. CBC User Guide. In Harvey J. Greenberg and J. Cole Smith, editors, *Emerging Theory, Methods, and Applications*, pages 257–277. INFORMS, September 2005.

[47] Christian Bliek1ú, Pierre Bonami, and Andrea Lodi. Solving mixed-integer quadratic programming problems with IBM-CPLEX: a progress report. In *Proceedings of the twenty-sixth RAMP symposium*, pages 16–17, 2014.

[48] EFD Ellison, M Hajian, H Jones, R Levkovitz, I Maros, G Mitra, and D Sayers. FortMP Manual. Brunel University: London. *Numerical Algorithms Group: Oxford. http://www. optirisk-systems. com/manuals/FortmpManual. pdf*, 2008.

[49] Bob Bixby. The gurobi optimizer. *Transp. Re-search Part B*, 41(2):159–178, 2007.

[50] Michael A Saunders. MINOS System Manual. Technical report, STANFORD UNIV CALIF SYSTEMS OPTIMIZATION LAB, 1977.

[51] Andreas Wächter. Short Tutorial: Getting Started With Ipopt in 90 Minutes. In Uwe Naumann, Olaf Schenk, Horst D. Simon, and Sivan Toledo, editors, *Combinatorial Scientific Computing*, Dagstuhl Seminar Proceedings, Dagstuhl, Germany, 2009. Schloss Dagstuhl - Leibniz-Zentrum fuer Informatik, Germany. ISSN: 1862-4405 Issue: 09061.

[52] Philip E Gill, Walter Murray, and Michael A Saunders. SNOPT: An SQP algorithm for large-scale constrained optimization. *SIAM review*, 47(1):99–131, 2005. Publisher: SIAM.

[53] Jorge Nocedal. KNITRO: an integrated package for nonlinear optimization. In *Large-Scale Nonlinear Optimization*, pages 35–60. Springer, 2006.

[54] János D Pintér. LGO—a program system for continuous and Lipschitz global optimization. In *Developments in global optimization*, pages 183–197. Springer, 1997.

[55] I. Quesada and I.E. Grossmann. An LP/NLP based branch and bound algorithm for convex MINLP optimization problems. *Computers & Chemical Engineering*, 16(10-11):937–947, October 1992.

Chapter 6

Industrial tools

This chapter presents the industrial implementations of the work proposed in this manuscript. All the methodologies presented in the previous chapters aims to contribute to the business needs. This implementation presents the associated tools to enable users to use the methodologies in practice. As a reminder, the schema 6.1 illustrates the different issues addressed in the tolerancing process and the associated chapters. The tool suite for tolerance management proposed for implementation of proposed methodologies is displayed in the purple frame.

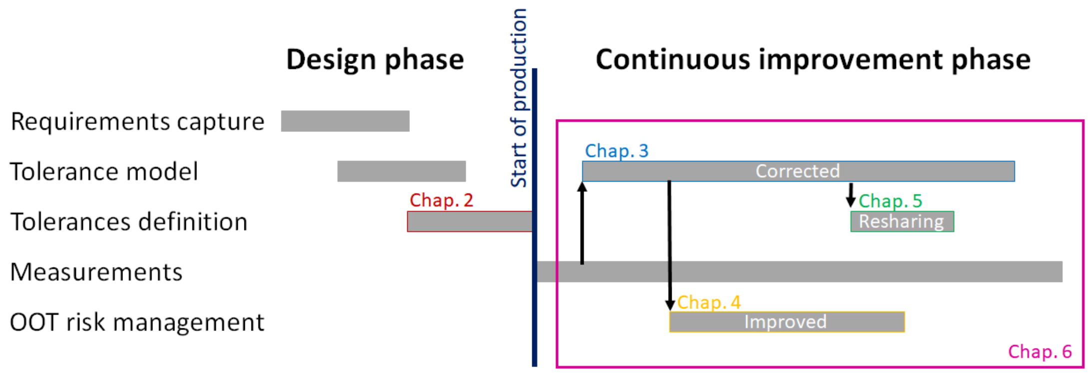

FIGURE 6.1: Tolerancing process mapping for implementation

6.1 Tolerance management at Airbus

In addition to a database system, it is important for Airbus to have a suite of tools to work on tolerance management. The DIVA platform detailed in the following section allows to study tolerances and to confront theory with reality thanks to the analysis of measurement data.

6.1.1 Tool for storing tolerance information

For several years, a tool and a database have been used at Airbus to capitalize on the stack chains of assemblies at all levels for all programs. Tolerance models are assumed to be linear and a 3D simulation software is directly connected to allow the calculation of linear influence coefficients.

This tool and database is the main working tool for tolerance analysts. During the development of a new assembly, the database is completed with the tolerance model, the tolerance limits associated with each contributor and information on the top level requirement. When it comes to the revision of a stack chain, the tool gathers the necessary elements for the tolerance study. All these elements are also the input data for the methodologies presented in the various chapters.

6.1.2 DIVA platform and objectives

DIVA (Data Integrator for Variation Analysis) is a complementary platform to the tolerance management tool, enabling the processing of feedback data made available during measurement campaigns.This allows tolerance analysts to improve their tolerance studies.

The platform brings together different applications developed for specific purposes. Among them, the works presented in this thesis related to the processing of measurement data are implemented. In particular, the tolerance model is corrected upstream as a prerequisite (as detailed in Chapter 3). The improved tolerance model is then used in the interface to manage acceptance criteria based on the industrial risk consideration (Chapter 4) and in the tolerance optimization module (Chapter 5).

6.2 Specific functionalities implemented in DIVA

This section focuses on the DIVA functionalities which are based on the previously introduced methodologies. The data and information on the screenshots are deliberately blurred for confidentiality reasons.

6.2.1 Data selection

A first module allows to select a contributor or a top level requirement to be studied. Several filters on industrial classifications allow to refine the study perimeter.

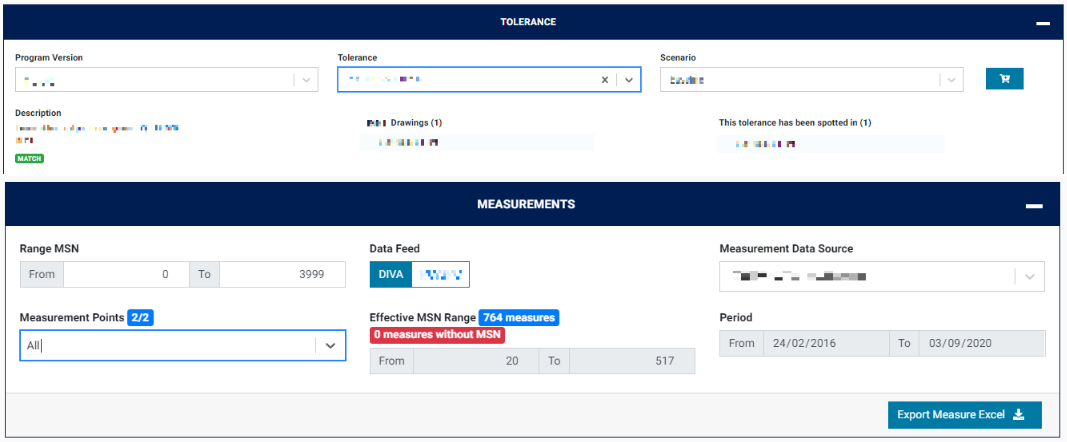

FIGURE 6.2: Tolerance study perimeter selection

After selecting a contributor or a top level requirement, the tool displays the different links with requirements or contributors involved in the same stack chain, and associated variability percentages.

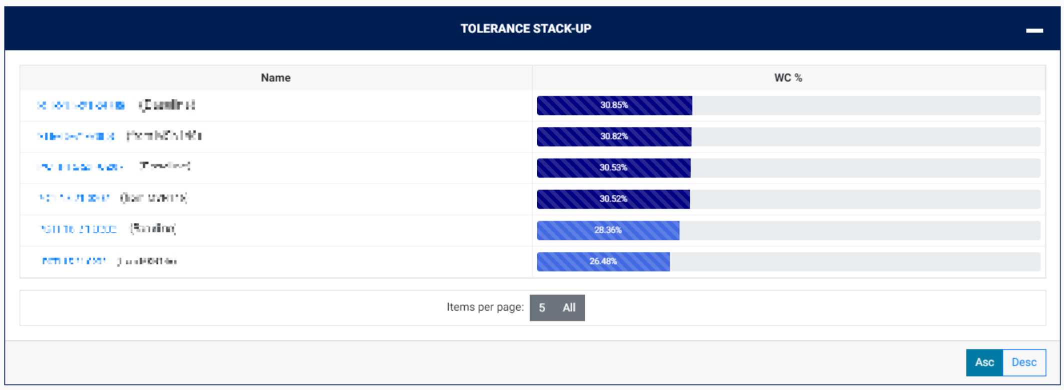

FIGURE 6.3: Example: Top levels requirement of a selected contributor

Then, measurement data histogram and several information about the considered characteristics are displayed. This example is for a top level requirement that will be also considered in the next section.

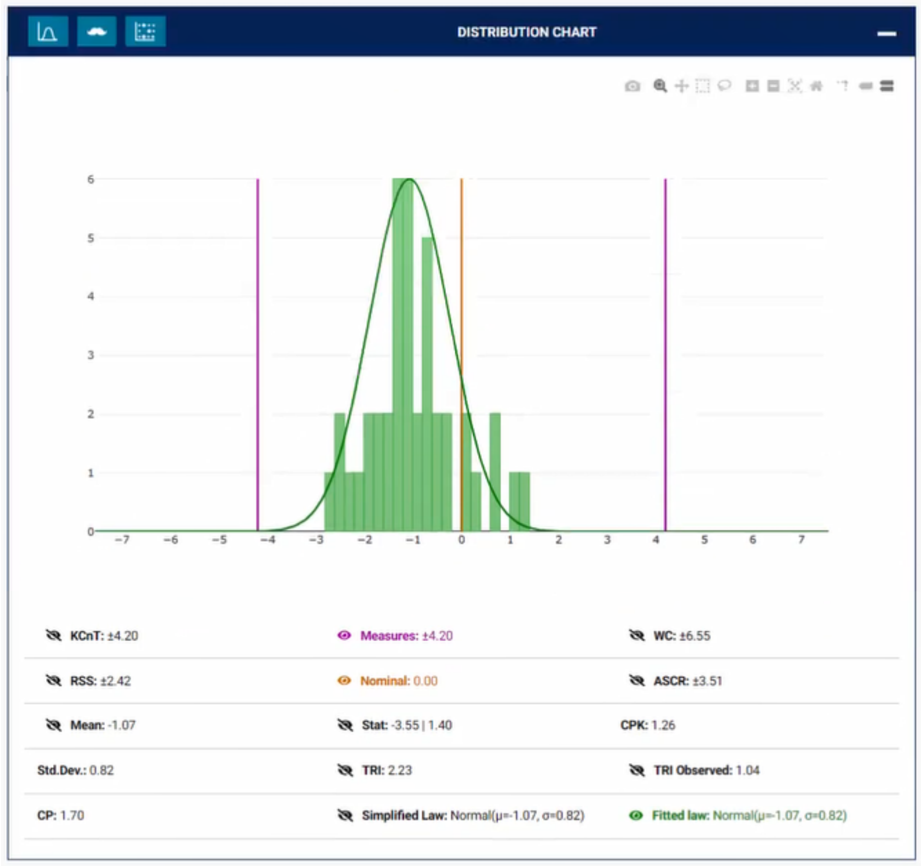

FIGURE 6.4: Example: Measurement data information in DIVA

In the top left corner of this window with the histogram, we can see the buttons allowing to access the correlation analysis of the different measurement points of a characteristic between them as well as their descriptive statistics.

6.2.2 Model correction

The model correction is a prerequisite, and the model improvement loop is run before the risk analysis for the acceptance criteria, the calculation of the distribution at the top level requirement and the optimization algorithms.

Top level requirement distribution - *open loop*

For this module, the analysis applies to a top level requirement. This allows the user to see improvements in the tolerance model, detailed in 3.4. The plot represents the density estimate of the top level requirement feature of the stack chain. The solid line is the result as expected during the design phase with the initial tolerance model.

At this stage, the measurement data are not taken into account. The stack chain contributors features are assumed to follow uniform distributions between tolerance limits (as in Chapter 2). It is an *open loop* approach.

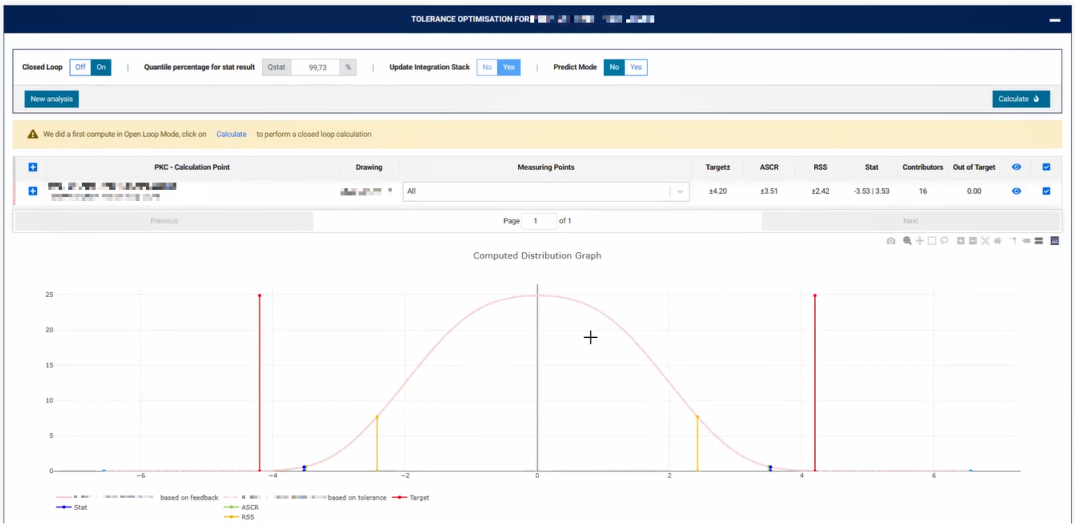

FIGURE 6.5: Output feature distribution density - open loop

By clicking on the button , the user accesses the details of the stack chain. In the next section, this information is updated, so the details display the corrected sign of the influence coefficients and the offset as described in Chapter 4.

Top level requirement distribution - *closed loop*

In a second step, the *closed loop* analysis takes into account the distributions for contributors coming from measurement data and the corrected tolerance model. In solid line, the density is the one estimated in this case. The corrected tolerance model is now able to explain the variability observed in Figure 6.4.

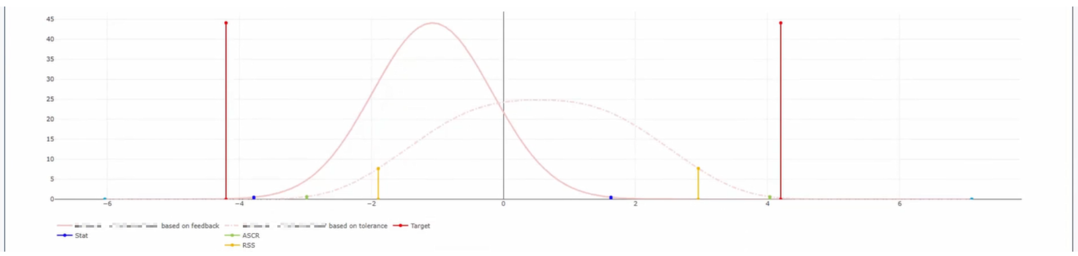

FIGURE 6.6: Output feature distribution density - closed loop

Top level requirement distribution - *predict* mode

Finally, the tool allows the user to switch to *predict* mode. The principle is to select a specific serial number (MSN) of an aircraft and to look specifically at the prediction on the top level requirement using the corrected model.

Available data from suppliers are retrieved from the database and replace the tolerances of the measured contributors. This has the effect of tightening the estimated distribution on the top level requirement since the measured contributors values no longer have any variability in the stack chain.

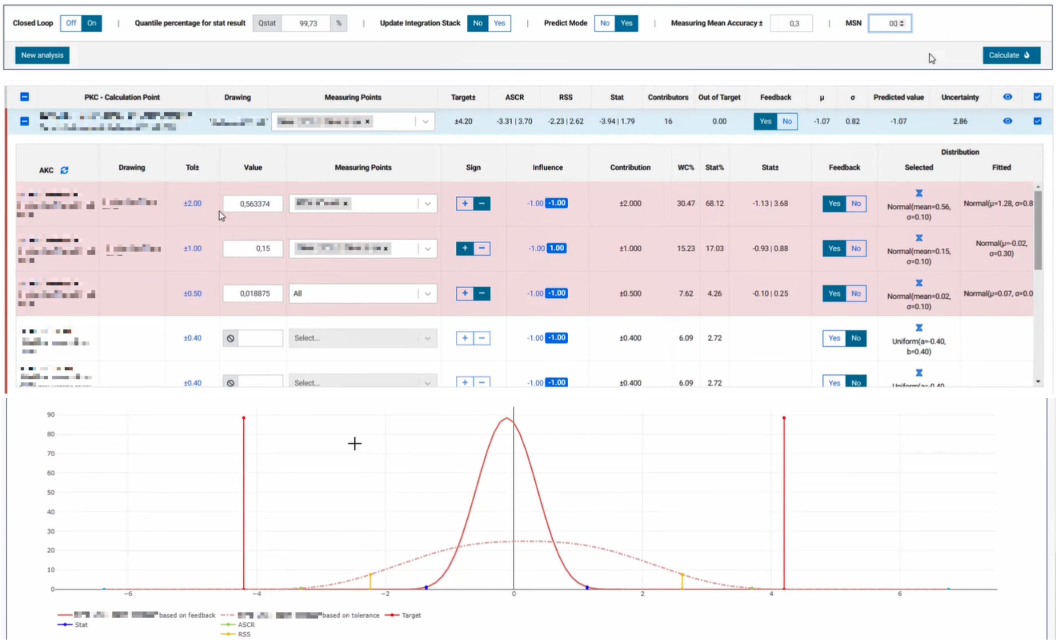

FIGURE 6.7: Output feature distribution density - *predict mode* for a MSN

An extension of this *predict* approach allows to consider an entire area of the aircraft, and to predict the distributions for several characteristics. This feature called *Enhanced Virtual Assembly* uses once again the corrected tolerance model.

6.2.3 Tolerance re-sharing optimization

The optimization for tolerance sharing is implemented in the same interface as Figures 6.5 and 6.7. There is also an interface to select the parameters λ_1, λ_2, λ_3 and a button in the top right corner to launch the optimization. As output, the new tolerance values of the stack chain are displayed.

A challenge in the implementation for industrial use is to consider that some contributors in the stack chain should not be modified, for example when feedback is not available. The optimization problem must therefore be restricted to a user-defined set of contributors.

6.2.4 Rigging optimization

The topic of choosing values for rigging contributors described by the optimization problem in the section 5.7 is also implemented in DIVA. The simple example on a top level requirement presented on the screenshots allows to see the application of the methodology on a simulated case close to a possible industrial case. In Figure 6.8, selected contributors for the study and their information are displayed. The example involves several contributors and two different physical parts in the assembly.

The focus is on one serial number (MSN) on which measurement data are available. The measured contributors thus have their measured value assigned and others contributor are either represented by uniform or normal random variables (for instance, if measurement data are available but not on the considered MSN).

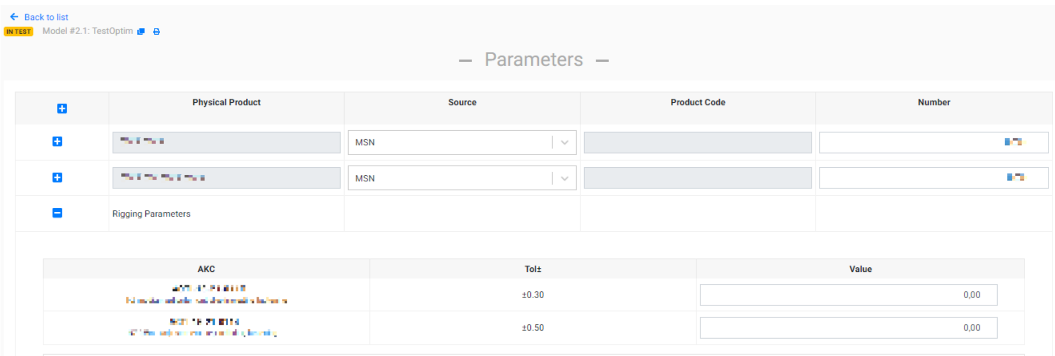

FIGURE 6.8: Rigging optimization example: contributors information

Figure 6.9 shows information about requirements associated to the contributors previously selected. In this case there are two top level requirements. The first one has a target of ± 4.2, and a out of tolerance risk of 9.08% risk. The prediction columns gives the predicted value of the top level requirement for the serial number (MSN) previously selected. It takes into account the measurement data for observed contributors. There are two rigging contributors, both with a negative coefficient influence equal to 1.

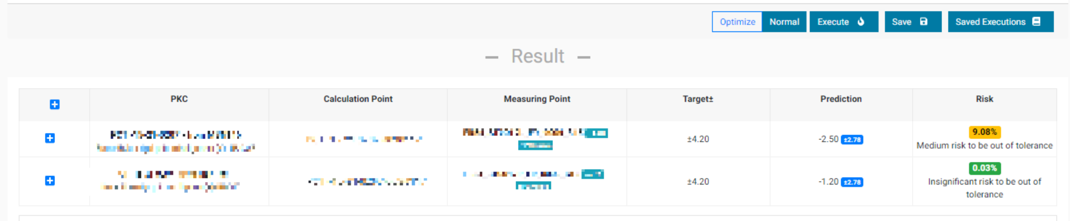

FIGURE 6.9: Example for a rigging optimization: top level requirement information

Figure 6.10 shows the distribution estimation of the first top level requirement, with the higher risk. The solid line is the one to be considered, this is the estimation based on feedback data, including the measures for the selected MSN. We can see that the negative side of the distribution is out of the tolerance target of the top level requirement, which corresponds to the 9.08% risk.

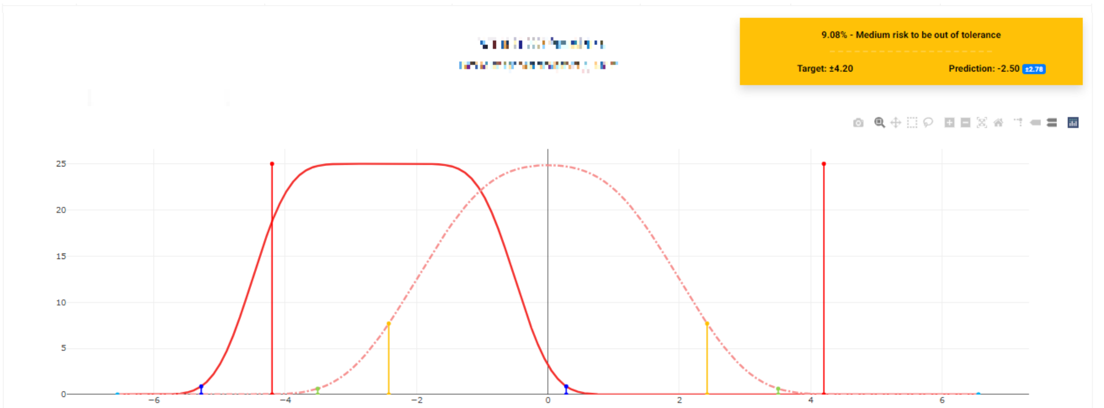

FIGURE 6.10: Distribution of the first top level requirement before optimization

After clicking on the *optimization* button, the result is displayed as in Figure 6.11. Both contributors have an assigned value at the maximum of their negative interval. Indeed, their influence coefficient is negative and it is indeed negative values that will reduce the out of tolerance risk for the top level requirement.

The rigging values change the prediction, which is shifted by $-1 \times -0.5 + -1 \times -0.3 = 0.8$. The top level requirement distribution taking into account these rigging values is displayed in Figure 6.12.

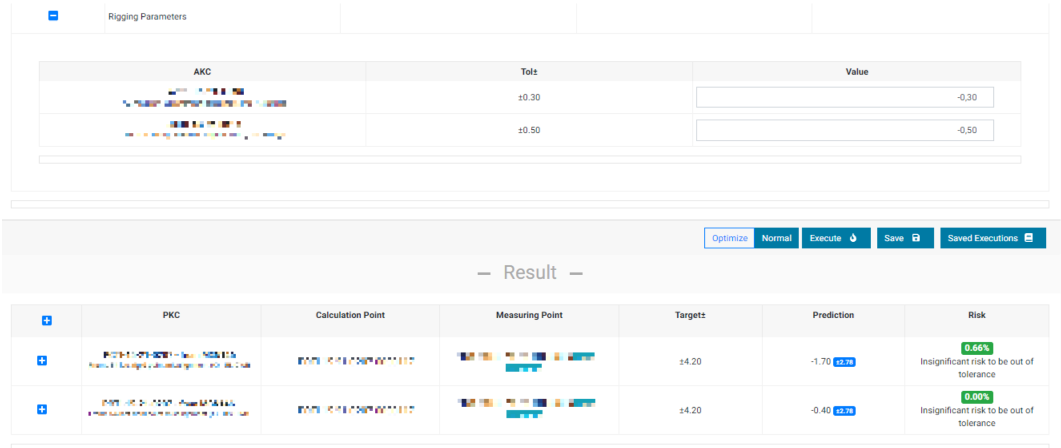

FIGURE 6.11: Result of the rigging optimization

We can see on Figure 6.10 that the distribution is offset and the out of tolerance probability is lower than in Figure 6.10. From 9.08%, the risk to be considered is now 0.66%.

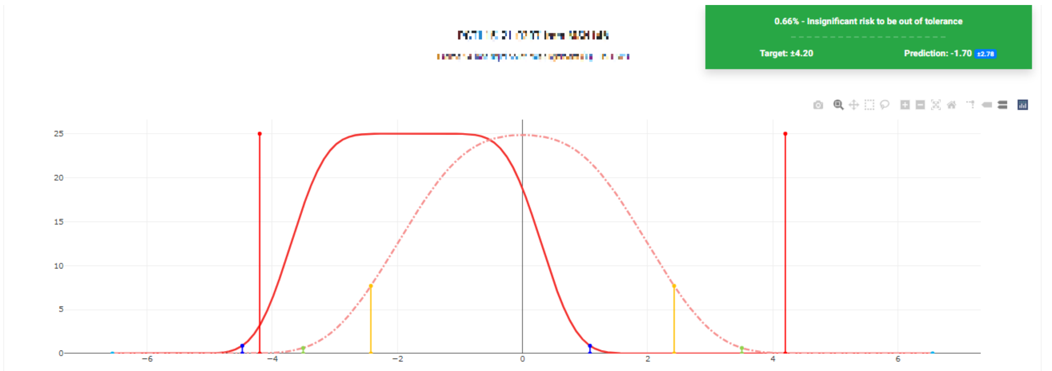

FIGURE 6.12: Distribution of the first top level requirement after optimization

6.2.5 Risk based acceptance criteria definition

In this part, the focus is on a contributor in a stack chain. For the definition of acceptance criteria, we have to consider a contributor involved in at least one assembly. Indeed, the risk is defined for a top level requirement. The availability of measurement data is not mandatory. However, the tolerance relaxation by acceptance criteria will be more favorable if the measurement data of the contributors involved with the considered one are available and better than expected. In practice, measurement data may be required by the business before considering acceptance criteria.

The interface implements exactly the methodology presented in Chapter 4. The visualization tool allows the selection of a set of risk levels in percentage or values in millimeters as acceptance criteria on the tolerances of the considered contributor. These criteria are either absolute or relative to the contributor's initial tolerance.

For the study of acceptance criteria of a contributor, several requirements may be involved. These are filled in the table and represented by curves of different colors. The values in percentage or in millimeters are displayed on the graph to facilitate the decision for the acceptance criteria. The plotted risk is the probability to be out of tolerance at the top level requirement if the contributor to which criteria are defined has the value of the x-axis. More details about this type of plot are available in the application section of Chapter 4.

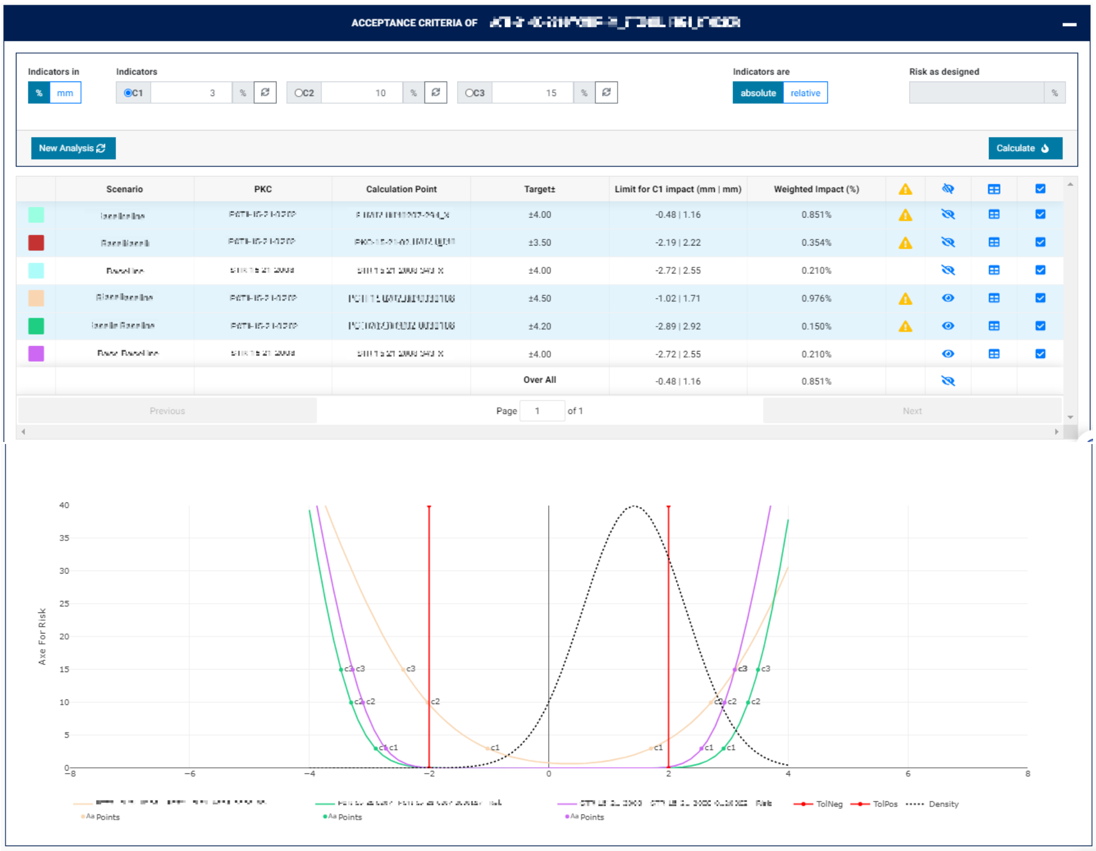

FIGURE 6.13: DIVA interface for risk based acceptance criteria

The table in Figure 6.13 gathers all the results, either the values in millimeters that correspond to the selected risk percentage or the risk percentage values associated with the values of the criteria selected by the user.

There is also a percentage value, called *weighted impact* which represents the risk weighted by the density of the contributors on which the acceptance criteria are to be defined. This is the probability consideration in the common impact/probability concept in risk management.

The tool allows to decide what is the appropriate level of risk and to apply the acceptance criteria. The decision is valid for a limited period of time because it is based on measurement data that may evolve. This tool automatically generates a report with all the information that justifies the decision.

Chapter 7

Conclusion and Perspectives

7.1 Conclusion

Tolerancing plays a major role in the manufacture of a product. It is an aspect that must be taken into account both during the development phase and during production. In the case of aeronautical structures, several challenges related to tolerancing have to be dealt with.

The work proposed in this manuscript is about the improvement of the statistical approach within an industrial context. The aim is to manage tolerances as well as possible in order to ensure the safety and quality of the product at all times, while ensuring performance and limiting the various costs involved. In a company such as Airbus, tolerancing involves various design offices and factories. Industrial processes, existing tools and software or the measurement availability and data flow are constraints to be taken into account to ensure that the proposed improvement will be implemented and bring benefit to the aerostructures production.

The first chapter established the general context of tolerancing and the specificities related to the statistical approach as well as to the particular case of aerostructure tolerancing. By going through the life cycle of a tolerance system, several improvements of the statistical tolerancing approach are proposed.

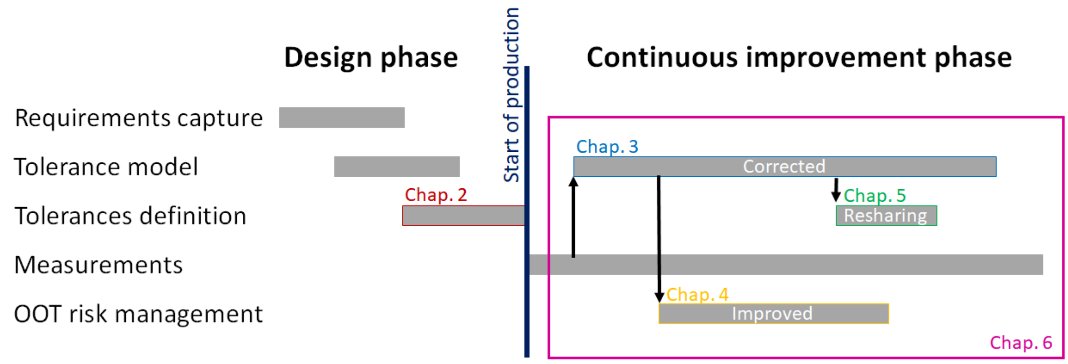

FIGURE 7.1: Tolerancing process mapping

First of all, tolerance analysis is discussed in Chapter 2. It is a question of defining the output variability of an assembly when the tolerance intervals of the contributors of an assembly are assumed to be known. The proposed statistical method is based on Chernov-type bounding of a sum of uniform distributions. This approach allows to define a variability limit on the output tolerance in a robust way. The interest of such a method lies in the ability to select and guarantee a precise scrap rate at the output of an assembly. The link between this method and the existing Airbus process is highlighted providing elements to further validate its use in industrial application.

The chapters 3, 4 and 5 focused on the evolution of a tolerance system when it is in production and when partial measurement data are available. A discussion of the tolerance model and its enhancement is provided in Chapter 3. An approach to improving the linear tolerance model is detailed, taking into account industrial constraints. The detailed process allows the improvement of tolerance processing in the serial phase and thus a better efficiency of tolerance studies when measurement data are available.

Chapter 4 presented a smart methodology for dynamic tolerance risk management through acceptance criteria. A formulation of the industrial risk is detailed and used as a decision aid. This approach requires new measurement data to allow the improvement of the tolerance model initially designed in the development phase. The knowledge of these data allows an evaluation of the industrial risk in real time on a process observed in factories.

Chapter 5 dealt with the sharing of variability in an assembly and its optimization. Indeed, a sharing is agreed in the design phase, but once again the measurement data allow to review and improve. A modeling of industrial costs allows the use of optimization techniques. The contribution of such an approach is to minimize non quality by limiting the cost of reviewing a tolerance sharing.

To apply methodologies based on measurement data, Chapter 6 presented the industrial tool used at Airbus. It is a platform that brings together several functionalities, and in particular model correction, risk acceptance criteria management and optimization of stack chain sharing. This implementation allows the techniques developed in this thesis to work with thousands of tolerances and measurements processed by the internal tool DIVA. In the area of tolerances, this industrial tool contributes to the development of the virtual twin. Tolerance management is enabled in a dedicated interface and this allows the continuous improvement of the tolerancing process.

The contributions of this thesis allowed to include new ways of working in the industrial process thanks to statistics-based methods. Industrial constraints such as the availability of measurements, the implementation of tools, or the historical culture of the technical teams were taken into account to build realistic solutions. These solutions have been implemented and are deployed or ready to be deployed for the actors working on tolerance challenges and current applications have already demonstrated added value for Airbus.

Associated publications and patents

Chapter 2 is published in the journal *The International Journal of Advanced Manufacturing Technology, 111(11), 3571-3581*, under the title *A Chernov bound for robust tolerance design and application.* A more general discussion on the end-to-end approach to tolerance entitled *A statistical approach for tolerancing from design stage to measurements analysis* was presented at the 16^{th} conference on *Computer Aided Tolerancing.* (CAT) in June 15-17, 2020. The methodologies developed in Chapters 2 and 4 are the subject of patent applications, referenced respectively under deposit numbers 1912668 and 2008847 at the French Patent Office (INPI).

7.2 Perspectives

In the continuation of this thesis work, several axes can be considered both in the industrial context and the academic background.

7.2.1 Tolerance model improvement

From an industrial point of view, the perspectives revolve around the refinement of methods to meet industrial needs even better. For example, this could involve adding to the tolerance model the consideration of variabilities induced by measurable external factors. Indeed, environmental factors such as humidity, temperature or pressure can have an influence on the dimensions. Currently, these factors are included in the overall dimensional uncertainty and are therefore not analyzed one by one. If measurement data for these effects were available, it would then be possible to assess and model their impact. Such a problem seems to be suited to machine learning techniques. On the one hand, this would allow the identification of external effects affecting the dimensions and their variabilities. On the other hand, these external effects could be quantified and then included in tolerance model as additional information. In output of an assembly, it would then be possible to evaluate more precisely the distribution of a top level requirement. Such an approach requires data on these detailed external effects that are temporally related to the dimensional measurement data. It is a challenge to record these data on a large scale, in different plants, and to record them together with the measurement data.

Concerning the correction of the linear tolerance model, methodologies about the correction of the linear coefficient signs are discussed in Chapter 3. It would be interesting to go further, to find and check the value of each coefficient. The limitation often encountered is that not all the contributors of an assembly are necessarily measured. This, together with the number of measurements points available for a measured contributors, induce an open question about the convergence of the integration stack calculation. A standard multivariate linear regression is therefore not applicable. However, a regression in a Bayesian framework could solve this problem by using a priori on the unmeasured contributors coming for instance from the distribution used in the tolerance model. Beyond the correction of the sign, this would allow, thanks to a Bayesian regression, to find the values of the influence coefficients for the linear model together with an associated credibility interval that would support the decision to update the model.

7.2.2 Resharing optimization challenges

Optimization for tolerance resharing could also be improved in different ways. First of all, the proof of the global optimality of the greedy algorithm has yet to be established. This is true for the cases analyzed so far, but formal proof is being studied.

Moreover, it is important for industrial needs to calibrate the parameters values of the optimization of the multi-objectives cost function. These parameters represent the relative importance of the different costs. For business needs, it is necessary to make sense of these values. A detailed study of the optimization results on Airbus stack chains is planned in order to calibrate these parameters as well as possible.

Finally, additional costs could be considered to refine the approach. Indeed, the optimization problem presented in Chapter 5 is based on three costs defined in accordance with Airbus business needs. Other costs related to, for example, over-quality or the price of the parts of an assembly could be taken into account in addition to those already considered in the optimization problem. To fit easily into the optimization problem already in place, the costs should be dependent on the initial tolerances and measurement data available. This is indeed the case for over-quality, but the consideration of external costs such as price requires a link function with the tolerances. Several works are already available as seen in the state of the art of Chapter 5 and could be combined with the approach presented.

7.2.3 Multi-level assemblies

The inter-relation aspect of tolerances could also be taken into account, either in the risk assessment or in the tolerance sharing. Indeed, a tolerance is often involved in several assemblies and present at different assembly levels. This means that a top level requirement can become a contributor to the next assembly step (Figure 7.2).

In the framework presented in the chapters of this manuscript, this means that the estimated distribution of a top level requirement can become the assumed distribution for a contributor. In this case, the contributor's measurement data should be considered together with the distribution estimated at the previous level. If the tolerance model is close enough to reality, it is hoped that these two distributions are similar. As such an estimated distribution is used in the risk calculation and in the costs for the optimization of the tolerance distribution, it would allow to manage a same dimension feature and its tolerance in the different assembly levels. This leads to academic challenges in the field of optimization, in particular about graph optimization.

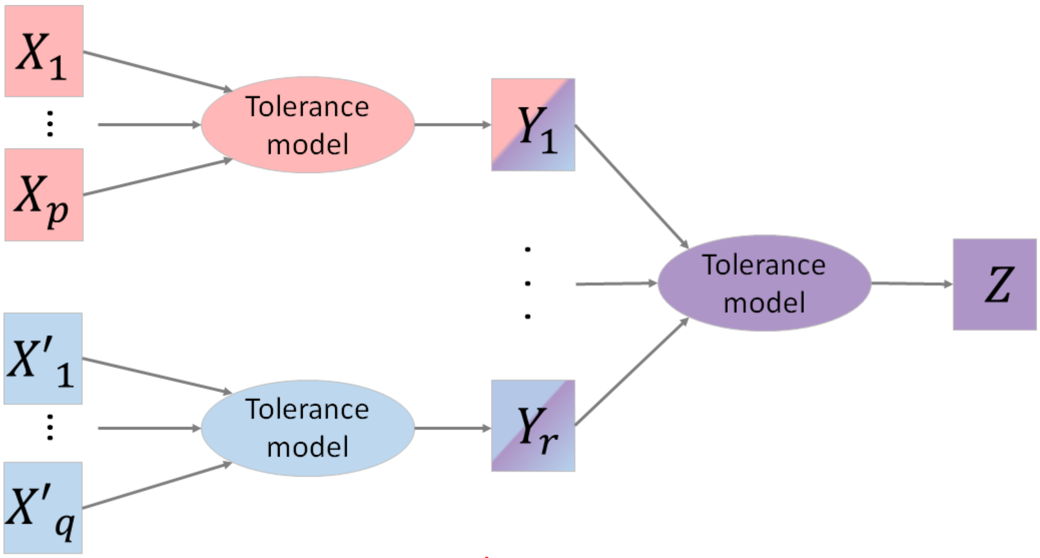

FIGURE 7.2: Multi-level assemblies models