

Perception du risque de conflit chez les contrôleurs aériens : le projet CREED

Article rédigé par Philippe AVERTY, Kévin GUITTET et Pascal LEZAUD
Written by Philippe AVERTY, Kévin GUITTET and Pascal LEZAUD

DGAC-DTI/SDER

Perception of risks of conflict by air traffic controllers: the CREED project

1 INTRODUCTION

The air traffic controller's work is composed of several tasks that generally overlap over time. The conjunction of these, when the supervised traffic becomes too complex, may lead to exceeding the controller's "cognitive resources". This progressively leads to a reduction in the resources devoted to the subsidiary tasks so as to concentrate them on the priority tasks, which are the detection and resolution of conflicts. Understanding this process is therefore a major stake in enabling the ATM system to adapt to the forecast growth of air traffic.

The study carried out at DTI-SDER within the CREED project was concerned exclusively with the conflict detection process. A set of experiments was therefore carried out to confront controllers with various potentially conflictual scenarios. The aim was to gather judgments concerning the possibility of a future loss of separation. The doubt inherent to this type of judgement, which is essential since it organises the controller's task, makes it impossible in most situations to make a purely binary choice, such as "conflict – no conflict". A range of answers was therefore

1 INTRODUCTION

Le travail du contrôleur du trafic aérien se compose de plusieurs tâches très intriquées dans le temps. Leur conjonction, lorsque le trafic supervisé devient trop complexe, peut conduire au dépassement des «ressources cognitives» du contrôleur. Celui-ci réduit alors progressivement les ressources affectées aux tâches annexes pour les concentrer sur les tâches prioritaires que sont la détection et la résolution des conflits. Comprendre ces processus représente donc un enjeu de taille pour permettre au système ATM de s'adapter à l'augmentation prévue du trafic aérien.

L'étude entreprise à la DTI-SDER au sein du projet CREED s'est intéressée exclusivement au processus de détection des conflits. Une expérimentation a donc été menée afin de confronter des contrôleurs à différents scénarii de conflits potentiels. L'objectif était de recueillir des jugements portant sur l'éventualité d'une perte future de séparation. Le doute inhérent à ce type de jugement, essentiel puisqu'il organise la tâche du contrôleur, rend également impossible, dans la majorité des situations, un choix purement binaire du type « conflit – non conflit ». Des modalités de réponses nuancées ont donc été proposées entre ces deux extrêmes. Enfin, les scénarii ont été présentés dans l'environnement radar propre à l'aéroport,

pour chaque recueil de données. La collaboration des contrôleurs des quatre grandes approches a permis de récolter des données très précieuses et de constituer ainsi une base a priori unique au monde.

Le traitement des données s'est décomposé en deux phases. Une analyse graphique a d'abord permis d'alimenter les réflexions de l'expert sur sa compréhension du processus de détection des conflits et de vérifier l'existence d'une certaine cohérence dans les jugements exprimés. Au delà des particularismes individuels bien présents et souvent soulignés dans la littérature, l'étude a ainsi démontré qu'une perception commune des conflits par les contrôleurs existe bel et bien, ainsi que la dimension collective du travail, le découpage de l'espace et la formation initiale semblent l'exiger. Un modèle statistique a ensuite été développé pour permettre de reproduire les jugements observés, et de les interpoler à d'autres situations de conflit potentiel. Ce modèle a également permis d'estimer et d'étudier l'hétérogénéité de la population des contrôleurs.

Cet article retrace donc les principales étapes du projet CREED : dans une première partie, nous discuterons des aspects cognitifs du travail du contrôleur et plus précisément du processus de détection des conflits. La problématique dégagée, nous détaillerons dans une seconde partie, l'expérimentation ayant permis de recueillir les jugements des contrôleurs. Les résultats seront présentés et argumentés dans une troisième partie, l'accent étant mis sur l'emploi des méthodes statistiques d'analyse graphique. La quatrième et dernière partie sera consacrée à la présentation des modèles logistiques construits et des connaissances alors obtenues. Des encarts devraient permettre au lecteur de mieux comprendre les outils utilisés dans cette partie. Enfin, la conclusion dévoilera les utilisations potentielles et les développements possibles du projet CREED.

2 ASPECTS COGNITIFS

Une difficulté majeure du contrôle aérien tient à son aspect « temps réel ». Ce qui est vrai à chaque niveau de la chaîne l'est tout particulièrement pour le contrôleur radar, dont la capacité peut, ponctuellement, être dépassée par les exigences du trafic supervisé. Cette capacité est généralement assimilée à ce qu'on appelle les « ressources cognitives » de l'individu, c'est-à-dire ce qu'il doit mobiliser « en lui-même » pour faire face aux différentes sollicitations de son travail : la prise d'informations, leur traitement, les actions, les activités de communication, etc. En salle de contrôle, la distribution de ces ressources est donc observable dans l'attention soutenue, permanente, vis-à-vis de l'image radar, au travers des interactions (verbales et non verbales) entre contrôleurs, de celles avec les pilotes, ainsi que dans la lecture et le renseignement des strips et l'utilisation du téléphone.

L'attention et l'activité mentale liée à la gestion des ressources restent les postes les plus consommateurs, en particulier dès que la charge de trafic devient forte ou très forte. Les tâches

proposées entre ces deux extrêmes. Enfin, les scénarios ont été présentés dans l'environnement radar spécifique de chaque jeu de données. La participation des contrôleurs des quatre grands aéroports a permis de recueillir des données précieuses et de constituer ainsi une base a priori unique au monde.

Le traitement des données s'est décomposé en deux phases. D'abord, une analyse graphique a permis d'alimenter les réflexions de l'expert sur sa compréhension du processus de détection des conflits et de vérifier l'existence d'une certaine cohérence dans les jugements exprimés. Au delà des particularismes individuels bien présents et souvent soulignés dans la littérature, l'étude a ainsi démontré qu'une perception commune des conflits par les contrôleurs existe bel et bien, ainsi que la dimension collective du travail, le découpage de l'espace et la formation initiale semblent l'exiger. Un modèle statistique a ensuite été développé pour permettre de reproduire les jugements observés, et de les interpoler à d'autres situations de conflit potentiel. Ce modèle a également permis d'estimer et d'étudier l'hétérogénéité de la population des contrôleurs.

plus précisément, le processus de détection des conflits. Une fois le problème clairement énoncé, dans la seconde partie nous donnerons des détails sur les jeux de données utilisés pour recueillir les jugements des contrôleurs. Les résultats seront présentés et discutés dans la troisième partie, en mettant l'accent sur les méthodes statistiques utilisées pour l'analyse graphique. La quatrième et dernière partie sera consacrée à la présentation des modèles logistiques construits et des connaissances alors obtenues. Des encarts devraient permettre au lecteur de mieux comprendre les outils utilisés dans cette partie. Enfin, la conclusion dévoilera les utilisations potentielles et les développements possibles du projet CREED.

2 COGNITIVE ASPECTS

One major difficulty of air traffic control resides in its "real-time" aspect. What is true at all levels in the chain is particularly so for the radar controller, whose capacity may occasionally be overloaded by the requirements of the supervised traffic. This capacity is generally assimilated with what is known as the individual's "cognitive resources", i.e. what he needs to mobilise "within himself" to face up to the various solicitations of his work: receiving information, processing it, taking action, communication activities, etc. In the control room, the distribution of these resources can be observed in the permanent, sustained

attention to the radar image, through the interactions (both verbal and non-verbal) between controllers and with the pilots, as well as through the reading and filling of the strips and the use of the telephone.	The attention and mental activity linked to resources management remain the most demanding areas, particularly as soon as the traffic load becomes heavy or very heavy. The subsidiary tasks then
---	---

annexes se voient alors progressivement réduites à leur expression minimale, et le contrôleur organique décharge le contrôleur radar d'un certain nombre d'actions secondaires. Parallèlement, le contrôleur radar privilégie l'attention allouée à l'image radar et aux échanges radio, et met en oeuvre deux processus cognitifs : la détection, et (si besoin) la résolution des conflits. On a coutume d'appeler « conflits » les situations susceptibles d'engendrer des valeurs de séparation entre avions inférieures au minimum admissible (cf. Encart n° 1).

Encart 1

**La séparation :
une question de point de vue ?**

Il n'est pas inutile de redéfinir un vocabulaire qui, pour être familier, n'en est pas moins imprécis. Plusieurs termes existent pour rendre compte des différents aspects de la séparation entre aéronefs, et pour la clarté du propos, il faut résister à la tentation de les substituer les uns aux autres. Nous expliquons donc le sens dans lequel nous les employons.

L'objectif du contrôle aérien est d'assurer qu'il existera en toute circonstance une séparation égale ou supérieure à la norme de séparation. Celle-ci est de 1000 pieds (ft) sur le plan vertical et, selon les performances des moyens radars, de 3, 5, ou 8 miles nautiques (NM) sur le plan horizontal. Il s'agit de minima admissibles. Deux avions ne devront jamais posséder une séparation inférieure à cette norme, simultanément sur les deux plans.

Toutefois, dans l'activité réelle, la perte de séparation (c.-à-d. une séparation observée inférieure à la norme) ne reste qu'une potentialité. En conséquence, et bien qu'elle soit dans la tâche la seule vraie référence, le contrôleur n'a jamais la possibilité d'en connaître la réalité, puisqu'il agit par anticipation pour éviter qu'elle se produise. L'incertitude qui entache la prévision des trajectoires empêche jusqu'à présent tout système de la calculer. On peut donc définir la détection comme le processus consistant à identifier, en procédant par extrapolation sur les positions futures (en temps et espace « 4D »), les configurations de trafic qui ont une probabilité suffisante de perte de séparation.

Cette détection peut être humaine ou artificielle. Dans le premier cas, il est fondamental de remarquer que le contrôleur procède par estimation, via sa perception — essentiellement visuelle, — de la probabilité de perte de séparation : les processus visuels sont alors au premier plan, appliqués à l'image radar. Observons que cette perception est, par nature, très liée au sujet. Lorsque le risque qu'il perçoit est significatif, on parle alors de conflit. Un conflit est donc un jugement formé par le contrôleur indiquant une possibilité suffisante de perte de séparation, dans le futur. Le terme de diagnostic est équivalent à celui de jugement.

Une première lecture des définitions précédentes permet d'avancer qu'un jugement de conflit entraînera une résolution, c.-à-d. une instruction de contrôle destinée à accroître suffisamment la valeur de la séparation naturelle pour que disparaisse le diagnostic de conflit. C'est effectivement le cas en général, mais pas systématiquement — à moins de convenir d'appeler conflit tout ce qui voit s'appliquer une résolution,

Box 1

**Separation:
a question of point of view?**

It may be useful to redefine some vocabulary that, although familiar, is nevertheless imprecise. Several terms exist to talk about various aspects of the separation between aircraft, and for the clarity of the explanations, we should resist the temptation to substitute them for each other.

Therefore, explained below is the meaning in which we use them. The aim of air traffic control is to ensure that in all circumstances there is a separation equal to or higher than the separation minima. This is 1000 feet (ft) on a vertical plane and, depending on the performance of the radar system, can be 3, 5 or 8 nautical miles (NM) on a horizontal plane. These are the acceptable minima. Two aircraft must never have a separation of less than this standard, simultaneously both vertically and horizontally.

However, in real activity, the loss of separation (i.e. an observed separation less than the minima) only remains a potential. Consequently, and although it is in the task the only real reference, the controller never has the possibility of knowing the reality, since he acts by anticipation to prevent it from happening. The uncertainty that harms trajectory predictions has so far prevented any system from being able to calculate this. We can therefore define detection as being the process that consists of identifying the traffic configurations that have a high enough probability of loss of separation, proceeding by extrapolation onto the future positions ("4D" – in time and space).

This detection may be human or artificial. In the first case, it is fundamental to note that the controller proceeds by estimation, via his perception – essentially visual – of the probability of loss of separation: the visual processes are then in the foreground, applied to the radar image. We can observe that this perception is, by nature, highly linked to the subject. When the risk he perceives is significant, then we talk about conflict. A conflict is therefore a judgement formed by the controller indicating a high enough possibility of loss of separation, in the future. The term diagnosis is equivalent to that of judgement.

A first look at the above definitions allows us to state that a judgement of conflict will lead to a resolution, i.e. a control instruction that aims to increase the natural separation value sufficiently, so that the conflict diagnosis disappears. This is indeed the case in general, but not systematically – unless we

mais ce n'est que déplacer le problème, avec à notre avis des difficultés de modélisation plus grandes encore.

Par exemple, pour des valeurs de séparation d'environ 6NM horizontalement et un même niveau de vol, il apparaît que dans la grande majorité des cas, une résolution est implémentée, signe que cette situation est interprétée comme conflit par les contrôleurs. Cependant, dans un pourcentage non négligeable des cas, aucune résolution n'est observable, malgré des paramètres objectifs strictement similaires. Généralement, il ne s'agit pourtant pas d'erreurs d'appréciation de la part des contrôleurs : en réalité la différence réside dans la décision du contrôleur de contenir le risque autrement qu'en altérant une trajectoire. Dans la mesure où la gestion par un maintien du doute existe, il est possible de contrôler (au sens littéral) la situation en portant une attention soutenue à cette configuration et à son évolution, de manière quasi-continue. De la sorte, pour une séparation estimée un peu faible mais suffisante, aucune action de contrôle pouvait n'être observée, sans que la sécurité puisse être mise en cause ni que la situation puisse être classée comme moins conflictuelle que celle pour laquelle une résolution était décidée. Toutefois pour qu'apparaisse ce type de « contrôle », il semble devoir être exigé que la charge de travail qu'il subit à ce moment là soit faible, voire même très faible. Cette caractéristique est parfaitement connue (et utilisée) des contrôleurs qualifiés, mais rend moins aisément utilisable la définition exacte du conflit.

agree to call "conflict" anything that applies to a resolution, but this is really just shifting the problem, with even greater modelisation difficulties, in our opinion.

For example, for separation values of approximately 6NM horizontally and at the same flight level, it appears that in the vast majority of cases a resolution is implemented - a sign that this situation is interpreted as a conflict by the controllers. However, in a significant percentage of cases, no resolution can be observed, in spite of the strictly similar objective parameters. Generally, however, these are not appreciation errors by the controllers: in reality the difference resides in the controller's decision to differently contain the risk other than by altering a trajectory. Insofar as the management by maintaining doubt exists, it is possible to control (literally) the situation by giving sustained attention to this configuration and to its evolution, in a quasi-continuous manner. In this way, for a separation that was estimated to be a bit tight but sufficient, no control action could be observed, without safety being jeopardized or without the situation being classified as less conflictual than the one for which a resolution was decided. However, for this type of "control" to appear, it seems to require that the workload at that time is light, or even very light. This characteristic is perfectly well known (and used) by qualified controllers, but it makes the exact definition of the conflict less easy to use.

2.1 La détection d'un conflit, c'est quoi finalement ?

Nous nous intéressons ici au premier des deux processus : la détection des conflits. Le plus souvent, celui-ci est modélisé de manière purement analytique. Des valeurs d'incertitudes préalablement fixées et associées aux paramètres géométriques décrivant la situation (par exemple 5% pour la vitesse sol (Vs) et 20% pour la vitesse verticale (Vz), avec un empan temporel de 8 minutes) permettent alors d'estimer une probabilité de conflit. Une paire d'avions est déclarée « en conflit » lorsque cette probabilité dépasse un certain seuil. Si l'on impose qu'un nombre raisonnable de « vrais conflits » soient identifiés, cette méthode fournit un nombre trop important de fausses alertes (5 à 10 fois le nombre de pertes de séparation effectives) pour être utilisable dans un cadre opérationnel, et ce quel que soit le point de vue adopté. Une classification aussi grossière est en effet trop peu rentable pour déterminer les cas où une action est à envisager, puisqu'un contrôleur identifierait lui-même bien moins de fausses alertes, et verra donc peu d'intérêt à utiliser une telle analyse. Ceci ne tient évidemment pas au fait que l'humain serait capable de plus de clairvoyance concernant l'avenir, ou qu'il intégrerait plus d'informations et de règles acquises par l'expérience : par exemple la connaissance de la compagnie aérienne, des conditions météorologiques et de certaines intentions de pilotage. La raison principale réside plutôt dans un fonctionnement propre, permettant au contrôleur de gérer le doute [Leroux] sur les positions futures, c.-à-d. de s'accommoder d'une ambivalence dans le pronostic conflit/non conflit, jusqu'à un certain stade de celui-ci.

become progressively reduced to a minimum, and the planner controller helps out the radar controller with a certain number of secondary actions. In the same way, the radar controller gives priority in his attention to the radar image and the radio exchanges, and uses two cognitive processes: conflict detection, and (if necessary) resolution. We are accustomed to calling "conflicts" situations that may lead to separations between aircraft that are less than the admissible minima (see Box N° 1).

2.1. What is conflict detection exactly?

Here we are concerned with the first of the two processes: conflict detection. Most frequently this is modelised in a purely

analytical way. Previously fixed uncertainty values, associated with geometrical parameters, describe the situation (e.g. 5% for the ground speed (Vs) and 20% for the vertical speed (Vz), with a time span of 8 minutes) and enable the probability of conflict to be estimated. A pair of aircraft is declared "in conflict" when this probability exceeds a certain threshold. As all of the "real conflicts" have to be identified, this method can give too many false alerts (5 to 10 times the number of real losses of separation) in an operational context, whatever the point of view used. Such a rough classification is indeed not profitable enough to determine the cases where action is to be

envisaged, since a controller himself would identify far fewer alerts, and would therefore not see much interest in such an analysis. This, of course, does not mean that a human being would be capable of more perceptiveness concerning the future, or that he would be able to integrate more information and rules acquired from experience, such as knowledge about the airline, the weather conditions and some of the pilot's intentions. Rather, the main reason resides in the precise way the controller works, which enables him to manage the doubt [Leroux] about future positions, i.e. to adapt to an ambivalence in the conflict/no conflict predictions, up to a certain point. This capacity is an essential part of the controller's expertise, and is acquired with experience. The calculations used by beginners are progressively abandoned in favour of an extension of doubt over time [Bisseret]. As a 4D trajectory cannot be predicted – except for integrating uncertainty margins that are too large – the controller develops this capacity to master the situation, both without taking action and without being certain that "it will make it" at the same time. This means "letting the situation develop" as is sometimes expressed in the control room by the controllers themselves.

Because he postpones the need to take action until later, it can be understood that the controller generates fewer false alerts than a calculation made at a fixed time horizon. This mode of operation is largely due to the system's lack of precision concerning trajectory forecasts. In other terms, a substantial improvement (e.g. with a future generation of FMS) would modify the present confidence that man has in the computer. Hence, for uncertainties of 1.4% in Vs and 10% in Vz, the number of alerts would only be twice the number of real losses of separation for a time horizon of 10 minutes [Alliot/Durand/Granger]. In these conditions, the information provided by the machine would start to be essential, and would probably lead to new modes of operation from the controllers, moving towards an increase in the use of calculations as opposed to using perceptions and doubt management. But at present, human processes prevail. Therefore, it is in our greatest interest to reach a level of sufficient knowledge of the detection process actually carried out by the human operator, concerning his management of uncertainty. This understanding is a necessary condition to develop pertinent computer tools for the controller in real situations, particularly

Cette capacité est une part essentielle de l'expertise du contrôleur, et s'acquiert avec elle. En effet, les calculs utilisés par les débutants sont progressivement abandonnés au profit d'une prolongation du doute dans le temps [Bisseret]. La trajectoire 4D ne pouvant être prédite — sauf à intégrer des marges d'incertitude trop importantes, — le contrôleur développe cette capacité de maîtriser la situation, à la fois sans agir ni être certain que « ça passe ». Il s'agit de « laisser mûrir » la situation, comme on l'entend parfois dire en salle par les contrôleurs eux-mêmes. Puisqu'il repousse à plus tard la nécessité d'une action, on comprend mieux que le contrôleur génère moins de fausses alertes qu'un calcul effectué à échéance temporelle fixe.

Un tel mode de fonctionnement est largement dû au manque de précision du système en ce qui concerne la prédiction de trajectoire. En d'autres termes, une amélioration substantielle (par exemple avec une future génération de FMS) modifierait la relation de confiance actuelle entre l'homme et le calculateur. Ainsi, pour des incertitudes de 1,4% en Vs et 10% en Vz le nombre d'alertes ne seraient que le double des pertes de séparation réelles pour un horizon temporel de 10 minutes [Alliot/Durand/Granger]. Dans de telles conditions, l'apport d'informations en provenance de la machine commencerait à devenir primordial, et induirait probablement des modes opératoires nouveaux chez les contrôleurs, dans le sens d'un accroissement de l'usage du calcul par rapport au perceptif et à la gestion du doute. Mais actuellement, ce sont les processus humains qui prévalent. Il y a donc tout intérêt à parvenir à une connaissance suffisante du processus de détection réellement effectué par l'opérateur humain, sur sa gestion de l'incertitude. Cette connaissance est en effet une condition nécessaire pour réaliser des aides informatiques pertinentes pour le contrôleur en situation réelle, en particulier en situation de forte charge. Le premier choix de cette étude a donc été d'adopter un point de vue « égocentré », c'est-à-dire qui prenne comme référence le jugement produit par le contrôleur. Il ne sera en particulier pas question ici d'une « validation » des jugements des contrôleurs selon leur écart aux normes de séparation théoriques (1000 ft / 3 à 8 NM), envisagées comme le niveau cible de compétence. Ce qui nous importe est la connaissance de la capacité experte réelle et observable de perception des conflits. En d'autres termes : on sait que les contrôleurs intègrent dans leurs propres jugements de risque de conflit une marge appropriée, supplémentaire à la norme, afin d'adapter l'objectif de la tâche aux processus cognitifs utilisés. Connaître cette marge, et éventuellement mettre au point un moyen de l'appréhender, constitue l'objectif du travail présenté ici.

2.2 Le rôle des automatismes

Les paramètres utilisés pour produire le jugement sont pour la plupart visibles sur l'écran radar. Ce sont essentiellement les positions des aéronefs et leur extrapolation à partir d'indices visuels. Il s'agit donc d'informations analogiques, à l'exception notable de l'altitude, qui existe sous forme alphanumérique (2ème ligne de l'étiquette). Le traitement de cette donnée par le contrôleur est donc de nature numérique (calcul)⁽¹⁾.

(1) En réalité, ce n'est pas si évident. En effet, la trajectoire d'un avion en évolution a un sens : ses points d'entrée et de sortie secteur (ou terrains de départ et d'arrivée) déterminent en particulier un intervalle à l'intérieur duquel l'expérience montre que l'avion doit se trouver. Cela correspond sensiblement à la notion de valeur par défaut des paramètres de vol utilisée dans [Leroux]. On peut donc faire l'hypothèse que l'ensemble [position + altitude + "identité" du vol] forme un motif (pattern), et peut être traité comme un tout. On peut a priori recourir à ce type de traitement – au lieu du calcul – pour connaître l'état futur probable d'un paramètre, et en particulier l'altitude.

Le traitement perceptif, visuel en l'occurrence, des autres paramètres pour extrapoler les positions 4D et aboutir au jugement de risque de conflit ne fait par contre aucun doute. Différentes aides existent qui améliorent notablement la précision de ce jugement : la comète (c.-à-d. les positions passées), le vecteur vitesse, etc. Elles sont intégrées dans les heuristiques perceptives acquises initialement par les contrôleurs au cours de leur formation pratique, puis enrichies et consolidées par l'expérience. Ce terme d'heuristique désigne l'art de trouver en utilisant des approches qui n'éliminent que progressivement les alternatives, conservant au bout du compte une gamme restreinte de solutions qui tendent vers celle qui serait optimale.

Le traitement des informations au moyen d'heuristiques perceptives présente deux caractéristiques. En premier lieu, elles sont particulièrement adaptées au traitement d'informations de type analogique, telles que celles disponibles sur l'image radar. Relativement à ce que nécessiterait un calcul, le temps de traitement associé est très court, pour une perte en précision paradoxalement faible voire inexistante (Bisseret). Avant d'être des heuristiques perceptives, les processus en question (extrapolations mentales, séquençement des actions, points de fixation du regard sur la carte radar, ...) sont des processus planifiés, dirigés consciemment. A ce stade, une fraction importante des ressources mentales du contrôleur leur est consacrée, pour l'essentiel plutôt sous forme d'attention que de calculs. Or les ressources humaines en terme de capacité de traitement de l'information sont limitées. Dans ces conditions, moins chacune des actions à mener utilisera de ressources, et plus il sera possible d'en mener simultanément, donc plus le contrôleur sera sûr et capacitif. Dans une activité où la composante temps réel est fondamentale, on comprend l'intérêt d'un très faible temps de traitement et d'une faible consommation de ressources mentales. La quantité de trafic qu'un contrôleur sera capable de gérer, à niveau d'effort constant, sera d'autant plus grande que son expertise lui aura permis d'automatiser les processus nécessaires à sa tâche.

Cette automatisation va se faire progressivement avec la pratique. Elle caractérise donc le contrôleur expérimenté (par opposition au débutant), lui permettant de traiter simultanément un nombre important d'avions avec une grande économie de charge mentale. Dans le domaine de la psychologie cognitive, la tâche du contrôleur aérien est l'exemple type de transformation d'une activité complexe et consciente en un processus largement automatisé. Bien sûr, chaque processus du traitement ne va pas être automatisable, loin s'en faut. Ce qui peut l'être n'en est que plus précieux.

On peut illustrer les effets de cette automatisation à partir des deux faits suivants, que tout un chacun peut observer en fréquentant les salles de contrôle :

Il est notoire que les contrôleurs parviennent parfois difficilement à accepter soit des outils opérationnels réellement nouveaux — même s'ils sont théoriquement plus performants — soit certaines modifications à ceux qui existent. En effet, ces

when there is a heavy load. The first choice of this study was therefore to adopt an "egocentric" point of view, meaning that the reference used is the judgement produced by the controller. There is no concern here, in particular, about a "validation" of controllers' judgements according to their deviation from the theoretical separation standards (1000 ft / 3 to 8 NM), which would be envisaged as the target level of competence. What is important is the understanding of the controllers' real, observable, expert capacity in the perception of conflicts. In other words, we know that controllers integrate into their own judgements an appropriate margin to cover the risk of conflict, in addition to the standard one, in order to adapt the aim of the task to the cognitive processes used. Understanding this margin, and possibly developing a means of getting a good grasp of it, is the aim of the work presented here.

2.2. The role of automatisms

The parameters used to produce the judgement are mostly visible on the radar screen. These are essentially the positions of aircraft and their extrapolation from visual indices. Hence, this is analogical information, with the remarkable exception of the altitude, which is in alphanumerical format (2nd line of the

label). This data is therefore processed numerically by the controller (calculation)⁽¹⁾. The perceptive process, visual in this case, of the other parameters in order to extrapolate the 4D positions and result in the judgement of a risk of conflict, on the contrary, does not involve any doubt. Different tools exist that notably improve the accuracy of this judgement: the comet (i.e. the past positions), the speed vector, etc. They are integrated into the perceptive heuristics that are initially acquired by controllers during their practical training, and then improved and consolidated through experience. The term of "heuristic" designates the art of finding, by using approaches that only eliminate the alternatives progressively, finally retaining a limited range of solutions that lead towards the optimum one. The processing of information using perceptive heuristics has two characteristics. Firstly, it is particularly adapted to processing analogical information, such as that available on the radar image. Compared to the time required to make a calculation, the associated processing time is very short, for a loss of precision that is paradoxically small, or even non-existent (Bisseret). Before being perceptive heuristics, the processes in question (mental extrapolations, sequencing of actions,

(1) In reality, this is not as obvious as it seems. The trajectory of a manoeuvring aircraft has a direction: its sector entrance and exit points (or departure and arrival airfields) determine, in particular, intervals within which experience shows that the flight parameters (altitude, speed, ...) should be, according to the aircraft type. This corresponds approximately to the notion of default value of the flight parameters used by [Leroux]. It is therefore possible to make the hypothesis that the whole (position + altitude + flight "identity") forms a pattern, and can be treated as a whole. This type of process can therefore be used — instead of a calculation — to know the future probable state of a parameter, and the altitude in particular.

concentrating on a spot on the radar map, etc.) are planned processes, which are consciously directed. At this stage, a large part of the controller's mental resources are devoted to these, largely in the form of attention rather than calculations. However, human resources in terms of the capacity to process information are limited. In these conditions, the less each of the actions to be performed uses the resources, the more actions it will be possible to perform simultaneously. The controller will thus be all the more sure and effective. In an activity where the real-time component is fundamental, the interest of having a very short processing time and a low consumption of mental resources is understandable. The quantity of traffic that a controller will be able to manage, with a constant level of effort, will be larger if his expertise allows him to automatise the processes required for his task. This automatization will occur progressively with practice. It therefore characterises the experienced controller (as opposed to a beginner), enabling him to process simultaneously a large number of aircraft with great savings in mental load. In the area of cognitive psychology, the air traffic controller's task is a typical example of the transformation of a complex, conscious

activity into a largely automatised process. Of course, each process cannot be automatised, far from it. The ones that can, therefore, are all the more precious. The effects of this automatization can be illustrated by the two following events, which anyone can observe in the control rooms: It is a well-known fact that controllers can sometimes have difficulty in accepting either entirely new operational tools – even if they are theoretically more effective – or certain modifications to ones that already exist. Most of the time these changes result in a modification to the arrangement of the information, or the sequencing of processing actions. This is enough to prevent the initial automatisms from being played out, or at least, from not being played out so well. The chain of events must be guided consciously again, and the processing of information in which the new tool is involved needs to be supervised, which is far more costly in terms of cognitive resources. These changes therefore require a period for the construction of new automatisms during which, for the same cognitive effort, the controller is less effective. For the same level of traffic, maintaining the same level of performance will require more effort by the controller; hence his reticence to modify his

changements ont la plupart du temps comme conséquence de modifier la disposition des informations, ou le séquençage des actions du traitement. Ceci suffit à empêcher les automatismes initiaux de jouer, ou tout au moins de ne plus jouer aussi bien. Il faut à nouveau guider consciemment l'enchaînement, superviser le traitement des informations dans lequel le nouvel outil est impliqué, ce qui est beaucoup plus coûteux en termes de ressources cognitives. Ces changements nécessitent donc une période de construction de nouveaux automatismes pendant laquelle, à coût cognitif égal, le contrôleur est moins performant. A quantité de trafic constante, maintenir son niveau de performance demandera plus d'efforts à ce contrôleur ; d'où sa réticence initiale à modifier son environnement de travail. Une deuxième conséquence de ces automatismes est le temps assez important nécessaire à l'obtention de la qualification de contrôle. En effet, l'ATC constitue une tâche a priori assez simple à effectuer, en particulier du fait du nombre très limité de paramètres sur lesquels agir (niveau, route, vitesse, taux). L'objectif est simple, et le nombre de leviers pour agir à la fois suffisant et succinct. En réalité, l'expertise ne réside pas dans la performance logico-déductive, mais dans l'acquisition, puis l'intégration des automatismes, ce qui explique la bien longue phase d'apprentissage pratique.

En second lieu, le traitement par heuristiques laisse la possibilité à de fortes différences individuelles de se développer. Il est clair qu'un traitement algorithmique, faisant localement usage de calculs arithmétiques, engendrerait des marges d'incertitude raisonnables et surtout calculables avec précision. Ce n'est pas le cas pour des heuristiques. La détection et la résolution des conflits montrent d'importantes disparités individuelles, c'est-à-dire d'un individu à l'autre (inter individuelles), ou pour un individu donné à qui on soumettrait une même situation à deux moments différents (intra individuelles). Pour nombre de situations de trafic, il n'est donc pas possible de prédire si elles seront identifiées à un moment ou à un autre comme conflictuelles. Ce fait peut paraître surprenant, puisque l'on considère, à juste titre, que le contrôle aérien nécessite une importante expertise et qu'il est tout sauf l'exécution d'instructions aléatoires. Le premier, A. Bissereet a montré qu'en matière de gestion des conflits, la précision des calculs n'est pas ce qui importe le plus au contrôleur. En d'autres termes, il est beaucoup moins préjudiciable pour lui d'accepter une part d'imprécision si cela le garantit davantage contre les omissions de diagnostic de conflit, que de recourir au calcul pour réduire l'incertitude au maximum, et de se faire « surprendre » de temps à autre par une erreur de diagnostic. En effet, précision ici n'est pas synonyme d'exactitude de diagnostic, dans la mesure où les paramètres de vol ne sont pas constants, et n'évoluent pas toujours selon des fonctions monotones. Ainsi⁽¹⁾, on ne peut parler de précision que sur la base des valeurs des paramètres présents (et immédiatement passés) et du comportement futur qu'on en a extrapolé. Dès lors qu'il y a un changement de sens dans l'évolution des valeurs (accélération au lieu de décélération, par ex.) il y a une possibilité de risque de conflit là où il n'y en avait pas initialement, sans que l'on puisse s'en prémunir a priori.

(1) A défaut de demander en permanence les valeurs à l'équipage (irréaliste), il faut contraindre l'aéronef par une instruction (imposition d'un cap, d'un niveau, etc...).

Par conséquent, le maintien du doute dans la représentation mentale du contrôleur est très salutaire. En différant autant que raisonnable la décision de classer telle configuration de trafic en conflit ou en non conflit, le contrôleur minimise les coûts d'une erreur d'appréciation. De plus, il peut plus facilement s'approcher de la résolution nécessaire et suffisante. Insistons encore : cette phase de doute est fondamentale en ATC, — le contrôle aérien, c'est l'art de douter [Leroux]. Autrement dit, c'est la gestion de ce doute, c'est-à-dire sa prolongation ou au contraire la décision de sa suppression qui constitue le cœur du savoir faire. Si les heuristiques perceptives sont bien adaptées au diagnostic des conflits, elles sont également en grande partie propres à l'individu, ce qui ne facilite pas, loin s'en faut, l'analyse et la modélisation du savoir faire des contrôleurs. Les études passées ont ainsi essentiellement montré des processus de détection/résolution des conflits qu'ils dépendaient du contrôleur et de la situation ! Les résultats exposés ci-après tempèrent ce constat en de mettant à jour un certain nombre de régularités, même si les importantes variations individuelles ont été à nouveau confirmées.

2.3 Quels sont les critères pour juger d'un conflit ?

La gestion du doute ne vise qu'à s'accommoder au mieux de l'avenir, et n'ignore jamais totalement l'évaluation des paramètres de vol à venir, les instructions de contrôle possibles et leurs effets. Ces paramètres restent donc la base à partir de laquelle sont formés les jugements des contrôleurs. Il apparaît que les marges nécessaires pour un jugement de « non conflit » sont plus importantes pour une séparation sur le plan vertical que pour une séparation horizontale. Globalement, elles sont d'environ 5 fois la norme (soit 5000') pour la première, et seulement 2 fois (env. 10 NM) pour la seconde. Il n'est donc pas surprenant que le paramètre « taux » (de montée ou de descente) ne soit guère utilisé pour le diagnostic, à l'exception peut être pour des paires d'aéronefs d'attitude identique lorsqu'ils sont en montée ou en descente.

Mais ce ne sont que des valeurs moyennes : les dimensions heuristiques et perceptives utilisées induisent, on l'a vu, un certain nombre de sources de variation individuelles. Les études passées mettent donc en évidence des différences de jugement qui peuvent être imputées aux causes suivantes :

- > **le moment du diagnostic** : les experts ne se sentent en mesure d'exprimer un jugement que lorsque les aéronefs se trouvent à environ 30NM du point d'intersection. Les jugements de non conflit sont formulés plus tardivement que les jugements de conflit ;
- > **l'attitude des avions** : on entend par là l'état de l'évolution verticale des aéronefs (montée / stable / descente). Même si l'existence d'un (au moins) aéronef évolutif permet d'espérer une résolution du conflit sur le plan vertical, la trajectoire 4D admet une incertitude significativement plus importante dans ce cas, que les heuristiques utilisables ne savent pas réduire véritablement. Le jugement de conflit (ou de non conflit) est donc moins aisé dès lors que l'un des aéronefs au moins est en évolution ;

working environment. A second sequence of these automatisms is the fairly long time-period needed to obtain the controller's qualification. ATC is, a priori, a fairly simple task to carry out, particularly due to the very limited number of parameters to be acted upon (level, route, speed and climb rate). The aim is simple, and the number of levers to act upon at one time are sufficient and brief. In reality, however, the expertise does not reside in the logico-deductive performance, but in the acquisition, then the integration of automatisms, which explains the very long practical learning phase. Furthermore, processing using heuristics leaves the possibility of large individual differences developing. It is clear that algorithmic processing, which locally uses arithmetical calculations, would give rise to uncertainty margins that are reasonable and, above all, that can be calculated with precision. This is not the case for heuristics. The detection and resolution of conflicts show large individual disparities, i.e. from one individual to another (inter individual), or for a given individual to whom the same situation is submitted at two different times (intra individual). For a number of traffic situations, it is therefore not possible to predict if they will be identified at one time or another as being conflictual.

This fact may appear to be surprising since it is considered - and rightly so - that air traffic control requires a lot of expertise and it is anything but the execution of random instructions. The first, A. Bisseret, showed, in the area of conflict management, that the precision of the calculations is not what is of prime importance for the controller. In other terms, it is far less detrimental for him to accept a certain degree of imprecision if it gives him better guarantees against omissions of conflict diagnosis, than to use calculations in order to reduce uncertainty to a maximum, and be "surprised" from time to time by a diagnostic error. Precision here, in fact, is not synonymous with diagnostic accuracy, insofar as the flight parameters are not constant, and do not always evolve monotonically. Therefore⁽¹⁾, one can only talk about precision on the basis of the parameter values at the present time (and in the immediate past), and future behaviour that has been extrapolated from this. As soon as there is a change of direction in the evolution of the values (e.g. acceleration instead of deceleration) there is a possible risk of conflict where there had not been initially, without anyone being able to be warned about it in advance. Consequently, maintaining the doubt in the controller's mental representation is very

(1) As it is not possible to request the flight crew for data permanently (unrealistic), the aircraft must be given an order by issuing an instruction (a required heading, level, etc...)

profitable. By deferring the decision to classify a certain traffic configuration into conflict or no conflict for as long as is reasonable, the controller minimises the costs of an error of appreciation. Furthermore, he can more easily approach the resolution that is necessary and sufficient. To insist once again: this phase of doubt is fundamental in ATC - air traffic control is the art of doubting [Leroux]. In other words, it is the management of this doubt, i.e. to prolong it or, on the contrary, to decide to suppress it, which makes up the heart of the know-how.

If perceptive heuristics are well adapted to the diagnosis of conflicts, they are also, to a large extent, specific to the individual, which does not facilitate the analysis and modelisation of the controllers' know-how, far from it. Previous studies have basically shown from conflict detection/resolution processes that they mostly depend on the controller and on the situation! The following results temper this observation by bringing to light a certain number of regularities, even if the significant individual variations have once again been confirmed.

2.3. What are the criteria for judging a conflict?

Doubt management only aims to adapt to the future as well as possible,

and never totally ignores the evaluation of the future flight parameters, or the possible control instructions and their effects. These parameters remain, therefore, the basis on which the controllers' judgements are formed. It would appear that the necessary margins for a "no conflict" judgement are higher for a vertical separation than for a horizontal separation. Globally, they are approximately 5 times the standard (i.e. 5000') for the former, and only twice (approximately 10 NM) for the latter. Hence, it is not surprising that the "rate" parameter (of climb or descent) is hardly ever used for diagnosis, except perhaps for pairs of aircraft with the same attitude when they are climbing or descending. But these are only average values: the heuristic and perceptive dimensions used, as we have seen, lead to a certain number of sources of individual variations. Previous studies have shown differences in judgement that can be attributed to the following causes:

- > **The moment of diagnosis:** the experts only feel able to express a judgement when the aircraft are at about 30NM from the point of intersection. "No conflict" judgements are formulated later than "conflict" judgements.
- > **Aircraft attitude:** this means the status of the aircrafts' vertical manoeuvres

- > **la géométrie du conflit (rattrapage, face-à-face, croisement) :** c'est, pour certains auteurs, une source de variation moins importante que l'attitude. Il a été néanmoins montré que plus l'angle de convergence augmente, plus l'imprécision s'accroît, avec un pic autour de 90°. Dans ce dernier cas, la séparation est parfois estimée sur le seul plan horizontal, donc indépendamment de l'attitude des avions ;
- > **le ratio des vitesses :** son effet semble incertain. On montre parfois qu'il sert à renforcer un jugement, lorsqu'il émerge.

Ces études abordent souvent la question de la détection des conflits sous l'angle du temps de réponse et de la précision (pourcentage de pertes de séparation avérées ultérieurement, relativement au nombre de paires d'avions déclarées en conflit). Ceci revient à évaluer la performance à la tâche, à l'aune de ce qui est prescrit (les normes de séparations, à strictement parler). Si ces normes prescrites ont toute leur utilité dans les domaines juridiques et techniques (performance des moyens radar), analyser sur leur base le savoir faire humain est un bon moyen de ne jamais le rendre intelligible. C'est pour cette raison que l'expérimentation menée dans le cadre de CREED ne s'y réfère pas, pas davantage que le modèle statistique final. Par ailleurs, il nous semble nécessaire d'ajouter deux sources de variation :

- > **la rectitude des trajectoires :** l'extrapolation visuelle du point d'intersection est plus aisée lorsque les routes des avions qui convergent sont rectilignes — et ce d'autant qu'un vecteur vitesse est disponible. Il l'est a priori moins lorsque la trajectoire qui reste à suivre jusqu'au point d'intersection est composée de plusieurs segments d'orientations magnétiques différentes ;
- > **la charge de travail contextuelle :** la dernière source de variation de la liste, mais sans doute pas la moins importante. Du fait de ses ressources cognitives limitées, le contrôleur n'a pas la capacité de gérer le doute de la même façon s'il n'existe qu'une faible charge de travail sur le secteur, ou si elle est très élevée. Cette observation est aisément faite à partir de toute position de contrôle réelle et démontre un fait important : la double dépendance qui existe entre les processus de détection et de résolution. En effet, on peut s'attendre à ce que, lorsqu'un conflit est détecté, une action de résolution soit entreprise. A l'inverse, on observe qu'une forte charge de travail fait parfois identifier comme conflits des configurations qui ne l'auraient pas été en situation de faible charge. Nous verrons comment nous avons tenté de limiter les effets de ce sens de la dépendance, afin d'obtenir un jugement portant exclusivement sur la détection.

On peut objecter que la liste des sources de variation n'est probablement pas close pour autant. Mais il n'est peut-être pas nécessaire d'être exhaustif. Ce qui compte, en réalité, c'est la représentation que le contrôleur se fait de la situation, en intégrant toute information pertinente provenant des supports à sa disposition. Cette image mentale à construire doit être fidèle à

la situation, du point de vue de l'anticipation à faire et de la décision à prendre. Le processus de détection des conflits est bien sûr une part importante de sa conscience de la situation (Situational Awareness ou SA [Endsley]) qui différencie cette notion globale de représentation ou d'image mentale selon trois niveaux. Le 1er niveau est celui de la perception simple des éléments spatiaux et temporels : la position des avions, une certaine idée de leur vitesse (liée au déplacement des plots ou à la lecture de la vitesse sol de l'étiquette), etc. Le 2nd niveau est celui de l'intégration des percepts du niveau précédent, initialement sans relations et qui peuvent à ce niveau former des motifs (patterns), c'est-à-dire induire un premier niveau d'organisation globale des trajectoires. Le 3ème niveau est celui du sens que possède la situation de trafic considérée. C'est en particulier le niveau où se projettent les différents éléments de l'environnement et leur évolution temporelle, les actions envisagées et l'estimation de leurs effets.

On peut reconnaître à ce concept de SA le mérite de fournir des critères concrets pour interpréter les différences entre novices et experts (les novices n'ayant globalement accès qu'au premier niveau). En outre, il a permis d'identifier la cause d'une proportion intéressante d'erreurs observées en situation réelle. Pour autant, il bute sur la détermination d'une liste exhaustive de ce qui est nécessaire pour juger toute situation de trafic, et du sens de ces informations qui peuvent selon les cas, appartenir à un niveau ou à un autre parmi les trois. Bien souvent, ce n'est qu'a posteriori, c'est-à-dire une fois qu'il a fait défaut, qu'un élément peut être identifié comme nécessaire. Une des raisons est bien sûr qu'une image mentale est un processus dynamique, complexe, bien différent d'un assemblage d'éléments disjoints. Même si elle est largement automatisée et souvent extrêmement rapide (< 200 ms), la prise d'information visuelle, on l'a vu, est un processus cognitif à part entière. Elle est un tri dans les informations disponibles, effectué en fonction du contexte et des attentes préalables. C'est donc un processus interactif, qui se prête mal à une décomposition simple.

Dans l'expérimentation décrite ci-après, nous avons tenté de contrôler au maximum les variables de l'environnement (autres que les variables explicatives) afin d'en réduire au mieux les effets. Parallèlement, nous avons cherché à reproduire un contexte de travail global qui soit identique pour tous les sujets et au plus près de celui de la salle.

Pour résumer, il n'existe pas de modèle simple du processus de détection des conflits. Le recours au calcul n'est que marginalement observé chez les experts, excepté peut être dans le cas d'un contexte inconnu. Le traitement est donc fait essentiellement à partir de motifs visuels, et présente une variabilité qui ne nuit pas à la sécurité, mais qui rend plus difficile la modélisation du savoir faire. Les jugements varient donc en fonction du problème et du contrôleur, au point de ne permettre d'en tirer, au premier examen, que des règles très globales et qualitatives.

(climb/stable/descent). Even if the existence of (at least) one climbing or descending aircraft gives the hope of resolving a conflict on the vertical plane, the 4D trajectory allows a significantly larger uncertainty in this case, which the usable heuristics cannot really reduce. The judgement of conflict (or no conflict) is therefore less easy when at least one of the aircraft is climbing or descending.

- > **Geometry of the conflict (overtaking, face-to-face, crossing):** for some authors this source of variation is less important than the attitude. However, it has been shown that the more the convergence angle increases, the higher the imprecision, with a peak at around 90°. In this last case, the separation is sometimes estimated just on the horizontal plane, thus independently of the aircraft attitude.
- > **Speed ratio:** its effect seems to be uncertain. It is shown sometimes that it is used to reinforce a judgement, when it emerges.

These studies often tackle the question of conflict detection from the angle of response time and precision (percentage of losses of separation proved later, in relation to the number of aircraft pairs declared to be in conflict). This is equivalent to evaluating the performance of the task,

according to what is prescribed (separation minima, strictly speaking). Although these prescribed standards are extremely useful in legal and technical areas (radar system performance), analysing human know-how on this basis is a good way of never making them intelligible. For this reason, the experiments carried out in the CREED context do not refer to them, and neither does the final statistical model. In addition, it would seem necessary to add two other sources of variation:

- > **Trajectory rectitude:** visual extrapolation of the intersection point is easier when the routes of the converging aircraft are rectilinear – and this is all the more true when a speed vector is available. A priori it is less so when the remaining trajectory up to the point of intersection is composed of several segments of different magnetic headings.
- > **Contextual workload:** the last source of variation on the list, but surely not the least important. Because of the controller's limited cognitive resources, he is not capable of managing the doubt in the same way if there is only a light workload in the sector, or if it is very heavy. This observation is easily made from any real control position, and shows up an important fact: the double dependency that exists between the detection and resolution

processes. It can be expected that a resolution action is taken when a conflict is detected. On the other hand, it is observed that a heavy workload sometimes causes configurations to be identified as conflicts, which would not have been so with a light workload. We will see how we tried to limit the effects of this sense of dependency, in order to obtain a judgement based exclusively on detection.

One may object that the list of sources of variation is probably not finished for all that. But perhaps it is not necessary to be exhaustive. What counts, in fact, is the representation that the controller himself makes of the situation, integrating all the pertinent information coming from the supports at his disposal. This mental image to be constructed must be faithful to the situation, from the point of view of the anticipation of what to do and the decision to make. The process of conflict detection is of course an important part of situational awareness or SA [Endsley], which differentiates this global idea of representation or mental image according to three levels. The first level is that of simple perception of the spatial and temporal elements: the position of the aircraft, a certain idea of their speed (linked to the plot movements or reading

the ground speed on the label) etc. The second level is that of integrating the percepts of the previous level, initially without relation, and which can at this level form patterns, i.e. induce a first level of global organisation of the trajectories. The third level is that of the meaning that the considered traffic situation has. This is, in particular, the level where the various elements of the environment and their evolution over time are projected, as well as the envisaged actions and an estimation of their effects. It must be recognised that this SA concept has the merit of providing concrete criteria for interpreting the differences between the novices and the experts (as the novices globally only have access to the first level). Furthermore, it has allowed identification of the cause of an interesting proportion of errors observed in real situations. However, it comes up against the difficulty of determining an exhaustive list of what is necessary in order to judge any traffic situation, and the meaning of this information, which may belong to one level or another of the three, according to the case. Very often, it is only later, meaning once it has been found to be lacking, that an element may be identified as necessary. One of the reasons is, of course, that a mental image is a dynamic, complex process, very different from an

Ainsi, qu'il perçoive ou non l'existence probable des normes de séparation, il a été observé que le contrôleur peut décider d'agir pour accroître sa marge. Cet accroissement est très variable selon le contrôleur, le contexte, le moment du recueil, etc. Le bruit induit est donc important, ce qui est très pénalisant pour la mesure concrète des paramètres du processus de détection. Une des contraintes que nous nous sommes fixées dans la mise au point de notre paradigme expérimental est de recueillir le jugement du contrôleur au plus près de sa perception, c'est-à-dire en minimisant les sources de variation de ce jugement (charge de travail contextuelle nulle, pas de résolution associée, etc.).

En dépit de ces constats, nous avons fait l'hypothèse qu'une forme d'accord entre contrôleurs doit exister concernant le jugement de conflit. L'importante dimension collective de ce travail (découpage de l'espace et découpage du vol dans le temps, entre contrôleurs) nous semble l'exiger. Il convient également de se rappeler que la formation pratique du contrôleur suppose en premier lieu d'apprendre à distinguer ce qui doit être considéré comme conflit de ce qui ne doit pas l'être. Une modélisation de ce « noyau commun » de jugement sera donc étudiée à partir des résultats expérimentaux.

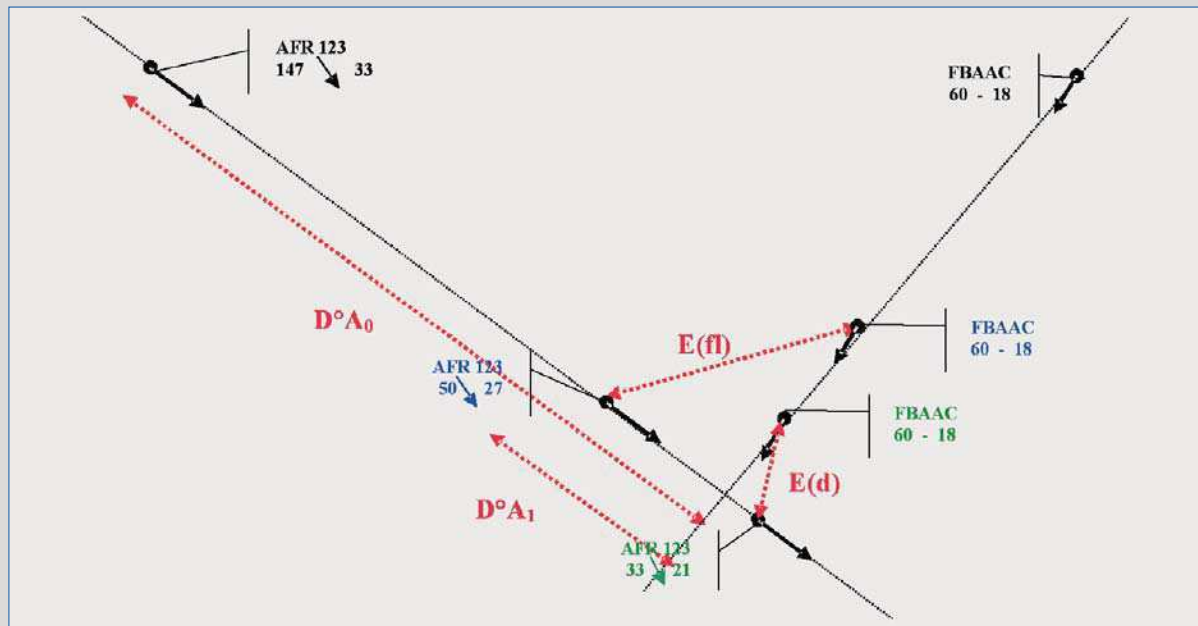
3 ASPECTS EXPÉRIMENTAUX

En accord avec ce qui précède, une expérimentation (projet CREED : Conflict Risk Evaluation based on Expert Detection) a été conçue pour évaluer cette capacité de détection réelle des contrôleurs. Elle devait permettre que s'exprime leur jugement sur un ensemble le plus large possible de conflits, afin d'identifier les valeurs des variables (ou sources de variations) qui y sont associées : quand est ce que « ça passe », et quand est ce que « ça ne passe pas », du seul point de vue du contrôleur ? L'aspect statistique de l'étude a été pris en compte dans l'élaboration du plan expérimental, ce qui a permis un traitement pertinent des données recueillies.

3.1 La construction des scénarios de trafic

Des séquences de trafic réalistes (nommées scénarii) ont été créées, à partir desquelles des sujets qualifiés se sont prononcés en termes de conflit ou de non conflit. Le travail de contrôle en-route présente des différences avec celui de contrôle d'approche : nous avons retenu comme condition expérimentale le contrôle d'approche, dont le trafic est fortement évolutif. Avec les moyens radar dont disposent les centres d'approche, les valeurs des séparations minimales exigées pour garantir la sécurité entre deux aéronefs sont de 3 NM pour une séparation sur le plan horizontal, et de 1000 ft pour une séparation sur le plan vertical.

Dans les scénarios figurent deux avions et deux seulement, dont les trajectoires se croisent. L'absence de tout autre trafic ou de toute tâche concurrente permet au sujet d'exprimer un jugement dans des conditions optimales, c'est-à-dire avec a



La figure ci-dessus met en scène deux aéronefs (AFR123 et FBAAC) dont les trajectoires – en noir – se coupent, et dont on peut lire certains paramètres (altitude et vitesse sol) sur « l'étiquette ». Ces aéronefs sont représentés à trois moments successifs sur leurs trajectoires, dans l'ordre chronologique par les couleurs noir, puis bleu, puis vert. Du fait de l'évolution verticale de l'un des deux appareils (AFR123 est en descente) le conflit – s'il existe – peut être résolu soit sur le plan horizontal, soit sur le plan vertical (c'est ce dernier cas ici). Nos trois variables explicatives sont donc :

- > E(d) (parfois appelé CPA, pour Closest Point of Approach), qui est la plus courte distance horizontale qui existe entre ces deux avions lorsqu'on déroule la totalité des deux trajectoires ;
- > E(fl), qui est la valeur de la séparation entre les deux aéronefs du fait de leur évolution sur le plan vertical.

Pour représenter la séparation verticale, nous n'avons pas choisi de retenir (comme c'est habituellement le cas) la valeur de la différence d'altitude lorsque les deux aéronefs sont distants de E(d). Au lieu de cela, la variable E(fl) figure la distance, sur un plan horizontal, entre les deux avions lorsque l'écart entre leurs altitudes respectives est de 1000 pieds, croissant. En d'autres termes, cette variable représente la distance entre les plots radar des deux aéronefs au moment où disparaît le risque de perte de séparation. Ce faisant, nous nous sommes rapprochés — et c'était bien l'objectif — de la représentation mentale construite par le contrôleur, et de sa prise en compte de la dimension verticale à partir d'une image radar standard (2D). A noter que l'empan $D^{\circ}A$ est également exprimé en NM.

- > DA (pour degré d'anticipation) qui est l'empan temporel existant entre le diagnostic et le survol du point d'intersection par (conventionnellement) le plus rapide des deux aéronefs.

The figure above shows two aircraft (AFR123 and FBAAC) whose trajectories – in black – intersect, and of which certain parameters can be read (altitude and ground speed) on the "label". These aircraft are represented at three successive moments on their trajectories, in chronological order in black, then blue and then green. Owing to the vertical manoeuvre of one of the two aircraft (AFR123 is descending) the conflict – if one exists – can be resolved either on the horizontal plane, or on the vertical plane (as is the case here). Thus our three explanatory variables are:

- > E(d) (also known as CPA, for Closest Point of Approach), which is the shortest horizontal distance between these two aircraft when both trajectories unfold completely;
- > E(fl), which is the separation value between the two aircraft due to their manoeuvring on the vertical plane.

To represent the vertical separation, we have opted not to retain the value of the difference in altitude (as is generally the case) when the two aircraft are separated by E(d). Instead of this, the variable E(fl) is the distance, on the horizontal plane, between the two aircraft when the spacing between their respective altitudes is 1000 feet, and increasing. In other words, this variable represents the distance between the radar plots of the two aircraft at the time when the risk of loss of separation disappears. In so doing, we come closer to the mental representation created by the controller and his processing of the vertical dimension from a standard radar image (2D) - as was indeed the objective. It must be noted that the span $D^{\circ}A$ is also expressed in NM.

- > DA (for degree of anticipation), which is the time span between the diagnosis and the overflying of the point of intersection by (conventionally) the faster of the two aircraft.

assembly of disjointed elements. Even though it is largely automatised and often extremely fast (< 200 ms), as we have seen, taking in visual information is a cognitive process in its own right. It is a sorting of the available information, made according to the context and the previous expectations. It is therefore an interactive process, which does not lend itself well to a simple breakdown.

In the experiment described below, we tried to control to a maximum the variables in the environment (other than the explicative variables) in order to reduce their effects as far as possible. At the same time we tried to reproduce an identical global working context for all the subjects, and as close as possible to that of the control room.

To sum up, a simple model of the conflict detection process does not exist. The use of calculations by the experts is only marginally observed, except perhaps in the case of an unknown context. The process is therefore essentially carried out from visual patterns, and presents a variability that has no negative effects on safety, but makes the modelisation of the know-how more difficult. Judgements vary according to the problem and the controller, to a point that only allows very global and qualitative rules to be drawn up, at first sight.

Therefore, whether the

controller perceives the probable existence of separation minima or not, it was observed that he may decide to act to increase his margin. This increase is highly variable according to the controller, the context, the time the data was gathered, etc. The "noise" induced is therefore significant, which is very disadvantageous for concrete measurement of the process parameters. One of the limitations we fixed in setting up our experimental paradigm was to gather the controller's judgement as close as possible to his perception, i.e. by minimising the sources of variation of this judgement (contextual workload nil, no associated resolution, etc).

In spite of these observations, we made the hypothesis that some form of agreement between controllers must exist concerning the judgement of conflict. The significant collective dimension of this work (dividing up the airspace and cutting up the flight over time, between controllers) seems to require it. It should also be recalled that the controller's practical training involves first of all learning to distinguish between what should be considered as a conflict, and what should not be. A modelisation of this "common core" of judgement will therefore be studied from the experimental results.

priori une absence de perturbation des processus cognitifs impliqués dans la tâche expérimentale. Les jugements exprimés dans ces conditions reflètent donc un maximum, du point de vue de la performance théorique¹, en matière de détection de conflit.

Les deux aéronefs présentés se déplacent sur l'écran, pendant environ 1 minute et 10 secondes. Cette durée a été fixée à la suite d'expérimentations préalables, avec l'objectif de laisser un temps à coup sûr suffisant pour que les sujets produisent un jugement qui les satisfasse.

Les trajectoires de ces deux aéronefs provenaient d'enregistrements radar (SNER) retouchés de façon à présenter les caractéristiques voulues. Disposer de trajectoires réelles permettait de fournir aux sujets des aéronefs dont les vitesses présentaient l'incertitude existant dans la « vraie vie ». L'intérêt était ici de dissuader les contrôleurs de toute extrapolation trop artificiellement précise (en particulier grâce au vecteur vitesse). Ils étaient ainsi amenés à conserver, dans le cadre de la détection, la gestion d'une d'incertitude identique à celle de leur activité réelle. Dans le même esprit, et grâce à la coopération des divisions techniques locales, les visualisations IRMA, les cartes radar et les outils étaient strictement les mêmes que ceux utilisés en salle.

Nous avons donc retenu trois variables, dites explicatives, pour réaliser ces scénarii (cf. encart n° 2). Elles sont susceptibles d'expliquer les variations de jugement exprimées, conformément au modèle et aux hypothèses sous-jacentes. Les deux premières variables proviennent de la définition même du conflit ; la troisième est tenue pour être sa source de variation extrinsèque la plus importante :

- > **Ed** : séparation minimale sur le plan horizontal. Nous avons retenu cinq niveaux pour la construction de nos scénarios de trafic : 0, 3, 6, 9, et 12 NM ;
- > **Efl** : séparation produite par l'évolution sur le plan vertical (et non "séparation verticale", cf. encart 2). Nous avons retenu également cinq niveaux pour cette variable : 0, 5, 10, 15 et 20 NM ;
- > **D°A** : empan temporel de prédiction. Il y a six niveaux pour cette variable : 6, 10, 15, 21, 28 et 36 NM.

Les scénarios de trafic proposés aux contrôleurs ont donc été créés selon la combinatoire générée par ces niveaux. Pour parvenir à ce que les trajectoires 4D se coupent concrètement selon les valeurs déterminées de Ed et Efl, les outils de prototypage rapide de la SDER (Rejeu, Twinkle, Digipilot) ont été utilisés. Les trajectoires ont pu être modifiées à volonté et très simplement, et pouvaient être ainsi « rejouées » sur IRMA.

Chaque scénario figurait un avion en descente et un avion stable (variable attitude), dont les trajectoires se coupaient à angle droit (variable géométrie). Les vitesses sol des deux aéronefs étaient toujours proches l'une de l'autre, dans un rapport compris entre 1 et 1,5 (variable vitesse). Étant fixées, ces variables étaient « contrôlées », c'est-à-dire qu'elles ne pouvaient plus constituer une source de variation des jugements.

(1) Et non du point de vue des modes opératoires réels. Il convient par exemple de remarquer qu'il n'est ici question ni de l'intervalle de temps pendant lequel il est souhaitable que les contrôleurs puissent détecter le conflit, ni de celui au cours lequel ils le détectent en réalité.

3.2 L'expérimentation et les résultats

L'expérimentation s'est déroulée sur quatre sites : Lyon St-Exupéry, Toulouse Blagnac, Marseille Provence et Bordeaux Mérignac. Sur l'ensemble de ces sites, 161 contrôleurs ont participé (75% d'hommes et 25% de femmes). Tous étaient qualifiés depuis trois années au minimum, ce qui excluait une distinction experts/novices.

Les sessions duraient environ 1 heure par sujet, ce qui correspond à la combinatoire des variables. Généralement, les sujets n'utilisaient pas la totalité du temps (1' 10" par scénario) et répondaient souvent au bout de 25" à 35".

Lors des consignes précédant l'expérience, il était bien insisté sur le fait que l'objectif était de n'exprimer « que ce que ses yeux percevaient », en se référant uniquement aux normes de séparation. Le but, déjà indiqué plus haut, était de minimiser l'ajout systématique et variable des marges liées à la mise en place d'une résolution. Le couplage « naturel » des deux processus, de détection et de résolution, a pour effet de bruyé considérablement les réponses, et de produire des jugements qui dissimulent le modèle de processus de détection proprement dit. Pour y parvenir, il était suggéré en début de tâche au contrôleur de se projeter non pas dans la peau du « contrôleur contrôlant », mais plutôt dans celle d'un testeur, ou bien d'un contrôleur en position miroir. L'intérêt était de proposer un comportement pertinent, naturel (c.-à-d. qui existe dans le métier réel), et qui invite le contrôleur à se soucier de l'existence (ou non) de la norme de séparation, mais sans avoir la capacité d'agir, de résoudre.



Figure 1 : L'interface de recueil
Figure 1: Data collection interface

Les jugements des contrôleurs ont été recueillis sur un écran tactile, avec une interface conçue pour l'occasion (cf. Figure 1). La moitié inférieure de cet écran comporte en particulier les huit touches (du vert au rouge) à utiliser pour exprimer le jugement, les résumés des plans de vol, et les touches de validation. Les huit touches sont réparties en deux sous-groupes de quatre : à gauche les quatre utilisables lorsque le jugement est celui de « conflit », à droite les quatre touches signifiant au contraire le « non conflit ». Les touches extrêmes correspondent à une

(1) And not from the point of view of real operational modes. It should be noted, for example, that there is no question here of a time period during which it is desirable for the controllers to detect the conflict, nor one during which they actually detect it.

3 EXPERIMENTAL ASPECTS

In keeping with the above, a set of experiments (CREED project: Conflict Risk Evaluation based on Expert Detection) was designed to evaluate the controllers' real detection capacity. It was to allow them to express their opinions on the widest possible range of conflicts, in order to identify the values of the variables (or sources of variations) associated with them: when does it "make it" and when does it "not make it", simply from the controller's point of view? The statistical aspect of the study was taken into consideration when drawing up the experimental plan, allowing the data gathered to be processed appropriately.

3.1. Construction of traffic scenarii

Realistic traffic sequences (named scenarii) were created, from which qualified subjects gave their opinions in terms of conflict or no conflict. The work of en-route control showed up differences from approach control: for the experimental conditions we chose approach control, which has a great deal of climbing or descending traffic. With the radar systems available in the approach centres, the separation minima required to ensure safety between two aircraft are 3 NM for horizontal separation, and 1000 ft for vertical separation.

In the scenarii there were two aircraft, and only two, whose trajectories crossed. The absence of any other traffic, or any concurrent task, allowed the subject to express a judgement in optimum conditions, i.e. with an a priori absence of disturbance of the cognitive processes involved in the experimental task. The judgements expressed in these conditions therefore reflect a maximum in the area of conflict detection, from the point of view of theoretical performance¹. The two aircraft presented moved on the screen for about one minute and ten seconds. This time period was fixed after previous experiments, with the aim of being sure to leave enough time for the subjects to produce a judgement with which they were satisfied. The trajectories of these two aircraft came from radar recordings (SNER), acquired in order to present the required characteristics. Using real trajectories allowed the subjects to be provided with aircraft whose speeds presented uncertainties existing in "real life". The interest here was to dissuade the controllers from making any too artificially precise extrapolations (in particular due to the speed vector). Thus, in the context of the detection, they were led to maintain the management of exactly the same degree of uncertainty as that of their real activity. In the

same spirit, and thanks to the cooperation of the local technical divisions, the IRMA displays, radar maps and tools were strictly the same as those used in the control room. We therefore chose three variables, known as explicative, to carry out these scenarii (see Box N° 2). They may well explain the variations in the expressed judgements, in accordance with the model and the underlying hypotheses. The first two variables come from the definition of the conflict itself; the third is considered to be its most significant extrinsic variation:

- > **Ed**: horizontal separation minima. We chose five levels for the construction of our traffic scenarii: 0, 3, 6, 9, and 12 NM;
- > **Efl**: separation produced by vertical evolution (and not "vertical separation", see Box N° 2). We also selected five levels for this variable: 0, 5, 10, 15 and 20 NM;
- > **D°A**: predicted time span. There are six levels for this variable: 6, 10, 15, 21, 28 and 36 NM.

The traffic scenarii proposed to the controllers were therefore created according to the combinations generated by these levels. In order to have the 4D trajectories intersect concretely according to the determined values of Ed and Efl, the SDER fast prototyping tools (Rejeu, Twinkle and Digipilot) were used. The trajectories could be modified very simply

upon request, and could thus be "played back" on IRMA. In each scenario there was a descending aircraft and a stable aircraft (attitude variable), whose trajectories intersected at right angles (geometry variable). The groundspeeds of the two aircraft were always close to each other, with a ratio of between 1 and 1.5 (speed variable). As they were fixed, these variables were "controlled", meaning that they could no longer become a source of variation of judgement.

3.2. Experiments and results

The set of experiments was carried out at four sites: Lyon St-Exupéry, Toulouse Blagnac, Marseille Provence and Bordeaux Mérignac. Over all these sites, 161 controllers participated (75% men and 25% women). All of them had been qualified for at least three years, which

confiance maximale des sujets dans le jugement exprimé, soit vert foncé dans le cas de « non conflit », rouge foncé si le diagnostic est « conflit ». Les touches du milieu (tonalité jaune) indiquent au contraire un doute important concernant le diagnostic, même si un choix a bien été effectué. En résumé, il s'agit donc d'une échelle unidimensionnelle, exprimant la probabilité d'une perte de séparation, telle que perçue par le contrôleur.

4 RÉSULTATS

Les résultats obtenus peuvent être décrits globalement, avant la réalisation du modèle statistique. Cette description se fait principalement à deux niveaux : les caractéristiques centrales et de dispersion des jugements des contrôleurs.

4.1 Les caractéristiques centrales

Nous utilisons comme caractéristiques centrales les médianes en raison du caractère ordonné des jugements : c'est le jugement « moyen », c'est-à-dire l'expression de l'individu le plus représentatif de l'échantillon sur la question posée (cf. encart 3 pour la définition précise de la médiane). Ces médianes sont représentées sur la figure ci-dessous selon le code des huit couleurs précédentes (rappelées à droite sur la figure). La couleur grisée correspond aux combinaisons des variables qui n'ont pas de sens d'un point de vue géométrique.

Chacun des six grands carrés composites ci-dessous renferme les jugements des contrôleurs pour un empan temporel donné. Cet empan est exprimé ici en milles nautiques. Avec des vitesses standard d'aéronefs en approche, l'ordre de grandeur de l'intervalle de temps va de 10 minutes (pour DA=36NM) à 1minute 30 secondes (DA=6NM) environ. Pour chacun, la variable E(d) est en abscisse, et E(fl) en ordonnée.

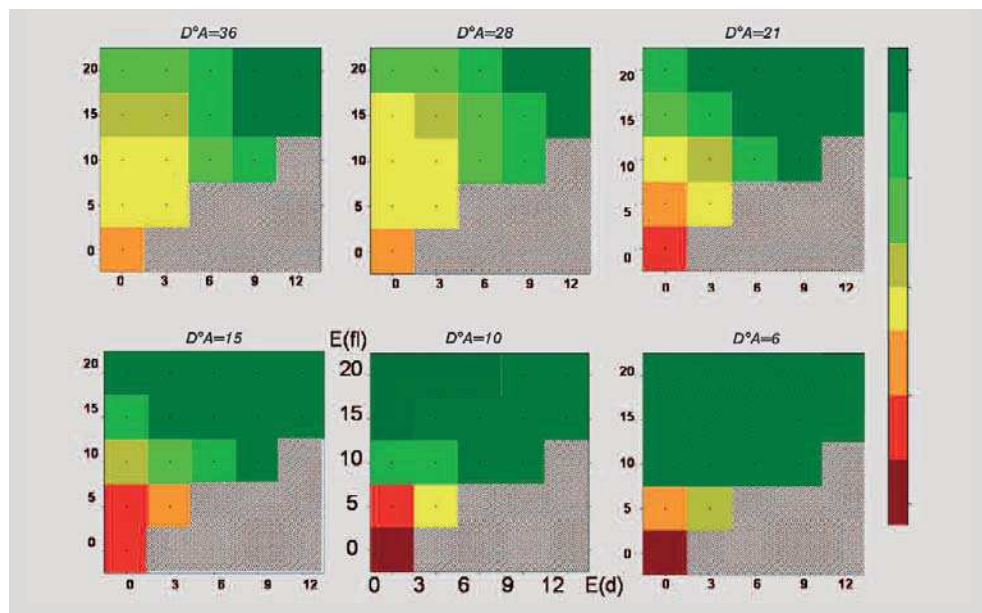


Figure 2 : valeurs médianes des jugements
Figure 2: Median values of the judgements

Petite introduction aux méthodes graphiques

Lorsqu'on dispose d'une série d'observations, il est souvent indispensable de la résumer par des indicateurs typiques dont le plus connu est la moyenne arithmétique. Il est cependant toujours insuffisant de résumer une série par un seul indicateur. Deux grandes classes d'indicateurs sont utiles : les caractéristiques de tendance centrale et les caractéristiques de dispersion.

Caractéristiques de tendance centrale

On considère un échantillon de n valeurs x_1, \dots, x_n d'une variable aléatoire. Si les observations sont rangées par ordre croissant, $x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$, la médiane est la valeur $M = x_{(n+1)/2}$ pour n impair. Si n est pair, on prendra

$$M = \frac{x_{(n)/2} + x_{(n+1)/2}}{2}$$

Ainsi, la médiane sera la valeur qui sépare l'échantillon en deux parties de taille égale, l'une correspondant aux valeurs inférieures à M et l'autre aux valeurs supérieures à M . Pour n impair, la médiane sera une valeur observée tandis que pour n pair, elle sera au milieu de deux valeurs observées. Remarquons que pour obtenir la médiane, il suffit de pouvoir ordonner les valeurs ; celles-ci n'étant pas nécessairement numériques.

La moyenne arithmétique est par définition

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

ce qui suppose les valeurs x_i numériques. Bien que fonction de toutes les observations, la moyenne est très sensible aux valeurs extrêmes, à la différence de la médiane. Pour une distribution parfaitement symétrique on a égalité entre la médiane et la moyenne.

Caractéristiques de dispersion

La dispersion est une caractéristique de la variabilité des valeurs observées. La mesure la plus évidente de cette variabilité est l'étendue ou intervalle de variation définie comme

$$W = |x_{\max} - x_{\min}|$$

Dépendant des valeurs extrêmes observées, c'est un indicateur instable. Pour pallier à cet inconvénient, on définit l'écart interquartile (EIQ). Nous avons vu que la médiane divise les données en deux ensembles de taille égale :

- > le quartile inférieur Q_1 , ou premier quartile, sera la valeur milieu du premier ensemble, dans lequel 25 % des valeurs sont inférieures à Q_1 et 75% lui sont supérieures ;
- > le quartile supérieur Q_3 , ou troisième quartile, sera la valeur milieu du deuxième ensemble, dans lequel 75 % des valeurs sont inférieures à Q_3 et 25% lui sont supérieures.

Le deuxième quartile n'est autre que la médiane. Le tableau donne un exemple pour un échantillon de taille paire.

Short introduction to graphical methods

When we have a set of observations, it is often indispensable to summarize it by typical indicators, of which the arithmetic mean is the most common. However it is insufficient to summarize a series by a single indicator. Two large classes of indicators are necessary: the characteristics of central tendency and characteristics of dispersion.

Characteristics of central tendency

We consider a sample of n values x_1, \dots, x_n of a random variable. If the observations are arranged in ascending order, $x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$, the median is the value $M = x_{(n+1)/2}$ for odd n . If n is even, we will consider

$$M = \frac{x_{(n)/2} + x_{(n+1)/2}}{2}$$

Thus, the median will be the value that divides the sample into two equal parts, one corresponding to the values less than M and the other to values greater than M . For odd n , the median will be an observed value, whereas for even n , it will be in the middle of two observed values. Note that to obtain the median, we simply need to be able to order the values; which need not necessarily be numerical.

By definition, the arithmetic mean is

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

which supposes numerical values x_i . Although it is a function of all observations, the mean is very sensitive to extreme values, unlike the median. For a perfectly symmetrical distribution, the median is equal to the mean.

Characteristics of dispersion

The dispersion is a characteristic of the variability of observed values. The most obvious measure of this variability is the range defined as

$$W = |x_{\max} - x_{\min}|$$

Depending on the extreme values observed, this is an unstable indicator. To compensate for this disadvantage, we define the interquartile range (IQR). We have seen that the median divides the data into two sets of equal size:

- > The lower quartile Q_1 , or first quartile, will be the middle value of the first set, in which 25% of the values are less than Q_1 and 75% greater
- > The upper quartile Q_3 , or third quartile, will be the middle value of the second set, in which 75% of the values are less than Q_3 and 25% greater.

The second quartile is the median itself. The table gives an example of a sample of even size.

Données Data	8,13,27,32,25,16,32,27,8,28,79,25,35,25,38,29,80,50,38,30,20,20,49,9
Données ordonnées Ordered data	8,8,9,13,16,20,20,25,25,25,27,27,28,29,30,32,32,35,38,38,49,50,79,80
Médiane Median	27,5
Quartile supérieur Upper quartile	36,5
Quartile inférieur Lower quartile	20

L'écart interquartile est la différence $Q_3 - Q_1$, il est plus robuste que l'étendue. Pour terminer, citons la bien connue variance σ^2 définie par

$$\sigma^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$$

Diagramme en boîte

Il est possible de représenter graphiquement les principales caractéristiques d'une variable aléatoire en utilisant les quartiles. On obtient alors le diagramme en boîte (box-plot en anglais). Il est constitué d'une boîte de largeur arbitraire et dont la longueur correspond à l'écart interquartile. On trace à l'intérieur la position de la médiane. La boîte est alors complétée par des « moustaches » correspondant aux valeurs adjacentes suivantes :

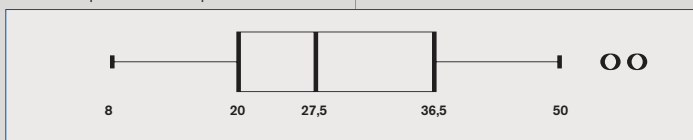
- > adjacente supérieure : plus grande valeur inférieure à $Q_3 + 1,5(Q_3 - Q_1)$;
- > adjacente inférieure : plus petite valeur supérieure à $Q_1 - 1,5(Q_3 - Q_1)$.

Les valeurs dites « extérieures » représentées par des O sont celles qui sortent des « moustaches ». Reprenons l'exemple précédent, nous obtenons $Q_3 - Q_1 = 16,5$ et

- > la plus grande valeur inférieure à $Q_3 + 1,5(Q_3 - Q_1)$ est 50 ;
- > la plus petite valeur supérieure adjacente à $Q_1 - 1,5(Q_3 - Q_1)$ est 8.

Il y a donc deux valeurs extérieures supérieures à 50 et aucune inférieure à 8. La Figure 11 fait alors apparaître deux observations déviantes

Un tel graphique permet en particulier de comparer plusieurs distributions entre elles : comparer les diagrammes en boîte est plus aisée que comparer des histogrammes. A titre d'exemple, la Figure 12 illustre la comparaison entre les réponses des contrôleurs pour $E_d = E_f = 0$ pour les différentes valeurs de DA.



The interquartile range is the difference $Q_3 - Q_1$, it is more robust than the range. To conclude, let us state the familiar variance σ^2 defined by

$$\sigma^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$$

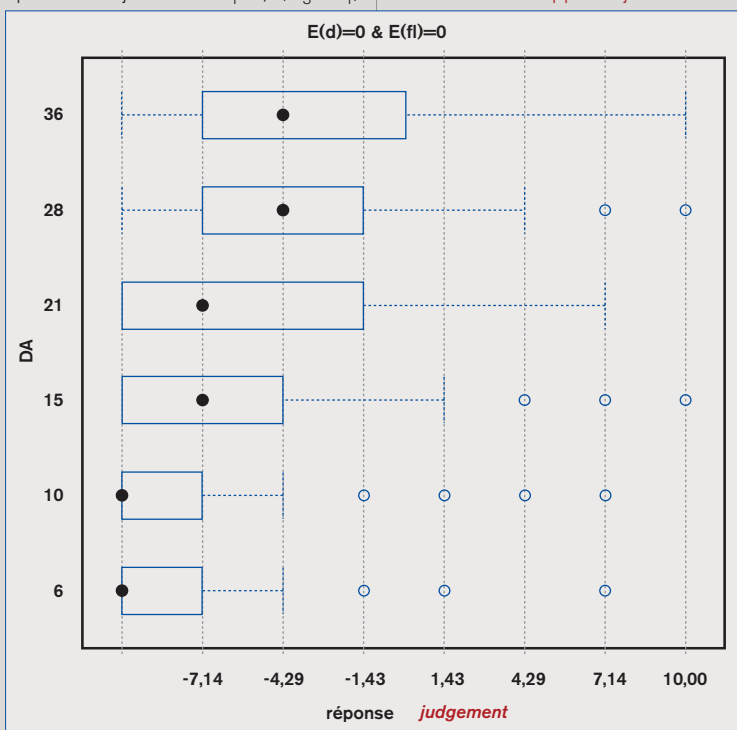
Box plot

It is possible to graphically represent the main characteristics of a random variable by using quartiles. We then obtain the box-plot. This consists of a box of arbitrary width and a length corresponding to the interquartile range. Within it, we plot the position of the median. The box is then completed by "whiskers" corresponding to the following adjacent values:

- > Upper adjacent: highest value less than $Q_3 + 1.5(Q_3 - Q_1)$
- > Lower adjacent: lowest value greater than $Q_1 - 1.5(Q_3 - Q_1)$.

The values known as "external" represented by Os are those that result from the "whiskers". Let us take up the previous example, we obtain $Q_3 - Q_1 = 16.5$ and

- > The largest value less than $Q_3 + 1.5(Q_3 - Q_1)$ is 50
- > The smallest upper adjacent value to $Q_1 - 1.5(Q_3 - Q_1)$ is 8.



Thus, there are two external values greater than 50 and none lower than 8. Figure 11 reveals two deviating observations

Such a graph is used, in particular, to compare several distributions: comparing box plots is easier than comparing histograms. As an example, Figure 12 shows the comparison between the replies of controllers for $E_d = E_f = 0$ for different values of DA.

Figure 11 : Diagramme en boîte de l'exemple
Figure 11 : Box plot of the example.

Figure 12 : Comparaison des diagrammes boîtes pour différentes valeurs de DA
Figure 12 : Comparison of box plots for different values of DA.

Diagramme quantile-quantile

Pour comparer la distribution empirique d'un échantillon à une distribution de référence, par exemple une gaussienne, nous utilisons un graphique basé sur la comparaison des quantiles de chacune des distributions. Nous avons introduit auparavant les premier et troisième quantiles ; la médiane pour sa part étant le second quantile. Cette notion se généralise de la façon suivante.

Soit (x_1, \dots, x_n) un échantillon et $(x_{(1)}, x_{(2)}, \dots, x_{(n)})$ l'échantillon ordonné, on appelle quantile empirique un nombre \hat{q}_p tel que la proportion des n observations inférieures à \hat{q}_p soit au moins égale à p et que la proportion des n observations supérieures à \hat{q}_p soit au moins égale à $1-p$. Ainsi

$$\hat{q}_p = \begin{cases} x_{(i+1)} & \text{si } i < np < i+1 \\ \frac{x_{(i)} + x_{(i+1)}}{2} & \text{si } i = np \end{cases}$$

Rappelons que la fonction de répartition empirique est la fonction $F_n(x)$ qui vaut i/n pour tout x dans l'intervalle $]x_{(i)}, x_{(i+1)}]$; plus généralement $F(x)$ calcule la proportion de valeurs x_i inférieures à x . On remarque alors que la fonction quantile est la fonction inverse de la fonction de répartition, puisque $F(\hat{q}_p) = p$, si np est une valeur entière et i/n si $i < np < i+1$.

Pour comparer la fonction empirique de notre échantillon à une fonction de répartition de référence, il est plus judicieux de comparer les quantiles, en raison de la difficulté de comparer deux histogrammes. Il suffit donc de porter sur un graphique les points (q_p, \hat{q}_p) , où q_p est le quantile de la fonction de répartition de référence. Les deux distributions seront d'autant plus proches que les points ainsi tracés seront alignés. La Figure 7 fournie un exemple où la fonction de répartition de référence est la loi gaussienne standard.

Quantile-quantile diagram

To compare the empirical distribution of a sample with a reference distribution, Gaussian for example, we use a graph based on the comparison of the quantiles for each of the distributions. Earlier we have introduced the first and third quantiles; the median being the second quantile. This concept is standardized as follows.

Let (x_1, \dots, x_n) be a sample and $(x_{(1)}, x_{(2)}, \dots, x_{(n)})$ the ordered sample, the empirical quantile is a number \hat{q}_p such that the proportion of n observations less than \hat{q}_p is at least equal to p and that the proportion of n observations greater than \hat{q}_p is at least equal to $1-p$. Thus

$$\hat{q}_p = \begin{cases} x_{(i+1)} & \text{si } i < np < i+1 \\ \frac{x_{(i)} + x_{(i+1)}}{2} & \text{si } i = np \end{cases}$$

We must remember that the empirical distribution function is the function $F_n(x)$ which is equal to i/n for all values of x within the interval $]x_{(i)}, x_{(i+1)}]$; more generally $F(x)$ calculates the proportion of values x_i less than x . We thus observe that the quantile function is the inverse of the distribution function since $F(\hat{q}_p) = p$, if np is an integer value and i/n if $i < np < i+1$.

When comparing the empirical function of our sample with a reference distribution function, it is wiser to compare the quantiles, owing to the difficulty in comparing two histograms. It is therefore sufficient to plot the points (q_p, \hat{q}_p) on a graph, where q_p is the quantile of the reference distribution function. The two distributions will be that much closer as the points thus plotted are aligned. Figure 7 is an example where the reference distribution function is the standard Gaussian law.

Chaque carré composite est formé d'une mosaïque de carrés élémentaires d'une seule couleur. Ces carrés élémentaires correspondent aux médianes des jugements des 161 contrôleurs, sur les différents scénarios de trafic.

Il apparaît à la lumière de ces résultats qu'une forte cohérence existe dans l'expression des jugements. Cette cohérence s'observe d'abord à travers l'influence des dimensions horizontale et verticale (variables $E(d)$ et $E(fl)$) sur le jugement. Dans chaque grand carré composite, on voit que l'accroissement (resp. la diminution) de la valeur de l'une ou de l'autre de ces variables conduit à une variation systématique du jugement, dans le sens d'une certitude plus grande de non conflit (resp. de conflit). Cette cohérence se confirme avec l'introduction de la troisième variable explicative $D^{\circ}A$, qui permet la comparaison des grands carrés composites successifs. On observe que pour des conflits jugés avec une anticipation importante (le carré $D^{\circ}A=36$), les sujets reconnaissent douter et émettent des jugements ambivalents. Le doute est néanmoins très bien différencié selon que la séparation finale sera un peu trop juste (carrés jaune) ou

excluded a distinction between experts and novices. The sessions lasted approximately one hour per subject, which corresponds to the combination of variables. Generally the subjects did not use up the whole time (1' 10" per scenario) and often replied after 25" or 35". During the instructions given before the experiment started, it was stressed that the aim was to express simply "what your eyes would perceive", referring only to the separation minima.

The aim, as indicated above, was to minimise the systematic and variable addition of margins linked to setting up a resolution. The effect of the "natural" coupling of the two processes – detection and resolution – was to produce "noise" in the replies, and to produce judgements that dissimulated the detection process model itself. In order to achieve this, it was suggested to the controller, early in the task, that he did not project himself into the role of the "controlling

controller", but rather into that of a tester, or a controller in a mirror position. The interest was to propose pertinent, natural behaviour (i.e. which exists in the real profession), and which invited the controller to be concerned about the existence (or otherwise) of the separation minima, but without being able to act on it, or to resolve it. The controllers' judgements were recorded using a touch screen, with an interface that had been specially designed for this (see Figure 1). On the lower half of this screen, there are notably the eight buttons (from green to red) to be used to express judgement, the flight plan summaries, and the validation buttons. The eight buttons are divided into two sub-groups of four: on the left are the four to be used when the judgement is that of a "conflict", and on the right the four buttons mean, on the contrary, "no conflict". The buttons at the far right and left correspond to the subject's maximum level of confidence in the judgement expressed, i.e. dark green in the case of "no conflict", and dark red if the diagnosis is "conflict". The middle buttons (shades of yellow) indicate, on the contrary, a significant doubt concerning the diagnosis, even if a choice was in fact made. To summarise, this is a one-dimensional scale expressing the probability of a loss of separation, as perceived by the controller.

4 RESULTS

The results obtained can be globally described, before making the statistical model. This description is made mainly on two levels: the central characteristics and the dispersion of the controllers' judgements.

4.1. Central characteristics

We use the medians as central characteristics because of the ordered character of the judgements: this is the "mean" judgement, i.e. the expression of the most representative individual in the sample to the question asked (see Box 3 for the precise definition of the median). These medians are represented on the figure below using the previous eight-colour code (recalled on the right of the image). The shaded area corresponds to combinations of variables that have no meaning from a geometrical point of view.

Each of the six composite squares below encloses the controllers' judgements for a given time span. This time span is expressed here in nautical miles. With standard speeds of approaching aircraft, the magnitude of the time interval goes from 10 minutes (for DA=36NM) to 1 minute 30 seconds (DA=6NM) approximately. For each, variable E(d) is on the abscissa, and E(fl) on the ordinate. Each composite square is formed by a mosaic of elementary squares of

assez importante (carrés vert clair). Cette tendance se confirme d'un D°A sur l'autre, le doute tendant progressivement à disparaître lorsque la valeur de D°A diminue. Cela s'observe graphiquement tant par l'augmentation du nombre de carrés verts (qui en outre deviennent progressivement plus foncés) à partir du coin supérieur droit que par la restriction du nombre de carrés rouge au seul carré inférieur gauche (en outre de plus en plus foncé), au fur et à mesure que les deux aéronefs se rapprochent du point d'intersection. Ce résultat est assez remarquable lorsque l'on connaît d'une part l'incertitude qui entache la prédiction de la trajectoire, et d'autre part la variabilité individuelle qu'autorise le traitement des informations par heuristiques visuelles. Tout se passe donc comme si la « communauté des contrôleurs » (si on admet qu'il s'agit de ce qui est représenté par la valeur médiane des jugements) était extrêmement clairvoyante et cohérente concernant les positions futures des avions considérés.

Pour autant, nombre de ces jugements traduisent une incertitude sur l'évolution de la situation (lorsque les réponses médianes correspondent aux couleurs intermédiaires de l'interface de recueil). Ce résultat est en accord avec les travaux précédents sur le sujet, qui consacrent le maintien du doute comme stratégie chez le contrôleur expérimenté. Il convient de noter que deux scénarios persistent à donner des réponses de doute jusqu'à un stade très avancé du conflit ; ceux construits avec les valeurs E(d)=0NM, E(fl)=5NM, D°A=6NM, et E(d)=3NM, E(fl)=5NM, D°A=6NM. L'interprétation qui semble s'imposer est que ces configurations correspondent à des situations qui, bien que proches du dénouement, ne présentent qu'une trop faible marge par rapport à la norme de séparation. Ceci explique que, bien qu'il n'y ait pas de perte de séparation à terme pour ces configurations, les contrôleurs répugnent à les classer « vert », c'est-à-dire clairement non conflit.

4.2 Les caractéristiques de dispersion

La cohérence soulignée précédemment concerne les caractéristiques centrales des jugements. Pour autant, rien n'indique si celles-ci proviennent d'avis unanimes de la part de la communauté des sujets ou si, au contraire, elles résultent de jugements très contrastés. Les jugements étant ordonnés du « conflit certain » au « non conflit certain », il est possible d'attribuer une valeur numérique à chacune des huit classes de jugements ; cela revient à transformer l'échelle de valeurs en une échelle numérique, les valeurs n'ayant pas de signification intrinsèque. La caractéristique de dispersion retenue est l'écart interquartile, noté EIQ, qui dans notre cas revient à mesurer l'écart des réponses entre le 41^{ème} et le 121^{ème} sujet, ceux-ci ayant été ordonnés selon leur jugement (cf. Encart n° 3 pour une définition précise de l'EIQ).

Ainsi, si on attribue conventionnellement la valeur -10 à la réponse « conflit certain » (rouge foncé sur l'interface de recueil, cf figure 1 p44) et la valeur +10 à la réponse « non conflit certain » (vert foncé), l'amplitude de variation des réponses peut donc présenter un maximum de 20 unités. Pour

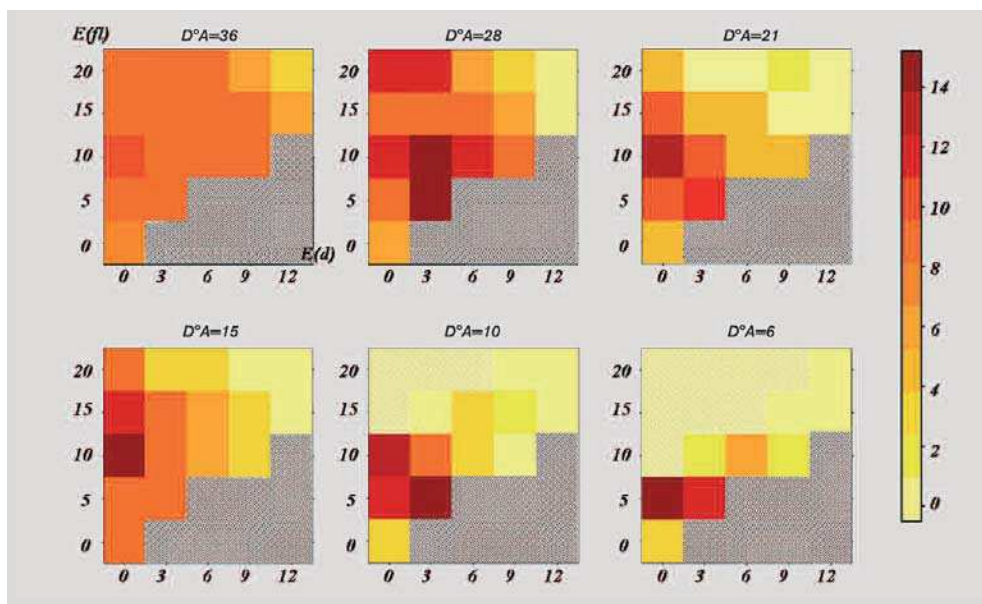


Figure 3 : variabilité des jugements
Figure 3: Variability of judgements

les jugements que nous avons recueillis, la dispersion des réponses est donc quelquefois considérable puisqu'on voit (cf. le code des couleurs, sur la droite de la figure) que l'EIQ peut dépasser 14 unités. Elle est par ailleurs loin d'être constante, et ses variations suivant $E(d)$, $E(f)$ et $D^{\circ}A$ rappellent les motifs observés sur les graphiques des médianes : des jugements « moyens » tranchés se traduisent généralement par des EIQ faibles. Ainsi, les scénarios de trafic présentant les EIQ les plus élevés correspondent aux combinaisons des trois variables explicatives pour lesquelles l'ambivalence ne parvient à être levée ni par l'intermédiaire des heuristiques individuelles, ni grâce aux outils opérationnels actuellement disponibles. L'analyse graphique des données recueillies dans le cadre de CREED a donc permis de vérifier l'importance du doute dans le processus de détection des conflits. Sans nier la variabilité individuelle des contrôleurs, elle a également permis d'identifier une cohérence globale dans l'expression de leurs jugements. Ce résultat très positif suggère que des outils d'aide au contrôle pourraient s'appuyer sur les sources de variations identifiées à ce stade de l'étude ($E(d)$, $E(f)$ et $D^{\circ}A$). De telles applications nécessiteraient néanmoins que l'on parvienne à estimer les jugements des contrôleurs pour toute situation de conflit potentiel. C'est l'objet des modèles statistiques développés au cours de la seconde phase du projet CREED.

5 MODELES STATISTIQUES

5.1 Modèle logistique ordonné

Nous abordons maintenant la partie relative au modèle statistique, qui devrait permettre de reproduire les jugements observés sur les scénarii d'origine et de les interpoler à toute situation de conflit potentiel.

En raison du caractère discret et ordonné des réponses proposées par les (et aux) contrôleurs, le formalisme retenu est celui du modèle logistique ordonné, présenté dans l'encart n° 4. En effet, le contrôleur aérien a jugé le risque de collision sur une échelle

one single colour. These elementary squares correspond to the medians of the 161 controllers' judgements on the various traffic scenarii. It appears in the light of these results that a strong coherency exists in the expression of the judgements. This coherency can be observed firstly through the influence of the horizontal and vertical dimensions (variables $E(d)$ and $E(f)$) on the judgement. In each large composite square, it can be seen that the increase (respectively the decrease) in the value of one or the other of these variables leads to a systematic variation in judgement, meaning a higher certainty of no conflict (respectively conflict). This coherency is confirmed with the introduction of the third explicative variable $D^{\circ}A$, which allows a comparison of the successive composite large squares. It can be

observed that for conflicts judged with a significant anticipation (square $D^{\circ}A=36$), the subjects recognise the doubt and give ambivalent judgements. Nevertheless, the doubt is well differentiated depending on whether the final separation will be a bit too close (yellow squares) or quite far (light green squares). This tendency is confirmed from one $D^{\circ}A$ to another, and the doubt tends to disappear progressively when the value of $D^{\circ}A$

decreases. This can be observed graphically, both for the increase in the number of green squares (which also become progressively darker), starting from the top right-hand corner, and for the restriction in the number of red squares to only the bottom left-hand square (which also becomes progressively darker), as the two aircraft get closer to the intersection point. This result is quite remarkable when we know, firstly, the uncertainty that marks trajectory forecasts, and secondly, the individual variability involved in processing data by visual heuristics. Everything happens as if the "controller community" (if we accept that this is what is represented by the median value of judgements) were extremely clairvoyant and coherent concerning the future position of the aircraft considered.

For all this, a number of judgements show an

uncertainty about the way the situation will develop (when the median replies correspond to the intermediate colours on the data collection interface). This result is in agreement with the previous works on the subject, which establish the maintenance of doubt as a strategy by the experienced controller. It should be noted that two scenarios persist in giving doubtful replies, even at a very advanced stage of the conflict: these are the ones constructed with the values E(d)=0NM, E(fl)=5NM, D°A=6NM, and E(d)=3NM, E(fl)=5NM, D°A=6NM. The interpretation that seems to stand out is that these configurations correspond to situations which, although they are close to the conclusion, only present too small a margin in relation to the separation minima. This explains the fact that, although there is no loss of separation eventually for these configurations, the controllers are loath to classify them as "green", meaning clearly no conflict.

4.2. Dispersion characteristics

The coherency previously underlined concerns the central characteristics of the judgements. All the same, nothing indicates if these come from the unanimous opinions from the community of subjects or if, on the contrary, they are the result of highly contrasting judgements. As the judgements are in order from "certain conflict" to "certain no conflict", it is possible to attribute a numerical value to each of the eight classes of judgements; this comes down to transforming the scale of values into a numerical scale, the values not having an intrinsic meaning. The dispersion characteristic used is the interquartile range, noted IQR, which, in our case, means measuring the range of replies between the 41st and the 121st subject, these having been ordered according to their judgement (see Box 3 for a precise definition of the IQR). Therefore, if we attribute conventionally the value -10 to the "certain conflict"

ordonnée à huit niveaux, depuis « airprox assuré » jusqu'à « c'est sans problème » en passant par « méfiance ». Le modèle logistique s'appuie sur une façon naturelle de représenter le processus de décision, consistant à définir une « variable latente » U continue, dont le niveau détermine la réponse finale du sujet. Si la valeur de U est inférieure à un certain seuil, noté α_1 , sa réponse sera « airprox assuré », si la valeur de U est supérieure à α_1 , mais inférieure à un autre seuil α_2 , la réponse sera celle immédiatement supérieure sur l'échelle des réponses possibles, et ainsi de suite. La variable latente U n'étant pas observable, elle doit donc être reconstruite. C'est à cette fin que les variables Ed, Efl et DA ont été définies. Ces variables, dites explicatives, sont censées expliquer au mieux la réponse du contrôleur. Cependant, en raison de sa nature particulière, la variable DA sera traitée différemment des autres. D'où la construction d'un modèle pour chaque valeur de DA : un modèle « global » sera obtenu ultérieurement par interpolation de ces modèles. D'autres sources de variations sont susceptibles d'affecter le jugement du contrôleur, mais ne pouvant être ou n'étant pas observés dans l'expérience, elles doivent néanmoins être incorporées dans U ; il faut ainsi décomposer U en une composante observée et une composante non observée. Usuellement, la composante non observée est considérée comme un bruit, donc comme une variable aléatoire, dont la loi détermine la probabilité des réponses. La composante observée est généralement modélisée par une combinaison linéaire (ou polynomiale) des variables explicatives ; pour une valeur de DA fixée elle sera ainsi formalisée par la relation linéaire

$$\beta'x = \beta_1 \times Ed + \beta_2 \times Efl + \beta_3 \times Ed \times Efl$$

où les coefficients inconnus β_1 , β_2 , et β_3 , sont appelés les régresseurs : notons la présence du terme Ed x Efl qui capture les influences croisées entre les deux variables Ed et Efl.

Chaque modèle est donc déterminé par dix coefficients qu'il convient d'estimer. La méthode retenue pour l'estimation est basée sur la maximisation d'une certaine fonction, appelée log-vraisemblance, qui dépend des données observées et des coefficients ; un algorithme d'optimisation est alors mis en œuvre. Les paramètres estimés sont présentés dans le Tableau 1

	DA=6	DA=10	DA=15	DA=21	DA=28	DA=36
α_1	1,506	0,862	-0,040	-0,704	-1,376	-1,793
α_2	2,399	1,898	1,124	0,525	-0,084	-0,237
α_3	2,840	2,316	1,668	1,078	0,589	0,413
α_4	3,181	2,653	2,155	1,564	1,239	1,157
α_5	3,733	3,185	2,751	2,215	1,850	1,804
α_6	4,050	3,603	3,255	2,826	2,346	2,376
α_7	4,747	4,299	4,143	3,595	3,137	3,178
β_1	0,232 **	0,405 ***	0,315 ***	0,281 ***	0,188 ***	0,231 ***
β_2	0,660 ***	0,400 ***	0,236 ***	0,158 ***	0,097 ***	0,077 ***
β_3	-0,036 ***	-0,026 ***	-0,006	-0,003	0,001	-0,002

Tableau 1 : Estimation des paramètres du modèle logistique ordonné : (codes de significativité : 0`***` 0.001`**` 0.01`*` 0.05`·` 0.1` ` 1)

Table 1: Estimation of the ordered logit model parameters: (significativity codes: 0`***` 0.001`**` 0.01`*` 0.05`·` 0.1` ` 1)

Le modèle logistique ordonné sans et avec effets mixtes

Face à certains types de données, il existe des outils plus performants pour l'analyse que la régression linéaire. C'est le cas notamment lorsque la variable dépendante (la variable que l'on cherche à expliquer) est une variable qualitative. On peut alors faire appel aux modèles de choix discrets, qui associent à chaque réponse possible (ou modalité) une « utilité » dépendant de ses caractéristiques et de celles du sujet considéré. En fonction de ces utilités, on calcule ensuite une probabilité de réponse pour chaque modalité. La calibration du modèle s'effectue en maximisant le produit des probabilités des réponses observées, la « vraisemblance ».

Le modèle logistique ordonné.

Lorsque les modalités sont ordonnées (on peut être par exemple « plutôt pas d'accord », « plutôt d'accord », « totalement d'accord »...), il est important d'en tenir compte au niveau de la modélisation. Une façon naturelle de représenter le processus de décision consiste à définir une « variable latente », continue et dont le niveau détermine la réponse finale du sujet. Le passage de la variable latente à la réponse effective s'effectue au moyen de seuils. La mention au baccalauréat, par exemple, est une variable qualitative déterminée par la moyenne obtenue à l'examen (et qui joue donc le rôle de variable latente), avec des seuils égaux à 10, 12, 14, 16... Le modèle logistique ordonné est un modèle de choix discret s'appuyant sur ce type de modélisation.

Dans le cas de CREED, le contrôleur aérien exprime son évaluation du risque de collision sur une échelle ordonnée à huit niveaux, depuis « airprox assuré » jusqu'à « c'est sans problème » en passant par « méfiance »... La variable latente (notée U) associée au risque de conflit n'est pas observable, et doit donc être reconstruite. C'est à cette fin que les variables E_d , E_f et DA ont été définies. Ces variables, dites explicatives, sont censées expliquer au mieux la réponse du contrôleur. Néanmoins d'autres facteurs susceptibles d'affecter son jugement ne peuvent être ou ne sont pas observés dans l'expérience. Il faut donc décomposer U en une composante observée et une composante non observée : $U = \beta'x + \varepsilon$. Usuellement, la composante non observée ε est considérée

The ordered logit model with and without mixed effects

In view of certain types of data, there are more efficient tools than linear regression for the analysis. This is particularly the case when the dependent variable (the variable that we are trying to explain) is a qualitative variable. We can then bring in discrete choice models, which associate every possible reply (or modality) with a "utility" depending on its characteristics and those of the subject considered. Depending on these utilities, we then calculate the reply probability for each modality. The model is calibrated by maximizing the product of probabilities of observed replies, the "likelihood".

The ordered logit model.

When the modalities are ordered (for example, we can "rather not agree," "somewhat agree," "completely agree," etc.), it is important to take it into account. A natural way of representing the decision process consists in defining a continuous "latent variable", whose level determines the final response of the subject. The transition from the latent variable to the effective response is done using thresholds. For example, the baccalaureate distinction is a qualitative variable determined by the mean obtained during the exam (and which thus plays the role of the latent variable), with thresholds equal to 10, 12, 14, 16... The ordered logit model is a discrete choice model based on this type of modelling.

In the case of CREED, the air traffic controller gives his risk of collision evaluation on a scale ordered at eight levels, from "airprox certain" to "no problem" passing through "caution"... The latent variable (marked U) associated with the risk of conflict cannot be observed, and thus has to be recreated. The variables E_d , E_f and DA were defined for this purpose. These variables, known as explanatory, are meant to explain the response of the controller as best as possible. Nevertheless, other factors likely to affect his judgement cannot be or are not observed in the experiment. Thus U must be broken down into an observed component and a non-observed component: $U = \beta'x + \varepsilon$. Ordinarily, the non-observed component ε is considered as a noise, thus a

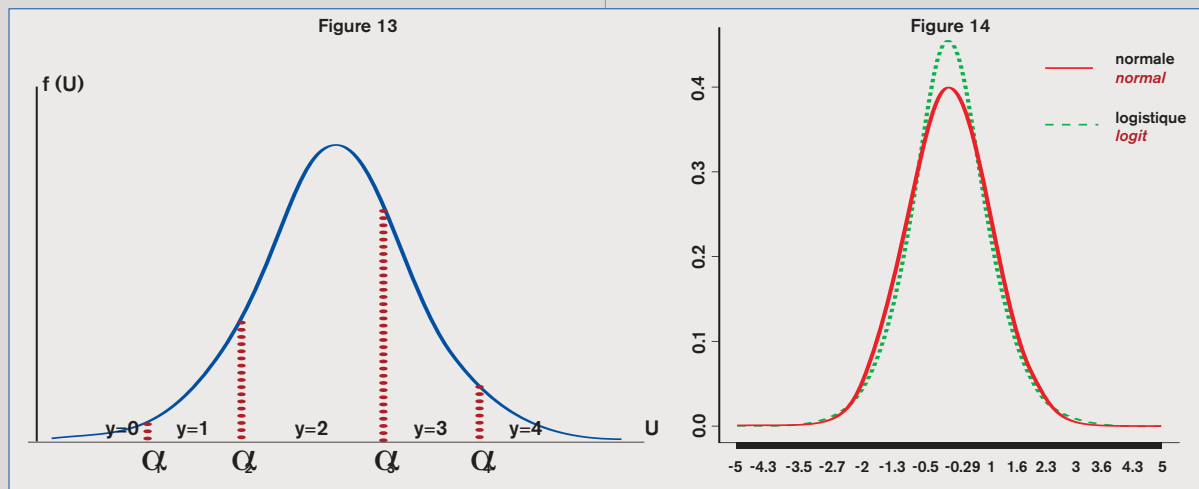


Figure 13 : distribution du jugement du contrôleur
Figure 14 : Comparaison des densités respectives de la loi logistique et la loi gaussienne (moyenne nulle et variance unité)
Figure 13: Distribution of the controller's judgement.
Figure 14: Comparison of the respective densities of the logistical law and the Gaussian law (zero mean and unit variance)

comme un bruit, donc une variable aléatoire dont la loi détermine la probabilité des réponses. Associant aux réponses du contrôleur les valeurs entières ordonnées $y = 0, \dots, 8$, la Figure 13 ci-dessous illustre la situation simplifiée avec seulement quatre seuils (et donc cinq réponses possibles). U est distribuée autour de $\beta'x$ selon la distribution de ε . La probabilité de la réponse $y=0$ (resp. $y=1$) est l'aire située sous la courbe pour les valeurs de $U \leq \alpha_1$ (resp. pour les valeurs $\alpha_1 \leq U \leq \alpha_2$)...

Dans le modèle logistique, le bruit est distribué selon une loi logistique dont la fonction de répartition est $F(t) = \text{Prob}(\varepsilon \leq t) = \text{esp}(t)/(1+\exp(t))$. Cette loi est proche d'une loi gaussienne, comme illustrée dans la Figure 14. Par exemple la probabilité de la réponse $y=0$ est :

$$\text{Prob}(y = 0) = \text{Prob}(U \leq \alpha_1) = \text{Prob}(\beta'x + \varepsilon \leq \alpha_1) = \text{Prob}(\varepsilon \leq \alpha_1 - \beta'x) = \frac{\exp(\alpha_1 - \beta'x)}{1 + \exp(\alpha_1 - \beta'x)}$$

et la probabilité de la réponse $y=1$:

$$\text{Prob}(y = 1) = \text{Prob}(\alpha_1 \leq U \leq \alpha_2) = \text{Prob}(\alpha_1 - \beta'x \leq \varepsilon \leq \alpha_2 - \beta'x) = \frac{\exp(\alpha_2 - \beta'x)}{1 + \exp(\alpha_2 - \beta'x)} - \frac{\exp(\alpha_1 - \beta'x)}{1 + \exp(\alpha_1 - \beta'x)}$$

Les probabilités des autres réponses sont obtenues de façon équivalente.

Le modèle logistique ordonné mixte

Reprenons l'exemple des mentions au baccalauréat. Pour un groupe d'élèves dont on contrôle les caractéristiques (et à difficulté d'examen donnée), le modèle logistique ordonné prévoit la fréquence de chaque mention, et donc également une mention médiane, par exemple (l'examen étant très facile) « Très Bien ». Supposons qu'un élève de ce groupe n'ait obtenu que la mention « Assez Bien ». Sans tirer de conclusion hâtive, on peut néanmoins penser que cet élève est moins studieux que l'élève « moyen » de son groupe, ou bien que le sujet d'examen a porté sur un domaine qu'il dominait moins que d'autres (comme par exemple la thermodynamique)... Comparer ces résultats avec ceux d'examens précédents permet de mieux cerner le profil de l'élève, et ce d'autant plus que leur nombre est important. A la limite, on pourrait introduire une ou plusieurs variables propres à l'élève (et appelées « effets fixes »), ou même définir un modèle par étudiant. Mais il faudrait pour cela disposer de très nombreuses observations. Dans le cadre de CREED, on a, le plus souvent et pour un DA fixé, moins de 7 observations par contrôleur. Un modèle à effets fixes serait donc impossible à estimer. La solution consiste à modéliser les variations dans la population (son hétérogénéité) à l'aide de variables aléatoires. On parle alors de modèle à « effets mixtes ».

Pour CREED, un modèle logistique mixte a été estimé pour chaque valeur de DA fixée. Nous avons autorisé chaque paramètre de la variable latente à être aléatoire, sans restriction sur la matrice de covariance des effets mixtes. La variable latente U devient donc $U = G_0 + (\beta_1 + G_1) \times Ed + (\beta_2 + G_2) \times Efl + (\beta_3 + G_3) \times Ed \times Efl + \varepsilon$ où (G_0, G_1, G_2, G_3) désigne un vecteur gaussien de moyenne nulle et de matrice de covariance Σ ; l'introduction de G_0 permet, en quelque sorte, de modéliser la prudence du contrô-

random variable whose law determines the probability of the replies. Associating the ordered integer values $y = 0, \dots, 8$ to the replies of the controller, Figure 13 below shows the simplified situation with only four thresholds (and thus five possible replies). U is distributed around $\beta'x$ according to the distribution of ε . The probability of the reply $y=0$ (respectively $y=1$) is the area under the curve for the values of $U \leq \alpha_1$ (reply for the values $\alpha_1 \leq U \leq \alpha_2$)...

In the logistical model, the noise is distributed according to a logistical law whose distribution function is $F(t) = \text{Prob}(\varepsilon \leq t) = \text{esp}(t)/(1+\exp(t))$. This law is similar to a Gaussian law, as shown in Figure 4. For example, the probability of the reply $y=0$ is:

$$\text{Prob}(y = 0) = \text{Prob}(U \leq \alpha_1) = \text{Prob}(\beta'x + \varepsilon \leq \alpha_1) = \text{Prob}(\varepsilon \leq \alpha_1 - \beta'x) = \frac{\exp(\alpha_1 - \beta'x)}{1 + \exp(\alpha_1 - \beta'x)}$$

and the probability of the reply $y=1$:

$$\text{Prob}(y = 1) = \text{Prob}(\alpha_1 \leq U \leq \alpha_2) = \text{Prob}(\alpha_1 - \beta'x \leq \varepsilon \leq \alpha_2 - \beta'x) = \frac{\exp(\alpha_2 - \beta'x)}{1 + \exp(\alpha_2 - \beta'x)} - \frac{\exp(\alpha_1 - \beta'x)}{1 + \exp(\alpha_1 - \beta'x)}$$

The probabilities of other replies are obtained similarly.

The ordered mixed logit model

Let us return to the example of baccalaureate distinctions. For a group of students whose characteristics we control (as well as the difficulty of the given exam), the ordered logit model predicts the frequency of each distinction, and thus also a median distinction, for example "Very Good" (the exam being very easy). Let us assume that one student of this group has got only the distinction "Quite Good." Without drawing any hasty conclusion, we can however think that this student is less studious than the "average" student of his group, or that the exam subject dealt with a field that he had not mastered as well as the others (thermodynamics, for example)... Comparing these results with those of previous exams helps determine better the student's profile, and this is that much better as their number is large. Ultimately, we could introduce one or more variables specific to the student (and known as "fixed effects"), or even define one model per student. But this would require many observations. As part of CREED, we have, most frequently and for a fixed DA, less than 7 observations per controller. A fixed effects model would therefore not be tractable. The solution consists in modelling the variations in the population (its heterogeneity) using random variables. We thus speak of "mixed effects" models.

For CREED, a mixed logistical model was evaluated for each fixed value of DA. We have allowed each parameter of the latent variable to be random, without restriction on the covariance matrix of mixed effects. The latent variable U therefore becomes $U = G_0 + (\beta_1 + G_1) \times Ed + (\beta_2 + G_2) \times Efl + (\beta_3 + G_3) \times Ed \times Efl + \varepsilon$ where (G_0, G_1, G_2, G_3) represents a Gaussian vector of zero mean and having the covariance matrix Σ ; the introduction of G_0 helps, in a way, model the controller's prudence by a

leur par une translation des seuils. Précisons que le vecteur gaussien G est supposé indépendant du bruit modélisant la composante non observée.

On peut remarquer que le nombre de paramètres du modèle augmente fortement, puisqu'aux dix paramètres du modèle logistique ordonné initial s'ajoutent les dix paramètres de la matrice Σ . L'introduction de ces paramètres est loin d'être gratuite. Elle permet en effet dans un premier temps de mieux comprendre les sources d'hétérogénéité des réponses (perception de la séparation verticale, prudence...), avant, dans un second temps et grâce à des méthodes bayésiennes (non décrites dans cet article), de s'intéresser aux caractéristiques individuelles des contrôleurs : il est alors possible d'étudier l'influence sur la perception des conflits de facteurs comme l'ancienneté du contrôleur, son sexe, son site d'origine...

Estimation du modèle mixte

Dans le cas du modèle logistique ordonné mixte, la log-vraisemblance (logarithme de la vraisemblance) dépend d'un vecteur gaussien et devient ainsi elle-même aléatoire. Elle ne peut donc en l'état être maximisée afin de déterminer les paramètres du modèle. Une solution classique qui permet de supprimer l'aléa consiste à remplacer la log-vraisemblance par son espérance et à en rechercher le maximum. Malheureusement, la fonction ainsi définie ne peut s'exprimer explicitement et doit être approchée numériquement.

Très schématiquement, le problème rencontré revient à estimer la valeur I_f de l'intégrale d'une certaine fonction f sur le cube unité de dimension d , soit

$$I_f = \int_{[0,1]^d} f(x) dx$$

Pour approcher I_f , on utilise une méthode de quadrature numérique de la forme

$$Q_N f = \sum_{k=1}^N a_k f(x_k);$$

où les points x_1, \dots, x_N appartiennent au cube unité et où les a_1, \dots, a_N sont des poids. La précision de l'approximation et la vitesse de calcul dépendent alors du choix, crucial, des x_i et des a_i .

On peut adopter la méthode de Simpson (points régulièrement espacés), mais cette approche est très sensible, en terme de temps de calcul, à la dimension d de l'espace, puisque l'erreur est en $N^{-4/d}$ où N désigne le nombre total de points en lesquels l'intégrand doit être évalué. Pour CREED, le modèle mixte, non restreint, fait intervenir 4 variables aléatoires indépendantes (donc $d=4$), et la précision souhaitée ainsi que le nombre d'itérations nécessaires à la convergence du modèle rendent cette option inacceptable en pratique.

La méthode de Monte Carlo est une autre stratégie possible. Dans sa version la plus simple, elle consiste à choisir des poids identiques $a_k=1/N$ et des points générés aléatoirement dans le

translation of thresholds. We must specify that the Gaussian vector G is assumed to be independent of the noise modelling the non observed component.

We observe that the number of parameters of the model greatly increases, because ten parameters of the matrix Σ are added to the ten parameters of the initial ordered logistical model. The introduction of these parameters is by no means without consequences. It actually allows us firstly to better understand the sources of heterogeneity of replies (perception of vertical separation, prudence, etc.), and subsequently, with the help of Bayesian methods (not described in this article), to consider individual characteristics of controllers: it is then possible to study the influence of factors such as the seniority of the controller, his gender, his original site etc. on the perception of conflicts

Estimate of the mixed model

In the case of the mixed ordered logistical model, the log-likelihood (logarithm of the likelihood) depends on a Gaussian vector and thus is itself random. Therefore it cannot be maximised in order to determine the parameters of the model. A standard solution used to do away with the variate consists in replacing the log-likelihood by its expected value and to look for the maximum. Unfortunately, the function defined in this way cannot be expressed explicitly and must be numerically approximated.

Schematically, the problem encountered amounts to estimating the value I_f of the integral of a certain function f over the unit cube of dimension d , hence

$$I_f = \int_{[0,1]^d} f(x) dx$$

To approximate I_f , we use a numerical quadrature method of the form

$$Q_N f = \sum_{k=1}^N a_k f(x_k);$$

where the points x_1, \dots, x_N belong to the unit cube and where the a_1, \dots, a_N are weights. The accuracy of the approximation and the calculation speed then depend on the crucial choice of x_i and a_i .

We can adopt Simpson's method (regularly spaced out points), but this approach is very sensitive, in terms of calculation time, to the dimension d of the space, as the error is in $N^{-4/d}$ where N is the total number of points at which the integral must be evaluated. For CREED, the unrestricted mixed model, involved 4 independent random variables (thus $d=4$), and the desired accuracy as well as the number of iterations required for the convergence of the model make this option unacceptable in practice.

The Monte Carlo method is another possible strategy. In its simplest form, it consists in selecting identical weights $a_k=1/N$ and randomly generated points in the unit cube according to a

cube unité selon une distribution uniforme. La loi des grands nombres assure que lorsque N tend vers l'infini, l'erreur décroît vers zéro ; le théorème central limite précise que les fluctuations sont asymptotiquement gaussiennes d'écart type

$$\frac{\sigma(f)}{\sqrt{N}},$$

où $\sigma^2(f) = \text{lf}^2 - (\text{lf})^2$ est la variance de f. L'aspect le plus remarquable de la méthode de Monte Carlo est qu'elle ne souffre pas de la malédiction de la dimension, l'erreur étant en $N^{-1/2}$.

Cependant, la méthode de Monte Carlo reste trop coûteuse pour estimer les paramètres du modèle logistique mixte du fait du très grand nombre d'itérations de l'algorithme d'optimisation. On lui préférera donc une quasi-méthode de Monte Carlo dans laquelle les poids sont égaux, comme dans la méthode de Monte Carlo, mais où les points x_1, \dots, x_N sont choisis de manière plus habile et telle qu'une erreur proche de N^{-1} soit possible. (Signalons néanmoins, qu'une constante pouvant dépendre exponentiellement de la dimension intervient dans cette erreur). La Figure 15 : illustre la différence entre la méthode de Monte Carlo et une quasi-méthode de Monte Carlo basée sur des suites de Halton (qui font partie des suites dites à faible discrédance) : informellement, la suite de Halton explore plus efficacement le cube qu'une suite uniformément distribuée (présence d'un nombre correct de points dans chaque sous-cube). Revers de la médaille, il est difficile d'obtenir de bonnes estimations de l'erreur sur le calcul de l'intégrale dans l'application des quasi-méthodes de Monte Carlo. En effet, il se peut que malgré un meilleur comportement asymptotique, l'erreur commise pour un nombre N fixé de points s'avère plus importante que celle obtenue par une simple méthode de Monte Carlo, et ceci d'autant plus que la dimension d est grande. Des approches mélangeant l'aléatoire et les suites à faible discrédance ont été développées pour tenter de pallier à ce défaut.

uniform distribution. The law of large numbers ensures that when N tends towards infinity, the error decreases to zero; the central limit theorem specifies that the fluctuations are asymptotically Gaussian with a standard deviation

$$\frac{\sigma(f)}{\sqrt{N}},$$

where $\sigma^2(f) = \text{lf}^2 - (\text{lf})^2$ is the variance of f. The most remarkable feature of the Monte Carlo method is that it is not bound by the dimension, the error being $N^{-1/2}$.

However, the Monte Carlo method is too cumbersome to estimate the parameters of the mixed logistical model owing to the very large number of iterations of the optimisation algorithm. We will therefore adopt instead a quasi-Monte Carlo method wherein the weights are equal, as in the Monte Carlo method, but where the points x_1, \dots, x_N are chosen more cleverly and such that an error close to N^{-1} is possible. (We must nevertheless point out that a constant that may depend exponentially on the dimension appears in this error). Figure 15 shows the difference between the Monte Carlo method and a quasi-Monte Carlo method based on Halton series (which are part of series known as low discrepancy series): informally, the Halton sequence explores the cube more effectively than a uniformly distributed series (presence of a correct number of points in each sub-cube). On the downside, it is difficult to obtain good estimations of the error on the calculation of the integral in the application of quasi-Monte Carlo methods. Actually, it may turn out that in spite of improved asymptotic behaviour, the error committed for a fixed number of points N is greater than that obtained by a simple Monte Carlo method, and this is that much greater as the dimension d is large. Approaches combining randomness and low discrepancy series were developed to try and overcome this defect.

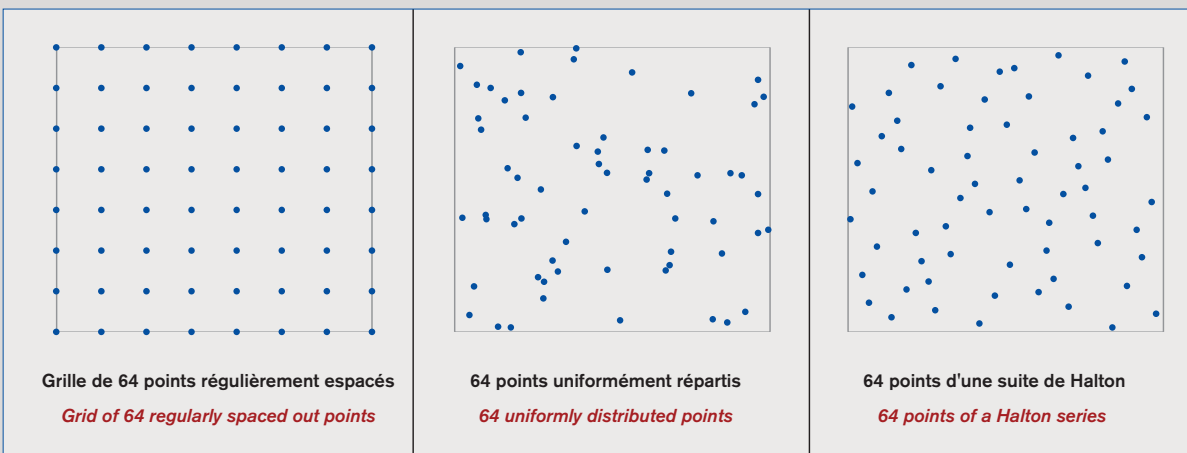


Figure 15 : Différentes grilles de points
Figure 15: Various grids of points

Dans CREED, l'emploi des suites de Halton (préconisé par les statisticiens utilisateurs des modèles logistiques mixtes) nous a permis de diminuer d'un facteur dix le nombre de points d'échantillonnage de la fonction log-vraisemblance. Le temps de calcul nécessaire à l'estimation des paramètres d'un modèle à DA constant a ainsi pu être ramené de 2h30 à 40'.

In CREED, the use of Halton series (recommended by statisticians using mixed logistical models) allowed us to reduce the number of sampling points of the log-likelihood function by a factor of ten. The calculation time required for the estimation of parameters of a model with constant DA was thus reduced from 2h30 to 40'.

On observe tout d'abord que le coefficient β_3 n'est significatif que lorsque DA est faible (ce qui explique qu'il puisse changer de signe pour DA=28). Bien plus intéressant, le rapport β_1 / β_2 varie linéairement en fonction de DA comme le montre la Figure 4. Ce rapport renseigne sur la variable qui est « la plus utilisée » lors du diagnostic « conflit – non conflit » : à l'extrême, s'il était égal à 0 (resp. à l'infini), le contrôleur ne baserait son jugement que sur Efl (resp. Ed). Ces deux variables étant exprimées dans une même unité et décrivant des distances horizontales, on peut estimer que c'est lorsque le rapport β_1 / β_2 est égal à 1 que Ed et Efl sont également discriminantes dans le jugement du contrôleur. Ce fait tendrait à prouver que celui-ci privilégie Efl pour DA inférieur à 12NM tandis qu'Ed devient prépondérant pour DA supérieur à 12NM. En réalité, on peut penser que ce constat est très lié aux scénarii proposés lors de l'expérimentation et pour lesquels, dans les cas de « non conflit », la sépara-

reply (dark red on the data collection interface) and value +10 to the "certain no conflict" (dark green), the amplitude in the variation of replies can therefore show a maximum of 20 units. For the judgements that we gathered, the dispersion of replies is therefore sometimes considerable, since it can be seen (colour code on the right of the figure) that the IQR can exceed 14 units. It is also far from constant, and its variations following E(d), E(fl) and D°A recall the patterns observed on the median graphs: definite "medium" judgements are generally expressed by a low IQR. Hence the traffic scenarii showing the highest IQR correspond to the combination of the three explicative variables for which the ambivalence cannot be lifted, neither by using individual heuristics, nor by using the operational tools that are currently available.

The graphic analysis of the data gathered in the context of CREED allowed the importance of doubt in the conflict detection process to be verified. Without denying the controllers' individual variability, it also identified a global coherency in the expression of their judgements. This very positive result suggests that tools to assist control could be based on the sources of variations identified at this stage of the study (E(d), E(fl) and D°A). Such applications would nevertheless require that we manage to estimate the controllers' judgements for all potentially conflictual situations. This is the purpose of the statistical models developed during the second phase of the CREED project.

5 STATISTICAL MODELS

5.1. Ordered logit model

Now we are going to tackle the part concerning the statistical model, which should allow the judgements observed in the scenarii to be reproduced and interpolated to any potentially conflictual situation.

Because of the discrete and ordered character of the replies proposed by (and to) the controllers, the formalism used is that of the ordered logistical model presented in Box N° 4. The air traffic controllers judged the risk of collision on an ordered scale of eight

Figure 4 : Variation du rapport β_1 / β_2 en fonction de DA (rouge), droite de régression (bleue)
Figure 4: Variation of relation β_1 / β_2 as a function of DA (red), regression line (blue)

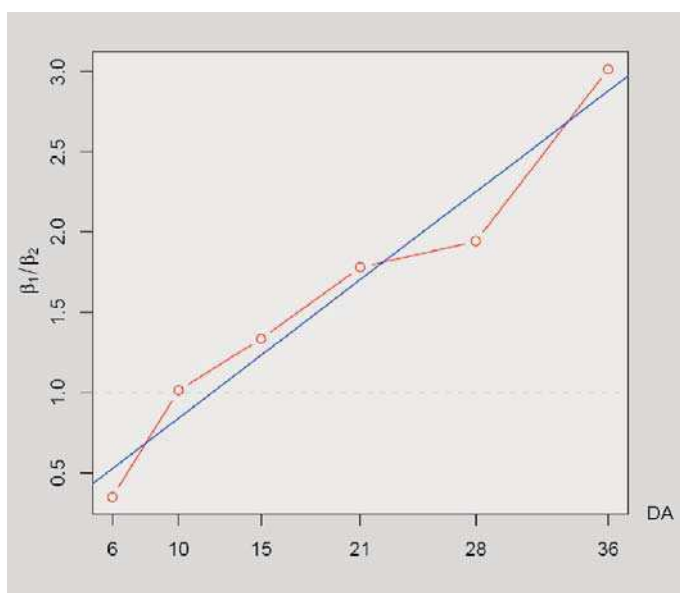
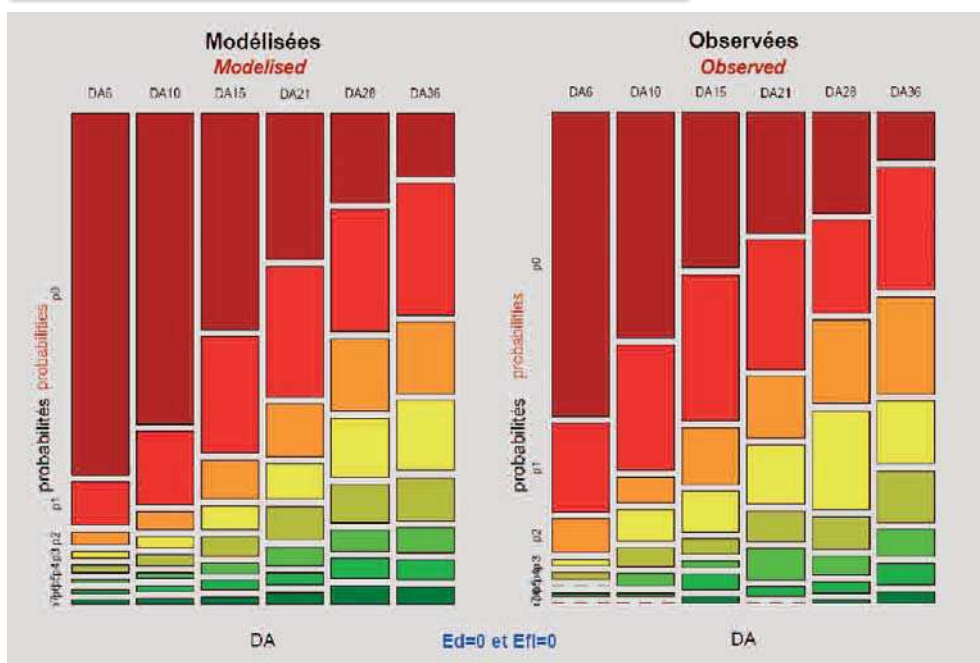


Figure 5 : Comparaison données observées/données prédites pour Ed=0 et Efl=0
Figure 5: Comparison of observed data/predicted data for Ed=0 and Efl=0



levels, going from "airprox certain", through "caution", to "no problem". The logit model is based on a natural way of representing the decision-making process, consisting of defining a continuous "latent variable", U, whose level determines the subject's final reply. If the value of U is less than a certain threshold, noted α_1 , the reply will be "airprox certain"; if the value of U is higher than α_1 , but less than another threshold α_2 , the reply will be the next one up on the scale of possible replies, and so on. As the latent variable U is not observable, it must therefore be reconstructed. It is for this reason that variables Ed, Efl and DA were defined. These variables, known as explicative, are supposed to give the best explanation of the controller's reply. However, due to its specific nature, the variable DA will be treated differently from the others. Hence the construction of a model for each value of DA: a "global" model will be obtained later through the interpolation of these models. Other sources of variations may affect the controller's judgement, but though they cannot be or were not observed in the experiment, they must nevertheless be incorporated into U, which must be broken down into an observed component and a non-observed component. Usually, the non-observed component is considered as a noise, thus as a random variable, whose

law determines the probability of the replies. The observed component is generally modelised by a linear (or polynomial) combination of the explicative variables; for a fixed value of DA, it will therefore be formalised by the linear relation: $\beta'x = \beta_1 \times Ed + \beta_2 \times Efl + \beta_3 \times Ed \times Efl$ where the unknown coefficients β_1 , β_2 , and β_3 , are called regressors: note the presence of the term Ed x Efl which captures the crossed influences between the two variables Ed and Efl. Each model is therefore determined by ten coefficients that need to be estimated. The method used for the estimation is based on the maximisation of a certain function, called log-likelihood, which depends on the data observed and the coefficients; an optimisation algorithm is then used. The estimated parameters are shown in Table 1. First of all, it can be observed that coefficient β_3 is only significant when DA is low (which explains that it can change signs for DA=28). Even more interesting, the relation β_1/β_2 varies linearly as a function of DA as shown in Figure 4. This relation provides information on the variable that is the "most used" during the "conflict – no conflict" diagnosis: in an extreme case, if it equalled 0 (respectively infinity), the controller would only base his judgement on Efl (respectively Ed). As these two variables are

tion se fait d'abord verticalement (les autres types de conflits étant évités autant que possible dans la vie réelle). Lorsque l'empan temporel est faible, les contrôleurs sont donc particulièrement sensibles à la dimension verticale. Ce biais est négligeable pour des empan plus importants. On retrouve alors la préférence attendue des contrôleurs pour les séparations intervenant dans le plan horizontal. Disposant de ces coefficients, nous pouvons alors calculer pour chaque valeur de Ed et Efl la probabilité d'observer une réponse donnée et comparer ces prédictions aux observations (cf. Figure 5).

5.2 Modèle logistique ordonné à effets mixtes

Le modèle logistique estimé fournit le comportement moyen d'un contrôleur ; la variabilité des réponses résulte du bruit ϵ , qui capte à la fois l'hétérogénéité de la population et la non prise en compte de l'ensemble des facteurs perceptifs utilisés pour juger du risque de conflit.

Pour analyser l'hétérogénéité de la population, liée par exemple à l'influence de certaines caractéristiques des individus (comme l'âge, le sexe, le site), il est possible de considérer chaque individu comme distribué aléatoirement dans la population des contrôleurs. Cette manière de procéder revient à formaliser l'inconnu à travers l'aléatoire, en remplaçant les paramètres du modèle par des variables aléatoires. Mais quelle distribution choisir ? Comme souvent pour ce type de modèles, la distribution gaussienne a été retenue. Une validation statistique de ce choix peut être faite a posteriori (cf. Figure 7). La variable latente U devient ainsi une variable aléatoire, dont la partie observée s'exprime sous la forme

$$G_0 + (\beta_1 + G_1) \times Ed + (\beta_2 + G_2) \times Efl + (\beta_3 + G_3) \times Ed \times Efl$$

où (G_0, G_1, G_2, G_3) désigne un vecteur gaussien de moyenne nulle. Alors que les variables G_1, G_2 et G_3 décrivent une plus ou moins grande sensibilité du sujet aux différentes variables, l'introduction de G_0 permet, en quelque sorte, de modéliser la prudence du contrôleur par une translation des seuils. Précisons que le vecteur gaussien G est supposé indépendant du bruit modélisant la composante non observée.

On peut remarquer que le nombre de paramètres du modèle augmente fortement, puisqu'aux dix paramètres du modèle logistique ordonné initial s'ajoutent les dix paramètres de la matrice de covariance du vecteur gaussien (matrice symétrique). L'introduction de ces paramètres est loin d'être gratuite ; elle permet en effet de mieux comprendre les sources d'hétérogénéité des réponses comme nous le verrons ci-dessous pour le paramètre β_3 .

Outre une complexité accrue, l'utilisation de coefficients bruités donne un caractère aléatoire à la fonction log-vraisemblance, qui rend l'estimation des paramètres du modèle plus délicate que dans le cas standard. Pour plus de détails, le lecteur intéressé se référera à l'encart n° 4, où les outils et méthodes numé-

riques utilisés au cours de cette étude sont introduits. On se contente ici de donner dans le Tableau 2 les valeurs des variances des variables gaussiennes G_0, G_1, G_2, G_3 .

expressed in the same unit and describe horizontal distances, it can dimension. This bias is negligible for longer time spans, and we recover the

Tableau 2 : Variances des effets mixtes
Table 2: Mixed effect variances

	DA=6	DA=10	DA=15	DA=21	DA=28	DA=36
var(G_0)	2,45	4,3	0,931	-0,843	-0,696	0,572
var(G_1)	0,148	0,124	0,0503	0,0236	0,0167	0,0159
var(G_2)	0,0292	0,0219	0,0070	0,0093	0,0104	0,00457
var(G_3)	0,0004	0,0005	0,0002	7,86 E-05	5,94 E-05	2,79 E-05

La très faible variance de G_3 , relativement à la valeur de β_3 , indique une très faible variabilité de ce paramètre dans la population ; la variable aléatoire G_3 peut vraisemblablement être supprimée, réduisant du même coup de quatre le nombre de paramètres du modèle. Cette hypothèse doit bien sûr être validée, ce qui a pu être fait en obtenant, dans un premier temps, les paramètres du nouveau modèle et en appliquant, ensuite, un test d'hypothèse (basé sur le critère d'Akaike). La valeur du test n'ayant pas permis pas de rejeter l'hypothèse faite, celle-ci a été conservée. Nous donnons dans le Tableau 3 les estimations des seuils, des régresseurs et les variances des effets mixtes pour ce modèle.

be estimated that when the relation β_1 / β_2 equals 1, then Ed and Efl are evenly discriminating in the controller's judgement. This fact tends to prove that this privileges Efl for DA less than 12NM whereas Ed becomes dominant for DA over 12NM. In reality, we may think that this observation controllers' expected preference for separations occurring on the horizontal plane. Disposing of these coefficients, we can then calculate for each value of Ed and Efl the probability of observing a given reply and compare these predictions with the observations. (Figure 5).

Tableau 3 : Paramètres pour le second modèle logistique à effets mixtes
Table 3: Parameters for the second logistical model with mixed effects

	DA=6	DA=10	DA=15	DA=21	DA=28	DA=36
α_1	1,895	1,079	-0,012	-0,720	-1,495	-2,052
α_2	2,847	2,320	1,286	0,569	-0,117	-0,361
α_3	3,307	2,814	1,902	1,176	0,623	0,369
α_4	3,666	3,206	2,454	1,720	1,350	1,224
α_5	4,254	3,822	3,135	2,471	2,054	1,985
α_6	4,595	4,306	3,728	3,212	2,648	2,678
α_7	5,361	5,120	4,815	4,194	3,634	3,668
β_1	0,219	0,468	0,280	0,281	0,194	0,256
β_2	0,736	0,476	0,274	0,186	0,109	0,087
β_3	-0,026	-0,027	0,004	0,002	0,005	-0,002
var(G_0)	1,688	2,772	0,845	0,024	0,474	0,632
var(G_1)	0,045	0,013	0,016	0,030	0,009	0,017
var(G_2)	0,019	0,012	0,006	0,006	0,004	0,004

Précisons pour terminer qu'il est possible, par une méthode dite bayésienne (cf. encart n° 4), d'obtenir une estimation (dans un certain sens) des paramètres propres à chaque sujet. Ceci permet d'observer l'influence sur le jugement du contrôleur de certains facteurs comme le site, le sexe et l'âge. Pour illustrer le type d'analyse qu'il est possible d'effectuer, étudions l'influence de l'âge. Les contrôleurs ont été regroupés en trois catégories : « jeune », « expert », « ancien » [Montjalet], cette classification tenant plus compte du facteur « expertise » du contrôleur que de son âge. Pour caractériser l'influence de l'expertise sur le jugement du contrôleur, nous représentons en Figure 6, pour chaque DA, les diagrammes en boîtes (cf. encart

is highly linked to the scenarii proposed in the experiments and for which, in cases of "no conflict", the separation is first done vertically (the other types of conflict being avoided as far as possible in real life). When the time span is short, the controllers are particularly sensitive to the vertical

5.2. Ordered logit model with mixed effects

The estimated logit model provides the controllers' average behaviour; the variability of the replies results from the noise ϵ , which captures both the heterogeneity of the population and the fact that all the perceptive factors

used to judge the risk of conflict are not taken into account.

In order to analyse the heterogeneity of the population, linked for example to the influence of certain individual characteristics (such as the age, the gender, or the site), it is possible to consider each individual as being randomly distributed in the controller population.

This method comes back to formalising the unknown by the random, replacing the model parameters by random variables. But what distribution should be chosen? As is often the case for this type of model, Gaussian distribution was used. A statistical validation of this choice can be made later (see Figure 7). The latent variable U then becomes a random variable, whose observed part is expressed in the following form:

$G_0 + (\beta_1 + G_1) \times Ed + (\beta_2 + G_2) \times Efl + (\beta_3 + G_3) \times Ed \times Efl$
 where (G_0, G_1, G_2, G_3) designates a zero mean Gaussian vector. Whereas variables G_1, G_2 and G_3 describe the subject's higher or lower sensitivity to the different variables, the introduction of G_0 allows the controller's prudence to be modelised

in a certain way, by a translation of the thresholds. We should state that the Gaussian vector G is supposed to be independent of the noise modelling the non-observed component. It can be observed that the number of parameters in the model strongly increases, since, in addition to the ten

n° 3) de l'échantillon des valeurs estimées de G_0 pour chaque contrôleur. On observe ainsi une grande variation de la médiane pour DA=6, 10 et 15. Néanmoins ceci reste, à l'heure actuelle, difficile à interpréter correctement en raison de la corrélation entre les effets mixtes. Quant au sexe et au site, leur influence semble quasi-nulle.

Cette approche fournit également un moyen de valider l'hypothèse gaussienne faite sur la distribution des individus, comme l'illustre la Figure 7 où sont représentés les coefficients pour chaque contrôleur (en noir) ; l'hypothèse gaussienne sera d'autant plus acceptable que ces points seront alignés sur la droite rouge.

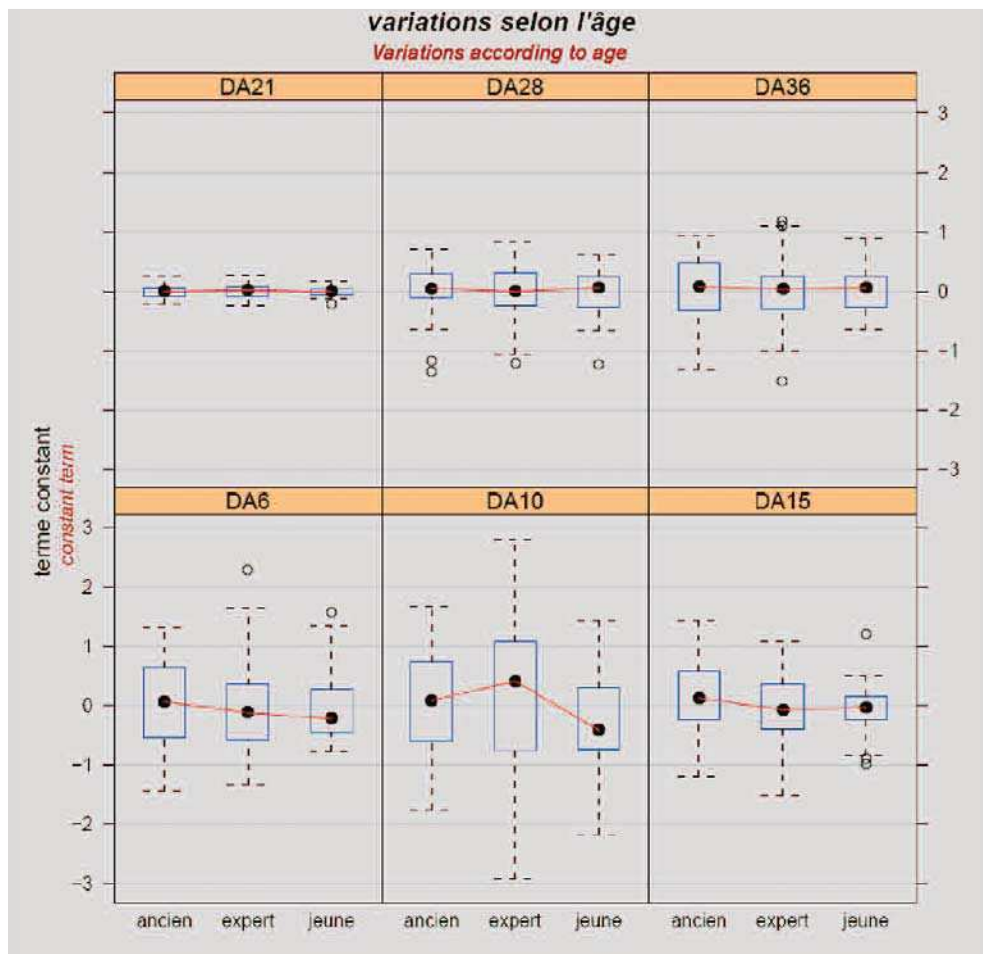


Figure 6 : Variabilité du seuil selon l'expertise du contrôleur
 Figure 6: Variability of the threshold according to the controller's expertise

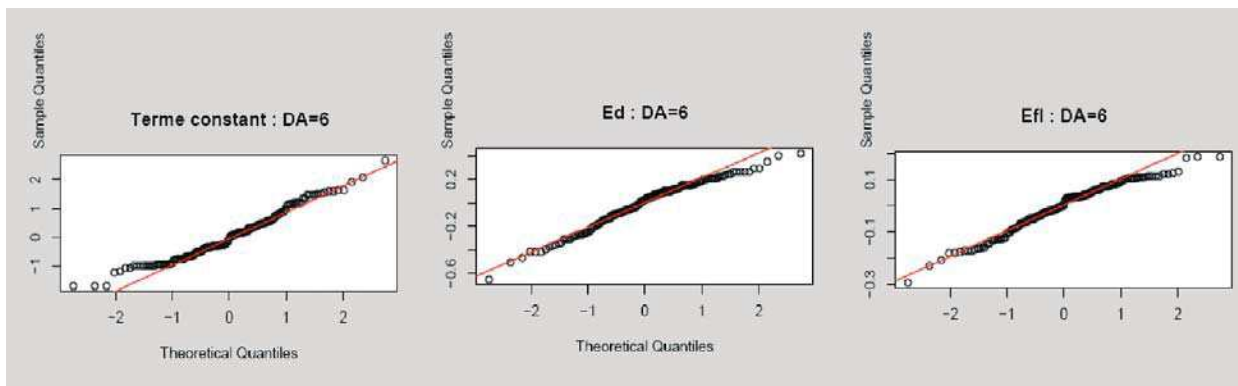


Figure 7 : Validation de l'hypothèse gaussienne pour la distribution des coefficients pour DA=6
 Figure 7: Validation of the Gaussian hypothesis for the distribution of coefficients for DA=6

5.3 Une illustration de l'hétérogénéité traduite par les effets mixtes.

L'estimation des caractéristiques propres de chaque contrôleur permet de comparer les (fréquences des) réponses données par deux contrôleurs confrontés à une situation identique. Ainsi, la Figure 8 représente, pour $DA=10$, la valeur de la médiane des réponses données par les contrôleurs 22M et 15T pour différentes configurations de croisement. Précisément, les courbes tracées représentent les valeurs des couples (E_d, E_{fl}) pour lesquelles la variable latente U est égale à l'un des seuils ; dans les domaines délimités par ces courbes, les médianes des réponses des contrôleurs sont donc constantes. On remarque que les courbes associées à 22M (continues) sont plus inclinées que celles associées à 15T, ce qui dénote une plus grande sensibilité à E_d (on peut rappeler que pour des valeurs de DA assez faibles, comme c'est le cas ici, le rôle de E_{fl} est plus important que celui de E_d). On peut également noter que les « iso-seuils » de 22M sont plus rapprochés que ceux de 15T : les jugements de 22M sont « plus tranchés » que ceux de 15T. Enfin, les courbes noires distinguent les croisements « qui passent » de ceux qui « ne passent pas ». On observe que ces courbes restent relativement proches, en particulier lorsque E_d est élevée (lorsque les trajectoires sont exactement convergentes, c.-à-d. pour $E_d=0$, le contrôleur 22M prend quand même près de 1.5 NM de marge de plus que le contrôleur 15T). Ainsi, même si les contrôleurs ne se comportent pas exactement de la même manière, leur perception de la différence entre croisements sûrs et croisements dangereux est assez similaire.

La comparaison de deux individus permet d'appréhender l'hétérogénéité de la population de manière plus parlante que l'analyse directe de la matrice de covariance des effets mixtes. Elle est également plus restrictive, l'étude de sujets deux à deux

parameters of the initial ordered logit model, there are the ten parameters of the covariance matrix of the Gaussian vector (symmetrical matrix). The introduction of these parameters is far from being unwarranted, as it enables us to better understand the sources of heterogeneity of the replies. An illustration is provided below for parameter β_3 .

In addition to an increased complexity, the use of random coefficients yields a random character to the log-likelihood function, which makes the estimation of the model's parameters more delicate than in standard cases. For further details, the interested reader can refer to Box N° 4, where the tools and numerical methods used during this study are introduced. On page 59, in Table 2, we give simply the values of the variances of the Gaussian

variables G_0, G_1, G_2, G_3 . The very low variance of G_3 , in relation to the value of β_3 , indicates a very low variability of this parameter in the population; the random variable G_3 can most likely be eliminated, eliminating at the same time four of the parameters of the model. This hypothesis, of course, had to be validated, which was done by obtaining, in the beginning, the parameters of the new model and then by applying a hypothesis test (based on Akaike criteria). As the value of test did not allow the hypothesis to be rejected, it was retained. We show in Table 3 the estimation of the thresholds, the regressors and the variances of the mixed effects for this model.

To finish, we would state that it is possible to obtain an estimate (in a certain way) of the specific parameters for each subject, using a Bayesian method (see Box N° 4). This allows observation of the influence on the controller's judgement of certain factors such as the site, the gender and the age. To illustrate the type of analysis that can be made, let us study the influence of age. The controllers were grouped into three categories: "young", "expert", "senior" [Montjalet]. This classification took into consideration the controller's "expertise" more than the age itself. In order to characterise the influence of the expertise on the controller's judgement, we show in

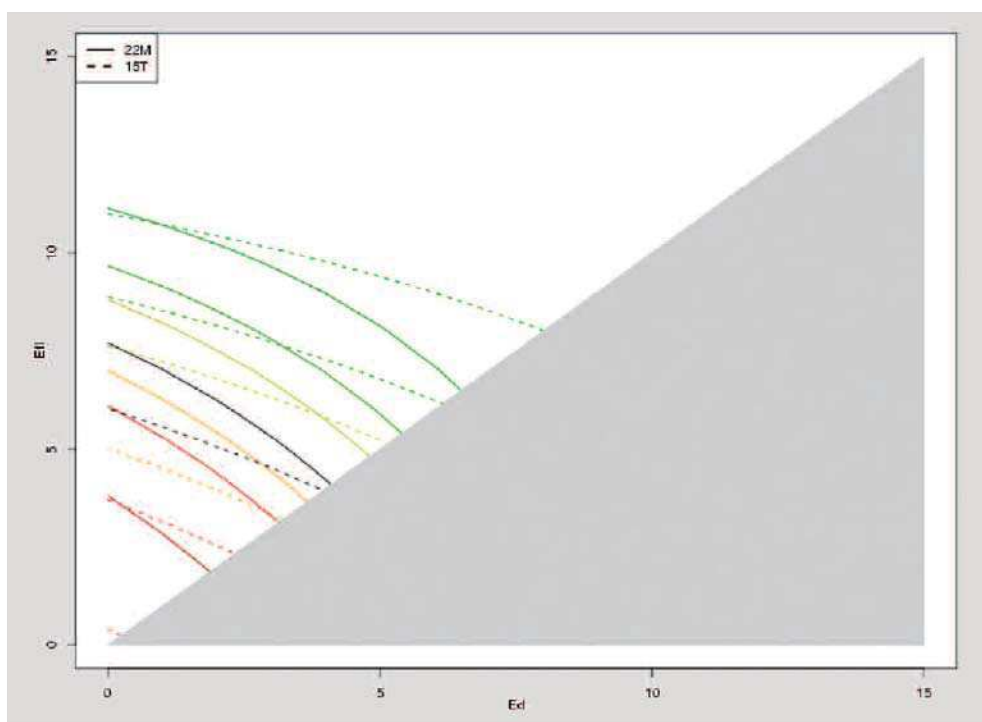


Figure 8 : Médiane des réponses données par les contrôleurs 22M et 15T pour différentes configurations de croisement.

Figure 8: Median of replies given by controllers 22M and 15T for different crossing configurations.

Figure 6 on page 58, for each DA, the diagrams in boxes (see Box N° 3) of the sample of estimated values of G0 for each controller. A wide variation of the median can be observed for DA=6, 10 and 15. However, at the present time, this still remains difficult to interpret correctly due to the correlation between the mixed effects. With regards to the gender and the site, their influence seems to be practically nil. This approach also provides a means to validate the Gaussian hypothesis made on the distribution of individuals, as illustrated in Figure 7, which shows the coefficients for each controller (in black); the Gaussian hypothesis is all the more acceptable when the points are aligned on the red line.

5.3. An illustration of heterogeneity expressed by the mixed effects

The estimation of each controller's own characteristics allows a comparison of the (frequency of) replies given by two controllers confronted with an identical situation. Thus, Figure 8 shows for DA=10, the value of the median of the replies given by controllers 22M and 15T for different crossing configurations. More precisely, the curves plotted show the values of the couples (Ed, Efl) for whom the latent variable U is equal to one of the thresholds; in the areas delimited by these curves,

the medians of the controllers' replies are therefore constant. It can be observed that the curves associated with 22M (solid lines) are steeper than those associated with 15T, which indicates a higher sensitivity to Ed (it should be recalled that for fairly low values of DA, which is the case here, the role of Efl is more significant than that of Ed). It can also be observed that the 22M "iso-thresholds" are closer together than those of 15T: the judgements of 22M are "more decisive" than those of 15T. Finally, the black curves distinguish the crossings that "make it" from those that "do not make it". We can see that these curves remain relatively close, particularly when Ed is high (when the trajectories converge exactly, i.e. for Ed=0, controller 22M takes a margin of almost 1.5 NM more than controller 15T). Therefore, even if the controllers do not behave in exactly the same way, their perception of the difference between safe crossings and dangerous crossings is fairly similar. The comparison of two individuals makes an analysis of the heterogeneity of the population more meaningful than a direct analysis of the mixed effects covariance matrix. It is also more restrictive, as the study of subjects two by two gives the potentially misleading impression that the heterogeneity is "unidirectional". In the CREED context this is however the case, since

donnant l'impression potentiellement trompeuse que l'hétérogénéité est « unidimensionnelle ». Dans le cadre de CREED, c'est pourtant bien le cas, puisque la corrélation des effets aléatoires est très importante. Il est ainsi presque possible d'ordonner les individus suivant leur mode de perception des conflits. C'est en fait de cette manière que les individus 22M et 15T, qui correspondent approximativement aux 1^{er} et 3^{ème} quartiles de la population, ont été sélectionnés.

Enfin, les effets mixtes permettent de décomposer la variance de la variable latente U et d'évaluer ainsi la part de la variabilité des réponses liée à l'hétérogénéité de la population. Cette part, qui dépend des caractéristiques du conflit, est comprise entre 0 et 20% pour des valeurs raisonnables de Ed et Efl (c.-à-d. pour des configurations de croisement qui ne soient ni trop sûres, ni trop dangereuses) dans le cas DA=6, et n'est jamais inférieure à 15% pour DA=10... La diversité des réponses observées lors de l'expérience ne doit donc pas, loin s'en faut, être imputée en intégralité à l'hétérogénéité des contrôleurs !

5.4 Modèle global (sans effets mixtes)

Jusqu'ici nous avons construit des modèles logistiques pour chacune des valeurs de la variable explicative DA. Pour pouvoir interpoler, il devient nécessaire d'intégrer cette variable dans un modèle plus général, que nous appellerons « modèle global ». Mais comment inclure cette variable, puisque dès le départ nous l'avons traité de façon différente des deux autres variables Ed et Efl ?

Nous avons opté pour l'approche synthétisée dans la Figure 9 :

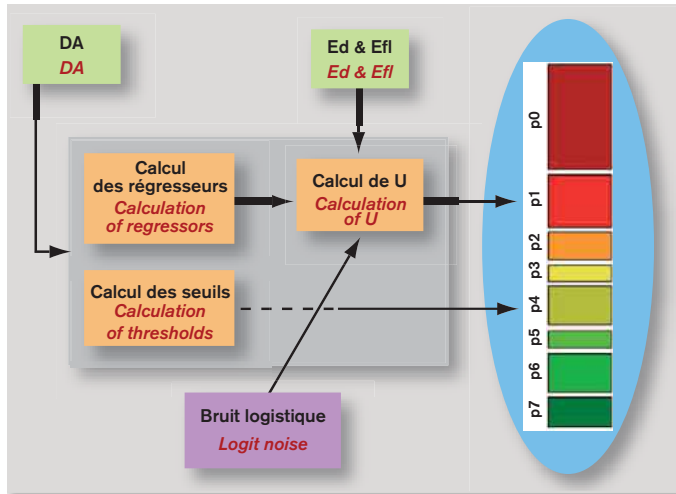


Figure 9 : Approche globale : DA, Ed et Efl étant fixées, les régresseurs et les seuils seront calculés, d'où la valeur de la partie observée de U. La fonction de répartition du bruit permet alors de calculer les probabilités de chaque réponse.
Figure 9: Global approach: as DA, Ed and Efl are fixed, the regressors and thresholds will be calculated, hence the value of the observed part of U. The noise distribution function allows calculation of the probabilities for each reply.

la variable DA étant fixée, les régresseurs et les seuils sont estimés, ce qui permet, pour des valeurs de Ed et Efl fixées, de calculer la partie observée de U. On peut ensuite, grâce à l'expression de la fonction de répartition du bruit et aux formules décrites dans l'encart n° 4, obtenir les probabilités de chaque réponse. Pour obtenir un modèle global, il faut donc trouver un moyen d'évaluer, pour des valeurs de DA différentes de celles déjà considérées, les sept seuils et les trois régresseurs. Une technique classique en statistique et appropriée à ce pro-

blème est la régression (non linéaire ici). Après avoir établi autant de modèles logistiques ordonnés (sans effets mixtes) que de valeurs de DA (DA=6, 10, 15, 21, 28, 36), nous disposons pour chaque paramètre d'un échantillon de six valeurs. La régression non linéaire permet alors d'exprimer chacun des paramètres comme une fonction de la variable DA, cette dernière devant être choisie a priori dans une certaine famille de fonctions, de façon à s'approcher au plus près des valeurs de l'échantillon. La Figure 10 montre les résultats des régressions, la courbe rouge étant la fonction obtenue.

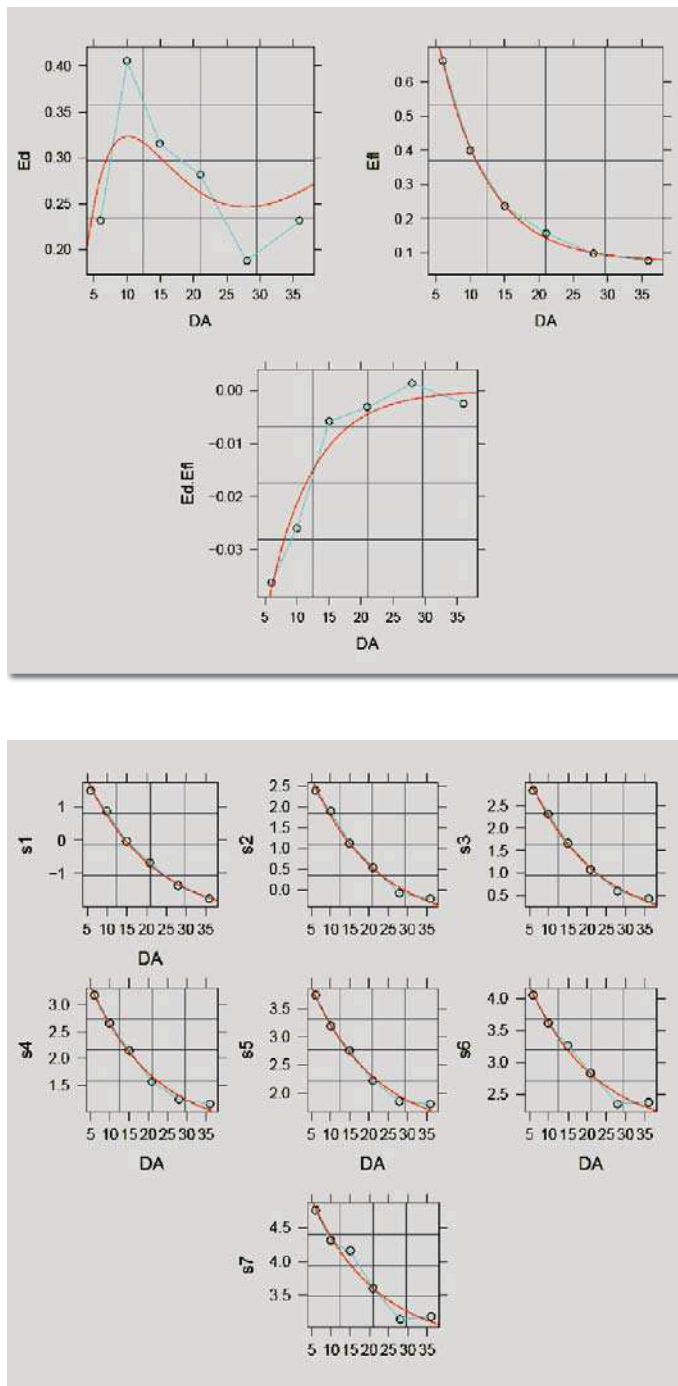


Figure 10 : Régression non linéaire de chaque paramètre : en rouge la fonction, en bleue les valeurs pour chaque modèle.

Figure 10: Non-linear regression of each parameter; in red the function, in blue the values for each model.

the correlation of the random effects is very important. It is therefore almost possible to put the individuals in order according to their conflict perception mode. It is, in fact, in this way that controllers 22M and 15T, who correspond approximately to the 1st and 3rd quartiles of the population, were selected. Finally, the mixed effects allow the variance of the latent variable U to be broken down and the share of the variability of replies linked to heterogeneity of the population to be evaluated. This share, which depends on the conflict characteristics, is between 0 and 20% for reasonable values of Ed and Efl (i.e. for crossing configurations that are neither too sure nor too dangerous) in the case DA=6, and never less than 15% for DA=10... The diversity of replies observed during the experiment should then not be attributed entirely to the heterogeneity of controllers – far from it!

5.4. Global model (without mixed effects)

So far we have constructed logit models for each of the values of explicative variable DA. In order to be able to interpolate, it becomes necessary to integrate this variable into a more general model, which we will call "global model". But how can we include this variable, since right from the start we treated it differently from the two other variables Ed and Efl? We opted for the

synthesized approach in Figure 9: as the variable DA is fixed, the regressors and thresholds are estimated. This allows the observed part of U to be calculated for fixed values of Ed and Efl. Next, using the expression of the noise distribution function and the formulae described in Box N° 4, we can obtain the probabilities for each reply. In order to obtain a global model, we therefore need to find a way to evaluate the seven thresholds and the three regressors, for DA values that are different from those already considered. A classical technique in statistics and appropriate to this problem is regression (non linear here). After having established as many ordered logit models (without mixed effects) as values of DA (DA=6, 10, 15, 21, 28, 36), we have a sample of six values for each parameter. Non-linear regression then allows each of the parameters to be expressed as a function of variable DA, which must be chosen a priori in a certain family of functions, in order to approach the values of the sample as closely as possible. Figure 10 shows the results of the regressions, the red curve being the function obtained. Each of the functions chosen uses three coefficients, thirty are needed to define the global model. Fortunately, as some of these are almost equal, it is possible to simplify the model by imposing these equalities, not without having

checked them previously by statistical tests (which must therefore not reject these equality hypotheses). It should be noted that regressor β_1 of Ed looks different from the other coefficients; this behaviour can be explained by the linear dependency in DA of the ratio β_1/β_2 . When this last restriction is imposed, the final global model "only" depends on 22 parameters.

6 CONCLUSION

The highly positive results obtained in the context of the CREED project clarified the knowledge of the cognitive mechanisms at the basis of conflict detection, highlighting in particular the importance of doubt management. In addition, the widespread (and rather frightening) idea according to which we observe a mode of control by controller has been challenged by the global coherency and sharpness that was identified graphically in the whole of the judgements expressed. In these conditions, describing statistically the behaviour of the "representative controller" is meaningful, and justifies the work of conflict detection modelisation undertaken in the second phase of the study. The models developed allowed the part of heterogeneity of the controller population to be measured in the variability of the replies observed, and to better understand the influence of the geometrical characteristics of the potential conflicts on their

perception. Although the results do not apply to all pairs of aircraft, for reasons of sources of variations not yet taken into account, the image of know-how in ATC has gained in precision. This statistical model also constitutes a significant step forward in the area of the design of operational tools. As it authorises the calculation of probabilities of judgement, it can participate in the standardisation and definition of complexity indicators capturing the propensity of a portion of airspace to generate conflicts (see [Gianazza] for a first example of such an application). It therefore opens up some interesting perspectives for progress in understanding the intrinsic complexity of airspace. Furthermore, the areas where, simultaneously, the "doubt" was at a maximum and the aircraft were close to each other, were identified with a new precision. These are traffic configurations that, in terms of resources mobilised by the controller, are "costly". Strategies are envisaged to guarantee that the future separation values of these configurations are increased, in order to facilitate doubt management and consequently to mobilise fewer resources. This is the aim of the European ERASMUS project, of which the DSN is one of the key players. The basic idea is that the gain obtained through the early modification of certain conflict parameters should result in traffic management that has a

Chacune des fonctions choisies faisant intervenir trois coefficients, il en faut trente pour définir le modèle global. Heureusement, certains de ceux-ci étant presque égaux, il est possible de simplifier le modèle en imposant ces égalités, non sans les avoir vérifiées au préalable par des tests statistiques (qui ne doivent donc pas rejeter ces hypothèses d'égalité). Notons que le régresseur β_1 de Ed présente une allure différente des autres coefficients ; ce comportement peut s'expliquer par la dépendance linéaire en DA du ratio β_1/β_2 . Lorsque l'on impose cette dernière restriction, le modèle global final ne dépend « plus que » de 22 paramètres.

6 CONCLUSION

Les résultats très positifs obtenus dans le cadre du projet CREED ont permis de préciser la connaissance des mécanismes cognitifs à la base de la détection des conflits, en soulignant notamment l'importance de la gestion du doute. Par ailleurs, l'idée couramment répandue (et un peu effrayante) suivant laquelle on observe un mode de contrôle par contrôleur a été mise à mal par la cohérence et l'acuité globales identifiées graphiquement dans l'ensemble des jugements exprimés. Dans ces conditions, décrire statistiquement le comportement du « contrôleur représentatif » a un sens, et justifie le travail de modélisation de la détection des conflits entrepris dans la seconde phase de l'étude. Les modèles développés ont permis de mesurer la part de l'hétérogénéité de la population des contrôleurs dans la variabilité des réponses observées et de mieux comprendre l'influence des caractéristiques géométriques des conflits potentiels sur leur perception. Bien que les résultats ne s'appliquent pas à toute paire d'aéronefs, en raison des sources de variations non encore prises en compte, l'image du savoir faire en ATC a donc gagné en précision.

Ce modèle statistique constitue également une avancée significative pour ce qui est de la conception d'outils opérationnels. Autorisant le calcul des probabilités de jugement, il peut en effet participer à l'étalonnage et la définition d'indicateurs de complexité capturant la propension d'une portion d'espace à générer des conflits (on trouvera dans [Gianazza] un premier exemple d'application dans ce sens). Il ouvre donc d'intéressantes perspectives de progrès dans la compréhension de la complexité intrinsèque de l'espace aérien.

Par ailleurs, les plages où, simultanément, le « doute » est maximal et les aéronefs proches l'un de l'autre ont pu être identifiées avec une précision nouvelle. Ce sont des configurations de trafic qui, en terme de ressources mobilisées chez le contrôleur, « coûtent cher ». Des stratégies sont envisagées pour garantir que les valeurs des séparations futures de ces configurations soient augmentées, afin de faciliter la gestion du doute et par conséquent de mobiliser des ressources moins importantes. C'est l'objectif du projet européen ERASMUS, dont la DSN est l'un des acteurs principaux. L'idée de base est que le gain obtenu grâce à la modification précoce de certains para-

mètres des conflits doit pouvoir aboutir à une gestion du trafic plus capacitive et plus sûre. Notre modèle statistique sera mis à contribution pour tester certaines hypothèses.

Il sera très intéressant de compléter ce modèle, afin qu'il soit en mesure de prédire les jugements probables des contrôleurs sur toute paire d'avions, en approche mais aussi en route. Un complément d'expérimentation assez important va être proposé dans ce sens. Par ailleurs, deux thèses de doctorat dont les sujets sont les stratégies cognitives des contrôleurs en situation de détection/résolution de conflit, puis leur implémentation informatique, ont débuté cette année pour deux étudiants de l'Université de Toulouse 2, qui ont pris comme point de départ les résultats du projet CREED.

Enfin, au delà des résultats obtenus et des applications envisagées à plus ou moins long terme, il convient de souligner que le projet CREED a su tirer parti de la complémentarité des chercheurs présents à la SDER. Le dialogue entretenu par ces acteurs pour modéliser parallèlement les aspects statistiques et cognitifs de l'ATC a contribué à éviter des dérives techniques dépourvues de sens. En retour, le modèle statistique a fourni des arguments permettant de valider les énoncés des sciences cognitives...

higher capacity and is safer. Our statistical model will be put to contribution to test certain hypotheses. It will be very interesting to complete this model so that it is able to predict the controllers' probable judgements on any pair of aircraft, in approach and also en route. A fairly significant additional set of experiments will be proposed to this end. In addition, two doctoral theses, in the areas of controllers' cognitive strategies in situations of conflict detection/resolution, and their computerised implementation, started this year for two students from the University of Toulouse 2, who used the

results of the CREED project as a starting point. Finally, beyond the results obtained and the applications that are planned in the shorter or longer term, it should be pointed out that the CREED project was able to take advantage of the complementarity of the researchers present at the SDER. The ongoing dialogue by these players to modelise simultaneously the statistical and cognitive aspects of ATC contributed to avoid technical drifts that would be devoid of meaning. In return, the statistical model provided arguments that allowed the validation of statements of cognitive sciences.

Article rédigé par Philippe AVERTY, Kévin GUITTET et Pascal LEZAUD - Written by Philippe AVERTY, Kévin GUITTET and Pascal LEZAUD

BIBLIOGRAPHIE BIBLIOGRAPHY

- Averty P. : *Conflict perception by ATCS admits doubt but not inconsistency*, in Proceedings of the 6th Air Traffic Management Research & Development Seminar, Baltimore (USA), 2005.
- Bisseret A. : *Utilisation de la théorie de la détection du signal pour l'étude de décisions opératives : effet de l'expérience des opérateurs*, Rapport I.R.I.A./C.E.N.A. C.O.7911-R60 ; 1979.
- Leroux M. : *Une approche cognitive pour la spécification d'aides à la décision pour les contrôleurs aériens*, Document CENA 93594 ; Avril 1993.
- Alliot J.M. ; Durand N. ; Granger G. : *A statistical analysis of the influence of vertical and ground speed errors on conflict probe*, in Proceedings of the 4th Air Traffic Management Research & Development Seminar, Santa Fe (USA), 2001.
- Gianazza D. ; Guittet K. : *Evaluation of air traffic complexity metrics using neural networks and sector status*, accepted at the 2nd International Conference on Research in Air Transportation (2006).
- Labrune Y. : *Traitement statistique des données du projet CREED*, Rapport de fin d'année de l'IUT Statistiques et Traitement Informatiques des Données, Université de Pau des Pays de l'Adour, 2004.
- Gille J. & Walschaerts M. : *Modélisation du risque de collision perçu par les contrôleurs aériens*, Rapport de projet de 5ième année Génie Mathématique et Modélisation, INSA Toulouse, 2005.
- Monjalet A. : *Évaluation de la perception des risques de conflits entre aéronefs chez les contrôleurs aériens*, Rapport Master1 Sciences Humaines Appliquées Option Modélisation, Université de Bordeaux II, 2005.