HAL
open science

Analyse descriptive des indicateurs de complexité du trafic aérien à partir des données IMAGE et COURAGE

Kevin Guittet, David Gianazza

To cite this version:

Kevin Guittet, David Gianazza. Analyse descriptive des indicateurs de complexité du trafic aérien à partir des données IMAGE et COURAGE. [Research Report] Ecole Nationale de l'Aviation Civile. 2005. hal-01020620

HAL Id: hal-01020620

https://enac.hal.science/hal-01020620v1

Submitted on 8 Jul 2014

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

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

Analyse descriptive des indicateurs de complexité du trafic aérien à partir des données IMAGE et COURAGE

Projet S2D2 (Sector Split and Dynamic Density)

Kevin Guittet David Gianazza
DSNA/DTI/SDER
Laboratoire d'Economie et d'Econométrie de l'Aérien
Laboratoire d'Optimisation Globale
7, avenue Edouard Belin 31055 Toulouse Cedex
Email : guittet,gianazza@recherche.enac.fr

Note DSNA/DTI/SDER NR05-905
Point d'avancement du 9 décembre 2005

Table des matières

1 Présentation des indicateurs **2**
1.1 Les indicateurs de complexité dans le cadre de S2D2 . . . 2
1.2 Les indicateurs basiques . . . 3
1.3 Les indicateurs de Chatterji et Sridhar . . . 4
1.4 Les indicateurs inspirés par Delahaye et Puechmorel . . . 5
1.4.1 La densité . . . 5
1.4.2 La divergence et la convergence . . . 6
1.4.3 La sensibilité . . . 6
1.5 Les indicateurs "utilisés" par RFM pour la PRU . . . 8
1.6 Les indicateurs inspirés de CREED . . . 9
1.6.1 Conclusion sur les indicateurs programmés dans S2D2 . . . 10

2 Les données existantes **11**
2.1 Récapitulatif des besoins du projet S2D2 . . . 11
2.2 Sources possibles de données . . . 11
2.3 Données plans de vol . . . 12
2.4 Données “espace aérien” . . . 13
2.5 Configurations de secteurs . . . 14
2.6 Données radar IMAGE . . . 14

3 Problèmes d’exploitation des données **16**
3.1 Description géographique des secteurs . . . 16
3.2 Centre / zone de qualification . . . 16
3.3 Référentiels et systèmes de coordonnées . . . 16
3.4 Problèmes de référence horaire . . . 17
3.5 « Corrélation »entre trace radar et plan de vol . . . 18
3.6 Ecarts entre trajectoires IMAGE et routes des plans de vol COURAGE . . . 18
3.7 Ecarts temporels entre traces radar IMAGE et plans de vol COURAGE . . . 19
3.8 Tentatives d’échantillonnage des trajectoires IMAGE . . . 20
3.8.1 Objectif . . . 20
3.8.2 Tentative de reconstruction des routes à partir des vecteurs vitesse . . . 20
3.8.3 Tentative de reconstitution des profils verticaux . . . 20
3.8.4 Conclusion sur l’échantillonnage des trajectoires IMAGE . . . 22
3.9 Récapitulatif des problèmes, et propositions . . . 22

4 Choix des données en entrée pour S2D2 **24**
4.1 Choix des données . . . 24
4.2 Choix d’un format intermédiaire . . . 25
4.3 Conclusion sur les données pour S2D2 . . . 25

5 Une analyse descriptive des indicateurs **26**
5.1 Statistiques descriptives 26
5.2 Analyse en composantes principales 26

Introduction

Cette note est un point d'avancement du projet S2D2 (Sector split and dynamic density), mené en collaboration entre le LEEA et le LOG. Elle présente une analyse descriptive d'un certain nombre d'indicateurs recueillis dans la littérature (notamment [1], [2], [3], [4], [5], [6]), dont l'objectif est de traduire la complexité du trafic aérien. Les indicateurs de complexité proposés dans de nombreuses études peuvent être plus ou moins abstraits, plus ou moins pertinents, et chacun ne représente jamais qu'une vision partielle du trafic. Les associer permet de mieux rendre compte de la complexité du trafic. C'est leur combinaison (souvent linéaire) que l'on désigne sous le terme *densité dynamique*.

Parmi la multitude d'indicateurs proposés, nous avons retenu vingt-trois, auxquels nous avons ajouté des indicateurs de flux entrant, avec différents horizons temporels. Les indicateurs sont calculés à partir de données de trafic archivées par les système IMAGE, pour les données radar, et par le système COURAGE, pour les autres données. La présente note est aussi l'occasion de soulever les nombreux problèmes (incohérence, incomplétude, mauvais choix de certaines structures de données), que nous avons rencontrés dans l'exploitation de ces données. Elle peut donc à ce titre dépasser le seul domaine d'intérêt de la complexité du trafic.

Le projet S2D2 cherche à dégager d'éventuelles corrélations entre l'état de regroupement des secteurs aériens sur les positions de contrôle et un certain nombre d'indicateurs, basiques (nombre d'avions, flux entrants), ou plus évolués, qui se veulent représentatifs de la complexité du trafic dans un secteur choisi. Souvent, cette complexité est subjective dans le sens où elle est estimée par les contrôleurs aériens, lors de simulations, ou sur du trafic réel. Certaines expérimentations prennent en compte des indicateurs physiologiques du stress du contrôleur face à un trafic donné. Ces expérimentations pointues sont relativement lourdes à mettre en place, et devraient porter sur un vaste ensemble de secteurs et une population significative de contrôleurs pour avoir une portée générale.

Sans remettre en cause ces démarches, nous proposons d'utiliser une mesure relativement objective, et simple à obtenir, de la charge de travail du contrôleur, à savoir les données archivées des affectations de secteurs aux positions de contrôle. Nous faisons ici l'hypothèse que les décisions de regroupement ou de "dégroupement"[1] de secteurs sont statistiquement significatives de la charge du contrôleur, tout en gardant à l'esprit les biais possibles (ouvertures de secteurs à des fins de formation, relève des équipes, incidents techniques, etc...).

La note est structurée comme suit : en premier lieu les indicateurs retenus sont présentés dans le chapitre 1, puis le chapitre 2 passe en revue les données existantes. Les problèmes rencontrés dans l'exploitation de ces données sont exposés au chapitre 3, et les choix retenus pour les données en entrée du projet S2D2 sont récapitulés dans le chapitre 4. L'ensemble de ces chapitres est relativement détaillé. Il est conseillé au lecteur pressé, intéressé malgrè tout par les problèmes inhérents aux données IMAGE et COURAGE, de s'attarder au moins sur les sections 3.9 et 4.3. Enfin, l'analyse descriptive des indicateurs fait l'objet du chapitre 5, dans laquelle une analyse en composantes principales permet de dégager les corrélations entre les différents indicateurs. La note se termine sur la conclusion et l'évocation des approches envisagées pour l'étude à venir des corrélations entre les données de regroupement de secteurs et les indicateurs.

[1] néolologisme employé pour désigner le fait de séparer un secteur de contrôle en deux secteurs plus petits

Chapitre 1

Présentation des indicateurs

Nous ferons référence par la suite à deux revues de la litterature sur les indicateurs de complexité : celle de Kopardekar [1], qui se focalise sur les études américaines, et celle d'Eurocontrol [2]. Dans l'article [3], qui fait suite à sa revue de la littérature, Kopardekar sélectionne un jeu de variables en régressant des notes de complexité attribuées par des contrôleurs sur du trafic enregistré (puis rejoué). Partir de cette sélection permet de réduire le nombre de variables à reprogrammer. En complément, on utilise également des indicateurs inspirés par l'article [5]. On pourrait bien sûr, dans de futurs développements, intégrer à la liste de variables d'origine le TLI [6], ou tout autre indicateur...

L'étude [2] constate que les indicateurs sont souvent décrits de manière imprécise dans les articles, ce qui rend délicate leur réutilisation. Pour nous être effectivement heurtés à cette difficulté, nous indiquons, pour chaque indicateur, sa définition exacte, ainsi que, le cas échéant, les raisons qui ont amené à des variations par rapport aux définitions recueillies dans la littérature.

1.1 Les indicateurs de complexité dans le cadre de S2D2

Pour le projet S2D2, nous avons recalculé différents indicateurs de complexité dans un module disponible sous CVS : DD_INDIC. Pour le moment, les indicateurs programmés appartiennent à trois catégories :

1. Les indicateurs "simples" : ces indicateurs sont calculés en parcourant l'ensemble des avions actifs. Pour chaque avion, une ou plusieurs variables sont évaluées. Les valeurs obtenues sont ensuite agrégées pour donner l'indicateur considéré.
2. Les indicateurs "doubles" : ces indicateurs sont calculés en parcourant l'ensemble des paires de l'ensemble des avions actifs. Pour chaque couple d'avions, une ou plusieurs variables sont calculées. Les valeurs obtenues sont ensuite agrégées (en deux étapes successives) pour donner l'indicateur considéré.
3. Les flux entrants : nombres d'avions dont l'arrivée dans le secteur se fera dans un laps de temps déterminé (et paramétrable).

Dans les catégories "simples" et "doubles", on retrouve la grande majorité des indicateurs calculés dans les études liées à la densité dynamique. La segmentation du calcul en étapes successives rend les calculs très modulaires et facilite l'implémentation de nouveaux indicateurs. De leur côté, les flux ne sont pas à proprement parler des indicateurs de complexité (ils ne peuvent en aucun cas expliquer la charge de travail instantanée du contrôleur) et à ce titre, ils ne sont pas utilisés pour les formules définissant des densités dynamiques. Pour autant, dans le cadre du projet S2D2, leur utilisation est nécessaire : les décisions de dégroupement/regroupement sont prises en fonction de la prévision de charge pour le secteur, et les flux entrants sont la meilleure information dont disposent les contrôleurs (au travers, notamment, de l'outil PRESAGE).

Les indicateurs calculés dans le module DD_INDIC sont les indicateurs retenus par Kopardekar dans sa revue de littérature, ainsi que d'autres indicateurs dérivés d'études conduites par des membres ou des

équipes de la SDER (ex CENA). Les indicateurs sont brièvement listés dans la table 1.1.

Indicateur	Références[1]	Type
nb	[7] [4] [1] [3]	S
nb2	[1] [3]	S
var_v	[4]	S
nb_des	[8] [7] [4]	S
nb_mon	[8] [7] [4]	S
sqrt_var_sur_moy	[4]	S
flux_5	Gia	F
flux_15	Gia	F
flux_30	Gia	F
flux_60	Gia	F
vprox_1	[4] [1] [3]	D
vprox_2	[4] [1] [3]	D
hprox_1	[4] [1] [3]	D
dens_p	[5]	D
des_ang	[5]	D
des_vit	[5]	D
div_p	[5]	D
conv_p	[5]	D
sensi_vg_d	[5]	D
sensi_hd_d	[5]	D
sensi_vz_d	[5]	D
sensi_vg_c	[5]	D
sensi_hd_c	[5]	D
sensi_vz_c	[5]	D
inter_vert	[9]	S
chgt_niv	[9]	S
inter_hori	[9]	D
ed_bon	[10], [11]	D
efl_bon	[10], [11]	D
ed_pb	[10], [11]	D
efl_pb	[10], [11]	D

TAB. 1.1 – Liste des indicateurs implémentés dans S2D2

1.2 Les indicateurs basiques

Dans cette famille, nous avons regroupé des indicateurs qui sont souvent utilisés dans les articles portant sur la densité dynamique :

- "nb" : nombre d'avions présents dans le secteur.
- "nb2" : carré du nombre d'avions présents dans le secteur.
- "var_v" : variance des vitesses.
- "nb_des" (resp. "nb_mon") : nombre d'avions en descente (resp. montée).
- "sqrt_var_sur_moy" : rapport de l'écart-type des vitesses sur la vitesse moyenne des avions présents dans le secteur.

Tous ces indicateurs sont "simples". Nous y avons ajouté quelques indicateurs de flux entrants :

- "flux_5"
- "flux_15"

– "flux_30"
– "flux_60"

Il s'agit simplement du nombre d'avions entrant dans le secteur entre les instants t et $t + \delta t$, l'horizon temporel δt étant de 5, 15, 30 ou 60 minutes.

1.3 Les indicateurs de Chatterji et Sridhar

L'article [4] définit de nombreux indicateurs originaux. Nous n'avons reprogrammés dans le cadre de S2D2 que ceux qui ont été retenus par Kopardekar dans [3]. Dans cette section, d_{ij} et h_{ij} représentent respectivement les distances horizontales et verticales entre les avions "i" et "j".

– "vprox_1" : cette mesure de proximité verticale s'écrit

$$\texttt{vprox}_1 = \frac{N}{\displaystyle\sum_{1\leq j\leq N}\left(\frac{\displaystyle\sum_{1\leq i\leq N} W_{ij}h_{ij}}{\displaystyle\sum_{1\leq j\leq N} W_{ij}}\right)} , \tag{1.1}$$

où les W_{ij} sont des poids définis par

$$W_{ij} = \begin{cases} \dfrac{1}{d_{ij}^2 + S_h^2 h_{ij}^2} & \text{si } i \neq j \ , \\ \\ 0 & \text{si } i = j \ . \end{cases} \tag{1.2}$$

"vprox_1" peut donc être décrit comme l'inverse de la "distance verticale moyenne (pondérée)". La pondération est inversement proportionelle au carré de la distance elliptique.

– "vprox_2" : cette mesure de proximité verticale s'écrit

$$\texttt{vprox_2} = \frac{\displaystyle\sum_{1\leq i\leq N} [K_i \neq \emptyset]}{\displaystyle\sum_{1\leq i\leq N} \min_{j\in K_i}\{h_{ij}\}} , \tag{1.3}$$

où $K_i = \{j \mid d_{ij} \leq r,\ j \neq i\}$ et r est un paramètre qui définit le voisinage horizontal d'un avion[2]. Le paramètre r utilisé par les auteurs dans [4] n'est pas indiqué. Dans cette étude, nous avons choisi "arbitrairement" $r = 5$ milles nautiques. On peut ainsi décrire "vprox_2" comme l'inverse de la "séparation verticale minimale moyenne pour les couples d'avions qui ne sont pas séparés horizontalement".

– "hprox_1" : cette mesure de proximité horizontale est définie par

$$\texttt{hprox}_1 = \frac{1}{\displaystyle\min_{1\leq i\leq N}\{\min_{j\in J_i}\{d_{ij}\}\}} , \tag{1.4}$$

où $J_i = \{j \mid h_i - \Delta/2 \leq h_j \leq h_i + \Delta/2,\ j \neq i\}$.
Comme dans le cas de r pour "vprox_2", Δ est un paramètre à fixer, et dont la valeur utilisée dans [4] n'est pas connue. Nous avons choisi $\Delta = 2000ft$, de telle sorte que "hprox_1" représente "l'inverse de la plus petite distance horizontale entre deux avions non-séparés vericalement".

[2]En réalité, le numérateur présenté dans [4] n'était pas clair ($[j \in K_i]$, sans somme sur j), et son écriture a été précisée. Sous cette forme, l'indicateur "vprox_2" peut bien être décrit comme la "average minimum vertical separation", comme c'est le cas dans [4].

Etrangement, bien que ces indicateurs aient été utilisés dans (au moins) deux études, ils présentent un défaut qui n'est mentionné nulle part. En effet, **les mesures de proximité verticales "vprox_1" et "vprox_2" peuvent être infinies**. C'est le cas notamment lorsque tous les avions évoluent au même niveau de vol. Avec les échantillons de trafic de taille faible et présentant un nombre relativement élevé d'avions utilisés dans [4] et [3], il est possible que ce problème n'ait pas été identifié ; mais une journée de trafic est suffisante pour le faire apparaître.

Enfin, on peut observer que ces indicateurs ("doubles") sont définis de manière à correspondre à des quantité physiques facilement identifiables (distances horizontales et verticales) et à ne pas croître systématiquement avec le nombre d'avions actifs. On peut ainsi espèrer qu'ils seront faiblement corrélés avec le volume du trafic dans le secteur considéré.

1.4 Les indicateurs inspirés par Delahaye et Puechmorel

L'article [5] présente différents indicateurs "géométriques" de complexité. Dans cette section, on ne retrouvera que ceux d'entre eux qui ne sont pas redondants avec d'autres variables définies dans [3].

Notations
Avant de définir les indicateurs géométriques inspirés par [5] et [12], on introduit quelques notations :

- La distance "elliptique" entre deux avions (i et j) est notée d_{ij}^{ob}. Cette distance correspond à la quantité $\sqrt{d_{ij}^2 + S_h^2 h_{ij}^2}$ utilisée dans la section précédente et dans [4].
- Le vecteur qui représente la distance entre ces deux avions est noté $\overrightarrow{X_i X_j}$ où X_i (resp. X_j) représente la position de l'avion i (resp. j) dans l'espace.
  On a donc $d_{ij}^{ob} = \|\overrightarrow{X_i X_j}\|$.
- La différence des vitesses des avions i et j est définie par $\overrightarrow{v}_{ij} = \overrightarrow{v}_j - \overrightarrow{v}_i$.
- La vitesse d'éloignement des avions i et j est désignée par v_{ij}. Cette vitesse (scalaire) est définie par $v_{ij} = \frac{d}{dt}[d_{ij}^{ob}]$. Elle vaut :

$$v_{ij} = \frac{< \overrightarrow{X_i X_j}, \overrightarrow{v}_{ij} >}{d_{ij}^{ob}} \ . \tag{1.5}$$

Par ailleurs, Delahaye et Puechmorel utilisent une fonction f pour pondérer les paires d'avions (l'agrégation des quantités calculées se faisant par somme pondérée sur l'ensemble des couples d'avions[3]. La fonction proposée est

$$f(d_{ij}^{ob}) = e^{-\alpha d_{ij}^{ob}} \ , \tag{1.6}$$

où α est un coefficient à déterminer. Mais dans [5], aucune valeur de α n'est suggérée. Dans le rapport [12], qui reprend les indicateurs de [5], les poids sont donnés (probablement de manière arbitraire) par

$$f(d_{ij}^{ob}) = \frac{e^{-\alpha (d_{ij}^{ob})^2} + e^{-\beta d_{ij}^{ob}}}{2} \ , \tag{1.7}$$

où $\alpha = 0.002$, $\beta = 0.01$ et où les distances sont exprimées en nautiques. C'est cette fonction que l'on a choisi d'utiliser dans cette étude, mais on pourrait choisir "toute autre" fonction décroissante de la distance.

1.4.1 La densité

La notion de densité diffère de celle utilisée par Kopardekar (pour qui il s'agit du nombre d'avions divisé par le volume du secteur). On parle ici de densité "là où est l'avion" :

$$Dens(i) = \sum_{j=1}^{N} f(d_{ij}^{ob}) \ . \tag{1.8}$$

[3] D'où l'on devine que les indicateurs considérées seront des indicateurs "doubles".

On remarque que la contribution de l'avion considéré est égale à 1. Pour obtenir l'indicateur final, on somme sur les avions du secteur considéré. A priori, on devrait observer une très forte corrélation avec le nombre d'avions dans le secteur (pour la limiter, on pourrait négliger la contribution d'un avion sur lui-même).

1.4.2 La divergence et la convergence

Les indicateurs Div et $Conv$ décrivent respectivement la divergence et la convergence des avions dans le secteur considéré. On retrouve la pondération en fonction des distances entre avions. Les indicateurs sont définis par :

$$Div(i) = \sum_{\substack{j=1 \\ j \neq i}}^{N} 1_{\mathbb{R}^-}(v_{ij}) \,.\, |v_{ij}| \, f(d_{ij}^{ob}) \;, \tag{1.9}$$

$$Conv(i) = \sum_{\substack{j=1 \\ j \neq i}}^{N} 1_{\mathbb{R}^+}(v_{ij}) \,.\, |v_{ij}| \, f(d_{ij}^{ob}) \;. \tag{1.10}$$

La différence entre les deux indicateurs tient donc aux avions pris en compte dans le calcul[4]. On peut penser que l'indicateur $Conv$ sera corrélé avec le nombre de conflits potentiels, bien que la convergence soit une propriété locale qui n'implique pas nécessairement l'apparition d'un conflit "au bout d'un certain temps".
En terme de complexité, l'indicateur Div est sans doute assez peu pertinent : une paire d'avions divergents représente a priori peu de travail pour le contrôleur. Par contre, il est possible que Div soit corrélé avec l'indicateur $Conv$ à un pas de temps antérieur (deux avions qui se seraient croisés divergent ensuite), et donc également avec $Conv$ en raison de la continuité du trafic.

1.4.3 La sensibilité

Il est clair que le nombre de conflits potentiels est une source de charge pour le contrôleur. Mais cette charge est a priori d'autant plus lourde que les conflits sont difficiles à résoudre, cette difficulté pouvant être raccordée à la sensibilité de la vitesse de convergence aux manoeuvres d'évitement possibles (changement de cap ou de vitesse). L'indicateur DOFI, introduit dans [1], essaie d'illustrer cette idée (voir en annexe). Une autre approche est adoptée dans [5], où deux familles d'indicateurs sont introduites. La première (Sd_+ et Sd_-) se base sur la vitesse de rapprochement entre avions, tandis que la seconde (St_+ et St_-) s'appuie la notion (discutable, voir en annexe) de "temps de conflit". Dans cette étude, seuls les indicateurs appartenant à la première famille sont recalculés...

Les indicateurs Sd_+ et Sd_-

Les indicateurs Sd_+ et Sd_- définis dans [5] estiment la sensibilité de la vitesse de rapprochement aux manoeuvres d'évitement. Il s'écrivent

$$Sd_-(i) = \sum_{\substack{j=1 \\ j \neq i}}^{N} 1_{\mathbb{R}^-}(v_{ij}) \, \|\overrightarrow{\nabla} v_{ij}\| f(d_{ij}) \;, \tag{1.11}$$

[4]Dans son code, F. Chatton considère que deux avions sont convergents si la distance a diminué entre les deux derniers états du simulateur. Il est bien plus pertinent de considérer la vitesse de rapprochement instantanée. C'est la méthode de calcul adoptée dans cette étude.

$$Sd_+(i) = \sum_{\substack{j=1 \\ j \neq i}}^{N} 1_{\mathbb{R}^+}(v_{ij}) \, \|\overrightarrow{\nabla} v_{ij}\| f(d_{ij}) \; . \tag{1.12}$$

où le gradient[5] est calculé sur les vitesses-sol v_i, v_j et les caps (définis par rapport au nord) θ_i et θ_j.

A l'origine, ces indicateurs sont définis pour du trafic 2D (d'où l'absence de composante du gradient en "taux de montée"), mais il y a un autre problème caché dans les définitions : en effet, s'il est normal de considérer ces variables, les régulations pouvant être opérées en cap comme en vitesse (et en niveau de vol), toutes les composantes du gradient de la distance n'ont pas la même dimension, et leurs valeurs dépendent du paramétrage des vitesses (i.e. des unités dans lesquelles elles sont exprimées). Il faudrait au moins procéder à une normalisation.

Une solution envisageable consiste à scinder chacun de ces indicateurs en trois indicateurs portant respectivement sur les manoeuvres en vitesse (variables v_i et v_j), en cap (variables θ_i et θ_j) et en niveau de vol (et donc en vitesses verticales). On peut alors normaliser chacun de ces indicateurs suivant les méthodes classiques. La table 1.2 présente les statistiques descriptives des indicateurs ainsi obtenus avant normalisation, et pour des vitesses horizontales et verticales exprimées respectivement en m/s et en ft/min, et des caps exprimés en degrés. Les ordres de grandeurs de ces indicateurs sont, comme on pouvait s'y attendre, très différents, et le gradient "non-pondéré" ne traduirait que la sensibilité en cap.

Indicateur	Min.	1er Qu.	Mediane	Moyenne	3ème Qu.	Max.
"sensi_vg_d"	0	0.019	0.423	0.955	1.34	15.5
"sensi_hd_d"	0	2.42	56.2	179	242	3297
"sensi_vz_d"	0	0.002	0.066	0.161	0.223	2.93
"sensi_vg_c"	0	0.026	0.46	1.00	1.36	23.8
"sensi_hd_c"	0	3.07	60.1	187	247	3362
"sensi_vz_c"	0	0.003	0.074	0.176	0.238	4.22

TAB. 1.2 – Statistiques descriptives pour les sensitivités "par composante"

Pour autant, la séparation des composantes du gradient, si elle permet de tenir compte de toutes les manoeuvres possibles, ne traduit pas la difficulté "globale" de résolution. Quant à la multiplication des indicateurs, elle rend plus délicate la sélection des variables lors de régressions[6], et augmente le caractère subjectif des conclusions d'une analyse en composantes principales (voir la section consacrée à l'ACP dans la suite de cette note). Une somme pondérée des composantes du gradient peut donc représenter une meilleure alternative que la scission de l'indicateur. Se pose alors la question des poids à utiliser, qui n'est pas simple.

En l'absence de données facilement utilisables pour suggérer des poids pertinents, la solution adoptée dans cette note consiste à considérer l'amplitude "normale" des manoeuvres d'évitements. Les valeurs choisies sont :

- 3 points de mach pour les manoeuvres en vitesse (horizontale),
- un taux de montée de 2000 ft/mn pour les manoeuvres verticales,
- un changement de cap de 30°.

Pondérées de cette manière, les moyennes des indicateurs "sensi_vg_d", "sensi_hd_d" et "sensi_vz_d" deviennent respectivement 12.35, 93.7 et 322. Ainsi, on voit que la composante liée à la vitesse sol devient faible (relativement aux deux autres), mais que la différence d'ordre de grandeur entre les autres composantes s'est atténuée, au profit de la sensibilité en taux de montée. Le déséquilibre qui subsiste est relati-

[5]Dans [5], le gradient est écrit sur la distance, mais c'est bien le gradient de la vitesse de rapprochement qui est calculé, ce que l'on constate lorsque les formules sont développées.

[6]C'est d'autant plus le cas ici que les composantes des densités sont très fortement corrélés : les termes diagonaux de la matrice de correlation des composantes de la sensibilité "divergente" sont 0.71, 0.75 et 0.87.

vement conforme aux solutions adoptées en général par les contrôleurs (même si la variabilité individuelle reste importante), au sens où les régulations en vitesse sont assez peu utilisées (au moins pour l'"en route"). La prépondération de la sensibilité en taux de montée sur la sensibilité en cap est plus discutable, et renvoie au choix relativement subjectif des poids. Au final, les indicateurs de sensibilité retenus "sensi_d" et "sensi_c" sont bien calculés à l'aide des gradients "pondérés".

Ces indicateurs de sensibilité, notamment en convergence, présentent une caractéristique déjà relevée dans [5] : une situation convergente "très sensible" est plus favorable qu'une situation convergente "peu sensible". Pour cette raison, une forte valeur de "sensi_c" ne traduira pas forcément une complexité très importante : un conflit très sensible, situation peu complexe, semblera, pour "sensi_c", équivalent à plusieurs conflits peu sensibles, situation complexe. Il faut donc interpréter "sensi_c" en tenant compte des autres indicateurs, et en particulier de "conv" (qui rend compte du nombre conflits). On peut aussi définir un indicateur d'insensibilité en convergence (différent de celui introduit en [5]), "insen_c" (on définit "insen_d" de façon similaire) :

$$\text{"insen_c"} = \frac{Conv^2}{Sd_+} \ . \tag{1.13}$$

Cet indicateur est croissant en "nombre de conflits" et en "insensibilité", et devrait donc bien décrire l'aspect de la complexité pour lequel les indicateurs de sensibilités initiaux avaient été introduits dans [5].

La séparation en deux groupes des avions convergents et divergents appelle la même remarque que celle formulée dans la section précédente : pour la complexité, les avions convergents semblent les plus intéressants, même s'il est possible d'observer une corrélation entre les indicateurs calculés sur les deux groupes d'avions.

Pour conclure sur les indicateurs inspirés par [5], on peut remarquer que l'agrégation des quantités calculées sur les couples d'avions se fait toujours par somme pondérée. Mais contrairement aux poids utilisés dans [4], il n'y a pas dans [5] le souci de décorréler l'indicateur final et le (carré du) nombre d'avions actifs. Les deux approches peuvent se justifier, et c'est la phase de validation des indicateurs qui doit décider de celle à retenir.

1.5 Les indicateurs "utilisés" par RFM pour la PRU

Trois indicateurs ont été calculés en s'inspirant des indicateurs utilisés par RFM pour la PRU (note NT05-153). Les définitions initiales n'ont pas été utilisées, pour une raison précise : les indicateurs de complexité utilisés pour S2D2 (et en général dans toutes les études autour de la charge de travail du contrôleur) portent sur les avions gérés par le contrôleur à l'instant considéré. La dimension temporelle utilisée pour la PRU (en particulier les cellules sont "4D") n'est pas pertinente ici. Dans le cadre de S2D2, il aurait donc fallu considérer les analogues instantanés des indicateurs décrits dans la note NT05-153, à savoir des indicateurs permettant de recalculer, par intégration en temps, les indicateurs d'origine. Bien que (parfois) possible, cette approche n'a pas été retenue dans cette étude. En effet, certaines quantités peuvent être pertinentes pour coller à la réalité de la tarification du contrôle, mais nettement moins quand il s'agit d'évaluer la charge de travail des contrôleurs. Ainsi, la note NT05-153 considère les nautiques parcourus (effectivement à la base de la formule des taxes de contrôle), dont l'analogue "instantané" est la somme des vitesses des avions contrôlés. Mais cette quantité est probablement, "à l'instant t", moins pertinente pour le contrôleur que le nombre d'avions contrôlés[7].

Les indicateurs définis pour la PRU ont donc inspiré les indicateurs suivants :

[7] Ce jugement est soutenu les résultats de [13].

– "inter_hori" (adaptation de l'indicateur "interactions horizontales") : pour un secteur et un instant donnés, il s'agit du rapport du nombre de "croisements" (sans tenir compte du sens de l'avion sur sa trajectoire) horizontaux potentiels d'angle supérieur à 20° sur le nombre d'avions. C'est un indicateur "double".
– "inter_vert" (adaptation de l'indicateur "interactions verticales") : pour un secteur et un instant donnés, on classe les avions en palier/montée/descente pour obtenir $n1$, $n2$ et $n3$. L'indicateur ("simple") est alors donné par

$$\texttt{inter_vert} = \frac{(n1n2 + n2n3 + n1n3)}{(n1 + n2 + n3)} \ . \tag{1.14}$$

– "chgt_niv" (adaptation de l'indicateur "changements de niveaux") : il s'agit de la vitesse verticale moyenne dans le secteur[8]. Il s'agit d'un indicateur "simple".

A noter que l'indicateur "inter_hori" adapté de [9] est extrèmement corrélé (corrélation de 0.94) avec le nombre d'avions "n". Le choix de l'angle de 20° choisi pour sa définition peut être mis en cause, car il ne sélectionne pas assez les paires d'avions en "interaction horizontale". Cette remarque n'est pas, a priori, valable pour [9] : en effet, même si l'indicateur *i*nstantanné calculé pour S2D2 s'inspire de celui introduit dans [9], il est finalement très différent.

Pour être complet sur la note NT05-153, on remarquera qu'à l'échelle du secteur, l'analogue instantané de la "densité ajustée" est simplement le nombre d'avions contrôlés, qui est déjà calculé par ailleurs. Enfin, deux indicateurs définis précédemment (variance des vitesses et ratio écart-type sur vitesse moyenne) décrivent l'hétérogénéité du trafic, qui apparaît dans la note sous la forme de l'écart-type des vitesses. Même si ces quantités ne sont pas strictement identiques, elles sont tellement corrélées que nous n'avons pas recalculé cet indicateur[9].

1.6 Les indicateurs inspirés de CREED

Les derniers indicateurs calculés pour le moment dans le module DD_INDIC sont directement inspirés du travail mené par P. Averty sur la perception des conflits par les contrôleurs aériens [10]. Pour P. Averty, le contrôleur ramène tous les paramètres d'un conflit "potentiel" (les deux avions considérés doivent être en rapprochement horizontal et vertical, et non pas en rapprochement "elliptique" comme pour les indicateurs définis dans [5]) à des distances horizontales. Les variables qui décrivent le conflit sont alors

– "Ed" : la distance horizontale minimale entre les avions.
– "Efl" : la distance horizontale au moment où les avions sont séparés de 1000 pieds (après le croisement vertical).
– "Da" : le "degré d'anticipation", qui est la distance qui sépare le plus rapide des avions du point d'intersection (dans le plan) des trajectoires.

Dans l'étude [10], l'auteur s'intéresse à la perception du risque de conflit. L'optique adoptée ici est un peu différente, puisque c'est la complexité du trafic que l'on cherche à modéliser. S'ils peuvent orienter la définition de l'indicateur de complexité, les résultats de [11] (qui reprend la base de données de [10]) ne pourront donc directement être utilisés ici. En attendant ces résultats, on a défini dans cette étude différents indicateurs qui réutilisent la vision "plan" de [10] : l'idée est que la complexité liée à un conflit potentiel est a priori grande lorsque les avions sont à la fois en convergence forte (donc "Ed" et "Efl" sont faibles), et qu'il reste peu de temps pour éviter le conflit ("DA" faible). Dans le pire des cas, si l'abordage est imminent, la complexité est infinie. D'où la forme initialement imaginée pouir S2D2 :

$$\frac{1}{\alpha\,\texttt{Da} + (1-\alpha)(\beta\,\texttt{Ed} + (1-\beta)\,\texttt{Efl})} \ , \tag{1.15}$$

[8]Là encore, au niveau du contrôleur, on regarderait plutôt le nombre d'avions en évolution verticale. Cet indicateur étant déjà calculé par la famille "var_vit", on a conservé ici la définition originale de la note NT05-153.

[9]Cela étant dit, le calcul est immédiat à partir de la variance des vitesses.

où α et β sont des paramètres compris entre 0 et 1.

Cette définition appelle trois remarques :

- Sous cette forme, β est à comparer au rapport des coefficients associés à "Ed" et "Efl" dans le modèle sans terme croisé de [11]. Pour S2D2, on a décidé par simplicité d'opter pour un β constant et de le choisir arbitrairement. Par ailleurs, une autre difficulté (qui fait l'objet de la seconde remarque) interdit la réutilisation fidèle de l'"index" défini dans [11]...
- En implémentant cet indicateur, on a constaté une difficulté : les situations étudiées dans [10] (et reprises dans [11]), présentent des avions qui se croisent presque perpendiculairement, et ces définitions ne posent pas de problème. Si les angles sont faibles, le point de croisement peut être situé en amont des trajectoires, et le dégré d'anticipation est négatif. Les indicateurs définis ci-dessous sont alors mal définis. Finalement, dans la version actuelle de DD_INDIC, le degré d'anticipation a été remplacé par la plus grande distance d'un des avions au point de séparation (horizontal) minimum, que l'on note ici "DaC". Les quantités "Da" et "DaC" sont représentées en annexe dans deux cas de conflits potentiels, dont l'un met en évidence le problème soulevé ci-dessus pour "Da".
- Dans son étude, P. Averty a évité de présenter au contrôleur des couples d'avions pour les lesquels la séparation verticale - de 1000 pieds - est postérieure au "croisement" des deux avions : ce cas est semble-t-il souvent jugé trop dangereux par les contrôleurs, et est donc résolu dès qu'il est identifié. Pour notre étude, il semble donc logique de traiter ce cas séparément, afin de permettre qu'il se traduise par une plus grande charge de travail pour le contrôleur.

Sous cette forme, β est à comparer au rapport des coefficients associés à "Ed" et "Efl" dans le modèle sans terme croisé de [11].

Les indicateurs de type "CREED" reprennent donc les définitions données en (1.15) en remplaçant "Da" par "DaC". Ils sont ensuite agrégés en sommant sur les conflits "bons" et "mauvais" (suivant qu'ils sont, ou non, du type présentés aux contrôleurs dans [10]). Dans DD_INDIC, ces indicateurs ("doubles")sont notés "creed_bon" et "creed_pb".

1.6.1 Conclusion sur les indicateurs programmés dans S2D2

Dans cette partie, nous avons décrit un certain nombre d'indicateurs de complexité du trafic aérien trouvés dans (ou adaptés de) la littérature. Tous ont été intégrés au module DD_INDIC, en vue de leur utilisation dans le cadre du projet S2D2. Il convient malgré tout de remarquer que des efforts restent à fournir pour affiner certains de ces indicateurs, les paramètres apparaissant dans leurs définitions ayant été fixés (plus ou moins) arbitrairement. Ceci étant dit, on peut sans fin créer de nouveaux indicateurs, il ne s'agit que d'une étape préliminaire devant permettre, in fine, de comprendre et d'expliquer la charge de travail du contrôleur ou la complexité du trafic (au sens du contrôleur). Ce qui passe d'abord par une estimation des indicateurs sur les données de trafic existantes...

Chapitre 2

Les données existantes

2.1 Récapitulatif des besoins du projet S2D2

Comme nous venons de le voir, les indicateurs à calculer sont nombreux et variés. Tous utilisent des informations relatives aux trajectoires des avions dans l'espace aérien. La plupart sont associés à une portion d'espace (secteur élémentaire ou regroupement de secteurs). Afin que l'analyse ait une certaine validité sur le plan statistique, il faudra disposer de données sur une durée suffisamment longue (un mois minimum).

Les informations indispensables pour le projet S2D2 sont, pour chaque avion, et toutes les minutes[1] :

- la position,
- l'altitude
- le secteur élémentaire dans lequel se trouve l'avion,
- le vecteur vitesse,
- le taux de montée/descente,
- éventuellement le type d'avion (calcul d'un indicateur *aircraft mix* ?).

Concernant le regroupement des secteurs sur les positions de contrôle, il est nécessaire de connaître l'état, à chaque instant de la journée, de chaque secteur élémentaire ou regroupement étudié :

- armé sur une position,
- "dégroupé" en n secteurs (pour les regroupements),
- regroupé à l'intérieur d'un secteur plus vaste,
- autre (état inexploitable).

Les informations de type "données radar" permettent de calculer les indicateurs présentés dans le chapitre précédent, que nous pourrons, dans un travail ultérieur, régresser sur l'état de regroupement. Il existe d'autres indicateurs, que nous envisageons de calculer plus tard, faisant appel, pour chaque secteur étudié (élémentaire ou regroupement), et pour chaque avion à l'intérieur du secteur, à la distance entre la position de l'avion et le bord du secteur. Pour ces indicateurs, il faut pouvoir définir les limites géographiques d'un regroupement de secteurs élémentaires. Comme l'analyse pourra porter sur un grand nombre de secteurs, il faudra trouver un algorithme calculant l'enveloppe, éventuellement non convexe, d'un ensemble de secteurs élémentaires. Par ailleurs, certains indicateurs font également appel au volume des secteurs.

2.2 Sources possibles de données

Les configurations de secteurs sont archivées par le système COURAGE, pour chacun des centres de contrôle, à partir des commandes de regroupement/dégroupement entrées par le chef de salle. C'est, à ce jour, notre seule source de données sur l'état de regroupement des secteurs. Les heures de changement de configuration sont données à la minute ronde.

[1] L'état des regroupement de secteurs est connu à la minute près

La description logique des secteurs (partition de l'espace de chaque centre de contrôle), et la description des limites géographiques sont également disponibles dans les données COURAGE.

Pour les informations de trajectoires, on peut envisager plusieurs sources possibles, chacune ayant ses avantages et ses inconvénients :

- la route et les niveaux prévus dans le plan de vol (archives COURAGE),
- les trajectoires recalculées par simulation, à partir des informations du plan de vol,
- les traces radar archivées par le système IMAGE (multi-radars, multi-centres),
- des enregistrements de plots multi-radars provenant du système de traitement radar d'un seul centre,
- des enregistrements de plots mono-radar.

Nous nous sommes focalisés en premier lieu sur les données COURAGE et IMAGE, décrites dans les sections qui suivent.

2.3 Données plans de vol

Les plans de vols proviennent des données archivées par le système COURAGE. Nous nous sommes appuyés sur un extrait d'une note provisoire du STNA ([14]) pour la lecture des fichiers plans de vol.

Il existe trois types de plans de vols archivés : la demande initiale, la demande finale, et le trafic réalisé. C'est ce dernier type de plans de vol que nous utiliserons par la suite. *A priori*, les heures de passage sur balise, ou travers balise, des plans de vols réalisés devraient être actualisés par les STR des différents centres, de même que les heures d'entrée dans les secteurs. Nous le vérifierons en comparant les temps d'entrée secteur des plans de vol réalisés avec les traces radar IMAGE.

Les lignes suivantes sont un extrait d'un fichier de plans de vol COURAGE :

```
00 DEBUT DU FICHIER COURAGE
01 VERSION 4.2
02 01-06-2003
03 15-05-2003
04 7629 7553 7488 7553
06 NCR CRG CRE MOD FLS DES FSH RET DEP SLT SLC EVL FAB FPA FSA FAC ABI IMP EX
05
11
20 JKK3192 EGNV LEPA 5713 -1 MD83 0
21 995 330 442
22 O N N O
31 SITET ETRAT DVL LISEU LGL SORAP BENAR VANAD AMB BALAN BEBIX TUGLI NARAK GA
32 -404 -401 -398 -396 -394 -391 -389 -386 -382 -375 -370 -366 -357 -354 -350
33 330 330 330 330 330 330 330 330 330 330 330 330 330 330 330 330 330 330 330
41 EG XS L2 X2 H2 QV
42 -409 -409 -392 -374 -364 -354
43 -404 -403 -382 -364 -354 -341
44 -333
45 1 2 10 13 14 18
46 1 9 12 13 17 18
13
20 =
21 1335 330 443
22 =
...
```

Les principales informations que nous utiliserons sont, pour chaque vol, la liste des balises de la route

prévue (ligne 31), les heures sur balises (ligne 32), les niveaux de vol prévus sur balises (ligne 33), la liste des secteurs traversés (ligne 41), les heures d'entrée "géographique" dans les secteurs (ligne 43 ; les heures "strips" de la ligne 42 ne sont pas utilisées pour l'instant dans les calculs d'indicateurs), et l'heure de sortie du dernier secteur (ligne 44). Les heures sont exprimées en minutes à partir de 0h.

2.4 Données "espace aérien"

Les données de description des secteurs et des centres de contrôle proviennent également des archives COURAGE.

Les données "centres" décrivent en fait les zones de qualification, et non les centres de contrôle. Voici par exemple quelques lignes extraites du fichier *centres* du 15/05/2003, décrivant la zone de qualification de Paris Ouest, avec la liste des secteurs élémentaires (ligne S), et les regroupements usuels de secteurs (lignes R) :

```
C PARIS-OUEST PW 12 24
S LN OG OT OY RT TB TE TH TN TP UK UZ
R APW 5 LN OT RT TH TP
R FPW 8 LN OT RT TB TE TH TN TP
R HPKZ 5 LN TH TP UK UZ
R NHK 3 LN TH UK
R OGRT 2 OG RT
...
```

Les données "secteurs" décrivent les limites géographiques de chaque secteur, dans un référentiel centré sur Paris (48.833^{o} N, 2.333^{o} E). Les coordonnées (x, y) sont dans un plan de projection stéréographique sur une sphère terrestre de rayon 6375 km (*cf* note provisoire [15]). Les distances dans le plan sont en kilomètres. Les secteurs sont des collections de tubes polygonaux dont les plancher et plafond sont donnés en niveaux de vol.

A titre d'exemple, le secteur TP de Paris (sans précision sur la zone de qualification), qui va du FL55 au FL265 (ligne S), est composé de plusieurs sous-secteurs, dont le premier est entre les niveaux FL55 et FL145 (cf première ligne D), le suivant entre les FL145 et FL195, et ainsi de suite. Chaque ligne commençant par un P donne les coordonnées x et y, en kilomètres, d'un point du polygone d'un sous-secteur.

```
S TP PARI F0061 Terminale Pontoise 00550265 04
D 0055 0145 07
E -65 74
P -108 32
P -76 111
P -52 111
P -47 106
P 2 37
P -77 28
P -91 30
D 0145 0195 12
E -157 76
P -187 40
P -187 51
P -184 133
P -159 147
...
```

2.5 Configurations de secteurs

Les regroupements et dégroupements de secteurs sont obtenus à partir des schémas d'ouvertures réels, archivés par le système COURAGE. Le détail du format des données est disponible dans [16] (extrait), et sans doute dans d'autres notes plus récentes de la DTI/SDSO (ex STNA).

Les principales informations contenues dans ces fichiers sont :

– un intervalle temporel, avec l'heure de début et de fin en minutes entières à partir de 0h, et,
– pour chaque position de contrôle armée, la liste des secteurs qui lui sont affectés.

Voici, à titre d'exemple, un extrait d'un fichier d'archive des regroupements :

```
00 Schema d'ouverture du centre de PARIS-OUEST pour la journee du 01-06-2003
01 Version 1.0
02 01-06-03
03 15-05-03
04 14
05 PW PARIS-OUEST
06 REALISE
10 ????? -187 312 1 1
11 PTE RPAW -1 12 TE TN TB TP TH LN UK UZ OG OY OT RT
10 ????? 312 348 2 2
11 PTE TENB -1 3 TE TN TB
11 PTP OPKZ -1 9 TP TH LN UK UZ OG OY OT RT
10 ????? 348 364 3 3
11 PTP HPKZ -1 5 TP TH LN UK UZ
11 POG ORGY -1 4 OT RT OG OY
11 PTE TENB -1 3 TE TN TB
...
```

2.6 Données radar IMAGE

Les données radar IMAGE sont élaborées à partir des images radar des cinq centres de contrôle en-route français. Nous ne disposons que d'un plot toutes les trois minutes environ, mais de façon non régulière : les plots ne sont pas à des heures fixes et l'intervalle des trois minutes n'est pas régulier.

Les traces radar IMAGE sont obtenues par simple concaténation des traces issues des systèmes de traitement radar (STR) différents centres. On constate, sur les exemples de la figure 2.1, qu'il peut exister, dans la même trace IMAGE, des positions très proches, en position et en temps, provenant de centres différents.

Une description des données IMAGE a été trouvée dans deux documents : [17], et [18]. Le format des données diffère selon les documents. Il semble que les données envoyées par les STR subissent une transformation avant archivage par le système IMAGE. Les informations sur les unités utilisées sont la plupart du temps incomplètes, et fragmentées entre les diverses documentations. Le format des données IMAGE archivées dont nous disposons est celui décrit dans [17], mais certains détails sur les unités ne sont donnés que dans [18]. Le tableau 2.2 récapitule les informations contenues dans les données IMAGE, avec leur unités.

Attention : les coordonnées sont obtenues par projection stéréographique sur un plan tangent à un ellipsoïde de Clarke, 1880, (d'après [17]). Le plan de projection est centré sur le point (47 N, 0 E). Ce n'est pas le même référentiel que pour les données COURAGE.

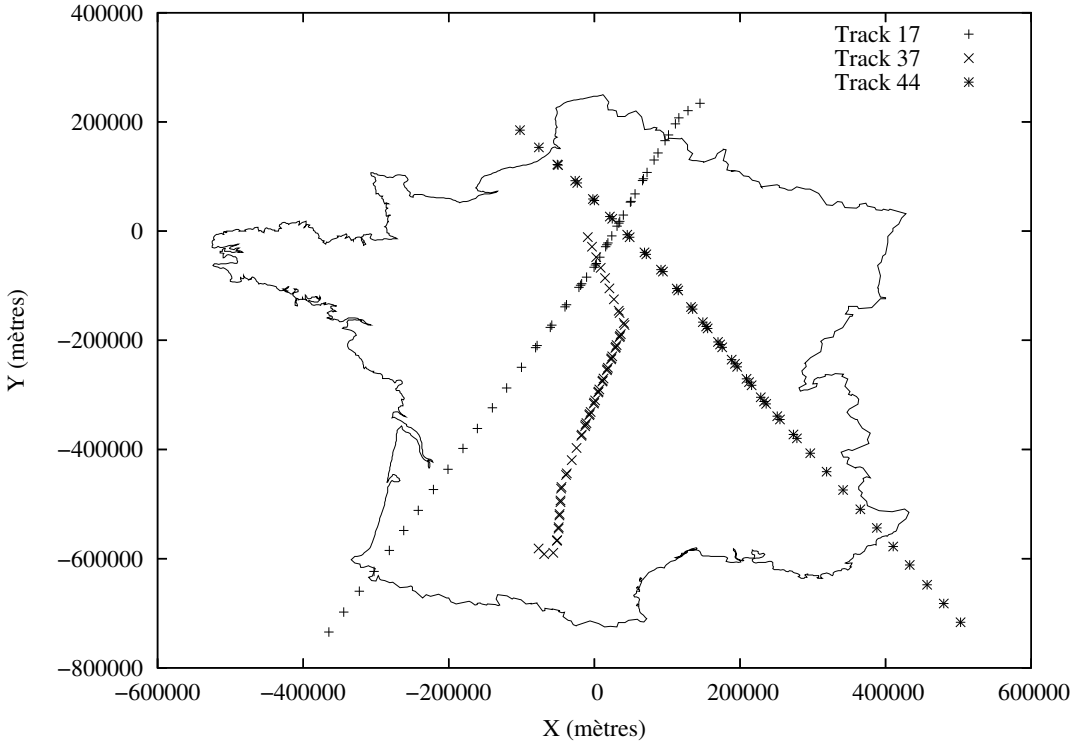

FIG. 2.1 – Exemples de traces radar IMAGE

Donnée	**Unité**
time	secondes, à partir de 0h, d'après [17]
mode A	-
latitude	degrés, d'après [17]
longitude	degrés, d'après [17]
speed_X	$1/16384^{\text{ième}}$ de NM/s (soit $0,22$ kts), d'après [18]
speed_Y	$1/16384^{\text{ième}}$ de NM/s (soit $0,22$ kts), d'après [18]
track_number	-
FL	Unité non spécifiée dans [17] . A priori ce sont des niveaux de vol (mais dans [18], le mode C est donné par incréments de 25 ft)
rate_cl_desc	$1/1024^{\text{ième}}$ de FL/s (soit $5,86$ ft/mn), d'après [18]
acft_type	-
dep_airport	-
arr_airport	-
callsign	-
sl_nb	-
sl_crossed	-

FIG. 2.2 – Données IMAGE, avec leurs unités

Chapitre 3

Problèmes d'exploitation des données

3.1 Description géographique des secteurs

Les limites géographiques des secteurs ne sont pas données sous forme de graphe, c'est-à-dire sous forme de listes de noeuds et d'arêtes, chaque arête reliant deux noeuds, et délimitant une portion de la frontière entre deux secteurs.

Les secteurs élémentaires sont décrits sous forme de collections de polygones indépendants, et on constate en pratique qu'ils ne forment pas une partition de l'espace aérien, alors qu'ils le devraient. Les arêtes des polygones délimitant des secteurs adjacents ne sont pas toujours exactement alignées, et il existe donc des "trous" entre certains secteurs, ainsi que des portions d'espace dans lesquels plusieurs secteurs élémentaires se recouvrent.

Ceci amène de nombreuses difficultés, que ce soit pour déterminer dans quel(s) secteur(s) se trouve un point donné, ou encore pour délimiter géographiquement un regroupement de secteurs (en particulier lorqu'il n'est pas convexe).

Une position donnée peut se trouver dans zéro, un, ou plusieurs secteur(s) élémentaire(s), alors que si les secteurs formaient une partition de l'espace, elle ne pourrait être que dans un seul secteur, ou sur une frontière. L'algorithme que nous utilisons renvoie une liste des secteurs d'appartenance. Pour l'instant, nous considérerons que la position est dans le premier secteur de la liste, si celle-ci n'est pas vide. Un raffinement ultérieur pourra consister à effectuer un filtrage sur plusieurs positions consécutives, afin de mieux déterminer le secteur d'appartenance.

Pour le calcul de la frontière d'un regroupement de secteurs (nécessaire pour pouvoir calculer la distance d'un avion à cette frontière), nous envisageons d'implanter un algorithme relativement complexe (expansion vectorielle de polygones, fusion de polygones, et réduction).

3.2 Centre / zone de qualification

Les fichiers *centres* décrivent en fait les zones de qualification (Paris/Est et Paris/Ouest, par exemple), et non les centres de contrôle (Paris,...). Les fichiers *secteurs* font référence au centre d'appartenance, mais pas à la zone de qualification.

Il s'agit d'un problème mineur, facilement contournable en écrivant un code un peu moins propre que ce qu'il pourrait être.

3.3 Référentiels et systèmes de coordonnées

Les données de position archivées par COURAGE et par IMAGE ne sont pas dans les mêmes systèmes de coordonnées.

Au vu des formules de projection données dans [19] et [15], les coordonnées (x, y) dans le système COURAGE sont celles d'une projection stéréographique *sur une sphère*. Il s'agit en effet de formules de

projection simplifiées, permettant d'effectuer la projection inverse sans itérations. Le rayon de la sphère diffère selon les sources : 6375 km dans [15], ou 6378 km dans [19]. Le point tangent du plan de projection à la sphère est (48.833° N, 2.333° E), coordonnées correspondant approximativement à Paris, selon [15], et seulement « Paris »dans [19]. La description la plus complète (celle de [15]) a été retenue.

Concernant les traces radar IMAGE, il est indiqué dans [17] que les coordonnées des positions sont obtenues par une projection stéréographique, en utilisant un ellipsoïde de Clarke (1880), avec un centre de projection au (47 N, 0 E). Cependant, les positions sont données en latitude et longitude, et non en coordonnées (x, y). Seules les vitesses sont données en coordonnées cartésiennes. Par la suite, on supposera donc que les positions sont exprimées en latitude et longitude *sur l'ellipsoïde de Clarke (1880)*, et les vitesses en coordonnées cartésiennes dans le plan de projection. Les formules de transformation de coordonnées ne sont pas fournies dans [17], ni [18]. Nous avons implanté celles décrites dans le guide de l'*European Petroleum Survey Group* ([20]).

3.4 Problèmes de référence horaire

```
05
11
20 FUA1138 EGAA LEPA 8797 0 B738 0
...
31 BHD SKESO GUR DIN GODAN TERPO NTS TUPAR CGC VELIN SECHE AGN TOU REVEL ROCA
32 1314 1319 1322 1330 1338 1340 1342 1350 1356 1362 1367 1371 1374 1378 1384
41 S6 JU NI C2 X2 N2 H2 QV
42 1309 1309 1323 1339 1353 1362 1364 1373
43 1314 1317 1333 1351 1363 1372 1374 1386
44 1394
...
13
...
31 =
32 1385 1390 1393 1401 1408 1410 1413 1420 1427 1433 1438 2 5 9 15 17 26
...
41 =
42 1380 1380 1394 1410 1424 1433 1435 4
43 1385 1388 1404 1422 1434 3 5 17
44 26
...
12
20 FUA1138 EGAA LEPA 8797 0 B738 2271
21 1399 350 462
22 =
31 =
32 1394 1399 1402 1409 1416 1418
...
41 S6 JU NU C3 X3 N2 H2 QV
42 1389 1389 1402 1415 1429 1438 1440 1449
43 1394 1397 1412 1427 1439 1448 1450 1462
44 1471
...
```

On constate un certain flou sur le choix de l'archivage des heures, y compris à l'intérieur d'un même

format de données. A titre d'exemple, comparons les lignes 43 (heures "géographiques" d'entrée secteur), du plan de vol initial, final, et réalisé, du FUA1138, le 27/06/2003. Dans la demande initiale (bloc 11), toutes les heures sont inférieures à minuit (1440 minutes). Dans la demande finale (bloc 13), le vol est un peu retardé, et, en conséquence, certaines heures d'entrée secteur sont dans la journée suivante, après minuit. Ces heures sont (implicitement) comptabilisées à nouveau à partir de 0, mais à partir des 0h de la journée suivante. Par contre, dans le plan de vol réalisé, les heures au-delà de minuit sont représentées par des nombres supérieur à 1440, ce qui est le plus logique, dans la mesure où cela permet de déterminer clairement que ces heures sont dans la journée suivante.

3.5 « Corrélation »entre trace radar et plan de vol

Les traces radar IMAGE contiennent un certain nombre d'informations sur le vol : l'indicatif, le type d'avion, et les aéroports de départ et de destination. Ces informations permettent de retrouver, dans les archives COURAGE, le plan de vol correspondant à chaque trace radar.

Malheureusement, nous avons constaté qu'il peut y avoir dans une même journée plusieurs plans de vol avec le même indicatif, le même type d'avion, et les mêmes départ et destination. Pour lever l'incertitude, un plan de vol ne sera corrélé à une trace radar que si la différence entre l'heure sur la première balise et l'heure du premier plot est inférieure à 1 heure.

3.6 Ecarts entre trajectoires IMAGE et routes des plans de vol COURAGE

En comparant les routes du plan de vol et celles effectivement suivie par les avions, on constate de nombreux écarts. La figure 3.1 illustre ces écarts, en reprenant les exemples de la figure 2.1.

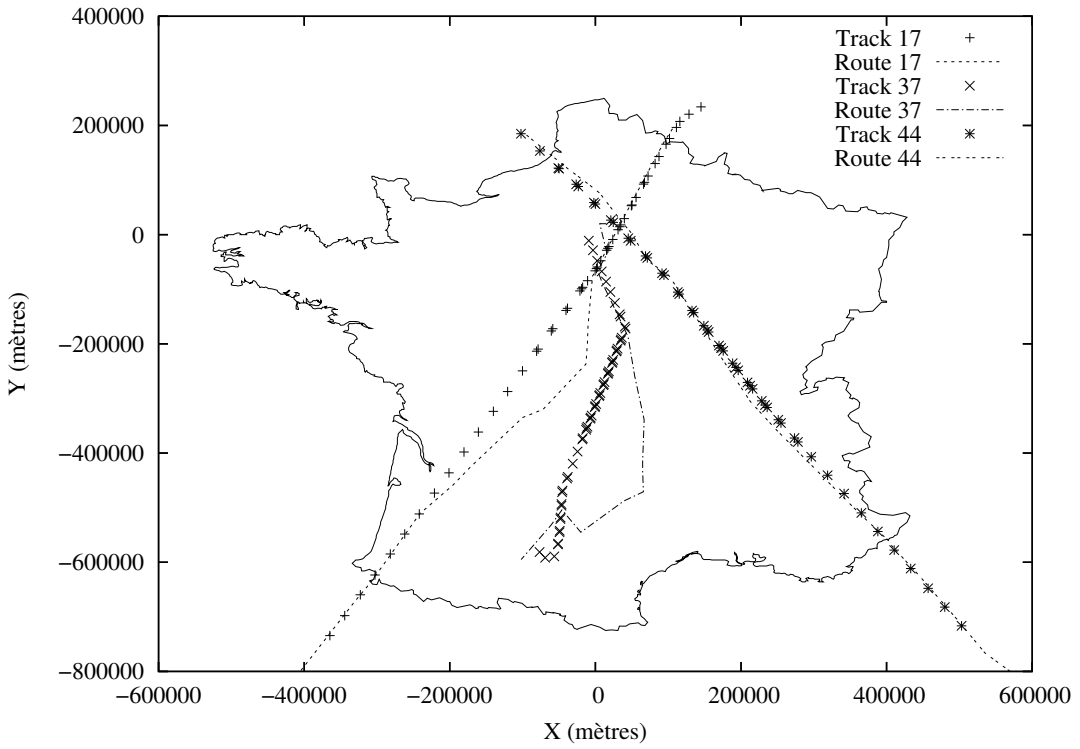

FIG. 3.1 – Exemple d'écarts entre traces radar IMAGE et route PLN

Les trajectoires réellement empruntées par les avions sont dans l'ensemble plus directes que celles initialement prévues dans le plan de vol. Les routes n'étant pas remises à jour en fonction des clairances des contrôleurs, il est normal et attendu qu'il existe des différences.

Ceci interdit toute utilisation de la route du plan de vol dans l'élaboration d'un vecteur vitesse, que ce soit pour calculer un indicateur de complexité, ou bien pour échantillonner les trajectoires avec un pas de temps inférieur à 3 minutes.

3.7 Ecarts temporels entre traces radar IMAGE et plans de vol COURAGE

Dans cette section, nous allons comparer les heures d'entrée secteur archivées dans les plans de vols réalisés (COURAGE), et les heures des plots radar IMAGE juste avant et juste après le passage d'une frontière de secteur.

Les heures des plans de vol réalisées sont normalement remises à jour par les systèmes de traitement radar. En conséquence, les différences entre l'heure PLN d'entrée dans un secteur, et les heures des plots IMAGE au passage de la frontière du secteur (dernier plot dans le secteur précédent, et premier plot dans le secteur) devraient logiquement être inférieures à l'écart maximum entre deux plots IMAGE.

En analysant les 8428 traces radar IMAGE que nous avons pu corréler avec un plan de vol de la journée du 27 juin 2003 (156 traces radar n'ont pu être corrélées), nous avons recensé 31483 transitions d'un secteur à un autre, dont 391 concernaient un plot radar de localisation douteuse (plot localisé simultanément dans plusieurs secteurs : *cf* le problème de la description géographique des secteurs), que nous avons laissé de côté.

Parmi les 31092 transitions restantes, 6626 étaient incohérentes (soit $21,3\%$) avec la liste des secteurs traversés, archivée dans le plan de vol. Pour 15934 transitions, soit $51,2\%$, l'heure d'entrée secteur donnée par le plan de vol n'était pas dans l'intervalle temporel défini par le dernier plot IMAGE dans le secteur précédent et le premier plot IMAGE dans le secteur. Nous avons donc seulement 8532 transitions d'un secteur à un autre qui sont à la fois cohérentes avec les secteurs du plan de vol, et pour lesquels les heures d'entrée secteur archivées sont relativement correctes. Cela ne représente que $27,4\%$ des transitions.

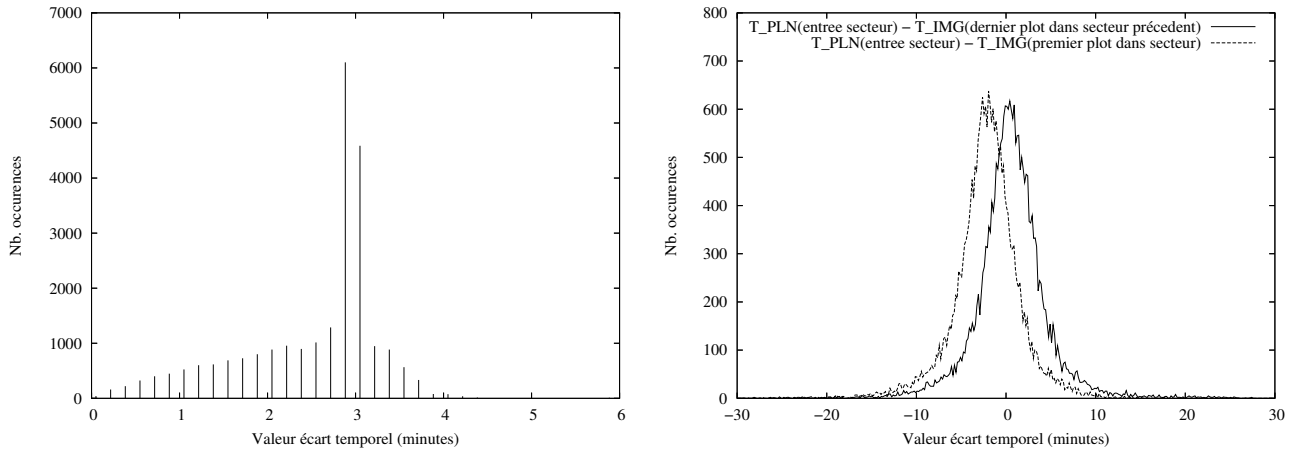

FIG. 3.2 – Distribution des écarts temporels, entre premier plot IMAGE dans le secteur et dernier plot IMAGE dans le secteur précédent (à gauche), ou entre heure d'entrée secteur PLN et heures des plots IMAGE (à droite)

Pour mieux détailler les écarts temporels, nous avons calculé d'une part la distribution des écarts entre plots IMAGE successifs, au moment des passages d'une frontière de secteur, et d'autre part, les distributions des écarts entre l'heure d'entrée secteur archivée dans le plan de vol et les heures des plots IMAGE, avant ou après passage de la frontière du secteur. Les distributions sont calculées en considérant des intervalles d'une largeur de 10 secondes, et en comptabilisant le nombre d'occurences pour lesquelles la valeur de l'écart temporel est comprise dans cet intervalle.

La partie gauche de la figure 3.2 montre, pour la journée du 27 juin 2003, la distribution des écarts entre plots IMAGE successifs au passage d'une frontière de secteur. Cette distribution présente un pic autour des 3 minutes, et l'écart n'excède que très rarement les 4 minutes.

La partie droite de la figure 3.2 montre deux courbes de distribution : l'une donnant les écarts temporels entre l'heure PLN[1] d'entrée dans le secteur et l'heure du dernier plot IMAGE dans le secteur précédent, et l'autre donnant les écarts entre l'heure PLN d'entrée secteur et l'heure du premier plot IMAGE dans le secteur. Les courbes de ces distributions sont relativement symétriques : les écarts sont également distribués autour d'une valeur moyenne. Pour la première, la moyenne des écarts est de $0,64$ minutes, et l'écart-type de $4,28$ minutes. La moyenne de la deuxième est de $-2,08$ minutes, avec un écart-type de $3,96$ minutes. Un calcul des quantiles sur ces distributions montre qu'il faut prendre une plage temporelle de 10 à 12 minutes autour des valeurs moyennes pour contenir 90% des occurences.

Ces distributions sont donc plus étalées que ce que l'on aurait pu attendre, sans que l'on ait pour l'instant d'explication plausible. Parmi les pistes possibles, citons l'algorithme de mise à jour des heures STPV par le STR, ou encore la nature des heures associées aux plots IMAGE : s'agit-il des heures auxquelles le système IMAGE récupère les plots fournis par les STR des différents centres, avec éventuellement un décalage par rapport à l'heure réélle du plot radar ? Ces points restent à élucider.

3.8 Tentatives d'échantillonnage des trajectoires IMAGE

3.8.1 Objectif

On souhaite calculer les différents indicateurs de complexité toutes les minutes, à la minute ronde. Or, la fréquence des plots IMAGE est plus proche des 3 minutes en moyenne, et les heures des plots ne tombent pas à la minute pile. L'objectif de l'échantillonage est donc de reconstruire les trajectoires avec un point toutes les minutes à partir d'une heure fixée.

En 3 minutes, un avion évoluant à 480 kts parcourt 24 NM. Pendant cette periode, les changements de cap, ou d'altitude, ou les variations de vitesse, peuvent amener l'avion à des positions assez éloignées de celles que fournirait une simple interpolation linéaire entre deux positions IMAGE connues.

C'est pourquoi nous allons tenter d'effectuer des interpolations un peu plus fines, en utilisant les informations de direction du vecteur vitesse pour échantillonner la route, et un modèle de performances avion pour les évolutions verticales.

3.8.2 Tentative de reconstruction des routes à partir des vecteurs vitesse

L'idée est d'essayer de détecter un éventuel virage ayant eu lieu entre deux points IMAGE successifs en examinant les directions des vecteurs vitesses associés à ces deux points. Si l'écart est supérieur à une valeur donnée, un nouveau point est calculé, par l'intersection des deux droites issues des vecteurs vitesses. Si ce nouveau point est situé entre les deux plots IMAGE, il est rajouté à la route.

La figure 3.3 montre le résultat de cette tentative, sur un exemple de trajectoire, en choisissant une valeur seuil de 30 degrés pour l'angle entre deux vecteurs vitesses successifs. On constate que la trajectoire reconstruite semble plus erratique que celle obtenue en reliant simplement les plots IMAGE.

C'est aussi le cas pour de nombreuses autres trajectoires de la même journée. Il semble donc que l'information de vecteur vitesse ne soit pas suffisante, ou pas suffisamment précise, pour reconstruire une approximation réaliste de la route suivie.

3.8.3 Tentative de reconstitution des profils verticaux

Sachant que les avions ne suivent pas des pentes linéaires de montée ou de descente, nous allons tenter d'utiliser le modèle tabulé de performances (BADA) des avions, afin de reconstituer des profils de montée ou de descente permettant d'interpoler des points intermédiaires avec plus de précision. C'est du moins ce que l'on espère.

[1] PLN : plan de vol

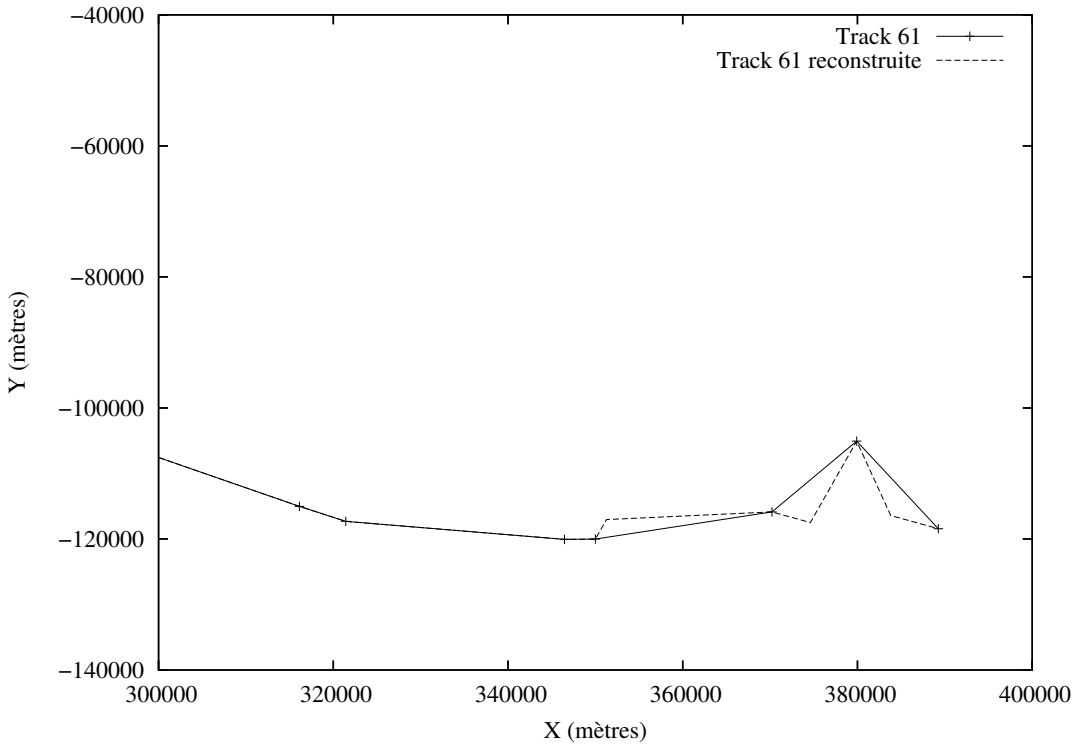

FIG. 3.3 – Exemple de reconstruction de route à partir des vecteurs vitesse

Une des difficultés de ce calcul est que les plots réels ne suivent généralement pas le profil nominal fourni par le modèle. Nous adoptons ici une méthode simple, permettant de reporter de façon linéaire, sur les points intermédiaires du profil, les écarts constatés entre le profil nominal et les plots IMAGE.

Pour chaque couple de plots IMAGE consécutifs, on compare l'écart d'altitude avec un seuil (fixé à 200 pieds). Si l'écart est inférieur ou égal au seuil, on admet qu'il s'agit de simples variations d'altitude dans la tenue d'un niveau de croisière. Dans le cas contraire, on considère les deux points comme faisant partie d'un segment d'évolution verticale, de montée ou de descente.

Dans ce dernier cas, on calcule, sur la distance séparant les deux points et en partant du premier plot IMAGE, le profil nominal d'évolution verticale, selon le modèle tabulé de performances avions. Puis, on calcule l'écart d'altitude entre le dernier point du profil nominal et le deuxième plot IMAGE. Cet écart est ensuite rabattu linéairement, en fonction de la distance à partir du premier plot, sur tous les points intermédiaires du profil nominal.

Malheureusement, dans un certain nombre de cas, cette méthode simple produit des profils aberrants, tel celui illustré sur la figure 3.4.

Cet exemple cumule plusieurs types de problèmes. Tout d'abord, les deux plots consécutifs proches, provenant de deux systèmes de traitement radar différents, situés au milieu de la descente entre le FL360 et le FL340, introduisent une petite étape en palier, qui peut être artificielle, ou pas. En regardant le détail des données IMAGE, on constate que le taux d'évolution verticale des deux plots au FL350 est nul. Il y a donc bien eu, dans ce cas précis, un palier intermédiaire au FL350.

On ne dispose finalement d'aucun plot IMAGE appartenant aux deux segments de descente entre le FL360 et le FL350, mais seulement de trois plots successifs appartenant à trois paliers différents. Il n'y a donc aucune chance, sur ce type de profil, de « caler »convenablement les segments évolutifs.

Par ailleurs, le report de l'erreur sur le profil nominal donne ici des profils peu réalistes. Seul le profil du dernier segment de descente pourrait éventuellement être corrigé, en considérant les vitesses verticales des plots : seul le dernier plot appartient effectivement à la descente, et l'on pourrait « caler »un profil nominal (ou un autre ?) sur cet unique point, et recalculer le *top of descent*.

Même si quelques améliorations sont possibles, on constate que la fréquence de renouvellement des

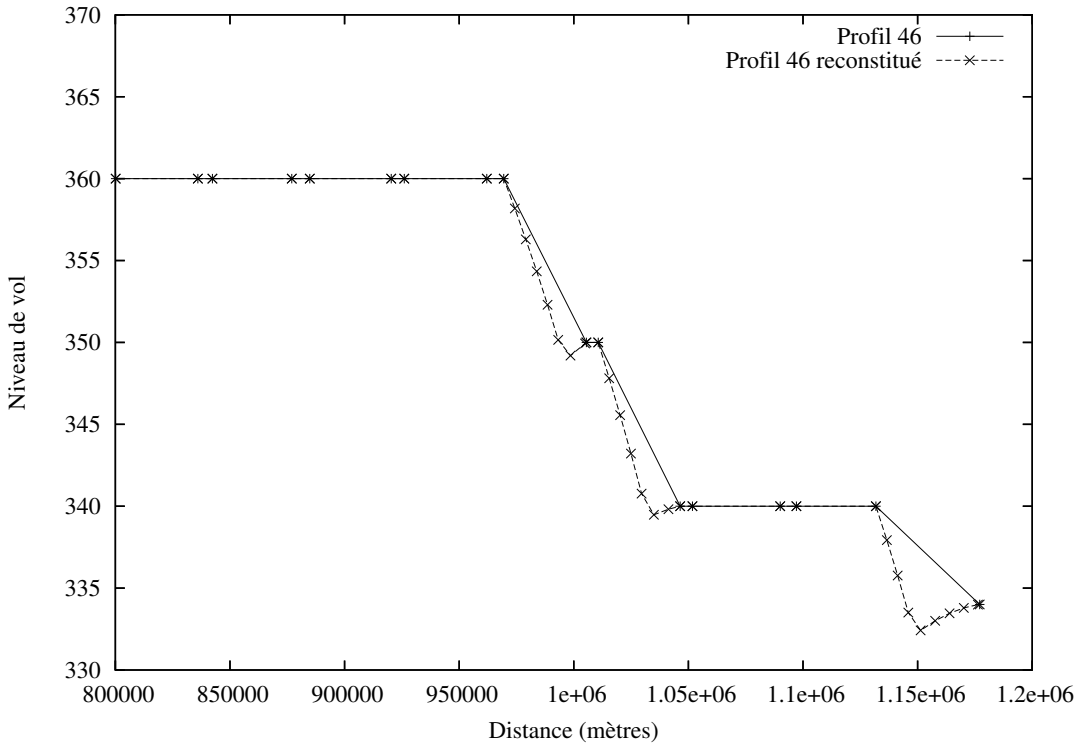

FIG. 3.4 – Exemple de reconstitution aberrante de profil vertical

plots IMAGE n'est pas suffisante pour reconstituer un profil réaliste. En l'état actuel des choses, on se contentera donc des profils verticaux obtenus en reliant les plots IMAGE par des segments de droite.

3.8.4 Conclusion sur l'échantillonnage des trajectoires IMAGE

En conclusion, nos tentatives de reconstitution fine, à partir des plots IMAGE toutes les 3 minutes, d'une route et d'un profil vertical, ont échoué. Au passage, on peut s'interroger sur l'utilisabilité des informations de vecteur vitesse, du moins en direction, contenues dans les plots IMAGE.

Pour obtenir des trajectoires échantillonnées toutes les minutes, on se contentera donc d'interpoler linéairement les informations nécessaires à partir des plots IMAGE, tout en gardant à l'esprit que cela peut introduire un certain biais sur le calcul des indicateurs.

Un travail ultérieur pourra consister à recalculer ses indicateurs sur des données radar d'un seul centre, plus précises (fréquence 12 s) mais disponibles sur peu de journées, et à évaluer les biais introduits par les interpolations des données IMAGE.

3.9 Récapitulatif des problèmes, et propositions

Ce travail a été l'occasion de faire quelques constatations et de lever un certain nombre d'interrogations sur les données, parfois très indirectement liées à notre problème initial de calcul d'indicateurs.

Concernant la description de l'espace aérien, il y aurait grand bénéfice à revoir la description géographique des secteurs, en la séparant en deux fichiers : l'un contenant des points, avec un identifiant et la localisation géographique, et l'autre décrivant les polygones des secteurs comme une suite d'arêtes, chaque arête étant définie par deux identifiants de points. Cela réglerait les imperfections existantes dans le partitionnement de l'espace (recouvrement de secteurs, zones « blanches »couvertes par aucun secteur), et faciliterait le calcul des frontières d'un regroupement de secteur (et accessoirement la détermination du voisinage d'un secteur). En l'état actuel des choses, il reste à développer un algorithme relativement complexe (expansion et fusion de polygones, puis réduction), afin d'obtenir l'enveloppe, souvent non convexe, de plusieurs secteurs.

Les données plan de vol archivées par COURAGE n'adoptent pas un format homogène pour les heures. Il serait logique d'adopter une même convention au sein d'un même format de données, par exemple en comptant négativement les heures dans la journée précédente, si besoin est, et en ayant toujours des suites croissantes d'heures pour une suite d'événements successifs, en évitant de repartir de 0 lorsqu'on passe dans la journée suivante, quitte à aller au delà des 1440 minutes contenues dans une journée.

Concernant les données IMAGE, la fréquence de renouvellement des plots (1 toutes les 3 minutes, environ) est insuffisante, en particulier pour retracer les évolutions verticales des avions. Les informations de direction du vecteur vitesse ne semblent pas très fiables, sans que l'on en connaisse la raison pour l'instant. Cela reste à confirmer, par exemple en examinant la dispersion des vecteurs vitesse sur les portions de vol en ligne droite.

La comparaison des données IMAGE et des plans de vols COURAGE fait apparaître deux phénomènes : d'une part des différences entre routes effectivement suivies et routes des plans de vol, ce à quoi l'on pouvait s'attendre, et d'autre part, une distribution relativement étalée des écarts temporels sur les heures d'entrée secteurs, et donc avec des écarts parfois importants. Sachant que les heures des plans de vols COURAGE réalisés sont normalement mises à jour par les systèmes de traitement radar des différents centres, il s'agit là d'une anomalie, qui reste sujette à investigation. Dans un premier temps, il est envisagé d'utiliser une autre source de données radar, pour confirmer ce phénomène.

Chapitre 4

Choix des données en entrée pour S2D2

4.1 Choix des données

Les plans de vols ne contiennent aucune indication de vitesse ou de taux d'évolution, et seules les positions sur les balises de la route sont connues. Ils ne sont donc pas exploitables tels quels. Une simulation permettrait de recalculer les positions et les informations utiles, toutes les minutes. En tout état de cause, les plans de vol ou les trajectoires simulées ne sont exploitables qu'à deux conditions : d'une part, que les trajectoires réelles soient suffisamment proches, en position et en altitude, des trajectoires prévues, et d'autre part, que les estimées temporelles soient suffisamment justes.

En théorie, les heures, et peut-être les niveaux de vol, associés aux balises du plan de vol « réalisé »devraient avoir été remises à jour par les STR[1] des différents centres de contrôle, et devraient donc être réalistes. Par contre, comme nous le verrons un peu plus loin, les trajectoires réelles ne sont pas toujours très proches des trajectoires prévues dans le plan de vol.

Les données multi-radar et multi-centres du système IMAGE contiennent la plupart des informations nécessaires, mais avec seulement un plot toutes les trois minutes environ. Les traces radar sont reconstruites à partir des images radar synthétiques produites par les STR des différents centres. Cette source de données présente l'avantage de couvrir tout l'espace français, et d'éliminer un certain nombre de traces radar (VFR, militaires) n'intéressant pas le contrôle en-route. L'inconvénient est qu'il faut reconstituer les trajectoires, à partir des traces IMAGE, afin d'obtenir les informations qui nous intéressent avec la fréquence voulue (toutes les minutes), ce qui introduit un biais dans le calcul des indicateurs.

Les plots mono-radar, et les plots multi-radars d'un seul centre, sont enregistrés toutes les 12 secondes environ, et il n'y a donc pas de problème de reconstruction de trajectoire. Par contre, l'enregistrement d'un seul radar ne couvre qu'une portion de l'espace aérien, et contient des données brutes, non lissées, ceci pour tous les vols ayant un transpondeur mode A/C, y compris les VFR et les vols militaires.

Par ailleurs, ces données ne sont pas disponibles sur une longue période. Nous disposons de quelques enregistrements mono-radar sur Paris (5, 6, 7, et 9 juin 2003), retravaillés par la SDER/RAD, avec des informations lissées pour les coordonnées (x, y) dans le repère CAUTRA, le cap, la vitesse, l'altitude, et la vitesse verticale. Une journée complète de plots multi-radars, du centre de Brest, est également disponible.

Au vu des différents éléments qui viennent d'être évoqués, nous avons décidé d'utiliser en premier lieu les données IMAGE, en cherchant à reconstruire des trajectoires et des profils de vol, échantillonnés toutes les minutes. Ces données sont plus proches de la réalité que les données plan de vol, et elles couvrent une zone géographique suffisamment grande, sur une période suffisamment longue : elles sont archivées chaque jour. Dans un deuxième temps, les données mono-radar, ou multi-radars d'un seul centre, pourraient être exploitées pour valider le calcul des indicateurs avec les données IMAGE.

[1] STR : Système de Traitement Radar.

4.2 Choix d'un format intermédiaire

Afin de faciliter un travail ultérieur sur d'autres sources de données (mono-radar ou multi-radar mono-centre), on introduit un format intermédiaire de données, contenant :

- la position (x et y en mètres)
- l'altitude en pieds
- la distance parcourue depuis le premier plot, en mètres
- le temps, en minutes depuis 0h
- la vitesse (Vx et Vy en noeuds)
- le taux de montee/descente en pieds par minute

Les données de positions et de vitesse sont dans le même référentiel que les données COURAGE.

4.3 Conclusion sur les données pour S2D2

Par rapport aux besoins exprimés pour le calcul d'indicateurs de complexité, il est actuellement possible de produire les données demandées, à partir des données COURAGE et IMAGE, sauf celles du type "distance à la frontière d'un regroupement de secteurs". Comme nous l'avons vu précedemment, cette donnée n'est pas simple à calculer, à cause de la mauvaise description géographique des secteurs élémentaires. Celle-ci rend délicate la détermination des frontières externes d'un groupe de secteurs. Il est envisagé de développer un algorithme en trois étapes : expansion des polygones définissant les frontières des secteurs élémentaires, puis fusion des polygones, et enfin réduction du polygone obtenu. Il reste donc un travail assez important à réaliser avant de pouvoir calculer l'indicateur basé sur la distance au bord du secteur.

Concernant la fréquence à laquelle les indicateurs sont calculés, nous avons choisi un taux de renouvellement d'une fois par minute. Les ouvertures de secteurs n'étant archivées que toutes les minutes, il n'a pas paru utile de calculer les indicateurs avec une fréquence plus importante.

Il faut par ailleurs garder à l'esprit les biais possibles introduits par l'interpolation linéaire effectuée sur les données IMAGE, dont la fréquence de renouvellement est notoirement trop faible, et pour lesquelles il n'est pas possible de faire une interpolation fine. Parmi les autres sources de biais possibles liés aux données, citons la qualité des vecteurs vitesses, qu'il reste à examiner de plus près.

Pour pouvoir estimer si ces biais liés à la qualité des données IMAGE sont importants ou non, il faudrait pouvoir comparer les indicateurs calculés à partir d'une autre source de données radar, sur une même journée. Nous disposons actuellement d'une journée complète (19/05/2005) de données radar au format MADREC, avec un plot toutes les 12 secondes. Malheureusement, la comparaison des différentes sources de données n'est pas possible, car nous n'avons pas, à l'heure actuelle, les données IMAGE pour cette même journée.

Ceci conclut le chapitre sur les données, et nous allons maintenant aborder l'analyse descriptive des indicateurs, calculés à partir de ces données, en gardant à l'esprit les biais éventuels que nous venons d'évoquer.

Chapitre 5

Une analyse descriptive des indicateurs

Les études qui étudient la densité dynamique essaient souvent d'utiliser une liste d'indicateurs de complexité qui décrit les différents aspects du trafic aérien : volume du trafic, hétérogénéité des avions, conflits... Par conséquent, une étude qui essaie d'intégrer des indicateurs provenants de sources éparpillées est confronté à de sévères problèmes de colinéarité. Ce phénomène est constaté dans [3], qui réalise une analyse en composantes principales (ACP) des données[1]. Kopardekar identifie 12 axes principaux (dans le cadre "normalisé" qui semble adopté dans [3], un axe est sélectionné si, suivant cet axe, la variance du nuage de points est plus grande que la variance initiale des variables utilisées) qu'il identifie subjectivement à des caractéristiques du trafic. Une présentation détaillée de cette ACP aurait été très intéressante, mais elle n'est pas incluse dans l'article. Dans cette section, nous présentons une ACP des indicateurs de complexité implémentés dans DD_INDIC, sur les secteurs armés le 1er juin 2003.

5.1 Statistiques descriptives

Les statistiques descriptives des indicateurs implantés dans DD_INDIC sont présentées, pour les secteurs armés du 01/06/2003 (une observation par minute), dans la table 5.1. Les observations pour lesquelles l'un au moins des indicateurs de proximité issu de [4] prenait une valeur infinie (252 occurences pour 51650 observations) ont été retirées de la base. Les ordres de grandeurs des différents indicateurs sont bien sûr très variables puisque ceux-ci n'ont pas (encore) été normalisés.

5.2 Analyse en composantes principales

Pour comprendre les relations entre indicateurs, on réalise une analyse en composantes principales. Les écarts-types des composantes principales sont représentées sur la figure 5.1.

On observe que la variance de la première composante principale est très élevée. Cette caractéristique s'apparente à un effet "taille" tel que décrit dans [21][2], et traduit le fait que presque tous les indicateurs augmentent statistiquement avec le nombre d'avions actifs. En comparaison, les autres composantes principales ont des variances très faibles. Si l'on décide de ne retenir comme significatifs que les composantes de variance supérieure à 1, on n'en compte que 6. Ce nombre est sensiblement plus faible que les 12 composantes identifiées dans [3], mais la différence peut s'expliquer par l'absence, à ce stade du projet S2D2, de critère lié à la géométrie des secteurs (volume, volume "utile", "structure de l'espace"...)[3].

[1] Par ailleurs, un rapide coup d'oeil aux résultats présentés dans l'étude [3] permet de deviner le problème, les signes de certains paramètres étant parfois fortement contre-intuitifs. A titre d'illustration : les coefficients associés aux indicateurs MET_density et WHEADVAR sont négatifs. Mais serait-ce encore le cas si la densité du trafic et le nombre d'avions en virage, qui interviennent quant à elles positivement dans la densité dynamique, n'étaient pas utilisées dans la régression ?

[2] De fait, on vérifie que tous les indicateurs, à de rares exceptions, sont corrélés positivement entre eux.

[3] Par ailleurs, il n'apparaît pas clairement dans [3] que seules les 23 variables retenues dans le modèle final ont été utilisées pour l'ACP...

Indicateur	Min.	1er Qu.	Mediane	Moyenne	3ème Qu.	Max.
"nb"	0	4	6	6.66	9	24
"nb2"	0	16	36	58.7	81	576
"var_v"	0	1.68e+01	3.95e+01	7.96e+02	1.10e+03	1.11e+04
"nb_des"	0	0	0	0.90	1	13
"nb_mon"	0	0	0	1.15e	2	13
"sqrt_var_sur_moy"	0	5.69e-02	8.91e-02	1.31e-01	1.79e-01	8.86e-01
"flux_5"	0	2	3	3.11	4	15
"flux_15"	0	6	9	9.32	12	38
"flux_30"	0	13	18	18.6	23	67
"flux_60"	0	26	36	37.1	45	117
"vprox_1"	0	2.01e-02	2.82e-02	3.65e-02	4.11e-02	4.01
"vprox_2"	0	0	0	3.63e-04	0	2.16
"hprox_1"	0	0	2.58e-02	3.87e-02	5.14e-02	15.9
"dens_p"	0	4.85	8.72	1.08e+01	1.45e+01	1.01e+02
"des_ang"	0	1.57e+01	1.10e+02	2.58e+02	3.36e+02	5.32e+03
"des_vit"	0	9.83	5.10e+01	1.08e+02	1.34e+02	3.00e+03
"div_p"	0	3.55	7.35e+01	2.31e+02	2.93e+02	4.30e+03
"conv_p"	0	4.72	8.01e+01	2.41e+02	3.09e+02	5.32e+03
"sensi_d"	0	9.77	2.01e+02	4.30e+02	5.96e+02	6.71e+03
"sensi_c"	0	1.28e+01	2.18e+02	4.64e+02	6.28e+02	1.04e+04
"insen_d"	0	7.24e-01	2.94e+01	1.75e+02	1.84e+02	1.81e+04
"insen_c"	0	1.01	3.29e+01	1.72e+02	1.93e+02	5.83e+03
"inter_vert"	0	0	4	7.62	1.00e+01	1.22e+02
"chgt_niv"	0	0	2.38e+02	4.25e+02	7.26e+02	5.20e+03
"inter_hori"	0	2	3.67	4.17	6	2.11e+01
"creed_bon"	0	0	0	4.44e-05	3.40e-05	1.86e-03
"cree_pb"	0	0	0	8.42e-05	7.90e-05	2.77e-03

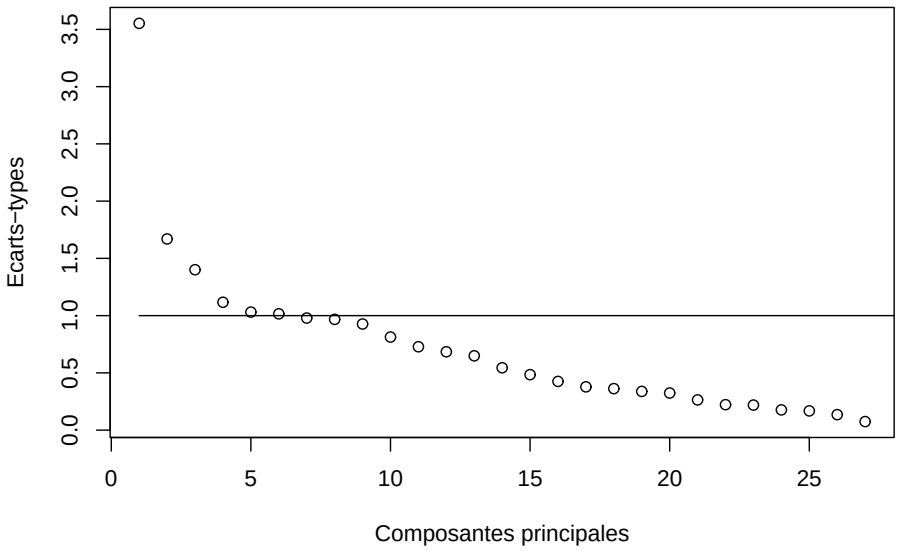

FIG. 5.1 – Ecarts-types des composantes principales

On peut essayer de relier les composantes principales aux caractéristiques du trafic. On obtient ainsi pour les premières composantes :

C1 : Presque toutes les variables présentent une forte corrélation avec cette composante qui s'apparente, on l'a dit, à un "facteur de taille". On peut noter que seuls deux indicateurs définis dans [4] sont ortogonaux à cet axe principal (corrélations inférieures à 0.05).
Comme dans [3], on peut décrire cette composante comme de "surveillance générale".

C2 : Les variables les plus corrélées avec cet axe (resp. 0.68, 0.69 et 0.70) sont, dans cet ordre, "sqrt_var_sur_moy" et "var_v" et "chgt_niv". Si l'on considère qu'une part de la variabilité en vitesse est liée précisément aux phases de transition, on peut considérer que cette composante est liée principalement aux avions en évolution verticale.

C3 : Les variables les plus corrélées avec cet axe sont, de très loin, les flux entrants.

C4 : Les variables les plus corrélées avec cet axe sont les variables liées directement au volume du trafic ("n", "inter_hori",...), tandis que les variables "chgt_niv" et "insen_c" et les indicateurs de type CREED lui sont fortement négativement corrélées. En "inversant" l'image, on peut penser que cet axe est lié à la convergence de flux.

C5 : Les variables les plus corrélées avec cet axe sont les "div" et "insen_p". Cet axe semble donc lié aux flux divergents.

C6 : Cet axe est très fortement corrélé avec les indicateurs de proximité verticale ([4]), et pourrait traduire la surveillance des séparations verticales (proche des minimas).

On voit que caractériser "intuitivement" les composantes principales devient rapidement un exercice délicat. Là encore, on peut regretter le manque de détails fournis dans [3]. Il convient également de souligner, comme dans [21], le caractère tautologique de l'interprétation des composantes principales : "on explique le résultat à l'aide des données qui ont servi à les obtenir". Dans le cadre considéré ici, ce phénomène est plus important encore : les études liées à la densité dynamique présentent souvent un pannel d'indicateurs censé couvrir les multiples aspects de la complexité du trafic. Les aspects en question ("densité", "hétérogénéité du trafic", "conflits potentiels"...) sont largement consensuels, et des indicateurs de différentes sources poursuivent donc des buts très similaires et sont souvent très corrélés les uns aux autres. Les utiliser conjointement dans une ACP contribue à leur donner un poids important, et les composantes principales ne font que traduire les idées des modélisateurs.

Conclusion, travaux futurs

Pour conclure sur les travaux préparatoires indispensables dans le cadre du projet S2D2, nous avons donc détaillé de manière formelle un certain nombre d'indicateurs de complexité du trafic aérien trouvés dans (ou adaptés de) la littérature, que nous avons implanté sous forme de librairie (module DD_INDIC, en langageOcaml). Ce module pourra par la suite être étendu, en intégrant de nouveaux indicateurs[4]. Il permet pour l'instant d'en calculer une trentaine, à partir de données de trafic mises sous un format spécifique, et ce dans le but de l'adapter ultérieurement à différentes sources de données.

Nous avons ensuite passé en revue les données existantes (archives IMAGE et COURAGE principalement), en soulignant un certain nombre d'incohérences, d'incompletudes, ou de mauvaix choix initiaux dans la structure des données, parfois préjudiciable à une bonne exploitation de celles-ci. Citons à titre d'exemple la difficulté à déterminer la frontière externe d'un regroupement de secteurs, à cause de la mauvaise description géographique des secteurs élémentaires.

Les problèmes que nous avons rencontré dans l'exploitation des données dépassent le seul cadre du projet S2D2, et sont susceptibles de concerner tout projet utilisant les mêmes données. Il sont détaillés en section 3, et récapitulés en 3.9, et concernent pour l'essentiel la mauvaise description des secteurs aériens, la fréquence insuffisante de renouvellement des plots IMAGE, et quelques interrogations sur la qualité des vecteurs vitesse, ainsi que des incohérences entre heures d'entrée secteur COURAGE et les heures des plots IMAGE au passage des frontières de secteurs, ces deux derniers points restant sujets à investigation. Nous avons également montré, à travers quelques tentatives, qu'il est illusoire de chercher à reconstituer finement les trajectoires des avions à partir des données IMAGE, des routes plan de vol COURAGE, et des données de performances avions. Nous avons donc finalement choisi d'interpoler linéairement les plots intermédiaires dont nous avions besoin, ce qui n'est pas très précis, compte tenu du temps séparant deux plots IMAGE consécutifs.

Il existe donc certainement un biais dans le calcul des indicateurs, lié à la mauvaise qualité des données ayant servi à les calculer, et notamment à la fréquence de renouvellement trop faible des plots IMAGE. Nous aurions aimé estimer l'importance de ces biais en calculant les indicateurs à partir d'autres sources de données (multi-radar mono-centre notamment). Nous disposons pour cela d'une journée complète (19/05/2005) de données radar au format MADREC, avec un plot toutes les 12 secondes. Malheureusement, nous n'avons pas, à l'heure actuelle, de données IMAGE pour cette même journée.

Cette étude conclut sur l'analyse descriptive des indicateurs, calculés sur une journée de trafic du mois de juin 2003. Les statistiques descriptives montrent la disparité des valeurs prises par les différents indicateurs, et donc l'importance qu'il y a à les normaliser avant toute tentative de corrélation avec les données de regroupement. L'analyse en composantes principales, sur des indicateurs normalisés de moyenne nulle et de variance 1, a permis quant à elle de préciser les corrélations entre les différents indicateurs, et de dégager 7 composantes (sur les indicateurs calculés) relativement indépendantes, qu'il n'est pas toujours aisé de rattacher explicitement à des caractéristiques du trafic.

A la suite de cette étude préliminaire des corrélations entre les différents indicateurs, il reste à étudier le liens entre ces derniers et les données de regroupement de secteurs. Ce travail, en cours, est envisagé selon deux approches : une méthode statistique avec une régression logistique, et une approche par réseaux de neurones. Ces approches pourront être appliquées soit à l'ensemble des indicateurs, ou bien aux combinaisons

[4]D'autres indicateurs sont d'ailleurs en voie d'implémentation (voir en annexes).

linéaires d'indicateurs correspondant aux axes principaux issus de l'analyse descriptive.

Bibliographie

[1] P. Kopardekar. Dynamic density : A review of proposed variables. Faa wjhtc internal document. overall conclusions and recommendations, Federal Aviation Administration, 2000.

[2] Cognitive complexity in air traffic control, a litterature review. Technical report, Eurocontrol experimental centre, 2004.

[3] P. Kopardekar and S. Magyarits. Measurement and prediction of dynamic density. In *Proceedings of the 5th USA/Europe Air Traffic Management R & D Seminar*, 2003.

[4] G.B. Chatterji and B. Sridhar. Measures for air traffic controller workload prediction. In *Proceedings of the First AIAA Aircraft Technology, Integration, and Operations Forum*, 2001.

[5] D. Delahaye and S. Puechmorel. Air traffic complexity : towards intrinsic metrics. In *Proceedings of the third USA/Europe Air Traffic Management R & D Seminar*, 2000.

[6] P. Averty, S. Athènes, C. Collet, and A. Dittmar. Evaluating a new index of mental workload in real atc situation using psychological measures. Note cena nr02-763, CENA, 2002.

[7] B. Sridhar, K. S. Sheth, and S. Grabbe. Airspace complexity and its application in air traffic management. In *Proceedings of the 2nd USA/Europe Air trafic Management R&D Seminar*.

[8] I. V. Laudeman, S. G. Shelden, R. Branstrom, and C. L. Brasil. Dynamic density : An air traffic management metric. Technical report, 1999.

[9] Note de synthèse sur l'indicateur de complexité pru. Technical report, DTI/SDER (ex CENA), 2005.

[10] P. Averty. Conflit perception by atcs admits doubt but not inconsistency. In *Proceedings of the 6th Air Traffic Management Research & Developpment Seminar*, 2005.

[11] P. Averty, K. Guittet, and P. Lezaud. *Work in progress, presented at an internal SDER seminar*. Technical report, DTI/SDER (former CENA), 2005.

[12] F. Chatton. Etudes de nouvelles métriques de complexité de la circulation aérienne. Master's thesis, Ecole Nationale de l'Aviation Civile (ENAC), 2001.

[13] E. Pfleiderer. Relationship of the aircraft mix index with performance and objective workload evaluation research (power) measures and complexity ratings. In *Proceedings of the 6th Air Traffic Management Research & Developpment Seminar*, 2005.

[14] J.L. Costes. Evolutions sur la génération du fichier pln pour courage, version provisoire. Technical report, STNA/7, 1995.

[15] Note anonyme non référencée. Projection utilisée dans courage, presage, et shaman pour l'affichage du fond de carte. Technical report, STNA/7.

[16] note STNA 95.2091. Spécifications évolutions previ-surf-courage. Technical report, STNA/7, 1995.

[17] CFMU. Etfms interface control document france. Technical report, Eurocontrol, 1999.

[18] Equipe STR. Dossier de description d'interface avec le système image. Technical report, STNA/7, 1999.

[19] P. Amaury and L. Crouzier. Dossier de conception préliminaire et de description détaillée courage, version provisoire. Technical report, STNA/7, 1993.

[20] European Petroleum Survey Group. Coordinate conversions and transformations including formulas. Technical report, EPSG, January 2004.

[21] G. Saporta. *Probabilités, analyse des données et statistique*. Technip, 1990.

[22] An evaluation of air traffic control complexity. Technical report, Wyndemere, 1996.

Annexes

Les indicateurs implémentables à moyen terme

Certains indicateurs, qui semblent pertinents, font appel à la définition de conflit potentiel, qui est rarement donnée de manière précise, ainsi qu'à des caractéristiques géométriques des secteurs, qui sont assez délicates à exploiter. A moyen terme, ces indicateurs devraient être intégrés à DD_INDIC.

On reprend ici le nom utilisé par Kopardekar dans [3] pour lister ces indicateurs :

- MET_boundary proximity ou WBPROX : nombre d'avions situés à une distance du bord du secteur inférieure à un certain seuil (e.g. 10 miles, dans [22]).
- MET_conflict neighbors : nombre d'avions situés à proximité d'un conflit potentiel (e.g. espacement inférieur à 10 miles horizontalement et 2000 pieds verticalement). Il faudrait la définition exacte d'un conflit potentiel.
- MET_conflict near boundary : nombre de conflits potentiels situés à une distance du bord du secteur inférieure à un certain seuil (e.g. 10 miles).
- MET_airspace structure ou WASP : différence au carré entre le cap de chaque avion dans le secteur et la direction de l'axe principal du secteur, pondérée par le "sector aspect ratio". On n'a pas de définition du "sector aspect ratio"...

On peut remarquer que ces indicateurs se focalisent sur les conflits potentiels et les avions situés au voisinage des frontières du secteur. En ce sens, ils semblent complémentaires des indicateurs déjà programmés, et donc particulièrement intéressants.

Enfin, les indicateurs utilisés pour le projet INTEGRA d'EUROCONTROL devraient également être prochainement intégrés à DD_INDIC, Y. Le Fablec ayant fourni son code Caml. Mais la documentation sur ces métriques de complexité est extrèmement concise...

Les indicateurs dont vous êtes le héros

Certains indicateurs qui semblent intéressants n'ont pas pu être reprogrammés, et ne sont pas implémentables actuellement. Il s'agit la plupart du temps d'indicateurs dont la définition n'était pas parfaitement claire. Pour en faire la liste, on reprend le nom utilisé par Kopardekar dans [3].

1. DOFI : cet indicateur est décrit comme l'"index des degrés de liberté", et est basé sur les manoeuvres possibles en situation de conflit.
2. MET_ac_count : nombre d'avions dans le secteur, indicateur utilisé dans [22]. Apparemment simple, cet indicateur est mal défini. Il semble en effet qu'on ait MET_ac_count $= N$, ce qui est impossible, puisque les deux indicateurs apparaissent dans une même régression dans [3], et qu'une collinéarité parfaite aurait été identifiée.
3. MET_density : rapport du nombre d'avions sur le "volume utilisable" du secteur (dans [22]). Le "volume utilisable du secteur" n'est pas défini. Il ne peut pas être égal au volume du secteur, sans quoi l'indicateur MET_density est égal à la densité (au sens de Kopardekar).
4. MET_heading variation ou WHEADVAR : littéralement, "count of number of bearing changes above a threshold value within the sector". La définition n'est pas claire, et l'indicateur semble être très corrélé avec le nombre d'avions en virage.

Temps de conflit, degré d'anticipation...

Certaines études (notamment [4] et [5]) utilisent la notion de "temps de conflit", défini par

$$t_c = \frac{d_{ij}^{ob}}{\|\overrightarrow{v}_{ij}\|} \ . \tag{5.1}$$

Si le terme "temps de conflit" semble transparent, et tente de traduire le temps dont dispose le contrôleur pour empêcher un éventuel abordage, il est en réalité bien difficile de déterminer ce que représente réellement cette grandeur : la différence des vitesses peut-être très éloignée de la vitesse de convergence, qui est loin d'être constante en général[5]...
Le "degré d'anticipation" DA est introduit dans [10] et représente une alternative visuelle au "temps de conflit". Le degré d'anticipation est représenté dans un cas favorable sur la figure 5.2. Mais deux avions peuvent être en convergence dans le plan sans que leurs trajectoires (toujours dans le plan) ne s'intersectent. Dans ce cas, illustré sur ma figure 5.3, "Da" est négatif.

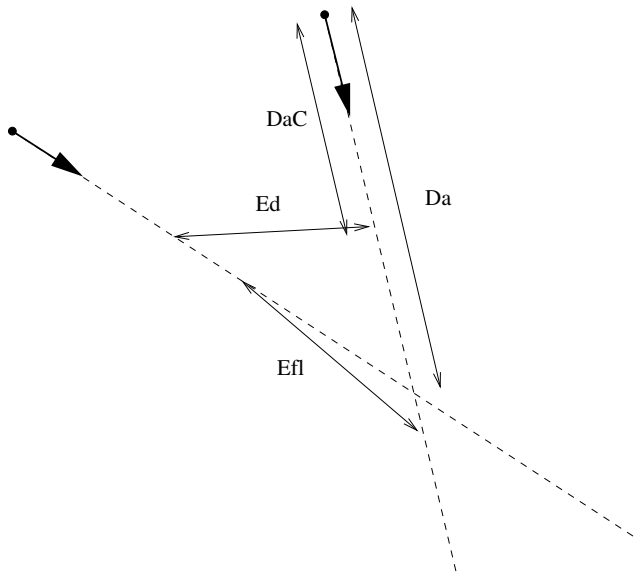

FIG. 5.2 – Un cas favorable pour le calcul de Da

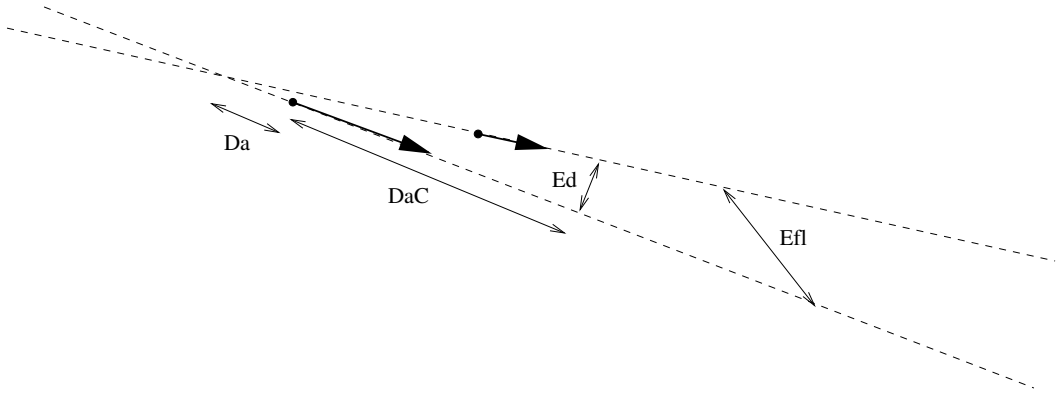

FIG. 5.3 – Un cas problématique pour le calcul de Da

Dans cette étude, pour éviter ces difficultés, nous avons utilisé une quantité un peu différente, un "degré d'anticipation corrigé" DaC, pour décrire le temps dont dispose le contrôleur pour empêcher un éventuel abordage. On définit d'abord le temps t_{ij} à partir duquel les trajectoires des avions i et j commencent à diverger. C'est la distance parcourue par l'avion le plus rapide pendant ce laps de temps qui donne DaC.

[5]La vitesse de convergence n'est jamais constante, sauf lorsque les trajectoires des avions convergent réellement. A noter que dans ce cas, la vitesse de rapprochement est égale à la différence des vitesses, et que le "temps de conflit" est bien le laps de temps jusqu'à l'abordage des deux avions considérés.

Enfin, si l'on note s_{ij} la distance qui sépare les avions à l'instant t_{ij},on a :

$$\begin{aligned} t_{ij}^2 &= -\frac{< \overrightarrow{X_iX_j}, \overrightarrow{v}_{ij} >}{\|\overrightarrow{v}_{ij}\|^2} \ , \\ &= -\frac{d_{ij}v_{ij}}{\|\overrightarrow{v}_{ij}\|^2} \ , \end{aligned} \tag{5.2}$$

et

$$s_{ij} = \sqrt{d_{ij}^2 - \|\overrightarrow{v}_{ij}\|^2 t_{ij}^2} \ . \tag{5.3}$$