



HAL
open science

Réseaux de neurones appliqués aux indicateurs de complexité et aux regroupements de secteurs aériens

David Gianazza

► **To cite this version:**

David Gianazza. Réseaux de neurones appliqués aux indicateurs de complexité et aux regroupements de secteurs aériens. 2006. hal-01020633

HAL Id: hal-01020633

<https://enac.hal.science/hal-01020633v1>

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.

Réseaux de neurones appliqués aux indicateurs de complexité et aux regroupements de secteurs aériens

Projet S2D2 (Sector Status and Dynamic Density)

David Gianazza

DSNA/DTI/SDER

Laboratoire d'Optimisation Globale

7, avenue Edouard Belin 31055 Toulouse Cedex

Email : gianazza@recherche.enac.fr

Note DSNA/DTI/SDER NR06-517

Point d'avancement du 28 avril 2006

1 Introduction

La présente note est un nouveau point d'avancement du projet S2D2 (Sector status and dynamic density), mené en collaboration entre le LEEA et le LOG. 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.

Le terme de *complexité du trafic aérien*, de plus en plus utilisé ces dernières années, recouvre des conceptions et des réalités très diverses. Pour la *Performance Review Unit* (PRU) d'Eurocontrol (cf note [1], du projet COCA), la complexité est abordée au moyen d'indicateurs agrégés sur une journée complète. L'étude citée se place à l'échelle géographique du centre de contrôle, en comparant différents centres d'Amérique du Nord et d'Europe, puis à l'échelle du secteur, en établissant une typologie des secteurs selon quelques indicateurs agrégés. Les secteurs utilisés correspondent à une ouverture maximum des positions de contrôle dans chaque centre (soit 82 secteurs en Europe, et 124 secteurs aux U.S.). Les regroupements de secteurs élémentaires effectivement armés ne sont pas exploités, ce qui se comprend puisque l'étude ne s'intéresse pas au domaine d'utilisation des secteurs de contrôle.

Dans de nombreuses autres études, la complexité est envisagée en relation avec la charge de travail du contrôleur à l'instant t . Elle est alors souvent subjective, dans le sens où elle est estimée par des contrôleurs aériens, lors de simulations ou sur du trafic réel, généralement rejoué ([2], [3]). Certaines expérimentations prennent en compte des indicateurs physiologiques du stress du contrôleur face à un trafic donné ([4], entre autres). 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 de concrétiser une idée évoquée dans [5] et d'utiliser une autre mesure de la charge de travail du contrôleur, certes moins précise dans le domaine d'utilisation normal des secteurs, mais relativement objective et simple à obtenir. L'idée consiste à exploiter les données archivées des affectations de secteurs aux positions de contrôle, en faisant l'hypothèse que les décisions de regroupement ou de "dégrouperment"¹ de secteurs sont statistiquement significatives de la charge de travail du contrôleur, et en gardant à l'esprit les biais possibles (ouvertures de secteurs à des fins de formation, relève des équipes, incidents techniques, etc...). La complexité est donc ici envisagée

¹néologisme employé pour désigner le fait de séparer un secteur de contrôle en plusieurs secteurs plus petits

à une échelle intermédiaire entre les indicateurs généraux de la PRU et la charge de travail instantanée. Nous cherchons une relation entre les indicateurs de complexité et le domaine d'utilisation des secteurs de contrôle, avec une granularité temporelle de l'ordre de la minute.

La précédente note ([6]) présentait une analyse descriptive de vingt-trois indicateurs de complexité du trafic aérien, collectés dans la littérature ([7], [8], [2], [3], [9], [4]), auxquels nous avons ajouté les flux entrant dans chaque secteur, avec des horizons temporels différents. Ces indicateurs étaient calculés sur des données de trafic français (archives IMAGE et COURAGE), et la note était également l'occasion de faire le point sur un certain nombre de problèmes dans l'exploitation de ces données.

Le travail présenté ici porte sur l'application des réseaux de neurones aux indicateurs de complexité du trafic aérien et aux états de regroupement des secteurs de contrôle. L'objectif est d'établir une relation entre les états de regroupement, que l'on suppose liés la charge de travail du contrôleur, et les indicateurs de complexité du trafic. On espère ainsi réussir à dégager l'influence de chaque type d'indicateur, ou de quelques combinaisons intéressantes.

Comme nous l'avons déjà évoqué, il existe déjà de nombreuses études portant sur la complexité du trafic aérien. Parmi celles se rapprochant le plus de notre démarche, citons G.B. Chatterji et B. Sridhar ([3]) qui utilisent un réseau de neurones sur des indicateurs et des "notes de complexité" concernant *un unique secteur*. Dans [2], P. Kopardekar et S. Magyarits font une régression linéaire sur des indicateurs et des notes de complexité, avec des échantillons de trafic concernant 36 secteurs de différents centres, en comparant différents modèles basés sur le simple comptage d'avions ou sur la *densité dynamique*, celle-ci étant vue dans cette étude comme une combinaison linéaire d'indicateurs. L'étude ne permettait cependant pas d'isoler l'influence respective des différents facteurs de la complexité du trafic.

Les deux études [3] et [2] cherchent à établir une relation entre des indicateurs de complexité (ou des combinaisons d'indicateurs) et la charge de travail exprimée sous forme de "note de complexité" donnée par un panel de contrôleurs à des échantillons de trafic. Il s'agit d'une estimation subjective, *a posteriori*, de la charge de travail.

L'originalité de notre démarche tient tout d'abord au fait que nous travaillons sur une mesure plus objective, même si elle est bruitée et moins précise, basée sur les regroupements de secteurs. D'autre part, nous calculons les indicateurs sur toutes les combinaisons de regroupements de secteurs utilisées sur la période considérée, ce qui nous donne plusieurs centaines de secteurs (au sens large). Nous disposons en conséquence d'une plus grande variété de secteurs géographiques et de situations de trafic. Les points de mesures sont plus nombreux et mieux répartis entre les différentes classes (secteur regroupé, armé, ou "dégroulé"), ce qui constituait, de l'aveu même des auteurs, une faiblesse dans les études [3] et [2]. Enfin, contrairement à [2], nous ne faisons pas d'hypothèse sur la linéarité de la relation entre indicateurs et charge de travail, et nous cherchons à mieux déterminer le poids relatif des facteurs de complexité.

La section 2 de la note rappelle les indicateurs que nous avons calculés (en reprenant largement la note [6]). Les données utilisées pour calculer les indicateurs sont rapidement présentées en section 3. La section 4 rappelle les principaux résultats de l'analyse en composantes principales que nous avons effectuée sur les indicateurs choisis. Dans la section 5, après quelques généralités sur les réseaux de neurones artificiels, nous présenterons les réseaux que nous avons utilisés et les critères choisis pour évaluer leurs résultats. Les résultats préliminaires et les considérations sur le choix des paramètres du réseau sont présentés en section 6. Les sections 7 et 8 présentent les résultats du processus de sélection des composantes principales et des indicateurs significatifs, ainsi que les taux de classifications correctes que l'on peut obtenir avec les combinaisons testées. Nous présenterons également une équation simple permettant de prédire l'état de regroupement à partir des valeurs des indicateurs retenus. En section 9, nous discuterons les résultats obtenus, en les comparant aux résultats d'autres études existantes. La dernière section regroupe les conclusions et les perspectives de travaux futurs.

2 Les indicateurs

Les indicateurs programmés par K. Guittet, que nous avons implantés sous forme de librairie Ocaml dans le module DD_INDIC, comprennent un sous-ensemble des indicateurs retenus par Kopardekar dans sa revue de littérature, ainsi que d'autres indicateurs, dérivés d'études conduites par Eurocontrol ou par des équipes de la SDER (ex CENA). Tous les indicateurs de la littérature n'ont pas été implantés, du fait de leur très grand nombre, mais aussi, souvent, par manque d'une description rigoureuse et complète.

Les indicateurs étudiés sont brièvement listés dans la table 1. Il s'agit de ceux déjà décrits dans la précédente note [6]. A ces indicateurs, nous avons ajouté comme variable explicative, le volume de chaque secteur considéré.

Indicateur	Références ²	Type
nb	[10] [3] [7] [2]	S
nb2	[7] [2]	S
var_v	[3]	S
nb_des	[11] [10] [3]	S
nb_mon	[11] [10] [3]	S
sqrt_var_sur_moy	[3]	S
flux_5	[5]	F
flux_15	[5]	F
flux_30	[5]	F
flux_60	[5]	F
vprox_1	[3] [7] [2]	D
vprox_2	[3] [7] [2]	D
hprox_1	[3] [7] [2]	D
dens_p	[9]	D
des_ang	[9]	D
des_vit	[9]	D
div_p	[9]	D
conv_p	[9]	D
sensi_d	[9]	D
insen_d	[9]	D
sensi_c	[9]	D
insen_c	[9]	D
inter_vert	[12]	S
chgt_niv	[12]	S
inter_hori	[12]	D
creed_bon	[13], [14]	D
creed_pb	[13], [14]	D

TAB. 1 – Liste des indicateurs implémentés dans S2D2

Afin que la présente note puisse être lue sans faire référence en permanence à la note [6], nous reprenons *in extenso* en annexe le texte de cette note détaillant les indicateurs.

3 Données

Les indicateurs sont calculés à chaque minute ronde de la journée, à partir des données IMAGE et COURAGE, comme décrit dans la note [6]. Rappelons notamment que les positions des avions (une par minute) sont obtenues en interpolant linéairement à partir des plots radar IMAGE (disponibles toutes les

trois minutes en moyenne). Il y a donc un certain biais dans le calcul des indicateurs, dû à la fréquence d'archivage, trop faible, des plots radar IMAGE.

Les archives des configurations de secteurs sont enregistrées toutes les minutes, ce qui explique notre choix pour la fréquence de calcul des indicateurs. Nous faisons ici l'hypothèse que les décisions de regroupement ou de "dégroupement" 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...).

Plusieurs mois d'enregistrements IMAGE et COURAGE sont disponibles. Au regard du volume de données que cela représente, nous nous restreindrons dans un premier temps à une seule journée de trafic, choisie dans une période chargée de l'année afin de limiter les biais sur les causes des dégroupements/regroupements. Une fois sélectionnés les indicateurs pertinents, il sera toujours possible par la suite d'effectuer l'apprentissage du réseau sur des échantillons de données plus grands. Les résultats de la présente note sont établis à partir des enregistrements du 1^{er} juin 2003.

Dans cette journée, 103 secteurs³ différents ont été armés à un moment ou à un autre de la journée. Pour chacun de ces secteurs, les indicateurs de complexité sont calculés à chaque minute de la journée, avec l'état de regroupement du secteur (*regroupé*, *armé*, ou *dégroupé*). L'ensemble des données ainsi obtenues est divisé en deux parties : environ soixante pour cent des données sont choisies aléatoirement et seront utilisés pour l'apprentissage, le reste étant réservé à la validation des prévisions du réseau (voir section 5).

Cette seule journée de trafic représente à elle seule un volume important de données, détaillé dans le tableau 2, avec une grande variété de secteurs géographiques et avec suffisamment de données pour chaque classe de regroupement.

	Total	Merged	Armed	Split
Apprentissage	71270	46.6%	27.0%	26.4%
Validation	47513	46.4%	27.0%	26.6%

TAB. 2 – Nombre de points de mesures disponibles pour le 1er juin 2003.

Avant de présenter les réseaux de neurones et leur résultats sur les indicateurs de complexité, rappelons tout d'abord succinctement les résultats de l'analyse en composantes principales (ACP), qui permet de dégager les axes principaux des corrélations existant entre les indicateurs.

4 Rappel des résultats de l'ACP

Les indicateurs recueillis dans la littérature et les flux entrants représentent 27 indicateurs de complexité. L'objectif de l'analyse en composantes principales est de détecter les colinéarités entre certains de ces indicateurs, afin de pouvoir par la suite travailler sur un nombre plus restreint de variables.

Nous rappelons ici les principaux résultats de l'ACP, déjà présentés dans [6], en les détaillant un peu plus et en les complétant par la liste des indicateurs représentatifs de chaque composante principale. L'ACP a permis de dégager 6 composantes principales, de valeurs propres supérieures à 1, qui recouvrent plus de 76% de la variance totale des données. Ces composantes s'interprètent comme suit :

C1 : Valeur propre 12,6, et 46,7% de la variance totale. Mis à part *vprox_1* and *vprox_2*, toutes les variables sont fortement (et positivement) corrélées avec cette composante, ce qui explique sa valeur propre élevée. Cette composante peut être vue comme un facteur de taille, et l'on peut suivre [2] en utilisant le terme de *surveillance générale*. Cette composante est fortement représentative du nombre d'avions présents dans le secteur.

C2 : Valeur propre 2,78, et 10% de la variance totale. Cette composante est fortement corrélée à *chg_niv*, σ_{gs}^2 et $\sigma_{gs}/\overline{gs}$ (resp. 0,70, 0,69 et 0,68). 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*.

³Un secteur étant ici soit un secteur élémentaire ou un groupe de secteurs élémentaires.

- C3 : Valeur propre 1,96, et 7,3% de la variance totale. Les variables les plus corrélées avec cet axe sont, de très loin, les *flux entrants*.
- C4 : Valeur propre 1,25, et 4,6% de la variance totale. 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*, et à l'*anticipation des conflits*.
- C5 : Valeur propre 1,06, et 3,9% de la variance totale. Les variables les plus corrélées avec cet axe sont les "div" et "insen_p". Cet axe semble donc lié aux *flux divergents*.
- C6 : Valeur propre 1,03, et 3,8% de la variance totale. Cet axe est très fortement corrélé avec les indicateurs de proximité verticale ([3]), et pourrait traduire la *surveillance des séparations verticales* (proche des minimas).

Composante	Variables représentatives
C_2	$chg_niv, \frac{\sigma_{gs}}{qs}, \sigma_{gs}^2, N_{cl}, N_{ds}$
C_3	$F_5, F_{15}, F_{30}, F_{60}$
C_4	$inter_hori, insen_c, Conv, creed_ok, creed_pb$
C_5	$insen_d, Div$
C_6	$vprox_1, vprox_2$

TAB. 3 – Variables représentatives des composantes principales

Le tableau 3 liste les indicateurs donnant la meilleure explication pour chacune des composantes C_2 à C_6 . Pour C_1 , nous prendrons simplement le nombre d'avions Nb . Pour C_2 et C_3 , les variables sélectionnées ont une corrélation, en valeur absolue, supérieure à 0,4 (corrélation avec l'axe de la composante). Les variables choisies sont donc hautement représentatives des composantes. Le choix des variables est moins facile pour la composante C_4 , où plusieurs des indicateurs ayant de hautes valeurs de corrélation ne sont pas très significatifs, car déjà présents dans les composantes précédentes. Il nous faudrait probablement raffiner la définition de certains des indicateurs de C_4 pour avoir une interprétation plus claire de cette composante. Nous avons dans l'immédiat retenu les variables avec les plus hautes corrélations, et qui n'apparaissent pas dans les composantes précédentes. Pour C_5 et C_6 , les variables retenues sont bien corrélées avec leur composante (0,4 à 0,7).

5 Les réseaux de neurones

5.1 Généralités

Les réseaux de neurones artificiels sont une classe d'algorithmes s'inspirant du modèle biologique du fonctionnement des neurones et des liaisons synaptiques, avec des unités (neurones) reliées entre elles par des liaisons auxquelles sont affectés des poids. Il existe une multitude de types de réseaux de neurones artificiels, adaptés à différents types de problèmes (cf [15], [16], [17]).

Au-delà des similitudes avec le modèle biologique, les réseaux de neurones sont des outils statistiques dont les fondements théoriques reposent sur les probabilités Bayésiennes ([18]). Dans le contexte de la régression ou de la classification de données, il s'agit de trouver une relation probabiliste entre un ensemble de variables explicatives x (les données en entrée), et une variable cible y (la donnée en sortie), en s'appuyant sur une combinaison de fonctions de base.

Nous allons dans la suite utiliser des réseaux de type *feed-forward*, également appelés perceptrons multi-couches (MLP) lorsque la fonction d'activation est logistique. Ces réseaux comportent une couche d'entrée, un certain nombre de couches cachées, et une couche de sortie. Chaque unité est totalement connectée à l'ensemble des unités de la couche précédente, et réalise une fonction d'activation permettant de calculer la sortie de l'unité, en fonction de la somme des sorties de la couche précédente, pondérée

par les poids des liaisons. La figure 1 illustre un tel réseau, avec une seule couche cachée.

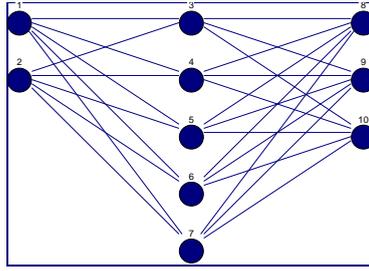


FIG. 1 – Exemple d’un réseau de type *feed-forward* avec une couche cachée

Considérons un réseau MLP avec une couche cachée. Le vecteur de sortie $y = (y_1, \dots, y_k, \dots, y_q)^T$ s’exprime comme une fonction du vecteur d’entrée $x = (x_1, \dots, x_i, \dots, x_p)^T$, de la façon suivante :

$$y_k = \Psi\left(\sum_{j=1}^q w_{jk} \Phi\left(\sum_{i=1}^p w_{ij} x_i + w_{0j}\right) + w_{0k}\right) \quad (1)$$

où les w_{ij} et w_{jk} sont les poids affectés aux connexions entre la couche d’entrée et la couche cachée, et entre la couche cachée et la couche de sortie, respectivement, et où les w_{0j} et w_{0k} sont des biais (également appelées seuils d’activation). Φ est une *fonction d’activation* s’appliquant sur la somme pondérée des sorties de la couche précédente (en l’occurrence, la couche d’entrée), et Ψ est une fonction s’appliquant, pour chaque neurone de sortie, à la somme pondérée des activations de la couche cachée. L’expression 5.1 peut se généraliser à des réseaux ayant plus d’une couche cachée.

L’erreur en sortie, c’est-à-dire la différence entre la valeur cible t et la sortie y calculée par le réseau, dépendra des paramètres du réseau (poids et biais). Une phase d’*apprentissage* est nécessaire, qui permet de déterminer les poids et les biais de façon à minimiser une fonction de l’erreur de sortie. L’apprentissage se fait sur un échantillon de données pour lesquels on connaît les valeurs cibles.

Dans le cas de la *régression*, la fonction minimisée durant l’apprentissage est généralement la somme des erreurs quadratiques, qui s’exprime de la façon suivante, avec N la taille des données d’apprentissage :

$$E(w) = \sum_{n=1}^N ||t^{(n)} - y^{(n)}||^2 \quad (2)$$

où $t^{(n)}$ est le $n^{\text{ièmes}}$ vecteur cible, et $y^{(n)}$ le $n^{\text{ièmes}}$ vecteurs de sortie

Dans le cas de la classification, il est généralement préférable de minimiser une fonction de log-vraisemblance du type *entropie croisée*. Le réseau comporte alors une unité de sortie par classe. La valeur en sortie d’une telle unité est la probabilité conditionnelle d’appartenance à une classe. Considérons un problème à C classes. La fonction de log-vraisemblance que l’on cherche à minimiser durant l’apprentissage est la suivante :

$$E(w) = - \sum_{n=1}^N \sum_{k=1}^C t_k^{(n)} \ln(y_k^{(n)}) \quad (3)$$

L’intérêt de cette fonction est qu’elle donne un poids relatif plus important aux données situées aux frontières des classes qu’à celles à l’intérieur de chaque domaine.

Diverses méthodes d’optimisation peuvent être utilisées pour minimiser la fonction $E(w)$. Citons notamment la *rétro-propagation*, dans laquelle les poids des connexions (et des biais) entre les couches du réseau subissent des modifications successives, en commençant par la couche de sortie. Ces modifications sont fonction de l’importance relative de chaque poids dans les variations de l’erreur de sortie. D’autres méthodes locales utilisant le gradient de la fonction d’erreur (BFGS, gradients conjugués, etc...) , sont également largement utilisées. On trouve également dans la littérature des méthodes d’optimisation globale,

telles que le recuit simulé ou les algorithmes évolutionnaires, généralement plus lentes que les autres méthodes, mais plus performantes quand la fonction d'erreur comporte de nombreux minima locaux.

Sous certaines conditions, les réseaux de neurones ont la propriété d'*approximation universelle paramétrable*, c'est-à-dire qu'ils permettent d'approximer n'importe quelle fonction suffisamment régulière, en n'utilisant qu'un nombre restreint de paramètres (les poids). Un piège connu des réseaux de neurones, et aussi d'autres méthodes statistiques de régression, est celui du sur-apprentissage : avec suffisamment de paramètres, il est toujours possible de trouver une bonne approximation de la sortie y à partir d'un jeu de données d'entrée fixé. Il existe de nombreux critères statistiques permettant d'évaluer si des résultats, apparemment bons, sont dûs à une bonne approximation du phénomène observé ou à un sur-apprentissage. Une façon de procéder consiste à tester le réseau de neurones sur un jeu de données différent, n'ayant pas servi à l'apprentissage. Nous avons choisi d'utiliser trois cinquièmes des données disponibles pour l'apprentissage, et les deux cinquièmes restants pour la validation.

5.2 Réseaux de neurones utilisés pour notre problème

Dans le cadre de notre étude, nous utilisons des réseaux MLP à trois couches, notés $I_x H_y O_z$ par la suite, pour un réseau comprenant x unités en entrée, y unités dans la couche cachée, et z unités en sortie. Les variables en entrée sont normalisées, en soustrayant la valeur moyenne de l'indicateur, et en divisant par l'écart-type.

De nombreux choix sont possibles quant aux fonctions Ψ et Φ , selon le problème considéré. Un choix classique pour Φ est la fonction sigmoïde, ou logistique :

$$\Phi(z) = \frac{1}{1 + e^{-z}} \quad (4)$$

C'est la fonction d'activation utilisée dans nos expérimentations. Comme nous nous intéressons à un problème de classification – affecter chaque vecteur d'entrée (une liste de valeurs d'indicateurs) à une classe représentant l'état du secteur (regroupé, armé, dégroupé) – nous avons choisi de minimiser l'entropie croisée. En conséquence, la fonction de transfert Ψ , appliquée par chaque unité de sortie, doit être la fonction *softmax* :

$$\Psi(z_k) = \frac{e^{z_k}}{\sum_{m=1}^C e^{z_m}} \quad (5)$$

Dans un premier temps, nous avons opté pour le logiciel SNNS ([19]), qui semblait permettre de tester de nombreux types de réseaux et offrir un grand choix d'options. Malheureusement, la version 4.1 ne permettait pas de minimiser l'entropie croisée. La documentation de la version 4.2 offrait cette option, mais la version Linux de SNNS s'est avérée boguée, ne permettant pas de sélectionner l'option qui nous intéressait.

Nous avons donc finalement utilisé le package `nnet` de l'environnement \mathbb{R}^4 . Dans ce package développé par le Pr B. D. Ripley, l'apprentissage se fait par une méthode de quasi-Newton (BFGS). Les paramètres de l'outil `nnet` sont les suivants : un intervalle (*range*, valeur par défaut 0.5) limitant le choix aléatoire des poids initiaux du réseau, un paramètre de dégradation des poids (0 par défaut), et le nombre maximum d'itérations. L'apprentissage s'arrête quand l'un des trois critères d'arrêt est satisfait, c'est-à-dire lorsque le critère optimisé (l'entropie croisée dans notre cas) est en dessous d'une valeur choisie pour un paramètre *abstol* ($1.0e - 4$ par défaut), ou lorsque l'écart avec la précédente valeur est inférieur à $1 - reltol$ (la valeur par défaut de *reltol* est $1.0e-8$), ou enfin si le nombre maximum de cycles est atteint.

Plusieurs combinaisons d'indicateurs de complexité, ou de composantes principales, seront testées. Le nombre d'unités d'entrée du réseau de neurones sera donc choisi égal au nombre d'indicateurs évalués (ou de composantes). Le nombre d'unités dans la couche cachée sera discuté un peu plus loin. La couche de sortie comprendra trois unités, une pour chaque classe (*regroupé*, *armé*, ou *dégroupé*).

Un vecteur cible $t^{(n)}$ de valeur $(1, 0, 0)$ signifiera que le secteur considéré était regroupé avec d'autres secteurs sur une même position de contrôle, au moment où la $n^{\text{ième}}$ mesure du vecteur d'indicateurs de

⁴cf <http://www.r-project.org/> pour plus de détails sur le langage R

complexité a été réalisée. De même, les configurations "secteur armé" et "secteur dégroupé" seront respectivement représentées par $(0, 1, 0)$ et $(0, 0, 1)$. Bien entendu, les vecteurs de sortie, calculé par le réseau, ne seront pas nécessairement composées de valeurs entière 0 ou 1. Ce seront des triplets (a, b, c) de valeurs réelles comprises entre 0 et 1, chaque valeur étant une probabilité d'appartenir à une classe. Le vecteur d'entrée $x^{(n)}$ sera affecté à la classe la plus probable.

5.3 Evaluation des sorties du réseau

Une difficulté classique dans l'utilisation des réseaux de neurones (ou d'autres méthodes de régression) est le *sur-apprentissage* : avec suffisamment de paramètres et de cycles d'apprentissage, il est toujours possible de très bien adapter les poids à un jeu de données fixé. L'expérimentateur peut adapter parfaitement le réseau à un échantillon choisi de données, et être ensuite déçu par la qualité des prévisions du réseau sur un nouvel échantillon. Ce phénomène est d'autant plus sensible que la taille des échantillons de données est faible par rapport au nombre de paramètres du réseau.

Pour vérifier que nous ne faisons pas de sur-apprentissage, nous procéderons systématiquement de la façon suivante : nous choisirons aléatoirement un sous-ensemble des données (deux cinquièmes) que nous utiliserons pour l'apprentissage, le reste des données étant réservé à la validation des prévisions du réseau.

Une façon de procéder, pour évaluer les sorties du réseau de neurones, consiste à comparer le critère $E(w)$ que l'on cherche à minimiser durant l'apprentissage. L'inconvénient est que cela ne traduit pas l'influence du nombre de poids et de biais du réseau sur les résultats. Un réseau avec peu de poids capturera moins bien les variations de la réponse à une entrée x qu'un réseau comportant plus de poids, qui sera par contre plus sujet au sur-apprentissage. Dans la suite de ce papier, nous comparons différents modèles statistiques, avec des ensembles de variables d'entrée de tailles différentes. Donc, même en fixant le nombre d'unités de la couche cachée, *le nombre de poids du réseaux n'est pas identique d'un modèle à l'autre*, ce qui biaise les résultats sur le critère $E(w)$.

D'où l'idée d'utiliser le critère d'Akaike (AIC, *Akaike Information Criterion*) ou des critères dérivés tels le BIC de Schwartz (*Bayesian Information Criterion*). Le critère d'Akaike ([20]) s'écrit $AIC = 2\lambda - 2\ln(L)$, où λ est le nombre de paramètres non ajustés du modèle (i.e. le nombre de poids et de biais du réseau, dans notre cas), et $\ln(L)$ est la log-vraisemblance. Ce critère s'appuie sur la théorie de l'information, et en particulier la distance de Kullback-Leibler (K-L) établissant une distance entre un modèle candidat et le modèle "vrai". Parmi les critères dérivés de l'AIC, nous avons retenu le critère de Schwartz, qui multiplie le nombre de paramètres par un facteur de pénalité dépendant de la taille de l'échantillon de données : $BIC = 2\lambda \ln(N) - 2\ln(L)$.

A noter que les critères AIC et BIC sont des critères *relatifs*, permettant de comparer entre eux des modèles statistiques relatifs à un même modèle "vrai". Comme ce dernier est généralement inaccessible, le terme le représentant dans la distance de Kullback-Leibler est considéré comme une constante, et abandonné, dans les critères AIC et BIC.

Dans notre cas, les critères s'expriment de la façon suivante, N étant la taille de l'échantillon de données :

$$AIC = 2\lambda - 2 \sum_{n=1}^N \sum_{k=1}^C t_k^{(n)} \cdot \ln(y_k^{(n)}) \quad (6)$$

$$BIC = 2\lambda \ln(N) - 2 \sum_{n=1}^N \sum_{k=1}^C t_k^{(n)} \cdot \ln(y_k^{(n)}) \quad (7)$$

Nous utiliserons la moyenne du critère d'Akaike $AIC_{avg} = \frac{AIC}{N}$ pour comparer les prévisions des réseaux de neurones sur les deux jeux de données (apprentissage et validation), de taille différentes. Les réseaux de neurones étant particulièrement gourmands en nombres de paramètres, comparés à d'autres méthodes statistiques, il nous a semblé judicieux d'utiliser plutôt le critère de Schwartz (BIC) pour comparer les différents modèles statistiques de prévisions de l'état de regroupement. Nous prendrons également la

moyenne $BIC_{avg} = \frac{BIC}{N}$, en gardant à l'esprit que ce critère ne permet pas de comparer les ensembles d'apprentissage et de test, puisque la pénalisation sur le nombre de paramètres n'est pas la même, les ensembles étant de taille différente.

Les résultats numériques des critères AIC et BIC n'étant pas très parlants, nous nous intéresserons également aux taux global de classifications correctes, ainsi qu'aux taux par classe. Il faut être conscient, cependant, que le taux de classifications correctes *n'est pas* le critère optimisé par le réseau de neurones. Il faut donc faire preuve d'une certaine prudence dans l'interprétation de ce type de résultats. Malgré tout, ils donnent des ordres de grandeur intéressants, et présentent l'avantage d'être immédiatement compréhensibles.

6 Résultats préliminaires, choix des paramètres

Avant de nous intéresser à la sélection des composantes et des indicateurs significatifs, vérifions l'efficacité du type de réseaux de neurones que nous avons choisi, et discutons du choix des paramètres du logiciel `nnet`.

La figure 2 montre l'évolution de l'entropie croisée pendant la phase d'apprentissage. Les variables en entrée du réseau sont les six composantes principales de l'ACP et le volume du secteur. La couche d'entrée comprend donc 7 unités. La couche cachée en comprend 15, et la couche de sortie 3, chacune représentant une classe (cf section 5). Le nombre maximum de cycles d'apprentissage était fixé à 1500. L'apprentissage s'est en fait terminé au bout de 1010 cycle, l'algorithme ne pouvant plus améliorer le critère d'entropie de façon significative. La courbe a l'allure classique d'une courbe d'apprentissage. Le réseau choisi minimise bien l'entropie croisée, de manière efficace.

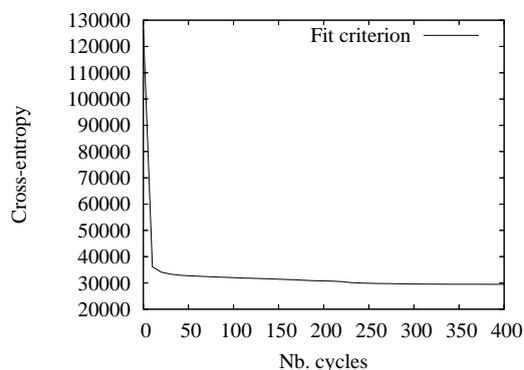


FIG. 2 – Evolution de l'entropie croisée selon le nombre de cycles d'apprentissage, avec les 6 composantes principales et le volume en entrée (réseau $I_2H_{15}O_3$)

Le choix des paramètres d'apprentissage, et la détermination de leur influence sur les résultats, est une question délicate. Dans cette étude, nous utiliserons les valeurs par défaut pour certains paramètres de `nnet` : les critères d'arrêt *abstol* et *reltol*, et le paramètre de dégradation des poids (*decay* = 0, par défaut).

L'intervalle dans lequel sont choisis les poids et les biais initiaux dépend des valeurs minimales et maximales prises par les variables en entrée, que nous avons normalisées. Une étude de paramètres, réalisée avec toutes les composantes et le volume en entrée, et avec 15 unités dans la couche cachée, montre que les meilleurs résultats (moyennés sur deux exécutions pour chaque valeur du paramètre d'intervalle) sont obtenus avec une valeur de paramètre *rang* = 0.4. C'est cette valeur que nous utiliserons dans la suite.

Il aurait pu être utile de faire d'autres études de paramètres, toujours en minimisant l'AIC ou le BIC, avec pour objectif de trouver le nombre optimal d'unités dans la couche cachée. Il faudrait cependant répéter cette démarche pour chaque ensemble de variables d'entrée testé, ce qui serait particulièrement fastidieux. Après quelques essais, nous nous sommes fixés à 15 unités dans la couche cachée.

7 Résultats sur les composantes

7.1 Sélection des composantes principales significatives

Afin de réduire le nombre de variables sur lequel il nous faut travailler, nous allons dans un premier temps nous intéresser aux composantes principales dégagées dans l'ACP, en y ajoutant le volume du secteur. Ce dernier n'était pas pris en compte dans l'ACP, mais, comme nous n'avons implanté aucun indicateur basé sur la géométrie du secteur, nous utiliserons le volume afin de nous rapprocher d'indicateurs du type "espace disponible autour des conflits", ou "distance à la frontière du secteur".

Une approche itérative est utilisée : nous considérerons d'abord la composante C_1 (représentative du nombre d'avions), puis en ajoutant le volume du secteur, puis les cinq autres composantes principales. Enfin, nous examinerons la totalité des 27 composantes, plus le volume.

Les critères AIC_{avg} et BIC_{avg} sont utilisés pour sélectionner le meilleur modèle. Un modèle (au sens statistique du terme) est un réseau de neurone appris, et un sous-ensemble de variables explicatives, dont nous pensons qu'il permettra d'expliquer au mieux l'état de regroupement des secteurs.

Pour chaque combinaison de variables explicatives, cinq exécutions de la phase d'apprentissage sont réalisées. En effet, la méthode d'apprentissage de `nnet` est une méthode d'optimisation locale, qui suit une pente de descente de la fonction d'erreur, à partir d'un point initial choisi au hasard (les poids et biais initiaux). La fonction d'erreur peut avoir plusieurs minima locaux. Même si ces minima peuvent être relativement proches, le résultat de l'apprentissage peut donc dépendre du point initial choisi. C'est pourquoi plusieurs exécutions conforteront nos résultats.

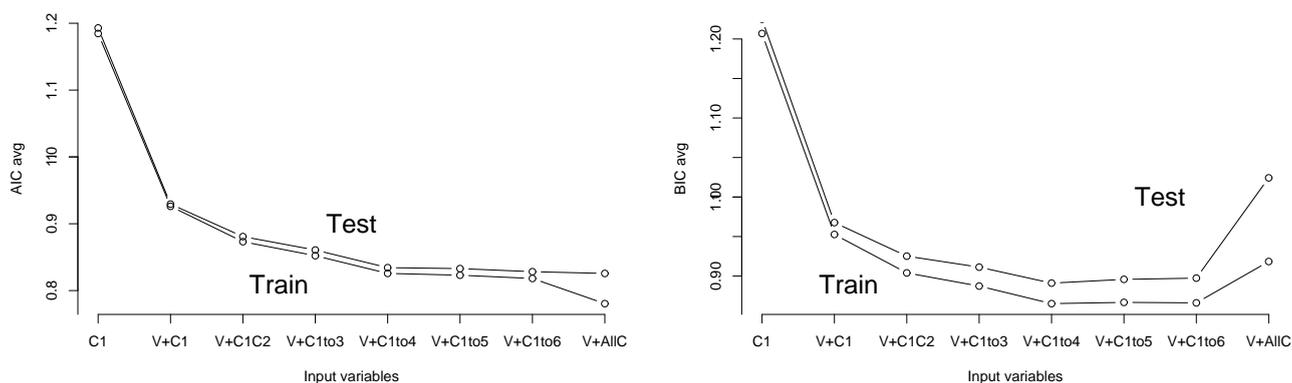


FIG. 3 – Evolution des critères AIC_{avg} (à gauche) et BIC_{avg} (à droite) pour les différentes combinaisons de composantes

La figure 3 montre, pour les données d'apprentissage (*train*) et de validation (*test*), l'évolution des critères AIC_{avg} (courbe de gauche) et BIC_{avg} (courbe de droite), au fur et à mesure que l'on ajoute des variables explicatives (composantes ou volume). Les valeurs moyennes sur les cinq exécutions sont présentées. Cette figure s'interprète comme suit : si le critère (AIC ou BIC) décroît de façon significative lorsqu'on ajoute une variable explicative, cela signifie que cette variable améliore la prévision de l'état de regroupement. Lorsque le critère croît ou stagne, cela signifie que le bénéfice apporté par la variable supplémentaire ne compense pas la complexité du nouveau modèle (augmentation du nombre de paramètres du réseau).

Remarquons, sur la figure de gauche, que l'évolution de l' AIC_{avg} est sensiblement identique sur les données d'apprentissage et de validation, à l'exception du dernier point, correspondant au modèle $\{V; C_1; \dots; C_{27}\}$. Sur ce point, l'AIC décroît pour les données d'apprentissage, alors qu'il stagne avec les données de validation, ce qui dénote un léger sur-apprentissage du réseau de neurones. Dans l'ensemble, on constate une grande cohérence entre les résultats sur les données d'apprentissage et de validation, ce qui montre que le réseau généralise correctement.

On constate, sur les deux figures, que l'introduction du volume V du secteur améliore considérablement les critères AIC et BIC, par rapport à la seule composante C_1 (facteur de taille, nombre d'avions). La composante C_2 , principalement liée à la variance des vitesses et aux évolutions verticales, améliore également les critères, de façon moins significative que le volume. C_3 (flux entrants) et C_4 (convergence des flux et anticipation des conflits) apportent aussi des améliorations, un peu moins importantes. Les composantes C_5 (flux divergents) et C_6 (surveillance des avions en proximité verticale) n'apportent aucun bénéfice supplémentaire.

Il peut sembler surprenant que la proximité des avions n'ait que peu d'influence sur l'explication de l'état de regroupement. Nous nous rallions à l'explication de [3], qui avait déjà constaté ce résultat : en situation normale, les avions horizontalement (resp. verticalement) proches ont déjà été séparés verticalement (resp. horizontalement), avant de se retrouver en situation de proximité (proche des normes de séparation). Ce sont donc des situations *déjà traitées* par le contrôleur. En revanche, l'anticipation de possibles situations futures de conflit (composante C_4) semble plus significative.

7.2 Taux de classifications correctes pour les composantes

Les tableaux 4 et 5 donnent les proportions de classification correctes faites par les réseaux de neurones (avec à chaque fois le meilleur réseau sur les cinq exécutions de la phase d'apprentissage), en utilisant respectivement les données d'apprentissage et les données de validation. La deuxième colonne présente le taux global, et les trois dernières colonnes donnent les taux spécifiques pour chaque classe. Les résultats sur les données d'apprentissage et de validation sont tout à fait cohérents, ce qui est le reflet des résultats de la section précédente.

Variables	Global	Regroupé	Armé	Dégroupé
$\{C_1\}$	72.91%	81.92%	44.90%	85.61%
$\{V; C_1\}$	79.34%	86.26%	59.01%	87.89%
$\{V; C_1; C_2\}$	80.70%	86.93%	60.79%	90.04%
$\{V; C_1; C_2; C_3\}$	80.84%	87.32%	60.23%	90.43%
$\{V; C_1; \dots; C_4\}$	82.03%	87.87%	63.02%	91.15%
$\{V; C_1; \dots; C_5\}$	81.80%	88.23%	61.77%	90.88%
$\{V; C_1; \dots; C_6\}$	81.83%	88.09%	62.16%	90.85%
$\{V; C_1; \dots; C_{27}\}$	83.36%	88.63%	65.61%	92.19%

TAB. 4 – Classifications correctes sur les données d'apprentissage

Variables	Global	Regroupé	Armé	Dégroupé
$\{C_1\}$	72.61%	81.64%	44.41%	85.40%
$\{V; C_1\}$	79.12%	85.56%	59.14%	88.12%
$\{V; C_1; C_2\}$	80.48%	86.58%	60.74%	89.82%
$\{V; C_1; C_2; C_3\}$	80.42%	86.71%	60.04%	90.08%
$\{V; C_1; \dots; C_4\}$	81.82%	87.78%	62.77%	90.74%
$\{V; C_1; \dots; C_5\}$	81.59%	88.19%	61.44%	90.49%
$\{V; C_1; \dots; C_6\}$	81.65%	87.98%	61.88%	90.64%
$\{V; C_1; \dots; C_{27}\}$	82.67%	88.04%	64.51%	91.68%

TAB. 5 – Classifications correctes sur les données de validation

Avec la seule composante C_1 en entrée, le réseau fournit déjà plus de 72% de prévisions correctes. Cependant, le taux pour la classe *secteur armé* n'est que de 45%. L'introduction du volume du secteur améliore nettement la prévision, avec plus de 79% pour le taux global, et environ 59% pour la classe *secteur*

armé. Avec le meilleur modèle trouvé lors de la sélection (les quatre premières composantes plus le volume), on obtient environ 82% de classifications correctes, et un taux de l'ordre de 63% pour les secteurs armés.

La dernière ligne des deux tables, avec toutes les composantes plus le volume, donnent de taux de classification encore meilleurs, alors qu'avec cet ensemble de variables en entrées, les critères AIC ou BIC stagnaient ou augmentaient. Ces résultats, apparemment meilleurs, ne signifient pas que le modèle regroupant toutes les composantes est plus performant que les autres. Ils sont simplement dûs au nombre de paramètres dans le réseau, nettement plus important dans ce cas.

Pour nous en convaincre, essayons un autre réseau pour le sous-ensemble $\{V; C_1; \dots; C_4\}$, avec un nombre de poids et de biais équivalent à celui du modèle $\{V; C_1; \dots; C_{27}\}$. Ce dernier s'appuyait sur un réseau $I_{28}H_{15}O_3$, comprenant 483 poids ou biais. Considérons maintenant, pour les quatre premières composantes et le volume, un réseau $I_5H_{53}O_3$, avec 53 unités dans la couche cachée. Ce réseau comprend 480 paramètres. Le tableau 6 montre que les taux de classification correctes avec ce réseau sont meilleurs que ceux obtenus avec toutes les composantes et le volume dans les tableaux 4 et 5.

Données	Global	Regroupé	Armé	Dégroupé
$\{V; C_1; \dots; C_4\}_{train}$	84.34%	89.74%	67.61%	91.88%
$\{V; C_1; \dots; C_4\}_{test}$	83.84%	89.32%	67.26%	91.09%

TAB. 6 – Taux de classification correctes pour le modèle $\{V; C_1; \dots; C_4\}$, avec un réseau $I_5H_{53}O_3$

Lorsque l'on s'intéresse aux résultats pour chaque classe, sur les trois tableaux, on constate que les taux pour les classes *regroupé* et *dégroupé* sont meilleurs que pour la classe *armé*. Ceci n'est pas une surprise, dans la mesure où le nuage de points, dans l'espace des variables, des mesures de la classes *armé* se situe "entre" les nuages représentant les deux autres classes. L'objectif du réseau de neurones est de déterminer les frontières entre ces nuages, ce qui présente une plus grande difficulté lorsqu'il y a deux frontières au lieu d'une seule.

8 Resultats sur les indicateurs

Le travail sur les composantes s'est avéré très utile sur le plan qualitatif, en permettant d'identifier les grandes classes d'indicateurs les plus significatifs. Cependant, pour ce qui est des applications futures, il n'est pas très pratique de calculer l'ensemble des 27 indicateurs, puis d'appliquer une matrice de transformation afin d'obtenir les composantes. Il serait utile de pouvoir prédire l'état de regroupement directement à partir d'un sous-ensemble d'indicateurs pertinents. C'est l'objectif de cette section.

8.1 Sélection des indicateurs significatifs

Nous allons sélectionner les indicateurs significatifs dans chaque composante du meilleur modèle $\{V; C_1; \dots; C_4\}$. Les ensembles initiaux de variables que nous allons considérer sont détaillés sur le tableau 3 de l'analyse en composantes principales. Notons que l'importance relative d'une variable au sein d'une composante (la corrélation linéaire avec l'axe de la composante) n'a pas nécessairement de rapport avec l'influence relative de cette variable sur la qualité de la prévision du réseau de neurones. En effet, la relation entre les variables explicatives et l'état de regroupement n'est pas *a priori* linéaire. Nous allons donc sélectionner à nouveau les variables selon le critère BIC.

Une procédure itérative est utilisée dans ce but. Nous avons choisi Nb , le nombre d'avions, pour représenter le facteur de taille du problème (composante C_1). Partant d'un ensemble basique de variables $\{V, Nb\}$, nous allons construire la meilleure séquence de variables pour C_2 , en choisissant à chaque étape la variable de C_2 qui donne la plus faible valeur du BIC. La courbe du BIC est alors tracée pour cette séquence, et le sous-ensemble de variables correspondant au minimum de cette courbe est ajouté à l'ensemble des variables pertinentes. Le processus est alors répété avec les variables de la composante suivante, en

prenant pour départ le nouvel ensemble de variables pertinente, et ainsi de suite jusqu'à ce que toutes les composantes utiles aient été traitées.

L'apprentissage des réseaux de neurones se fait sur les données d'apprentissage, et le BIC est calculé sur les données de validation. Cinq exécutions de la phase d'apprentissage sont effectuées pour chaque combinaison de variables testée. La meilleure valeur du BIC sur ces cinq exécutions est retenue et utilisée dans le processus de sélection.

Dans la section sur la sélection des composantes, les valeurs moyennes et les valeurs minimum donnaient les même résultats. Ici, nous avons constaté de petites différences entre les deux, probablement dues au fait que certains indicateurs d'une même composante peuvent donner des résultats très proches (comme chg_niv , $\frac{\sigma_{gs}}{gs}$, et σ_{gs}^2 dans la composante C_2 par exemple). Considérant que le meilleur réseau de neurones est le plus proche du "modèle réel", c'est celui que nous utiliserons dorénavant.

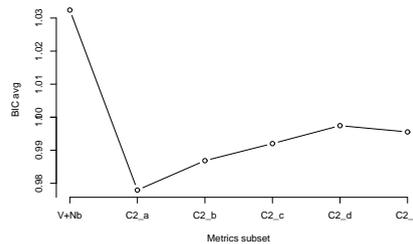


FIG. 4 – Evolution du BIC_{avg} pour la composante C_2 , avec la séquence chg_niv , $\frac{\sigma_{gs}}{gs}$, σ_{gs}^2 , N_{cl} , N_{des}

La figure 4 montre l'évolution du critère BIC_{avg} lorsque l'on ajoute successivement les variables de la composante C_2 , dans l'ordre de la meilleure séquence trouvée par la procédure de minimisation du BIC : chg_niv , $\frac{\sigma_{gs}}{gs}$, σ_{gs}^2 , N_{cl} , N_{des} . Les labels de l'axe des abscisses de cette figure sont détaillés dans le tableau 7.

Ensemble	Indicateurs
C2_a	V, Nb, chg_niv
C2_b	$V, Nb, chg_niv, \frac{\sigma_{gs}}{gs}$
C2_c	$V, Nb, chg_niv, \frac{\sigma_{gs}}{gs}, \sigma_{gs}^2$
C2_d	$V, Nb, chg_niv, \frac{\sigma_{gs}}{gs}, \sigma_{gs}^2, N_{cl}$
C2_e	$V, Nb, chg_niv, \frac{\sigma_{gs}}{gs}, \sigma_{gs}^2, N_{cl}, N_{ds}$

TAB. 7 – Sous-ensembles de la sequence présentée en figure 4 (comp. C_2)

Le minimum de la courbe est atteint pour la combinaison $\{V, Nb, chg_niv\}$. Cela ne veut pas dire que les autres variables de C_2 ne sont pas significatives : les résultats avec $\{V, Nb, \sigma_{gs}^2\}$ et $\{V, Nb, \frac{\sigma_{gs}}{gs}\}$ étaient très proches du meilleur résultat. La variance des vitesses et le ratio de l'écart-type des vitesses sur la vitesse moyenne sont donc simplement redondants avec la vitesse verticale moyenne (chg_niv). C'est également le cas pour le nombre d'avions en descente et le nombre d'avions en montée dans le secteur, bien que ces indicateurs soient légèrement moins performants que les autres.

Ces résultats montrent la pertinence des indicateurs liés à la variance des vitesses et aux changements d'altitude, et sont en contradiction avec d'autres résultats présentés dans [21]. Dans cette étude, Masalonis, Callahan, et Wanke appliquaient une régression logistique à des notes subjectives de complexité et des indicateurs de *densité dynamique*. La contribution de chaque indicateur était évaluée, pour chaque centre de contrôle, en retirant l'indicateur du modèle testé. En particulier, ils constataient que la vitesses et les changements d'altitude pouvaient être retirés sans dégrader significativement les prévisions faites par leur modèle.

Nos résultats vont à l'encontre de ces affirmations, puisque nous constatons que la variance des vitesses et les changements d'altitude améliorent le critère BIC. Nous n'avons pas d'explication évidente à cela. Il

est possible que, dans [21], le volume du secteur suffise à capturer le fait que certains secteurs sont dédiés à un certain type de trafic : les secteurs d'arrivée ou de départ, avec des avions évolutifs, sont généralement de plus petite taille que les secteurs purement en-route. Comme, pour notre part, nous utilisons des secteurs ou des groupes de secteurs de tailles diverses, et dans des états de regroupement variés (regroupé, armé, ou dégroupé), il n'y a pas de relation directe entre le volume du secteur et la nature du trafic dans ce secteur. La variance des vitesses et les changements de niveaux redeviennent alors pertinents dans ce contexte. Mais, comme nous n'avons pas les détails de l'étude [21], il est difficile d'être trop affirmatif.

Considérons maintenant les variables de la composante C_3 , en appliquant la même procédure, et en partant du nouvel ensemble de variables $\{V, Nb, chg_niv\}$.

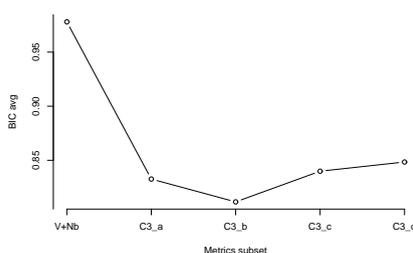


FIG. 5 – Evolution du BIC_{avg} pour la composante C_3

Ensemble	Indicateurs
C3_a	V, Nb, chg_niv, F_{60}
C3_b	$V, Nb, chg_niv, F_{60}, F_{15}$
C3_c	$V, Nb, chg_niv, F_{60}, F_{15}, F_5$
C3_d	$V, Nb, chg_niv, F_{60}, F_{15}, F_5, F_{30}$

TAB. 8 – Sous-ensembles de la séquence présentée en figure 5 (comp. C_3)

La figure 5 montre l'évolution du critère BIC_{avg} pour la meilleure séquence trouvée pour la composante C_3 (cf tableau 8). Le minimum est atteint pour le sous-ensemble $\{V, Nb, chg_niv, F_{60}, F_{15}\}$.

Le flux entrant à 15 minutes fait partie des indicateurs pertinents. Cela est cohérent avec le fait que les strips sont distribués sur les positions de contrôle environ 15 minutes avant que l'avion n'entre effectivement dans le secteur.

Nous avons par contre été assez surpris de constater que le flux entrant avec un horizon temporel d'une heure appartenait également à l'ensemble des variables significatives. Ce flux entrant est probablement moins directement lié aux variations du nombre d'avions, et il est possible qu'il permette de les lisser.

Le flux d'avions entrant dans les 5 prochaines minutes n'apporte par contre aucune amélioration à la prédiction. Les expérimentations montrent qu'en ajoutant F_5 à l'ensemble $\{V, Nb, avg_vs\}$, le critère BIC_{avg} augmente, alors que tous les autres flux améliorent la valeur du critère. Le flux F_5 est donc soit une variable fortement redondante avec les précédentes, soit pas du tout significative.

Ensemble	Indicateurs
$C4_a$	$V, Nb, chg_niv, F_{60}, F_{15}, inter_hori$
$C4_b$	$V, Nb, chg_niv, F_{60}, F_{15}, inter_hori, creed_pb$
$C4_c$	$V, Nb, chg_niv, F_{60}, F_{15}, inter_hori, creed_pb, creed_ok$
$C4_d$	$V, Nb, chg_niv, F_{60}, F_{15}, inter_hori, creed_pb, creed_ok, Conv$
$C4_e$	$V, Nb, chg_niv, F_{60}, F_{15}, inter_hori, creed_pb, creed_ok, Conv, insen_c$

TAB. 9 – Sous-ensembles de la séquence présentée en figure 6 (comp. C_4)

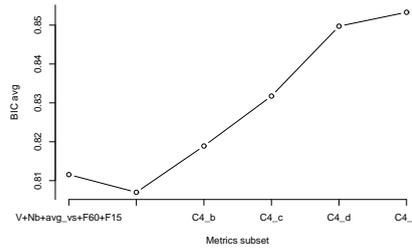


FIG. 6 – Evolution du BIC_{avg} pour la composante C_4

La figure 6 et le tableau 9 donne les résultats pour la composante C_4 . Le minimum pour cette composante est atteint lorsqu'on ajoute *inter_hori*. L'ajout des autres variables dans la meilleure séquence n'améliore pas le critère.

Il est intéressant de noter que les indicateurs représentatifs de l'*anticipation des conflits* pour les avions convergents, dans le plan horizontal, n'apporte aucun bénéfice supplémentaire lorsqu'on les ajoute à *inter_hori* et aux précédentes variables significatives. Ceci, et le fait que les indicateurs d'*anticipation de conflits* et de *convergence de flux* étaient déjà regroupés au sein d'une même composante par l'ACP, nous fait penser que la caractéristique principale captée par cet ensemble d'indicateurs (C_4) est le fait que les trajectoires des paires d'avions se croisent, ou non, à l'intérieur du secteur. L'anticipation elle-même des conflits ne semble pas être bien représentée, peut-être parce qu'on ne s'intéresse qu'aux situations de convergence dans le plan horizontal, et pas aux autres situations de conflit.

Pour conclure sur la sélection des indicateurs, la méthode que nous proposons aboutit à un sous-ensemble de seulement 6 indicateurs utiles $\{V, Nb, avg_vs, F_{60}, F_{15}, inter_hori\}$, parmi les 27, plus le volume du secteur, que nous avons programmé. Cette méthode nous a permis d'établir une relation directe (equation 5.1, avec les poids et les biais issus de l'apprentissage) entre ces 6 indicateurs et la probabilité d'appartenance à une classe d'état de regroupement.

8.2 Relation entre indicateurs et état de regroupement

Voici le détail de l'équation permettant de prédire l'état de regroupement à partir des indicateurs les plus pertinents. Nous donnons également les quelques lignes de code R permettant de calculer en pratique les sorties du réseau de neurones, avec le détail des matrices de poids.

L'équation 5.1 peut s'écrire sous la forme suivante, pour le meilleur réseau que nous avons trouvé :

$$(y_1, y_2, y_3)^T = \Psi\left(\sum_{j=1}^{15} w_{jk} \Phi\left(\sum_{i=1}^6 w_{ij} x_i + w_{0j}\right) + w_{0k}\right) \quad (8)$$

où le vecteur en entrée $(x_1, \dots, x_6)^T$ est l'ensemble des meilleurs indicateurs de complexité que nous avons sélectionné ($\{V, Nb, avg_vs, F_{60}, F_{15}, inter_hori\}$), et où le vecteur de sortie $y = (y_1, y_2, y_3)^T$ est le triplet des probabilités conditionnelles d'appartenance aux classes d'état de regroupement. Cela signifie que y_1 est la probabilité d'appartenir à la classe *regroupé*, sachant que le vecteur d'entrée est $(x_1, \dots, x_6)^T$. De même pour y_2 et y_3 , avec les classes *armé* et *dégroupé*, respectivement. Ψ est la fonction *softmax* (cf 5.2), et Φ est la fonction sigmoïde, ou logistique (cf 5.2).

Le code R donnant le triplet de probabilités (y_1, y_2, y_3) est donné sur la figure 7. Dans ce code, Mh and Mo sont des matrices de paramètres (poids et biais) alloués respectivement aux connections entre la couche d'entrée la couche cachée, et entre la couche cachée et la couche de sortie. Les biais sont gérés dans le réseau en ajoutant des neurones spécifiques, de sortie constante 1. Les biais sont alors de simples poids alloués aux connections entre de tels neurones et les unités de la couche suivante.

Le tableau 10 donne le détail des valeurs de Mh. La première colonne contient les biais w_{0j} . La seconde

```

f.softmax <- function(z)
  {exp(z)/sum(exp(z))};
f.logistic <- function(z)
  {1/(1+exp(-z))}

f.predict <- function(input) {
  f.softmax(Mo %*% c(1,
  f.logistic(Mh %*%
(c(1,input))))))
}

```

FIG. 7 – Code R pour la prévision de l'état de regroupement

	[,1]	[,2]	[,3]	[,4]	[,5]	[,6]	[,7]
[1,]	2.6700	0.6035	-3.2242	1.5125	2.84665	0.17517	1.99546
[2,]	1.3664	5.5459	1.2125	-0.5923	-2.27209	0.69194	-1.43059
[3,]	0.3902	0.6342	1.2399	-0.0718	0.54612	-0.03906	-0.92772
[4,]	0.5531	0.2000	0.3522	-0.0218	0.21848	-0.00494	-0.17895
[5,]	12.7531	-12.9649	3.8234	-0.1762	-0.05529	0.15634	-2.54624
[6,]	-2.5944	-0.5692	3.0737	-1.5888	-2.70640	-0.20288	-1.86718
[7,]	-62.6510	-112.4401	0.8557	0.3602	2.27203	-0.58279	-1.27146
[8,]	248.0508	696.6207	1.4739	-9.8654	0.63017	-0.11527	-4.27899
[9,]	-16.6421	-43.5024	-3.2625	-0.5521	-0.95558	-0.08815	3.29767
[10,]	-24.9896	138.5643	2.9702	-1.5819	0.66453	-0.50509	-4.00290
[11,]	-14.1631	14.4098	-3.6297	0.1643	0.17842	-0.16329	2.39568
[12,]	1.3726	5.4969	1.0101	-0.5810	-2.41572	0.70270	-1.25898
[13,]	-54.6044	-146.4793	-1.4978	-0.8585	-0.74717	-0.26718	2.03936
[14,]	-10.0972	0.6481	-4.0344	0.2354	-5.43306	0.54850	3.24417
[15,]	2.3865	1.9788	-0.3361	0.0353	0.20662	-0.02091	0.50016

TAB. 10 – Matrice Mh des poids et biais des connections entre la couche d'entrée et la couche cachée

donne les poids des connections entre l'entrée x_1 (volume V) et les 15 unités de la couche cachée, et ainsi de suite avec les autres colonnes.

Le tableau 11 décrit la transposée Mo^T (pour des questions de présentation) de la matrice Mo . La première colonne de Mo (première ligne de Mo^T) contient également des biais, et les 15 autres donnent les poids des connections entre les unités de la couche cachée et celles de la couche de sortie.

8.3 Taux de classifications correctes pour les indicateurs

Les tableaux 12 et 13 donnent les pourcentages de classifications correctes obtenus respectivement sur les données d'apprentissage et de validation, en ajoutant successivement les indicateurs les plus pertinents issus du processus de sélection. Les résultats obtenus sur les deux ensembles de données sont très similaires, ce qui était déjà le cas pour les composantes (cf section 7.2).

On constate que ces résultats sont sensiblement équivalents à ceux sur les composantes, voire un peu

	[,1]	[,2]	[,3]
[1,]	75.33756	195.42101	-269.39478
[2,]	46.33752	35.77978	-83.11712
[3,]	-77.57276	-52.77819	129.61192
[4,]	67.61727	-3.31538	-64.50745
[5,]	-200.23995	18.92759	181.49234
[6,]	-58.88296	-210.66951	270.42091
[7,]	47.64013	34.81970	-82.17057
[8,]	3.90553	0.69906	-4.47156
[9,]	-11.08421	-6.53394	17.31517
[10,]	7.13780	-5.45104	-2.09130
[11,]	-4.38653	2.88566	1.17121
[12,]	-59.72134	-210.62732	269.54483
[13,]	77.51679	51.23199	-127.99521
[14,]	-14.74073	-3.60412	18.72713
[15,]	240.72415	110.87175	-350.40952
[16,]	38.76275	-25.18119	-14.04946

TAB. 11 – Transposée de la matrice Mo des poids et biais des connections entre la couche cachée et la couche de sortie

Indicateurs	Global	Regroupé	Armé	Dégroupé
Nb	73.17%	83.96%	47.61%	80.36%
V, Nb	77.95%	87.87%	54.66%	84.34%
V, Nb, chg_niv	79.28%	87.85%	56.18%	87.82%
V, Nb, chg_niv, F_{60}	82.42%	88.05%	65.17%	90.15%
$V, Nb, chg_niv, F_{60}, F_{15}$	83.43%	89.33%	65.57%	91.29%
$V, Nb, chg_niv, F_{60}, F_{15}, inter_hori$	83.47%	89.07%	65.58%	91.91%

TAB. 12 – Evolution du taux de classifications correctes (données d'apprentissage)

Indicateurs	Global	Regroupé	Armé	Dégroupé
Nb	73.10%	83.54%	47.61%	80.53%
V, Nb	78.03%	87.69%	55.02%	84.33%
V, Nb, chg_niv	79.37%	87.54%	56.50%	88.13%
V, Nb, chg_niv, F_{60}	82.18%	87.76%	64.50%	90.26%
$V, Nb, chg_niv, F_{60}, F_{15}$	83.12%	89.02%	64.55%	91.55%
$V, Nb, chg_niv, F_{60}, F_{15}, inter_hori$	83.24%	89.14%	64.56%	91.75%

TAB. 13 – Evolution du taux de classifications correctes (données de validation)

meilleurs lorsqu'on prend tous les indicateurs, par rapport à toutes les composantes. Notons que l'on a plus d'indicateurs que l'on n'avait de composantes, et qu'en conséquence notre modèle contient plus de paramètres, ce qui peut expliquer cette amélioration (cf discussion sur l'influence du nombre de poids et de biais dans 5.3, avec l'exemple dans 7.2).

9 Discussion des résultats

Reprenons maintenant les principaux résultats dégagés par cette étude, en les comparant à d'autres études sur les indicateurs de complexité.

9.1 Nombre d'avions, volume du secteur, utilité des indicateurs supplémentaires

Tout d'abord, nous avons constaté que le nombre d'avions présents à l'instant t dans le secteur permet à lui seul d'obtenir environ 73% de classifications correctes sur l'état de regroupement des secteurs.

Le volume du secteur, que nous n'avons pas pris en compte dans l'ACP (note [6]) est également une variable explicative importante, avec 78% de classifications correctes en utilisant le nombre d'avions et le volume. L'utilité de cette variable explicative est d'autant plus importante que nous n'avons implanté aucun indicateur s'appuyant sur la géométrie du secteur ("distance au bord du secteur", "espace disponible pour la résolution de conflit", ...). Le codage de ces indicateurs pourrait constituer l'objet de travaux futurs.

L'ajout d'autres indicateurs, relatifs à la variance des vitesses et aux évolutions verticales, aux flux entrants, et à la convergence des flux, améliore encore les résultats. Cependant, ce gain supplémentaire est relativement modeste, de l'ordre de 5% supplémentaires sur les données de validation.

Des études précédentes ([3] et [2] en particulier) constataient déjà que les modèles utilisant le seul nombre d'avions n'étaient pas très performants dans la prévision de la charge de travail du contrôleur. Les gains constatés en ajoutant d'autres indicateurs étaient cependant plus importants que dans notre étude, en particulier sur les situations de trafic moyennement ou fortement chargées.

Dans [3], Chatterji et Sridhar affichent des taux de prédiction (sur les données de validation) allant de 16% à 54% pour la classe correspondant à une charge moyenne, et entre 0% et 100% pour la classe correspondant à une forte charge. Cependant, les résultats sur les données de validation ne sont pas cohérents avec ceux obtenus sur les données d'apprentissage. De leur propre aveu, il n'est pas possible de tirer des

conclusions raisonnables de ces résultats, étant donné qu'ils ne disposaient que de peu de points de mesures pour les situations de trafic moyennement ou fortement chargées.

Le papier de Kopardekar et Magyarits ([2]) indiquait également que le modèle utilisant la *densité dynamique*, vue dans cette étude comme une combinaison linéaire d'indicateurs, semblait être plus performant que celui basé sur le seul comptage d'avions. Les résultats montraient des différences significatives du critère R^2 de la régression, dans la comparaison des différents modèles. Kopardekar et Magyarits manquaient également de points de mesure, mais pour les situations de trafic très faibles, ou bien fortement chargées.

Les deux études ([3], [2]) s'appuyaient sur des notes subjectives, données par des contrôleurs aériens, à divers échantillons de trafic. Il semble que seules les périodes durant lesquelles les secteurs étaient armés aient été prises en compte⁵. Il est possible que l'utilisation de notes subjectives, en se focalisant sur le domaine normal d'utilisation du secteur, produise un "effet loupe" sur les phénomènes observés, et accentue les différences entre les différents modèles testés dans [3] et [2].

Une autre explication, plus simple et sans doute plus plausible, des différences relativement plus faibles que nous constatons entre nos différents modèles, réside dans la nature de nos données. Nous disposons d'un grand nombre de points de mesure, pour tous les états possibles (regroupé, armé, dégroupé), avec un grand nombre de points au centre de chaque classe, pour lesquels même le plus mauvais modèle donne de bons résultats de classification. Par exemple, s'il n'y a que 2 avions dans un secteur, celui-ci doit certainement être regroupé avec d'autres secteurs. À l'inverse, s'il y en a par exemple 60, le secteur a sans aucun doute été dégroupé précédemment. Nous aurions certainement constaté de plus grands écarts dans les résultats en calculant les taux de classification uniquement à partir de mesures collectées autour des instants de dégroupement ou de regroupement de secteurs.

Notons que nos points de mesure sont plus nombreux et mieux répartis entre les différentes classes que dans les autres études que nous avons citées. Nous disposons d'un éventail plus étendu de secteurs géographiques et de situations de trafic. En contrepartie, nous n'avons qu'une estimation beaucoup moins fine de la charge de travail dans le domaine normal d'utilisation des secteurs (état "armé").

9.2 Variance des vitesses, évolutions verticales

D'après nos résultats, les composantes ou les indicateurs liés à la variance des vitesses et aux évolutions verticales des avions apportent un gain dans l'explication des états de regroupement de secteurs. En supposant que ces derniers soient liés à la charge de travail du contrôleur, on peut donc dire que les évolutions verticales et la variance des vitesses sont un élément de cette charge de travail. Notons que ces indicateurs sont corrélés entre eux, et regroupés au sein d'une même composante.

Les travaux de [3] montraient que, sur l'unique secteur qu'ils considéraient, les évolutions verticales n'avaient pas d'influence sur la charge de travail, bien que les contrôleurs interviewés aient majoritairement déclaré qu'il s'agissait bien d'un facteur de complexité. En utilisant plusieurs secteurs, nous pouvons mieux discerner l'influence de ce facteur. Le fait que ces indicateurs ne soient pas pertinents sur un secteur donné, mais le deviennent sur plusieurs secteurs, laisse à penser qu'ils caractérisent essentiellement la spécialisation des secteurs : "arrivée" ou "départ" avec du trafic évolutif et de fortes variations des vitesses, ou bien *en-route* pur avec essentiellement du trafic stable en niveau et en vitesse.

Nos résultats vont également à l'encontre de ceux de l'étude de Masalonis, Callahan, et Wanke, comme nous l'avons déjà dit à la section 8. Il est possible qu'en travaillant uniquement sur le domaine d'utilisation normale du secteur (état "armé"), le volume suffise à lui seul à caractériser la spécialisation du secteur. Cela expliquerait pourquoi ils peuvent ôter la variance des vitesses et les changements d'altitude de leur modèle, sans en dégrader les performances.

Pour notre part, nous utilisons des secteurs ou des groupes de secteurs de tailles diverses, et dans des états de regroupement variés (regroupé, armé, ou dégroupé), et nous mesurons les indicateurs également en-dehors des limites d'utilisation normale du secteur. Le volume du secteur est donc très variable et il n'y a donc pas de relation directe entre le volume du secteur et la nature du trafic dans ce secteur. La variance des

⁵Cela n'est pas explicite dans les publications, mais le fait qu'ils manquent de points de mesures, notamment pour les situations chargées, nous incite à penser qu'ils ne considéraient que les secteurs dans l'état "armé", et pas "regroupé", ou "dégroupé"

vitesse et les changements de niveaux redeviennent alors pertinents. On peut toutefois se demander si une simple classification statique des secteurs en fonction de la nature du trafic ("arrivée", "départ", "en-route") ne conviendrait pas tout aussi bien.

9.3 Surveillance des proximités horizontales ou verticales

L'utilisation de la composante principale représentative de la surveillance des avions proches verticalement (C_6) s'avère sans effet, voire contre-productive. L'indicateur de proximité horizontale ($hprox_I$) n'apparaît pas dans l'ACP comme étant fortement corrélé à un axe significatif.

La surveillance des situations de proximité (verticale ou horizontale) ne semble donc pas être un facteur décisif. Etant donné que les couples d'avions horizontalement (resp. verticalement) proches ont déjà été séparés verticalement (resp. horizontalement), en situation normale (sauf incident), avant de se retrouver en situation de proximité, on peut concevoir que cet indicateur ne soit pas significatif.

Les travaux présentés dans [3] faisaient déjà ce constat, que nos résultats confirment sur des échantillons de données beaucoup plus larges et représentatifs : Chatterji et Sridhar ne disposaient que de données sur un seul secteur, et d'un faible nombre d'instances de trafic moyennement ou fortement chargé.

9.4 Flux entrants

L'utilisation des flux entrants, et plus particulièrement ceux à 15 et à 60 minutes d'horizon temporel, améliore les prévisions du réseau de neurones. Le flux d'horizon 30 minutes est probablement redondant avec ces deux flux. Le flux d'avions entrant dans les 5 prochaines minutes ne semble pas pertinent lorsqu'on l'ajoute au nombre d'avions présents dans le secteur, au volume de celui-ci, et à la vitesse verticale moyenne.

Dans [5], une rapide étude des distributions des valeurs du nombre d'avions et des flux entrants, autour des instants de dégroupements, montrait que les distributions "avant dégroupement" et "après dégroupement" étaient légèrement décalées pour le nombre d'avions, et sensiblement confondues pour les flux entrants. Cette étude sommaire soulignait l'importance du nombre d'avions dans les décisions de dégroupement, et la moindre importance des flux. Cependant, les indicateurs étaient étudiés isolément, et non pas en les combinant, et l'étude se focalisait sur des intervalles temporels autour des dégroupements, c'est-à-dire sur la zone frontière entre les nuages de points associés respectivement aux secteurs armés et dégroupés.

L'étude des combinaisons d'indicateurs, en considérant la totalité des points de mesures, nous a permis de mieux distinguer l'influence des flux entrants. Les résultats trouvés avec le réseau de neurones, sur la pertinence des flux à 15 minutes, sont tout à fait cohérents avec la pratique opérationnelle des contrôleurs, pour lesquels le nombre de strips est un élément important de décision. Les strips présents sur le tableau matérialisent à la fois le nombre d'avions présents dans le secteur, et les avions qui vont y entrer dans les dix, quinze prochaines minutes.

Par contre, il paraît surprenant au premier abord que le flux à 60 minutes apparaisse également comme un indicateur significatif. Nous pensons bien que le nombre d'avions présents dans le secteur, avec ses fortes variations au cours du temps, devrait être combiné avec un flux entrant plus stable afin de lisser ces variations et de mieux s'approcher des critères réels de décision de regroupement/dégroupement. Il semble que le seul flux à 15 minutes n'y suffise pas, et qu'il faille ajouter celui à 60 minutes, dont les variations sont plus amorties.

D'un point de vue plus général, il est évident que l'anticipation temporelle joue un rôle dans la décision de regrouper ou dégroupier un secteur. Mais on voit qu'il faut aussi tenir compte de la variabilité des indicateurs dans le temps. On conçoit aisément qu'une trop forte variabilité ne permette pas de prendre des décisions fiables de dégroupement/regroupement. Dans les perspectives de travaux futurs, il pourrait être utile de s'intéresser à la prévision temporelle d'autres indicateurs que le nombre d'avions, en les moyennant éventuellement sur une certaine durée afin d'en amortir les variations.

9.5 Convergence de flux, anticipation de conflits

La composante représentant la *convergence des flux* et l'*anticipation des conflits* apporte également un gain dans l'explication des états de regroupement. Les résultats que nous avons obtenus, tant dans l'analyse en composantes principales que dans la prévision des états de regroupement, tendent à montrer que la caractéristique principale de cette composante (C_4) est la convergence, ou la non-convergence des paires d'avions dans le secteur⁶, voire même l'aspect structurel lié au fait que les routes se croisent dans le secteur ou non.

L'aspect *anticipation de conflits* semble secondaire, sans doute parce que les indicateurs que nous avons considérés ne s'intéressent qu'aux conflits entre avions convergents, dans le même plan horizontal. D'autres indicateurs de *détection de conflits* (pas seulement pour les avions convergents, et pas seulement dans le plan horizontal) pourraient s'avérer utiles, et donner lieu à une composante plus orthogonale à celles issues des indicateurs que nous avons implantés pour l'instant. Cela pourrait être le sujet de travaux futurs.

L'indicateur le plus significatif parmi ceux de la composante C_4 , pour l'explication des états de regroupement, est *inter_hori*, c'est-à-dire le rapport du nombre de "croisements" horizontaux potentiels d'angle supérieur à 20° sur le nombre d'avions. Cet indicateur ne tient pas compte du sens de l'avion sur sa trajectoire, ni de la vitesse de rapprochement des avions. Il ne traduit donc pas véritablement la convergence, dans le sens d'un rapprochement des avions (comme *Conv* par exemple), mais plutôt le fait que les flux d'avions se croisent à l'intérieur du secteur, ou suivent des routes plus ou moins parallèles (en face-à-face ou en rattrapage).

Pour être plus clair, faisons le distinguo entre plusieurs notions différentes :

- la *structure du réseau de routes* (statique dans le temps) à l'intérieur du secteur (points de croisement, routes parallèles, ...). Nous n'avons pas implanté d'indicateur statique représentatif de la structure du réseau dans le secteur,
- la *structure générale des flux d'avions* évoluant dans le secteur, représentée par *inter_hori*. Contrairement à la notion précédente, celle-ci évolue dynamiquement dans le temps, mais elle ne tient pas compte de la convergence ou de la divergence des paires d'avions, mais seulement de l'angle des trajectoires,
- la *convergence* des paires d'avions (indicateurs *Conv*, *insen_c*). Cette notion est plus fine que les précédentes, dans la mesure où l'on examine si les avions sont effectivement en rapprochement,
- l'*anticipation des conflits*, pour les avions qui convergent dans le plan horizontal (indicateurs CREED).

Avec les indicateurs que nous avons implantés, la composante C_4 recouvre les trois dernières notions. On peut penser que la structure du réseau de routes est également plus ou moins en relation avec ces dernières. Il semble, pour l'instant et avec les données relativement bruitées dont nous disposons, que l'aspect "structure générale des flux" prédomine dans l'explication des états de regroupement, par rapport à la convergence des flux et à l'anticipation des conflits. Nous n'avons pas testé si un indicateur statique, représentant la structure des routes dans le secteur, ne donnerait pas d'aussi bons résultats.

9.6 Divergence des flux

La composante sur la *divergence des flux* ne permet pas d'améliorer pas les résultats obtenus avec les autres composantes. Il peut y avoir diverses raisons à cela. La divergence des flux peut ne pas être un facteur significatif. Ou bien elle peut être pertinente, tout en étant redondante avec les composantes précédentes, en particulier avec celle liée à la convergence des flux⁷.

Les résultats avec la composante C_4 (cf section précédente) montraient en effet que l'aspect "structure des flux", qui concerne aussi bien des avions convergents ou divergents, était prépondérant. Si l'on cherche à analyser plus finement, en comparant par exemple l'influence relative sur les états de regroupement de l'indicateur *inter_hori*, représentatif de la structure générale des flux, de l'indicateur *Conv* pour la conver-

⁶Notons que cette notion de convergence des paires d'avions est liée à la fois au nombre d'avions, et à la convergence des routes (statiques), ce qui explique les petits soucis d'interprétation de cette composante dans l'ACP.

⁷En effet, une situation de divergence n'est en général que le reflet décalé dans le temps d'une précédente situation de convergence.

gence des flux, et enfin Div pour la divergence, on constate (voir tableau 14) que le premier diminue encore la valeur du critère BIC_{avg} par rapport à l'ensemble de référence $\{V, Nb, chg_niv, F_{60}, F_{15}\}$, alors que les deux autres le dégradent.

Ensemble	BIC_{avg}
$\{V; Nb; chg_niv; F_{60}; F_{15}\}$	0.812
$\{V; Nb; chg_niv; F_{60}; F_{15}; inter_hori\}$	0.807
$\{V; Nb; chg_niv; F_{60}; F_{15}; Conv\}$	0.860
$\{V; Nb; chg_niv; F_{60}; F_{15}; Div\}$	0.843

TAB. 14 – Comparaison des indicateurs représentatifs de structure des flux, de la convergence, et de la divergence.

Il semble donc que la divergence des flux, tout comme la convergence, soit moins pertinents que la structure générale des flux. Celle-ci peut évoluer au cours du temps dans un secteur donné, puisqu'il peut y avoir à certains moments plutôt des vols dont les routes se croisent, et à d'autres moments des vols en routes parallèles.

A ce stade, il n'est pas possible de trancher si cet aspect dynamique est utile ou non à l'explication des états de regroupements, ou si nous pourrions nous contenter d'indicateurs statiques indiquant pour chaque secteur si les routes principales sont plus ou moins parallèles, ou se croisent dans le secteur.

9.7 Aspects structurels et dynamiques

A part le volume du secteur, tous les indicateurs que nous avons testés évoluent dynamiquement, en fonction des avions présents dans le secteur, de leurs trajectoires, de leurs évolutions verticales. L'idée sous-jacente est que la charge de travail instantanée dépend de l'évolution de ces paramètres dans le temps. C'est sans aucun doute le cas en ce qui concerne le nombre d'avions et les flux entrants.

Le volume du secteur est par contre un élément structurel de l'espace aérien, statique à l'échelle temporelle où nous nous situons. Nous avons vu que son utilisation parmi les variables explicatives améliorerait nettement les résultats. Dans la discussion sur les indicateurs des composantes C_2 (variance des vitesses, évolutions verticales) et C_4 (convergence des flux ; cf section précédente), nous en sommes venus à nous demander si l'aspect dynamique est réellement utile dans la relation entre ces indicateurs et l'explication des états de regroupement, ou si nous pourrions remplacer ces indicateurs par des indicateurs statiques, représentatifs de la structure de l'espace aérien.

Il apparaît important de faire la lumière sur ces questions. En effet, pourquoi calculer des indicateurs sophistiqués si l'on peut se contenter, *pour l'application qui nous intéresse*, du nombre d'avions, des flux entrants, du volume du secteur, et d'un ensemble d'indicateurs statiques relatifs au type de secteur :

- en-route, arrivée, départ,...
- route parallèles, ou sécantes.

auxquels on pourrait en rajouter d'autres (nombre de points de croisement, etc...)? Si tel était le cas, cela simplifierait le calcul des indicateurs, mais augurerait plutôt mal d'une utilisation ultérieure de méthodes du type *modèle dynamique de choix discret*, que nous envisageons pour une étude plus fine de la charge instantanée à partir des données de regroupement de secteurs.

En l'état actuel de notre étude, il n'est pas possible de trancher avec certitude sur cette question, du moins avec les données relativement bruitées dont nous disposons, et avec la méthode que nous avons utilisée. Il nous faudrait en premier lieu des données de meilleure qualité, que l'on pourrait obtenir en archivant la cause des changements de configuration des salles de contrôle. Ensuite, deux approches différentes, mais non exclusives, pourraient être explorées : d'une part tester des indicateurs statiques comme ceux que nous venons d'évoquer, ou des indicateurs "moyennés" (centres de *clusters* calculés sur les nuages de points des valeurs des indicateurs), et d'autre part tester une méthode prenant mieux en compte l'aspect dynamique des états de regroupements et des indicateurs et vérifier si elle améliore les résultats.

En tout état de cause, il faut se garder de tirer des conclusions hâtives sur des données aussi bruitées que les nôtres, et il n'est pas impossible que l'influence de certains indicateurs apparaisse plus nettement sur des données de meilleure qualité.

10 Conclusion et perspectives

En conclusion, la méthode originale que nous proposons, utilisant les données de regroupement de secteurs pour évaluer la pertinence des indicateurs de complexité, semble tout à fait praticable. Elle présente l'avantage, par rapport aux autres méthodes existantes, de s'appuyer sur des données faciles à obtenir, relativement objectives, et avec suffisamment de points de mesure et de variété géographique, même si elles sont peut-être plus bruitées que des mesures physiologiques, ou des notes subjectives de complexité.

L'application des réseaux de neurones aux indicateurs de complexité et aux regroupements de secteurs donne des résultats globalement assez bons (plus de 83% de classifications correctes), compte-tenu des biais et des bruits inhérents à nos données (diversité des causes réelles des dégroupements et regroupements, extrapolation des données radar IMAGE, etc...). On peut penser que dans la période choisie (mois de juin), qui est habituellement chargée, une bonne proportion des regroupements ou dégroupements de secteurs sont effectivement liés à la charge de travail, ce qui explique ces résultats. L'analyse des relations entre les indicateurs de complexité et la charge de travail pourrait toutefois être encore améliorée *si l'on connaissait plus précisément la raison de chaque changement de configuration de la salle de contrôle* (formation, maintenance, relève, trafic,...).

Parmi les indicateurs testés, il semble que l'on puisse se contenter de quelques-uns, relativement simples, pour expliquer les états de regroupements des secteurs, du moins pour les applications pratiques concernant le domaine d'utilisation de ces derniers, avec une échelle temporelle de l'ordre de la minute. Les indicateurs pertinents sont, à part le volume, ceux issus des composantes C_1 à C_4 de l'ACP. Plus précisément, notre processus de sélection a retenu les variables explicatives suivantes : le nombre d'avions, le volume du secteur, la vitesse verticale moyenne, les flux entrants à 15 et 60 minutes, et la proportion d'avions dont les trajectoires se croisent avec un angle supérieur à 20 degrés.

Ces résultats ne signifient pas que les autres indicateurs soient tous non-pertinents dans l'explication des états de regroupement. Cela peut être le cas pour certains d'entre eux, mais d'autres peuvent être simplement redondants avec les indicateurs sélectionnés. Par exemple, il est possible d'utiliser la variance des vitesses à la place de la vitesse verticale moyenne sans dégrader significativement la qualité prédictive du réseau. Ce n'est pas le cas si l'on utilise par exemple le flux entrant à 5 minutes à la place des autres flux sélectionnés.

Rappelons que les résultats que nous avons trouvés ne constituent pas une vérité définitive. Comme souvent en statistiques, il peut s'avérer que l'introduction d'autres variables explicatives (ou l'utilisation de données moins bruitées) fasse apparaître comme significatives des variables qui ne l'étaient pas précédemment, ou inversement. Citons l'exemple simple du volume du secteur. Si nous ne l'avions pas associé aux autres indicateurs, bon nombre de ceux-ci seraient certainement apparus comme nettement moins pertinents, voire pas du tout.

En dehors de la sélection des indicateurs les plus pertinents, nous avons également établi une relation simple entre ces derniers et les probabilités conditionnelles d'appartenance aux classes représentant les états de regroupement. Cette relation permet de prédire si le secteur est regroupé, armé, ou dégroupé, en fonction des valeurs des indicateurs sélectionnés.

Concernant la méthode employée, il apparaît que les réseaux de neurones sont un outil tout à fait adapté à notre problème, et probablement meilleur qu'une simple régression linéaire, puisque nous ne faisons pas d'hypothèse de linéarité entre les variables explicatives et la variable cible. D'autres méthodes statistiques, comme la régression logistique, sont envisagées et pourraient donner de bons résultats. Enfin, nous pourrions nous intéresser de plus près aux aspects dynamiques de la charge de travail du contrôleur, et d'autres méthodes devraient alors être envisagées (modèles dynamiques de choix discret).

En dehors de ces aspects méthodologiques, les perspectives de travaux futurs sont multiples. Nous pourrions notamment tester des indicateurs statiques relatifs à la typologie ou la géométrie des secteurs, ou rajou-

ter d'autres indicateurs dynamiques susceptibles d'apporter des améliorations, en introduisant par exemple des éléments supplémentaires sur la géométrie des flux dans les secteurs, sur la détection des conflits, ou encore sur la prévision temporelle des valeurs des indicateurs. Rappelons encore que les résultats présentés ne portent que sur les indicateurs que nous avons implantés, et qu'il en existe de nombreux autres dans la littérature, qui pourraient améliorer les résultats de cette étude. Signalons au passage la difficulté à implanter certains de ces indicateurs, définis de façon floue ou incomplète.

Parmi les applications directes, à court/moyen terme, de notre travail, nous envisageons l'élaboration de schémas optimaux d'ouvertures de secteurs ([22], [23]). A partir des trois probabilités produites par le réseau de neurones, nous souhaiterions obtenir un indicateur de charge de travail, ainsi que des valeurs seuils (correspondant aux décisions de regroupement ou de dégroupement), qui viendraient remplacer les flux entrants et les capacités actuellement utilisés dans la prévision pré-tactique de schémas d'ouvertures.

Références

- [1] G. M. Flynn, C. Leleu, and L. Zerrouki. Traffic complexity indicators and sector typology analysis of u.s. and european centres. Technical report, Eurocontrol, 2003.
- [2] 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.
- [3] 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.
- [4] 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.
- [5] D. Gianazza. *Optimisation des flux de trafic aérien*. PhD thesis, Institut National Polytechnique de Toulouse, 2004.
- [6] K. Guittet and D. Gianazza. Analyse descriptive des indicateurs de complexité du trafic aérien à partir des données image et courage. Note nr05-905, DSNA/DTI/SDER, Décembre 2005.
- [7] P. Kopardekar. Dynamic density : A review of proposed variables. Faa wjhtc internal document. overall conclusions and recommendations, Federal Aviation Administration, 2000.
- [8] Cognitive complexity in air traffic control, a litterature review. Technical report, Eurocontrol experimental centre, 2004.
- [9] 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.
- [10] 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 traffic Management R&D Seminar*.
- [11] I. V. Laudeman, S. G. Shelden, R. Branstrom, and C. L. Brasil. Dynamic density : An air traffic management metric. Technical report, 1999.
- [12] Note de synthèse sur l'indicateur de complexité pru. Technical report, DTI/SDER (ex CENA), 2005.
- [13] P. Averty. Conflit perception by atcs admits doubt but not inconsistency. In *Proceedings of the 6th Air Traffic Management Research & Developpment Seminar*, 2005.
- [14] P. Averty, K. Guittet, and P. Lezaud. *Work in progress, presented at an internal SDER seminar*. Technical report, DTI/SDER (former CENA), 2005.
- [15] C. M. Bishop. *Neural networks for pattern recognition*. Oxford University Press, 1996. ISBN : 0-198-53864-2.
- [16] B. D. Ripley. *Pattern recognition and neural networks*. Cambridge University Press, 1996. ISBN : 0-521-46086-7.

- [17] I. Rivals, L. Personnaz, G. Dreyfus, and J.L. Ploix. *Modélisation, classification et commande par réseaux de neurones : principes fondamentaux, méthodologie de conception et illustrations industrielles*. Lavoisier Technique & Documentation, 1995.
- [18] M. I. Jordan and C. Bishop. *Neural Networks*. CRC Press, 1997.
- [19] Snnns user manual, version 4.1. Technical report, University of Stuttgart.
- [20] W. H. Greene. *Econometrics analysis*. Prentice Hall International. ISBN : 0-13-015679-5.
- [21] A. J. Masalonis, M. B. Callaham, and C. R. Wanke. Dynamic density and complexity metrics for realtime traffic flow management. In *Proceedings of the 5th USA/Europe Air Traffic Management R & D Seminar*, 2003.
- [22] D. Gianazza and J. M. Alliot. Optimization of air traffic control sector configurations using tree search methods and genetic algorithms. In *Proceedings of the 21st Digital Avionics Systems Conference*, 2002.
- [23] D. Gianazza, J. M. Alliot, and G. Granger. Optimal combinations of air traffic control sectors using classical and stochastic methods. In *Proceedings of the 2002 International Conference on Artificial Intelligence*, 2002.
- [24] F. Chatton. Etudes de nouvelles métriques de complexité de la circulation aérienne. Master's thesis, Ecole Nationale de l'Aviation Civile (ENAC), 2001.
- [25] E. Pfeleiderer. 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 & Development Seminar*, 2005.

ANNEXE : Détail des indicateurs

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.

Les indicateurs de Chatterji et Sridhar

L'article [3] définit de nombreux indicateurs originaux. Nous n'avons reprogrammés dans le cadre de S2D2 que ceux qui ont été retenus par Kopardekar dans [2]. 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

$$vprox_1 = \frac{N}{\sum_{1 \leq j \leq N} \left(\frac{\sum_{1 \leq i \leq N} W_{ij} h_{ij}}{\sum_{1 \leq j \leq N} W_{ij}} \right)}, \quad (9)$$

où les W_{ij} sont des poids définis par

$$W_{ij} = \begin{cases} \frac{1}{d_{ij}^2 + S_h^2 h_{ij}^2} & \text{si } i \neq j, \\ 0 & \text{si } i = j. \end{cases} \quad (10)$$

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

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

$$vprox_2 = \frac{\sum_{1 \leq i \leq N} [K_i \neq \emptyset]}{\sum_{1 \leq i \leq N} \min_{j \in K_i} \{h_{ij}\}}, \quad (11)$$

où $K_i = \{j \mid d_{ij} \leq r, j \neq i\}$ et r est un paramètre qui définit le voisinage horizontal d'un avion⁸. Le paramètre r utilisé par les auteurs dans [3] 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".

⁸En réalité, le numérateur présenté dans [3] 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 [3].

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

$$\text{hprox}_1 = \frac{1}{\min_{1 \leq i \leq N} \{ \min_{j \in J_i} \{ d_{ij} \} \}} , \quad (12)$$

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 [3] 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".

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 [3] et [2], 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 quantités physiques facilement identifiables (distances horizontales et verticales) et à ne pas croître systématiquement avec le nombre d'avions actifs. On peut ainsi espérer qu'ils seront faiblement corrélés avec le volume du trafic dans le secteur considéré.

Les indicateurs inspirés par Delahaye et Puechmorel

L'article [9] 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 [2].

Notations

Avant de définir les indicateurs géométriques inspirés par [9] et [24], 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 [3].
- 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 :

$$v_{ij} = \frac{\langle \overrightarrow{X_i X_j}, \overrightarrow{v}_{ij} \rangle}{d_{ij}^{ob}} . \quad (13)$$

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'avions⁹). La fonction proposée est

$$f(d_{ij}^{ob}) = e^{-\alpha d_{ij}^{ob}} , \quad (14)$$

où α est un coefficient à déterminer. Mais dans [9], aucune valeur de α n'est suggérée. Dans le rapport [24], qui reprend les indicateurs de [9], les poids sont donnés (probablement de manière arbitraire) par

$$f(d_{ij}^{ob}) = \frac{e^{-\alpha (d_{ij}^{ob})^2} + e^{-\beta d_{ij}^{ob}}}{2} , \quad (15)$$

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.

⁹D'où l'on devine que les indicateurs considérées seront des indicateurs "doubles".

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" :

$$Dens(i) = \sum_{j=1}^N f(d_{ij}^{ob}). \quad (16)$$

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).

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 :

$$Div(i) = \sum_{\substack{j=1 \\ j \neq i}}^N 1_{\mathbb{R}^-}(v_{ij}) \cdot |v_{ij}| f(d_{ij}^{ob}), \quad (17)$$

$$Conv(i) = \sum_{\substack{j=1 \\ j \neq i}}^N 1_{\mathbb{R}^+}(v_{ij}) \cdot |v_{ij}| f(d_{ij}^{ob}). \quad (18)$$

La différence entre les deux indicateurs tient donc aux avions pris en compte dans le calcul¹⁰. 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.

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 rattachée à la sensibilité de la vitesse de convergence aux manoeuvres d'évitement possibles (changement de cap ou de vitesse). L'indicateur DOFI, introduit dans [7], essaie d'illustrer cette idée. Une autre approche est adoptée dans [9], 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, cf annexe de la note [6]) 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 [9] estiment la sensibilité de la vitesse de rapprochement aux manoeuvres d'évitement. Il s'écrivent

$$Sd_{-}(i) = \sum_{\substack{j=1 \\ j \neq i}}^N 1_{\mathbb{R}^-}(v_{ij}) \|\vec{v}_{ij}\| f(d_{ij}), \quad (19)$$

¹⁰Dans 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.

$$Sd_+(i) = \sum_{\substack{j=1 \\ j \neq i}}^N 1_{\mathbb{R}^+}(v_{ij}) \|\vec{\nabla} v_{ij}\| f(d_{ij}) . \quad (20)$$

où le gradient¹¹ 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 15 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. 15 – Statistiques descriptives pour les sensibilités "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 régressions¹², 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 relativement 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 [9] : 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,

¹¹Dans [9], 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.

¹²C'est d'autant plus le cas ici que les composantes des densités sont très fortement corrélés : les termes diagonaux de la matrice de corrélation des composantes de la sensibilité "divergente" sont 0.71, 0.75 et 0.87.

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 [9]), "insen_c" (on définit "insen_d" de façon similaire) :

$$\text{"insen_c"} = \frac{\text{Conv}^2}{\text{Sd}_+} . \quad (21)$$

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 [9].

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 [9], 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 [3], il n'y a pas dans [9] le souci de décorrélérer 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.

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 contrôlés¹³.

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

- "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 n_1 , n_2 et n_3 . L'indicateur ("simple") est alors donné par

$$\text{inter_vert} = \frac{(n_1n_2 + n_2n_3 + n_1n_3)}{(n_1 + n_2 + n_3)} . \quad (22)$$

- "chgt_niv" (adaptation de l'indicateur "changements de niveaux") : il s'agit de la vitesse verticale moyenne dans le secteur¹⁴. Il s'agit d'un indicateur "simple".

A noter que l'indicateur "inter_hori" adapté de [12] est extrêmement 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 [12] : en effet, même si l'indicateur instantané calculé pour S2D2 s'inspire de celui introduit dans [12], 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

¹³Ce jugement est soutenu les résultats de [25].

¹⁴Là 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.

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 indicateur¹⁵.

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 [13]. 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 [9]) 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 [13], 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 [14] (qui reprend la base de données de [13]) 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 [13] : 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 pour S2D2 :

$$\frac{1}{\alpha Da + (1 - \alpha)(\beta Ed + (1 - \beta) Efl)} , \quad (23)$$

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 [14]. 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 [14]...
- En implémentant cet indicateur, on a constaté une difficulté : les situations étudiées dans [13] (et reprises dans [14]), 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 degré 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 [14].

Les indicateurs de type "CREED" reprennent donc les définitions données en (23) 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 [13]). Dans DD_INDIC, ces indicateurs ("doubles") sont notés "creed_bon" et "creed_pb".

¹⁵Cela étant dit, le calcul est immédiat à partir de la variance des vitesses.