

**Revue statistique du programme
de surveillance du balbuzard
dans la zone d'entraînement à
basse altitude au Labrador et à
l'est du Québec**

Préparée pour :

Institut pour la surveillance et la recherche environnementales

Case postale 1859, succursale B
114, chemin Hamilton River
Happy Valley-Goose Bay, Labrador
A0P 1E0

Par :

Jacques Allard, Ph.D.
Analyse statistique Atlantique Inc.
21, rue Colburne
Moncton N.-B. E1C 8Z9

Téléphone : 506-866-2655
Courriel : mathstat@nbnet.nb.ca

2008 09 19

1	INTRODUCTION	4
2	HISTORIQUE DU PROGRAMME DE SURVEILLANCE DU BALBUZARD	5
2.1	Information révisée	5
2.2	Chronologie de la surveillance du balbuzard dans la ZEBA.....	5
3	DESCRIPTION DU PROGRAMME DE SURVEILLANCE EN 1999-2007.....	7
3.1	Méthode d'échantillonnage.....	7
3.2	Résultats observés.....	7
3.2.1	Occupation des nids (activité des nids)	7
3.2.2	Réussite de nidification et productivité de nidification.....	7
3.3	Analyse statistique.....	8
3.3.1	Emploi de plusieurs méthodes statistiques au cours des années	8
3.3.2	Seuils d'importance.....	8
3.3.3	Intervalles de confiance	9
4	MISE EN REVUE DU PROGRAMME DE SURVEILLANCE EN 1999-2007	10
4.1	Étude par observation	10
4.2	L'interprétation du taux d'occupation de nids	10
4.3	Méthode d'échantillonnage.....	13
4.4	Hypothèse alternative.....	14
4.5	Puissance et seuil d'importance statistique.....	14
4.5.1	Erreur de type I (faux positifs).....	14
4.5.2	Erreur de type II (faux négatifs) et puissance	15
4.6	Tests multiples.....	16
4.7	Tests statistiques pour les comparaisons sur une seule année	17
4.7.1	Impact de l'approche d'échantillonnage par quotas sur l'estimation du taux d'occupation de nids	17
4.7.2	Test sur les taux d'occupation de nids et du succès de nidification.....	17
4.7.3	Variable discrète : Productivité de nidification	20

5	MODÈLES SUR PLUSIEURS ANNÉES.....	25
5.1	Modèle de succès de nidification selon la présence/absence d'activité d'entraînement à basse altitude.....	25
5.2	Modèle de succès de nidification selon l'intensité d'exposition ou d'activité d'entraînement à basse altitude.....	26
5.3	Modèle de productivité de nidification selon la présence/absence d'activité d'entraînement à basse altitude.....	27
5.4	Modèle de productivité de nidification selon l'intensité d'exposition ou d'activité d'entraînement à basse altitude.....	27
6	ANALYSE DES RÉSULTATS EN 1999-2007.....	29
6.1	Réussite de nidification.....	29
6.2	Productivité de nidification.....	31
6.3	Taux d'occupation de nids.....	32
7	CONCLUSION.....	34
7.1	Interprétation des résultats actuels.....	34
7.2	Circonstances spéciales.....	34
7.3	Surveillance périodique.....	35
7.4	Organisation de la base de données.....	35
7.5	Tests statistiques.....	35
7.5.1	Analyses sur une seule année.....	35
7.5.2	Analyses sur plusieurs années et l'impact de l'intensité.....	35
7.5.3	Probabilité d'erreurs de type I et de type II.....	36
7.5.4	Intervalles de confiance.....	36
8	REMERCIEMENTS.....	37

1 Introduction

Le mandat de l'Institut pour la surveillance et la recherche environnementales (ISRE) vis-à-vis la zone d'entraînement à basse altitude (ZEBA) au Labrador et à l'Est du Québec inclut la responsabilité « de fournir des vérifications indépendantes des effets environnementaux » liés à l'entraînement à basse altitude (EBA). Dans ce mandat, l'Institut a entrepris la surveillance des activités de reproduction du balbuzard pendant plusieurs années.

Le document actuel inclut une mise en revue de la méthodologie statistique utilisée jusqu'à date dans le Programme de surveillance du balbuzard et une mise en revue statistique des résultats accumulés. Il inclut aussi des recommandations sur la méthodologie statistique et l'importance de ne pas arrêter le programme.

2 Historique du programme de surveillance du balbuzard

2.1 Information révisée

L'information suivante a été utilisée :

- Un aperçu de l'histoire de la surveillance de nids de balbuzards dans la ZEBA.
- Une base de données des observations de nids de balbuzards entre 1994 et 2002 (Nest_outcomes_1994-2002.xls).
- Une base de données plus large des observations de nids de balbuzards entre 1999 et 2001 (avec les observations tirées de l'article précédent) (Nest_outcomes_1999-2001.xls).
- Une base de données des observations de nids de plusieurs espèces, y compris le balbuzard, en 1998 (Osprey_nests_98_mod.xls).
- Des rapports annuels sur le programme « Surveillance des balbuzards dans la zone d'entraînement à basse altitude au Labrador » entre 1999 et 2007.
- Un rapport préliminaire de l'atelier du balbuzard qui a eu lieu en 2003 à St John's, T.-N.-L., sur la surveillance post-migratoire des balbuzards dans la zone d'entraînement à basse altitude au Labrador.
- Une description de plusieurs programmes de surveillance fournie au réviseur par l'ISRE.

2.2 Chronologie de la surveillance du balbuzard dans la ZEBA

L'ISRE fait des rapports sur plusieurs études sur la nidification du balbuzard. Le Tableau 1 présente un résumé des études et les résultats décrits dans plusieurs documents remis au réviseur.

Tableau 1. Historique sur la surveillance du balbuzard vis-à-vis l'activité d'entraînement à basse altitude		
Période	Description	Résultats
1991	Le Ministère de la Défense nationale (MDN) débute un programme annuel de surveillance des nids de balbuzards pour identifier les nids actifs dans un programme d'atténuation qui établit une zone d'exclusion de 2,5 nm autour des nids.	
1994	Une comparaison de la réussite de nidification et de la survie entre 5 nids à l'intérieur et 5 nids à l'extérieur de la zone d'exclusion de 2,5 nm.	On ne retrouve pas de différences statistiques importantes; il n'y a pas de rapport sur la probabilité d'erreur de type II.
1995, 1996	La réussite de nidification et la survie est estimée encore une fois.	Les résultats ne sont présentés.
1996-1998	Il y a plusieurs études sur la réussite de nidification en fonction de la fréquence de vols ou la taille de la zone d'exclusion.	Les résultats ne sont présentés.
1999-2007	Étude systématique de l'activité de nidification, la réussite de nidification et la productivité des nids dans une zone secondaire de la ZEBA et dans une zone correspondante en dehors de la ZEBA.	Les résultats de cette étude sont à la base de la mise en revue actuelle et sont décrits et discutés en détail ci-dessous.

L'ISRE a aussi organisé un atelier sur le programme de surveillance du balbuzard en octobre 2003.

3 Description du programme de surveillance en 1999-2007

Le programme de surveillance en 1999-2007 semblerait représenter le programme le plus complet parmi ceux qui ont été indiqués. Maintenant, nous décrivons ce programme en détail.

3.1 Méthode d'échantillonnage

L'échantillonnage s'effectue par hélicoptère. La méthode d'échantillonnage est décrite ci-dessous dans le rapport sur le programme de surveillance du balbuzard dans la ZEBA en 2007 :

« Avant le départ, on a choisi une route qui maximiserait les nids rencontrés (aléatoire), tout en minimisant la distance/temps de voyage (et les coûts associés). L'objectif était de retrouver un échantillon de 30 nids actifs dans chaque zone d'étude. Il faut signaler que cet échantillon peut inclure de nouveaux nids retrouvés pendant les études. Ceci étant dit, il se peut que les nids utilisés en 2007 n'aient pas été inclus dans les années précédentes de surveillance. »

M. Perry Trimper, le gestionnaire de projet, a remis plus de détails au réviseur.

Le trajet d'échantillonnage est axé sur des sites qui avaient des nids antérieurement et commence avec les sites les plus rapprochés de Goose Bay. Tous les nouveaux sites qui étaient inconnus antérieurement sont inclus dans l'échantillon de cette année et ont été ajoutés à la base de données pour les années plus récentes. L'échantillonnage continue jusqu'à ce que 30 nids actifs soient retrouvés, allant au-delà des sites dans la base de données, s'il y a lieu.

3.2 Résultats observés

Trois résultats ont été analysés dans le programme de surveillance du balbuzard dans la ZEBA :

- L'activité de nidification, définie comme la présence de signes d'élevage dans un nid. Nous préférons utiliser l'expression « occupation du nid ».
- La réussite de nidification, définie uniquement pour les nids actifs, est la présence d'au moins un jeune à l'envol.
- La productivité de nidification, définie uniquement pour les nids actifs, est le nombre de jeunes à l'envol.

3.2.1 Occupation des nids (activité des nids)

Pour un nid disponible, l'occupation d'un nid se définit par la présence/l'absence d'indices liés à l'activité d'élevage à partir du nid. Pour chaque nid, l'occupation de nids représente une variable dichotomique (binaire). Le taux d'occupation de nids (taux d'activité de nids) est signalé.

3.2.2 Réussite de nidification et productivité de nidification

Les deux résultats, la réussite de nidification et la productivité de nidification, ont une forte relation. La réussite d'un nid se définit avec un (1) ou plusieurs jeunes à l'envol, tandis qu'un nid échoué se définit avec aucun jeune à l'envol. La productivité d'un nid

actif se définit comme le nombre de jeunes à l'envol observés. D'habitude, ce chiffre est entre 0 et 3, car 4 est rarement observé dans les zones étudiées ici.

3.3 Analyse statistique

3.3.1 Emploi de plusieurs méthodes statistiques au cours des années

Dans le programme de surveillance en 1999-2007, la différence entre les taux d'occupation de nids et la réussite de nidification a été testée en utilisant la méthode exacte de Fisher à intervalle unilatéral jusqu'en 2005 et une analyse de régression logistique (sur une variable nominale qui représente la zone de contrôle / d'expérimentation) par la suite. Le changement n'a pas reçu beaucoup d'importance.

Nous remarquons plusieurs erreurs et inconsistances dans les sections statistiques des rapports. Voici des exemples d'erreurs ou d'inconsistances :

- Dans le rapport de 2003, le Tableau 2, la valeur P de la méthode exacte de Fisher à intervalle unilatéral pour le taux d'occupation de nids en 2003 (activité de nidification) est signalée à 0,109 : il semblerait que le mauvais intervalle a été utilisé dans l'essai – la bonne valeur est 0,957.
- Dans le rapport de 2004, le Tableau 4, les résultats indiqués pour 1999-2003, tout comme ceux de la méthode exacte de Fisher à intervalle unilatéral, sont différents par rapport à ceux qui ont été indiqués en 2003. En fait, contrairement à l'énoncé dans la méthodologie et les en-têtes de colonnes, les valeurs P du rapport de 2004 sont les suivantes : pour le succès de nidification en 2004 seulement, la valeur P de la méthode exacte de Fisher à intervalles bilatéraux (0,209) est signalée; pour tous les autres taux d'occupation de nids (activité de nidification) et les cas de réussite de nidification, la valeur P de régression logistique unilatérale est signalée.
- Dans le rapport de 2007, le Tableau 4, la valeur P de régression logistique à intervalle bilatéral pour le taux d'occupation de nids (activité de nidification) en 2007 est signalée à 0,578. La bonne valeur est 0,766, tel qu'indiqué dans l'Annexe B, section 2. (La mauvaise valeur semble être celle de l'analyse de « toute l'année » illustrée dans l'Annexe B, section 1.)

La productivité a été essayée en utilisant un test T à deux échantillons. Bien que ce ne soit pas spécifié dans la méthodologie, les vérifications aléatoires nous font croire que la version à variance équivalente (homoscédastique) du test a été utilisée.

La documentation indique que d'autres méthodes ont été utilisées pendant les années, y compris l'analyse de la variance. Les rapports du programme de surveillance en 1999-2007 ont aussi inclut des résultats provenant d'autres analyses qui se trouvent dans les annexes.

3.3.2 Seuils d'importance

Le seuil d'importance statistique choisi est de $\alpha=0,20$, ce qui mène à une probabilité d'erreur de Type I (faux positifs) de 20 %. Il semblerait que ce choix a été pris pour limiter la probabilité d'erreur de Type II (faux négatifs).

3.3.3 Intervalles de confiance

Malgré le fait que les estimations des paramètres (taux d'occupation de nids, réussite de nidification et la productivité de nidification) ont été remises, les intervalles de confiance ne sont pas fournis, ni pour les paramètres, ni pour les différences de paramètres entre les valeurs expérimentales et de contrôle.

4 Mise en revue du programme de surveillance en 1999-2007

Dans cette section, nous repasserons le programme de surveillance en 1999-2007, ainsi que ses résultats.

4.1 Étude par observation

L'étalon-or dans le monde des recherches sur l'effet d'une expérience est « l'étude sur échantillon aléatoire et contrôle », où les sujets ont des tâches aléatoires pour un groupe expérimental et pour un groupe de contrôle. Les tâches aléatoires garantissent que toute différence observée dans le résultat entre les groupes est causée soit par l'expérience ou le hasard. La théorie sur la probabilité est utilisée pour quantifier la probabilité d'un résultat observé uniquement par hasard (le niveau d'importance statistique observé ou la « valeur P »). Si cette probabilité est petite (0,05 est le seuil de probabilité traditionnel), le chercheur aura une tendance à conclure que la différence observée est causée par l'expérience et que la relation est une de cause-effet.

Dans plusieurs cas, il est impossible d'entreprendre une expérience contrôlée aléatoire. Dans ces cas, on entreprend souvent une « étude par observation ». Une comparaison entre un groupe de contrôle existant (les sujets qui ne reçoivent pas l'expérience) et un groupe expérimental existant (les sujets qui reçoivent l'expérience) est un exemple d'une étude par observation. Le programme de surveillance du balbuzard dans la ZEBa est une étude par observation.

Les études par observation peuvent être affectées par des biais. Il y a un biais lorsqu'un facteur qui diffère de l'expérience provoque une différence dans le résultat. Un biais peut être évident (le chercheur est au courant du biais) ou caché (le chercheur n'est pas au courant du biais). Les biais évidents peuvent parfois être corrigés (par exemple, en utilisant une analyse stratifiée). Les biais cachés ne peuvent pas être corrigés.

Dans le programme de surveillance du balbuzard dans la ZEBa, le groupe expérimental consiste de sites de nidification dans la zone secondaire de la ZEBa et le groupe de contrôle consiste de sites de nidification dans une zone adjacente qui se trouve à l'est de la zone expérimentale. Les rapports du programme de surveillance du balbuzard dans la ZEBa n'indiquent pas qu'il y a eu des biais évidents ou des corrections des biais évidents. Cependant, on peut supposer qu'il y avait des biais cachés. Par exemple, les zones expérimentales et de contrôle sont adjacentes et la zone de contrôle se trouve immédiatement à l'est de la zone expérimentale. La relation entre l'est et l'ouest dans les groupes suggère qu'il pourrait y avoir des différences systématiques dans les patrons climatiques (y compris les vents). La géographie, y compris la configuration des rivières et des lacs (sources alimentaires), les formations rocheuses et les terrains boisés (disponibilité de sites de nidification) pourrait aussi être systématiquement différente.

4.2 L'interprétation du taux d'occupation de nids

Nous considérons que l'occupation de nids peut être difficile à interpréter. Un déclin important d'occupation de nids indiquerait sûrement qu'il y a eu un changement important et soudain dans le nombre de paires de nidification. Cependant, réciproquement, il se peut qu'un déclin progressif dans le nombre de paires de nids ne reflète pas un déclin progressif d'occupation de nids.

Nous estimons qu'il serait difficile de représenter un modèle adéquat de la relation entre le taux de changement dans la population et le taux d'occupation de nids. Le taux d'occupation de nids dépend de la durée de vie des nids (la probabilité de survie d'un nid occupé ou non occupé pendant 1, 2... ans) et de la tendance de la population pour bâtir des nouveaux nids. Le problème se complique davantage à cause des facteurs externes : par exemple, les nids bâtis sur des affleurements rocheux ou des structures d'origine humaine vont probablement avoir des durées de vie plus longues par rapport aux nids bâtis au sommet des épinettes plus hautes. Par conséquent, un changement du taux d'occupation de nids pourrait être causé par un changement dans la durée de vie d'un nid (par exemple, un changement dans les vents) ou par un changement dans les tendances de population pour bâtir des nids.

À notre connaissance, l'information détaillée requise pour construire un modèle comme celui-ci n'est pas disponible.

Ces difficultés s'accumulent si on compare deux groupes, puisque les différences de paramètres du modèle peuvent devenir des facteurs déconcertants.

En combinant les problèmes d'échantillonnage (section suivante) et les problèmes qui peuvent être interprétés, nous concluons que le taux d'occupation de nids ne sera sûrement pas un indicateur très utile vis-à-vis l'impact des vols à basse altitude sur la reproduction du balbuzard.

Le réviseur principal a indiqué que les rapaces ont une tendance à occuper des « territoires » (par exemple, un affleurement rocheux au-dessus d'un bon endroit pour la pêche) et que la statistique intéressante serait le « taux d'occupation territorial », ce qui suggère que le taux d'occupation de nids aurait pu être pris implicitement en tant que statistique de substitution pour le taux d'occupation territorial.

Bien sûr, même en tant que substitution pour le taux d'occupation territorial, le taux d'occupation de nids souffre des problèmes décrits ci-dessus. Cependant, l'observation du réviseur suggère que le nombre de nids actifs dans certaines zones choisies à l'avance (par exemple, des territoires) pourrait être un meilleur indice de la population. Ces zones correspondraient à des petites zones occupées historiquement, comme une partie d'une berge de fleuve (disons, une profondeur d'un kilomètre) ou une formation rocheuse spécifique qui peut être dépistée systématiquement à tous les ans. Cette approche serait similaire à l'échantillonnage de poissons marins en utilisant des postes fixes (par exemple, la somme normalisée de poissons provenant du chalutage qui se fait au même endroit à tous les ans). Cependant, « l'échantillonnage d'un poste fixe » peut mener à des conclusions erronées lorsque la population se déplace même légèrement dans cette zone d'observation.

Détails techniques

Pour illustrer la difficulté dans l'interprétation du taux d'occupation de nids, nous avons entrepris une simulation de deux populations, une stable et une en déclin. Les deux populations ont commencé avec environ 400 paires dans des nids et 800 nids (c'est-à-dire, une occupation de 50 %). La quantité de paires par nid est restée à environ 400 dans la population stable, tandis qu'il y a eu un déclin d'environ 10 % par an dans la population en déclin. Nous avons fait des suppositions raisonnables sur la durée de vie des nids non utilisés et sur la propension de paires dans les nids qui bâtissent des nouveaux nids. La figure 1 illustre le résultat d'une simulation unique de 30 ans pour chaque population.

Pour la population stable, le nombre de paires dans des nids, le nombre de nids et l'occupation de nids reste relativement constante. Pour la population en déclin, le nombre de paires dans des nids et le nombre de nids deviennent négligeables après 20 ans. Cependant, le taux d'occupation de nids reste plus ou moins pareil (mais en moyenne, il est au-dessous de 10 % par rapport à la population stable).

Ce résultat est causé par l'effondrement éventuel de nids non utilisés : dans une population en déclin, le nombre de nids disponibles subira aussi un déclin et le taux d'occupation de nids restera constant.

Si tout le reste est équivalent [ceteris paribus], le taux d'occupation de nids est lié au taux de changement dans l'effectif de la population (et non pas à l'effectif de la population en soi).

Ceteris paribus, si une population commence un déclin continu à un rythme constant, l'occupation de nids montrera un déplacement unique, une seule fois, à une valeur inférieure. Par la suite, ce chiffre restera à cette valeur inférieure.

Ceteris paribus, si l'effectif de la population change tout à coup à un niveau inférieur et reste à ce niveau inférieur, l'occupation de nids montrera une courte baisse (qui n'est probablement pas détectée) et retournera au niveau précédent pas la suite.

Similairement, même si la population A et la population B ont une densité ou des chiffres différents, les taux d'occupation de nids peuvent être les mêmes.

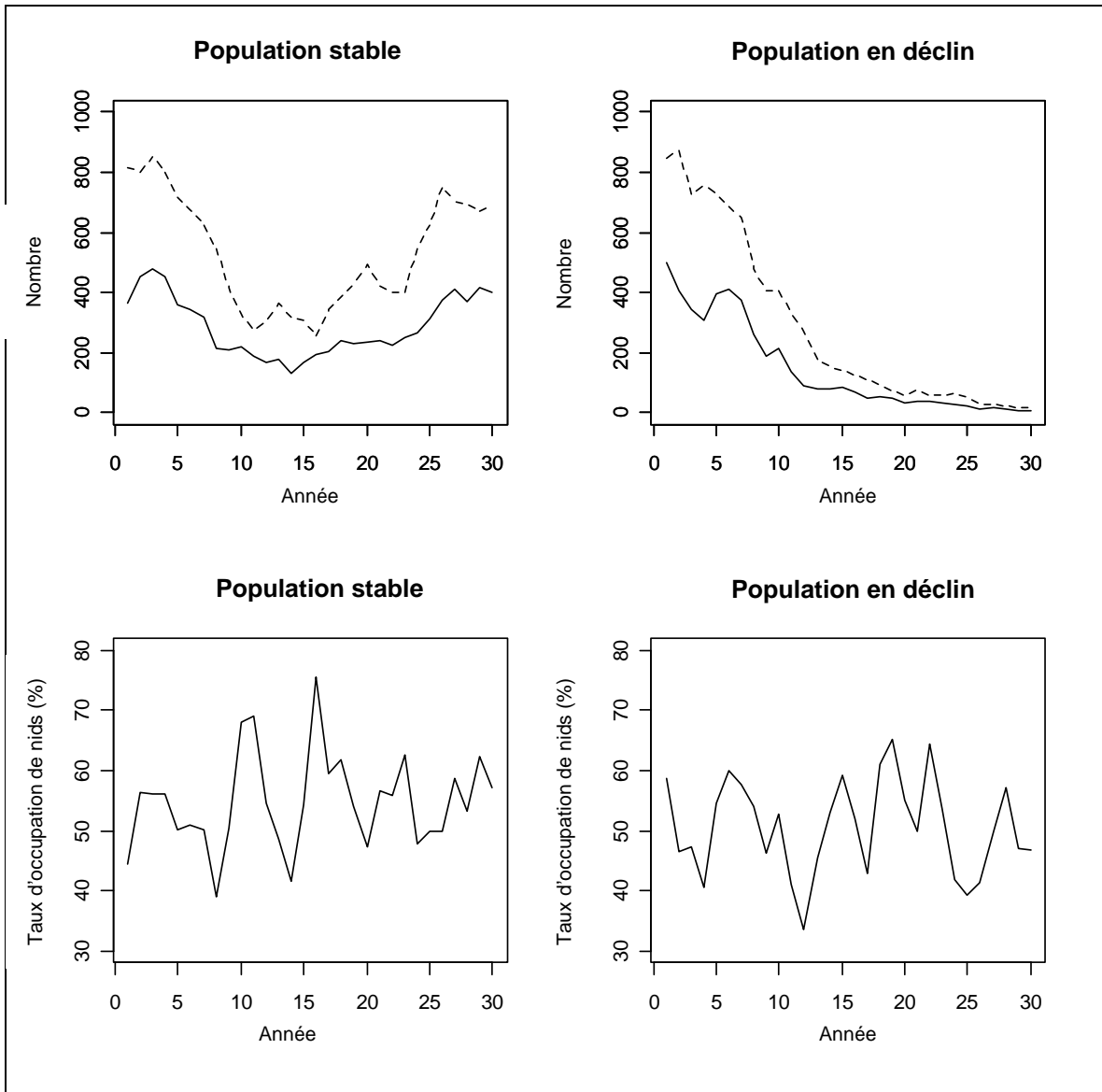


Figure 1. Les résultats d'une simulation de taille de population sur 30 ans (nombre de paires dans des nids – la ligne solide dans les panneaux du haut), le nombre de nids disponibles (ligne pointillée dans les panneaux du haut) et le taux d'occupation de nids (panneaux du bas) pour une population stable et pour une population subissant un déclin de 10 % par an en moyenne. Le taux d'occupation de nids est très similaire pour les deux populations.

4.3 Méthode d'échantillonnage

Il est clair que la méthode d'échantillonnage n'équivaut pas à l'échantillonnage aléatoire. L'échantillonnage a un biais envers les sites utilisés antérieurement et les sites plus proches de Goose Bay (en fait, les sites d'échantillonnage en 2007 sont dans la moitié nordique de la zone expérimentale et la zone de contrôle, juste au sud de Goose Bay).

Nous constatons que les nids dans la zone expérimentale qui se trouvent plus proches de Goose Bay vont probablement être plus exposés aux vols à basse altitude, puisque les avions doivent décoller et atterrir à Goose Bay. [En fait, ceci pourrait être avantageux

pour cette étude, puisque ceci permet d'amplifier les effets des vols à basse altitude : les nids qui se trouvent plus loin de Goose Bay vont probablement être exposés à moins de vols.]

Tandis que ces biais semblent s'appliquer de la même façon pour le groupe expérimental et le groupe de contrôle, il pourrait y avoir des différences subtiles : par exemple, il fut mentionné qu'il y a plusieurs autres sites disponibles en tant que zones de contrôles par rapport à la zone expérimentale; ceci voudrait dire que les nids choisis sont plus proches de Goose Bay dans la zone de contrôle par rapport à la zone expérimentale (il semblerait que ceci ne s'applique pas aux échantillons en 2007) ou que l'échantillon de contrôle peut inclure plus de « vieux » sites par rapport à l'échantillon expérimental, puisque le « quota » peut être atteint plus rapidement dans la zone de contrôle.

Un des objectifs dans la procédure d'échantillonnage est de minimiser l'emploi de carburants. Ceci suggère que les nids avec la meilleure probabilité d'occupation seraient ciblés et retrouvés bien avant les nids non occupés. Par conséquent, le taux d'occupation de nids pourrait être surestimé.

En général, nous considérons que ces caractéristiques de la méthode d'échantillonnage vont probablement avoir un impact important sur les estimations du taux d'occupation de nids et un impact moins important sur les estimations du taux de réussite de nidification et de la productivité du nid.

Nous ne pouvons pas déterminer si la méthode d'échantillonnage a un impact sur les comparaisons entre la zone expérimentale et la zone de contrôle.

Nous concluons que la méthode d'échantillonnage n'est pas similaire à l'échantillonnage aléatoire simple.

4.4 Hypothèse alternative

Le choix d'une hypothèse alternative n'est pas uniforme d'un rapport à l'autre. Nous considérons que l'hypothèse alternative devrait correspondre à un impact négatif causé par l'activité d'entraînement à basse altitude : une diminution du taux d'occupation de nids, une diminution du taux de réussite de nidification et une diminution de la productivité de nidification.

Nous considérons que les tests statistiques devraient être des tests unilatéraux, dans le sens qui correspond à une hypothèse alternative décrivant un impact négatif de l'entraînement à basse altitude sur l'activité reproductrice.

4.5 Puissance et seuil d'importance statistique

4.5.1 Erreur de type I (faux positifs)

Le niveau d'importance empirique (valeur P) et le seuil d'importance jouent un rôle essentiel dans l'interprétation des tests statistiques. Le seuil d'importance de 0,05 est largement utilisé. Cependant, je considère que les recommandations suivantes donnent un point de vue plus nuancé du seuil d'importance :

« Comme règle générale, on peut affirmer qu'a priori, lorsque quelqu'un a une attitude « neutre » par rapport à un type de contradiction spécifique, on commence à soupçonner du niveau $\alpha=0,20$, on est plus ou moins convaincu de sa réalité au niveau 0,05 et on a plus de confiance au niveau 0,01. Habituellement, la conviction antérieure d'un expérimentateur vis-à-vis la possibilité d'un type de contradiction spécifique a un effet sur son attitude. A priori, si l'hypothèse alternative était plausible, l'expérimentateur aurait beaucoup plus de confiance au niveau d'importance 0,05 du résultat par rapport à la contradiction de toutes les expériences antérieures. »

Statistics for Experimenters: An Introduction to Design, Data Analysis, and Model Building, Box, Hunter, Hunter, Wiley-Interscience, 1978.

Selon cette échelle, le choix d'un seuil d'importance de $\alpha=0,20$ pour l'étude du balbuzard correspond au rejet de l'hypothèse nulle lorsqu'une divergence rend quelqu'un « légèrement soupçonneux ». En supposant qu'en réalité, il n'y a pas de divergence, un seuil d'importance de $\alpha=0,20$ veut aussi dire que l'on peut arriver à une fausse conclusion sur une divergence (de type I) 20 % des fois.

Le choix du seuil d'importance a été attribué au désir de réduire la probabilité d'une erreur de type II, ce qui nous mène à conclure qu'il n'y a de divergence quand en réalité, il y en a une (faux négatif).

Le choix de l'erreur de type I devrait dépendre de la conséquence du rejet incorrect de l'hypothèse nulle. Si le rejet de l'hypothèse nulle mène à un très grand coût ou une intervention à risque élevé, on voudra garder la probabilité de l'erreur de type I à un niveau bas.

Étant donné l'information qui nous est disponible, les conséquences d'une erreur de type I (conclure incorrectement que l'activité de l'entraînement à basse altitude a un effet sur la reproduction du balbuzard) sont importantes : elles peuvent correspondre à des dépenses inutiles pour des études à l'avenir et à la fermeture du programme d'entraînement à basse altitude. Nous considérons que $\alpha=0,20$ n'est pas convenable.

Nous recommandons que lorsque l'on interprète les tests statistiques, il faut utiliser les directives ci-dessus de Box, Hunter, Hunter. Spécifiquement, nous recommandons que le seuil d'importance des tests avec des données pour un an soit d'environ $\alpha = 0,05$.

4.5.2 Erreur de type II (faux négatifs) et puissance

Il est important de connaître la puissance des tests statistiques utilisés pour une hypothèse alternative bien choisie. La probabilité de l'erreur de type II, β , pour une alternative donnée est $1-Puissance$ pour cette alternative. Donc, connaître le « diagramme de puissance » d'un test équivaut à connaître la probabilité d'une erreur de type II pour toutes les hypothèses alternatives.

Bien que le diagramme de puissance donne plus d'information que la probabilité d'erreur de type II pour une seule alternative, il est souvent pratique d'utiliser la probabilité d'erreur de type II pour une seule alternative pour décrire les propriétés d'un test.

Une « hypothèse alternative bien choisie » devrait refléter un changement biologique « important ». La probabilité d'erreur de type II correspondante devrait refléter l'importance du changement biologique. Bien qu'il y ait plusieurs approches, nous en suggérons deux :

- L'hypothèse alternative pourrait être un effet de l'entraînement à basse altitude supérieur sur une fraction de la variation temporelle naturelle du paramètre. Par exemple, si le taux de réussite de nidification varie entre 20 % et 80 % entre chaque année dans la zone de contrôle, l'hypothèse alternative serait que l'effet de l'entraînement à basse altitude soit la moitié de la variation naturelle, c'est-à-dire $(80-20)/5 = 30\%$. Étant donné que ce changement est inférieur à la variation naturelle, nous proposons une probabilité d'erreur de type II entre $\beta=0,05$ et $\beta=0,10$ (c'est-à-dire, une puissance entre 0,95 et 0,90).
- L'hypothèse alternative pourrait être l'effet de l'entraînement à basse altitude supérieur sur l'effet le plus petit que l'on estime irréversible. Par exemple, si l'on considère que la population du balbuzard dans la ZEBa pourrait accroître subitement après un déclin de 50 %, mais pas avec un déclin plus prononcé, l'effet de l'entraînement à basse altitude de 50 % serait une hypothèse alternative naturelle. Étant donné que ce type de changement (qui aboutit dans la disparition de la population locale) serait très important au niveau biologique, nous proposons que la probabilité de l'erreur de type II β soit 0,05 ou moins (c'est-à-dire, une puissance de 0,95 ou plus) et environ 0,01 de préférence.

Nous recommandons que la probabilité d'erreur de type II ciblée, β , soit choisie selon l'importance des changements biologiques reflétés par l'hypothèse alternative choisie.

4.6 Tests multiples

Si l'hypothèse nulle est valide, à chaque fois qu'un test statistique avec un seuil d'importance α est utilisé, la probabilité de faire une erreur de type I est α . Si un test statistique est utilisé indépendamment k-fois et l'hypothèse nulle est valide dans tous les cas, la probabilité « côté famille » de faire au moins une erreur de type I est de $1 - (1 - \alpha)^k$. Par exemple, si $\alpha=0,05$ et $k=9$, $1 - (1 - \alpha)^k = 0,37$ et si $\alpha=0,20$ et $k=9$, $1 - (1 - \alpha)^k = 0,87$. Donc, si k augmente, la probabilité de faire au moins une erreur de type I augmente.

Les analyses sur la surveillance du balbuzard en 1999-2007 présentent ce type de situation : pour chaque résultat, un test unique s'applique 9 fois (une fois par an). En utilisant $\alpha=0,20$, on obtient un taux d'erreur de type I côté famille de 0,87, une valeur très élevée pour la séquence de tests sur 9 ans. Ceci veut dire que si l'activité d'entraînement à basse altitude n'avait pas du tout d'impact, on se tromperait probablement dans notre conclusion à la fin de la séquence de 9 ans. Ceci est une situation clairement indésirable.

Les statisticiens reconnaissent que ces « multiples comparaisons » peuvent donner une fausse apparence d'importance. Un des remèdes populaires est la méthode Bonferroni (connue aussi sous le nom de correction Bonferroni) qui consiste d'utiliser un seuil d'importance à test unique de α/k . Cette méthode garantit que les multiples comparaisons aient une probabilité d'erreur de type I générale d'au moins l'équivalent de α , ce qui devient la probabilité d'erreur côté famille. Il existe plusieurs autres méthodes similaires.

Si la probabilité d'erreur de type I côté famille est utilisée sur 9 ans de données provenant du programme sur la surveillance du balbuzard en 1999-2007, ceci exige que le seuil du test unique soit de $0,20/9 = 0,022$. Avec cette condition, seulement 2007

montre des différences importantes (il faut signaler qu'il n'y a pas eu d'activité d'entraînement à basse altitude en 2006 et 2007).

Souvent, on trouve que l'approche Bonferroni et d'autres approches similaires sont trop conservatrices, car elles ont une puissance peu élevée. Des méthodes plus récentes montrent des meilleurs rendements (il y a beaucoup de littérature récente, disons après 1990, à ce sujet).

En utilisant le modèle statistique sur plusieurs années qui réduit le nombre d'hypothèses, vous éviterez les tests multiples. Ces modèles seront expliqués ci-dessous.

En combinant notre constatation sur l'importance de la section secondaire antérieure avec l'augmentation de l'impact de multiples comparaisons, le nombre d'années échantillonnées augmente et nous concluons que le choix du test annuel avec un seuil d'importance de $\alpha=0,20$ n'est pas statistiquement valable.

Si les séquences de tests statistiques pour une seule année sont présentées, nous recommandons qu'une « correction à multiples comparaisons » soit utilisée pour interpréter les résultats des tests.

4.7 Tests statistiques pour les comparaisons sur une seule année

4.7.1 Impact de l'approche d'échantillonnage par quotas sur l'estimation du taux d'occupation de nids

L'approche « quota » (c'est-à-dire, d'échantillonner jusqu'à ce que 30 nids actifs soient retrouvés) n'est pas habituelle, mais c'est une approche similaire à la méthode « d'échantillonnage binomial inverse » dans l'échantillonnage capture-recapture. En supposant que l'on parle de l'échantillonnage aléatoire simple, la quantité totale de nids qui doivent être observés pour obtenir 30 nids occupés suit une distribution binomiale négative avec les paramètres p = proportion de nids occupés et $r = 30$. Nous avons vérifié que l'estimateur a – au plus – un très petit biais et que l'erreur normalisée du théorème de la limite centrale habituelle de l'estimateur est satisfaisante.

4.7.2 Test sur les taux d'occupation de nids et du succès de nidification

Les rapports du programme de surveillance du balbuzard en 1999-2007 ont utilisé la méthode exacte de Fisher à intervalle unilatéral jusqu'en 2005 et la régression logistique à partir de 2006. Donc, les données antérieures ont été testées à nouveau en utilisant la régression logistique.

La méthode exacte de Fisher est un test sur un tableau de corrélation du type suivant :

		X		
		A	B	Total
Y	U	n_{11}	n_{12}	$n_{1.}=n_{11}+n_{12}$
	V	n_{21}	n_{22}	$n_{2.}=n_{21}+n_{22}$
Total		$n_{.1}=n_{11}+n_{21}$	$n_{.2}=n_{12}+n_{22}$	

Il fut prouvé que la méthode exacte de Fisher à intervalles bilatéraux est uniformément la méthode la plus puissante parmi les tests sans biais pour ce type de problème (Lehmann, 1986, cité par Good, 2005).

Si l'hypothèse nulle est vraie ($p_0 = p_1 = 0,60$), la probabilité d'une erreur de Type I n'est pas exactement 0,05 parce que le nombre de valeurs potentielles du niveau d'importance empirique est fini et relativement petit ($\min(n_{1,1}, n_{1,2}, n_{2,1}, n_{2,2}) + 1$). Un processus de répartition aléatoire après les calculs peut donner un test avec une probabilité d'erreur de type I = α , mais c'est utilisé très rarement.

La méthode exacte de Fisher à intervalle unilatéral est le test exact de permutation. Nous avons calculé la puissance de la méthode exacte de Fisher à intervalle unilatéral pour plusieurs valeurs de p_0 , p_1 et la taille de l'échantillon $n_0 = n_1 = n$.

Puisque nous avons choisi de décrire l'impact de l'entraînement à basse altitude proportionnellement au taux de succès de nidification original, l'impact est plus difficile à détecter pour les taux de succès de nidification originaux plus petits (par exemple, un impact de -10 % sur un taux de réussite de 25 % est seulement 2,5 %).

Les résultats apparaissent dans les tableaux suivants. Nous constatons que la probabilité d'erreur de type I semble être plus petite dans tous les cas par rapport au seuil d'importance choisi : encore une fois, ceci est à cause du fait que les comptes se font en nombres entiers et 0,05 ou 0,10 ne se rapprochent pas à des nombres finis avec des niveaux d'importance empiriques possibles pour un ensemble donné de distributions marginales.

Si le taux de réussite non perturbé (c'est-à-dire, la zone de contrôle) est de $p_0 = 25\%$ (un événement relativement rare), l'impact de l'entraînement à basse altitude de -50 % donne un taux de réussite dans la zone perturbée de 12,5 %. Pour toute taille d'échantillon raisonnable, la puissance de détection d'un changement si petit dans le test est basse.

Les taux de réussite non perturbés de $p_0 = 50\%$ ou $p_0 = 75\%$ sont plus intéressants. Dans ces cas, avec les tailles d'échantillons = 30, nous considérons que la puissance du test est adéquat pour un impact d'entraînement à basse altitude de 50 % ou plus, c'est-à-dire, couper le taux de réussite de nidification en deux.

Il faut aussi constater que, tel que prévu, le choix de α a un effet sur la puissance du test. Par exemple, en utilisant une taille d'échantillon de 30, un taux de réussite non perturbé de $p_0 = 75\%$ et un taux d'impact de -30 %, la puissance du test augmente de 0,45 ($\beta = 0,55$) à 0,63 ($\beta = 0,37$) lorsque α change de 0,05 à 0,10.

Nous considérons qu'avec 9 ans de données recueillies, les analyses des taux de réussite de nidification pour une seule année ne sont pas désirables. Cependant, si ces analyses allaient s'effectuer de toute façon, nous recommandons d'utiliser la méthode exacte de Fisher à intervalle unilatéral et que la puissance du test soit indiquée et considérée au moment de prendre les décisions.

Les mêmes observations et recommandations s'appliquent si l'on effectue des analyses du taux d'occupation des nids.

Tableau 3. Puissance de la méthode exacte de Fisher à intervalle unilatéral sur la présence d'un effet négatif causé par l'activité d'entraînement à basse altitude sur la réussite de nidification, selon l'impact en tant que pourcentage du taux de réussite (c'est-à-dire, une réduction relative de 10 % change un taux de réussite de 50 % à 45 %) pour $\alpha = 0,05$ et $\alpha = 0,10$ et les tailles d'échantillons = 30.

Tailles d'échantillons = 30/groupe $\alpha = 0,05$		Taux de réussite non perturbé (p_0)			
		25 %	50 %	75 %	90 %
Pas d'impact	(α)	0,02	0,03	0,02	0,01
-10 %	$p_1 = 0,9 \times p_0$	0,04	0,06	0,09	0,14
-20 %	$p_1 = 0,8 \times p_0$	0,06	0,13	0,25	0,43
-30 %	$p_1 = 0,7 \times p_0$	0,10	0,22	0,45	0,72
-40 %	$p_1 = 0,6 \times p_0$	0,15	0,36	0,70	0,91
-50 %	$p_1 = 0,5 \times p_0$	0,21	0,52	0,85	0,98
Tailles d'échantillons = 30/groupe $\alpha = 0,10$		Taux de réussite non perturbé (p_0)			
		25 %	50 %	75 %	90 %
Pas d'impact	(α)	0,06	0,07	0,06	0,05
-10 %	$p_1 = 0,9 \times p_0$	0,08	0,15	0,17	0,27
-20 %	$p_1 = 0,8 \times p_0$	0,12	0,26	0,39	0,60
-30 %	$p_1 = 0,7 \times p_0$	0,19	0,40	0,63	0,84
-40 %	$p_1 = 0,6 \times p_0$	0,26	0,56	0,84	0,95
-50 %	$p_1 = 0,5 \times p_0$	0,35	0,71	0,94	0,99

Tableau 4. Puissance de la méthode exacte de Fisher à intervalle unilatéral sur la présence d'un effet négatif causé par l'activité d'entraînement à basse altitude sur la réussite de nidification, selon l'impact en tant que pourcentage du taux de réussite (c'est-à-dire, une réduction relative de 10 % change un taux de réussite de 50 % à 45 %) pour $\alpha = 0,05$ et $\alpha = 0,10$ et les tailles d'échantillons = 50.

Tailles d'échantillons = 50/groupe $\alpha = 0,05$		Taux de réussite non perturbé (p_0)			
		25 %	50 %	75 %	90 %
Pas d'impact	(α)	0,03	0,02	0,03	0,02
-10 %	$p_1 = 0,9 \times p_0$	0,06	0,09	0,15	0,26
-20 %	$p_1 = 0,8 \times p_0$	0,09	0,19	0,41	0,66
-30 %	$p_1 = 0,7 \times p_0$	0,17	0,35	0,71	0,92
-40 %	$p_1 = 0,6 \times p_0$	0,25	0,57	0,91	0,99
-50 %	$p_1 = 0,5 \times p_0$	0,39	0,79	0,98	1,00
Tailles d'échantillons = 50/groupe $\alpha = 0,10$		Taux de réussite non perturbé (p_0)			
		25 %	50 %	75 %	90 %
Pas d'impact	(α)	0,07	0,07	0,07	0,05
-10 %	$p_1 = 0,9 \times p_0$	0,11	0,16	0,26	0,40
-20 %	$p_1 = 0,8 \times p_0$	0,18	0,31	0,53	0,79
-30 %	$p_1 = 0,7 \times p_0$	0,27	0,50	0,81	0,96
-40 %	$p_1 = 0,6 \times p_0$	0,39	0,71	0,95	1,00
-50 %	$p_1 = 0,5 \times p_0$	0,55	0,87	0,99	1,00

Tableau 5. Puissance de la méthode exacte de Fisher à intervalle unilatéral sur la présence d'un effet négatif causé par l'activité d'entraînement à basse altitude sur la réussite de nidification, selon l'impact en tant que pourcentage du taux de réussite (c'est-à-dire, une réduction relative de 10 % change un taux de réussite de 50 % à 45 %) pour $\alpha = 0,05$ et $\alpha = 0,10$ et les tailles d'échantillons = 100.

Tailles d'échantillons = 100/groupe $\alpha = 0,05$		Taux de réussite non perturbé (p_0)			
		25 %	50 %	75 %	90 %
Pas d'impact	(α)	0,04	0,04	0,03	0,03
-10 %	$p_1 = 0,9 \times p_0$	0,09	0,13	0,26	0,49
-20 %	$p_1 = 0,8 \times p_0$	0,18	0,35	0,69	0,94
-30 %	$p_1 = 0,7 \times p_0$	0,31	0,65	0,94	1,00
-40 %	$p_1 = 0,6 \times p_0$	0,47	0,85	1,00	1,00
-50 %	$p_1 = 0,5 \times p_0$	0,69	0,98	1,00	1,00
Tailles d'échantillons = 100/groupe $\alpha = 0,10$		Taux de réussite non perturbé (p_0)			
		25 %	50 %	75 %	90 %
Pas d'impact	(α)	0,08	0,08	0,08	0,07
-10 %	$p_1 = 0,9 \times p_0$	0,16	0,22	0,38	0,61
-20 %	$p_1 = 0,8 \times p_0$	0,29	0,49	0,80	0,97
-30 %	$p_1 = 0,7 \times p_0$	0,43	0,78	0,98	1,00
-40 %	$p_1 = 0,6 \times p_0$	0,62	0,93	1,00	1,00
-50 %	$p_1 = 0,5 \times p_0$	0,81	0,99	1,00	1,00

4.7.3 Variable discrète : Productivité de nidification

Il y a trois distributions du nombre de jeunes à l'envol observés dans la zone de contrôle ou la ZEBA dans le tableau ci-dessous, ainsi qu'une distribution mixte et à variance élevée. Les distributions observées sont typiques pour la productivité basse, moyenne et élevée. La distribution mixte est un mélange de 50-50 des distributions basses et élevées conçues pour obtenir un cas de variance élevée. Nos écarts-types sont légèrement différents par rapport à ceux que l'on retrouve dans les rapports annuels à cause de quelques ambiguïtés dans les bases de données brutes.

Tableau 6. Distributions de densité théoriques du nombre de jeunes à l'envol utilisés dans les simulations.				
Nombre	Probabilité			
	Productivité basse (ZEBA 2005)	Productivité moyenne (Contrôle 2006)	Productivité moyenne (Composite)	Productivité élevée (Contrôle 2007)
	Probabilité			
0	0,70	0,25	0,37	0,04
1	0,20	0,31	0,16	0,11
2	0,07	0,34	0,22	0,39
3	0,03	0,09	0,25	0,46
4	0,00	0,00	0,00	0,00
Moyenne	$\mu = 0,43$	$\mu = 1,28$	$\mu = 1,35$	$\mu = 2,27$
Écart-type	$\sigma = 0,82$	$\sigma = 0,94$	$\sigma = 1,21$	$\sigma = 0,76$

Les distributions de productivité basse et élevée sont fortement asymétriques. Pour les comparaisons annuelles, le rapport des variances observées varie entre 1,1 et 2,1.

Nous avons examiné les tests à intervalle unilatéral pour la productivité moyenne en utilisant les distributions basse, élevée et composite. Pour les tailles d'échantillons qui équivalent à 30 ou plus, le test T à intervalle unilatéral et à variances sans équivalent et l'estimation Welch-Satterthwaite des degrés de liberté et un test de distribution aléatoire donnent des résultats similaires. Pour les données dans cette étude, les résultats de la variance équivalente et de la version de variance sans équivalent sont très similaires dans presque tous les cas. Étant donné les distributions observées, les rapports de variance et les tailles d'échantillons, cette conclusion est uniforme aux caractéristiques bien fondées du test T.

Les tableaux suivants montrent la puissance du test pour plusieurs distributions non perturbées et plusieurs niveaux d'impact.

Tout comme la réussite de nidification, tout impact de l'entraînement à basse altitude dans une situation de productivité basse est difficile à établir, puisque cet impact absolu sera petit.

Les cas de productivité moyenne et élevée, $\mu = 1,35$ et $\mu = 2,27$, sont les plus intéressants. Nous considérons que pour les comparaisons sur une seule année, la puissance du test est acceptable pour l'impact de l'entraînement à basse altitude de -30 % ou plus, même avec la taille d'échantillon actuel de 30 dans chaque zone.

Nous considérons qu'avec 9 ans de données recueillies, les analyses sur la productivité de nidification moyenne sur une seule année ne sont pas désirables. Cependant, si ces analyses allaient s'effectuer de toute façon, nous recommandons d'utiliser un test T de variance sans équivalent et à intervalle unilatéral et que la puissance du test soit indiquée et considérée au moment de prendre les décisions.

Tableau 7. Puissance d'une variation Welch du test T à intervalle unilatéral et à variances sans équivalent sur la présence d'un effet négatif de l'activité d'entraînement à basse altitude sur la productivité de nidification, avec l'impact en tant que pourcentage de la productivité (c'est-à-dire, une réduction relative de 10 % change une quantité moyenne de jeunes à l'envol de 2 à une quantité moyenne de 1,80) pour $\alpha = 0,05$ et $\alpha = 0,10$ et des tailles d'échantillons = 30.

Tailles d'échantillons = 30/groupe $\alpha = 0,05$			Moyenne de productivité de nidification non perturbée (H_0)		
			$\mu_0 = 0,43$	$\mu_0 = 1,35$	$\mu_0 = 2,27$
Réduction relative de la moyenne de productivité à cause de l'activité d'entraînement à basse altitude	Pas d'impact	(α)	0,05	0,05	0,05
	-10 %	$\mu_1 = 0,9 \times \mu_0$	0,08	0,11	0,22
	-20 %	$\mu_1 = 0,8 \times \mu_0$	0,12	0,22	0,52
	-30 %	$\mu_1 = 0,7 \times \mu_0$	0,17	0,38	0,79
	-40 %	$\mu_1 = 0,6 \times \mu_0$	0,25	0,52	0,94
	-50 %	$\mu_1 = 0,5 \times \mu_0$	0,36	0,72	0,99
Tailles d'échantillons = 30/groupe $\alpha = 0,10$			Moyenne de productivité de nidification non perturbée (H_0)		
			$\mu_0 = 0,43$	$\mu_0 = 1,35$	$\mu_0 = 2,27$
Réduction relative de la moyenne de productivité à cause de l'activité d'entraînement à basse altitude	Pas d'impact	(α)	0,11	0,10	0,10
	-10 %	$\mu_1 = 0,9 \times \mu_0$	0,14	0,20	0,36
	-20 %	$\mu_1 = 0,8 \times \mu_0$	0,20	0,34	0,67
	-30 %	$\mu_1 = 0,7 \times \mu_0$	0,27	0,52	0,88
	-40 %	$\mu_1 = 0,6 \times \mu_0$	0,39	0,66	0,98
	-50 %	$\mu_1 = 0,5 \times \mu_0$	0,50	0,83	1,00

Tableau 8. Puissance d'une variation Welch du test T à intervalle unilatéral et à variances sans équivalent sur la présence d'un effet négatif de l'activité d'entraînement à basse altitude sur la productivité de nidification, avec l'impact en tant que pourcentage de la productivité (c'est-à-dire, une réduction relative de 10 % change une quantité moyenne de jeunes à l'envol de 2 à une quantité moyenne de 1,80) pour $\alpha = 0,05$ et $\alpha = 0,10$ et des tailles d'échantillons = 50.

Tailles d'échantillons = 50 / groupe $\alpha = 0,05$			Moyenne de productivité de nidification non perturbée (H_0)		
			$\mu_0 = 0,43$	$\mu_0 = 1,35$	$\mu_0 = 2,27$
Réduction relative de la moyenne de productivité à cause de l'activité d'entraînement à basse altitude	Pas d'impact	(α)	0,05	0,05	0,06
	-10 %	$\mu_1 = 0,9 \times \mu_0$	0,08	0,14	0,32
	-20 %	$\mu_1 = 0,8 \times \mu_0$	0,15	0,29	0,72
	-30 %	$\mu_1 = 0,7 \times \mu_0$	0,24	0,52	0,94
	-40 %	$\mu_1 = 0,6 \times \mu_0$	0,35	0,74	0,99
	-50 %	$\mu_1 = 0,5 \times \mu_0$	0,49	0,90	1,00
Tailles d'échantillons = 50 / groupe $\alpha = 0,10$			Moyenne de productivité de nidification non perturbée (H_0)		
			$\mu_0 = 0,43$	$\mu_0 = 1,35$	$\mu_0 = 2,27$
Réduction relative de la moyenne de productivité à cause de l'activité d'entraînement à basse altitude	Pas d'impact	(α)	0,10	0,11	0,11
	-10 %	$\mu_1 = 0,9 \times \mu_0$	0,16	0,23	0,46
	-20 %	$\mu_1 = 0,8 \times \mu_0$	0,25	0,42	0,83
	-30 %	$\mu_1 = 0,7 \times \mu_0$	0,36	0,67	0,97
	-40 %	$\mu_1 = 0,6 \times \mu_0$	0,48	0,83	1,00
	-50 %	$\mu_1 = 0,5 \times \mu_0$	0,64	0,95	1,00

Tableau 9. Puissance d'une variation Welch du test T à intervalle unilatéral et à variances sans équivalent sur la présence d'un effet négatif de l'activité d'entraînement à basse altitude sur la productivité de nidification, avec l'impact en tant que pourcentage de la productivité (c'est-à-dire, une réduction relative de 10 % change une quantité moyenne de jeunes à l'envol de 2 à une quantité moyenne de 1,80) pour $\alpha = 0,05$ et $\alpha = 0,10$ et des tailles d'échantillons = 100.

Tailles d'échantillons = 100 / groupe $\alpha = 0,05$			Moyenne de productivité de nidification non perturbée (H_0)		
			$\mu_0 = 0,43$	$\mu_0 = 1,35$	$\mu_0 = 2,27$
Réduction relative de la moyenne de productivité à cause de l'activité d'entraînement à basse altitude	Pas d'impact	(α)	0,05	0,05	0,05
	-10 %	$\mu_1 = 0,9 \times \mu_0$	0,11	0,20	0,52
	-20 %	$\mu_1 = 0,8 \times \mu_0$	0,22	0,49	0,95
	-30 %	$\mu_1 = 0,7 \times \mu_0$	0,36	0,77	1,00
	-40 %	$\mu_1 = 0,6 \times \mu_0$	0,55	0,94	1,00
	-50 %	$\mu_1 = 0,5 \times \mu_0$	0,74	0,99	1,00
Tailles d'échantillons = 100 / groupe $\alpha = 0,10$			Moyenne de productivité de nidification non perturbée (H_0)		
			$\mu_0 = 0,43$	$\mu_0 = 1,35$	$\mu_0 = 2,27$
Réduction relative de la moyenne de productivité à cause de l'activité d'entraînement à basse altitude	Pas d'impact	(α)	0,10	0,10	0,10
	-10 %	$\mu_1 = 0,9 \times \mu_0$	0,19	0,30	0,66
	-20 %	$\mu_1 = 0,8 \times \mu_0$	0,34	0,63	0,97
	-30 %	$\mu_1 = 0,7 \times \mu_0$	0,50	0,87	1,00
	-40 %	$\mu_1 = 0,6 \times \mu_0$	0,69	0,97	1,00
	-50 %	$\mu_1 = 0,5 \times \mu_0$	0,85	1,00	1,00

5 Modèles sur plusieurs années

5.1 Modèle de succès de nidification selon la présence/absence d'activité d'entraînement à basse altitude

Nous explorons le type de modèle suivant :

- A. Il faut coder l'intensité de vols de la manière suivante (basé sur un nombre approximatif de vols).

Année	Données sur les vols		Activité de vols		Intensité de vols (avions à réaction x1000)	
	Nombre de missions	Nombre d'avions à réaction	ZEBA	Contrôle	ZEBA	Contrôle
1999	1728	2970	1	0	3	0
2000	2758	4780	1	0	5	0
2001	3045	5306	1	0	5	0
2002	2431	4333	1	0	4	0
2003	n/d	n/d	1	0	3	0
2004	n/d	n/d	1	0	2	0
2005	n/d	n/d	1	0	1	0
2006	0	0	0	0	0	0
2007	0	0	0	0	0	0

Les données pour 2003 à 2005 ne sont pas disponibles. Cependant, l'emploi de la ZEBA a subi un déclin progressif pendant ces années. Nous avons estimé les intensités de vols manquants en utilisant une interpolation linéaire entre les valeurs de 2002 et 2006.

- B. Nous considérons que l'Année est une variable nominale pour les raisons suivantes : (1) en ce moment, il n'y a pas de preuves convaincantes qu'il existe une structure temporelle et (2) nous nous intéressons uniquement dans l'effet d'entraînement à basse altitude. (Les données suggèrent qu'un modèle cyclique peut s'appliquer. Cependant, une séquence d'observations plus longue serait essentielle pour se fier à un modèle cyclique.)
- C. Il faut créer un modèle de régression logistique avec le succès de nidification pour avoir une variable-réaction. L'Année et l'Activité de vols seraient deux variables nominales explicatrices :

$$\text{Log}(p_{\text{Réussite}}/p_{\text{Échec}}) \sim \text{Constante} + \text{Année} + \text{Activité de vols}$$

- D. Il faut considérer le coefficient de l'Activité de vols en tant que paramètre d'intérêt.

La puissance de cette analyse sur l'effet d'activité de vols est la suivante.

Tableau 11. Puissance du modèle de régression logistique sur la présence d'un effet négatif de l'activité d'entraînement à basse altitude sur le succès de nidification selon le nombre d'années d'observation et l'impact en tant que pourcentage du taux de réussite (c'est-à-dire, une réduction relative de 10 % change un taux de réussite de 70 % à 63 %) pour $\alpha = 0,05$ et $\alpha = 0,10$.

Taille d'échantillon = 30 par groupe x année $\alpha = 0,05$		Nombre d'années d'observation						
		9	10	12	15	20	25	50
Réduction relative du taux de réussite à cause de l'activité d'entraînement à basse altitude	Pas d'impact (α)	0,051	0,054	0,055	0,059	0,050	0,050	0,050
	-10 %	0,34	0,35	0,39	0,49	0,54	0,65	0,88
	-20 %	0,76	0,82	0,87	0,93	0,98	0,99	1,00
	-30 %	0,97	0,98	0,99	1,00	1,00	1,00	1,00
	-40 %	1,00	1,00	1,00	1,00	1,00	1,00	1,00
	-50 %	1,00	1,00	1,00	1,00	1,00	1,00	1,00
Taille d'échantillon = 30 par groupe x année $\alpha = 0,10$		Nombre d'années d'observation						
		9	10	12	15	20	25	50
Réduction relative du taux de réussite à cause de l'activité d'entraînement à basse altitude	Pas d'impact (α)	0,100	0,107	0,104	0,110	0,097	0,106	0,096
	-10 %	0,48	0,50	0,53	0,62	0,68	0,77	0,93
	-20 %	0,84	0,90	0,93	0,97	0,99	1,00	1,00
	-30 %	0,99	0,99	1,00	1,00	1,00	1,00	1,00
	-40 %	1,00	1,00	1,00	1,00	1,00	1,00	1,00
	-50 %	1,00	1,00	1,00	1,00	1,00	1,00	1,00

Notes techniques

La puissance a été calculée de la manière suivante pour chaque nombre d'années et le niveau d'impact de l'entraînement à basse altitude désiré. Le taux de réussite de nidification non perturbé est distribué uniformément entre 25 % et 85 %. Le taux de réussite de nidification perturbé par l'entraînement à basse altitude a été calculé (c'est-à-dire, 28 % - 2.8 %). La régression logistique a été utilisée pour 2 500 ensembles de données créés par un processus pseudo-aléatoire et la statistique du test afin d'obtenir le coefficient de la variable tactile de l'impact de l'entraînement à basse altitude enregistré. Ensuite, la puissance estimée a été calculée pour les valeurs désirées de α .

5.2 Modèle de succès de nidification selon l'intensité d'exposition ou d'activité d'entraînement à basse altitude

Ce modèle est similaire au modèle antérieur, sauf pour l'emploi du nombre d'avions à réaction (ou le nombre de missions – les deux variables ayant un rapport très étroit) en tant que variable explicative.

$$\text{Log}(p_{\text{Succès}}/p_{\text{Échec}}) \sim \text{Constante} + \text{Année} + \text{Intensité de vols}$$

L'ISRE considère qu'il pourrait y avoir des mesures d'exposition au bruit de l'entraînement à basse altitude. Le modèle peut utiliser telle mesure :

$$\text{Log}(p_{\text{Succès}}/p_{\text{Échec}}) \sim \text{Constante} + \text{Année} + [\text{mesure d'exposition}]$$

5.3 Modèle de productivité de nidification selon la présence/absence d'activité d'entraînement à basse altitude

En considérant à nouveau les distributions de productivité basse, moyenne et élevée pour le nombre de jeunes à l'envol illustrées dans le Tableau 6 (mais non pas pour la distribution composite artificielle), nous observons qu'elles semblent s'apparier mieux à une distribution binomiale avec les paramètres $n=3$ et $p=\mu/3$. Pour les petites valeurs de μ , la distribution binomiale $B(3, \mu/3)$ et la distribution Poisson (μ) sont similaires et s'apparient aussi bien; cependant, pour $\mu >$ environ 1,75, la distribution Poisson ne s'apparie pas essentiellement à cause des chiffres qui correspondent aux environs de 3 jeunes à l'envol.

Les écarts-types du nombre de jeunes à l'envol par rapport à l'année/étendue de la zone vont de 0,77 à 1,20 et la moitié sont entre 0,90 et 1,07. Donc, le rapport de variance maximal est de 2,42. Étant donné la taille d'échantillon de 30 pour chaque année/zone, nous considérons que ceci est un niveau acceptable d'hétéroscédasticité pour une simple analyse de la variance.

Donc, nous proposons de modeler la productivité de nidification en utilisant une simple analyse du modèle de variance.

$$\text{Nombre de jeunes à l'envol} \sim \text{Constante} + \text{Année} + \text{Activité de vols}$$

5.4 Modèle de productivité de nidification selon l'intensité d'exposition ou d'activité d'entraînement à basse altitude

Nous proposons de modeler la productivité de nidification en utilisant une simple analyse du modèle de variance.

$$\text{Nombre de jeunes à l'envol} \sim \text{Constante} + \text{Année} + [\text{mesure d'exposition}]$$

Ce type de modèle correspond à une approximation, car on peut sûrement supposer que l'impact de l'entraînement à basse altitude est une réduction proportionnelle du nombre de jeunes à l'envol (c'est-à-dire, multiplicative) et il n'est pas normal de retrouver des erreurs.

Cependant, le nombre d'observations est très large ($2 \times 30 \times [\text{nombre d'années}]$). Les résultats des simulations indiquent que la proportion de faux positifs se rapproche énormément au α choisi, au moins pour le α entre 0,05 et 0,10 et la puissance est satisfaisante pour un nombre d'années qui équivaut à 9 ou plus pour l'impact de l'entraînement à basse altitude de -20 % ou plus.

Nous constatons aussi que la puissance du test n'augmente pas énormément après 9 ans si nous considérons un impact de -30 % ou plus et qu'elle augmente légèrement pour l'impact plus petit.

Tableau 12. Puissance du modèle d'analyse de la variance sur l'effet de l'activité d'entraînement à basse altitude sur la productivité de nidification, selon le nombre d'années d'observation et l'impact en tant que pourcentage du taux de réussite (c'est-à-dire, une réduction relative de -10 % change une moyenne de 2,0 jeunes à l'envol à 1,8).

Taille de l'échantillon = 30 par groupe × année $\alpha = 0,05$		Nombre d'années d'observation						
		9	10	12	15	20	25	50
Réduction relative du taux de succès à cause de l'activité d'entraînement à basse altitude	Pas d'impact (α)	0,056	0,052	0,050	0,049	0,050	0,045	0,049
	-10 %	0,32	0,32	0,38	0,43	0,57	0,65	0,92
	-20 %	0,79	0,84	0,88	0,95	0,99	1,00	1,00
	-30 %	0,97	0,99	1,00	1,00	1,00	1,00	1,00
	-40 %	1,00	1,00	1,00	1,00	1,00	1,00	1,00
	-50 %	1,00	1,00	1,00	1,00	1,00	1,00	1,00
Taille de l'échantillon = 30 par groupe × année $\alpha = 0,10$		Nombre d'années d'observation						
		9	10	12	15	20	25	50
Réduction relative du taux de succès à cause de l'activité d'entraînement à basse altitude	Pas d'impact (α)	0,102	0,109	0,109	0,099	0,102	0,095	0,096
	-10 %	0,44	0,45	0,51	0,56	0,68	0,75	0,96
	-20 %	0,87	0,90	0,93	0,97	0,99	1,00	1,00
	-30 %	0,99	0,99	1,00	1,00	1,00	1,00	1,00
	-40 %	1,00	1,00	1,00	1,00	1,00	1,00	1,00
	-50 %	1,00	1,00	1,00	1,00	1,00	1,00	1,00

6 Analyse des résultats en 1999-2007

6.1 Réussite de nidification

Le tableau suivant montre les résultats de l'application de la méthode exacte de Fisher à intervalle unilatéral au taux de réussite de nidification. Avec l'application d'une correction aux multiples comparaisons, on peut déduire qu'il n'y a pas de différence importante au seuil de 0,05.

Sans l'application d'une correction aux multiples comparaisons, la seule différence importante au seuil de 0,05 est pour 2007, une année où il n'y a pas eu d'activité d'entraînement à basse altitude dans la ZEBa.

Bien que le taux de succès de nidification soit bas, la puissance du test est naturellement basse car tout impact de l'entraînement à basse altitude serait petit et difficile à détecter. Nous considérons que dans tous les autres cas, la puissance du test pour l'impact de l'entraînement à basse altitude de -30 % est acceptable (~0,60-0,75).

Tableau 13. Test de permutation selon l'année pour le taux de réussite de nidification. L'hypothèse alternative signale que le taux de réussite de nidification est plus bas dans la ZEBa.

Année	Zone de contrôle		ZEBa		État de la ZEBa	Valeur P	Puissance pour $\alpha=0,05$ et l'impact de l'EBA est -30 %
	Taille de l'échantillon	Taux de réussite de nidification	Taille de l'échantillon	Taux de réussite de nidification			
1999	30	0,83	30	0,90	Activité	0,873	0,58
2000	30	0,33	30	0,53	Activité	0,966	0,13
2001	30	0,43	30	0,30	Activité	0,211	0,18
2002	30	0,43	30	0,50	Activité	0,781	0,18
2003	30	0,83	30	0,93	Activité	0,949	0,65
2004	30	0,87	30	0,70	Activité	0,105	0,65
2005	30	0,38	30	0,30	Activité	0,356	0,15
2006	29	0,77	30	0,53	Pas d'activité	0,052	0,48
2007	30	0,96	30	0,72	Pas d'activité	0,019	0,85

Les résultats de l'application d'une seule régression logistique sur le succès de nidification vis-à-vis l'activité de vols et l'intensité de vols apparaissent respectivement dans le Tableau 14 et le Tableau 15.

Dans une régression logistique, $\log(p_{\text{Réussite}}/p_{\text{Échec}})$ est modélisé comme une combinaison linéaire de paramètres. Pour les variables explicatives nominales (année, activité de vols), il y a un codage. Dans le tableau suivant, le coefficient de l'*activité de vols* = VRAI est le plus important. On estime que la valeur est de 0,04, avec un intervalle de confiance de 95 % [-0,39,0,47]. Donc, nous estimons que $\log(p_{\text{Réussite}}/p_{\text{Échec}})$ est 0,04 supérieur s'il y a de l'activité d'entraînement à basse altitude (*Activité de vols* = VRAI) plutôt que lorsqu'il n'y en a pas (*Activité de vols* = Faux). Ceci veut dire que la probabilité

du succès de nidification est aussi *supérieure*. Bien sûr, nous ne pouvons pas rejeter l'hypothèse nulle que l'entraînement à basse altitude ne cause pas d'effets contre l'alternative que l'entraînement à basse altitude cause un effet négatif.

Les autres coefficients décrivent l'effet annuel par rapport à l'année de base, 1999.

Similairement pour l'intensité de vols, où cette fois-ci, le coefficient de l'**Intensité de vols**, 0,04, indique que $\log(p_{\text{Réussite}}/p_{\text{Échec}})$ *augmente* d'environ 0,04 pour chaque 1000 avions à réactions dans l'air. L'intervalle de confiance de 95 % pour le coefficient du paramètre est de [-0,08,0,16].

Tableau 14. Régression logistique du succès de nidification sur l'activité de vols. Le coefficient d'activité de vols n'a pas de différence importante entre 0 et positif. En examinant que l'entraînement à basse altitude a un impact négatif, nous ne rejetons pas l'hypothèse nulle ($p = 0,574$).

	Coefficient	Intervalle de confiance	s.e.	Wald z	P (bilatéral)	P (unilatéral)
Interception	1,85	[1,07,2,63]	0,40	4,68	0,000	
Année=2000	-2,14	[-3,04,-1,24]	0,46	-4,65	0,000	
Année=2001	-2,42	[-3,32,-1,52]	0,46	-5,20	0,000	
Année=2002	-2,01	[-2,91,-1,11]	0,46	-4,36	0,000	
Année=2003	0,15	[-0,93,1,23]	0,55	0,28	0,783	
Année=2004	-0,59	[-1,55,0,37]	0,49	-1,19	0,234	
Année=2005	-2,54	[-3,46,-1,62]	0,47	-5,42	0,000	
Année=2006	-1,23	[-2,17,-0,29]	0,48	-2,57	0,010	
Année=2007	-0,22	[-1,28,0,84]	0,54	-0,41	0,683	
Activité de vols=VRAI	0,04	[-0,39,0,47]	0,22	0,19	0,853	0,574

Tableau 15. Régression logistique du succès de nidification sur l'intensité de vols. Le coefficient d'intensité de vols n'a pas de différence importante entre 0 et positif. En examinant que l'entraînement à basse altitude a un impact négatif, nous ne rejetons pas l'hypothèse nulle ($p = 0,731$).

	Coefficient	Intervalle de confiance	s.e.	Wald z	P (bilatéral)	P (unilatéral)
Interception	1,82	[1,06,2,58]	0,39	4,67	0,000	
Année=2000	-2,18	[-3,10,-1,26]	0,47	-4,68	0,000	
Année=2001	-2,46	[-3,38,-1,54]	0,47	-5,23	0,000	
Année=2002	-2,02	[-2,92,-1,12]	0,46	-4,39	0,000	
Année=2003	0,15	[-0,93,1,23]	0,55	0,28	0,783	
Année=2004	-0,57	[-1,53,0,39]	0,49	-1,15	0,249	
Année=2005	-2,50	[-3,42,-1,58]	0,47	-5,30	0,000	
Année=2006	-1,20	[-2,12,-0,28]	0,47	-2,53	0,012	
Année=2007	-0,19	[-1,23,0,85]	0,53	-0,35	0,727	
Intensité de vols	0,04	[-0,08,0,16]	0,06	0,62	0,539	0,731

6.2 Productivité de nidification

Le tableau suivant montre une analyse de la moyenne de jeunes à l'envol observés.

Avec l'application d'une correction aux multiples comparaisons, on peut déduire qu'il n'y a pas de différence importante au seuil de 0,05.

Sans l'application d'une correction aux multiples comparaisons, on observe des différences importantes au seuil 0,05 en 2004 et en 2007. Tel qu'indiqué, 2006 et 2007 sont des années où il n'y a pas eu d'activité d'entraînement à basse altitude dans la ZEBA. En fait, la moyenne de jeunes à l'envol dans la ZEBA semble diminuer (par rapport à la zone de contrôle) car le nombre de vols est réduit.

Tableau 16. Analyse statistique selon l'année de la productivité de nidification.

Année	Zone de contrôle			ZEBA			État de la ZEBA	Différence ZEBA moins Contrôle	SE	Différence relative	Valeur P	Puissance pour $\alpha=0,05$ et l'impact de l'EBA est -30 %
	n	Moyenne de jeunes à l'envol	DS	N	Moyenne de jeunes à l'envol	DS						
1999	30	1,57	0,94	30	1,77	0,90	Activité	0,20	0,24	13 %	0,799	0,45
2000	30	0,57	0,90	30	1,03	1,13	Activité	0,47	0,26	82 %	0,959	0,20
2001	30	0,70	0,88	30	0,60	1,04	Activité	-0,10	0,25	-14 %	0,344	0,24
2002	30	0,93	1,17	30	1,00	1,14	Activité	0,07	0,30	7 %	0,588	0,28
2003	30	1,90	0,77	30	1,73	1,03	Activité	-0,17	0,24	-9 %	0,241	0,57
2004	30	1,97	1,07	30	1,43	0,96	Activité	-0,53	0,26	-27 %	0,024	0,60
2005	29	0,72	1,01	30	0,43	0,77	Activité	-0,29	0,23	-40 %	0,109	0,22
2006	30	1,33	1,05	30	1,00	0,96	Pas d'activité	-0,33	0,26	-25 %	0,102	0,38
2007	26	2,27	0,83	29	1,66	1,20	Pas d'activité	-0,61	0,28	-27 %	0,017	0,80

Nous ne pouvons pas entreprendre l'analyse proposée de la variance sur la productivité de nidification parce que les données brutes ne sont pas disponibles pour les années 1999-2002.

6.3 Taux d'occupation de nids

Malgré le fait que nous considérons que le taux d'occupation de nids est difficile à interpréter et ces estimations sont peut-être partielles à cause de la méthode d'échantillonnage, nous entreprenons l'analyse statistique pour que tout soit complet.

Dans l'analyse selon l'année, aucune des différences ne sont importantes du point de vue statistique. Dans la régression logistique, l'hypothèse nulle n'est pas rejetée et le coefficient de la variable appropriée est compatible avec un effet **positif** de l'entraînement à basse altitude sur les taux d'occupation de nids.

Tableau 17. Test de permutation selon l'année pour le taux d'occupation de nids. L'hypothèse alternative indique que le taux d'occupation est inférieur dans la ZEBA.

Année	Zone de contrôle		ZEBA		État de la ZEBA	Valeur P
	Taille de l'échantillon	Taux d'occupation de nids	Taille de l'échantillon	Taux d'occupation de nids		
1999	98	0,60	99	0,58	Activité	0,409
2000	82	0,39	77	0,45	Activité	0,837
2001	74	0,47	75	0,41	Activité	0,285
2002	112	0,31	81	0,42	Activité	0,954
2003	53	0,60	40	0,75	Activité	0,957
2004	44	0,73	44	0,75	Activité	0,686
2005	40	0,73	49	0,63	Activité	0,243
2006	52	0,62	61	0,51	Pas d'activité	0,170
2007	53	0,53	58	0,50	Pas d'activité	0,457

La régression logistique est le tableau suivant. Le coefficient d'**Activité de vols = VRAI** et d'**Intensité de vols** sont positifs et l'hypothèse nulle n'est pas rejetée dans les analyses. Les détails se trouvent dans les tableaux suivants.

Tableau 18. Régression logistique de l'occupation de nids sur l'activité de vols. Le coefficient d'activité de vols n'a pas de différence importante entre 0 et un chiffre positif. Si l'on teste que l'entraînement à basse altitude a un impact négatif, nous ne rejetons pas l'hypothèse nulle ($p = 0,716$).

	Coefficient	Intervalle de confiance	s.e.	Wald z	P (bilatéral)	P (unilatéral)
Interception	0,31	[0,00,0,62]	0,16	1,92	0,055	
Année=2000	-0,67	[-1,10,-0,24]	0,22	-3,12	0,002	
Année=2001	-0,59	[-1,02,-0,16]	0,22	-2,68	0,007	
Année=2002	-0,94	[-1,35,-0,53]	0,21	-4,49	0,000	
Année=2003	0,34	[-0,17,0,85]	0,26	1,30	0,195	
Année=2004	0,68	[0,13,1,23]	0,28	2,41	0,016	
Année=2005	0,36	[-0,17,0,89]	0,27	1,35	0,177	
Année=2006	-0,08	[-0,57,0,41]	0,25	-0,30	0,761	
Année=2007	-0,25	[-0,74,0,24]	0,25	-1,02	0,309	
Activité de vols=VRAI	0,11	[-0,14,0,36]	0,13	0,79	0,431	0,716

Tableau 19. Régression logistique de l'occupation de nids sur l'intensité de vols. Le coefficient d'intensité de vols n'a pas de différence importante entre 0 et un chiffre positif. Si l'on teste que l'entraînement à basse altitude a un impact négatif, nous ne rejetons pas l'hypothèse nulle ($p = 0,826$).

	Coefficient	Intervalle de confiance	s.e.	Wald z	P (bilatéral)	P (unilatéral)
Interception	0,31	[0,02,0,60]	0,15	2,01	0,044	
Année=2000	-0,71	[-1,14,-0,28]	0,22	-3,23	0,001	
Année=2001	-0,62	[-1,05,-0,19]	0,22	-2,79	0,005	
Année=2002	-0,95	[-1,36,-0,54]	0,21	-4,56	0,000	
Année=2003	0,34	[-0,17,0,85]	0,26	1,30	0,195	
Année=2004	0,70	[0,15,1,25]	0,28	2,46	0,014	
Année=2005	0,40	[-0,13,0,93]	0,27	1,48	0,140	
Année=2006	-0,08	[-0,55,0,39]	0,24	-0,32	0,747	
Année=2007	-0,26	[-0,73,0,21]	0,24	-1,05	0,295	
Intensité de vols	0,03	[-0,03,0,09]	0,03	0,94	0,347	0,826

7 Conclusion

7.1 Interprétation des résultats actuels

Les données de l'étude en 1999-2007 nous mènent à conclure que l'activité d'entraînement à basse altitude n'a probablement pas d'effet négatif sur la réussite de nidification ou la productivité de nidification chez le balbuzard. Les variations entre la ZEBa et la zone de contrôle ne sont pas différentes du point de vue statistique; de plus, notre estimation ponctuelle de l'impact est essentiellement nulle en relation à la variation naturelle.

La même conclusion statistique s'applique au taux d'occupation de nids. Cependant, nous considérons que l'interprétation de toute différence entre les taux d'occupation de nids est difficile à faire, surtout à long terme, et le processus d'échantillonnage actuel va très probablement présenter des biais dans l'estimation.

Nous considérons que la puissance du test statistique entrepris en utilisant toute la base de données est assez pour appuyer ces conclusions.

Nous considérons que le biais d'échantillonnage le plus évident (c'est-à-dire, les nids proche de la base militaire) augmenterait sûrement l'impact de l'entraînement à basse altitude, s'il y a lieu.

D'autres sources de preuves semblent appuyer cette conclusion. Aucune de ces études entreprises avant 1999 ne semblent avoir indiqué un impact causé par l'entraînement à basse altitude. Les preuves provenant d'autres sources (construction routière, entretien des lignes sous tension – consultez le site Web de l'ISRE) indiquent que les balbuzards peuvent tolérer un niveau assez élevé de perturbation pendant leur activité de nidification. La productivité de nidification semble être dans la portée signalée dans d'autres sites en Amérique du Nord (0,56 à 1,3 selon D.M.Bird, site Web de l'ISRE).

Étant donné les données de surveillance et d'autres observations, nous concluons que tout impact de l'entraînement à basse altitude sur la reproduction du balbuzard va sûrement être largement inférieur à la variation naturelle du succès de reproduction.

Étant donné que le programme de surveillance en 1999-2007 est une étude par observation, il faut considérer la possibilité de biais qui annulent précisément l'impact de l'activité d'entraînement à basse altitude. Il est probable que ces biais, s'ils existent, peuvent être découverts en accumulant plus de données en utilisant le même concept par observation, même avec une augmentation des tailles d'échantillons.

Basé sur les preuves examinées, nous nous attendons à ce qu'une suite du programme de surveillance actuel NE produise PAS plus d'information sur l'impact de l'entraînement à basse altitude, s'il y en a pour commencer.

7.2 Circonstances spéciales

Malgré le fait que nous considérons que les résultats du programme de surveillance devraient être analysés dans un modèle unique et que les données obtenues entre 1999 et 2007 ne montrent pas d'effets causés par l'entraînement à basse altitude, nous constatons que des circonstances spéciales (comme l'introduction d'une activité radicalement nouvelle) peuvent mériter un programme d'un an. Dans ce cas, nous recommandons l'emploi de la méthode exacte de Fisher et le test T pour les données avec des variances sans équivalent.

7.3 Surveillance périodique

On pourrait considérer la surveillance périodique (comme par exemple, la surveillance à tous les 5 ans). Nous considérons que ce programme de surveillance donnerait des résultats peu fiables à cause de la fluctuation naturelle de la population.

7.4 Organisation de la base de données

Si le programme de surveillance doit être poursuivi, nous recommandons que l'information soit conservée dans une seule base de données ou feuille de calcul pour garantir l'uniformité et pour que les analyses sur plusieurs années soient entreprises efficacement. Spécifiquement, les données en 1999-2007 devraient être transférées à une seule base de données.

7.5 Tests statistiques

Si le programme de surveillance doit être poursuivi, nous recommandons que les méthodes statistiques suivantes soient utilisées.

7.5.1 Analyses sur une seule année

En supposant que la taille de l'échantillon est supérieure ou égale à 30 pour chaque zone, nous recommandons qu'une comparaison sur une seule année des taux de réussite de nidification soit entreprise en utilisant un test de permutation exact à intervalle unilatéral (la version à intervalle unilatéral de la méthode exacte de Fisher).

Le même test devrait être utilisé pour comparer les taux d'occupation de nids, si cette comparaison prendra suite.

En supposant que la taille de l'échantillon reste supérieure ou égale à 30 pour chaque zone, nous recommandons d'entreprendre une comparaison sur une seule année de la productivité de nidification en utilisant le test T pour la population avec des variances inégales.

Dans tous les cas, il faudrait fournir la puissance du test.

Si l'on considère une séquence sur plusieurs années de comparaisons sur une seule année, il faudrait appliquer une correction pour plusieurs comparaisons.

7.5.2 Analyses sur plusieurs années et l'impact de l'intensité

Nous recommandons que l'analyse sur plusieurs années du taux de réussite de nidification soit administrée avec la régression logistique, l'année et la présence/l'absence de vols à basse altitude qui sont utilisés en tant que variables explicatives nominales.

Nous recommandons que l'analyse sur plusieurs années de la productivité de nidification soit administrée par l'analyse de la variance, l'année et la présence/l'absence de l'intensité de vols à basse altitude qui sont utilisés en tant que variables explicatives nominales.

Nous recommandons que l'analyse sur plusieurs années du taux de réussite de nidification soit administrée avec l'utilisation de la régression logistique et de l'année en tant que variables explicatives nominales et une mesure de l'intensité de vols à basse altitude en tant que covariation, si cette mesure est disponible.

Nous recommandons que l'analyse sur plusieurs années de la productivité de nidification soit administrée par l'utilisation de l'analyse de la variance et de l'année en tant que variables explicatives nominales et une mesure de l'intensité de vols à basse altitude en tant que covariation.

7.5.3 Probabilité d'erreurs de type I et de type II

Étant donné les conséquences d'une erreur de type I, nous recommandons d'utiliser le niveau d'importance $\alpha = 0,05$.

Le choix de probabilité β de l'erreur de type II doit dépendre de l'hypothèse alternative sélectionnée. Si l'hypothèse alternative correspond à un impact de l'entraînement à basse altitude relativement bas (par exemple, une réduction de 30 % du taux de réussite), nous recommandons que β soit entre 0,05 et 0,10. Si l'hypothèse alternative correspond à un impact de l'entraînement à basse altitude très important (par exemple, un impact qui mène à la disparition de la population locale), nous recommandons que β soit tout au plus 0,05 et préférablement 0,01.

7.5.4 Intervalles de confiance

Nous recommandons que – dans les futures analyses – les intervalles de confiance pour les différences dans les paramètres entre les groupes soient obtenus et qu'ils accompagnent les résultats des tests statistiques.

8 Remerciements

L'auteur voudrait remercier Dr David Schneider pour sa révision perspicace des deux versions préliminaires du document.

Références

Rosenbaum, Paul R. *Observational Studies*, Springer, 2^e édition, 2002, 375 p.

Good, Phillip I. *Permutation, Parametric, and Bootstrap Tests of Hypotheses*, Springer, 3^e édition, 2004, 376 p.

Stewart-Oaten, A., Bence, J.R., Osenberg, C.W. « Assessing effects of unreplicated perturbations: No simple solutions », *Ecology* 73, 1992, p. 1396-1404.