



Accord sur la conservation des albatros et des pétrels

Deuxième réunion du Comité consultatif

à Brasilia, au Brésil, du 5 au 8 juin 2006

**Examen des méthodes de surveillance des tendances
telles qu'appliquées aux populations d'oiseaux de mer**

Auteur : Secrétariat

Examen des méthodes de surveillance des tendances telles qu'appliquées aux populations d'oiseaux de mer

(Cette communication a été fournie sous contrat à l'ACAP)

Chris Wilcox

Pelagic Fisheries and Ecosystems [Pêches et écosystèmes pélagiques]

CSIRO Marine and Atmospheric Research [Recherche marine et atmosphérique]

Introduction et champ d'application

L'estimation de l'abondance et de la distribution est un thème central en écologie, et l'application de méthodes statistiques à cette question a tissé un fil de continuité à travers l'histoire de cette science. Il existe un ensemble important de travaux sur ce sujet, qui s'est élargi rapidement avec le développement de la littérature dans le domaine de la biologie de conservation. Ce document a pour objet de présenter un examen sommaire des philosophies sur lesquelles reposent les méthodes disponibles, de traiter des problèmes que les analyses des données disponibles devront aborder, et enfin de suggérer une méthode qui pourrait être utile pour l'analyse.

Philosophie de l'analyse

Il existe au moins deux approches possibles de l'analyse des données de recensement d'oiseaux et, bien que ces approches produisent toutes les deux des résultats utiles, il est important de choisir une méthode en connaissance de cause car elles présentent toutes les deux des avantages et des inconvénients. Un choix s'impose entre la modélisation des recensements en tant que phénomènes en soi et la modélisation de la dynamique démographique sous-jacente, à partir de laquelle il est possible de faire les recensements. Les analyses antérieures des tendances des oiseaux de mer ont adopté l'approche de la modélisation des recensements en tant que tels (par exemple, Woehler, Cooper *et al.* 2001) et ont réussi à produire de bons ajustements aux données et des résultats utiles. En général, cette approche part de la présomption que les données de recensement sont soit négatives binomialement soit distribuées selon la loi de Poisson, puis utilise une expression liée soit aux paramètres de forme ou d'échelle (binomiale négative) soit à la moyenne de distribution (Poisson) qui inclut le temps et d'autres covariables présentant éventuellement un intérêt. Il existe diverses méthodes pour analyser les données dans ce contexte, qui dépendent essentiellement de la forme fonctionnelle choisie pour les variables explicatives. Si l'on privilégie les prédicteurs linéaires, on pourra utiliser un modèle linéaire généralisé ou GLM [de l'anglais « generalised linear model »] (Venables & Ripley 1999), et pour les situations où il n'est pas possible de postuler la linéarité, on pourra utiliser un modèle additif généralisé ou GAM [« generalised additive model »] (Venables & Ripley 1999). Après l'exploration à l'aide d'un GAM, un polynôme adéquatement ordonné pourrait être substitué dans un GLM. D'autres extensions sont possibles à l'aide de modèles mixtes linéaires généralisés ou GLMM [« generalised linear mixed model »] dans lesquels certains des facteurs sont des effets fixes et d'autres des

termes aléatoires (Pinheiro & Bates 2000). Ces extensions sont particulièrement utiles dans les cas où il se produit une erreur d'observation et peuvent être utilisées pour modéliser le processus causant l'erreur, permettant ainsi de supprimer une source d'erreur de la tendance temporelle faisant l'objet de l'étude.

Une autre approche de la modélisation phénoménologique des recensements est de prendre comme point de départ explicite le processus biologique et de modéliser les recensements en tant que résultats de la croissance ou du déclin de la population à l'étude. Dans ce cas, on spécifie le modèle sur la base du cycle biologique et des processus dynamiques de la population, on estime ses paramètres sur la base des données du recensement, puis on déduit la tendance en ce qui concerne la taille de la population, c'est-à-dire le taux de croissance démographique. Comme pour l'approche par la modélisation fondée sur le recensement, ces modèles vont du très simple au très complexe. Dennis (1991) propose une méthode pour estimer les paramètres dans un modèle de population très simple à partir des données de recensement. On trouvera une présentation très lucide de ce modèle et de ses extensions, ainsi que des instructions pour estimer les paramètres, dans Morris & Doak (2002). Ce modèle est fondé sur une approximation de la diffusion et, bien que cette approche repose sur des suppositions manifestement simplificatrices, elle est bien acceptée dans la littérature biologique et a été largement utilisée. À l'autre extrémité de cette échelle, on trouve des modèles d'évaluation intégrée dans lesquels le modèle biologique sous-jacent peut comprendre les taux et les variations en matière de survie, de croissance et de fécondité pour chaque classe d'âge ou de taille, qui sont estimés sur la base de l'ensemble des données disponibles, notamment les chiffres de population, les données démographiques comme les enquêtes sur la nidification, ainsi que les valeurs disponibles dans la littérature. Tuck *et al.* (2001) appliquent une méthode analogue pour évaluer les incidences de la mortalité due à la capture accessoire des pêches sur les albatros, et ces méthodes ont été largement utilisées ailleurs dans la gestion des pêches.

Il y a un chevauchement considérable de l'approche utilisant la modélisation des recensements et celle utilisant la modélisation des populations. En fait, les versions les plus simples de chacun de ces deux modèles consistent essentiellement à ajuster une expression linéaire sur l'ensemble des données de recensement. La différence essentielle est que, dans un cas, le modèle s'intéresse aux résultats du processus (c'est-à-dire les recensements proprement dits), tandis que dans l'autre, il s'intéresse au processus du passage d'un recensement à l'autre (c'est-à-dire le processus biologique qui sous-tend les recensements). Cette différence dans la méthode d'analyse a une incidence sur son aptitude à correspondre au processus et aux applications potentielles des paramètres évalués dans le modèle. Les deux exemples suivants clarifieront ces distinctions. Une variable aléatoire de Poisson peut prendre des valeurs entières entre 0 et l'infini. Si l'on utilise un modèle fondé sur la loi de Poisson pour représenter les recensements d'une population d'oiseaux de mer qui dispose d'un nombre limité de sites de nidification, le recensement réel ne peut pas dépasser le nombre de sites de nidification, bien que le modèle statistique attache une probabilité non nulle aux valeurs supérieures à ce plafond. Ce manque de souplesse de l'approche phénoménologique résulte essentiellement de l'adoption d'un processus aléatoire (le processus de Poisson) qui est semblable du point

de vue phénoménologique mais différent du point de vue mécaniste du processus biologique réel (naissances et morts d'individus). La deuxième raison, probablement plus importante, d'utiliser un modèle qui repose sur un processus biologique, même simple, est que ses paramètres peuvent être interprétés de manière significative. Ainsi, dans le cas d'un modèle ajusté sur une population en déclin, comme il est possible de lier les paramètres aux quantités qui peuvent être modifiées sur le terrain, ces paramètres peuvent être riches en renseignements sur les mesures à prendre pour freiner le déclin. Par exemple, dans l'exemple des sites de nidification limités, il sera peut-être possible de distinguer entre la perte d'habitats appropriés (c'est-à-dire un nombre moindre de sites de nidification) et un changement du taux de croissance de la population comme causes de changements démographiques au fil du temps. Selon la méthodologie utilisée pour évaluer les paramètres du mode biologique, il y a également plusieurs autres aspects utiles qu'il convient de signaler. Dans les modèles biologiques, il est possible de comparer les estimations, pour des paramètres similaires, (par exemple, la fécondité per capita) entre des modèles de complexité différente, et de transférer des renseignements provenant d'espèces riches en données vers des espèces pauvres en données. Pour cela, il est nécessaire d'évaluer les modèles dans un contexte bayésien, examen qui dépasse le cadre de cette communication. Ces méthodes sont toutefois déjà bien développées dans les pêches et il serait relativement facile de les appliquer aux types de modèle suggérés ici.

Un autre aspect à prendre en considération pour le choix d'une méthode d'analyse est la définition exacte de statut et sa quantification. Il existe au moins deux façons d'aborder ce problème : l'évaluation du statut peut se faire sur la base de la tendance historique des recensements ou elle peut se faire sur la base de la projection des recensements dans l'avenir. Pour une analyse historique, il s'agit généralement de déterminer s'il existe une pente négative dans la tendance temporelle des recensements dans le cas d'une régression de Poisson, ou si le taux estimé de croissance démographique est négatif, dans le cas d'un modèle fondé sur un processus. Un autre modèle, préférable au précédent, est de comparer l'ajustement des données à un modèle indiquant une tendance à l'ajustement d'un modèle qui suppose une abondance constante ou des recensements utilisant le critère AIC [Akaike's Information Criterion], car cette comparaison permet d'éviter les problèmes de non-détection et de puissance dont on a beaucoup parlé dans le cadre de la surveillance fondée sur des tests statistiques d'hypothèse nulle. On a récemment suggéré que l'approche par la projection était préférable à l'utilisation de la tendance historique pour évaluer la situation de conservation (Staples, Taper *et al.* 2005). Il s'agit d'utiliser le modèle paramétré sur la base des données de recensement pour prédire l'abondance ou le recensement futurs. On fixe ensuite un seuil pour le niveau d'abondance acceptable, et la tendance surveillée est la probabilité d'un dépassement de seuil avec le temps. Cette approche a été préconisée pour plusieurs raisons : 1) elle projette dans l'avenir, et fait généralement ressortir les problèmes plus tôt que les méthodes qui étudient les tendances historiques ; 2) la situation de conservation a une interprétation claire dans le contexte de la politique en matière de protection de la biodiversité. 3) le risque relatif (c'est-à-dire l'augmentation du risque d'une année à l'autre) s'est avéré à l'épreuve des erreurs de spécification du modèle (par exemple, non-inclusion de la dépendance à la densité) ; 4) cette méthode nécessite relativement peu de données et ne dépend par conséquent pas de

séries à long terme (Staples, Taper *et al.* 2005). Staples *et al.* (Staples, Taper *et al.* 2005) présentent un exposé convaincant de cette approche et appliquent une approximation de diffusion, décrite dans la section suivante, qui démontre qu'elle obtient de bons résultats à cet égard dans diverses situations.

Méthodologie de l'analyse

Il existe divers programmes statistiques qui peuvent être utilisés pour réaliser l'analyse des données de recensement dans une approche phénoménologique, et les méthodes sont bien documentées et généralement acceptées dans la littérature scientifique. Bien qu'il y ait assurément des problèmes avec les suppositions sur lesquelles repose l'analyse, par exemple, la fidélité de la conformité de la distribution des observations à la loi de Poisson, il est généralement possible d'obvier à ces problèmes soit en modifiant le modèle statistique soit en transformant les données. Le programme TRIM (<http://firmy.publikuj.cz/EBCC/index.php?ID=13>), qui a été conçu en vue de l'analyse des données de recensement, comprend plusieurs fonctionnalités très utiles et serait un programme satisfaisant pour réaliser une analyse phénoménologique.

Il existe également des méthodes bien développées pour évaluer les paramètres des modèles qui reposent sur un processus biologique, bien qu'il n'existe pas de programme standard communément utilisé pour le paramétrage analogue au programme TRIM. Les méthodes de calcul sont toutefois bien connues et facilement applicables dans un langage statistique clair ou sur une feuille calcul. La feuille de calcul ci-jointe estime un modèle démographique reposant sur une approximation de diffusion pour un ensemble simple de données de recensement. Toutefois, pour concilier les estimations d'erreur d'une part, et le taux d'évolution démographique et la variation du processus d'évolution démographique d'autre part, il est plus facile de réaliser l'analyse dans un langage statistique. R est un langage statistique et logiciel gratuits qui fonctionne sans problème sur n'importe quelle plate-forme. La page suivante contient des instructions et le code pour réaliser l'analyse présentée dans Morris & Doak (2002) pour un modèle indépendant de la densité (voir chapitre 3, p. 51) Morris & Doak (2002) présentent différentes applications des estimations de paramétrage et des mesures diagnostiques qui peuvent être utilisées, ainsi que des extensions de modèle. Il est possible de modifier sans difficulté à ces fins le code R fourni dans la section suivante.

Il y a quelques différences méthodologiques qu'il importe de signaler en comparant les deux approches analytiques. Le programme TRIM comporte une autre formulation de la vraisemblance pour le modèle de Poisson, pour expliquer la non-indépendance de recensements successifs (une violation des suppositions des GLM standard). Par comparaison, les modèles de dynamique démographique présentés ici sont une classe de processus Markov de premier ordre, c'est-à-dire que les valeurs du recensement suivant dépendent des valeurs précédentes. À la différence de l'approche GLM de Poisson, les modèles Markov du processus biologique supposent que les observations successives sont corrélées – et modélisent en fait le processus qui crée cette corrélation. Ces simples modèles biologiques supposent toutefois que le taux de changement de l'abondance est

constant, en d'autres termes, que le taux de ce changement est constant, bien que les recensements puissent changer d'une année à l'autre. Il existe des méthodes pour trouver les discontinuités du taux de changement, telles que présentées par Morris & Doak (2002), et les extensions simples ne sont pas difficiles à appliquer. Les tendances complexes du taux de changement à travers l'espace et le temps, ou accompagnées d'autres conditions, nécessitent toutefois des analyses plus complexes. À cet égard, le programme TRIM est plus souple du fait qu'il accepte des covariables sans difficulté.

Travaux cités

- Dennis, B., P. L. Munholland, *et al.* (1991). "Estimation of Growth and Extinction Parameters for Endangered Species." Ecological Monographs **61**(2): 115-143.
- Morris, W. F. & D. F. Doak (2002). Quantitative Conservation Biology Theory and Practice of Population Viability Analysis. Sunderland, MA, Sinauer Associates.
- Pinheiro, J. C. & D. M. Bates (2000). Mixed-Effects Models in S and S-PLUS. New York, Springer.
- Staples, D. F., M. L. Taper, *et al.* (2005). "Risk-based viable population monitoring." Conservation Biology **19**(6): 1908-1916.
- Tuck, G. N., T. Polacheck, *et al.* (2001). "Modelling the impact of fishery by-catches on albatross populations." Journal of Applied Ecology **38**(6): 1182-1196.
- Venables, W. N. & B. D. Ripley (1999). Modern applied statistics with S-PLUS. New York, Springer.
- Woehler, E. J., J. Cooper, *et al.* (2001). A statistical assessment of the status and trends of Antarctic and Subantarctic seabirds. Hobart, SCAR Bird Biology Subcommittee, Scientific Committee of CCAMLR: 43.

Échantillon de code R

Ce code estime le taux de croissance démographique « mu », l'intervalle de confiance autour de la moyenne et la variance "sigma.sq" pour un modèle d'approximation de la diffusion indépendant de la densité, tel qu'exposé dans Morris & Doak 2002.

Pour exécuter le code R, il faut créer le fichier "data for R.csv" dans la feuille de calcul Excel ci-jointe en remplaçant les données sur la feuille de « sample data » [données d'essai] du cahier "sample analysis" [analyse d'essai]. Changez le format de la feuille de calcul en sélectionnant "export to R" et sauvegardez-la comme fichier .CSV dans le répertoire à partir duquel vous souhaitez travailler. Veillez à insérer toutes les formules de façon à ce que toutes les données soient incluses. Puis lancez R et choisissez le répertoire où les "data for R.csv" sont sauvegardées comme répertoire de travail. Insérez le code ci-dessous dans la fenêtre R. Tapez "mu", "mu.confidence.interval" ou "sigma.sq" et les valeurs respectives seront affichées à l'écran.

```
raw.data <- read.csv("data for R.csv",header=T)
as.data.frame(raw.data)
attach(raw.data)
DA.est <- lm(trans.ln.change~0 + trans.interval)
Mu <- coefficients(DA.est)
names(Mu) = NULL
sigma.sq <- anova(DA.est)$"Mean Sq"[2]
Mu.confidence.interval <- confint(DA.est,1)
```