ESTIMATION D'UN INDICE ANNUEL POUR LE RELEVÉ DES OISEAUX NICHEURS

Modèle de l'indice annuel

Si y_{ij} représente le dénombrement observé sur le parcours i pendant l'année j et u_{ij} est une variable indicatrice montrant si l'observation a été faite, c.-à-d. :

$u_{ij} = 1$ si une observation a été faite sur le parcours i pendant l'année j
$u_{ij} = 0$ dans le cas contraire

Le modèle ne peut fonctionner pour toutes les observations. Chaque parcours est divisé en blocs d'observations effectuées dans des conditions comparables. Chacun de ces blocs correspond aux dénombrements relevés que ce soit par observateur, par moment de l'année, par moment de la journée ou par condition météorologique. L'espèce doit avoir été aperçue au moins une fois dans un bloc pour que le modèle fonctionne avec les données. En outre, si une espèce n'est jamais observée sur aucun parcours pendant une année donnée, cette année ne peut être incluse dans le modèle.

$v_{ij} = 1$ si une observation a été faite sur le parcours i pendant l'année j et peut être utilisée pour l'analyse
$v_{ij} = 0$ dans le cas contraire

Si w_i est le secteur que le parcours i représente dans la population.

On présumera que le y_{ij} aura une distribution de Poisson avec une valeur attendue de λ_{ij}, c.-à-d. :

$$P(y_{ij} = k) = \frac{\lambda_{ij}{}^{k} \exp(-\lambda_{ij})}{k!}$$

On présumera que la valeur attendue aura la structure suivante :

$$\lambda_{ij} = \exp(\mu + \alpha_i + \beta_j + \gamma_{i(k)})$$

où
μ représente la moyenne globale
α_i représente l'effet du parcours i
β_j représente l'effet de l'année j

$\gamma_{i(k)}$ représente l'effet des conditions d'observation.

L'effet des conditions d'observation comprend les effets de tous les facteurs utilisés pour diviser les données en blocs comparables. Cela comprend les effets des observateurs et d'autres termes comme les conditions météorologiques, le moment de la journée ou la date.

Multiplier les termes de probabilité les uns par les autres et prendre les logarithmes donne l'équation de probabilité :

$$L = \sum_{i}^{n} w_i \sum_{j}^{m} v_{ij} \left\{ y_{ij} (\mu + \alpha_i + \beta_j + \gamma_{i(k)}) - \exp(\mu + \alpha_i + \beta_j + \gamma_{i(k)}) - \ln(y_{ij}) \right\} \qquad (1)$$

Les estimations maximales de probabilité des paramètres sont la solution de l'ensemble d'équations suivant :

$$\sum_{i}^{n} w_i \sum_{j}^{m} v_{ij} \left\{ y_{ij} - \exp(\hat{\mu} + \hat{\alpha}_i + \hat{\beta}_j + \hat{\gamma}_{i(k)} \right\} = 0 \qquad (2)$$

$$\sum_{j}^{m} v_{ij} \left\{ y_{ij} - \exp(\hat{\mu} + \hat{\alpha}_i + \hat{\beta}_j + \hat{\gamma}_{i(k)} \right\} = 0 \qquad (3)$$

$$\sum_{i}^{n} w_i v_{ij} \left\{ y_{ij} - \exp(\hat{\mu} + \hat{\alpha}_i + \hat{\beta}_j + \hat{\gamma}_{i(k)} \right\} = 0 \qquad (4)$$

$$\sum_{j \varepsilon k}^{m} v_{ij} \left\{ y_{ij} - \exp(\hat{\mu} + \hat{\alpha}_i + \hat{\beta}_j + \hat{\gamma}_{i(k)} \right\} = 0 \qquad (5)$$

où le cumul de l'équation (5) s'étend sur plusieurs années en ce qui concerne les observations du bloc de conditions k.

Estimation des paramètres du modèle

Les estimateurs maximaux de probabilité ne peuvent être rédigés sous forme fermée et les estimations sont créées à l'aide d'un mécanisme itératif.

Les estimations initiales ont été fixées à $\hat{\alpha}_i^{(0)} = 0$, $\hat{\beta}_j^{(0)} = 0$, $\hat{\gamma}_{i(k)}^{(0)} = 0$ et

$$\hat{\mu}^{(0)} = \frac{\sum_{i}^{n} w_i \sum_{j}^{m} v_{ij} y_{ij}}{\sum_{i}^{n} w_i \sum_{j}^{m} v_{ij}}$$ (c.-à-d. la moyenne pondérée de secteur de toutes les observations)

Étant donné les estimations à l’étape g ($\alpha_i^{(g)}, \beta_j^{(g)}, \gamma_{i(k)}^{(g)}$), les estimations pour la prochaine étape sont calculées de la façon suivante :

$$\sum_{j}^{m} v_{ij} \left\{ y_{ij} - \exp(\hat{\mu}^{(0)} + \hat{\alpha}_i^{(g+1)} + \hat{\beta}_j^{(g)} + \hat{\gamma}_{i(k)}^{(g)} \right\} = 0 \qquad (6)$$

$$\sum_{i}^{n} w_i v_{ij} \left\{ y_{ij} - \exp(\hat{\mu}^{(0)} + \hat{\alpha}_i^{(g+1)} + \hat{\beta}_j^{(g+1)} + \hat{\gamma}_{i(k)}^{(g)} \right\} = 0 \qquad (7)$$

$$\sum_{j \in k}^{m} v_{ij} \left\{ y_{ij} - \exp(\hat{\mu}^{(0)} + \hat{\alpha}_i^{(g+1)} + \hat{\beta}_j^{(g+1)} + \hat{\gamma}_{i(k)}^{(g+1)} \right\} = 0 \qquad (8)$$

Le côté gauche des équations (6) à (8) ne comporte qu’une seule variable inconnue et est une fonction monotone décroissante. On trouve la solution en calculant d’abord l’équation au niveau actuel du paramètre, puis en augmentant ou en diminuant de un jusqu’à ce que la solution soit bornée. La solution bornée est ensuite raffinée au moyen d’une recherche binaire.

Lorsque l’itération est terminée, on calcule le changement dans les estimations des paramètres

$$C^{(g)} = \sum_{i}^{n} (\alpha_i^{(g)} - \alpha_i^{(g+1)})^2 + \sum_{j}^{m} (\beta_j^{(g)} - \beta_j^{(g+1)})^2 + \sum_{i}^{n} \sum_{k} (\gamma_{i(k)}^{(g)} - \gamma_{i(k)}^{(g+1)})^2 \qquad (9)$$

Les itérations se poursuivent jusqu’à $C^{(g)} < tol$, où tol est un niveau de tolérance à la convergence. La section A5 ci-dessous traite du choix d’une valeur de tolérance convenable.

Dans l’algorithme précédent, l’estimation $\hat{\mu}^{(0)}$ n’est jamais mise à jour par les étapes d’itération. Cela s’explique par le fait qu’il s’agit d’un paramètre redondant qui peut être fixé arbitrairement. Lorsque les itérations ont convergé, les estimations sont ajustées aux contraintes $\sum_{k} \hat{\gamma}_{i(k)} = 0$ pour tous les i, et $\sum_{j}^{m} \beta_j = 0$ comme suit. Si G représente la dernière itération, alors les effets des blocs de conditions sont ajustés pour totaliser zéro dans chaque parcours de la façon suivante :

$$\bar{\gamma}_i^{(G)} = \sum_{k} \hat{\gamma}_{i(k)}^{(G)} / K_i \qquad \hat{\gamma}_{i(k)} = \hat{\gamma}_{i(k)}^{G} - \bar{\gamma}_i^{(G)} \qquad \hat{\alpha}_{i(k)} = \hat{\alpha}_{i(k)}^{G} + \bar{\gamma}_i^{(G)}$$

Les effets des blocs sont ensuite ajustés de façon à ce que chacun totalise zéro grâce aux calculs suivants :

$$\bar{\beta}_j^{(G)} = \sum_{j}^{m} \hat{\beta}_j^{(G)} / m \qquad \hat{\beta}_j = \hat{\beta}_j^{(G)} - \bar{\beta}_j^{(G)} \qquad \hat{\mu} = \hat{\mu}^{(0)} + \bar{\beta}^{(G)}$$

Les ajustements conservent le même dénombrement prévu pour chaque observation.

Indices annuels et tendances

L'indice annuel de la population est calculé de la façon suivante :

$$I_j = \left(\frac{\sum_i w_i v_{ij}}{\sum_i w_i u_{ij}} \right) \exp(\hat{\mu} + \hat{\beta}_j) \qquad (10)$$

Le terme entre parenthèses de l'équation (10) réduit l'indice pour refléter le fait que le modèle peut être appliqué seulement à une partie de l'ensemble de données. Il s'agit d'une estimation du dénombrement qui aurait été effectué pendant l'année j si tous les parcours avaient été effectués dans des conditions moyennes.

L'ajustement du nombre de parcours observés dans une région donnée peut avoir une incidence sur l'interprétation des données. Le tableau A1 montre un exemple. Dans cet exemple, trois des quatre parcours peuvent être utilisés avec le modèle. Toutefois, pour le parcours 3, seules les données du premier bloc peuvent être utilisées dans l'analyse. Si l'ajustement du nombre de parcours n'était pas inclus, le fait que l'espèce est disparue du parcours serait ignoré dans l'analyse. Une autre incidence relative à l'ajustement du nombre de parcours est de tenir compte de la proportion du nombre de parcours sur lesquels l'espèce est observée, puisque les parcours sont abandonnés et adoptés à mesure que des bénévoles se retirent ou sont recrutés pour le relevé. Dans l'exemple du tableau A1, le parcours 4 est adopté au cours de la 7^{e} année. Le fait que la proportion du nombre de parcours où l'espèce n'est pas observée passe de 100 p. 100 au début à seulement 50 p. 100 dans les dernières années est pris en compte par le terme entre parenthèses de l'équation (9).

Pour l'année où l'espèce n'est jamais observée, l'indice annuel est fixé à zéro. Il s'agit de l'estimation maximale de pondération dans cette situation, mais l'inclusion de cette année dans le processus d'adaptation du modèle aurait soulevé des difficultés pour l'algorithme informatique.

On peut placer dans l'indice annuel une estimation de la tendance dans l'échelle d'entrée de la façon suivante :

$$\tau = \frac{\sum_j (x_i - \overline{x}) \ln(I_j)}{\sum_j (x_i - \overline{x})^2} \qquad (11)$$

où x_j représente la j^e année réelle d'observation et $\bar{x}$ représente la moyenne de toutes les années d'observations. Le cumul s'étend sur toutes les années pendant lesquelles l'indice annuel est différent de zéro.

Estimation de l'écart-type

On estime l'écart-type de l'indice annuel au moyen d'une méthode du jack-knife. Dans cet algorithme, des pseudo-répétitions sont créées en éliminant un parcours à la fois et en recalculant l'indice annuel et la tendance. La variabilité des pseudo-répétitions sert à estimer la variance de l'indice annuel et de la tendance. Dans le calcul mis en œuvre dans le programme, tous les parcours sont utilisés (y compris les parcours où l'espèce n'a jamais été observée) pour calculer la variance. Parce que le nombre de parcours varie d'une année à l'autre, le nombre de pseudo-répétitions varie d'un indice annuel à l'autre.

La matrice complète de variance-covariance pourrait être estimée au moyen de la méthode du jack-knife, mais parce que le nombre d'observations varie d'année en année, le nombre d'observations utilisées pour créer chaque entrée du tableau sera différent. De telles complexités peuvent limiter l'utilité de la matrice estimée. L'estimation de la matrice complète de variance-covariance n'a pas été mise en oeuvre.

Exactitude de l'adaptation du modèle

Puisque les estimations maximales de probabilité ne peuvent être rédigées sous forme fermée, les estimations sont calculées au moyen d'algorithmes afin de trouver la solution de diverses équations. Les algorithmes supposent une routine de recherche itérative telle que décrite à la section A2 ci-dessus. Le critère de convergence est donné par l'équation (9). L'utilisation d'une valeur plus petite pour la tolérance offre une estimation plus exacte, toutefois l'algorithme sera plus long à exécuter.

Afin de choisir une tolérance convenable, l'algorithme a été exécuté avec la tolérance fixée à 10^{-10} (c.-à-d. : log(tol)=-10). On présume que cela fournit une estimation exacte du meilleur modèle adapté. Le programme a alors été exécuté avec le log(tol) =-4, -5,…,-9, puis les indices annuels résultants et leurs écarts-types ont été comparés à ceux obtenus avec log(tol)=-10.

$$rel_j(\log(tol)) = \frac{\left| I_j(\log(tol)) - I_j(-10) \right|}{I_j(-10)}$$

où $I_i(\log(tol))$ représente l'indice annuel de l'année j estimé avec le critère de convergence tol. Un calcul semblable a été effectué pour l'écart-type estimé.

Le modèle a été adapté à 10 espèces pour une période de 32 ans, et la valeur maximale de l’erreur absolue relative est indiquée dans le tableau A2. On peut constater que l’exactitude augmente à mesure que la tolérance diminue, mais qu’il y a peu de changement pour log(tol) <-8. Fixer log(tol)=-8 donne des estimations exactes jusqu’à au moins 4 décimales pour l’indice annuel et jusqu’à 3 décimales pour l’écart-type.

Tableau A1 : Ajustement pour le nombre de parcours inclus dans l'analyse.

	Année							
Parcours	1	2	3	4	5	6	7	8
1	x	x	x	-	x	-	x	x
2	x	-	x	x	x	x	x	x
3	x	x	-	x	0	0	0	0
4							0	0

x représente un dénombrement observé de l'espèce différent de zéro
0 représente un dénombrement observé de zéro
- représente une observation manquante
Un ensemble d'observations encadré indique que les observations ont été effectuées dans les mêmes conditions, c.-à-d. par le même observateur.

Tableau A2 : Différence absolue relative maximale entre les indices annuels et leurs écarts-types adaptés aux différents niveaux de tolérance

	----------Erreur relative maximale----------	
log(tol)	Indice annuel	Écart-type pour l'indice annuel
-4	0,0250	0,0556
-5	0,0213	0,0233
-6	0,0213	0,0127
-7	0,0068	0,0053
-8	0,0001	0,0017
-9	0,0001	0,0015