



## CERTIFICAT EN ANALYSE DE DONNÉES POUR L'ÉCOLOGIE ET LA GESTION DE LA BIODIVERSITÉ

### Cas pratique : modèle linéaire généralisé sous R

Jean-Yves Barnagaud : [jean-yves.barnagaud@ephe.psl.eu](mailto:jean-yves.barnagaud@ephe.psl.eu)

**Objectifs :** Construire un modèle linéaire généralisé à partir d'une question biologique, l'écrire formellement sous forme d'une équation, vérifier les conditions d'application et en tirer les conséquences, l'implémenter sous R, le vérifier, l'interpréter et représenter les résultats. Faire preuve d'imagination / de flexibilité dans l'application des méthodes afin de répondre au mieux à la question posée avec les données disponibles, sans sacrifier la rigueur statistique.

**Cas d'étude 1:** Nous réanalysons des données d'écologie comportementale issues d'une étude sur l'olfaction des baleines (Bouchard et al., PLOS One, 2019). La question posée était : les baleines répondent-elles à un stimulus olfactif évoquant leur ressource alimentaire ?

**Hypothèse :** Les baleines sont attirées par les odeurs du krill et du diméthyl-sulfide (DMS, un hydrolysat de plancton), qui évoquent leur ressource alimentaire.

**Données :** Les données sont stockées dans *baleines.txt*. Le dispositif expérimental implique un voilier qui circule dans une zone connue pour être fréquentée par des baleines (ici, uniquement des baleines à bosse). Quand un groupe de baleines est repéré (généralement à plusieurs centaines de mètres), un dispositif composé d'un flotteur et d'un conteneur immergé est mis à l'eau. Le voilier s'écarte à quelques centaines de mètres. A partir du moment où le groupe de baleines franchit la limite d'une zone-test de 50m, on considère qu'il s'est approché du dispositif : on chronomètre alors le temps passé dans les 50 m autour du dispositif (time\_z12). Le conteneur peut être chargé d'argile (témoin), de krill ou de DMS (odeurs assimilables à l'alimentation des baleines), ou rester vide (deuxième témoin). L'expérience est répliquée dans une zone de reproduction (Madagascar) et deux zones d'alimentation (Islande et Antarctique).

**Cas d'étude 2:** Le gestionnaire de la réserve naturelle des salins d'Hyères (agglomération de Toulon) a souhaité savoir si les modalités de gestion de l'eau dans ses salins avaient une influence sur la dynamique des espèces d'oiseaux ciblées par son plan de gestion. Depuis plusieurs années, un suivi décadaire des oiseaux est effectué sur la réserve par un ornithologue, et des mesures de salinité, d'oxygène, de température de l'eau et de niveau d'eau sont prises sur des bassins-test du salin. L'étude originale concerne plusieurs espèces et

deux salins (Barnagaud et al., Biological Conservation, 2019), mais nous allons nous concentrer sur la Mouette rieuse et un seul salin (le plus grand, le Salin des Pesquiers).

Hypothèse : La Mouette rieuse étant une espèce d'eau douce, on attend que ses effectifs baissent quand la salinité augmente. Un excès d'oxygénation et une température trop élevée limitent la disponibilité en ressource alimentaire et devraient aussi faire baisser les effectifs de mouettes. En revanche, on s'attend à ce que le niveau d'eau, qui joue un rôle protecteur pour les colonies, augmente les effectifs de mouettes.

Données : Les données sont stockées dans *mouette\_rieuse\_hyeres.txt*. Elles ont été préalablement filtrées afin d'enlever un certain nombre de données de salinité et d'oxygénation manquantes (donc qui ne seraient de toute façon pas incluses dans une analyse par modélisation), mais leur structure n'a pas été modifiée. On dispose de la date du comptage, du bassin concerné (le salin est constitué d'un grand nombre de ces bassins, mais seuls 14 sont suivis), de la concentration de l'eau en sel (g/L), de l'oxygène dissous (mg/L), du niveau d'eau (cm), de l'année du suivi et de la date ré-exprimée en date julienne.

### **Tâches à effectuer :**

- Comprendre la question et la reformuler comme un modèle statistique
- Explorer les données, identifier leur structure, les difficultés qui pourraient se poser
- Ajuster éventuellement la stratégie de modélisation au regard de ces difficultés
- Implémenter le modèle sous R, vérifier les conditions d'application
- Interpréter les résultats et conclure sur les hypothèses
- Préparer une restitution orale de 10 minutes maximum. Cette fois, la restitution devra se présenter sous la forme d'une présentation des résultats pour un colloque d'écologie. Il s'agira donc pour vous de formuler question, modèle et résultats du point de vue de l'écologue, en minimisant l'information purement technique et en la ré-exprimant autant que possible au regard de la biologie. Vous devrez donc identifier ce qui est nécessaire et suffisant de savoir sur l'analyse statistique que vous avez effectuée, et de transcrire l'ensemble en un raisonnement d'écologue.

Conseils : Ayez un regard critique sur vos données : si une difficulté surgit, dites-vous qu'elle peut être surmontée ; ce sont votre compréhension biologique du sujet et votre imagination qui vous permettront de trouver la solution. Faites les vérifications nécessaires des conditions d'application des modèles que vous utilisez. Un modèle de régression est robuste à des déviations substantielles à ses hypothèses, mais jusqu'à un certain point. Ne négligez pas la restitution : le travail de reformulation d'une analyse statistique n'est pas si évident que ça, surtout quand vous ne travaillez pas sur votre sujet habituel. Posez-vous la question de ce que vous voudriez savoir à propos de ces analyses en tant qu'auditeur dans un colloque, et tenez-y vous.