

Suivi scientifique d'espèces animales

Aspects méthodologiques

essentiels pour l'élaboration de protocoles de suivis

Provence
Alpes
Côte d'Azur



*Note méthodologique
à l'usage des gestionnaires de sites Natura 2000*



Rédaction : Aurélien BESNARD
Coordination : Jean-Marc SALLES

Commanditaire : DREAL PACA – Pôle Natura 2000

Juin 2010

Suivi scientifique d'espèces animales

Aspects méthodologiques essentiels pour l'élaboration de protocoles de suivis

Note méthodologique à l'usage des gestionnaires de sites Natura 2000

DREAL PACA – Pôle Natura 2000

(contact : scientifique.natura@developpement-durable.gouv.fr)



Rédacteur
principal

Aurélien BESNARD

Maître de conférence à l'Ecole Pratique des Hautes Etudes – laboratoire biologie et écologie des vertébrés, à Montpellier. Expert dans le développement de protocoles de suivis d'espèces et l'analyse statistique de données en écologie et dynamique des populations animales. Il intervient en tant qu'enseignant sur les aspects monitoring d'espèces dans le cadre de l'EPHE mais aussi à l'ATEN. Il s'investit particulièrement sur le lien entre le monde de la recherche et celui de la gestion en collaborant régulièrement avec des programmes Life, les parcs nationaux, les réserves nationales mais aussi l'ONCFS, l'ONF et certains bureaux d'études naturalistes.

Page web : http://www.cefe.cnrs.fr/ebv/staff/Aur%C3%A9lien_Besnard.htm

Coordinateur et
rédacteur
secondaire

Jean-Marc SALLES

Ingénieur écologue, ECO-MED (Ecologie et Médiation), bureau d'études spécialisé dans l'expertise du patrimoine naturel (habitats, flore, faune). Coordinateur scientifique Natura 2000 auprès de la DREAL PACA. A participé à la mise en œuvre de divers programmes de suivi scientifique, notamment d'oiseaux et mammifères marins dans les Terres Australes et Antarctiques Françaises.

Relecteurs

Olivier GIMENEZ (CEFE/CNRS), Vincent RIVIERE (ECO-MED)

Citation :

Besnard A. & J.M. Salles, 2010. *Suivi scientifique d'espèces animales. Aspects méthodologiques essentiels pour l'élaboration de protocoles de suivis. Note méthodologique à l'usage des gestionnaires de sites Natura 2000.* Rapport DREAL PACA, pôle Natura 2000. 62 pages.

SOMMAIRE

Préambule	5
1. Introduction : contexte historique	8
2. Rappels de quelques concepts fondamentaux	9
2.1. Concepts d'écologie et de biologie des populations.....	9
2.2. Concepts de biostatistique.....	13
3. Limites et biais des méthodes usuelles	16
3.1. Rappel sur les inventaires	16
3.2. Rappel sur les Indices d'abondance	17
3.3. La notion de détectabilité	18
4. Echantillonner	22
4.1. Pourquoi échantillonner	22
4.2. Notions de bases en échantillonnage.....	22
4.3. Définir la zone d'étude / la notion de favorable.....	23
4.4. Les grands types de plan d'échantillonnage.....	25
4.4.1. Echantillonnage classique	25
4.4.2. Echantillonnage stratifié	26
4.4.3. Adaptive sampling	27
4.5. Les contraintes et pseudo-aléatoire	29
4.6. Réplicats temporels et échantillonnage.....	29
5. Les méthodes d'analyse de données et leurs hypothèses	30
5.1. Les recensements	30
5.2. Le « Distance Sampling ».....	32
5.3. Les Captures-Marquage-Recaptures.....	37
5.4. Les méthodes d' « <i>occupancy</i> » ou présence-absence.....	43
5.5. Choisir une méthode de suivi	45
6. Mettre en évidence un impact : Le BACI	49
7. Stratégie générale pour construire un protocole	51
8. Cas d'école	54
8.1. Cas 1 : mauvaise prise en compte de l'écologie de l'espèce.....	54
8.2. Cas 2 : intérêt d'une pré-étude	55
8.3. Cas 3 : choix entre « <i>Distance Sampling</i> » et CMR	56
8.4. Cas 4 : ajuster l'effort de prospection par l'analyse régulière des données collectées	58
Annexe 1. Bibliographie	60
Annexe 2. Logiciels d'analyse de données	62

TABLE DES FIGURES

Figure 1 :	Schématisation d'un écosystème forestier, système complexe d'interrelations entre des éléments physiques (biotope) et biologiques (biocénoses) (source : internet).	9
Figure 2 :	De l'écocomplexe à l'habitat. Exemple du massif des Alpilles (sources : PNRA).....	10
Figure 3 :	Illustration de la variance, dispersion des données autour de la moyenne.....	14
Figure 4 :	Exemple d'intervalle de confiance.....	15
Figure 5 :	Augmentation de l'effectif observé au cours du temps.....	19
Figure 6 :	Variation de la mesure de l'effectif en fonction de différents scénarii de détectabilité..	20
Figure 7 :	Echantillonnage simple. Placement aléatoire de 10 transects au sein d'une zone d'étude	25
Figure 8 :	Echantillonnage stratifié. Placement aléatoire de 12 transects au sein d'une zone d'étude, répartis pas secteurs	27
Figure 9 :	Adaptive sampling.....	28
Figure 10 :	Echantillonnage aléatoire avec exclusion des zones non occupées	26
Figure 11 :	Protocole de recensement par quadrat rectangulaire ou circulaire.....	30
Figure 12 :	La détection peut varier selon les types de milieux présents dans le quadrat.....	31
Figure 13 :	Protocole de « Distance Sampling » : « Line transect » et « Point transect ».....	33
Figure 14 :	Exemples de télémètres lasers disponibles sur le marché	34
Figure 15 :	Illustration de différentes techniques de marquage	39
Figure 16 :	CMR. Comparaison entre deux populations, close et non close	40
Figure 17 :	CMR. Matrice des histoires de capture par individu	41
Figure 18 :	Présence-absence. Matrice des histoires de capture par espèce.....	43
Figure 19 :	Présence-absence. Exemples types d'histoires de capture par espèce	44
Figure 20 :	Positionnement des différentes méthodes en fonction de divers critères de choix.....	46
Figure 21 :	Protocole BACI. Comparaison d'un site témoin et d'un site altéré.	50

PREAMBULE

La DREAL PACA (Direction Régionale de l'Environnement, de l'Aménagement et du Logement de la région Provence-Alpes-Côte d'Azur) anime et coordonne la mise en œuvre du réseau écologique européen Natura 2000. A l'échelle de la région PACA, ce réseau est constitué de 128 sites Natura 2000, couvrant plus de 30% de la surface régionale, soit près d'1 million d'hectares. A ce titre, la DREAL PACA encadre la réalisation des documents d'objectifs (DOCOB), qui définissent les modalités de gestion écologique propres à chaque site Natura 2000. Elle a élaboré en 2002, puis 2007, un cahier des charges régional, permettant un cadrage méthodologique pour la réalisation d'inventaires naturalistes (faune, flore et habitats d'intérêt communautaire), conduits par chaque opérateur local de site Natura 2000, pour l'élaboration des DOCOB.

Dans la continuité de ce travail, la DREAL PACA souhaite apporter un cadrage régional aux opérateurs et animateurs locaux pour les suivis biologiques à réaliser sur les sites Natura 2000 en vue de suivre l'évolution des habitats et espèces remarquables et d'évaluer périodiquement l'efficacité des mesures de gestion mises en œuvre. Bien que la finalité de ces suivis soit évidente de prime abord, elle se heurte à des interrogations d'ordre :

- scientifique : quels habitats / espèces doit-on suivre en priorité ? Quels protocoles doit-on utiliser pour garantir des résultats fiables et robustes (modalités, fréquence dans le temps, taille de l'échantillon représentatif, etc.) ?
- Budgétaire : comment optimiser l'utilisation des crédits publics, forcément limités par rapport aux stricts besoins scientifiques, quant à eux illimités ?

La fiabilité et la robustesse des résultats constituent un objectif majeur. En effet, contrairement à un inventaire naturaliste dressant un état des lieux à un instant T, un protocole de suivi biologique doit s'inscrire dans une démarche de veille régulière, sur le moyen et le long terme. Ainsi, afin que les résultats collectés l'année 1 puissent être à nouveau exploitables l'année 1+ N , le protocole doit être bien pensé afin de limiter au maximum les biais pouvant survenir au cours du temps (variabilité de l'expert réalisant le suivi, adaptation de l'échantillonnage aux péripéties du moment, années « vides », évolution des méthodes d'analyse biostatistique, etc.). Un retour d'expérience montre qu'une mauvaise anticipation conduit le plus souvent à collecter des données trop éphémères dans l'espace et dans le temps, c'est-à-dire ne permettant pas une extrapolation à une plus vaste échelle géographique, ni une réutilisation plusieurs années par la suite. Afin de pouvoir suivre correctement la mise en œuvre du DOCOB, orienter efficacement les actions de gestion, ainsi que de garantir une bonne utilisation des deniers publics, il est donc crucial d'utiliser des protocoles permettant d'obtenir des résultats fiables et robustes, dans l'espace et dans le temps.

Pour construire une stratégie régionale permettant d'établir des priorisations, la DREAL PACA propose une démarche en 4 phases :

- Phase 1 : définition du cadre méthodologique général -> élaboration d'une note méthodologique décrivant l'état de l'art en matière de protocoles scientifiques. Cette note technique doit constituer le socle scientifique indispensable avant d'engager toute réflexion sur les sites Natura 2000 de PACA.
- Phase 2 : application au contexte particulier de la région PACA -> sur la base de cette note méthodologique et des listes priorisées d'habitats / espèces (niveau régional et national), identification des suivis à privilégier en PACA, selon quels types de protocoles, sur quels sites Natura 2000 et avec quel budget approximatif. Elaboration

par la DREAL d'une note stratégique, définissant les priorités régionales en fonction du ratio « pertinence scientifique / coût » de chaque suivi.

- Phase 3 : concertation avec les experts scientifiques du CSRPN. Validation ou adaptation de la stratégie régionale.
- Phase 4 : diffusion au réseau -> une fois la stratégie validée, élaboration et diffusion aux opérateurs / animateurs Natura 2000 d'une feuille de route, par site ou par DOCOB, déclinant les suivis à conduire en priorité.

Le présent rapport constitue la note méthodologique générale de la phase 1. Après un rappel de quelques concepts fondamentaux, ce rapport présente les grands types de protocoles disponibles à ce jour pour suivre des populations animales, leurs conditions d'utilisation (fonction de différentes variables telles que : l'abondance de l'espèce étudiée, sa détectabilité, les contextes biogéographique et topographique du site, etc.), ainsi qu'une analyse critique de leurs avantages / inconvénients. En fin de rapport, un chapitre est consacré aux critères de choix du « meilleur » protocole, choix souvent difficile car fonction de critères multiples.

Remarque : cette note était initialement envisagée pour couvrir le champ complet des habitats et espèces (flore et faune). Compte-tenu des grandes différences d'approche entre ces différents compartiments biologiques, il est apparu plus opportun, dans un premier temps, de cibler les méthodes appliquées à la faune. Pour **la flore et les habitats Natura 2000**, une réflexion est en cours avec les CBN alpin et méditerranéen, qui fera ultérieurement l'objet d'une synthèse. En première approche, le paragraphe relatif à la méthode BACI peut toutefois être extrapolé à l'ensemble des compartiments biologiques.

* * * * *

Avant-propos de l'auteur (A. Besnard)

Il est important ici de rappeler qu'il existe un certain nombre d'ouvrages de synthèse présentant de manière souvent très pointue les techniques de suivis de la faune et la flore. Ces ouvrages sont parfois centrés sur des groupes taxonomiques particuliers, d'autres sont plus généraux. Notons cependant que la très grande majorité de cette bibliographie disponible est en anglais. Une série de documents cependant, importante dans le domaine, est celle produite par Valérie FIERS et publiée par l'Atelier Technique des Espaces Naturels et par Réserves Naturelles de France, sous forme de cahiers techniques (cf. bibliographie en annexe 1). Cette collection propose une synthèse bibliographique conséquente ainsi qu'une synthèse des techniques utilisées pour de nombreux taxons. Elle constitue en cela une collection précieuse à consulter. Cependant, tous ces documents s'intéressent d'une manière générale aux techniques et méthodes de collecte de données et non à leur exploitation ultérieure. **Au cours des quarante dernières années, les méthodes d'analyse des données spécifiquement issues de suivis d'espèces animales ou végétales ont très largement évolué.** Ces évolutions ont notamment permis de montrer un certain nombre de biais existant dans des techniques de suivis jusqu'alors considérées comme pertinentes. De plus, avec ce recul de plusieurs années sur ces méthodes d'analyse, les personnes en charge des analyses de données (souvent des biostatisticiens) ont pu clairement mettre en évidence les limites même de ces nouvelles méthodes, **limites largement liées aux hypothèses statistiques qu'elles posent.** Ainsi leur évolution a notamment été guidée par la nécessité de relâcher certaines hypothèses trop contraignantes des méthodes classiques, pour analyser finement des jeux de données issus de suivis existants. Pourtant, malgré ces nombreux développements / raffinements, ces méthodes posent toujours un certain nombre d'hypothèses incontournables à l'heure actuelle. Le nombre de méthodes pertinentes pour analyser les données de suivis reste aussi assez restreint, comme nous le verrons par la suite. Ainsi mon expérience m'a souvent confronté à des jeux de données dont la méthode de collecte semble pertinente mais qui, après examen détaillé, se

révèle ne respecter les hypothèses statistiques d'aucune des méthodes « disponibles ». Dans cette situation, deux solutions existent :

- attendre qu'un statisticien développe la méthode d'analyse adéquate, mais cela peut prendre très longtemps, voire ne jamais se réaliser ;
- analyser les données avec les méthodes existantes et pointer les biais possibles.

Pour éviter ce genre de situation, je suis convaincu que la mise en place d'un protocole de suivi doit se réaliser en pleine connaissance des outils statistiques disponibles pour analyser les données collectées, connaissance des méthodes mais aussi donc des hypothèses statistiques sur lesquelles elles reposent. En effet, ces hypothèses se traduisent souvent par des contraintes sur le terrain. La phase de construction d'un protocole de suivi pertinent est donc un nécessaire aller-retour entre connaissance sur l'espèce et le site, budget disponible, technique de terrain potentielle, méthode d'analyse statistique disponible et précision souhaitée des résultats. Dans cette phase de construction des protocoles, l'omission d'un de ces facteurs (et de ses interactions évidentes avec les autres) conduit régulièrement à des catastrophes (effort de terrain conséquent pour aucun résultat au final, ou un résultat bien en deçà des attentes). La **note méthodologique** proposée ici porte donc bien son nom, **il ne s'agit pas de refaire à nouveau une synthèse des techniques de collecte de données mais de présenter de manière didactique, succincte et synthétique, les grandes méthodes d'analyse de données issues de suivis d'espèces, qui posent souvent des problèmes particuliers**. Ainsi mon objectif est de présenter les méthodes mais surtout leurs hypothèses statistiques fondamentales qui, comme nous le verrons, contraignent les manières de travailler sur le terrain et donc les techniques de collecte. Je présente donc ici le panel, finalement assez réduit, des méthodes d'analyse permettant d'estimer des effectifs de populations animales.

A ma connaissance, il s'agit du premier document de ce type en français. Il propose une version rédigée de cours que je donne depuis plusieurs années sur le sujet. Il est aussi le résultat de l'identification d'un manque flagrant de connaissance des opérateurs en charge des suivis d'espèces, sur ces aspects et sur leur importance lors des nombreuses consultations de conseils à l'analyse de données que je réalise dans le cadre de mes fonctions universitaires.

1. Introduction : contexte historique

Faire un historique de la biologie de la conservation serait ici long et fastidieux et s'éloignerait de nos préoccupations opérationnelles. Cependant, il est intéressant d'explorer pourquoi les questionnements autour des suivis scientifiques robustes émergent avec une telle force aujourd'hui dans les espaces protégés. **Historiquement, les biologistes de la conservation s'intéressaient majoritairement à des espèces que l'on pourrait qualifier d'emblématiques.** Ces espèces étaient majoritairement des oiseaux ou de grands mammifères, dont l'écologie était bien connue mais aussi aisés à suivre techniquement (forte détectabilité, capture et marquage relativement simple...). De ce fait, les experts avaient souvent une intuition assez bonne sur l'évolution des effectifs de ces espèces et sur les menaces qui pesaient sur leurs populations. De plus, lorsque cela n'était pas le cas, il était souvent relativement aisé de mettre en œuvre des protocoles de suivis afin de mieux comprendre le fonctionnement de ces populations.

Aujourd'hui l'objectif est la conservation de la biodiversité dans son ensemble. Passons outre le fait que cette idée pose un certain nombre de questions sur la définition même de la biodiversité, ainsi que des problèmes philosophiques ou éthiques sur l'interventionnisme que cela induit. Le sous-produit de cette volonté est que les « conservationnistes » s'intéressent aujourd'hui à un **panel d'espèces très large**, avec des caractéristiques biologiques, écologiques et des dynamiques de populations très différentes. Certaines espèces sont très mal connues, d'autres sont très complexes à suivre. La diversité des situations est très grande. Les experts restent rares sur certains groupes et leurs avis peuvent largement diverger.

Un autre phénomène majeur est à prendre en compte, celui de l'émergence de la **culture de l'évaluation des politiques publiques** dans les milieux de la conservation. Née vers 1930 aux Etats-Unis, l'évaluation des politiques publiques s'est fortement développée en Europe à partir de 1960, essentiellement dans les pays anglo-saxons. En France, elle s'installe de manière formelle avec le décret du 22 janvier 1990, affirmant que « *l'évaluation d'une politique publique consiste à comparer ses résultats aux moyens qu'elle met en œuvre, qu'ils soient juridiques, administratifs ou financiers, et aux objectifs initialement fixés. Elle doit aboutir à un jugement partagé sur l'efficacité de cette politique [...]* ». Les évaluations des actions de gestion ont été souhaitées par les pouvoirs publics, mais aussi plus directement par les gestionnaires d'espaces et d'espèces protégés dans un souci d'optimisation de l'utilisation de leurs budgets : les suivis doivent fournir les informations nécessaires aux évaluations menées dans les espaces protégés, afin de déterminer si les actions mises en œuvre sont efficaces.

Cette culture de l'évaluation constitue un axe fort au sein de la directive Habitats et de l'outil Natura 2000, et ce à plusieurs échelons géographiques :

- au niveau européen : synthèses régulières de la mise en œuvre du réseau Natura 2000, séminaires biogéographiques évaluant la complétude scientifique du réseau et signalant les éventuelles insuffisances,
- au niveau de chaque état membre : évaluation tous les 6 ans de l'état de conservation des habitats et espèces d'intérêt communautaire,
- au niveau régional et départemental : synthèses régulières par les services de l'Etat et le CSRPN,
- au niveau du site : suivi scientifique de l'évolution des habitats et espèces, évaluation de l'efficacité des mesures de gestion, évaluation de la mise en œuvre du DOCOB.

Toutes ces évolutions conduisent aujourd'hui à promouvoir des **stratégies de suivi permettant de produire des données robustes, quantitatives et homogènes**, permettant d'aller au-delà du simple dire d'experts lorsque ce dernier s'avère trop subjectif ou incertain du fait de la complexité des situations à analyser. Cette obligation de résultats implique une obligation de moyens, donc des arbitrages pouvant conduire à concentrer les efforts de suivis sur les habitats et espèces les plus pertinents.

2. Rappels de quelques concepts fondamentaux

2.1. Concepts d'écologie et de biologie des populations

Considérons un territoire quelconque d'une superficie relativement vaste, pouvant être un site Natura 2000, un parc naturel, un massif montagneux, etc. Celui-ci comprend divers **habitats** naturels et semi-naturels, définis par leurs conditions physico-chimiques (humidité, relief, géologie, nature et composition chimique des sols, microclimat, etc.), leur composition floristique (phytosociologie) et leur histoire, de déterminisme naturel (ex : orogénèse, érosion) et/ou anthropique (ex : façonnage par le pastoralisme, la sylviculture, etc.).

Ces différents habitats peuvent être regroupés en grands types de milieux, pouvant être rattachés à un **écosystème** particulier. Par exemple, différents types de formations boisées (se distinguant selon l'espèce d'arbre dominante) peuvent être regroupés sous l'item « forêts », et constituent l'écosystème forestier du site. Toutefois, la notion d'écosystème est un concept plus fonctionnel que géographique, rendant parfois difficile son zonage sur une carte (ex : transition graduelle entre une zone humide et une zone sèche, ou autres écotones).

Plusieurs écosystèmes peuvent être regroupés au sein d'**écocomplexes**, définis comme étant un assemblage d'écosystèmes inclus dans une même entité géographique. Sur une carte, un écocomplexe sera le plus souvent attribué à une entité géographique présentant une certaine unité. Par exemple telle vallée, telle chaîne montagneuse, telle plaine. Ce découpage permet d'identifier plusieurs secteurs au sein d'un vaste site naturel.

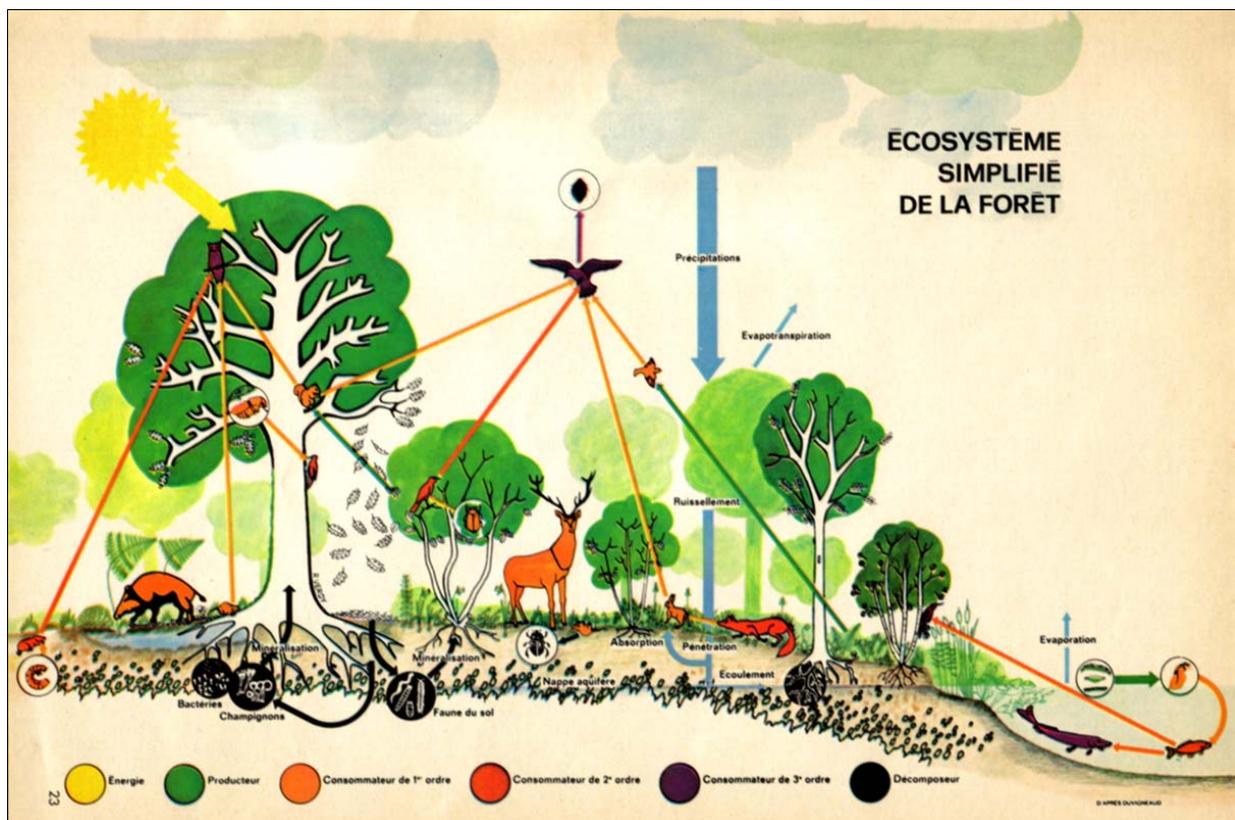


Figure 1 : Schématisation d'un écosystème forestier, système complexe d'interrelations entre des éléments physiques (biotope) et biologiques (biocénoses) (source : internet).

Si l'on considère les **espèces végétales et animales**, les mêmes concepts peuvent être appliqués.

Considérons une espèce quelconque. Sur notre vaste territoire, cette espèce comprend généralement plusieurs centaines voire milliers d'**individus**, dont la répartition spatiale peut être soit homogène, soit hétérogène. Dans le cas d'une espèce sexuée, ces individus doivent se reproduire pour permettre à l'espèce de perdurer sur le site. Ils forment alors une **population**, groupe d'individus appartenant à la même espèce et pouvant échanger leurs gènes au sein d'un territoire donné. La notion de population est un concept à dimension variable, pouvant être utilisé aussi bien pour désigner la population de la petite vallée X que la population européenne d'une même espèce. Lorsque ce terme est employé dans la littérature, l'échelle à laquelle il se rapporte est rarement sinon jamais précisée et perd ainsi de son sens biologique originel. C'est pourquoi des termes plus spécifiques sont parfois utilisés : les termes de micropopulation et de métapopulation. La **micropopulation** (ou noyau de population) désigne le groupe d'individus d'une même espèce à une échelle fine (rivière ou tronçon de rivière, vallée, crête sommitale, etc.) alors qu'une **métapopulation** désigne l'assemblage de plusieurs noyaux de populations entretenant des liens fonctionnels (ex : chiroptères se reproduisant au printemps-été en de multiples petites colonies dispersées, mais se rassemblant l'hiver en quelques immenses colonies, permettant ainsi un brassage génétique).

D'un point de vue biologique, le concept de population intègre une dimension fonctionnelle : celle d'une plus forte probabilité pour un individu de pouvoir se reproduire avec son voisin qu'avec un individu de la même espèce vivant à plus grande distance. Cette notion de distance est à considérer au sens figuré, car elle n'est pas forcément d'ordre géographique. Ainsi par exemple, deux poissons vivant dans deux rivières parallèles distantes de 100 m auront une plus faible probabilité de se rencontrer et d'échanger leurs gènes que deux poissons vivant dans la même rivière mais distants de 500 m. La notion d'obstacle à la dispersion des individus est donc importante, car plus un groupe d'individus est isolé et déconnecté d'autres groupes, plus les échanges génétiques seront faibles et plus cet isolat génétique risque de développer des caractéristiques qui lui sont propres, donc présenter un fort enjeu local de conservation (ex : Vipère d'Orsini au sommet du Mont Ventoux, noyau de population isolé dont des analyses génétiques ont montré l'originalité). C'est ce phénomène qui, à plus grande échelle et sur un grand pas de temps, est à l'origine de la formation des espèces. On parle alors de **spéciation**. La notion de distance dépend également d'un deuxième paramètre important : la mobilité des individus de l'espèce. Plus l'espèce est mobile, plus les individus pourront facilement se rencontrer et échanger leurs gènes (ex : oiseaux, chiroptères, grands mammifères). Dans ce cas, la notion de population risque de n'être biologiquement pertinente qu'à un vaste échelon géographique (régional, national, voire continental), bien qu'elle soit généralement utilisée par défaut à un moindre échelon pour désigner les individus présents au sein d'un site, un site Natura 2000 par exemple. Dans ce cas, la notion de population présente plus un caractère descriptif que fonctionnel. *A contrario*, pour une espèce à faible mobilité (ex : certains insectes, plantes, mollusques, micromammifères...), le moindre obstacle à la dispersion des individus (ex : barrage sur une rivière, relief, destruction d'habitat servant de corridor biologique...), peut constituer une barrière difficilement franchissable et favoriser la création d'isolats génétiques, donc de micropopulations. C'est pourquoi certains secteurs géographiques enclavés sont parfois très riches en espèces ou sous-espèces endémiques locales (ex : insectes endémiques de certaines vallées des Alpes-Maritimes, Criquet de Crau, etc.). Ce même phénomène peut parfois s'observer pour des espèces inféodées à des milieux relictuels, dont la capacité de dispersion est importante mais néanmoins insuffisante pour coloniser de nouveaux secteurs, faute de milieux propices dans leur rayon d'action (ex : plantes relictuelles glaciaires dans certaines tourbières de montagne).

On voit donc que la notion de population, très utilisée par les biologistes, peut recouvrir différentes réalités selon l'espèce et l'échelle géographique considérées.

La dynamique et la génétique des populations sont deux « branches » de la biologie des populations qui visent à simplifier l'appréhension de cette complexité, par la recherche de

patrons et de lois fonctionnelles pouvant s'appliquer à l'ensemble des espèces et de leurs populations.

La **dynamique des populations** cherche à décrire et comprendre les fluctuations d'effectifs des individus d'une population (augmentation, régression, stabilité). Chaque population peut être caractérisée par diverses variables démographiques populationnelles, ou « traits d'histoire de vie ». Par exemple : la fécondité, l'espérance de vie, les taux de natalité (ou recrutement) et de mortalité, le sex-ratio, l'âge de maturité sexuelle, etc. Ces différents paramètres permettent de modéliser l'évolution d'une population. Deux grands types de modèles populationnels peuvent être distingués :

- **population fermée** : on considère que la population étudiée est isolée de son environnement extérieur, donc que 100% des individus sont présents dans un espace bien délimité et qu'ils ne peuvent pas en sortir. Il n'y a donc ni migration (individus sortants), ni immigration (individus entrants). C'est le cas le plus simple et le plus facile à modéliser. Exemples : une population de poissons dans un lac, de cervidés dans un massif forestier clos, d'insectes sur une île très isolée, etc.
- **population ouverte** : on considère que des échanges d'individus (migration et immigration) peuvent avoir lieu avec d'autres populations, formant ainsi une métapopulation. Ce type de situation est plus difficile à modéliser car implique de connaître plus de variables (notamment l'importance de l'immigration et émigration qui sont difficiles à obtenir). C'est pourtant le cas correspondant à la majeure partie des populations naturelles, notamment des populations migratrices. Exemples : oiseaux et chiroptères en général, papillons, libellules, amphibiens, etc.

La **génétique des populations** cherche quant à elle à décrire et comprendre la diffusion et la variabilité des gènes au sein d'une population. Ainsi par exemple, lorsqu'un individu naît avec une mutation lui conférant un avantage sélectif (ex : couleur de peau plus foncée lui permettant de mieux se fondre dans son milieu, donc d'être moins repérable par les prédateurs), on dit qu'il dispose d'une meilleure fitness. Sa probabilité de se reproduire et de transmettre ce gène, ou plus exactement cet allèle (= variante d'un gène, par exemple variante « marron » du gène « couleur des yeux ») sera forte, donc cet allèle va se diffuser au sein de la population. Appliqué à l'ensemble des individus constituant une population, donc partageant un génome collectif, un même gène peut présenter une multitude de variantes alléliques. On parle alors de gène polymorphe. La génétique des populations cherche à mesurer les fréquences de ces allèles, influencées par les « pressions évolutives » que sont la sélection naturelle, la dérive génétique, les taux de mutation et de migration. Les fluctuations de fréquence des allèles sont la première étape de l'évolution. En effet, la fixation de certains allèles peut conduire à des adaptations. Par la suite, l'accumulation d'adaptations différentes au sein de populations distinctes peut conduire au processus de spéciation.

Ainsi, plus une population se trouve isolée et/ou en limite de son aire de répartition naturelle, plus elle présente une fragilité, tant démographique que génétique, mais plus également elle est soumise à de fortes pressions sélectives, à même de faire émerger une variabilité jusque là inédite, par phénomène de dérive génétique. Ce phénomène a bien été étudié sur certaines îles où, à partir d'une population de départ très réduite (parfois limitée à un seul couple reproducteur !), des espèces introduites (chat haret, rat noir...) ont pu recréer une population conséquente (plusieurs centaines voire milliers d'individus) et disposant de caractères génétiques inédits, en l'espace d'une cinquantaine d'années seulement, alors qu'une telle évolution prendrait normalement plusieurs centaines voire milliers d'années.

En biologie de la conservation, ces isolats génétiques bénéficient d'un intérêt particulier car constituent un enjeu local de conservation très fort. C'est également le cas dans les principes fondateurs de la politique Natura 2000, bien que, malheureusement, la liste des espèces inscrites à l'annexe 2 de la directive Habitats comprend peu d'espèces endémiques ou sub-endémiques, pour la région PACA tout du moins. Toutefois, en préservant les habitats et leur fonctionnalité, l'outil Natura 2000 prend en compte l'ensemble de la biodiversité qu'ils hébergent, donc participe efficacement à la conservation d'une multitude d'espèces, certaines parfois encore inconnues de la science.

2.2. Concepts de biostatistique

Pour comprendre le fonctionnement souvent complexe des systèmes naturels, l'écologie s'est appuyée depuis ses origines sur les mathématiques, et plus particulièrement sur le calcul de probabilités. Les biomathématiques, ou biostatistiques, tendent à une compréhension du monde réel. Cette compréhension passe par la mise en place d'un modèle, prenant en compte un certain nombre de paramètres considérés comme causes d'un phénomène. Ce modèle constitue un objet mathématique, dont l'étude permet une meilleure compréhension du phénomène étudié, éventuellement une prédiction qualitative ou quantitative quant à son évolution future.

Aussi complexe qu'il soit, un modèle doit toujours être considéré comme une représentation simplifiée et partielle du monde réel. Une des précautions d'usage de tout scientifique est donc de garder un œil critique sur les résultats fournis par un modèle et d'être conscient de ses limites méthodologiques, afin de ne pas tirer de fausses conclusions. Pour ce faire, le scientifique se doit de toujours discuter de la pertinence des hypothèses simplificatrices de son modèle. Toutefois, en biologie des populations, la représentation simplifiée d'une réalité forcément complexe permet, dans la plupart des cas, d'obtenir des résultats relativement robustes et utiles aux gestionnaires d'espaces naturels.

Les concepts de base sont les suivants :

Individu : un individu statistique est une entité élémentaire sur laquelle on peut appliquer un tirage au sort. Un individu statistique peut correspondre à un individu biologique (un oiseau, une plante) mais pas nécessairement. Ce peut être également un emplacement géographique, par exemple un quadrat ou un point d'écoute, lorsqu'on effectue un tirage aléatoire des placettes/points qui seront échantillonnés dans le cadre d'un protocole.

Population : une population statistique représente un ensemble d'individus statistiques. Tout l'objet de la biostatistique est de prédire l'état d'une population globale inconnue à partir de mesures faites sur une partie des individus, constituant une population partielle connue. Si l'on considère que les individus étudiés sont représentatifs (voir ci-dessous les notions d'échantillon et de tirage aléatoire), on pourra faire l'hypothèse que les paramètres mesurés sur la population connue correspondent à ceux de la population globale inconnue, avec un certain intervalle de confiance.

Echantillon : l'échantillon désigne un fragment d'un ensemble prélevé pour juger de cet ensemble. Dans la plupart des cas, il s'agit d'une collection d'éléments, d'individus statistiques, prélevés de façon particulière de la population statistique afin de tirer des conclusions sur cette dernière. Cependant, il arrive parfois que tous les éléments de la population soient étudiés. On parle alors d'échantillonnage exhaustif ou de recensement. L'effectif de l'échantillon peut varier de l'unité à l'effectif N de la population statistique. L'effort d'échantillonnage sera mesuré par la taille n de l'échantillon et par la fraction d'échantillonnage, égale à n/N .

Tirage aléatoire : pour que des résultats puissent être généralisés à la population statistique, l'échantillon doit être représentatif de cette dernière, c'est-à-dire qu'il doit refléter fidèlement sa composition et sa complexité. Seul l'échantillonnage aléatoire assure la représentativité de l'échantillon mais on verra plus loin que la notion de stratification conduit à relativiser l'équivalence entre aléatoire et représentativité. Contrairement à une idée reçue, **la taille de l'échantillon n'affecte pas la représentativité, car elle ne fait que varier l'intervalle dans lequel la vraie valeur d'un paramètre de la population a de fortes chances de se trouver** (= intervalle de confiance). En outre, contrairement à une pratique assez répandue en écologie, **la sélection comme stations d'échantillonnage de sites jugés subjectivement représentatifs du milieu étudié, ne constitue pas un échantillon aléatoire et peut fortement biaiser les résultats** comme il le sera clairement montré avec des exemples précis dans la suite du document.

Variable : une variable est une caractéristique mesurée ou observée sur chacun des éléments de l'échantillon, ou sur des entités prédéfinies qui se rattachent aux unités d'échantillonnage (ex : milieu environnant). Plusieurs catégories de variables peuvent être définies, dont trois sont essentielles :

- les variables quantitatives, mesurables par une unité de mesure sur une échelle ordinale (c'est-à-dire ordonnée intrinsèquement. Ex : nombre d'œufs par nid, distance en mètres, poids en kilogrammes, etc.),
- les variables semi-quantitatives ou variables de rang, lorsqu'un découpage en plusieurs classes permet une mesure selon une échelle ordinale (ex : stades larvaires classés du premier au dernier, classes d'âges, etc.),
- les variables qualitatives, lorsque les données recueillies sur une échelle nominale (et non ordinale) sont susceptibles de classement dans des catégories (qualitatives) collectivement exhaustives et mutuellement exclusives (ex : sexe, couleur du pelage, etc.).

Moyenne : elle exprime la valeur qu'aurait chacun des individus de la population (ou de l'échantillon) s'ils étaient tous identiques et sans changer la dimension globale de la population.

Variance : elle permet de caractériser la dispersion des valeurs par rapport à la moyenne. On peut interpréter la variance comme la moyenne des carrés des écarts à la moyenne (rigoureusement: l'espérance des carrés des écarts à l'espérance; vulgairement : moyenne des carrés moins le carré des moyennes). Ainsi, entre deux populations présentant la même espérance, celle ayant une plus grande variance apparaîtra comme plus dispersée (nuage de points plus étalé). Le fait que l'on prenne le carré de ces écarts à la moyenne évite que des écarts positifs et négatifs ne s'annulent.

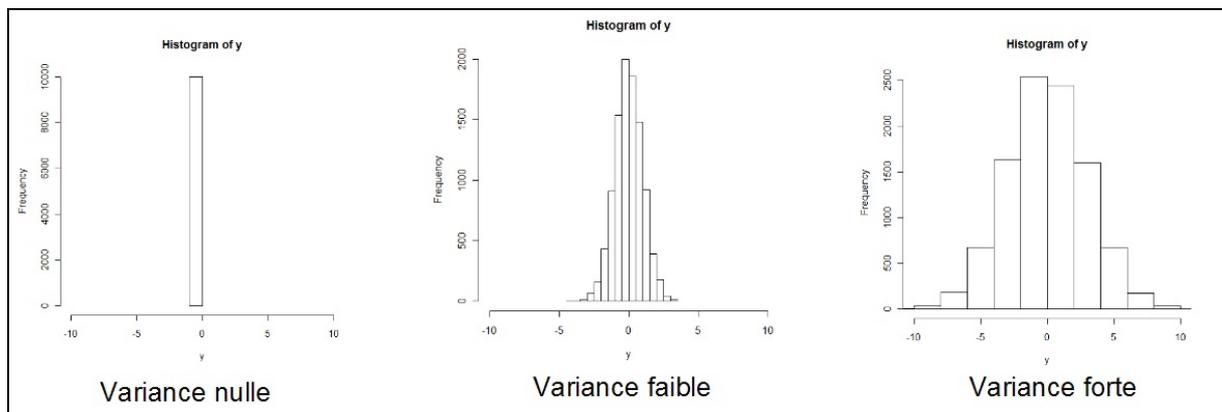


Figure 3 : Illustration de la variance, dispersion des données autour de la moyenne

Ecart-type : il s'agit là encore d'une mesure de la dispersion d'une série de valeurs autour de leur moyenne. Il est égal à la racine carrée de la variance. L'écart-type est une mesure capitale pour comparer plusieurs moyennes. Ainsi, deux échantillons (ex : noyaux de population alpin et méditerranéen d'une même espèce) affichant chacun une moyenne différente mais avec un large écart-type (forte dispersion des données), ne pourront pas être qualifiés comme statistiquement différents. L'écart-type est de même unité que la variable étudiée.

Coefficient de variation : il s'agit là encore d'une mesure de la dispersion d'une série de valeurs autour de leur moyenne. Il est égal à l'écart-type divisé par la moyenne. Il a pour caractéristique d'être indépendant de l'unité de la variable et donc de faciliter les comparaisons entre des variables d'ordres de grandeurs très différents par exemple.

Intervalle de confiance : on parle d'intervalle de confiance lorsque l'on donne un intervalle qui contient, avec un certain degré de confiance, la valeur à estimer. Le degré de confiance est en principe exprimé sous la forme d'une probabilité. Par exemple, on dit souvent qu'un intervalle

de confiance à 95% (ou au seuil de risque de 5%) a une probabilité égale à 0,95 de contenir la valeur du paramètre que l'on cherche à estimer. En pure rigueur statistique, cette conclusion n'est pas vraie mais ce détail dépasse le cadre de ce document. Ainsi, lorsqu'on effectue un sondage (tirage au hasard d'un sous-ensemble d'une population), l'estimation d'une quantité d'intérêt donnée est soumise au hasard et correspond rarement exactement à la valeur réelle de la quantité que l'on cherche à estimer. En présentant pour l'estimation non pas une valeur mais un encadrement, on quantifie l'incertitude sur la valeur estimée. Plus l'intervalle de confiance est de taille petite, moins il y a d'incertitude sur la valeur estimée. Si la variable étudiée suit une loi normale, l'intervalle de confiance à 95 % est égal à la moyenne plus ou moins [1,96 x écart-type].

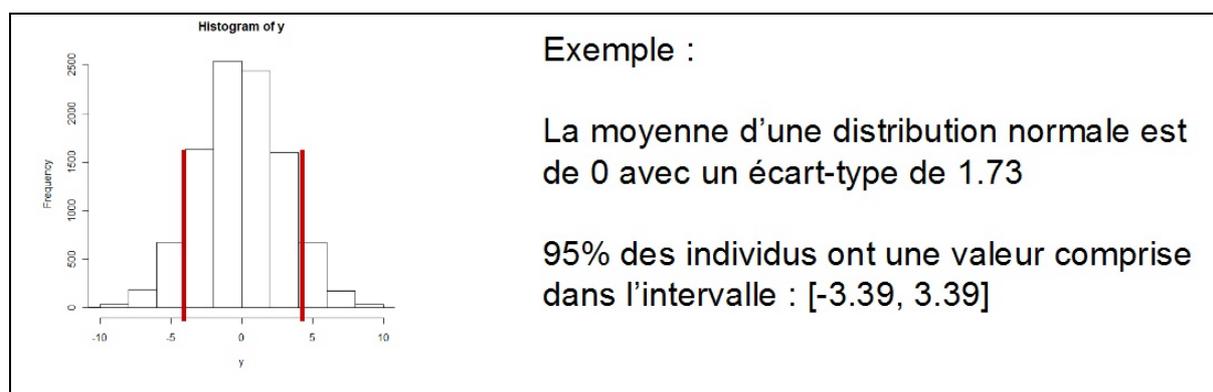


Figure 4 : Exemple d'intervalle de confiance.

Loi normale : la loi normale est l'une des distributions de probabilité les plus connues et intuitives. Elle a été introduite par le mathématicien Abraham de Moivre en 1733, qui l'a utilisée afin d'approcher des probabilités associées à des variables aléatoires binomiales possédant un paramètre n (= effectif) très grand. Cette loi a été mise en évidence par Gauss au XIXe siècle et permet de modéliser de nombreuses études biométriques. Sa densité de probabilité dessine une courbe dite courbe en cloche ou courbe de Gauss, centrée sur sa moyenne et symétrique de part et d'autre de cette moyenne. Elle est généralement associée à des variables continues.

Loi de Poisson : cette loi, introduite au début du XIXe siècle par le magistrat français Siméon-Denis Poisson, s'applique aux phénomènes accidentels ou très rares, où la probabilité p d'occurrence de l'événement est très faible. En écologie, cette loi est souvent utilisée pour compter des événements ou des individus qui sont distribués aléatoirement dans l'espace et dans le temps. Par exemple, pour dénombrer des individus dans des quadrats lorsque la répartition des organismes est aléatoire. Dans la pratique, la loi de Poisson est utilisée lorsque la probabilité est très faible ($p < 0,05$) et le nombre de sondages très élevé ($n > 50$), car pour voir apparaître les événements rares il faut multiplier les épreuves. Exemple : probabilité de capture d'oiseaux rares lors de sessions de capture au filet japonais.

3. Limites et biais des méthodes usuelles

De manière assez régulière deux grandes classes de méthodes sont utilisées par les naturalistes et les gestionnaires ces trente dernières années : les inventaires et les indices d'abondance ponctuels ou kilométriques. Ces deux méthodes sont bien connues. Elles présentent l'avantage d'être simples à mettre en œuvre sur le terrain avec un minimum de technicité et surtout d'être simples à utiliser en termes statistiques (calculs simples du type « moyenne » et « écart-type » réalisables à l'aide d'Excel par exemple). Cependant ces deux grandes méthodes ne sont pas exemptes de limites bien au contraire, limites qui ont souvent été pointées depuis plusieurs années et qui devraient conduire à limiter leur utilisation à quelques situations bien précises ou tout du moins à interpréter leurs résultats avec précaution.

3.1. Rappel sur les inventaires

Les inventaires sont utilisés depuis l'antiquité pour décrire les faunes et flores présentes sur des sites. Ils consistent à visiter une ou plusieurs fois un ou plusieurs sites choisis et de dresser la liste de toutes les espèces rencontrées. Parfois (et même souvent) ils sont limités à un groupe biologique donné souvent pour des contraintes de temps (flore ou faune, oiseaux ou reptiles/amphibiens...). En pratique les inventaires sont donc une opération relativement peu coûteuse et simple à mettre en œuvre si tant est que l'on dispose des compétences naturalistes pour reconnaître les espèces.

La limite des inventaires se trouve dans ce que l'on veut leur faire dire. Nous disposons à l'issu d'un inventaire sur un site d'une liste d'espèces. Malheureusement il est très difficile, voire impossible, d'atteindre l'exhaustivité dans cette liste (surtout lorsque l'on travaille sur un grand nombre de taxons en même temps) à moins d'un effort considérable et encore. Par ailleurs, outre le fait que l'exhaustivité ne soit que très rarement atteinte, le pourcentage d'espèces manquantes n'est pas quantifiable. En effet, avoir une liste de 10 espèces vues sur un site ne permet pas d'inférer sur le nombre d'espèces ratées même en connaissant l'effort de prospection réalisé.

Ce premier résultat pointe le problème majeur de tous les protocoles de suivis d'espèces : la détection. En effet la difficulté majeure rencontrée lorsque l'on travaille sur les plantes ou animaux sur le terrain est que les individus ou les espèces ne sont pas tous détectables avec la même facilité et ne sont donc pas nécessairement toutes détectés. Un grand nombre de facteurs vont influencer cette détection des espèces, par exemple : leur biologie et écologie en premier lieu (rythme d'activité saisonnier (=phénologie) ou journalier, localisation des zones plus ou moins denses en végétation, etc.), mais il existe aussi un effet observateur potentiellement très fort (expérience relative, *a priori* sur les espèces et familiarité plus ou moins forte avec certaines, fatigue, temps de prospection réalisé, etc.).

Quelles sont les conséquences de ces problèmes de détection pour l'utilisation des inventaires ? En conservation les inventaires vont le plus souvent être mis en œuvre pour comparer soit des sites entre eux à un instant donné (étude synchronique), soit pour décrire l'évolution d'un site années après années (étude diachronique). On comprend bien dès lors que le fait de ne pas pouvoir quantifier le pourcentage d'espèces présentes mais ratées dans les jeux de données collectées ne permet pas de répondre à cette question. Nous pourrions très bien avoir détecté 20 espèces une année donnée et 25 espèces quelques années après sur le même site alors même que le site aurait globalement perdu des espèces, il aurait suffi pour cela que les observateurs passent à une période un peu plus favorables à quelques espèces, ou qu'ils aient acquis plus d'expérience ou même que le temps de prospection ait changé. On pointe ici la première limite des inventaires.

La deuxième limite des inventaires vient de leur nature non quantitative en ce qui concerne chaque espèce. Nous avons bien une information sur le nombre d'espèces mais aucune sur le

nombre d'individus de chaque espèce. On comprend aisément que ce type de donnée peut malheureusement complètement masquer des modifications majeures des populations d'espèces présentes. La population d'une espèce présente et détectée lors de deux sessions à dix ans d'intervalle sur un site pourrait s'être effondrée que notre inventaire inclurait toujours cette espèce.

Quelle pertinence pour les inventaires alors ? Les inventaires devraient être mis en œuvre avec des experts performants, sur des groupes biologiques assez restreints, avec un effort quantifié et conséquent, et surtout être interprétés avec précaution. Il semble que les inventaires soient plus intéressants pour mettre en évidence la présence d'espèces patrimoniales que pour comparer des sites dans l'espace ou dans le temps. Ils peuvent être aussi pertinents dans le cadre d'une forme de pré-étude d'une zone de manière à obtenir une vision globale même imprécise d'un site afin de définir par la suite les priorités en termes d'études fines.

3.2. Rappel sur les Indices d'abondance

La deuxième grande classe de méthodes couramment utilisées dans le milieu naturaliste sont les indices d'abondance ponctuels ou kilométriques. Ils consistent à suivre une ou plusieurs espèces quantitativement avec des protocoles standardisés. Dans le cas des indices d'abondance ponctuels, très largement utilisés sur les passereaux, l'observateur se place sur un site à un point fixe (point d'écoute) et note le nombre d'individus vus ou entendus d'une ou plusieurs espèces durant une période de temps fixée par avance (10, 15 ou 20 minutes le plus souvent). Les indices kilométriques consistent eux à définir un parcours de longueur fixe et de noter tous les indices de présence ou les individus vus ou entendus sur ce trajet (crottes, intersection avec des traces, etc).

On voit rapidement que ces méthodes lèvent une des limites des inventaires : elles fournissent des résultats quantitatifs sur les espèces (nombre d'individus vus). Popularisées par J. Blondel dans les années soixante-dix, elles sont aujourd'hui l'outil le plus utilisé par les bureaux d'études et les gestionnaires du fait de leur simplicité de mise en œuvre et pour leur facilité de manipulation en terme statistiques (comparaisons de moyenne très facilement réalisées sous Excel par exemple).

Cependant ces méthodes ne sont pas exemptes d'hypothèses sous-jacentes rarement énoncées, et encore plus rarement détaillées de manière critique. Si ces hypothèses ne sont pas respectées, les résultats peuvent être largement biaisés et leur interprétation limitée voire erronée. Tout comme pour les inventaires, le point crucial de ces méthodes repose sur les problèmes de détection.

La grande force des indices d'abondance est de partir du postulat que l'on ne détecte jamais 100% des individus présents. On détecte une fraction inconnue de la population présente, donc on ne peut pas s'intéresser à l'abondance d'une espèce sur un site donné. Mais les indices d'abondance vont être utilisés pour comparer des sites entre eux ou suivre des sites dans le temps. Cette comparaison implique une hypothèse : la détection est constante dans le temps et l'espace. En effet si nous voulons comparer deux indices d'abondance entre eux, il faut poser l'hypothèse que la fraction de la population détectée dans les deux cas était la même (par exemple on détecte 50% des individus dans le deux cas, donc si on a un site à 50 individus vus et un à 100 individus vus, alors le deuxième site héberge une population deux fois plus grande que le premier site). Si cette hypothèse n'est pas respectée les indices d'abondance ne sont pas pertinents (si sur un site on détecte 25% des individus et sur un autre 50% des individus, alors il est logique d'obtenir 50 et 100 individus alors même que les populations sont de même taille).

Pourtant comme nous l'avons vu pour les inventaires, la détection est un facteur extrêmement variable (météo, activité des animaux, expérience observateur, fatigue, etc) et il faut donc discuter de ces facteurs. D'une manière générale les naturalistes connaissent bien ces biais et

standardisent au maximum leurs prospections (période de l'année, horaires fixes, toujours le même observateur, etc.) ce qui effectivement a pour conséquence de souvent de maximiser le taux de détection et surtout de limiter sa variabilité entre sites ou dans le temps.

Malgré toutes ces précautions il reste deux grandes classes de situations très difficiles voire impossibles à contrôler, qui peuvent fortement biaiser les résultats des indices d'abondance. Le premier concerne le changement dans la détection, le deuxième des changements dans l'activité des individus. Ces biais sont détaillés dans les deux encarts des pages suivantes. L'un montre qu'une augmentation de l'indice peut résulter d'une augmentation de l'expérience de l'observateur ou de sa plus grande familiarité avec le site suivi. L'autre montre que l'activité des individus peut être fortement modulée avec la densité locale. Dans ces deux cas, l'hypothèse comme quoi la détection est constante entre les deux sites suivis n'est pas vérifiée, les résultats des comparaisons d'indices sont donc à prendre avec beaucoup de précaution.

Notons pour finir qu'il est impossible à partir des indices d'abondance que la détection soit constante. L'intuition du naturaliste aura valeur d'argument. Mais ces arguments devront être clairement exposés avec les résultats pour que le lecteur puisse se faire sa propre opinion. Notons enfin que si l'intuition du naturaliste a valeur d'argument nous sommes, malgré des données chiffrées, dans une situation de « dire d'expert », situation que les indices d'abondance avaient pour mission d'éviter.

3.3. La notion de détectabilité

Nous venons donc de voir que le problème majeur posé à la fois par les inventaires et les indices d'abondance, quoique partiellement résolu avec cette deuxième méthode, est celui de la détection des individus. Ce problème est, avec les difficultés d'échantillonnage que nous verrons par la suite, la grande difficulté de tout suivi de population animale et végétale. Ce problème est clairement identifié depuis longtemps par les méthodologistes qui travaillent à améliorer les méthodes pour obtenir des estimations fiables des abondances ou des tailles de population. Ces méthodologistes ont fourni un travail conséquent de réflexion, de développement de nouvelles méthodes et de test de ces méthodes sur le terrain au cours de trente dernières années. Aujourd'hui il existe tout un panel de méthodes dont le but est d'estimer la détection des individus et donc de pouvoir corriger le nombre d'individus simplement observés pour en déduire le nombre d'individus réellement présents. Ces méthodes posent deux difficultés pour les naturalistes et les gestionnaires : elles sont plus complexes mathématiquement et par voie de conséquence demandent plus de technicité lors des analyses de données (logiciels souvent spécifiques et majoritairement en anglais) ; elles demandent généralement un effort plus grand sur le terrain donc nécessitent de disposer de budgets plus importants (mais pas toujours). Dans la suite de ce document nous présenteront de manière opérationnelle ces nouvelles méthodes en insistant notamment sur les hypothèses statistiques sur lesquelles elles reposent et dont le respect impose des précautions particulières dans la mise en place de protocoles.

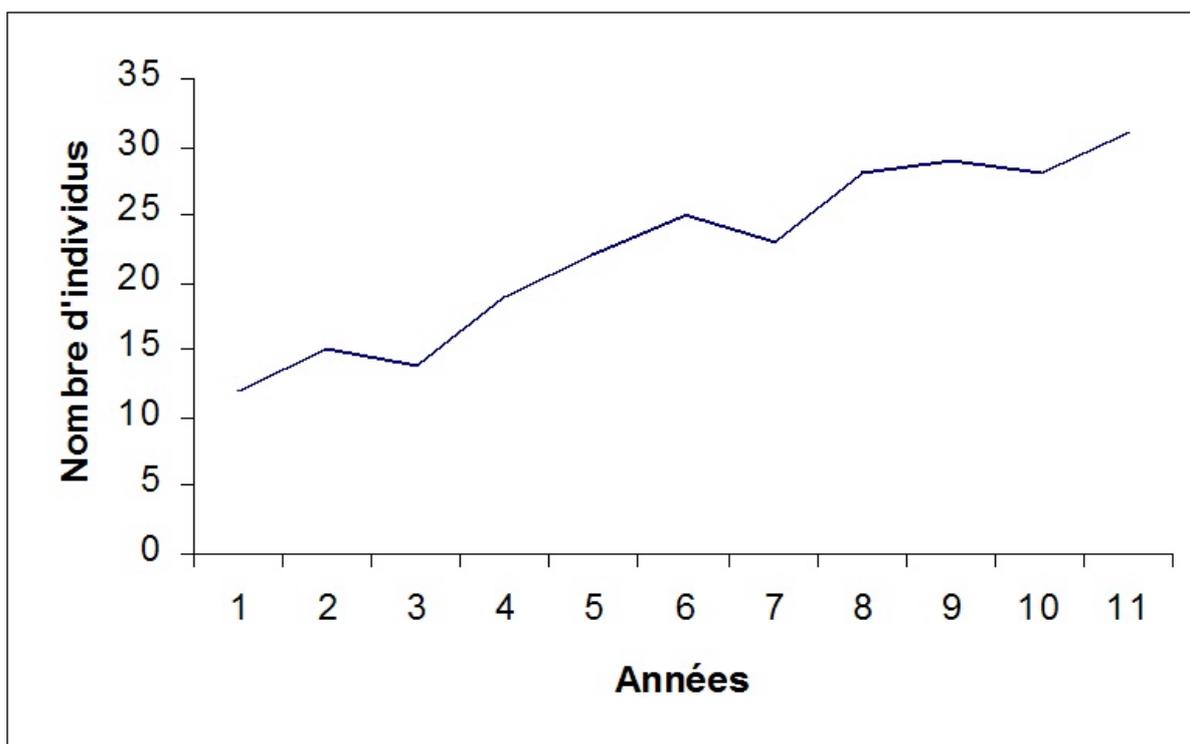


Figure 5 : Augmentation de l'effectif observé au cours du temps.

La figure présentée ci-dessus fait clairement apparaître une augmentation du nombre d'individus observés. Une conclusion naïve serait d'en conclure une augmentation de la population de l'espèce. Pourtant on peut aussi attribuer cet accroissement non pas à celui de la population mais à une amélioration de la détection des individus. On peut envisager au moins trois situations simples pour lesquelles cette courbe serait obtenue malgré une stabilité de la population :

- l'effort de terrain a augmenté au cours du temps, par exemple par le recrutement d'observateurs supplémentaires ou par un temps de terrain plus important,
- la qualité de la prospection s'est améliorée, par exemple du fait que les observateurs connaissent mieux l'espèce et donc sont plus efficaces,
- la détection de l'espèce est de plus en plus facile, par exemple si la végétation change et rend la détection plus aisée au cours du temps.

Ces trois explications reposent toutes sur le principe que la détection des individus présent n'est pas de 100%, impliquant que les comptages sont un mélange permanent entre effectifs réels et détection effective.

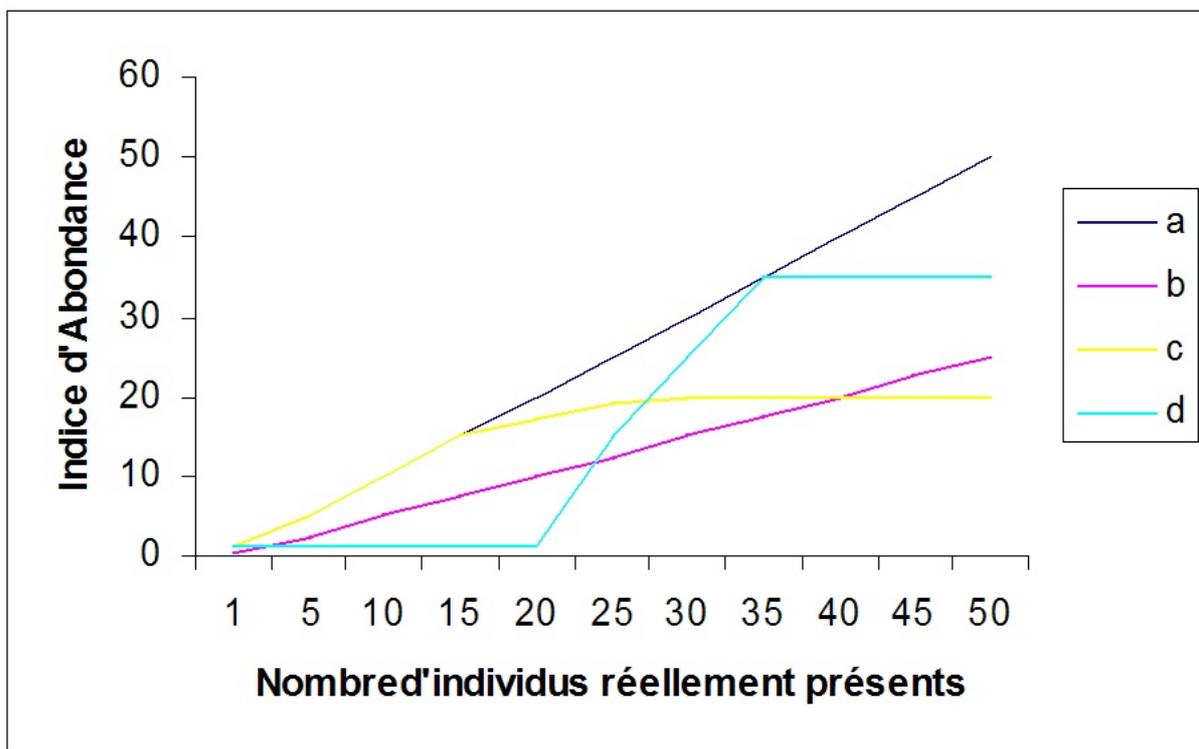


Figure 6 : Variation de la mesure de l'effectif en fonction de différents scénarii de détectabilité.

- a) détection de 100%
- b) détection constante mais inférieure à 100%
- c) détection avec une valeur plateau
- d) détection avec un seuil de déclenchement et une valeur plateau

La figure ci-dessus présente quatre courbes que l'on obtient fréquemment lorsque que l'on s'attache à comparer des indices d'abondance relevés sur le terrain et les abondances réelles en individus.

- La courbe a est la plus simple à interpréter : elle montre un accroissement strictement linéaire entre l'indice d'abondance et l'abondance réelle ; de plus les deux mesures sont identiques. Dans cette situation notre indice d'abondance est tout simplement l'abondance réelle. Nous sommes donc dans la situation d'un comptage exhaustif des individus présents. Cette situation comme nous l'avons vu est relativement rare.
- La courbe b montre un accroissement linéaire de l'indice d'abondance avec l'abondance réelle, mais l'indice augmente moins vite que l'abondance réelle. Nous sommes ici dans une situation relativement optimale pour laquelle la détection n'est pas de 100% mais est constante, quelque soit le nombre d'individus réellement présents. Ici on voit que l'indice augmente deux fois moins vite que la densité réelle : on peut donc estimer la détection à 50% (on ne voit qu'un individu sur 2 présents, donc 1 chance sur 2 d'observer un individu présent).

... / ...

- La courbe c présente une situation beaucoup moins optimale et pourtant très fréquemment observée sur le terrain. On constate ici que l'indice augmente comme le nombre d'individus réellement présents lorsque ce nombre est petit, puis stagne alors que l'abondance réelle continue d'augmenter. On peut observer ce genre de courbe lorsqu'on travaille par exemple sur les traces de grands mammifères. Lorsque le nombre d'individus présents devient grand, l'utilisation de l'espace par ces individus est modifiée et l'indice ne mesure plus la même chose que lorsque les individus sont peu nombreux. On peut aussi observer ces courbes dans le cas de recensement de mâles chanteurs d'oiseaux, soit parce que la capacité de l'observateur à « compter » est limitée par le mélange de nombreux chants, soit parce que les oiseaux eux-mêmes régulent leur chant lorsqu'ils sont trop nombreux pour être clairement perçus. On voit dans cette situation que l'indice d'abondance sera un très mauvais indicateur de l'abondance réelle.
- Enfin la dernière situation, encore plus complexe, peut être moins courante mais néanmoins possible, est celle présentée par la courbe d. L'indice n'augmente pas tant que le nombre d'individus est faible puis augmente très fortement à partir d'un certain seuil puis se stabilise alors que l'abondance réelle continue d'augmenter. Cette situation peut se rencontrer là encore avec les comptages de mâles chanteurs, lorsque les individus sont très peu nombreux : il n'y a que peu de compétition et donc peu de stimulation sociale conduisant les mâles à chanter, mais à partir d'un certain seuil de densité les mâles entrent effectivement en compétition et cette situation crée la stimulation, conduisant tous les mâles à chanter. Enfin l'indice d'abondance atteint son maximum comme vu sur la courbe c.

Pour conclure cette première partie il convient de rappeler que toutes les méthodes statistiques ont leurs limites. Il est important de noter que toutes les méthodes dont nous parlerons par la suite sont éprouvées statistiquement, toutes « marchent » d'un point de vue mathématique. Il n'y a pas de mauvaise méthode. Il n'y a qu'une mauvaise adéquation entre les caractéristiques du terrain ou de l'espèce étudiée et la méthode statistique choisie. Comme nous l'avons déjà vu pour les inventaires et les indices d'abondance, les méthodes statistiques posent un certain nombre d'hypothèses. Pour qu'elles soient utilisables dans un contexte particulier, il faut soit vérifier que ces hypothèses sont vérifiées, soit admettre qu'elles sont vérifiées. La rigueur scientifique d'un suivi dépendra de cette étape. La science, l'accroissement des connaissances se constitue par étapes successives. Nul ne peut reprocher d'avoir fait un inventaire si le budget disponible pour réaliser une étude à l'aide d'une méthode *a priori* plus pertinente n'était pas disponible. Cependant il faudra s'efforcer systématiquement de pointer les limites de l'étude menée, s'efforcer d'identifier les biais potentiels et de proposer des solutions permettant de s'affranchir de ces biais. Cette démarche aura le double avantage d'informer les lecteurs sur ces biais et du fait que vous en êtes conscient. Le lecteur pourra se faire sa propre opinion sur les résultats obtenus en conscience de ses limites et ne pourra donc rejeter le travail en bloc. Par ailleurs elle permettra éventuellement de redéfinir des demandes budgétaires supplémentaires si les financeurs jugent que les résultats actuels, au vu des limites énoncées, ne sont pas suffisants.

4. Echantillonner

4.1. Pourquoi échantillonner

Une des difficultés lorsque l'on désire suivre une population animale réside dans la définition des limites des populations mais ce problème complexe n'a pas de conséquence directe sur le suivi des populations sur des sites Natura 2000 pour lesquels on désire obtenir une estimation de la taille de la population **sur le site** en question. Cependant il est important de bien garder à l'esprit qu'un site Natura 2000 est une division réalisée sur des critères humains qui sont pour la plupart déconnectés de ceux des populations animales ou végétales. Ceci a pour conséquence qu'il pourra parfois être complexe d'interpréter des variations locales d'abondance sur le site Natura 2000 malgré de gros efforts de terrain pour la simple raison que certains phénomènes se produisent en dehors du site et que ce site est perméable. Nous verrons par la suite l'importance des témoins dans la mise en évidence d'effet de mesures de gestion, témoins qui, s'ils sont bien choisis, permettent effectivement de contrer une partie de ces problèmes de perméabilité des sites.

La deuxième grande difficulté lorsqu'on étudie une population au sein même d'un site clairement identifié et défini réside dans la réalisation d'un suivi exhaustif de l'ensemble de la zone d'étude. Pour certaines espèces dont les populations sont de petite taille et pour des sites Natura 2000 de taille limitée, cet objectif d'exhaustivité sera atteignable (par exemple suivi d'aires de reproduction d'Aigles de Bonelli) mais dans une grande majorité de cas il ne le sera pas. Deux stratégies sont alors possibles. La première consiste à réduire la question. Au lieu d'essayer d'estimer la taille de la population sur l'ensemble de la zone on peut choisir de ne l'estimer que sur une toute petite zone, celle qui paraît la plus intéressante. Malheureusement dans cette situation on limite les possibilités sur le long terme pour avoir des résultats intéressants et, dans le cas de populations d'espèces relativement mobiles, on peut obtenir des résultats très peu pertinents. La deuxième grande stratégie consiste à suivre un certain nombre de plus petites zones choisies de manière à ce qu'elles soient représentatives de l'ensemble de la zone d'étude. Les résultats obtenus sur ces petites zones pourront alors être extrapolés à l'ensemble du site dans un deuxième temps. Cette deuxième démarche consiste donc à suivre un échantillon de la zone d'étude complète. La démarche de choisir les petites sous-unités à suivre est une étape cruciale de tout montage de protocole de suivi de population. Cette démarche s'appelle l'échantillonnage. Elle n'est pas spécifiquement limitée à l'écologie mais intervient dans de très nombreux domaines (sondages par exemples). Il s'agit d'un concept issu des statistiques et il a été très étudié et développé par les statisticiens. Nous détaillons ci-dessous les termes essentiels pour bien comprendre ce qu'est un échantillonnage, présentons rapidement les grandes généralités liés à l'échantillonnage puis les grands types de plans d'échantillonnage que l'on utilise généralement pour suivre les populations animales et végétales.

4.2. Notions de bases en échantillonnage

Une des petites sous-unités, petite zone qui sera suivie se nomme un individu statistique. L'ensemble de ces petites sous-unités se nomme l'échantillon. L'ensemble de toutes les petites sous-unités possibles (c'est-à-dire toute la zone d'étude) se nomme la population statistique.

Le point principal à retenir lorsque l'on définit une stratégie d'échantillonnage est que la meilleure manière d'avoir une vision représentative du système étudié est de réaliser un tirage aléatoire des sous-unités suivies. Cela peut paraître contre-intuitif pour des personnes ayant une bonne connaissance de leur zone d'étude, pourtant c'est la seule démarche valable statistiquement pour garantir des résultats non biaisés. Si les sous-unités sont tirées aléatoirement et qu'elles sont en nombre suffisant, **alors l'aléatoire assure la neutralité par rapport à la zone, et le nombre assure la représentativité.**

Un biais très fréquent lorsque l'on étudie des populations et de ne se concentrer que sur les zones où l'espèce est la plus abondante car c'est la zone où nous obtiendrons le plus d'informations. Cette démarche est à proscrire lorsque l'on cherche à estimer la taille d'une population et surtout à la suivre dans le temps. Les risques associés à ce type de protocole sont détaillés dans la partie suivante (§ 3.3. définir la zone d'étude).

Une autre grande généralité sur l'échantillonnage est que plus on augmente le nombre de sous-unités suivies, plus les résultats seront précis. A l'extrême, si on suit toutes les sous-unités possibles, on aura alors un suivi exhaustif de la zone. Cela signifie aussi que si l'on suit peu de sous-unités, elles seront représentatives de la zone puisque tirées aléatoirement mais elles ne représenteront qu'une petite partie des possibles et la confiance en leur extrapolation sera faible, d'où une précision plus faible de nos résultats.

La précision d'une estimation en statistique se mesure par l'intervalle de confiance de cette estimation. Par exemple, on dira qu'il y a en moyenne 10 individus par sous-unité avec un intervalle de confiance compris entre 5 et 15, à 95% de probabilité. Ceci signifie qu'il y a en moyenne 10 individus par sous-unité étudiée mais que la moyenne réelle de toutes les sous-unités possibles se situe avec 95% de chance entre 5 et 15 individus. **Un point crucial à retenir est qu'une moyenne n'a que peu de valeur de résultat dans un rapport. Ce qui importe, c'est l'intervalle de confiance, car c'est lui qui donne la confiance que l'on a dans le résultat énoncé.** Vos conclusions ne seront pas les mêmes si vous avez une moyenne de 50 individus par sous-unité avec un intervalle de confiance compris entre 10 et 90, que s'il est compris entre 48 et 53.

Attention à ne pas mal interpréter les liens entre taille de l'échantillon (nombre de sous-unités suivies), biais et précision. Si la méthode que vous utilisez pour suivre votre population est biaisée, c'est-à-dire qu'elle fournit des résultats erronés, alors l'augmentation de la taille de l'échantillon ne réduit pas ce biais, elle améliore la précision de la valeur mais le biais reste de même ordre. Par exemple si vous sous-estimez de 25% la taille de la population sur chaque sous-unité, alors multiplier le nombre de sous-unités ne résoudra pas le problème.

4.3. Définir la zone d'étude / la notion de favorable

Définir la zone d'étude peut présenter certains problèmes. Le premier est d'ordre écologique : il s'agit de définir les limites de la population. En effet, il existe une très forte variabilité des populations et de leur fonctionnement. Ainsi, comprendre ce qui se passe sur une zone donnée nécessite souvent d'avoir accès aux informations sur ce qui se passe à l'extérieur du site si la population est de taille très importante. Par exemple, dans le cas de suivis d'espèces migratrices, la chute importante d'une population ne pourra pas nécessairement être enrayerée par l'amélioration des sites de reproduction ou des sites de halte migratoire si les causes de la chute sont sur les zones d'hivernage. Dans le cas d'une population à très large répartition, il peut en être de même. L'amélioration d'un site peut ne pas avoir de conséquence sur la dynamique globale de la population. Nous n'aborderons pas ici plus en détail ce problème complexe, car dans le cadre de suivis de sites Natura 2000 (par exemple), la zone d'étude est globalement définie. Cependant, comme nous l'avons vu précédemment, il est souvent impossible de suivre exhaustivement la zone. Il faudra donc au sein même d'un site Natura 2000 déterminer des sous-unités à suivre.

Une première règle d'or est qu'il faut essayer de prospecter le plus largement possible. Ainsi pour s'assurer d'avoir une bonne représentativité de ce qu'il se passe sur le site, il faudra éviter de ne se concentrer que sur une petite partie du site mais au contraire essayer de répartir l'effort sur l'ensemble du site.

Lorsque l'on a une connaissance préalable plus ou moins fine de la répartition des individus sur la zone d'étude, il est fréquent pour le naturaliste de ne choisir de prospecter que les zones à forte densité de manière à maximiser le nombre de données collectées et de compter le

maximum d'individus. Malheureusement, cette démarche est à proscrire lorsque l'on tente d'avoir une estimation de la taille de la population et de son évolution sur l'ensemble du site. En effet, cette démarche pose de nombreux problèmes. Le premier est évidemment d'obtenir une estimation de la taille de la population sur l'ensemble de la zone. Le fait de ne suivre que les zones à forte densité n'informe pas nécessairement sur le nombre de ces zones (outre celles déjà connues) et surtout n'informe pas sur la surface de zones occupées à faibles densités, qui peuvent parfois être très grandes et par conséquent abriter une très forte proportion de la population. Le deuxième biais de cette méthode concerne l'évolution dans le temps des populations et il est crucial de bien le comprendre. Le fait de ne suivre que les zones à forte densité a généralement pour conséquence qu'on ne peut observer que des décroissances ou, au moins, une stabilité des effectifs. En effet, des zones très favorables à l'espèce peuvent soit devenir défavorables (on observe donc une décroissance), soit rester stables de même que leur population. Pourtant dans cette situation, nous ne disposons d'aucune information sur les zones initialement moins favorables. Ces zones peuvent être devenues favorables et donc présenter d'importantes hausses d'effectifs. Elles peuvent aussi s'être encore plus dégradées et donc présenter une décroissance des effectifs, même si les zones favorables, elles, sont restées stables.

Prenons deux exemples.

Exemple 1 : suivi du lagopède alpin dans les Pyrénées.

Le lagopède alpin est étudié depuis une trentaine d'années dans les zones centrales des Pyrénées, qui abritent de bonnes populations. Ces populations sont stables dans l'ensemble. On pourrait donc conclure à un bon état de santé des populations. Pourtant, le message d'une forte régression de cette espèce est largement véhiculé. En fait, ce constat vient d'une vision empirique et est notamment tiré du fait que les populations des zones périphériques des Pyrénées semblent décroître voire disparaître. Malheureusement, le protocole utilisé ne permet pas de s'en assurer. Il ne permet donc pas de bien comprendre le risque auquel est soumise cette espèce.

Exemple 2 : suivi de tétras sur des places de chant.

Cet exemple est issu d'une étude menée aux USA sur une espèce de Tétrras. Le suivi de cette espèce se réalise sur des places de chant connues sur une grande zone géographique. Cette zone a été largement prospectée dans le cadre d'un programme initial visant à décrire sur la situation de l'espèce dans cette région. Les places de chant sont repérées lors de cette première phase. Elles ont par la suite été suivies annuellement avec un protocole très rigoureux (passages multiples pendant la saison, météo calibrée...) et cela pendant une vingtaine d'années. Les auteurs de cette étude montrent avec ces suivis que l'espèce est en décroissance forte et notamment dans les zones d'implantation de puits de forage de gaz. Ce résultat peut paraître cohérent, pourtant il est fortement critiquable. En effet, seuls les sites connus à une date donnée sont suivis. Cette espèce est connue pour être fidèle à ses places de chant mais ceci est connu lorsque l'habitat est relativement stable. Lorsque l'habitat se dégrade, il est probable que cette espèce puisse abandonner des places de chant devenues « défavorables » et s'installer ailleurs (à quelques centaines de mètres par exemple, voire beaucoup plus loin si nécessaire). Par conséquent, il est abusif de dire que l'espèce a régressé. La conclusion rigoureuse est que l'espèce a abandonné une partie de ces places de chant et notamment celles localisées dans les zones de forage. Cependant, le protocole ne s'intéressant pas du tout à l'émergence possible de nouvelles places de chant au cours de la période d'étude, nous ne pouvons pas rejeter l'hypothèse selon laquelle les oiseaux se sont simplement déplacés et que les effectifs sont stables (ils pourraient même être en augmentation). La conséquence fréquente d'un suivi sur les meilleurs sites est une vision pessimiste de la situation car ces sites ont tendance à se dégrader.

Ces deux exemples illustrent bien le risque très important qu'il existe à ne suivre que les zones les plus favorables. Il faut de fait suivre l'ensemble des zones potentiellement occupées, même par quelques individus, et **surtout anticiper sur les changements potentiels futurs de manière à ne pas exclure de l'échantillonnage des sites peu favorables aujourd'hui mais qui pourraient dans quelques années devenir favorables**, pour une raison quelconque (abandon des pratiques agricoles, remise en état des sites...). Au mieux, dans une stratégie d'échantillonnage, on pourra exclure les zones pour lesquelles l'absence est certifiée (quitte à la vérifier sur un faible nombre de sous-unités) et pour lesquelles la stabilité dans le temps est assurée (ex : zones urbaines, zones d'altitude...).

4.4. Les grands types de plan d'échantillonnage

4.4.1. Echantillonnage classique

Un plan d'échantillonnage peut être très simple. Il peut consister à sélectionner aléatoirement un certain nombre de sous-unités au sein de la zone d'étude, petites sous-unités qui seront suivies par la suite. Il peut s'agir de transects, de points d'écoutes, de quadrats, etc.

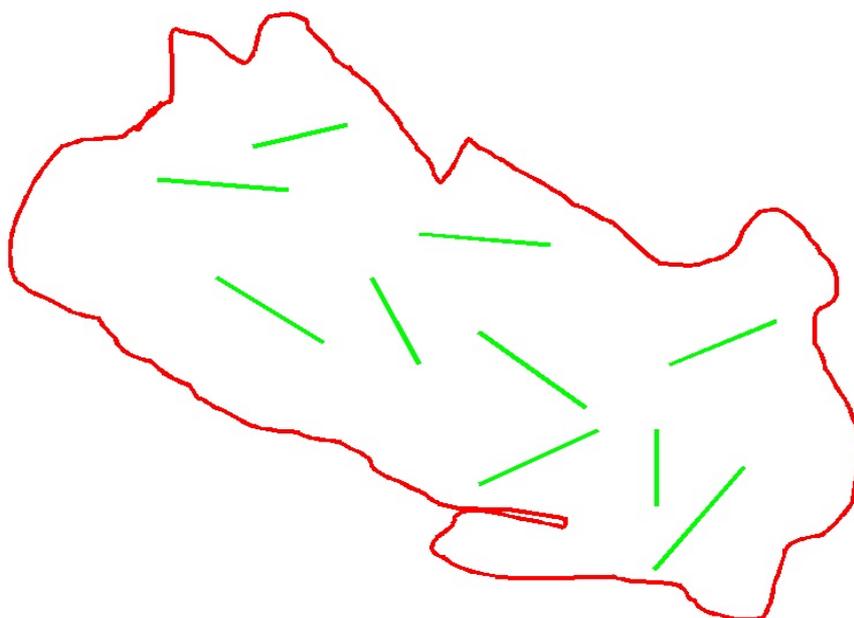


Figure 7 : Echantillonnage simple. Placement aléatoire de 10 transects au sein d'une zone d'étude

L'effort alloué à chaque sous-unité (leur surface dans le cas de quadrats, leur longueur pour des transects, le temps passé pour des points d'écoute...) et leur nombre dépendra du budget-temps disponible pour l'étude. Il faudra évidemment trouver un compromis entre suivre suffisamment les sous-unités pour avoir des résultats qui ont du sens (avoir une bonne détection par exemple) et en suivre suffisamment pour que l'échantillon complet soit représentatif de la zone d'étude. Pour cela, il n'existe pas de règle fixe mais nous fournirons quelques conseils dans la partie décrivant les méthodes d'analyses de données car chaque méthode va poser des contraintes particulières en termes de taille d'échantillon.

Techniquement, la sélection aléatoire de sous-unités est aisée. Dans le cas de quadrats ou de parcelles par exemple, on commence par découper la zone d'étude en quadrats/parcelles sous SIG (ou sur photos ou carte à la main). La forme des parcelles/quadrats a généralement peu d'importance mais on verra par la suite qu'il faudra s'efforcer qu'ils fassent tous la même taille pour la simplicité des calculs. Chaque quadrat est ensuite numéroté. A l'aide d'un logiciel tel que R (logiciel gratuit, cf. lien en annexe) ou même avec Excel, on tire alors aléatoirement une suite de chiffre qui correspondront aux quadrats qui seront suivis.

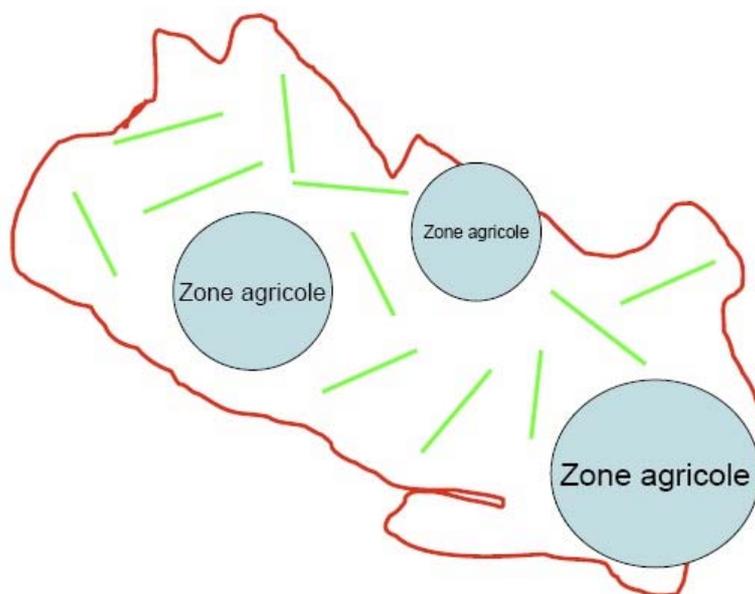


Figure 8 : Echantillonnage aléatoire avec exclusion des zones non occupées (ici les zones agricoles)

4.4.2. Echantillonnage stratifié

Dans de nombreuses situations, l'abondance de l'espèce étudiée varie très fortement selon des caractéristiques de l'habitat. Souvent les naturalistes connaissent par avance, de par leur expérience ou de par la littérature, quels sont les habitats les plus favorables à l'espèce et ceux les moins favorables. Lorsque dans un échantillon, on mélange des données issues de sites à forte densité et des sites à faible densité (grande variabilité des abondances mesurées sur les sites), alors la précision sur la moyenne du nombre d'individus par site est très faible (grande variance). Dans la situation où l'on peut regrouper, **en amont** de l'étude (attention à ne jamais réaliser ceci après la collecte des données), les différentes sous-unités suivies en quelques classes qui présenteront des abondances différentes, alors on peut construire un échantillonnage dit « stratifié ». Ce type d'échantillonnage consiste à définir les catégories de sites, à les identifier sur le terrain (ou sur SIG) puis à tirer aléatoirement les sites qui seront suivis en fixant le nombre de sous-unités que l'on veut suivre dans chaque catégorie ou strate. Ce type d'échantillonnage présente deux intérêts : il permet d'améliorer la précision des estimations de l'abondance par strate mais aussi globalement, il permet de faire varier l'effort de terrain selon les strates. En effet, le nombre de sous-unités suivies de chaque strate n'a pas besoin d'être le même. On peut alors imaginer de suivre plus intensément les sites à forte densité (afin d'avoir une bonne précision des estimations sur ces sites) et moins intensément les sites à faible densité (puisqu'ils apporteront de toute façon moins d'information sur la taille globale de la population).

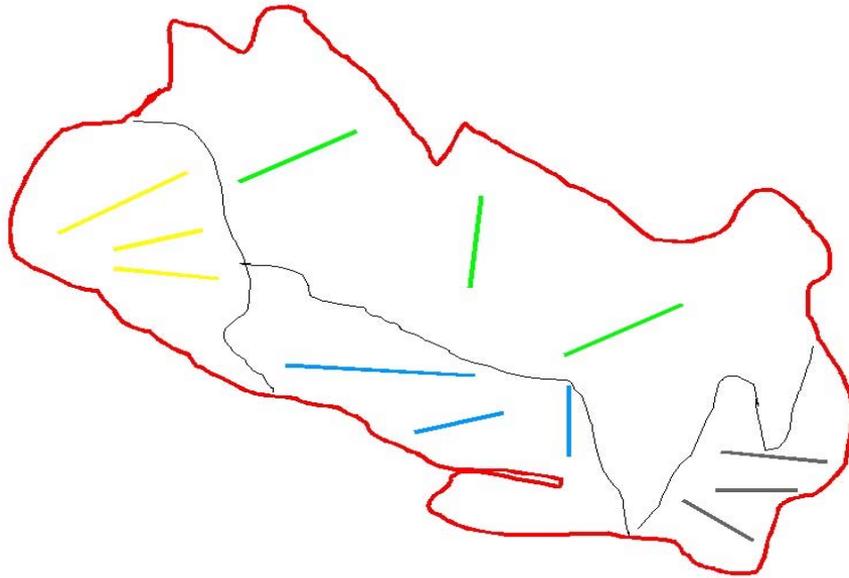


Figure 9 : Echantillonnage stratifié. Placement aléatoire de 12 transects au sein d'une zone d'étude, répartis pas secteurs

Prenons un exemple très concret, même si un peu caricatural, permettant d'illustrer l'importance de la stratification. Nous nous intéressons à une **espèce de chauve-souris** présente dans le sud de la France. Cette espèce est connue pour occuper depuis des décennies cinq grandes grottes, chacune abritant environ 10.000 individus. Par ailleurs, cette espèce occupe aussi de toutes petites grottes, très nombreuses (pour fixer les idées nous dirons qu'environ 10.000 petites grottes sont occupées). Chaque petite grotte abrite cependant uniquement quelques individus, disons cinq en moyenne. Les naturalistes suivent les cinq grandes grottes depuis des années. Si une des grandes grottes devient défavorable pour une raison quelconque et que cette grotte est désertée par les chauves-souris, alors nous observons une baisse de 20% des effectifs de cette espèce (50.000 à 40.000). La chute est importante et impose de mettre en place des mesures de protection plus ou moins coûteuses. Cependant, si en parallèle un autre naturaliste s'intéresse aux petites grottes et observe un accroissement de cinq à six individus en moyenne par grotte dans un échantillon de petites grottes, alors cet expert parle lui d'une augmentation de la population de 20%. Que conclure ? En fait la population de cet exemple théorique est bien évidemment stable, les 10.000 individus qui ont abandonné la grande grotte se sont simplement redistribués dans les petites. Mais chaque groupe de naturalistes se basant sur une population statistique de grottes différentes (les grandes vs les petites), leurs conclusions sont totalement opposées. Une solution pour cette situation serait de construire deux strates : une strate avec les grandes grottes qui pourrait être suivie exhaustivement (les cinq) et une strate avec les petites grottes qui serait suivie par échantillonnage (par exemple une centaine). Ce plan d'échantillonnage assure de pouvoir extrapoler les résultats à l'ensemble de la population. Notons que cet exemple montre aussi les limites au suivi des sites les plus favorables à l'espèce, limites que nous avons déjà abordées auparavant.

4.4.3. Adaptive sampling

Lorsque l'on cherche à suivre une espèce rare et/ou une espèce dont la population est répartie de manière agrégée dans l'espace (regroupement de nombreux individus sur quelques petits sites, au contraire d'une répartition homogène dans l'espace), on comprend aisément qu'un échantillonnage classique risque de conduire à ne suivre que des sous-unités sur lesquelles l'espèce est absente, à moins de pouvoir suivre un très grand nombre de sous-unités. Pour remédier à ce problème, les méthodologistes ont développé l'*adaptive sampling*. Il consiste à tirer aléatoirement un certain nombre de sous-unités que l'on va suivre. Pour le moment, rien de

nouveau. Cependant, l'originalité de ce type de protocole réside dans le fait que, lors du terrain, de nouvelles sous-unités suivies seront ajoutées à l'échantillon selon les résultats obtenus sur les sous-unités sélectionnées initialement. Ainsi ce type d'échantillonnage propose que, lorsque la sous-unité suivie est positive pour la présence de l'espèce, on prospecte les sous-unités adjacentes. Si une ou plusieurs nouvelles sous-unités sont elles aussi occupées par l'espèce, alors on prospecte là aussi les sous-unités adjacentes, etc. Lorsqu'une sous-unité suivie est négative pour la présence de l'espèce, on prospecte alors la sous-unité suivante de l'échantillon retenue en tout premier lieu. Il en est de même lorsque l'on a suivi plusieurs sous-unités adjacentes et que toutes sont négatives. On comprend bien que ce type de démarche permet de mieux suivre les espèces agrégées, car l'effort sera augmenté sur les zones de présence.

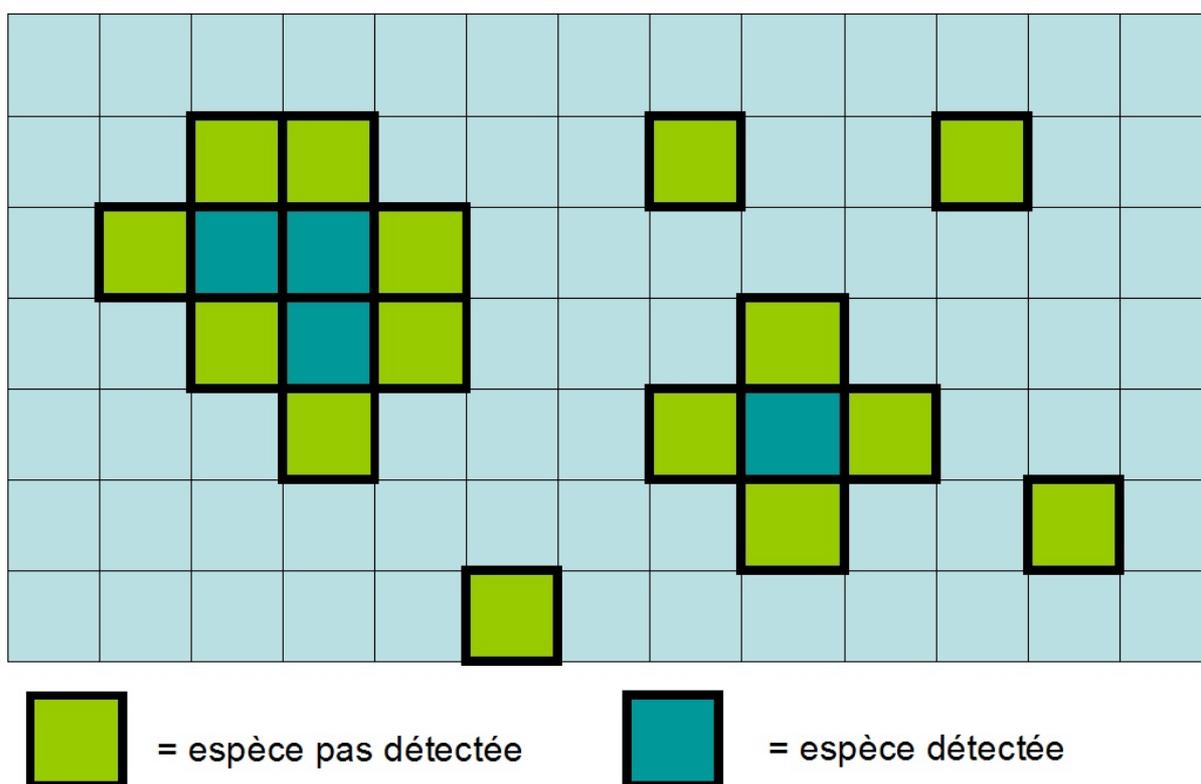


Figure 10 : Adaptive sampling

Attention, en pratique il faut limiter l'utilisation de ce type de protocole aux espèces dont on est sûr qu'elles sont très rares ou très localisées. En effet, si ce n'est pas le cas, vous risquez de devoir prospecter un nombre considérable de sous-unités. En effet, une sous-unité positive entraîne plusieurs nouvelles prospections qui seront alors positives si l'espèce n'est pas vraiment agrégée et ces nouvelles prospections entraînent à leur tour de nouvelles prospections, etc. Certains auteurs ont suggéré d'adopter des règles d'arrêt de la prospection (par exemple en fixant un nombre maximum de sous-unités à suivre) mais il semble que ce type de protocole entraîne un biais dans les estimations des abondances. Pour que ce type de protocole soit valable, il faut donc normalement aller jusqu'au bout de la démarche, y compris en suivant toutes les sous-unités sélectionnées initialement.

4.5. Les contraintes et pseudo-aléatoire

En pratique, il est souvent impossible d'avoir accès à toutes les sous-unités potentielles. Par exemple, on peut être confronté à des problèmes d'autorisation d'accès de par les propriétaires ou tout simplement à des problèmes d'accessibilité des sites (maquis impossible à prospecter en transects, zones de montagne inaccessibles ou trop dangereuses, etc.). Dans ces situations, il faut alors faire avec le réel et ne plus inclure dans l'échantillon que des sous-unités effectivement prospectables. A première vue, ce type de contrainte pourrait ne pas poser de problème puisqu'il est totalement indépendant de la volonté de l'observateur et de sa connaissance de l'espèce. Cependant, il faudra se poser une question très importante : est-ce que les zones auxquelles j'ai accès sont représentatives de l'ensemble de la zone vis-à-vis de l'espèce étudiée. Si ça n'est pas le cas, alors il faudra être prudent dans l'extrapolation des résultats à la surface totale.

Exemple : suivis de tortues d'Hermann en Corse :

Les maquis corses sont très difficiles voire impossibles à prospecter du fait que la végétation est très dense et que les terrains sont souvent en forte pente. Dans cette situation, il a été convenu de suivre les populations de tortues en suivant les sentiers de chèvres dans le maquis. Il a été procédé à un tirage aléatoire des sentiers à suivre (tirage aléatoire parmi le réellement prospectable). La question ouverte étant : est-ce que ce que l'on observe en densité au bord des sentiers est représentatif de ce qui se passe dans le maquis ?

4.6. Réplicats temporels et échantillonnage

Lorsqu'on désire suivre une population d'espèce sur un moyen ou long terme pour déterminer son évolution (c'est généralement ce que l'on cherche à faire, en dehors d'un simple état des lieux à un instant donné), il se pose la question de l'échantillonnage à chaque nouvelle étape du suivi. Faut-il garder les mêmes sites (ou sous-unités) ou re-sélectionner aléatoirement de nouveaux sites à chaque fois. La réponse n'est pas évidente en premier lieu. La statistique voudrait que l'on tire de nouveaux sites à chaque fois. Cette procédure permet en effet d'avoir systématiquement une vision objective de l'état de la population à un instant donné et de comparer *a posteriori* ces états. Cependant, dans cette situation, on ne s'intéresse qu'au bilan : on a une estimation de l'abondance à une date donnée et une autre quelques mois/années après, que l'on compare, mais aucune information sur le pourquoi du changement s'il y en a.

Le suivi régulier des mêmes sites ne pose pas le même problème. En effet, il va assurer d'estimer la dynamique des sites (croissance ou décroissance locale) et éventuellement la mettre en relation avec des changements sur ces sites (perturbation, mise en place d'une gestion, etc.) si ces informations ont été relevées. Avec ce deuxième type de protocole, on a donc accès à plus d'information sur le fonctionnement des populations, ce qui est très pertinent. Cependant, il faut noter que les sites suivis sont toujours les mêmes, donc que les données collectées d'une année à l'autre ne sont pas statistiquement indépendantes (ce qui peut avoir pour conséquence de biaiser les intervalles de confiance des estimations). Les analyses statistiques usuelles posent souvent cette hypothèse d'indépendance des données mais des méthodes plus récentes permettent de modéliser cette dépendance (nommée autocorrélation temporelle). Ces méthodes sont plus complexes à mettre en œuvre mais elles existent. Lorsque ce choix est possible, il vaudrait donc mieux privilégier le suivi des mêmes sites d'une année sur l'autre plutôt qu'un tirage aléatoire chaque année.

5. Les méthodes d'analyse de données et leurs hypothèses

Quatre grands types de méthodes sont ici présentés :

- Recensements
- Distance Sampling
- Capture Marquage Recapture (CMR)
- Occupancy, ou présence-absence

5.1. Les recensements

Les recensements sont la technique la plus souvent utilisée pour l'estimation des abondances ou d'une taille de population. Ils posent l'hypothèse cruciale que la détection des individus est de 100%, c'est-à-dire que tous les individus présents sur un site donné sont observables et comptés lors de la session de terrain. Ils peuvent prendre de nombreuses formes. Par exemple, on peut envisager de compter des plantes à fleur sur des quadrats d'un mètre sur un mètre, des arbres sur des parcelles-échantillon de plusieurs hectares, des mollusques sur un transect linéaire de largeur définie en bords de mer, etc.

Dès lors que l'on a obtenu des dénombrements sur les sous-unités de l'échantillon, de même taille, les calculs de l'abondance ou de la densité moyenne ainsi que de leurs intervalles de confiance sont relativement aisés et peuvent se réaliser à l'aide d'un logiciel convivial tel que Excel. Lorsque l'on travaille avec un échantillonnage de type « stratifié » (voir ci-dessus), les calculs se complexifient dans leur forme mais restent très simples à réaliser avec Excel.

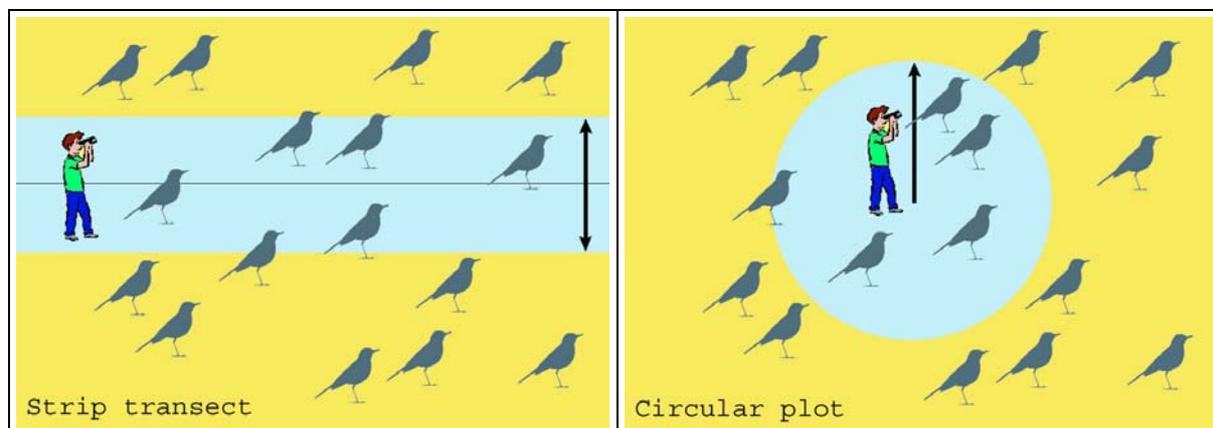


Figure 11 : Protocole de recensement par quadrat rectangulaire ou circulaire

D'une manière générale on s'efforcera de travailler sur des zones de mêmes formes et de même taille. Le respect de ces contraintes, qui peut être selon les conditions plus ou moins difficile, a pour triple avantage de réduire les erreurs dans la prospection des sites, de simplifier les calculs ultérieurs et souvent de réduire les intervalles de confiance des estimations. Souvent, on préconise de travailler sur des cercles, des carrés ou des rectangles. L'avantage du cercle est qu'il ne nécessite de définir qu'un seul point pour la zone à échantillonner mais, lorsque l'on désire suivre des sous-unités de grande taille, les limites du cercle peuvent devenir difficile à déterminer. Son deuxième avantage est qu'il minimise le périmètre par rapport à la surface choisie et donc minimise les erreurs d'assignement (cas des individus qui sont en bordure, pour lesquels on ne sait pas s'il faut les inclure ou les exclure du décompte). Le carré a un plus petit périmètre par unité de surface par rapport au rectangle, donc minimise là encore les erreurs d'assignement. Malheureusement en pratique la définition d'un cercle, d'un carré ou d'un rectangle n'est pas toujours possible ou demande des efforts importants. On pourra alors

choisir de définir des sous-unités de formes variées au sein du même échantillon, par exemple du fait de contraintes topographiques. En théorie ceci ne pose pas de problème majeur pour les calculs tant que la taille (surface) des zones est strictement identique. Dans le cas contraire, les estimations de la densité moyenne ou de la taille de la population restent relativement aisées mais celles de leurs intervalles de confiance ne l'est plus du tout. Le calcul des intervalles de confiance implique par exemple de vérifier que l'écart-type des estimations est proportionnel à la surface échantillonnée et, lorsque ça n'est pas le cas, des méthodes plus complexes doivent être mises en œuvre. On comprendra bien qu'à moins de fortes contraintes de terrain, il faudra largement privilégier le suivi de sous-unités de même taille.

La détermination de la taille des sous-unités à suivre et de leur nombre est une étape délicate. En général, à budget constant, on pourra étudier une certaine surface donnée qui pourra se décliner selon un gradient allant d'une seule sous-unité très grande à de très nombreuses sous-unités de petites tailles (sauf contraintes de déplacement importantes). On a vu qu'augmenter la taille de l'échantillon, c'est-à-dire le nombre de sous-unité, avait pour conséquence de réduire les intervalles de confiance, ce qui peut être très intéressant. Cependant de trop petites zones ont des périmètres importants par rapport à la surface étudiées et donc augmentent le risque d'erreur d'assignement. Par ailleurs, des petites zones favorisent les comptages nuls (zéro individu) et l'on sait qu'un grand nombre de comptages nuls a tendance à biaiser les estimations d'intervalles de confiance. **Une petite règle empirique est que la taille des sous-unités doit être suffisamment grande pour qu'au moins la moitié des comptages soient non-nuls. Enfin pour des raisons d'ordre statistique, il peut être intéressant de décompter sur trente sous-unités ou plus (respect des conditions de la loi « normale » ou loi de « Laplace-Gauss »).** Le respect de cette hypothèse simplifie le calcul des intervalles de confiance.

Malgré leur simplicité, les recensements posent l'hypothèse de comptages exhaustifs des individus occupant les sous-unités suivies. Dès lors que cette hypothèse n'est pas respectée, les recensements sont biaisés. Comme nous l'avons vu, augmenter le nombre de sous-unités suivies réduira les intervalles de confiance des estimations obtenues, mais le biais restera le même. Cette hypothèse d'exhaustivité limite fortement leur utilisation à des espèces peu mobiles, plutôt de grande taille, dans des milieux assez ouverts et sur des surfaces de tailles limitées.

Détection imparfaite

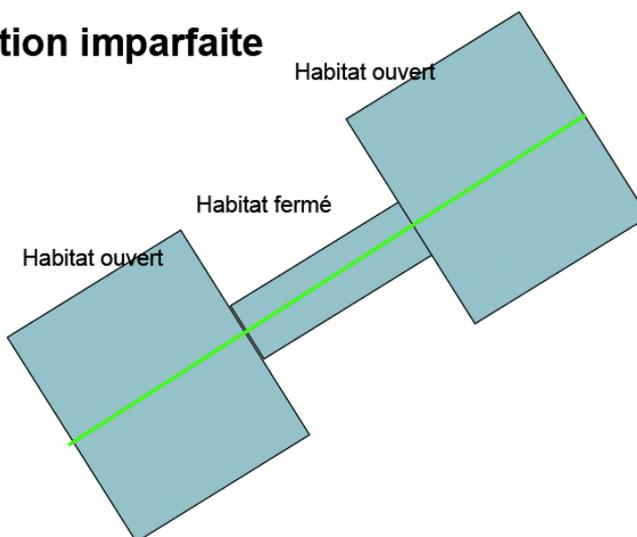


Figure 12 : La détection peut varier selon les types de milieux présents dans le quadrat

Dans quelques cas particuliers, il est possible d'échantillonner l'ensemble de la zone d'étude, permettant ainsi un recensement très précis, souvent à l'individu près. C'est le cas notamment pour le **comptage de grands mammifères** au sein d'un vaste espace clos, tel que certains massifs forestiers. Dans ce cas, une technique de comptage consiste à rassembler de nombreux observateurs, qui forment une ligne avec une distance régulière entre chaque observateur, suffisante pour permettre aux animaux de passer. En ratissant la zone, chaque observateur note les animaux franchissant la ligne sur sa droite (ou sa gauche). Cette technique est régulièrement utilisée pour des recensements à but cynégétique et permet même de sexer et âger les animaux.

Signalons également l'existence de techniques de **comptage simultané** : un grand nombre d'observateurs prospecte un même site, le même jour à la même heure. Chaque observateur (ou petit groupe d'observateurs) se voit affecté la prospection d'un secteur déterminé. A l'issue du comptage, le cumul des observations simultanées sur l'ensemble des secteurs du site permet d'estimer l'effectif global, en tenant compte des doubles comptages (lorsque des animaux se déplacent d'un secteur à un autre). Cette technique est régulièrement utilisée en ornithologie (ex : recensement annuel de l'Outarde canepetière en plaine de Crau, du Grand-duc d'Europe sur l'ensemble d'un massif voire même d'un département, comptage d'oiseaux migrateurs sur plusieurs cols distants de plusieurs kilomètres pour étudier la répartition du flux migratoire, etc.).

On est donc ici dans des formes de recensement qui posent l'hypothèse que tous les individus ont été vus, ce qui même dans des situations *a priori* simples comme ici n'est pas vraiment assuré.

5.2. Le « Distance Sampling »

Dès lors que la possibilité de détecter tous les individus présents sur un site n'est plus certaine (on parlera alors d'une probabilité de détection des individus présents sur le site étudié, inférieure à 1 ou inférieure à 100%), les recensements peuvent s'avérer largement biaisés et ne sont donc plus applicables sans risque. Comme nous le verrons ultérieurement, l'utilisation de méthodes dites de « Capture-Marquage-Recapture » (notées en abrégées « méthodes de CMR ») peut s'avérer pertinente dans ces conditions limitées de détection des individus présents, mais les protocoles associés à ces méthodes nécessitent souvent un investissement très important sur le terrain (plusieurs sessions de terrain successives) et imposent un certain nombre de contraintes techniques parfois insurmontables (notamment la capture répétée des mêmes individus). Lorsque la probabilité de détection des individus présents est inférieure à 100% et/ou lorsque le marquage des individus est trop contraignant en termes de temps à passer sur le terrain pour obtenir un échantillon suffisant pour appliquer les méthodes usuelles de CMR, on pourra s'orienter vers les méthodes dites de « *Distance Sampling* » (que l'on pourrait traduire par « *méthode d'échantillonnage par la distance* ») dont le fondement théorique repose sur l'hypothèse que la probabilité de détection des individus présents sur un site est fonction de la distance à l'observateur. Un naturaliste de terrain, même inexpérimenté, comprendra intuitivement cette hypothèse fondamentale des méthodes de « *Distance Sampling* ».

Posons l'hypothèse qu'une partie des individus présents sur la zone étudiée n'est pas détectée. L'intuition du naturaliste permet aussi de supposer que plus un individu est loin de l'observateur, plus le risque de le rater est grand. Cette intuition conduit à formuler l'hypothèse, en termes plus mathématique, que la probabilité de détection des individus présents est une fonction, *a priori* décroissante, de leur distance à l'observateur. Si l'observateur note les distances d'observation de tous les individus détectés au cours de son étude, on peut comprendre intuitivement que le jeu de données ainsi collecté inclura moins de « longues » distances que de « courtes » distances. Pourtant, si les individus réellement présents sur la zone d'étude sont répartis de manière plus ou moins homogène, ce qui est souvent le cas, leurs distances à l'observateur sont elles aussi homogènes et le jeu de données récolté devrait inclure autant de « longues » distances que de « courtes » distances. Ainsi, la différence entre

le nombre de distances « courtes » et le nombre de distances « longues » au sein du jeu de données réellement collecté sur le terrain, semble inclure les informations sur la baisse de la probabilité de détection des individus avec leur distance à l'observateur. La répartition de ces distances au sein du jeu de données devrait donc permettre d'estimer les probabilités de détection en fonction de la distance. Ces probabilités de détection étant estimées, il est aisé de les utiliser pour estimer le nombre d'individus réellement présents sur la zone (si on estime à 50% le nombre d'individus que l'on détecte à 50 mètres et que l'on a réellement observé 10 individus, alors on peut estimer à $10/0.5$ soit 20 le nombre d'individus réellement présents à cette distance). Le fait que l'échantillon de distances d'observations collectées sur le terrain fournisse les données nécessaires à l'estimation de la densité a conduit les méthodologistes à nommer ces méthodes « *Distance Sampling* ».

Les méthodes de « *Distance Sampling* » sont de deux formes :

- « *Line transect* » : parcours d'une ligne de longueur donnée et observation de part et d'autre de la ligne,
- « *Point transect* » : observation durant un temps déterminée à partir d'un point fixe.

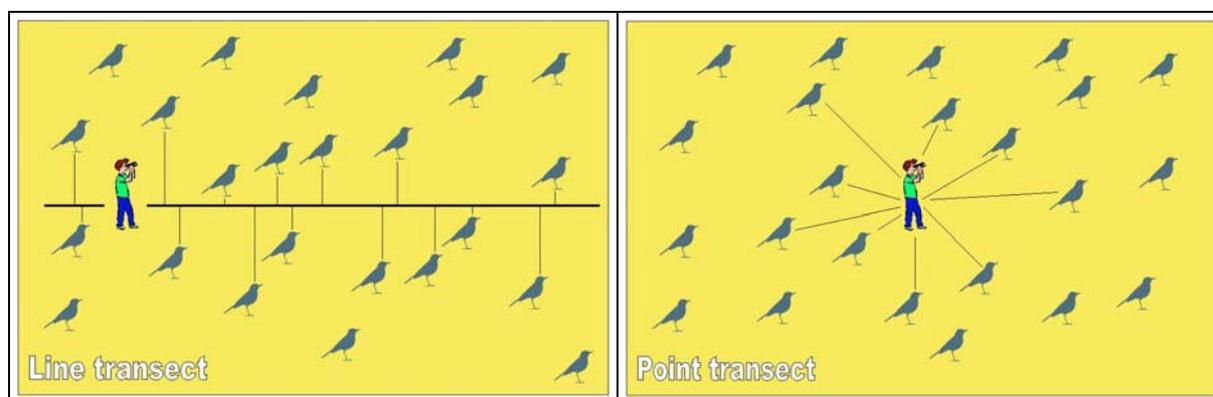


Figure 13 : Protocole de « Distance Sampling » : « Line transect » et « Point transect »

En « *line transect* » les distances à mesurer sont les distances perpendiculaires à la ligne parcourue. Il est souvent plus courant de détecter les individus bien avant d'arriver au point où ils seraient placés perpendiculairement à la ligne. Il est donc souvent plus aisé de noter la distance d'observation. Ceci ne pose pas de problème particulier si et seulement si l'angle d'observation par rapport à la ligne est noté lui aussi puisqu'un simple calcul de trigonométrie permettra de calculer la distance perpendiculaire. Certaines méthodes ont été développées pour estimer les densités directement à partir des distances et des angles d'observations, mais ces méthodes nécessitent des hypothèses plus restrictives et souvent peu plausibles comparées aux méthodes se basant sur les distances perpendiculaires. Notons que le calcul des distances perpendiculaires impose de bien connaître à l'avance le trajet qui sera parcouru.

En « *point transect* », on parle de distance radiale, c'est-à-dire la distance entre l'observation et l'observateur sans prendre en compte les angles.

Le développement récent d'outils électroniques tels que les télémètres laser facilite grandement la mesure de distances sur le terrain. Certains constructeurs proposent aujourd'hui des jumelles avec télémètre intégré.



Figure 14 : Exemples de télémètres lasers disponibles sur le marché

« Line transect » ou « point transect » ? avantages / inconvénients

Le choix de travailler en « line transect » ou en « point transect » peut être imposé par des contraintes liées au terrain. Par exemple sur des espèces sensibles au dérangement, on aura plutôt tendance à travailler en « point transect » de manière à limiter les déplacements dérangent. De même dans des habitats difficiles à parcourir on utilisera les « point transects ». Par contre on privilégiera les « line transects » sur des espèces assez rares, de manière à maximiser la surface prospectée, ou sur des espèces qui sont détectées lors de leur fuite par exemple. Il est important de noter qu'on ne peut pas dans le même protocole mélanger des données issues de « point transect » et de « line transect » du fait de contraintes méthodologiques. Par conséquent si l'on choisit de faire du « point transect », toutes les observations réalisées pendant les déplacements entre les points sont perdues. Cependant, lorsque l'on a le choix de construire un protocole en « line transect » ou « point transect », il faut réfléchir à différents atouts et limites de chaque méthode. Un des avantages du « point transect » est qu'il ne nécessite pas de concentrer ses efforts sur le parcours et souvent les distances sont plus faciles à mesurer (point de repère, immobilité de l'observateur...). Toutefois il est important de noter que le « point transect » est souvent bien moins précis que le « line transect » pour le même nombre d'observations et qu'il est très sensible aux déplacements des individus. Sa moindre précision vient du fait que la surface prospectée n'est pas constante entre deux distances. Alors qu'en « line transect » la surface prospectée entre deux distances (bande de prospection) est constante quelle que soit la distance (par exemple cette surface est identique entre 10 et 15 mètres ou 20 et 25 mètres). En « point transect » la surface prospectée entre deux distances augmente avec la distance à l'observateur. En effet la surface de l'anneau entre 10 et 15 mètres est bien plus petite que la surface de l'anneau entre 20 et 25 mètres. Ceci a pour conséquence que la surface prospectée à proximité de l'observateur est très petite et donc que le jeu de données collecté inclut nécessairement très peu de courtes distances. Or, d'un point de vue méthodologique, on sait que ce sont ces courtes distances qui sont les plus importantes pour estimer l'abondance. Par conséquent leur faible nombre en « point transect » a pour conséquence directe une faible précision. En conclusion, sauf contrainte forte comme celles vues précédemment, il est recommandé pour une plus grande précision de réaliser des « line transects ».

L'utilisation des méthodes de « *Distance Sampling* » repose sur le respect de trois hypothèses, leur non-respect entraînant des biais majeurs dans les résultats obtenus.

Hypothèse 1 : Les individus présents sur la ligne ou sur le point sont tous détectés

Lorsque cette hypothèse n'est pas respectée, la densité estimée est biaisée négativement. Lorsque l'on sait que cette hypothèse n'est *a priori* pas respectée, deux solutions sont envisageables :

La première consiste à améliorer le protocole pour s'assurer que la détection soit exhaustive sur la ligne. Par exemple, il est possible d'affecter un observateur spécifiquement à la ligne tandis qu'un deuxième observateur « prospecte » l'ensemble de la zone si la détection est difficile sur la ligne.

La deuxième solution consiste à estimer par une autre méthode la détection sur la ligne (par exemple à l'aide de deux observateurs indépendants sur le même transect) et ainsi de corriger l'estimation obtenue (le logiciel « DISTANCE » usuellement utilisé pour les analyses de données issues de protocoles de « *Distance Sampling* », permet de préciser cette détection sur la ligne lorsque l'on obtient son estimation par une autre méthode).

Hypothèse 2 : Les animaux ne doivent pas se déplacer avant leur détection.

Si les déplacements sont aléatoires vis-à-vis de l'observateur et que leur vitesse est réduite en comparaison à celle de l'observateur (de l'ordre de la moitié de la vitesse de l'observateur), alors le biais peut être généralement négligé. Si les mouvements sont effectués par les animaux en réponse à la présence de l'observateur, alors les estimations de densité peuvent être très fortement biaisées. Le biais est négatif (sous-estimation de la densité) lorsque les animaux fuient l'observateur. Au contraire, il est positif lorsque les animaux sont attirés par l'observateur (curiosité, repasse acoustique d'enregistrements de chants/cris, etc.).

Un comportement de fuite vis-à-vis de l'observateur peut généralement se détecter sur les histogrammes d'observation en fonction des distances, mais pas dans le cas d'une attraction. Lorsqu'on suppose qu'un tel comportement de fuite ou d'attraction est possible pour l'espèce étudiée, les solutions méthodologiques sont inexistantes pour débiaiser les résultats. Les solutions pratiques sont quant à elles relativement limitées, elles consistent à attendre un peu afin que les individus reviennent à leur position initiale (territoires) pour les protocoles de « *Points Transects* », ou à regarder loin devant lors des protocoles de « *Line Transect* » afin de détecter les individus avant leur déplacement. Pour les espèces disposant de capacités de déplacement très importantes et dont la discrétion au cours de ces déplacements est de mise, alors les méthodes de « *Distance Sampling* » peuvent s'avérer totalement inutilisables.

Hypothèse 3 : Les distances sont mesurées avec précision et sans biais

Les méthodes de « *Distance Sampling* » sont relativement robustes lorsque les erreurs de mesure sont aléatoires. En effet, des erreurs aléatoires ont pour conséquence logique de se compenser (une fois un individu sera attribué à une distance trop importante, la fois suivante un individu sera attribué à une distance plus faible, etc.). Cependant, on comprendra aisément que la méthode se révèle particulièrement sensible à des biais systématiques des mesures. Ces biais systématiques sont fréquents sur les mesures des longues distances, qui sont souvent plus difficiles à estimer que de courtes distances (notons que les longues distances sont souvent sous-estimées même par des naturalistes chevronnés).

Par ailleurs, les méthodes de « *Distance Sampling* » sont aussi très sensibles aux erreurs liées aux arrondis. Ainsi, tout naturaliste de terrain sait bien qu'on utilise fréquemment des mesures du type 5, 10, 15,..., 40, 50, 100 et rarement les mesures 7, 9, 13...41, 48, 53, 98 qui correspondent pourtant à la réalité. Notre faible capacité d'estimation des distances est la cause principale de ce phénomène que les anglophones nomment « *heaping* » et qui peut avoir des conséquences catastrophiques sur l'estimation des densités.

Lorsque les mesures des distances exactes s'avèrent difficiles, ou lorsque des arrondis sont systématiquement effectués, il est possible de regrouper les distances en classes de distances, qui sont parfaitement utilisables pour obtenir des estimations de densités non-biaisées, pourvu que ces classes soient constituées de manière pertinente.

De nombreux outils ou méthodes existent pour estimer les distances avec plus ou moins de précisions et qui sont plus ou moins pertinentes selon les espèces étudiées. Il conviendra lors de l'élaboration des protocoles de bien s'interroger sur la méthode la plus pertinente pour la situation étudiée, ainsi que du degré de précision nécessaire, à mettre en relation avec le coût financier et le coût en terme de temps à passer sur le terrain des différentes méthodes.

Enfin, une dernière hypothèse souvent posée est à examiner, celle des doubles comptages du même individu. Lorsque ces doubles comptages sont effectués sur le même transect, par exemple lorsque des individus sont détectés à deux emplacements différents suite à leur déplacement discret pendant le parcours du transect, alors l'estimation est biaisée. Par contre, lorsque le même individu est compté au cours du parcours de deux transects différents, ces transects constituent différentes unités de comptage (ou unités statistiques) et ces doubles comptages ne biaisent pas les estimations de densité ou faiblement, pourvu qu'ils soient en quantité relativement limitée.

Notons que pour obtenir des estimations dont la précision sera raisonnable (par exemple de l'ordre d'un facteur deux entre les limites inférieures et supérieures de l'intervalle de confiance), les études préalables semblent montrer qu'il faut obtenir au minimum 60-80 observations, ce qui peut être particulièrement difficile dans le cas d'espèces rares ou très difficiles à détecter par exemple. Cette taille d'échantillon n'est malheureusement pas une règle fixe car elle dépend largement de la variabilité de la densité de l'espèce étudiée. Si cette variabilité est grande dans l'espace alors il faudra probablement plus d'observations. Si elle est faible, moins d'observations pourront suffire. Nous en reparlerons par la suite dans ce document mais il est très intéressant d'analyser vos données au fur et à mesure de leur collecte pour pouvoir facilement optimiser votre effort de terrain en fonction de vos objectifs en termes de précision.

5.3. Les Captures-Marquage-Recaptures

Les méthodes dites de « Capture-Marquage-Recapture », notées par la suite CMR, reposent tout comme les méthodes de « *Distance Sampling* » sur l'hypothèse que tous les individus ne sont pas détectés lors d'un passage sur le terrain. Elles se proposent elles aussi d'estimer la probabilité de détecter les individus, pour ensuite corriger les comptages afin d'obtenir une estimation non-biaisée par la détection de l'abondance sur un site.

Ces méthodes se basent généralement sur l'identification individuelle des animaux (mais aussi des plantes dans quelques situations) suivis sur le terrain. Dans quelques situations particulières que l'on abordera ci-dessous, on pourra se contenter d'identifier des groupes d'individus. D'une manière générale, il faudra donc pouvoir distinguer les individus et les reconnaître au cours du temps. Pour quelques espèces, ceci peut être possible sans poser de marque voire même sans capture. Par exemple en utilisant des marques de blessures sur les cétacés ou les patrons de coloration chez certains amphibiens tels que le Sonneur à ventre jaune ou le Triton crêté. Mais ces situations sont relativement rares et peuvent conduire aussi à certaines erreurs d'identification, si l'on ne prend pas assez de précaution (double identification, photo standardisé des individus...). Dans la très grande majorité des cas, il faudra capturer les individus, les marquer et ensuite relire leurs marques lors de plusieurs sessions successives, soit par recapture (par exemple les bagues métalliques chez les oiseaux lisibles seulement lorsque l'oiseau est en main), soit par simple lecture à distance (par exemple les bagues darvik chez les oiseaux, les marques par fanions colorés sur des grands mammifères ou des rapaces, marques lisibles à plusieurs dizaines voire centaines de mètres à l'aide d'un télescope). Cette contrainte de capture/marquage limite l'utilisation de ces méthodes à des espèces peu sensibles à la capture et à la manipulation, pour lesquelles on dispose de méthodes de capture efficaces et pour lesquelles existent des systèmes de marquage inoffensifs (ne modifiant pas la biologie des individus, leur succès reproducteur, leur taux de mortalité, leur capacité de déplacement...) et durables (qui ne se perdent pas ou ne se dégradent pas au cours de la saison de terrain).

Les techniques de capture sont très diverses et souvent spécifiques à des groupes d'espèces (filet pour les oiseaux, épuisettes pour les poissons, pièges pour les micromammifères, capture à la main pour les tortues terrestres...). Il en est de même pour les types de marquages (bagues chez les oiseaux, transpondeurs chez les poissons, amputation chez les reptiles [de plus en plus rares heureusement]...). Dans quelques rares cas (grands carnivores), l'identification des individus peut être obtenue par analyse génétique des fèces trouvées sur le terrain. Nous n'aborderons pas ici ces aspects qui sont largement disponibles dans la bibliographie et en consultant les spécialistes des espèces suivies. Notez cependant que les techniques de capture et de marquages évoluent fortement ces derniers temps, notamment du fait de la miniaturisation des éléments de type transpondeur ou radio-émetteur. Il peut donc être intéressant de se tenir au courant de ces avancées et d'évaluer systématiquement la pertinence d'utilisation de ce type de nouveaux outils lorsque l'on construit un nouveau protocole.

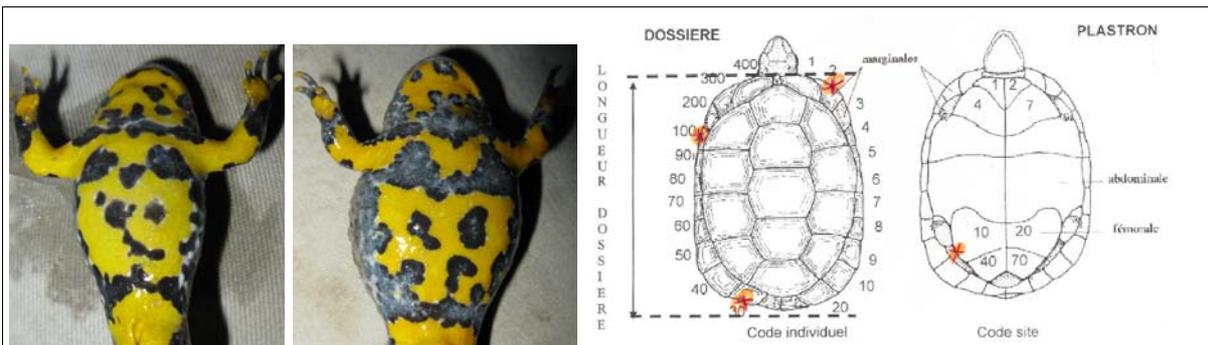
Pour les aspects purement méthodologiques, il existe un certain nombre d'hypothèses dans les méthodes de CMR permettant d'estimer les tailles de populations. Ces hypothèses doivent être respectées pour que les estimations ne soient pas biaisées. Leur respect impose d'une manière générale des contraintes dans la manière de mener le terrain.

La première hypothèse est que le terrain doit être conduit en un minimum de 2 sessions, idéalement plus. Lors de chaque session, l'observateur doit prospecter de manière **aléatoire l'ensemble de la zone d'étude** sur laquelle il veut estimer la taille de la population. Par exemple, il doit réaliser une prospection aléatoire de quadrats de 5 hectares pour des Tortues d'Hermann, ou poser des lignes de pièges de manière aléatoire à chaque session pour capturer des micromammifères, ou poser des nasses de manière aléatoire dans des mares pour

recenser des cistudes. Nous verrons par la suite quelles sont les conséquences du non respect de ces deux contraintes.

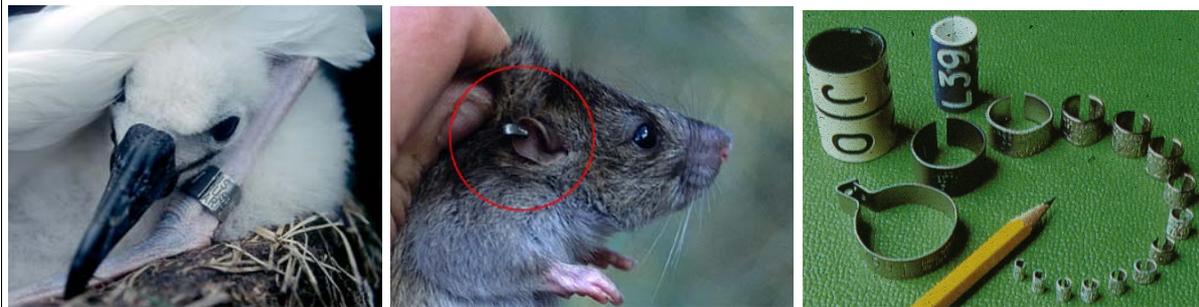
La deuxième hypothèse fondamentale est que **les marques ne doivent pas être perdues ou ne doivent pas se dégrader et devenir illisibles** durant tout le temps où les sessions de terrain sont conduites. Contrairement à ce que l'on pourrait croire, les pertes de marques sont très fréquentes et les conséquences sur les estimations sont extrêmement importantes. En effet, s'il y a perte de marques, on aura tendance à sur-estimer la taille de la population comme nous le verrons ci-dessous.

La troisième hypothèse est que **la population étudiée doit être « close »**, c'est-à-dire qu'elle n'échange pas d'individus avec l'extérieur par émigration ou immigration et qu'il n'y a ni natalité ni mortalité dans l'intervalle de temps entre la première et la dernière session de terrain. Cette hypothèse peut paraître difficile à respecter. Cependant rien ne contraint à réaliser des sessions très espacées les unes des autres. Il est tout à fait possible en théorie de réaliser des sessions de terrain très proches dans le temps, par exemple des passages tous les jours sur des tortues, toutes les deux heures pour des papillons, etc. Ce type de protocole, que l'on adaptera à la biologie de l'espèce et à ses capacités de déplacement notamment, permet en général de limiter les échanges avec l'extérieur. Il permet aussi de s'affranchir de tous les problèmes de mortalité/natalité.



Marques corporelles

- taches ventrales, Sonneur à ventre jaune - Photos : V.Rivière
 - limage d'écailles, Tortue Cistude - Photos : CEN L.R.



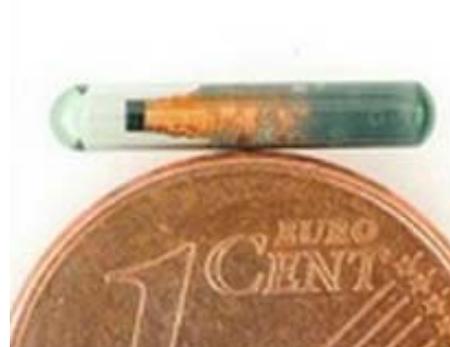
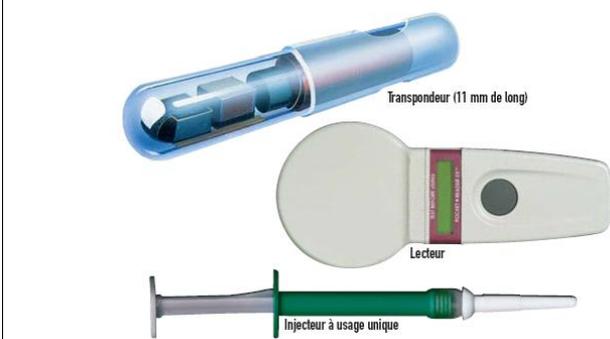
Bague métal (albatros, rat) – Photos : JM.Salles + internet



Bague darvik (skua)
 Photos : JM.Salles



Fanions couleur (busard, chamois)
 Photos : JM.Salles + M.Cornillon



Transpondeur - Photos : internet

Figure 15 : Illustration de différentes techniques de marquage

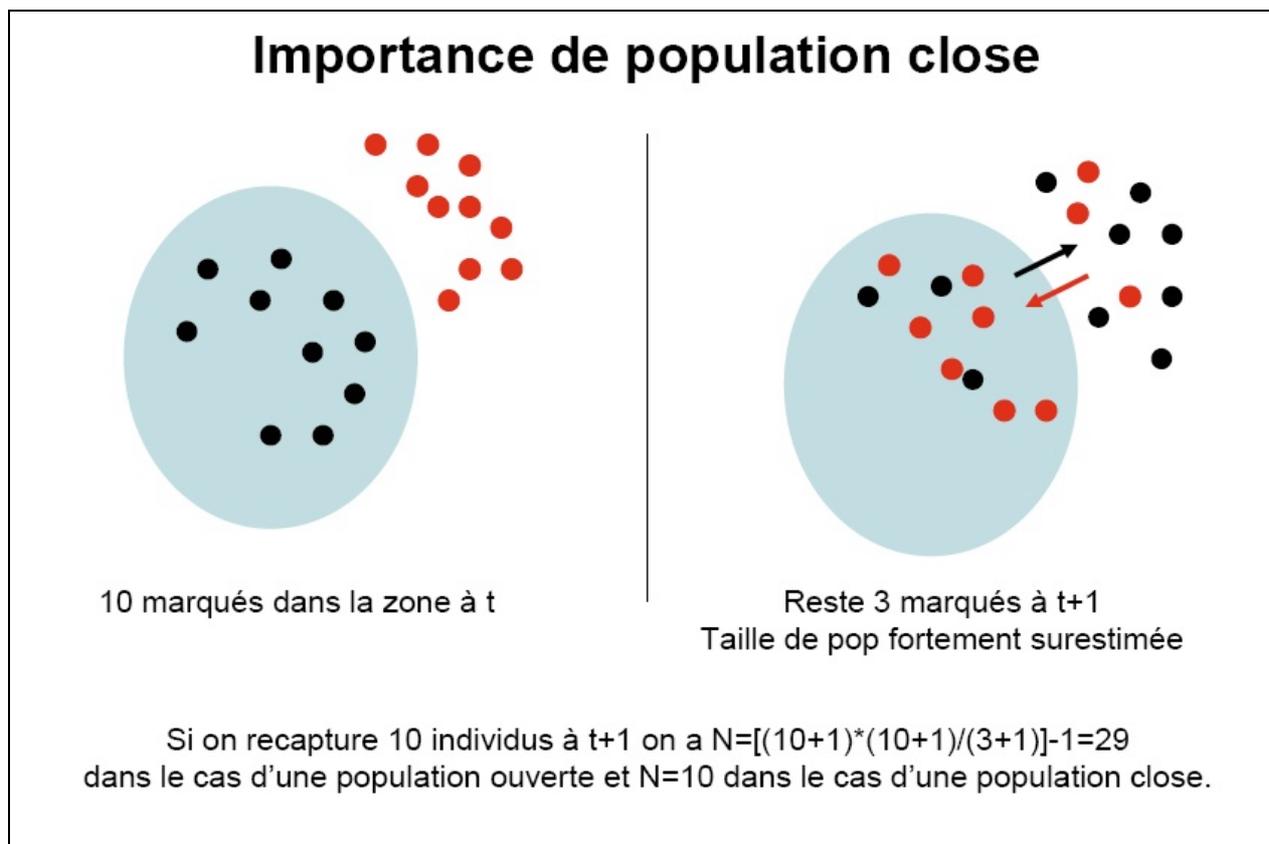


Figure 16 : CMR. Comparaison entre deux populations, close et non close

Pour mieux comprendre l'importance du respect de ces hypothèses, il faut rentrer un minimum dans le détail de la méthode. Les estimations de CMR reposent sur le principe de dilution des marques. Prenons pour illustrer ce principe un exemple très simple avec seulement deux sessions de terrain réalisées à un jour d'intervalle sur une zone clairement délimitée, sur une population de Tortue d'Hermann. Nous prospectons la zone de manière aléatoire le premier jour et nous capturons 15 Tortues d'Hermann. Celles-ci sont toutes marquées individuellement avec un numéro à la peinture et relâchées à leur place initiale. Le lendemain, un autre observateur prospecte à son tour la zone et capture 12 tortues dont 4 sont déjà marquées. Le fait que le deuxième observateur soit neutre par rapport aux captures du premier observateur (il ne connaît pas la localisation des tortues marquées) permet d'affirmer qu'il avait autant de chance de trouver une tortue marquée qu'une tortue non-marquée (nous verrons plus loin que ce n'est pas toujours le cas). La proportion de tortues marquées dans celles qu'il a capturées lors du deuxième passage est donc la proportion de tortues marquées sur l'ensemble du site prospecté. Une simple règle de trois permet donc de calculer la taille de la population notée N :

$$N = (\text{Nb tortues capt. 1}^{\text{er}} \text{ pass.}) \cdot (\text{Nb tortues capt 2}^{\text{ème}} \text{ pass.}) / (\text{Nb tortues capt. les deux fois})$$

Dans le détail le calcul doit être légèrement corrigé de la manière suivante :

$$N = [(\text{Nb tortues capt. 1}^{\text{er}} \text{ pass.} + 1) \cdot (\text{Nb tortues capt 2}^{\text{ème}} \text{ pass.} + 1) / (\text{Nb tortues capt. les deux fois} + 1)] - 1$$

Ce calcul se nomme le « Lincoln-Peterson index » et est un outil très simple et très pratique pour estimer une taille de population avec seulement deux passages mais aussi la probabilité de détection des individus par chaque observateur. Ici ces probabilités de détection vont être un simple ratio entre le nombre de tortues capturées par chaque observateur et l'estimation de la taille totale de la population. Comme nous l'avons abordé précédemment, une estimation n'a

que peu de valeur en soit, car elle requière pour être interprétée un intervalle de confiance. Nous ne détaillerons pas ici le calcul de cet intervalle de confiance qui peut être facilement retrouvé sur le web par exemple, mais sachez qu'il peut être calculé très simplement sous Excel. Dans le cas de notre exemple sur les Tortues d'Hermann, la taille de la population est estimée à 41 individus et l'intervalle de confiance à 95% de cette estimation est [19-61].

Dans cette situation avec seulement deux sessions, nous voyons qu'un simple marquage de groupe (un point de peinture sur les tortues de la première session) aurait été suffisant. Mais cela n'est pas le cas lorsque l'on réalise un plus grand nombre de passages, ce qui est recommandé pour obtenir des estimations plus précises. Lorsque l'on réalise plusieurs passages, les individus doivent être tous identifiés individuellement de manière à construire leurs « histoires de captures ». Ces « histoires de captures » sont le point crucial des méthodes de CMR. Elles sont en fait un simple tableau résumant les données brutes obtenues sur le terrain. Ce tableau se construit avec tous les individus en ligne (un individu par ligne) et toutes les sessions en colonne (une colonne par session). Lorsqu'un individu a été vu/capturé à une session donnée, cette case individu-session est remplie avec un 1 ; lorsqu'il n'a pas été vu/capturé la case est remplie avec un 0.

Session	1	2	3	4	5	6	
individu #1	1	0	1	1	0	1	Avec 1 = observé 0 = non-observé
individu #2	0	0	0	1	1	1	
individu #3	0	1	1	1	0	1	
...							
individu #4	0	0	1	1	0	0	

Figure 17 : CMR. Matrice des histoires de capture par individu

Le fait que nous posions l'hypothèse que la population est close permet d'affirmer que tous les zéros dans ces histoires de capture correspondent au fait que l'individu était présent mais qu'il n'a pas été capturé/déecté. Le nombre de 1 dans le tableau divisé par le nombre total de cases du tableau (nombre de sessions fois le nombre d'individus) fournit donc une indication sur le taux de détection des individus. Ce taux de détection est usuellement appelé « taux de capture » en CMR. Dès lors que l'on a accès à ce taux de capture, on peut l'utiliser pour corriger le nombre d'individus vus, afin d'estimer le nombre d'individus réellement présents. En pratique, les calculs ne se font pas à la main mais à l'aide de logiciels, dont le plus connu, gratuit, nommé CAPTURE, implémenté dans un autre logiciel gratuit nommé MARK (malheureusement en anglais tout comme son manuel d'utilisation).

Le respect de l'hypothèse de population close est primordial tant pour le Lincoln-Peterson Index que pour les analyses avec le logiciel CAPTURE. En effet, on comprend aisément que si des individus marqués sortent de la zone d'étude ou que des non-marqués entrent dans la zone entre deux sessions de capture, la proportion d'individus encore marqués dans la population lors d'une session donnée ne correspond plus du tout à la proportion qu'il y aurait eu dans une population fermée ; nous ne pouvons donc plus estimer la taille de la population. En théorie, des échanges importants, aléatoires et non-directionnels (c'est à dire autant d'individus qui sortent que d'individus qui rentrent) nous permettent avec un protocole usuel d'obtenir la taille de l'ensemble de la population environnante. Mais dans la pratique il est rare que les échanges soient complètement aléatoires (par exemple les individus loin de la zone d'étude ont moins de chance de rentrer dans la zone que les individus en bordure de notre zone d'étude). Par

conséquent, il est très important de respecter au maximum cette hypothèse. Dans le cas de populations pour lesquelles les échanges sont très importants avec l'extérieur de la zone d'étude, alors les CMR ne seront pas pertinentes. Une solution possible est d'agrandir la zone de prospection (quitte à réduire l'effort par unité de surface) pour englober l'ensemble de la population environnante.

Le respect de l'hypothèse de **prospection aléatoire et complète** de la zone est aussi important. En effet, si vous ne prospectez pas systématiquement toute la zone, vous obtenez des zéros qui ne sont pas des non-détections d'individus présents mais des zéros qui sont des non-détections d'individus dans des zones non prospectées. Il ne s'agit pas du même phénomène et vos estimations sont de fait biaisées. La prospection aléatoire est importante car si vous prospectez mieux les zones sur lesquelles vous savez qu'il y a des individus marqués, vous aurez tendance à les recapter systématiquement et donc à croire que le taux de capture est fort, ce qui va en contrepartie sous-estimer la taille de la population (en fait vous estimez la taille de la population que vous « connaissez »). Si au contraire vous évitez les zones sur lesquelles vous avez déjà marqué des individus, vous aurez tendance à systématiquement capturer des nouveaux individus donc à croire que votre population est immense car vous obtiendrez un taux de recapture très faible. Il est parfois frustrant et donc difficile de respecter cette hypothèse de prospection aléatoire, surtout sur des espèces difficiles à détecter. Les naturalistes auront en effet souvent l'envie de trouver des individus et donc de retourner sur les emplacements où ils savent qu'il y a des individus. Pourtant vous aurez compris qu'il faut absolument éviter ce comportement. Pour cela, l'idéal est d'envoyer des prospecteurs différents à chaque session.

Dans la pratique il est important de noter que plus le taux de capture/recapture est fort, plus la précision sera grande. Il est aussi important de savoir que la pression de capture peut être différente selon les sessions, du moment qu'elle couvre l'ensemble de la zone de manière aléatoire. Ainsi des différences de météorologie, de qualité d'observateurs, etc. selon les sessions, peuvent être facilement prises en compte dans les modèles. Il suffira en effet de préciser au logiciel CAPTURE que l'on suppose des variations de la capture dans le temps. De fortes différences de taux de capture auront cependant tendance à réduire la précision des estimations.

Enfin les développements des méthodes de CMR permettent de gérer des problèmes plus complexes tels que l'hétérogénéité du taux de capture entre individus (par exemple s'il y a des individus globalement très actifs et d'autres globalement peu actifs, ils auront des probabilités probablement différentes d'être capturés), des différences de taux de capture entre groupes d'individus (par exemple entre sexes ou entre classes d'âge) mais aussi les problèmes dits de trap-dépendance. La trap-dépendance est un phénomène lié au fait que des individus déjà capturés n'ont pas la même probabilité d'être recapturés que des individus jamais capturés. C'est par exemple le cas lorsque l'on réalise du piégeage qui peut être traumatisant ; il est dans ce cas fort probable qu'un individu qui a été capturé une première fois sera plus difficile à recapter par la suite qu'un individu « naïf » jamais capturé. Dans le cas où l'on utilise des pièges avec appâts sur une espèce peu sensible à la capture et à la manipulation (ex : micromammifères), on peut au contraire avoir des individus qui reviennent plus facilement dans les pièges car ils savent qu'ils y trouvent de la nourriture facilement. Hétérogénéité de capture, trap-dépendance, variabilité de la capture dans le temps, sont des problèmes gérés par le logiciel CAPTURE, qui permet par ailleurs de tester s'ils existent dans votre jeu de données. Sachez cependant que si ces phénomènes existent, ils ont tous pour conséquence de biaiser vos estimations s'ils ne sont pas pris en compte dans le modèle et ils vont systématiquement baisser la précision lorsqu'ils sont pris en compte. Par ailleurs, il n'existe pas de méthode pour estimer les tailles de population lorsqu'il y a simultanément dans le jeu de données de la trap-dépendance, de l'hétérogénéité du taux de capture et de la variabilité dans le temps de ces taux de capture.

5.4. Les méthodes d' « occupancy » ou présence-absence

Ces méthodes dites d'«occupancy », traduites par « présence-absence » en français, sont toutes récentes (première publication de cette méthode en 2003). Elles partent du constat que tout comme les individus ne sont pas détectables avec 100% de chance sur un site, les espèces ne le sont pas non plus. Lorsque l'on s'intéresse par exemple à la répartition d'une espèce dans l'espace, on va généralement se pencher vers les atlas de répartition. Ces atlas présentent généralement uniquement les présence-absences des espèces. Pourtant dans ces atlas, un zéro dans un carré peut être une absence réelle mais peut aussi être une présence manquée par les observateurs. Pour certaines espèces communes, il y a peu de chance que les atlas soient biaisés. Mais pour les espèces rares ou les espèces très difficiles à détecter, il est fortement probable que des zéros soient en fait des sites sur lesquelles l'espèce n'a pas été vue malgré sa présence.

Lorsque l'on veut conduire des études à de larges échelles ou sur des espèces rares ou difficiles à détecter, il va être impossible d'utiliser des méthodes comme le « *Distance Sampling* » ou les CMR, qui demanderaient des efforts considérables pour obtenir suffisamment de données nécessaires à l'obtention d'une bonne précision. Pourtant il faudrait s'assurer de corriger les répartitions observées pour les problèmes de détection.

Pour cela les méthodologistes ont utilisé tout le corpus théorique et les outils développés pour les CMR, mais en utilisant des données issues de suivis de sites au lieu de suivis d'individus. La méthode consiste à sélectionner un grand nombre de sites à suivre et de les prospector un certain nombre de fois au cours d'une saison de terrain (en pratique 2 visites par site peuvent suffire mais souvent un minimum de 3 sera nécessaire). Tout comme pour les CMR, les sites doivent être considérés comme « clos », c'est-à-dire qu'ils ne sont pas abandonnés ou colonisés au cours de la saison de terrain. Ainsi ils sont soit occupés pendant tout le temps du terrain, soit inoccupés mais ne changent pas d'état. Dès lors, on peut construire des histoires de détection de l'espèce équivalentes aux histoires de capture des CMR, avec cette fois ci les sites en lignes et les passages en colonne. Tout comme pour les CMR, les zéros qui sont observés sur des sites pour lesquels il y a au moins un 1 signifient que l'espèce est présente (le 1) mais qu'elle a été ratée (les 0). On peut donc estimer à partir de ce type de données la probabilité de détection de l'espèce. Ceci peut se faire avec le logiciel gratuit mais lui aussi en anglais nommé PRESENCE.

SITE	SAISON			
	1	2	...	T
1	101	000	...	000
2	111	111	...	111
3	000	000	...	000
.				
.				
.				
S	101	010	...	001

Figure 18 : Présence-absence. Matrice des histoires de capture par espèce

Ces méthodes ont été largement développées en peu de temps. Elles permettent aujourd'hui d'estimer la probabilité d'occupation moyenne d'une espèce sur une zone, mais aussi sa probabilité de détection. Dès lors que l'on dispose de cette probabilité de détection, nous pouvons en déduire la probabilité qu'un site sur lequel l'espèce n'a jamais été vue soit pourtant occupé. En effet, si la probabilité de détection de l'espèce est faible, la probabilité qu'elle ait été ratée plusieurs fois de suite sur un site sur lequel elle est pourtant présente peut être forte. L'intérêt de cette méthode est donc de fournir pour tous les sites « négatifs » (c'est-à-dire où l'espèce n'a jamais été détectée) un degré de confiance dans l'absence de l'espèce.

10010010	site occupé mais espèce souvent ratée
11111011	site occupé et espèce souvent détectée
00000000	Site occupé et toujours raté ou site non-occupé par l'espèce ?

Figure 19 : Présence-absence. Exemples types d'histoires de capture par espèce

Par ailleurs ces méthodes permettent aujourd'hui de tester des effets de variables externes sur la probabilité d'occupation et sur la probabilité de détection des sites. Par exemple, on peut examiner si la probabilité qu'un site soit occupé est dépendante du type de végétation, d'un mode de gestion ou de la présence d'eau, etc. De même on pourra tester sur la probabilité de détection des effets de la météo, de la végétation, de l'observateur, etc. Un sous-produit extrêmement intéressant de ces tests est que, si on arrive à mettre en évidence un facteur de l'environnement qui explique bien la probabilité d'occupation de l'espèce, alors on peut estimer la probabilité qu'un site soit occupé sans le prospector. Par exemple, si l'on montre que les sites avec de l'eau ont 90% de chances d'être occupés par la Tortue d'Hermann dans le Var, alors n'importe quel site avec de l'eau dans le Var a 90% de probabilité d'être occupé. On pourrait donc à partir de ce genre de résultat projeter sur l'ensemble de la plaine des Maures la probabilité d'occupation des sites à partir d'une simple cartographie de la présence d'eau (il s'agit ici d'un exemple fictif mais ceci fonctionne particulièrement bien pour l'impact des incendies sur cette espèce).

En pratique, il est possible de quitter le site dès que l'espèce a été détectée, ce qui permet de maximiser le terrain et de visiter un maximum de sites. Par contre, il est très important de retourner sur les sites positifs un certain nombre de fois car c'est la proportion de 0 et de 1 sur ces sites qui va fournir l'information pour estimer la probabilité de détection. Il est aussi extrêmement important de continuer à prospector les sites « négatifs » avec le même nombre de sessions que les sites positifs. Ceci peut être très frustrant et très démotivant, mais cette rigueur est cruciale pour obtenir des informations précises. Il peut même être recommandé d'augmenter le nombre de visites sur les sites négatifs dans le cas d'espèces difficiles à détecter, pour augmenter le degré de confiance dans l'absence de l'espèce.

Tout comme pour les CMR, plus la probabilité de détection de l'espèce sera élevée, plus la précision sur l'occupation sera forte. Il est malheureusement impossible de fournir une règle sur le nombre de sites à suivre et le nombre de visites à effectuer car cela dépendra fortement de cette probabilité de détection. Il existe un petit logiciel gratuit, nommé GENPRES (toujours en anglais) qui permet de tester différents scénarii à partir du moment où vous avez déjà une idée de la valeur de la détection et du pourcentage de sites globalement occupés sur la zone que vous voulez étudier. Ce logiciel construit ce que l'on nomme des « tests de puissance », dont nous parlons plus bas dans ce document.

Enfin il est à noter que les développements plus récents de ces méthodes permettent aujourd'hui d'estimer non plus simplement des probabilités d'occupation par une espèce mais les abondances moyennes sur les sites à partir soit de simples données de présence-absence, soit à partir de données de comptages. Pour cela, les protocoles sont les mêmes (visites répétées sur des sites) mais posent des hypothèses statistiques supplémentaires. Ces hypothèses ne sont pas toujours respectées et malheureusement nous ne disposons pas encore, contrairement aux CMR, d'outils pour tester si oui ou non les données collectées les respectent. Par ailleurs, les intervalles de confiances des abondances estimées par ces méthodes sont souvent très larges. Ces protocoles nécessitent par conséquent le suivi d'un grand nombre de sites au cours de nombreuses sessions. Nous ne détaillerons donc pas plus ces aspects ici, mais sachez que cette possibilité existe et que ces méthodes sont en train de fortement s'améliorer. Il vous est donc recommandé de consulter des spécialistes des méthodologies, dans le cas où vous voudriez vous orienter vers ces méthodes pour lesquelles nous avons encore peu de recul.

5.5. Choisir une méthode de suivi

Il existe donc quatre grandes classes de méthodes pour suivre les populations d'espèces. Ces quatre méthodes posent toutes des hypothèses particulières, qui peuvent d'emblée limiter leur utilisation pour telle ou telle espèce dans tel ou tel contexte. Pourtant, du fait qu'elles soient relativement flexibles, plusieurs peuvent dès lors être envisagées pour une situation donnée. Il conviendra alors de déterminer celle qui sera optimale, c'est-à-dire celle qui fournira des estimations non-biaisées avec une précision suffisante pour la question posées et surtout avec un minimum de budget temps.

Comment choisir la méthode ? Les graphiques présentés ci-dessous montrent la complexité de déterminer avec certitude quelle est la méthode optimale. En effet, cela dépendra de la détectabilité de l'espèce étudiée, de la surface de la zone étudiée et du budget disponible, mais aussi de la sensibilité de l'espèce en termes de dérangement, de ses capacités de déplacement, de l'expérience des observateurs recrutés, de la nécessité de répéter le suivi sur plusieurs sites ou dans le temps, etc. Ce choix de la méthode est donc multifactoriel et il est impossible de donner une règle simple pour le réaliser.

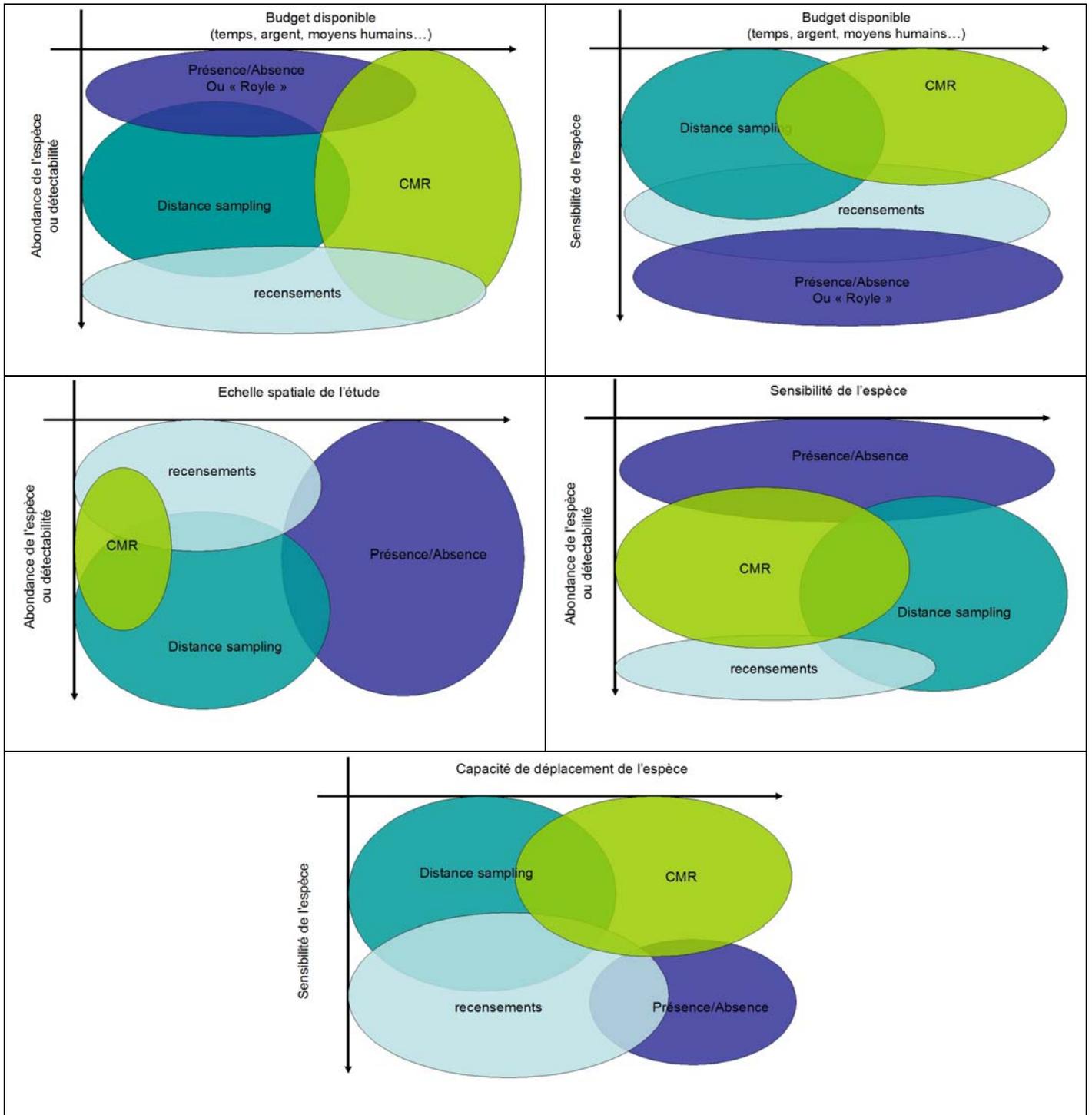


Figure 20 : Positionnement des différentes méthodes en fonction de divers critères de choix

Commentaire : cette figure, présentant un certain nombre de cadres, a pour volonté d'illustrer le fait que le choix d'une méthode de suivi d'espèce est multifactoriel. En effet, si l'on trace un graphique avec deux axes représentant certaines contraintes des protocoles (contraintes de budget, de sensibilité de l'espèce, d'abondance, de détection, de capturabilité, d'échelle spatiale...), nous pouvons relativement aisément définir les espaces approximatifs de combinaison de ces contraintes (les patates), qui conduisent à un choix relativement évident d'une méthode. Notons cependant que le choix peut être multiple (chevauchement des patates). Par ailleurs, la chose la plus importante à retenir est que lorsqu'on change les axes de contraintes, l'organisation des patates est fortement modifiée. Ainsi les contraintes à explorer sont très nombreuses et dépendent fortement de la question posée, de l'espèce, du site, de l'expérience des observateurs et du budget disponible. Dans le détail, le nombre d'axes à déterminer pourrait rapidement avoisiner la dizaine voire la vingtaine. Cette multiplicité des contraintes ne permet malheureusement pas de déterminer quelle sera à coup sûr la méthode pertinente pour votre étude. Le problème serait le même si l'on cherchait à construire un tableau, sous forme de clef dichotomique.

Ceci illustre aussi bien la limite que peut poser la mise en œuvre de **protocoles « recettes de cuisine »** à la suite de la lecture d'un article sur la même espèce par exemple. En effet, ni votre question, ni votre budget, ni votre expérience, ni votre site ne sera le même que celui de l'article retenu. Dans ce contexte, l'utilisation du même protocole ne vous assurera en aucun cas d'avoir les résultats escomptés. Dans le même ordre d'idée, la mise en œuvre de **protocoles nationaux à l'échelle locale** n'assure généralement pas d'avoir de bons résultats. En effet, les protocoles nationaux sont construits sur le principe de trouver les plus petits dénominateurs communs à l'ensemble des opérateurs de terrain. Ces protocoles posent aussi des questions nationales, avec des degrés de précision souvent très différents de ceux que vous désirez. Enfin ces protocoles sont menés sur de très grandes surfaces et sur des tailles d'échantillon très grandes, ce qui permet d'avoir une précision intéressante. Ceci ne signifie pas de manière manichéenne que la technique mise en œuvre par le protocole national ne sera pas pertinente pour votre situation, mais cela reste à explorer et à comparer avec des alternatives. Il en est de même pour les stratégies d'échantillonnage.

Pour résumer les grandes tendances, on peut toutefois dégager quelques grandes généralités :

Les recensements : ils doivent se limiter aux espèces détectables à 100% donc plutôt de tailles importantes, peu mobiles et sur des surfaces de tailles limitées.

Le « Distance Sampling », bien adapté aux espèces dont la détection n'est pas de 100% mais nécessite malgré tout un nombre d'observations relativement important pour être exploitable. A utiliser par conséquent sur des espèces relativement abondantes. Ne nécessitant pas de manipulations, elles peuvent être utilisées sur de nombreuses espèces et sur des surfaces relativement larges. Elles posent cependant des problèmes pour les espèces cryptiques (détection inférieure à 100% sur la ligne ou sur le point) ou pour les espèces qui sont sensibles à la présence de l'observateur (déplacement en réponse à l'observateur). Enfin elles sont mal adaptées pour les espèces se déplaçant fortement (ex: papillons) ou de manière très discrète (ex : amphibiens).

Les « captures-marquage-recapture » : il s'agit de la méthode qui fournit généralement la meilleure précision sur les estimations des tailles de population. Malheureusement, étant relativement lourde à mettre en œuvre, son utilisation est généralement réservée à des programmes de suivis d'espèces, par des gestionnaires et/ou chercheurs, dotés de moyens à la fois en temps et en budget (ex : programme LIFE, espèce bénéficiant d'un plan national d'action...). En effet, cette méthode impose généralement la capture et le marquage physique des individus, ce qui peut poser des problèmes éthiques ou plus simplement d'autorisation réglementaire. Elle impose aussi de travailler sur des espèces capturables et marquables, ce qui n'est pas toujours évident. Enfin, elle nécessite des efforts de terrain importants du fait de la nécessité de recapter les individus, si possible plusieurs fois et en nombre suffisant, donc se limite à des études menées généralement sur des petites surfaces. Enfin, les estimations des CMRs étant très sensibles au fait que la population soit statistiquement « close », cette méthode se limite à des espèces qui sont assez fidèles à leur site au cours de la saison de terrain.

Les méthodes de « présence-absence » : ces méthodes récentes ne fournissent pas d'estimation d'effectifs mais des informations sur l'occupation de l'espace (expansion ou contraction de l'occupation par exemple). Elles sont donc particulièrement utiles pour les espèces rares pour lesquelles on s'intéressera plus à la répartition qu'au nombre d'individus. L'utilisation de ces méthodes peut ne pas permettre de mettre en évidence des évolutions importantes de densités (par exemple si les tailles de populations baissent de manière générale mais que la surface occupée reste quant à elle stable). Elles nécessitent aussi souvent de travailler sur un grand nombre de sous-unités d'échantillonnage, de manière répétée dans le temps et sur lesquelles les populations seront fidèles au cours d'une saison de terrain (hypothèse de site « clos »). Elles ne seront donc pas très adaptées pour des espèces qui « colonisent » et « abandonnent » les sites rapidement par exemple.

6. Mettre en évidence un impact : Le BACI

Usuellement, les protocoles de suivis d'espèces ont vocation comme leur nom l'indique à être menés au minimum sur le moyen terme et souvent sur le long terme, afin de mettre en évidence des changements dans la taille des populations étudiées. Ces suivis peuvent avoir pour simple objectif de vérifier l'état des populations (croissance, décroissance) mais ils peuvent aussi avoir comme ambition de vérifier la pertinence de mesures de gestion ou au contraire l'impact négatif de certaines perturbations, d'origine anthropique ou non.

La description simple de l'évolution d'une taille de population dans le temps ne nécessite pas de précaution particulière en terme de protocole, autres que celles vues précédemment sur la détection et l'échantillonnage. Au contraire, la mise en évidence de l'impact d'une gestion ou d'une perturbation impose quant à elle des contraintes fortes que nous détaillons ici. Ces contraintes se regroupent sous le terme de **BACI** pour « **Before-After Control Impact** » qui sont une classe particulière de protocoles d'échantillonnage spécifiquement conçus pour étudier l'impact de perturbation ou de mesures de gestion.

Prenons pour être plus parlant un exemple : nous suivons une population dans une zone qui est soumise à une gestion conservatoire. La question est donc de mettre en évidence un effet positif des mesures de gestion prises. La taille de la population est estimée tous les ans pendant quelques années sur le site géré. Puis la mesure de gestion conservatoire est mise en œuvre et la population continue à être suivie sur le long terme. Si la population augmente, alors la conclusion immédiate sera que l'action de gestion conservatoire a été pertinente et efficace. Au contraire, si la population est stable voire décroît, nous rejeterons la conclusion selon laquelle notre mesure de gestion est efficace et stopperons son utilisation.

Pourtant, nous ne pouvons pas tirer ces conclusions. En effet, pour pouvoir faire la démonstration claire que l'action de gestion a ou n'a pas de répercussions sur la population, il est indispensable de la comparer avec un site qui n'est pas géré. On appelle ces sites des « témoins » car ils sont le témoin de ce qui se passe sans action. Si les deux populations, du site géré et témoin, présentent globalement les mêmes patrons d'évolution des tailles de population, alors nous pouvons douter que notre action de gestion a eu un impact quelconque sur la population. Si au contraire le site témoin est stable et que la population du site géré est en croissance, alors nous pouvons suspecter que notre action de gestion est efficace. De la même manière, si la population du site témoin décroît alors que la population gérée reste stable, alors nous sommes tentés de conclure que notre action de gestion est efficace.

Cependant, cette situation est encore trop simpliste pour pouvoir mettre en évidence sans ambiguïté l'impact des mesures de gestion. En effet, il est tout à fait possible qu'un site témoin et qu'un site géré évoluent dans des directions opposées, par la simple fait du hasard ou pour d'autres raisons que nous n'aurions pas identifiées mais qui sont indépendantes de la mesure de gestion (ex : pour un oiseau insectivore, forte variabilité locale de la biomasse en insectes-proies induite par des conditions microclimatiques particulières). La solution pour contrer cet argument est de multiplier les expériences. En effet, si le hasard peut expliquer facilement des différences de dynamiques sur deux sites, l'un témoin et l'autre géré, il ne le pourra pas par exemple sur 10 sites. Ainsi, si 5 sites gérés sont en croissance et que 5 sites témoins sont stables ou en décroissance, nous pourrions conclure que la mesure de gestion a été efficace. Bien évidemment, nous ne pouvons travailler ici à budget constant, car le suivi de 8 sites par exemple imposera évidemment 4 fois plus d'effort que le suivi de 2 sites. Cependant, c'est la seule solution pour montrer clairement que les mesures de gestion ont un impact. Sans cet effort, aucune conclusion fiable ne pourra être tirée des données collectées qui sont donc des données perdues et des années perdues.

Il y a des situations où l'on souhaite montrer l'impact d'une perturbation ou d'une mesure de gestion *a posteriori*, c'est-à-dire sans disposer des données antérieures. C'est souvent le cas lorsque l'on étudie des perturbations rares et imprévisibles, telles que les incendies ou les

pollutions accidentelles, pour lesquelles il est rare que les sites altérés soient les mêmes que ceux qui auraient éventuellement été étudiés précédemment. Dans cette situation, un protocole usuel consiste à mesurer l'abondance sur un ou idéalement plusieurs sites témoins et à la comparer à celle mesurée sur le site perturbé. Lorsque tous les sites témoins ont des abondances plus faibles ou plus fortes que le site perturbé, on conclue généralement que la perturbation (dégradation ou amélioration) a eu un impact. Pourtant, pour pouvoir tirer cette conclusion, il faut mesurer le degré de variabilité naturelle de l'abondance entre sites et si possible dans le temps. En effet, si l'abondance est extrêmement variable selon les sites et/ou dans le temps, alors il est tout à fait probable que le site perturbé soit différent des autres sites simplement du fait du hasard. La multiplication des sites « témoins » et « altérés » permet normalement de tirer plus facilement des conclusions sur l'impact de perturbations.

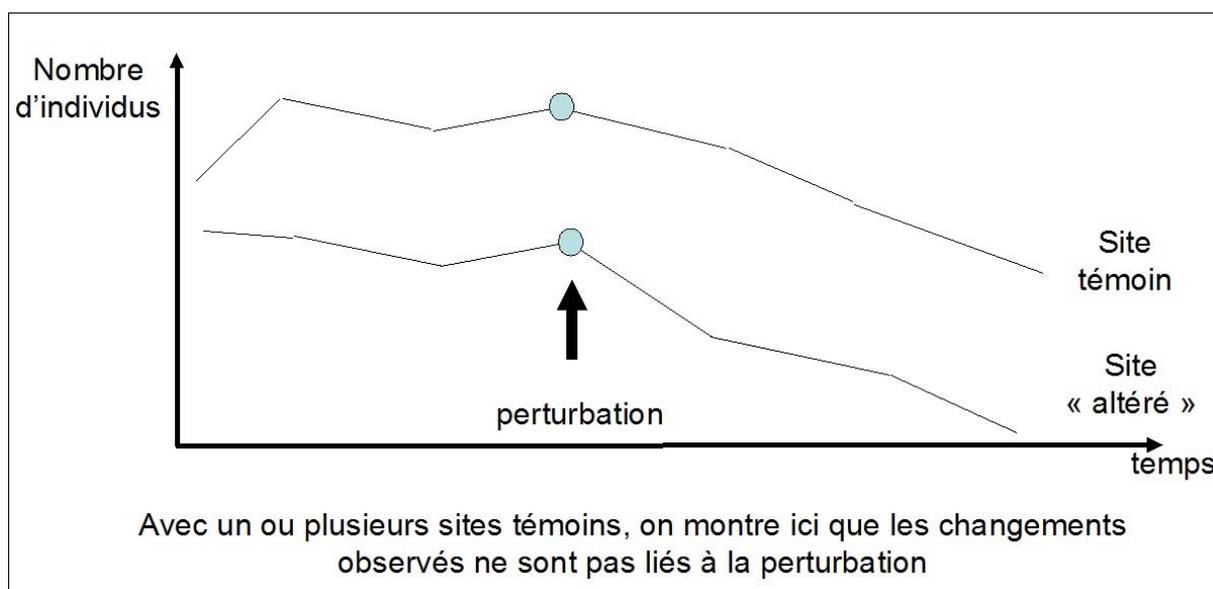


Figure 21 : Protocole BACI. Comparaison d'un site témoin et d'un site altéré.

En conclusion, un protocole BACI, c'est-à-dire un protocole idéalement construit pour mettre en évidence l'effet d'une perturbation/modification sur les populations d'une espèce, impose de suivre plusieurs sites, des sites « témoins » et des sites « altérés/modifiés », de manière répétée dans le temps, avant et après la perturbation/modification. Tout protocole qui négligera une partie de ces contraintes sera sujet à caution.

7. Stratégie générale pour construire un protocole

Commençons par un point essentiel que nous n'avons pas abordé pour le moment : celui de la **construction de la question** qui vous intéresse. Cette étape est souvent un préalable même à la mise en route d'une réflexion sur la nécessité d'un suivi. Pourtant, le fait qu'elle précède souvent la mise en œuvre d'un protocole ne signifie pas nécessairement que cette étape est un compartiment indépendant de la construction en soit du protocole. En effet, dans les différentes étapes que nous verrons ci-dessous, la définition précise de la question d'intérêt est toujours présente en toile de fond. Ceci signifie aussi que cette question doit toujours rester présente à l'esprit. Un protocole est toujours construit pour répondre à une question donnée. Il est pourtant fréquent que la volonté soit, soit de répondre à plusieurs questions identifiées avec un même protocole, soit qu'on utilise des données collectées pour répondre à une question afin d'apporter des réponses à une autre question qui se serait posée ultérieurement. Ces deux cas sont à bannir car il est très rare que les résultats soient au rendez-vous. Dans le cas où vous voulez répondre à plusieurs questions avec le même protocole, le risque est clairement de ne pas porter assez d'effort sur chaque question ou que le protocole soit un compromis désastreux entre différentes contraintes. Dans le cas où l'on veut répondre à des questions nouvelles, il est rare que le résultat soit concluant. Toujours est-il que dans les différentes étapes de mise en œuvre d'un protocole, il peut être pertinent de réexaminer la question initiale, notamment lorsque les résultats des pré-études et des tests de puissance démontrent par avance des difficultés à répondre avec certitude à la question posée initialement. **Un aller-retour nécessaire entre question-méthode-protocole-données-analyse est souvent très riche pour faire avancer la connaissance globale d'un système.**

La construction d'un protocole pertinent ne peut faire l'économie d'une bonne connaissance préalable de la biologie/écologie de l'espèce étudiée, mais aussi de sa répartition et de son abondance sur le site. Cela peut paraître paradoxal dans le sens où le protocole mis en place a justement pour objectif de connaître l'abondance de l'espèce. Pourtant, on a vu tout au long de ce document qu'il fallait disposer d'information sur la répartition spatiale pour construire un plan d'échantillonnage adapté, mais aussi sur les abondances pour déterminer quelle méthode d'analyse de données serait la plus pertinente.

Synthétiser les informations existantes : une des questions centrales et des plus complexes à laquelle le montage d'un protocole doit s'atteler est celle de l'effort de terrain réalisable, mais aussi sa répartition. Pour cela, la première étape est bien entendu de collecter toutes les informations disponibles sur l'espèce dans la région étudiée. Recherche bibliographique, discussions avec les spécialistes, discussions avec les naturalistes de la région sont les trois grands axes pour acquérir des connaissances générales sur le fonctionnement des populations de l'espèce étudiée d'une manière générale, mais aussi des connaissances sur sa répartition et ses abondances locales.

Conduire des « pré-études » : lorsque l'espèce a été peu étudiée en amont sur le site et/ou que son écologie est mal connue, il conviendra d'envisager de décaler le protocole définitif d'un an au moins, le temps d'acquérir le minimum de connaissances nécessaires à la construction d'un protocole pertinent. En fait, il s'agit de monter un premier protocole, plus ou moins léger selon les besoins dont l'étude dispose, à mener la première année pour se fixer les idées. On parle alors de « pré-étude ». Cette pré-étude peut se construire sur un protocole du même type de celui envisagé sur le plus long terme, mais il peut aussi être différent. Par exemple, nous pouvons envisager un échantillonnage avec de très nombreuses sous-unités, mais prospectées assez brièvement pour avoir une vision globale de la présence sur les sites. Au contraire, on peut envisager de construire un protocole proche de celui qui sera mené par la suite, afin d'obtenir des premières estimations et de calculer quel effort il faudra fournir en terme d'échantillonnage pour obtenir le degré de précision désiré (voir ci-dessous « test de puissance »).

Construire des tests de puissance : les tests de puissance sont un outil qui permet d'optimiser le temps de terrain. Leur usage est encore assez limité en écologie bien qu'ils apportent des informations essentielles dans les étapes de construction des protocoles. Le test de puissance consiste à utiliser les informations déjà disponibles sur l'espèce, soit par la bibliographie mais idéalement les données issues de pré-étude, pour simuler/générer des faux jeux de données avec des échantillons de tailles croissantes. Cette opération est répétée de très nombreuses fois (souvent 1000 fois pour chaque taille d'échantillon, voire plus) et les jeux de données ainsi construits sont analysés avec les méthodes prévues par l'étude. Les résultats de ces analyses permet d'estimer quelle sera la précision que l'on obtiendra en fonction de l'effort de terrain que l'on souhaite réaliser. Ces tests permettent donc d'optimiser le temps de terrain en évitant d'en faire trop si l'on s'aperçoit que la précision est bonne pour un échantillon de taille inférieure à ce qui était prévu. Ils permettent aussi de ne pas prévoir une taille d'échantillon trop petite pour la précision. Réalisés en amont des demandes de financements ou lors de commandes d'études, ils peuvent permettre de montrer avant même le début de l'étude que la précision ne sera par exemple pas suffisante au regard du budget alloué pour mettre en évidence les variations de densités ou d'autres paramètres. Malheureusement la construction de tests de puissance est une opération technique complexe qui nécessite des connaissances assez poussées en programmation et en mathématique. Notre conseil pratique est donc de consulter des spécialistes qui pourront vous aider à construire ces tests, au moins sur les thématiques pour lesquels les enjeux en termes de conservation ou financiers sont importants (espèces patrimoniales, études sur le long terme, protocoles très lourds...).

Analyser les données régulièrement : un aspect très important mais très rarement mis en œuvre lorsqu'on développe un suivi de population est de bien analyser les données au fur et à mesure qu'elles sont collectées. Il peut s'agir d'une analyse au cours de la saison de suivi ou d'une année sur l'autre, les deux apporteront des informations importantes notamment en l'absence de tests de puissance. Par exemple, l'analyse des données pendant la phase de collecte est souvent peu coûteuse en termes de temps (faire quelques graphiques, quelques comptages, voire quelques analyses statistiques) mais peuvent être cruciales. Ainsi il n'est pas rare de faire trop de terrain par rapport à l'objectif déterminé. On sait que l'intervalle de confiance d'une estimation va décroître de manière exponentielle négative avec l'effort, c'est-à-dire que l'ajout d'une sous-unité d'échantillonnage va réduire fortement les intervalles de confiance au début de l'étude mais de plus en plus faiblement au fur et à mesure que leur nombre total s'accroît. Il est donc parfois inutile de poursuivre les prospections et il vaudrait mieux investir le temps restant dans une étude complémentaire ou sur un tout autre sujet. Au contraire, il n'est pas rare d'arrêter l'étude un tout petit peu trop tôt. L'examen des intervalles de confiance après chaque nouvelle session de terrain est donc très riche en enseignement et permet d'optimiser ses efforts. Enfin, l'examen des données au fur et à mesure peut informer sur l'adéquation du protocole. Par exemple, en intersaison il n'est pas rare que les conditions changent sur le terrain. Un protocole qui était pertinent les premières années de l'étude peut ne plus l'être. Attendre la fin de l'étude, notamment quand elle est menée sur plusieurs années, pour analyser les données revient souvent prendre un risque inutile car le temps d'analyse des données est souvent bien moindre que celui nécessaire sur le terrain pour collecter les données.

Ajustement/ modification d'un protocole : lorsque l'on constate en cours d'étude que le protocole n'est pas, ou n'est plus optimal, soit parce que la population et le site ont changé, soit parce que les questions ont changé ou soit parce que les méthodes disponibles ont évolué, alors il est possible de modifier le protocole. Il faudra cependant tenter de conduire pendant un certain temps l'ancien et le nouveau protocole en parallèle, de manière à déterminer si les données collectées par les deux protocoles peuvent être comparées. Ceci est particulièrement important dans le cadre de suivis à long terme, afin de ne pas perdre la possibilité de comparaison entre des séries de données antérieures dont la collecte a parfois été extrêmement coûteuse.

Enfin, il est important pour conclure ce document de rappeler que les méthodes de suivi de population ont très largement évolué au cours des trente dernières années, tant d'un point de vu théorique (CMR, Distance Sampling...) que pratique (transpondeurs, télémètres lasers...).

Ces méthodes et techniques sont en constante évolution encore aujourd'hui, comme le montre l'émergence toute récente des méthodes dites de présence-absence. Par conséquent, il est important d'interagir avec des spécialistes à la fois des espèces étudiées mais aussi des méthodes de suivis de population, qui seront au courant des dernières avancées à vous conseiller.

8. Cas d'école

8.1. Cas 1 : mauvaise prise en compte de l'écologie de l'espèce

Objectif initial : Estimation d'une taille de population de Cistude d'Europe (*Emys orbicularis*) sur un étang de grande taille.

Protocole choisi : L'estimation de cette taille de population a été réalisée avec un protocole de Capture-Marquage-Recapture à l'aide d'une dizaine de nasses cylindriques posées tout au long de l'été sur le pourtour de l'étang, espacées régulièrement et relevées tous les jours.



Résultat : L'examen des données de CMR issues de ce protocole montre qu'il existe une forte trap-dépendance positive. Il s'agit d'un biais suggérant que les individus déjà capturés ont une plus forte probabilité d'être recapturés que les individus non capturés auparavant. Une estimation de taille de population est obtenue à l'aide du logiciel CAPTURE en prenant en compte ce problème de trap-dépendance. Cependant la confiance dans les résultats est faible en l'absence d'explication de la provenance de ce biais et donc de son impact. Les intervalles de confiance avec ces méthodes sont très larges et il semble que la taille de la population soit largement sous-estimée par rapport à l'attendu des naturalistes connaissant bien la zone. La pertinence des résultats est donc limitée.

Problème identifié : Une réflexion approfondie sur la cause de cette trap-dépendance suggère que le protocole est en cause, tout du moins qu'il n'est pas bien adapté à l'espèce. En effet, les cistudes d'Europe sont des animaux relativement territoriaux. La pose de pièges pour leur capture est pertinente et semble être la bonne méthode pour estimer la taille de cette population. Cependant le problème provient du fait de laisser les pièges toujours au même emplacement d'une session de capture à l'autre. En effet, de par la territorialité des animaux et du faible nombre de pièges en relation avec la taille de l'étang, chaque piège est placé sur un territoire (ou à l'intersection de deux territoires au mieux). Par conséquent, les pièges ont tendance à systématiquement capturer les mêmes individus, à savoir ceux dont le territoire inclus un piège. Tous les individus dont les territoires n'incluent pas de pièges ne sont quant à eux jamais, ou très rarement capturés. Le protocole ainsi constitué permet d'estimer la taille de la population dans le rayon d'action des pièges et non la taille de la population sur l'ensemble de l'étang.

Solution : deux solutions peuvent exister dans cette situation de piégeages d'animaux *a priori* territoriaux. La première consiste à placer un grand nombre de pièges de manière à ce que tous les individus soient potentiellement capturables. Cependant cette solution impose des contraintes fortes (plus de matériel, plus de temps pour relever les pièges...) qui ne sont pas toujours gérables. La deuxième solution consiste à déplacer systématiquement les pièges et cela de manière totalement aléatoire sur le pourtour de l'étang (idéalement sur l'ensemble de la zone potentiellement occupée par l'espèce). Ce placement aléatoire des pièges permet ainsi de respecter l'hypothèse d'indépendance des événements de capture, hypothèse posée dans les méthodes de type CMR (elle mime en quelque sorte une prospection aléatoire du site d'une session à l'autre telle qu'elle pourrait être réalisée sur une espèce terrestre comme la tortue d'Hermann par exemple). Cependant là encore, le temps de travail doit être revu à la hausse puisque le déplacement systématique des pièges est une opération couteuse en temps.

Point important à retenir : il faut en général bien connaître la biologie de l'espèce étudiée ainsi qu'avoir un minimum de recul sur la zone d'étude pour mettre en place un protocole pertinent. Le choix d'une méthode statistique et du protocole de terrain qui va permettre de

collecter les données doit se faire en examinant l'adéquation entre les hypothèses de cette méthode et la biologie de l'espèce considérée.

8.2. Cas 2 : intérêt d'une pré-étude

Objectif initial : Proposer un protocole pertinent pour étudier l'évolution des populations de Lagopède alpin dans les Pyrénées.

Protocole choisi : Nous avons vu dans le texte principal que cette espèce a fait l'objet d'un suivi annuel depuis plusieurs années dans les Pyrénées. La méthode mise en œuvre consistait à estimer sur des points d'écoute le nombre de mâles chanteurs, sur des placettes fixes dans le temps, situées dans la zone centrale de Pyrénées.



Résultat : Les analyses semblent montrer que la population de lagopèdes alpins est stable dans le temps depuis une vingtaine d'années.

Problème identifié : Les résultats de cette étude posent un certain nombre de problèmes.

Le premier étant que l'échantillonnage ne couvrant pas l'ensemble des Pyrénées, les conclusions sont limitées à la zone centrale pour laquelle il semble que la dynamique soit différente des zones périphériques. En effet, en zone centrale les populations semblent stables alors qu'elles semblent baisser fortement en périphérie (résultat souvent classique pour des populations en marge d'aire de répartition donc plus sensibles aux perturbations et modifications de l'habitat). Mais les populations périphériques ne sont pas suivies donc nous n'avons pas vraiment d'argument pour le démontrer. Il s'agit ici clairement d'un problème de stratégie d'échantillonnage.

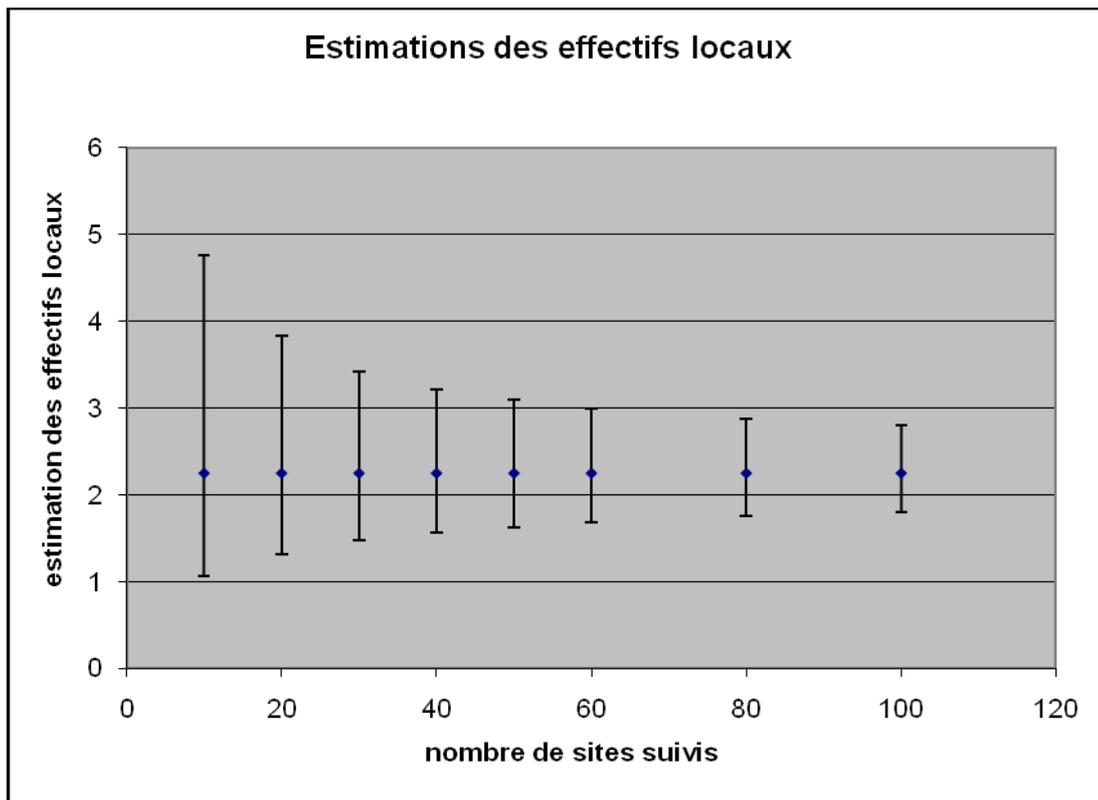
Le deuxième problème provient du fait que le nombre de mâles chanteurs compté par les observateurs est une valeur fixe, sans intervalle de confiance, qui ne permet donc pas d'appréhender le degré de précision de l'étude.

Enfin le troisième problème est que la détection de cette espèce n'est pas connue (détection ici étant un mélange entre l'émission du chant et sa détection réelle pour l'auditeur). Il s'agit ici d'une limite liée à la méthode choisie pour la collecte des données.

Solution : Le suivi d'une telle espèce étant assez sensible et aussi potentiellement couteux, notamment du fait des difficultés d'accès aux sites, il a été choisi de proposer une « pré-étude » sur une année, sur plusieurs sites, afin de déterminer quel serait le protocole optimal et quel en serait la puissance. Etant donné les connaissances sur l'espèce et sur le terrain, il a été retenu la méthode proposée par Royle, qui permet d'estimer les abondances sur des points d'écoute et leur intervalle de confiance à partir de comptages répétés au cours d'une saison (raffinement des méthodes de « présence-absence »). Le protocole retenu a été conduit sur 3 grands sites totalisant une trentaine de points d'écoute visités 3 fois.

L'analyse des données collectées permet de déterminer qu'une demi-heure d'écoute par point est suffisante, alors qu'auparavant les observateurs restaient 1h30 sur le site. Par ailleurs, les résultats montrent clairement que la détection (ici l'activité de chant des oiseaux) n'est pas de 100% et que même en conditions météorologiques idéales elle est variable selon les jours.

Les données collectées sont par la suite utilisées pour des simulations dont le principe ici est de simuler aléatoirement un jeu de données plus conséquent, à savoir constituer des jeux de données avec plus de sites suivis. Le principal résultat de ces simulations est présenté sur la figure ci-dessous.



Cette figure présente l'évolution des intervalles de confiance que l'on peut attendre en augmentant l'effort (le nombre de points d'écoute). Sachant que le budget disponible permet de prospecter environ 20 points tous les ans, nous pouvons d'ores et déjà remarquer qu'avec un suivi annuel de 20 points, les intervalles de confiance sont bien trop larges pour espérer détecter des baisses même fortes de la population. La recommandation est alors de prospecter au cours de 3 ans successifs des séries de 20 sites différents, ce qui fournit un échantillon de 60 sites au final. L'opération est dès lors répétée dans le temps. Au bout de 6 années de terrain, nous disposons donc de 2 visites sur chaque point, espacées de 3 ans. Nous disposons aussi de 60 sites sur 2 années, soit d'un échantillon de 120 points d'écoute. Avec un tel échantillon, on voit que les intervalles de confiance pourraient permettre de détecter des baisses même relativement faibles.

Point important à retenir : Bien faire attention à l'échantillonnage de la zone d'étude sur laquelle on veut extrapoler les résultats. Ne pas hésiter à mettre en œuvre des pré-études dont les résultats serviront à construire des tests permettant de mieux calibrer le protocole et d'avoir une première estimation de la précision à l'issue de l'étude.

8.3. Cas 3 : choix entre « Distance Sampling » et CMR

Objectif initial : Estimer la taille de la population de Tortue d'Hermann dans la plaine des Maures (Var).

Protocole choisi : Une sélection aléatoire de transects est réalisée à l'aide d'un logiciel de SIG sur l'ensemble de la zone d'étude, à l'exclusion des zones pour lesquelles l'absence de l'espèce est certaine (vignes, zones urbaines), ces zones étant aussi exclues de la surface d'extrapolation ultérieure des résultats. Ces transects sont parcourus par un observateur qui note les distances de détection de toutes les tortues observées.



Résultat : Les données ainsi collectées sont analysées à l'aide du logiciel DISTANCE puisqu'il s'agit d'un simple protocole de « Distance Sampling ». Les milieux étant assez différents, les

transects sont attribués à des strates dont on teste un impact éventuel sur la détection mais aussi sur la densité. Par ailleurs, certaines tortues étant détectées à l'ouïe (mouvement des tortues dans des feuilles), d'autres à la vue, la méthode de détection est aussi incluse sous forme de post-stratification, sachant que la détection à l'ouïe est souvent plus lointaine que celle à vue. On obtient avec ces données une estimation de la densité en tortues par type d'habitat et l'extrapolation de ces densités sur l'ensemble de la plaine des Maures fournit une estimation de la taille de la population totale. Cette estimation ayant un coefficient de variation de l'ordre de 8%, elle est considérée comme assez précise.

Pourtant une seconde étude, menée par la même équipe et sur le même plan d'échantillonnage mais avec un protocole de CMR (utilisant des quadrats au lieu de transects), estime une taille de population trois fois supérieure à celle de la méthode de « *Distance Sampling* ». Une telle incohérence des résultats est extrêmement surprenante et a des conséquences majeures en termes de conservation.

Problème identifié : Un rapide travail de prospection avec deux observateurs fournit une partie de la réponse. En effet, dans les milieux occupés dans la plaine des Maures, à savoir majoritairement des maquis, les tortues sont extrêmement cryptiques. Il n'est ainsi pas rare de rater une tortue, même située au pied de l'observateur. Il suffit pour cela qu'elle soit un peu cachée sous un bosquet. Ceci viole à l'évidence l'hypothèse fondamentale des méthodes de « *Distance Sampling* » qui stipulent que la détection au niveau de l'observateur doit être de 100%. Un deuxième problème a pu être identifié dans des zones plus ouvertes pour lesquelles la détection sur la ligne est *a priori* plus forte. Il s'agit de l'activité des tortues. En effet les tortues ne sont pas actives tous les jours, même lorsque les conditions météorologiques sont favorables. Par exemple, s'il y a eu plusieurs jours de beau temps avant le passage sur le terrain, certaines tortues auront pu se nourrir suffisamment pour ne plus avoir à être actives pendant plusieurs jours. Des travaux d'observation en enclos suggèrent d'ailleurs qu'environ 30% des tortues présentes sont actives à une date donnée. Notons que les tortues qui ne sont pas actives ne sont pas non plus détectables car elles se cachent sous des buissons voire dans des anfractuosités du sol. Dans ces conditions, la méthode de « *Distance Sampling* » ne permet à l'évidence que d'estimer la population active.

Solution : Seule la méthode de CMR permet d'estimer la vraie taille de la population. Les recensements, le « *Distance Sampling* » et les nouvelles méthodes de présence-absence se basent toutes sur des comptages d'animaux à une date donnée. Toutes ces méthodes estimeront donc une taille de population de tortues actives à une date donnée, soit environ 1/3 de la taille de la population totale. Les méthodes de CMR étant basées sur une identification individuelle, les passages successifs permettent en général de détecter lors d'un passage des individus qui n'étaient pas actifs lors des passages précédents. Cette situation que l'on peut comparer à une forme d'émigration temporaire de la zone d'étude (les individus ne sont pas disponibles à la capture à une date donnée car ils ne sont pas actifs, mais le sont à d'autres dates) a pour conséquence que le taux de capture estimé est en fait un mélange entre la probabilité de détecter un individu actif et la probabilité qu'un individu soit actif. Le taux de capture prend donc en compte le paramètre d'activité des individus, la taille de la population est dès lors estimée sans biais.

Point important à retenir : A nouveau il faut bien connaître la biologie ou l'écologie de l'espèce que l'on étudie. Dans cette étude, tout a été fait pour que les résultats du « *Distance Sampling* » soient précis (stratification, post-stratification) et ils le sont. Malheureusement les résultats sont biaisés par le comportement de l'espèce, qui ne permet pas de respecter les hypothèses de base de la méthode d'analyse choisie.

8.4. Cas 4 : ajuster l'effort de prospection par l'analyse régulière des données collectées

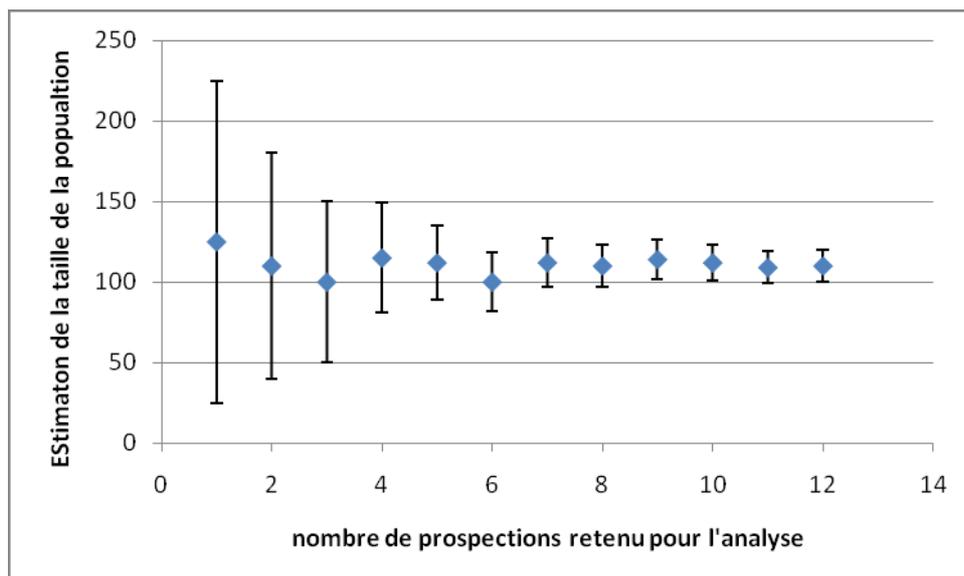
Objectif initial : Estimer la taille de population de Cistude d'Europe dans un étang.

Protocole choisi : La méthode de CMR est retenue. C'est la méthode la plus standard pour cette espèce. Les captures sont réalisées avec des nasses qui sont déplacées régulièrement (tous les 3 jours) de manière aléatoire comme il l'a été suggéré dans le cas 1 ci-avant. Les nasses sont relevées tous les jours. Les individus capturés sont marqués et relâchés. Les captures effectuées sur les 3 jours sans déplacement sont toutes considérées comme provenant d'une seule et même session (une session est donc l'agrégation de 3 relevés). Ce protocole est mené pendant 3 mois.



Résultat : L'analyse des données collectées permet d'estimer une taille de population avec une très bonne précision. Elle est estimée au finale comme étant de 110 individus et son intervalle de confiance est entre 104 et 116. Ceci semble donc montrer que le protocole est pertinent.

Problème identifié : Pourtant une analyse détaillée de ces résultats montre un profil intéressant très fréquent. En effet, on constate en analysant les données cumulées semaines après semaine que l'estimation de la taille de la population se précise très fortement au cours des 6 premières semaines de terrain. Mais ces analyses montrent aussi qu'après 6 semaines de terrain, le gain de précision est extrêmement faible. Cette relation est quelque chose de bien connu et général lors de la conduite de protocoles. En effet, on sait par la théorie statistique que les intervalles de confiance vont décroître de manière exponentielle négative avec l'effort. Ceci signifie qu'un même accroissement d'effort aura un impact fort dans des gammes faibles d'effort mais un impact faible lorsque l'effort est déjà important. Pour être spécifique à cette étude, ce protocole aurait pu durer moitié moins de temps (et donc coûter moitié moins cher), la précision des résultats aurait été la même.



Cette figure présente les estimations de la taille de la population de cistude et ses intervalles de confiance en fonction du nombre de sessions de terrain réalisées. Classiquement l'intervalle de confiance décroît de manière exponentielle négative avec le nombre de sessions. Ici on constate que le terrain aurait pu s'arrêter autour de 7/8 sessions sans grosse perte de précision.

Solution : Ce résultat illustre bien que les tests de puissance et/ou les analyses des données au fur et à mesure de leur collecte sont importants. En effet, analyser même de manière sommaire les données collectées au fur et à mesure ne nécessite en général que quelques minutes de manipulations sur les logiciels, ce qui est incomparablement moins que le temps généralement passé sur le terrain. Dans ce cas extrême, une analyse toute les semaines aurait nécessité au plus fort 10 minutes de manipulation par semaine, soit une heure de travail au bout des 6 semaines. Cette heure d'analyse aurait permis de gagner 5 ou 6 semaines de terrain, ce qui est considérable.

Point important à retenir : Analyser les données au fur et à mesure n'est pas une perte de temps, cela peut permettre de montrer que l'objectif est atteint plus rapidement que prévu, donc d'éviter un effort de terrain inutile. A l'inverse, cela peut aussi conduire à proposer une ou deux sessions de terrain en plus pour avoir la précision désirée.

Annexe 1. Bibliographie

Seules les références les plus importantes sont ici détaillées.

En Anglais :

Malheureusement la grande majorité des livres sur les méthodologies de monitoring d'espèces sont en anglais.

Braun C.E., 2005. *Techniques for wildlife investigations and management (Sixth edition)*. The Wildlife Society, Bethesda, Maryland, USA, 355 p.

- ⇒ Une somme d'information considérable sur les techniques de suivis de la faune. La bible pour les aspects techniques.

Borchers D. L., Zucchini W and Buckland S.T., 2002. *Estimating Animal Abundance: Closed Populations*. Springer-Verlag, 332 pp.

- ⇒ Un ouvrage complet et très didactique sur les méthodes détaillées dans cette note méthodologique, idéal pour aller un cran plus loin sur la compréhension et la manipulation de ces méthodes.

Goldsmith F. B., 1994. *Monitoring for conservation and ecology*. Chapman and Hall New York, New York, USA.

- ⇒ Un livre posant de bonnes questions et apportant de bonnes réponses sur la manière de construire des protocoles pour différents groupes biologiques. Ouvrage assez général, de lecture très aisée.

Williams B.K., Nichols J.D., Conroy M.J., 2002. *Analysis and management of animal populations*. Academic Press, San Diego, California, USA, 1040 p.

- ⇒ La bible pour les aspects méthodologiques. Lecture un peu ardue car insiste sur les aspects mathématiques, mais contient des chapitres plus généraux sur la philosophie des monitoring ainsi que sur la démarche de définition des questions, points qui sont fondamentaux.

En Français :

En Français, il existe le très bon travail de synthèse réalisé par Valérie Fiers et publié par RNF sous forme de 4 livrets regroupés dans un coffret « Etudes scientifiques en espaces naturels – méthodes et expériences ».

Fiers V. et al., 1998. Observatoire du patrimoine naturel des réserves naturelles de France. Analyse et bilan de l'enquête 1996. Quétigny, Réserves Naturelles de France : 200 p.

Fiers V. et coll., 2004. Guide pratique. Principales méthodes d'inventaire et de suivi de la biodiversité. Réserves Naturelles de France : 262 p.

Fiers V. et coll., 2005. Etudes scientifiques. Recueil d'expériences dans les réserves naturelles de France. Réserves Naturelles de France : 222 p.

Fiers V. et coll., 2005. Bibliographie. Etudes scientifiques en espaces naturels. Réserves Naturelles de France : 108 p.

Un cahier technique a également été publié par l'ATEN :

Fiers V. et coll., 2003. Etudes scientifiques en espaces naturels. Cadre méthodologique pour le recueil et le traitement de données naturalistes. Cahier technique de l'ATEN numero 72. Réserves Naturelles de France. Montpellier : 96 p.

- ⇒ Ces documents sont plutôt axés sur les aspects techniques que sur les méthodes d'analyse de données mais ils offrent une vision complète des outils disponibles et proposent aussi une large introduction très didactique sur la définition des questions scientifiques.

Blandin P., 2009. *De la protection de la nature au pilotage de la biodiversité*. Editions Quae, Versailles, 124 p.

- ⇒ Un livre un peu annexe mais très intéressant, expliquant l'évolution de la manière de penser la protection de la nature en France.

Annexe 2. Logiciels d'analyse de données

Pour le tirage aléatoire

R : <http://www.r-project.org/>

Logiciel gratuit

Pour le « *Distance Sampling* »

DISTANCE : <http://www.ruwpa.st-and.ac.uk/distance/>

Logiciel gratuit, manuel très détaillé avec des exemples très progressifs permettant de se former en autonomie sur son utilisation. En anglais malheureusement.

Pour les Captures-Marquages-Recaptures :

MARK : <http://www.phidot.org/software/mark/download/>

Logiciel avec prise en main assez rapide, interface conviviale, manuel en anglais très complet.

Pour la Présence-absence :

PRESENCE et GENPRES: <http://www.mbr-pwrc.usgs.gov/software/presence.html>

Ces deux logiciels sont récents et en plein développement. Prise en main relativement complexe car pas de manuel en dehors de quelques pages html pour acquérir les bases sur la constitution des fichiers et leur ouverture. Il y a cependant depuis peu un lien sur cette page avec des exercices progressifs qui peuvent permettre aux personnes vraiment intéressées de s'attaquer à ces logiciels en autonomie.