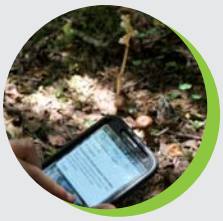
Conservatoire d'espaces naturels Rhône-Alpes

RHÔNE-ALPES

GUIDE MÉTHODOLOGIQUE

Définir, mettre en œuvre et analyser des suivis scientifiques en espaces naturels

SOMMAIRE



p4

LES ÉTAPES DU SUIVI SCIENTIFIQUE

p 12

LES PRINCIPES DE L'ANALYSE STATISTIQUE

p 18

CE QU'IL FAUT RETENIR

PRÉAMBULE

Définir, mettre en œuvre et analyser des suivis scientifiques en espaces naturels

Il est illusoire de penser produire un guide méthodologique pouvant couvrir l'immense diversité des questions qu'un gestionnaire d'espace naturel se pose et la diversité des méthodes d'échantillonnage à disposition, et encore moins de lui donner des recettes infaillibles pour analyser ses données de manière pertinente et facile. Il est cependant possible de donner à voir quels sont les liens entre les différentes étapes de la mise en œuvre d'un suivi et d'illustrer concrètement les conséquences des choix à faire à différents niveaux ainsi que les questions multiples à ne pas oublier avant de se lancer.

Ce guide a donc pour objectif de :

- **décrire les étapes de mise en place d'un suivi scientifique et notamment les liens entre plan d'échantillonnage et analyse des données ;**
- **donner à voir les principaux écueils qu'un gestionnaire d'espace naturels pourra rencontrer lors de la phase d'analyse de ses données ;**
- **rappeler de manière simple les fondamentaux de la statistique.**

Soyez rassurés, lecteurs, tout va bien se passer !

Définir, mettre en œuvre et analyser des suivis scientifiques en espaces naturels

Rédaction : Delphine Danancher (Cen Rhône-Alpes),*.

Référence de ce document : Danancher D. et al.*, Définir, mettre en œuvre et analyser des suivis scientifiques en espaces naturels, CEN Rhône-Alpes, 16p.

Merci à Pascal Faverot, Camille Chappat, Frédéric Didier (CEN Rhône-Alpes) et Bernard Bal (CEN Haute-Savoie) pour leur aide pour l'organisation de cette journée et la mise en forme de ce document.

** voir la liste des autres contributeurs en quatrième de couverture.*

INTRODUCTION

Mettre en œuvre des suivis scientifiques, pour un gestionnaire, c'est à la fois tenter de comprendre les phénomènes naturels à tous les niveaux (espèces, habitats, écosystèmes, dynamiques temporelles…) et les décrire. C'est aussi évaluer son travail, les objectifs qu'il se sera fixés dans le cadre d'un plan de gestion pour ajuster ses actions *(Cf guide d'élaboration des plans de gestion CT 88 ATEN, Figure 1).*

Les suivis scientifiques sont des opérations de collecte de données répétées dans le temps selon un protocole standardisé (c'est-à-dire lui-même stable dans le temps) visant à détecter les variations temporelles et/ou spatiales des variables étudiées.

Les suivis scientifiques sont à distinguer des inventaires qui sont des opérations de collecte de données quantitatives ou qualitatives, ponctuelles dans le temps, visant à décrire un patron d'abondance ou de répartition d'espèces.

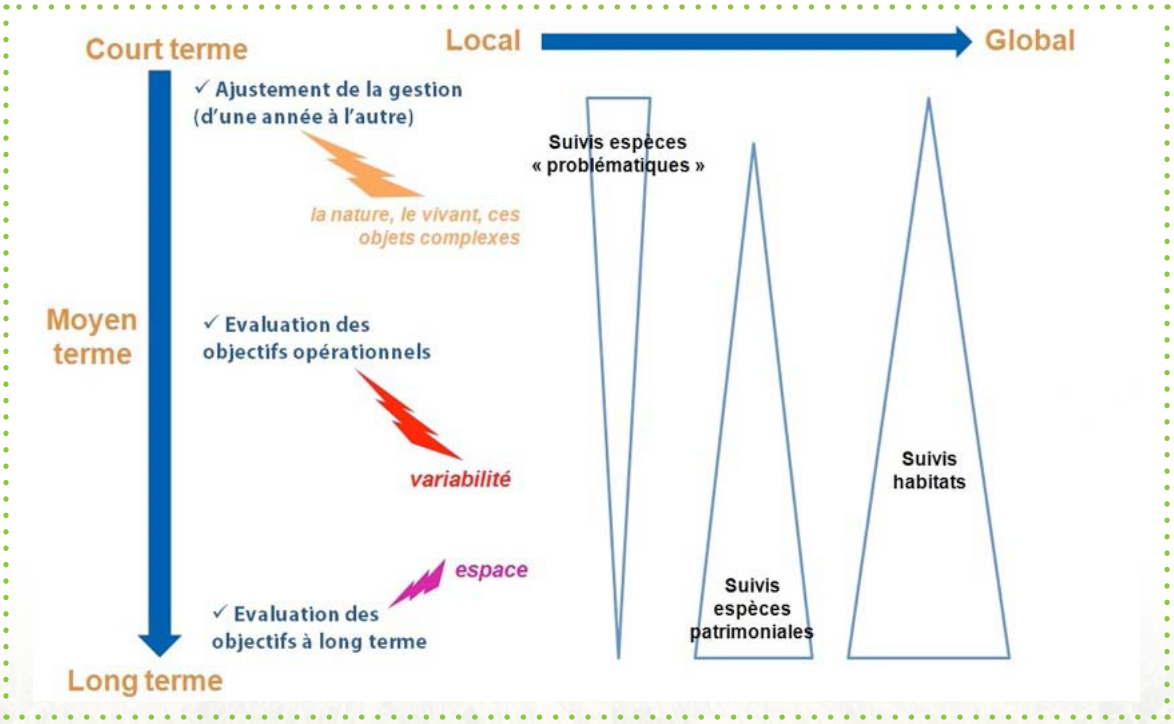


Figure 1 : Schématisation des objectifs des suivis pour un gestionnaire d'espaces naturels (Anne Wolff).

LES ÉTAPES DU SUIVI SCIENTIFIQUE

Toutes les étapes d'un suivi scientifique sont en lien les unes avec les autres et doivent être planifiées avant de passer à sa mise en œuvre *(figure 2)*.

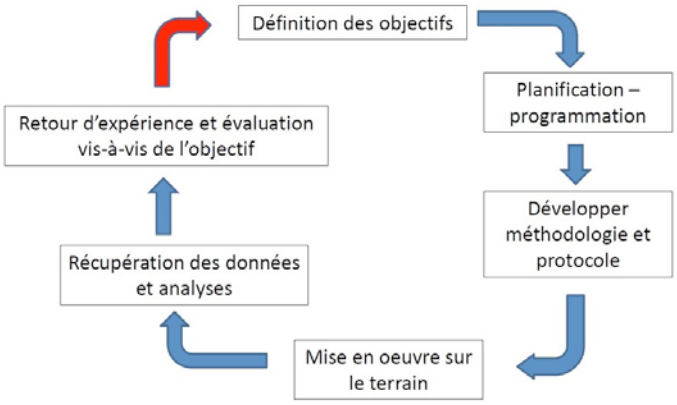


Figure 2 : Les étapes d'un suivi scientifique (Aurélien Besnard).

Définir un objectif précis pour le suivi

Un suivi réussi, c'est avant tout une question bien posée, bien définie, et pour laquelle il est possible d'apporter une réponse. Pour y parvenir, les trois points suivants sont nécessaires.

- **Bien appréhender le contexte et les enjeux du suivi**

Dans le cadre de la gestion des espaces naturels, les objectifs d'un suivi sont variés. Il peut servir à alerter sur des risques (extinction, invasion, etc). On parle alors de notion de veille avec éventuellement définition d'un seuil d'alerte (sur des bases scientifique ou via l'expertise) pour déclencher des politiques publiques (interdiction de chasse, pâturage, ouverture du milieu, régulation des niveaux d'eaux, régulation fréquentation, limitation des invasives…). Les données de base peuvent aussi être exploitées pour des classements d'espèces (UICN, listes rouges).

Ces suivis peuvent aussi avoir pour objectif d'évaluer des mesures de gestion, des politiques publiques ou l'impact de perturbations.

Pour que les résultats de ces suivis puissent être partagés par les différents acteurs gravitant autour des espaces protégés, il est nécessaire que les objectifs le soient aussi. Il faut donc avoir en tête clairement le contexte et la justification de l'opération de suivi, quels sont les acteurs qui posent cette question et surtout à quoi et à qui serviront les données collectées, à quelle échéance ?

- **Faire le bilan des connaissances sur l'objectif**

D'autres se sont peut-être posé la même question avant vous, parfois ont trouvé des réponses à cette question, réponses que vous pouvez exploiter (il n'est pas nécessaire de ré-inventer l'eau chaude) ou ont commis des erreurs que vous pourrez éviter dans la mise en place du dispositif de suivi (plan d'échantillonnage, variable ciblée, matériel de relevé, etc). Le bilan bibliographique est donc indispensable en amont des réflexions sur la mise en place d'un suivi. Ce bilan peut se faire au travers de recherches bibliographiques dans la littérature internationale mais aussi dans la littérature grise (non publiée comme les rapports de stages par exemple). Il peut être complété par la consultation de spécialistes de la question écologique qui est posée (il peut être alors utile d'avoir des avis contradictoires pour se forger sa propre idée). Il est important aussi d'avoir en tête l'utilité de consulter les spécialistes également sur les aspects méthodologiques (au sens « statistiques » ici) dès la préfiguration du protocole envisagé, car ils pourront aider à déterminer le type d'analyse de données à réaliser, ce qui peut avoir des conséquences sur la manière de collecter les données.

- **Passer d'un objectif à une question précise**

En tant que scientifique et gestionnaire, il faut dans la mesure du possible essayer de sortir de la simple description pour tendre vers une démarche de type hypothético-déductive. Un grand mot pour simplement dire qu'à partir de votre questionnement initial vous devez poser une hypothèse (qu'on appellera « hypothèse nulle » lors de la phase d'analyse) que vos données récoltées sur le terrain (ou suivant un protocole expérimental pour les plus ambitieux) devront valider ou rejeter lors de l'analyse des données (exemple 1).

EXEMPLE 1

La question du gestionnaire :

Faucher est-il efficace pour éliminer le solidage qui envahit mon site?
Le gestionnaire souhaite pouvoir évaluer l'atteinte de cet objectif dans 5 ans, à la fin de son plan de gestion.

Le paramètre mesuré : il est necessaire de mesurer une variable analysable statistiquement, par exemple un nombre de pieds de solidage rapporté à une surface, c'est-à-dire une densité de pied.

L'hypothèse nulle : il n'y a pas de différence de densité de pieds de solidage sur les secteurs que je fauche et ceux que je ne fauche pas.

Ce que l'on apprend avec cette question sur le protocole à mettre en place : il faudra des zones que je faucherai (zones « traitées »), et d'autre non fauchées dans un site donné (zones « témoins »). On comprend qu'il faut éviter les facteurs confondants (ici par exemple des facteurs qui influencent la production du nombre de pieds de la plante) et qui risquent de masquer l'effet étudié (le rôle de la fauche). On cherchera donc à comparer la zone fauchée et la zone non fauchée en recherchant le plus d'homogénéité entre ces deux conditions, mais comme dans le milieu naturel cette homogénéité n'existe pas, on multipliera dès lors le nombre de zones testées (plusieurs secteurs fauchés et plusieurs secteurs non fauchés) dans le site étudié et on les disposera au hasard pour inclure cette hétérogénéité dans les deux types de zones.

Evidemment, si le gestionnaire souhaite savoir si l'effet (la fauche) est aussi efficace pour contrôler le solidage dans différents types de sites, il faudra mettre en place ce suivi sur chacun d'entre eux (avec des secteurs fauchés et non fauchés), et lister les autres variables pouvant interagir avec la fauche pour influencer les densités de pieds.

C'est seulement si cette hypothèse de départ est claire qu'il sera possible de définir correctement quelles sont les variables à mesurer (donc votre protocole) et de mener à bien une analyse sensée.

Ceci reste vrai même si vous ne conduisez pas de test statistique mais une analyse plus qualitative des résultats d'un suivi (exemple 2).

EXEMPLE 2

La question du gestionnaire :

Je fais du broyage tous les ans sur une pelouse sèche pour limiter l'embroussaillement et préserver les espèces de milieux ouverts. Est-ce que la gestion que je pratique permet réellement de limiter l'embroussaillement de ma pelouse sèche ?

Le paramètre mesuré : Sur le terrain je délimite à l'aide d'un GPS toutes mes surfaces embroussaillées. Je fais cette mesure 3 fois avec à chaque fois un intervalle de 2 ans. Je fais le rapport entre cette surface embroussaillée et la surface totale en pelouse sèche.

Hypothèse nulle : la fauche n'impacte pas le taux d'embroussaillement de mon site.

J'ai besoin de définir un seuil d'embroussaillement, *a priori*, au-dessus duquel je jugerai que ma pratique ne permet pas de maintenir un milieu ouvert. Je dois donc aller plus avant dans la définition de ce qu'est un milieu ouvert pour moi. Si au bout de 6 ans, plus de 20% de la pelouse sèche reste embroussaillée (seuil que je fixe *a priori*, d'après mon expertise), je conclurai que la fauche n'a pas permis la réouverture du milieu.

Définir une technique d'échantillonnage bien adaptée à la question posée

Quelle est la durée d'un suivi ?

Ce qui détermine la durée d'un suivi est souvent - et c'est légitime - l'échéance à laquelle une réponse est attendue (révision de plan de gestion dans 5 ans, fin administrative d'un programme, etc). Cette nécessité doit être mise en balance avec les connaissances que vous avez sur le processus que vous souhaitez mesurer et l'échelle de temps à laquelle vous êtes en capacité à mesurer un changement qui a du sens (exemples 3A et 3B). D'une manière générale, plus on fait de passages sur le terrain, plus l'analyse sera précise car plus on aura de données pour estimer les paramètres. Cependant, déceler un changement, une tendance, implique de positionner ces passages sur une échelle de temps pertinente pour le phénomène observé.

EXEMPLE 3A

Suite à une restauration hydrologique, à partir de quand puis-je considérer que la diversité floristique sera stabilisée et qu'une comparaison de l'état après restauration à l'état initial aura un sens. Ces informations (vitesse de réaction au changement) doivent être recherchées lors de l'analyse bibliographique.

EXEMPLE 3B

Dans le cadre d'un suivi d'une population qui décline très progressivement, le fait de passer tous les ans pendant 5 ans peut amener à conclure qu'il n'y a pas eu de changement, simplement parce que ce dernier est faible et masqué par les variations inter-annuelles. Un suivi s'étalant sur 10 ans aurait à contrario pu permettre de détecter la chute des effectifs.

Un suivi doit être évalué régulièrement pour vérifier son adéquation aux objectifs. Il est important de prévoir des échéances d'évaluations intermédiaires (faisabilité de terrain, précision des résultats, etc). Ces analyses peuvent donner de premières tendances exploitables mais surtout alerter en cas de problème, qu'il soit méthodologique (le protocole n'est pas du tout adapté) ou en lien avec les résultats (les premières tendances montrent que la gestion que je pratique ne va pas dans le bon sens).

Budgétiser le suivi

Avoir une réflexion sur les moyens humains (compétences, réalisation en interne ou externe...) et financiers que le suivi que vous planifiez devra mobiliser est essentiel. C'est un gage de pérennité. Suite à cette analyse, il sera alors possible de s'assurer que la question retenue est en adéquation avec les moyens disponibles. Mieux vaut renoncer à un suivi financièrement impossible à mettre en œuvre ou à maintenir sur le long terme que de se lancer tête baissée et devoir s'arrêter au milieu du gué, sans résultat exploitable.

Développer le protocole

Quelle variable mesurer ?

Lorsqu'un suivi est mis en place, ce que l'on suit pour tenter de le mesurer, de l'évaluer ou de l'expliquer est appelé variable « focale » (ou variable expliquée). Cette variable centrale, si les choses sont bien faites, se trouve dans la question que vous posez, dans votre hypothèse. Cette variable peut être quantitative (ex : la taille ou le poids d'un individu, le nombre d'individus) ou qualitative (ex. le sexe ou la couleur des fleurs d'un individu). Pour ce dernier cas, on appellera « modalités » les différentes valeurs que la variable qualitative pourra prendre (ex : pour le sexe d'un individu : mâle, femelle).

Les éléments concrets que l'on va suivre (les individus en eux-mêmes, des parcelles dans lesquelles se trouvent les individus, etc) sont ce qu'on appelle des unités statistiques (ou individus statistiques). Ces individus statistiques doivent être répliqués pour avoir un jeu de données suffisamment grand pour permettre des analyses statistiques fiables.

Lorsque l'on tente d'expliquer un phénomène observé sur la variable focale, il est possible d'ajouter d'autres variables que l'on dit « explicatives » (la pluviométrie, la température, etc). Ainsi, un suivi prend du sens et n'est plus simplement descriptif (j'ai significativement plus d'individus de l'espèce X sur le site Y que sur le site Z) mais il acquiert un pouvoir explicatif (j'ai significativement plus d'individus de l'espèce X sur le site Y que sur le site Z parce que la ressource trophique y est plus importante). La collecte de ces variables a cependant un coût (humain et donc financier) et il est alors nécessaire de se poser la question de leur utilité, de leur exploitabilité. Ceci est d'autant plus vrai que, plus on augmente le nombre de variables dans une analyse, plus il est en général nécessaire d'augmenter la taille de l'échantillon pour identifier leurs effets respectifs. Il est généralement conseillé de prévoir un minimum de 10 unités d'échantillonnage par variable dans les modèles. Ainsi, pour connaitre le nombre d'échantillons nécessaires à une étude statistique fiable, il suffit de multiplier le nombre de variables par 10.

Pour la variable expliquée comme pour les variables explicatives, le degré de finesse de la mesure doit être discuté (le comptage d'une espèce de plante dans une parcelle, par exemple, tient compte de tous les individus présents, ou seulement des pieds fleuris, etc.). Cette bonne définition de la variable mesurée permet d'assurer la bonne répétabilité entre les observateurs dans le temps.

Le plan d'échantillonnage

On comprend bien que la mise en place d'un suivi ne va pas pouvoir se faire sur toute la surface d'une zone sur laquelle on étudie un phénomène. L'étude doit donc être réalisée sur des unités spatiales plus petites mais en s'assurant que les résultats obtenus soient transférables à la zone entière. C'est la réalisation d'un plan d'échantillonnage qui va permettre ce changement d'échelle.

La première question à se poser va donc être la définition de la surface (taille et localisation) sur laquelle va être déployé cet échantillonnage (il va définir ce qu'on appelle la « population statistique », qui est donc un échantillon de la « population » totale étudiée). Cette décision aura des conséquences évidentes sur la quantité de travail à fournir sur le terrain mais aussi des conséquences plus subtiles sur la phase d'analyse statistique et d'interprétation des résultats *(exemples 4A et 4B, voir également le chapitre sur les principes de l'analyse statistique)*.

EXEMPLE 4A

J'évalue l'évolution de la diversité spécifique en suivant les milieux ouverts d'une zone humide et je ne peux donc pas tirer des conclusions à l'échelle de mon site car il contient aussi des milieux boisés et des milieux aquatiques.

EXEMPLE 4B

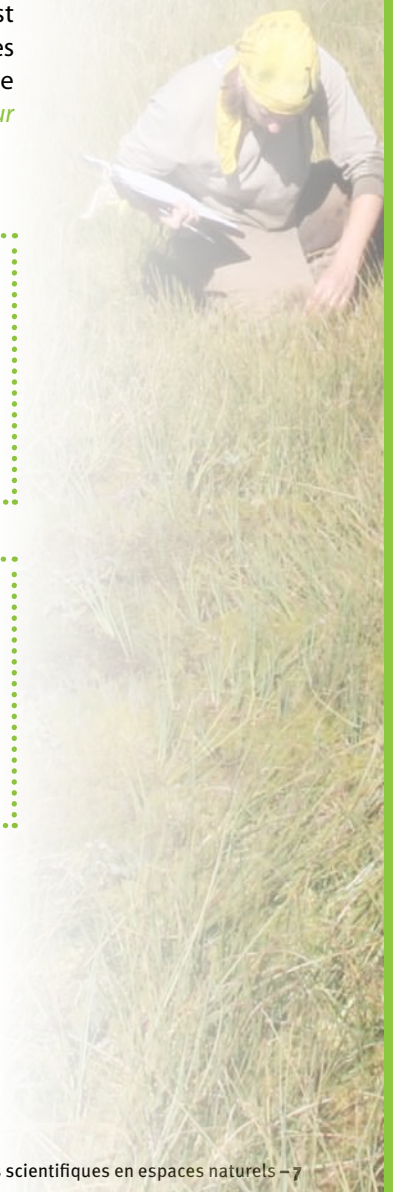
Je teste l'efficacité de mes pratiques de gestion sur un seul site. Je n'ai pas d'information sur la transférabilité de mes résultats à d'autres contextes.

Les méthodes d'échantillonnage sont diverses. Elles peuvent nécessiter parfois une connaissance préalable de la population que l'on étudie :

- **Echantillonnage aléatoire :** chaque unité statistique (le quadrat, le transect, le point d'écoute, etc) de la population statistique a la même chance d'être choisi que tous les autres éléments de la population visée (exemple 5).

EXEMPLE 5

Je définis une maille sur mon terrain d'étude. Je tire au sort 20 mailles dans lesquelles je fais un comptage d'individus. Je peux ainsi estimer le nombre d'individus du terrain d'étude sans les avoir tous comptés.



• **Echantillonnage aléatoire stratifié :** la population est divisée en sous-groupes (strates) homogènes. Un échantillon indépendant est prélevé dans chacune de ces strates (exemples 6A et 6B). Cette méthode nécessite donc d'avoir des connaissances préalables sur le terrain d'étude et/ou la variable étudiée.

EXEMPLE 6A

Dans le cadre d'un suivi d'une espèce patrimoniale à l'échelle d'un territoire pour évaluer l'état d'une population, il est possible de décider de faire plus de relevés dans les sites où la présence en abondance de l'espèce est connue par rapport aux sites pour lesquels on sait que les effectifs sont plus faibles (sachant que ces derniers apporteront de toute façon moins d'information sur l'état de la population).

EXEMPLE 6B

Je veux faire des relevés phytosociologiques en utilisant le type d'habitat corine comme strate. Il est possible de décider de faire un nombre de relevé constants quelle que soit la surface de l'habitat échantillonné. On parlera dans ce cas d'échantillonnage pondéré car les habitats couvrant le plus de surface auront une densité de relevé plus faible que les habitats les moins représentés.

• **Adaptative cluster sampling :** notamment lorsque l'on étudie une population de petite taille ou de faible densité et agrégée dans l'espace, si l'on utilise un échantillonnage aléatoire simple, beaucoup de points de relevés ne contiendront aucune observation et quelques points contiendront de très forts effectifs, ce qui conduit à des estimations très peu précises. L'adaptative cluster sampling permet d'être plus efficace pour un même effort de terrain en exploitant des connaissances préalables sur la répartition des individus. Il commence comme précédemment par un tirage aléatoire mais permet l'ajout de points/quadrats à proximité des points où une observation est réalisée (exemple 7). Il s'agit donc d'un mode d'échantillonnage qui s'ajuste sur le terrain, contrairement aux autres modes décrits plus haut pour lesquels les décisions sont prises a priori. Il est particulièrement adapté aux populations avec des individus répartis en taches.

EXEMPLE 7

Je fais un suivi par maille pour compter des individus et je tire au sort 20 mailles sur lesquelles je compterai. Quand sur une maille l'espèce est présente, j'ajoute à sa périphérie 4 autres mailles adjacentes qui feront partie de mon échantillon et si dans ces mailles adjacentes certaines sont positives, j'ajoute à nouveau les mailles adjacentes de cette dernière.

• **Two-phase sampling :** s'utilise pour les espèces rares et agrégées, comme l'adaptive cluster sampling. Cette démarche repose sur un échantillonnage en deux phases (d'où son nom), la première visant à définir les strates, la deuxième à réaliser un échantillonnage stratifié et pondéré dans les strates définies dans la première étape.

Il est très important de bien comprendre que selon le phénomène observé, l'échelle pertinente d'observation n'est pas la même (figure 3).

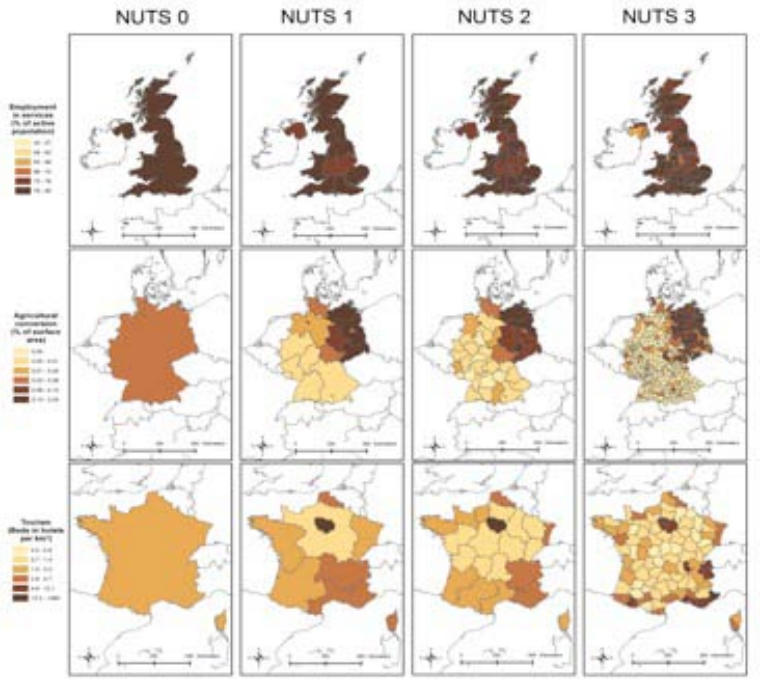


Figure 3: Cartographie, en ligne, de différents phénomènes (emploi dans les services, déprise agricole, fréquentation touristique,) en colonne, à différentes échelles d'observations (du national au départemental). Ici, l'intensité et l'homogénéité spatiale de différents phénomènes peuvent varier en fonction des échelles d'observation.
A grande échelle, ces phénomènes semblent homogènes mais des niveaux de découpage plus fins de la surface d'observation font apparaitre des différences entre les phénomènes en termes d'intensité et de répartition au sein du territoire observé (Aurélien Letourneau).

Une variable peut paraitre homogène à une échelle d'observation donnée alors qu'une échelle plus fine laissera voir sa variabilité. Ainsi, les plans de gestion ou la mise en œuvre de politiques devront prendre en compte la variation à différentes échelles de ces phénomènes pour être appliqués, selon les besoins, de manière homogène ou différenciée au sein du territoire concerné.

Par ailleurs, la perception des objets du territoire (par exemple, de corridors écologiques) varie en fonction de l'échelle d'étude, et l'absence de prise en compte de cette échelle peut conduire à des incompréhensions, voire de mauvaises interprétations. Il est donc toujours intéressant de jouer sur les échelles et de faire parler les études locales au regard du contexte régional, comme de décliner des enjeux régionaux au regard de contextes locaux (Figure 4). Il faut bien être conscient que ces variations selon les échelles spatiales existent aussi selon l'échelle temporelle.

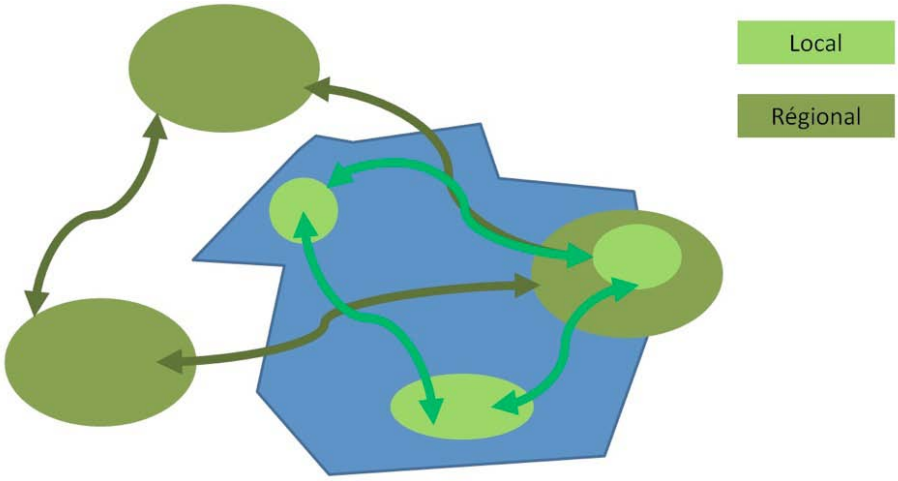


Figure 4 : Emboitement des échelles locales et régionales pour la notion de corridor (Aurélien Letourneau)

Lors de la mise en place d'un suivi, la question du besoin de disposer ou non d'un (ou de plusieurs) témoin(s) se pose souvent. Généralement lorsque l'on souhaite démontrer l'effet d'un facteur (modalité de gestion, caractéristique du milieu, etc), une situation témoin est recherchée (site identique sans gestion, gradient de caractéristiques du milieu, etc.). Ce témoin doit être, autant que faire se peut, identique à la zone de suivi, exception faite de la variable d'intérêt (exemple 8). Lorsque l'on met en place un tel dispositif de suivi, on se rapproche d'une démarche expérimentale. Bien sur, n'étant pas en laboratoire, il est impossible de contrôler les processus qui régissent l'écosystème étudié. Il est également impensable de suivre toutes les variables à l'œuvre dans la dynamique du système étudié, donc prudence dans l'interprétation car il peut y avoir des facteurs confondants !

EXEMPLE 8

Je veux savoir si le pâturage ovin est efficace pour lutter contre l'embroussaillement. Je mets en place un suivi de la dynamique arbustive sur plusieurs zones (quadrats) dans une zone pâturée. Je définis aussi des quadrats non soumis au pâturage (défend). Je fais les relevés de végétation dans l'ensemble de ces quadrats. A terme, je comparerai les deux types de quadrats en considérant que la seule différence entre eux est la présence ou pas de moutons. Je pourrai tenir compte de facteurs confondants (par exemple la qualité ou l'hygrométrie du sol) si je les ai mesurés.

Idéalement tous ces éléments qui constituent un protocole doivent être synthétisés de manière rédigée dans un document avec une date et un numéro de version. Ce document est la mémoire du suivi, en cas de changement de personnel notamment, mais aussi en cas de malentendu ou d'incertitude. Il peut évoluer en incrémentant des versions qui incluent alors une page précisant les modifications apportées à la version précédente.

Dès cette phase de mise en place de l'échantillonnage, vous décidez implicitement des méthodes d'analyse que vous pourrez utiliser et il est important d'en être conscient. Le nombre d'échantillons, le nombre de variables qui sont suivis décideront de la puissance des analyses (c'est-à-dire de la capacité du protocole à répondre à la question posée avec l'effort de mesure consenti). La nature des données (variables continues (comme la taille des individus en cm) ou discrètes (comme des comptages) ou

variables catégorielles (comme les couleurs) décidera également des tests statistiques qui seront utilisables. La répartition des données sera également importante (loi normale ou non, répartition unimodale ou bimodale).

Il peut être important de se faire accompagner par un spécialiste des analyses de données comme beaucoup de gestionnaires ont l'habitude de le faire pour la détermination ou l'écologie des espèces. Il est également important de réaliser ces choix au regard du budget dont on dispose (ou mieux de chercher un budget au regard des objectifs de suivi, mais cela n'est pas toujours possible).

Collecter les données

Le protocole, pour plus de facilité (rapidité de saisie, utilisation de référentiel pour décrire les observations, etc) et de sécurité (fiabilité et exhaustivité des données, respect des formats, répétabilité entre observateurs, etc.) doit être converti en fiche opérationnelle de terrain. Cette fiche doit être explicite et peut également inclure des informations sur les conditions du relevé (date évidemment, heure, conditions météo notamment durée, etc). Ces dernières permettent par exemple d'écarter a posteriori des relevés dont les résultats sont aberrants, ou bien d'estimer des valeurs de détection pondérées par ces variables. Il est important, lorsque plusieurs observateurs interviennent, de bien se mettre d'accord sur la manière de remplir ces fiches.

Il est temps de décider de qui mesure. Avec quelle compétence (formation) ? Avec quel matériel ? Il faut alors penser à la disponibilité du matériel et des véhicules, aux enjeux de sécurité (travail de nuit, plusieurs observateurs, etc), aux risques de juxtaposition (période, charge de travail..) avec les autres opérations que vous conduisez. Ainsi aurez vous un regard critique sur la faisabilité en regard de l'échantillonnage prévu.

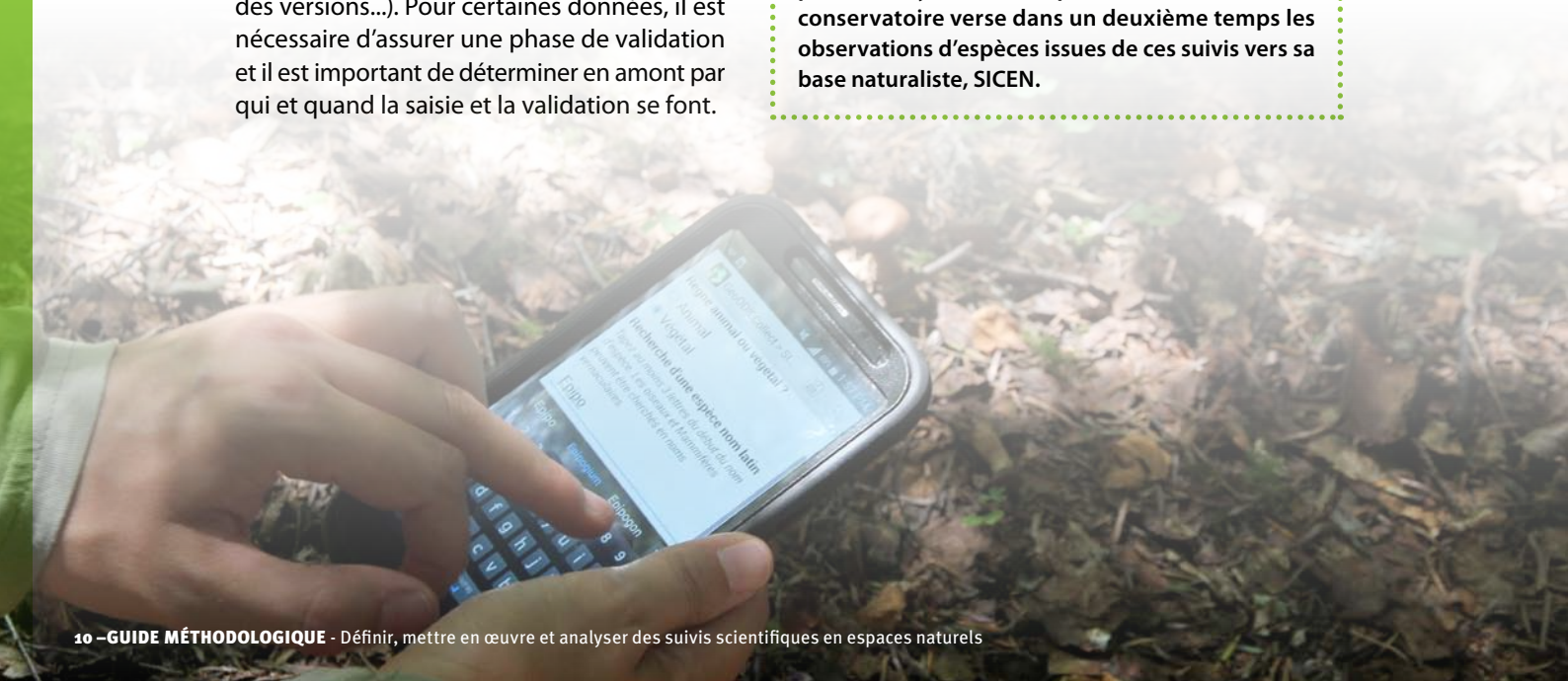
Stocker les données

Si l'on dispose déjà d'une base de données, il est nécessaire de se demander si les données collectées pourront l'intégrer sans problème et dans le cas contraire de réfléchir au format de base de données qui sera nécessaire pour bancariser les observations et la métadonnée qui les accompagne. Généralement les bases de données naturalistes stockent des données d'observation d'espèces (ou d'habitats) accompagnées de l'identité de l'observateur, de la date et du lieu d'observation. Elles peuvent ne pas permettre la saisie des données de contexte nécessaires à l'interprétation (conditions de relevé, description écologique des secteurs échantillonnés, localisation précise des points de relevés…)

Lorsque l'on ne dispose pas d'outil de saisie et de stockage des données, il faut se poser la question de qui peut la construire, avec quelles compétences (expert géomaticien…) (exemple 9). Une base de données permet l'exploitation des données mais assure également leur pérennité (sauvegarde, gestion des versions...). Pour certaines données, il est nécessaire d'assurer une phase de validation et il est important de déterminer en amont par qui et quand la saisie et la validation se font.

EXEMPLE 9

Pour les suivis RhoMéO flore le conservatoire d'espaces naturels de Rhône-Alpes a construit un outil de saisi spécifique qui permet la mobilisation de la donnée naturaliste et des données associées pour l'analyse et l'interprétation des suivis. Le conservatoire verse dans un deuxième temps les observations d'espèces issues de ces suivis vers sa base naturaliste, SICEN.



Analyser les données

La phase d'analyse des données, la plus redoutée par un certain nombre d'entre nous, pose un certain nombre de questions qui auront pu être en partie traitées si les recommandations décrites plus haut ont bien été prises en compte. En effet beaucoup des décisions prises en amont ont d'ores et déjà scellé les possibilités d'analyse. Ainsi, cette étape ne doit pas être réfléchie après la mise en œuvre des suivis de terrain mais bien pendant leur conception, à un moment où il est encore possible de faire évoluer le protocole (voire la question !).

Il est important de se demander s'il sera nécessaire de faire appel à un(e) spécialiste pour analyser les données ou si c'est vous qui allez le faire et à quelle échéance.

Lors de l'étape d'analyse des données, il est possible d'utiliser :

- **la statistique descriptive :** l'analyse descriptive permet de résumer un jeu de données en un ensemble de descripteurs synthétisant les données. Ces descripteurs offrent par exemple une représentation graphique des résultats qui permet souvent de voir rapidement leurs principales caractéristiques. Les indicateurs descriptifs les plus utilisés sont :
  - les indicateurs de tendance centrale : moyenne, médiane…
  - les indicateurs de dispersion : écart type, variance…

Dans le cas d'un nombre important de variables (et c'est souvent le cas en écologie), les analyses multivariées permettent d'étudier les relations entre ces variables dans un espace multi-dimensionnel mais avec des présentations simplifiées basées sur la sélection des variables participant le plus au phénomène étudié.

- **la statistique inférentielle :** ce sont les outils qui permettent de généraliser les résultats observés sur notre échantillon (par exemple dans les quadrats) à la population d'origine (le secteur étudié) dont il est issu et que nous n'avons pas pu mesurer dans son ensemble. C'est ce que nous appelons les tests statistiques (ou tests d'hypothèses) qui servent à valider ou invalider nos hypothèses statistiques (exemple 10).

EXEMPLE 10

Un suivi de la taille de la végétation en fonction de la nature du substrat. Des échantillons collectés sur le substrat 1 et d'autres sur le substrat 2. L'hypothèse nulle (Ho) est : il n'y a pas de différence significative de taille entre la végétation poussant sur le substrat 1 et celles poussant sur le substrat 2 dans la population.

A noter que ces tests peuvent être réalisés avec une ou deux variables à la fois (test univarié ou bivarié) mais également en approche multivariée, cette dernière ne devant pas être uniquement vue comme une approche descriptive.

Les résultats des analyses de données vont devoir être restitués à des interlocuteurs différents, spécialistes ou non et il est important de bien déterminer le niveau de détail à fournir pour chaque interlocuteur.

Evaluer un suivi

Un suivi peut gagner en qualité s'il est confronté à des avis externes (un collègue, un conseil scientifique, etc). Il est important de ne pas attendre la toute fin d'un suivi avant de l'évaluer et de produire une analyse intermédiaire (analyse de données des premières années par exemple) qui peut permettre un ajustement précoce de la méthode.

De même, il peut être intéressant de faire une « pré-étude » (petites études en amont du suivi pendant un ou deux ans) qui permettra de mieux comprendre le fonctionnement du système, d'orienter les choix méthodologiques (tailles des unités, un ou plusieurs passages, répétabilité observateurs, détection…), d'évaluer la pertinence des choix méthodologiques/techniques, la faisabilité du suivi et d'alimenter des tests de puissance ; il peut alors s'imposer de se remettre autour de la table si les retours terrain/analyses de données tendent à laisser penser que le suivi ne répondra pas aux objectifs. Ces pré-études peuvent apparaître comme une perte de temps sur le moment mais elles permettent souvent d'optimiser largement les suivis et donc de gagner en précision et en fiabilité sur le long terme (ou d'économiser des moyens humains/financiers).

LES PRINCIPES DE L'ANALYSE STATISTIQUE

Les outils de statistique sont là pour décrire de manière formalisée (selon un modèle) les variations naturelles que nous observons et pour permettre de faire la différence entre ce qui est du au hasard et ce qui répond à un patron que nous pouvons expliquer. Le résultat de leur application permet de prendre des décisions. Ils correspondent à un langage international bien établi qui fait d'eux également un outil de communication.

Bien que la mise en pratique des tests statistiques arrive en fin de parcours du suivi scientifique, l'analyse des données doit être pensée et formalisée en même temps que le protocole de collecte des données et la formulation de l'hypothèse. Trop souvent, dans le monde de la gestion des milieux naturels, seuls les protocoles de collecte de données sont produits et rédigés. Très peu de méthodologies de suivi fournissent des connaissances ou des recommandations sur l'analyse des données associées.

Analyser ses données, ce n'est pas forcément sortir une batterie de tests complexes, c'est avant tout savoir les regarder, les décrire, étape de première importance qui pourtant, la plupart du temps, est négligée. Avant de sauter sur une analyse de type inférentiel, il est important de passer par une analyse descriptive qui dans bien des cas apporte nombre de réponses à un gestionnaire. Enfin, trouver le bon test pour ses données, c'est choisir une méthode qui correspond à leur structure, mais c'est la choisir en connaissant bien son domaine d'application et surtout ses limites.

Les statistiques les plus couramment utilisées dans le contexte « environnemental » reposent sur des estimations de probabilités. C'est donc prendre des décisions pour lesquelles on pèse le risque de se tromper. L'exemple le plus connu est le seuil de probabilité auquel on rejette ou non une hypothèse et qui est le plus souvent fixé à 5% en écologie (1% en médecine). Il signifie concrètement que l'on fixe à 5% le seuil de risque de se tromper en rejetant notre hypothèse alors qu'elle est vraie, donc de la rejeter à tort.

Dans ce chapitre, il nous sera impossible de décrire, pour chaque type de données, la méthode d'analyse à suivre et il s'agira plutôt d'apprendre à regarder ses données avec un œil critique et avisé, de comprendre l'impact de notions clefs (taille d'échantillon, puissance d'un test, structure des données) pour permettre à un gestionnaire de faire les choix les plus avisés possible et surtout de se faire accompagner par un expert pour ses suivis les plus ambitieux. Nous essayerons donc de répondre à un ensemble de questions simples sur les fondamentaux de l'analyse de données, car comprendre les effets de choix que nous faisons est fondamental pour prendre à l'avenir les bonnes décisions. Pour rendre le propos plus simple, nous utiliserons les exemples d'une analyse bivariée (une variable expliquée et une variable explicative, la plus connue étant la régression linéaire). Nous aborderons cependant rapidement les analyses multivariées qui permettent une analyse simultanée des relations existant entre différentes variables.

Pourquoi nous dit-on toujours qu'il est important d'avoir beaucoup d'échantillons pour faire une analyse correcte ?

Plus la taille de l'échantillon que nous suivons est importante, plus les caractéristiques de cet échantillon seront proches des caractéristiques de la population source. Une simple moyenne calculée sur des tailles d'échantillon différente pourra varier parfois de manière importante (exemple 11). Le chiffre magique de 30 réplicats nous est souvent fourni par la science comme étant la limite inférieure à respecter car elle permet en général de décrire de manière fiable la population source à partir d'un échantillon.

EXEMPLE 11

Soit une population de 50 individus pour lesquels on mesure un niveau de parasitisme noté sur une échelle de 0 (pas de parasite) à 10 (complètement parasité). Le taux moyen de parasitisme calculé sur les 50 individus est de 10.4. Si je calcule cette même moyenne du niveau de parasitisme sur un échantillon de 30 individus tirés au hasard parmi les 50, la moyenne est de 10.4. Si je refais ce calcul avec seulement 10 individus elle augmente à 11.8 et avec 5 individus elle tombe à 9.2 seulement.

Qu'est ce que la puissance d'un test ? Comment l'augmenter ?

Lorsque je fais un test statistique, je veux qu'il me permette de rejeter mon hypothèse si celle-ci est fausse et de l'accepter si elle est vraie. Je ne veux pas faire l'erreur d'accepter une hypothèse fausse, (risque appelé béta) ou de rejeter une hypothèse vraie (risque dit alpha). La difficulté est généralement la suivante : je n'ai que mon échantillon à analyser et je ne connais pas la vraie valeur dans ma population. Si je ne trouve pas de différence statistique entre deux situations que je veux comparer, je ne peux pas savoir si cela est dû à l'absence de différence entre les populations que je compare ou simplement au fait que mon test n'est pas assez puissant pour voir cette différence. La puissance d'un test statistique est la capacité de ce test à rejeter une hypothèse alors qu'elle est fausse (1-béta pour les plus matheux). Globalement, et c'est assez intuitif, plus la différence que l'on souhaite montrer est importante, plus le test sera puissant, et plus cette différence est ténue, moins ce sera facile de le faire. Un des moyens les plus simples pour augmenter la puissance d'un test statistique, c'est-à-dire pour augmenter sa capacité à se rapprocher de la vérité, est d'augmenter la taille de mon échantillon ! Il existe des tests de puissance qu'il est possible de mettre en œuvre avant de réaliser un suivi *(figure 5)*. Ils permettront de savoir combien d'objets statistiques (individus, placettes, etc) vous devez inclure dans votre étude pour que vos résultats soient fiables.

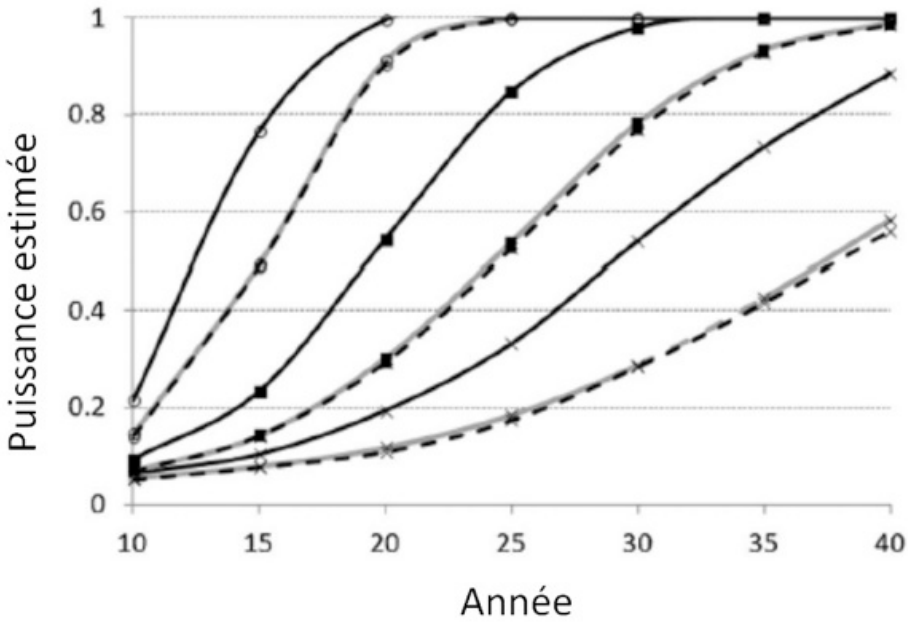


Figure 5 : Quel effort de terrain consacrer pour détecter les tendances d'une population de Tortue d'Hermann ? Trois techniques de suivi des populations ont été testées : la Capture-marquage-recapture (ligne grise), le distance sampling (ligne noire) et le N-Mixture (ligne pointillée). On veut évaluer la puissance de chaque méthode si la tendance annuelle de la population est de 1 % (croix), 2 % (carré) ou 5 % (cercle). On considère qu'une méthode doit parvenir à une performance de puissance de l'ordre de 0.8 (axe des ordonnées). Il faudra donc au minimum 15 ans de suivi (axe des abscisses) pour détecter une tendance de 5% avec le distance sampling… et 30 ans de suivi pour détecter une tendance annuelle de 2% avec la CMR et le N-Mixture, des recommandations au combien valable pour un gestionnaire ! *(d'après Couturier et al., 2013, Estimating Abundance and Population Trends When Detection Is Low and Highly Variable: A Comparison of Three Methods for the Hermann's Tortoise, The Journal of Wildlife Management 77(3) : 454–462; 2013)*

Quel est l'impact des erreurs de mesure sur le terrain ou des erreurs de saisie dans mon analyse ? Que puis-je faire pour le limiter ?

Au retour du terrain et lorsque la saisie des données est terminée, la première chose à faire est de les regarder afin de voir si certaines d'entre elles ne sont pas aberrantes (exemple 12).

EXEMPLE 12

J'ai fait des mesures de température, toutes sont comprises entre 15 et 20°C. Une seule mesure est à 27°C. S'agit-t-il d'une erreur de mesure (mon enregistreur a mal fonctionné), de saisie des données (je vérifie sur ma fiche de terrain) ou d'une valeur réellement observée ?

Si une valeur présente dans votre jeu de données est aberrante il est important de la retirer car elle peut avoir un impact fort sur le reste de votre analyse et fausser vos résultats. Enlever cette donnée n'est pas un problème si votre jeu de données est suffisamment grand. On peut aussi explorer l'impact de cette valeur sur les conclusions de l'analyse.

Est-ce que pour faire une analyse, il faut vraiment toujours que mes données soient normales ?

La normalité des données doit se regarder sous deux angles :

- celui de la variable expliquée ;
- celui des résidus.

La loi normale est une loi de probabilité et probablement la loi la plus utilisée pour décrire les phénomènes naturels. Elle se traduit par une courbe de Gauss que l'on appelle communément la courbe en cloche (unimodale c'est-à-dire un seul pic de données et symétrique) lorsqu'on représente en ordonnée la fréquence d'observation de la variable que l'on suit (en abscisse) (exemple 13). Il existe des tests statistiques pour vérifier la normalité des données mais déjà regarder cette courbe est très informatif : Est elle bien symétrique ? Est-elle centrée autour de la moyenne ?

EXEMPLE 13

Je suis le poids des graines d'une plante et, pour cela, j'échantillonne trois sites sur lesquels je sais qu'elle est présente (Figure 6). Lorsque je regarde l'ensemble de mes données, elles ne suivent pas la loi normale (courbe non symétrique) mais si je les regarde sur les 3 sites suivis, de manière indépendante, je me rends compte que pour chacun d'eux la distribution des données a une forme différente. J'observais des données que je pensais homogènes dans leur ensemble alors qu'elles ne le sont pas. Ici la non-normalité de mes données venait simplement d'un mélange de variables.

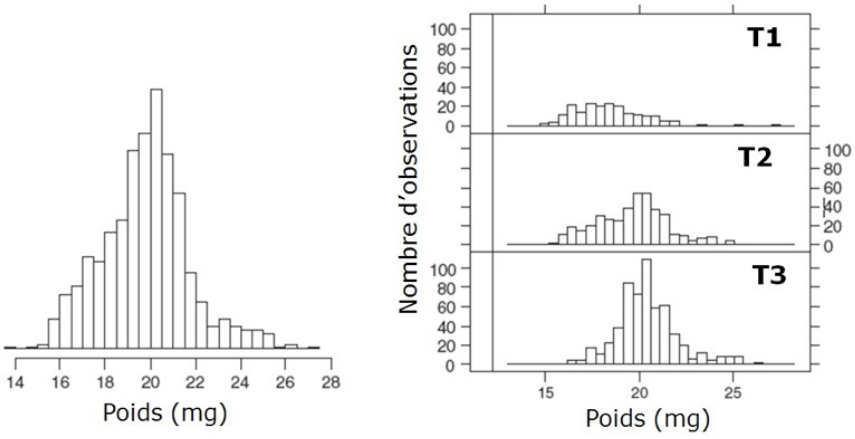


Figure 6 : Distribution du point des graines (en mg) pour les trois sites d'observation cumulés, T1, T2 et T3 (à gauche) et pour chacun des sites pris séparément (à droite).

L'élément de normalité le plus important à respecter est la normalité des résidus. Un résidu est la part de hasard, d'aléa, que le modèle n'explique pas. Si l'on prend l'exemple d'une simple régression linéaire, les points sont les observations, la droite de régression est le modèle. Il permet de prédire pour chaque valeur de la variable explicative une valeur de la variable expliquée qui est au plus proche de la réalité mais tout de même un peu différente. Cet écart entre la valeur réelle observée sur le terrain et la valeur prédite par le modèle est ce qu'on appelle un résidu *(figure 7)*. Lorsque vous regardez vos résidus, vous ne devez pas observer de patron particulier dans leur répartition. Pour simplifier ils doivent se répartir de manière homogène de part et d'autre de votre courbe (les résidus positifs ne doivent pas être tous en début de distribution et les négatifs en fin par exemple) et avoir tous une gamme de taille comparable (la valeur des résidus ne doit pas être très importante en début de distribution et très faible en fin par exemple).

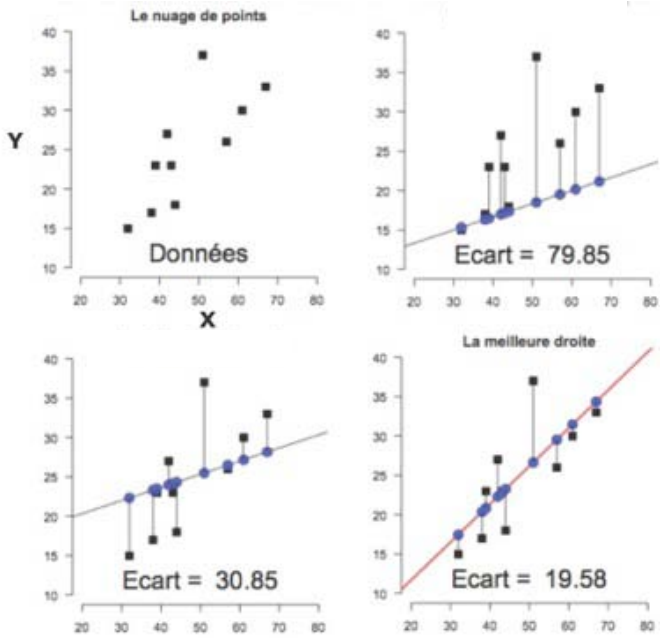


Figure 7 : X est la variable explicative et Y la variable à expliquer. On cherche un modèle simple permettant de prédire Y en fonction de X. La valeur Ecart est la somme des carrés des distances entre chaque observation et sa projection sur le modèle (ce sont les erreurs du modèle matérialisées par des segments verticaux). Le meilleur modèle en régression linéaire est celui qui minimise cette valeur
(Sylvain Doledec).

S'ils ne sont pas homogènes, c'est soit que vous avez ignoré une variable explicative qui structure fortement vos données, soit que le modèle que vous avez choisi (linéaire dans cet exemple) n'est pas le bon : il ne limite pas au maximum les résidus, et vous observez ce que l'on appelle de la sur-dispersion (écart trop grand pour une partie des données observées par rapport au modèle choisi). Votre modèle doit toujours minimiser la taille des résidus, c'est-à-dire expliquer la plus grande part possible des variations que vous observez. Vos résidus correspondent à ce qui n'est pas expliqué. Cette part de non-expliqué doit se distribuer normalement, ce qui confirmera que ce qui reste inexpliqué par votre modèle est vraiment du au hasard, à une variabilité naturelle de vos observations (différences de croissance entre des individus liée à leurs caractéristiques propres par exemple).

Dans cette phase d'exploration de vos données, dans le cas ou vous utilisez un modèle multivarié (plusieurs variables explicatives) un dernier élément reste à vérifier : celui de l'homogénéité des variances des variables explicatives. Cet estimateur de la dispersion des données doit être comparable, c'est-à-dire dans la même gamme de valeurs, entre toutes les variables explicatives. Il s'agit d'un pré-requis généralement difficile à respecter mais vers lequel il faut tendre. Il reste cependant possible de s'accommoder de différences de variances, soit en utilisant des tests adaptés soit en transformant les données pour homogénéiser les variances.

Pour conclure, il est important d'avoir à l'esprit qu'il existe tout un panel d'outils statistiques pour permettre l'analyse des données qui, par nature, n'ont pas une distribution normale. En particulier, dans le domaine des analyses paramétriques le modèle linéaire généralisé (GLM) permet d'étudier des variables réponses qui suivent d'autres lois que la loi normale (par exemple, loi de poisson, binomiale, négative binomiale). Par ailleurs, les tests dits non paramétriques permettent aussi de travailler sur tous les types de distributions et avec des échantillons de relativement petite taille. Ils sont souvent utilisés par les gestionnaires, car plus simples à mettre en œuvre, mais il faut cependant être conscient qu'à conditions égales (mêmes tailles d'échantillons par exemple), les outils non paramétriques sont parfois moins puissants que les outils paramétriques.

La normalité n'est donc pas une obligation pour faire des analyses mais doit être respectée lorsque la méthode utilisée l'exige. Il est donc important de se renseigner sur les conditions d'utilisation d'une méthode d'analyse avant de la mettre en œuvre.

L'analyse sera-t-elle plus intéressante et précise si je mets plus de variables explicatives pour le phénomène que j'observe ?

Plus on a de variables explicatives dans un modèle, plus il faudra un grand nombre d'échantillons pour conserver une puissance correcte pour l'analyse (voir plus haut). Les variables explicatives apportent évidemment des informations intéressantes mais il faut trouver un compromis entre leur intérêt et le surcroît de travail qu'elles occasionnent.

De plus, il peut arriver que des variables explicatives soient en fait corrélées entre elles (ce que l'on appelle la colinéarité des variables explicatives) c'est-à-dire que la distribution d'une de ces variables explicatives est en lien avec la distribution d'une autre variable explicative. Elles apporteront finalement d'une certaine manière la même information que vous les preniez toutes les deux en compte ou une seule d'entre elles. Une des hypothèses de la statistique est que les variables explicatives que l'on introduit dans un même modèle sont indépendantes. Il existe des méthodes qui permettent de « nettoyer » votre modèle de ces variables corrélées pour ne garder que les explicatives qui ont vraiment du sens (« Variance Inflation factor » pour des régressions linéaires multiples par exemple).

Il est également important de veiller, dans son jeu de données, à détecter les éventuelles phénomènes d'interaction entre les variables et de décomposer son jeu de donnée afin de voir si un lien observé sur le jeu de données dans son ensemble n'est pas, en fait, la résultante de phénomène différents observés sur des échelles infra (exemple 14).

EXEMPLE 14

On analyse le lien largeur du pétale et longueur du pétale pour 3 espèces d'iris *(figure 8)* **et l'on observe que ces deux paramètres augmentent ensemble de manière linéaire (graphique du haut). Cependant, si on fait la même représentation graphique en séparant les 3 espèces, on voit que le lien entre ces deux variables diffère entre les espèces (graphique du bas).**

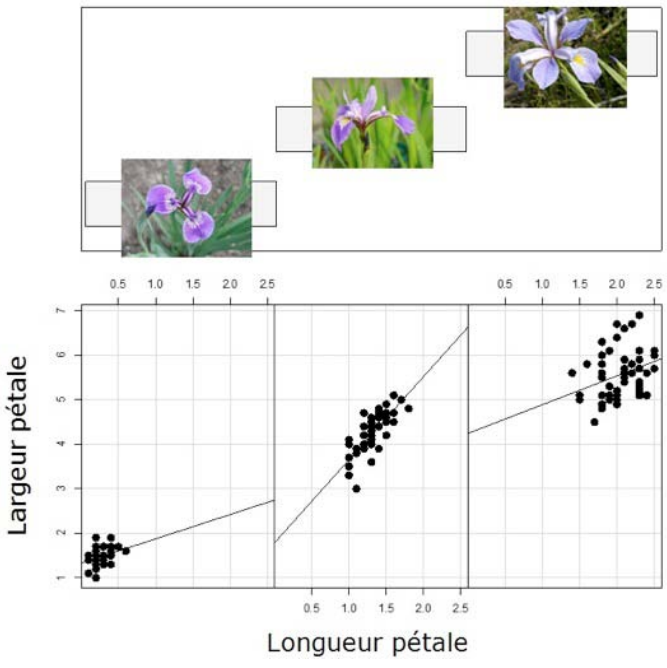


Figure 8 : Relation entre la largeur et la longueur du pétale pour 3 espèces d'iris *(en haut)* **et pour chacune des 3 espèces prises séparément** *(en bas)*, **Sylvain Doledec.**

Nous venons de voir qu'il pouvait exister des liens entre les variables que nous suivons, mais il est également possible qu'il y ait des liens entre les observations faites pour une seule variable. Les observations que nous faisons dans le cadre d'un suivi sont à la fois réparties dans le temps et dans l'espace. Lorsque l'on fait un suivi dans le temps, il faut tenir compte de la potentielle auto-corrélation temporelle entre les observations au sein d'une même variable explicative : c'est-à-dire le lien qui peut exister entre l'observation à un temps t et celle à un temps t+1. De même, si par exemple on met en œuvre des relevés floristiques le long d'un transect, il est possible que les observations soient corrélées spatialement entre-elles : la deuxième observation ressemble plus à la première qu'à n'importe quel autre point pris au hasard dans la zone par exemple. Il existe des méthodes permettant de prendre en compte cette auto-corrélation dans les analyses (calcul de l'indice de Moran pour l'aspect spatial par exemple).

Plus globalement, l'étape de pré-analyse des données est au moins aussi importante que l'analyse des données en elle-même. Elle l'est car elle évite de faire des erreurs, mais surtout parce qu'elle donne déjà de très bonnes informations sur les phénomènes que l'on observe et que l'on tente de décrire. Les représentations graphiques des données se suffisent parfois à elles-mêmes et la mise en œuvre de tests n'est pas une obligation.

Les approches multivariées sont-elles seulement descriptives ?

L'approche multivariée permet d'analyser les relations existant entre différentes variables en une seule fois sans avoir à tester leurs relations deux à deux. Elle n'est pas seulement descriptive, elle permet aussi le test d'hypothèses. Les tests mis en œuvre pour analyser la significativité ne sont simplement pas les mêmes que ceux développés en statistiques inférentielles. Il s'agit ici de tests de permutation. Ces derniers correspondent à un tirage au hasard des lignes des tableaux de données afin d'obtenir une distribution qui simule le hasard. Si la situation observée n'est pas due au hasard, elle se trouvera en dehors de ces valeurs simulées par permutation.

L'AFC (analyse factorielle des correspondances) est la méthode la plus connue chez les gestionnaires mais elle n'est pas la seule méthode qui existe. Comme précédemment le choix d'un type d'analyse multivariée que l'on pourra pratiquer dépend des données dont on dispose :

- **AFC** : s'applique à des tableaux croisés (appelés tables de contingence) de deux variables qualitatives. Cette méthode est très utilisée pour montrer de la corrélation entre espèces et relevés (utile pour réordonner un tableau par exemple) ;
- **ACP** (analyse en composantes principales) : utile pour les données quantitatives avec des unités différentes. C'est une méthode qui permet de construire une variable de synthèse qui résume au mieux toutes les variables analysées, c'est-à-dire qui est en corrélation maximale avec toutes ces variables ;
- **ACM** (analyse des correspondances multiples) : utile lorsque l'on dispose de plusieurs variables qualitatives. Elle permet d'obtenir une variable de synthèse qui séparera les données en groupe d'individus homogènes entre eux avec la plus grande différence possible entre les divers groupes ainsi formés.

Il existe également des méthodes permettant de relier deux tableaux, par exemple un tableau de données floro-faunistiques et un tableau de variables environnementales. Les ACP et AFC sur variables instrumentales (AFCVI, ACPVI) permettent une approche explicative d'un tableau par un autre, tandis que les analyses de co-inertie permettent d'identifier la covariance entre les deux tableaux. On peut ajouter à cette batterie, l'analyse RLQ et le test du 4^{e} coin associé qui permet de relier un tableau de données environnementales, un tableau de données floro-faunistiques et un tableau de traits des espèces.

CE QU'IL FAUT RETENIR

Un gestionnaire a besoin d'analyser ses données parce qu'il a constamment besoin d'évaluer son travail. Il évalue l'atteinte des objectifs à long terme qu'il s'est fixés dans le cadre d'un plan de gestion. Il évalue également l'atteinte de ses objectifs opérationnels via l'évaluation de sa gestion pour la faire évoluer si nécessaire et vérifier qu'il préserve bien les enjeux de son site.

Cependant, évaluer pour un gestionnaire n'est pas toujours faire des analyses statistiques. Un gestionnaire évalue le niveau de réalisation des actions qu'il a planifiées (nombre d'actions mises en œuvre par rapport au nombre d'actions prévues). Il est également amené à décrire un état attendu pour son site et à fixer des seuils empiriques (via de l'expertise) à partir desquels il considérera que ses objectifs sont atteints (la restauration d'une pelouse sèche par débroussaillage sera considérée comme réussie si l'on obtient un taux d'embroussaillement final inférieur à 20% par exemple). Une description fine d'un état attendu de manière concertée avec les acteurs du site et avec la définition de seuils de changement d'état avant la mise en œuvre de la gestion est également une bonne manière d'analyser les sites que nous gérons. Elle n'est pas en contradiction avec l'utilisation de tests statistiques, elle est simplement complémentaire, car l'utilisation de ces derniers ne se justifie par toujours.

Dans tous les cas, que l'on fasse des analyses statistiques ou non, les recommandations de ce guide s'appliquent, lorsqu'on s'engage dans la collecte et l'interprétation de données sur du moyen ou long terme, bref sur un suivi. La plus importante d'entre elles est certainement de suivre une méthode claire, dont on organisera toutes les étapes avant même de commencer. Pour toute question que l'on se pose, il est nécessaire que cette dernière soit formulée de manière claire et précise, et de disposer d'une méthode pour y répondre également bien pensée. Enfin il est important de savoir très tôt comment les données dont on dispose seront traitées et de formaliser la méthode d'analyse des observations, ou du moins d'interprétation. La plupart du temps, lorsqu'on parle de suivis scientifiques dans le monde de la gestion, on parle simplement de protocoles de collecte de données. Il est important de prendre conscience qu'un protocole de collecte de données n'a pas d'utilité s'il n'est pas accompagné d'une méthode d'analyse. Ceci est vrai, que cette dernière soit extrêmement simple ou particulièrement sophistiquée. Pour cela, il est impératif de penser à la phase d'analyse des données très en amont de la mise en œuvre des suivis car la plupart des choix qui sont faits lors de la détermination du plan d'échantillonnage ou du recueil des données aura des conséquences importantes sur les possibilités d'analyse.

C'est un comble de dire cela en conclusion de ce guide, mais un de ses apports majeurs, nous l'espérons, sera de faire comprendre qu'une analyse descriptive des données peut être parfois plus intéressante que la mise en place de tout un attirail statistique, surtout quand celui-ci est mal maîtrisé. Il est important de prendre conseil auprès d'experts en analyse de données, voire de réaliser les suivis en collaboration avec eux, de la même manière que vous avez l'habitude de consulter des experts naturalistes. Il est primordial de regarder ses données sous de nombreux angles avant de mettre en œuvre des tests. Cette phase de pré-analyse est un préalable indispensable, qui à la fois fait gagner du temps et évite de se tromper du tout au tout.

Pour ceux qui vont faire de la statistique malgré tout, cette synthèse explique certaines des contraintes qu'elle nous impose (taille d'échantillon, normalité des données..), montre que la statistique repose sur des hypothèses mathématiques et que de leur respect dépendra la qualité des résultats.

Planifier son analyse c'est aussi prendre conscience qu'il faudra y consacrer du temps, temps qui parfois ne sera pas négligeable si vous être peu expérimenté dans le domaine.

Comme pour le reste de ses activités, dans le domaine des suivis, un gestionnaire doit prioriser les questions scientifiques qu'il se pose, s'investir à fond sur les suivis qu'il juge prioritaires et éviter de les multiplier alors que les ressources dont il dispose dans ce domaine sont généralement limitées. Rien n'est pire que d'avoir investi beaucoup de temps et d'énergie dans la collecte de données dont on ne peut rien faire au final.

Enfin, il est toujours important de savoir faire parler une étude locale au regard d'un contexte plus large (régional par exemple) et également dans l'autre sens, de savoir décliner des enjeux régionaux au regard du contexte local. Pour les suivis, comme pour le reste, il est important de savoir regarder ce qui se passe en dehors de notre espace de travail (quels impacts sur mon site viennent de l'extérieur ? Quelle part de ce que j'observe ne dépend pas de ma gestion ?...). Ce jeu sur les échelles spatiales (dont on peut faire le parallèle sur les échelles temporelles !) peut notamment s'opérer plus facilement parce que vous mutualisez vos suivis avec des partenaires, ce qui vous donnera une vision plus large que votre seul site. Il est en effet toujours important de se questionner sur la spécificité de vos résultats, sur leur transférabilité (validité dans d'autres cas), sur leur généralisation (validité dans tous les cas).

CONSEILS DE LECTURE SUR LE SUJET :

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.

D. Poinsot, 2004. Statistiques pour statophobes. [en ligne : *http://perso.univ-rennes1.fr/denis.poinsot]*

Outil d'analyse de données :

Logiciel R téléchargeable gratuit pour lequel de nombreuses aides sont disponibles : *https://www.r-project.org*

Ce guide a été rédigé pour faire suite à une journée technique organisée le 18 décembre 2015 à l'Agrapole de Lyon par le CEN Rhône-Alpes et le conseil scientifique des CEN de Rhône-Alpes sur la mise en œuvre et l'analyse des suivis de la végétation.

Il s'inspire des interventions des différents participants à cette journée d'échange et à été rédigé avec leur aide :

Aurélien Besnard *(MCF EPHE, CEFE Montpellier)*

Cédric Dentant *(chargé de mission, Parc national des Écrins)*

Sylvain Doledec *(CNRS, UMR 5023 – LEHNA, Lyon 1)*

Perrine Gauthier *(IR CNRS CEFE Montpellier)*

Aurélien Letourneau *(IR CNRS CEFE Montpellier)*

Claude Miaud *(DR CNRS CEFE Montpellier)*

Gilles Pache *(CBNA)*

John Thompson *(DR CNRS CEFE Montpellier)*

Thomas Tully *(MC, UMR 7618, Université Pierre et Marie Curie)*

Dominique Vallod *(Enseignante-chercheuse, ISARA)*

Anne Wolff *(Chargée d'étude, CEN Rhône-Alpes)*

Avec la collaboration de **Véronique Bonnet** (*CBNA*) et du **conseil scientifique des Conservatoires de Rhône-Alpes** lors de la conception de cette journée d'échange notamment pour la recherche des intervenants.

AUTOMNE 2016

agence de l'eau RHÔNE MÉDITERRANÉE CORSE établissement public de l'État
La Région Auvergne-Rhône-Alpes
isaralyon Une école d'ingénieurs au cœur de la vie
Conservatoire d'espaces naturels Rhône-Alpes
Conservatoire d'espaces naturels Savoie
Avenir Conservatoire d'espaces naturels Isère
Asters Conservatoire d'espaces naturels Haute-Savoie