



## **Paysage et régulation des bio-agresseurs : approches méthodologiques et connaissances actuelles**

**par**

Claire Lavigne -UR PSH, Plantes et systèmes de culture horticoles, INRA, Avignon

Sandrine Petit -UMR Agroécologie, INRA, Dijon

**Document issu d'une présentation faite le 7 janvier 2017 au GT1 du RMT « agriculture et biodiversité »**

### Plan de la présentation

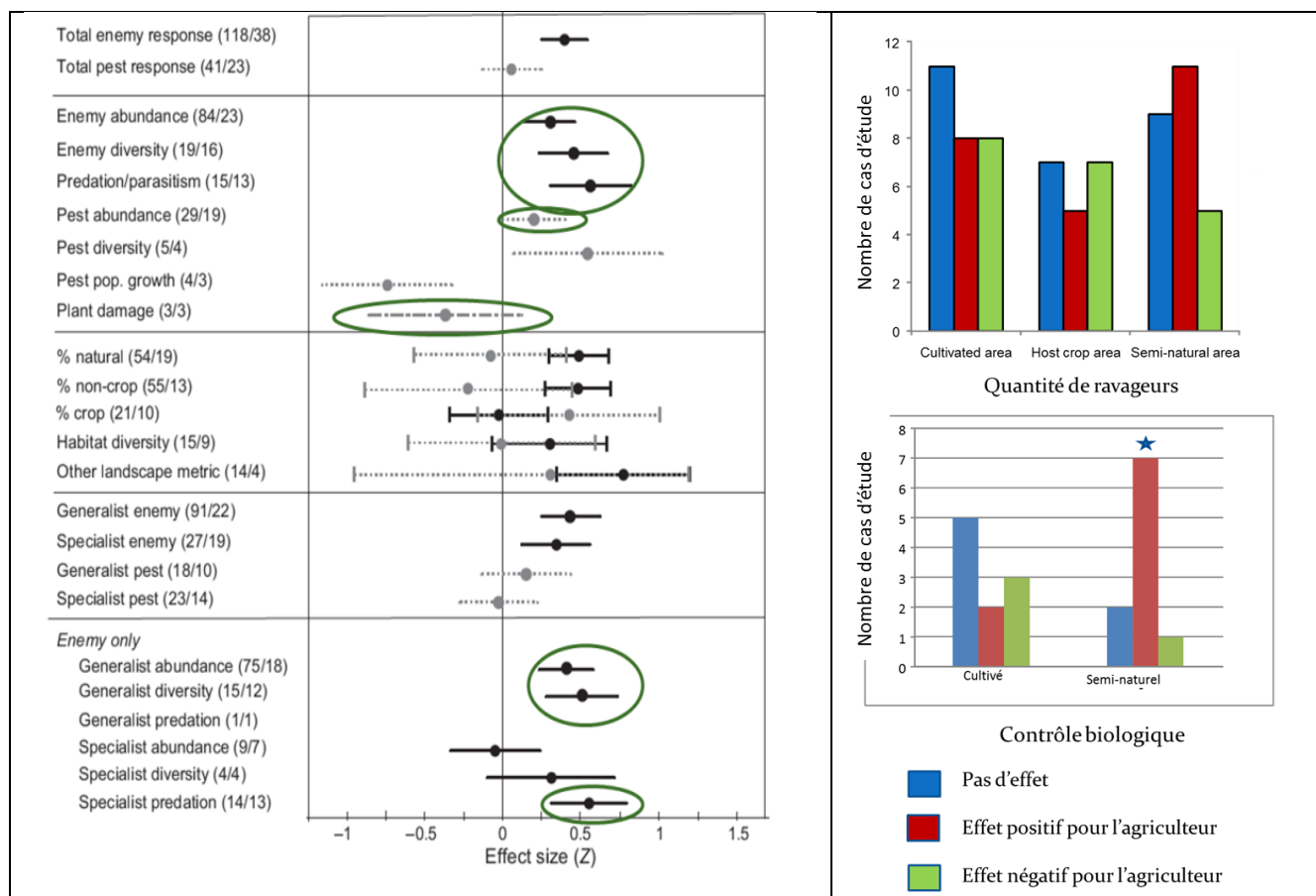
- Introduction
  - Principe du cadre d'analyse (~buffers)
  - Deux cas d'études
- Le paysage
  - Que représenter?
  - Sources de données
  - La numérisation
  - Synthétiser les données de paysage – quelles métriques?
  - Choix des parcelles
- Les mesures de contrôle biologique (rapidement, déjà discuté)
  - Type de données
  - Les sources de variation à contrôler
- Analyses des données
  - Réfléchir à toutes les variables autres que paysage (pratiques? Taille des parcelles...)
  - Méthodes statistiques
- Quelques résultats de la littérature

## PREAMBULE

Avant de se lancer dans une analyse de l'effet du paysage sur le contrôle biologique, il est intéressant d'avoir en tête quelques résultats de la littérature. Cela aide à réfléchir aux caractéristiques des situations (paysages, ravageurs, auxiliaires) où on peut attendre un effet de la composition du paysage sur le contrôle biologique.

### Quelques résultats sur les effets de la composition du paysage sur le contrôle biologique

Une méta-analyse très citée de la littérature (Chaplin-Kramer et al. 2011 dans *Ecology Letters*) étudie l'effet de la complexité du paysage (mesurée soit par la proportion d'habitats semi-naturels soit comme l'inverse de la quantité de cultures) sur différentes mesures liées au contrôle biologique. Sur la figure, les résultats concernant les ennemis naturels sont en trait plein noir, les résultats concernant les ravageurs en petits pointillés gris et les dégâts en gros pointillés gris. Quand les points et leurs intervalles sont à droite de la ligne verticale, l'effet de la complexité du paysage est positif. On voit sur la figure dessous à gauche que s'il y a en moyenne un effet positif de cette complexité sur les ennemis naturels, l'effet n'est en moyenne pas significatif sur les ravageurs ou les dégâts. De même la figure à droite (Veres et al. 2013 dans *Agriculture Ecosystem Environment*) montre un effet positif en moyenne sur le parasitisme ou la prédation des surfaces en habitats semi-naturels mais pas sur les ravageurs. Ceci soulève des questions sur les situations où la composition du paysage diminue bien les ravageurs (il y en a !) et celles où il n'a pas d'effet ou un effet inverse.



D'après Chaplin-Kramer et al. 2011, *Ecology Letters* et d'après Veres et al. 2013 *Agriculture, Ecosystem, Environment*

### Quelques indications sur les situations où on peut attendre un effet du paysage sur le contrôle biologique

Une étude récente publiée par Tscharrntke et al. dans *Biological conservation* en 2016 s'intéresse aux différentes causes possibles d'une absence d'effet du paysage sur le contrôle biologique. Elles sont illustrées sur la figure suivante tirée de cette publication.

- (a) Il n'y a pas d'ennemis naturels dans la région
- (b) Les habitats semi-naturels sont une source de ravageurs
- (c) Les cultures fournissent plus de ressources que les habitats semi-naturels
- (d) L'habitat naturel est insuffisant en composition, proximité, quantité...
- (e) Les pratiques agricoles ne permettent pas l'établissement des ennemis naturels

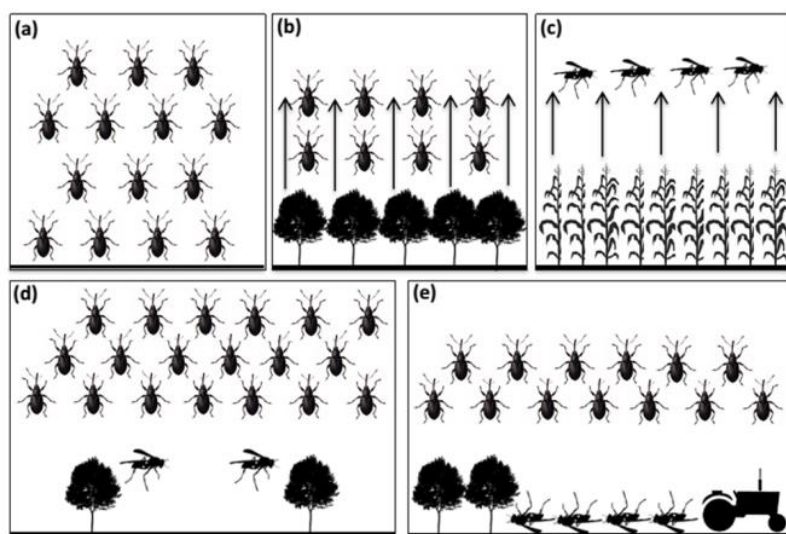
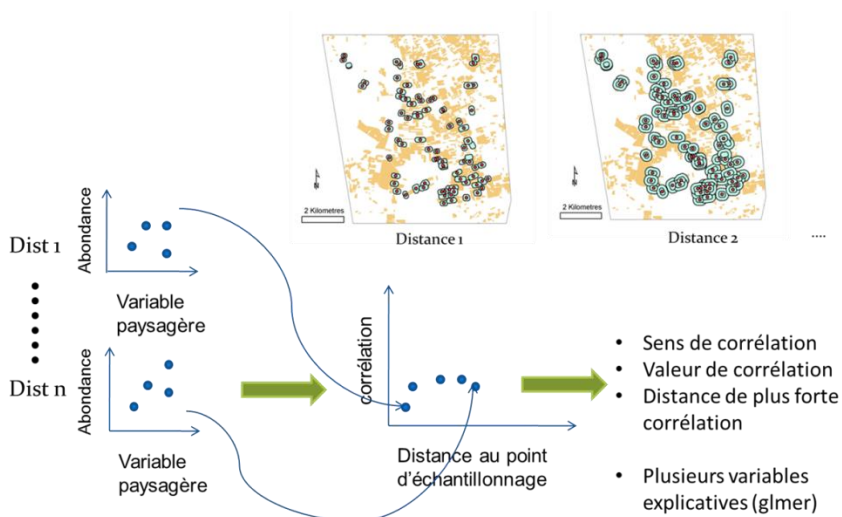


Figure tirée de Tscharrntke et al. (2016) *Biological conservation*.

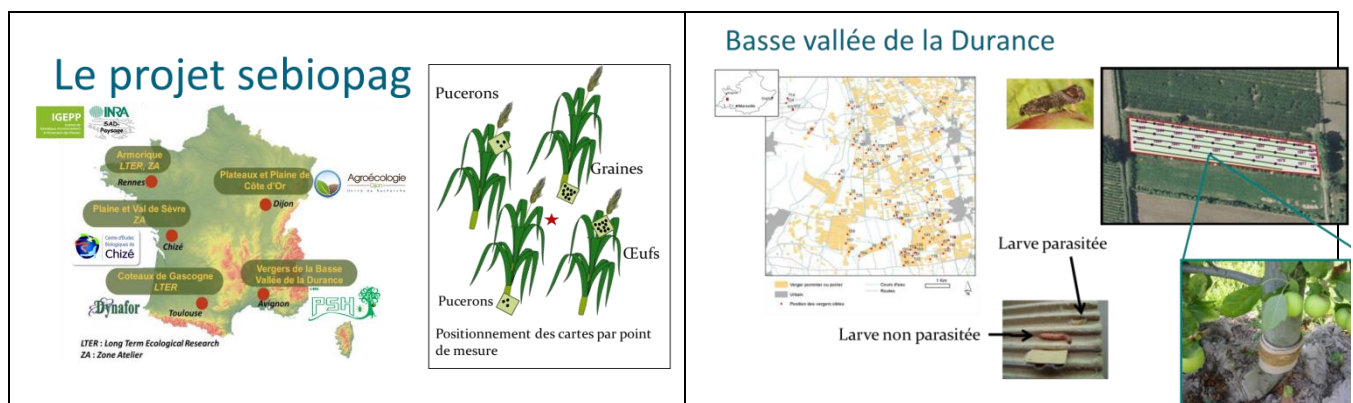
### INTRODUCTION AUX METHODES D'ANALYSE

Le principe des analyses proposées dans ce document consiste à mettre en relation des variables biologiques (abondances de ravageurs, taux de prédation....) avec des caractéristiques du paysage mesurées dans des 'buffers' dessinés autour des parcelles ou des points d'échantillonnage.

Ces relations (ici des corrélations) peuvent être calculées pour plusieurs tailles de buffers (distances). On regarde ensuite des caractéristiques de ces corrélations : sens positif ou négatif, valeur, distance de plus forte corrélation... Lorsqu'il y a plusieurs variables explicatives locales ou paysagères, d'autres modèles statistiques sont possibles (modèles linéaires généralisés par exemple) (illustration).



Les données ayant servi de base pour cet exposé proviennent en partie du projet sebiopag (<http://sebiopag.inra.fr/>) qui vise notamment à comprendre l'impact de la structure du paysage sur la prédation de proies sentinelles, et des suivis en basse vallée de la Durance qui visent notamment à comprendre l'impact de la structure du paysage sur l'abondance du carpocapse des pommes en vergers ainsi que sur sa prédation et son parasitisme (illustration).



## LE PAYSAGE

### Que représenter ?

En théorie:

Ce qui est a priori important pour le processus écologique étudié :

- Habitats des ravageurs / auxiliaires; pesticides; corridors; sources lumineuses?...
- Une description fonctionnelle (adaptée aux rôles des éléments paysagers sur le processus biologique)

En pratique:

Ce qui est a priori important biologiquement et:

- qui peut être cartographié correctement (pratiques agricoles?)
- qui varie correctement entre les points d'échantillonnage
- en regroupant ce qui peut avoir la même fonction (ex: différents milieux ouverts pour une espèce forestière?)

Souvent dans la littérature: des types de cultures (pérennes, annuelles, céréales...), des éléments semi-naturels, des surfaces artificialisées...

*Exemples :*

Carpocapse : vergers pommes poires + arbres isolés hôtes (noyers, cognassiers...)

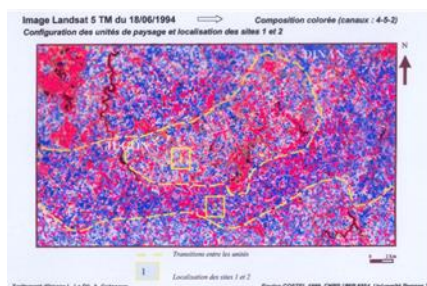
Adventices: historique des rotations...

Syrphes : ressources florales d'après leurs préférences, bois pour hibernation...

### **Les sources de données**

Elles sont très diverses. Le site de l'IGN est très riche (<http://www.ign.fr/>). Certaines données sont gratuites pour les organismes de recherche publique ou d'enseignement, ou, moyennant le montage d'un dossier, pour les projets d'intérêt public (ex : projet CASDAR). On peut citer aussi la base européenne Corine Land Cover qui est libre (<http://www.statistiques.developpement-durable.gouv.fr/donnees-ligne/li/1825.html>). La base Corine Land Cover est disponible pour les années : 1990, 2000, 2006 et 2012 ; l'occupation du territoire est détaillée en 5, 15 ou 44 postes. Les échelles couvertes sont : France métropole, grandes régions, régions, départements, communes. Les données satellites demandent des traitements de télédétection et donc des collaborations. Malgré toutes ces données disponibles, il faut faire des vérifications / mises à jour sur le terrain. Les résultats sont parfois assez différents. Quelques exemples de données sont illustrés ci-dessous.

Images satellitaires



Photographie aérienne (géoportail + IGN)



Cartes (géoportail + IGN)

RPG



Données de terrain!!

### **La numérisation des données**

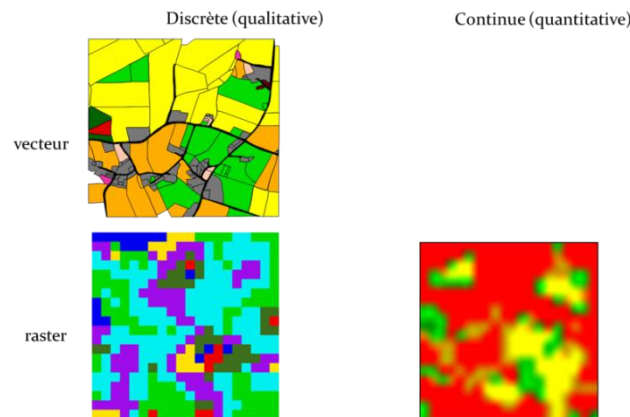
Elle consiste à dessiner les cartes à partir des photos dans un logiciel de SIG. Les données peuvent être représentées sous forme de polygones et lignes (mode vecteur) ou sous forme de pixels (mode raster). En général à chaque polygone ou pixel on affecte une occupation du sol représentée par une couleur (variable qualitative). On peut aussi choisir de représenter pour chaque pixel une variable quantitative (par exemple la diversité des habitats dans le pixel, une intensité de rayonnement infra-rouge (lié à la photosynthèse)...) (illustration).

Lors de la numérisation se posent deux questions qui doivent être résolues en commun par les partenaires du projet, en plus du choix des objets à représenter :

-quelle sera l'étendue de la surface numérisée ?

-quel sera le grain ? en mode vecteur cela revient à choisir la surface/longueur minimale d'un élément numérisé (par exemple est-ce qu'un arbre isolé est considéré comme un bois ? à partir de quelle surface boisée considère-t-on que l'on a un bois) ; en mode raster cela revient au choix de la taille du pixel.

De nombreux logiciels SIG existent. Un logiciel libre couramment utilisé est QGIS.

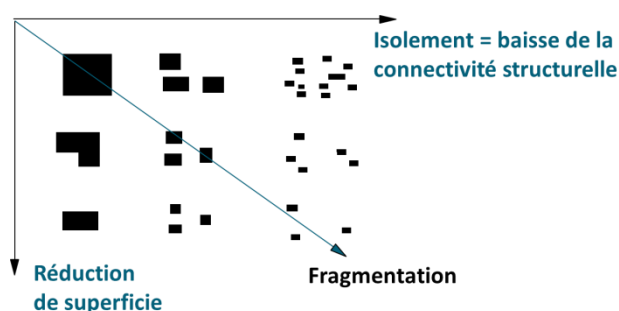


### Synthétiser les données de paysage – quelles métriques?

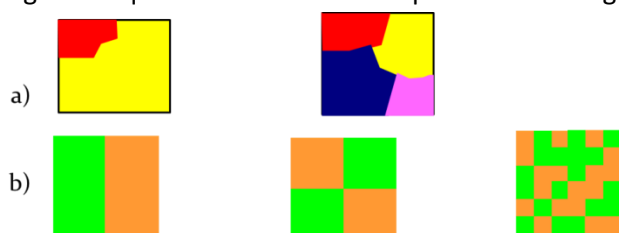
Une première série de métriques décrit **la composition** du paysage, c'est-à-dire la variété et l'abondance des types de classes constituant le paysage. L'abondance est généralement décrite en proportion de la surface du paysage (par exemple : 45% de forêt de feuillus). La variété peut être décrite en termes de nombre d'occupation du sol ou de diversité (indice de Shannon par exemple).

Une seconde série de métriques décrit **la configuration** du paysage, c'est-à-dire les caractéristiques géométriques et l'agencement des éléments.

Si on considère un seul type d'habitat, on peut considérer le nombre de taches de cet habitat, leur taille moyenne, la distance entre taches. Ces informations renseignent sur la **fragmentation** de l'habitat (illustration). La distance entre taches renseigne également sur la **connectivité** des habitats qui peut être structurelle (on regarde les distances physiques entre taches) ou fonctionnelle (on regarde entre quelles taches l'organisme étudié peut se déplacer facilement).



Si l'on considère l'ensemble des occupations du sol, on peut s'intéresser à **l'hétérogénéité** du paysage. Elle augmente quand le nombre d'occupation du sol augmente (a) et quand le nombre de patches augmente (b).





Il existe une grande diversité de métriques qui peuvent être regroupées en 8 classes. Certaines peuvent s'appliquer au niveau de la parcelle, d'autres à une classe d'habitat, d'autres à l'ensemble du paysage. On peut utilement regarder l'aide des logiciels libres FRAGSTAT ou CHLOE pour le calcul de ces indices. Le plus difficile est de choisir les métriques pertinentes !

1. Surface, densité, bordure (Area, density, edge metrics)
2. Forme (Shape metrics)
3. Noyau (Core area metrics)
4. Isolement, proximité (Isolation, proximity metrics)
5. Contraste (Contrast metrics)
6. Contagion, interspersion (Contagion, interspersion metrics)
7. Connectivité (Connectivity metrics)
8. Diversité (Diversity)

### ***Choix des parcelles à échantillonner***

Lorsque cela est possible, le choix des parcelles doit être guidé par les principes suivants :

- Les placer sur un gradient de paysage (ou quelques valeurs de variables contrastées) : par exemple dans des paysages plus ou moins boisés si c'est l'effet des bois que l'on regarde.
- Les placer suffisamment éloignées les unes des autres pour éviter les chevauchements de buffers et pour que les parcelles soient relativement indépendantes.
- Les choisir assez nombreuses. Lorsqu'un seul facteur du paysage est étudié on trouve en général des études autour de 15-30 parcelles dans des projets qui travaillent sur une région avec des parcelles sur un gradient paysager et éventuellement plusieurs années de suivi. Mais il y a de plus en plus d'études nationales ou internationales qui analysent des jeux d'au moins une centaine de parcelles. La question n'est pas tout à fait la même, dans un cas on cherche à savoir si localement le paysage a un effet, dans l'autre on cherche si cet effet existe de manière générale, dans des situations climatiques et géographiques différentes. Ce qu'il faut garder à l'esprit c'est que les parcelles doivent être d'autant plus nombreuses que l'effet attendu du paysage est faible et qu'il y a des risques de facteurs confondants incontrôlés (cf dessous).
- Eviter les principaux effets confondants (c'est-à-dire qu'il y ait une covariation entre deux variables de paysage). Par exemple si les parcelles qui sont dans les zones boisées sont également les plus petites, il ne sera pas possible de savoir si les différences observées entre parcelles sont dues à leur taille ou à la présence de bois dans le paysage.
- Eviter les sources de "bruit" (si possible : une espèce cultivée, une même taille des parcelles, un même mode de conduite....)

### **LES MESURES DE CONTROLE BIOLOGIQUE**

Il y a un très grand nombre de mesures plus ou moins directes du contrôle biologique qui ont déjà été discutées dans d'autres ateliers. Certaines sont listées dessous. Les couleurs marquent le type de variable (**nombre**, **proportion**, **ratio**...) qui aura une importance pour l'analyse statistique car elles ne suivent pas toutes des lois normales. Par ailleurs il peut être important de réfléchir aux covariables à mesurer également pour aider à l'interprétation des résultats (par exemple un taux de parasitisme va sans doute dépendre de l'abondance du ravageur, de même un taux de positif à la détection PCR va dépendre aussi de l'abondance de la proie...).

- **Abondances** d'un taxon d'auxiliaire
- Diversité (**richesse spécifique**, shannon) d'auxiliaires
- **Ratio** auxiliaires/ravageurs
- Taux de parasitisme (covar: abondance ravageur?)
- Taux de prédation de proies sentinelles (covar: abondance de proies)
- Taux de positifs PCR (covar: abondance de proies)
- **Nombre** de ravageurs (covar: pratiques)
- **Variation** du nombre de ravageurs (ratio entre deux dates? Différence entre deux dates?)



Larves de carpocapse  
saine et parasitée



Oeufs sentinelles de carpocapse

## ANALYSES DES DONNEES

Selon le type de données, la structure des paysages, le nombre de parcelles, le fait qu'on mesure plusieurs fois ou pas la même parcelle... les analyses peuvent différer. Nous donnons juste quelques indications et points de vigilance ci-dessous.

### Principe

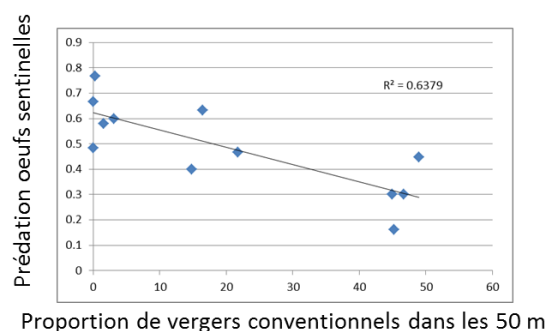
Le principe est de mettre en relation une mesure biologique avec la variable du paysage d'intérêt et d'autres variables mesurées qui pourraient avoir un effet sur la mesure biologique (cf. les covariables citées précédemment).

$$\text{Mesure biologique} \sim \text{autres variables} + \text{paysage}$$

La manière dont on les met en relation fait généralement appel à des corrélations ou des régressions (linéaires ou non...) même si d'autres méthodes existent.

### Exploration des données

C'est la première étape. Pour chaque variable (de contrôle biologique, du paysage ou autre) il est recommandé de regarder sa gamme de variation, sa distribution (est-ce qu'il y a beaucoup de 0 par exemple). On peut aussi explorer les données par exemple en traçant les relations entre variable de contrôle biologique et variable du paysage ou en comparant la moyenne du contrôle biologique pour différentes modalités d'une variable (par exemple le taux de prédation est-il en moyenne le même entre des parcelles avec et sans haie ?). Cela aide à se faire une idée de l'information qu'il y a dans les données.





Exemple de relation d'exploration de données qui donne une idée de facteur explicatif possible : lien entre pourcentage de prédation d'œufs de carpocapse sentinelles en verger et proportion de paysage couvert par des vergers conventionnels dans les 50m du bord de la parcelle.

### ***Test des corrélations entre variables explicatives***

Un point important de cette étape est également de tester si plusieurs des variables explicatives sont corrélées si on n'a pas pu l'éviter au moment du choix des parcelles (cf. effets confondants). Si plusieurs variables sont fortement corrélées, il vaut mieux n'en mettre qu'une dans l'analyse statistique et dans l'interprétation être prudent sur le lien réel. Par exemple si les parcelles plus petites sont celles qui sont dans des paysages plus boisés (corrélation positive entre proportion de bois et taille des parcelles) on peut trouver une relation positive entre bois et taux de prédation qui est en fait due à la taille des parcelles (ou l'inverse).

### ***Regarder la distribution de la variable à expliquer / des résidus***

Ce point est un peu plus technique. Certains tests statistiques font des hypothèses sur les variables (les analyses de variance classiques par exemple font l'hypothèse que la variable a une distribution normale et que la variance est homogène entre traitements). Certains tests sont plus ou moins sensibles à ces hypothèses. La conformité avec les hypothèses s'analyse généralement en regardant les résidus des modèles. Si ceux-ci sont trop 'bizarres' il peut être intéressant de faire des transformations des données. Par exemple (arc sinus d'une proportion stabilise la variance ; quand le nombre  $n$  d'individus par point de donnée est  $< 5$ , une transformation ( $\log n+1$ ) peut être efficace. On peut aussi déclarer d'autres distributions de données (modèles linéaires généralisés).

Enfin les relations entre variables peuvent ne pas être linéaires.

### ***Prendre en compte la structure des données***

Comme souvent en statistique, il est important de se demander quel est le niveau de répétition dont on dispose pour tester un facteur, c'est-à-dire de combien de données indépendantes on dispose pour tester l'effet de ce facteur. Voyons un exemple où on veut tester l'effet d'un facteur F (deux modalités A, B). Si on mesure plusieurs points dans une parcelle de type A, ces différentes mesures ne sont pas indépendantes : les pratiques agricoles sont les mêmes, la variété cultivée aussi, si les points sont assez proches la même population d'auxiliaires peut être responsable de la prédation... ; elles ont donc beaucoup plus en commun que le facteur A et ne sont pas des répétitions de la modalité « A ». Pour tester l'effet F il faudrait donc plusieurs parcelles de type A. Pour résumer l'ensemble des données d'une même parcelle (par exemple l'abondance d'un ravageur) on peut faire une moyenne de l'ensemble des mesures de la parcelle, ou plus efficacement, mettre un effet aléatoire 'parcelle' dans des modèles de régression (ou toutes analyse de variance). Ceci permet de décrire l'emboîtement spatial (plusieurs données au sein d'une même parcelle) et le fait que les données au sein d'une parcelle ne sont pas indépendantes. Le même type de question peut se poser quand une même parcelle est échantillonnée plusieurs fois puisque, de la même façon, des données mesurées à plusieurs dates sur une même parcelle ne sont sans doute pas indépendantes (une parcelle très attaquée à une date, a, par exemple, plus de chances de l'être à la date suivante).

## **CONCLUSIONS**

Pour conclure ce bref aperçu des questions méthodologiques pour mener à bien des études sur le lien entre paysage et contrôle biologique, avant de se lancer il vaut mieux :

- Bien choisir les variables par rapport à la question biologique

- Echantillonner dans un nombre suffisant de parcelles si possible choisies par rapport à cette question en faisant attention aux facteurs confondants
- Réfléchir en amont aux analyses statistiques (plus faciles avec un 'bon' dispositif)