

## Directive pour l'évaluation biologique des produits phytosanitaires

### MISE EN PLACE ET ANALYSE DES ESSAIS D'ÉVALUATION BIOLOGIQUE

#### Champ d'application spécifique

Cette norme destinée à être utilisée en association avec les Normes OEPP de la série PP 1 (Directives pour l'évaluation biologique des produits phytosanitaires) fournit des conseils détaillés sur la mise en place et l'analyse des essais d'évaluation biologique.

#### Approbation et amendement spécifiques

Approbation initiale en septembre 1989.  
Révision approuvée en septembre 1998.

Cette directive a pour but de fournir des renseignements d'ordre général concernant la mise en place et l'analyse des essais pour l'évaluation biologique des produits phytosanitaires pour lesquels les Directives OEPP pour l'évaluation biologique des produits phytosanitaires donnent des instructions plus détaillées pour chaque combinaison hôte-ravageur. La mise en place d'un essai est étudiée en premier lieu (dispositif expérimental, disposition et taille des parcelles, rôle et implantation des témoins non traités). La nature des observations à réaliser est ensuite examinée (types de variables, modes d'observation). Enfin, des suggestions sont faites sur l'analyse statistique des résultats d'un essai et d'une série d'essais (estimation des effets, choix du test statistique, transformation des variables). Des exemples d'échelles utilisées dans les directives OEPP sont présentés à l'annexe I.

#### 1. Dispositif expérimental

##### 1.1 Types de dispositifs

Les directives OEPP pour l'évaluation biologique des produits phytosanitaires prévoient que les traitements expérimentaux concernent les "produit(s) à étudier, produit(s) de référence, et le témoin non traité; parcelles disposées selon un dispositif statistique adéquat". Il est également envisagé que les produits soient étudiés à des doses et/ou à des dates

d'application différentes. Ceci concerne surtout l'utilisation de la dose double dans les essais de sélectivité des herbicides.

Les dispositifs monofactoriels conviennent pour les essais si le(s) produit(s) à étudier, le(s) produit(s) de référence et le témoin non traité peuvent être considérés comme plusieurs valeurs d'un même facteur, et qu'aucun autre facteur ne doit être utilisé. Par contre, si, par exemple, l'effet de chaque produit dans un essai d'évaluation d'un herbicide doit être étudié à différentes doses, un dispositif factoriel doit être utilisé, avec en général toutes les combinaisons de traitements possibles pour les deux facteurs représentés. Des interactions importantes entre les facteurs peuvent ainsi être mises en évidence et estimées.

Les principaux dispositifs aléatoires susceptibles d'être utilisés sont la randomisation totale et les blocs randomisés. Ces dispositifs sont illustrés ci-dessous à partir d'un exemple monofactoriel comprenant 8 traitements (5 produits à étudier, 2 produits de référence et un témoin non traité) et 4 répétitions.

##### 1.1.1 Randomisation totale

La randomisation totale (Figure 1) est le dispositif potentiellement le plus puissant du point de vue statistique (probabilité maximale de détecter une différence significative si elle existe), parce qu'il permet de conserver le maximum de degrés de liberté

**Figure 1.** Randomisation totale. Chaque traitement, numéroté de 1 à 8, est répété quatre fois; les traitements sont attribués complètement au hasard aux 32 parcelles

8	7	3	7	2	3	7	4
1	2	6	2	2	3	4	6
8	6	5	4	4	8	1	5
1	5	5	8	1	7	3	6

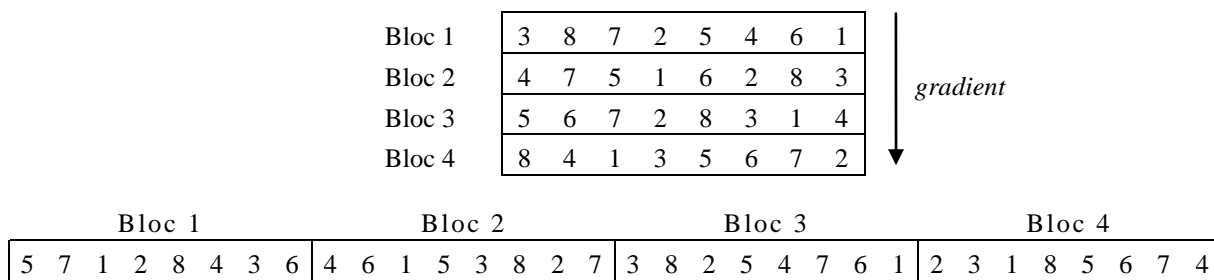
pour la variance résiduelle. Il n'est cependant utilisable que si l'on est certain que le site de l'essai est très homogène. Si l'hétérogénéité entre les diverses parties du site d'essai devient importante, la variance résiduelle aura un niveau inacceptable. Il convient alors d'avoir recours à un dispositif qui tient compte de cela, tel qu'un dispositif en blocs randomisés.

### 1.1.2 Blocs randomisés

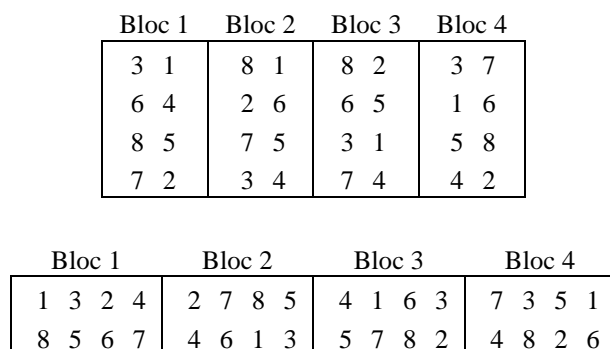
Un bloc est un groupe de parcelles dans lesquelles l'environnement est homogène. Dans ce dispositif, les blocs sont créés délibérément pour que les parcelles qui les composent soient aussi semblables que possible. Chaque traitement figure une et une seule fois dans

chaque bloc. Les traitements sont distribués de manière aléatoire à l'intérieur des blocs, qui jouent le rôle de répétitions. L'attribution des traitements dans chaque bloc doit être aléatoire et être faite séparément pour chaque bloc. Dans les exemples ci-dessous (Figure 2 à 4), il y a 4 blocs et 8 traitements. La disposition des blocs a pour but de contrôler l'hétérogénéité du terrain (par ex. pente, direction du travail lors du semis ou de la plantation, exposition, degré d'infestation etc.) ou des conditions lors de l'essai (application des traitements, notations). La disposition des blocs nécessite donc une connaissance préliminaire du site d'essai.

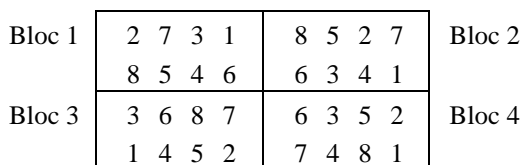
**Figure 2.** Disposition possible des blocs et des parcelles pour des blocs randomisés dans un essai de plein champ. L'environnement présente un gradient le long du champ et on en tient compte, soit en disposant les blocs dans le sens du gradient, soit en les plaçant côte-à-côte. Dans chaque cas, les parcelles au sein de blocs disposés perpendiculairement au gradient sont influencées de la même manière par la variable environnement.



**Figure 3.** Disposition possible des blocs et des parcelles pour des blocs randomisés dans un essai de plein champ. Alternative pour un dispositif en blocs randomisés dans le cas où il n'existe pas de gradient environnemental visible mais la distance maximale entre les parcelles dans un même bloc est relativement grande et laisse supposer une hétérogénéité. Les 8 parcelles sont relativement proches dans un rectangle de 4 par 2, et les blocs sont côte-à-côte.



**Figure 4.** Autre exemple du cas de la figure 3. De nouveau, les 8 parcelles sont relativement proches dans un rectangle de 4 par 2, mais les blocs sont disposés selon un rectangle de 2 par 2.



Les blocs ne sont pas forcément adjacents. Une bonne connaissance préliminaire d'un champ peut être utilisée pour disperser les blocs dans l'ensemble du champ pour tenir compte d'une hétérogénéité observée précédemment (Figure 5).

On peut tenir compte d'une hétérogénéité complexe si l'on connaît bien le terrain et qu'on peut être confiant que les conditions vont sûrement se maintenir pour l'expérimentation à venir; il n'est alors même plus nécessaire que les parcelles d'un même bloc soient adjacentes. Par exemple, les blocs peuvent être éclatés pour tenir compte d'une infestation irrégulière de nématodes. Dans la figure 6, les parcelles du bloc 1 ont été placées délibérément à des points d'infestation visiblement faible, et les parcelles du bloc 2 à des points d'infestation visiblement forte.

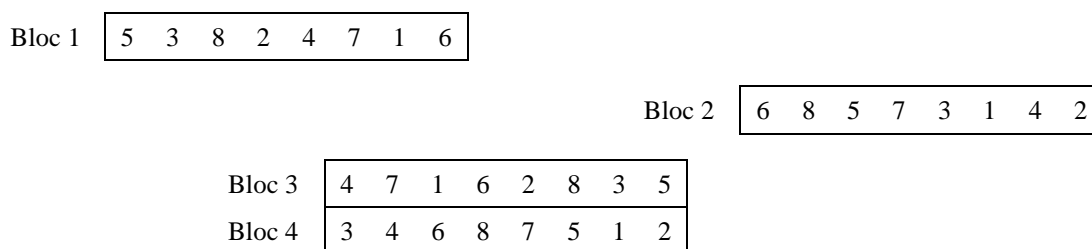
Le choix du dispositif, et le cas échéant la taille et l'orientation des blocs utilisés, dépendent de l'hétérogénéité perçue dans le site d'essai (par ex. sol, pente, exposition, infestation par un organisme nuisible, cultivar etc.). Ces paramètres ne sont jamais totalement uniformes, et un dispositif en blocs randomisés dans une zone modérément hétérogène fournit généralement davantage d'informations utiles sur la performance d'un produit qu'un dispositif totalement randomisé dans un site qu'on pense homogène mais qui par la suite se révèle hétérogène. La disposition des blocs dépend également de la dimension et de la forme des parcelles (voir ci-dessous).

Des essais en blocs randomisés conduits dans plusieurs régions à conditions agroclimatiques distinctes et/ou au cours de différentes années ou périodes de végétation peuvent parfois être considérés comme faisant partie d'une série d'essais. Dans l'analyse de variance, il est alors nécessaire de distinguer la variance entre sites de la variance entre blocs, et également d'évaluer l'interaction site-traitement qui pourra être particulièrement intéressante. Pour chaque nouvel essai, les traitements doivent de nouveau être attribués aléatoirement dans chaque bloc.

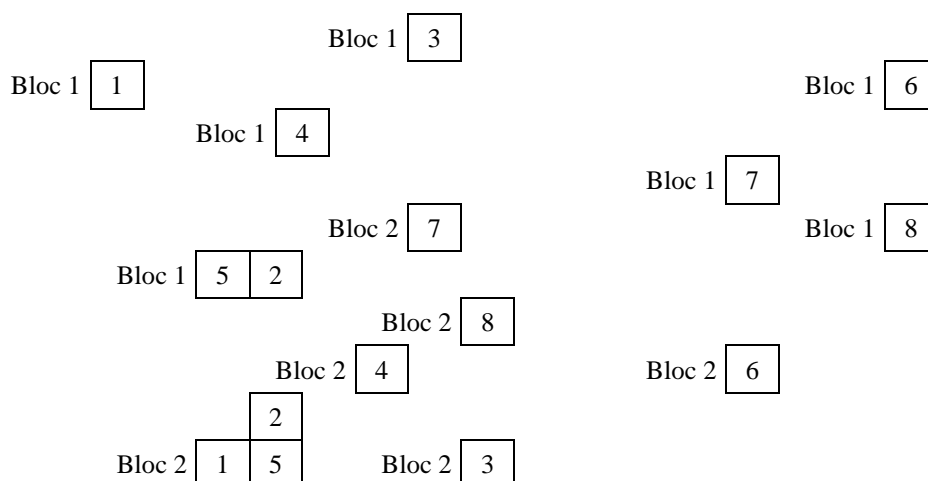
### 1.1.3 Split plot

Lorsqu'un essai multifactoriel est indispensable, on utilise généralement un dispositif en blocs randomisés dans lequel chaque combinaison des traitements apparaît une fois dans chaque bloc. Cependant, il arrive qu'on ne puisse pas randomiser totalement un des facteurs dans toutes les parcelles d'un bloc. Par exemple, supposons un essai à deux facteurs, le produit (avec quatre niveaux, de 1 à 4) et le matériel de culture (avec trois niveaux, A, B, C), et supposons que les parcelles sont relativement petites. La taille du matériel utilisé pour l'opération culturale en question peut ne pas permettre une randomisation totale sur les douze parcelles de chaque bloc. Un dispositif split-plot est alors recommandé: les sous-parcelles sont associées, dans chaque bloc, par groupes de quatre pour former

**Figure 5.** Disposition possible des blocs et des parcelles pour des blocs randomisés dans un essai de plein champ. Blocs dispersés dans l'ensemble du champ, selon une hétérogénéité observée précédemment.



**Figure 6.** Disposition possible des blocs et des parcelles pour des blocs randomisés dans un essai de plein champ. Blocs disposés dans le champ en fonction d'une hétérogénéité complexe observée précédemment.



trois parcelles entières par bloc. Le facteur "opération culturale" est randomisé dans ces trois parcelles, et le facteur "produit" est randomisé, séparément, dans les sous-parcelles de chaque parcelle entières (voir l'exemple à la figure 7). Ce dispositif nécessite une analyse de variance légèrement plus complexe, à deux strates, chacune ayant un carré des écarts moyen que l'on peut utiliser pour tester l'effet des différents facteurs et leur interaction.

#### 1.1.4 Disposition systématique

Les dispositifs systématiques non randomisés ne conviennent jamais pour les essais d'évaluation biologique, sauf dans des cas très particuliers (par exemple essais variétaux sur la sélectivité des herbicides). En règle générale, ce type de dispositif n'est utilisable qu'à des fins de démonstration.

### 1.2 Puissance des tests

Lors de la mise en place d'essais, il est important de considérer la puissance requise pour tous les tests statistiques qui seront effectués. La puissance est la probabilité de détecter une différence entre les traitements si cette différence existe. La puissance dépend d'un certain nombre de paramètres, dont:

- la précision des résultats (variation résiduelle)
- le nombre de répétitions

Il faut choisir un dispositif qui donne une bonne chance de détecter, avec une signification statistique, une différence importante dans la pratique pour le contraste étudié. On peut aussi souhaiter que les intervalles de confiance des estimations des traitements ne dépassent pas un écart préalablement défini. Avant d'entreprendre un essai, il faut choisir entre la performance d'un seul essai ou d'une série d'essais. Pour cela, il faut répondre à la question suivante: "le facteur à étudier interagit-il avec l'environnement ?".

- si la réponse est OUI, préférer une série d'essais conduite sur plusieurs sites. Le nombre de sites est prioritaire sur le nombre de répétitions à un même site. L'interprétation des résultats doit reposer sur une analyse des résultats groupés de l'ensemble de la série d'essais. L'analyse des résultats d'un seul site ne doit pas être présentée, sauf si elle est nécessaire à des fins particulières. Trois répétitions par site suffisent généralement.
- si la réponse est NON, un seul essai peut suffire et il doit comporter au moins quatre répétitions.

En général, des résultats d'essais précédents peuvent indiquer la variabilité probable des observations. Si ces

données existent, il est possible de déterminer le dispositif et la dimension de l'essai nécessaires pour la puissance requise. Il est parfois possible de le faire à partir de considérations théoriques. Par exemple, pour des données suivant une loi binomiale, une limite supérieure peut être déterminée pour la variabilité des proportions. Différents systèmes informatisés ou graphiques sont disponibles pour aider à déterminer le nombre de répétitions nécessaires. Ils utilisent l'amplitude de la différence devant être estimée, ou le niveau de signification exigé pour cette différence, et la précision attendue. La section suivante donne des règles générales simples.

### 1.3 Nombre de degrés de liberté en fonction du nombre de traitements et de répétitions

Pour permettre une analyse statistique valide, le nombre de degrés de liberté (dl) doit être suffisamment élevé. Dans un essai avec 8 traitements et 4 répétitions dans un dispositif en blocs randomisés, il y a 21 dl résiduels. Ce résultat est obtenu comme suit: dl totaux ( $32 - 1 = 31$ ) moins dl traitements ( $8 - 1 = 7$ ) moins dl blocs ( $4 - 1 = 3$ ), c'est à dire  $31 - 7 - 3 = 21$ . Pour 4 traitements et 4 répétitions, le nombre de dl résiduels est de  $15 - 3 - 3 = 9$ , ce qui est généralement considéré comme trop faible pour être utile. Le nombre de dl résiduels peut être augmenté en augmentant le nombre des répétitions. Cela implique souvent l'utilisation de blocs supplémentaires dans un dispositif en blocs randomisés. Le nombre de dl résiduels souhaitable dépend du degré de précision recherché (puissance); en cas de doute, obtenir l'avis d'un spécialiste. En général, l'expérience des essais d'évaluation biologique montre qu'il est déconseillé de mettre en place un dispositif avec moins de 12 dl résiduels; davantage sont nécessaires lorsque la variable mesurée est très variable (par ex. la densité d'un insecte).

Le choix du dispositif expérimental influence également le nombre de dl résiduels. C'est le dispositif totalement randomisé qui donne le maximum de dl. Le dispositif en blocs randomisés utilise certains de ses dl pour tenir compte de l'hétérogénéité du milieu (par ex. selon un gradient). Le split-plot utilise des dl pour tenir compte de plusieurs composants de variation éventuels. L'expérimentateur doit essayer de conserver le nombre maximal de dl pour estimer la variation résiduelle, tout en choisissant un dispositif optimal permettant de minimiser cette variation en tenant compte de toutes les sources connues d'hétérogénéité.

**Figure 7.** Exemple de split-plot. Les deux facteurs du traitement sont: le produit (1,2,3,4, randomisés dans les sous-parcelles des parcelles entières) et la méthode de culture (A,B,C, randomisés dans les parcelles entières de chacun des deux blocs).

Bloc 1					Bloc 2			
1A	2A	3A	4A	parcelle entière 1	2B	4B	3B	1B
3C	4C	1C	2C	parcelle entière 2	2A	3A	4A	1A
2B	3B	1B	4B	parcelle entière 3	1C	3C	2C	4C

### 1.4 Parcelles expérimentales: taille, forme, nécessité de bordures

En règle générale, les parcelles seront rectangulaires et de la même dimension dans un même essai ainsi que pour une série d'essais donnée. La précision augmente avec la dimension de la parcelle, mais seulement dans une certaine limite, car la variabilité des conditions du sol et d'infestation tend également à augmenter. Des parcelles rectangulaires longues et étroites conviennent bien pour la récolte mécanisée. Des parcelles presque carrées diminuent le risque d'interférence entre les parcelles.

Dans les directives de l'OEPP, les dimensions des parcelles sont préconisées pour les combinaisons particulières d'organisme nuisible/plante-hôte. Dans le cas où les interférences entre parcelles sont probables, les parcelles auront une dimension plus grande (parcelle brute) et les observations se feront seulement sur une partie centrale plus restreinte (parcelle nette). En règle générale, les directives de l'OEPP suggèrent des dimensions de parcelle nette, la dimension de parcelle brute étant alors laissée à l'appréciation de l'expérimentateur qui doit déterminer quelles sont les zones de protection nécessaires en considérant toutes les sources potentielles d'interférences entre les parcelles pour chaque essai ou série d'essais. Une source courante d'interférences est la dispersion du produit phytosanitaire (par ex. dérive des produits pulvérisés ou des vapeurs, mouvements latéraux sur/dans le sol) en dehors de la parcelle, ce qui contamine les parcelles adjacentes. Ce phénomène est particulièrement important pour les pulvérisations appliquées sur de grandes plantes. Noter par contre que

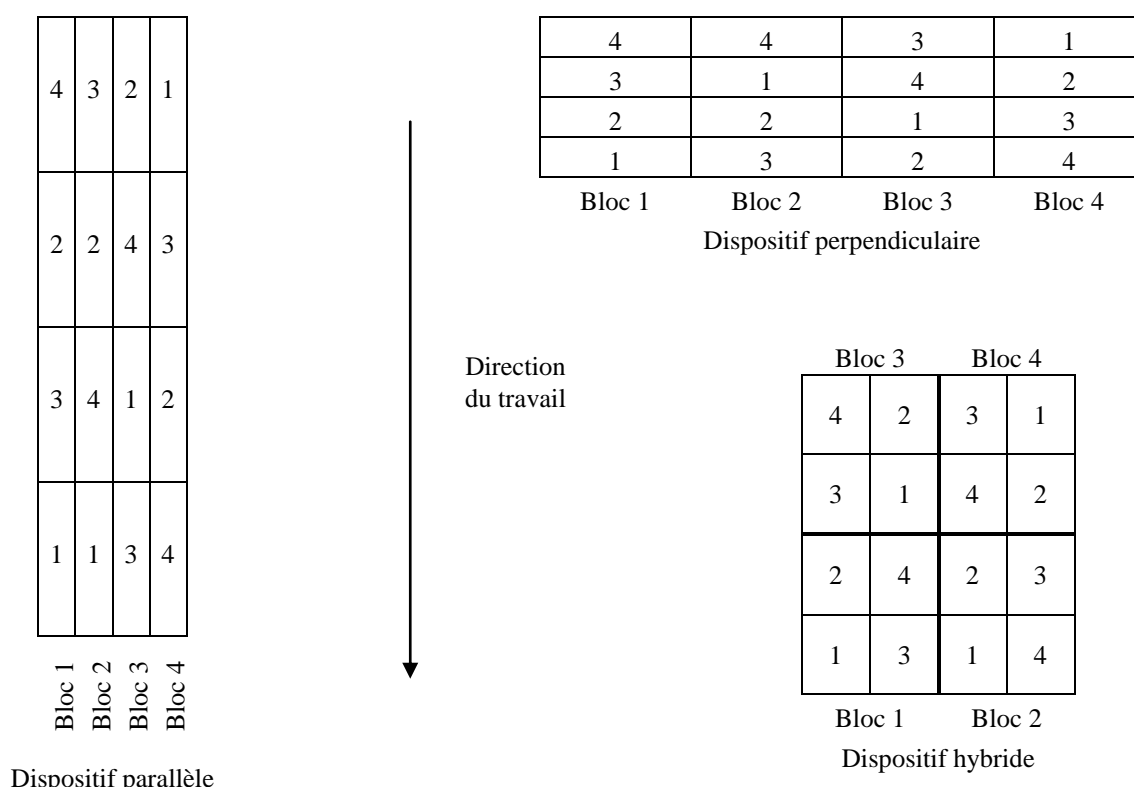
l'erreur expérimentale augmente avec la taille des zones de protection.

Une autre source courante d'interférences est la dispersion des organismes nuisibles (par ex. champignons portés par le vent ou insectes très mobiles) à partir des parcelles non traitées ou des parcelles où la lutte est peu efficace. Cette dispersion peut à la fois augmenter la population de l'organisme nuisible dans les parcelles des traitements les plus efficaces et la diminuer dans les parcelles des moins efficaces. De même, si un produit phytosanitaire est étudié sur une culture dans laquelle la lutte intégrée est pratiquée, les effets négatifs sur les prédateurs et les parasites peuvent être masqués par leur migration entre les parcelles.

Selon le matériel utilisé pour la récolte, il peut être nécessaire d'augmenter la dimension de la parcelle nette au-delà de la dimension nécessaire pour effectuer les observations.

Les parcelles peuvent être disposées perpendiculairement ou parallèlement à la direction de travail (semis ou plantation). Le dispositif perpendiculaire (voir figure 8) présente l'avantage d'exposer uniformément toutes les parcelles d'un bloc à une erreur éventuelle pendant le travail (façons culturales, semis, etc.). Par contre, le traitement et la récolte sont plus difficiles. Le dispositif parallèle présente des avantages pratiques pour l'application des traitements et la récolte, mais il présente un plus grand risque d'hétérogénéité le long de blocs très longs. Le dispositif hybride peut éventuellement faire office de compromis.

**Figure 8.** Dispositifs en blocs randomisés similaires, mais avec une disposition des parcelles différente par rapport à la direction du travail.



## 1.5 Rôle et implantation des témoins non traités

### 1.5.1 Utilité du témoin non traité

La principale caractéristique des "témoins non traités" est qu'ils ne reçoivent aucun des traitements phytosanitaires en cours d'étude. Les témoins reçoivent toutes les interventions qui sont appliquées de façon uniforme à l'ensemble de l'essai, en particulier les façons culturales et les traitements visant les organismes nuisibles non étudiés. Noter que le témoin non traité ne reçoit normalement aucun traitement contre l'organisme à étudier. Cependant, dans certaines circonstances, il peut être utile d'inclure dans le témoin non traité certaines opérations appliquées aux autres traitements. Par exemple, lorsque les autres traitements reçoivent les produits dans une solution aqueuse distribuée par un pulvérisateur dans la parcelle, le témoin non traité peut inclure le passage d'un pulvérisateur contenant uniquement de l'eau. Il s'agit d'imiter autant que possible les opérations réalisées dans les autres traitements, à l'exception de l'application du produit.

Le témoin est indispensable, sous une forme ou une autre. Son objectif principal est de confirmer la présence d'un niveau d'infestation suffisant; en l'absence d'une infestation adéquate, l'efficacité ne peut pas être démontrée et les résultats n'ont pas de sens. Cette confirmation peut être qualitative (présence d'espèces dominantes, type de flore, adventices...) ou quantitative (respect de seuils maximaux ou minimaux, répartition dans l'espace).

Le témoin non traité ne joue normalement pas le rôle de point de référence pour l'essai; c'est le rôle du produit de référence. En particulier, le témoin ne fait généralement pas l'objet d'une comparaison statistique avec les traitements dans le cas où on utilise un test de signification pour évaluer l'efficacité. En général, on sait avant le début de l'essai qu'il y aura une différence entre les traitements et le témoin non traité, et tester la signification de cette comparaison revient à tester une hypothèse nulle dont on sait déjà qu'elle est fautive; le test est donc superflu et n'est pas valide. C'est en fait l'efficacité du produit de référence qui est utilisée comme référence dans les tests statistiques.

Cependant, selon le sujet et le type d'essai, les témoins peuvent jouer un rôle utile, et parfois plusieurs rôles à la fois, dont les suivants.

- une aide pour les observations. Une estimation visuelle des dégâts ou de l'infestation peut parfois être effectuée en termes relatifs par rapport à un témoin.
- l'utilisation de la technique du "témoin adjacent" pour mesurer et tenir compte de la répartition des parcelles dans l'espace.
- l'observation du développement de l'organisme nuisible (émergence, envol, libération de spores, etc.), en particulier pour déterminer les dates d'application ou d'observation.
- une réserve d'inoculum pour assurer que le niveau d'inoculum ne tombe pas trop bas ou ne devient pas

trop hétérogène (dans les cas extrêmes, ceci peut équivaloir à une inoculation artificielle).

- une assistance pour interpréter les résultats des essais. Par exemple, une différence significative entre deux traitements peut ne pas avoir la même importance selon le niveau d'infestation.
- l'amélioration de la compréhension des résultats de l'analyse en les exprimant sous une autre forme, ou en permettant leur représentation graphique (par ex. transformation de la mortalité en taux d'efficacité).
- la possibilité d'observations supplémentaires, en particulier le rendement quantitatif et qualitatif, qu'il peut être intéressant de relier aux autres résultats de l'essai.
- enfin, et exceptionnellement, formation d'un terme de comparaison pour les traitements à l'étude si aucun produit de référence n'est disponible. Ceci peut être le cas, par exemple, lorsque le type de produit ou son usage sont nouveaux ou lorsque l'utilisation de tous les produits de référence potentiels a été interdite. Exceptionnellement, ce rôle est alors similaire au rôle du produit de référence, bien que son interprétation soit très différente. Les témoins peuvent alors être comparés aux différents traitements à l'aide de tests statistiques, de la même manière que pour le produit de référence dans les essais habituels.

### 1.5.2 Types d'implantation des témoins

Il est possible d'envisager quatre types d'implantation des témoins.

#### *Témoins inclus*

Les témoins sont considérés comme les autres traitements. Les parcelles témoins sont de la même forme et de la même dimension que les autres parcelles, et elles sont randomisées au hasard dans l'essai.

#### *Témoins imbriqués*

Les parcelles témoins sont disposées systématiquement dans l'essai. Leur dimension et leur forme ne sont pas forcément les mêmes que pour les autres parcelles de l'essai. Les observations réalisées dans ces parcelles sont d'une autre nature et ne doivent pas être prises en compte dans l'analyse statistique. Le but du dispositif est de garantir une distribution plus homogène de l'effet d'une zone adjacente non traitée que ce qu'il est possible d'obtenir avec un témoin inclus. Différents dispositifs peuvent être utilisés; les parcelles peuvent être placées entre les blocs ou entre les parcelles traitées à l'intérieur des blocs (voir figure 9).

#### *Témoins exclus*

Les parcelles témoins sont placées à l'extérieur de la zone de l'essai (et n'en sont pas adjacentes), dans une zone où les conditions sont très proches de celles de la zone d'essai. Les répétitions ne sont pas indispensables mais peuvent être utiles si la zone n'est pas homogène.

Les observations réalisées dans ces parcelles ne doivent pas être prises en compte dans l'analyse statistique.

### Témoins adjacents

Chaque parcelle est divisée en deux sous-parcelles égales, dont une (au hasard) n'est pas traitée. Les observations sont faites de la même manière dans les deux sous-parcelles. Les observations réalisées dans ces parcelles ne doivent pas être prises en compte dans l'analyse statistique sauf si on tient compte du fait que le dispositif est une forme de split-plot. Dans un dispositif split-plot, la variabilité à l'intérieur des parcelles peut être différente de la variabilité entre les parcelles; par conséquent, l'analyse de variance doit comporter deux strates d'erreur. L'avis d'un spécialiste peut être nécessaire pour interpréter les résultats.

### 1.5.3 Choix du type d'implantation du témoin non traité

Le choix du type de témoin dépend de son rôle (ou de ses rôles) dans l'essai. Le dispositif avec témoins inclus a très souvent été utilisé dans les essais d'efficacité, et il est encore fréquemment utilisé en pratique, mais ce n'est pas forcément celui qui convient le mieux. Le système de décision suivant donne des indications.

- (1) Si le témoin est utilisé dans un test statistique, il est indispensable d'utiliser un dispositif avec témoin inclus. Cependant, cela n'est qu'exceptionnellement nécessaire (voir plus haut). Un autre type de témoin

dit normalement être utilisé. On doit alors tenir compte de l'hétérogénéité des parcelles.

- (2) Si l'hétérogénéité est forte, un dispositif avec témoin adjacent convient. Si l'hétérogénéité est faible ou modérée, on doit alors tenir compte des interférences éventuelles entre les parcelles témoins et les parcelles adjacentes.
- (3) Si les parcelles témoins ne sont pas susceptibles d'interférer avec les parcelles adjacentes, un dispositif avec témoin imbriqué convient. Si les parcelles témoins sont susceptibles d'interférer avec les parcelles adjacentes, un dispositif avec témoins exclus doit être utilisé.

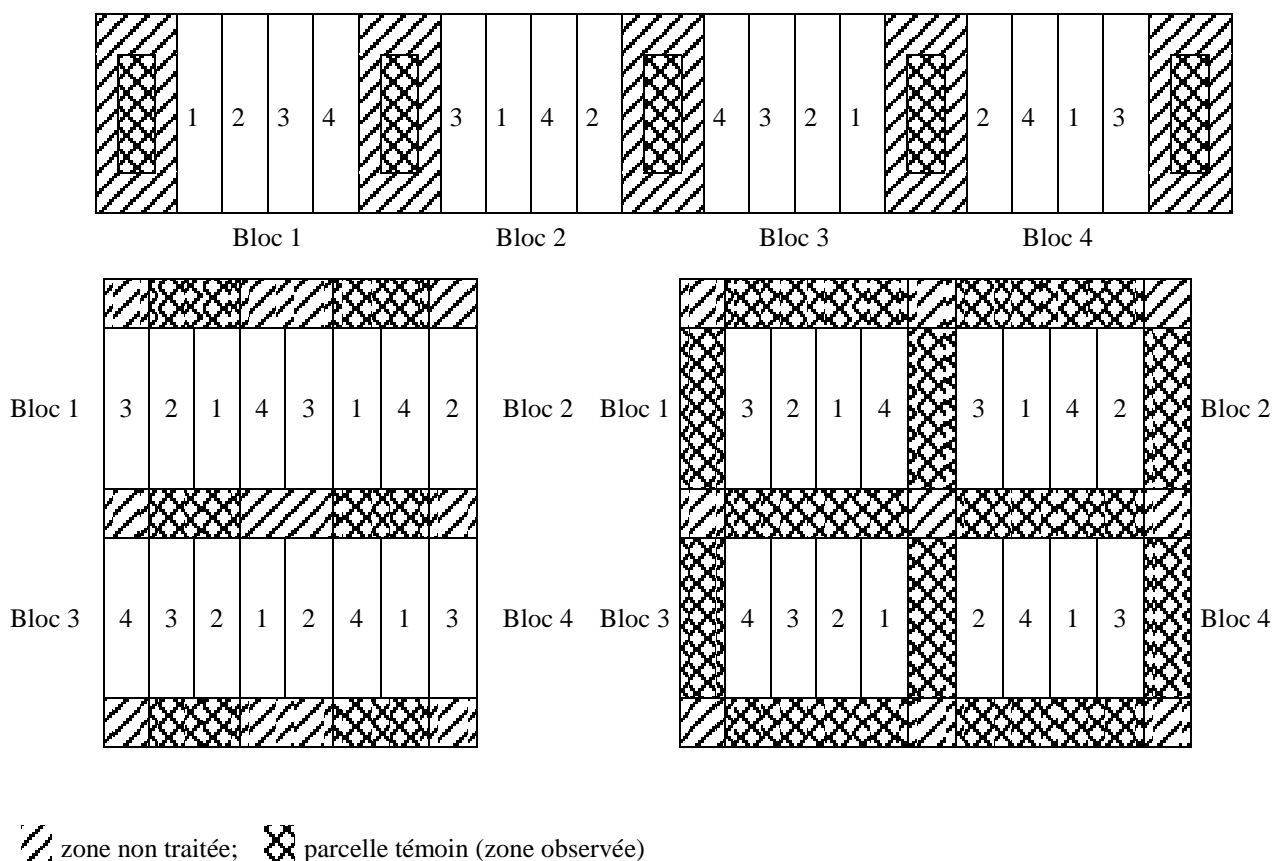
## 2. Principes de l'évaluation des effets des produits phytosanitaires

Lors de la notation des effets des produits phytosanitaires au cours d'un essai d'évaluation biologique, des variables sont notées par des modes d'observation.

### 2.1 Variables

La nature de la variable est importante car elle influence en général la méthode statistique choisie pour interpréter les résultats. Plusieurs grandes catégories de variables peuvent être distinguées:

Figure 9. Exemple de l'utilisation de parcelles témoins imbriquées pour un essai en blocs randomisés avec 4 blocs et 4 traitements.



- (1) variables binaires (par ex. oui/non, présence/absence). Elles peuvent donner des variables binomiales, qui représentent le nombre de fois qu'un état a été observé pour un nombre donné d'observations (par ex. le nombre de plantes infestées sur 20 choisies par sondage dans une parcelle).
- (2) variables nominales. Variables ayant des valeurs d'importance équivalente ne pouvant être classées dans un ordre particulier (non ordinales), par ex. espèces d'organismes nuisibles, couleurs. Les valeurs de ces variables sont généralement indiquées par des mots.
- (3) variables ordinales. Variables correspondant à des classes que l'on peut mettre dans un certain ordre mais qui ne sont pas mesurées. Elles sont généralement qualitatives, et les classes ne peuvent alors être ordonnées que les unes par rapport aux autres (par exemple: mauvais, modéré, bon; grades descriptifs des dégâts sur les feuilles). Dans d'autres cas, les valeurs peuvent être exprimées numériquement (elles pourraient être mesurées précisément, mais ne le sont pas pour des raisons pratiques), par ex. catégories de couverture par les adventices, ou catégories d'infestation par les pucerons.
- (4) variables quantitatives. Variables pouvant être mesurées, et qui le sont en pratique, par ex. rendement, hauteur des végétaux, nombre de larves. Elles peuvent être discrètes si elles sont exprimées seulement par des entiers (par ex. résultats des comptages) ou continues (par ex. poids, tailles). Les variables quantitatives peuvent également résulter d'opérations mathématiques. Des différences ou des sommes peuvent être obtenues (par ex. la différence entre une valeur avant et après traitement). Des valeurs relatives peuvent être calculées, sous forme de proportions ou de ratios. Une proportion représente le quotient "quantité partielle/quantité totale" et est comprise entre 0 et 1 (il s'agit d'une variable quantitative continue bornée). En pratique, une proportion est souvent une fréquence relative: "fréquence dans une classe/fréquence totale", c'est à dire une variable binomiale ou multinomiale. Elle est souvent exprimée sous forme de pourcentage (c'est à dire une valeur comprise entre 0 et 100). Les ratios n'ont pas de limite supérieure, par ex. (valeur initiale – valeur finale)/valeur initiale. Ils peuvent également être exprimés en % (auquel cas il est possible d'obtenir une valeur supérieure à 100). Les proportions doivent toujours être préférées aux ratios en tant que variables pour les analyses statistiques, et elles doivent toujours être utilisées sauf raison pertinente. Noter qu'une valeur exprimée en pourcentage peut être une proportion, ou un ratio, ou même une variable binomiale, et il est important de distinguer ces cas pour les analyses statistiques.

## 2.2 Modes d'observation des variables

Les variables doivent être mesurées aussi précisément que possible. Afin de noter les variables, il est possible d'avoir recours à quatre modes d'observation.

### 2.2.1 Mesure

Une mesure détermine la valeur de manière objective. Le résultat d'une mesure peut être une variable continue (poids, taille) ou discrète (comptage). Pour les essais au champ, lorsque la mesure ne concerne pas l'ensemble de la parcelle, elle doit porter sur un échantillon dont la taille et le mode d'échantillonnage ont été déterminés au préalable en fonction de la précision requise.

### 2.2.2 Estimation visuelle

L'estimation visuelle détermine les valeurs de manière subjective, mais avec les mêmes échelles et la même gamme de valeurs que les mesures. Elle concerne généralement des variables continues (par ex. surface couverte par les adventices) mais peut également concerner les comptages dans le cas de grands nombres (par ex. nombre de lésions foliaires). En général, une estimation est plus facile à faire par rapport à une référence (ou un témoin non traité) qu'en valeur absolue. Il faut souligner qu'on obtient le même type de variable qu'avec des mesures. Les résultats sont donc exprimés dans les mêmes unités.

L'observateur doit avoir bénéficié d'un bon apprentissage pour faire les estimations et il doit avoir recours à un étalonnage préalable. Si ces conditions sont remplies, l'estimation peut être très exacte. La précision peut toutefois dépendre de la valeur de la variable observée. Par exemple, si on estime le pourcentage de surface foliaire atteinte, les valeurs extrêmes seront mieux déterminées que les valeurs intermédiaires. Ceci a été pris en compte lors de la conception des diverses aides à l'estimation et échelles (voir ci-dessous). Une transformation statistique des données estimées peut également être nécessaire.

### 2.2.3 Classement

Le classement situe chaque individu par rapport aux autres (lui donne un rang). Le résultat d'un classement est une variable ordinale qualitative. Si le nombre d'échantillons à comparer est peu élevé, leur classement pour une variable donnée peut être effectué au champ. Cependant, l'analyse par des méthodes non paramétriques de variables classées n'a pas la puissance des méthodes paramétriques (telle que l'analyse de variance) et ne permet pas d'obtenir d'estimations précises pouvant être facilement comparées pour différents essais. Le classement doit donc être évité dans la mesure du possible pour les essais au champ.



### 2.2.4 Notation

Effectuer une notation consiste à affecter un objet à une classe définie sans ambiguïté. L'ensemble de ces classes est généralement appelé "échelle", surtout lorsque la variable sous-jacente est ordinale, ce qui est souvent le cas. La notation est également utilisée pour les variables binaires et nominales. Elle est par définition subjective. Elle s'applique à de nombreux types de variables: ordonnées ou non, continues ou discrètes. Elle peut être utilisée pour les variables qualitatives ainsi que pour les variables quantitatives trop onéreuses à mesurer. Son principal intérêt réside dans le fait qu'elle est rapide et non destructrice. Elle peut également être utilisée pour caractériser une parcelle entière par une seule notation. Le nombre de paliers sur une échelle détermine sa sensibilité. Cette dernière ne doit être ni trop faible (sinon les résultats n'apportent qu'une information limitée), ni trop élevée (l'échelle devient alors peu pratique à utiliser).

Les échelles sont adaptées à des usages spécifiques et en général il n'existe pas d'échelle unique et universelle pour un type de variable donné. Les directives OEPP donnent de nombreux exemples d'échelles dont l'usage est recommandé (voir annexe I) et qui permettent la notation de diverses combinaisons organisme nuisible/culture. Des règles simples ont généralement été suivies lors de l'élaboration de ces échelles, surtout pour définir leurs extrêmes. Le point le plus bas sur une échelle ordinale (pas d'effet) doit recevoir la note 1 (et non 0 qui est réservé aux observations manquantes dans de nombreux systèmes de notation) et le point le plus élevé doit correspondre à la valeur la plus forte de l'effet, avec des valeurs intermédiaires par ordre croissant.

Dans la pratique, prendre les précautions suivantes:

- (a) établir l'échelle à l'avance;
- (b) faire la notation sans connaître le plan de traitement;
- (c) travailler bloc par bloc;
- (d) examiner toutes les parcelles dans la même direction pour éviter des différences d'éclairage.

Le tableau 1 résume les différentes modes d'observation et les différents types de variables obtenues.

### 2.3 Utilisation des échelles pour l'estimation visuelle et la notation

L'estimation visuelle et la notation sont souvent confondues. Ces opérations sont superficiellement semblables, mais leurs résultats diffèrent: l'estimation visuelle produit une série de valeurs estimées pour une variable quantitative discrète ou continue. La notation donne les nombres par classes. Les classes d'une échelle de notation sont souvent représentées par des nombres séquentiels (par ex. 1 à 9), mais cela ne veut pas dire que les intervalles entre les valeurs de l'échelle sont identiques. Si les intervalles varient, il n'est pas conseillé d'entreprendre une analyse statistique sans l'avis d'un spécialiste, ou d'estimer les paramètres statistiques sans réflexion préalable. Les résultats calculés doivent être interprétés avec beaucoup de précaution. Les valeurs des échelles pourraient aussi bien être représentées par les lettres de l'alphabet, ce qui montre bien leur caractère de variable ordinale et les dangers d'une approche trop simpliste.

Par contre, les échelles peuvent aussi être utilisées comme aides à l'estimation visuelle ("variable ordinale avec intervalles"). Si les valeurs des échelles correspondent aux valeurs réelles d'une variable quantitative (par ex. sur une échelle visuelle du % de surface foliaire atteinte), l'observateur attribue des valeurs de l'échelle ou interpole des valeurs intermédiaires selon son jugement. Les valeurs obtenues, transformées si nécessaires, peuvent alors être considérées comme des estimations d'une variable continue et être analysées à l'aide des procédures statistiques habituelles. Il est important de souligner que si l'observateur dispose des ressources nécessaires (temps, main-d'œuvre, expérience) pour faire des estimations plus précises, ou même des mesures, les données pourront être analysés avec une meilleure précision et une plus grande puissance. En revanche, rien ne sert d'effectuer une estimation relativement précise (par ex. % de surface foliaire atteinte) pour ensuite lui substituer la valeur correspondante (bien moins précise) d'une échelle. Les avantages de la notation par échelle (vitesse et simplicité) n'existent que si l'observateur note directement dans une classe d'échelle (il doit y être entraîné) sans essayer d'estimer de manière plus précise.

**Tableau 1.** Différents modes d'observation et types de variables

	MESURE	ESTIMATION VISUELLE	CLASSEMENT	NOTATION
Binaire				X
Nominale				X
Ordinale			X	X
Discrète	X	X		
Continue limitée	X	X		
Continue non limitée	X	X		

## 2.4 Qualité d'un mode d'observation

Les modes d'observation peuvent être distingués par un certain nombre de qualités:

- "précision" - absence de biais, dans le contexte statistique;
- "fiabilité" - variabilité (ou variance) faible;
- "exactitude" - combinaison de la précision et de la fiabilité.
- "sensibilité" - réaction du mode d'observation à une modification faible de la valeur de l'unité expérimentale;
- "répétitivité" - un même observateur attribue la même valeur (ou des valeurs très proches) à des unités expérimentales identiques;
- "reproductibilité" - des observateurs différents attribuent la même valeur (ou des valeurs très proches) à la même unité expérimentale.

Ces qualités importantes détermineront le choix des modes d'observation selon les objectifs, en particulier dans les séries d'essais.

## 3. Analyse statistique des résultats de l'essai

### 3.1 Principes

Les sections suivantes visent à esquisser la bonne pratique statistique dans l'analyse de données. Il ne s'agit pas, et il ne peut pas s'agir, d'une prescription valable pour toutes les analyses et toutes les situations. Les expérimentateurs ne doivent jamais sous-estimer l'utilité de l'avis d'un statisticien. L'opinion d'une personne très expérimentée dans l'analyse des données obtenues au champ doit être préférée à celle d'un théoricien. Il est très important que les expérimentateurs comprennent l'avis qu'on leur donne et il est souvent préférable qu'ils effectuent une analyse simple qu'ils peuvent exposer et défendre avec confiance plutôt que d'accepter des conseils les conduisant à une analyse qu'ils ne comprennent que partiellement. La bibliographie à la fin de cette directive peut être utile. Elle mentionne plusieurs textes qui présentent les principes de la bonne pratique statistique plutôt que de donner des recettes statistiques à appliquer aveuglément.

### 3.2 Structure de base et séquence de l'analyse

Les Directives OEPP pour l'évaluation biologique des produits phytosanitaires mentionnent que "les résultats seront en principe soumis à une analyse statistique, par des méthodes qui doivent être précisées". La procédure à suivre peut être illustrée par un essai caractéristique dans lequel plusieurs produits à étudier sont appliqués à une dose unique et comparés avec un produit de référence en présence d'un témoin non traité. L'efficacité du produit est évaluée par une variable quantitative mesurée. L'objectif de l'essai est de

comparer les produits à étudier au produit de référence, et en particulier d'identifier les plus efficaces. La séquence de l'analyse, pour un essai conduit conformément à la directive OEPP pertinente, est la suivante:

1. l'essai est-il réaliste, c'est à dire capable de fournir des données utiles ? Cela est le cas seulement si l'infestation par l'organisme nuisible est suffisamment forte;
2. les résultats sont-ils cohérents ? Le produit de référence donne-t-il les résultats attendus par comparaison avec le témoin non traité ?
- 3 si ces deux conditions sont satisfaites, on peut comparer les produits à étudier au produit de référence, et éventuellement comparer les produits entre eux. L'analyse a pour objectif premier d'estimer l'amplitude des différences entre les traitements et de fournir une estimation de la variabilité des estimations, souvent sous la forme d'une erreur-type ou indice similaire. Un objectif secondaire est de comparer les traitements à l'aide d'un test statistique approprié;
- 4 si on ne trouve aucune différence significative entre les produits à étudier et le produit de référence, il est utile de vérifier que l'essai est assez puissant pour détecter une différence significative. Autrement dit, l'essai est-il assez puissant pour estimer les différences entre les traitements avec une erreur-type suffisamment faible ? Si ce n'est pas le cas, on ne peut pas tirer de conclusion de l'essai.

Des systèmes identiques peuvent être mis au point pour d'autres essais d'évaluation biologique, dont en particulier le cas spécial de la sélectivité des herbicides, et le cas exceptionnel où il n'existe pas de produit de référence et où les traitements doivent être comparés au témoin non traité (voir section 1.5.1).

### 3.3 Choix de la méthode d'analyse

La méthode d'analyse est globalement déterminée par le type de variable. Si la variable est quantitative (binaire, binomiale, discrète ou continue), une méthode statistique paramétrique est être utilisée, généralement basée sur une analyse de variance. Si la variable est qualitative, des méthodes non paramétriques conviennent. L'analyse de variance repose sur deux hypothèses: l'additivité des effets et l'homogénéité de la variance (normalité). On suggère parfois d'utiliser des méthodes non paramétriques si ces hypothèses ne sont pas satisfaites. Cependant, la non additivité et la non normalité peuvent très souvent être améliorées et ne justifient pas forcément l'utilisation de méthodes non paramétriques. Celles-ci n'ont pas la puissance de leurs équivalents paramétriques et doivent normalement être utilisées seulement pour certaines formes de notations ou de données ordinales.

### 3.3.1 Additivité des effets

Il est important de considérer si les effets sont susceptibles de s'additionner sur l'échelle d'analyse de la variable de réponse. Par exemple, si la variable est la densité de population d'un insecte, les effets du traitement tels qu'un insecticide ou une attaque d'un champignon seront probablement multiplicatifs, et toucheront une proportion de la population. Alternativement, si la variable est la proportion d'adventices tuées par un herbicide, alors les effets seront probablement additifs, non pas sur l'échelle naturelle mais, par ex., sur une échelle probit ou logit. Deux méthodes sont couramment utilisées pour transformer l'échelle naturelle en une échelle plus réaliste: l'utilisation de transformations et l'utilisation de modèles linéaires généralisés.

### 3.3.2 Homogénéité de la variance

Les transformations peuvent résoudre le problème de l'additivité des effets mais elles ne garantissent généralement pas l'homogénéité des variances. Celle-ci doit être vérifiée indépendamment, même si l'additivité est normalement plus importante. Pour les données discrètes sous forme de comptages, une transformation logarithmique garantit généralement à la fois l'additivité et l'égalité des variances. Pour les données binaires, les données binomiales ou les données sous forme de proportions, une transformation logit, probit ou une transformation complémentaire log-log garantira en général l'additivité, mais pas forcément l'égalité des variances.

## 3.4 Analyse de variance

### 3.4.1 Tableau des moyennes

Une analyse de variance est recommandée, le cas échéant après une transformation. Elle prendra d'abord la forme d'un tableau de la moyenne de chaque traitement, avec une estimation de la variabilité des moyennes, normalement sous forme d'une erreur-type ou d'un intervalle de confiance. Ce tableau souligne l'amplitude des effets et est recommandé pour surmonter le problème bien connu que la signification biologique n'est pas égale à la signification statistique; l'amplitude et l'importance des effets peuvent être fortes mais ne pas être significatives en raison de la puissance insuffisante de l'analyse ou du test.

Dans le tableau d'analyse de variance, il faut faire attention à attribuer les unités aux strates correctement, avec une structure des traitements et des blocs convenant au dispositif choisi. En particulier, il faut se prémunir contre le problème bien connu de la pseudo-répétition, qui survient lorsqu'on ne tient pas compte du fait que les traitements n'ont pas été totalement randomisés dans toutes les unités d'échantillonnage mais ont été appliqués à des groupes d'unités.

### 3.4.2 Tests $F$ et contrastes orthogonaux

En plus de la présentation des tableaux des moyennes et des erreurs-types, des tests statistiques, normalement des test  $F$ , peuvent être appliqués à l'ensemble des

données. Un test d'ensemble sur tous les traitements ne constitue pas une preuve de l'efficacité, sauf dans les cas les plus simples, car il est en général influencé par les données des témoins non traités. Au contraire, on recommande fortement que la somme des carrés des traitements soit divisée entre les composants d'intérêt biologique en définissant des contrastes orthogonaux (indépendants). Par exemple, dans le premier exemple où les 8 traitements ont été comparés, il y avait 5 produits à étudier, 2 produits de référence et un témoin non traité. Les huit traitements donnent 7 dl dans les sommes des carrés des traitements. Des contrastes pertinents sont: le témoin non traité par rapport à la moyenne des sept autres traitements (1 dl); le premier produit de référence par rapport au second produit de référence (1 dl); la moyenne des produits de référence par rapport à la moyenne des produits à étudier (1 dl); les différences entre les moyennes des produits à étudier (4 dl). Parmi ces contrastes, les deux premiers ont pour but d'éliminer les variations relativement peu intéressantes du point de vue biologique, et les deux dernières aident à mettre en évidence les objectifs réels de l'essai. Chaque contraste fournit une valeur distincte de  $F$ , qui peut être utilisée pour tester formellement des hypothèses intéressantes, qui, dans cet exemple, pourraient être qu'en moyenne, les produits à étudier ne sont pas meilleurs que les produits de référence, et qu'il n'y a pas de différence entre les produits à étudier. L'interprétation de la première de ces hypothèses peut être influencée par la différence entre les produits de référence, surtout si elle est importante. Si des contrastes non orthogonaux doivent être testés, par exemple les cinq contrastes séparés (1 dl chacun) entre la moyenne de chaque produit à étudier et un produit de référence spécifique, alors un test  $F$  (ou un test  $t$  le cas échéant) doit également être appliqué, en utilisant le carré des écarts résiduel moyen de l'analyse de variance.

Les contrastes et les hypothèses doivent de préférence être précisés par avance, au stade de la conception du dispositif, et utilisés avec modération. Les tests ne doivent pas être effectués simplement parce qu'une analyse préliminaire *post-hoc* a mis en évidence des différences qui semblent importantes et pourraient être significatives en cas de test. La régularité est souvent un meilleur indicateur de la présence d'un effet réel que des tests de signification isolés, en particulier si la puissance du test est faible. Par exemple, si un produit à étudier se révèle plus efficace qu'un produit de référence à chacun de onze sites distincts, mais pas de manière significative dans aucun, le bon sens permet de dire que les résultats sont réguliers (et d'ailleurs, un test binomial démontre que la probabilité d'obtenir un résultat aussi régulier s'il y avait en fait aucune différence réelle entre les traitements est inférieure à 0,001).

### 3.4.3 Note sur les procédures de comparaison multiple

L'utilisation des procédures de comparaison multiple a été proposée dans la première édition de cette Norme (OEPP/EPP, 1990). Cependant, une opinion moderne (Perry, 1986) est que ces procédures ne conviennent

pas dans de nombreux cas où elles étaient auparavant utilisées pour analyser les essais d'évaluation biologique. Elles ne conviennent pas en particulier pour les données structurées de façon factorielle (par ex. dans la comparaison de différentes doses pour différents traitements). Les procédures de comparaison multiple font actuellement l'objet de discussions pour déterminer dans quels cas elles sont valides et aucun consensus international n'a encore été trouvé. Pour cette raison, l'utilisation de procédures de comparaison multiple pour interpréter les résultats d'essais échangés au niveau international peut poser problème, et il serait préférable d'utiliser d'autres procédures statistiques <sup>1</sup>.

### 3.5 Modèles linéaires généralisés

Les modèles linéaires généralisés sont une forme de régression qui généralise l'analyse de variance. Ils sont supérieurs aux transformations dans la mesure où ils traitent séparément et simultanément le problème de l'additivité des effets et de l'égalité des variances (non normalité). Ils permettent de spécifier directement la distribution de la variable. Par exemple, pour les comptages d'insectes, un modèle de ce type peut spécifier une fonction logarithmique (pour tenir compte des effets multiplicatifs) et une distribution de Poisson pour les comptages (pour tenir compte directement du problème d'égalité de variances et de non normalité). Alternativement, une variable binomiale peut être analysée en utilisant une fonction logit (pour permettre l'additivité) et en précisant une distribution binomiale (correspondant aux données, qui peut se présenter sous la forme de  $r$  plantes malades sur  $n$  traitées). Il existe de nombreux parallèles entre l'analyse de la déviation qui résulte de l'utilisation d'un modèle linéaire généralisé, et l'analyse traditionnelle de la variance. En particulier, les concepts de sommes des carrés, de degrés de liberté, de contrastes orthogonaux, de test du  $\chi^2$  et de test  $F$ , et des moyennes estimées avec des erreurs-types ont tous des parallèles spécifiques dans les modèles linéaires généralisés et ceux-ci peuvent être fournis pour étude.

### 3.6 Modèles d'effets aléatoires

Cette Norme a considéré les traitements comme des effets fixes. Certains expérimentateurs, dans certains essais (en particulier les essais d'uniformité) peuvent souhaiter considérer les effets du traitement comme un échantillon aléatoire d'une population inconnue plus large; cette méthode est appelée modélisation des effets aléatoires. Pour ces modèles, la technique moderne du REML (estimation résiduelle par la vraisemblance maximale) est recommandée. Le REML peut également être utilisé lorsqu'on veut comparer plusieurs laboratoires ou sites, pour estimer les composantes de la variance, ou lorsqu'un dispositif ne

peut pas être analysé par analyse de variance car il comporte tellement de valeurs manquantes qu'il est déséquilibré. Il existe là aussi des similitudes entre les concepts et les valeurs du REML et ceux de l'analyse de variance; cependant l'avis d'un expert est généralement nécessaire.

### 3.7 Données ordinales

Les méthodes modernes d'analyse des données catégorielles sont décrites dans le livre d'Agresti (1984), mais un avis statistique est nécessaire pour les utiliser à bon escient. Par ailleurs, il est parfois nécessaire de traiter des variables entières comme des variables ordinales, si leur gamme de variation n'est pas suffisamment importante pour qu'on puisse les considérer comme continues et que l'essai reste néanmoins valide.

### 3.8 Données qualitatives et méthodes non paramétriques

Des méthodes non paramétriques peuvent être utilisées pour les données qui sont réellement qualitatives (par exemple les données nominales) ainsi que pour certaines données classées, ou pour les données qui ne suivent pas une distribution paramétrique bien connue (comme les distributions normale, binomiale, beta, gamma ou de Poisson). Cependant, la quantité d'informations que ce type d'analyse peut communiquer est souvent limitée. En particulier, ce type d'approche se concentre sur les tests au détriment de la fonction plus importante d'estimation de l'amplitude des effets. Par ailleurs, les méthodes non paramétriques manquent de puissance.

Une très bonne description des tests traditionnels est donnée par Siegel (1956) et explique clairement quel test convient pour quelle série de données. Des approches plus modernes utilisent des techniques informatisées comme les tests de randomisation. Les méthodes de randomisation peuvent être très utiles lorsqu'on doit se méfier des approches paramétriques, par ex. si les données sont très éloignées de données normales, ou si elles comportent une forte proportion de zéros (dans le cas où l'essai est néanmoins considéré valide). D'autres méthodes informatisées non paramétriques sont recommandées pour améliorer les estimations, ou pour mieux calculer la variabilité d'une estimation; elles incluent les méthodes de "bootstrapping" et de "jackknifing", et nécessitent souvent l'avis d'un spécialiste.

## 4. Analyse statistique d'une série d'essais

Les essais conduits conformément aux directives et qui donnent des résultats réalistes et cohérents (voir plus haut) peuvent être regroupés pour une analyse statistique unique. Les produits à étudier sont comparés dans leur ensemble, comme décrit plus haut, avec le produit de référence et éventuellement entre eux, mais la fonction principale de l'analyse est de mesurer et de

<sup>1</sup> La présente édition de cette Norme adopte sur ce point une position transitoire. Une troisième édition est prévue pour tenir compte d'un futur consensus international sur les procédures de comparaison multiple dans ce contexte.

tester les interactions éventuelles entre les produits et les sites, c'est à dire de vérifier que les différences entre les produits se répètent aux différents sites.

## Remerciements

L'OEPP remercie vivement le professeur J.N. Perry pour la révision de cette norme.

## Bibliographie

- AGRESTI A. (1984) Analysis of Ordinal Categorical Data. Wiley; New York
- B.B.A. (1980) Richtlinie für Planung, Durchführung und Auswertung von Versuchen mit Pflanzenbehandlungsmitteln: 1. Versuchsplanung; 2. Versuchsdurchführung. Biologische Bundesanstalt für Land- und Forstwirtschaft, Braunschweig (DE)
- B.B.A. (1982) Richtlinie für Planung, Durchführung und Auswertung von Versuchen mit Pflanzenbehandlungsmitteln: 3. Auswertung des Einzelversuches; 4. Sachregister, Tabellen. Biologische Bundesanstalt für Land- und Forstwirtschaft, Braunschweig (DE)
- BINNS, M.R. & NYROP, J.P. (1992) Sampling insect populations for the purpose of IPM decision making. *Annual Review of Entomology* **37**, 427-453
- BRYAN-JONES J. & FINNEY D.J. (1983) On an error in "Instructions to authors". *HortScience* **18**, 279-282
- C.E.B. (1981) Rôle et implantation des témoins sans traitement dans les essais de produits phytosanitaires. ANPP-DT4. J. ARNOUX
- C.E.B. (1983) Principes d'appréciation des effets des produits phytosanitaires dans les essais de plein-champ. ANPP-DT5. Y. RIBRIOUX
- C.E.B. (1986) Utilisation des tests statistiques dans l'interprétation des essais de produits phytosanitaires. ANPP-DT6. J. ARNOUX - J.P. GOUET
- C.E.B. (1990) Les réseaux d'essais. ANPP-DT9. J.P. GOUET
- C.E.B. (1990) Les unités expérimentales. ANPP-DT10. J.P. GOUET
- COCHRAN W.G. & COX G.M. (1957) Experimental Design (2nd ed.). Wiley, New-York
- COX D.R. (1958) Planning of Experiments. Wiley, New-York
- CRAWLEY M.J. (1993) GLIM for Ecologists. Blackwell Scientific Publications, Oxford
- CROWDER M.J. & HAND D.J. (1990) Analysis of Repeated Measures. Chapman and Hall, London
- CULLIS B.R. & GLEESON A.C. (1991): Spatial analysis of field experiments – an extension to two dimensions. *Biometrics* **47**, 1449-1460
- DAGNELIE P. (1969) Théorie et Méthodes Statistiques. Vol. 2, Duculot, Gembloux
- DYKE, G.V. (1988) Comparative Experiments with field crops. Griffin, London
- FINNEY D.J. (1971) Probit Analysis (3rd ed.) Cambridge University Press
- FINNEY D.J. (1978) Statistical Method in Biological Assay. (3rd ed.), Griffin, London
- FINNEY D.J. (1980) Statistics for Biologists. Chapman & Hall, London
- GOUET J.P. & PHILIPPEAU G. (1992) Comment interpréter les résultats d'une analyse de variance? Editions ITCF
- GOUET J.P. (1974) Les Comparaisons de Moyennes et de Variances. Application à l'Agronomie. Institut Technique des Céréales et des Fourrages
- HOLLANDER M. & WOLFE D.A. (1973) Non-parametric Statistical Methods. Wiley, London
- HUGHES G. & MADDEN L.V. (1992) Aggregation and incidence of disease. *Plant Pathology* **41**, 657-660
- HURLBERT S.H. (1984) Pseudoreplication and the design of ecological field experiments. *Ecological Monographs* **54**, 187-211
- LITTLE T.M. (1978) If Galileo published in HortScience. *HortScience* **13**, 504-506
- McCULLAGH P. & NELDER J.A. (1983) Generalized Linear Models. Chapman & Hall, London
- MEAD R. & CURNOW R.N. (1983) Statistical Methods in Agriculture and Experimental Biology. Chapman & Hall, London
- MEAD R. (1988) The design of experiments: Statistical principles for practical applications. CUP, Cambridge
- NELDER J.A. (1971) Contribution to the discussion of the paper by O'Neill and Wetherill. *Journal of the Royal Statistical Society series B* **36**, 218-250
- PATTERSON H.D. & THOMPSON R. (1971) Recovery of inter-block information when block sizes are unequal. *Biometrika* **58**, 545-554
- PEARCE S.C., CLARKE G.M., DYKE G.V. & KEMPSON R.E. (1988) Manual of Crop Experimentation. Griffin, London
- PERRY J.N. (1986) Multiple-comparison procedures: a dissenting view. *Journal of Economic Entomology* **79**, 1149-1155
- PERRY J.N. (1989) Review: population variation in Entomology: 1935-1950. I. Sampling. *The Entomologist* **108**, 184-198
- PERRY J.N. (1994). Sampling and applied statistics for pests and diseases. *Aspects of Applied Biology* **37**, 1-14
- PERRY J.N. (1997) Statistical aspects of field experiments. pp. 171-201 in: Methods in Ecological and Agricultural Entomology - Chapter 7. (Eds. D.R. Dent & M.P. Walton). CABI, Wallingford.
- PLACKETT R.L. (1981) The Analysis of Categorical Data. (2nd ed), Griffin, London
- PREECE, D.A. (1982) The design and analysis of experiments: what has gone wrong? *Utilitas Mathematica* **21A**, 201-244
- RASCH D., HERRENDÖRFER G., BOCK J., VICTOR N. & GUIARD V. (1996)

Verfahrensbibliothek, Versuchsplanung und -auswertung, Band I. R. Oldenbourg Verlag GmbH, München

RASCH D., HERRENDÖRFER G., BOCK J., VICTOR N. & GUIARD V. (1998) Verfahrensbibliothek, Versuchsplanung und -auswertung, Band II. R. Oldenbourg Verlag GmbH, München

ROBINSON, D.L. (1987) Estimation and use of variance components. *The Statistician* **36**, 3-14

SIEGEL S. (1956) Nonparametric Statistics for the Behavioral Sciences. McGraw Hill, New York

YATES F. (1937) The Design and Analysis of Factorial Experiments - Technical Communication No.35 of the Commonwealth Bureau of Soils, Harpenden, England. C.A.B., Farnham Royal, Surrey, England

## Annexe I

*Exemples d'échelles utilisées dans les directives OEPP pour l'évaluation biologique des produits phytosanitaires*

### 1. Nominale

Décoloration des feuilles de pomme de terre (Norme OEPP PP 1/135(2) Directive sur la phytotoxicité):

chlorose;

nervures jaunes;

taches jaunes;

couleur générale vert foncé ou vert pâle;

albinisme.

### 2. Ordinale sans intervalles définis quantitativement

Evaluation des racines de chou pour *Plasmodiophora brassicae* (Norme OEPP PP 1/39(2) Directive pour l'évaluation biologique des fongicides contre *Plasmodiophora brassicae*):

(1) aucune hernie;

(2) légères hernies, pour la plupart sur les racines latérales;

(3) hernie moyenne, sur la racine principale et/ou les racines latérales;

(4) forte hernie de la racine principale et/ou des racines latérales.

Evaluation des laitues pour *Botrytis cinerea* (Norme OEPP PP 1/54(2) Directive pour l'évaluation biologique des fongicides contre *Botrytis* spp. sur cultures légumières):

(1) plantes indemnes;

(2) attaque légère, infection uniquement des pétioles de la base;

(3) attaque moyenne, lésion de la tige mais ne l'encerclant pas;

(4) forte attaque, lésion de la tige l'encerclant complètement, laitue non commercialisable.

### 3. Echelles ordinales avec intervalles définis basés sur des nombres

Attaque par *Venturia inaequalis* sur pomme (Norme OEPP PP 1/5(3) Directive pour l'évaluation des fongicides contre *Venturia inaequalis* et *V. pirina*):

(1) pas d'attaque;

(2) 1-3 taches d'infection par fruit;

(3) > 3 taches d'infection par fruit.

Nombre de lésions sur betteraves dues à *Scutigerella immaculata* (Norme OEPP PP 1/45(2) Directive pour l'évaluation biologique des insecticides contre *Scutigerella immaculata*):

(1) pas de lésion;

(2) 1-2 lésions;

(3) 3-5 lésions;

(4) > 5 lésions.

Certaines échelles sont basées en partie sur un nombre et en partie sur une surface, par ex.:

(1) feuille saine;

(2) 1-2 taches par feuille;

(3) >2 taches par feuille;

(4) >1/3 de la surface foliaire infectée.

### 4. Echelles ordinales avec intervalles définis basés sur des variables continues

Evaluation de l'infection du blé par *Pseudocercospora herpotrichoides* (directive no 28)

(1) pas de symptôme;

(2) moins de 50% de la circonférence attaquée;

(3) plus de 50% de la circonférence attaquée;

(4) 100% de la circonférence de la tige attaquée.

Généralement une telle échelle sera (au moins en partie) logarithmique.

Surface foliaire de l'olivier attaquée par *Spilocaea oleagina* (Norme OEPP PP 1/81(2) Directive pour l'évaluation biologique des fongicides contre *Spilocaea oleagina*).

(1) pas d'attaque;

(2) 0-10% de la surface foliaire attaquée;

(3) 10-25% de la surface foliaire attaquée;

(4) 25-50% de la surface foliaire attaquée;

(5) 50-100% de la surface foliaire attaquée.

Bien que ces échelles soient en apparence logarithmiques, il n'y a en fait pratiquement jamais de progression logarithmique constante entre la valeur centrale d'une classe et la valeur centrale suivante. Ainsi, bien qu'en théorie il soit possible d'analyser les notes d'une échelle comme des variables continues correspondant à la simple transformation logarithmique des variables originales, ce cas ne se présente presque jamais car les échelles ne sont pas réellement logarithmiques. D'autre part, l'attribution de la valeur 1 à la classe zéro n'est pas homogène avec le reste de l'échelle.

Un autre point à souligner est que les classes sont définies par les intervalles d'une variable continue. Dans le cas des feuilles d'olivier atteintes par *S. oleagina* (ci-dessus) l'observateur regarde la feuille et décide de l'attribuer dans la classe 3 ou 4. Il ne procède pas en regardant la feuille, décidant qu'elle est atteinte à 35%, puis la plaçant dans la classe 4. S'il faisait ainsi, il pourrait tout aussi bien faire une estimation directe et ne pas utiliser d'échelle, car il obtiendrait ainsi des informations plus utiles et mieux analysables. Il en ressort que la présentation des intervalles sous la forme 0-10, 11-25, 26-50, 51-100 n'a pas de sens et est inadaptée. Si l'observateur trouve une feuille qui lui semble atteinte à environ 50%, il devra décider subjectivement si elle est à mettre dans la classe 4 ou 5, sans essayer de différencier entre 50 ou 51 (ce qui est manifestement impossible).

Dans de rares cas, des catégories descriptives sont mélangées avec des intervalles définis:

Evaluation des feuilles de pommiers pour *Podosphaera leucotricha* (Norme OEPP PP 1/69(2) Directive pour l'évaluation biologique des fongicides contre *Podosphaera leucotricha*):

- (1) pas de tache;
- (2) attaque légère (taches éparses);
- (3) attaque moyenne à importante (taches couvrant jusqu'à 50% de la surface foliaire);
- (4) attaque très importante (plus de 50% de la surface foliaire atteinte; bords des feuilles commençant à s'enrouler et à se dessécher).

##### 5. Echelles ordinales avec classes définies par leur valeur centrale

Ce type d'échelle est à considérer comme une aide à l'estimation. Les échelles visuelles sont les plus fréquentes (exemple: pour *Cercospora beticola*, *Peronospora tabacina*, etc.).

Ces échelles servent généralement à estimer le % de surface foliaire atteinte, et ont fait l'objet d'un calibrage soigné.

Les degrés sont généralement choisis de manière à couvrir la gamme d'attaque prévue (exemple: 1, 5, 10, 25, 50) et de manière à permettre l'interpolation, plutôt que dans une séquence régulière pratiquement logarithmique (qui serait préférable pour faire des notations).

Dans le cas de la Norme OEPP PP 1/2(3) Directive pour l'évaluation biologique des fongicides contre *Phytophthora infestans* sur pomme de terre, les points de l'échelle sont illustrés par des termes descriptifs plutôt que par une échelle visuelle et les descriptions ont fait l'objet d'un calibrage soigné par rapport au % de surface foliaire atteinte. Néanmoins, la nature de l'échelle reste la même:

- (0) pas d'infection;
- (1) jusqu'à 10 taches par plante, ou taches peu sévères réparties sur toute la plante;
- (5) environ 50 taches par plante ou jusqu'à 1 feuille attaquée sur 10;

- (10) jusqu'à 4 feuilles infectées sur 10, les plantes ont encore une forme normale;
- (25) pratiquement chaque feuille est atteinte mais les plantes ont encore une forme normale; la parcelle peut sembler verte bien que toutes les plantes soient atteintes;
- (50) toutes les plantes sont atteintes et environ la moitié de la surface foliaire est abîmée; la parcelle semble verte mais est tachetée de brun.