

L'Opération Sources en Belgique: premières analyses statistiques des résultats et légitimité des cartes "nitrates"*

J. M. WATTECAMPS

Fondation Universitaire Luxembourgeoise, 185 Avenue de Longwy, B-6700 Arlon, Belgique

J. BERNIER

Ecole Nationale du Génie rural et des Eaux et Forêts, 19 Avenue du Maine, F-75732 Paris Cedex 15, France

L. DE BACKER

Université Catholique de Louvain, Faculté des Sciences Agronomiques, Département de Génie Rural, Place Croix du Sud, B-1348 Louvain-la-Neuve, Belgique

J. J. BOREUX

Fondation Universitaire Luxembourgeoise, 185 Avenue de Longwy, B-6700 Arlon, Belgique
e-mail: boreux@ful.ac.be

Résumé L'Opération Sources du WWF (Worldwide Fund for Wildlife) est une action éducative de sensibilisation à la qualité des eaux souterraines. Elle implique principalement les élèves de cinquième et sixième primaires de Belgique. Depuis 1994, des centaines de dosages des ions nitrates ont été effectués dans un grand nombre de sources et de puits à l'aide de bandelettes colorimétriques. A l'issue de chaque campagne, les points d'eau surveillés ont été reportés sur une carte de Belgique où leur couleur indique leur qualité eu égard à la concentration en nitrates. L'objectif pédagogique étant atteint, la validation de ces cartes en tant qu'image de la réalité pose le problème de la vérification du lien entre la concentration réelle (mg l^{-1}) du point d'eau visité et sa couleur sur la carte. A cette fin, des vérifications ont été réalisées sur 308 points d'eau choisis au hasard dans l'ensemble des sources et puits visités. L'analyse statistique révèle que la relation entre le test colorimétrique et le contrôle réalisé en laboratoire est hautement significative. Ce résultat justifie des investigations ultérieures, notamment la mobilisation de l'information régionale pour la qualification d'un point d'eau peu surveillé et la comparaison des cartes dans le temps.

Operation Sources in Belgium: preliminary statistical analyses of the results and validation of the "nitrates" map

Abstract The Opération Sources of the WWF (Worldwide Fund for Wildlife) in Belgium is an environmental educational campaign focused on the groundwater quality. It involves pupils in the fifth and sixth grades of primary school in Belgium. Since 1994, several hundred analyses of nitrate have been carried out using colorimetric test strips at numerous springs and wells throughout Belgium. At the end of each campaign, the point sources of water surveyed were recorded on a map of Belgium, using different colours to indicate their range of nitrate concentration. The educational objective having been fulfilled, in order to validate these maps it was necessary to verify for each visited point source the relationship between its actual nitrate concentration (mg l^{-1}) and its colour on the map. To this end, verification was carried out on 308 point sources chosen at random from those tested. The statistical analysis shows that the relationship between the results of the colorimetric *in situ* tests and corresponding laboratory control checks is highly significant. This result warrants

*Données récoltées dans le cadre de l'Opération Sources du WWF en Belgique (campagnes 1994–1996).

further investigations, notably the use of regional information to qualify a poorly surveyed source point and also to check the temporal changes of the maps.

INTRODUCTION

L'Opération Sources est une campagne éducative portant sur la qualité des eaux souterraines. Elle a été lancée en 1994 par le Fond Mondial pour la Nature (WWF) et concerne l'ensemble du territoire belge. D'une part, l'Opération Sources a pour but d'informer la jeunesse, de la sensibiliser et de la responsabiliser en l'invitant à des engagements significatifs en faveur de la préservation et de l'amélioration des ressources en eau douce. D'autre part, cette campagne a également pour ambition de contribuer à une gestion rationnelle de l'environnement par une surveillance régulière des teneurs en nitrates dans l'eau de sources et de puits.

Grosso modo la Belgique (Fig. 1) peut être divisée en deux parties: la Wallonie ($16.9 \times 10^3 \text{ km}^2$) occupe la partie sud du pays et la Flandre ($13.6 \times 10^3 \text{ km}^2$), la partie nord. Lors de chaque campagne, les élèves des écoles primaires visitent plus de 600 points d'eau dont 350 sources, essentiellement situées en Wallonie, et 250 puits pour la plupart situés en Flandre en raison de son faible relief.

La teneur en nitrates du point d'eau visité est mesurée à l'aide d'un simple test colorimétrique qui fournit la classe de concentrations à laquelle il appartient. Pour ce qui concerne les points d'eau surveillés il convient d'emblée de faire quatre remarques importantes:

- ils représentent à peine 3% du nombre total des points d'eau que l'on peut grossièrement évaluer à 20 000 unités en Belgique;
- ils ne sont pas choisis au hasard mais bien en fonction de la proximité de l'établissement scolaire et/ou de leur accessibilité physique ou légale;
- toutes les émergences n'ont pas le même poids: leur débit, par exemple, est une donnée essentielle pour apprécier leur représentativité des eaux souterraines;
- les élèves peuvent être amenés à visiter des points d'eau géographiquement proches mais appartenant à des circuits hydrauliques très différents.

En raison de ces remarques préliminaires la valeur scientifique des données recueillies dans le cadre de l'Opération Sources est discutée ici. Particulièrement, la question est posée de la pertinence de la représentation des points d'eau visités sur une carte de Belgique où leur couleur indique la classe de concentrations en nitrates à laquelle ils ont été affectés (Fig. 1). En effet, dans la mesure où de campagne en campagne une telle carte fournirait une image convenable de la réalité, elle synthétiserait remarquablement la qualité des eaux souterraines en Belgique et son évolution dans le temps. L'argument est d'autant plus fondé si l'on considère qu'un test à l'aide d'une bandelette colorimétrique réalisé *in situ* est 50 fois moins coûteux qu'un titrage au laboratoire (ce chiffre ne considère que le coût unitaire du dosage hors les frais consécutifs à la collecte et à l'acheminement des échantillons).

Cette contribution se divise en trois parties. La première partie présente l'Opération Sources, notamment son concept de base, le protocole expérimental, la nature des données et leur codification et justifie la méthode d'analyse statistique retenue. La seconde partie détaille l'analyse statistique de la liaison entre les tests colorimétriques réalisés *in situ* et les contrôles réalisés au laboratoire. La troisième partie examine l'influence de quelques anicroches reconnues sur la sensibilité des

résultats. Nous concluons en abordant le thème de la mobilisation de l'information régionale pour juger de la qualité d'un point d'eau peu surveillé. Ce dernier aspect, plus technique, fera d'ailleurs l'objet d'une publication ultérieure.

L'OPÉRATION SOURCES

Motivation et concept de base

L'Opération Sources présente un intérêt considérable sur le plan éducatif. Ainsi en 1996, 1769 classes primaires étaient inscrites à l'Opération Sources. Potentiellement c'est donc plus de 30 000 élèves qui ont pu participer activement au projet. Grâce à cette initiative, des centaines de milliers de personnes ont été sensibilisées à la problématique de la qualité des eaux souterraines par le biais des médias. Selon le WWF, le tirage cumulé des quotidiens et hebdomadaires ayant présenté la campagne de mars 1996 s'élevait à 4 500 000 exemplaires.

L'idée est simple. Deux fois l'an, en mars (période de crue) et en octobre (période d'étiage), les classes de cinquième et sixième années primaires (enfants de 10–12 ans) vont mesurer la teneur en nitrates dans l'eau des sources et des puits situés à proximité de leur école. Le test est réalisé à l'aide de bandelettes colorimétriques fournies par le WWF. A l'issue de chaque campagne de mesures, tous les résultats sont reportés sur une carte de Belgique—la carte nitrates—diffusée dans les écoles et les journaux nationaux (Fig. 1). Cette carte situe les points d'eau visités et les représente à l'aide d'un code couleur propre à la classe de concentrations en nitrates mesurée *in situ*. De plus, elle distingue les sources des puits. Naturellement, les enseignants ne manquent pas d'exploiter chaque campagne de mesures. L'Opération Sources est en fait conçue comme un tremplin afin que chaque classe s'engage dans un projet pédagogique dont l'objectif principal est la sensibilisation des jeunes à une gestion soutenable de l'environnement, notamment de la ressource "eau".

Le protocole expérimental

La surveillance des points d'eau et son exploitation à des fins pédagogiques sont expliquées aux enseignants dans un dossier didactique fourni à chaque classe participante (Wattecamps & Michiels, 1995).

Ce dossier présente la problématique de la pollution des eaux souterraines par les nitrates. Il propose des activités en vue de sensibiliser les enseignants et les enfants à cette question. Le dossier insiste particulièrement sur la prévention de la pollution et incite les classes à s'engager à leur niveau afin d'agir en faveur de la ressource en eau douce.

Une attention toute particulière a été apportée aux "fiches techniques" qui précisent les protocoles à suivre pour réaliser les différentes observations et mesures. Ainsi, sur le terrain, pour chaque point d'eau visité, les classes doivent:

1. localiser le point d'eau sur une carte topographique IGN (Institut Géographique National) à l'échelle 1:25 000;
2. identifier la nature du point d'eau (source ou puits) et décrire ses caractéristiques principales;

3. observer et décrire l'environnement dans un rayon de 200 m en amont et en aval du point d'eau visité (surtout l'amont);
4. effectuer le test nitrate à l'aide des bandelettes colorimétriques et procéder aussitôt à son décodage visuel (i.e. sans l'aide d'un réflectomètre);
5. prélever un échantillon d'eau dans une éprouvette stérile (fournie par le WWF) et l'envoyer immédiatement au laboratoire à des fins de contrôle (uniquement pour un petit nombre de classes choisies au hasard);
6. rédiger un rapport standardisé et l'envoyer au WWF dans les meilleurs délais.

Remarquons que toutes les mesures sont réalisées dans une période de 15 jours se situant de part et d'autre des dates pivots du 22 mars—date retenue en raison de sa portée symbolique puisqu'il s'agit de la "Journée Mondiale de l'Eau"—et du 17 octobre. Au reste, ces périodes correspondent (théoriquement) respectivement à des hauts et bas niveaux piézométriques.

Limitations Le dossier pédagogique demande expressément que la classe soit partagée en trois groupes d'élèves, chaque groupe devant procéder au test colorimétrique indépendamment des deux autres. Cette procédure a non seulement pour but d'impliquer tous les élèves mais permet aussi, au moins *a priori*, d'écarter les éventuelles mesures aberrantes. L'analyse des résultats va montrer que l'appréciation de la couleur prise par l'élément sensible de la bandelette-test fait fréquemment l'objet d'un consensus extra groupes.

Ceci étant dit, reprenons point par point le protocole expérimental. La localisation des points d'eau visités et le renvoi dans les délais du rapport standardisé (points 1 et 6) ne présentent plus de difficultés majeures. Le travail réalisé peut être considéré comme bon à excellent dans 90% des cas. A ce jour, l'identification et la caractérisation des points d'eau visités (point 2) n'ont pas fait l'objet d'une analyse suffisante pour être correctement évaluées. En accompagnant les classes sur le terrain, on constate que la description de l'amont de la source (point 3) reste très sommaire et diversement appréciée par les élèves. Lors des deux premières campagnes réalisées en 1994, l'acheminement des échantillons de contrôle vers le laboratoire (point 5) a posé beaucoup de problèmes mais les campagnes suivantes ont permis d'affiner la procédure, notamment pour ce qui concerne la conservation des échantillons. Cependant dans 33% des cas, il subsiste un risque non maîtrisable de dénitrification. Nous verrons dans la seconde partie de cet article si la dénitrification des échantillons peut modifier la signification de la relation entre les tests colorimétriques et les analyses de contrôle en laboratoire.

Pour ce qui concerne le point 4 relatif au test colorimétrique réalisé *in situ*, il est l'objet même de cet article.

Nature des données et hypothèses de raisonnement

Les six premières campagnes ont permis de visiter plus de 4500 points d'eau mais à l'heure d'écrire cette note, seules les quatre premières campagnes présentent des contrôles au laboratoire (Tableau 1).

Dès ici nous voulons insister sur le fait que la concentration en nitrates mesurée au laboratoire est une variable quantitative continue (i.e. le résultat est un nombre réel)

Tableau 1 Opération Sources: nombre de points d'eau testés et contrôlés lors de chaque campagne.

Campagne	Testés	Contrôlés
Mars 1994	790	99
Octobre 1994	762	92
Mars 1995	898	81
Octobre 1995	628	36
Mars 1996	861	
Octobre 1996	620	
Total	4459	308

alors que le test colorimétrique pratiqué *in situ* est fondamentalement une variable qualitative catégorielle (i.e. le résultat est une couleur) même si une échelle numérique lui est associée. Au demeurant, comme le but poursuivi est la validation de l'Opération Sources en tant qu'outil d'aide à la décision, le test colorimétrique doit être considéré comme la réponse au titrage au laboratoire qui tient donc le rôle de la variable explicative. En d'autres termes et comme le montre le Tableau 1, la validation des cartes nitrates actuellement disponibles ne repose que sur 308 couples (X,Y) où les variables aléatoires X et Y représentent respectivement la concentration en nitrates (mg l^{-1}) mesurée au laboratoire et la concentration révélée *in situ* par le test colorimétrique. La relation entre ces deux variables étant au centre de nos préoccupations, nous la noterons:

$$R = R(X,Y) \quad (1)$$

Dès maintenant et pour alléger le texte, nous utiliserons communément le mot "source" pour désigner indifféremment une source ou un puits.

L'hypothèse d'homogénéité Remarquons d'emblée que les 308 sources testées et contrôlées ne représentent que 10% des sources testées par les élèves. C'est peu pour étendre les résultats obtenus à l'ensemble du territoire belge, d'autant plus que les sources visitées ne représentent qu'à peine 3% du nombre estimé de points d'eau du pays.

Pourtant, nous supposons que la relation (1) peut être appliquée à toute source du territoire belge. En d'autres termes, nous n'aborderons pas les facteurs systématiques d'hétérogénéité spatiale du milieu surveillé.

L'hypothèse de raison Du point de vue de la précision, la relation (1) fait intervenir deux types de "mesures" radicalement différentes.

- (a) Le test colorimétrique: l'expertise repose sur le fait que les ions nitrate (NO_3^-) modifient la couleur de l'élément sensible de la bandelette-test. La "mesure" consiste donc à évaluer *in situ* une densité de couleur par comparaison avec six couleurs de référence proposées par le fabricant. Selon le fabricant des bandelettes-test, l'affectation de la source surveillée à une classe de concentration en nitrates (mg l^{-1}) repose sur la correspondance suivante:

No. de la classe	1	2	3	4	5	6
Concentration (mg l^{-1})	0-10	10-25	25-50	50-100	100-250	250-500

Nous discuterons tout à l'heure de certains problèmes pratiques liés à l'affectation d'une source à une classe mais il n'est pas douteux que l'évaluation de la couleur

par l'élève peut être soumise à une incertitude notable. Celle-ci s'ajoute à celles provenant d'une part des erreurs éventuelles dans l'élaboration du tableau de correspondance du fabricant (vraiment très faibles si le fabricant est sérieux) et d'autre part des erreurs de conservation et de manipulation des bandelettes par l'élève. Au total le test colorimétrique produit une information qui doit être considérée, au moins *a priori*, comme entachée d'une incertitude non négligeable.

- (b) Le titrage au laboratoire: Certaines classes scolaires tirées au sort remplissent une éprouvette stérile avec l'eau de source qu'elles viennent de tester. Cet échantillon est bientôt envoyé vers un laboratoire agréé où la concentration en nitrates peut être déterminée à 0.5 mg l^{-1} près. Bien entendu, la prise d'échantillon, les conditions d'envoi, les durées de conservation variables ou trop longues et les dénitrifications éventuelles peuvent invalider la représentativité de l'échantillon d'eau prélevé envers l'état réel de la source.

Nonobstant, le titrage en laboratoire apparaît moins incertain que la mesure *in situ*. C'est pourquoi, nous ferons l'hypothèse que l'état réel de la source visitée correspond à la concentration en nitrates donnée par le laboratoire.

Choix d'une méthodologie pour l'étude de la relation $R = R(X, Y)$

Sous ces deux hypothèses, l'étude de la relation (1) pose deux questions essentielles:

- Comment mesurer l'intensité et la signification de la relation (1)?
- Comment varie-t-elle? Est-elle sensible à certains facteurs, e.g. saisonniers?

On peut penser avoir le choix entre deux approches:

1. représenter chaque classe du test colorimétrique par une concentration caractéristique de la classe en question, p. ex. la médiane de la classe en mg l^{-1} , et traiter les couples d'observations par les méthodes classiques de régression statistique;
2. catégoriser les titrages en laboratoires et analyser un tableau de contingence.

A première vue une régression pourrait sembler avoir un sens (le coefficient de corrélation linéaire égale 0.70). En fait il n'en est rien et la régression est, ici, très fallacieuse. D'une part, parce que les données colorimétriques, exprimées en médiane de classe (mg l^{-1}) ou pas, restent des catégories. D'autre part, nous avons vu que les titrages au laboratoire et les données bandelettes sont soumis à des erreurs d'importance et de nature très différentes. Il ne faut pas oublier que le modèle de régression classique demande que les variables explicatives (ici il n'y en a qu'une notée X) soient sans erreur ou que celles ci soient comparables. Une solution consisterait bien sûr à inverser le rôle des variables mais on serait alors en opposition avec l'objectif poursuivi qui est la validation du test colorimétrique et pas la validation du titrage au laboratoire (notons encore une fois le caractère dissymétrique des distributions intéressantes pour le problème).

La seconde approche est basée sur la "catégorisation" de l'ensemble des données disponibles. Cette approche réduit légèrement l'information sur les données "laboratoire" mais présente l'avantage de répondre correctement au deux questions posées par la relation R .

La catégorisation des observations Avant de passer à cette étude proprement dite, il convient de faire l'inventaire des problèmes rencontrés par le modélisateur dans

l'élaboration du tableau de données à analyser. Nous allons voir que l'affectation d'un résultat à une classe de concentrations impose de faire un choix et de poser une nouvelle hypothèse.

- (a) Le problème de dépendance extra groupes: Nous avons déjà eu l'occasion de souligner une première difficulté qui vient de la dépendance entre les tests colorimétriques réalisés sur la même source. En effet, conformément aux indications du dossier pédagogique, pour chaque point d'eau visité l'instituteur répartit ses élèves en trois groupes d'observateurs. On espère obtenir ainsi trois observations indépendantes et donc une certaine validation de l'observation visuelle de la couleur prise par l'élément sensible de la bandelette. Cependant sur les 308 sources ayant donné lieu à un contrôle en laboratoire, 72% montrent des résultats identiques, ce qui laisse penser que les trois groupes d'élèves échangent des messages. Il en résulte une dépendance entre les observations. Quant aux 28% restants ils sont caractérisés, soit par des observations contradictoires (16%), soit par des observations absentes (12%). En conséquence on ne retiendra qu'une seule donnée de terrain par point d'eau visité.
- (b) Le problème de l'encodage de l'information "bandelette": Une seconde difficulté vient de l'encodage lui-même. Seulement 39% des observateurs—soit 120 sur 308—présentent le résultat du test conformément aux instructions données dans le dossier pédagogique, à savoir un intervalle de concentrations selon la correspondance proposée par le fabricant. Il est fâcheux que les 61% restants (188 sur 308) n'aient fourni qu'une limite d'intervalle ou une valeur nulle. Les observateurs qui ont donné une limite d'intervalle semblent avoir privilégié les valeurs par excès, du moins au cours de la première campagne de 1994.

Dès lors, nous ferons temporairement l'hypothèse que les valeurs individuelles correspondent aux limites supérieures des classes (cette hypothèse se révélera par ailleurs difficile à maintenir).

Cette hypothèse pose le problème du traitement des valeurs 0. Nous leur affecterons une classe particulière, notée zéro. En outre, comme aucune des 308 sources contrôlées n'appartient à la classe 6 (250–500), nous les noterons dorénavant 0, 1, 2, 3, 4, 5.

La catégorisation des contrôles Les concentrations sont fournies le plus souvent au 0.1 mg l^{-1} près. Bien que leur précision soit légèrement inférieure à ce chiffre nous conserverons la décimale. Pour les ranger dans les mêmes classes que les observations par bandelettes nous admettons que les intervalles de concentration sont fermés à droite (limite supérieure incluse) et ouverts à gauche (limite inférieure non incluse). Cette convention pose alors le problème de l'affectation à une classe des valeurs égales ou proches de 0. Nous rangerons dans la classe zéro toutes les concentrations inférieures à 0.1 mg l^{-1} . En définitive on raisonnera sur les correspondances suivantes (Tableau 2).

Tableau 2 Opération Sources: règles pour l'affectation d'un résultat à une classe de concentrations.

No. de la classe	0	1	2	3	4	5
Bandelette (120)	–	0–10	10–25	25–50	50–100	100–250
Bandelette (188)	0	10	25	50	100	250
Laboratoire (308)	[0,0.1]	[0.1,10]	[10,25]	[25,50]	[50,100]	[100,250]

Tableau 3 Opération Sources: données disponibles pour l'étude de la relation R . Les classes bandelettes sont en lignes (variable Y) et les classes labo en colonnes (variable X).

	x0	x1	x2	x3	x4	x5	Total
y0	5	22	2	2	0	0	31
y1	1	28	19	6	3	1	58
y2	1	9	29	24	0	0	63
y3	0	3	11	46	17	1	78
y4	0	0	2	13	44	7	66
y5	0	0	0	1	6	5	12
Total	7	62	63	92	70	14	308

L'étude de la relation (1) tire parti des données présentées dans le Tableau 3. Ce tableau, d'effectif $n = 308$, présente dans la case i,j ($0 \leq i,j \leq 5$) le nombre de sources rangées dans la classe j par le labo (variable X , indice j) et dans la classe i par les élèves suite au test colorimétrique (variable Y , indice i).

ETUDE STATISTIQUE DE LA RELATION R

Les données se présentent donc sous la forme d'un tableau de contingence (Tableau 3). Dès maintenant et pour la suite nous utilisons le système de notations conventionnel dans lequel: n_{ij} désigne l'effectif observé dans la case croisée i,j ; $n_{i\cdot}$ et $n_{\cdot j}$ désignent respectivement les totaux par lignes et colonnes (on remplace par un point l'indice qui varie); enfin n représente la taille de l'échantillon étudié ($n = 308$).

Existence de la contingence

On rappelle que la contingence est une mesure de la déviation totale par rapport à l'indépendance. Il est assez naturel de s'interroger sur la réalité d'une telle déviation avant de tenter de la quantifier. Le Tableau 4 donne donc les effectifs attendus \hat{n}_{ij} ($0 \leq i,j \leq 5$), c'est-à-dire les effectifs qu'on devrait observer si le contrôle au laboratoire (variable X , indice j) était indépendant du test colorimétrique (variable Y , indice i). On les estime à l'aide de la relation suivante:

$$\hat{n}_{ij} = \frac{n_{\cdot j} n_{i\cdot}}{n} \quad (2)$$

Tableau 4 Opération Sources: effectifs attendus sous l'hypothèse d'indépendance. Les classes bandelettes sont en lignes (variable Y) et les classes labo en colonnes (variable X).

	x0	x1	x2	x3	x4	x5	Total
y0	0.7	6.2	6.3	9.3	7.0	1.4	31
y1	1.3	11.7	11.9	17.3	13.2	2.6	58
y2	1.4	12.7	12.9	18.8	14.3	2.9	63
y3	1.8	15.7	16.0	23.3	17.7	3.5	78
y4	1.5	13.3	13.5	19.7	15.0	3.0	66
y5	0.3	2.4	2.5	3.6	2.7	0.5	12
Total	7	62	63	92	70	14	308

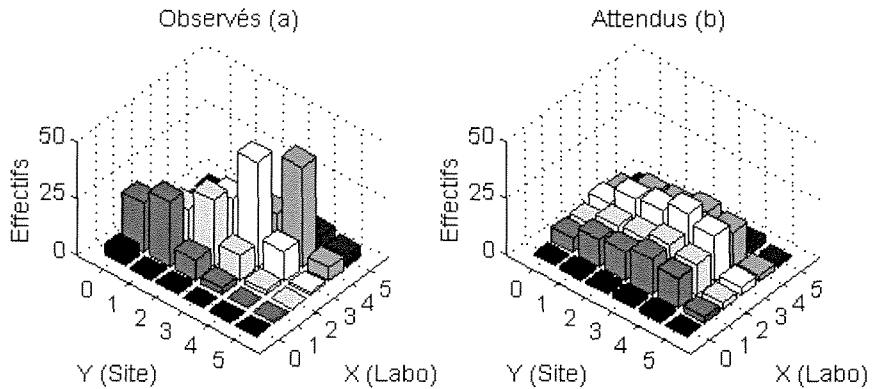


Fig. 2 Opération Sources: la représentation 3D des effectifs observés et des affectifs attendus sous l'hypothèse d'indépendance montre l'existence d'une contingence.

La Figure 2 montre les effectifs observés et attendus sous l'hypothèse d'indépendance. On devine bien sûr un lien entre le test colorimétrique et la concentration réelle de la source en nitrates (le contraire ôterait toute crédibilité à l'Opération Sources). Evidement, la rigueur scientifique exige une mesure de ce lien.

Test d'indépendance du χ^2

Ce test d'indépendance mesure en fait la distance qui sépare les Tableaux 3 et 4. Le carré de cette distance est mesuré par:

$$\chi^2 = \sum_{i=0}^5 \sum_{j=0}^5 \frac{(n_{ij} - \hat{n}_{ij})^2}{\hat{n}_{ij}} \quad (3)$$

Pour un seuil de signification α choisi, on compare la valeur (3) avec la valeur théorique $\chi_{v,\alpha}^2$ extraite de la loi du χ^2 à v degrés de liberté.

Le nombre de degrés de liberté v représente le nombre de cases d'un tableau de contingence à r lignes et c colonnes que l'on peut remplir arbitrairement sous la contrainte des marges. Dès lors: $v = (r - 1)(c - 1)$.

Si le χ^2 calculé est inférieur au $\chi_{v,\alpha}^2$ de la table, on considère que les deux variables sont indépendantes. Dans le cas contraire on a deux possibilités:

- les deux variables X et Y sont réellement indépendantes et le résultat obtenu (3) est dû à l'échantillonnage aléatoire (l'échantillon des sources contrôlées est plutôt singulier puisque, *a priori*, il avait moins de α chances sur cent d'être observé);
- l'échantillon observé représente bien sa population parente (les quelques trente mille sources belges) ce qui conduit naturellement au rejet de l'hypothèse d'indépendance entre les variables X et Y .

Presque tous les praticiens "oublient" la première possibilité et gèrent la décision comme suit:

$$\begin{cases} \chi^2 \leq \chi_{\alpha, \nu}^2 \Rightarrow H \\ \chi^2 > \chi_{\alpha, \nu}^2 \Rightarrow K \end{cases} \quad (4)$$

où H et K désignent respectivement l'hypothèse nulle "les variables X et Y sont indépendantes" et l'hypothèse alternative "les variables X et Y sont dépendantes".

L'habilité ne réside évidemment pas dans le calcul de (3) ni dans l'utilisation automatique des inégalités (4).

- Ainsi, la première possibilité indiquée ci-dessus doit être envisagée et ce n'est pas la statistique qui permet de l'exclure. Le praticien doit pouvoir argumenter que l'échantillon dont il dispose est représentatif de l'ensemble de la population étudiée. Pour ce qui nous concerne, l'hypothèse d'homogénéité nous permet de le faire.
- De la même manière, la règle de décision par tout ou rien est trop absolue dans le cas où le χ^2 calculé est très inférieur à $\chi_{\alpha, \nu}^2$. En effet, dans un tel cas on conclut a fortiori à l'indépendance des caractères étudiés sur l'ensemble de la population. Or un tel χ^2 calculé signifie que les effectifs observés sont extrêmement proches des effectifs attendus. Pourtant, si les caractères étudiés sont réellement indépendants et si les individus observés ont véritablement été choisis aléatoirement, cela signifie qu'on dispose d'un échantillon qu'on avait très peu de chances d'obtenir! On peut donc suspecter un problème majeur dans la prise d'échantillon, p. ex. le caractère aléatoire simple de l'échantillon n'est pas acquis.
- Dans le cas où le test révèle une dépendance entre les variables X et Y , le χ^2 calculé (3) n'est plus une valeur particulière d'une loi du χ^2 à ν degrés de liberté. On sait donc qu'il y a une liaison mais il reste à en évaluer la force.
- Enfin, rappelons aussi que ce test est peu fiable lorsque certains effectifs estimés (2) sont petits (habituellement inférieur à cinq). On peut régler ce problème en regroupant plusieurs catégories en une seule pourvu que ce regroupement ait un sens.

Eu égard à cette dernière remarque, l'examen des effectifs attendus (Tableau 4) impose un regroupement de certaines classes de concentrations (Tableau 2). Il est assez naturel que ce regroupement s'effectue selon les diagonales des Tableaux 3 et 4 car cela revient à considérer comme constante l'erreur d'estimation en classe mesurée par la différence $Z = Y - X$. En d'autres termes, on affecte dans la case, notée z_k , toutes les sources qui sont rangées dans la classe i par les élèves (variable Y) et dans la classe $j = i + k$ par le labo (variable X) ($0 \leq i, j \leq 5$; $-5 \leq k \leq 5$) (Tableau 5). Les deuxième et

Tableau 5 Regroupement selon l'erreur d'estimation en classe Z mesurée par l'écart entre la classe bandelette (variable Y) et la classe labo (variable X).

Z	-5	-4	-3	-2	-1	0	1	2	3	4	5	Total
O	0	1	5	9	89	157	40	7	0	0	0	308
E	1.4	9.7	25.3	41.5	57.6	64.1	52.4	34.2	17.5	3.9	0.3	308
		11.1	25.3	41.5	57.6	64.1	52.4	34.2	21.7			308

La ligne O correspond aux valeurs observées ("observed" (Tableau 3)) et la ligne E aux valeurs attendues ("expected" (Tableau 4)).

La dernière ligne regroupe les classes extrêmes de façon à ce que les effectifs théoriques soient tous supérieurs à 5 (test du χ^2).

troisième lignes de ce tableau présentent respectivement les effectifs observés et les effectifs attendus sous l'hypothèse d'indépendance des variables X et Y . La dernière ligne du Tableau 5 est obtenue en regroupant les classes extrêmes toujours pour éviter des effectifs attendus inférieurs à 5.

A partir du Tableau 5, on trouve $\chi^2 = 249.0$. Avec $\nu = 7$ degrés de liberté, la p -value ou probabilité de dépassement d'une telle valeur si l'hypothèse d'indépendance est vraie est quasi nulle. Ces résultats montrent donc une dépendance significative entre les observations sur sites et leur contrôle au laboratoire.

Mesure de l'intensité de la liaison entre les variables X et Y

Il existe plusieurs mesures de l'association sur tableau de contingence (Agresti, 1990). Ces mesures sont souvent difficiles à interpréter. Un paramètre commode et d'interprétation immédiate est la proportion π des sources qui seraient correctement classées ($Z = 0$) par les deux méthodes si on avait accès à la population définie par l'ensemble des sources et puits du territoire belge, c'est à dire:

$$\pi = P(Z = 0) \quad (5)$$

On note que du point de vue des incertitudes d'échantillonnage, sur n contrôles réalisés, le nombre de sources A pour lequel le titrage au laboratoire confirme le test *in situ* (i.e. $Z = 0$) est une variable aléatoire distribuée selon la loi binomiale de paramètres π , n :

$$p_m = P(A = m) = C_n^m \pi^m (1 - \pi)^{n-m} \quad (6)$$

L'espérance et la variance de A sont respectivement données par $n\pi$ et $n\pi(1 - \pi)$. Dès lors, la statistique $R = A/n$ est un estimateur non biaisé de la proportion π car $E(R) = \pi$. De plus, son efficacité augmente avec n car $\text{Var}(R) = \pi(1 - \pi)/n$.

Conformément à une convention d'écriture bien établie, une estimation de la proportion théorique π est une valeur particulière, disons r (en minuscule), de la statistique R (en majuscule). En divisant les effectifs du Tableau 5 par la taille de l'échantillon ($n = 308$) on trouve une distribution empirique de la statistique R et une distribution virtuelle de R correspondant à l'hypothèse d'indépendance entre les mesures colorimétriques et la réalité.

La proportion $r_O = 0.51$ est la proportion empirique tandis que la proportion $r_E = 0.21$ correspond à l'hypothèse d'indépendance (Tableau 6).

Toutes les valeurs possibles r de R forment la distribution d'échantillonnage de R . L'approximation normale dit que cette distribution d'échantillonnage varie autour de π selon une loi approximativement Normale, d'espérance μ et d'écart-type σ :

Tableau 6 Estimation de la proportion π pour les sources: correctement classées par le test colorimétrique et par le labo ($Z = 0$); mal classées, le test colorimétrique sous-estimant la concentration réelle ($Z < 0$); mal classées, le même test surestimant la réalité ($Z > 0$). L'estimation r_O concerne les observations tandis que l'estimation r_E concerne les valeurs attendues sous l'hypothèse d'indépendance.

Z	< 0	0	> 0
r_O	0.34	0.51	0.15
r_E	0.44	0.21	0.35

$$R \cong N(\mu, \sigma) \Rightarrow \begin{cases} \mu = \pi \\ \sigma = \sqrt{\pi(1-\pi)/n} \end{cases} \quad (7)$$

Cette approximation est donc d'autant plus robuste que n est grand. En pratique, on ne connaît évidemment pas l'écart-type σ puisque son expression fait intervenir le paramètre inconnu π . C'est pourquoi on le remplace par l'erreur standard, notée SE , obtenue en remplaçant π par son estimation empirique r :

$$SE = \sqrt{r(1-r)/n} \quad (8)$$

Sur cette base on peut construire des intervalles de confiance. En effet la distribution Normale permet de calculer deux limites de tolérance r_1 et r_2 (symétriques en probabilité) pour les variations aléatoires de la statistique R avec un risque α fixé:

$$P(r_1 \leq R \leq r_2) = 1 - \alpha \quad (9)$$

Ces limites sont:

$$r_1 = r - u_{\alpha/2} SE \quad \text{et} \quad r_2 = r + u_{\alpha/2} SE \quad (10)$$

où $u_{\alpha/2}$ est la valeur normale centrée réduite dont la probabilité de dépassement est égale à $\alpha/2$ et où SE est l'erreur standard (8).

A partir des relations (10) on construit un intervalle de confiance à $1 - \alpha\%$ (en l'occurrence 90%) pour la proportion empirique r . Par le calcul on obtient les valeurs particulières $r_1 = 0.46$ et $r_2 = 0.56$. La valeur d'indépendance $r_E = 0.21$ étant très nettement en dehors de cet intervalle, on a une autre présentation de la très haute signification de l'association entre les concentrations observées sur le terrain et les contrôles en laboratoire.

Nous avons présenté en détail la technique statistique d'analyse sur les données du Tableau 3 car nous allons l'utiliser systématiquement pour la suite.

Analyse pour les tests colorimétriques correctement codés

On rappelle que parmi les 308 tests colorimétriques contrôlés, seulement 120 ont été codés conformément aux instructions données, i.e. affectés à une classe de concentrations. Pour ces données il n'y a pas lieu de créer une classe zéro. Toutefois pour comparer cette analyse avec celle réalisée sur les données du Tableau 3, nous avons gardé les 31 valeurs nulles dans la classe zéro. Le seul titrage inférieur à 0.1 mg l^{-1} a donc été affecté à cette même classe (Tableau 7).

Tableau 7 Données correctement codées plus classe zéro. Les classes bandelettes sont en lignes (variable Y) et les classes labo en colonnes (variable X).

	x0	x1	x2	x3	x4	x5	Total
y0	5	22	2	2	0	0	31
y1	1	19	4	1	2	0	27
y2	0	3	5	6	0	0	14
y3	0	3	6	20	5	0	34
y4	0	0	2	10	19	2	33
y5	0	0	0	1	6	5	12
Total	6	47	19	40	32	7	151

En refaisant la même analyse que précédemment, le χ^2 calculé sur le tableau des écarts Z est égale à 134.8 avec 8 degrés de liberté donnant encore une fois une p -value quasiment nulle. Le test colorimétrique est donc significatif. Pour ce qui concerne la proportion π des points d'eau qui seraient correctement classés ($Z = 0$), son estimation empirique est $r_O = 0.48$ et l'intervalle de confiance 0.41–0.55 ($\alpha = 10\%$) ne contient pas la proportion correspondant à l'hypothèse d'indépendance $r_E = 0.19$.

Les conclusions apparaissent très semblables au cas précédent, l'association entre le test colorimétrique et le contrôle au labo est certainement significative. Changer le risque alpha, p. ex. prendre $\alpha = 0.05$ ou 0.01, ne modifie en rien les conclusions. En d'autres mots, la méthode d'analyse employée jusqu'ici est insensible au codage incorrect pratiqué par une majorité des élèves (61%). Pour tenter de mettre en évidence les différences qui doivent exister entre les Tableaux 3 ($n = 308$) et 7 ($n = 151$) nous pouvons comparer les distributions conditionnelles des tests colorimétriques à classe labo fixée:

$$P(Y = i | X = j) = \frac{P(Y = i \cap X = j)}{P(X = j)} \cong \frac{n_{ij}}{n_{\cdot j}} \quad (11)$$

Les résultats sont présentés dans les Tableaux 8 ($n = 308$) et 9 ($n = 151$). Nous reviendrons bientôt sur cette comparaison.

Tableau 8 A partir du Tableau 3 ($n = 308$), probabilité conditionnelle du test colorimétrique disposant du classement du labo sensé représenté la réalité.

	x0	x1	x2	x3	x4	x5
y0	0.71	0.35	0.03	0.02	0.00	0.00
y1	0.14	0.45	0.30	0.07	0.04	0.07
y2	0.14	0.15	0.46	0.26	0.00	0.00
y3	0.00	0.05	0.17	0.50	0.24	0.07
y4	0.00	0.00	0.03	0.14	0.63	0.50
y5	0.00	0.00	0.00	0.01	0.09	0.36

Tableau 9 A partir du Tableau 7 ($n = 151$), probabilité conditionnelle du test colorimétrique disposant du classement du labo sensé représenté la réalité.

	x0	x1	x2	x3	x4	x5
y0	0.83	0.47	0.11	0.05	0.00	0.00
y1	0.17	0.40	0.21	0.03	0.06	0.00
y2	0.00	0.06	0.26	0.15	0.00	0.00
y3	0.00	0.06	0.32	0.50	0.16	0.00
y4	0.00	0.00	0.11	0.25	0.59	0.29
y5	0.00	0.00	0.00	0.03	0.19	0.71

SENSIBILITÉ DE LA RELATION R À DES FACTEURS DE VARIATION SYSTÉMATIQUES

En vue de la solution des problèmes d'estimation de la "concentration réelle" en nitrates d'une source qui sera basée sur l'utilisation des distributions conditionnelles (Tableaux 8 ou 9), il importe de s'assurer que la relation $R = R(X, Y)$, résumée par ces distributions, n'est pas entachée de biais résultant de facteurs externes.

Outre la succession des campagnes dans le temps nous ne traiterons que les facteurs éventuels de biais résultant de certaines pratiques identifiées dans la collecte des données colorimétriques et des mesures au labo.

Homogénéité des laboratoires agréés

La base de données utilisée indique que trois laboratoires (WWF, UIA et la FUL) ont effectués des contrôles. Pour chacun d'entre eux, on calcule les répartitions correspondantes de la variable d'erreur Z (Tableau 10).

Tableau 10 Comparaison des écarts Z en fonction du laboratoire agréé ayant effectué des contrôles.

Z	-4	-3	-2	-1	0	1	2	Total
WWF	0	1	0	13	18	9	0	41
UIA	1	2	4	27	62	9	2	107
FUL	0	2	5	49	77	22	5	160
Total	1	5	9	89	157	40	7	308

L'hypothèse nulle est que le classement des sources selon le critère Z est indépendant du laboratoire agréé. Après regroupement des classes, le χ^2 de comparaison donne la valeur 5.1 pour 4 degrés de liberté, soit une p -value de 0.28. Cette valeur de probabilité est trop grande pour rejeter l'hypothèse nulle qui traduit l'homogénéité des laboratoires agréés du point de vue du titrage en nitrates.

En ce qui concerne le critère R , on trouve dans les trois cas:

$\alpha = 10\%$	Critère r	Inf. r_1	Sup. r_2
WWF	0.44	0.31	0.57
UIA	0.58	0.50	0.66
FUL	0.48	0.42	0.55

Comme les trois intervalles de confiance à 90% se recourent mutuellement, le résultat du test du χ^2 est confirmé par cette comparaison.

Pour les codages incorrects: lecture par excès?

On peut calculer les répartitions en nombres et en pourcentages des écarts $Z = Y - X$, tant pour les tests colorimétriques mal codés par les élèves (une valeur), mais en supposant que les valeurs données sont les limites supérieures des intervalles, que pour les tests correctement codés (affectation de la source visitée à un intervalle de concentrations en nitrate). Cette comparaison exige qu'on écarte les valeurs zéro puisque cette classe n'existe pas pour les résultats correctement codés.

L'hypothèse nulle stipule que le classement des sources selon le critère Z est indépendant du type de codage. Après regroupement des classes, le χ^2 de comparaison donne la valeur 21.0 avec 3 degrés de liberté. La p -value est 10^{-4} ce qui est indicatif d'une différence très significative entre les bons et les mauvais codages. Or ici la statistique R ne confirme pas ce résultat. Un examen plus attentif montre que ce sont

les sources mal classées qui expliquent le rejet de l'hypothèse nulle par le test du χ^2 . Une interprétation plausible de ces résultats est que les mauvais codages ont été donnés tantôt par excès, tantôt par défaut et ce avec des fréquences non négligeables mais malheureusement inconnues. Remarquons que l'hypothèse contraire, i.e. les mauvais codages sous-estiment la concentration réelle, n'améliorerait certainement pas les choses. En effet cette hypothèse correspondrait au décalage d'une classe de Z de la première ligne du Tableau 11. Bien entendu ces interprétations supposent qu'en dehors de l'imprécision du codage, les deux groupes d'élèves ne se distinguent pas par la qualité de lecture des bandelettes.

Tableau 11 Comparaison des écarts Z en fonction des tests colorimétriques mal codés (i.e. une valeur) et bien codés (i.e. un intervalle).

Z	-4	-3	-2	-1	0	1	2	Total
Mal	1	1	6	50	84	15	0	157
Bien	0	2	1	17	69	25	6	120
Total	1	3	7	67	153	40	6	277

Quoi qu'il en soit cette situation est très dommageable et on peut penser à n'utiliser que les tests colorimétriques correctement codés augmentées des valeurs 0 mais ce parti pris réduit singulièrement les données utilisables qui n'étaient déjà pas si nombreuses. Nous y reviendrons.

Etat des échantillons contrôlés au laboratoire

Il était possible d'étudier de façon assez sommaire une hétérogénéité éventuelle de la qualité des échantillons arrivés dans les laboratoires par des informations succinctes du type: renseignements insuffisants, dénitrification possible ou certaine, arrivée tardive au laboratoire, etc. pouvant affecter la représentativité de la mesure. Les données ont donc été classées en deux catégories, la première avec la mention "Douteux", sans précision de type, la seconde avec la mention "Conforme". Les résultats sont montrés dans le Tableau 12.

Tableau 12 Comparaison des écarts Z en fonction de la qualité de l'échantillon d'eau parvenu au laboratoire pour le contrôle du test colorimétrique.

Z	-4	-3	-2	-1	0	1	2	Total
Douteux	1	1	2	31	51	14	3	103
Conforme	0	4	7	58	106	26	4	205
Total	1	5	9	89	157	40	7	308

Les deux répartitions sont très proches l'une de l'autre. Le χ^2 de comparaison est égal à 0.61 avec 3 degrés de liberté. La p -value est 0.89 ce qui indique que rien dans ces données ne permet de rejeter l'hypothèse que les deux distributions sont identiques. Bien entendu il faut souligner qu'une mention quelconque n'est pas nécessairement un bon critère de qualité de l'observation mais il n'est pas possible d'être plus précis avec les données disponibles. Nous admettons donc que de ce point

de vue de qualité toutes les observations sont homogènes (jusqu'à plus ample information éventuellement contradictoire sur le plan de la signification statistique).

Répartition des campagnes dans le temps

La comparaison des distributions des écarts Z pour les 4 campagnes est donnée dans le Tableau 13.

Tableau 13 Comparaison des écarts Z en fonction des dates des quatre campagnes (M pour mars et O pour octobre).

Z	-4	-3	-2	-1	0	1	2	Total
M94	0	3	2	29	54	11	0	99
O94	0	1	2	34	43	10	2	92
M95	1	1	4	23	35	13	4	81
O95	0	0	1	3	25	6	1	36
Total	1	5	9	89	157	40	7	308

L'hypothèse nulle est que la variable Z est indépendante de la date. Cette hypothèse traduit l'homogénéité des campagnes dans le temps. Le χ^2 de comparaison est (après regroupement des classes de Z d'effectifs théoriques trop faibles) égal à 14.2 pour 6 degrés de liberté.

Avec un seuil de signification du test α égale à 5%, on est conduit à rejeter l'hypothèse nulle car $\chi^2_{6,0.05} = 12.59$ alors que si on choisit $\alpha = 1\%$ on doit l'accepter puisque $\chi^2_{6,0.01} = 16.81$. La p -value ou probabilité d'observer $\chi^2 = 14.2$ si l'hypothèse nulle est vraie est égale à 3% et tombe donc dans la zone des seuils usuels de signification à 1% ou 5%. On peut donc suspecter une hétérogénéité des campagnes qu'il semble raisonnable d'imputer à une hétérogénéité de l'attitude des élèves vis à vis du codage selon les campagnes. En effet le même test réalisé sur les données correctement codées (Tableau 7) ne permet pas de rejeter l'hypothèse d'homogénéité des campagnes. De campagne en campagne, le dossier pédagogique semble mieux suivi.

QUELQUES ÉLÉMENTS DE CONCLUSION SUR LES COMPARAISONS

On rappelle que la relation R exprime le lien entre la concentration en nitrates relevée *in situ* par le test colorimétrique et cette même concentration déterminée au laboratoire. Un lien fort et significatif confère aux cartes nitrates un caractère opérationnel. Par conséquent, dans la mesure où les résultats des tests précédents peuvent être considérés comme fiables, la conclusion d'ensemble est la suivante. Si on se limite aux données définies sans ambiguïté d'affectation aux classes, la relation R est stable au cours des campagnes et semble indépendante de diverses causes de variabilité éventuelles accessibles par l'information disponible.

Cependant le manque de précision sur le report des classes de concentration par les élèves est un défaut majeur qui nous oblige donc à n'utiliser que les seules données

fournies par intervalles et les valeurs 0. Ceci est très dommageable par la réduction de moitié des observations réellement utilisables.

De ce fait les trop petits échantillons statistiques traités par sous ensembles laissent planer un doute sur la qualité des tests. Nous voulons parler de la puissance statistique de ces tests qui apparaît limitée ici. Il s'agit du pouvoir de détection (en termes de probabilités) d'une hétérogénéité éventuelle. Une seule indication partielle à cet égard sera donnée ici. Le dernier test d'hétérogénéité des campagnes sur les 151 données par intervalle a nécessité des regroupements. Il en résulte une représentation très sommaire des distributions des Z et malgré cela le nombre d'observations est encore très réduit pour certaines campagnes. Une puissance limitée de ce test en découle.

Les commentaires ci dessus n'ont pas pour objet de trop relativiser l'ensemble des conclusions précédentes mais de montrer l'importance d'une analyse plus précise qui sera tentée ultérieurement.

Malgré tout la conclusion importante de ces comparaisons sera retenue pour la suite: la relation d'étalonnage R doit être estimée sur les seules données d'observations bandelettes par intervalles de concentration augmentées des données 0. Cette relation très hautement significative est représentée par le Tableau 9 des distributions conditionnelles (équation (11)).

On notera que cette relation R , ainsi estimée sur des fréquences brutes dans un échantillon statistique assez limité, présente des variations de "probabilités" qui manquent de continuité pour certaines classes de titrage-labo. C'est le tribut payé à l'absence d'un modèle quantitatif plus précis qui sera nécessaire ultérieurement.

La relation R est importante pour aborder une série de problèmes qui feront l'objet d'un article ultérieur. Ces problèmes sont tous liés à l'estimation de la concentration ou des concentrations de nitrates dans une source ou un ensemble de sources appartenant à une région donnée à partir des seuls tests colorimétriques.

- Comment estimer au mieux la concentration de nitrates réelle d'une source à partir de la seule observation d'une bandelette plongée dans cette source? Une question annexe en découle: peut-on utiliser l'information sur l'ensemble de la région à laquelle appartient la source pour affiner l'estimation locale?
- Comment mesurer le gain d'information apporté par cette seule observation et par la mobilisation de l'information régionale?
- Comment estimer, non pas une concentration individuelle, mais la répartition des concentrations réelles d'un ensemble de sources appartenant à une région homogène?

L'importance de ces problèmes de caractérisation régionale de la qualité des sources tient à l'utilisation de ces indicateurs pour le suivi de la "pollution régionale" au cours du temps. L'objectif final est en effet de répondre à la question suivante: quelle est l'évolution dans le temps de la qualité des sources en Belgique? Ces différents problèmes seront analysés dans un travail ultérieur.

REFERENCES

- Agresti, A.(1990) *Categorical Data Analysis*. Wiley, New York, USA.
 Wattecamps, J. M. & Michiels, L. (1995) *Un Regard Clair Sur L'eau Des Sources: Dossier pédagogique de l'Opération Sources* (deuxième edn). Fond Mondial pour la Nature, WWF Belgium, Bruxelles, Belgique.