

SIMULATION DE STRATEGIES D'ECHANTILLONNAGE

APPLICATION A LA LOIRE ET A LA NOE SECHE

HUBERT P. *, ADAMSKI M. *, MEYBECK M. **

*CENTRE D'INFORMATIQUE GEOLOGIQUE. ECOLE DES MINES DE PARIS
35 RUE ST HONORE 77305 FONTAINEBLEAU - FRANCE

**LABORATOIRE DE GEOLOGIE APPLIQUEE UNIVERSITE P. ET M. CURIE
4, PLACE JUSSIEU 75230 PARIS CEDEX 05 - FRANCE

1 - PROBLEMATIQUE : BILANS VRAIS, BILANS ESTIMES

Nous ne reviendrons pas ici sur l'intérêt d'établir des bilans quantitatifs des exportations d'eau, de matières dissoutes ou de matières en suspension à l'exutoire d'un bassin versant. Nous nous intéresserons uniquement à la définition, à l'estimation et à la précision de tels bilans.

L'estimation d'un bilan comporte d'une part un échantillonnage, d'autre part des méthodes de calcul appliquées aux mesures issues de cet échantillonnage. A une estimation, ou plutôt à une procédure d'estimation, il convient d'associer deux éléments :

- d'une part, ce que nous pourrions appeler un coût ou plus généralement une quantité d'efforts consentis pour réaliser le plan d'échantillonnage de mesure et de calcul.
- d'autre part, une qualité d'estimation liée au biais et à la précision de l'estimation du bilan, qui est une variable aléatoire, relativement au bilan vrai qui demeure fondamentalement inaccessible.

Ces deux aspects ne sont évidemment pas indépendants. Un effort limité ne peut conduire qu'à une médiocre estimation qui ne peut être améliorée que par un accroissement des efforts.

On se rend compte aujourd'hui que l'échantillonnage et les mesures coûtent cher. Il convient donc, lorsque les données sont déjà disponibles, d'en tirer a posteriori le meilleur parti possible. De même, lorsqu'il s'agit de définir, ex nihilo, une procédure permettant, avec des ressources limitées, d'estimer un bilan, il convient d'utiliser au mieux ces ressources. Dans les deux cas, se pose un problème d'estimation auquel nous tenterons ici d'apporter quelques réponses.

1.1 - Flux et bilans

Considérons une section de rivière. A un instant donné, il existe en chaque point de cette section une vitesse V et une concentration c . Le flux instantané de l'élément considéré peut s'exprimer comme :

$$F(t) = \int \int_S v(x,y,t) c(x,y,t) ds \quad (1)$$

Il s'agit d'une intégrale double étendue à la section mouillée S . Produit d'une vitesse par une concentration et une surface, ce flux s'exprime donc en unités de masse par unité de temps.

Le bilan de l'élément considéré entre les moments t_1 et t_2 s'exprimera donc comme :

$$B_{t_1}^{t_2} = \int_{t_1}^{t_2} F(t) dt \quad (2)$$

ou plus exactement :

$$B_{t_1}^{t_2} = \int_{t_1}^{t_2} \left[\iint_{S(t)} v(x,y,t) c(x,y,t) ds \right] dt \quad (3)$$

puisque, en raison des variations de débit la section mouillée évolue au cours du temps.

1.2 - Problème de l'hétérogénéité spatiale dans la section mouillée

On ne dispose évidemment jamais de l'enregistrement en continu de la vitesse et de la concentration en chaque point d'une section. Dans la section de mesure la vitesse est échantillonnée à l'occasion des jaugages qui permettent d'établir une courbe de tarage (relation hauteur-débit) de la section. C'est donc le débit qui sera disponible à partir de l'enregistrement en continu de la hauteur, ou à partir de la lecture périodique ou épisodique d'une échelle. Par contre on ne procède généralement pas à l'exploration du champ des concentrations. On admet, sans doute parfois abusivement, que le cours d'eau est homogène à un instant donné. On dispose exceptionnellement d'enregistrements en continu des concentrations en un point de la section. Plus fréquemment, on dispose de concentrations d'échantillons ponctuels dans le temps et dans l'espace. On dispose enfin, parfois, de concentrations d'échantillons composites réalisés à partir d'échantillons ponctuels. (Il faut remarquer que la réalisation de tels échantillons, proportionnellement au débit par exemple, est sous tendue par un mode d'estimation du bilan).

Dans le cas d'un enregistrement en continu du débit et d'une concentration, on estimera le bilan par :

$$\hat{B}_{t_1}^{t_2} = \int_{t_1}^{t_2} Q(t) C(t) dt \quad (4)$$

$$\text{où} \quad Q(t) = \bar{v}(t) \cdot S = \iint_S v(x,y,t) ds \quad (5)$$

apparaît comme le produit de la vitesse moyenne dans la section à l'instant t par la surface de la section.

C(t) est généralement la concentration en un point particulier x_0, y_0 mesurée à l'instant t.

$$C(t) = c(x_0, y_0, t) \quad (6)$$

qui est utilisée comme estimateur de la moyenne c(t) dans la section

$$\bar{c}(t) = \frac{1}{S} \iint_S c(x,y,t) ds \quad (7)$$

$$c(t) = \hat{c}(t) = c(x_0, y_0, t) \quad (8)$$

Faute d'exploration du champ des concentrations on ne sait généralement rien

de la qualité de cet estimateur. Des biais et des incertitudes considérables peuvent être introduits à ce niveau, fonctions de la structure, inconnue, du champ de concentration. Il faut remarquer, de plus, que généralement :

$$\int \int_S v(x,y,t) c(x,y,t) ds \neq \bar{v}(t) \bar{c}(t) S \quad (9)$$

Là encore, l'absence de connaissance sur la covariance des fonctions v et c nous interdit de porter un jugement sur la qualité de l'estimateur (équation 4) du bilan vrai (équation 3).

1.3 - "Bilan vrais" à partir de prélèvements ponctuels

Dans le cas de prélèvements ponctuels régulièrement répartis selon un pas de temps Δt , le bilan du pas de temps n_1 au pas de temps n_2 est estimé par :

$$\hat{B}_{n_1}^{n_2} = \Delta t \sum_{n=n_1}^{n=n_2} Q(n) C(n) \quad (10)$$

Les problèmes évoqués plus haut sont évidemment présents ici, avec en plus le fait que Q et C sont les estimations de moyennes spatio-temporelles du débit et de la concentration et non plus des estimations de moyennes spatiales. Il s'agit pourtant là de l'information la plus complète à laquelle nous puissions généralement prétendre. Lorsque l'intervalle de temps Δt est suffisamment petit pour qu'on puisse admettre que Q et C y varient peu, on admet que l'estimation (10) représente le bilan vrai, et c'est à elle que nous comparerons d'autres estimations, établies à partir de sondages.

Le protocole conduisant à de telles estimations comporte deux règles :

- une première règle pour tirer les échantillons,
- une seconde règle pour faire les estimations à partir des mesures réalisées sur les échantillons.

La seconde règle, relative au calcul, ne va pas de soi et nous voudrions montrer, sur un exemple, quelles peuvent être ses conséquences, en particulier le biais qu'elle peut introduire.

On considère une période de n intervalles de temps Δt consécutifs. Soient Q_i et c_i le débit et la concentration durant la période i où l'on suppose que Q et c restent constants. La durée pendant laquelle une concentration et un débit peuvent être considérés comme constants dépendra de l'hydrologie du cours d'eau et de la variable considérée. Sur les grands bassins (superficies $> 10000 \text{ km}^2$) et pour les variables non impliquées dans les cycles nyctéméraux (c'est-à-dire à l'exclusion de paramètres tels que O_2 ou pH) on peut considérer que $Q(t)$ et $c(t)$ sont constants sur une journée.

On peut alors définir la moyenne et l'écart-type du débit et de la concentration sur la période :

$$\bar{Q} = \frac{\sum_{i=1}^{i=n} Q_i}{n} \quad \bar{c} = \frac{\sum_{i=1}^{i=n} c_i}{n}$$

$$\sigma_Q^2 = \frac{\sum_{i=1}^{i=n} (Q_i - \bar{Q})^2}{n} \quad \sigma_C^2 = \frac{\sum_{i=1}^{i=n} (C_i - \bar{C})^2}{n}$$

La meilleure estimation que l'on puisse faire du "bilan vrai" de l'espèce dont on mesure la concentration est :

$$\hat{B} = \sum_{i=1}^{i=n} Q_i C_i$$

Q_i et C_i peuvent s'écrire comme la somme de leur moyenne et d'un terme d'écart à leur moyenne :

$$\begin{aligned} Q_i &= \bar{Q} + q_i \quad ; \quad C_i = \bar{C} + c_i \\ \hat{B} &= \sum_{i=1}^{i=n} Q_i C_i = \sum_{i=1}^{i=n} (\bar{Q} + q_i) (\bar{C} + c_i) \\ &= \sum_i (\bar{Q} \bar{C} + \bar{Q} c_i + q_i \bar{C} + q_i c_i) \\ &= n \bar{Q} \bar{C} + \bar{Q} \sum_{i=1}^{i=n} c_i + \bar{C} \sum_{i=1}^{i=n} q_i + \sum_{i=1}^{i=n} q_i c_i \\ &\quad \quad \quad = 0 \quad \quad \quad = 0 \\ &= n \left[\bar{Q} \bar{C} + \sum_{i=1}^{i=n} \frac{q_i c_i}{n} \right] \end{aligned}$$

$(\sum_{i=1}^{i=n} q_i c_i) / n$ n'est autre que la covariance de Q et C , proportionnelle au coefficient de corrélation linéaire r de Q et C

$$r = \frac{\text{cov}(Q, C)}{\sigma_Q \sigma_C}$$

\hat{B} peut donc s'écrire :

$$\hat{B} = n \left[\bar{Q} \bar{C} + \text{cov}(Q, C) \right]$$

ou
$$\hat{B} = n \left[\bar{Q} \bar{C} + r \sigma_Q \sigma_C \right]$$

1.4 - Bilans avec un échantillonnage épisodique

Admettons que nous possédions sur une période de n intervalles de temps Δt un débit moyen de chaque intervalle :

$$Q_i \quad i = 1, 2, \dots, n$$

et la concentration d'un intervalle particulier k. Quelle estimation peut-on alors faire du flux pendant cette période ?

a) Extrapolation de la concentration

Cette solution utilise "toute" l'information disponible.

On calcule le débit moyen sur n intervalles et on lui applique la concentration connue :

$$\hat{B}_1^k = n \bar{Q} C_k$$

L'espérance de cette estimation est la moyenne des estimations réalisées avec $k=1, k=2, \dots$, jusqu'à $k=n$:

$$\bar{B}_1 = \sum_{k=1}^{k=n} \frac{n \bar{Q} C_k}{n} = \bar{Q} \sum_{k=1}^{k=n} C_k = n \bar{Q} \bar{C}$$

estimation biaisée puisque la meilleure estimation possible est :

$$\hat{B} = n \left[\bar{Q} \bar{C} + r \sigma_Q \sigma_C \right]$$

L'estimateur \hat{B}_1^k surestime le bilan si r est négatif (effet de dilution avec le débit). Il le sous-estime si r est positif (effet de concentration avec le débit).

b) Extrapolation des flux

On calcule le flux de l'intervalle k et on l'étend à toute la période considérée :

$$\hat{B}_2^k = n Q_k C_k$$

L'espérance de cette estimation est :

$$\bar{B}_2 = \sum_{k=1}^{k=n} \frac{n Q_k C_k}{n} = \sum_{k=1}^{k=n} Q_k C_k$$

On retrouve l'expression de B. Cette seconde estimation n'est pas biaisée.

L'estimateur \hat{B}_2^k , non biaisé, est a priori meilleur que l'estimateur \hat{B}_1^k .

Si l'on compare l'ensemble échantillonnage plus calcul à un appareil de mesure, le biais est l'analogie d'une erreur systématique et fait référence à la justesse de l'appareil de mesure. Il est évidemment nécessaire de considérer également la précision de cet appareil de mesure, qui peut être appréciée par l'écart-type associé à chaque estimateur.

Pour l'estimateur \hat{B}_1^k il vient

$$\begin{aligned} \sigma_{\hat{B}_1}^2 &= \sum_{k=1}^{k=n} \frac{(\hat{B}_1^k - \bar{B}_1)^2}{n} \\ &= \sum_{k=1}^{k=n} \frac{(n \bar{Q} C_k - n \bar{Q} \bar{C})^2}{n} \\ &= n^2 \bar{Q}^2 \sum_{k=1}^{k=n} \frac{(C_k - \bar{C})^2}{n} \\ &= n^2 \bar{Q}^2 \sigma_C^2 \end{aligned}$$

L'expression de l'écart-type de l'estimation \hat{B}_2^k est plus complexe. Nous n'en donnerons qu'une expression approchée sans développer les calculs qui y conduisent

$$\sigma_{\hat{B}_2}^2 = n^2 \left[\bar{Q}^2 \sigma_C^2 + 2 \bar{Q} \bar{C} r \sigma_Q \sigma_C + \bar{C}^2 \sigma_Q^2 \right]$$

On peut remarquer que dans le cas où Q et C sont indépendants ($r=0$), la précision de l'estimateur \hat{B}_1^k est meilleure que celle de l'estimateur non biaisé \hat{B}_2^k , car alors $\sigma_{\hat{B}_1} < \sigma_{\hat{B}_2}$.

La méthode d'extrapolation des flux est en général plus juste mais d'une précision moindre que la méthode d'extrapolation de la concentration. La justesse de cette dernière dépend alors de la nature d'une corrélation concentration-débit. Si une telle corrélation existe on peut alors améliorer considérablement l'estimation des bilans.

c) Reconstitution des concentrations

Dans les intervalles de temps où les concentrations sont inconnues, on reconstitue les concentrations C_i à l'aide soit de corrélation concentration-débit $c=f(Q)$, ou même de variation dans le temps $c=f(t)$, pour chaque intervalle de temps Δt sur lequel le bilan est construit.

Les régressions concentrations-débits ont fait l'objet d'innombrables études. La plupart du temps les corrélations sont du type $C = aQ^b$ avec :

$1 < b < 3$ pour les MES et les formes particulières (P. mercure, DDT, etc..)

- $0 < b < 1$ pour les éléments provenant du lessivage des sols (Corg dissous, certaines formes de N et P).
- $-1 < b < 0$ pour les éléments majeurs provenant de l'altération, et pour les NO_2^- dans certains cas.

Dans d'autres cas des régressions polynomiales plus complexes ont été mises en évidence, impliquant une dilution pour les faibles débits puis un lessivage lors des très fortes crues.

De telles régressions nécessitent bien sûr une étude préliminaire poussée de la variabilité des concentrations. Dans le cas des nutriments il apparaît que de telles régressions ne peuvent qu'exceptionnellement être mises en évidence et utilisées. Ainsi, dans de petits bassins les régressions générales $C=f(Q)$ sont rares : soit les concentrations sont indépendantes des débits, soit elles sont organisées à l'échelle de chaque crue.

Lorsque les rivières sont très influencées par des lacs situés en amont, on peut caractériser des évolutions saisonnières des concentrations. Il est alors aisé de reconstituer les concentrations à chaque instant.

Le bilan approché s'écrit alors :

$$\hat{B}_3 = \sum_{i=1}^{i=n} Q_i C_i$$

Où C_i est la concentration reconstituée correspondant au débit Q_i . Le biais est alors minimisé.

1.5 - Echantillonnage aléatoire

Nous rappellerons ici quelques résultats classiques de la théorie de l'échantillonnage.

Echantillonnage aléatoire

Un échantillon de n individus prélevés dans une population de n individus fournit une estimation non biaisée de la moyenne de la population.

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{i=n} x_i$$

Dans le cas des bilans dans les rivières, les individus échantillonnés sont alors les flux de matière à un instant donné.

L'écart-type de cet estimateur est alors :

$$s_{\bar{x}} = \frac{s}{\sqrt{n}} \sqrt{1 - \rho}$$

S est une estimation de l'écart-type de la population

$$s^2 = \sum_{i=1}^{i=n} \frac{(x_i - \bar{x})^2}{n-1}$$

et $\rho = n/N$ est un facteur de correction tenant compte du fait que la population échantillonnée est finie. Si $n=N$ l'échantillon se confond avec la population et $s_{\bar{x}}$ s'annule.

Echantillonnage aléatoire stratifié

La population étudiée est divisée en strates repérées plus bas par l'indice h. Dans le cas des bilans dans les rivières, l'échantillonnage peut être stratifié de plusieurs manières : par classe de débit, par saison, par classe de MES, etc...

n_h individus sont prélevés dans la strate h, et la moyenne de la strate est estimée par :

$$\bar{x}_h = \sum_{i=1}^{i=n_h} \frac{x_i}{n_h}$$

L'estimation de la moyenne de la population est alors :

$$\bar{x} = \sum_h w_h \bar{x}_h$$

où w_h , rapport de l'effectif réel n_h de la strate h à l'effectif total n de la population, représente le poids de la strate h.

Cette estimation est non biaisée et son écart-type s'exprime comme :

$$s_{\bar{x}} = \left[\sum_h w_h^2 \frac{s_h^2}{n_h} (1 - \rho_h) \right]^{-1/2}$$

s_h^2 est la variance estimée de la strate h

$$s_h^2 = \sum_{i=1}^{i=n_h} \frac{(x_i - \bar{x}_h)^2}{n_h - 1}$$

β_h est analogue, pour la strate h, au β défini plus haut pour la population.

2 - APPLICATIONS A LA LOIRE ET A LA NOÉ SÈCHE

Sur deux bassins de tailles très différentes, la Loire et la Noé Sèche, pour lesquels on disposait de données très nombreuses permettant de réaliser des "bilans vrais", nous avons testé différents protocoles de surveillance : échantillonnage aléatoire annuel avec 12, 24, 36 et 48 prélèvements par an, échantillonnage aléatoire mensuel avec 12, 34, 36 et 48 prélèvements par an, et échantillonnage régulier (avec le même effort de surveillance) qui est le plus classiquement employé.

Pour chaque protocole de très nombreuses simulations de bilans (100 dans la mesure du possible) ont permis d'estimer la précision des bilans effectués avec un échantillonnage limité par rapport au bilan vrai pour quatre variables de base : nitrates, ammoniac, Ptotal sur eau brute, orthophosphates.

Ces travaux ont été réalisés dans le cadre du GRECO-Lacs du CNRS (GRECO-Lacs, 1985 ; 1987) et ont en particulier fait l'objet de travaux de mémoire (Adamski, 1985 ; Gondo, 1986).

2.1 - La Noé Sèche

Le bassin versant de la Noé Sèche se situe dans les Cotes du Nord à trois kilomètres au Nord-Ouest de la commune de Quintin. La Noé Sèche est un affluent du Gouët. La superficie du bassin versant est évaluée à 600 hectares. Ce bassin versant a été étudié par le SRAE Bretagne (1984), sur un cycle annuel du 3 mars 1983 au 29 février 1984. Un préleveur automatique délivrant un échantillon moyen sur trois jours sur lequel sont mesurées les concentrations (P-PTOT, P-PO₄, N-NO₃, N-NH₄), alors que les débits sont connus au jour le jour.

2.2 - La Loire à Orléans

Il s'agit là d'une banque de données communiquée par l'Agence de Bassin Loire-Bretagne (Monsieur P. Cruzet). Les données initiales ont été prises quotidiennement sur un échantillon instantané intégré sur quelques heures : elles s'échelonnent du 6 janvier 1980 au 4 mars 1984 et concernent le débit, les formes du phosphore (P-PTOT, P-PO₄) et les formes de l'azote (N-NO₃, N-NH₄), mais présentent quelques lacunes.

Nous avons construit à partir du fichier initial un nouveau fichier de valeurs moyennes journalières du 1 novembre 1980 au 31 octobre 1984 (4 années) en procédant de la façon suivante :

- si plusieurs observations de concentration sont disponibles en jour donné, on retient la moyenne de ces observations.
- si un débit journalier manque, il est estimé par interpolation linéaire entre les deux observations les plus proches qui l'encadrent (figure 1a).
- si une concentration journalière manque, elle est estimée par interpolation linéaire à partir d'une relation locale entre concentration et débit, en utilisant les deux concentrations les plus proches qui l'encadrent (figure 1b).

2.3 - Les protocoles d'échantillonnage

a) Echantillonnage aléatoire annuel (EAA)

Ce protocole consiste à réaliser un bilan annuel à partir de N échantillons

prélevés aléatoirement au cours de l'année. On localisera donc par tirage au sort N journées dans une année et on estimera alors le bilan par :

$$\hat{B} = \frac{365}{N} \sum_{n=1}^{n=N} Q_n C_n$$

où Q_n et C_n sont respectivement le débit moyen et la concentration moyenne du $n^{\text{ième}}$ jour tiré au sort. Ce protocole a été simulé pour 12, 24, 36 et 48 échantillons par an.

b) Echantillonnage aléatoire mensuel (stratifié) (E. A. M)

Ce protocole est analogue au précédent mais au niveau mensuel. Il s'agit donc, par rapport au précédent protocole, d'un échantillonnage stratifié, les différents strates étant les douze mois de l'année. Nous avons simulé ce protocole en tirant au sort k jours par mois ($k = 1, 2, 3, 4$) durant une année. Le protocole conduit donc à estimer le bilan annuel avec 12, 24, 36 et 48 prélèvements, ce qui permet de le comparer, à effort d'échantillonnage égal, au protocole précédent.

c) Echantillonnage régulier (pour la Loire seulement) (E. R)

Là encore nous avons étudié un protocole utilisant $12k$ prélèvements par an ($k = 1, 2, 3, 4$). k prélèvements sont effectués chaque mois à des dates régulièrement réparties (par exemple les 2, 12 et 22 de chaque mois pour $k = 3$).

Les protocoles EAA et EAM, fondés sur des tirages au sort, ont été appliqués aux données de la Loire en 1981 et aux données de la Noé Sèche. Chaque protocole, pour chaque bassin et pour chaque effort d'échantillonnage a été appliqué 100 fois, fournissant à chaque fois une estimation du bilan auquel nous pouvons associer une erreur d'estimation puisque nous connaissons le "bilan vrai". On a calculé la moyenne et l'écart-type de cet échantillon d'erreurs d'estimation pour caractériser empiriquement, dans les deux bassins considérés, le biais et la précision du protocole auquel il est associé.

Le protocole ER n'a été appliqué qu'aux données de la Loire pour l'année 1981. Compte tenu de sa nature nous n'avons pas pu l'appliquer 100 fois mais seulement 30 fois pour $k = 1$ (en prélevant le 1er de chaque mois, puis le 2, etc...), 14 fois pour $k = 2$ (en prélevant le 1er et le 15 de chaque mois, puis le 2 et le 16, etc...), 10 fois pour $k = 3$ et 7 fois pour $k = 4$. Ces tirages limités en nombre par rapport aux protocoles précédents sont évidemment soumis à des fluctuations d'échantillonnage importantes.

Il faut d'abord noter que ces différents protocoles fournissent des estimations non biaisées des bilans (la moyenne des bilans estimés n'est pas significativement différente du bilan vrai).

En ce qui concerne la précision des bilans nous avons reporté les résultats bruts sur le tableau 1. (Il s'agit là pour chaque protocole, pour chaque élément et chaque bassin de l'écart type de l'erreur d'estimation exprimé en %). Ces mêmes résultats sont représentés graphiquement sur la figure 2 pour les échantillonnages de la Noé Sèche, sur la figure 3 pour les échantillonnages de la Loire. Sur chaque figure un diagramme particulier est consacré à chaque forme étudiée (P-Ptot, P-PO₄, N-NO₃, N-NH₄).

Les ensembles de points relatifs à chaque protocole (pour chaque station et

chaque élément) ont été reportés sur des diagrammes log-log. Ils sont sensiblement organisés selon des droites (à l'exception du protocole E. R. sur la Loire pour lequel nous disposons d'un nombre restreint d'échantillonnage et pour lequel on pouvait donc s'attendre à des fluctuations importantes). Sur les figures les points empiriques sont reliés par des pointillés et on a tracé en trait plein une droite d'ajustement sur chaque diagramme.

Les pentes de ces droites sont voisines de -0.5 pour la Loire et de -0.7 pour la Noé Sèche. La valeur de -0.5 montre que, en accord avec les résultats théoriques exposés en 1.5, la précision d'un échantillonnage aléatoire varie comme $1/\sqrt{n}$, n étant l'effectif de l'échantillon. En ce qui concerne la Noé Sèche, la valeur de -0.7 indique que la précision s'améliore plus rapidement avec l'effectif de l'échantillon que ne le laisse supposer la théorie.

Nous avons également pu comparer le niveau général des différents protocoles. Dans tous les cas l'échantillonnage aléatoire mensuel (stratifié) donne de meilleurs résultats que l'échantillonnage aléatoire annuel. De même, l'échantillonnage régulier est toujours supérieur à l'échantillonnage aléatoire mensuel.

De façon générale, la meilleure précision est obtenue pour les nitrates qui s'étaient révélés les plus réguliers à travers leurs corrélogrammes. Logiquement, les bilans en ammoniacque sont aussi les moins précis alors que cet élément s'était révélé le moins régulier au niveau des corrélogrammes (GRECO-Lacs, 1987).

Nous avons comparé les précisions obtenues empiriquement à des précisions théoriques, basées sur la théorie de l'échantillonnage aléatoire et utilisant les valeurs numériques correspondant aux flux d'eau (moyenne et écart-type par an ou par mois selon le type d'échantillonnage). Aussi bien pour la Noé Sèche que pour la Loire la précision des bilans en $P-PO_4$, $P-Ptot$ et $N-NH_4$ est plus faible que la précision calculée à partir des données de flux d'eau. Pour les bilans de $N-NH_3$, ces précisions sont comparables. Les écarts observés s'expliquent de deux façons. D'une part, nous avons négligé au niveau de la théorie les effets de la persistance qui sont pénalisants pour la précision des bilans. D'autre part, nous nous sommes basés sur les flux d'eau, dont le régime peut être différent des flux des formes du phosphore et de l'azote.

On peut s'apercevoir, à l'issue de cette étude, que nous avons négligé l'étude des flux, ce qui est paradoxal pour une étude concernant les bilans. C'est hélas une habitude bien établie de ne considérer comme objet d'étude que les débits (qui sont bien des flux d'eau) et les concentrations et de s'intéresser aux relations débit concentration, toujours décevantes lorsque l'échelle de temps de l'échantillonnage dépasse celle de l'événement hydrologique. A l'avenir, c'est sur une étude approfondie de nouveaux objets, les séries chronologiques de flux, qu'il faudrait s'appuyer pour interpréter les données existantes et planifier les campagnes d'acquisition de données nouvelles, en liaison avec les moyens disponibles et les objectifs de précision souhaités.

BIBLIOGRAPHIE

ADAMSKI, M. (1985) : Traitement de données hydrologiques et échantillonnage, mémoire de DEA, Université P. et M. Curie et Ecole des Mines de Paris, 120 p.

GONDO, J. (1986) : Etude statistique des données de qualité dans la Loire à Orléans, mémoire de DEA, Université P. et M. Curie et Ecole des Mines de Paris, 87 p.

GRECO-Lacs (1985) : Méthodologie de l'établissement des bilans d'apports en éléments nutritifs aux systèmes lacustres, rapport d'activité 1984-1985, M. Meybeck coordinateur, Ecole Normale Supérieure, Paris, 49 p.

GRECO-Lacs (1987) : Méthodologie de l'établissement des bilans d'apports en éléments nutritifs aux systèmes lacustres, rapport d'activité n°2 1985-1986, M. Meybeck coordinateur, IBM-ENS, Montrouge, 129 p.

SRAE Bretagne (1984) : Qualité des eaux en domaine rural, bilan des études sur la Noé Sèche (22), SRAE-Bretagne, Rennes.

Tableau 1 : Loire à Orléans et Noé Sèche. Précision (1) des bilans empiriques de nutriments en %, en fonction de l'effectif d'échantillonnage (12, 24, 36 et 48 éch./an)

	12/an		24/an		36/an		48/an	
	Loire	Noé	Loire	Noé	Loire	Noé	Loire	Noé
<u>P-total</u>								
E.A.A.	30.9	31.2	22.6	18.6	15.5	13.1	14.1	11.4
E.A.M.	22.0	17.9	14.1	10.5	12.6	9.3	11.3	6.9
E.R.	15.2 (30)	/	12.2 (14)	/	5.6 (10)	/	4.8 (7)	/
<u>P-PO₄</u>								
E.A.A.	25.3	31.4	21.0	20.2	16.1	12.4	11.3	12.0
E.A.M.	18.3	16.9	14.4	11.1	10.1	9.3	8.8	7.0
E.R.	13.0 (30)	/	10.7 (14)	/	2.4 (10)	/	2.9 (7)	/
<u>N-NO₃</u>								
E.A.A.	24.7	21.6	19.4	11.8	13.0	9.5	10.6	8.3
E.A.M.	14.3	9.9	10.7	6.0	8.2	4.2	7.6	3.3
E.R.	11.3 (30)	/	5.4 (14)	/	1.6 (10)	/	1.3 (7)	/
<u>N-NH₄</u>								
E.A.A.	28.9	37.6	19.5	22.2	18.4	20.1	15.7	14.3
E.A.M.	19.1	24.6	14.2	18.1	11.1	15.1	9.2	12.8
E.R.	16.2 (30)	/	11.7 (14)	/	7.8 (10)	/	2.9 (7)	/

(1) la précision est ici appréciée par l'écart-type de l'erreur relative des bilans "tirés" par rapport au bilan "vrai" suivant trois protocoles :

E.A.A. Echantillonnage aléatoire annuel

E.A.M. Echantillonnage aléatoire mensuel

E.R. Echantillonnage régulier (ex. 24 éch. le 3 et le 17 de chaque mois)

o à partir de 100 tirages sauf indication contraire (x)

Légende des figures

Figure 1 : Reconstitution des débits (Q_{t+1}) (a) et des concentrations (C_{t+1}) (b) dans une rivière par extrapolation linéaire entre 2 valeurs (C_t , Q_t ; C_{t+2} , Q_{t+2}) aux temps t et $t+2$. (Cas de la Loire à Orléans).

Figure 2 : Noé Sèche. Précision (écart-type de l'erreur relative en %, par rapport au flux exact) des bilans empiriques des flux de nutriments en fonction de l'effectif de l'échantillonnage pour 12, 24, 36 et 48 échantillons/an. EAA : échantillonnage aléatoire annuel ; EAM : échantillonnage aléatoire mensuel.

Figure 3 : Loire à Orléans. Précision (écart-type de l'erreur relative en %, par rapport au flux exact) des bilans empiriques des flux de nutriments en fonction de l'effectif de l'échantillonnage pour 12, 24, 36 et 48 échantillons/an. EAA échantillonnage aléatoire annuel ; EAM : échantillonnage aléatoire mensuel ; ER : échantillonnage régulier.

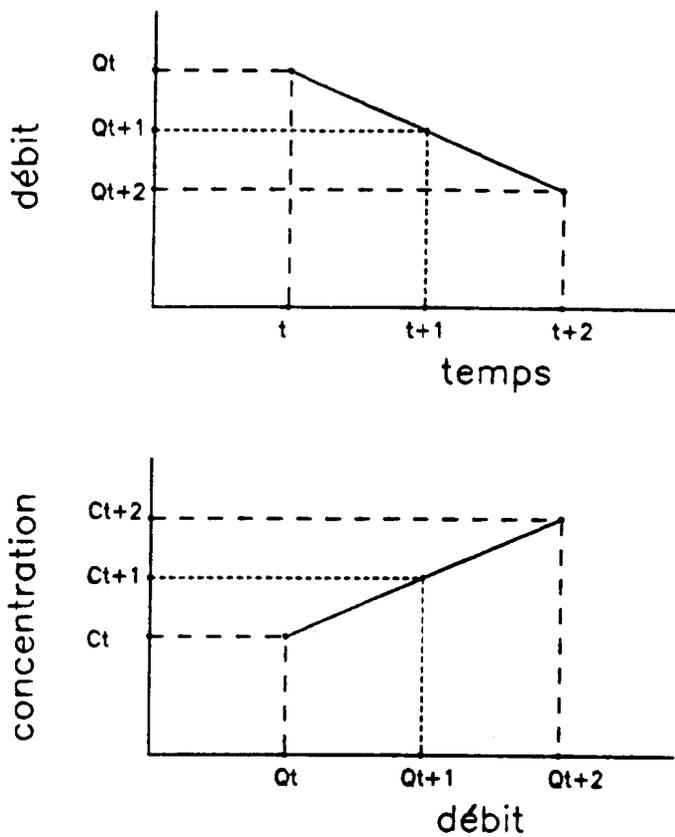


Fig. 1

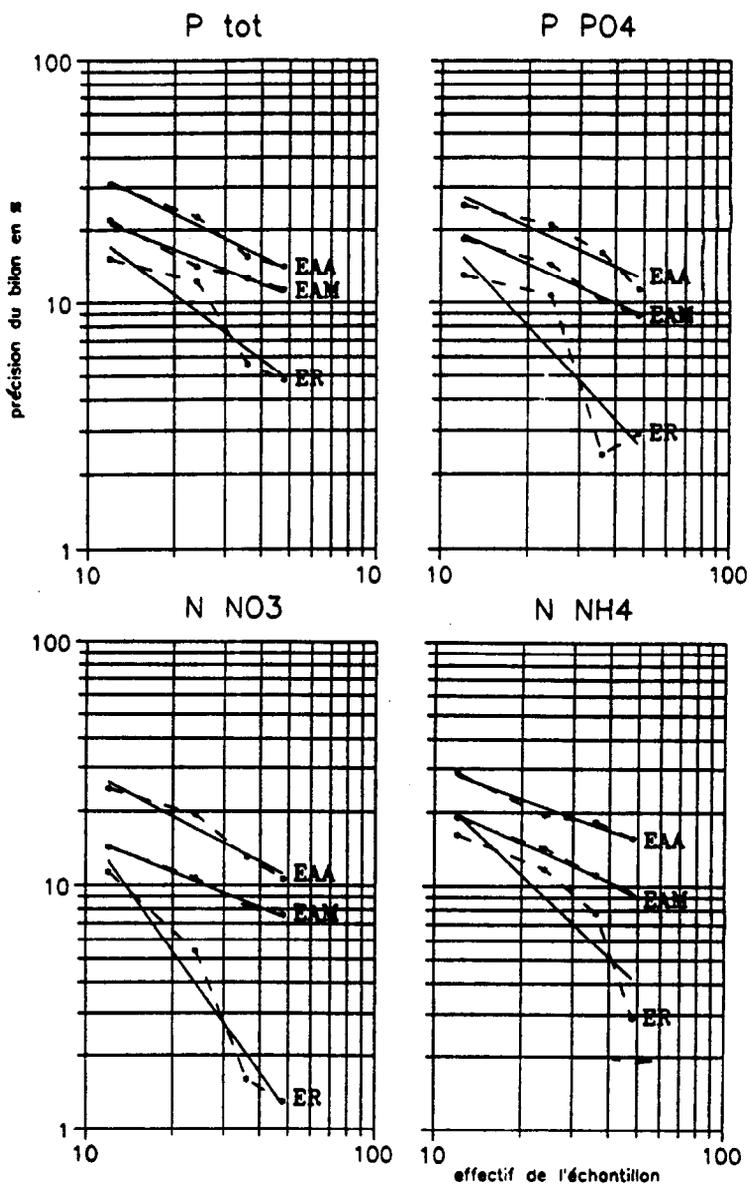


Fig. 2

NOE SECHE

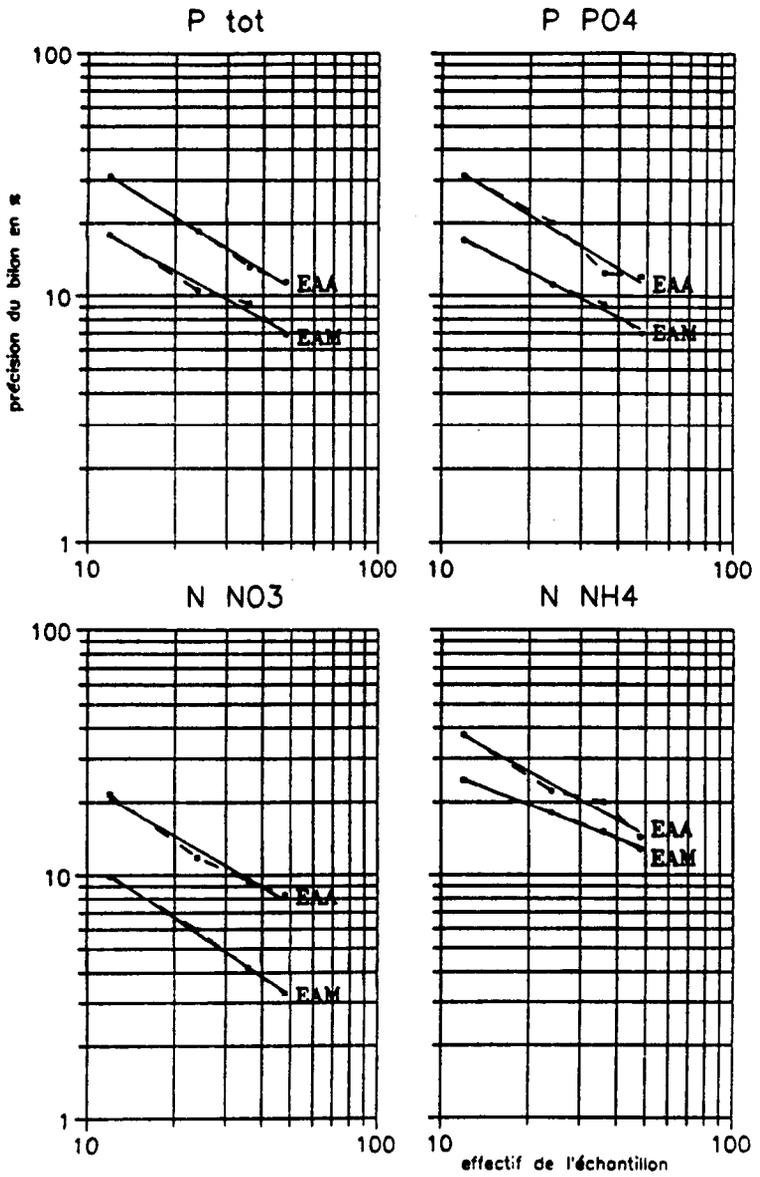


Fig. 3

LOIRE

DISCUSSION

J. COLOMBANI regrette la présence de lacunes dans l'échantillon des débits, lacunes qui ont été comblées dans le fichier de travail. Par interpolation linéaire entre les débits avant et après les lacunes comme cela a été fait, on ne peut pas s'attendre à une grande précision dans l'estimation des volumes écoulés.

Un intervenant précise que, par exemple, sur la Noé sèche, les études n'ont pas été lancées pour mener un programme scientifique rigoureux mais sous la pression des problèmes qui se posaient en termes aigus : eutrophisation, rejets agricoles, nitrification... Ce n'est que progressivement que les experts sont arrivés à la notion de bilan physico-chimique et donc qu'ils ont eu besoin d'un bilan hydrologique complet. Aujourd'hui, la Noé sèche, par la qualité de ses données physico-chimiques, reste une exception en France. Les scientifiques ont réussi à connaître tous les apports en nitrates des agriculteurs bien qu'ils ne purent connaître leur lieu précis d'épandage sur les champs. Cette finesse de la connaissance va de pair avec la petitesse du bassin - environ 600 ha - où tout phénomène accidentel peut avoir un poids important dans le bilan du bassin. A l'inverse, la taille du bassin de la Loire a un effet écrêteur incontournable. La stratégie d'échantillonnage ne peut être la même sur un petit ou un grand cours d'eau. Il y a, à côté du problème précédent, un autre problème de stratégie d'échantillonnage. Les scientifiques français qui se sont penchés sur la pollution des eaux se sont d'abord intéressés surtout aux nitrates, dans l'Ouest agricole. Par là même, ils ont quelque peu négligé les phosphates provenant des lessives qui contribuent à l'eutrophisation des cours d'eau.

G. CAUWET souligne l'intérêt pour la modélisation d'un travail comme celui qui a été présenté par M. ADAMSKI. Il rappelle qu'il a mené des études sur le carbone organique dissous immédiatement en amont de l'estuaire de la Loire, c'est-à-dire bien à l'aval de la région d'Orléans où a travaillé M. ADAMSKI.

M. ROCHE pense que le fait que les travaux présentés par M. ADAMSKI soient basés sur des données qui présentent de nombreuses lacunes est surtout gênant sur un petit bassin versant aux apports irréguliers, avec une fâcheuse incidence sur le calcul de la variance. De façon expérimentale, rappelle M. ROCHE, dans une étude déjà ancienne, il a essayé de tenir compte de la dispersion d'un échantillon lacunaire de la manière suivante : il divisa son échantillon en un certain nombre de classes et il créa autant de réservoirs de concentration qu'il y avait de débits. Il suffit pour un débit donné de voir à quelle classe de concentration il appartient. Pour affiner, il est possible d'inclure une variation saisonnière à cette relation concentration-débit.