ARTICLES

La base de données « Charges critiques » : Gestion et acquisition des données

The critical load database : Management and acquisition of data

Anne PROBST*, Jean-Paul PARTY**, Olivier ERTZ***

NDLR : Article issu du Séminaire ADEME « Pollution atmosphérique et charges critiques : Bilan et perspectives des recherches menées en France », 15 mai 2001, Paris (MATE).

Résumé

Dans le but d'assurer une meilleure gestion et une meilleure diffusion des données concernant les charges critiques acides et azotées pour le territoire français, après une étude de faisabilité, une première application de système de gestion de base de données a été proposée. Dans un deuxième temps, une phase d'évolution de la base (BDCC) a été menée à bien. Son développement a été assuré à partir de la première structure dite de niveau 1 à l'aide du SGBDR Oracle 7.3 pour Windows NT. D'autre part, outre l'extension de la base, un développement déterminant a permis le couplage entre la base de données BDCC et un système d'information géographique (Map Info 4.5).

Dans ce travail, les possibilités de développement d'Oracle ont permis d'apporter des modifications à la structure définie initialement et d'étendre la possibilité de la base par ajout de nouvelles tables d'une part, et par ajout de nouvelles fonctions statistiques indispensables à l'exploitation des données de la base, d'autre part. L'évolution de la base permet en outre de pouvoir importer et exporter par une procédure conviviale les fichiers sous format Excel. Une mise en sécurité des données de la base a été réalisée. Dans chaque cas, il s'agissait de rendre plus efficace le traitement, la gestion et l'acquisition des nouvelles données au sein du système. Ainsi, contrainte par les besoins au niveau européen et par le développement informatique de la base, la nouvelle structuration des tables et ses avantages dans la gestion des données sont présentées. Enfin, le développement du couplage entre la base de données et le SIG est illustré par un exemple d'application de liaison entre les données de la base et des cartes existantes de valeurs seuils importées à partir d'un autre SIG (Arc-Info). Les possibilités de croisement de données existantes et d'importation de cartes sont nombreuses.

Abstract

A first application of a data base management system was improved in order to ensure a better management of the critical loads data for France. The development was ensured starting from the first structure known as of level 1 using the RDBMS Oracle 7.3 for Windows NT. In addition to the extension of the base, the link between the data base BDCC and a GIS (Map Info 4.5) has been performed. In this work, the possibilities of Oracle allowed to modify the initially defined structure and to extend the base by addition of new tables on one hand, and by addition of new statistical functions essential to the exploitation of the data of the base, on the other hand. The evolution of the database allows to import and export by a user-friendly procedure the files under Excel format. A setting in data security of the base was carried out. In each case, it was a question of making the treatment, the management and the acquisition of the new data within the system more effective. Thus, on the basis of the requirements at the European level and on the constraints for the data-processing development of the database, the new structuring of the tables and its advantages in the data management are presented. Finally, the link between the data base and the GIS Map-Info was performed. An example of the link between the database and maps of threshold values imported from another GIS (Arc-Info), is shown. The crossing possibilities of existing data and importation of maps are thus numerous.

* *Laboratoire des Mécanismes de Transferts en Géologie, UMR 5563 CNRS-UPS, 38, rue des 36 Ponts, 31400 Toulouse.*
** *SOL-CONSEIL, 2, rue de Roppenheim, 67000 Strasbourg.*
*** *Stagiaire DESS Informatique. Centre de Géochimie de la Surface, 1, rue Blessig, 67084 Strasbourg Cedex.*

Contexte et objectifs du projet BDCC

Dans le domaine de la pollution de l'air, les objectifs de réduction des émissions de polluants ont été négociés au sein de la Convention de Genève sur la pollution transfrontalière à longue distance (UN/ECE – Convention on Long-Range Transboundary Air Pollution). Ces objectifs requièrent une base scientifique afin d'apprécier de la façon la plus objective possible les risques et les impacts sur les écosystèmes. Pour répondre à ce besoin, l'ADEME (Agence de l'environnement et de la maîtrise de l'énergie) a mis en place et animé les travaux français concernant les charges et les niveaux critiques depuis le début des années 90.

Ainsi, depuis 1992 une étude détaillée des facteurs de l'environnement a permis de définir des zones sensibles aux apports atmosphériques acides sur le territoire français. Les premiers essais de calcul et de cartographie des charges critiques d'acidité ont été menés à partir de 1994, successivement :

- pour les sols des Vosges [1-3] ;
- pour les eaux de surface des Vosges [4], puis des Ardennes [5] et enfin de certaines régions du Massif central [3, 6, 7] ;
- enfin, deux séries de données pour l'ensemble de la France ont été fournies en 1996 et 1998 au RIVM, centre coordonnateur de la cartographie des charges critiques pour l'Europe [8-10].

Pour pallier la déficience de la France pour la mise à jour des données au niveau européen, une base de données environnementale orientée « charges critiques » (BDCC) a été initiée. Plusieurs étapes successives à ce projet ont été proposées :

- une étude préalable de faisabilité a permis de déterminer les objectifs, la stratégie, l'organisation des informations et les outils à retenir [11] ;
- une première partie de structure de base de données avec pour objectif une réponse plus efficace aux demandes du RIVM a été réalisée [8] ;
- une seconde partie de structure de la base indispensable à la première a été réalisée [12]. Il s'agissait de faire évoluer la structure proposée dans le premier niveau en créant de nouvelles tables prêtes à accepter des nouvelles données de base et de construire un lien dynamique avec un SIG.

Nous présentons ici la conception de la base de données et son évolution au cours du temps jusqu'à son état actuel en liaison avec un SIG. Les résultats obtenus pour la France sur les charges critiques acides et azotées sont présentés dans Party et al. [13, 14].

Objectifs de la conception de la BDCC

La conception de la BDCC consistait en premier lieu en la mise en place d'une structure susceptible, d'une part, de stocker les données de base nécessaires et, d'autre part, de permettre un calcul automatisé et transparent des charges critiques pour les sols, les eaux de surface et la végétation des milieux naturels ainsi que l'édition informatique des résultats finaux requis notamment par le RIVM.

Glossaire

- **ARC-INFO :** Logiciel standard de SIG, structure vecteur (Standard software for GIS, vector structure).
- **CATAENAT :** Charge acide totale sur les écosystèmes naturels terrestres (Sous-réseau du Réseau RENECOFOR).
- **Charge critique :** Quantité maximale de polluants admissible par un écosystème avant que des effets néfastes n'apparaissent sur son environnement (Maximum amount of pollutant below which significant harmful effects do not occur on the environment).
- **EMEP50 :** European Monitoring and Evaluation Programme (Grille Quadrillage référence au pas de 50 km sur l'Europe pour le calcul des charges critiques).
- **IDRISI :** Logiciel de SIG, structure raster.
- **Lambert II étendu :** Système de coordonnées référencées utilisé en France.
- **ODBC :** Open Database Connectivity (système ouvert de connectivité aux bases de données) : interface supportée par ORACLE.
- **ONF :** Office national des forêts (National Forest Office, France).
- **ORACLE :** Nom du SGBDR.
- **Raster (trame) :** Modèle reposant sur de petites cellules juxtaposées.
- **RENECOFOR :** Réseau national de suivi à long terme des écosystèmes forestiers (National Network for the Monitoring of Forest Ecosystem, France).
- **RIVM :** Service de la santé publique et de l'environnement aux Pays-Bas (National Institute of Public Health and the Environment in the Netherlands).
- **SGBDR :** Système de gestion de base de données de type relationnel.
- **SIG :** Système d'information géographique (Geographic Information System, GIS).
- **SQL :** Langage non procédural de requêtes à une base de données (Structured Query Language).
- **UTM :** Système de coordonnées international (Universal Transverse Mercator).
- **Vecteur :** Modèle reposant sur des primitives graphiques (segments).

Plusieurs contraintes fortes ont permis d'emblée de structurer cette étude :

- la pérennité de la prise en compte des données françaises par le RIVM est liée à la rapidité d'échange des données. Les besoins du RIVM doivent pouvoir être traités de façon rapide et fiable et la France doit pouvoir répondre de façon simple et transparente à ces besoins ; ceci implique en particulier chaque année une mise à jour des éléments nécessaires au calcul des charges critiques ;

- l'acquisition des données de base doit être prévue sur le long terme pour deux raisons : la première est d'inventorier et d'informatiser les nombreuses données utilisables dans le domaine des charges critiques qui sont encore disséminées dans différents laboratoires du territoire français sans être toujours informatisées, et de les intégrer « en temps réel » ; la seconde est de pouvoir utiliser, en collaboration avec les détenteurs, les résultats en cours d'acquisition au travers de différents réseaux d'observation français : Réseau européen du Département Santé des forêts, du ministère de l'Agriculture, le réseau RENECOFOR de l'ONF [15], l'Observatoire de la qualité des sols du ministère de l'Environnement... ;

- il apparaît enfin nécessaire de réfléchir dès à présent sur les évolutions possibles de la problématique charges critiques avec la valorisation effectivement des acquis de ces 10 dernières années [3, 9, 16]. De plus, les nouveaux projets de calcul des charges critiques en cours (calcul des charges critiques d'acidité par modélisation dynamique, calcul des charges critiques en métaux lourds, en particulier) contribueront à fournir des données encore très rares et qu'il y aura lieu d'exploiter avec le maximum d'efficacité.

Un schéma initial s'est donc dégagé de lui-même assez rapidement. Il s'agissait de concevoir et de mettre en place une base de données :

– dans un but finalisé de cartographie à l'échelle européenne ;

– dans un projet de valorisation de données existantes (ou en cours d'acquisition) issues de plusieurs organismes et d'études pluridisciplinaires au travers d'une extension et d'un partage des connaissances à l'échelle française.

Rappels des exigences au niveau européen.

1. Les charges critiques sont estimées de façon générale pour le territoire européen au travers d'un maillage systématique de 50 × 50 km (réseau EMEP50) puis extrapolées par le RIVM à une grille élargie de 150 × 150 km (Figure 1).

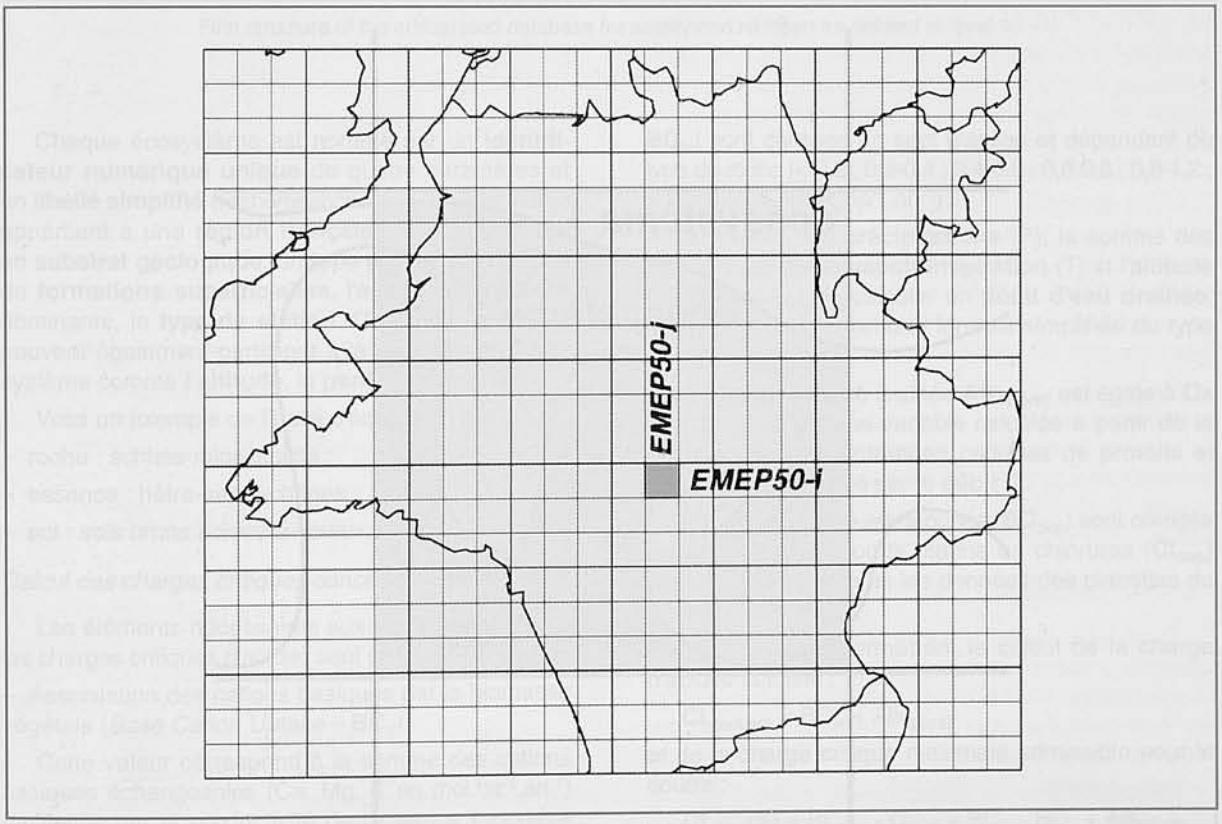


Figure 1.
Quadrillage EMEP50 (50 × 50 km) sur le territoire français.
Exemple d'une maille repérée par ses coordonnées horizontale et verticale EMEP50-i et EMEP50-j.
EMEP50 (50 × 50 km) grid on France.
Example of a referenced grid using vertical and horizontal coordinates EMEP50-i and EMEP50-j.

2. Pour chaque maille, les charges critiques d'acidité sont calculées à partir des données propres aux écosystèmes sous-jacents. La figure 2 montre, par exemple, que la maille EMEP50-i/EMEP50-j sera renseignée par trois calculs de charges critiques propres aux trois écosystèmes sous-jacents.

3. Les fichiers de sortie à l'attention du RIVM doivent comporter 17 variables concernant les charges critiques d'acidité et d'azote pour tous les écosystèmes en présence ; soit succinctement : les coordonnées géographiques en degrés décimaux pour la longitude et la latitude ; deux indices de localisation de la maille du réseau EMEP dans les directions horizontale et verticale ; les surfaces des écosystèmes considérés ; 11 variables concernant les charges critiques et enfin les types d'écosystèmes. Ceci peut être schématisé dans le tableau suivant :

maille EMEP n° 1	écosystème n° 1	n variables « charges critiques »
maille EMEP n° 1	écosystème n° 2	n variables « charges critiques »
maille EMEP n° 1	écosystème n° 3	n variables « charges critiques »
maille EMEP n° 2	écosystème n° 1	n variables « charges critiques »
maille EMEP n° 2	écosystème n° 2	n variables « charges critiques »

La base de niveau 1

L'étude de faisabilité d'une base de données des charges critiques pour la France [11] a décomposé le projet en deux niveaux développés successivement dans le temps. Le projet initial a permis de développer le premier niveau [9, 17, 18], puis par la suite le niveau 2 ainsi que le lien entre la base de données et le SIG dont l'utilisation a permis l'élaboration des premières cartes de charges critiques pour la France [10].

Structure de la base au niveau 1

Le premier développement a conduit à une structuration de la base de données « charges critiques » au moyen de tables. À chacune des tables correspond une liste minimale de variables.

Table 1 – Dictionnaire actuel des écosystèmes

Les écosystèmes utilisés pour le calcul des charges critiques doivent être identifiés par un certain nombre de critères. La table « Dictionnaire actuel des écosystèmes » regroupe ces différents paramètres.

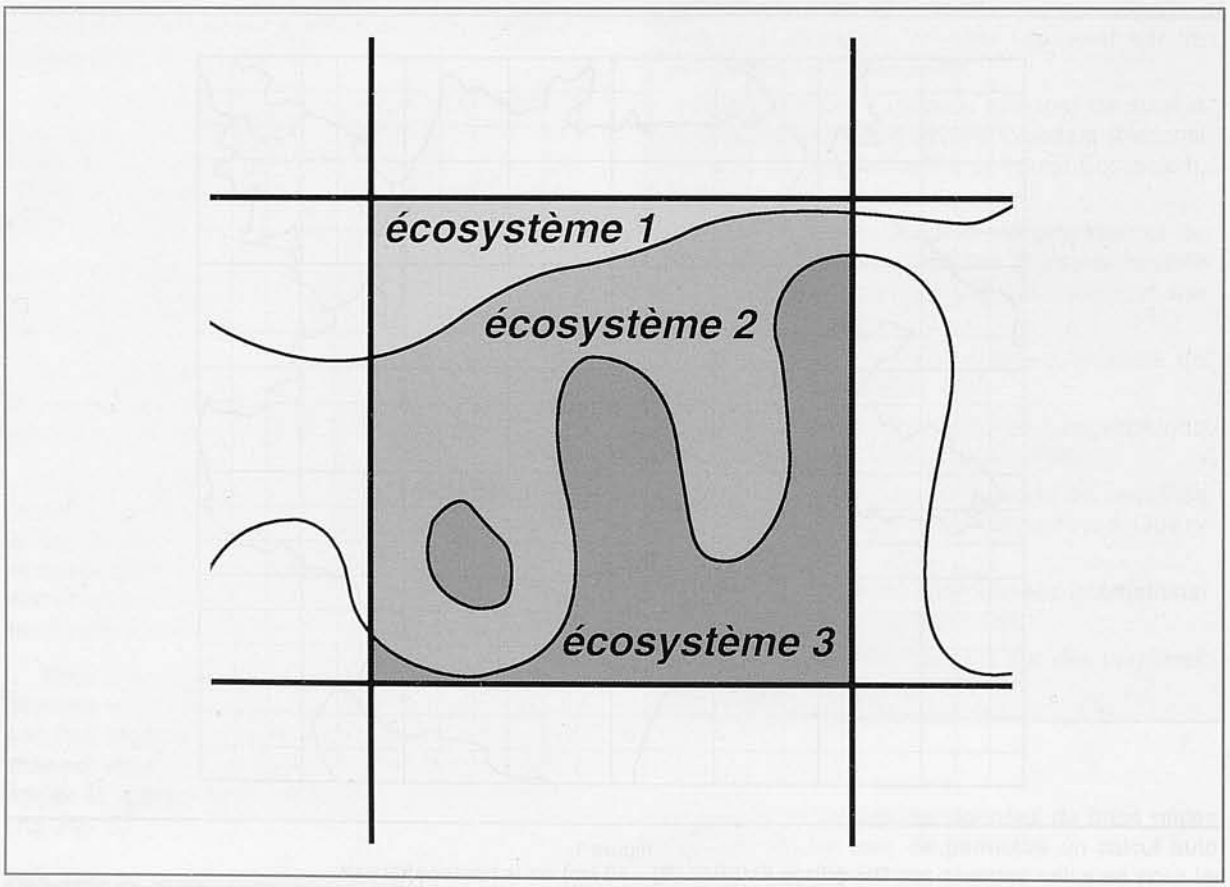


Figure 2.
Exemple de superposition d'une maille EMEP50 sur un fond de carte « écosystème ».
Example of the EMEP50 grid overlaying the ecosystem map.

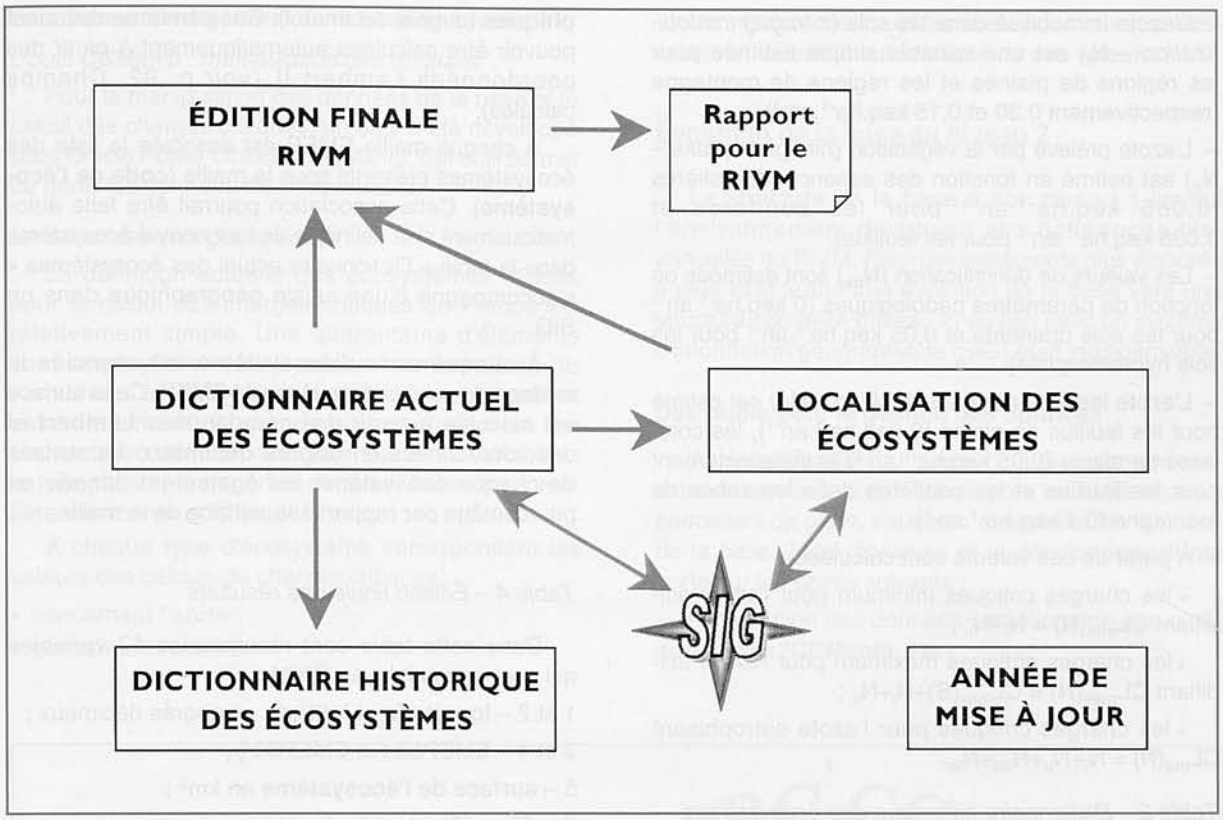


Figure 3.
Premier schéma de la structure de la base de données pour les charges critiques d'acidité et d'azote tel qu'il avait été défini au niveau 1.
First structure of the critical load database for acidity and nitrogen as defined at level 1.

Chaque écosystème est nommé par un **identificateur numérique unique** de quatre caractères et un **libellé simplifié** (roche/essence/sol). L'écosystème appartient à une **région française**. Il est défini par un **substrat géologique**, un **type de sol**, la présence de **formations superficielles**, l'**essence forestière** dominante, le **type de station**. D'autres paramètres peuvent également participer à la définition de l'écosystème comme **l'altitude**, la **pente**, **l'exposition**...

Voici un exemple de libellé d'écosystème :

- roche : schiste-micaschiste ;
- essence : hêtre-sapin-chênes ;
- sol : sols bruns acides à ocreux.

Calcul des charges critiques concernant l'acidité

Les éléments nécessaires aux calculs concernant les charges critiques d'acidité sont décrits ci dessous :

– Assimilation des cations basiques par la biomasse végétale (*Base Cation Uptake* – $\mathbf{BC_u}$).

Cette valeur correspond à la somme des cations basiques échangeables (Ca, Mg, K en $mol.ha^{-1}.an^{-1}$) multipliée par la productivité de l'essence forestière ($m^3.ha^{-1}.an^{-1}$). Les valeurs actuellement utilisées proviennent de travaux de recherche menés sur le cycle des éléments [11, 19].

– Les valeurs d'altération (*Base Cation weathering* – $\mathbf{BC_w}$) sont divisées en sept classes et dépendent du type de roche (< 0,2 ; 0,2-0,4 ; 0,4-0,6 ; 0,6-0,8 ; 0,8-1,2 ; 1,2-2,0 ; > 2,0 $keq.ha^{-1}.an^{-1}$).

– La somme des **précipitations** (P), la somme des pertes d'eau par **l'évapotranspiration** (T) et l'**altitude** (a), permettent de calculer un **débit d'eau drainée**, en $m^3.ha^{-1}.an^{-1}$, selon une formule simplifiée du type $Q = (1-a) \times P-T$.

– L'alcalinité critique drainée $\mathbf{Alk_{le(crit)}}$ est égale à $Qx(H^+ + Al^{+++})$. C'est une variable calculée à partir de la somme des concentrations critiques de protons et d'aluminium multiplié par le **débit Q**.

– Les cations basiques déposés ($\mathbf{BC_{dep}}$) sont corrigés par rapport aux apports marins en chlorures ($\mathbf{Cl_{dep}}$) et sont estimés d'après les données des placettes du réseau CATAENAT.

– Ces variables permettent le calcul de la charge d'acidité actuelle :

$$\mathbf{CL_{(act.ac)}} = BC_w + Alk_{le(crit)}$$

et de la charge critique maximale admissible pour le soufre :

$$\mathbf{CL_{max}(S)} = BC_{dep} - Cl_{dep} + BC_w - BC_u + Alk_{le(crit)}$$

Partie calcul des charges critiques concernant l'azote

Voici les éléments nécessaires aux calculs concernant les charges critiques d'azote :

– L'azote immobilisé dans les sols (*nitrogen immobilization* – N_i) est une variable simple estimée pour les régions de plaines et les régions de montagne (respectivement 0,30 et 0,15 keq.ha^{-1}.an^{-1}).

– L'azote prélevé par la végétation (*nitrogen uptake* – N_u) est estimé en fonction des essences forestières (0,065 keq.ha^{-1}.an^{-1} pour les conifères et 0,038 keq.ha^{-1}.an^{-1} pour les feuillus).

– Les valeurs de dénitrification (N_{de}) sont estimées en fonction de paramètres pédologiques (0 keq.ha^{-1}.an^{-1} pour les sols drainants et 0,05 keq.ha^{-1}.an^{-1} pour les sols hydromorphes).

– L'azote lessivé (*nitrogen leaching* – N_{le}) est estimé pour les feuillus de plaine (0 keq.ha^{-1}.an^{-1}), les conifères de plaine (0,05 keq.ha^{-1}.an^{-1}) et indistinctement pour les feuillus et les conifères dans les zones de montagne (0,1 keq.ha^{-1}.an^{-1}).

– À partir de ces valeurs sont calculées :

• les charges critiques minimum pour l'azote acidifiant $\mathbf{CL_{min}(N)} = N_i+N_u$;

• les charges critiques maximum pour l'azote acidifiant $\mathbf{CL_{max}(N)} = CL_{max}(S)+N_i+N_u$;

• les charges critiques pour l'azote eutrophisant $\mathbf{CL_{nut}(N)} = N_i+N_u+N_{de}+N_{le}$.

Table 2 – Dictionnaire historique des écosystèmes

La liste actuelle des écosystèmes utilisés dans le calcul des charges critiques n'est pas un ensemble fini. De nouveaux écosystèmes peuvent être introduits dans la table précédente et certains des écosystèmes existants modifiés. On peut, par exemple, vouloir faire une distinction dans l'écosystème « pin maritime/sable pur/podzol » en introduisant la précision « podzol à alios induré » ou « podzol meuble ». Toutes les modifications susceptibles de conduire à une meilleure estimation des charges critiques sont évidemment souhaitables, mais il apparaît raisonnable de conserver une trace des systèmes utilisés antérieurement. De plus, un enregistrement annuel des écosystèmes utilisés doit être fait.

La table « Dictionnaire historique des écosystèmes » comporte le **code actuel de l'écosystème**, **l'ancien code de l'écosystème**, le **libellé de l'ancien écosystème**, le **nom de l'auteur** de la modification, la **date** de modification, la **raison** (forte) de la modification.

Table 3 – Localisation

Cette table est la plus importante pour la base de données puisque c'est à ce niveau que sont croisés le quadrillage EMEP50 et les écosystèmes définis sur le territoire français. Cette table donne la structure finale du fichier résultat.

Elle comporte les coordonnées des mailles du réseau EMEP (**EMEP50-i et EMEP50-j**) pour lesquelles des informations sont disponibles. Ces coordonnées en X et en Y sont également disponibles dans le système **Lambert II étendu** (données kilométriques), **UTM** et en **coordonnées géographiques** (degrés décimaux). Ces dernières devraient pouvoir être calculées automatiquement à partir des coordonnées Lambert II (*voir* p. 92, Champs calculés).

À chaque maille EMEP est associée la liste des écosystèmes présents sous la maille (**code de l'écosystème**). Cette association pourrait être faite automatiquement si la définition de tout nouvel écosystème dans la table « Dictionnaire actuel des écosystèmes » s'accompagne d'une saisie géographique dans un SIG.

À chaque code d'écosystème est associée la **surface** de celui-ci dans la maille EMEP. Cette surface est calculée à partir des coordonnées **Lambert** et des coordonnées en **degrés décimaux**. La surface de chaque écosystème est également donnée en pourcentage par rapport à la surface de la maille.

Table 4 – Édition finale des résultats

Dans cette table sont résumés les 17 variables qui sont transmises au RIVM :

1 et 2 – **longitude** et **latitude** en degrés décimaux ;

3 et 4 – **EMEP50-i** et **EMEP50-j** ;

5 – **surface de l'écosystème** en km^2 ;

6 – $\mathbf{CL_{max}(S)}$ en eq.ha^{-1}.an^{-1} ;

7 – $\mathbf{CL_{min}(N)}$ en eq.ha^{-1}.an^{-1} ;

8 – $\mathbf{CL_{max}(N)}$ en eq.ha^{-1}.an^{-1} ;

9 – $\mathbf{Cl_{nut}(N)}$ en eq.ha^{-1}.an^{-1} ;

10 – $\mathbf{BC_{dep} - Cl_{dep}}$ en eq.ha^{-1}.an^{-1} ;

11 – $\mathbf{BC_u}$ en eq.ha^{-1}.an^{-1} ;

12 – $\mathbf{BC_w}$ en eq.ha^{-1}.an^{-1} ;

13 – $\mathbf{Alk_{le(crit)}}$ en eq.ha^{-1}.an^{-1} ;

14 – N_i en eq.ha^{-1}.an^{-1} ;

15 – N_u en eq.ha^{-1}.an^{-1} ;

16 – $N_{le(crit)}$ en eq.ha^{-1}.an^{-1} ;

17 – **type d'écosystème**.

Outre la consultation ponctuelle des résultats compilés, cette table permet une édition informatique automatique dans le format requis par le RIVM.

Table 0 – Mise à jour de la base de données

Tous les éléments présentés ci-dessus sont dépendants d'une table « Mise à jour ». Celle-ci fournit la dernière **date de mise à jour** de la base de données, le nom de la ou des **personnes responsables** l'ayant effectuée ainsi que celui de **l'organisme responsable**.

Cette table sert également à un certain nombre de vérifications pratiques indispensables comme le **nombre d'enregistrements par carré EMEP**, le **nombre total d'écosystèmes pris en compte**, la **surface totale des écosystèmes pris en compte** dans : 1. les calculs de charges critiques d'acidité, 2. les calculs de charges critiques pour l'azote et 3. l'ensemble des calculs.

Fonctionnement général de la base en niveau 1

L'outil CCManip : manipulation des données

Pour la manipulation des données de la base et le calcul des charges critiques, un outil a été développé sous Oracle Power Objects : CCManip. Cet outil permet de gérer les données de chaque table.

Le calcul des charges critiques

La définition actuelle des écosystèmes utilisés pour le calcul des charges critiques en France est relativement simple. Une quarantaine d'éléments sont employés et définis par un nombre limité de variables : à une région française correspondent essentiellement de grands massifs forestiers qui peuvent être définis par quatre éléments variables : le type de sol, le matériau parental, l'essence forestière dominante et l'altitude.

À chaque type d'écosystème correspondent les valeurs des calculs de charges critiques :

- concernant l'azote ;
- concernant l'acidité.

Un module de l'outil permet, pour chaque écosystème, un calcul automatique de ces valeurs.

Évolution de la base au niveau 2

La structure de la base à son niveau 1 limitait l'environnement de travail aux échéances bisannuelles du RIVM. Pour des traitements plus élaborés à plus long terme une évolution de la base vers une structure complète (niveau 2) en lien avec un système d'information géographique (SIG) était indispensable.

Des outils pour la gestion des données

Après réalisation du niveau 1, il était donc nécessaire de reprendre ou de développer les outils logiciels permettant de gérer, visualiser, et exploiter les données de la base. Ainsi, l'analyse et le développement ont porté sur les points suivants :

– manipulation des données : amélioration des outils disponibles (**CCManip**, etc.) ;

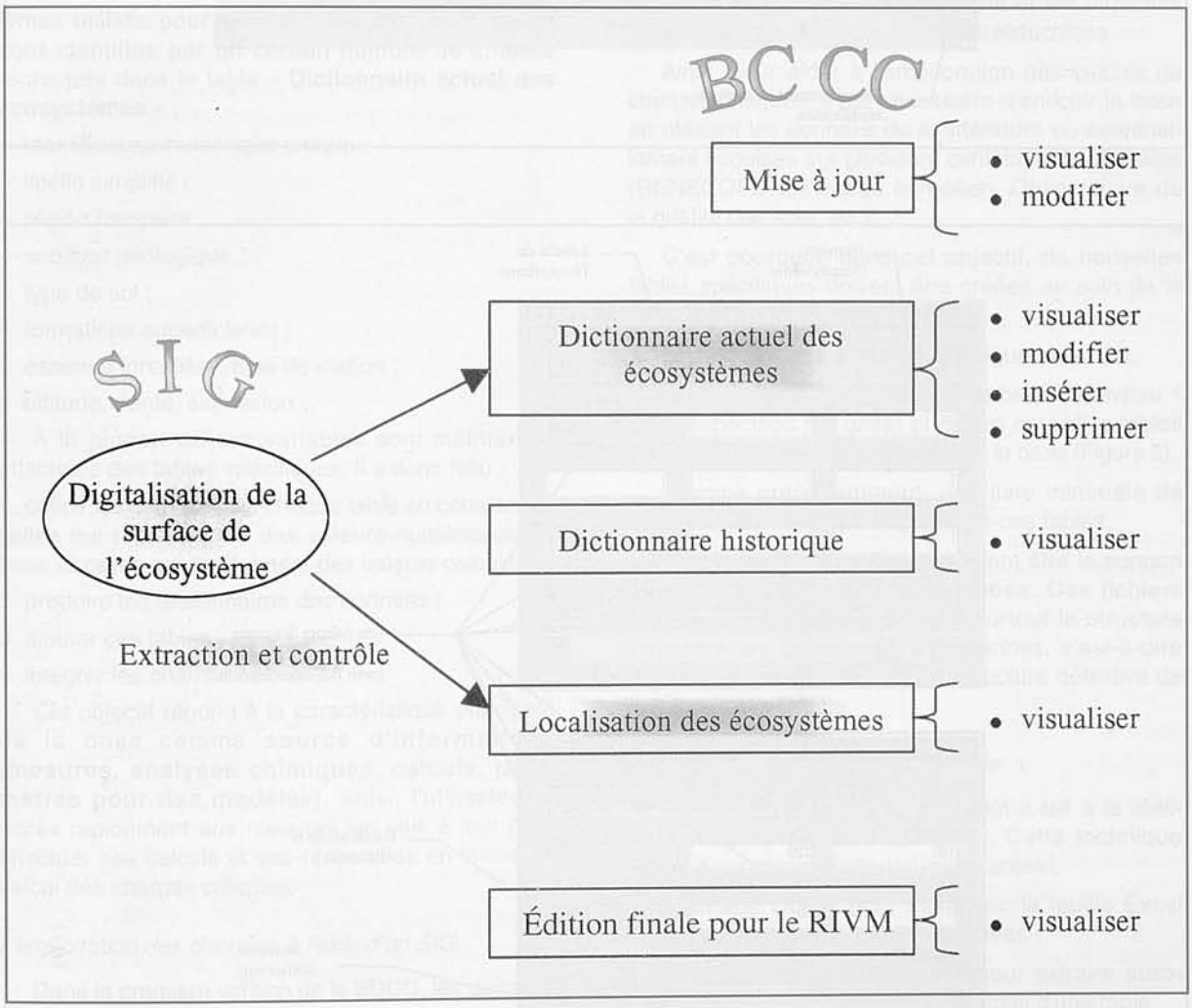
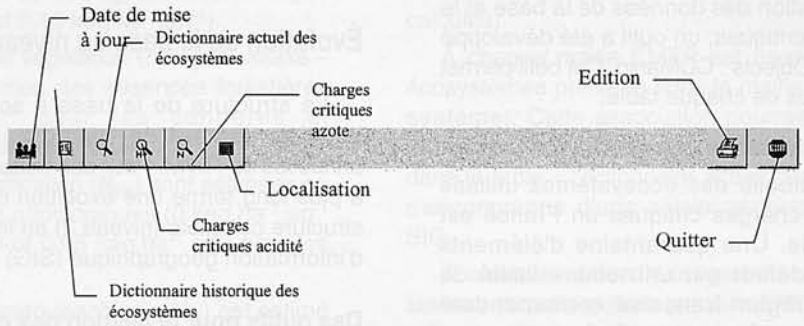
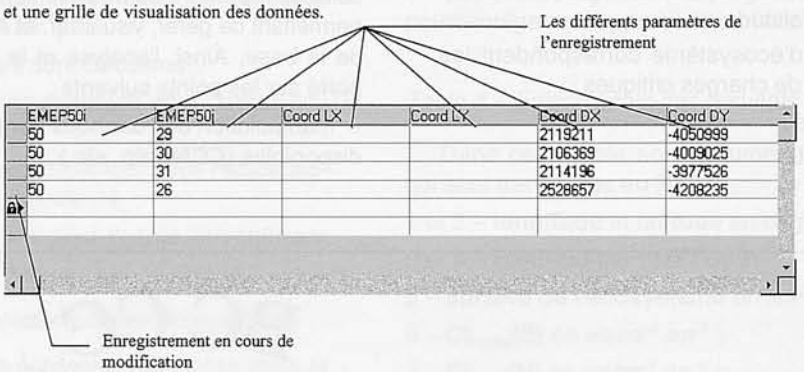


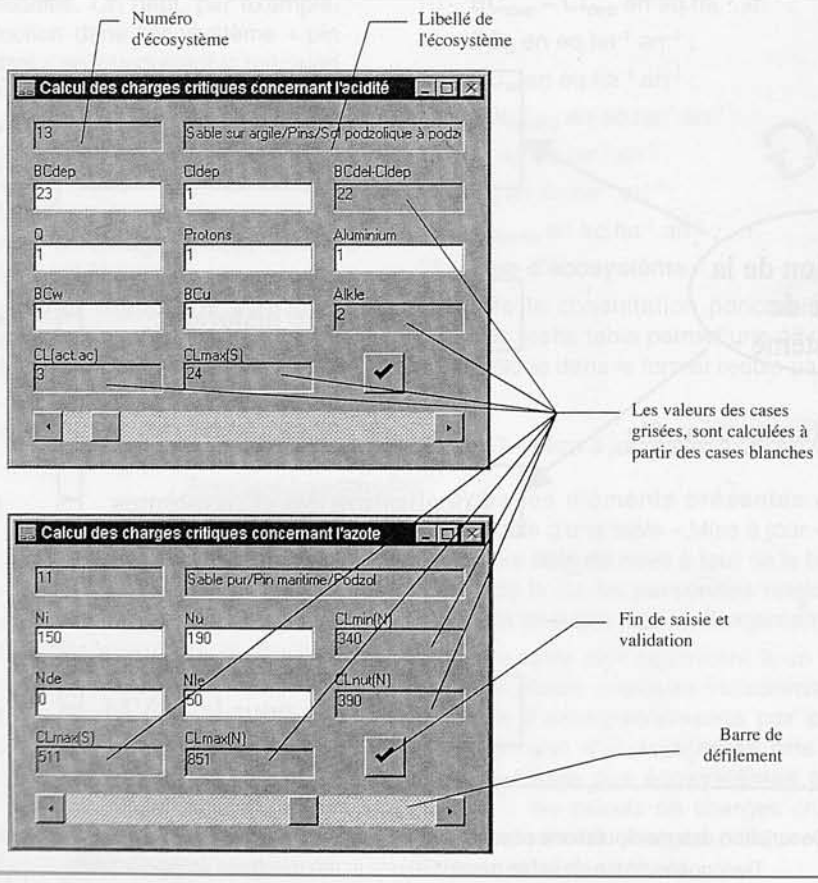
Figure 4.
Description des manipulations possibles dans la base de données à l'aide de CCManip.
Description of the possible manipulations in the database using CCManip.

L'application présente une barre d'outils



et une grille de visualisation des données.





– extraction des données : analyse puis définition de **méthodes d'extraction** avec les outils existants ;

– insertion des données : analyse des besoins et développement de l'outil **CCImport** ;

– calcul automatique des **données statistiques** : tables moyenne, médiane, écart type ;

– sauvegarde des données : analyse du niveau de sécurité souhaité. Restructuration de la base, puis développement de l'outil **CCMaj** et de la fonction automatique de sauvegarde **CCMajTrigger** pour la base.

Évolution de la base : le niveau 2

Le premier développement de la base BDCC a tenu compte d'une contrainte majeure de réalisation, la prévision d'ouverture vers une structure plus complexe (le niveau 2). La base de niveau 1 constituait donc un bon support pour cette deuxième phase.

Cette phase peut se découper en deux objectifs :

L'enrichissement de la base BDCC

Conformément au premier schéma, les écosystèmes utilisés pour le calcul des charges critiques sont identifiés par un certain nombre de critères regroupés dans la table « **Dictionnaire actuel des écosystèmes** » :

– identificateur numérique unique ;
– libellé simplifié ;
– région française ;
– substrat géologique ;
– type de sol ;
– formations superficielles ;
– essence forestière, type de station ;
– altitude, pente, exposition...

À la plupart de ces variables sont maintenant attachées des tables spécifiques. Il a donc fallu :

1. définir les colonnes de chaque table en considérant celles qui représentent des valeurs numériques de base et celles qui contiennent des valeurs calculées ;
2. produire les dictionnaires des données ;
3. ajouter ces tables ;
4. intégrer les champs calculés.

Cet objectif répond à la caractéristique principale de la base comme **source d'informations (mesures, analyses chimiques, calculs, paramètres pour des modèles)**. Ainsi, l'utilisateur a accès rapidement aux mesures les plus à jour pour effectuer ses calculs et ses recherches en terme de calcul des charges critiques.

L'exploitation des données à l'aide d'un SIG

Dans la première version de la BDCC, les données sont exploitées via CCManip et par extraction des données vers le tableur Excel, par exemple. Cependant, toutes les informations de la BDCC sont des données géoréférencées, c'est-à-dire localisées par des coordonnées géographiques sur le globe terrestre. Souvent on utilise les coordonnées en degrés décimaux, mais il existe de nombreuses autres normes (projections planes Lambert, UTM...). Ces données peuvent donc être cartographiées et visualisées grâce à un système d'information géographique. Puisqu'une grande part de l'intérêt de la base de données repose sur le couplage entre les données géoréférencées contenues dans la base et celles numérisées dans un SIG, il est nécessaire d'étudier le lien entre la base Oracle et un SIG. Cette étude s'inscrit dans une démarche cartographique et statistique basée sur la mesure, le croisement et l'analyse de données spatialisées.

Nous avons donc examiné les différents SIG existants et étudié alors les possibilités offertes pour un lien avec le SGBD Oracle.

Structure de la base complète : niveau 2

L'objet de la BDCC étant, de façon générale, d'améliorer les connaissances au niveau français, de nouveaux écosystèmes devront être pris en compte à des niveaux de plus en plus fins et les variables utilisées seront de moins en moins réductrices.

Ainsi, pour aider à l'amélioration des calculs de charges critiques, il est nécessaire d'enrichir la base en utilisant les données de la littérature ou éventuellement acquises sur plusieurs centaines de placettes (RENECOFOR, Réseau européen, Observatoire de la qualité des sols, etc.).

C'est pourquoi, dans cet objectif, de nouvelles tables spécifiques doivent être créées au sein de la structure de la BDCC.

Soit le nouveau schéma (*voir* Figure 5, p. 92).

Les contraintes définies pour la base de niveau 1 sont respectées (en grisé) et quatre nouvelles tables de données (à gauche) enrichissent la base (Figure 5).

Comme précédemment, une liste minimale de variables correspond à chacune de ces tables.

De simples fichiers Excel peuvent être le support de base de ces nouvelles données. Ces fichiers contiennent les données, mais surtout la structure des données organisées en colonnes, c'est-à-dire directement ce qui constitue la structure définitive de la future table Oracle.

Migration de la structure d'une table

Database Designer se prête tout à fait à la réalisation rapide de cette opération. Cette technique utilise deux caractéristiques importantes :

– l'interface ODBC pour interroger la feuille Excel comme une table de base de données ;

– le « Reverse Engineering » : pour extraire **automatiquement** la structure en colonnes d'une table.

Dans notre cas, il s'agit d'extraire les noms des colonnes du fichier Excel.

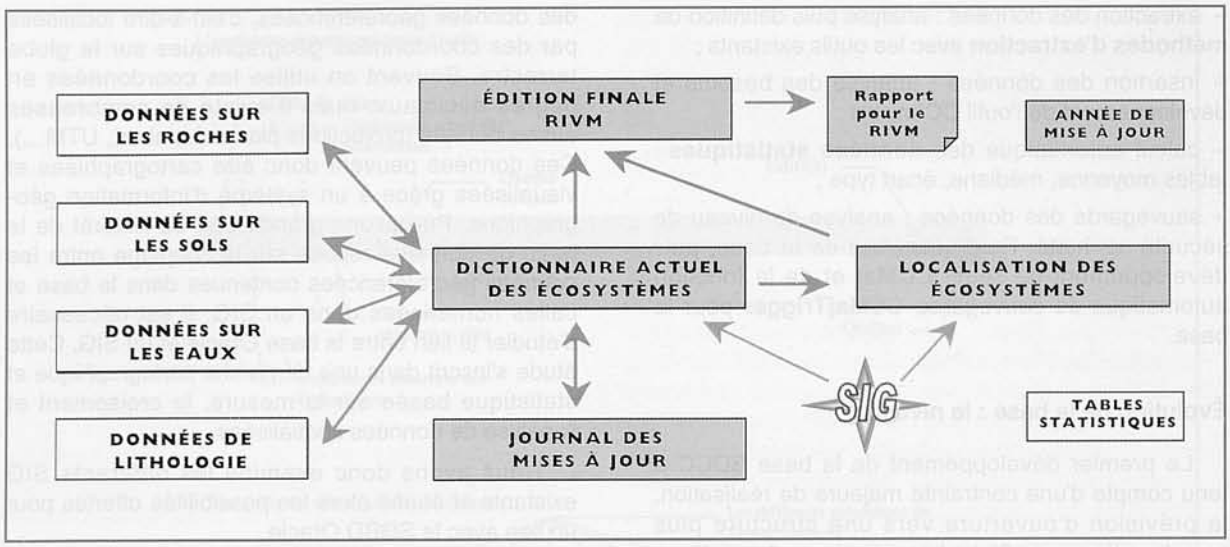


Figure 5.
Schéma de la structure de la BDCC au niveau 2.
Structure of the database at level 2.

Reverse Engineering

Pour une migration de structure, il est tout d'abord nécessaire de :

- purger le fichier en ne conservant que les données de base (les champs calculés peuvent être ajoutés plus tard, *voir* Champs calculés) ;
- supprimer les graphiques ou tout autre élément de la feuille qui n'est pas une donnée de base ;
- vérifier les intitulés de colonnes ;
- sauvegarder le fichier au format CSV : format texte délimité par des tabulations.

On peut alors utiliser le driver ODBC Text et la fenêtre de connexion ODBC du Reverse Engineer Wizard du Designer.

Affiner la structure

Cette étape consiste à utiliser les fonctionnalités du Designer pour créer un modèle de base par l'intermédiaire d'une interface graphique conviviale : création de tables, de colonnes, de vues, de contraintes d'intégrité...

Grâce à ces fonctions, cette étape consiste à :

- (re)modifier le cas échéant les intitulés de colonnes ;
- spécifier la clé de la table si nécessaire ;
- modifier le format des colonnes.

Création de la table

Cette opération utilise une dernière fonctionnalité du Designer, la connexion native avec la base Oracle. On utilise à cet effet le module « Generation Wizard ». Son utilisation permet d'envoyer automatiquement les requêtes de création de la structure que l'on vient de spécifier, et ce, en contrôlant les erreurs.

Grâce à **Oracle Database Designer**, on peut donc facilement transformer toutes les données issues de fichiers Excel.

On voit ici qu'un des intérêts de la base BDCC réside dans le regroupement, la centralisation et l'uniformisation des informations. Auparavant, toutes les informations pouvaient être éparpillées dans des fichiers de différents formats sur divers supports. Ainsi, les données de base sont rassemblées sous le même format et dans ce sens, la BDCC aide à un calcul plus sûr des charges critiques.

Champs calculés

L'analyse des tables a montré que de nombreux champs peuvent être ajoutés en tant que champs calculés. En effet, prenons quelques exemples :

– Pour produire systématiquement le diagramme de De La Roche pour chaque analyse totale de ***sol***, il nous faut disposer de paramètres qui sont calculés à partir des pourcentages des éléments majeurs. Ces calculs de paramètres sont répétitifs et peuvent aisément être automatisés.

– En ce qui concerne les ***eaux***, des données d'analyses chimiques sont stockées en mmol/l, et il est parfois nécessaire d'avoir l'indication en mg/l. Ce sont des calculs qui s'automatisent facilement.

L'utilisation du « trigger » répond précisément à ces calculs automatiques.

– Les automatismes de ce type sont déclenchés sur la table *nom_table* en cas d'insertion ou modification sur des lignes de la table.

– Le champ calculé est affecté du résultat de la fonction f. La fonction précise le calcul à effectuer en utilisant le champ de référence et des paramètres.

– Exemple :

Y : champ calculé (quantité en mg/l)

X : champ de référence (quantité en mmol/l)

M : paramètres (masse molaire).

f : la fonction

Calcul : **Y = f(X,M) = MX**

Étude du lien avec un SIG

Cette étude s'est inscrite dans une démarche cartographique et statistique basée sur la mesure, le croisement et l'analyse de données spatialisées [20, 21]. Plus précisément, elle a consisté à déterminer les possibilités de lien non natif de la base Oracle avec la version SIG MapInfo Pro 4.5 [22].

Aspect logiciel

Rappelons qu'une grande part de l'intérêt de la base de données repose sur le couplage entre les données géoréférencées contenues dans la base et celles numérisées dans le SIG. Ce SIG devait pouvoir traiter à la fois des fichiers « raster » et des fichiers « vecteurs », ce qui a limité le choix :

– IDRISI traite des fichiers images (ou « rasters ») et permet des superpositions de fichiers vecteurs ;

– Arc-Info réalise des traitements à la fois images et vecteurs ;

– MapInfo traite les vecteurs et peut afficher des fichiers images.

IDRISI ne satisfaisait pas complètement les contraintes. Arc-Info présentait tous les avantages, et serait l'outil parfait si son utilisation n'était pas complexe.

MapInfo est un produit de ADDE, c'est un produit avant tout « géodécisionnel ». La validité de ce choix a été renforcée par l'accord technologique MapInfo – Oracle *(7 avril 1999)*, les deux acteurs s'étant mis d'accord pour déployer les analyses spatiales dans les applications d'intelligence économique, de « datawarehouse » et de commerce électronique. Oracle a choisi d'intégrer la technologie MapInfo dans les modules spatiaux de ses bases de données.

Aspects techniques du lien

MapInfo permet de charger une table Oracle via le driver ODBC MapInfo pour Oracle. Il charge la table ODBC depuis la base de données distante dans une table MapInfo, appelée dans ce cas *table liée*. Cette table conserve des liens avec la table ODBC correspondante sous Oracle. La plupart des opérations applicables aux tables MapInfo normales peuvent être exécutées sur les tables liées : affichage, modification, copie, enregistrement... Il est cependant impossible d'en modifier la structure.

Chargement d'une table liée

MapInfo permet de charger toute une table Oracle mais aussi seulement une partie en spécifiant une requête SQL précise. On peut ainsi, par exemple, faire une sélection par colonnes et par lignes.

Enregistrement d'une table liée

On peut donc visualiser et modifier une table Oracle via le SIG. Cependant, dans le cas d'une version multi-utilisateurs, l'enregistrement de modifications peut générer des conflits entre les données résidant dans la base Oracle et les nouvelles données que nous souhaitons y enregistrer. En effet, il est possible que d'autres utilisateurs aient effectué des modifications dans la table utilisée depuis son chargement dans MapInfo.

Pour répondre à cela, MapInfo possède un mode de résolution de conflits qui prend en compte trois instances de l'enregistrement concerné par la mise à jour :

– l'état initial de l'enregistrement dans la base Oracle, au moment de son extraction ;

– l'état actuel de l'enregistrement, tel qu'il apparaît dans MapInfo après modification éventuelle par l'utilisateur ;

– l'état actuel de l'enregistrement dans la base Oracle, tel qu'il apparaît dans la base Oracle au moment de la mise à jour.

Il y a conflit lorsque l'état initial de l'enregistrement dans la base Oracle est différent de son état au moment de la mise à jour.

Ainsi, ce mode permet de choisir si la mise à jour doit utiliser la valeur de la base Oracle ou celle de MapInfo.

MapInfo propose aussi un mode de résolution par défaut.

Ce mode renforce la fiabilité de fonctionnement du lien MapInfo-Oracle. Cependant, le fonctionnement n'est pas totalement transparent et c'est à l'utilisateur de gérer les mises à jour. En revanche, en mode mono-utilisateur, aucun problème ne se pose.

Visualisation des données sous forme de carte

C'est là un objectif majeur pour l'évolution de la base.

Pour cela, MapInfo fournit la fonction « Rendre une table ODBC distante cartographiable dans MapInfo ». Cette fonction fournit à MapInfo les informations nécessaires pour cartographier la table distante :

– les colonnes contenant les coordonnées nécessaires pour cartographier la table ;

– la projection utilisée ;

– la colonne utilisée comme index spatial ;

– le symbole à utiliser pour les objets.

MapInfo stocke ces informations dans une table Dictionnaire de cartes sous Oracle.

Croisement

Afin d'effectuer les futurs découpages en écosystèmes du territoire français, il est indispensable que le SIG puisse croiser deux couvertures spatiales et déterminer les nouveaux polygones émanant du

croisement. MapInfo permet de faire cette opération que l'on appelle croisement pour ArcInfo mais que l'on qualifie plutôt « d'intersection évoluée » pour MapInfo.

Bilan technique

MapInfo 4.5 n'a pas un accès natif à Oracle, ce qui ne rend pas toutes les opérations transparentes. Cependant depuis la dernière version 5.0, cet accès est natif.

Le lien SIG-Oracle pourrait sans doute encore être amélioré avec :

– le système d'index spatial Oracle SDO ;

– les logiciels issus de l'accord MapInfo-Oracle, utilisables pour notre problématique.

Exploitation des données

MapInfo permet un éventail très riche d'analyses des données : sectorisation, analyses thématiques, sélection des données, extraction.

Remarque : Il est intéressant de noter que MapInfo fournit une interface graphique simple pour créer des requêtes SQL sans écrire de ligne de code SQL.

De même, MapInfo permet d'extraire les données en les sauvegardant à des formats standard comme Excel.

Exemples d'utilisation du SIG

La carte ci-dessous a été réalisée avec ArcInfo [18] et importée dans MapInfo.

Cette carte peut servir de support de travail avec les données propres à la base de données.

Lors du couplage MapInfo – BDCC, le centroïde d'un écosystème est repéré dans le quadrillage EMEP50 par un point correspondant. C'est un exemple simple de couplage entre les données géoréférencées contenues dans la base et les données numérisées dans le SIG.

Conclusion pour la BDCC et perspectives d'évolution

Dans l'état actuel, la BDCC peut être utilisée comme nouvel outil de travail et de recherche pour le

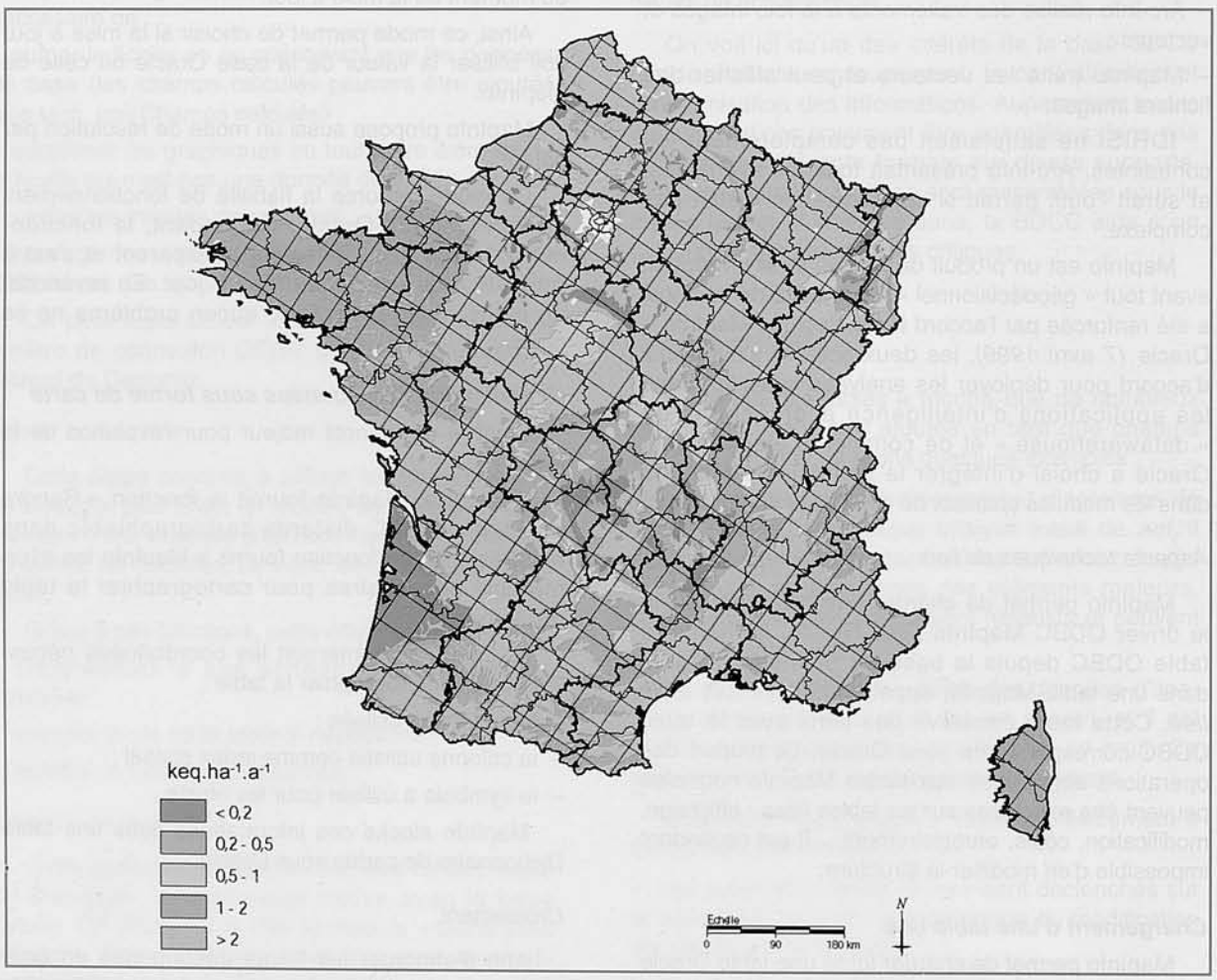


Figure 6.
Carte de représentation des classes d'altération des sols (BC_W).
Map of soil weathering rates (BC_W).

calcul des charges critiques. Le projet BDCC se situe à un état avancé du niveau 2, à savoir :

- la BDCC est une concentration d'informations pour les travaux sur les charges critiques : elle répond aux objectifs de regroupement, de centralisation et d'uniformisation des informations ;
- tous les outils existants et développés fournissent un environnement efficace de travail sur les charges critiques : CCManip, CCMaj, CCImport, Database Designer...

De plus, le lien avec le SIG est un aspect majeur de la base. L'étude réalisée sur le SIG MapInfo est concluante : MapInfo permet des représentations cartographiques assez aisées du contenu de la BDCC. À présent, la base peut évoluer en tenant encore plus précisément compte de l'aspect SIG :

- pour l'amélioration des connaissances au niveau de la France, de nouveaux écosystèmes devront être pris en compte à des niveaux de plus en plus fins. Pour cela, de nouvelles données devront être incluses. Le SIG apporte à l'environnement de travail sur les charges critiques des facilités et des fonctionnalités de nature à faciliter les mises à jour. À long terme, la BDCC devrait comporter 10 000 à 20 000 enregistrements de localisation avec environ 200 écosystèmes définis sur le territoire français ;
- pour le rassemblement des cartes existantes caractérisant les facteurs du milieu (topographie, hydrographie, géologie, pédologie, précipitations, forêts...). Leur croisement amènera à des découpages du territoire sur tel ou tel aspect des écosystèmes à étudier, et le SIG permettra de les localiser et d'en affiner la représentation ;
- pour le couplage SIG – BDCC et les études statistiques qui seront opérées.

Cependant, cette base doit rester évolutive : sa structure doit pouvoir être modifiée et des outils supplémentaires y être ajoutés si nécessaire.

Le travail important accompli nécessite encore des améliorations si on veut à l'avenir être totalement efficace dans la gestion actuelle des données de la base. Il y a lieu de :

- procéder à un générateur de requête sous Map-info pour pouvoir importer les données dans les nouvelles tables créées (roches, sols, eaux...) à partir du SIG ;
- modifier le niveau d'importation du niveau 2 de façon, d'une part, à importer facilement n'importe quelle autre table (évolution future pour les programmes charges critiques métaux lourds...) et, d'autre part, à faire une importation simplifiée des données dans chaque table du niveau 2, celles-ci étant indépendantes. Cela implique la création d'un module particulier ;
- pour une importation simplifiée des données, il y a nécessité de créer des champs spécifiques pour chaque table ;
- ensuite, il y aura lieu d'importer et de gérer de nouvelles données issues de la « littérature scientifique ».

Enfin, de nombreuses « sorties cartographiques » sur SIG seront alors possibles en fonction des nouvelles données existantes.

Cette dernière partie est indispensable à l'élaboration de cartes de charges critiques plus précises reposant sur des bases scientifiques solides. Elle nécessite également la visualisation via le lien dynamique avec un SIG.

La BDCC est un système complet permettant d'obtenir automatiquement les valeurs de charges critiques (selon les paramètres choisis par l'utilisateur). Cependant, cet aboutissement vise le long terme car de nombreux modèles sont utilisés pour le calcul des charges critiques. Or ces modèles et les nouvelles données doivent encore être importés dans la base, ce qui est un travail important nécessitant une analyse fine. Ainsi, il est évident qu'une évolution est aujourd'hui indispensable pour rendre la base « opérationnelle en terme d'échéance RIVM » pour :

- stocker dans les nouvelles tables dédiées les nouvelles données scientifiques originales servant à alimenter et à faire évoluer les paramètres calculés de la première partie de la base ;
- croiser les nouvelles couches de données acquises, paramétrer dans la base les modèles en fonction des besoins de calcul ;
- enfin, fournir les nouvelles cartes associées de calcul des charges critiques.

En particulier, il s'agit de faire évoluer la base dans le but d'appliquer les modélisations dynamiques de calcul pour l'acidité et introduire les nouvelles tables et modèles pour permettre la cartographie des charges critiques en métaux lourds, ce qui constitue la priorité actuelle au niveau européen.

Cette volonté d'amélioration qualitative implique, d'une part, la valorisation des données de la base via l'association avec un SIG et, d'autre part, une montée en puissance de la base tant dans sa structure (nouvelles tables et nouveaux modèles que ce soit pour l'acidité, l'azote ou les métaux lourds) que dans son alimentation en données sur l'ensemble des paramètres de l'environnement nécessaires.

Ceci montre qu'une interface constante est nécessaire pour la pérennité et l'évolution d'une telle base entre une personne qualifiée pour la gestion informatique et cartographique de l'outil et les chercheurs compétents impliqués dans les programmes « charges critiques ». C'est à cette condition essentielle que cette base peut être opérationnelle et productrice de données en temps réel pour la France à l'échelle européenne.

Remerciements

Les travaux sur les charges critiques ont pour l'essentiel été financés par l'ADEME et le ministère de l'Environnement.

Mots clés

Charges critiques. Base de données. SGBDR. Oracle. Système d'information géographique, SIG. Map-Info.

Keywords

Critical load. Database. RDBMS. Oracle. Geographic Information System, GIS. Map-Info.

Références

1. Party JP, Probst A, Dambrine E. Détermination et cartographie des charges critiques en polluants atmosphériques dans les Vosges. Rapport scientifique ADEME, année 1992, mars 1993 : 68 p. + annexes.
2. Party JP, Probst A, Dambrine E, Thomas AL. Détermination et cartographie des charges critiques en polluants atmosphériques dans les Vosges. Rapport scientifique définitif, Contrat ADEME n° 3930013, programme « Charges critiques », année 1993/1994, 1995 : 51 p.
3. Party JP. Acidification des sols et des eaux de surface des écosystèmes forestiers français : facteurs, mécanismes et tendances. Taux d'altération sur petits bassins versants silicatés. Application au calcul des charges critiques d'acidité. Thèse de Doctorat. Univ. Louis Pasteur, Strasbourg 1999 : 249 p.
4. Party JP, Probst A, Dambrine E, Thomas AL. Critical loads of acidity to France : sensitivity areas in the north-eastern France. Water, Air and Soil Pollution 1995 ; 85 (1-4) : 2407-12.
5. Février C, Party JP, Probst A. Acidité des eaux de surface et charges critiques dans le massif des Ardennes françaises. CR Acad Sci 1999 ; 328 : 29-35.
6. Hissler C. Charges critiques d'acidité pour les eaux de surface du Massif central. Cas des massifs du Morvan et du Limousin. DEA Physique et chimie de la Terre. ULP Strasbourg 1997 : 32 p.
7. Guillet B, Didon-Lescot JF, Bourguignon S, Nguyen S, Boivent B, Vidal M. Participation au programme Charges critiques. L'évaluation des charges critiques acides sur les bassins versants du Mont-Lozère et extension à la région sud cévenole. Rapport de fin de contrat de recherche, Dpt des Sciences de la Terre, Univ. Orléans 1997 : 42 p. + annexes.
8. Party JP, Probst A, Clément A, Kreyer E. Calcul des charges critiques acides et azotées. Établissement d'une structure de base de données à l'échelle de la France. Rapport scientifique intermédiaire, ministère de l'Environnement, novembre 1997 : 23 p.
9. Party JP, Probst A, Thomas AL, Dambrine E, Février C, King D. Les charges critiques en France. Impacts et conséquences à long terme des polluants atmosphériques sur les écosystèmes naturels terrestres et aquatiques. Rapport scientifique définitif, ADEME 1997 : 59 p.
10. Party JP, Probst A, Thomas AL, Dambrine E. Charges critiques en polluants atmosphériques en France : écosystèmes vulnérables, écosystèmes menacés. Les Cahiers du DSF, 1-1999, [La santé des forêts (France) en 1998], Min. Agri. Pêche (DERF), Paris 1999 : 80-4.
11. Badeau V, Party JP, Probst A, Dambrine E. Étude de la faisabilité d'une base de données charges critiques pour la France. Rapport ADEME, SOL-CONSEIL, CGS-CNRS Strasbourg, INRA 1997 : 33 p.
12. Probst A, Ertz O, Party JP, Clément A. Base de données charges critiques : extension et acquisition des données, lien avec un SIG. CGS-CNRS, ADEME, Rapport définitif, 2000 : 21 p.
13. Party JP, Probst A, Thomas AL, Dambrine E. Charges critiques d'acidité en polluants atmosphériques en France : conséquences vis-à-vis des sols et des peuplements forestiers. Pollution Atmosphérique 2001 ; 172 : 519-27.

14. Party JP, Probst A, Thomas AL, Dambrine E. Calcul et cartographie des charges critiques azotées en France : application de la méthode empirique. Pollution Atmosphérique 2001 ; 172 : 531-44.
15. Ponnette Q, Ulrich E, Brêthes A, Bonneau M, Lanier M. RENECOFOR – Chimie des sols dans les 102 peuplements du réseau, ONF, Dépt des recherches techniques 1997 : 427 p.
16. Party JP, Probst A, Février C. Calcul des charges critiques d'acidité et d'azote sur le massif des Vosges ; extension au territoire français. Dossier documentaire du séminaire « Pollution de l'air et charges critiques pour les sols, l'eau et la végétation », ADEME, Paris 19 novembre 1996 : 41-62.
17. Party JP, Probst A. Calcul des charges critiques acides et azotées. Établissement d'une structure de base de données à l'échelle française. Rapport scientifique définitif, ADEME, contrat 9793016, décembre 1998 : 93 p.
18. Probst A, Party JP. National focal center report : France. In : Calculation and Mapping of critical thresholds in Europe. Status report 1999, Posch M, De Smet PAM, Hettelingh JP, Downing RJ (eds), RIVM report N° 259101009, ISBN N° 90-6960-083-8, 86-90, 1999 : 165 p.
19. Dambrine E, Probst A, Party JP. Détermination et cartographie des charges critiques de polluants atmosphériques pour les écosystèmes naturels, en particulier forestiers. Pollution atmosphérique, Numéro spécial « Charges critiques » 1993 : 21-8.
20. Thomas AL. Analyse spatiale de la sensibilité à l'acidification des écosystèmes du massif gréseux vosgien, Thèse de Doctorat, ENGREF-INRA Orléans 1998 : 135 p.
21. Thomas AL, King D, Dambrine E, Party JP, Probst A. A spatial study of the relationships between streamwater acidity and geology, soils and relief (Vosges, north-eastern France). J Hydrology 1999 ; 217 : 35-45.
22. Rouet P. Les données dans les systèmes d'information géographique, Éditions Hermes 1991 : 278 p.