

ANNEXE 2 : Extraction des unités fonctionnelles du modèle

Cette annexe est composée de 3 parties :

- l'objectif de l'annexe 2A est de rappeler les concepts liés aux Systèmes d'Information Géographique et de montrer les intérêts et les limites de leur couplage avec les modèles hydrologiques ;
- l'annexe 2B résume la procédure d'extraction des sous-unités développée sous le SIG Arc-Info;
- l'annexe 2C permet d'illustrer la méthode de correction du MNT mise en œuvre pour améliorer la qualité des réseaux hydrographiques extraits.

2A : Concepts liés aux Systèmes d'information Géographique et intérêt du couplage « SIG - Modèle hydrologique/qualité de l'eau »

2A-1. Définition du SIG

La définition suivante est classiquement utilisée pour décrire ce qu'est un SIG : "c'est un système capable de stocker, partager, consulter et manipuler les objets représentés sur les cartes et les plans avec leurs descriptions géométriques, ainsi que toute l'information qui leur est attachée". (LAURINI et MILLERET RAPPORT, 1993).

Souvent réduit au seul logiciel, l'outil est alors privilégié au détriment des objectifs. L'excès inverse tend à vouloir remplacer des disciplines telle la géographie ou la cartographie.

Pour se démarquer de ce concept flou et pour le compléter, de nouveaux termes apparaissent :

- le SIT (Systèmes d'Information Territorial) (BEDARD, 1987) ou le SIE (Système d'Information pour Environnement) pour dénoter la finalité ;
- géomatique ou géo-spatiale pour définir une nouvelle science ;
- Information Géographique pour remettre en avant l'élément fondamental du système ;
- le Système d'Information à Référence Spatiale (SIRS) au sens premier du Système d'Information (SI) avec une problématique de structuration des données, d'organisation des moyens et des ressources humaines, de développement d'outils sans chercher à s'appropriier des champs disciplinaires connexes telle la géographie (PRELAZ DROUX, 1995).

2A-2. Les concepts associés aux SIG

2A-21. Les bases de données

Une base de données est une collection de données non-redondantes représentant les besoins d'un ensemble d'utilisateurs (LAURINI et THOMPSON, 1992). Une base de données de Système d'Information Géographique est un ensemble structuré de données graphiques et non-graphiques permettant de décrire un plan d'information, de définir des relations spatiales et des caractéristiques d'un site donné (ANTENUCCI et al., 1991). La construction de la base de données dans un SIG est l'étape la plus onéreuse, la plus longue et celle qui pose le plus de problèmes d'intégration au sein du SIG. Les anglo-saxons distinguent le terme de *data base* en deux mots qui désigne l'ensemble de données numériques, structurées qui peuvent être développées, mises à jour et extraites pour des applications variées. Le terme *database* en un seul mot correspond à un ensemble de données structurées au sein d'un logiciel de gestion des systèmes de bases de données (ANTENUCCI et al., 1991).

Le volume d'information contenu au sein d'une base de données nécessite une organisation qui optimise les traitements ultérieurs des données. Les informations sont classiquement organisées en tableaux qui contiennent les caractéristiques de chaque entité. Il existe quatre structures d'organisation des données, apparues chronologiquement : la structure hiérarchique, la structure en réseau, la structure relationnelle et la structure orientée objet.

Ces quatre structures sont présentées ci-dessous avec leurs contraintes et avantages principaux (cf. Fig. 2A-1).

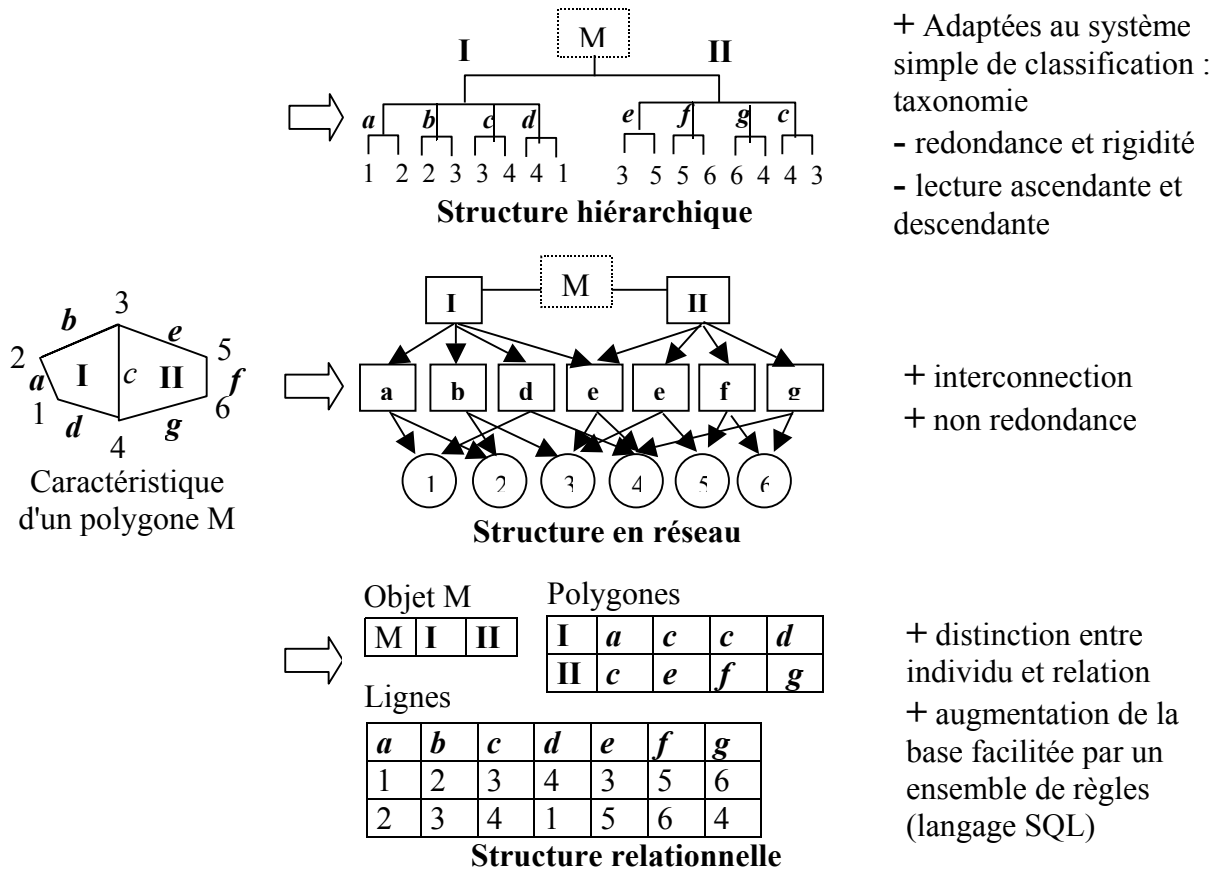


Figure 2A-1 : Structure hiérarchique, en réseau et relationnelle d'après (BURROUGH et MCDONNEL, 1998)

La quatrième structure, orientée objet, utilise le formalisme du langage de programmation orientée objet (JACOBSEN et al., 1992). L'utilisation de cette structure permet de combiner la rapidité des approches hiérarchiques et de types réseaux avec la flexibilité des approches relationnelles (BURROUGH et MCDONNEL, 1998).

En conclusion, l'approche hiérarchique convient aux bases de données où l'on souhaite segmenter l'information spatiale en différents thèmes de gestion. La structure en réseaux est adaptée à la gestion des liens entre les entités "lignes" et "polygones" contenues dans la base. L'approche relationnelle est utilisée préférentiellement pour extraire des informations selon leurs caractéristiques ou pour créer de nouveaux attributs aux éléments présents dans la base. Enfin, la structure orientée objet est requise pour les bases où l'on souhaite privilégier les interactions entre objets (BURROUGH et MCDONNEL, 1998).

Les recherches actuelles en matière de base de données dans les SIG visent à obtenir des structures de plus en plus flexibles qui permettent de prendre en compte l'évolution spatiale et temporelle des entités contenues dans les bases de données (PARENT, 2000).

2A-22 Types de données manipulées au sein d'un SIG

Les SIG se distinguent donc des systèmes de gestion de bases de données classiques par le fait qu'ils intègrent des données graphiques et non-graphiques (figure 2A-2) :

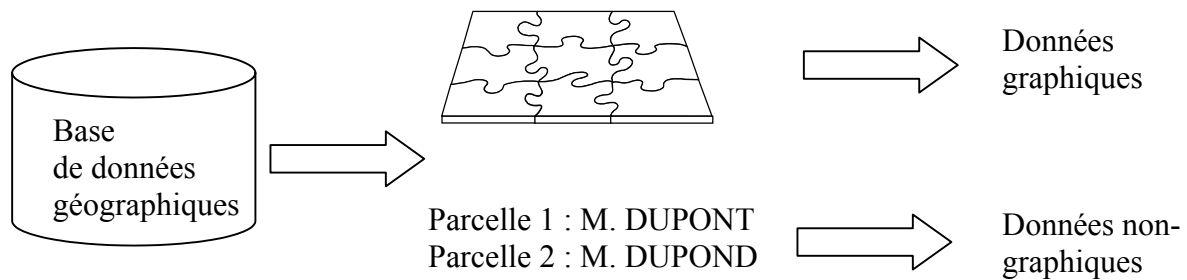


Figure 2A-2 : Caractéristiques d'une base de données d'un SIG

Les données graphiques contiennent des éléments cartographiques sous forme numérique. Les données graphiques sont représentées grâce à 5 types d'éléments : le point, la ligne, la surface, le pixel et l'annotation (BURROUGH, 1986). Les données non-graphiques contiennent les caractéristiques et les qualités des données graphiques et leurs relations (ANTENUCCI et al., 1991).

2A-23. Mode de représentation des données à référence spatiale

Le mode de représentation des données à référence spatiale (ou données graphiques) implique la saisie, le codage, le stockage de l'information et influe sur le type et l'efficacité des traitements associés.

Six modes de représentation peuvent être définis (GOODCHILD, 1993a) :

- **le semis de points irréguliers** : la base de donnée contient alors un ensemble de points localisés en x, y et z qui renseignent une variable (exemple des stations météorologiques) ;
- **le semis de points réguliers** : les éléments considérés sont ici répartis régulièrement, sur une grille carrée ou rectangulaire (cas des Modèles Numériques de Terrain) ;
- **la ligne** : la base de données contient une suite de lignes (définies par des couples x, y), caractérisées par une valeur z donnée (cas d'un réseau hydrographique) ;
- **le polygone** : surface fermée par des arcs, tout point de l'espace n'appartient qu'à un seul polygone (limites de sous-bassins versants) ;
- **le maillage rectangulaire ou carré régulier** : l'espace est divisé en une grille régulière, chaque cellule de la grille est caractérisée par une valeur unique (images satellitales) ;

- **le maillage triangulaire irrégulier** : l'espace est divisé en une suite de triangles irréguliers possédant chacun une valeur unique (mode de représentation possible de Modèle Numérique de Terrain).

Le terme anglo-saxon de **raster** regroupe le semis de points réguliers et le maillage rectangulaire ou carré. Les autres modes de représentation correspondent au mode **vecteur**. Les trois premiers modes ne permettent pas de représenter une variable en tout point de l'espace. Une interpolation est alors nécessaire (GOODCHILD, 1993a).

Le choix du mode de représentation (vecteur ou raster) va conditionner le type de traitements ultérieurs au sein du SIG et nécessite donc une réflexion sur l'adéquation du mode et de la donnée représentée.

Les avantages et contraintes des deux modes sont résumés ci-dessous (BURROUGH et MCDONNEL, 1998) (tableau 2A-1) :

	Vecteur	Raster
Avantages	<ul style="list-style-type: none"> - bonne représentation de la structure des données - structure compacte - topologie explicite permettant une bonne analyse en réseau - projection géographique et translation aisée - qualité de représentation aux différentes échelles - l'extraction, la mise-à-jour et la généralisation des données sont possibles 	<ul style="list-style-type: none"> - structure simple de données - accès aux attributs simplifiés - facilité d'analyse spatiale (filtre) - modélisation simplifiée du fait de la forme régulière des cellules - technologie bon marché - supports idéaux pour de nombreux types de données
Inconvénients	<ul style="list-style-type: none"> - structures complexes de données - intersection nécessitant puissance de calcul - visualisation et impression nécessitant important temps d'affichage - contenu homogène de chaque polygone rendant l'analyse spatiale impossible - simulation difficile car chaque polygone a une forme propre 	<ul style="list-style-type: none"> - volume des données - perte d'information due au choix des cellules - moindre qualité esthétique - projections géographiques nécessitant algorithmes adaptés pour éviter distorsion de la grille

Tableau 2A-1 : Avantages et contraintes des deux modes de représentation vecteur-raster (BURROUGH et MCDONNEL, 1998)

Le choix du mode de représentation dépend donc de la nature de la donnée et du type d'analyse à mener. Ainsi, la propagation d'une information dans un réseau se traitera plus facilement dans un mode vecteur. Par contre, une étude surfacique se fera préférentiellement en mode raster. Le raster serait ainsi plus adapté à l'interprétation du **où** et le vecteur à l'interprétation du **pourquoi** (TOMLIN, 1990).

Les fonctionnalités des SIG permettent actuellement de manipuler à la fois des données de vecteur et raster. La conversion d'un mode à l'autre est réalisable (cf. Fig. A2-3) :

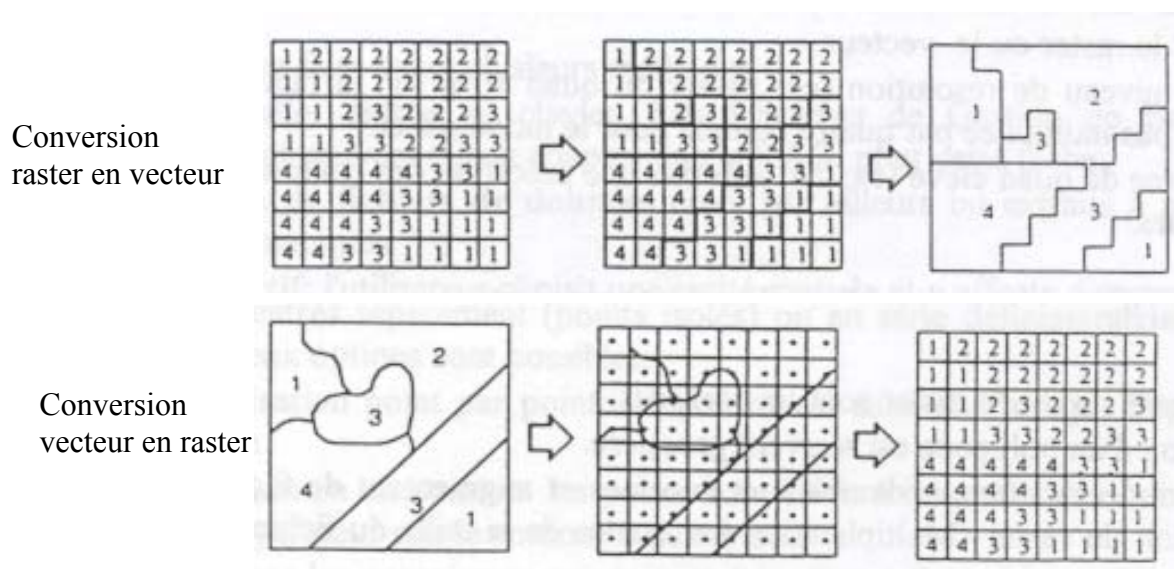


Figure 2A-3 : conversion raster-vecteur (BARKER, 1988)

La conversion du mode raster à celui de vecteur implique l'utilisation d'algorithmes de décision complexes. L'aspect cartographique des informations obtenues en vecteur présente le même défaut que les données en mode raster (aspect crénelé des contours) (BARKER, 1988). La conversion du mode vecteur vers le mode raster nécessite des algorithmes de décision simples mais induit une perte de précision géographique (ANTENUCCI et al., 1991).

En conclusion, le mode de représentation des données est lié au domaine d'étude et à la nature des phénomènes analysés. Dans une problématique de gestion des ressources en eau, les variables contrôlant les écoulements (géologie, topographie) sont le plus souvent continues dans l'espace. Le mode raster est donc le mieux adapté pour prendre en compte ce continuum (VIEUX et al., 1990). Toutefois, l'utilisation d'outil permettant de travailler simultanément avec les deux modes est un avantage. En effet, le recours au mode vecteur est performant pour la modélisation des réseaux hydrographiques (BRUNSTEIN, 1996).

2A-24. Système de référence de localisation

La représentation des données graphiques implique de préciser le système de localisation. On distingue un système mathématique (ou direct) et textuel (ou indirect), qui est le plus souvent administratif (adresse postale ou cadastrale).

Le système mathématique consiste à déterminer la position d'un point sur la terre. La géodésie, science qui analyse la forme de la terre (DONNAY, 1994), propose ainsi plusieurs systèmes de projection. Le système de Mercator est universellement accepté (figure 4). Chaque point du globe est ainsi identifié par sa latitude et sa longitude. Le système Lambert est lui classiquement utilisé en France (figure 2A-4). Il est lui-même composé de 4 sous-systèmes I, II, III et IV selon la zone considérée du Nord au sud de la métropole et permet de localiser un point en x et y.

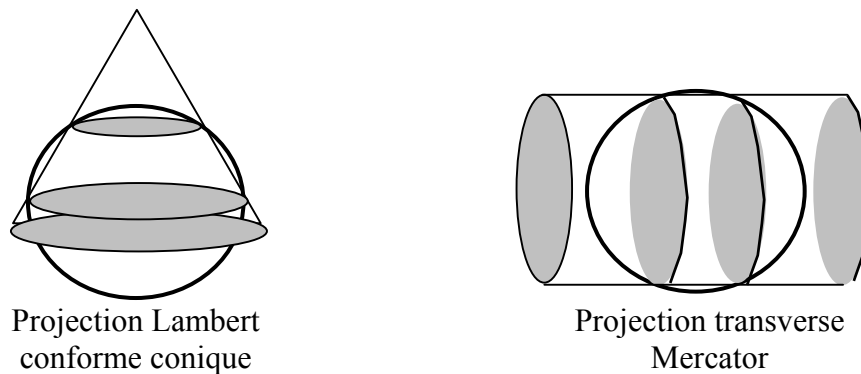


Figure 2A-4 : Système de projection Lambert et Mercator

Le système textuel est classiquement représenté par le code postal qui ne permet pas une représentation cartographique. La parcelle cadastrale permet le lien entre système direct et indirect car chaque parcelle est identifiée par un numéro et par une surface qui peut être localisée en x et y.

2A-25. Les grandes fonctions d'analyse des données à référence spatiale

La finalité des SIG est de permettre la sélection, l'extraction et l'analyse des données à référence spatiale (BURROUGH et MCDONNEL, 1998). Les paragraphes suivants illustrent les fonctions d'analyse qui permettent de générer de nouvelles informations par des opérateurs logiques (ET, OU, SAUF, SI) et des fonctions mathématiques.

Ces fonctions peuvent être classées en trois types d'opérations en partant du principe que chaque opération se rapporte à une seule localisation même si plusieurs couches d'information sont utilisées (TOMLIN, 1990) :

- les **fonctions locales** : chaque nouvelle valeur est le résultat d'un calcul sur une ou plusieurs valeurs existantes sur cette même localisation (superposition de plusieurs couches d'information);
- les **fonctions focales** : chaque valeur est obtenue par une fonction des localisations voisines selon une proximité donnée;
- les **fonctions aréales** : chaque valeur est obtenue par une fonction des localisations voisines selon un zonage donné.

Chaque type d'opérations correspond à des problématiques diverses.

On distingue parmi les fonctions locales, celles qui requièrent une seule information (altitude), de celles qui nécessitent plusieurs thèmes (occupation du sol et type de sol).

Les fonctions qui ne font appel qu'à un seul type d'information sont classiquement les opérations de réaffectation (modification de la discrétisation en classes d'aptitude) et les fonctions utilisant une valeur numérique pour modifier l'information (soustraction, addition, division). Ce type d'opération revient par exemple à passer d'une altitude par rapport au

niveau de la mer à une hauteur par rapport à un niveau de base en soustrayant localement l'altitude du nouveau niveau de base. Les fonctions trigonométriques peuvent être utilisées pour représenter les variations cycliques dans le temps.

Les opérations nécessitant plusieurs thèmes d'information vont permettre d'effectuer des combinaisons locales entre couches d'information (calcul du minimum local). Ce type de traitement permet par exemple de générer des cartes de risques en croisant des facteurs de vulnérabilité différents.

On distingue au sein des fonctions focales, celles qui impliquent un voisinage immédiat ou étendu. Avec un voisinage immédiat, chaque valeur du pixel de la grille (mode raster) va être fonction des quatre ou huit localisations adjacentes. Ces traitements concernent le filtrage de l'information (lissage, extraction de motifs homogènes).

Avec un voisinage étendu, la fenêtre d'analyse est étendue et une direction peut être définie. Dans l'hypothèse de la recherche de zones vulnérables aux fumées d'une industrie, la direction des vents dominants peut être prise en compte (LAURENT, 1996).

Ces fonctions permettent d'étudier la propagation d'une information en fonction de la distance. Celle-ci peut être la distance euclidienne ou fonction d'un coût de propagation. C'est le cas du chemin de l'eau qui suit la direction de plus grande pente.

Dans les fonctions aréales, la valeur de chaque localisation est fonction d'une valeur associée à la zone ou à l'aire contenant cette localisation. Cette zone peut être définie comme un ensemble de localisations pas forcément contiguës. Ce type de fonction permet ainsi de calculer des statistiques d'une couche d'information (occupation du sol) en fonction d'un zonage défini dans une autre couche d'information (limites de sous-bassins versants).

En conclusion, les opérations de type local ne s'intéressent pas aux relations de voisinage et peuvent donc être réalisées dans un tableur ou par un système de gestion de base de données classique (TOMLIN, 1990). Par contre, les relations de voisinage (opérations focales et aréales) ne peuvent être étudiées que grâce à des systèmes de traitement où le système de référence de localisation est explicitement codé (SIG, logiciel de télédétection, d'analyse d'image, etc ...) (LAURENT, 1996).

La classification proposée ci-dessus par TOMLIN concerne exclusivement les données en mode raster (TOMLIN, 1990). Le classement suivant permet d'intégrer les traitements classiques utilisés à la fois pour les données en mode vecteur et raster (BURROUGH et MCDONNEL, 1998). Il distingue les fonctions d'analyse de données discrètes et de données continues dans l'espace.

Le mode d'analyse des données discrètes préserve les propriétés des objets manipulés : forme, localisation et topologie. Seules les opérations de reclassement peuvent conduire à des fusions d'objets qui présentent les mêmes attributs après traitement. Ce mode de traitement en mode raster est à rapprocher des opérations locales définies précédemment. En mode vecteur, l'utilisation d'opérateurs logiques (ET, OU, SAUF) permet de générer de nouvelles informations par croisement de différents thèmes. Comme pour le mode raster, ce type de traitement permet de générer des cartes de risque en croisant les différents facteurs de vulnérabilité. La particularité de cette opération en mode vecteur réside dans la prolifération de petits polygones sans existence réelle. La détermination d'une taille minimum d'objets est

alors nécessaire pour conserver une information pertinente et éliminer les artefacts liés à ce type d'opération.

Le deuxième type de traitements concerne les opérations qui impliquent une distance entre deux informations. Le terme anglo-saxon pour ce type d'opération est le "buffering". Un nouvel attribut va être défini pour chaque objet selon son appartenance à une zone spécifiée. Ce type de traitement est classiquement utilisé pour définir les zones constructibles en fonction de la proximité d'une rivière.

Le dernier type d'opération de données discrètes, consiste en l'analyse de connectivité. Ce type de traitement concerne les données en mode vecteur et est à rapprocher des analyses de réseau. Les opérations classiquement utilisées consistent à définir le parcours d'un réseau d'arcs qui répond à des critères donnés. C'est par exemple la détermination du chemin le plus sûr (ou le plus court) d'un réseau routier du point A au point B en fonction des attributs de chaque tronçon de route.

En conclusion, ces fonctions permettent donc de manipuler des données discrètes dans l'espace. En mode raster, ces fonctions sont définies comme locales (TOMLIN, 1990). En mode vecteur, elles permettent de répondre aux questions : à quelle distance tel objet est situé d'une ressource ou d'un danger potentiel et quel est le chemin optimal pour parcourir un réseau?

Les fonctions d'analyse de données continues concernent exclusivement les données raster (BURROUGH et MCDONNEL, 1998). Ces opérations concernent les fonctions focales et aréales précédemment définies. L'utilisation de fenêtres d'analyse permet de caractériser chaque pixel comme étant une fonction des pixels voisins. Ce type d'opérations est utilisé pour extraire des informations dérivées (pente, convexité, directions d'écoulement) d'un modèle numérique de terrain. Des opérations similaires sur des données raster ont été définies par les recherches sur les automates cellulaires ou chaque pixel de la grille est en relation avec ses voisins (LANGLOIS et PHIPPS, 1997).

Les opérations d'analyse de données sont intégrées de façon plus ou moins exhaustives au sein des SIG selon le mode de données manipulées (raster-vecteur) et leur degré de complexité (ANTENUCCI et al., 1991). La difficulté de l'utilisation d'un SIG dans une thématique donnée comme la gestion de la qualité des eaux de surface consiste à combiner ces fonctions pour apporter l'information adéquate.

2A-26. Les méthodes d'analyse spatiale

L'analyse de données à référence spatiale nécessite préalablement de définir la finalité de ces traitements. TOMLIN distingue ainsi l'approche analytique et l'approche synthétique en analyse cartographique (TOMLIN, 1990). L'approche analytique vise à décomposer les données en niveaux de plus en plus fins en terme de signification. L'approche synthétique consiste à recomposer les données pour les utiliser dans d'autres contextes.

L'auteur propose aussi une distinction entre l'analyse spatiale descriptive et prescriptive (TOMLIN, 1990). Ces deux types d'analyse vont être fonction du type de question posée. L'analyse descriptive vise à déterminer la probabilité d'apparition d'un phénomène (intensité de l'érosion en fonction de facteurs défavorisants). L'analyse prescriptive consiste à répondre aux questions de localisation optimale (où dois-je implanter un pluviomètre pour contrôler la pluviométrie d'un bassin versant donné).

2A-3. Précision et qualité au sein du SIG

Ce paragraphe vise à présenter les concepts d'erreur et de précision ainsi que leur prise en compte dans les bases de données à référence géographique. L'impact des erreurs sur la qualité de l'analyse spatiale et de la modélisation est examiné. Les erreurs contenues dans la base de données proviennent de manipulations hasardeuses mais sont aussi dues aux choix du type de données et aux outils de traitements (BURROUGH et MCDONNEL, 1998).

L'erreur est définie comme une chose fautive, erronée par rapport à une norme (différence par rapport à un modèle ou au réel). L'estimation de l'erreur se fait ainsi par différence entre la valeur exacte d'une grandeur et sa valeur calculée ou mesurée. Du point de vue des bases de données géographiques, l'erreur peut être définie comme "la non concordance entre les objets encodés dans la base de données géographiques et les entités du monde réel que ces objets sont supposés représenter" (VEREGIN, 1995). Les termes liés au concept d'erreur, dans les bases de données à référence géographique, peuvent être définis (GOODCHILD, 1993a) :

- **exactitude** : relation entre une mesure et la réalité qu'elle se propose de représenter ;
- **précision** : degré de détail dans le stockage ou dans la manipulation de la mesure lors des calculs ;
- **résolution** : taille réelle du plus petit objet qui peut être discerné sur une carte ;
- **échelle** : notion intimement liée à la résolution puisqu'il y a une limite physique au dessin d'un objet sur une carte, cette limite étant jugée en général à 0.5 mm (ce qui correspond par exemple à 12.5 m pour une carte au 1/25 000^{ième}).

La manipulation de données à référence géographique au sein des SIG implique l'évaluation de la qualité de ces dernières. Plusieurs sources d'erreurs, pouvant altérer la qualité de la base de données, sont distinguées (BURROUGH et MCDONNEL, 1998) :

- **la justesse du contenu** : on distingue la justesse du contenu pour des données thématiques (une parcelle est considérée non cultivée alors qu'elle est cultivée en maïs) ou des données continues (mauvaise altitude pour un sommet de montagne) ;
- **l'erreur de mesures** : celle-ci peut résulter de mauvaises observations ;
- **le facteur humain** : il intervient lorsqu'une interprétation est effectuée (analyse de photos aériennes) ;
- **les erreurs de laboratoire** : elles comprennent les erreurs de manipulation et d'interprétation des analyses de terrain qui conduisent à l'information cartographique ;
- **la justesse de la localisation** : primordiale au sein des SIG, elle peut être affectée par le support utilisé (données scannées à partir d'une carte déformée) le type d'information (les données cadastrales sont obtenues avec une meilleure justesse qu'un découpage en zone de végétation) ;
- **la variabilité spatiale de l'information** : les cartes thématiques segmentent artificiellement le milieu (par nature hétérogène) en une suite d'éléments supposés homogènes.

De même que les erreurs précitées peuvent altérer la qualité de la base de données, l'intégrité de la base de données peut être mise en cause par divers facteurs (BURROUGH et MCDONNEL, 1998) :

- **L'âge des données** : la mise en place de SIRS implique souvent l'utilisation de données dont l'acquisition n'est pas concomitante. Or excepté le domaine géologique, la validité des informations diminue avec le temps ;
- **L'uniformité d'acquisition** : il est théoriquement préférable d'obtenir une couverture homogène (en terme de nomenclatures et de protocoles) d'un thème donné. L'acquisition sur plusieurs années d'une couverture thématique peut conduire à des changements de protocoles d'acquisition de l'information ;
- **L'échelle et la résolution** : il est important de disposer d'une échelle adéquate pour traiter un problème donné (une grande échelle conduit à une surcharge d'information alors qu'une petite échelle ne fournit pas suffisamment d'information) ;
- **la densité des observations** : le plan d'échantillonnage doit être pertinent pour prendre en compte la variabilité du milieu étudié ;
- **la pertinence** : l'utilisation d'information indirecte (image satellitale) pour obtenir une classification thématique (occupation du sol) implique de juger de la pertinence du support d'information utilisé ;
- **le format des données** : l'intégrité d'une base de données est fonction du type de support de la donnée (supports numériques divers qui nécessitent des lecteurs adaptés), de la structure de la donnée (topologie), de l'exportation et importation de la donnée (propre ou non à chaque logiciel utilisé) et enfin du format des données non graphiques (système de référence de localisation : échelle, système de projection; et format des données) ;
- **accessibilité** : certaines données peuvent relever du secret militaire; la multiplication de bases de données mondiales permet d'obtenir des informations sur des pays où l'information locale est limitée ;
- **le coût et la propriété** : l'acquisition de certaines données peut engendrer un coût important (digitalisation); la manipulation de données implique souvent la régularisation pour l'utilisateur auprès des services à l'origine de la donnée (copyright) ;
- **l'erreur numérique** : ce point met en exergue la capacité des logiciels à manipuler les données avec le niveau de précision requis (erreur d'arrondis, stockage de données, précision dans le système de coordonnées géographiques).

La manipulation d'objets (parcelles, zones "homogènes" d'occupation du sol, limites de bassins et positions de réseaux hydrographiques) pose le problème de la notion de limite. Chacun de ces objets est représenté dans la base de données avec précision mais pas forcément avec justesse. Concernant la question des limites des "objets" manipulés, on peut distinguer les erreurs dues aux conversions vecteur-raster d'une part et les erreurs liées à la digitalisation et aux intersections d'objets d'autre part (BURROUGH et MCDONNEL, 1998).

Le passage de la représentation des données du mode vecteur au mode raster implique une perte systématique d'information (ANTENUCCI et al., 1991). Plus la résolution est fine plus la perte est moindre mais plus le volume des données est important. Ce passage du mode vecteur au mode raster conduit à l'existence de pixels mixtes qui recouvrent plusieurs objets aux attributs différents. Des approches statistiques ont été proposées pour étudier la qualité de la conversion des données d'un mode de représentation à l'autre (GOODCHILD, 1980).

L'obtention par digitalisation est une source supplémentaire d'erreurs dans la base de données. La première source d'erreurs est la carte qui sert de support à la digitalisation . Les sources d'erreurs sont celles définies ci-dessus. La deuxième source d'erreurs peut provenir de la qualité de l'opération de digitalisation (KEMP, 1993).

Enfin les opérations d'intersection entre données en mode vecteur et/ou raster peuvent être à l'origine d'erreurs pour les données issues de ces croisements. La notion de limite est ici au centre du problème. L'existence d'une zone d'incertitude autour d'une ligne a été proposée pour pondérer les résultats d'intersection (BLAKEMORE, 1984). Concernant la problématique d'inclusion d'un point dans un polygone il est ainsi possible de distinguer cinq cas de figure : le point est assurément inclus, assurément exclu, probablement inclus mais pouvant être exclu, probablement exclu mais pouvant être inclus ou bien de position indéfinie.

L'intersection de plusieurs couches d'information en mode vecteur provoque l'apparition de polygones sans signification thématique. Le nombre de ces petits polygones "parasites" semble plus lié à la complexité de la géométrie des objets qu'au nombre d'objets croisés (GOODCHILD, 1978). Cet auteur montre que le nombre de polygones "parasites" est plus important si les sources d'information sont corrélées. C'est ainsi le cas d'un croisement de couches résultant de plusieurs digitalisations à partir d'un même support.

En conclusion, la qualité d'une base de données géographiques est fonction de l'existence d'erreurs à trois niveaux (PURICELLI et al., 1997) :

- niveau de gestion des données (structure des données) .
- niveau de la géométrie des données (modèle de représentation des données) .
- niveau sémantique (signification des objets).

La présence de ces erreurs est à l'origine de la propagation des erreurs dans les opérations d'analyse des données à référence spatiale.

2A-4. Intérêt de l'intégration « SIG - Modèle hydrologique/qualité de l'eau »

La description et le contrôle de l'occupation du sol, des pollutions diffuses et des migrations de nutriments vers le milieu aquatique sont devenus de véritables axes de recherche (GRAYSON et al., 1993).

Cette évolution entraîne une prise en compte accrue de la distribution spatiale des phénomènes. Dans ce contexte, l'utilisation des modèles de qualité d'une part et des SIG d'autre part, apparaît fortement complémentaire. La liaison d'un modèle et d'un SIG peut ainsi permettre d'améliorer les sous-unités en nombre et en définition. En effet, "le facteur limitant le plus la modélisation hydrologique est non pas la caractérisation mathématique des processus physiques mais l'aptitude à spécifier les valeurs des paramètres du modèle" (MAIDMENT, 1993a).

Plusieurs niveaux d'utilisation des SIG dans le cadre de modélisation hydrologique peuvent être distingués. MAIDMENT en distingue quatre de complexité croissante (MAIDMENT, 1993a) :

- Utilisation de SIG comme outil de visualisation des résultats du modèle : Le SIG permet alors d'améliorer la compréhension et la diffusion des résultats issus du modèle (REJESKI, 1993). Les SIG sont ainsi utiles pour cartographier les risques de contamination des aquifères comme dans le modèle américain DRASTIC (ENGEL et al., 1996, NAPOLITANO et FABRI, 1996) ;
- Utilisation de SIG comme outil d'évaluation des paramètres d'entrées du modèle :
Le SIG est utilisé en amont du modèle pour renseigner les différentes sous-unités spatiales prises en compte par le modèle. Le SIG peut ainsi permettre de définir des zones homogènes en terme de réponses hydrologiques à l'intérieur du bassin ou en terme de flux d'azote produits. Ainsi le zonage en HRU (Hydrologic Responses Units) qui permet de discrétiser le bassin en classe de similarité hydrologique, peut être défini par un SIG (JETON et LARUE-SMITH, 1993). Ces HRU sont obtenues à partir de différentes couches d'information contenues dans le SIG : la pente, l'altitude, la pédologie et la végétation (FLÜGEL, 1996). Concernant la qualité de l'eau, l'application du concept de CHRU (CHEMICAL Response Unit) repose sur un croisement des mêmes couches d'informations (BENDE, 1997). La segmentation en sous-bassins versants peut être réalisée dans l'environnement de programmation d'un SIG (HELLWEGGER et MAIDMENT, 1999).
Pour évaluer les paramètres d'entrée du modèle, le SIG peut être associé à d'autres outils d'analyse. Ainsi, la télédétection peut permettre, sur le bassin versant, de caractériser l'occupation des sols (PUECH et al., 1995) et l'état hydrique de surface des sols (GINESTE, 1998). L'utilisation de modèle numérique de terrain permet d'obtenir une information sur la morphologie du bassin. La pente est une des informations qui conditionnent la localisation des zones contributives de TOPMODEL (BEVEN et WOOD, 1983) ;
- Modélisation au sein d'un SIG :
L'utilisation du SIG dans un but de modélisation est envisageable tant que la variabilité temporelle n'intervient pas. Par exemple, l'écoulement théorique moyen annuel et les charges annuelles en polluants peuvent être estimés sur un bassin versant en utilisant un SIG (LAURENT, 1996). Ce type de modélisation permet de fournir un premier diagnostic sur les sites de production des masses d'azote et de vulnérabilité des biefs associés ;
- Couplage d'un modèle de qualité à un SIG :
Le SIG est utilisé pour déterminer les paramètres d'entrée du modèle (MANGUERRA et ENGEL, 1998) et permet de visualiser les résultats renvoyés par le modèle après traitement des données (PANICONI et al., 1999). Les résultats des modèles basés sur la résolution d'équations aux différences finies ou aux éléments finis peuvent ainsi être visualisés dans l'espace à l'issue des simulations.

2A-5. Limites de l'utilisation conjointe SIG - modèle

L'utilisation conjointe SIG - modèle permet de prendre en compte la variabilité spatiale au sein de la modélisation. Toutefois, un certain nombre de limites est à considérer.

L'utilisation du SIG en aval du modèle de qualité implique de pouvoir quantifier les erreurs des plans d'information utilisés (CHOUDHRY et MORAD, 1998). Les erreurs contenues dans la base de données proviennent de manipulations hasardeuses mais sont aussi dues aux choix du type de données et aux outils de traitement (BURROUGH et MCDONNEL, 1998). Le concept d'erreur dans les bases de données géographiques est développé dans l'annexe 4.

MAIDMENT considère que la critique la plus importante en terme d'utilisation conjointe SIG-modèle hydrologique est la signification des variables (MAIDMENT, 1993a). On peut dissocier ce point de vue en deux aspects distincts :

- Cette critique concerne d'une part le passage d'une donnée ponctuelle (profils de sol) à une donnée surfacique (carte pédologique). A quelles échelles les données expérimentales ont été collectées et à quelles échelles sont utilisés les plans dérivés ?
Une réflexion doit donc être menée pour juger de la signification hydrologique des données collectées au sein du SIG pour une problématique donnée (PUECH et al., 1995) ;
- Elle concerne d'autre part la perte d'information due à la résolution des plans d'information. De nombreux travaux portent ainsi sur la sensibilité des modèles hydrologiques induite par la résolution des données (VIEUX, 1993), (FARAJALLA et VIEUX, 1995, MOLNAR et JULIEN, 2000). L'évaluation de la perte d'information selon la résolution et l'échelle d'acquisition des données peut être menée en utilisant la théorie de l'information (COLOSIMO et MENDICINO, 1996). Les indices d'entropie de Shannon, de dominance et de contagion sont ainsi utilisés pour étudier l'adéquation de l'échelle des données au problème posé (LAURENT, 1996), (SINGH, 1999).

La variabilité des phénomènes pris en compte dans les modèles est une limite à l'utilisation conjointe des SIG et des modèles. En effet, les SIG permettent classiquement de représenter l'environnement par un système à deux dimensions (x et y). Une troisième dimension (l'altitude ou la profondeur) peut être prise en compte par les modèles numériques de terrain (BLASZCZYNSKI, 1993), (MAIDMENT, 1993b). Toutefois, la prise en compte d'une quatrième dimension fait défaut au sein des SIG (MORRIS et al., 2000). En effet, le temps n'est pas représentable sur une couche d'information. Or, l'étude des processus hydrologiques implique la prise en compte de la variable temps pour étudier la dynamique des phénomènes (MAIDMENT, 1993a). Les potentialités de couplage entre modèles hydrologiques et de qualités des eaux et SIG dépendent donc de la structure spatio-temporelle adoptée dans le modèle. TIM montre ainsi que le potentiel d'intégration d'un SIG avec différents modèles est très inégal (TIM, 1995). La création de SIG spécialement adaptés à la modélisation des mécanismes de l'eau semble offrir une meilleure prise en compte de la dynamique des écoulements (RISSONS, 1995).

2A-6. Perspectives de développement des SIG

Les SIG ont connu un développement considérable durant les dix dernières années tant en terme de technologie (ANTENUCCI et al., 1991), de méthodes d'analyse (BURROUGH et MCDONNEL, 1998), et de couplage avec des modèles environnementaux (GOODCHILD, 1993b). La recherche actuelle vise à développer un SIG plus "intelligent" (BURROUGH et MCDONNEL, 1998), plus flexible. Plusieurs voies de recherches sont actuellement explorées :

- **recherches sur l'accessibilité du SIG** : il y a une dizaine d'années, la prise en main d'un SIG nécessitait plusieurs semaines voire plusieurs mois. Certains SIG commerciaux nécessitent encore une phase d'apprentissage importante. La tendance actuelle est au développement de SIG modulaires adaptés à des applications précises (analyse de réseaux, gestion de base de données, traitement d'images et représentation en trois dimensions par exemple). La notion "d'interopérabilité" (accès à des opérateurs divers et variés) est ainsi au centre des recherches et vise à étendre la disponibilité des bases de données géographiques (BERGMANN et al., 2000). Cette volonté d'accessibilité à l'information géographique passe aussi par de nouveaux formats d'échange de données (importation/exportation) (ARIKAWA et KUBOTA, 2000) ;
- **recherches sur la structure des bases de données** : la gestion des bases de données et les questions de précision, d'erreur et de qualité qui s'y rattachent, permettent l'émergence d'un ensemble de normes sur les métadonnées (données sur les données) qui offrent la possibilité aux utilisateurs de juger de la pertinence d'une information par rapport à leur problématique (AFNOR, 1999). Le développement des bases de données se fait actuellement avec une structure orientée objet qui augmente la flexibilité de la base et facilite l'intégration de la dimension temporelle des processus (PARENT, 2000, VOGES et BERMAN, 2000). Cette structure orientée objet est aussi utilisée pour créer des bases de données dites "multi résolution" ou "multi échelle" qui permettent une représentation multiple d'un objet en fonction de l'échelle d'étude (par agrégation, désagrégation, fusion d'objets présents à une échelle donnée) (MAGILLO et DE FLORIANI, 2000, SPACCAPIETRA et al., 2000) ;
- **recherches sur la représentation des objets** : l'analyse des relations entre objets spatiaux (croisement, proximité, contiguïté, connexité) est limitée par une représentation logique des limites d'objets (vrai, je suis à l'intérieur, faux, je suis à l'extérieur). Cette représentation ne tient pas compte de l'incertitude sur les limites d'objets. Les recherches actuelles visent à appliquer la théorie du flou pour décrire un objet tout en intégrant l'incertitude sur ses limites (BURROUGH et MCDONNEL, 1998), (CAZEMIER et al., 1998), (MOLENAAR, 1998).

Les thèmes actuels de recherche concernent donc la modularité, la flexibilité, l'interopérabilité qui visent à ouvrir les bases de données géographiques au plus grand nombre. Les problématiques de qualité des bases de données, de structuration et de manipulation de ces données demeurent.

2B : Méthode d'extraction automatique des unités fonctionnelles (biefs de rivières et sous-bassins versants associés) à partir d'un MNT

2B-1. Définition du Modèle Numérique de Terrain

La recherche des limites d'un bassin versant passe par l'exploitation de données altitudinales. Le développement de l'informatique a progressivement permis de stocker l'information "élévation" sous forme matricielle. Le **Modèle Numérique de Terrain** ou **MNT** offre ainsi pour chaque maille, les coordonnées en x et en y auxquelles est associée la valeur en z. Le terme de pixel est retenu pour qualifier chaque maille du MNT. Ainsi chaque pixel porte une information en z comme le montre la figure 2B-5.

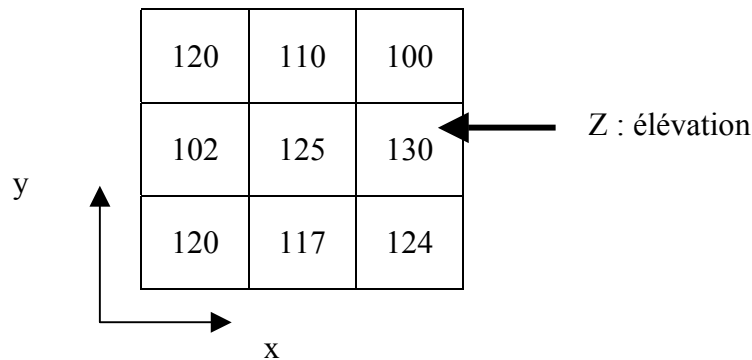


Figure 2B-5 : MNT sous forme maillé ou raster

La précision et la validité des MNT dépendent du mode d'acquisition de l'information, de la variabilité locale de l'altitude et des méthodes d'interpolation utilisées pour générer le MNT. L'évaluation de la qualité du MNT est délicate car elle suppose une référence qui permettrait une comparaison. Or souvent, la référence est un autre MNT lui même perturbé.

Sur les zones d'étude le modèle numérique de terrain provient de la BD Alti® de l'IGN. La précision en z de ce produit se situe entre 1 et 3 m selon ses concepteurs. Le MNT présente une résolution de 50 m. Ceci implique donc que chaque maille de 50 mètres de côté est caractérisée par une altitude moyenne. Son système de projection géographique est le Lambert II étendu.

2B-2. Méthodes classiques d'extraction des rivières et limites de bassins

2B-21. Comblement des points bas

Le MNT brut présente classiquement des points bas desquels l'eau ne peut virtuellement pas s'écouler. Si de telles zones, dites endoréiques, peuvent parfois exister comme dans les contextes karstiques, elles sont souvent dues à des imprécisions du MNT utilisé. Le traitement de ces zones est alors nécessaire pour permettre les écoulements vers l'aval. Une étape de comblement des points bas du MNT est alors classiquement réalisée, comme l'illustre la figure 2B-6.

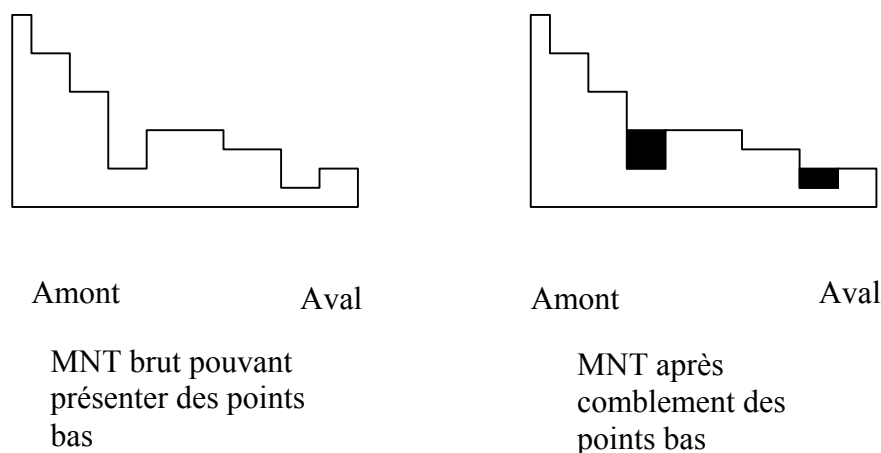


Figure 2B-6 : Traitement préliminaire des points bas du MNT

Une fois ce traitement préliminaire réalisé, le MNT peut être utilisé pour obtenir des informations hydrologiques dérivées.

2B-22. Détermination de la pente

En hydrologie, la connaissance de la pente d'une zone donnée est souvent nécessaire pour comprendre la dynamique des écoulements.

Le critère pente va ainsi permettre d'estimer le risque de ruissellement sur la zone étudiée. La notion de pente doit être associée à une échelle donnée. En effet, la perception de la pente est différente si l'on se place au niveau du m², de la parcelle ou du versant (PUECH, 1996).

La figure 2B-7 met en évidence la nécessité de préciser l'échelle prise en compte pour étudier la pente.

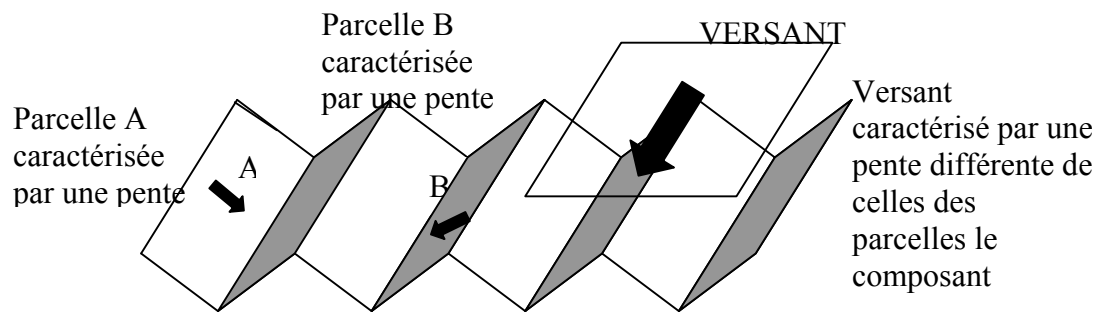


Figure 2B-7 : Variabilité du degré de pente en fonction de l'échelle spatiale prise en compte (de la parcelle au versant).

Le calcul de la pente est obtenu par le passage d'un filtre sur le MNT. Ces filtres sont des fenêtres glissantes dont le centre passe successivement sur chaque maille du MNT et qui y calculent un gradient selon la valeur des pixels voisins. Le nombre de pixels voisins pris en compte peut être variable, mais plus le voisinage pris en compte est important, plus le calcul est long.

La répartition de la pente sur le bassin versant peut être représentée par une courbe hypsométrique qui exprime le pourcentage de superficie au-delà d'une certaine élévation.

L'allure de la courbe permet de caractériser le cycle d'érosion du bassin étudié. Ainsi, une courbe convexe sera caractéristique d'un bassin où l'érosion est encore intense, alors qu'une forme concave indiquera une stabilisation des processus d'érosion (LLAMAS, 1993).

2B-23. Extraction du réseau hydrographique

Le réseau hydrographique peut être extrait du MNT. La connaissance de la géométrie du réseau hydrographique apporte une information sur la dynamique du ruissellement superficiel et par conséquent sur le temps de réponse d'un bassin à un événement pluvieux.

Pour déterminer la position du réseau hydrographique à partir du MNT, un seuil d'apparition des rivières va être déterminé. Un pixel du MNT sera considéré comme rivière s'il draine une surface amont supérieure au seuil choisi. La valeur de surface amont choisie pour considérer l'apparition du réseau va donc en conditionner la complexité.

La figure 2B-8 illustre l'importance du choix de la valeur de surface amont nécessaire à l'apparition du réseau :

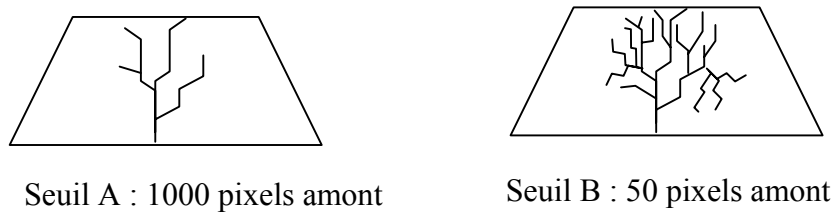


Figure 2B-8 : Complexité du réseau déterminé à partir d'un MNT en fonction du seuil d'apparition des rivières : le seuil A est plus élevé que le seuil B

2B-24. Détermination des limites des bassins versants

Le découpage de l'espace en zones hydrologiques indépendantes doit être envisagé préalablement à toute analyse de la qualité de l'eau (BELAMIE, 1983). Le traitement des informations apportées par le MNT permet de déterminer le découpage de l'espace en bassins versants (CLUIS et QUENTIN, 1993a).

Les bassins obtenus correspondent aux bassins versants topographiques puisque seul l'information altitude est utilisée. Dans des zones karstiques caractérisées par des résurgences ou dans des zones urbaines, il convient de corriger les limites du bassin versant pour prendre en compte la totalité des zones amont drainées correspondant à l'exutoire choisi. Cette correction fait bien sûr appel à des informations exogènes (cartes géologiques et plans d'occupation des sols).

2B-3. Automatisation de la phase d'extraction des unités fonctionnelles à l'aide du logiciel Arc-Info

2B-31. Introduction

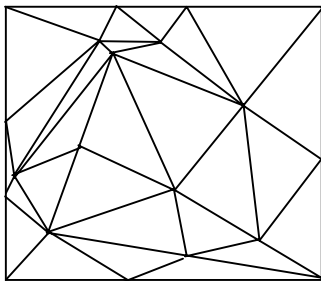
Des informations utiles pour une étude hydrologique peuvent donc être tirées d'un MNT. Le logiciel de traitement utilisé pour analyser l'information portée par le MNT est Arc-Info. Chacun des calculs présentés précédemment est réalisé successivement à la demande de l'utilisateur. Ces commandes sont optimisées et le temps de calcul n'excède pas quelques secondes même pour des plans d'information de plusieurs centaines de km².

Pour étudier plus finement les zones drainées à chaque confluence du réseau hydrographique, l'utilisateur doit successivement repérer chaque exutoire potentiel et lancer le calcul pour obtenir les limites des sous-bassins versants correspondant à une confluence donnée. Selon l'importance du réseau, cette méthode peut rapidement devenir fastidieuse. Une automatisation de ces tâches a donc été réalisée.

Une routine conçue dans le cadre de ce travail permet d'extraire automatiquement les unités fonctionnelles que sont les biefs de rivières et les sous-bassins associés à partir de la structure du réseau hydrographique (sources et confluence) et de la localisation des rejets ponctuels. Le langage utilisé pour développer cette routine est l'AML (Arc Macro Language). Ce langage informatique présente l'avantage d'utiliser toutes les fonctions et commandes développées dans le logiciel Arc-Info tout en permettant de créer ses propres applications.

2B-32. Format des MNT

Deux formats principaux de MNT sont utilisés sous Arc-Info : le format TIN et le format maillé ou raster.



Le format **TIN** permet de représenter le relief sous forme de facettes triangulaires construites à partir d'un nuage de points de mesures. Ce maillage irrégulier permet de réduire la redondance dans les zones planes (MAIDMENT, 1993a).

120	118	120	123
123	115	114	120
106	114	110	110
100	108	105	108

Le format maillé ou **raster** est la structure utilisée quasi systématiquement dans la modélisation. Ce format est particulièrement adapté au croisement avec d'autres sources d'information sous formes numériques (MAIDMENT, 1993b).

Figure 2B-9 : Formats des MNT utilisés dans le logiciel Arc-Info

Le format du MNT de l'IGN, utilisé dans cette étude, est de type raster.

2B-33. Comblement des points bas et calcul de la pente

Sous le logiciel Arc-Info, l'étape classique de comblement des points bas doit être réalisée avant toute utilisation hydrologique du MNT. La fonction pré-programmée SINK permet ainsi de combler les points bas. Des particularités de la zone étudiée peuvent conduire à des pré-traitements supplémentaires du MNT comme le creusement du MNT pour passer outre des obstacles générés lors de la conception du MNT.

Après ces pré-traitements, des plans dérivés présentant un intérêt dans une étude hydrologique peuvent être obtenus à partir du MNT. La pente est ainsi une des premières informations calculées à partir du MNT qui permet de caractériser le bassin versant.

Sous Arc-Info, le gradient pente est calculé par le passage d'un filtre 3 x 3 caractérisé par les coefficients de pondération suivant :

$$x = \frac{1}{8} \times \begin{pmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{pmatrix} \quad y = \frac{1}{8} \times \begin{pmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{pmatrix} \quad \text{Équation A2-1}$$

La pente peut être calculée en degré ou en pourcentage. Dans cette étude, le calcul en pourcentage est retenu. La pente s'exprime en $m.m^{-1}$.

2B-34. Détermination des directions d'écoulement

Sous Arc-Info, la détermination des directions d'écoulement est réalisée ainsi : selon la direction de la plus forte pente, schématisée par la figure 2B-10, le pixel recevra une valeur donnée :

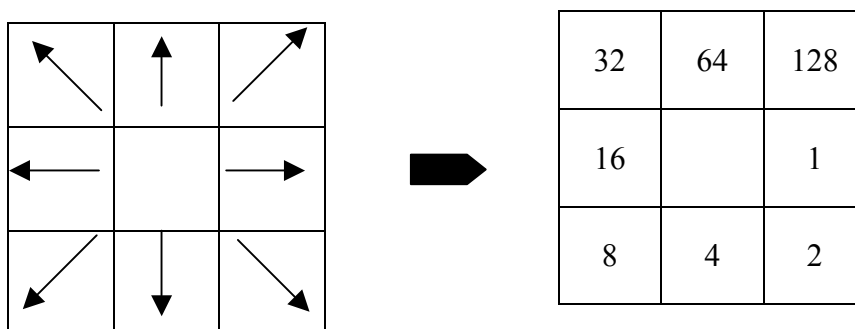


Figure 2B-10 : Codage des pixels selon la direction d'écoulement

Chaque pixel reçoit donc une des 8 valeurs présentées dans la figure 5 selon la direction de plus forte pente. Ainsi selon la direction de plus grande pente déterminée à partir du MNT on obtient la matrice direction d'écoulement comme le montre la figure 2B-11.

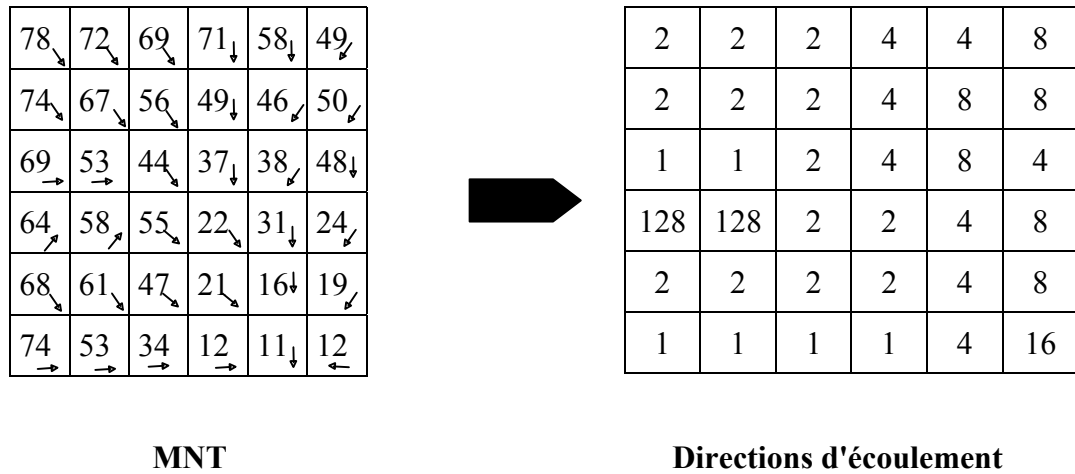


Figure 2B-11 : Obtention du plan direction d'écoulement à partir du MNT

Ce plan dérivé du MNT est particulièrement utile pour appréhender les directions privilégiées qui pourront être empruntées lors d'un ruissellement de surface.

2B-35. Calcul des valeurs cumulées

Le calcul des surfaces drainées se fait d'après les directions d'écoulement obtenues précédemment. Chaque pixel du MNT se voit affecté du nombre de pixels situés en amont selon les directions d'écoulement.

Pour un pixel donné, la valeur du nombre de pixels amont va traduire l'importance de la zone drainée.

La figure 2B-12 illustre un exemple classique de détermination des valeurs d'accumulation (pour améliorer la lisibilité de la figure, les directions d'écoulement sont représentées par des flèches et non codées par les valeurs présentées par la figure 2B-11).

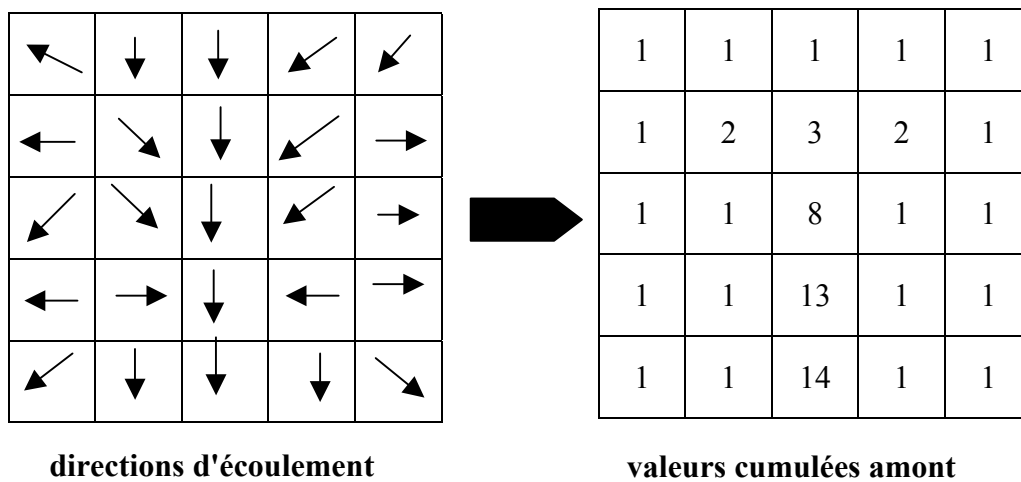
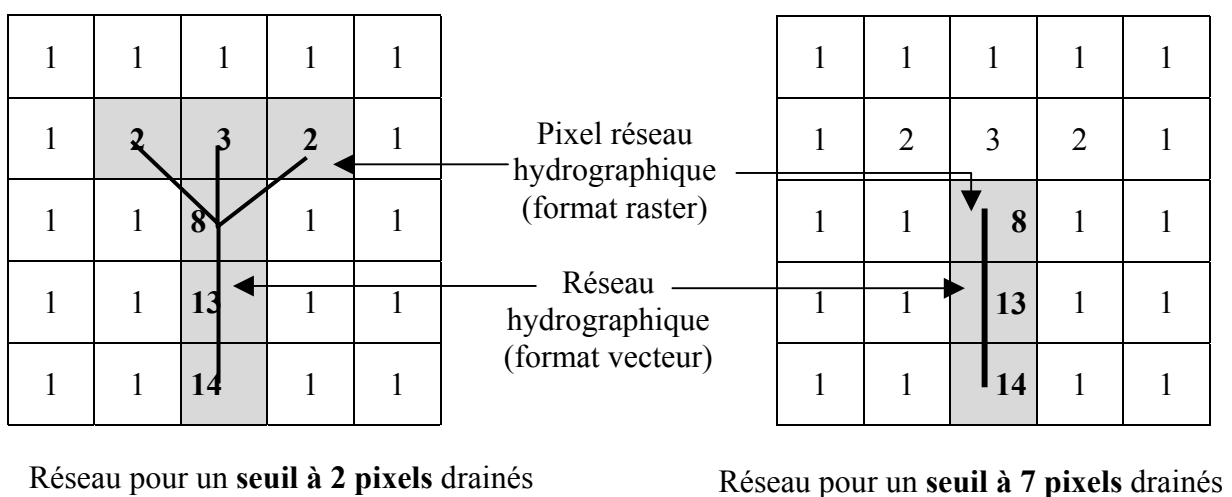


Figure 2B-12 : Obtention des valeurs cumulées à partir des directions d'écoulement

Les informations valeurs cumulées et directions d'écoulement permettent d'extraire un réseau hydrographique de la zone étudiée.

2B-36. Extraction automatique du réseau hydrographique

Le calcul prend donc en compte la double information plan des directions d'écoulement et plan des valeurs cumulées. Seuls les pixels qui présentent une surface amont suffisante, traduite par les valeurs cumulées, seront considérés comme appartenant au réseau hydrographique. résultat est schématisé par la figure 2B-13.



Réseau pour un seuil à 2 pixels drainés

Réseau pour un seuil à 7 pixels drainés

Figure 2B-13 : Détermination du réseau hydrographique selon le seuil de pixels drainés

Sous format raster, le réseau sera constitué par l'ensemble des pixels dont la valeur cumulée est supérieure ou égale au seuil. Ce réseau peut être vectorisé, auquel cas le linéaire qui en résulte passe par les centres des pixels "réseau hydrographique".

La couverture du plan réseau hydrographique, en mode vecteur, possède une table d'attributs qui contient les caractéristiques de chacun des arcs du réseau. Chaque arc du réseau est compris entre deux nœuds : un nœud de départ et un nœud d'arrivée comme l'illustre la figure 2B-14.

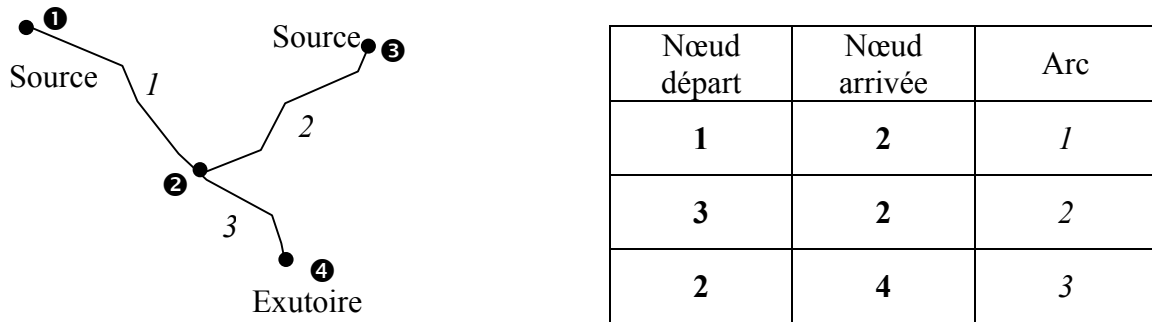


Figure 2B-14 : Codage du réseau hydrographique sous Arc-Info

Ce plan étant dérivé du MNT comme nous l'avons vu dans la première partie, la longueur du réseau peut être calculée en nombre de pixels comme le montre la figure 2B-15.

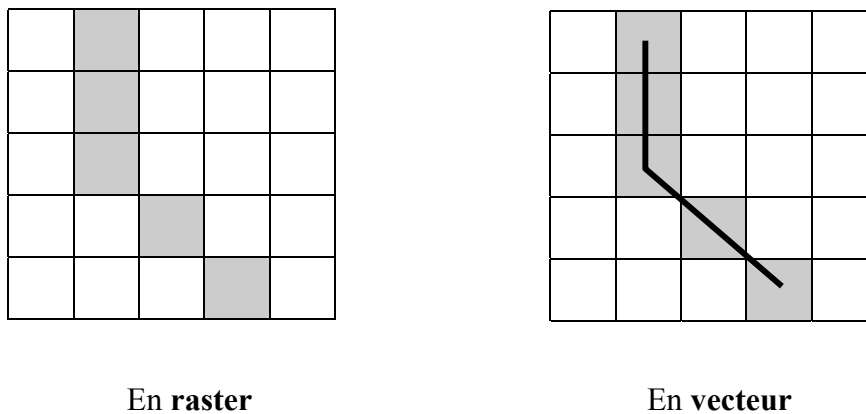


Figure 2B-15 : Détermination du réseau sous forme raster et vecteur

Après avoir déterminé le plan réseau hydrographique, la détermination des limites des bassins versants constitue un plan dérivé indispensable dans une thématique hydrologique.

2B-37. Détermination des limites des bassins versants

Les limites des bassins versants sont calculées à partir du plan d'information direction d'écoulement. Sous Arc-Info, l'utilisateur choisit le ou (les) pixel(s) exutoire(s) pour le(s)quel(s) il souhaite connaître la zone amont. Le résultat est schématisé par la figure 2B-16.

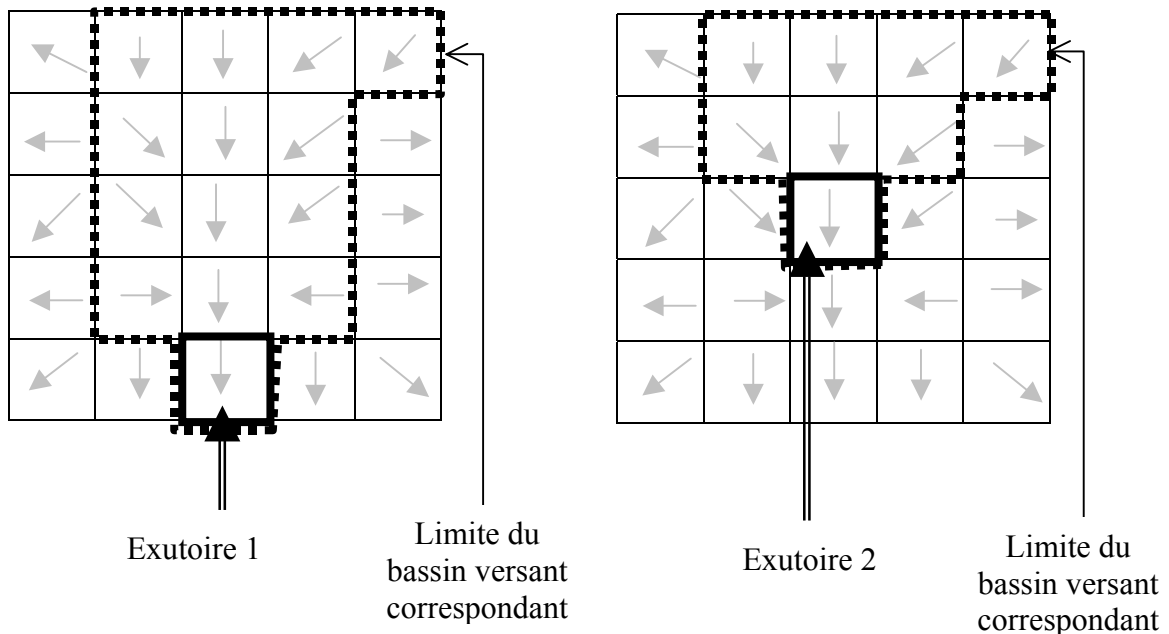


Figure 2B-16 : Détermination des limites de bassins versants

La détermination des limites des sous-bassins versants a nécessité une phase de programmation plus importante à l'aide de routines écrites en AML. En effet, cette détermination de limites topographiques implique la recherche des points exutoires à partir desquels on souhaite déterminer les sous-bassins amont. Ces différents points caractéristiques du réseau vont être traités différemment selon qu'il s'agisse des confluences ou bien des sources et de l'exutoire.

La première étape consiste à travailler sur la table d'attributs du plan réseau hydrographique pour déterminer les sources et l'exutoire. Ces nœuds particuliers peuvent être sélectionnés car ils ne sont jamais nœuds d'arrivée et nœuds de départ dans la table. Pour ces nœuds particuliers, le bassin versant correspondant va être tracé à partir du pixel qui contient le nœud comme le montre la figure 2B-17.

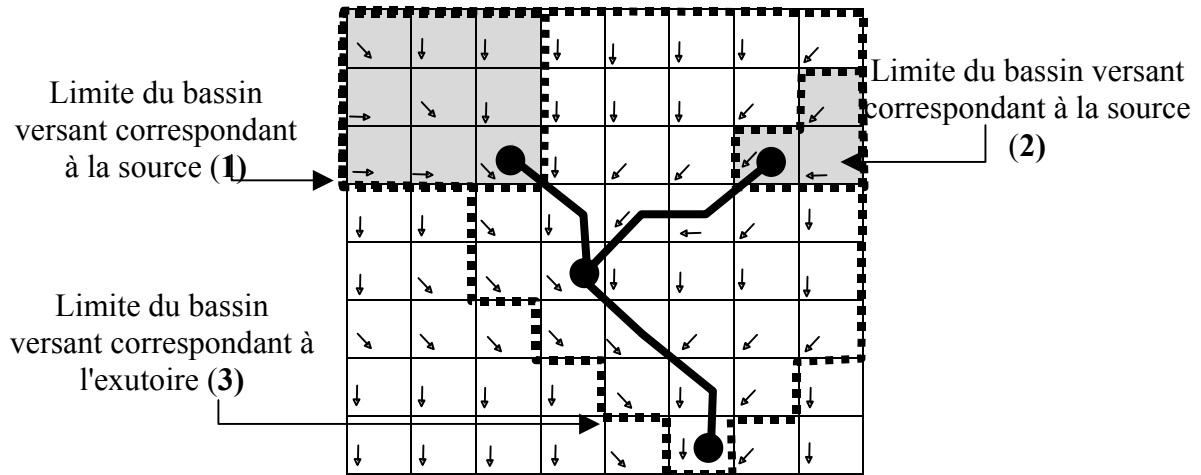


Figure 2B-17 : Obtention des bassins versants correspondant aux sources et à l'exutoire d'un réseau

La deuxième étape du traitement consiste à déterminer les sous-bassins versants des biefs amont pour une confluence donnée comme le montre la figure 2B-18

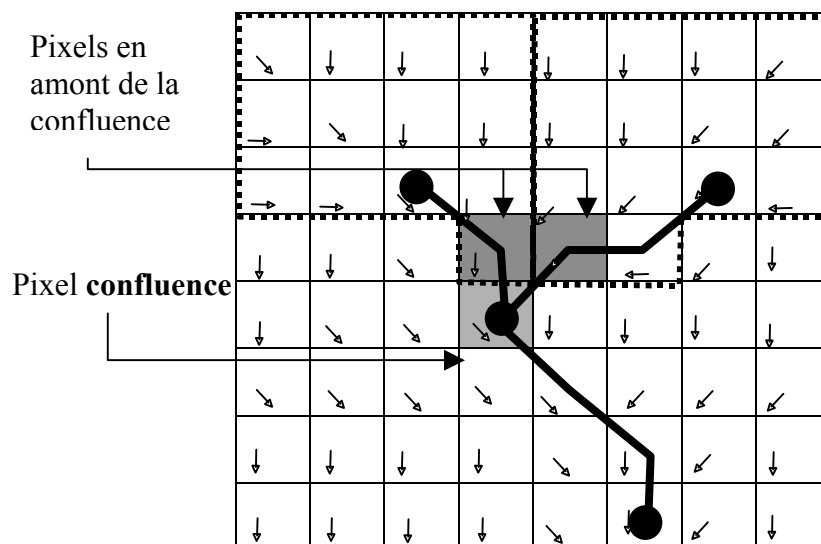


Figure 2B-18 : Obtention des sous-bassins versants correspondant aux biefs amont d'une confluence

Selon cette règle de calcul, le pixel confluence appartiendra au sous-bassin versant du bief aval.

Pour automatiser le traitement des confluences, plusieurs tests sont réalisés sur la confluence et ses 8 pixels voisins :

- test 1 : le pixel testé est-il une rivière ?
- test 2 : le pixel testé est-il la confluence ?
- test 3 : le pixel testé est-il en aval de la confluence ?
- test 4. Le pixel testé est-il sur un bief aboutissant à une autre confluence ?

Si le pixel, parmi les 9 testés, ne répond à aucun de ces tests par l'affirmative, alors il est retenu et sert d'exutoire au sous-bassin versant du bief amont testé. Ainsi si deux biefs amont aboutissent à la confluence, deux sous-bassins seront déterminés.

A l'issue de ces deux traitements sur les sources et l'exutoire d'une part et sur les confluences d'autre part, on obtient sous forme raster les limites du bassin versant et de ses sous-bassins.

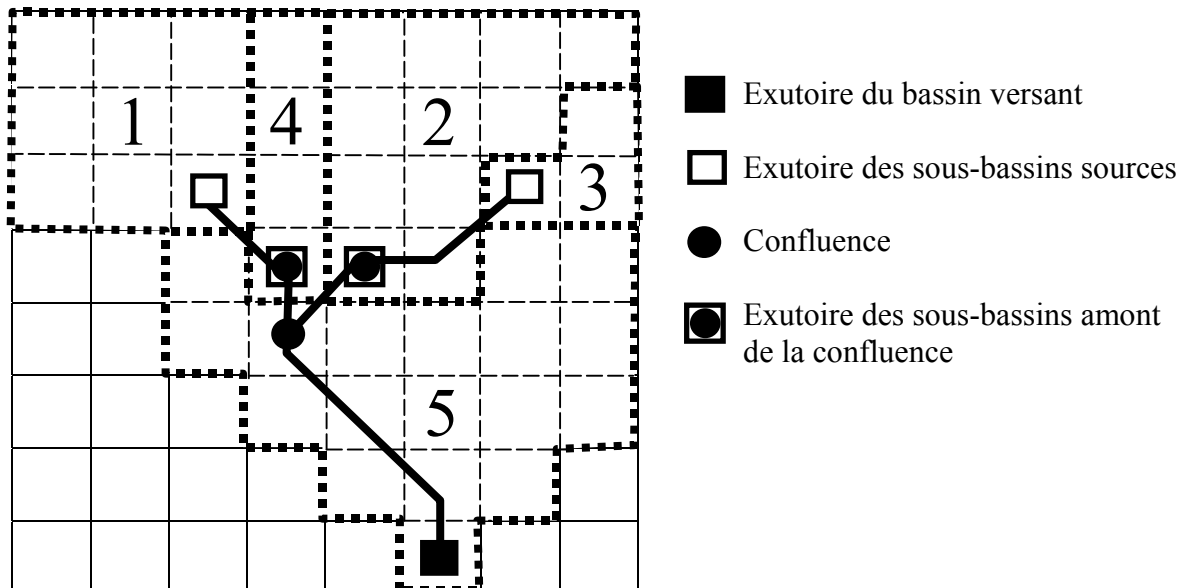


Figure 2B-19 : Obtention des limites du bassin versant et de ses sous-bassins

L'intégration des rejets ponctuels est effectuée par croisement géographique comme le montre la figure 2B-20.

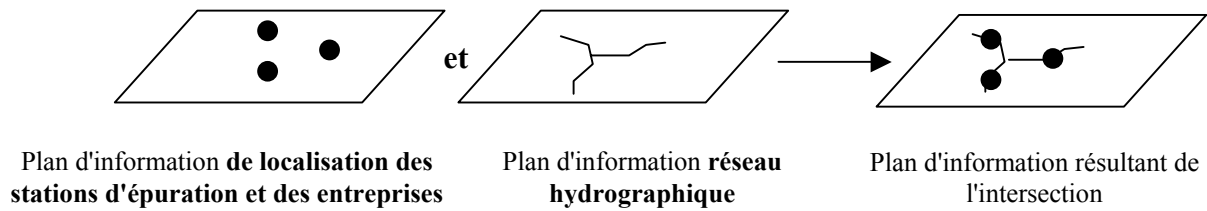


Figure 2B-20 : Croisement de deux plans d'information : localisation des entreprises potentiellement polluantes et réseau hydrographique

L'étape suivante consiste à déterminer les limites du nouveau sous-bassin en amont du point de rejet comme l'illustre la figure 2B-21.

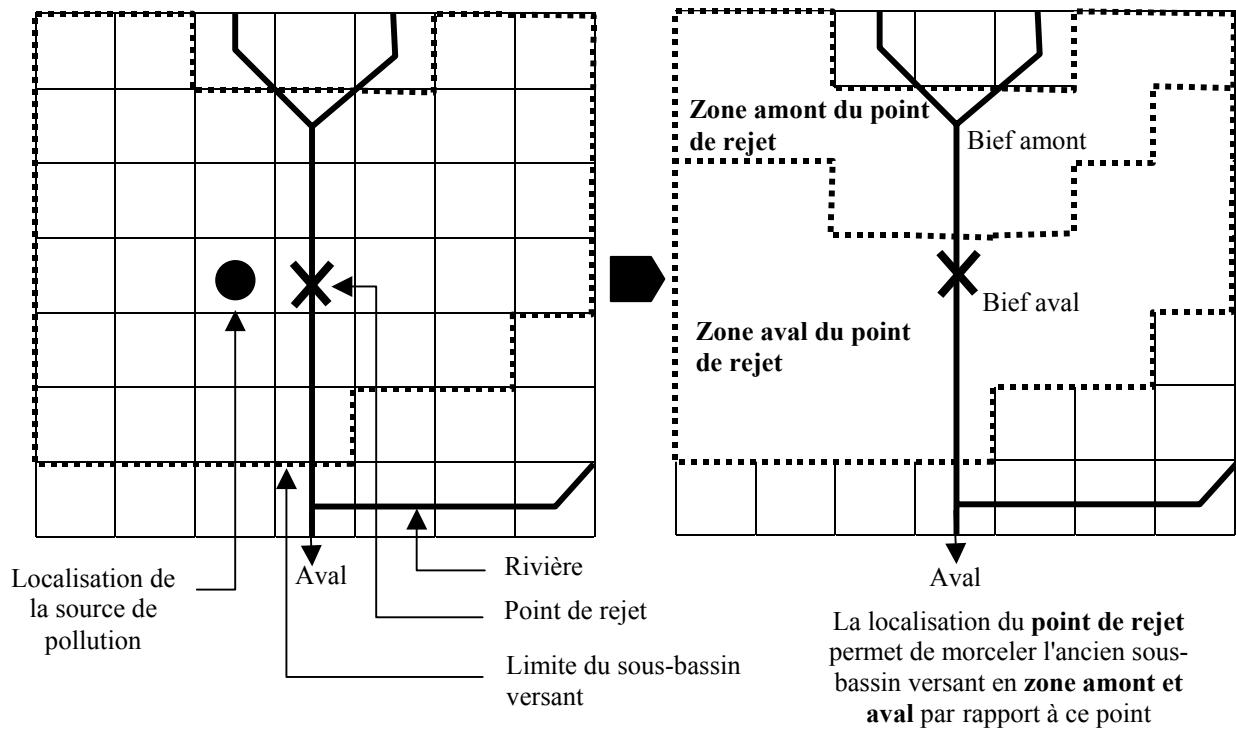


Figure 2B-21 : Intégration des rejets ponctuels pour l'obtention des unités fonctionnelles de production (sous-bassins) et de transport (biefs)

Le bief initial, concerné par un rejet ponctuel, est donc divisé en un bief amont et aval en fonction de la position du point de rejet. De même, après ce traitement il est possible d'identifier une zone en amont et en aval du point de rejet.

2B-4 Conclusion

La création d'une routine automatique permet d'extraire à partir du MNT les unités fonctionnelles de production (sous-bassins versants) et de transport (biefs) à partir de la structure du réseau (sources et confluences) et la position des rejets ponctuels (stations d'épuration).

2C Méthode de correction du MNT pour améliorer la qualité des réseaux hydrographiques extraits

2C-1. Introduction

L'objectif de cette annexe est d'illustrer la méthode de correction du MNT pour améliorer la qualité du réseau hydrographique extrait qui sert ensuite de base à la création de l'architecture.

Pour obtenir une cohérence entre unités de production et de transport, la même source d'information (le MNT) est utilisée. Cependant, pour obtenir une architecture spatiale cohérente avec la connaissance des écoulements sur le terrain, la qualité du réseau hydrographique issu du MNT doit être contrôlée. On entend ici par qualité d'extraction uniquement l'écart entre le réseau extrait du MNT et le réseau de référence utilisé pour décrire le bassin (sont-ils superposés ? l'un divague-t-il plus que l'autre ? l'écart entre les deux réseaux est-il constant ?).

Le réseau hydrographique de référence utilisé dans ce travail est celui de la BD Carto® de l'IGN.

2C-2 Méthode de diagnostic de la qualité d'extraction du réseau hydrographique

2C-21 Méthode de diagnostic

La qualité de l'extraction du réseau dépend de la qualité propre du MNT, de la topographie de la zone d'étude et des algorithmes utilisés. La qualité du MNT sur la zone est de 1 à 3 m en altitude pour une résolution de 50m (Bd Alti® de l'IGN). Les bassins étudiés présentent une morphologie particulière avec une zone amont élevée (de 150 à 300 m) et un relief contrasté. Les zones aval sont relativement monotones et de faible altitude (proche du niveau de la mer). L'extraction de réseau hydrographique est délicate dans ces zones planes et la qualité du réseau est par conséquent soumise à caution.

L'algorithme utilisé est dénommé D8 car la détermination des directions d'écoulement se fait suivant un flux unidirectionnel en 8-connexité. Le point faible de cet algorithme est la discrétion des directions d'écoulement en huit directions d'écoulement séparées par 45°, ce qui peut introduire de fort biais dépendant de la grille d'orientation du MNT.

Une méthode de diagnostic de qualité d'extraction des rivières a été proposée en systématisant ces biais (CHARLEUX-DEMARGNE, 2001) : le MNT est transformé par rotations successives de 1 degré, d'angle compris entre 0° (MNT initial) et 89°, ce qui induit 90 plans de réseaux hydrographiques. Ces 90 plans sont transformés par rotation inverse puis

additionnés, pour obtenir, en chaque pixel, le nombre de fois où il appartient au réseau : ce plan peut alors se traduire en terme de probabilité d'appartenir au réseau. L'analyse de ce plan de mesure probabiliste permet de repérer les zones où le calcul du réseau hydrographique est stable et celles où le tracé est plus aléatoire, révélateur d'une qualité d'extraction insatisfaisante.

2C-22 Résultats

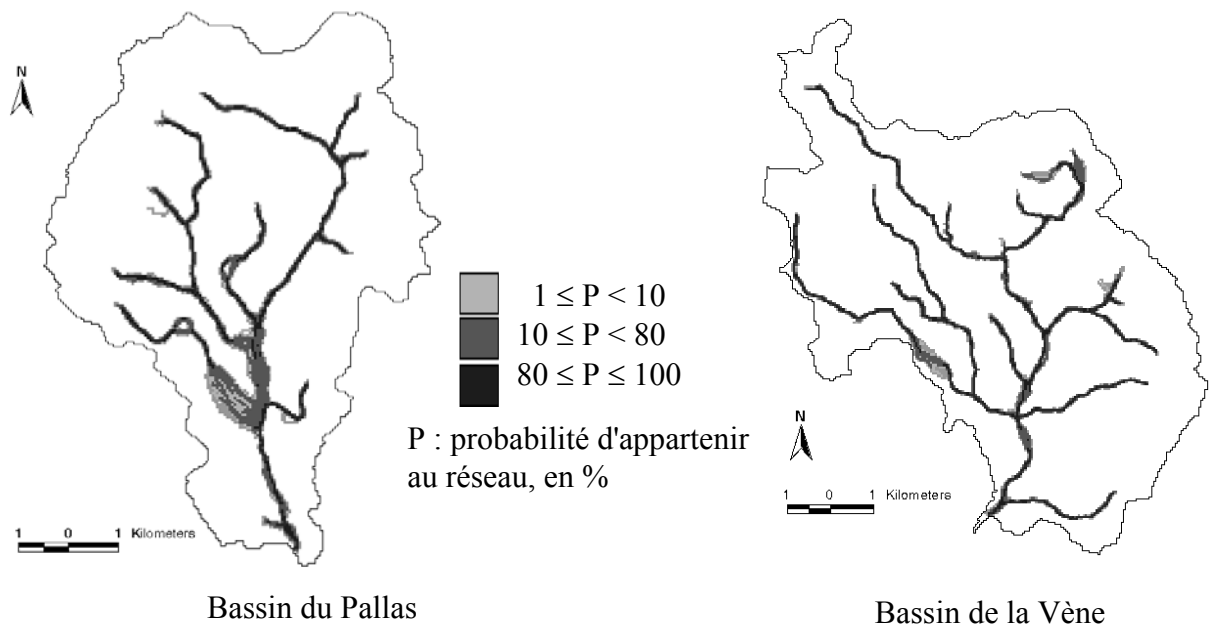


Figure 2C-22 : Cartographie de la qualité d'extraction du réseau hydrographique

Les zones d'instabilité du tracé du réseau, avec une faible probabilité d'appartenance au réseau, se révèlent être les zones au relief monotone et les zones de confluence.

2C-23 Modification du MNT pour améliorer l'extraction du réseau

La modification du MNT est un sujet sensible. En effet, cela consiste à générer une nouvelle information à partir du MNT et d'une donnée exogène. La modification doit donc être justifiée par l'objectif du traitement. Nous souhaitons obtenir un réseau issu du MNT proche géométriquement du réseau de l'IGN que nous considérerons dans la suite du travail comme étant le réseau de référence. La méthode de diagnostic de la qualité du réseau permet de quantifier l'effet d'une modification du MNT.

La modification envisagée consiste à incruster le réseau de référence (IGN) sur la surface du MNT pour contraindre les écoulements selon l'information que nous jugeons la plus fiable. Dans ces zones d'instabilité, on peut alors mettre en œuvre des techniques d'amélioration, dont l'impact est évalué grâce à ce même diagnostic de qualité. La solution testée ici est basée sur la modification du MNT provenant de la BD Alti® de l'IGN, en incrutant un réseau hydrographique de référence, celui de la BD Carto®. Pour tous les pixels n'appartenant pas au réseau, on ajoute une valeur forfaitaire dz à l'altitude du MNT. La comparaison des

mesures probabilistes de réseau, relatives aux MNT modifiés suivant différentes valeurs dz , permet de choisir la valeur dz optimale, qui est fonction du type de terrain et du MNT lui-même.

Le graphique 2C-23 montre l'effet d'une incrustation sur un MNT.

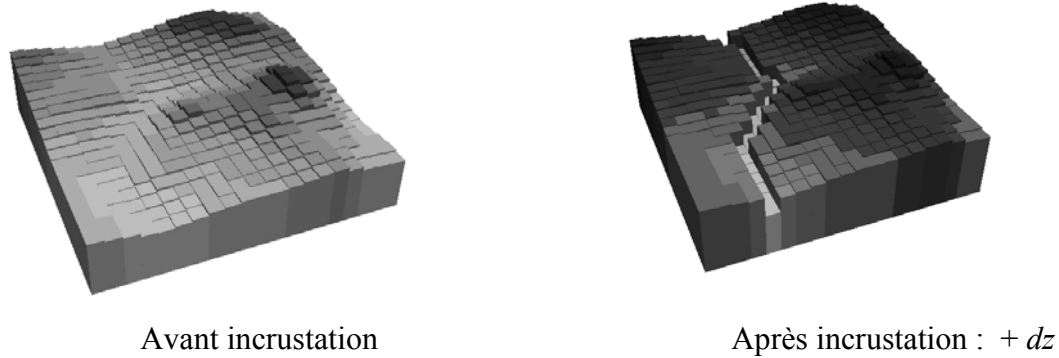


Figure 2C-23 : Effet de l'incrustation du réseau de référence sur le MNT

Le résultat de cette étude permet de choisir la profondeur d'incrustation pour obtenir le réseau le plus stable.

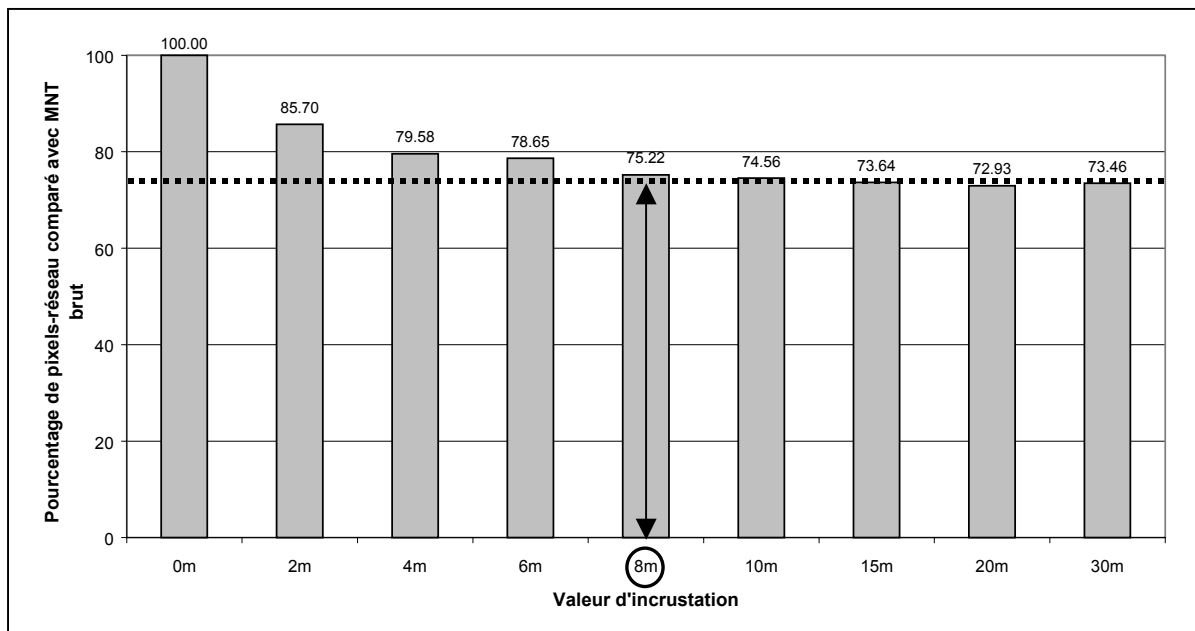


Figure 2C-3 : Comparaison du % de pixels-réseaux ($P > 1$) obtenu pour différentes valeurs dz sur le bassin du Pallas.

Plus le réseau est stable, plus le nombre de pixels-réseaux diminue. Sur le bassin du Pallas, la valeur de dz qui apporte une amélioration significative est de 8 m. On considère que l'amélioration est stable si l'évolution du nombre de pixels-réseaux est inférieure à 1% entre 2 valeurs de dz .

Le graphique 2C-25 permet de fixer le dz optimum à 4m sur le bassin de la Vène.

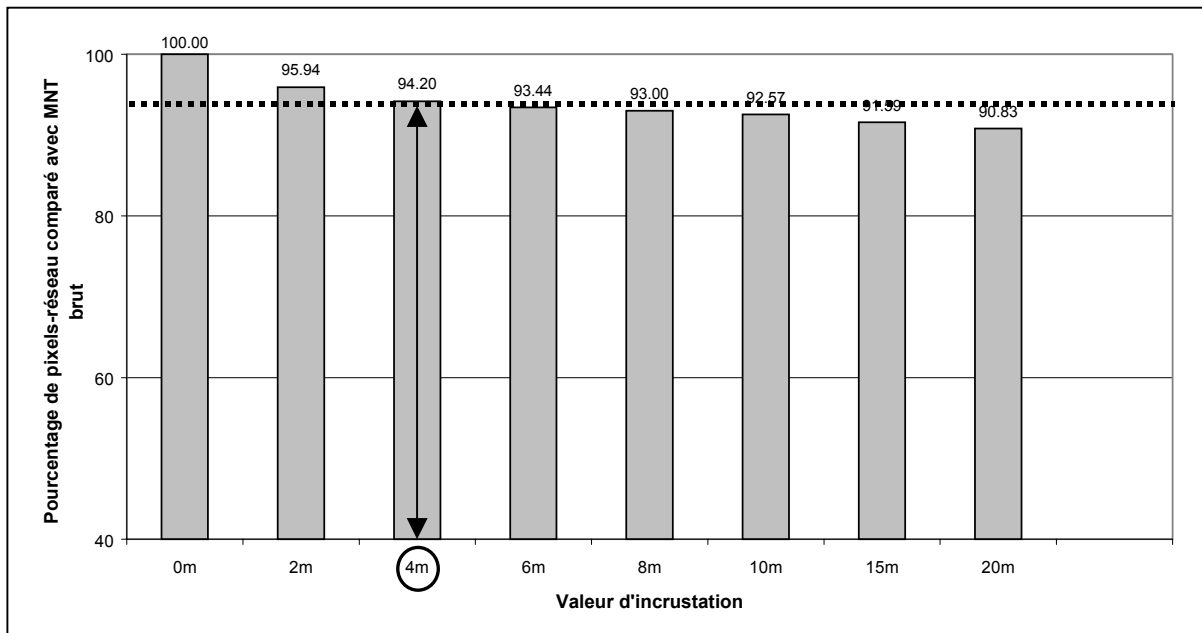


Figure 2C-25 : Comparaison du % de pixels-réseaux ($P > 1$) obtenu pour différentes valeurs dz sur le bassin de la Vène.

La modification du MNT peut être également envisagée dans le cas où le MNT ne permet pas l'écoulement pourtant constaté sur le terrain.

Ce cas de figure s'observe sur l'étang de l'Or et est dû à la source d'information du MNT. Celui-ci est en effet généré par interpolation entre des points cotés de la Bd Alti® de l'IGN. Le bassin versant de l'étang de l'Or présente la particularité d'être parcouru d'est en ouest par un canal (canal Philippe LAMOUR). Celui-ci est surélevé par rapport au bassin lui-même. Cependant, pour permettre l'écoulement des rivières du nord vers le sud, ce canal passe en siphon sous le lit des rivières du bassin. Cette particularité fonctionnelle n'apparaît pas sur le MNT et le canal apparaît comme une barrière aux écoulements.

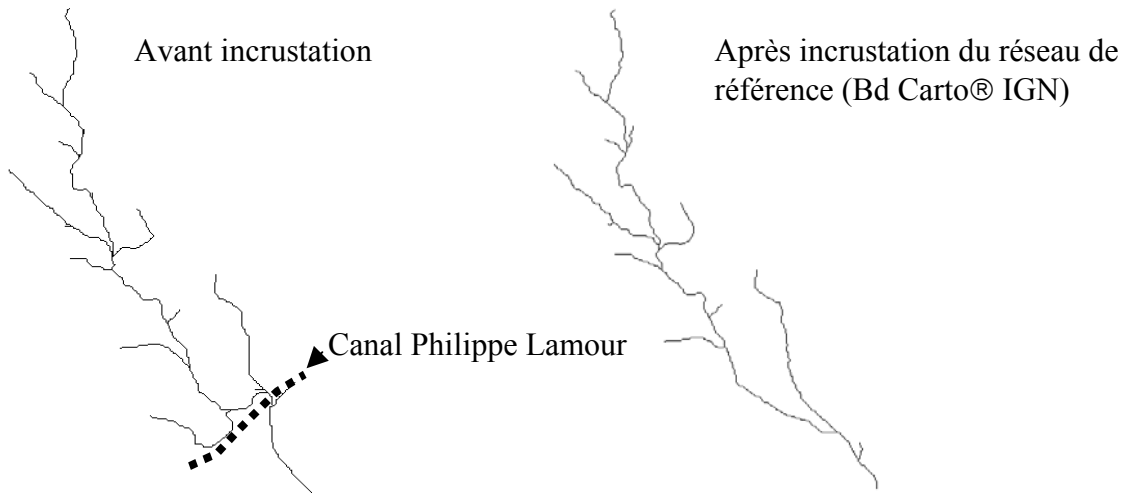


Figure 2C-26 : Effet de l'incrustation d'un réseau de référence sur la qualité du réseau extrait (Rivière du Salaison)

La phase de modification du MNT permet d'extraire un réseau dont la géométrie est similaire au réseau de référence.

Là encore la méthode de diagnostic permet de déterminer la valeur dz d'incrustation optimale pour retrouver un écoulement en accord avec le réseau hydrologique. Le graphique 2C-27 permet ainsi de fixer la valeur d'incrustation dz optimale sur du bassin du Salaison.

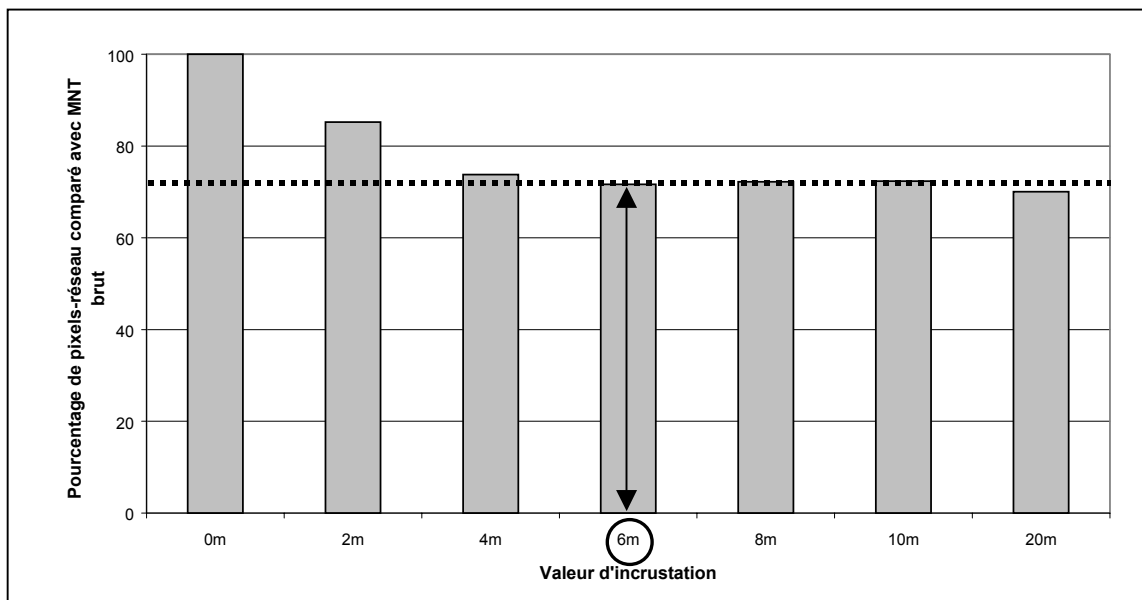


Figure 2C-27 : Comparaison du % de pixels-réseaux ($P > 1$) obtenu pour différentes valeurs dz sur le bassin du Salaison.

La valeur d'incrustation optimale dz est donc de 6m.

2C-3. Conclusion

En conclusion, la modification du MNT par l'incrustation d'un réseau de référence permet :

- d'une part de stabiliser le réseau extrait surtout dans les zones planes où le MNT ne permet pas d'extraire correctement le réseau hydrographique ;
- et d'autre part de rétablir des écoulements conformes au fonctionnement hydrologique sur un bassin, en cas d'obstacles artificiels intégrés au MNT (cas de l'étang de l'Or).

Cette méthode est appliquée sur l'ensemble des deux sites d'étude pour permettre une extraction des unités hydrologiques