

LES DONNEES MAILLEES (OU " RASTER ")

Par B. Allouche (CERTU)

Sont désignées couramment comme données maillées (ou " raster", terminologie non correcte en français) toutes données géographiques représentées selon un maillage régulier de l'espace.

Contrairement aux données géographiques en mode vecteur qui décrivent des objets à l'aide de points, de lignes ou de polygones, les données maillées permettent une description systématique de toute la surface d'un territoire selon un découpage régulier. C'est sur cette maille, ou encore pixel pour les anglo-saxons, qu'est rattachée le codage de l'information.

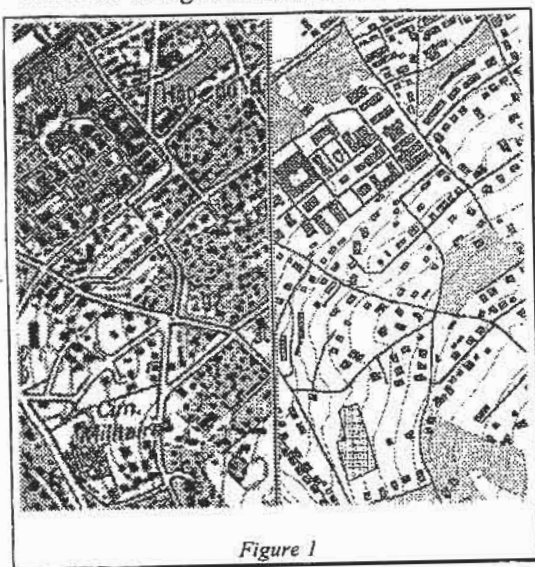


Figure 1

L'exemple montré en figure 1 illustre le principe de la description sous forme vecteur et de la description sous forme maillée, pour des objets géographiques similaires.

La structuration des données sous forme maillée est aujourd'hui peu utilisée dans les SIG, sauf pour du simple affichage. Il existe plusieurs raisons à cela :

- d'ordre technologique, freinant la disponibilité de ces données ;
- d'ordre méthodologique, freinant l'usage de ces données ;
- d'ordre pratique, rendant difficile l'intégration dans un SIG.

Ces difficultés vont être explicitées ci-après.

Les tendances technologiques

Si, sur le principe, les données maillées existent depuis longtemps, elles sont restées confinées à un usage relativement spécialisé (production de

données par photo-interprétation, télédétection, vectorisation) faute de moyens techniques suffisants pour en tirer le meilleur parti. Cette situation évolue rapidement depuis quelques années au rythme des améliorations technologiques et de l'élargissement de l'offre.

Amélioration des conditions de mise en œuvre

Le principal obstacle à l'usage des données a été levé avec l'arrivée d'éléments de stockage et de traitements plus performants. Les données maillées exigent en effet des capacités mémoires bien au dessus des applications classiques de la micro-informatique. Ainsi, une donnée raster exige 10 à 100 fois plus de mémoire qu'une base de données vectorielle à surface égale et à gamme d'échelle comparable,

L'évolution technologique bénéficie donc aux utilisateurs, en permettant une mise en œuvre avec une meilleure efficacité de ces données numériques, sur des territoires plus importants.

Augmentation de la résolution

Les producteurs bénéficient également d'avancées technologiques leur permettant de proposer aujourd'hui des données plus fines, dont la fabrication est fiabilisée et simplifiée.

C'est ainsi que l'on voit apparaître des outils de production à grand rendement offrant des possibilités nouvelles en matière de modèles numériques de terrain (10 cm de précision en altimétrie et selon des mailles de l'ordre du mètre; figure 2) ou des images à très haute résolution (10 cm obtenu directement par caméra numérique).

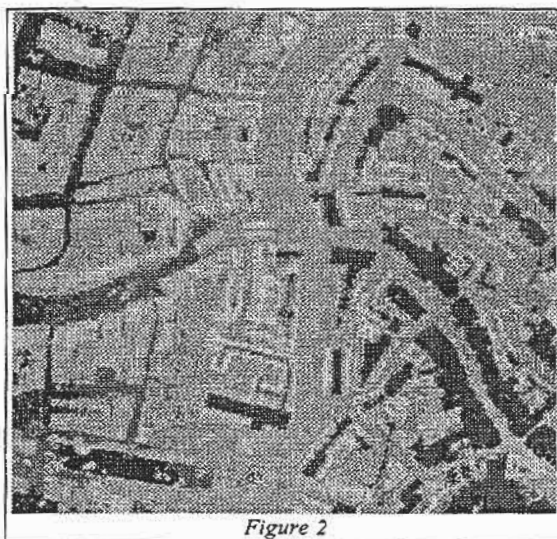


Figure 2

Les résolutions métriques, voire sub-métriques, gagnent aussi le domaine spatial civil et seront largement disponibles à travers le monde d'ici un à deux ans (fournisseurs américains, indiens, mais aussi européens avec SPOTS et ses images à 2,5 m de maille au sol).

Abaissment des coûts de production

Ces avancées technologiques améliorant le recueil de l'information se traduisent également par des produits plus élaborés et à un coût moindre. Par exemple, l'offre en matière de données orthophotographiques se développe en s'ouvrant à un marché plus vaste compte tenu de baisses significatives des coûts de production. Cela favorise l'émergence d'un marché plus concurrentiel.

L'IGN s'apprête à commercialiser des orthophotoplans numériques départementaux avec une maille de 50 cm, et envisage pour un avenir plus lointain d'employer une caméra numérique pour proposer des produits d'une qualité encore meilleure avec des mailles très fines (20 cm) sur de plus faibles étendues (villes par exemple).

La société GéoSys propose également des couvertures des principales villes avec des images numériques sur catalogue, mises à jour annuellement.

Cette liste d'offre n'est évidemment pas exhaustive (sociétés Eurosense, Ortea, InterAtlas, Géo-image...).

Les difficultés d'exploitation

Bien que d'apparence souvent plus lisible pour l'œil, les données maillées présentent quelques difficultés d'exploitation dans un SIG.

En effet, sauf à représenter le terrain tel qu'il est (cas des images numériques du territoire utilisées comme simple "fond de carte"), ces données nécessitent souvent un travail d'interprétation des informations portées par la maille.

En outre, le principe du maillage est peu compatible avec une notion de base de données attributaires. On peut en effet encore difficilement mettre en œuvre des informations attributaires complexes associées à chaque maille.

Enfin, les logiciels permettant une intégration de données maillées dans un SIG n'offrent pas de souplesse réelle de manipulation, en particulier pour le croisement des couches raster entre elles. Il est par exemple difficile de gérer des images dont les mailles ont des tailles différentes. Par

ailleurs, en terme de restitution visuelle ou cartographique, une seule "couche" est visible. Certains logiciels offrent la possibilité de rendre transparente une couleur, mais cette solution n'est applicable qu'à des cartes rasters peu chargées, en général monochromes.

L'exploitation croisée avec des données vecteur n'est généralement pas disponible sur les logiciels actuels, malgré l'intérêt énorme que peuvent représenter de tels croisements.

A ces difficultés générales d'exploitation, dues essentiellement aux outils logiciels, s'ajoutent d'autres difficultés liées à la nature de l'information elle-même. On peut ainsi distinguer quatre catégories de données maillées, comportant chacune leur limite méthodologique :

les données raster traduisant une grandeur naturellement continue :

On doit entendre ici des informations qui ont un sens en chaque point du territoire : une donnée d'altitude, de pente, d'orientation, des mesures de pression atmosphérique, de bruit, des temps d'accès, des cartes de distance, des valeurs de contrainte ...

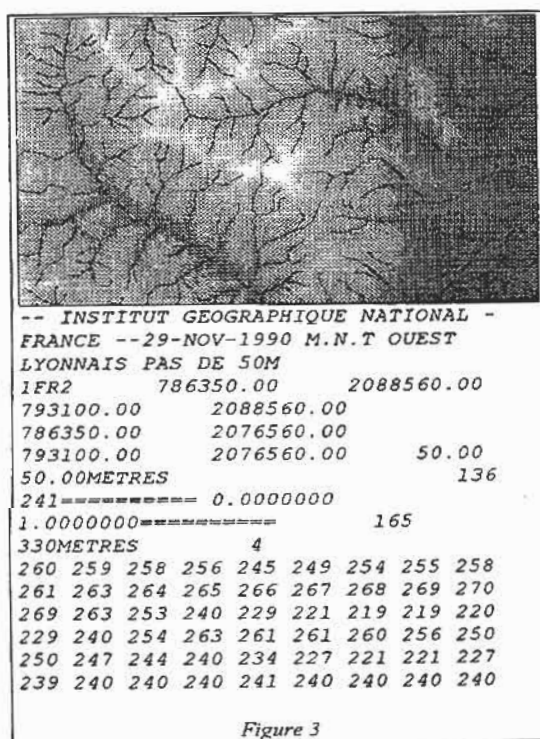


Figure 3

Le mode maillé est pour ce type d'information le moyen le plus pertinent de représentation. La valeur portée par la maille est une valeur directement liée au phénomène représenté (par exemple, l'altitude pour le relief, figure 3). Il n'y a pas d'ambiguïté pour exploiter ces données.

Les difficultés viendront souvent des logiciels SIG qui ne permettent pas d'intégrer ces données dans toute leur richesse, ou ne permettent pas d'en tirer l'information utile.

les données nécessitant une interprétation :

On classera ici les photographies aériennes ou satellites. Ce sont des images " naturelles " du territoire sous forme numérique.

Dans ce cas l'information portée par la maille correspond à une intensité lumineuse renvoyée par le sol. Toute la difficulté consiste à déterminer à quoi correspond cette intensité lumineuse : s'agit-il de forêt, de bâti, d'eau, de roche, de surfaces labourées ?... (figure 4).

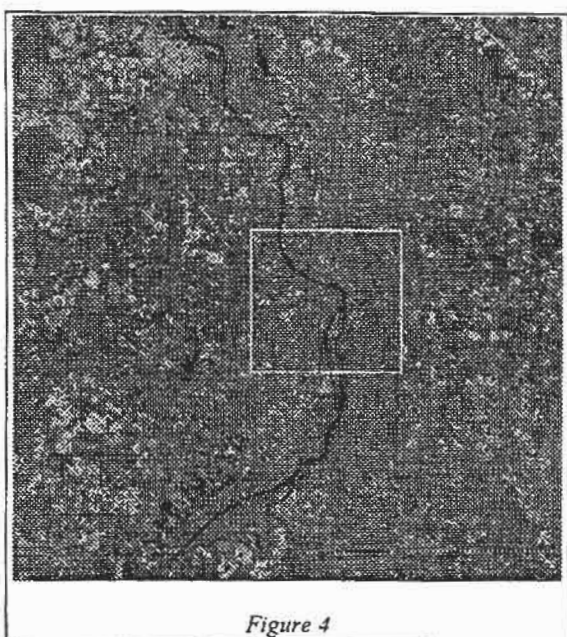


Figure 4

En fait, le codage de l'information n'est pas univoque et nécessite une interprétation faisant souvent appel à d'autres sources d'information (texture de l'image, connaissance des lieux, ...).

Pour cela, trois voies sont couramment utilisées :

- soit l'interprétation totalement automatique, utilisant au mieux les techniques du traitement d'image : cette solution est somme toute ancienne, dispose de nombreuses références, mais souffre d'une évolution technique permanente et d'une non stabilisation des procédés ;
- soit une interprétation totalement réalisée par un opérateur : il faut alors disposer d'une réelle compétence, et la reproductibilité de l'interprétation est toujours un problème ;

- enfin, une voie de plus en plus pratiquée consiste à utiliser les potentiels des deux solutions

précédentes, dans ce qui est appelé, l'interprétation assistée par ordinateur : on prépare différentes vues de l'image à partir des techniques automatiques, et elles sont interprétées par un opérateur.

Ces données sont donc des sources d'information très riches et facilement actualisables, mais elles nécessitent en contrepartie un haut niveau de savoir-faire pour être exploitées.

les données portant une information déjà interprétée :

Chaque maille porte un code faisant référence à une information de façon univoque. Le code fait référence à une nomenclature (figure 5) bien précise (classification statistique du territoire selon la nomenclature CLUSTER proposée par EUROSTAT, par exemple).

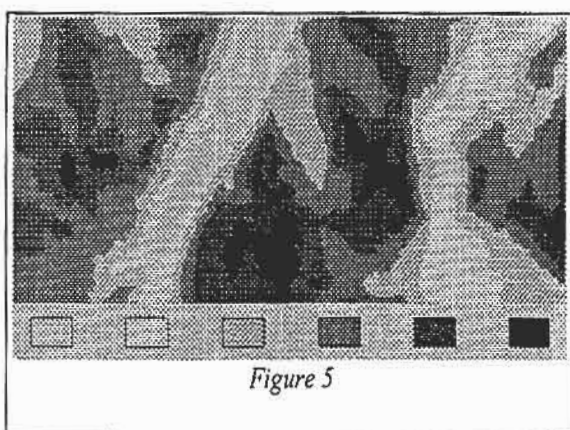


Figure 5

Afin de représenter des objets très proches les uns des autres, on est soit obligé de diminuer globalement la taille de la maille, ce qui augmente de manière importante la taille des fichiers. Une autre solution consiste à utiliser des techniques plus sophistiquées, telles que les quad-tri, qui consiste à couper une maille en quatre mailles plus petites, et ceci uniquement là où cela est utile. Malheureusement, elles sont peu souvent implémentées dans les logiciels SIG.

Le principal frein à leur utilisation est le peu de données qui existent sous cette forme aujourd'hui.

les données reproduisant une information déjà interprétée, existant sur papier :

Il s'agit des cartes scannées, telles que les cartes routières, le cadastre, ou les cartes géologiques.

Extraire des informations géographiques d'une carte scannée en s'appuyant sur la légende graphique présente beaucoup de difficultés : en effet une couleur correspond rarement à un poste de la légende.

Des techniques de reconnaissance de caractères, de vectorisation automatique, alliées à des classifications de couleur, peuvent dans certains cas aider à extraire une partie de l'information contenue dans la carte scannée, afin d'en produire une information vectorielle. Leur emploi reste cependant difficile, sauf par des professionnels.

Les cartes scannées peuvent donc être utilisées comme référentiel visuel, comme fond de plan (elles présentent un aspect familier aux utilisateurs), comme source d'information (numérisation sous forme vecteur d'éléments graphiques contenus dans la carte), mais ne peuvent que rarement être utilisées comme une information pouvant être croisée directement avec une autre.

Les difficultés pratiques

En faisant abstraction des raisons qui motivent (démontrent) l'intérêt d'utiliser des données maillées et des difficultés méthodologiques qui freinent leur usage réel, il demeure un certain nombre de réalités pratiques qui expliquent aussi une grande part de ce décollage tardif de l'usage des données raster dans les SIG.

La constitution

Pour les données du type image, en l'absence (encore récente) d'une offre sur étagère d'images numériques géométriquement correctes sur des territoires étendus, les besoins des utilisateurs ne peuvent être satisfaits que par des prestations ponctuelles coûteuses ou par l'emploi d'outils et de méthodes réclamant des savoir-faire peu répandus.

Réalisée à partir de clichés, une couverture orthophotographique numérique d'un territoire nécessite au moins trois étapes importantes et délicates :

numériser les clichés; les scanners actuels permettent une numérisation facile, avec des gammes de qualité allant d'une qualité bureautique (trop faible pour une utilisation géographique) jusqu'à des qualités métrologiques avec du matériel très coûteux.

les corriger géométriquement individuellement, en tenant compte des déformations introduites par l'appareil de prise de vue, celles inhérentes au relief (ortho-rectification), celles introduites lors de la numérisation ;

assembler (mosaïquer) les différentes images en gommant les bords de feuille, les superpositions, et en corrigeant les différences de saturation de

couleur, de contraste, pour ne faire qu'une seule image continue et homogène.

Autant dire que peu d'utilisateurs de SIG se sont lancés dans ces productions en dehors d'organismes dont c'est le métier (IGN, producteurs privés...).

Depuis quelques temps, la mise à disposition d'outils d'ortho-rectification laisse la possibilité d'effectuer cette opération pour des utilisateurs moins chevronnés. Cependant ceci restera limité à de petites productions.

Le stockage

La principale difficulté technique soulevée par l'utilisation des données raster est la consommation importante d'espace mémoire, en particulier pour en assurer la conservation (stockage).

Si le problème est largement connu depuis au moins dix ans avec l'utilisation de bases comme la BDAlti de l'IGN ou encore les images satellite qui représentent quelques dizaines de méga-octets (Mo) de mémoire pour une couverture départementale, le problème prend une autre dimension avec l'arrivée de données métriques voire submétriques.

Dans le cas d'une orthophotoplan couleur au pas de 0,50 m, au format TIFF à 216 couleurs, cette information s'élèverait à au moins 60 000 Mo (soit près de 100 CDROM) pour un département.

Dans le cas du plan cadastral (à l'origine ne comportant que des informations en noir et blanc) à l'échelle du 1/1000, une numérisation de qualité, moyenne impose 5 points (mailles) par millimètre, ce qui exige au minimum pour une surface des 5000 km² (un département moyen) de stocker environ 15 000 Mo d'information.

Des techniques de compression de données avec perte d'information permettent des gains

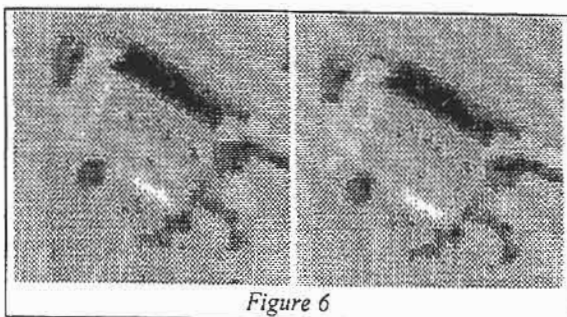


Figure 6

importants, ce qui permet d'atteindre des volumes plus habituels.

Le format de compression le plus courant et parmi les plus performants est dénommé JPEG (figure 6 : à gauche non compressé ; à droite JPEG compressé 20 fois). Il est bien adapté à des images "naturelles", beaucoup moins pour des images au trait (cadastre par exemple). Mais des méthodes de compression adaptées à chaque cas existent :

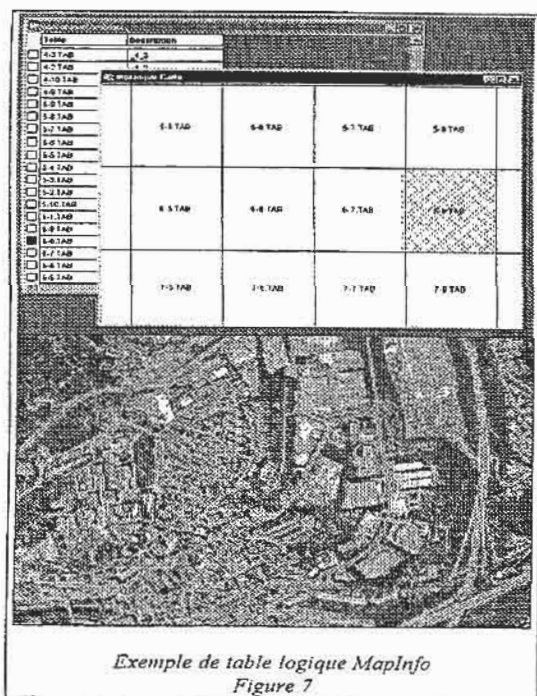
c'est un domaine encore en pleine évolution technologique (compression par ondelette, fractale, etc...).

Reste aux outils SIG à savoir les lire directement, avec des temps de réponse raisonnables. En effet, le stockage compressé ne règle pas le problème du chargement en mémoire de ces données, dans le format propriétaire des SIG.

Le tuilage

Afin d'éviter la manipulation d'une telle quantité d'information d'un seul tenant, on découpe l'image en "tuiles" (on parle également de "dalles") dont les dimensions sont choisies a priori : pour le producteur cela facilite la gestion des livraisons ; pour l'utilisateur cela permet de ne charger à un instant donné en mémoire que les "tuiles" réellement utiles.

La plupart des logiciels permettant de manipuler de larges couvertures territoriales autorisent une gestion en "tuiles" de la donnée raster : cette gestion des "tuiles" devient transparente à l'utilisateur (ERDAS, Multiscopie, ERMapper... MapInfo avec l'emploi de tables logiques) et



permet une utilisation confortable de ces données très lourdes à manipuler.

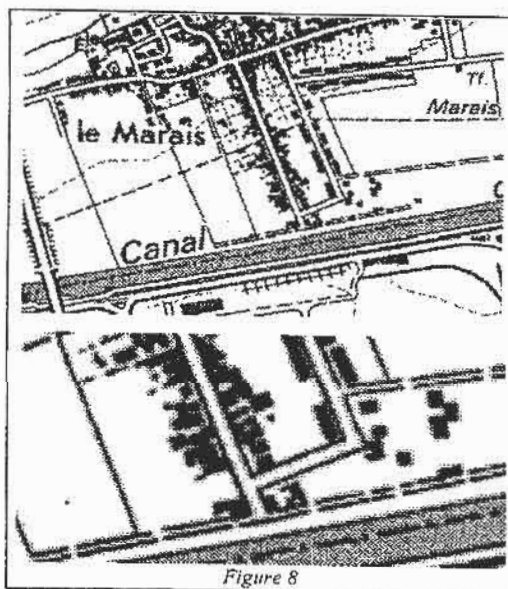
Les images satellitaires, les cartes scannées (fournies selon un "tuilage" correspondant aux coupures des cartes papier au 1/25 000), les orthophoto-plans ("tuiles" de 1 km²) sont ou seront fournies selon ce procédé.

Maille, précision, résolution

La taille de la maille est une information importante pour caractériser ce type de données. C'est même souvent la seule information caractérisant la finesse de la donnée : "pas" d'un modèle numérique de terrain, "pixels" d'une image satellite

Or, la taille de la maille ne donne d'indication que sur l'agrandissement maximum possible avant apparition des mailles (voir figure 8). En aucun cas la taille de la maille ne permet d'estimer une précision ou une résolution de la donnée.

Ainsi, une carte au 1/25.000 peut être scannée à 300 points par pouce ou à 1200, sans modifier la précision du résultat, qui est celle de la carte d'origine, parfois dégradée par le scannage. De même, une orthophotographie peut avoir une maille de 0,50 m au sol, mais n'avoir une précision annoncée que de quelques mètres ; cette dernière découle en effet de la rectification géographique indispensable pour passer de la photo à une donnée géographique.



Par ailleurs, la résolution, entendue comme la possibilité de différencier des objets sur l'image, sera très variable selon le type de donnée maillée. Sur une carte scannée avec un pas de numérisation adapté, la résolution sera celle de la carte et pas plus. Sur une photo aérienne, elle

sera dépendante du contraste entre deux objets limitrophes, et de leur forme : on peut voir sur une image à maille au sol de 0,50 m la signalisation horizontale sur une route, alors qu'elle a une largeur bien inférieure à la maille. A l'inverse, la frontière entre deux objets ne présentant qu'un faible contraste, et aux limites non régulières, ne pourra être déterminée qu'avec au moins deux mailles.

Par rapport aux données vecteur, on peut donc limiter l'échelle d'affichage des données maillées en jouant sur la taille de la maille. Cependant, les difficultés dues à l'utilisation de données de précision ou de résolution trop différentes restent les mêmes que pour des données vecteurs.

Conclusion

Toutes les difficultés qui viennent d'être évoquées ne doivent pas faire oublier les avantages majeurs que représentent la complémentarité des informations raster par rapport à des informations sous forme vecteur.

Les données vecteur et maillées peuvent constituer des ensembles très complémentaires et de plus en plus de professionnels adoptent cette solution. Ainsi, parmi les producteurs d'information, l'IGN propose des orthophotoplans numériques sur l'ensemble du territoire et la DGI a décidé d'une numérisation exhaustive des plans cadastraux sous forme d'image numérique. Parmi les utilisateurs, on peut citer les grandes collectivités locales et les principaux ministères

techniques, qui emploient de plus en plus des couvertures " image " de leur territoire, sous forme photo ou cartes scannées.

Le mode raster permet de mieux représenter des phénomènes à évolution continue, et reste l'outil le plus adapté pour les modélisations de phénomènes physiques ou assimilés. Le mode vecteur reste le plus adapté pour la saisie d'objets à limite franche, ainsi que pour toute application de gestion nécessitant des bases relationnelles. La possibilité de réaliser des traitements entre ces deux modes de données (quelles sont les valeurs des mailles se trouvant sur telle zone vecteur, trouver les zones vecteurs positionnées sur les mailles de telle valeur) devrait à terme permettre de choisir le mode de représentation le plus adapté pour chaque donnée, au mieux des usages souhaités.

Les données maillées permettent des approches différentes en terme de représentation du territoire. Elles ouvrent la voie à des modélisations plus fines que le mode vecteur (co-visibilité, impact de nuisances en fonction de l'éloignement, analyse multi-critère...), à condition de savoir quoi modéliser... et de s'en donner la peine !

Ces possibilités encore peu exploitées devraient dans l'avenir contribuer à un rééquilibrage entre le mode vecteur et le mode maillé.