

## DE LA QUALITE DE L'INFORMATION GÉOGRAPHIQUE À LA PERFORMANCE D'UNE FONCTION OPERATIONNELLE

SAPORITIN.

*GEO212, Geospatial information, PARIS, FRANCE*

### POSITION DU PROBLÈME

Les systèmes utilisateurs militaires (SIC, SIOC, Systèmes d'armes ou de navigation) sont devenus des consommateurs réguliers et exigeants de données géographiques. Pour les plus élaborés d'entre eux, la qualité de l'information géographique conditionne leur fonctionnement et leurs performances opérationnelles.

En parallèle, les efforts de la géographie militaire, notamment en France dans le cadre du programme DNG3D et de ses deux grands contrats de production (Géobase Défense et Topobase Défense) ou des réalisations du BGHOM, ont abouti à la constitution d'un patrimoine complémentaire et varié (Ortho-images, MNT, données vectorielles, cartes papier, etc.), sur de très larges zones du globe, et ce avec un niveau de qualité croissant.

Cependant, le dialogue reste limité entre les organismes en charge de la production et ceux qui spécifient les systèmes utilisateurs, et on apprécie aujourd'hui mal ce que la qualité de l'information géographique disponible a comme impact réel sur les performances de ces systèmes utilisateurs. Comment faire le lien entre les besoins des systèmes opérationnels et les productions de données ?

L'enjeu pour les utilisateurs opérationnels de l'information géographique est de concevoir des systèmes consommateurs de données géographiques, dont les performances sont connues et maîtrisées à un niveau de qualité de l'information géographique donné. Les producteurs d'information cherchent de leur côté à ajuster la production et la qualification des données géographiques pour répondre aux besoins des systèmes, sans faire de « sur qualité ».

### REFORMULATION DU PROBLÈME

Les organismes comme les instituts cartographiques nationaux, civils ou militaires ont pour objectif premier est de mettre à disposition des utilisateurs une cartographie généraliste indispensable. Mais ces données sont-elles capables de répondre à des besoins spécifiques de fonctions opérationnelles ? Peut-on établir un lien entre qualité de l'information géographique et niveau de performance de la fonction opérationnelle ?

Ce lien est bidirectionnel :

- Quel niveau de performance de la fonction permettent-elles d'atteindre avec un niveau de qualité de l'information donnée ? On parle de lien ou question direct(e).
- Pour atteindre un niveau de performance de la fonction quel est le niveau de qualité des données géographiques attendu ? On parle de lien ou question indirect(e).

De nombreuses questions induites émergent alors :

- Qu'est-ce qu'une fonction opérationnelle ?
- La performance d'une fonction opérationnelle... comment s'exprime-t-elle ?
- Qualité de l'information géographique... de quoi parle-t-on ?
- Comment établir cette fonction directe et indirecte ?

### L'APPROCHE MÉTHODOLOGIQUE

#### Introduction

L'approche méthodologique insufflée par le cahier des charges de l'étude s'appuie sur plusieurs phases :

- Analyser et définir la fonction opérationnelle, ses critères de performance et l'ensemble des paramètres, géographiques et non géographiques en entrée de la fonction.
- Définir leurs critères de qualité et les niveaux de qualité plausibles en fonction de leur méthode de « production ».
- Définir comment chacun de ces paramètres conditionnent les performances de la fonction et modéliser ces fonctions de transfert, puis implémenter ces fonctions de transfert dans le dispositif de simulation.
- Définir l'impact du niveau de qualité de ces paramètres sur les performances de la fonction :

1. Produire des paramètres de qualité variable et réalistes

2. Dérouler la fonction opérationnelle sur ces différents jeux de données « dégradés »

3. Mettre en évidence les paramètres dont les variations de qualité ont les impacts les plus forts sur les performances de la fonction

- Construire la relation « qualité des paramètres » - « performances de la fonction ».

**Définition de la fonction, de ses critères et niveaux de performance**

Dans notre étude, la fonction nous est imposée. Elle est définie de la façon suivante : « Etude de l'emploi de données vectorielles 2D et de MNT pour la fonction suivi d'un itinéraire pré calculé pour un convoi terrestre ». L'analyse sémantique renvoie principalement à la notion de calcul d'itinéraire, limité à des engins terrestres. Par extension, on ne considère que des déplacements sur axes et exclue les déplacements tout terrain.

Cela n'est pas suffisant et il faut approfondir l'analyse pour distinguer différentes sous-fonctions que nous rangeons en 2 catégories :

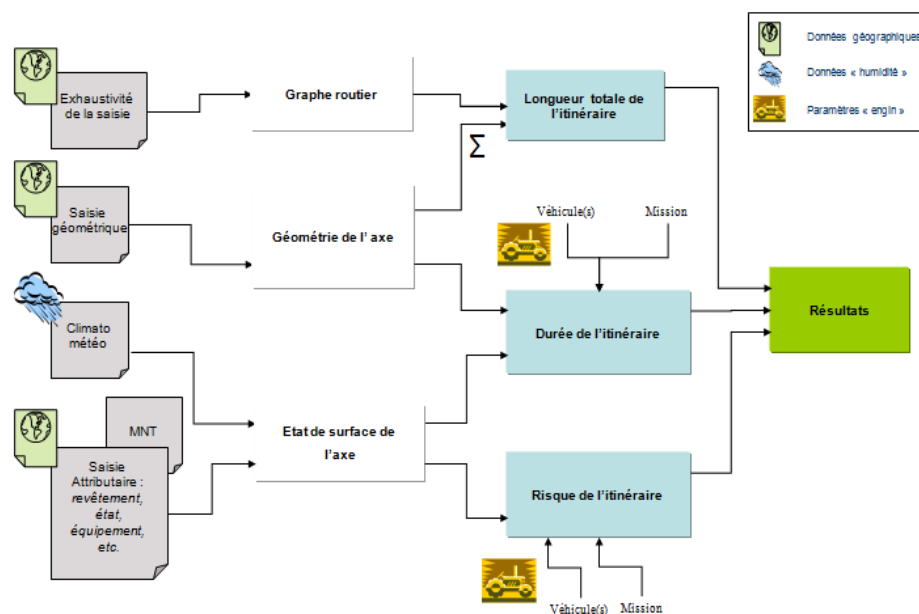
- Les fonctions opérationnelles : celles qui produisent le résultat attendu par l'utilisateur,
- Les fonctions de paramètres : celles qui fournissent l'ensemble des paramètres utilisés par les fonctions opérationnelles. Par rapport à l'enjeu de l'étude qui cherche à établir la relation entre qualité de l'information géographique et performances de la fonction (nous y reviendrons), les fonctions de paramètres doivent être elles-mêmes éclatées en fonctions de paramètres géographiques (qu'on peut confondre avec les données géographiques) et les fonctions de paramètres environnementaux et de paramètres de mission.

Cet approfondissement doit ensuite s'attacher à deux points essentiels :

- définir les critères de performances de la fonction,
- identifier les dépendances qui existent entre les différents paramètres et chacun des critères de performance. Ce point sera essentiel pour modéliser la fonction et fixer les niveaux de performance acceptables.

Dans notre cas, la fonction est ramenée à un calcul d'itinéraire optimisé selon 3 critères de performance que sont la longueur, le temps et le risque associé à l'itinéraire. Ils sont exprimés par un ratio entre la valeur retournée par la fonction compte tenu de l'incertitude sur les paramètres d'entrée et la valeur vraie.

Ensuite, les dépendances entre chacun des paramètres (notamment les données géographiques) et chacun des critères sont établies, comme le montre le schéma ci-dessous. Cette phase est importante car elle prépare à la modélisation mathématique de la fonction et permet de limiter le nombre de paramètres en entrée du calcul de chaque critère de la fonction. Par exemple (voir figure ci-dessous), le critère de longueur de l'itinéraire n'est influencé que par la géométrie élémentaire des tronçons et par l'exhaustivité du graphe.



Par contre, définir des niveaux de performance acceptables pour chacun de critères est difficile. Est-il acceptable de faire 20% de distance supplémentaire par rapport à la prévision/calcul ? 30 ? 40 ? Où est la limite ?

Pour mener à bien cette tâche les approches classiques d'analyse top-down, analyse fonctionnelle, etc. sont pertinentes.

Dans un cas où aucune définition n'est donnée, un travail de recueil de besoin et d'analyse fonctionnelle serait nécessaire auprès des utilisateurs finaux.

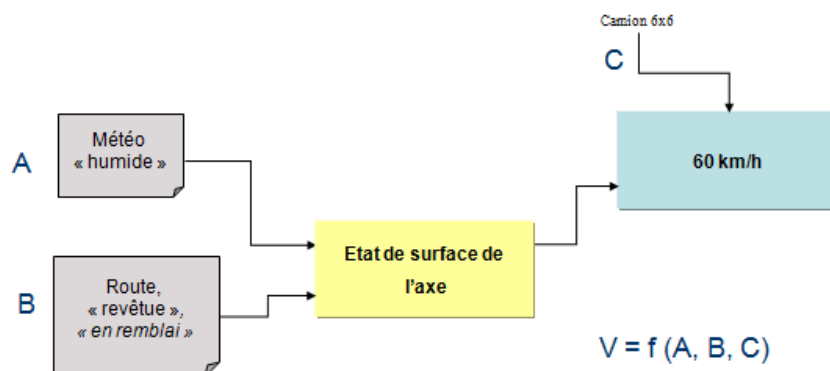
### Modélisation de la fonction

*Définition des fonctions de transferts des paramètres vers les critères de performance,*

Il résulte de l'analyse de la fonction, 3 critères d'optimisation : la distance, le temps, et le risque. Ainsi, étant donné deux points sur un graphe routier, on cherche à optimiser l'un des trois « critères » suivants :

- « distance »,
- « temps »,
- « risque ».

L'objectif de la phase de modélisation est de définir les fonctions mathématiques qui, à partir des paramètres en entrée, calculent la distance, le temps et le risque de chaque tronçon. Par exemple, comme le montre l'exemple ci-dessous, on définit le modèle de calcul de la vitesse en fonction de certains paramètres, notamment les caractéristiques géographiques des axes. Les calculs s'appuient sur les dépendances identifiées lors de la phase précédente. Dans cet exemple on illustre qu'un camion 6x6, se déplaçant sur une route revêtue et en remblai, a une vitesse moyenne de 60 km/h.



Cette fonction permet de calculer une vitesse pour chacun des tronçons du graphe. Ces « impédances » attribuées à chacun tronçon sont utilisées par l'algorithme de recherche d'itinéraire optimisé. En sortie de l'algorithme d'optimisation, on obtient un résultat numérique (distance en km, temps en minutes ou risque « sans unité ») et la liste des tronçons qui composent l'itinéraire optimal, comme l'illustre la figure ci-dessous.

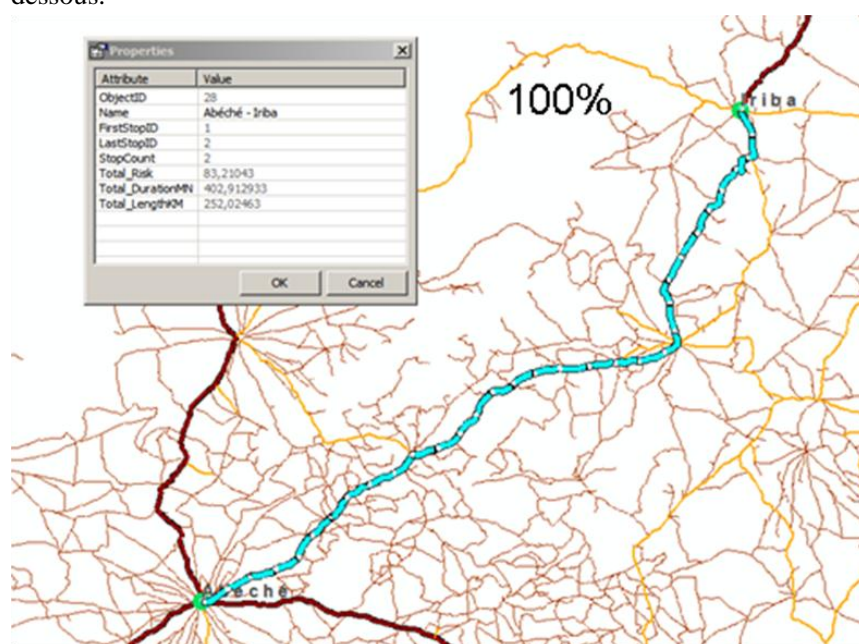


Figure 3 : Itinéraire Abéché – Iriba optimisé sur la longueur

La fonction « distance » dépend principalement du graphe et des caractéristiques géométriques de l'axe (forme, pente, ..). A partir des données géographiques de la Défense, il nous a été possible de réaliser simplement un premier niveau de modélisation de cette fonction prenant en compte juste le graphe. Nous n'avons pas considéré de contraintes sur les véhicules.

La fonction « temps » fait intervenir la longueur des tronçons et la vitesse sur ces tronçons. Cette dernière n'est pas une donnée native des données géographiques et son calcul est complexe car dépendant de nombreux paramètres d'état du tronçon et de la météo, ainsi que du véhicule et même du comportement du pilote. D'autres travaux sur la mobilité (PEA ECORS mené par le BRGM au profit du Ministère de la Défense) ont mis en évidence le besoin d'une connaissance très fine des propriétés mécaniques du sol (portance, adhérence, rugosité, etc.), d'humidité, etc., pour parvenir à calculer des vitesses fiables, par ailleurs très dépendantes des caractéristiques de véhicules. Certaines de ses propriétés ne peuvent être acquises que par des mesures in situ.

Le modèle utilisé repose sur le type de la route (route principale, route secondaire, piste) et sur l'attribut Revêtement (RST) – voir les coefficients retenus dans la figure ci-dessous. Ce dernier attribut est sémantiquement pauvre (uniquement deux valeurs possibles) et pas toujours bien renseigné. Par ailleurs, les coefficients de ce modèle tout en cherchant à être réalistes n'ont pas été étalonnés et nécessiterait un étalonnage. En outre le temps de passage d'un gué est pris en compte de façon forfaitaire, en affectant un temps de passage de 7, 14 ou 21 mn en fonction du type de tronçon.

### Caractéristiques de l'axe → coefficient :

- Type d'axe (route principale, secondaire, chemin) Chemin carrossable : x 70%
- Pente moyenne du tronçon de x 80% à x 10% selon la pente
- Revêtement Route non revêtue ou revêtement inconnu : x 80%

### Valeur standard :

- 60 Km/h si « Route & Revêtue & Plate »

### Cas particulier des gués :

- La longueur n'intervient pas
- La valeur standard est un temps
- Seul le type d'axe détermine le coefficient

### *Ces pondérations se combinent*

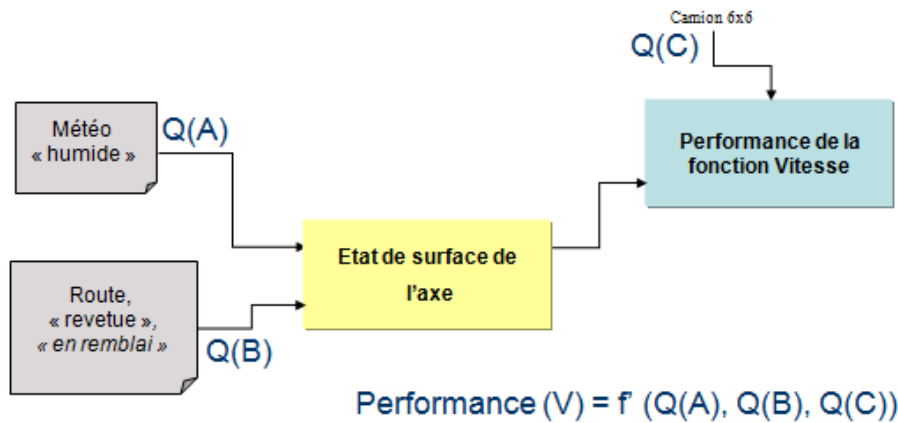
La fonction « risque » cherche à modéliser le risque, en prenant en compte non seulement les données géographiques, mais aussi les données météorologiques (ou plus précisément leur impact). La dépendance de la fonction aux données géographiques a été réalisée en utilisant les informations géographiques à notre disposition dans les données géographiques de Défense, à savoir le type de la route, la précision du tracé (attribut ACC) et la résistance aux intempéries (attribut WTC).

L'étalonnage des coefficients de cette fonction est lui aussi difficile et n'a pas été réalisé.

Les travaux menés dans le PEA ECORS et ceux conduits ici illustrent qu'un compromis entre réalisme et complexité de la modélisation de la fonction est toujours nécessaire. D'un côté on cherche à implémenter des fonctions complexes auxquelles on ne sait pas fournir des paramètres, de l'autre, on implémente des fonctions trop simplistes pour être pertinentes. Trouver le bon équilibre reste un enjeu fort de cette phase de modélisation.

### Impact de la qualité de l'information géographique sur la fonction

On cherche maintenant à étudier en quoi la qualité d'un paramètre en entrée impacte la performance de la fonction. On peut résumer cette question par le schéma suivant.



Cette tâche s'appuie sur les dépendances entre les critères de performance et les composantes géographiques isolées lors de la phase d'analyse de la fonction. En effet, le fait qu'on ait mis en évidence que le calcul de l'itinéraire le plus court (critère distance) ne dépend que de la géométrie des tronçons et de l'exhaustivité du graphe va nous permettre de déduire que seules la précision géométrique des tronçons et l'exhaustivité du graphe sont impliquées. On limite ainsi les relations à analyser.

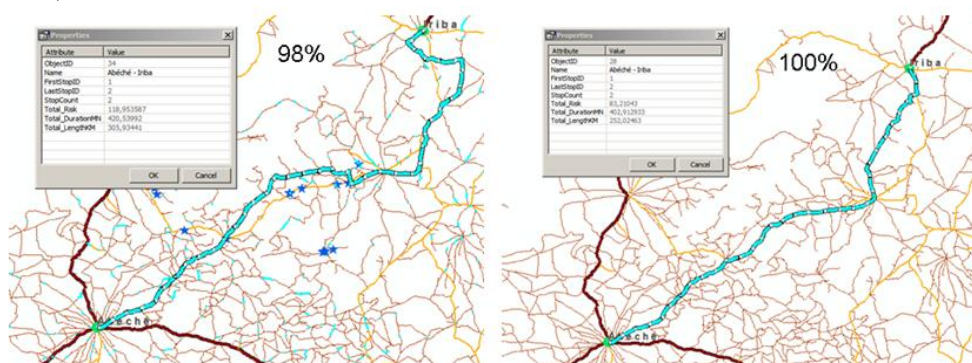
Par contre, cette démarche se heurte au fait qu'on ne sait pas exprimer mathématiquement (i.e. modéliser)  $Q(x)$ . Ceci tient au fait que trop de paramètres influencent la production des données géographiques (paysage, sources d'image, opérateur, etc.) et que  $Q(x)$  varie spatialement au sein d'un même jeu de données. En outre,  $Q$  peut dépendre de la population considérée. Par exemple, le taux d'omission sur les routes diffère du taux d'omission sur les pistes alors que les deux populations rentrent dans le calcul de la fonction.

Enfin, on sait au mieux calculer  $Q(X)$  pour une population limitée (par exemple on sait déterminer un taux de confusion – précision sémantique - pour un attribut d'une classe donnée), or le calcul de la fonction dépend lui de la qualité de  $n$  attributs. On ne sait pas le généraliser pour donner la précision sémantique de la population.

#### Méthode de contournement pour analyser cette dépendance

Cette méthode consiste à simuler des jeux de données dérivés d'un jeu de données « maître » ayant un état connu, puis de générer un lot de données de moindre ou de meilleure qualité, en faisant varier un ou plusieurs paramètres qualité, puis à injecter ces jeux de données en entrée du calcul de chemin optimisé.

Dans l'exemple ci-dessous, on calcule le chemin le plus court sur un jeu de données dont l'exhaustivité a été dégradée. On simule une dégradation de l'exhaustivité sans destruction physique des tronçons (on injecte un flag dans la base de données pour interdire 2% des tronçons du graphe – les tronçons en bleu dans la carte à gauche). A titre de comparaison, figure le même calcul pour un graphe « non dégradé » (à droite).



La dégradation s'appuie sur 3 facteurs :

- La population qui sera impactée,
- Le critère qualité,
- Les « lois et procédures » de dégradation/amélioration.

Un algorithme de dégradation/amélioration a été implémenté pour chacun des 3 critères qualité qui influencent les performances selon le graphe des dépendances défini : Précision géométrique, précision sémantique et exhaustivité.

La dégradation est injectée au travers d'un tirage aléatoire dans l'espace d'étude car comme mentionné précédemment, nous ne savons pas modéliser la variabilité de la qualité. Les résultats montrent d'ailleurs que le tirage aléatoire a un impact non négligeable selon l'itinéraire car les itinéraires ne sont pas tous identiques vis-à-vis du graphe routier de l'étude.

Les niveaux de dégradations résultent d'une connaissance a priori de niveaux de qualité plausibles et de contrôles qualité supplémentaires sur les jeux de données utilisés comme souches des simulations.

La limite de cette méthode est qu'elle ne permet de mener que des tests de sensibilité « atomiques » qui sont de plus fortement impactés par le tirage aléatoire. On ne sait donc pas analyser l'effet « combiné » des différents critères qualité qui caractérisent un jeu de données réel.

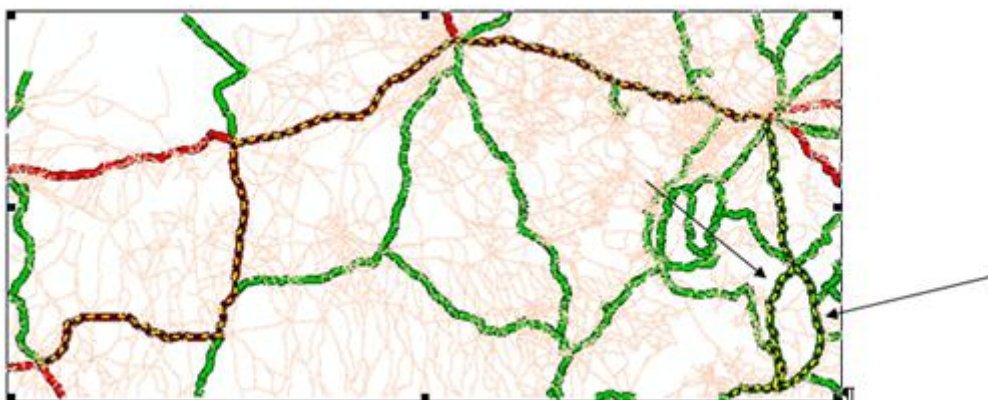
Pour limiter l'impact du tirage aléatoire, l'étude a mis en œuvre, via des simulations Monte Carlo, un niveau supplémentaire de tests, consistant, pour une même dégradation à « produire » 100 jeux de données. On déroule ensuite l'algorithme d'optimisation sur ces 100 jeux. On obtient alors deux éléments complémentaires et pertinents :

- La moyenne et l'écart type du critère de performance (par exemple le temps moyen de parcours de l'itinéraire et l'écart type de ce temps) à un niveau de dégradation donné, (ci-dessous, on voit pour les 5 itinéraires testés, l'impact de la variation de la précision sémantique – taux de confusion injecté de 10 % sur l'attribut WTC - sur le risque). On observe que les valeurs moyennes de performance à un niveau de dégradation donné, ont un écart type souvent faible et sont conformes au niveau de performance constaté sur un seul tirage aléatoire.

WTC	Risque (référence)	Moyenne	Ecart-type	Ecarttype (après bruitage) / Risque (référence)
<b>Abèché - Iriba</b>	80,8	78	5	6,2 %
<b>Adré - Sélibé</b>	74,6	74,6	0	0 %
<b>Am Yoma - Troatoua</b>	172,2	182	8,8	5,1 %
<b>Mongo - Touhanga</b>	337,2	337,2	0,6	0,2 %
<b>Oum Hadjer - Biltine</b>	65,3	62,4	11,3	17,3 %

10% de bruitage sur l'attribut WTC

- une représentation cartographique des 100 itinéraires optimisés (graphique ci-dessous). On observe une variation spatiale des itinéraires optimisés importantes, qui indique qu'à un même niveau de performance, les chemins optimisés peuvent suivre des directions significativement variables.



Enfin, une troisième méthode d'analyse et d'évaluation de l'impact de la qualité de l'information géographique a été implémentée. Elle consiste à faire varier la qualité non pas sur l'ensemble du réseau mais de façon concentrée sur un itinéraire choisi, qui dans un processus de décision aurait été retenu. Elle s'appuie de la même façon sur une méthode Monte Carlo et sur 100 tirages.

Ainsi la variation de qualité de l'information impacte entièrement sur l'itinéraire retenu. On réduit ainsi l'effet du tirage aléatoire qui « disperse » ou au contraire « concentre » la dégradation. On obtient là aussi une performance moyenne et son écart type, permettant d'apprécier directement l'impact d'un critère qualité sur un itinéraire donné.

S'elles n'apportent pas des réponses à toutes les questions posées, ces méthodes d'analyse d'impact enrichissent nettement les résultats et fournissent des éléments plus clairs d'aide à la décision en présentant les options d'itinéraires possible à un niveau de performances proches ou l'impact d'un critère qualité sur un itinéraire choisi. Cet intérêt a été souligné par des utilisateurs auxquels les résultats étaient présentés.

## LES TESTS

### Données utilisées

Les données géographiques disponibles pour cette étude couvrent une superficie de 15 degrés carré sur le Sud du Tchad. Nous avons eu trois jeux de données :

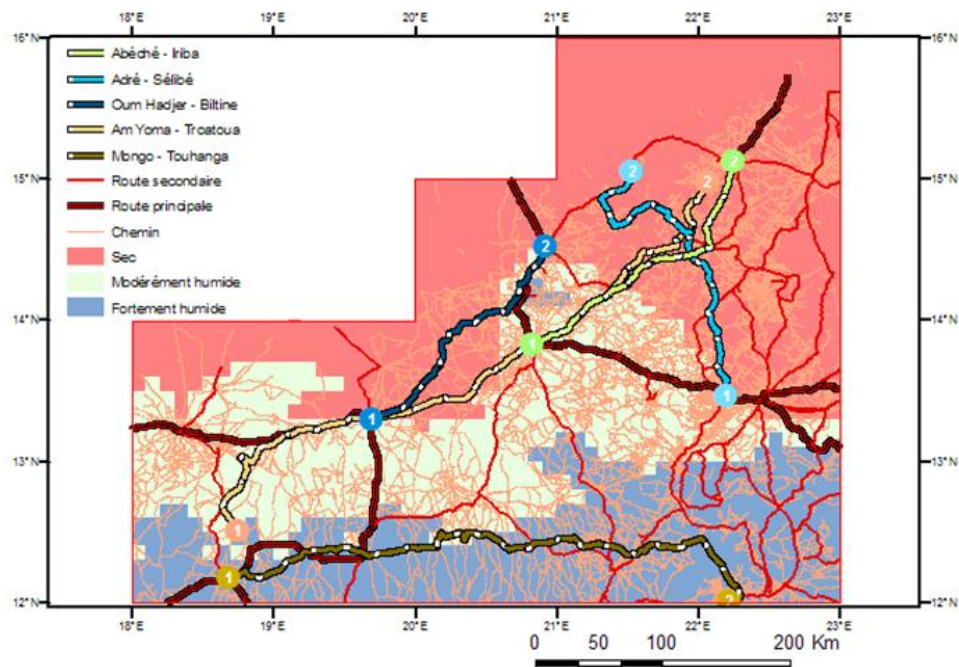
- VMAP1 cartographique,
- VMAP2i,
- VMAP1 réfecté issu d'une généralisation du VMAP2i.

Les échelles équivalentes du VMAP1 et du VMAP2 sont respectivement le 250 000ème et le 50 000ème.

Les 5 itinéraires ont été déterminés en fonction de la diversité de la zone suivant les paramètres influençant la fonction, ainsi que celle des données météorologiques disponibles sur la zone.

Les localités de départ et d'arrivée sont pour la plupart des agglomérations relativement importantes, à l'exception de Sélibé, Am Yoma, Troatoua et Touhanga. Trois des itinéraires sont orientés suivant la direction d'extension maximale de la zone :

- Abéché-Iriba est presque entièrement en montagne ;
- Am Yoma – Troatoua, le plus long de tous, est réparti entre plaine et montagne ;
- Oum Hadger – Biltine est entièrement en plaine.
- Adré – Sélibé, orienté perpendiculairement à cette direction, est situé de telle façon que deux trajets très différents l'un de l'autre sont possibles, l'un par la plaine, l'autre par la montagne.
- Mongo – Touhanga est un itinéraire Ouest-Est, dont la ligne directe à vol d'oiseau est entièrement située en zone très humide.



La zone de test et les 5 itinéraires avec en fond la carte d'humidité en saison humide

Au total, ce sont plus de 1500 jeux de données dégradés et de calculs d'itinéraires qui ont été réalisés, en jouant sur les 3 points suivants :

- Les 3 produits VMAP1 cartographiques, VMAP1 réfecté et VMAP2i.
- Pour les 3 critères : précision géométrique, précision sémantique et exhaustivité.

•Selon des niveaux de dégradations réalistes selon chaque critère.

### **Sensibilité de la fonction à la précision géométrique**

Les tests effectués ont mis en évidence que la géométrie élémentaire des tronçons n'a quasiment aucun impact sur les performances de la fonction, même pour la fonction « distance ».

Les tests de dégradations de la géométrie ont mis en évidence l'existence d'un seuil de rupture : il est extrêmement improbable que l'altération de la longueur des tronçons dépasse 2,5%, car les distorsions morphologiques induites seraient inacceptables. Comme des dégradations géométriques d'amplitude inférieure à ce seuil ont une influence minime sur la performance de la fonction, nous concluons que l'influence de la géométrie sur la performance de la fonction est négligeable.

### **Sensibilité de la fonction à l'exhaustivité**

L'impact de l'exhaustivité est nettement plus significatif que celui de la géométrie. Nous nous sommes principalement focalisés sur le paramètre d'omission même si un test d'ajout de franchissements (gués) a été effectué.

L'influence d'une perte (ou dégradation) d'exhaustivité sur la performance de la fonction est conditionnée par les caractéristiques intrinsèques du réseau. Si le réseau est à la fois dense et isotrope, des itinéraires quasiment équivalents seront identifiés pour contourner les ruptures que les dégradations auraient créées (toutefois jusqu'à un certain niveau de dégradation de l'exhaustivité).

Paradoxalement, les réseaux qui sont les plus denses et pour lesquels on trouve un résultat malgré des pourcentages d'omission élevé, sont ceux qui induisent les plus grandes amplitudes sur la longueur totale du trajet (jusqu'à 175%). En revanche, moins le réseau est dense, plus le niveau des pertes d'exhaustivité qui provoquent la rupture (impossibilité de trouver une solution) est bas. De même, si le réseau est anisotrope (localisé par exemple dans des fonds de vallée orientées du Nord au Sud), les itinéraires orientés de façon différente seront très fortement perturbés par des pertes d'exhaustivité.

Par ailleurs, l'étude a montré que l'exhaustivité des franchissements (dans la zone d'étude, ce sont principalement des gués qui sont omis mais ceci resterait tout à fait valide pour d'autres types : ponts, bacs, etc...) était cruciale pour que la fonction (i.e. l'optimisation des coûts) ne favorise des itinéraires irréalistes c'est-à-dire faussement « sécurisés » par l'absence de ces franchissements.

Par ailleurs, les tests sur l'exhaustivité dans l'étude de sensibilité ont montré qu'une sur-densité des réseaux n'apporte pas une amélioration des performances pour le critère de performance « distance ».

Enfin les tests ont montré que l'exhaustivité accidentelle avait plus d'impact que la généralisation.

### **Sensibilité de la fonction à la précision sémantique**

Dans les tests réalisés, la sémantique a un impact assez limité sur les performances des fonctions « temps » et « risque ». Ce constat doit cependant être pris avec beaucoup de précaution. On dispose, dans les produits géographiques utilisés pendant l'étude, d'attributs sémantiques « pas très adaptés » et leur prise en compte dans le modèle est sujette à discussion (peu de retour d'expérience sur cette modélisation).

Par exemple, on suppose (de par le modèle choisi) que le fait de rouler sur une route, avec RST « revêtu » permet de rouler à 60 km/h, alors qu'avec un RST « non revêtu », on ne peut rouler qu'à 48 km/h, ce qui est tout à fait simpliste. Le PEA ECORS mené en parallèle a mis en évidence que de nombreux attributs étaient nécessaires (adhérence, rugosité, portance, etc.) dont on ne dispose pas dans les produits VMAP. Donc on est « condamné » à sur-interpréter les informations sémantiques données par les VMAP. Dit autrement l'étalonnage de la fonction, et au-delà la maîtrise du modèle de calcul, reste un sujet majeur.

Les valeurs attributaires des données géographiques influent sur le résultat de la fonction, et donc sur sa performance, par l'intermédiaire des modèles qui calculent les indicateurs de coût sur lesquels portent l'optimisation des itinéraires (risque et temps de parcours). Un seul modèle a été utilisé au cours des tests pour calculer chacun de ces indicateurs.

Le modèle de fonction implémenté joue un rôle déterminant. Il est un amplificateur ou au contraire un amortisseur de l'impact de la qualité sur les performances de la fonction. Dans le cas de la fonction risque, le modèle fait varier le risque de 1 à 16 selon les attributs des axes et les conditions météo/humidité, alors que pour la vitesse on a des variations de l'ordre de 1 à 5.

On constate une impossibilité à ramener la sensibilité à un attribut à une seule courbe. A fortiori on ne sait pas aujourd'hui combiner les n courbes de sensibilité d'un même itinéraire selon les différents attributs qui impactent sur la fonction (par exemple F\_code et RST impactent la vitesse de déplacement, et génèrent deux schémas similaires à celui présenté ci-dessus).

En l'état, ce constat remet en cause l'idée de courbe d'impact unique reliant la qualité de l'information et la performance de la fonction pour un produit vecteur et une fonction comme celle étudiée. La



conséquence la plus forte est l'incapacité totale à définir la fonction inverse, c'est-à-dire de déterminer la qualité nécessaire et suffisante pour atteindre un niveau de performance de la fonction.

### **LES ENSEIGNEMENTS**

3 classes d'enseignements émergent de ces travaux.

Le premier touche aux données elles même et aux contrôles qualité. Il en ressort les constats suivants :

- L'exhaustivité de certaines classes est primordiale : Ponts et gués (les franchissements) ont un impact majeur. Les omissions et artefacts sur ces classes induisent des calculs totalement faux et des itinéraires dangereux. Ceci rejoint le concept de « vulnérabilité » décrit par Gleyze dans ses travaux de thèse.
- La densification du réseau routier doit être faite à un juste niveau en production : la sur-densification du réseau n'apporte pas d'augmentation de performance. Ceci est un enseignement important pour la production, dans son souci de ne pas faire de sur-qualité.
- La précision sémantique, compte tenu du modèle de fonction implémenté a moins d'impact.
- La précision géométrique n'a aucun impact sur les performances de la fonction.

Le deuxième enseignement touche aux apports opérationnels dans l'aide à la décision. Le dispositif mis en œuvre apparaît comme un outil utile en tant qu'aide au choix d'un itinéraire. Plusieurs raisons l'expliquent :

- Il offre une restitution des meilleurs résultats suivant différents critères de performance et sans se limiter au meilleur résultat seulement.
- Il présente les résultats (itinéraires retournés) dans leur environnement géographique et non seulement en termes de performance numérique.
- Il permet grâce au calcul d'impédance pour chaque tronçon une présentation graphique du poids de chaque tronçon sur les itinéraires (risques locaux forts).
- Enfin les méthodes d'analyse de l'impact de la qualité permettent :
  1. Une appréciation de l'impact des incertitudes des données géographiques et de la variabilité de la performance.
  2. Une représentation graphique de la fiabilité de chaque itinéraire (via les cartographies des itinéraires retournés par les méthodes Monte Carlo).
  3. De focaliser l'impact de l'incertitude sur un itinéraire choisi, en indiquant l'intervalle des performances de la fonction.

Le troisième est méthodologique :

- L'impact de la modélisation mathématique de la fonction est énorme. Selon le poids qu'on impute à un paramètre, on amplifie ou au contraire réduit l'impact de la qualité de ce paramètre géographique sur les performances de la fonction.
- Enfin, il n'est pas possible d'exprimer de façon synthétique la qualité d'un jeu de données (ce serait une combinaison des niveaux de qualité de chaque critère ?), en conséquence de quoi il n'est pas possible de faire un lien direct entre la qualité d'un jeu de données (avec sa complexité et ses nombreux indicateurs) et un niveau de performance de la fonction. Et a fortiori il est impossible de fixer un niveau de qualité en production pour atteindre un niveau attendu de performance de la fonction.

Quoique les enseignements de ces travaux soient riches, ils n'apportent que des réponses éclatées à la question initiale. Et en l'état nous pensons que la méthode suivie ne peut pas être améliorée pour appréhender la complexité des données vecteur et la combinatoire qui en découle au niveau du modèle de données, des critères qualité et des impacts multiples sur les performances de la fonction. Les alternatives devront donc s'attacher à simplifier cette complexité tout en gardant des modèles et fonctions réalistes.

Parmi les quelques pistes, on peut mentionner le fait d'agréger les informations aujourd'hui portées par chaque tronçon pour caractériser de façon plus simple des axes entiers (Ex : la route entre A et B est « sans danger » quelles que soient les conditions) ou encore l'idée d'aller vers des critères qualité synthétiques plus macroscopiques (à l'instar de la densité du réseau étudié pendant l'étude de sensibilité de l'exhaustivité).

Quoiqu'il en soit beaucoup reste à faire.

### **REMERCIEMENTS**

Les travaux mentionnés dans cet article ont été réalisés sur financement de la Direction Générale de l'Armement (DGA) du Ministère de la Défense dans le cadre du Programme d'Etudes Amont CALIFE Lot 2 confié à Thalès Communications et Géo212. Les données de test utilisées sont la propriété du Ministère de la Défense.

### **BIBLIOGRAPHIE**

- [Abbas 94] I. Abbas, « Bases de données vectorielles et erreur cartographique. Problèmes posés par le contrôle ponctuel ; une méthode alternative fondée sur la distance de Hausdorff », Thèse de doctorat, Université Paris 7, 1994
- [Bolstad 90] P.V. Bolstad, P. Gessler & T.M. Lillesand, « Positional uncertainty in manually digitized map data », *International Journal of Geographical Information Systems*, 4 (4), 1990.
- [Bonin 02] O. Bonin, « Modèle d'erreurs dans une base de données géographiques et grandes déviations pour des sommes pondérées ; application à l'estimation d'erreurs sur un temps de parcours », Thèse de doctorat, Université Paris 6, 2002.
- [Bonin 05] O. Bonin, « Modélisation des erreurs de position et d'attributs dans les bases de données géographiques », *Actes des Journées de Méthodologie Statistique*, 2005.
- [Fischer 91] P. Fischer, « Modeling soil map-unit inclusions by Monte Carlo simulation », *International Journal Geographical Information Systems*, vol 5, n°2, pp193-208, 1991.
- [GLEYZE 01] J.-F. Gleyze : Les dommages induits par les coupures du réseau routier - Eléments de recherche pour l'évaluation de la fiabilité d'un réseau routier, in *Actes de la conférence Risques d'accidents et risques environnementaux dans les transports routiers*, Besançon (F), 15 pages, laboratoire COGIT, IGN-SR-01-042-S-COM-JFG, octobre 2001.
- [GLEYZE 05] J.-F. Gleyze : La vulnérabilité structurelle dans les réseaux de transport dans un contexte de risque, Rapport de thèse, soutenue le 22 septembre 2005, laboratoire COGIT , UFR géographie Paris 7-, septembre 2005.
- [LAUR-SAPO 10] Thierry Laurençot et Nicolas Saporiti, rapport final du PEA CALIFE, DGA 2010.
- [Vauglin 97] F. Vauglin, « Modèles statistiques des imprécisions géométriques des objets géographiques linéaires », Thèse de doctorat, Université Marne-la-Vallée, 1997.