

ESTIMATION DE L'IMPRÉCISION DES MESURES GÉOMÉTRIQUES DE LONGUEUR ET DE SURFACE

Proposition d'un modèle destiné aux utilisateurs de données géographiques

par Jean-François Girres

Laboratoire COGIT

Institut national de l'information géographique et forestière

73 avenue de Paris, 94165 SAINT-MANDÉ

jean-francois.girres@ign.fr

De nombreuses applications SIG reposent sur des mesures de longueur ou de surface calculées à partir de la géométrie des objets d'une base de données géographiques, comme les calculs d'itinéraires routiers par exemple. Cependant, aucune information relative à l'imprécision de ces mesures n'est aujourd'hui communiquée à l'utilisateur. En effet, la majorité des indicateurs de précision géométrique proposés porte sur les erreurs de positionnement des objets, mais pas sur les erreurs de mesure. Dans ce contexte, cet article présente les résultats d'un travail de thèse, dont l'objectif est de mettre au point des méthodes d'estimation de l'imprécision des mesures géométriques de longueur et de surface, afin de renseigner un utilisateur dans une logique d'aide à la décision. Pour répondre à cet objectif, un modèle permettant d'estimer les impacts de règles de représentation et de processus de production sur les mesures de longueur et de surface est proposé. Une première réflexion sur la combinaison des impacts pour produire une estimation globale de l'imprécision de mesure est également étudiée, montrant l'intérêt d'une telle approche. Enfin, le modèle proposé est implémenté au sein du prototype EstIM (Estimation de l'imprécision des mesures).

Introduction

De nombreux systèmes d'aide à la décision sollicitent aujourd'hui l'information géographique par le biais de mesures de longueur et de surface. C'est, par exemple, le cas des calculs d'itinéraires routiers, qui nécessitent de connaître la longueur des tronçons de route. Ces mesures de longueur et de surface sont calculées directement à partir de la géométrie des objets d'une base de données géographiques vectorielles, c'est pourquoi nous les appelons mesures géométriques. Or, nous savons que ces bases de données constituent une simplification du monde réel, où, par commodité, les objets sont structurés géométriquement sous forme de points, de lignes et de polygones. Ainsi, nous pouvons considérer que ces mesures géométriques sont impactées par cette simplification inévitable du monde réel. Dans la pratique, l'utilisateur ne s'en soucie généralement pas, préférant assumer l'imperfection des

données géographiques et son impact sur les mesures géométriques. Cependant, il paraît fondamental d'estimer l'imprécision des mesures géométriques, afin d'informer l'utilisateur des risques encourus lors de la manipulation des données. Mais une telle estimation constitue une entreprise complexe, du fait de la diversité des processus accumulés dans l'erreur finale de mesure.

La modélisation de l'impact de l'imprécision géométrique a fait l'œuvre de nombreuses contributions ces dernières années. Si de nombreux modèles ont été proposés afin de modéliser cet impact en terme de positionnement des objets (voir par exemple Chrisman, 1982 ; Caspary et Scheuring, 1992 ; Shi et Liu, 2000 ; Hunter et Goodchild, 1996), il n'en va pas de même en ce qui concerne la modélisation de l'imprécision des mesures géométriques. Les principaux modèles développés dans ce domaine sont ceux de Chrisman et Yandell (1988), Griffith (1989), Kiiveri

(1997) et Leung *et al* (2004). Ces différents modèles proposent d'estimer la variance des mesures géométriques de longueur ou de surface, en partant d'une hypothèse commune, relative à la propagation de l'imprécision de positionnement de chaque point de la géométrie de l'objet sur l'ensemble de la mesure géométrique calculée. Ces modèles présentent cependant un certain nombre de limites. Tout d'abord, la dimension altimétrique des objets n'est pas prise en compte, ce qui implique une sous-estimation systématique des mesures calculées. De plus, ces modèles considèrent que la répartition des erreurs de mesure est homogène dans l'ensemble de la base de données étudiée. Ainsi, deux objets identiques, peu importe leur localisation dans l'espace, présenteront une imprécision de mesure identique. Pourtant, des expérimentations simples prouvent le contraire. Par exemple, des tronçons de routes localisés en zone de montagne présentent généralement une imprécision de mesure plus importante que ceux localisés dans des reliefs de plaine. Enfin, ces modèles ne prennent pas en compte les spécificités des processus de production mis en œuvre pour saisir la géométrie des objets, et des erreurs associées, en fonction de la précision des données sources (cartes, photographies aériennes, images satellites) ou du matériel utilisé (acquisition par GPS différentiel ou directe, appareil de restitution photogrammétrique ou PC classique).

Ainsi, cet article présente les résultats d'un travail de thèse, dont l'objectif est d'élaborer un modèle permettant à un utilisateur d'estimer l'imprécision des mesures géométriques de longueur et de surface. Pour cela, nous proposerons tout d'abord un cadre conceptuel, permettant d'identifier les composantes de l'erreur de mesure géométrique, et ainsi de mieux comprendre les différents processus accumulés. Nous présenterons ensuite la mise en œuvre du modèle proposé, permettant à un utilisateur de données géographiques d'estimer l'imprécision des mesures géométriques.

1 Cadre conceptuel des composantes de l'erreur de mesure

La modélisation de l'imprécision des mesures géométriques de longueur et de surface suppose de prendre en compte l'ensemble des processus pouvant potentiellement les impacter. Nous présentons dans cette section les différents processus identifiés, permettant d'établir un cadre conceptuel des composantes de l'erreur de mesure géométrique. Pour cela, l'ensemble des opérations permettant de représenter les entités du monde réel sous forme d'objets géographiques a été étudié. Deux

processus principaux ont ainsi été différenciés : le processus de simplification du monde réel, qui définit les règles de représentation des objets, et le processus de production de la géométrie des objets, qui intègre les moyens d'acquisition et les données sources.

1.1 Règles de représentation du monde réel

La représentation du monde réel sous forme d'objets géographiques implique un certain nombre de simplifications qui influent directement sur les mesures géométriques de longueur et de surface. Trois impacts sont étudiés dans le cadre de ce travail de thèse : la projection cartographique, la non-prise en compte du terrain, et l'approximation polygonale des courbes.

1.1.1 Projection cartographique

Pour représenter la surface de la Terre dans un plan (qui demeure la forme de représentation la plus commode sur cartes papier ou sur un écran d'ordinateur), différentes phases de simplification sont opérées, tout d'abord par une approximation de la forme de la Terre sous forme d'ellipsoïde, celui-ci étant ensuite projeté sur un plan à l'aide d'une projection cartographique. L'impact de cette dernière, en particulier dans le cadre de projections conformes (comme la projection Lambert 93), implique une déformation de la surface terrestre, ayant pour effet de modifier les mesures de longueur et de surface. Ainsi, pour estimer l'impact de la projection cartographique, nous proposons de calculer des mesures géométriques sur l'ellipsoïde de référence, en déterminant en chaque point l'altération linéaire impliquée par la projection utilisée.

1.1.2 Non-prise en compte du terrain

De nombreuses bases de données géographiques n'intègrent pas de dimension altimétrique dans la modélisation géométrique des objets représentés. Ceci a pour impact de sous-estimer systématiquement les mesures géométriques de longueur et de surface calculées. Pour pallier cette absence, nous proposons d'enrichir d'une information altimétrique les objets des bases de données n'en disposant pas. Ainsi, en utilisant des modèles numériques de terrain existants, et en appliquant des mesures adaptées en 2D5, l'enrichissement géométrique des objets permet d'estimer l'impact de la non-prise en compte du terrain.

1.1.3 Approximation polygonale des courbes

Dans le cas où les entités du monde réel présentent des formes courbes (comme les virages d'une

route par exemple), la représentation géométrique des objets sous forme de ligne polygonale génère une sous-estimation systématique des mesures géométriques de longueur. C'est notamment le cas des routes présentant une forte sinuosité. Afin d'estimer l'impact de l'approximation polygonale des courbes sur les mesures de longueur, nous proposons de reconstruire une géométrie courbe réaliste sur les objets le nécessitant, par le biais de méthodes d'interpolation polynomiales existantes, comme les splines cubiques d'Hermite par exemple.

1.2 Processus de production des objets géographiques

De nombreux systèmes d'acquisition sont aujourd'hui disponibles pour saisir la géométrie des objets géographiques. Dans le cadre de ce travail de thèse, deux impacts des processus de production ont été étudiés : l'erreur de pointé de l'opérateur, et la généralisation cartographique.

1.2.1 Erreur de pointé de l'opérateur

L'erreur de pointé de l'opérateur est une erreur accidentelle, qui intervient lors de processus de restitution manuels, comme les relevés photogrammétriques, la digitalisation de photographies aériennes, ou encore lors de relevés topographiques. Cette erreur est généralement tolérée, voire compensée par le biais de méthodes d'ajustement, comme les méthodes par moindres carrés. L'erreur de localisation en chaque point des objets de la géométrie, généralement modélisée par le biais d'une loi normale indépendante en x et en y , a pour effet de se propager sur l'ensemble de la géométrie, et d'ainsi impacter les mesures géométriques de longueur et de surface. Afin d'estimer son impact, une approche par simulation de l'erreur de pointé est proposée, permettant de la reproduire de manière réaliste, et ainsi d'étudier la sensibilité des mesures géométriques calculées en aval. Pour cela, la connaissance de l'échelle moyenne de saisie demeure un préalable indispensable, puisqu'elle conditionne directement l'amplitude de l'erreur de localisation de chaque point de la géométrie.

1.2.2 Généralisation cartographique

L'impact de la généralisation cartographique doit également être pris en compte lorsque les objets d'une base de données sont saisis à partir de cartes (comme c'est par exemple le cas de la BDCARTO de l'IGN), ou bien lorsque ceux-ci sont dérivés d'une autre base, par le biais de processus de généralisation automatique (Ruas, 2002). L'impact de la généralisation cartographique se manifeste par des modi-

fications importantes de la forme, position et taille des objets, ce qui influence les mesures géométriques calculées. Pour estimer cet impact, nous proposons d'exploiter une approche par tests. Cette approche se base sur des comparaisons entre des bases de données généralisées à différentes échelles et une base de données de référence (n'ayant pas subi l'impact de la généralisation cartographique). L'objectif de cette approche est d'élaborer un modèle statistique, permettant d'estimer une longueur de référence en fonction d'une longueur calculée à partir d'objets généralisés. Pour cela, différentes hypothèses relatives aux connaissances des processus de généralisation sont intégrées. Par exemple, pour les géométries linéaires, nous considérons que l'erreur de longueur varie en fonction de la classe d'objets étudiée, leur échelle de représentation et leur contexte géographique. Les résultats montrent que les erreurs constatées sur les tronçons de routes localisés en zone de montagne présentent des erreurs de longueur plus importantes que leurs homologues localisés en zones urbaines ou rurales (Girres, 2011).

Après avoir présenté les composantes potentielles de l'erreur de mesure au sein d'un cadre conceptuel global, et les méthodes permettant d'estimer leur impact, nous présenterons dans la section suivante leur intégration au sein d'un modèle général d'estimation de l'imprécision des mesures géométriques, dédié à l'utilisateur de données géographiques.

2 Modèle d'estimation de l'imprécision des mesures géométriques

Cinq modèles d'impacts, associés aux règles de représentation et aux processus de production présentés précédemment, ont été développés dans le cadre de cette thèse. Ces modèles ont été intégrés au sein d'un modèle général permettant d'estimer l'imprécision des mesures géométriques sans données de référence (fig. 1).

2.1 Principe général d'estimation

2.1.1 Acquisition des connaissances sur les données

Étant donnée la non-présence systématique de données de référence pour estimer l'imprécision des mesures géométriques, nous proposons d'élaborer un modèle basé sur l'acquisition de connaissances sur les données évaluées. En effet, les cinq modèles d'estimation des impacts des règles de représentation et des processus de production supposent un paramétrage adapté aux données évaluées. Pour

cela, nous proposons d'enrichir les données évaluées de connaissances acquises par le biais de différentes sources d'information (données externes, connaissances de l'utilisateur, métadonnées, méthodes d'exploration des données).

Par exemple, pour estimer l'impact de la projection cartographique, des grilles d'altération linéaire sont produites et utilisées. Pour l'impact de la non-prise en compte du terrain, des modèles numériques de terrain ou d'élévation peuvent être aisément mobilisés. En ce qui concerne la modélisation de l'impact de l'erreur de pointé, la connaissance de l'échelle de saisie (déterminée par l'utilisateur ou par des méthodes d'estimation automatique) est un préalable indispensable. Enfin, pour estimer l'impact de la généralisation cartographique, la délimitation du contexte géographique des objets peut être réalisée par le biais de données de zonage (extraites de bases de données d'occupation du sol par exemple). L'échelle de représentation des objets généralisés peut, quant à elle, être déterminée par l'utilisateur ou par le biais de méthodes d'estimation automatiques.

L'intégration de ces informations offre la possibilité d'enrichir la qualification des données évaluées, et ainsi de paramétrer le modèle afin de pallier l'absence de données de référence.

2.1.2 Communication des résultats de l'estimation

Pour communiquer à l'utilisateur les résultats de l'estimation de l'imprécision des mesures géométriques sur les données évaluées, deux stratégies de communication sont envisagées. La première repose sur une estimation différenciée des différents impacts modélisés par les cinq modèles associés aux règles de représentation et aux processus de production. Cette stratégie offre un intérêt pédagogique évident, et permet d'autant plus d'estimer le poids porté par chacune des composantes accumulées dans l'erreur finale de mesure. Cependant, dans un second temps, nous envisageons de communiquer une estimation globale de l'imprécision des mesures géométriques calculées, prenant en compte les différentes causes d'erreurs associées aux données géographiques évaluées, en définissant un intervalle de confiance autour de la mesure géométrique estimée.

2.2 Combinaison des impacts : principe et expérimentation

De nombreuses méthodes de combinaison des impacts peuvent être envisagées. Dans le cadre de ce travail, nous avons choisi de développer des méthodes standard et d'évaluer les plus efficaces sur

des cas réels. Ainsi, trois méthodes de combinaison ont été proposées afin d'estimer une erreur globale Δt à partir des différents impacts Δi . La première repose sur une agrégation simple (une addition) des différents impacts. La seconde repose sur une agrégation quadratique des impacts. Enfin, la troisième repose sur la sélection de l'impact le plus significatif, c'est-à-dire l'impact maximal. Bien que nous ayons conscience des limites de ces méthodes de combinaison, nous considérons qu'elles fournissent néanmoins une approximation réaliste de l'erreur de mesure géométrique globale.

Pour évaluer la méthode de combinaison la plus efficace, une expérimentation a été menée. L'expérimentation proposée porte sur la comparaison des longueurs de vingt itinéraires tirés aléatoirement du réseau routier de la BDCARTO, en comparaison avec leurs homologues extraits de la BDTOPO enrichie d'une géométrie courbe (afin de prendre en compte l'impact de l'approximation polygonale des courbes en plus de celui du terrain). Pour évaluer la méthode de combinaison la plus performante, une analyse par moindres carrés est utilisée. L'objectif consiste à déterminer la méthode de combinaison des impacts permettant de minimiser l'erreur de longueur résiduelle estimée, en comparaison avec l'erreur de référence observée (qui correspond à 1,288% de la longueur calculée sur les itinéraires de la BDCARTO). Les résultats montrent que pour les itinéraires de la BDCARTO, la méthode de combinaison par agrégation quadratique s'avère être la plus performante avec une erreur résiduelle estimée de 0,975%, en comparaison des méthodes par agrégation simple (1,329%) et par sélection de l'impact le plus significatif (1,001%). Cette expérimentation montre surtout que la méthode par agrégation quadratique permet de réduire de manière significative l'erreur de référence.

Bien que la méthode par agrégation simple ait pour impact d'augmenter l'erreur résiduelle estimée, cette méthode demeure intéressante dans le cadre de la formulation d'un intervalle de confiance, étant donné que dans 85% des cas, la longueur estimée avec cette méthode de combinaison est supérieure à la longueur de référence de la BDTOPO enrichie. Ainsi, suite à cette expérimentation, pour le réseau routier de la BDCARTO, nous proposons de formuler un intervalle de confiance de la manière suivante :

- Longueur minimale estimée = Longueur initiale calculée.
- Longueur estimée = Longueur initiale + agrégation quadratique des impacts.

- Longueur maximale estimée = Longueur initiale + agrégation simple des impacts.

Une telle formulation permet de fournir à l'utilisateur une estimation globale de l'imprécision de la mesure de longueur calculée à partir du réseau routier de la BDCARTO. Une expérimentation similaire a été réalisée sur les données TOP100 (utilisées pour produire les cartes touristiques TOP100), montrant que dans le cas des données fortement généralisées, la méthode la plus efficace pour minimiser l'erreur résiduelle estimée est la méthode par agrégation simple des impacts.

2.3 Le prototype EstIM

Afin de mettre en œuvre le modèle d'estimation proposé dans le cadre de ce travail de thèse, le prototype EstIM (Estimation de l'imprécision des mesures) a été développé sous la forme d'une extension du logiciel libre OpenJUMP. Le prototype EstIM se présente sous la forme d'un formulaire général, permettant de sélectionner les différents impacts que l'utilisateur souhaite évaluer. Pour faciliter le paramétrage du modèle, différentes méthodes d'acquisition des connaissances sur les données évaluées (estimation des échelles de saisie ou de représentation des données, délimitation du contexte géographique des objets) ont été intégrées dans l'interface homme-machine. Enfin, la communication à l'utilisateur a été envisagée de manière textuelle, en listant chacune des contributions à l'erreur finale par les différents modèles d'impacts, et sous forme d'estimation globale

en utilisant les méthodes de combinaison des impacts proposées précédemment (fig. 2).

2.4 Conclusion et perspectives

Cet article a permis de présenter les résultats d'un travail de thèse relatif à la conception d'un modèle d'estimation de l'imprécision des mesures géométriques de longueur et de surface pour un utilisateur de données géographiques. L'originalité du modèle proposé est qu'il cherche à estimer l'imprécision de mesure sans données de références, mais en cherchant à évaluer les impacts liés aux causes de l'erreur de mesure. En ce sens, ce travail permet de disposer d'une meilleure compréhension des composantes de l'erreur globale de mesure, à travers un cadre conceptuel différenciant règles de représentation et processus de production affectant les mesures géométriques. Également, une première réflexion sur la combinaison des impacts a été proposée, montrant que même des méthodes simples de combinaison permettent de réduire l'erreur de mesure initiale, et ainsi d'approximer une mesure réelle. Enfin, ce travail de modélisation a été implémenté au sein d'un prototype de logiciel opérationnel, permettant à un utilisateur d'estimer de manière conviviale l'imprécision des mesures calculées. Malgré ces contributions, l'enrichissement du modèle d'impacts supplémentaires (comme l'imprécision des acquisitions GNSS) doit être envisagé, de même que l'affinement du modèle en prenant en compte les spécificités des objets étudiés et en caractérisant davantage leur contexte géographique.

Bibliographie

- Caspary W. et Scheuring R.**, 1992, "Error-band as Measures of Geographic Accuracy", dans *Proceedings of EGIS'92*, p. 226-233.
- Chrisman N.R.**, 1982, "A Theory of Cartographic Error and its Measurement in Digital Data Bases", dans *Proceedings of AUTOCARTO*, 5, p. 159-168.
- Chrisman N.R. et Yandell B.S.**, 1988, "Effects of Point Error on Area Calculations: A Statistical Model", *Surveying and Mapping*, n°48, p. 241-246.
- Girres J.F.**, 2011, "An evaluation of the impact of cartographic generalization on length measurements computed from linear vector databases", dans *Proceedings of ICC'11*.
- Griffith D.A.**, 1989, "Distance calculations and errors in geographic databases", dans Goodchild, M. F. et Gopal, S.: *The Accuracy of Spatial Databases*, p. 81-90, Ed. Taylor & Francis.
- Hunter G.J. et Goodchild M.F.**, 1996, "A New Model For Handling Vector Data Uncertainty in Geographic Information Systems", *URISA Journal*, n°8, p. 51-57.
- Kiiveri H.T.**, 1997, "Assessing, representing and transmitting positional uncertainty in maps", *International Journal of Geographical Information Science*, n°11, p. 33-52.

Leung Y., Ma J.H. et Goodchild M.F., 2004, "A general framework for error analysis in measurement-based GIS Part 4: Error analysis in length and area measurements", *Journal of Geographical Systems*, n°6, p. 403-428.

Ruas A., 2002, « Échelle et niveau de détail », dans Ruas, A.: *Généralisation et représentation multiple*, p. 26-44, Ed. Hermes, Lavoisier.

Shi W. et Liu W., 2000, A stochastic process-based model for the positional error of line segments in GIS, *International Journal of Geographical Information Science*, n°14, p. 51-66.

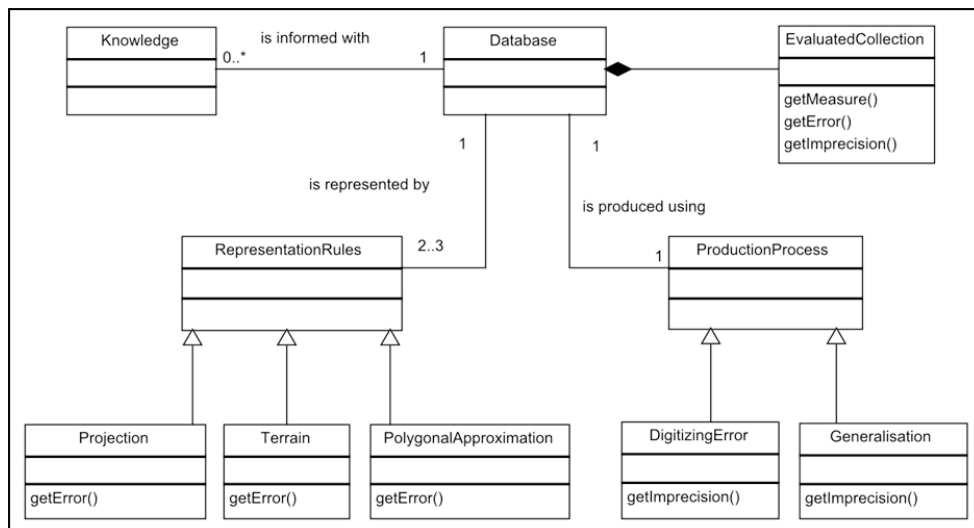


Figure 1 : Diagramme de classe UML du modèle d'estimation

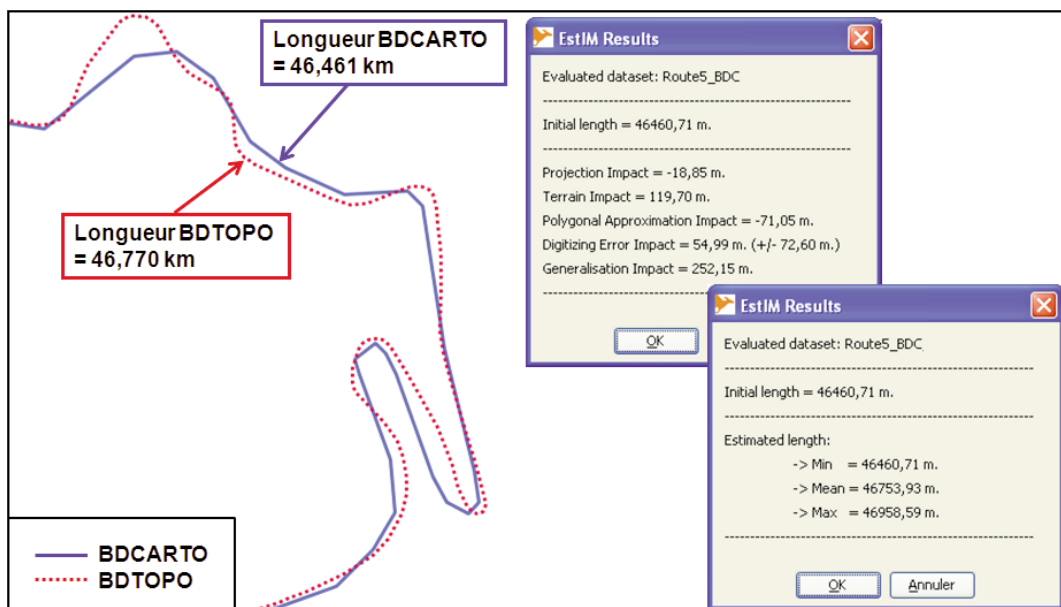


Figure 2 : Résultats d'estimation par le prototype EstIM sur une route de la BDCARTO (en bleu) par comparaison avec son homologue de la BDTOPO (en pointillés rouges)