

AU SUJET DES INDICATEURS RETENUS POUR LA QUALIFICATION DES DONNÉES GÉOGRAPHIQUES



Qualifier les données géographiques
Un décryptage de la norme ISO 19157

Critère de précision de position

L'usage des données ouvertes et géolocalisées et la précision d'usage croissent et à venir nous rend progressivement producteurs et utilisateurs de données géographiques.

Les activités régaliennes et les politiques publiques s'appuient sur de l'information multi-sources et la qualité des données produites ou utilisées devient un enjeu indissociable. Pour autant, tout le monde ne dispose pas des moyens des producteurs institutionnels de données et il paraît utile de formuler des recommandations et des méthodes plus adaptées au contexte de chacun, pour qualifier les données géographiques, communiquer sur les résultats obtenus voire savoir les interpréter. C'est l'objectif que s'est fixé le Cerema en proposant cette collection de fiches, à l'attention des producteurs et des usages.

Cette fiche présente le critère de qualité sur la précision de position. Elle est utile pour les données géographiques produites à vocation d'un usage à petite, moyenne ou grande échelle.

1. Définition de la norme

La précision de position se définit comme la précision de la position des entités ou des points d'un système de référence spatial. Elle se compose de deux composantes :

- la précision absolue (ou absolue) : mesure des valeurs de coordonnées absolues par rapport au référentiel spatial ou géométrique en tant que telles ;
- la précision relative (ou relative) : mesure des positions relatives des entités dans un jeu de données par rapport à leurs positions vraies, respectives ou mesurées au sein que telles.

Fiche n° 10 - décembre 2017

collection | Connaissances

-> critère de précision de position



Qualifier les données géographiques
Un décryptage de la norme ISO 19157

Critère de précision thématique

L'usage des données ouvertes et géolocalisées et la précision d'usage croissent et à venir nous rend progressivement producteurs et utilisateurs de données géographiques.

Les activités régaliennes et les politiques publiques s'appuient sur de l'information multi-sources et la qualité des données produites ou utilisées devient un enjeu indissociable. Pour autant, tout le monde ne dispose pas des moyens des producteurs institutionnels de données et il paraît utile de formuler des recommandations et des méthodes plus adaptées au contexte de chacun, pour qualifier les données géographiques, communiquer sur les résultats obtenus voire savoir les interpréter. C'est l'objectif que s'est fixé le Cerema en proposant cette collection de fiches, à l'attention des producteurs et des usages.

Cette fiche propose d'expliquer un des critères qualité de la norme ISO 19157 retenus dans cette collection de fiches, la précision thématique. Elle s'attache à qualifier le contenu des attributs présentés par les objets ainsi que la justesse de classement des objets dans leur famille de rattachement. C'est un critère qualité important chaque fois que l'application des données dépasse le simple fait d'identifier la seule présence d'un objet sur le territoire, que ce soit pour des représentations cartographiques ou pour des exploitations statistiques par exemple.

Il est donc nécessaire de maîtriser l'évaluation de la précision thématique et de connaître les indicateurs permettant de partager cette évaluation entre producteurs et utilisateurs.

1. Définition de la norme

La précision thématique se définit comme la précision des attributs quantitatifs, et la justesse des attributs non quantitatifs, du classement des entités et de leurs relations.

Elle se compose de trois éléments de qualité des données :

- la justesse du classement : comparaison des classes attribuées aux entités ou à leurs attributs par rapport à l'univers du discours (par exemple, monde réel ou données de référence) ;
- la justesse des attributs non quantitatifs : mesure permettant d'établir si les valeurs d'un attribut non quantitatif sont correctes ou pas ;
- la précision des attributs quantitatifs : précision de la valeur d'un attribut par rapport à la valeur vraie ou reconnue comme vraie.

Fiche n° 09 - décembre 2017

collection | Connaissances

-> critère de précision thématique



1

le critère de précision de position

PROBLEMATIQUE :

Quel indicateur pour qualifier le positionnement d'une couverture polygonale ? (hors travaux topographiques)

1 - critère de précision de position



page 5

On recommandera donc d'utiliser les mesures suivantes en fonction du contexte et par ordre de priorité :

- classe de précision au sens de l'arrêté du 16 septembre 2003 ;
- valeur moyenne des incertitudes.

L'inté
basée
comp
quad

4. Méthode de contrôle

Ce paragraphe s'applique à décrire l'ensemble des

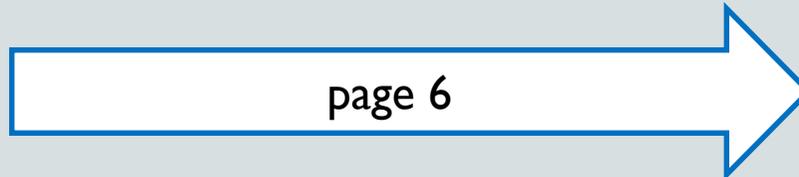
■ Obj

- **L'emploi des 2 indicateurs retenus peut apparaître contraignant, en particulier pour qualifier le positionnement d'une couverture polygonale**
 - notamment si les données géographiques ne relèvent pas de travaux topographiques
 - et lorsque le travail de mesure pour la qualification apparaît disproportionné



d'où la question : est-il possible de retenir un autre indicateur ?

1 - critère de précision de position



- Dans la fiche méthodologique :
- la question de la qualification du positionnement d'objets linéaires et surfaciques est abordée
- ...mais la qualification est toujours renvoyée à la mesure du positionnement de points



• caractérisée d' « **exercice délicat** »



• des objets surfaciques qualifiés de « **périmètres flous** »

■ **Objets linéaires et surfaciques**

Identifier le long d'une ligne, ou d'un périmètre des points caractéristiques, est un exercice délicat.

Plusieurs solutions sont possibles en fonction des configurations, qui nécessitent des actions préparatoires plus ou moins complexes :

- réduire la précision de position d'un réseau à la précision de position des sommets de ce réseau dès lors que leur identification et leur localisation ne présentent pas d'ambiguïté (privilégier les carrefours en X ou en T d'un réseau routier plutôt que les jonctions en biseau) ;
- relever, dans le jeu de données et dans la source de contrôle des points représentant des changements de cap (pour un réseau électrique par exemple constitué de lignes brisées, pour une limite administrative, pour un périmètre basé sur le découpage parcellaire...) ou des inversions d'inflexion pour un réseau constitué de lignes courbes (réseau routier ou ferroviaire, courbe de niveau, zone d'aléa...).

Configurations à éviter : points de contrôle flous (réseau à contrôler en rouge), source de contrôle (noir)

Configurations à privilégier : points de contrôle identifiables

4.2 In

La plupart sont co

d'un réf

saisie su

moins p

don't on

géomé

En cas

partielle

BD PARC

de posit

tement

ce cont

naires n

les part

avec ce

Dans le

tograph

produit

conditi

effectu

contrôle

la précis

Pour ra

pour la

référen

Photogr

métriqu

Levé GP

PARCELL

Orthoph

Carte 1/

satellite

BD CAR

1 - critère de précision de position

➤ Un contrôle certes difficile pour un objet surfacique, pourtant une méthode adaptée (simplifiée ?) devrait pouvoir être proposée

Des réponses possibles :

➤ **Le dénombrement des écarts à la norme**

- écart : écart maximum de positionnement constaté sur chacun des polygones - ou parties de polygone - d'un échantillon soumis à contrôle
- norme : spécifications d'un produit déterminant a priori un(des) seuil(s), limite(s) que les incertitudes de positionnement ne devraient pas dépasser



1 - critère de précision de position

Le dénombrement des écarts à la norme

➤ soit comme cité dans la fiche méthodologique,

la mesure n°30 de la norme 19157 (page 10)
« nombre d'incertitudes
au dessus d'un seuil donné »

OU

la mesure n°31 de la norme 19157 (page 2)
« taux d'erreurs de position au-dessus
d'un seuil donné »

METHODE

*mise en œuvre sur un
échantillon de polygones
(ou parties de polygone) :*

- ▶ analyse de la qualité du positionnement par rapport à un seuil préalablement défini (dans les spécifications) pour chacun des polygones contrôlés
- soit pas d'écart à la norme
- sinon mesure de l'écart maximum constaté sur ce polygone

nombre ou % poly présentant un écart < seuil

nombre ou % poly (écart > seuil)

0

seuil

écart maximum
constaté sur chacun
des polygones contrôlés

1 - critère de précision de position

Le dénombrement des écarts à la norme

➤ Une autre réponse possible :

**un indicateur cadrant un peu plus
la dispersion des incertitudes**

(version simplifiée de la méthode mise en œuvre
pour des travaux topographiques / Arrêté du 16 septembre 2003)

**le classement des écarts à la norme
selon deux seuils**

METHODE

*mise en œuvre sur un
échantillon de polygones
(ou parties de polygone) :*

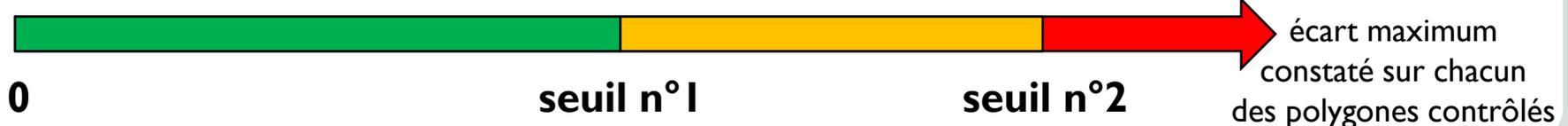
▶ *analyse de la qualité du
positionnement par rapport à
deux seuils préalablement
définis (dans les spécifications)
pour chacun des polygones
contrôlés*

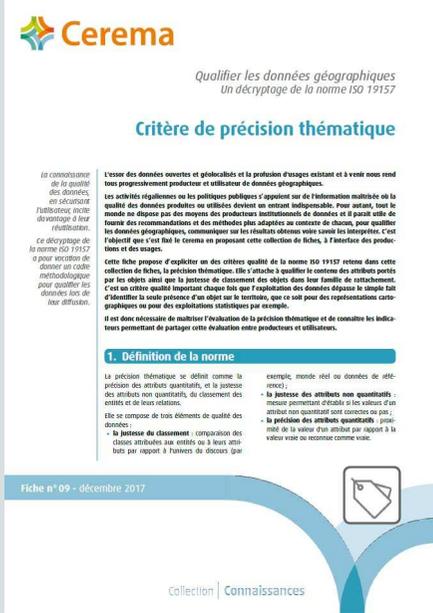
- *soit pas d'écart à la norme*
- *sinon mesure de l'écart
maximum constaté sur ce
polygone*
- *rejet si un écart maximum
est supérieur au seuil n°2*

nbre ou % poly (écart < seuil n°1)

nbre ou % poly (écart > seuil n°1)

et pas de poly avec écart > seuil n°2





2

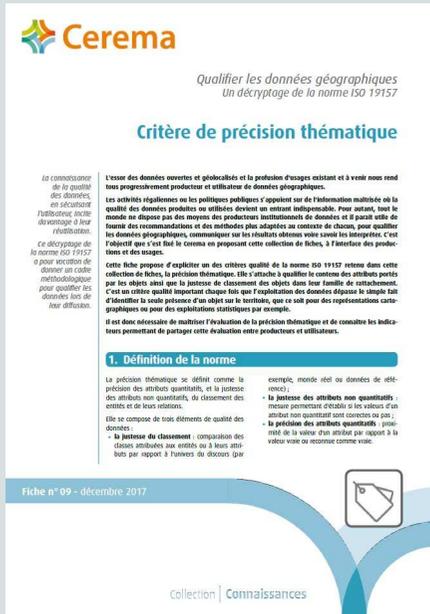
le critère de précision thématique

PROBLEMATIQUE :

Quel indicateur pour qualifier la justesse d'un classement ?

2 - critère de précision thématique

INDICATEURS RETENUS



page 4

La qualification du lot de données sera indiquée :

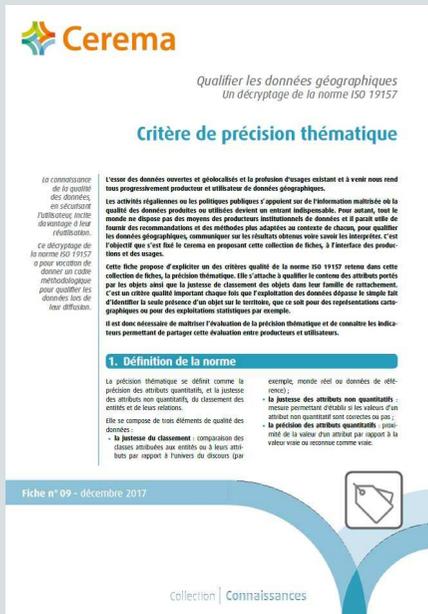
- prioritairement par le taux global de bon classement (...)
- secondairement par le coefficient kappa

Par contre, dans le projet de Registre CNG des “métadonnées relatives à la qualité des données géographiques” :
sont mentionnées les deux mesures qui seraient intégrées dans les métadonnées,

- la matrice relative de classement erroné (RMCM)
- le coefficient kappa

d'où la question : pourquoi cette différence entre indicateurs retenus et mesures intégrées dans les métadonnées ?

2 - critère de précision thématique



Pour résumer, dans la présente contribution sur les moyens de qualifier la justesse de classement, sont abordés les points suivants :

- l'importance de mettre en avant la « mesure de base » que représente la « *matrice de classement erroné* » (MCM, ou « *matrice de confusion* »), au lieu de la RMCM
- la recommandation de prioriser l'utilisation - comme indicateur - du « *taux global de bon classement* » ou « *taux d'accord global* »
(simple et robuste)
- la non recommandation de l'utilisation du « *coefficient kappa* »
(interprétation non évidente, résultats non comparables, comportements imprévisibles, ...)
- De plus, dans la fiche méthodologique, des erreurs ou imprécisions sont relevées, notamment en termes de définition et d'utilisation de la « *matrice « relative » de classement erroné* » (RMCM, ou « *matrice de confusion « relative »* »)

- **En conclusion :**
- **Plus généralement,**
- **sont ici posées des questions de restriction du choix des indicateurs**

- **dans les fiches méthodologiques : les indicateurs retenus ou recommandés**

- **dans les métadonnées : les indicateurs possiblement retranscrits dans les métadonnées**

- **...alors qu'il est utile de répondre à l'ensemble des différents types de productions de données géographiques,**
- **sachant que les exigences en termes de qualité (et de contrôle de la qualité) sont toutes aussi diverses et variées.**

Merci de votre attention

