RX-06

Un inventaire automatique dont le total brut était faux de deux ordres de grandeur

Retour d'expérience. Un algorithme correct, appliqué à une donnée correcte, a produit un inventaire dont 66 % du volume était un artefact d'infrastructure.

Famille Retours d'expérience	*Généricité* GÉNÉRIQUE	*Statut* EXPLORÉ — production du corpus, non validée par une décision
Écrit le 17 septembre 2026	*Licence* CC BY-SA 4.0	

L'attendu

Inventorier les dépressions fermées du secteur — mares, fosses, micro-cuvettes — à partir du modèle numérique de terrain LiDAR HD de l'IGN, pour pré-localiser les habitats humides à prospecter. Méthode standard : remplir les dépressions par inondation prioritaire, soustraire la surface remplie du terrain d'origine, et compter ce qui reste.

Le déroulé

L'algorithme a rendu **180 dépressions totalisant 3,5 millions de mètres cubes** sur 576 hectares. Soit, rapporté à la surface concernée, une profondeur moyenne de près de quatre mètres sur soixante-dix-neuf hectares de plateau granitique. Le chiffre est absurde, mais il fallait le rapporter à quelque chose pour le voir : pris seul, « 180 dépressions » ne choque pas.

Deux contrôles ont suffi.

La forme. La plus grande dépression couvre 45 hectares dans une boîte englobante de 1 180 × 1 972 mètres qu'elle n'occupe qu'à 19 % — un ruban sinueux, logé entre 540

et 552 mètres, c'est-à-dire au point le plus bas de l'emprise. À elle seule, elle porte 58 % du volume. Une dépression fermée est compacte ; un ruban de deux kilomètres est un fond de vallée.

Le recoupement. Rastérisation des ponts, de la voie ferrée et des plans d'eau cartographiés dans OpenStreetMap, puis test de contact avec chaque dépression :

	part du volume total
jouxte un pont ou la voie ferrée	**66 %**
recouvre un plan d'eau cartographié	15 %

Le mécanisme est connu et il est structurel : **un modèle numérique de terrain restitue les remblais mais pas les ouvrages qui les traversent.** Tout busage, tout pont, toute digue ferroviaire barre la vallée dans le modèle. L'eau, dans le calcul, ne passe pas — elle s'accumule jusqu'à déborder par-dessus le remblai, et l'algorithme enregistre une retenue de plusieurs centaines de milliers de mètres cubes là où il y a une buse de quatre-vingts centimètres.

Après filtrage sur la compacité, la population de véritables dépressions fermées est de **59 objets, 2,7 hectares, 24 000 m³** — deux ordres de grandeur sous le chiffre brut.

Pourquoi rien ne l'a signalé

L'algorithme est correct. L'inondation prioritaire fait exactement ce qu'elle promet : elle trouve toute cuvette d'où l'eau ne peut sortir sans monter. Un fond de vallée barré par un remblai *est* une telle cuvette, dans le modèle.

La donnée est correcte. Le MNT LiDAR HD est complet sur l'emprise — vérifié, aucun pixel manquant — et restitue fidèlement la surface du sol, remblai compris. Il ne lui est pas demandé de connaître les buses.

L'erreur est dans la correspondance entre le modèle et le monde, et cette correspondance n'est écrite nulle part dans la chaîne. Personne ne la déclare, donc personne ne la vérifie.

Et le format de sortie masque le problème. Un total en mètres cubes n'a pas d'échelle intuitive. Ce qui a alerté, c'est de le ramener à une profondeur moyenne par hectare — une grandeur dont on sait à quoi elle ressemble.

Les leçons

1. **Un total agrégé ne se publie pas sans sa distribution.** Ici, un seul objet portait 58 % du volume. Toute somme dominée par un ou deux éléments décrit ces éléments, pas la population. **Regarder qui pèse, avant de publier combien.**

2. **Ramener tout agrégat à une grandeur dont on connaît l'ordre de grandeur.** « 3,5 millions de m³ » ne dit rien ; « quatre mètres de profondeur moyenne sur un plateau » se réfute en une seconde. Le contrôle de vraisemblance ne marche que sur des grandeurs familières, et la conversion est presque toujours possible.

3. **La forme d'un objet trahit sa nature avant toute mesure.** Compacité, allongement, taux de remplissage de la boîte englobante : trois nombres qui distinguent une mare d'un fond de vallée sans rien savoir d'hydrologie. Quand un inventaire mélange des natures, la géométrie sépare souvent mieux que le seuil.

4. **Croiser avec une source indépendante, même approximative.** OpenStreetMap n'est pas une référence topographique, et le test de contact est un test de coïncidence, non une preuve. Il a suffi : quand 66 % d'un volume touche des ouvrages qui couvrent une fraction infime de la surface, la coïncidence n'en est pas une. **Un croisement imparfait vaut mieux qu'un contrôle absent.**

5. **Un modèle physique a des hypothèses tacites sur le monde.** L'inondation prioritaire suppose que l'eau ne circule que par la surface. Tout ouvrage souterrain — buse, drain, karst, canalisation — viole cette hypothèse sans que rien ne le signale. **Écrire les hypothèses tacites du modèle est la seule façon de les tester.**

Les changements

- Tout inventaire automatique publie, avec son total, la **part portée par ses trois plus gros éléments**.
- Tout agrégat est **converti en une grandeur familière** avant publication, et le contrôle de vraisemblance porte sur cette conversion.
- Les objets d'un inventaire sont **caractérisés par leur forme** — compacité, allongement — et les classes de forme sont examinées séparément.
- Tout résultat issu d'un modèle physique s'accompagne de la liste de ses **hypothèses tacites sur le monde**, et d'au moins un test de la plus fragile.

- Un croisement avec une source indépendante précède la publication, même quand cette source est moins précise que la donnée principale.

Portée

Générique. Le même piège guette tout inventaire produit par un algorithme correct sur une donnée correcte : le comptage de bâtiments sur un modèle qui confond hangar et couvert d'arbres, la détection de haies qui compte les alignements de pylônes, la mesure de surfaces imperméabilisées qui range l'ombre avec l'asphalte, l'extraction de réseaux hydrographiques qui prend un fossé de route pour un ruisseau.

Un algorithme juste, sur une donnée juste, peut rendre un résultat faux — quand le modèle et le monde ne se correspondent pas là où personne ne l'a écrit.

Fiches liées

RX-03 · RX-04 · RX-05 · PA-02 · ANALYSE-LIDAR

CoMons, Collectif du Mons — RX-06, Un inventaire automatique dont le total brut était faux de deux ordres de grandeur. Fiche du Grain, CC BY-SA 4.0. Version en ligne : comons.actitude.org/grain-rx-06-inventaire-faux-de-deux-ordres.html