

INFOS PRESSE

Grâce à l'IA, l'IGN vigie du territoire

L'IA : une alliée de choix pour cartographier l'Anthropocène

L'Anthropocène appelle à un **effort accru et rapide de description des transformations subies par les territoires du fait des activités humaines.**

Pour fournir des référentiels cartographiques sur des **enjeux écologiques majeurs** (suivi de l'occupation des sols, potentiel de biodiversité, état de santé des forêts, recul du trait de côte...), les innovations numériques et en particulier les techniques d'intelligence artificielle, sont des **leviers technologiques indispensables** qui se sont imposés très tôt à l'IGN dans l'automatisation de la production. L'institut s'est ainsi doté, en 2022, d'une [feuille de route](#) « Démocratiser l'IA » pour déployer massivement cette technologie.

Désormais, l'IA occupe une place centrale dans 10 productions de l'institut et son usage se généralise dans ses chaînes de production. Les **technologies de machine learning, de deep learning et d'IA générative** ont accéléré la cartographie à partir de données de télédétection (vues aériennes, satellites, Lidar...). Les algorithmes obtenus par les méthodes d'apprentissage machine sont décisifs pour accélérer la production de données et montrer les changements en cours.

En plus des équipes de recherche ayant pris le virage de l'IA depuis plus de dix ans, l'IGN renforce ses compétences et connaissances en IA et compte aujourd'hui **30 ingénieurs spécialisés**, contribuant aux activités d'innovation, de développement et de production. L'Institut investit également dans les compétences d'avenir via son école l'ENSG Géomatique qui donne plus de place à la data et à l'IA, et qui s'accompagne de l'expansion du corps des ingénieurs des sciences géographiques et du numérique.

La capacité générique d'observation de l'institut est en augmentation grâce à un **mixte technologique** permettant de croiser données satellitaires, aériennes, LiDAR HD, etc. et données collaboratives. La multiplicité des acteurs publics et privés produisant de l'information géographique rend aussi indispensable un travail en commun, en réseau et en animation de communautés pour développer des solutions innovantes sur la base de données d'autorité, en réponse à des enjeux sociétaux.

Parce que la technologie est aussi un enjeu démocratique, l'IGN est engagé dans une **politique d'ouverture de ses données, méthodes et outils relatifs à l'IA** : partage des données d'apprentissage, des modèles entraînés, etc. Avec 20 milliards de pixels, le jeu de données FLAIR était lors de sa mise à disposition par l'Institut l'un des trois plus riches jeux de données ouverts pour l'occupation des sols au monde, de même ampleur que DeepGlobe (Facebook, MIT) et LoveDA (Université de Wuhan). De son côté, le *dataset* PureForest est le plus grand jeu de données LiDAR au monde sur les essences forestières (lire p.6).

Ce partage est porteur de retombées importantes pour les communautés scientifiques et techniques, facilite la prise en main de données aux structures complexes (LiDAR, séries temporelles d'images) et, surtout, réduit les volumes de données d'apprentissage nécessaires au développement de nouveaux systèmes IA. Or, ce besoin en données massives constitue encore l'un des principaux déterminants de la performance et du coût de ces systèmes.

Exemples de grandes productions IGN s'appuyant sur l'IA

- **L'OCS GE: pour détecter et suivre l'artificialisation du sol**

En s'appuyant sur ses données socles, l'IGN opère¹ la production d'un référentiel à grande échelle, tous les trois ans, pour la description de l'occupation du sol sur le territoire national : l'OCS GE (occupation du sol à grande échelle). Au service de l'État et des collectivités, cette cartographie montre en détail l'occupation des sols selon plusieurs éléments clés pour en distinguer les zones imperméables, agricoles, forestières. Elle permet de quantifier et de qualifier l'évolution des territoires et leur artificialisation nette.

Depuis 2019, l'institut emploie la télédétection d'objets (habitations, végétation, etc.) par IA pour accélérer la production et multiplier les usages. Des modèles obtenus par apprentissage profond analysent les images aériennes et extraient une 1^{ère} estimation de la couverture des sols. Ces cartes de "prédiction", d'une haute résolution de 20 cm par pixel sont diffusées sous le nom de [CoSIA](#) (Couverture du Sol par Intelligence Artificielle).

Ces premiers résultats font l'objet de traitements ultérieurs et de croisements avec des données existantes (données forestières, foncières ou relatives aux aides agricoles) pour produire la donnée finale et mesurer l'artificialisation.

Quelques chiffres :

- *D'importants volumes de données d'apprentissage (~ 2000km²) de grande qualité doivent être produits pour l'entraînement des modèles, mais ceux-ci sont sans commune mesure avec la superficie du territoire national dans son intégralité.*
- *Les performances du modèle IA et des traitements géomatiques associés permettent d'atteindre une bonne classification des couvertures et des usages du sol supérieurs à 73 % pour les premiers départements produits.*

- **[CarHab](#), cartographie des habitats naturels et semi-naturels terrestres**

Initiée dans le cadre de la stratégie nationale pour la biodiversité 2011-2020, CarHab est une modélisation cartographique nationale des habitats naturels et semi-naturels pour les écosystèmes terrestres de métropole et d'Outre-mer, à l'échelle du 1 : 25 000^e, avec pour objectif de couvrir la France entière d'ici 2026 (avec mise à jour tous les 6 ans). Ce programme partenarial² porté par le ministère en charge de l'écologie vise à répondre aux enjeux de conservation de la biodiversité, d'aménagement du territoire et de gestion durable des ressources naturelles. CarHab fait appel à une modélisation par *machine learning* mobilisant des données de végétation existantes, des techniques d'analyses d'images. Une phase de prospection terrain permet d'affiner le modèle.

- **Le programme national [LiDAR HD](#)**

Dans le cadre du programme national LiDAR HD, l'IGN produit et diffuse une cartographie 3D de l'intégralité du sol et du sursol de la France en données LiDAR, précise, détaillée et homogène. Les données diffusées sont des nuages de points recalés, bruts ou classifiés, et des

¹ Ce référentiel est produit par l'IGN en collaboration avec le Cerema et l'INRAE à la demande du ministère de la Transition écologique et de la Cohésion des territoires pour répondre à l'exigence du plan « Zéro Artificialisation Nette ».

² L'IGN, les conservatoires botaniques nationaux, PatriNat, l'Unité mixte de recherche Environnement-Ville-Société (EVS-ISTHME), l'Office français de la biodiversité (OFB), avec l'appui du CesBio, du BRGM et du Cerema.

modélisations numériques 3D. Grâce à un process automatique combinant les méthodes usuelles de classification, le croisement avec des bases de données existantes, et l'IA (algorithmes de *deep learning* utilisant des données d'apprentissage), les nuages de points 3D acquis sont classifiés en plusieurs classes (sol, eau, végétation, bâtiments, ponts, sursol pérenne). Les nuages de points donnent ensuite lieu à la production de MNT (modèles numériques de terrain), MNS (modèles numériques de surface) et MNH (modèles numériques de hauteur).

Quelques chiffres :

- *Les données produites vont représenter un volume total de **3 pétaoctets** (soit 3 millions de giga-octets).*
- **10 points par mètre carré** en moyenne
- **80 % du territoire français couvert (mi-2024)**

Exemples de nouveaux grands projets qui vont s'appuyer sur l'IA

La BD Forêt et le masque forêt – non forêt

La BD forêt V2 était produite de façon traditionnelle grâce à la photo-interprétation humaine. Une étude des besoins auprès de multiples utilisateurs a montré la nécessité d'aller vers une production de données plus détaillée, et des mises à jour plus fréquentes. Fort de son expérience en *deep learning*, l'IGN a initié des travaux préliminaires qui ont montré toute la puissance de cette technologie dans la discrimination des essences. L'IGN s'engage aujourd'hui auprès du Ministère en charge de l'agriculture pour une production de cette BD Forêt France entière en trois ans environ.

Une première étape a été de produire le masque forêt – non forêt dont une première version est en diffusion. La seconde version comprendra la détermination des landes, données très sensibles pour la prévention des risques d'incendies. Ce masque est une donnée d'entrée pour permettre l'identification des zones soumises à OLD (Obligations légales de débroussaillage).

Une seconde étape a été de réaliser une expérimentation basée sur l'usage du LIDAR HD pour la discrimination des essences, permettant de générer un data set, nommé « PureForest », inédit dans le monde scientifique (voir ci-dessous).

L'observatoire des haies

Ce projet a pour objectif de mieux cartographier les haies afin de les préserver. Cela répond à une directive européenne sur la biodiversité. Une première base de données avait été produite en concaténant des données de sources variées. Le besoin d'améliorer la complétude, la géométrie, la fréquence de production a été rapidement ressenti pour répondre aux enjeux des politiques publiques. L'évolution de la méthode était donc nécessaire ; les premiers démonstrateurs réalisés en *deep learning*, en s'appuyant sur des orthophotographies, mais aussi sur le Lidar permettent d'une part de produire les données sur les haies mais également d'identifier les arbres isolés.

Le bâti de la BD Uni

La BD Uni est une description vectorielle 3D (structurée en objets) des éléments du territoire et de ses infrastructures, de précision métrique, exploitable à des échelles allant du 1 : 2 000 au 1 : 50 000. Des études ont montré le potentiel du *deep learning* pour le dessin des périmètres de chaque bâtiment. Des échanges techniques avec la direction générale des finances publiques sur l'algorithme Frame Field a permis de faire progresser les résultats. La mise en production devrait se faire à court terme.

Exemples de jeux de données ouverts massifs et de qualité pour l'IA

L'IGN veille à contribuer aux écosystèmes de la géomatique et de l'IA et ouvre ses actifs technologiques : 5 *datasets* et 9 modèles IA ont été partagés sur HuggingFace et 7 bibliothèques ouvertes liées aux systèmes IA ont été créées sur le GitHub IGNF (entraînements de modèles, construction de *datasets*, évaluation des systèmes...).

- [FLAIR \(French Land cover from Aerospace ImageRy\)](#)

Des jeux de données d'apprentissage massifs ont été ouverts aux communautés de recherche en IA dans le cadre des challenges FLAIR, compétitions scientifiques et techniques sur des problèmes de recherche pour l'analyse de l'occupation des sols par intelligence artificielle. Les méthodes produites et s'évaluant sur ces jeux de données contribuent à améliorer les résultats de classification des données d'occupation du sol, à augmenter la capacité de généralisation du modèle IA, mais aussi à combiner les données issues de prises de vues aériennes et satellites (Sentinel 2). Ces défis ont également permis de soutenir les travaux de recherche et d'innovation dans la cartographie des territoires par la publication de codes et données et d'engager une nouvelle collaboration technique à l'étranger.

Quelques chiffres :

- *FLAIR : 20 milliards de pixels, 19 classes de couverture des sols, 100K images soit 980km² d'images annotées dans 55 domaines différents. FLAIR#2 fusionne avec 50K acquisitions Sentinel-2.*
- *A sa publication, l'un des trois plus riches jeux de données ouverts pour l'occupation des sols au monde, de même ampleur que DeepGlobe (Facebook, MIT) et LoveDA (Université de Wuhan)*
- *Intérêt pour ces données : FLAIR#1 305 inscrits et 1668 soumissions ; Flair#2 152 inscrits et 1045 soumissions.*

- [PureForest](#)

PureForest est un *dataset* combinant des données LiDAR acquises dans le cadre du programme national LiDAR HD et des orthophotographies aériennes à grande échelle pour la classification des espèces d'arbres dans les forêts monospécifiques.

Quelques chiffres :

- *Avec 135 00 images LiDAR de 50 m par 50 m, couvrant 449 forêts de 40 départements français sur un total de 339 km², PureForest est le plus grand jeu de données LiDAR au monde sur les essences forestières.*