ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume XII-4/W2-2026
10th International Conference on Smart Data and Smart Cities (SDSC), 29 September–1 October 2026, Sofia, Bulgaria

# Towards a Reproducible Workflow for Urban Vegetation Stratification in Lyon Using Aerial Imagery and LiDAR

Arthur Villarroya-Palau[1], John Samuel[1,2], Ludovic Darmet[3], Gilles Gesquière[1]

[1] LIRIS, UMR 5205, Universite Claude Bernard Lyon 1, Université Lumière Lyon 2, CNRS, INSA Lyon, Lyon, France
[2] CPE Lyon, Villeurbanne, France
[3] TelesCoop, France
arthur.villarroya-palau@univ-lyon2.fr, john.samuel@cpe.fr,ludovic@telescoop.fr, gilles.gesquiere@univ-lyon2.fr

**Keywords:** Urban vegetation, Semantic segmentation, LiDAR, Aerial imagery, Reproducibility.

## Abstract

As cities confront rising heat stress, biodiversity loss, and stormwater pressures, urban vegetation has become essential urban infrastructure for climate adaptation and human well-being. Reliable and regularly updated vegetation inventories are therefore essential for urban planning and environmental monitoring. However, detailed vegetation mapping in cities remains difficult because urban scenes are heterogeneous, structurally complex, and highly dynamic. This paper reviews the principal data sources and segmentation approaches relevant to urban vegetation mapping, with a particular focus on the Lyon metropolitan area. We compare optical imagery, LiDAR point clouds, and existing open resources including COSIA, FLAIR-HUB, LiDAR HD, Myria3D, FRACTAL, and the Armature 2 vegetation dataset. The originality of the paper lies in a reproducible, open, city-scale fusion workflow that converts existing optical segmentation and LiDAR height products into a three-stratum vegetation map. Rather than proposing a new segmentation network, the workflow operationalises multimodal evidence: optical imagery provides dense and frequently updateable vegetation extent, while LiDAR supplies the vertical information needed to separate herbaceous, shrub, and tree layers.

## 1. Introduction

Urban trees, shrubs, and herbaceous surfaces deliver a broad spectrum of ecosystem services that are directly relevant to urban resilience and public health. They mitigate urban heat islands, sequester carbon (Nowak and Crane, 2002), filter air pollutants (Janhäll, 2015, Klingberg et al., 2017), attenuate noise, support biodiversity, and promote psychological restoration among city dwellers (Tyrväinen et al., 2014). These benefits depend not only on the total extent of green cover but also on the structural composition of the vegetation. A dense tree canopy provides shade and evapotranspiratory cooling; a shrub layer along streets reduces particulate deposition at pedestrian height; low herbaceous cover contributes to stormwater infiltration. Stratified vegetation inventories that distinguish these layers are therefore essential inputs for evidence-based urban planning.

This work is also framed by the IA.rbre project[1], a collaborative platform for producing, analysing, and visualising territorial data in support of urban climate-adaptation decisions.

Producing such inventories at city scale, however, remains a methodological challenge. Field surveys are time-consuming and cannot be repeated at the frequency demanded by urban dynamics. Remote sensing offers the necessary spatial coverage, but urban scenes are spectrally heterogeneous (Herold et al., 2004): building materials, shadows, impervious surfaces, bare soil, and vegetation of different heights coexist within small areas, producing ambiguous signatures. Deep learning has substantially improved segmentation performance in remote sensing (Audebert et al., 2018, Garioud et al., 2023), yet no single-modality approach achieves robust vertical stratification. Optical imagery captures surface texture and colour but provides little direct information on height; LiDAR measures vertical structure precisely but is acquired only intermittently and struggles with low, flat herbaceous surfaces (Lenoir et al., 2022).

[1] `https://iarbre.fr/`

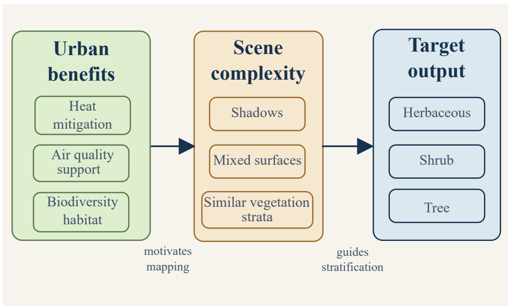



Figure 1. High-level framing of the problem addressed in this paper: urban vegetation inventories are motivated by ecosystem services, complicated by heterogeneous urban scenes, and expected to deliver reproducible stratified outputs.

This paper makes two principal contributions. First, it provides a structured comparison of French and Lyon-specific resources for urban vegetation mapping, with attention to their reproducibility, temporal currency, and suitability for vegetation stratification. Second, it implements and evaluates an open multimodal workflow that fuses FLAIR-HUB optical segmentation with LiDAR-derived height rasters to produce herbaceous, shrub, and tree layers at city scale. The contribution is therefore not a new neural architecture, but a reproducible operational synthesis of existing open data and tools for planning-oriented vegetation inventories. Figure 1 frames the broader problem these contributions address.

More specifically, the service rationale spans cooling, air-quality regulation, biodiversity support, and planning-oriented monitoring. The corresponding methodological difficulty is that dense urban scenes mix shadows, sealed surfaces, bare soil, and vegetation strata within short spatial distances, so the same planimetric footprint may correspond to very different vegetation forms. The workflow proposed in this paper is designed to respond to that tension with a compact but reproducible three-stratum output: herbaceous cover, shrubs, and trees.

## 2. Urban Vegetation Mapping: Challenges and Context

Urban vegetation delivers multiple ecosystem services, and many of them depend on vegetation structure rather than on green cover alone. This is why stratified inventories that distinguish herbaceous cover, shrubs, and trees are more informative than a simple vegetation/non-vegetation map.

The operational challenge is that intermediate layers are particularly difficult to identify remotely. Tree crowns produce distinctive spectral and textural patterns in very high resolution imagery and are detectable with specialised tools (Weinstein et al., 2020). Herbaceous surfaces produce strong spectral signals relative to impervious ground. Shrubs and hedges (Degerickx et al., 2020, Pu et al., 2024), by contrast, occupy the spectral middle ground between the two, and their geometry is highly variable: a low, rounded shrub and a dense young-tree stand may produce nearly identical pixel distributions in an orthophoto.

These difficulties are compounded by the requirement for methodological reproducibility. Most of existing urban vegetation inventories were produced with proprietary software, undocumented parameter choices, or visual-interpretation workflows that are difficult to transfer to new areas or years. The Lyon metropolitan area illustrates this situation clearly: it possesses high-resolution optical and LiDAR acquisitions, as well as a detailed reference dataset from prior research (Bellec, 2018), but these resources have not yet been integrated into a documented, open, and extensible workflow.

## 3. Data Sources for Urban Vegetation Inventories

This section reviews the principal data families used for urban vegetation mapping and clarifies why no single source is sufficient for robust stratification. Table 1 summarises their typical spatial resolution, principal strengths, and operational limitations before the following subsections discuss each source in more detail.

### 3.1 Optical imagery

Public satellite missions such as Landsat-8/9 and Sentinel-2 (Li and Chen, 2020) provide long temporal series and broad spatial coverage, which are valuable for regional or multi-temporal monitoring. In practice, Landsat multispectral imagery is typically used at 30 m resolution, while Sentinel-2 provides 10 m and 20 m optical bands depending on the channel. These resolutions are generally too coarse for detailed urban mapping of individual trees or shrub patches. Commercial systems such as Pléiades, SPOT, and WorldView offer sub-metric detail but are costly and operationally constrained.

Aerial orthophotos occupy a strategic position for urban applications. In France, IGN's *BD ORTHO* (IGN, 2026a) delivers RGB and colour-infrared imagery at 5–20 cm resolution, making it the primary input for fine-scale urban vegetation mapping. The FLAIR-HUB framework (Garioud et al., 2023, IGNF, 2026a) builds on these very high resolution aerial acquisitions but is not limited to *BD ORTHO* alone: more broadly, FLAIR is presented

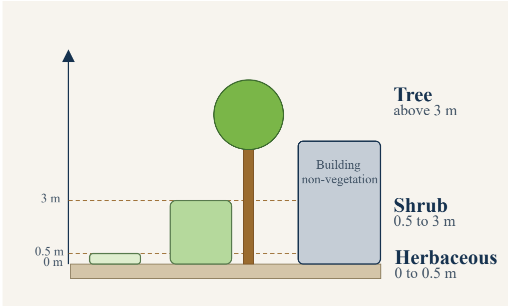



Figure 2. Cross-sectional overview of the three vegetation strata used in the workflow. Height thresholds at 0.5 m and 3 m structure the automated assignment, while optical information is still needed to distinguish tall vegetation from buildings.

as a country-scale land-cover dataset derived from multi-source optical imagery. It therefore provides not only pre-trained segmentation models, but also open datasets and reproducible experimental configurations for land-cover classification across France.

### 3.2 LiDAR point clouds

Airborne LiDAR provides a three-dimensional description of the urban surface through laser range measurements. From point clouds, a digital terrain model (DTM), a digital surface model (DSM), and a normalised height model (nDSM) can be derived. Height is central for distinguishing vegetation strata, but it is not the only cue available in LiDAR processing: point classification also relies on attributes such as return intensity and echo structure. In the LiDAR HD documentation, for example, asphalt is associated with low intensity values, trees with intermediate values, and grass with high values, although these values remain acquisition-dependent. In practice, heights below 0.5 m typically correspond to herbaceous cover, 0.5–3 m to shrubs, and heights exceeding 3 m to trees and structures (Figure 2).

In France, the LiDAR HD programme (IGN, 2026c) delivers a national high-density dataset that is gradually being completed across all territories. Myria3D (IGNF, 2026c) provides an open deep-learning framework for semantic segmentation of LiDAR HD point clouds, extending the label space beyond simple geometric classes. FRACTAL offers a large annotated benchmark (Gaydon et al., 2024) derived from LiDAR HD acquisitions, which supports evaluation of spatial generalisation.

The primary limitation of LiDAR in this context is thematic ambiguity at ground level. LiDAR captures vertical structure effectively, yet it does not reliably distinguish asphalt or other sealed surfaces from areas covered by low vegetation or grass when elevation differences are minimal. This limitation reinforces the need for multimodal fusion: LiDAR is essential for separating shrubs and trees through height information, while optical imagery remains necessary to identify low vegetation correctly and avoid confusion with non-vegetated ground surfaces.

### 3.3 Vector reference data

Vector datasets describe urban vegetation through explicit objects: tree-inventory points, shrub polygons, and green-space boundaries. In practice, these layers are often produced directly by cities or metropolitan services for operational purposes such as tree management, green-space maintenance, or urban planning. They are useful for GIS integration and for linking

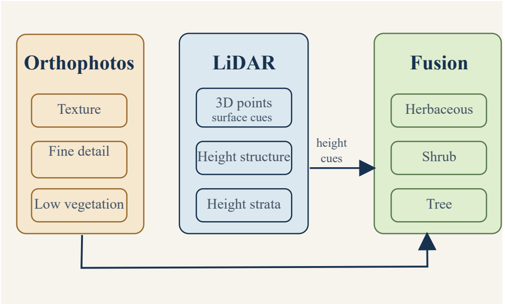



Figure 3. Overview of source complementarity. Orthophotos provide dense planimetric evidence for vegetation extent, whereas LiDAR provides the height information required for vertical stratification.

remotely-sensed features to semantic attributes. Their quality, however, depends entirely on the upstream mapping workflow, which is often not openly documented to non-expert users.

### 3.4 Comparative overview

Figure 3 and Table 1 together illustrate the fundamental complementarity between the two principal sources. Orthophotos are strong for horizontal delineation and herbaceous detection; LiDAR is strong for vertical discrimination of shrubs and trees. Neither source alone achieves robust stratified mapping, which motivates the fusion approach described in Section 7.

In practical terms, the optical branch contributes fine spatial detail, texture, and colour cues that are especially effective for low vegetation and small fragmented patches. The LiDAR branch contributes the terrain-normalised height signal needed to separate intermediate and tall vegetation layers. The article therefore treats the two sources as deliberately asymmetric inputs: orthophotos establish vegetation extent, while LiDAR resolves the vertical organisation that orthophotos alone cannot infer reliably.

## 4. Segmentation Approaches

This section focuses primarily on segmentation methods applied to aerial orthophotos, with complementary remarks on semantic segmentation in LiDAR point clouds where it directly supports vegetation stratification.

### 4.1 Classical approaches

The normalised difference vegetation index (NDVI) is one of the most widely used indicators for vegetation detection (Pettorelli et al., 2005). It is computed from reflectance measured in the near-infrared (NIR) and red (RED) spectral bands:

$$\mathrm{NDVI} = \frac{NIR - RED}{NIR + RED}. \tag{1}$$

In this expression, $NIR$ denotes the reflectance in the near-infrared band and $RED$ denotes the reflectance in the red visible band. Vegetation typically reflects strongly in the near-infrared and absorbs red light for photosynthesis, so higher NDVI values generally indicate greener and denser vegetation. For aerial orthophotos, this requires colour-infrared acquisitions rather than RGB imagery alone. NDVI is robust for regional monitoring but has well-documented limitations in urban areas: shadows, roofing materials, asphalt, and bare soil can all depress or inflate the signal. More critically, NDVI provides no information on vegetation height, making it unsuitable for separating the three strata of interest.

Classical classification methods such as maximum likelihood, minimum distance, and unsupervised clustering such as K-means and ISODATA (Phiri and Morgenroth, 2017) offer exploratory utility but struggle with the class complexity of urban scenes. Object-based image analysis (OBIA) improves spatial coherence by grouping pixels into meaningful objects and combining spectral, textural, and geometric descriptors (Baatz and Schäpe, 2000). OBIA is effective with very high-resolution imagery but sensitive to scale parameters and often implemented in proprietary environments, which limits reproducibility.

### 4.2 Deep learning approaches

Deep learning has transformed semantic segmentation for Earth observation. Fully convolutional networks established end-to-end dense prediction (Long et al., 2015), and U-Net improved boundary reconstruction through encoder-decoder skip connections (Ronneberger et al., 2015). These encoder-decoder architectures remain competitive baselines, particularly when training data are limited. Multimodal extensions that jointly process optical and elevation channels have further demonstrated the benefit of height information for urban scene understanding (Audebert et al., 2018).

Vision transformers have since become the dominant segmentation backbone. The Vision Transformer (ViT) (Dosovitskiy et al., 2021) applies self-attention globally over image patches. The Swin Transformer (Liu et al., 2021) introduces a hierarchical architecture with shifted local windows, improving computational efficiency and multi-scale context. SegFormer (Xie et al., 2021) combines a lightweight hierarchical transformer encoder with a simple MLP decoder, achieving strong performance with fewer parameters and no need for positional encoding at inference. These architectures are particularly suitable for complex urban scenes because their self-attention mechanisms capture long-range contextual dependencies between heterogeneous surface types. The FLAIR-HUB framework (Garioud et al., 2023) builds on these advances and has demonstrated state-of-the-art segmentation of French aerial orthophotos across 19 land-cover classes.

Promptable foundation-model approaches also deserve mention, especially for satellite imagery. CLIPSeg (Lüddecke and Ecker, 2022), for example, uses a CLIP-based text encoder with a transformer decoder to produce binary segmentation masks from free-text prompts rather than from a fixed closed set of classes. This flexibility is attractive for exploratory remote-sensing tasks and weakly supervised workflows, even if such models are currently less common than dedicated supervised architectures for high-resolution orthophotos.

For 3D data, deep learning methods such as PointNet++ (Qi et al., 2017) and RandLA-Net (Hu et al., 2020) enable semantic segmentation of large LiDAR point clouds. RandLA-Net underlies the Myria3D framework and has been validated on the LiDAR HD dataset. More recently, SATree (Du et al., 2026) addressed individual-tree instance segmentation directly in 3D LiDAR point clouds by jointly predicting crown/stem/background semantics, stem-centred heat values, and centroid offsets before graph-based grouping of points into tree instances. Tree crown detection in optical imagery has also advanced significantly through purpose-built deep learning tools (Weinstein et al., 2020), though transferring these to dense urban scenes requires careful domain adaptation.


https://doi.org/10.5194/isprs-annals-XII-4-W2-2026-203-2026  

| Data type | Typical resolution | Main strength | Main limitation | Examples |
|---|---|---|---|---|
| Public satellite imagery | 10–30 m | Long time series and broad coverage | Too coarse for detailed urban analysis | Sentinel-2, Landsat |
| Commercial satellite imagery | 0.3–1.5 m | Fine spatial detail | Cost and acquisition constraints | Pléiades, WorldView, SPOT |
| Aerial orthophotos | 5–20 cm | Very fine planimetric detail; supports open benchmarks and models | No direct 3D information | BD ORTHO; FLAIR-HUB as derived benchmark/models |
| LiDAR point clouds | 8–20 pts/m$^2$ | Accurate vertical structure and height strata | Intermittent acquisition; costly processing | LiDAR HD, local LiDAR |
| Vector inventories | Object-based | Direct GIS integration | Depend on prior interpretation or extraction | SHP, GPKG, GeoJSON |

Table 1. Main data sources used for urban vegetation mapping.

### 4.3 Evaluation metrics and the stratification problem

Segmentation quality in urban mapping must be assessed with metrics that explicitly account for class imbalance. Overall accuracy is misleading when large background classes dominate. Precision, recall, Dice/F1-score, and per-class Intersection over Union (IoU) are more informative. For vegetation stratification specifically, recall and IoU for the shrub category are the most diagnostic metrics, because shrubs are the stratum most frequently absorbed into adjacent classes by models not designed for vertical discrimination.

Figure 2 illustrates the three strata and the critical role of height thresholds. The 0.5 m boundary separates herbaceous ground cover from the shrub layer, while the 3 m boundary separates shrubs from trees and from tall structures. Buildings produce height returns identical to those of trees, which is why LiDAR height alone is insufficient and must be combined with spectral information.

These threshold values are not merely cartographic conventions. In our case they reflect both the working interpretation used in the Lyon Armature 2 vegetation mapping context and broader operational conventions inspired by the ASPRS low/medium/high vegetation classes. They therefore provide a transparent rule set that can be audited, adjusted, and transferred to other acquisitions. The ambiguous case lies above 3 m, where both tree crowns and building roofs generate strong returns; here the orthophoto branch supplies the spectral evidence needed to preserve only vegetated surfaces in the final map.

## 5. Existing Resources in the French and Lyon Contexts

### 5.1 COSIA and FLAIR-HUB

COSIA (*Cartographie de l'Occupation des Sols par Intelligence Artificielle*) (IGN, 2026b) is a national land-cover product produced by IGN from very high-resolution aerial imagery and elevation data. Despite its name, it does not rely on automated inference alone: the production chain also includes manual review and corrections, as well as cross-checking with other reference sources such as *BD TOPO*. It therefore provides operationally consistent coverage across France, but COSIA is not designed for fine urban vegetation stratification: vegetation classes are broadly defined, and intermediate strata such as shrubs and hedges are not individually distinguished.

FLAIR-HUB (Garioud et al., 2023) is the methodological complement to COSIA. It provides fully open datasets, pre-trained segmentation models, and documented experimental configurations for aerial-image segmentation. FLAIR-HUB models trained on millions of labelled patches across France achieve strong generalisation to new urban areas. Their principal limitation for our purpose is that the training taxonomy merges shrubs with low vegetation or omits them entirely, which means direct inference does not produce the three-stratum output required for detailed urban vegetation inventories.

### 5.2 LiDAR HD, Myria3D, and FRACTAL

LiDAR HD (IGN, 2026c) provides the national high-density point cloud reference for France and is the key resource for height-based vegetation analysis. Myria3D (IGNF, 2026c) complements this ecosystem with an open framework for semantic segmentation of LiDAR HD clouds, primarily based on RandLA-Net (Hu et al., 2020), and provides a reproducible baseline for 3D classification. FRACTAL offers a large annotated benchmark derived from LiDAR HD data that enables evaluation of spatial generalisation across diverse urban and peri-urban landscapes.

These three resources are valuable for characterising urban structure and vegetation height at the national scale. Their shared limitation is temporal sparsity: LiDAR acquisitions are not refreshed annually, which restricts height-based stratification to the years in which data are available.

### 5.3 Armature 2's Lyon vegetation dataset

The multi-temporal vegetation dataset produced for the project Armature 2 (LabEx IMU) (LabEx IMU, 2026) from the doctoral research of Bellec (Bellec, 2018) for the Lyon metropolitan area constitutes a uniquely valuable local reference. It covers the period 1984–2015, distinguishes herbaceous vegetation, shrubs, small trees, and large trees at very high spatial resolution, and provides a validated semantic taxonomy for the Lyon context. This stratum-level classification makes it indispensable for fine-tuning and validation.

The principal limitations are temporal (the series ends in 2015) and methodological: the classification workflow relies on OBIA routines implemented in commercial software, which reduces exact reproducibility and makes it difficult to extend the series to recent years. The proposed workflow addresses both points by providing an open, documented alternative.

## 6. Critical Analysis

The preceding review reveals a strong structural complementarity between optical segmentation and LiDAR analysis. Optical imagery from BD ORTHO and FLAIR-HUB provides dense, regularly updated surface mapping and is particularly effective for delineating herbaceous surfaces, small green patches, and tree crowns in plan view. It cannot, however, distinguish a dense shrub from a tree crown based on spectral information alone. LiDAR provides precisely the structural information needed to resolve this ambiguity, but its temporal sparsity and limited performance for flat, low surfaces constrain its autonomous use.

The national open resources also exhibit complementary limitations. COSIA provides operational coverage and benefits from hybrid production procedures that combine automated mapping with external corrections, but it offers limited stratum-level vegetation detail for our purpose. FLAIR-HUB provides openness and reproducibility but was not optimised for the three-stratum taxonomy required by urban vegetation managers. The Armature 2 dataset provides detailed local reference data but terminates in 2015 and depends on a proprietary workflow.

These observations converge on a hybrid strategy: use FLAIR-HUB as the optical backbone, and exploit LiDAR for vertical discrimination. This is the strategy formalised in Section 7.

## 7. Proposed Workflow

### 7.1 Overview

The implemented workflow is a reproducible Docker-based chain released as a software repository (Villarroya-Palau et al., 2026). It downloads the orthophoto and LiDAR tiles for a Lambert-93 bounding box, mosaics the orthophotos, rasterises the LiDAR point cloud, runs FLAIR-HUB inference, reweights the vegetation classes, and finally fuses the optical and LiDAR outputs. The objective is to preserve the spatial continuity and update frequency of orthophoto segmentation while improving vegetation stratification through explicit structural information. Figure 4 summarises the overall processing chain.

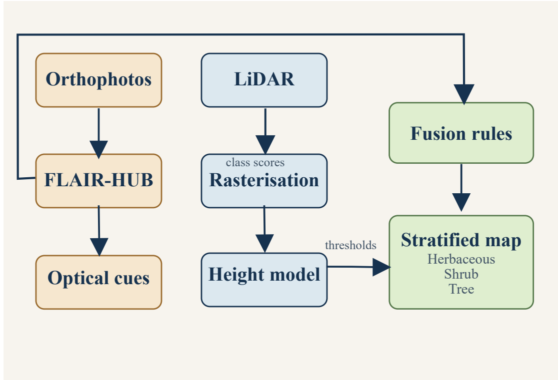



Figure 4. High-level view of the proposed multimodal fusion workflow. The optical branch provides class probabilities, the LiDAR branch provides height information, and explicit fusion rules combine both sources into a stratified vegetation map.

### 7.2 Optical segmentation

The current baseline uses the RGB-only FLAIR-HUB model `IGNF/FLAIR-HUB_LC-A_RGB_swinlarge-upernet` at pinned revision `336bd84` (IGNF, 2026b). More precisely, only the aerial modality is activated, with channels `[1,2,3]`, so the workflow uses RGB orthophotos rather than colour-infrared imagery at this stage.

Input orthophoto tiles are first resampled from their native 5 cm resolution to 50 cm, mosaicked over the study area, and then processed by the official `flairhub_zonal` inference entry point. In the current baseline configuration, inference is run with 512×512 pixel windows, a margin of 128 pixels, batch size 1, and an output probability raster written at 1 m resolution. During inference, class-probability maps are retained rather than only the final argmax labels.

The vegetation mask is then derived from the 19-class FLAIR-HUB output by keeping only four vegetation-related COSIA classes: herbaceous vegetation (8), brushwood (14), deciduous (12), and coniferous (13). These are remapped to a compact vegetation label set used downstream. In the baseline experiment, the class weights are all equal to 1.0, so there is no additional confidence threshold in the optical branch: the decision is the argmax over the reweighted probability cube, and all non-target classes are set to the ignore value 255. The workflow nevertheless keeps the reweighting step explicit so that later experiments can change class weights without rerunning the full inference chain.

### 7.3 LiDAR processing

The LiDAR branch starts from classified LAZ tiles and rasterises the irregular point cloud onto a regular grid with the same target cell size as the FLAIR output, namely 1 m in the current baseline. For each grid cell, the workflow records the maximum elevation to build the surface model, the class of the highest point for the semantic class raster, and a terrain estimate derived while excluding vegetation classes but retaining buildings. This choice is deliberate: vegetation is removed because dense canopies can prevent reliable returns below them, whereas buildings are kept so that vegetation located on roofs can still be assigned a meaningful height above its supporting surface. The resulting gaps in the terrain raster are then iteratively filled by neighbourhood averaging until no empty pixels remain. Tile rasters are finally mosaicked over the full study area.

The height raster used for fusion is the difference between the digital surface model and this terrain model. Excluding vegetation from the terrain estimate avoids situations where the surface and terrain values become artificially similar inside dense tree cover, which would otherwise underestimate vegetation height. In practice, this also means that LiDAR is not meshed first and only later sampled on the orthophoto grid; instead, the points are directly aggregated into raster cells during preprocessing. Because the orthophotos are used as model input at 50 cm but FLAIR probabilities are exported at 1 m, the effective alignment for fusion is between the LiDAR rasters and the 1 m FLAIR output raster. Small residual edge mismatches are absorbed by padding arrays to a common shape before fusion.

For the LiDAR-driven vegetation map itself, the current baseline uses two explicit thresholds from the workflow configuration: 0.30 m to separate low from medium vegetation, and 5.0 m to separate medium from high vegetation. LiDAR vegetation is considered only where the class raster belongs to the vegetation-related classes $\{3, 4, 5, 8\}$; all other cells remain invalid at this stage unless additional options are activated.

### 7.4 Fusion strategy

The implemented fusion is rule-based. First, the LiDAR branch produces a vegetation map using the LiDAR class raster and the 0.30 m / 5.0 m thresholds. Second, the optical vegetation mask is projected onto the same height raster, but with a more conservative low threshold: 0.3 m for herbaceous versus shrub separation, and 5.0 m for shrub versus tree separation. This means that optical-only fallback pixels are classified as herbaceous when height $< 0.3$ m, shrub when $0.3 \leq h < 5.0$ m, and tree when $h \geq 5.0$ m.

The final fused raster keeps the LiDAR result wherever LiDAR provides a valid vegetation class. Wherever LiDAR is invalid, the workflow falls back to the FLAIR-derived vegetation map. In the baseline configuration, this fallback is not limited to herbaceous surfaces only: the option `flair_only_herbaceous` is disabled, so the optical branch can fill all LiDAR gaps. This is operationally important for recovering vegetation where LiDAR coverage is sparse or where the semantic LiDAR class is missing.

The result is a stratified raster with three vegetation classes used for evaluation and mapping. In the baseline workflow, non-vegetation and unresolved pixels are kept outside the vegetation map rather than being forced into a vegetation class. This makes the implementation more conservative than a pure optical segmentation and explains why the operational thresholds in the code are slightly different from the conceptual 0.5 m / 3 m thresholds used earlier to explain the stratification principle.

From a computational point of view, the workflow is asymmetric. LiDAR download, rasterisation, mosaicking, and fusion can run on CPU, whereas FLAIR-HUB inference with the Swin-Large backbone is the most demanding stage and benefits strongly from an NVIDIA GPU with CUDA support. The shipped configuration therefore provides both CPU and GPU Docker services and keeps the inference batch size at 1 by default. We do not report benchmark runtimes here because wall-clock time depends strongly on the selected bounding box, the number of input tiles, and the available hardware.

## 8. Results

The evaluation uses the 2018 Grand Lyon *vegetation stratifiée* raster as reference, derived from the Armature 2 / Bellec workflow discussed in Section 5.3. This reference is highly valuable because it provides local stratum-level labels, but it is not a field-measured ground truth. The comparison is therefore affected by both reference uncertainty and temporal mismatch: vegetation may have changed between the acquisition dates of the reference, orthophotos, and LiDAR data. For this reason, disagreement between the fused map and the reference may reflect model error, reference artefacts, or genuine land-cover change.

Figure 5 should therefore be read as a comparison against this Armature-derived reference rather than against an absolute ground truth measured in the field. More specifically, Figure 5(a) shows the baseline without reweighting, Figure 5(b) the low-reweight configuration, and Figure 5(c) the harder reweighting variant. It is also important to stress that these confusion matrices are row-normalised accuracies: they show how each reference class is redistributed among predictions, so diagonal values correspond to per-class recall. They are not IoU matrices.

Figure 6 makes these limitations more concrete. Figure 6(a) illustrates a locally implausible vegetation pattern in the Armature reference, while Figure 6(b) shows a shadowed lawn area that appears to be misclassified. These cases do not invalidate the usefulness of the dataset, but they do mean that disagreement maps and confusion scores should be interpreted as differences with respect to an operational reference, not as direct evidence of absolute model error everywhere.

Across the three tested configurations, the no-reweight baseline already separates trees reasonably well, with a tree recall of 94.3%, but the shrub class remains much less stable: only 44.8% of shrub reference pixels are correctly recovered, while 24.2% are absorbed into the *other* class and 17.4% into trees. Introducing low reweighting improves the balance between herbaceous and shrub classes. Herbaceous recall increases from 58.3% to 73.2%, and shrub recall increases from 44.8% to 53.5%, while tree recall remains very high at 93.6%. The hard reweighting variant pushes this trend further: herbaceous recall rises to 77.8% and shrub recall to 63.7%, but with a stronger loss of precision through additional confusion with the *other* class and neighbouring vegetation strata.

| Configuration | Mean IoU | Mean Precision | Mean Recall | Mean Dice |
|---|---|---|---|---|
| No reweighting | 0.567 | 0.706 | 0.708 | 0.703 |
| Low reweighting | 0.571 | 0.693 | 0.727 | 0.707 |
| Hard reweighting | 0.571 | 0.682 | 0.747 | 0.708 |

Table 2. Comparison of the three fused configurations from the archived metrics reports. Means are computed over the vegetation strata only, excluding *other*.

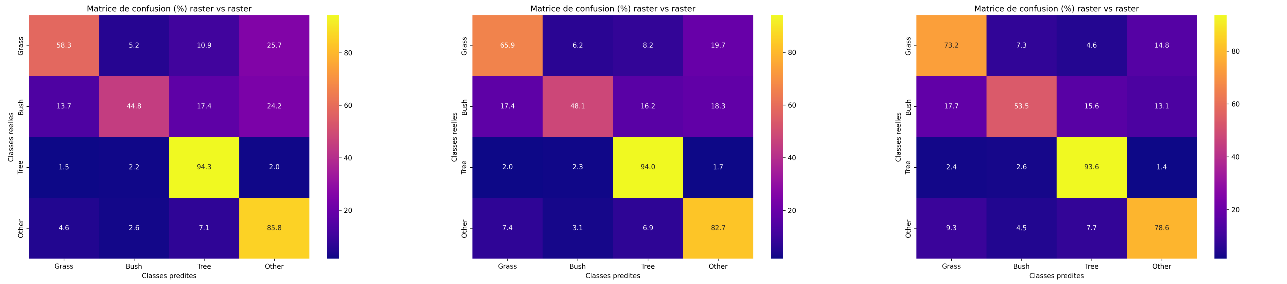
The spatial comparison further confirms that all variants reproduce the dominant tree masses and broad herbaceous surfaces visible in the reference. The main remaining difficulty concerns intermediate vegetation, especially around tree edges, narrow strips, and mixed urban surfaces where shrub pixels can still be fragmented or absorbed into neighbouring classes. Visually, the hard reweighting variant recovers more low vegetation but also spreads vegetation labels more aggressively, whereas the low reweighting configuration remains the most balanced compromise.

Table 2 confirms the same trade-off at the aggregate level. Mean IoU rises from 0.567 without reweighting to 0.571 with both low and hard reweighting. Mean recall increases monotonically from 0.708 to 0.727 and then 0.747, while mean precision decreases from 0.706 to 0.693 and 0.682. In other words, hard reweighting is the most recall-oriented variant, but low reweighting offers the most balanced improvement over the baseline. At class level, this is driven mainly by better recovery of herbaceous vegetation, whose IoU increases from 0.470 to 0.483 and then 0.507, while shrub IoU remains almost unchanged around 0.30 despite the recall gains.

## 9. Discussion and Conclusion

Urban vegetation inventories require both thematic detail and methodological reproducibility. The analysis presented in this paper shows that no single source currently satisfies both requirements in complex urban environments such as Lyon. Orthophotos provide dense and regularly updated surface information, but stratification remains difficult from optical cues alone. LiDAR adds the vertical structure needed to separate vegetation layers, yet it is more sporadic in time, heavier to process, and less informative for very low vegetation when used without spectral context. Existing operational products partly fill these gaps, but they often rely on proprietary processing chains, hybrid manual corrections, or class systems that do not directly match the three-stratum need addressed here.

Within that broader landscape, the proposed workflow should be read as a practical synthesis of the preceding analysis rather than as an isolated technical contribution. It combines the reproducibility and coverage of open orthophoto-based segmentation with explicit LiDAR-based height rules to recover herbaceous, shrub, and tree strata in a transparent way. The results against the Armature-derived reference confirm both the value and the limit of this strategy: tree detection is already strong, herbaceous vegetation improves with optical reweighting, but shrub delineation remains the main bottleneck because it sits at the interface between surface appearance and vertical structure.

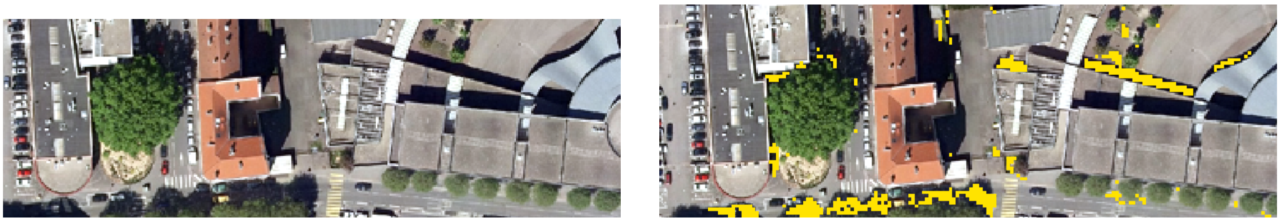



(a) LiDAR + orthophoto fusion without probability reweighting.

(b) LiDAR + orthophoto fusion with low vegetation reweighting.

(c) LiDAR + orthophoto fusion with hard vegetation reweighting.

Figure 5. Row-normalised confusion matrices for three fused configurations. Each row sums to 100%, so the values express class-conditional accuracy/recall rather than IoU.



(a) Example of a locally implausible Armature pattern, illustrating that the reference can itself contain hallucinated vegetation structures.

(b) Example of a shadowed lawn area misclassified in the Armature reference, showing that dark radiometric conditions can also bias the label map.

Figure 6. Illustration of two limitations of the Armature-derived reference used for evaluation. These examples show that some apparent errors in the comparison may come from the reference itself rather than from the fused prediction alone.

(a) Example orthophoto.

(b) Armature-derived reference overlay on the orthophoto.

(c) Fused prediction without reweighting.

(d) Fused prediction with low reweighting.

(e) Fused prediction with hard reweighting.

Figure 7. Visual comparison between the Armature-derived reference and the fused predictions. Yellow corresponds to low vegetation, light green to intermediate vegetation, and dark green to tree cover.

Future work should first broaden validation beyond the present Lyon reference by testing additional sectors, acquisition years, and, where possible, other cities. Because the current evaluation relies on a 2018 operational raster rather than field observations, targeted on-the-ground checks remain necessary for ambiguous shrub, hedge, and shadowed lawn areas. Participatory or crowd-sourced observations could complement municipal inventories by flagging recent vegetation change, although such data would require quality control before use as validation. A second priority is to quantify computational scalability, especially GPU inference time, memory use, and tile throughput as the bounding box expands. Finally, more explicit ablation against alternative multimodal strategies, including optical-only baselines, LiDAR-only baselines, and learned optical-elevation fusion, would better position the rule-based workflow relative to state-of-the-art approaches.

## Acknowledgements

The IA.rbre project received funding from Banque des Territoires under the call for projects “Démonstrateurs d’IA frugale au service de la transition écologique dans les territoires” (DIAT – France 2030).

## References

Audebert, N., Le Saux, B., Lefèvre, S., 2018. Beyond RGB: Very high resolution urban remote sensing with multimodal deep networks. *ISPRS Journal of Photogrammetry and Remote Sensing*, 140, 20–32.

Baatz, M., Schäpe, A., 2000. Multiresolution segmentation: An optimization approach for high quality multi-scale image segmentation. *Angewandte Geographische Informationsverarbeitung XII*, Wichmann, 12–23.

Bellec, A., 2018. Dynamiques spatiales, temporelles et ecologiques de la vegetation de la Metropole de Lyon, 1984–2015. PhD thesis, Universite de Lyon.

Degerickx, J., Hermy, M., Somers, B., 2020. Mapping Functional Urban Green Types Using High Resolution Remote Sensing Data. *Sustainability*, 12(5), 2144. https://www.mdpi.com/2071-1050/12/5/2144.

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., Houlsby, N., 2021. An image is worth 16x16 words: Transformers for image recognition at scale. *International Conference on Learning Representations*.

Du, S., Stoter, J., Kooij, J. F., Nan, L., 2026. SATree: Structure-aware tree instance segmentation from 3D LiDAR point clouds. *Urban Forestry & Urban Greening*, 120, 129414. https://linkinghub.elsevier.com/retrieve/pii/S1618866726001548.

Garioud, A., Gonthier, N., Astruc, L., Corbin, A., Dufourg, H., Ristorcelli, T., Mallet, C., 2023. FLAIR: A country-scale land cover semantic segmentation dataset from multi-source optical imagery. *Advances in Neural Information Processing Systems*, 36.

Gaydon, C., Daab, M., Roche, F., 2024. Fractal: An ultra-large-scale aerial lidar dataset for 3d semantic segmentation of diverse landscapes.

Herold, M., Roberts, D. A., Gardner, M. E., Dennison, P. E., 2004. Spectrometry for urban area remote sensing—Development and analysis of a spectral library from 350 to 2400 nm. *Remote Sensing of Environment*, 91(3-4), 304–319. https://linkinghub.elsevier.com/retrieve/pii/S0034425704000768.

Hu, Q., Yang, B., Xie, L., Rosa, S., Guo, Y., Wang, Z., Trigoni, N., Markham, A., 2020. Randla-net: Efficient semantic segmentation of large-scale point clouds. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 11108–11117.

IGN, 2026a. BD ORTHO. Research data, https://www.data.gouv.fr/datasets/bd-ortho-r. Accessed 25 August 2026.

IGN, 2026b. COSIA: Cartographie de l'occupation des sols par intelligence artificielle. Research data, https://geoservices.ign.fr/documentation/donnees/vecteur/cosia. Accessed 25 August 2026.

IGN, 2026c. LiDAR HD. Research data, https://geoservices.ign.fr/lidarhd. Accessed 25 August 2026.

IGNF, 2026a. FLAIR: AI challenges on remote sensing. Software project repository, https://github.com/IGNF/FLAIR. Accessed 25 August 2026.

IGNF, 2026b. FLAIR-HUB_LC-A_RGB_swinlarge-upernet. Software model version, revision 336bd84, https://huggingface.co/IGNF/FLAIR-HUB_LC-A_RGB_swinlarge-upernet. Accessed 25 August 2026.

IGNF, 2026c. Myria3D: Aerial LiDAR HD semantic segmentation with deep learning. Software project repository, https://github.com/IGNF/myria3d. Accessed 25 August 2026.

Janhäll, S., 2015. Review on urban vegetation and particle air pollution: Deposition and dispersion. *Atmospheric Environment*, 105, 130–137.

Klingberg, J., Broberg, M., Strandberg, B., Thorsson, S., Pleijel, H., 2017. Influence of urban vegetation on air pollution and noise exposure: A case study in Gothenburg, Sweden. *Science of the Total Environment*, 599–600, 1728–1739.

LabEx IMU, 2026. Armature 2. Research project data page, https://imu.universite-lyon.fr/armature-2-282507.kjsp. Accessed 25 August 2026.

Lenoir, J., Gril, E., Durrieu, S., Horen, H., Laslier, M., Lembrechts, J. J., Zellweger, F., Alleaume, S., Brasseur, B., Buridant, J., Dayal, K., De Frenne, P., Gallet-Moron, E., Marrec, R., Meeussen, C., Rocchini, D., Van Meerbeek, K., Decocq, G., 2022. Unveil the unseen: Using LiDAR to capture time-lag dynamics in the herbaceous layer of European temperate forests. *Journal of Ecology*, 110(2), 282–300. https://besjournals.onlinelibrary.wiley.com/doi/10.1111/1365-2745.13837.

Li, J., Chen, B., 2020. Global Revisit Interval Analysis of Landsat-8 -9 and Sentinel-2A -2B Data for Terrestrial Monitoring. *Sensors*, 20(22). https://www.mdpi.com/1424-8220/20/22/6631.

Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., Guo, B., 2021. Swin transformer: Hierarchical vision transformer using shifted windows. *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 10012–10022.

Long, J., Shelhamer, E., Darrell, T., 2015. Fully convolutional networks for semantic segmentation. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 3431–3440.

Lüddecke, T., Ecker, A. S., 2022. Image segmentation using text and image prompts. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 7086–7096.

Nowak, D. J., Crane, D. E., 2002. Carbon storage and sequestration by urban trees in the USA. *Environmental Pollution*, 116(3), 381–389.

Pettorelli, N., Vik, J. O., Mysterud, A., Gaillard, J.-M., Tucker, C. J., Stenseth, N. C., 2005. Using the satellite-derived NDVI to assess ecological responses to environmental change. *Trends in Ecology & Evolution*, 20(9), 503–510.

Phiri, D., Morgenroth, J., 2017. Developments in Landsat land cover classification methods: A review. *Remote Sensing*, 9(9), 967.

Pu, Y., Wilmshurst, J. F., Guo, X., 2024. Separating Shrub Cover From Green Vegetation in Grasslands Using Hyperspectral Vegetation Indices. *Canadian Journal of Remote Sensing*, 50(1), 2347630. https://www.tandfonline.com/doi/full/10.1080/07038992.2024.2347630.

Qi, C. R., Yi, L., Su, H., Guibas, L. J., 2017. Pointnet++: Deep hierarchical feature learning on point sets in a metric space. *Advances in Neural Information Processing Systems 30*, 5099–5108.

Ronneberger, O., Fischer, P., Brox, T., 2015. U-net: Convolutional networks for biomedical image segmentation. *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, 234–241.

Tyrväinen, L., Ojala, A., Korpela, K., Lanki, T., Tsunetsugu, Y., Kagawa, T., 2014. The influence of urban green environments on stress relief measures: A field experiment. *Journal of Environmental Psychology*, 38, 1–9.

Villarroya-Palau, A., Samuel, J., Darmet, L., Gesquière, G., 2026. UD-IArbre-Research: Reproducible workflow for urban vegetation stratification in Lyon. Software repository, https://github.com/VCityTeam/UD-IArbre-Research. Accessed 25 August 2026.

Weinstein, B. G., Marconi, S., Bohlman, S., Zare, A., White, E., 2020. DeepForest: A Python Package for RGB Deep Learning Tree Crown Delineation. *Methods in Ecology and Evolution*, 11(12), 1743–1751.

Xie, E., Wang, W., Yu, Z., Anandkumar, A., Alvarez, J. M., Luo, P., 2021. SegFormer: Simple and efficient design for semantic segmentation with transformers. *Advances in Neural Information Processing Systems*, 34, 12077–12090.