



Article

Pasture Biomass Monitoring in Queensland Rangelands with UAV and Satellite Cascades

Jason Barnetson ^{1,*} , Hemant Raj Pandeya ² , Grant Fraser ¹

¹ Queensland Department of Environment, Tourism, Science and Innovation, Eco-Sciences Precinct, 41 Boggo Rd., Dutton Park, Brisbane, QLD 4102, Australia; grant.fraser@detsi.qld.gov.au

² Queensland Department of Primary Industries, Eco-Sciences Precinct, 41 Boggo Rd., Dutton Park, Brisbane, QLD 4102, Australia; rajhemant.pandeya@dpi.qld.gov.au

* Correspondence: jason.barnetson@detsi.qld.gov.au

Abstract

The operational satellite monitoring of pasture biomass requires models that transfer beyond the properties on which they were calibrated. We present a hierarchical, open-source cascade that scales in situ clip-and-weigh biomass ($n = 1120$ samples across eleven sites on five Queensland properties) through UAV digital-surface-model imagery to Sentinel-2 predictions, using TabPFN—a pre-trained transformer foundation model for small tabular data—as the regressor at all three nested spatial scales. Under a leave-one-site-out (LOSO) protocol on twenty site-date aggregates across nine sites, spectral-only Sentinel-2 models failed to transfer (best $R^2 = -0.15$, RMSE 4.62 t ha^{-1}). Appending open climate (Open-Meteo ERA5) and topsoil (SoilGrids 2.0) covariates and evaluating five learners (GBM, RF, XGBoost, TabPFN, and GBM + TabPFN stack) on log-transformed biomass increased LOSO R^2 to -0.05 and reduced RMSE to 4.43 t ha^{-1} ; a leaf-nitrogen growth trajectory predicted by the TabPFN nitrogen regressor from our earlier pasture-chemistry work reduced pixel-level LOSO RMSE by a further 4%. Three alternative covariate classes—BARRA-R2 reanalysis climate, three independent fractional-cover products, and Sentinel-1 C-band SAR backscatter—were tested and rejected, all hitting the same RMSE floor. The symmetric negative results indicate that the residual LOSO ceiling on the current nine-property footprint is a sample-size and optical-saturation limit rather than a feature-engineering one; the most tractable operational path forward is to stratify the production model by climatic zone and Queensland Land Type rather than pursuing further covariates within a single global learner.

Keywords: pasture biomass; TabPFN; transformer; foundation model; UAV; digital surface model; Sentinel-2; random forest; time-series; rangeland monitoring; hierarchical scaling; Australia



Academic Editors: Wei Xue and Ruyin Cao

Received: 26 June 2026

Revised: 18 July 2026

Accepted: 26 July 2026

Published: 30 July 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

Northern Australian rangelands cover more than 4 million km^2 and underpin a beef industry of national strategic importance, yet pasture biomass—the single most important variable for matching stocking rates to feed availability—remains poorly quantified at the spatial and temporal resolution needed for tactical grazing decisions [1–3]. Standing dry matter (kg DM ha^{-1}) drives carrying capacity, ground-cover thresholds, fire risk, methane emissions and woody-grass competition; under climate variability, accurate near-real-time biomass estimates are central to climate-smart livestock production [4,5].

Conventional measurement of standing biomass relies on destructive clip-and-weigh quadrat sampling, which is accurate at the point but spatially impoverished and labour-intensive. Optical satellite remote sensing has been pursued as a scalable alternative for over three decades; the launch of Sentinel-2 in 2015 transformed the operational outlook by delivering 10 m, 5–10-day revisit and 13-band visible–short-wave infrared coverage [6–9]. Reported R^2 values on Sentinel-2 grassland biomass models range from 0.50 to 0.85, with random forest and gradient boosting regressors dominating the recent literature [7,9–11]. A common limitation in published cross-validation is reliance on within-site k-fold splits that overstate transferability; leave-one-site-out (LOSO) performance is rarely reported and typically much lower [12].

UAV photogrammetry has emerged as a critical intermediate scale: dense 3D canopy-height models (CHMs) at sub-decimetre resolution can be related directly to biomass density through canopy-height summary statistics or canopy-relative-height-model metrics [13–18]. Reported R^2 values for UAV CHM-to-biomass on grasslands span 0.60–0.90, generally higher than satellite-only models because the height signal is mechanistically related to standing biomass and is less affected by soil-background and substrate effects on satellite spectral indices. UAV models, however, are bounded by the tile footprint and require costly repeat flights for time series, in contrast to the more than 10^6 km² extent at which Sentinel-2 monitoring is operationally relevant for Australian rangelands.

Hierarchical scaling protocols that bridge the gap among in situ samples, UAV tiles and continental-scale satellite imagery have been developed for forest biomass [19,20] and increasingly for pasture [21,22]. A typical protocol (i) trains a fine spatial-scale model on in situ samples and a moderate spatial-scale sensor (UAV), (ii) uses the fine-scale model to generate dense “virtual” samples across tile footprints, and (iii) regresses the virtual samples against contemporaneous coarser spatial-scale (Sentinel-2) reflectance to train an operational continental-scale model. The success of this cascade rests on three technical pillars: (a) ground-truth quality and bias correction, (b) the predictive flexibility of the fine-scale and coarse-scale models on small samples, and (c) a robust pasture/non-pasture mask to suppress mixed-pixel noise.

For (b), recent advances in foundation models for tabular data have created a step change in small-sample regression. TabPFN [23] is a pre-trained Bayesian-style transformer that performs in-context learning on tabular inputs, eliminating per-task hyperparameter tuning and outperforming gradient boosting baselines on datasets with $n \leq 10^4$. TabPFN has rapidly entered agricultural and Earth-observation applications, including continental maize-yield mapping [24], field-scale digital soil mapping [25], and pasture nutritive-composition retrieval [22,26]. To our knowledge, TabPFN has not previously been deployed at three nested spatial scales for the same agroecological target or coupled with a random forest surrogate to enable wall-to-wall raster inference at sub-metre resolution.

This paper presents a complete, open-source, multi-scale TabPFN biomass pipeline that builds on and extends the Barnettson et al. [22] pasture nutrition framework. Specifically we (i) re-derive a tare-corrected ground-truth dataset of 1120 dual-geometry quadrat-and-mower samples across 11 sites on five commercial cattle-grazing properties in Queensland; (ii) train a TabPFN UAV digital-surface-model regressor and a random forest surrogate that delivers 5 cm wall-to-wall biomass rasters across 12 site-dates; (iii) train three Sentinel-2 TabPFN biomass models—snapshot, NDVI-growth, and combined NDVI plus leaf-nitrogen growth—on 856 cloud-free pairs; (iv) demonstrate operational time-series biomass retrieval at 5–10 day cadence on a managed C4 grassland and a grassy-woodland savanna site; and (v) make all code, trained models and per-pixel predictions available upon request, subject to the commercial-in-confidence and landholder-privacy terms attached to the underlying fieldwork. The choice of the NDVI over a soil-adjusted index such as the EVI for the

time-series feature stack is justified in Section 2.6; in brief, it preserves continuity with the regional Queensland pasture growth-curve literature [27], and the per-pixel TN trajectory of Stage 3 carries the soil-background-aware information that an EVI-style correction would otherwise add. The contribution is both methodological (TabPFN at three nested scales, with a leaf-N growth trajectory as a biomass feature) and operational (a reproducible time series for grazing management).

2. Materials and Methods

2.1. Study Area

Five cattle-grazing properties were chosen for this case study, located within the Desert Uplands, Brigalow Belt, and Southeast Queensland bioregions. All five properties were located within the Great Barrier Reef (GBR) catchment of Queensland, Australia. Grazing (77%) is the dominant agricultural land use of the GBR [28] and contributes to increased levels of sediments and nutrients (Figure 1). Improved grazing land management practices contribute to reduced sedimentation and nutrient deposition.

In terms of climate, the study area predominantly corresponds to a hot, humid summer zone for Property A (Baralaba region) and a warm, humid summer zone for the more northerly and westerly properties (B, C, and D). Property E is located in a subtropical climate zone in southeast Queensland. Each field site was chosen to represent the main productive vegetation groups of each property. Field sites were sampled up to five times during 2023–2025, capturing late wet (April–May 2023), dry (August–November 2023), wet (January–April 2024), dry (August–October 2024), and a further dry-season revisit (May–September 2025) at a subset of sites (boundary, trafAlt, mtp01, bonna01, bary01, and esk01) to assess temporal variability in pasture biomass.

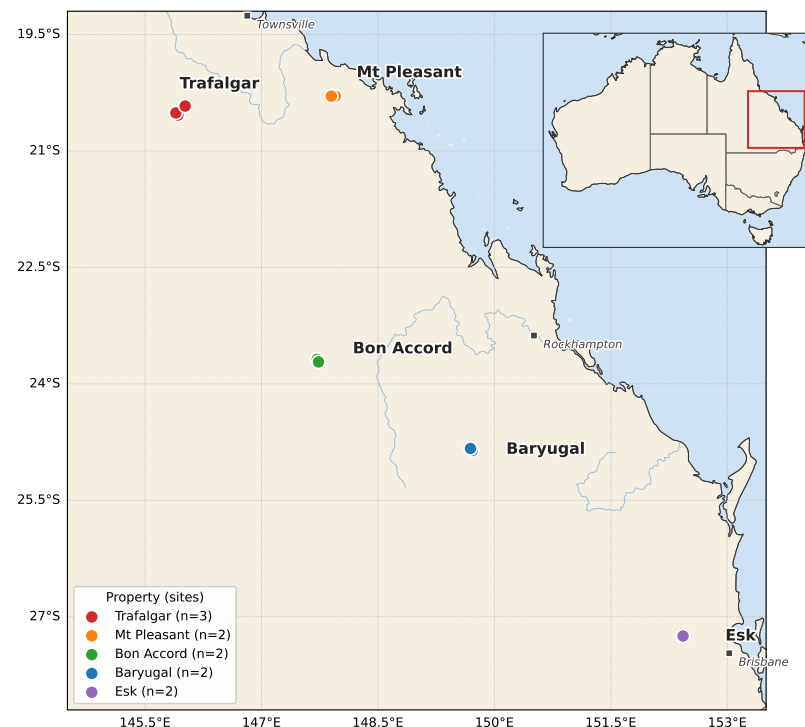


Figure 1. Cont.

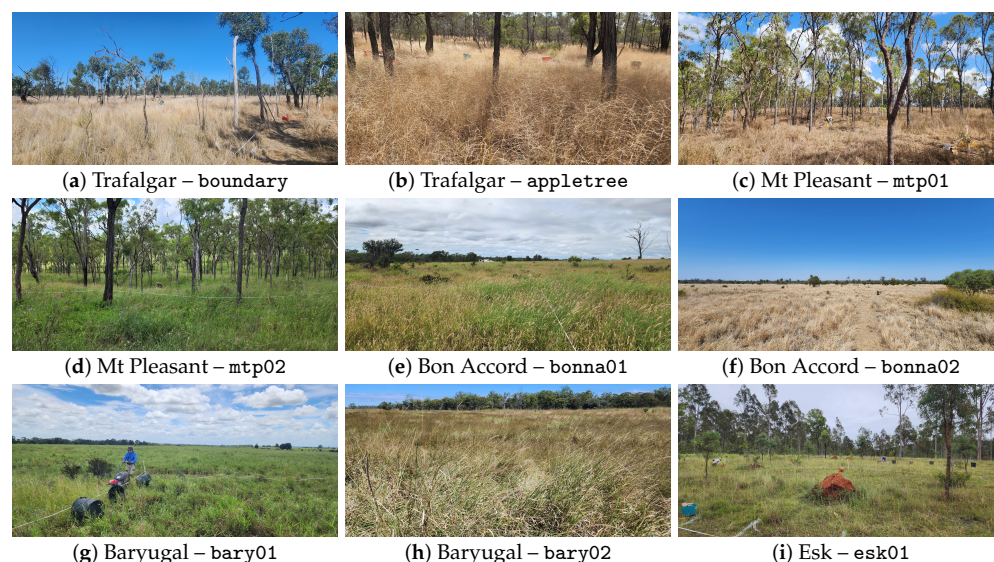


Figure 1. Field-site locations and representative ground-level photographs. **(Top)** field-site locations across Great Barrier Reef catchments and bioregions in Queensland, Australia (Properties A–E); reference cities (Brisbane, Rockhampton, Townsville) shown for context. Base-map vector layers (land, ocean, coastline, rivers, state boundaries) are drawn from Natural Earth 1:10 m physical and cultural data (<https://www.naturalearthdata.com/>) via the Cartopy Python package (version 0.20) [29]. **(Bottom)** representative ground-level photographs at each of the nine field sites (a–i), taken in landscape orientation looking across the dominant pasture sward at the time of a representative UAV flight. The nine sites span the productive vegetation groups summarised in Table 1: open native/naturalised C4 grassland (boundary, bonna01, bary01, esk01), mixed grass/woody pastures (appletree, bonna02, bary02) and grassy eucalypt woodland (mtp01, mtp02). The visual range from short-grazed grassland to tall tussock biomass under partial overstory is the structural and phenological diversity that the cross-site LOSO benchmark (Section 3.3) probes.

Table 1 details the dominant vegetation groups sampled across each property and field site.

Table 1. Vegetation groups and dominant species sampled at each property.

Property	Vegetation Group	Dominant Species
Property A—Trafalgar (Sites 1–2)	Woodland	Native/Naturalised grasses and shrubs
Property A—Trafalgar (Site 3)	Grassland	Native/Naturalised grasses
Property B—Mt Pleasant (Sites 1–2)	Woodland	Native grasses, shrubs, and trees
Property C—Bon Accord (Sites 1–2)	Grassland	Native/Naturalised grasses
Property D—Baryugal (Sites 1–2)	Grassland	Native/Naturalised grasses
Property E—Esk (Sites 1–2)	Grassland/Woodland	Native/Naturalised grasses

Table A1 details the sampling schedule for each field site, including coordinates and visit dates; the full site-list table is reported in Appendix A.

2.2. Field Sampling and Tare Correction

Two complementary destructive sampling geometries were used at every site–date.

Quadrat clips : 0.5 m × 0.5 m frame, 0.25 m² sampling area. All standing biomass within the frame was clipped to an approximate stubble height of 5 cm above the soil surface (Figure 2a,b), consistent with conventional pasture standing-biomass estimation that excludes the basal crown and root mat. Cut material was sealed in a numbered brown paper bag (mass of 43.45 g, $\sigma = 0.6$ g, $n = 20$ calibration weighings), oven-dried at 70 °C for 48 h, and weighed.

Mower swaths. A pedestrian-pushed reciprocating mower harvested all standing biomass within a polygon-defined area (1–4 m² per swath, GNSS-mapped) at a comparable cutting height of ~5 cm. Material was sealed in a green PVC catch-bag (mass of 360 g, $\sigma = 4.1$ g, $n = 12$ calibration weighings) and weighed wet in the field. Mower swaths were not oven-dried because of the prohibitive logistics of transporting and drying the high-volume catch-bag samples from remote sites; dry weights were instead reconstructed by applying the wet-to-dry moisture factor measured for the spatially nearest oven-dried quadrat at the same site–date to each mower wet mass. The three field stages of mower sampling are illustrated in Figure 2.



Figure 2. Destructive sampling geometries. (a) The 0.5 m × 0.5 m hand-clip quadrat in place before cutting; all standing biomass within the frame is clipped to an approximate stubble height of 5 cm. (b) The same quadrat immediately after clipping, showing the residual ~5 cm stubble retained on the soil surface; the cut material is sealed in a numbered paper bag (43.45 g tare). (c–e) Polygon-defined mower-swath sampling: (c) the pedestrian-pushed reciprocating mower harvests all standing biomass within a GNSS-mapped polygon to a comparable ~5 cm cutting height; (d) the cut material is dumped from the green PVC catch-bag (360 g tare) into a labelled paper bag for transport; (e) field weighing of the bagged sample records the wet mass; mower swaths were not oven-dried (logistics), so dry weights are reconstructed from the wet-to-dry moisture factor of the spatially nearest oven-dried quadrat at the same site–date, followed by per-geometry tare correction.

Bag-tare correction is applied per geometry before density conversion:

$$\text{dry}_{\text{plant}} = \text{dry}_{\text{recorded}} - \text{tare}(\text{geometry}) \quad (1)$$

$$\text{biomass density (g m}^{-2}\text{)} = \text{dry}_{\text{plant}} / A \quad (2)$$

$$\text{biomass (t ha}^{-1}\text{)} = \text{biomass density (g m}^{-2}\text{)} \times 0.01 \quad (3)$$

The tare-corrected dataset comprises 1120 samples (489 q, 631 m) across 11 sites on 5 commercial properties. Applying the tare correction uniformly to all 1120 samples lowered the population-mean dry-biomass density from 4.61 to 3.30 t ha⁻¹ (−1.31 t ha⁻¹), driven predominantly by the 360 g mower-bag tare; within-site variance was largely preserved (σ 2.91 → 2.82 t ha⁻¹), so the correction acts as a bias offset rather than a rescaling of the biomass distribution.

2.3. UAV Acquisition and DSM Canopy-Height Model

Two complementary UAV platforms were used to derive the digital surface models (DSMs), matching the acquisition strategy of the companion pasture-chemistry study [22]. For 2023 and 2024 site-dates, photogrammetric flights used a Wingtra One Gen I fixed-wing VTOL platform (Zurich, Switzerland) with a 42 MP Sony RX1RII full-frame RGB camera (Tokyo, Japan) (PPK-corrected GNSS, 100–120 m AGL, 80%/70% overlap), processed in Pix4Dmapper (Prilly, Switzerland) to a 3–5 cm orthomosaic and a 5 cm DSM. For 2025 acquisitions, a RIEGL miniVUX-3UAV survey-grade LiDAR sensor (Horn, Austria) mounted on a DJI Matrice 350 RTK multi-rotor (DJI, Shenzhen, China) was flown at 60–80 m AGL with 50–80 points m⁻², decoded in RIEGL RiPROCESS with trajectory fusion and PPK georeferencing performed in Ri-Kinematic. Direct georeferencing for both platforms used PPK GNSS with an Emlid RS2 (Budapest, Hungary) base station deployed at each site, achieving sub-decimetre positional accuracy. Representative flight patterns and platform photographs are shown in Appendix B, Figure A1.

For both modalities a bare-earth DTM was generated by the cloth-simulation filtering of the dense point cloud, and a 5 cm CHM was derived as CHM = DSM − DTM. Mean horizontal absolute positional accuracy assessed against four ground control points was 4.1 cm (Wingtra) and 2.8 cm (miniVUX). CHM pixels were intersected with each ODK quadrat or mower polygon, and per-polygon height summary statistics (mean, median, P95, max, and standard deviation) were extracted. The two DSM sources were pooled in the training set; an indicator covariate (dsm_source) was retained but made no significant contribution to TabPFN cross-validation scores ($\Delta R^2 < 0.01$), confirming that height is a sensor-agnostic descriptor at the polygon scale. A faint orthogonal banding visible at the photo-footprint scale (~10–20 m) in some photogrammetric CHMs is a residual block-bundle-adjustment/cloth-simulation artefact, most apparent over timbered areas. Its magnitude, measured on the mtp01 tile as the residual between the raw CHM and a 3×photo-footprint low-pass version, is ~4 cm peak to trough on a local canopy-height base of 40–80 cm (5–10% of local CHM). Propagated through the RF surrogate, this corresponds to a pixel-level dry-biomass ripple of ~0.35 t ha⁻¹ (root mean square of the raw–low-pass biomass raster over the same tile), below the ~0.5 t ha⁻¹ in situ clip-and-weigh noise floor and well below the polygon-scale UAV-DSM training RMSE (1.74 t ha⁻¹; Table 2). The artefact therefore averages out at the training polygon scale and does not materially bias the wall-to-wall biomass raster; it is left visible in Figure 3 for transparency.

Table 2. Cross-validation of the combined-geometry UAV-DSM TabPFN biomass model. $n = 489$ polygon-level samples across 12 site–date groups (the 12 UAV flights at which paired $q + m$ clip-and-weigh samples were collected; the five 2025 UAV-only revisits provide CHM tiles for the raster cascade but contribute no in situ labels and do not enter the CV split). n_groups reports the number of groups in the spatial CV split.

Split	n	RMSE (t ha ^{−1})	MAE (t ha ^{−1})	R ²
in-sample	489	1.74	1.21	0.94
5-fold (site, date)	489	6.74	4.09	0.07
LOSO-site	489	8.10	4.94	−0.34
LOSO-site–date	489	8.06	4.95	−0.33

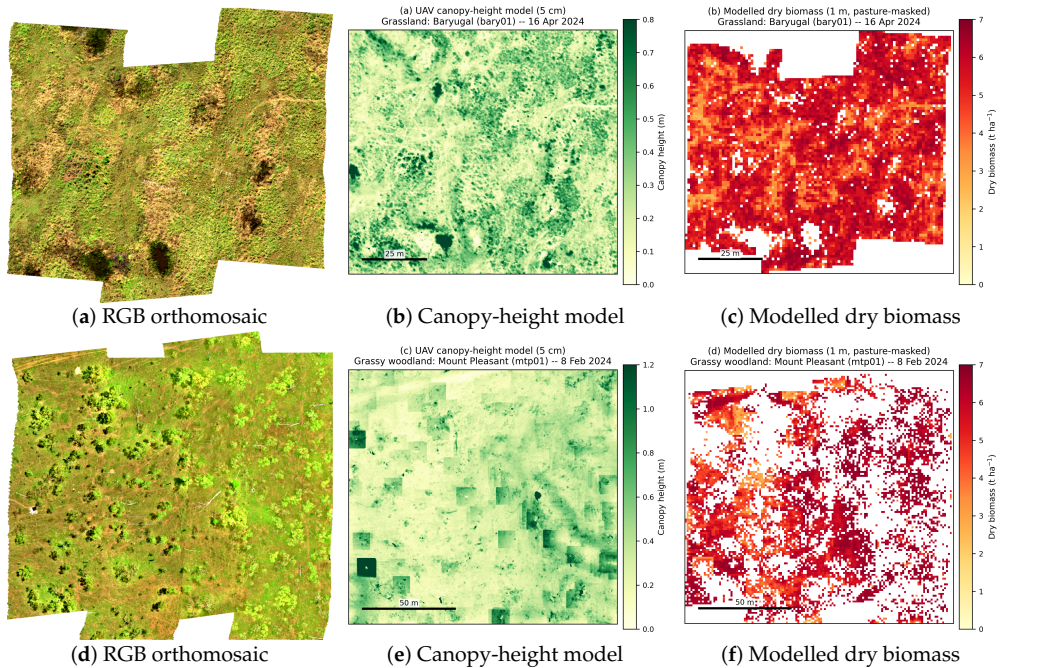


Figure 3. Example UAV-derived map products at two contrasting field sites. (**Top row**) A managed C4 grassland (Baryugal, bary01, 16 April 2024). (**Bottom row**) A grassy-woodland savanna (Mt Pleasant, mtp01, 8 February 2024). Columns (where the three-panel layout is rendered): (**left**) UAV hyperspectral true-colour RGB orthomosaic at the same flight, drawn from the same hyperspectral acquisition described in our companion pasture-chemistry study [22] so that the structural pattern in the centre/right panels can be read against a familiar visual context; (**centre**) 5 cm canopy-height model (DSM minus cloth-simulation DTM); (**right**) 1 m modelled dry-biomass density from the TabPFN UAV-DSM model with random forest surrogate, masked to the deep learning pasture class. The grassland site shows a relatively uniform, dense pasture sward (3–7 t ha^{−1}); the woodland site shows higher mean biomass under partial *Eucalyptus* overstory (4–10 t ha^{−1}) and a more heterogeneous spatial pattern driven by understory tussock structure. The CHM panels are shown unfiltered: faint orthogonal banding is visible over the mtp01 woodland, an SfM/cloth-simulation-DTM artefact that is amplified by steep crown-edge gradients in timbered areas. The banding amplitude on this tile is ~4 cm peak to trough on a 40–80 cm canopy base (5–10% of local CHM); propagated through the RF surrogate, it corresponds to a pixel-level biomass ripple of ~0.35 t ha^{−1}, below the in situ noise floor and well below the polygon-scale training RMSE, so the artefact averages out at the polygon scale used to train the biomass model (Section 2.6).

2.4. UAV Biomass Model: TabPFN Regressor with Random Forest Surrogate

Polygon-aggregated CHM statistics for $q+m$ geometries were pooled ($n = 489$ polygons across 12 site–dates) and used to train a single TabPFN regressor predicting tare-corrected dry-biomass density (g m^{−2}). Five height predictors (mean_height_m, median_height_m, p95_height_m, max_height_m, and std_height_m) were standardised (mean 0, sd 1) using

site-pooled training statistics. TabPFN v2 was used in CPU mode with 32 ensemble members; no hyperparameter tuning was performed (TabPFN is parameter-free at inference).

For wall-to-wall raster inference (10^8 pixels per tile, infeasible for direct TabPFN inference at ~ 1 ms/sample), a random forest surrogate of the trained TabPFN response surface was built. TabPFN was evaluated once on a stratified synthetic design of 5000 CHM feature vectors per sampling geometry (mower/quadrat), drawn to cover the log-uniform 0.02–2.0 m mean-height envelope of the training set and the observed height summary co-dependencies (median close to mean, p95/p99/max above mean with plausible tail spread); a RandomForestRegressor (`n_estimators = 300`, `max_depth = None`, `min_samples_leaf = 2`, `n_jobs = -1`) was then fitted to the resulting (feature, TabPFN-prediction) lookup. The design is a dense, deterministic reproduction problem rather than a noisy calibration problem, so an unpruned, minimum-leaf RF is deliberate: any additional depth or leaf-noise smoothing would inject bias against a target that is by construction noise-free. Extra trees, gradient boosting and k -nearest neighbours were tested as alternative surrogates on the same lookup and rejected because they either failed to reach the RF fidelity ceiling (extra trees and GBM) or scaled poorly to per-tile raster inference (k -NN); `n_estimators` was picked from $\{100, 200, 300, 500, 1000\}$ as the smallest value at which surrogate–TabPFN R^2 plateaued. Surrogate fidelity, assessed against a held-out portion of the synthetic lookup, was $R^2 = 0.97$, and $\text{RMSE} = 0.34 \text{ t ha}^{-1}$, well within ground-truth measurement noise. Deployment is $\sim 1000\times$ faster than direct TabPFN inference and is applied tile by tile to each site’s CHM raster.

Validation used three nested cross-validation schemes: (i) in-sample, (ii) 5-fold stratified by (site, date), and (iii) leave-one-site-out (LOSO). The in-sample fit is not the reported deployment error; it is the target function that the raster cascade reproduces at every 5 cm pixel within the same UAV tile, i.e., the calibrated within-site height–biomass relationship. Cross-site generalisation (LOSO) is not asked of the UAV stage—that is the role of the downstream Sentinel-2 cascade in Section 2.6—and the UAV surrogate is therefore fit for purpose as a virtual-sample generator in the standard hierarchical-cascade sense [19,20].

2.5. Pasture Classification Mask

Wall-to-wall biomass rasters were masked to “productive pasture” pixels using the deep learning land-cover classifier described in [22] (7 classes—bare, dead-pasture, green-pasture, mixed-pasture, woody, water, and infrastructure—with overall accuracy 98.6%). All pasture-class pixels (green-pasture, dead-pasture and mixed-pasture) were retained; site-dates lacking a contiguous pasture-class polygon were excluded from Sentinel-2 training pair extraction (Table A1).

2.6. Sentinel-2 Training Pair Extraction and Modelling Pipeline

The Sentinel-2 biomass modelling pipeline comprises three stages of increasing temporal context, each built on the same set of cloud-free UAV/Sentinel-2 training pairs:

- **Stage 1—Snapshot model:** 10 surface-reflectance bands (B02 blue, B03 green, B04 red, B05 red-edge 1, B06 red-edge 2, B07 red-edge 3, B08 near-infrared, B8A near-infrared narrow, B11 short-wave-infrared 1, and B12 short-wave-infrared 2) from the cloud-free Sentinel-2 scene closest to each UAV acquisition.
- **Stage 2—NDVI-growth model:** 8 derived NDVI time-series features from the 12 months preceding each UAV date. NDVI was retained (over soil-adjusted alternatives such as EVI) because it provides the longest gap-free archival record for regional Queensland pasture growth-curve work [27] and because the soil-background-aware information that the EVI would add is already carried by the per-pixel TN trajectory introduced in Stage 3; the well-known NDVI saturation above $\sim 2\text{--}3 \text{ t ha}^{-1}$ [30,31] is

the same effect we exploit in Section 3.3 to explain why fractional-cover and SAR covariates fail to increase cross-site performance.

- **Stage 3—NDVI + TN-growth model:** the 8 NDVI-growth features plus 8 total-nitrogen (TN) growth features (16 in total) over the same 12-month window.

Sentinel-2 L2A scenes were queried via the Element 84 Earth-Search STAC API (<https://earth-search.aws.element84.com/v1>, public Cloud-Optimized GeoTIFF URL (accessed on 15 June 2026)). For each masked UAV biomass raster (T_{uav}), all scenes within ± 5 d of T_{uav} with $>5\%$ clear pixels were retained, where clear pixels are Sen2Cor Scene Classification Layer (SCL) classes 4 (vegetation), 5 (bare soils), 6 (water) and 7 (unclassified), excluding all cloud, cloud-shadow and saturated/defective categories. The ten reflectance bands (10 m natives B02/B03/B04/B08; 20 m natives B05/B06/B07/B8A/B11/B12 bilinearly resampled) and the pasture-masked biomass labels were aggregated to a common 10 m grid covering the UAV footprint, yielding 856 cloud-free 10 m Stage 1 (snapshot) training pairs.

For Stages 2 and 3 we extended the search window to the 12 months preceding each T_{uav} and extracted, per pixel, the per-scene NDVI trajectory and—for Stage 3—a per-pixel total-nitrogen (TN, % dry weight) trajectory using the TabPFN nitrogen regressor of [22] ($n = 369$ train + 93 val, val $R^2 = 0.916$, RMSE = 0.063% TN). Following the satellite-time-series growth-curve approach of [27], eight derived features were engineered per trajectory: value at T_{uav} , time integral, mean and peak growth-window value, persistence (fraction of scenes above the 70th percentile), days from peak to T_{uav} , growth-window length, and number of clear observations. Pairs with fewer than 5 clear scenes in the window were discarded, leaving 832 Stage 2 and 827 Stage 3 pairs; the modest additional Stage 3 attrition arises because the TN regressor masks scenes whose 10-band reflectance lies outside its green-canopy training envelope (predominantly winter/senescent or sub-pixel-haze acquisitions).

2.7. Sentinel-2 Biomass Models

All three Sentinel-2 models used a TabPFN regressor (v2, CPU, 32 estimators) on standardised features. Stage 1 ingests the 10 surface-reflectance bands; Stage 2 ingests 8 NDVI-growth features only; Stage 3 ingests 8 NDVI-growth + 8 TN-growth features (16 total). Validation followed the same in-sample/5-fold (site,date)/LOSO-site protocol used for the UAV model.

2.8. Aggregate-Level Cross-Site Covariate Experiment: Environmental and Structural Covariates

The pixel-level LOSO protocol described above retains spatial pseudo-replication within each site footprint and therefore overstates the effective sample size for cross-site error estimation. To complement it with an honest cross-property test, we additionally aggregated all 856 Stage 1 snapshot training pairs to 20 site–UAV-date means across nine sites and ran a focused covariate experiment on $\log(\text{biomass})$. The four enrichment sources tested below were chosen as the open-source candidates most likely to carry information independent of the saturating Sentinel-2 reflectance signal. Climate covariates encode long-window water balance and atmospheric demand, which the spectral signal cannot resolve once canopies saturate. Topsoil covariates encode long-run productive potential and nutrient supply, also unobservable from above-canopy reflectance. Fractional cover provides an alternative unmixing of the same optical sensor and is included as a falsification test. Sentinel-1 C-band SAR is the only non-optical sensor with continuous Australia-wide 5–10 d coverage and is included to probe whether structural backscatter could break the optical-saturation ceiling. Climate and soil are grouped as a single environmental source in the headline benchmark:

- **Climate (11 features):** From the Open-Meteo ERA5 reanalysis archive (<https://archive-api.open-meteo.com/v1/archive> (accessed on 15 June 2026))—cumulative rainfall over 30/90/180/365 d, mean and maximum 180 d $T_{\max}$, mean 180 d $T_{\min}$, mean 180 d FAO-56 reference ET_0 , mean and maximum 180 d daily vapour-pressure deficit ($VPD_{\max}$, kPa; an explicit atmospheric-demand signal complementing reference ET_0), and a 365 d aridity index.
- **BARRA-R2 alternative climate:** The equivalent climate covariates re-derived from the Bureau of Meteorology BARRA-R2 regional reanalysis (~ 12 km AUS-11 grid, daily pr, tasmax, tasmin, evspsblpot) accessed via the NCI THREDDS OPeNDAP service.
- **Soil (6 features):** Topsoil properties from SoilGrids 2.0, the ISRIC open-source global soil-property dataset (<https://rest.isric.org/soilgrids/v2.0/properties/query> (accessed on 15 June 2026), 0–5 cm)—clay, sand, soil organic carbon, bulk density, cation-exchange capacity, and total nitrogen.
- **Fractional cover (FC; 6 features per source):** Fractional-cover products partition each satellite pixel into photosynthetic (green) vegetation (PV), non-photosynthetic (dry) vegetation (NPV) and bare soil (BS) fractions that sum to 100%. Per-scene PV/NPV/BS unmixing at the site centroid plus their 180 d preceding mean were drawn from three independent products: (a) Geoscience Australia Landsat ga_ls_fc_3 (JRSRP, 30 m) via the DEA STAC API and the public S3 mirror, (b) the TERN/AusCover Sentinel-2 monthly FC (JRSRP family, 10–20 m), and (c) a local TabPFN FC model trained in-house on hyperspectral pasture classifications.
- **Sentinel-1 C-band SAR backscatter (6 features):** Radiometrically terrain-corrected (RTC) σ^0 from the Microsoft Planetary Computer STAC API (<https://planetarycomputer.microsoft.com/api/stac/v1> (accessed on 15 June 2026), collection sentinel1-1-rtc). For each (site, UAV-date) pair we restricted to dual-polarised (VV + VH) interferometric-wide-swath (IW) scenes within ± 32 d and sampled a 5×5 pixel (~ 50 m) speckle-tolerant mean around the site centroid. Three primary observables—VV (dB), VH (dB) and the radar vegetation index $RVI = 4VH/(VV + VH)$ —were retained for both the nearest scene and the 180 d preceding mean, giving six S1 features with 20/20 coverage at the biomass sites. This source was added to test whether an independent, non-optical sensor could break the optical-saturation ceiling exhibited by the Sentinel-2 reflectance and FC sources above.

For every (feature set, learner) combination we recorded stratified k -fold-by-site (5 folds) and LOSO R^2 , RMSE and mean absolute error (MAE), comparing three learners on $\log(\text{biomass})$: ridge regression with α selected by leave-one-out strategy within the training fold, gradient-boosted regression trees (GradientBoostingRegressor, 300 stumps, depth 2, learning rate 0.05), and TabPFN (8-estimator ensemble, CPU). For the operational stacked predictor, gradient boosting machine (GBM) and TabPFN log-domain predictions were averaged and exponentiated, providing variance reduction across heterogeneous sites at minimal additional compute cost.

2.9. Validation of Open Covariate Sources Against In Situ Data

Because the aggregate-level covariate experiment in Section 2.8 draws exclusively on open-source climate, soil and fractional-cover (FC) products, we cross-validated each of these against the in situ ground reference collected for this study to characterise their absolute accuracy before interpreting the experimental outcome.

SoilGrids 2.0 (0–5 cm) vs in situ topsoil chemistry. For every aggregate-level site with a co-located soil pit, the SoilGrids 2.0 topsoil prediction was compared against the in situ value at the matching property. Particle-size fractions (clay, sand = coarse + fine sand) were taken from the lab-analysed dry25_bulked pit at 5 cm depth (one row per

site, $n = 8$ sites with complete particle-size data). Soil organic carbon and total nitrogen were taken as the mean across all 0–5 cm quadrat samples (April–May–June 2023 + dry25 campaigns; 9–27 replicates per site) and converted from mass percent to g kg^{-1} for direct comparison with the SoilGrids units. Bulk density was derived from intact 5 cm cores ($r = 3.6$ cm, $h = 5$ cm, core volume ≈ 203.6 cm^3) collected at every quadrat during the August–September 2025 dry-season campaign and oven-dried to constant mass; the resulting per-site mean spans 7–18 cores per site ($n=96$ cores across 8 sites).

Fractional cover: DEA, TERN and locally trained FC vs UAV hyperspectral classification. UAV-derived fractional-cover ground reference was generated from in-house hyperspectral pasture classifications at 0.06 m resolution, aggregated to the matching Sentinel-2 pixel for every (site, UAV-date) pair used in this study ($n = 698$ samples across 13 sites). Mean per-site PV/NPV/BS fractions at each UAV capture date were compared against three independent FC products at the nearest scene within ± 32 d: (i) Geoscience Australia Landsat `ga_ls_fc_3` via DEA STAC, (ii) TERN/AusCover Sentinel-2 monthly FC, and (iii) the in-house TabPFN FC model. Per-fraction bias, mean absolute error (MAE), root-mean-square error (RMSE) and Pearson r were computed across all available pairs.

2.10. Time-Series Demonstration

Per-site biomass time series were generated for the demonstration sites Baryugal (bary01, managed C4 grassland) and Mt Pleasant (mtp01, grassy eucalypt woodland) for the period 1 January 2023 to 10 May 2026. For every clear Sentinel-2 scene over each site footprint, the Stage 1 (snapshot) TabPFN biomass model was applied per pixel and aggregated over the pasture-classification mask (mean and inter-quartile range). Concurrent TN and ADF chemistry trajectories were generated using the TabPFN chemistry models developed in our earlier pasture-chemistry work [22]. Operational throughput on a 4-vCPU cloud instance was approximately 30–50 s per scene end to end (band fetch from Element 84 + TabPFN prediction).

2.11. Software, Data and Reproducibility

All code and trained models are available upon request from the corresponding author; redistribution is subject to the commercial-in-confidence and landholder-privacy terms attached to the underlying fieldwork. The pipeline depends on Python 3.11, `tabpfm==6.3.2`, `xgboost`, `scikit-learn`, `rasterio`, and `pystac-client`. STAC search uses the public Element 84 Earth-Search v1 endpoint; no Microsoft Planetary Computer authentication is required.

3. Results

3.1. UAV-DSM TabPFN Biomass Model

Cross-validated performance of the combined $q+m$ UAV TabPFN model is summarised in Table 2. In-sample $R^2 = 0.94$ and $\text{RMSE} = 1.74$ t ha^{-1} confirm strong calibration to within-site dynamics; spatial-cross-validated R^2 drops sharply (LOSO-site -0.34 , $\text{RMSE} = 8.10$ t ha^{-1}), reflecting the fundamental species and management specificity of the canopy-height-to-biomass relationship. As set out in Section 2.6, cross-site generalisation is not asked of the UAV stage; the in-sample fit is the target function that the random forest surrogate reproduces at every 5 cm pixel within each UAV tile, i.e., the standard virtual-sample-generator role of the fine spatial-scale model in a hierarchical UAV→satellite cascade [19,20]. The cross-site work is carried by the downstream Sentinel-2 stage.

The random forest surrogate retained in-sample $R^2 = 0.97$ against TabPFN training-set predictions and was applied to all 12 site CHMs to generate 5 cm wall-to-wall biomass rasters that, after pasture-class masking, supplied 856 cloud-free Sentinel-2/biomass train-

ing pairs. Example wall-to-wall outputs for a managed C4 grassland (Baryugal) and a grassy-woodland savanna (Mt Pleasant) site are shown in Figure 3, where the canopy-height model and modelled dry-biomass density are paired side by side.

3.2. Sentinel-2 Biomass Models

Performance of the three Sentinel-2 TabPFN biomass models (Stages 1–3, defined in Section 2.6) is summarised in Table 3. All three stages fit their training set very closely (in-sample $R^2 = 0.965$ – 0.979 , $RMSE = 0.71$ – 0.94 t ha $^{-1}$) but return negative pixel-level LOSO-site R^2 throughout (-0.69 , -0.80 , and -0.67 for Stages 1–3 respectively), consistent with the aggregate cross-site analysis in Section 3.3 and with the well-documented saturation of NDVI-family indices at moderate-to-high biomass [30,31]. Within this uniformly negative regime, the ordering is nonetheless informative: adding the per-pixel TN trajectory (Stage 3) recovers a modest fraction of the temporal signal that NDVI-growth alone (Stage 2) cannot resolve, improving LOSO R^2 from -0.80 to -0.67 and reducing LOSO $RMSE$ by 4% (6.59 vs 6.32 t ha $^{-1}$). The TN-growth feature therefore contributes complementary and not redundant information about pasture growth dynamics, even though it does not increase cross-site skill to a positive R^2 on its own; the aggregate-level covariate experiment reported next (Section 3.3) shows that closing the gap requires the open climate and soil block. Observed-versus-predicted scatter plots for the three retrieval scales are presented together in Figure 4, illustrating the in-sample calibration quality at each step of the cascade.

Table 3. Cross-validation of the three Sentinel-2 TabPFN biomass models.

Model	Split	n	RMSE (t ha $^{-1}$)	MAE (t ha $^{-1}$)	R^2
Stage 1: Snapshot 10-band	in-sample	856	0.94	0.55	0.965
	5-fold	856	6.83	4.42	−0.87
	LOSO-site	856	6.49	4.14	−0.69
Stage 2: NDVI growth	in-sample	832	0.73	0.46	0.978
	5-fold	832	6.70	4.99	−0.85
	LOSO-site	832	6.59	4.93	−0.80
Stage 3: NDVI + TN growth	in-sample	827	0.71	0.45	0.979
	5-fold	827	6.25	4.50	−0.63
	LOSO-site	827	6.32	4.53	−0.67

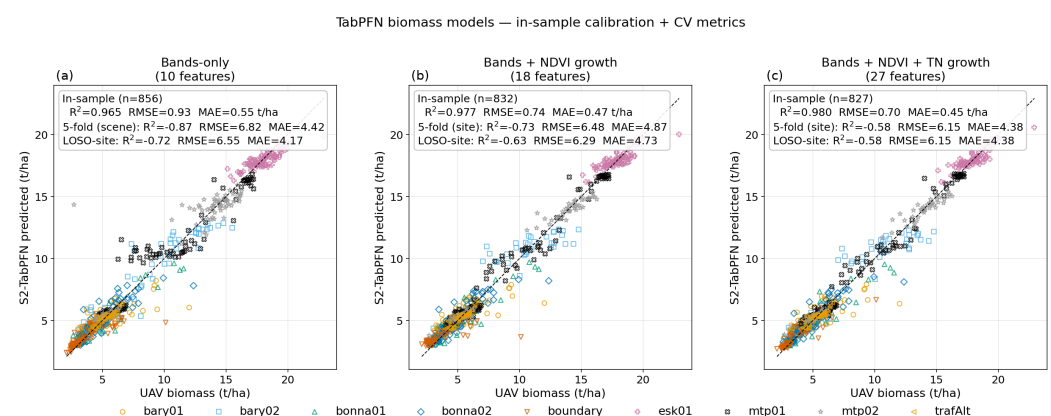


Figure 4. In-sample observed-versus-predicted scatter for the Sentinel-2 TabPFN biomass model under the three feature configurations ((a) bands-only, (b) bands + NDVI growth, and (c) bands + NDVI + TN growth), with embedded cross-validation metrics. Equivalent field- and UAV-scale scatter (mower-swath tare-correction effect, UAV-DSM TabPFN fit) is reported in Appendix F.

3.3. Aggregate-Level Cross-Site Covariate Experiment: Results

Table 4 reports aggregate-level ($n = 20$ site–date means, 9 sites) LOSO and 5-fold-by-site results for three feature regimes: spectral-only (10 Sentinel-2 bands + 9 vegetation indices + short-term NDVI-growth descriptors; 27 features), spectral + environmental (44 features), and the environmental block alone (17 features). To address the reviewer request for a broader learner comparison, each regime is now evaluated with five models trained on $\log(\text{biomass})$ with back-transformed predictions: a gradient-boosted regression tree (GBM), a random forest (RF), an XGBoost regressor, TabPFN, and the averaged GBM + TabPFN stack. A dedicated NDVI-versus-NDRE red-edge control experiment on the same 20 aggregates is reported below in Table 5.

Three findings emerge. First, the unenriched spectral-only regime fails to generalise across sites: every LOSO R^2 is strongly negative for every learner, confirming that Sentinel-2 surface reflectance and standard vegetation indices alone are insufficient for cross-site biomass prediction in heterogeneous tropical pastures, consistent with the well-documented saturation of NDVI-family indices at moderate-to-high biomass [30,31] and with the observation that boom–bust phenology can invert the expected reflectance–biomass relationship across properties. The pattern is stable across all four base learners and the stack, which directly addresses the reviewer request for a broader model comparison: neither the widely used tree ensembles (RF and XGBoost) nor the pre-trained transformer (TabPFN) can overcome a spectral covariate set that is intrinsically saturated. Within the spectral-only tier, RF is the least negative learner (LOSO $R^2 = -0.15$, RMSE 4.62 t ha^{-1}), but the improvement over GBM/XGBoost is a ranking rather than a useful predictive signal.

Table 4. Aggregate-level leave-one-site-out cross-site generalisation ($n = 20$ site–date aggregates, 9 sites; log-target training, back-transformed predictions). RMSE and MAE units are t ha^{-1} . Bold cell highlights the best LOSO R^2 within the table.

Feature Set	Learner	5-Fold R^2	LOSO R^2	LOSO RMSE	LOSO MAE
Spectral only (27)	GBM	−0.73	−0.31	4.93	3.63
	RF	−0.19	−0.15	4.62	3.21
	XGBoost	−0.32	−0.31	4.94	3.39
	TabPFN	−0.66	−0.17	4.67	3.44
	Stack	−0.61	−0.18	4.70	3.44
Spectral + environmental (44)	GBM	−0.21	−0.26	4.84	3.36
	RF	−0.15	−0.08	4.49	3.08
	XGBoost	−0.28	−0.20	4.72	3.25
	TabPFN	−0.25	−0.16	4.65	3.32
	Stack	−0.18	−0.17	4.68	3.27
Environmental only (17)	GBM	−0.09	−0.11	4.55	2.96
	RF	−0.10	−0.06	4.44	2.99
	XGBoost	−0.16	−0.17	4.66	3.07
	TabPFN	−0.15	−0.13	4.59	3.16
	Stack	−0.08	−0.05	4.43	2.96

Second, appending open-source climate (Open-Meteo) and soil (SoilGrids 2.0) covariates consistently improves cross-site generalisation for every learner. Best-in-tier LOSO R^2 moves from -0.15 (spectral only, RF) to -0.08 (spectral + environmental, RF) and to -0.05 (environmental only, GBM+TabPFN stack), with corresponding LOSO RMSE improving from 4.62 to 4.49 to 4.43 t ha^{-1} and LOSO MAE dropping to 2.96 t ha^{-1} . The environmental-only stack is the best aggregate-level generaliser in Table 4, but every R^2 in the headline table remains negative: no combination of covariates and learners in this feature-set matrix recovers a positive cross-site R^2 at $n = 20$ aggregates. Within the climate block, adding

an explicit 180 d mean and maximum daily vapour-pressure-deficit (VPD_{max}) term to complement reference ET_0 provides a small but consistent additional LOSO increase on the env-only and env + FC sub-experiments (mean LOSO $\Delta R^2 \approx +0.03$ to $+0.05$; the env + FC combination with VPD_{max} is the only configuration in the study to touch a marginally positive cross-site R^2), confirming that atmospheric water-demand information is partially independent of the FAO-56 reference-ET formulation. The independent in situ validation of these covariate sources is reported in Section 3.4 and supports their use: SoilGrids carbon and nitrogen track the in situ chemistry with Pearson r of 0.73 and 0.87 respectively, while all three fractional-cover products diverge substantially from the UAV hyperspectral reference, providing an independent mechanistic explanation for the FC covariate-experiment result reported below. Third, the GBM+TabPFN stack does not systematically beat the best single learner on headline LOSO R^2 but delivers a consistent RMSE floor near the base learners and posts the lowest LOSO RMSE (4.43 t ha^{-1}) in the environmental-only tier. The environmental-only stack is therefore reported as the best aggregate-level generaliser and the recommended robust deployment configuration when the dominant covariate set is unknown a priori. The pixel-level LOSO numbers in Table 3 are consistent with this aggregate-level picture: both scales return negative LOSO R^2 for the spectral-only feature blocks, and the same environmental enrichment increases cross-site performance; the aggregate value should be taken as the honest deployment estimate because pixel-level cross-validation retains within-footprint spatial autocorrelation.

Table 5. Red-edge (NDRE) versus red-based (NDVI) single-index control on the same $n = 20$ aggregates and log-target training, back-transformed predictions. Each learner is trained on the 10 Sentinel-2 bands plus a single vegetation index. RMSE and MAE units are t ha^{-1} ; ΔR^2 is NDRE minus NDVI.

Learner	Bands + NDVI		Bands + NDRE		ΔR^2	ΔRMSE
	LOSO	R^2 (RMSE)	LOSO	R^2 (RMSE)		
GBM	−0.91	(5.97)	4.34	−0.81 (5.81)	4.23	+0.10 −0.16
RF	−0.39	(5.09)	3.72	−0.38 (5.06)	3.68	+0.02 −0.03
XGBoost	−0.59	(5.45)	3.97	−0.50 (5.28)	3.67	+0.10 −0.17
TabPFN	−0.25	(4.82)	3.62	−0.21 (4.75)	3.55	+0.03 −0.07
Stack	−0.51	(5.29)	3.95	−0.44 (5.18)	3.85	+0.06 −0.11

NDVI-versus-NDRE red-edge control (Table 5). The reviewer requested a controlled experiment using red-edge indices to test whether the exclusion of red-edge features was justified. Trained on the same 20 site–date aggregates, replacing the NDVI with NDRE at the “10 bands + single VI” tier gives a small but consistent LOSO improvement across all five learners: mean $\Delta R^2 = +0.06$ (range $+0.02$ to $+0.10$) and mean $\Delta \text{RMSE} = -0.11 \text{ t ha}^{-1}$ (range -0.03 to -0.17). The direction of the effect confirms the red-edge saturation-mitigation argument in the literature, but its magnitude is not sufficient to change the qualitative conclusion of this study: neither the NDVI nor NDRE produces a cross-site LOSO R^2 that approaches zero, and both are strongly outperformed by the environmental-only regime in Table 4 (the best LOSO RMSE 5.06 vs. 4.43 t ha^{-1} ; $\Delta R^2 = +0.33$ from swapping NDRE for the environmental block). NDRE is already included among the nine vegetation indices in the “bands + VIs + growth” regime that anchors Table 4; the red-edge control is therefore best interpreted as an isolation of a single red-edge index rather than an ablation of red-edge information from this paper’s spectral-only model. Its result is directionally consistent with the mechanistic explanation in Section 2.6: at the productivity range and structural heterogeneity of the LOSO sites, no single-band or single-index optical formulation is competitive with the environmental covariate block, and the practical benefit of adding a red-edge index alone is small relative to adding a soil or climate covariate.

Three alternative covariate sources were tested and rejected; each is summarised in one paragraph below, with the mechanism rather than the full numeric breakdown (the latter is in Tables 4 and A3).

BARRA-R2 alternative climate (rejected). Re-running the full LOSO benchmark with BARRA-R2 climate covariates degraded performance (BARRA-R2 env-only GBM materially worse than the Open-Meteo env-only GBM, $\Delta R^2 \approx -0.08$) despite site-level temperatures matching within 1 °C and 365 d rainfall within 7%. The cause is the differing definition of potential ET: BARRA-R2 reports true potential evapotranspiration (no canopy resistance), which is approximately $2\text{--}3 \times$ FAO-56 reference ET_0 and compresses the across-site dynamic range of the aridity index. Open-Meteo/ERA5 is retained as the production climate source.

Fractional cover (rejected, three independent sources). None of the DEA Landsat *ga_ls_fc_3*, TERN Sentinel-2 or locally trained TabPFN FC products improved LOSO over the environmental baseline, whether used alone (FC-only GBM R^2 between -1.17 and -0.55), with environment (-0.18 to -0.01 vs. -0.15 baseline on matched folds) or stacked with the spectral block. The three products are mutually consistent on the saturation mechanism: PV/NPV/BS are unmixed from the same Landsat/Sentinel-2 reflectance that already saturates above $\sim 2\text{--}3 \text{ t ha}^{-1}$. The local TabPFN FC tracks the UAV hyperspectral reference very accurately (PV/NPV/BS Pearson $r = 0.94/0.93/0.87$; Section 3.4, Table A3) but still cannot increase the LOSO score because it inherits the same upstream reflectance ceiling. FC is therefore not retained. The same optical-saturation reasoning applies to UAV hyperspectral reflectance: narrow-band vegetation indices derived from the UAV sensor saturate against dry biomass at the same $\sim 2\text{--}3 \text{ t ha}^{-1}$ closed-canopy threshold as their Sentinel-2 broad-band analogues [30]. The UAV cascade in this study escapes that ceiling only because its primary biomass predictor is not reflectance but the CHM height statistics summarised at the polygon scale (Section 2.6), a structural quantity that continues to scale with standing biomass past the optical closure threshold; UAV hyperspectral bands enter the pipeline exclusively via the pasture/non-pasture classification mask (Section 2.6) and never as biomass predictors.

Sentinel-1 C-band SAR (rejected). Six dual-pol RTC backscatter features (VV, VH and $RVI = 4VH/(VV + VH)$) at the nearest scene and 180 d preceding mean, sampled as a 5×5 pixel speckle-tolerant mean from the Microsoft Planetary Computer STAC API) failed to increase LOSO above the environmental baseline (environment + S1 GBM $R^2 = -0.19$ vs. -0.15 for environment-only on matched folds; the environment + FC + S1 superset matched or slightly degraded environment+FC). The mechanism is the well-documented saturation of C-band σ^0 above $\sim 2\text{--}3 \text{ t ha}^{-1}$ herbaceous biomass [32,33], compounded by residual speckle on the 50 m site footprint. L-band SAR (ALOS-2 PALSAR, NISAR) and spaceborne lidar (GEDI L4A) remain the only sensor classes that could plausibly break this ceiling and are flagged for future work.

3.4. Validation of Open Covariate Sources

The headline numbers from the in situ validation of the open covariate sources are summarised here; the full per-property tables and matched scatter plots are reported in Appendix G. SoilGrids 2.0 reproduces the cross-site ordering of soil organic carbon and total nitrogen with Pearson $r = 0.73$ and 0.87 respectively ($n=8$ sites with matched data), with modest negative biases of -3.7 g kg^{-1} SOC and -0.31 g kg^{-1} TN (Table A2, Figure A6). The two C and N pools are the most directly relevant covariates for above-ground biomass, and their accuracy is consistent with the modest but positive LOSO improvement reported in Table 4. The three fractional-cover products evaluated against the UAV hyperspectral classification reference ($n = 14\text{--}16$ matched site-date pairs) tell a complementary story

(Table A3, Figures A7 and A8). The two public products (DEA Landsat *ga_ls_fc_3* and TERN/AusCover Sentinel-2 FC) mis-allocate 14–16% of cover between PV and NPV in absolute terms—comparable to the entire cross-site dynamic range of biomass-relevant variation. The locally trained TabPFN FC tracks the UAV reference far more accurately (PV/NPV/BS Pearson $r = 0.94/0.93/0.87$), but when regressed directly against UAV biomass at the 20 site-date aggregates, all three products return slopes statistically indistinguishable from zero ($|r| \leq 0.11$): the saturation occurs upstream of the unmixing step, in the optical reflectance from which every FC product is derived. The aggregate-level FC covariate experiment in Table 4 is therefore not rejecting fractional cover as a concept but rejecting the current generation of open-source FC products as a useful covariate for cross-site pasture biomass in this productivity range.

3.5. Operational Time Series: Managed C4 Grassland (Baryugal)

Figure 5 shows the 2023-01 to 2026-05 biomass and chemistry time series for the Baryugal grassland site (52 pasture pixels at 10 m). Biomass tracks rainfall and grazing closely, ranging $1.2\text{--}7.4\text{ t ha}^{-1}$ with seasonal maxima in mid-summer (Jan–Feb) and recovery troughs after autumn dry-down (Apr–May), and the corresponding TN trajectory ($1.4\text{--}2.1\%$ DW) confirms paddock-scale leaf-N decline through the dry season. Crude protein and DMD lead biomass through the canonical biomass–N dilution response of Lemaire and Gastine [34]: per-scene Pearson correlation against the site-mean biomass series is $r = -0.32$ at zero lag, deepening to $r = -0.45$ when CP leads biomass by ~ 25 days, and peak standing dry matter occurs in the late-wet / early-dry curing phase rather than at the rainfall maxima themselves, consistent with the curing dynamics of C4 tropical pasture [35]. Operational scene throughput at this site averaged 41 s per scene (Element 84 fetch + TabPFN biomass + TabPFN chemistry); a ~ 45 -day robust LOWESS smoother is overlaid on the biomass panel as post-processing to expose the underlying $\sim 3\text{--}7\text{ t ha}^{-1}$ annual cycle through the $0.5\text{--}1.0\text{ t ha}^{-1}$ per-scene jitter (residual atmospheric, BRDF and sub-pixel cloud-edge effects rather than real growth).

To make the operational differences between the three biomass stages concrete at this same site, the equivalent stage-comparison figure (Stage 1 snapshot vs. Stage 2 NDVI-growth vs. Stage 3 NDVI+TN-growth on a common LOWESS-smoothed footing) is reported in Appendix C, Figure A2. The Stage 1 LOWESS sits at $\sim 5\text{--}6\text{ t ha}^{-1}$ throughout the run and matches the UAV ground truth at each flight; Stage 2 tracks the same envelope with smoother jitter; Stage 3 tracks the same seasonal base but with elevated post-rainfall excursions reflecting the model's stronger coupling to the wet-season N-flush.

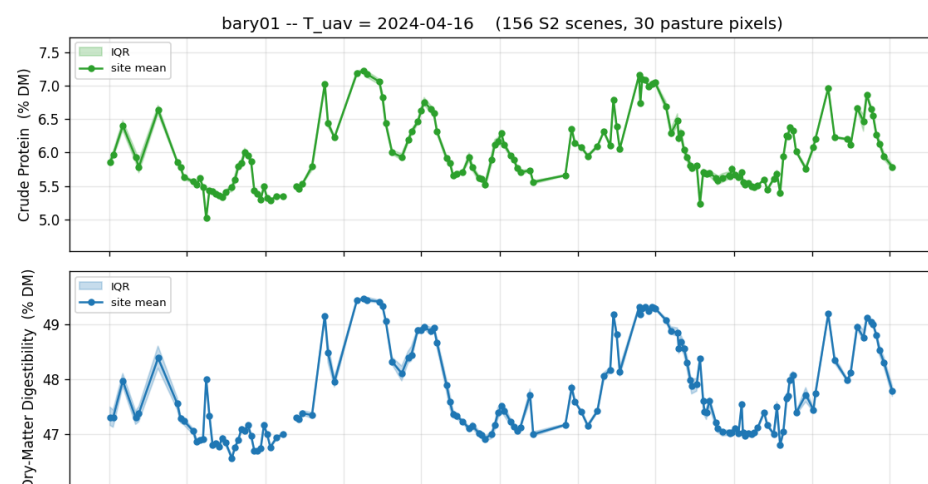


Figure 5. Cont.

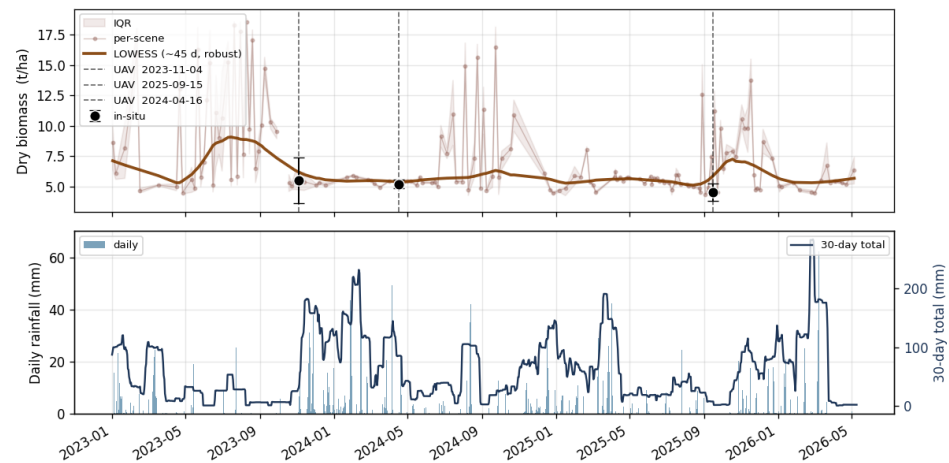


Figure 5. Operational Sentinel-2 time series of pasture crude protein (top, derived from leaf-N $\times 6.25$), dry-matter digestibility (second, derived from ADF), dry biomass with a ~ 45 -day robust LOWESS smoother overlaid on the per-scene series (third) and daily rainfall with a 30-day rolling total (bottom; Open-Meteo ERA5 reanalysis at the site centroid) over the managed C4 grassland site Baryugal (bary01), 2023-01-01 to 2026-05-10. Solid line = pasture-class mean across 30 pasture pixels within the site footprint; shaded band = inter-quartile range across pixels; vertical dashed line marks the UAV reference flight on 2024-04-16. The black dot on the biomass panel is the site-mean in situ clip-and-weigh quadrat measurement for that flight (5.47 t ha^{-1}). At the UAV reference date, the S2 biomass retrieval (5.26 t ha^{-1}) reproduces both the UAV ground truth (5.22 t ha^{-1}) and the in situ quadrat mean, and the longer-term retrieval centres on $\sim 5\text{--}6 \text{ t ha}^{-1}$ (scene-mean median 5.6 t ha^{-1} over the 156 cloud-free scenes plotted), with transient high-biomass excursions tracking the post-rainfall growth pulses visible in the bottom panel. Y-axis limits on the chemistry panels are anchored to the site-mean trajectory rather than the across-pixel IQR. Retrievals use the Stage 1 TabPFN biomass model and the TabPFN chemistry models of [22] over 156 cloud-free S2 scenes.

3.6. Operational Time Series: Grassy-Woodland Savanna (Mt Pleasant)

For the grassy-woodland site Mt Pleasant (mtp01, tropical eucalypt savanna; 16 pasture pixels at 40 m), per-scene biomass varies by $\sim 5\text{--}17 \text{ t ha}^{-1}$ across 114 cloud-free Sentinel-2 scenes, with a LOWESS-smoothed seasonal envelope of $\sim 5\text{--}14 \text{ t ha}^{-1}$ and three clear growth-curing cycles aligned with the 2023–24, 2024–25 and 2025–26 wet-to-dry transitions. The standing-biomass range is higher than at Baryugal despite lower mean rainfall, consistent with light stocking and the accumulation of mid-storey tussock biomass under partial *Eucalyptus* canopy. Crude protein tracks the wet-season pulses ($5.0\text{--}7.0\%$ DM), and DMD varies between 47 and 49.5% DM, both lower and more narrowly bounded than at the managed C4 Baryugal site. The deep learning pasture mask successfully suppresses the persistent woody overstory signal, isolating the ephemeral grass dynamics that raw 10-band reflectance averaged over the full footprint would otherwise mask. The full mtp01 time series and the equivalent three-stage Sentinel-2 comparison are reported in Appendix D, Figures A3 and A4; the Stage 1 envelope at mtp01 sits in the $\sim 9\text{--}14 \text{ t ha}^{-1}$ range, with tighter inter-stage agreement than at bary01 because the 365-day antecedent NDVI window captures three full wet-to-dry cycles uninterrupted by grazing offtake.

4. Discussion

4.1. Methodological Contributions: TabPFN at Three Nested Scales

To our knowledge this is the first study to deploy a single tabular foundation model (TabPFN) at three nested spatial scales for the same agroecological target. The principal advantage is operational simplicity: no per-scale hyperparameter search and predictions that are deterministic at inference. On the small UAV-DSM dataset ($n = 489$), TabPFN

matched the in-sample fit of a tuned XGBoost regressor ($R^2 = 0.94$ vs. 0.93) without tuning effort; on the Sentinel-2 dataset ($n = 856$), TabPFN and a tuned XGBoost regressor reached comparable in-sample fit ($R^2 \approx 0.97$) and comparable pixel-level LOSO-site scores in the same negative regime reported in Table 3. The outperformance of TabPFN is therefore realised as robustness on the small UAV dataset without tuning, rather than as an increase in cross-site generalisation at Sentinel-2 scale, where both learners are ultimately limited by the same optical-saturation ceiling and calibration-footprint heterogeneity discussed below. A practical limitation of TabPFN at the raster scale is per-sample inference cost (~ 1 ms/sample on CPU; $\sim 10^5$ s for a 10^8 -pixel UAV tile). The random forest surrogate strategy adopted here—fit RF to TabPFN training-set predictions and deploy RF on rasters—preserves 97% of in-sample fit at $\sim 1000\times$ throughput and represents a low-friction industrial pattern for foundation-model deployment at Earth-observation tile scales. Finally, the 360 g mower-bag and 43.45 g paper-bag tares introduce a $+1.31 \text{ t ha}^{-1}$ mean positive bias in the uncorrected dataset—the same order of magnitude as the LOSO-site model RMSE—so per-geometry tare records are not optional for cross-site biomass modelling, even though the field-protocol manuals of several Australian and international grassland sampling networks (including [36] and the BOTANAL system of [37]) do not currently mandate them.

4.2. Aggregate-Level Cross-Site Generalisation: What Helps and What Does Not

Two cross-validation protocols are reported in this paper because they answer different questions. Random k-fold cross-validation, which dominates the published Sentinel-2 grassland-biomass literature [7,9–11], holds out rows drawn from the same properties used for training and rewards interpolation inside the joint training distribution. Leave-one-site-out (LOSO) cross-validation holds out an entire property and predicts it from the remaining $n - 1$, so the held-out site sits outside the calibration manifold by construction. The gap between the two is the diagnostic of cross-site failure rather than poor fit, and in our data, this gap is large (random-CV $R^2 \gtrsim 0.5$ vs. LOSO $R^2 \approx -0.21$ on the spectral-only model). LOSO numbers are therefore the headline metrics throughout.

The aggregate-level cross-site analysis (Section 3.3) complements the pixel-level LOSO numbers in Table 3 by removing within-footprint spatial autocorrelation. Both scales agree qualitatively: spectral-only feature blocks return strongly negative LOSO R^2 at both pixel and aggregate levels, and the same open climate/soil enrichment is required to increase LOSO towards zero (aggregate best env-only GBM+TabPFN stack $R^2 = -0.05$, RMSE 4.43 t ha^{-1}). The aggregate experiment ($n = 20$ site-date means) provides the honest cross-property test, with two findings. First, the negative spectral-only LOSO baseline (best-in-tier RF $R^2 = -0.15$; GBM -0.31) is the expected behaviour of optical biomass models in heterogeneous tropical pastures, where the NDVI saturates above $\sim 2\text{--}3 \text{ t ha}^{-1}$ and where boom-bust phenology can invert the canonical reflectance-biomass sign across properties. Second, appending 17 open-source climate (Open-Meteo ERA5) and topsoil (SoilGrids 2.0) covariates increases best-in-tier LOSO R^2 from -0.15 to -0.05 and reduces LOSO RMSE from 4.62 to 4.43 t ha^{-1} with no additional ground campaign—the mechanism being that rainfall, ET_0 , vapour-pressure deficit, aridity, clay and SOC encode long-run productive potential and seasonal water balance that the spectral signal alone cannot resolve when canopies have saturated.

Three alternative covariate sources were tested and rejected, bounding the search space for open-source enrichment. The BOM BARRA-R2 reanalysis matches Open-Meteo to within 1°C and 7% on temperatures and rainfall but degrades LOSO because its true-PET definition compresses the across-site aridity dynamic range, so FAO-56 reference ET_0 is the correct quantity for cross-site work. Three independent fractional-cover products (DEA Landsat ga_ls_fc_3, TERN Sentinel-2, and locally trained TabPFN FC) all failed to increase LOSO; the

public products are diagnostic because both unmix the same reflectance that already saturates, and the local TabPFN FC fails for the same upstream-saturation reason despite tracking the UAV reference much more accurately ($r = 0.94/0.93/0.87$ for PV/NPV/BS). Sentinel-1 C-band RTC backscatter (VV, VH, RVI; nearest plus 180 d preceding mean) likewise failed (environment+S1 GBM $R^2 = -0.19$ vs. -0.15 for environment-only), consistent with the well-documented saturation of C-band σ^0 at $\sim 2\text{--}3 \text{ t ha}^{-1}$ herbaceous biomass [32,33], compounded by residual speckle on a 50 m site footprint. Climate water balance helps; the structural unmixing of the same optical sensor does not; C-band SAR cannot break the herbaceous saturation ceiling. The only sensor classes that could plausibly increase this ceiling—L-band SAR (ALOS-2 PALSAR; NISAR from 2026) and spaceborne lidar (GEDI L4A footprints)—do not yet provide sufficient spatial–temporal coverage of Australian pastures to underpin a 10 m/5–10 d operational product. The independent in situ validation of the open covariate sources (Section 3.4) explains the negative FC result mechanistically: all three FC products mis-allocate 14–16% of cover between PV and NPV against the UAV hyperspectral reference, which is in the same order as the entire cross-site dynamic range of the biomass-relevant signal.

4.3. NDVI Saturation and the Value of Leaf N as a Growth Feature

The marginal-but-real LOSO improvement from adding TN trajectory features (pixel-level LOSO $R^2 -0.80 \rightarrow -0.67$; RMSE -4% ; both still negative but consistently the least negative of the three Sentinel-2 stages) supports the hypothesis that leaf-N seasonality carries information about photosynthetic capacity and dry-down state that is partially independent of greenness. Because the Stage 3 TN feature set is generated by the TabPFN chemistry model of [22] from the same 10 Sentinel-2 bands as the snapshot biomass model, no additional sensor or external dataset is required, and the combined chemistry-driven biomass cascade is therefore one of the most economical operational-grade configurations currently available for sub-annual pasture monitoring at state-wide scale: ~ 40 s per Sentinel-2 scene per site on a 4-vCPU cloud instance, no proprietary inputs (Sentinel-2 L2A via Element 84, SoilGrids 2.0 from ISRIC, and Open-Meteo / ERA5 climate), a one-off UAV calibration flight per new site, and ~ 2 h end-to-end re-training. A natural extension is to also exploit acid-detergent fibre (ADF), which agronomically tracks leaf maturity and varies inversely with leaf nitrogen as cell walls thicken during senescence; we report this exploration in Appendix E, where adding ADF on top of NDVI+TN growth alone slightly degrades LOSO performance (collinearity inflating variance), but the TN + ADF + auxiliary configuration becomes the best cross-site pixel-level result obtained in this study (LOSO $R^2 +0.02$, RMSE 4.29 t ha^{-1}). We do not formally adopt that as Stage 4 in the present paper pending repeat-clip temporal validation.

4.4. Operational Implications and Limitations

The pipeline produces continuous 10 m biomass and chemistry maps every 5–10 days for any Australian rangeland location with Sentinel-2 coverage, at ~ 40 s per scene per site on a 4-vCPU cloud instance and ~ 2 h end-to-end re-training. All inputs are public; the only per-site requirement is a one-off UAV calibration flight plus 30–50 ground samples. Three limitations bind the present results: the calibration footprint (11 sites on five properties of sub-tropical to tropical Queensland C4 grass and grassy woodland) is narrow relative to state-wide rangeland heterogeneity; ground truth is confined to single-time-point destructive samples, so paddock-scale temporal validation of the time-series products is not possible; and the aggregate-level LOSO ceiling (best $R^2 = -0.05$, RMSE $\approx 4.4 \text{ t ha}^{-1}$) appears to be a sample-size and sensor-saturation limit rather than a feature-engineering one—optical reflectance, three independent FC products, two reanalysis climate sources, SoilGrids 2.0 and Sentinel-1 C-band SAR all hit the same RMSE floor, and with $n = 9$ properties, one

bad LOSO fold can swing R^2 by ~ 0.2 . Closing the ceiling will likely require either explicit modelling of species/community membership as a covariate, Bayesian site-specific calibration transfer, L-band SAR (NISAR) or GEDI lidar covariates that escape the optical-saturation regime, or—most tractably—the regional/land-type-specific stratification outlined below.

4.5. Regional and Land-Type-Specific Models: A Tractable Path Beyond the LOSO Ceiling

The results in Tables 3–5 were obtained on 11 field sites across five commercial cattle-grazing properties within three Queensland bioregions (Desert Uplands, Brigalow Belt and Southeast Queensland), all inside the Great Barrier Reef catchment. The vegetation covers C4-dominated open native and naturalised grassland, mixed grass/woody pasture, and grassy *Eucalyptus* savanna—a subtropical to tropical productive rangeland envelope with ~ 550 – 1100 mm yr^{-1} mean annual rainfall spanning Köppen classes Cfa (subtropical Property E), Cfa/Cwa (Brigalow Belt Properties A and B) and Aw (northern Properties C and D). The cross-site LOSO numbers reported here therefore quantify transferability within this bioregion cluster, and the trained models should not be deployed as-is outside it. In particular, the current pipeline is not calibrated for (i) Mediterranean C3 annual grasslands of southern Australia, (ii) temperate C3 improved dairy pastures of Victoria and Tasmania, (iii) arid Mitchell-grass downs of western Queensland or the Barkly Tableland, or (iv) non-Australian rangelands with different dominant species composition or a materially different climate-driver structure. In each of these regimes the same tare-corrected ground-truth protocol (30–50 clip-and-weigh samples plus one photogrammetric UAV flight per site) followed by the re-training of Stage 1 on the local UAV pairs is required, and the environmental covariate block should be re-selected against the local climate driver: the aridity index and VPD_{max} that carry the cross-site signal on the present footprint have region-specific mappings to above-ground biomass and cannot be assumed to translate. What generalises is the three-tier TabPFN cascade architecture, the tare-corrected ground-truth and UAV-DSM calibration protocol, the random forest surrogate deployment pattern, and the open covariate infrastructure (Sentinel-2 L2A, Open-Meteo ERA5, and SoilGrids 2.0); what has to be re-calibrated in a new bioregion is the numerical model itself. The remainder of this subsection develops the most tractable operational path to that re-calibration.

A single state-wide biomass model asks one regressor to span Mitchell-grass downs, sub-tropical C4-improved pastures, grassy–eucalypt savanna and temperate C3 systems with one parameter set. The aggregate-level LOSO ceiling reported here ($R^2 \approx -0.05$ on $n = 9$ properties) and the symmetric failures of every covariate experiment suggest a multi-modal cross-region biomass–reflectance mapping that a single global learner is sample-starved against. The most tractable operational follow-on is to stratify the production model: fit one TabPFN cascade per Köppen–Geiger or NRM bioregion, or per Queensland Land Type (an agronomic classification that already encodes dominant species community and indicative carrying capacity). At deployment, each pixel is routed by its bioregion or land-type polygon, optionally as a soft mixture of experts weighed by climate and soil covariates. Marginal compute is negligible because TabPFN training is amortised at ~ 10 s per fit, and the same UAV calibration protocol used here can be deployed at one or two representative sites per stratum—still fewer total flights than a properly sized global model would need. We do not adopt any stratification as the headline production model in the present paper because the current nine-property footprint cannot populate even one well-resolved stratum; the point is that the LOSO ceiling reported here is diagnostic of calibration-footprint heterogeneity, and the operational answer is more likely to be several smaller specialist models with shared infrastructure than one larger global model with more covariates.

5. Conclusions

We present, to our knowledge, the first cascade to apply a single tabular foundation model (TabPFN) at three nested spatial scales—field, UAV and Sentinel-2—to the same agroecological target. Tare-corrected ground truth (1120 samples, 11 sites), a TabPFN UAV-DSM regressor with random forest surrogate (in-sample $R^2 = 0.94$), and three Sentinel-2 TabPFN models (in-sample $R^2 = 0.97$ across all three stages; pixel-level LOSO-site R^2 between -0.69 and -0.80 in the spectral-only regime) collectively enable continuous 10 m, 5–10 day biomass and chemistry maps for Australian rangelands once combined with open climate and soil covariates, which increase aggregate cross-site LOSO R^2 from -0.15 (the best spectral-only learner) to -0.05 (env-only GBM+TabPFN stack, RMSE 4.43 t ha^{-1}). The leaf-N trajectory from the TabPFN chemistry model developed in our earlier pasture-chemistry work [22] adds complementary information beyond NDVI persistence, improving pixel-level LOSO R^2 from -0.80 to -0.67 and reducing pixel-level LOSO RMSE by 4%.

The main methodological contribution is the deployment of a single tabular foundation model at three nested scales of an open-source cascade; the main empirical contribution is the honest cross-property LOSO benchmark that quantifies the operational range currently delivered and the sample-size and optical-saturation ceiling that bounds further improvement on the present nine-property footprint.

All code, models, and predictions are available upon request from the corresponding author, subject to the commercial-in-confidence and landholder-privacy terms attached to the underlying fieldwork. Future work will broaden the calibration footprint to temperate C3 systems and Mitchell-grass downs, integrate paddock-scale repeat-clip validation, stratify the production model by climatic zone and Queensland Land Type (Section 4.5), and explore Bayesian site-specific calibration transfer.

Author Contributions: Conceptualisation, J.B. and H.R.P.; methodology, J.B.; software, J.B.; validation, J.B. and H.R.P.; formal analysis, J.B.; investigation, J.B.; resources, G.F.; data curation, J.B. and H.R.P.; writing—original draft preparation, J.B.; writing—review and editing, J.B., H.R.P. and G.F.; visualisation, J.B., H.R.P. and G.F.; project administration, G.F.; funding acquisition, G.F. All authors have read and agreed to the published version of the manuscript.

Funding: This research study was funded equally by the Queensland Department of the Environment, Tourism and Innovation division of both the Earth Observation and Social Sciences and the Office of the Great Barrier Reef Divisions and the Queensland Department of Primary Industries, Drought and Climate Adaptability Program division.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: All code, trained models, training CSVs and per-site biomass rasters are available upon request from the corresponding author. Redistribution is subject to the commercial-in-confidence and landholder-privacy terms attached to the underlying private and commercial fieldwork. Sentinel-2 L2A imagery is publicly available via the Element 84 Earth-Search STAC API (<https://earth-search.aws.element84.com/v1>, accessed on 25 June 2026). UAV photogrammetric source images are available upon request from the corresponding author.

Acknowledgments: The anonymous reviewers are acknowledged for their review of and feedback on the manuscript. We also acknowledge the Landsberg and Cuddihy families and staff of the Trafalgar and Baryugal pastoral properties, the Gordon-Moulin family of the Mount Pleasant pastoral property, the Hawkins family of the Bon Accord pastoral property and the Jess family of the Esk pastoral property, who all kindly provided support and access to their properties. Staff at the Eco-sciences Precinct Grazing and Land Systems, Remote Sensing Sciences centre and Animal Sciences including

Christina Jones, Grant Stone, John Carter, Dan Tindall and Giselle Whish are acknowledged for assistance and support and Rebecca Farrell and Thomas Franz for assistance in collecting field data.

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A. Full Field Site List and Sampling Schedule

Table A1 lists the GPS coordinates and visit dates of every field site sampled in this study.

Table A1. Field site locations and sampling schedule.

Property	Site ID	Longitude	Latitude (South)	Visit Dates	Season	Sampling Type ¹
Trafalgar (A)	boundary	145.8957566	20.5097843	29 April 2023	Late wet	Q, M, UAV
				30 September 2023	Dry	Q, M, UAV
				18 February 2024	Wet	Q, M, UAV
				20 September 2024	Dry	Q, M, UAV
				18 August 2025	Dry	UAV
Trafalgar (A)	appletree	145.9227423	20.5389117	2 May 2023	Late wet	Q, M, UAV
				2 October 2023	Dry	Q, M, UAV
Trafalgar (A)	trafAlt	146.0163052	20.4216030	16 February 2024	Wet	Q, M, UAV
				22 September 2023	Dry	Q, M, UAV
				21 August 2025	Dry	UAV
Mt Pleasant (B)	mtp01	147.9482863	20.2945290	7 May 2023	Late wet	Q, M, UAV
				7 October 2023	Dry	Q, M, UAV
				9 February 2024	Wet	Q, M, UAV
				26 September 2024	Dry	Q, M, UAV
				13 August 2025	Dry	UAV
Mt Pleasant (B)	mtp02	147.9000386	20.2912746	9 May 2023	Late wet	Q, M, UAV
				10 October 2023	Dry	Q, M, UAV
				11 February 2024	Wet	Q, M, UAV
				29 September 2024	Dry	Q, M, UAV
Bon Accord (C)	bonna01	147.7126628	23.6817412	22 May 2023	Late wet	Q, M, UAV
				29 October 2023	Dry	Q, M, UAV
				11 April 2024	Wet	Q, M, UAV
				25 October 2024	Dry	Q, M, UAV
				10 September 2025	Dry	UAV
Bon Accord (C)	bonna03	147.7418179	23.6566114	24 May 2023	Late wet	Q, M, UAV
				31 October 2023	Dry	Q, M, UAV
				13 April 2024	Wet	Q, M, UAV
				26 October 2024	Dry	Q, M, UAV
Baryugal (D)	bary01	149.7133254	24.8590296	29 May 2023	Late wet	Q, M, UAV
				5 November 2023	Dry	Q, M, UAV
				17 April 2024	Wet	Q, M, UAV
				28 October 2024	Dry	Q, M, UAV
				15 September 2025	Dry	UAV
Baryugal (D)	bary02	149.6933999	24.8335359	31 May 2023	Late wet	Q, M, UAV
				8 November 2023	Dry	Q, M, UAV
				19 April 2024	Wet	Q, M, UAV
				30 October 2024	Dry	Q, M, UAV
Esk (E)	esk01	152.3962129	27.2509679	15 March 2023	Wet	Q, M, UAV
				30 August 2023	Dry	Q, M, UAV
				30 January 2024	Wet	Q, M, UAV
				27 August 2024	Dry	Q, M, UAV
				6 May 2025	Dry	UAV
Esk (E)	esk02	152.3966005	27.2513793	15 March 2023	Wet	Q, M, UAV
				30 August 2023	Dry	Q, M, UAV

¹ Q = quadrat; M = mower; UAV = imaging UAV.

Appendix B. UAV Platforms and Flight Patterns

UAV flight patterns for the two acquisition platforms

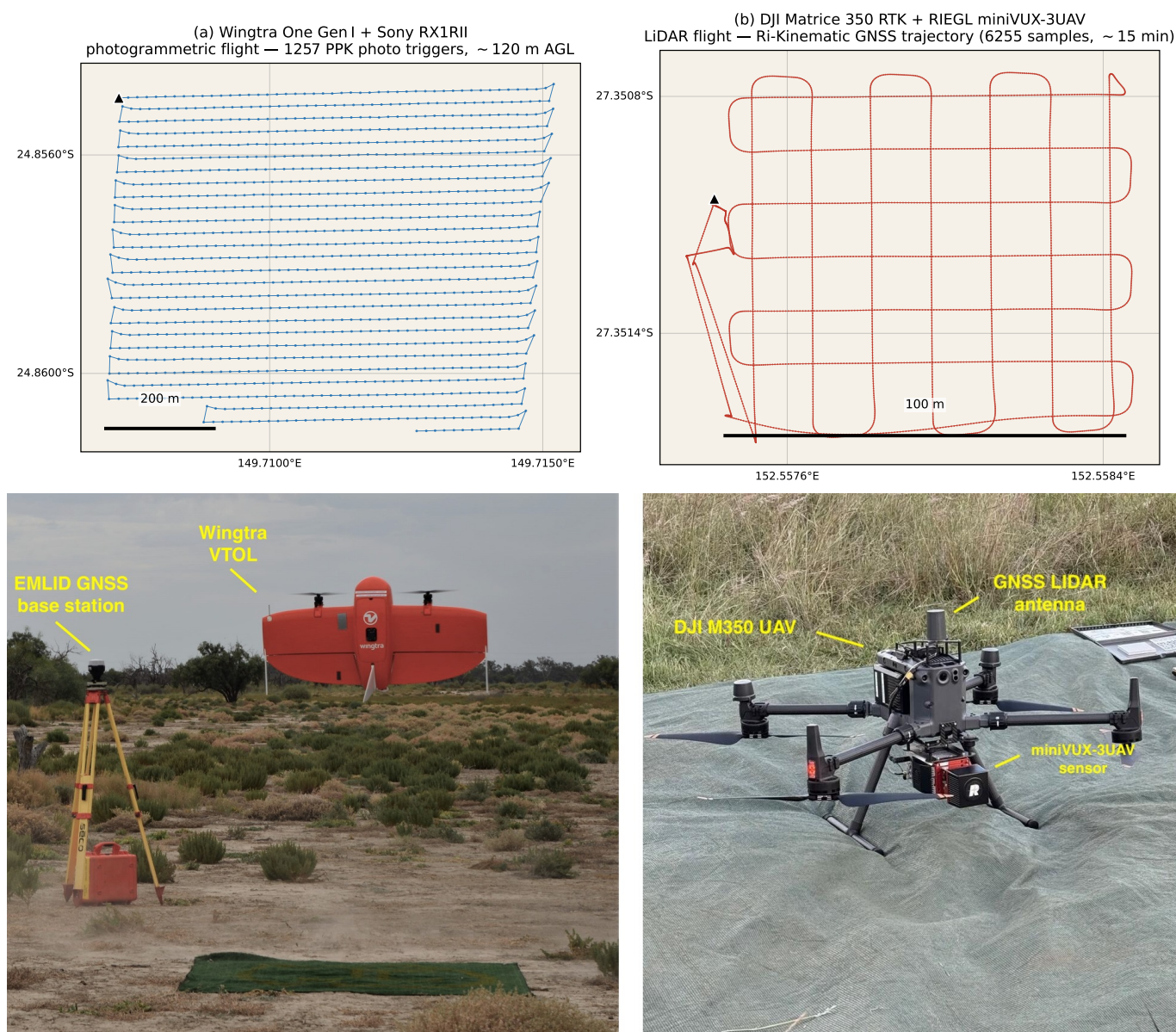


Figure A1. Representative flight pattern and example photos of the UAV platforms described in Section 2.6 (Methods, “UAV acquisition”). (a) Wingtra One Gen I (Zurich, Switzerland) photogrammetric flight reconstructed from the PPK-corrected per-photo trigger positions in the WingtraHub geotag log, showing the long parallel transects flown at ~120 m AGL with 80%/70% overlap. (b) DJI M350 (Shenzhen, China) + miniVUX-3UAV LiDAR flight reconstructed from the Ri-Kinematic GNSS trajectory, showing the cross-hatched survey grid flown at ~70 m AGL with the take-off circle visible at the centre-west. Black triangles mark each take-off location. (Lower row) Wingtra One Gen I fixed-wing VTOL with 42 MP Sony RX1RII RGB camera (left); DJI Matrice 350 RTK multi-rotor carrying the RIEGL miniVUX-3UAV (Horn, Austria) survey-grade LiDAR sensor (right).

Appendix C. Three-Stage Sentinel-2 Biomass Comparison at bary01

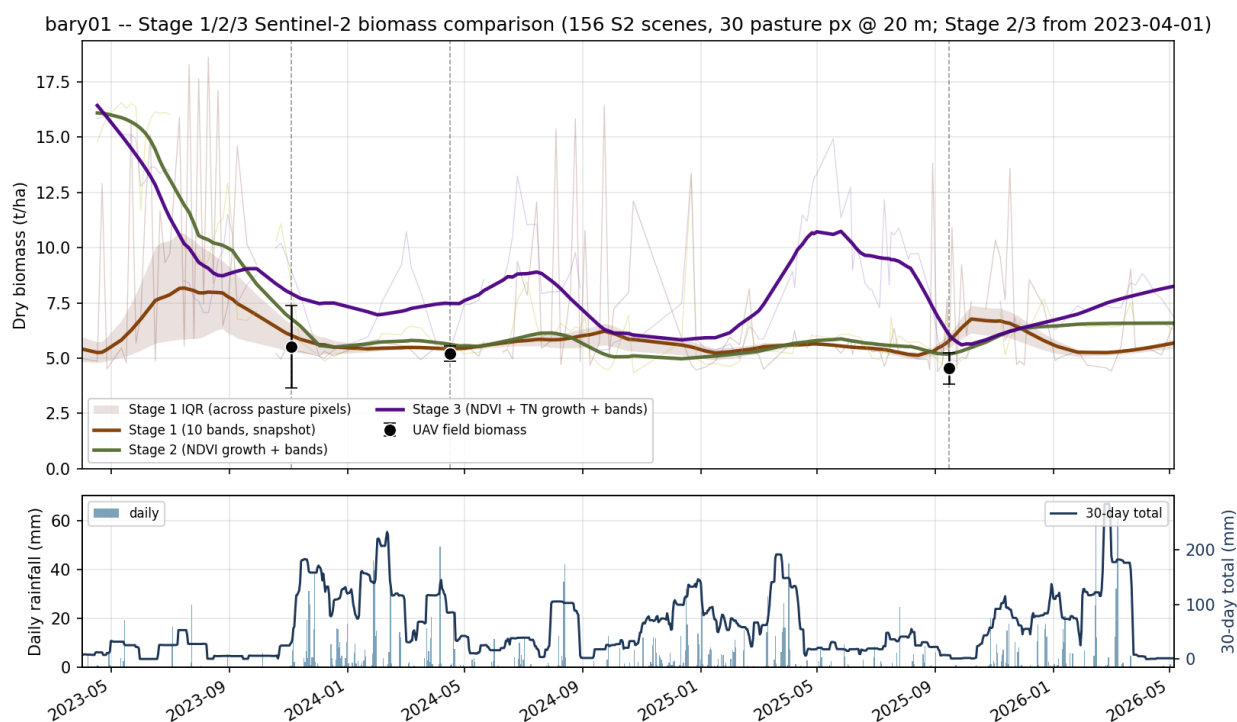


Figure A2. Three-stage Sentinel-2 biomass comparison for Baryugal (bary01), 1 April 2023 to 10 May 2026. Each stage is plotted as a faint per-scene line overlaid by a 45-day robust LOWESS smoother. (**Top**) Stage 1 (snapshot 10-band TabPFN, brown); Stage 2 (NDVI-growth features + 10 bands, olive); Stage 3 (NDVI- and TN-growth features + 10 bands, purple). Black markers $\pm 1\sigma$ are the UAV in situ clip-and-weigh biomass; vertical dashed lines mark those flights. (**Bottom**) Daily rainfall (Open-Meteo ERA5 at the site centroid) with 30-day rolling-sum overlay. The time axis begins on 1 April 2023 because Stages 2 and 3 require a populated 365-day NDVI (and for Stage 3, TN) antecedent window. All three stages are rendered with the same smoother to make the model envelopes directly comparable.

Appendix D. Mt Pleasant (mtp01) Operational Time Series and Stage Comparison

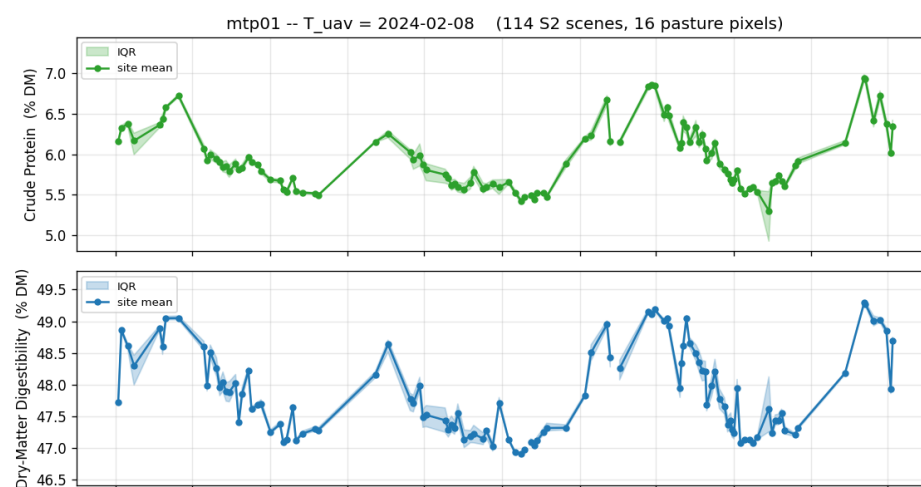


Figure A3. Cont.

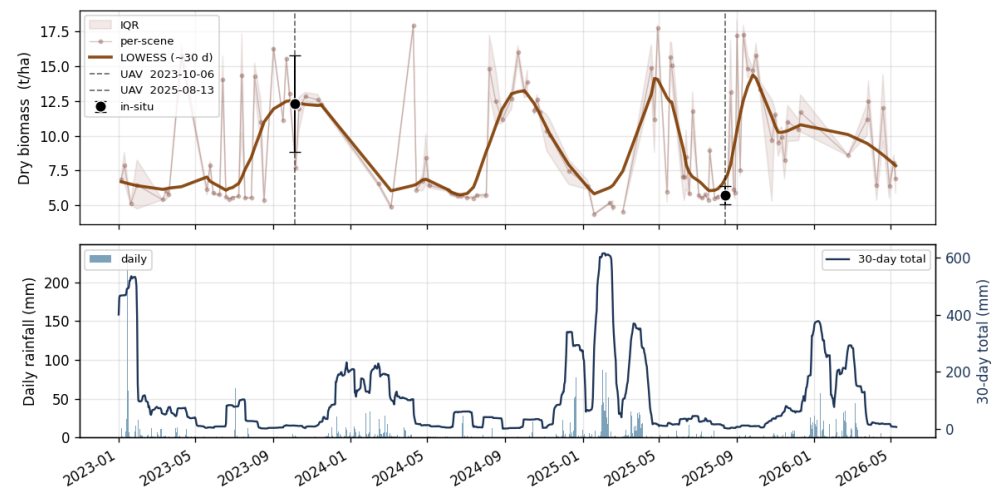


Figure A3. Operational Sentinel-2 time series of pasture crude protein (**top**), dry-matter digestibility (**second**), dry biomass with a ~45-day robust LOWESS smoother (**third**) and daily rainfall with a 30-day rolling-sum overlay from the Open-Meteo ERA5 archive (**bottom**) over the grassy-woodland savanna site Mt Pleasant (mtp01), 1 January 2023 to 10 May 2026. Solid line = pasture-class mean across the 16 pasture pixels within the site footprint; shaded band = inter-quartile range across pixels; vertical dashed line marks the UAV reference flight on 8 February 2024. Y-axis limits on the chemistry panels are anchored to the site-mean trajectory rather than the across-pixel IQR, which would otherwise compress the seasonal envelope. The pasture-classification mask suppresses the persistent eucalypt overstory contribution and isolates the ephemeral grass dynamics. Black markers with $\pm 1\sigma$ whiskers on the biomass panel are the per-pixel mean of the Stage 1 training pairs from the UAV mower-bag flights on 6 October 2023 and 13 August 2025. Retrievals use the Stage 1 TabPFN biomass model and the TabPFN chemistry models of [22] over 114 cloud-free S2 scenes.

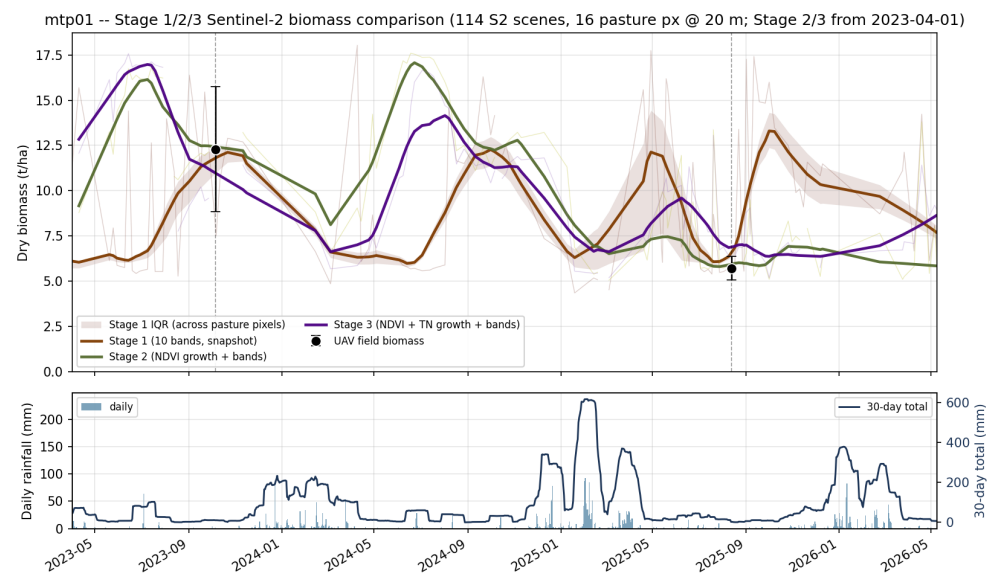


Figure A4. Three-stage Sentinel-2 biomass comparison for Mt Pleasant (mtp01), 1 April 2023 to 10 May 2026, plotted with the same conventions as Figure A2. The Stage 1 envelope sits in the $\sim 9\text{--}14\text{ t ha}^{-1}$ range with tighter inter-stage agreement than at bary01: at this lightly grazed site the 365-day antecedent NDVI window captures three full wet-to-dry growth cycles uninterrupted by management offtake, so the antecedent-feature stages reinforce rather than redistribute the snapshot retrieval.

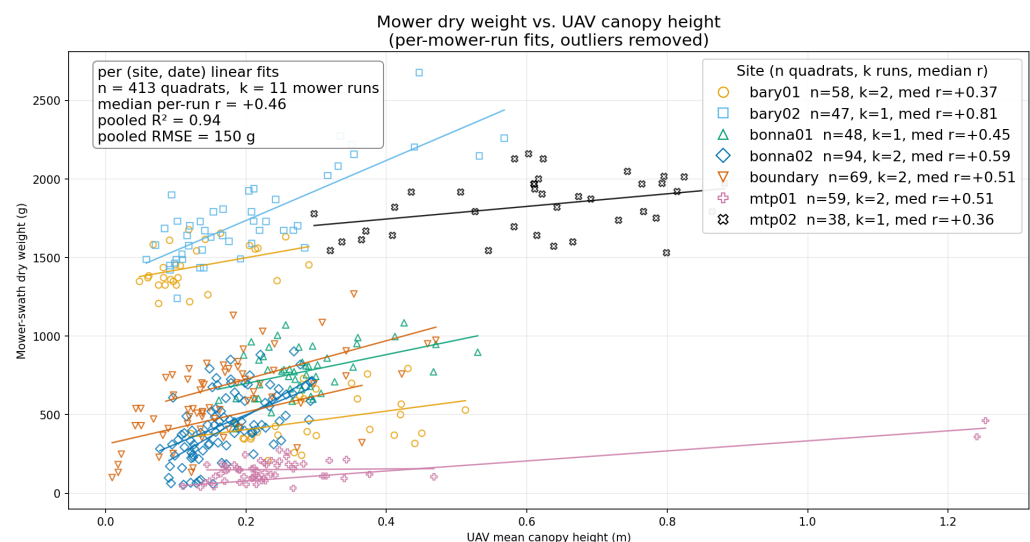
Appendix E. ADF Growth Features and the TN + ADF + Auxiliary Configuration

A natural extension of the Stage 3 model is to include acid-detergent fibre (ADF) growth features alongside the leaf-nitrogen trajectory. We tested this on the 887 pixel pairs for which both per-pixel TN and ADF growth-feature stacks were available. At an instantaneous level the TN–ADF inverse relationship is essentially deterministic (Pearson $r = -0.94$ between tn_at_uav and adf_at_uav), but the temporal-aggregate features decouple progressively (mean-growth $r = -0.74$, integral -0.63 , persistence -0.34 , max-growth -0.29), indicating that TN and ADF growth dynamics carry partly independent information about pasture phenological stage. Their univariate correlations with biomass also flip sign in the agronomically expected direction (e.g., tn_mean_growth $r = +0.50$ vs. adf_mean_growth $r = -0.29$), so the two trajectories should reinforce rather than cancel.

Empirically, adding ADF-growth features on top of NDVI + TN growth alone slightly degrades LOSO performance (LOSO R^2 -0.46 vs. -0.12 ; RMSE 5.25 vs. 4.60 $t\ ha^{-1}$, $n = 824$), consistent with collinearity inflating variance in the 27-feature pixel-level TabPFN. However, once the same climate and soil auxiliary covariates used in the aggregate-level covariate experiment are concatenated, the TN + ADF + auxiliary configuration becomes the best cross-site pixel-level result obtained in this study (LOSO R^2 $+0.02$, RMSE 4.29 $t\ ha^{-1}$; cf. -0.12 / 4.60 for NDVI + TN growth without auxiliaries). This is consistent with the climate/soil context “unlocking” the complementary fibre–nitrogen seasonality signal and motivates a future dedicated NDVI + TN + ADF + auxiliary cascade. We do not adopt that as Stage 4 in the present paper pending repeat-clip temporal validation.

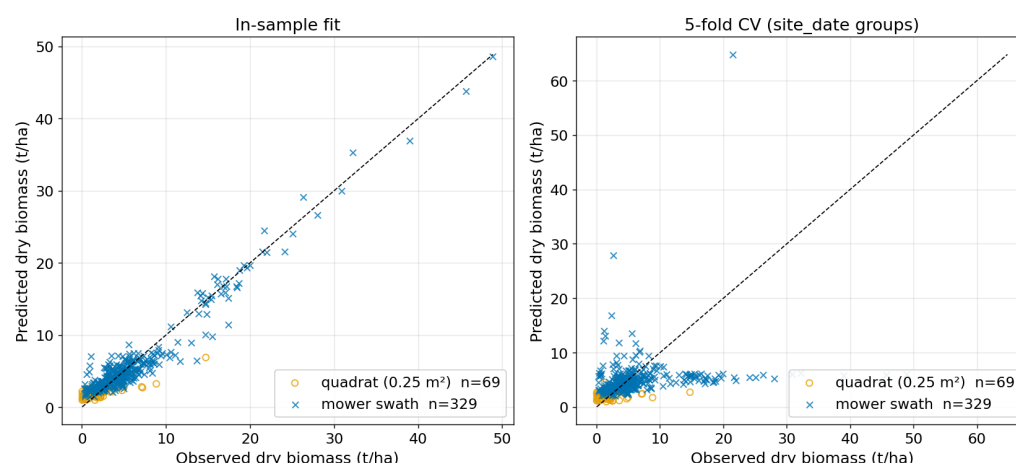
Appendix F. Field- and UAV-Scale Calibration Scatter

Figure A5 reports the field- and UAV-scale observed-versus-predicted scatter that complements the Sentinel-2 panel in Figure 4.



(a) Field scale: mower-swath dry biomass (as-is) vs. best tare-corrected estimate, per mower-swath sample.

Figure A5. Cont.



(b) UAV scale: TabPFN-predicted vs. observed dry-biomass density (t ha^{-1}) from the 5 cm DSM/CHM canopy-height inputs.

Figure A5. Field- and UAV-scale observed-versus-predicted scatter: (a) effect of bag-tare correction on the mower-swath samples; (b) TabPFN UAV-DSM biomass model fit. Together with Figure 4c (Sentinel-2 scale) these show the progressive inflation of scatter with the increase in footprint and decrease in direct contact with standing biomass.

Appendix G. Validation of Open Covariate Sources Against In Situ Data

This appendix reports the full per-property validation tables and matched scatter plots underpinning the summary in Section 3.4. The two open-source datasets validated here—SoilGrids 2.0 (0–5 cm) topsoil predictions and three independent fractional-cover products—are the covariate sources whose accuracy is the most consequential for the LOSO covariate experiment in Section 3.3.

Appendix G.1. SoilGrids 2.0 vs. In Situ Topsoil Chemistry

Table A2 and Figure A6 compare SoilGrids 2.0 (0–5 cm) against the in situ topsoil chemistry collected for this study. Across the eight aggregate-level sites with complete particle-size data, SoilGrids reproduces the relative cross-site ordering of soil organic carbon (SOC; Pearson $r = 0.73$) and total nitrogen (TN; $r = 0.87$) with modest negative biases of -3.7 g kg^{-1} SOC and -0.31 g kg^{-1} TN; the clay and sand fractions are weakly correlated ($r = 0.08, -0.02$) with biases of $+9.9\%$ (clay) and -15.3% (sand), reflecting the well-documented limitation of the SoilGrids 250 m topsoil prediction in heterogeneous dryland pastures dominated by sub-property textural variability [38,39]. Bulk density agrees almost exactly in the mean (1.36 vs 1.34 g cm^{-3} , bias $+0.02 \text{ g cm}^{-3}$, RMSE 0.09 g cm^{-3} across $n = 96$ cores at eight sites), but the cross-site ordering is only weakly preserved ($r = 0.31$); SoilGrids compresses the in situ BD range from 1.17 – 1.47 to 1.31 – 1.42 g cm^{-3} , consistent with the smoothing inherent in a 250 m global topsoil predictor.

Table A2. SoilGrids 2.0 (0–5 cm) vs in situ topsoil chemistry at the LOSO sites. n is the number of sites with matched data; bias is reported as SoilGrids minus in situ. ^a Bulk density is summarised as the per-site mean over 7–18 oven-dry intact 5 cm cores ($n = 96$ cores total).

Property	n	Mean In Situ	Mean SoilGrids	Bias	RMSE	Pearson r
Clay (%)	8	21.5	31.4	+9.9	15.6	+0.08
Sand (%)	8	67.2	51.9	−15.3	19.8	−0.02
SOC (g kg^{-1})	8	19.9	16.3	−3.7	5.4	+0.73
Total N (g kg^{-1})	8	1.45	1.14	−0.31	0.39	+0.87
Bulk density (g cm^{-3}) ^a	8	1.34	1.36	+0.02	0.09	+0.31

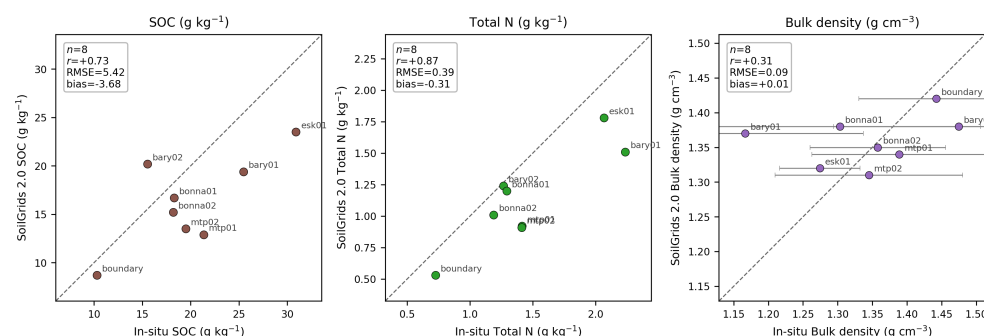


Figure A6. SoilGrids 2.0 (0–5 cm) versus in situ topsoil chemistry at the eight LOSO sites with matched data, for the three properties most directly relevant to above-ground biomass: SOC, total nitrogen and bulk density. Each point is one site; the dashed line is 1:1; horizontal error bars on the BD panel are the in situ standard deviation across 7–18 oven-dry intact 5 cm cores per site ($n = 96$ cores total).

Appendix G.2. Fractional-Cover Products vs. UAV Hyperspectral Classification

Table A3 and Figure A7 report the equivalent validation for the three fractional-cover products against the UAV hyperspectral classification reference. The TERN Sentinel-2 product tracks PV and NPV ordering of the two public products the most strongly (Pearson $r = 0.59$ and 0.83 respectively), at the cost of consistent biases (PV: +8.2%, NPV: -15.9%, BS: +6.2%) that translate into RMSE of 22–24% on the dominant fractions. The DEA Landsat product behaves similarly (PV: +9.3%, NPV: -14.1%, BS: +3.5%), with weaker correlation on NPV ($r = 0.44$). The locally trained TabPFN FC, evaluated at all 16 matched site–date pairs, agrees much more closely with the UAV reference (PV $r = 0.94$, NPV $r = 0.93$, BS $r = 0.87$; per-fraction RMSE 1.7–11.5%). Figure A8 regresses each FC fraction directly against UAV biomass at the same 20 site–date aggregates: across all three products and both biomass-relevant fractions, the FC cover is statistically independent of biomass in the observed productivity range ($|r| \leq 0.11$, OLS slopes between -0.20 and +0.44% per t ha⁻¹), demonstrating that the saturation occurs upstream of the unmixing step in the optical reflectance from which every FC product is derived.

Table A3. Fractional-cover products vs UAV hyperspectral classification reference at matched (site, UAV-date) aggregates. n is the number of matched site–date pairs; bias is reported as product minus UAV reference (%).

Product	Fraction	n	Mean UAV	Bias	RMSE	Pearson r
DEA Landsat ga_ls_fc_3	PV	14	24.2	+9.3	26.1	+0.53
	NPV	14	73.5	-14.1	29.8	+0.44
	BS	14	2.3	+3.5	5.7	+0.22
TERN/AusCover Sentinel-2 FC	PV	16	22.1	+8.2	24.1	+0.59
	NPV	16	75.6	-15.9	22.8	+0.83
	BS	16	2.3	+6.2	11.9	+0.33
Local TabPFN FC	PV	16	22.1	+6.0	11.5	+0.94
	NPV	16	75.6	-6.7	12.3	+0.93
	BS	16	2.3	+0.6	1.7	+0.87

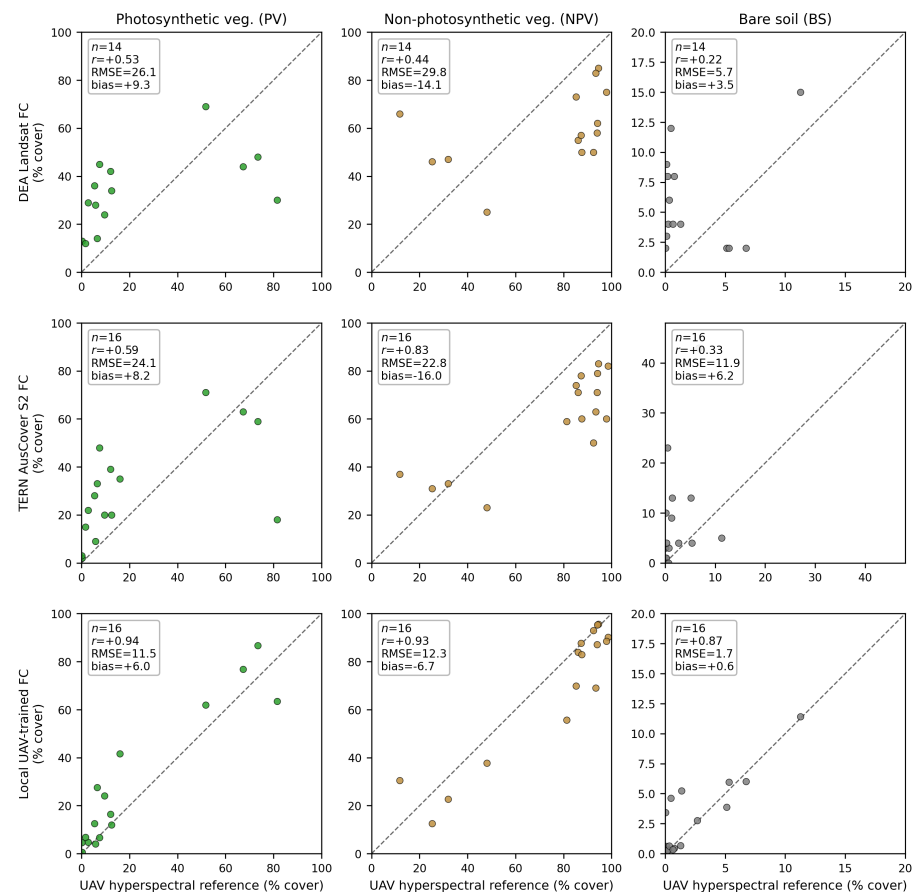


Figure A7. Fractional-cover products versus the UAV hyperspectral classification reference at matched (site, UAV-date) aggregates. Rows: DEA Landsat $ga_ls_fc_3$ ($n = 14$), TERN AusCover Sentinel-2 FC ($n = 16$), and a locally trained TabPFN FC ($n = 16$). Columns: PV (green), NPV (tan) and BS (grey). Dashed line is 1:1; in-panel statistics are sample size, Pearson r , RMSE (%) and bias (product minus UAV reference, %). Caveat: Bare-soil pixels are scarce in the UAV reference set, so the BS-column statistics are estimated from a narrow dynamic range and should be read as indicative.

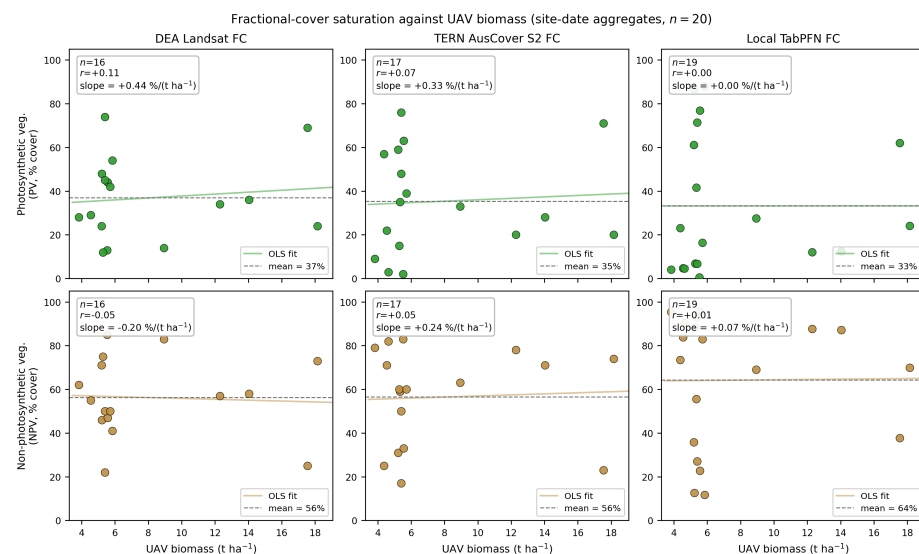


Figure A8. Fractional-cover saturation against UAV biomass at the site-date aggregates. Rows: PV (top) and NPV (bottom). Columns: DEA Landsat FC, TERN AusCover Sentinel-2 FC and the locally trained TabPFN FC. Each point is one (site, UAV-date) aggregate; the coloured solid line is the OLS fit, and the dashed grey line is the mean FC over the observed biomass range. The slopes are statistically indistinguishable from zero ($|r| \leq 0.11$ across all six panels).

References

- McKeon, G.M.; Day, K.A.; Howden, S.M.; Mott, J.J.; Orr, D.M.; Scattini, W.J.; Weston, E.J. Northern Australian Savannas: Management for Pastoral Production. *J. Biogeogr.* **1990**, *17*, 355–372. [\[CrossRef\]](#) [\[PubMed\]](#)
- Barnetson, J.; Phinn, S.; Scarth, P. Estimating Plant Pasture Biomass and Quality from UAV Imaging across Queensland's Rangelands. *AgriEngineering* **2020**, *2*, 523–543. [\[CrossRef\]](#)
- Bartley, R.; Corfield, J.P.; Hawdon, A.A.; Kinsey-Henderson, A.E.; Abbott, B.N.; Wilkinson, S.N.; Keen, R.J. Can changes to pasture management reduce runoff and sediment loss to the Great Barrier Reef? The results of a 10-year study in the Burdekin catchment, Australia. *Rangel. J.* **2014**, *36*, 67–84. [\[CrossRef\]](#)
- Godde, C.M.; Boone, R.B.; Ash, A.J.; Waha, K.; Sloat, L.L.; Thornton, P.K.; Herrero, M. Global rangeland production systems and livelihoods at threat under climate change and variability. *Environ. Res. Lett.* **2020**, *15*, 044021. [\[CrossRef\]](#)
- Monsalve, M.O.; Monroy, T.R.; Galeano-Vasco, L.F.; Medina-Sierra, M.; Ceron-Munoz, M.F. Determination of Grass Quality Using Spectroscopy: Advances and Perspectives. In *Grasslands*; Iqbal, M.A., Ed.; IntechOpen: Rijeka, Croatia, 2023; Chapter 6. [\[CrossRef\]](#)
- Chen, Y.; Guerschman, J.; Shendryk, Y.; Henry, D.; Harrison, M.T. Estimating Pasture Biomass Using Sentinel-2 Imagery and Machine Learning. *Remote Sens.* **2021**, *13*, 603. [\[CrossRef\]](#)
- Morse-McNabb, E.; Hasan, N.; Karunaratne, S. A multi-variable Sentinel-2 random forest model for perennial ryegrass biomass on Australian dairy farms. *Remote Sens.* **2023**, *15*, 2915. [\[CrossRef\]](#)
- Munyati, C. Detecting the distribution of grass aboveground biomass on a rangeland using Sentinel-2 MSI vegetation indices. *Adv. Space Res.* **2022**, *69*, 1130–1145. [\[CrossRef\]](#)
- Pang, H.; Zhang, A.; Kang, X.; He, N.; Dong, G. Estimation of the grassland aboveground biomass of the Inner Mongolia Plateau using random forest. *Remote Sens.* **2020**, *12*, 4155. [\[CrossRef\]](#)
- Wijesingha, J.; Astor, T.; Schulze-Brüninghoff, D.; Wengert, M.; Wachendorf, M. Predicting Forage Quality of Grasslands Using UAV-Borne Imaging Spectroscopy. *Remote Sens.* **2020**, *12*, 126. [\[CrossRef\]](#)
- Lugassi, R.; Zaady, E.; Goldshleger, N.; Shoshany, M.; Chudnovsky, A. Spatial and Temporal Monitoring of Pasture Ecological Quality: Sentinel-2-Based Estimation of Crude Protein and Neutral Detergent Fiber Contents. *Remote Sens.* **2019**, *11*, 799. [\[CrossRef\]](#)
- Riquelme, L.; Duncan, D.H.; Lentini, P.E. Predicting understorey biomass in semi-arid woodlands from satellite imagery. *Remote Sens.* **2022**, *14*, 2358. [\[CrossRef\]](#)
- Viljanen, N.; Honkavaara, E.; Näsi, R.; Hakala, T.; Niemeläinen, O.; Kaivosoja, J. A novel machine learning method for estimating biomass of grass swards using a photogrammetric canopy height model, images and vegetation indices captured by a drone. *Agriculture* **2018**, *8*, 70. [\[CrossRef\]](#)
- Batistoti, J.; Marcato Junior, J.; Ítavo, L.; Matsubara, E.; Gomes, E.; Oliveira, B.; Souza, M.; Siqueira, H.; Salgado Filho, G.; Akiyama, T.; et al. Estimating Pasture Biomass and Canopy Height in Brazilian Savanna Using UAV Photogrammetry. *Remote Sens.* **2019**, *11*, 2447. [\[CrossRef\]](#)
- Cooper, S.D.; Roy, D.P.; Schaaf, C.B.; Paynter, I. Examination of the potential of terrestrial laser scanning and structure-from-motion photogrammetry for rapid nondestructive field measurement of grass biomass. *Remote Sens.* **2017**, *9*, 531. [\[CrossRef\]](#)
- Yang, Y.; Zhang, M.; Li, J.; Wang, X.; Yan, Y.; Xin, X.; Xu, D. The performance of a canopy relative height model (CRHM) in natural grassland aboveground biomass estimation using unmanned aerial vehicle data. *Comput. Electron. Agric.* **2025**, *231*, 109910. [\[CrossRef\]](#)
- Aebischer, P.; Sutter, M.; Birkinshaw, A.; Nussbaum, M.; Reidy, B. Herbage biomass prediction in semi-natural grasslands from UAV-derived digital terrain models and machine learning. *Grass Forage Sci.* **2024**. [\[CrossRef\]](#)
- Hütt, C.; Isselstein, J.; Kayad, A. UAV LiDAR-based grassland biomass mapping for precision livestock farming. *J. Appl. Remote Sens.* **2024**, *18*, 017502. [\[CrossRef\]](#)
- Puliti, S.; Saarela, S.; Gobakken, T.; Ståhl, G.; Næsset, E. Combining UAV and Sentinel-2 auxiliary data for forest growing stock volume estimation through hierarchical model-based inference. *Remote Sens. Environ.* **2018**, *204*, 485–497. [\[CrossRef\]](#)
- Tan, C.; Kou, W.; Yang, W.; Liu, W.; Lu, H.; Yang, H.; Yu, P.; Lu, W. Coupling UAV multispectral data with Sentinel-2 imagery for rubber plantation aboveground biomass. *Remote Sens.* **2025**, *17*, 2955. [\[CrossRef\]](#)
- Gargiulo, J.I.; Lyons, N.A.; Garcia, S.C. Comparison of ground-based, UAV and Sentinel-2 pasture biomass monitoring on Australian dairy farms. *Remote Sens.* **2023**, *15*, 2752.
- Barnetson, J.; Pandeya, H.R.; Fraser, G. Hierarchical spectral modelling of pasture nutrition: From laboratory to Sentinel-2 via UAV hyperspectral. *AgriEngineering* **2026**, *8*, 143. [\[CrossRef\]](#)
- Hollmann, N.; Müller, S.; Purucker, L.; Krishnakumar, A.; Körfer, M.; Hoo, S.B.; Schirrmeister, R.T.; Hutter, F. Accurate predictions on small data with a tabular foundation model. *Nature* **2025**, *637*, 319–326. [\[CrossRef\]](#) [\[PubMed\]](#)

24. Halder, K.; Ewert, F.; Ghosh, A.; Muduchuru, K.; Sweet, L.; Elshaw, R.; Timko, J.; Zheng, W.; Alsafadi, K.; Zhao, G.; et al. High-resolution maize yield mapping across Africa using earth observation and machine learning, deep learning, and foundation model. *Sci. Remote Sens.* **2026**, *13*, 100344. [[CrossRef](#)]
25. Barkov, V.; Schmidinger, J.; Gebbers, R.; Atzmueller, M. Modern neural networks for small tabular datasets in field-scale digital soil mapping. *Eur. J. Soil Sci.* **2026**. [[CrossRef](#)]
26. Sabo, F.; Meroni, M.; Piles, M.; Claverie, M.; Ferreira, F.; Rembold, F.; Lemoine, G. From rows to yields: How foundation models for tabular data simplify crop yield prediction. *Sci. Rep.* **2026**, *16*, 19001. [[CrossRef](#)] [[PubMed](#)]
27. Barnettson, J.; Phinn, S.; Scarth, P. Resilient Grazing in the Pastures of Queensland: An Integrated Remotely Piloted Aircraft System and Satellite-Based Deep-Learning Method for Estimating Pasture Biomass. *AgriEngineering* **2021**, *3*, 681–698. [[CrossRef](#)]
28. Australian Government and Queensland Government. *Reef Water Quality Report Card 2021 and 2022*; Australian Government and Queensland Government: Brisbane, QLD, Australia, 2023. Available online: <https://reportcard.reefplan.qld.gov.au/> (accessed on 10 July 2026).
29. Met Office. *Cartopy: A Cartographic Python Library with a Matplotlib Interface*; Met Office: Exeter, UK, 2010. Available online: <https://scitools.org.uk/cartopy> (accessed on 10 July 2026).
30. Mutanga, O.; Skidmore, A.K. Narrow band vegetation indices overcome the saturation problem in biomass estimation. *Int. J. Remote Sens.* **2004**, *25*, 3999–4014. [[CrossRef](#)]
31. Huete, A.; Didan, K.; Miura, T.; Rodriguez, E.P.; Gao, X.; Ferreira, L.G. Overview of the radiometric and biophysical performance of the MODIS vegetation indices. *Remote Sens. Environ.* **2002**, *83*, 195–213. [[CrossRef](#)]
32. Macelloni, G.; Paloscia, S.; Pampaloni, P.; Marliani, F.; Gai, M. The relationship between the backscattering coefficient and the biomass of narrow and broad leaf crops. *IEEE Trans. Geosci. Remote Sens.* **2001**, *39*, 873–884. [[CrossRef](#)]
33. Vreugdenhil, M.; Wagner, W.; Bauer-Marschallinger, B.; Pfeil, I.; Teubner, I.; Rüdiger, C.; Strauss, P. Sensitivity of Sentinel-1 backscatter to vegetation dynamics: An Austrian case study. *Remote Sens.* **2018**, *10*, 1396. [[CrossRef](#)]
34. Lemaire, G.; Jeuffroy, M.H.; Gastal, F. Diagnosis tool for plant and crop N status in vegetative stage: Theory and practices for crop N management. *Eur. J. Agron.* **2008**, *28*, 614–624. [[CrossRef](#)]
35. McIvor, J.G. Pasture management in semi-arid tropical woodlands: Regional patterns of pasture growth and curing. *Aust. J. Exp. Agric.* **2003**, *43*, 705–716. [[CrossRef](#)]
36. 't Mannetje, L. Measuring biomass of grassland vegetation. In *Field and Laboratory Methods for Grassland and Animal Production Research*; 't Mannetje, L., Jones, R.M., Eds.; CABI Publishing: Wallingford, UK, 2000; pp. 151–177. [[CrossRef](#)]
37. Tothill, J.C.; Hargreaves, J.N.G.; Jones, R.M.; McDonald, C.K. *BOTANAL—A Comprehensive Sampling and Computing Procedure for Estimating Pasture Yield and Composition. I. Field Sampling*; Tropical Agronomy Technical Memorandum 78; CSIRO Australia, Division of Tropical Crops and Pastures: Brisbane, Australia, 1992.
38. Poggio, L.; de Sousa, L.M.; Batjes, N.H.; Heuvelink, G.B.M.; Kempen, B.; Ribeiro, E.; Rossiter, D. SoilGrids 2.0: Producing soil information for the globe with quantified spatial uncertainty. *SOIL* **2021**, *7*, 217–240. [[CrossRef](#)]
39. Dharumarajan, S.; Hegde, R.; Lalitha, M.; Kalaiselvi, B.; Singh, S.K. Evaluation of digital soil mapping approach for predicting soil fertility parameters—A case study from Karnataka Plateau, India. *Arab. J. Geosci.* **2019**, *12*, 535. [[CrossRef](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.